50th IFF Spring School 2019

Scattering!
Soft, Functional and Quantum Materials

Lecture Notes

JÜLICH
Forschungszentrum

Lecture Notes of the
50th IFF Spring School 2019

Manuel Angst, Thomas Brückel,
Stephan Förster, Karen Friese,
Reiner Zorn (Eds.)

Scattering!
Soft, Functional and Quantum Materials

This Spring School was organized by
the Institutes JCNS, PGI, ICS and IAS
of the Forschungszentrum Jülich
on 11 until 22 March 2019.

In collaboration with Universities and
research institutions.

Contents

E <u>Topical Applications</u>

F <u>Complementary Techniques</u>

Preface

Imagine what we would know – or better: would not know – about the microscopic structure and dynamics of condensed matter if geniuses like Ernest Rutherford (Nobel Prize 1908), Max von Laue (Nobel Prize 1914), son and father William Lawrence and William Henry Bragg (Nobel Prize 1915), Clifford G. Shull (Nobel Prize 1994) and Bertram N. Brookhouse (Nobel Prize 1994) and many others would not have invented scattering as a most powerful tool for condensed matter science; scattering with x-rays, neutrons and electrons, which "tell us where atoms are and how they move" (C. G. Shull). Would we be able to build smartphones, produce high-performance plastic materials, modern energy harvesting and energy storage devices or cure diseases with carefully designed pharmaceuticals without the understanding of the microscopic world gained through scattering methods? Much of the amenities we have become accustomed to are based on research with scattering methods. Information on the atomic length scale is provided mainly through x-ray, neutron and electron scattering- and microscopy. To achieve a deeper look into this fascinating microscopic world, large scale facilities have been constructed based on synchrotron x-ray- and neutron radiation sources. With the x-ray free electron laser in Hamburg, Germany, and the European Spallation Source in Lund, Sweden, Europe hosts and will host worldwide leading facilities in this important research field. Groundbreaking research is being performed at such facilities in a very broad range of research areas in physics, chemistry, life science, geoscience, material science and engineering. The range of materials, structures, phenomena, and processes, which can be studied, is unlimited. Experimental methods have been developed which span an incredible range of length- and time-scales from picometer to meter and from femtoseconds to hours. Doing experiments at these large-scale facilities is an especially exciting aspect of research for young scientists. Not only do they obtain unique microscopic information on structure, excitations and dynamics of condensed matter, but from the start they are familiarized with cutting edge technology and with work in an international collaboration.

The goal of this Spring School is to enable students and young researchers to apply the entire tool box of advanced scattering methods to their topical research in order to provide them with microscopic information, which they cannot obtain otherwise. To this end, we teach the basics of scattering, introduce the various radiation sources with the particular properties of the corresponding radiation, familiarize the students with the appropriate methods and instruments and show how modern scattering methods can be applied to provide essential and unique contributions to the solution of grand challenges facing our modern societies such as energy supply, health, earth and environment, cultural heritage, transport and information technology.

The IFF Spring School 2019 is structured in six sections:

__*Basics-Scattering*__: This section provides a solid introduction into the basics of scattering methods, starting from the interaction processes of X-rays and neutrons with matter via scattering theory and the application of correlation functions.

__*Basics-Materials*__: An overview is given of our current understanding of the structure of crystals and complex fluids, the dynamics of disordered systems and large molecules, and the collective excitations in crystalline matter.

__*Sources and Instrumentation*__: Building on this firm foundation, we then proceed to introduce present-day synchrotron radiation and reactor-based neutron sources and corresponding instrumentation as well as modern and future pulsed sources, such as free electron lasers and neutron spallation sources. The two European flagship projects, XFEL and ESS, are highlighted as well as the most advanced development of High Brilliant Compact Neutron Sources, which will provide a new quality of research with neutrons particularly for nanostructured and biological materials.

__*Techniques*__: Out of the enormous range of modern scattering techniques, some of the most advanced are presented in detail. In this block, we focus on techniques applicable to topical investigations, such as nanostructures, functional and biological materials, quantum materials and correlated electron systems and latest development for in situ operando studies of energy materials. To prepare students for research at these future sources, innovative scattering techniques at pulsed spallation sources or free electron lasers are introduced.

__*Topical Applications*__: Examples of topical applications in various fields of science allow the students to connect what they have learned to their own research. We point out the application of scattering methods for the grand challenges in information technology, energy materials and life sciences.

__*Complementary Techniques*__: The course is rounded off by a comparison of scattering methods with complementary techniques, such as real space imaging or electronic structure spectroscopies.

The preparation of the School benefited from the long track record of the Forschungszentrum Jülich of advanced research in the field of condensed matter science employing scattering methods and the operation of neutron scattering facilities and beamlines at synchrotron radiation sources. Neutron scattering was at the basis of the foundation of the former Institut für Festkörperforschung IFF (Institute of Solid State Research) in 1969, which is the eponym for this Spring School, which has been held every year since then, now in 2019 for the 50th time. After a restructuring in 2011, research with neutrons is carried on by the Jülich Centre for Neutron Science (JCNS), which operates instruments at some of the most forefront neutron large scale facilities worldwide and has research foci in "Soft Matter" and in "Quantum Materials and Collective Phenomena". Research with synchrotron radiation is performed in JCNS, the Peter Grünberg Institute (PGI) and the Institute for Complex Systems (ICS), which emerged from the IFF in the mentioned restructuring process. PGI owns and operates several synchrotron beamlines. Finally, the Ernst Ruska-Center for Microscopy and Spectroscopy with Electrons (ER-C) allows complementary studies using electrons as a probe. The knowledge on the application of neutrons, synchrotron radiation and electrons

assembled in these institutes sets the stage for this Spring School and allows us to present the students with the most forefront developments in the field.

Thanks to Drs. Manuel Angst, Philipp Ebert, Olaf Holderer, Emmanuel Kentzinger, Ulrich Rücker and Astrid Schneidewind for organizing the rehearsals and coordinating the content of the various lectures.

We are especially glad that several distinguished colleagues from external universities and research laboratories have agreed to contribute to the program of the School:

- ***Chapman***, *Henry, Prof.,*   *DESY, Hamburg, Germany*
- ***Disch***, *Sabrina, Dr.,*   *Universität zu Köln, Germany*
- ***Feidenhans'l***, *Robert, Prof.,* *X-FEL, Hamburg, Germany*
- ***Müller***, *Martin, Prof.,*   *HZG, Geesthacht, Germany*
- ***Schreyer***, *Andreas, Prof.,*   *ESS, Sweden*
- ***Strobl,*** *Markus, Prof.,*   *PSI, Switzerland*
- ***Techert***, *Simone, Prof.,*   *DESY, Hamburg, Germany*
- ***Xiao***, *Yinguo, Prof.,*   *Peking University, Shenzhen, China*

The public evening lecture on occasion of the 50th anniversary of the IFF Spring School is given by:

- ***Tolan***, *Metin, Prof.,*   *TU Dortmund, Germany*

After our course has presented everything you always wanted to know about scattering, Metin Tolan shows us with great humor the limits of the laws of physics in his talk on "Shaken, not stirred! - James Bond in the Focus of Physics"

Without the participation of all these colleagues, the program would not be as interesting, versatile, and attractive. 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. We are very grateful to the board of directors of the 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. Finally, our special thanks go to Ms. Barbara Daegener for the general management, the organization, public outreach and the compilation of the lecture notes. Without her enormous commitment, this School would not have been possible.

Manuel Angst, Thomas Brückel, Stephan Förster, Karen Friese and Reiner Zorn
March 2019

A 1   Scattering – Introduction and Overview

Th. Brückel & S. Förster
Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

---

1   What is Scattering?

Scattering is one of the most important processes in nature. It gives rise to manifold phenomena and can be used as a powerful tool for science and industry. But what actually is scattering? A general definition of scattering could read as follows:

Scattering is the physical process in which radiation is being deflected by an object from straight propagation.

Scattering can occur whenever there are localized non-uniformities in the medium through which the radiation propagates. In the quantum regime, according to the wave–particle duality, radiation can be described by moving particles or by a wave. Therefore, the above definition includes particle-particle collisions, but also reflections or refractions at a surface or interface. There can be single or multiple scattering events in the media.

Scattering is the basic mechanism which gives rise to many important phenomena such as:
- electrical resistivity in solids, which is due to scattering of electrons from defects or thermal vibrations of the atoms
- light emission in fluorescence lamps through inelastic scattering of electrons by gas atoms
- the blue color of the sky through so-called elastic Rayleigh scattering of the sun light, which has a strong wavelength dependence of λ^{-4}
- cosmic rays scattered in the atmosphere (aurora borealis!)
- scattering of the charged particles inside particle accelerators, an important factor determining the beam emittance, which is the average spread of particle coordinates in position-momentum phase space
- emission of bremsstrahlung x-rays from anodes bombarded by electrons, which lose energy through inelastic collisions
- neutrons scattered and thermalized in moderators of nuclear reactors giving rise to a certain energy spectrum

Besides being at the origin of many phenomena, scattering is one of the most important tools in many fields of science as it can provide information about the internal structure of the medium from which the radiation is scattered. Single scattering events are particularly useful due to their simple theoretical description. Examples include:
- in the geosciences, seismological studies of the propagation and deflection of elastic waves through the earth are the primary tool for underground exploration (e.g. to detect petroleum bearing formations) and the mapping of the earth's interior
- the scattering of radar waves is being used e.g. for air traffic control or the detection of weather formations
- nuclear- and particle physics uses the scattering of high energy elementary particles (electrons, protons etc.) from accelerators to investigate the structure of the nuclei or nucleons etc.
- medical ultrasound investigations are an important tool to obtain information on internal body structures or blood flow through scattering of high frequency sound waves
- and last but not least: nearly all information which we humans as individuals collect on a day-to-day basis about the world in which we live, comes from light scattering and imaging through our eyes!

In this Spring School, we concentrate on the application of scattering methods for condensed matter or biological materials. Here the aim is to understand the microscopic atomic structure of matter, i.e. to find out where the atoms are located inside our samples and also how they move. We want to relate this information to the properties and functionalities of the materials under investigation. To this end, scientists mimic the mentioned process of obtaining information through visual perception in well controlled scattering experiments: they build a source of radiation, direct a beam towards a sample, detect the radiation scattered from a sample, i.e. convert the signal into an electronic signal, which they can then treat with computers. In most cases one wants an undisturbed image of the object under investigation and therefore chooses the radiation, so that it does not influence or modify the sample. Scattering is therefore a non-destructive and very gentle method, if the appropriate type of radiation is chosen for the experiment.

If the energy or wavelength of the scattered particles or waves, respectively, is the same as before the scattering process, one speaks of *elastic scattering or diffraction*. Scattering without analysis of the energy of the scattered beam gives information about the position of the atoms inside the sample. If the energy of the radiation is changing during the scattering process, one speaks of *inelastic scattering*, which can provide additional information about the movement of atoms in the sample. The analysis of the energy of the scattered radiation with respect to the energy of the incident radiation is called *spectroscopy*.

Which type of radiation should we chose for such investigations? It is quite intuitive to understand that if we want to measure the distance between the atoms, we need a "ruler" of comparable lengths. The distance between atoms is in the order of 0.1 nm = 10^{-10} m = 0.0000000001 m. Since this is such an important length scale in condensed matter science, it has been given its own unit: 0.1 nm = 1 Ångstrøm = 1 Å. If we compare the wavelength of light with this characteristic length scale, it is 4000 to 7000 times longer and therefore light is not appropriate to measure distances between atoms. In the electromagnetic spectrum, x-rays have a well-adapted wavelength of about 1 Å for studies on such a microscopic scale. They also have a large penetration power as everybody knows from the medical x-ray images. The alternative to x-rays are so-called thermal neutron beams, i.e. beams of neutrons being emitted from a moderator at ambient temperature. Such neutrons also have wavelengths in the Å range, but in addition energies which are comparable to the energies of elementary excitations in solids. As neutral particles neutrons penetrate deep into matter. Finally, electrons can be used but as charged particles they suffer from reduced penetration and multiple scattering events. We will mainly concentrate on x-ray and neutron scattering throughout this course.

2   A brief history of x-ray and neutron scattering

> *"If I have seen further it is by standing on the shoulders of giants.*
> *(Sir Isaac Newton, 1643 - 1727)".*

A discovery made more than 100 years ago revolutionized mankinds' understanding of condensed matter: the observation of interference patterns obtained with x-rays scattered by a single crystal [1]. In 1914 Max von Laue received the Nobel prize in physics for the interpretation of these observations. One cannot overestimate the impact of this discovery: it was the first proof that atoms as the elementary building blocks of condensed matter are arranged in a periodic manner within a crystal; at the same time the experiment proved the wave nature of x-rays. The importance of x-ray diffraction for condensed matter research was immediately recognized at the beginning of the 20th century as evidenced by the award of two

successive Nobel prizes in physics, one 1914 to Max von Laue *for his discovery of the diffraction of x-rays by crystals*" and a second one 1915 to William and Lawrence Bragg "*for their services in the analysis of crystal structure by means of x-rays*" [2]. Both, Laue and the Bragg's, could build on earlier experiments by Geiger and Marsden [3, 4] and interpreted by Ernest Rutherford [5] which proved - again by scattering, this time with alpha particles, - that the atom was composed of a nucleus with a diameter in the femtometer (10^{-15} m) range, while the surrounding electron cloud has a typical extension of 1 Å = 0.1 nm = 10^{-10} m. While this seems trivial to us nowadays, this was a breakthrough discovery at the time since alternate models for the atomic structure with a more continuous distribution of positive and negative charges had been discussed and only scattering methods could provide the final proof of the now well accepted structure of the atom consisting of a tiny nucleus and an extended electron cloud. Since these early experiments, a lot of scattering investigations on condensed matter systems have been done. The overwhelming part of our present-day knowledge of the atomic structure of condensed matter is based on x-ray structure investigations, complemented by electron and neutron diffraction. Electrons due to the strong Coulomb interaction with the atoms suffer multiple scattering events, which make a quantitative evaluation to obtain atomic positions much more difficult. As a probe in condensed matter, electrons made their impact mainly with microscopy techniques (see lecture F6 by Rafal Dunin-Borkowski and Joachim Mayer). Ernst Ruska was awarded the Nobel prize in physics in 1986 "*for the design of the first electron microscope*".

Entirely new possibilities became apparent with the discovery of the neutron by James Chadwick [6, 7]. He received the Nobel prize in physics in 1935. However, for neutrons to become a valuable probe in condensed matter research, they had to be available in large quantities as free particles. This was only possible with the advent of nuclear reactors, where nuclear fission is sustained as a chain reaction. The first man-made nuclear reactor, Chicago Pile-1, was built beneath the west stands of Stagg Field, a former squash rackets court of the campus of the University of Chicago. The reactor went critical on December 2, 1942. The experiment was led by Enrico Fermi, an Italian physicist, who was awarded the Nobel prize in physics in 1938 for this work on transuranium elements. The reactor was a rather crude construction based on a cubical lattice of graphite and uranium oxide blocks. It had no provision for cooling, but two rudimentary manual emergency shutdown systems: one man to cut with an axe a rope, on which a neutron absorbing cadmium rod was suspended, which would drop into the reactor and stop the chain reaction; and a team of three guys standing above the pile ready to flood it with a cadmium salt solution. Considering what was known about nuclear fission at the time it is no wonder that the standing joke among the scientists working there was: *if people could see what we are doing with a million and a half of their dollars, they would think we are crazy. If they knew, why we are doing it, they would know we are* [8]. Out of this very crude first experiment, which was only possible within the World War II Manhattan Project, the modern sophisticated research reactors with their extremely high safety standards developed. In contrast to their big brothers, the nuclear power plants, these reactors are mainly used for isotope production and neutron scattering experiments. The two pioneers of neutron diffraction and inelastic neutron scattering, Clifford G. Shull and Bertram M. Brockhouse, respectively, received the Nobel prize in physics in 1994, many years after the first neutron diffraction experiments, which were performed at Oak Ridge National Lab in 1946. The work of Clifford Shull clearly demonstrated the different contrast mechanisms of neutron scattering compared to x-ray scattering, which in particular allows one to make light elements like hydrogen visible and to distinguish different isotopes like hydrogen and deuterium [9]. But Shull also demonstrated that neutrons, due to their nuclear magnetic moment, could not only be used to determine the arrangement of atoms in solids but

they could also be used to determine the magnetic structure e.g. of antiferromagnetic materials [10]. While Shull studied "*where the atoms are located*" and eventually how the magnetic moments are arranged in the solid, Brockhouse observed for the first time "*how the atoms are moving*" in the solid. He developed the so-called triple-axis spectroscopy, which enables the determination of the dispersion relations of lattice vibrations and spin waves [11].

Since the early work in x-ray and neutron scattering sketched above, many years have passed, new radiation sources such as synchrotron radiation sources, x-ray free electron lasers or neutron spallation sources have been developed, experimental methods and techniques have been refined and the corresponding theoretical concepts established. For the further development of modern condensed matter research, the availability of these probes to study the structure and dynamics on a microscopic level is absolutely essential. It comes as no surprise therefore that scattering methods have been employed in ground-breaking work which led to recent high-level prizes. Examples include the discovery of topologically stabilized magnetic vortices with possible applications in future spintronics or the discovery of magnetic monopole excitations in frustrated magnetic materials (Europhysics Prizes 2016 and 2012, respectively). The Nobel prize in chemistry 2011 was awarded to Dan Shechtman "*for the discovery of quasi-crystals*". By means of electron diffraction, Shechtman discovered icosahedral symmetry in aluminum manganese alloys. The observed tenfold symmetry is not compatible with translational symmetry in three dimensions. While the icosahedral symmetry was discovered with electron diffraction, the question where the atoms are located requires the collection of many weak quasicrystal reflections and the analysis of their intensities, which is only possible with x-ray and neutron diffraction. A higher dimensional reciprocal space approach had to be developed to explain the diffraction pattern of such quasicrystals. Another outstanding piece of work in x-ray diffraction is the Nobel prize in chemistry 2009, which was awarded jointly to Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath "*for studies of the structure and function of the ribosome*". Ribosomes translate DNA information into life by producing proteins, which in turn control the chemistry in all living organisms. By means of x-ray crystallography the Nobel awardees were able to map the position for each and every one of the hundreds of thousands of atoms that make up the ribosome. 3D models that show how different antibiotics bind to the ribosome are now used by scientists in order to develop new drugs.

Here we have given a brief and absolutely incomplete historical summary of the development of scattering methods. This Spring School is devoted to modern applications of this powerful tool. In this introductory overview, we will now give a short summary of which information we can obtain from scattering experiments, compare the two probes x-rays and neutrons briefly, discuss techniques and applications, giving an outlook into the bright future of the field which the two new European facilities promise: the European X-Ray Free Electron Laser X-FEL (www.xfel.eu) and the European Spallation Source ESS (europeanspallationsource.se) and finally explain how large-scale facilities for this type of research are organized.

3   Introduction to scattering

3.1   Scattering cross section

Let us look at a scattering experiment in condensed matter science in the so-called *Fraunhofer- or far-field-approximation*, where we assume that the incident and scattered waves can be described as plane waves with wavelengths λ and λ' (strictly monochromatic)

and propagation direction $\boldsymbol{k}/k$ and $\boldsymbol{k'}/k'$, respectively. Let us define the so-called *scattering vector*

$$\boldsymbol{Q} = \boldsymbol{k'} - \boldsymbol{k} \tag{1}$$

where $\boldsymbol{k}$ and $\boldsymbol{k'}$ are the wave vectors of the incident and scattered radiation, respectively:

$$\underline{k} = \frac{\boldsymbol{k}}{k} \cdot \frac{2\pi}{\lambda}; \underline{k}' = \frac{\boldsymbol{k'}}{k} \cdot \frac{2\pi}{\lambda'} \tag{2}$$

$\hbar\boldsymbol{Q}$ represents the momentum transfer during scattering, since according to de Broglie, the momentum of the particle corresponding to the wave with wave vector $\boldsymbol{k}$ is given by $\boldsymbol{p}=\hbar\boldsymbol{k}$. The magnitude of the scattering vector can be calculated from wavelength λ and scattering angle 2θ (between $\boldsymbol{k}'$ and $\boldsymbol{k}$) as follows

$$Q = |\boldsymbol{Q}| = \sqrt{k^2 + k'^2 - 2kk'\cos 2\theta} \Rightarrow Q = \frac{4\pi}{\lambda}\sin\theta \tag{3}$$

A scattering experiment comprises the measurement of the intensity distribution as a function of the scattering vector. The scattered intensity is proportional to the so-called *cross section*, where the proportionality factors arise from the detailed geometry of the experiment. For a definition of the scattering cross section, we refer to Figure 1.
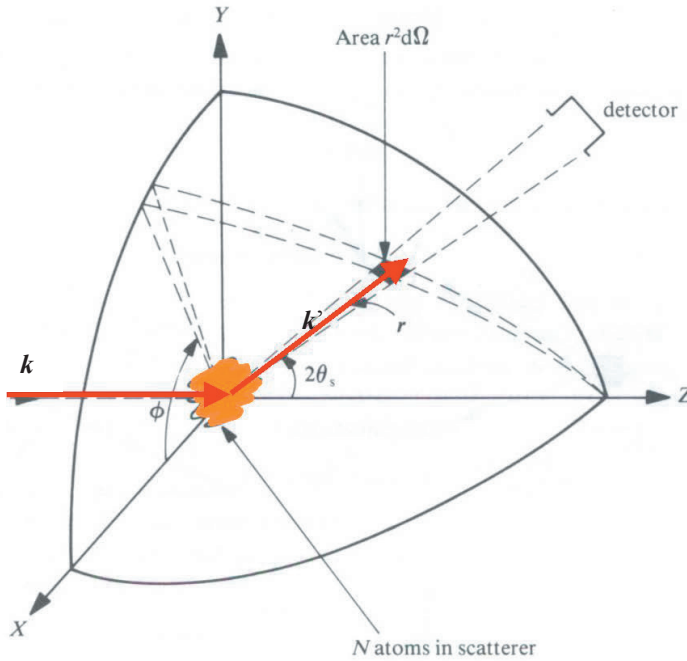
Fig. 1: *Geometry used for the definition of the scattering cross section.*

If n' particles are scattered per second into the solid angle $d\Omega$ seen by the detector under the scattering angle 2θ and into the energy interval between E' and $E' + dE'$, then we can define the so-called *double differential cross section* by:

$$\frac{d^2\sigma}{d\Omega dE'} = \frac{n'}{jd\Omega dE'} \tag{4}$$

Here j refers to the incident beam flux in terms of particles per area and time. If we are not interested in the change of the energy of the radiation during the scattering process, or if our detector is not able to resolve this energy change, then we will describe the angular dependence by the so-called *differential cross section:*

$$\frac{d\sigma}{d\Omega} = \int_0^\infty \frac{d^2\sigma}{d\Omega dE'} dE' \tag{5}$$

Finally the so-called *total scattering cross section* gives us a measure for the total scattering probability independent of changes in energy and scattering angle:

$$\sigma = \int_0^{4\pi} \frac{d\sigma}{d\Omega} d\Omega \tag{6}$$

For a diffraction experiment, our task is to determine the arrangement of the atoms in the sample from the knowledge of the scattering cross section $d\sigma / d\Omega$. The relationship between scattered intensity and the structure of the sample is particularly simple in the so-called *Born approximation*, which is often also referred to as *kinematic scattering approximation* (see lecture A2). In this case, refraction of the beam entering and leaving the sample, multiple scattering events and the extinction of the primary beam due to scattering within the sample are being neglected (these effects are also dealt with in lecture A2). Following Figure 2, the phase difference between a wave scattered at the origin of the coordinate system and at position r is given by

$$\Delta\Phi = 2\pi \cdot \frac{\left(\overline{AB} - \overline{CD}\right)}{\lambda} = \boldsymbol{k'} \cdot \boldsymbol{r} - \boldsymbol{k} \cdot \boldsymbol{r} = \boldsymbol{Q} \cdot \boldsymbol{r} \tag{7}$$
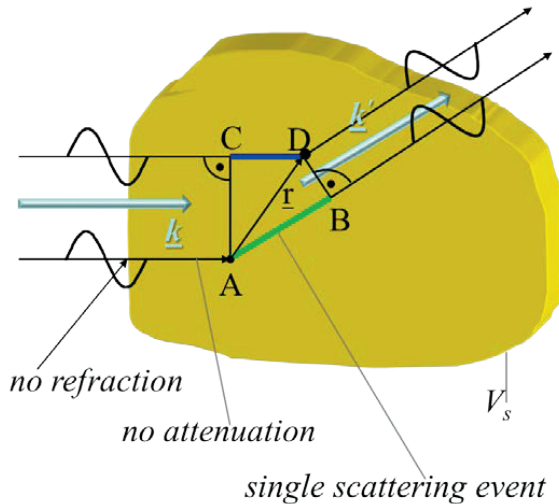
Fig. 2: *A sketch illustrating the phase difference between a beam scattered at the origin of the coordinate system and a beam scattered at the position r. The yellow body represents the sample from which we scatter.*

The probability for a scattering event to occur at position r is proportional to the local interaction potential $V(r)$ between radiation and sample. For a coherent scattering event (interference of scattered waves), the total scattering amplitude is given by a linear

superposition of the waves scattered from all points within the sample volume V_s, i.e. by the integral

$$A(\boldsymbol{Q}) \sim \int\limits_{V_s} V(\boldsymbol{r}) \cdot e^{i\boldsymbol{Q}\cdot\boldsymbol{r}} d^3r \tag{8}$$

This equation demonstrates that the scattered amplitude is directly connected to the interaction potential by a simple Fourier transform: *scattering is a probe in reciprocal space, not in direct space and gives direct access to thermodynamic ensemble averages!*

A knowledge of the scattering amplitude for all scattering vectors $\underline{\boldsymbol{Q}}$ allows us to determine via a Fourier transform the interaction potential uniquely. This is the complete information on the sample, which can be obtained by the scattering experiment. Unfortunately nature is not so simple. On one hand, there is the more technical problem that one is unable to determine the scattering cross section for <u>all</u> values of momentum transfer $\hbar\boldsymbol{Q}$. The more fundamental problem, however, is given by the fact that normally the amplitude of the scattered wave is not measurable. Instead only the scattered intensity

$$I(\boldsymbol{Q}) \sim |A(\boldsymbol{Q})|^2 \tag{9}$$

can be determined. Therefore the phase information is lost and the simple reconstruction of the scattering potential via a Fourier transform is no longer possible. This is the so-called *phase problem* of scattering. There are ways to overcome the phase problem, i.e. by the use of reference waves. Then the potential V(**r**) becomes directly accessible. The question, which information can be obtained from a scattering experiment despite the phase problem, will be addressed below and in subsequent lectures.

Which wavelength do we have to choose to obtain the required real space resolution? For information on a length scale L, a phase difference of about $Q \cdot L \approx 2\pi$ leads from the primary beam *(Q = 0)* to the interference maximum. According to (3) $Q \approx 2\pi/\lambda$ for practical scattering angles *(2θ ~ 60°)*. Combining these two estimates, we end up with the requirement that the wavelength λ has to be in the order of the real space length scale L under investigation. To give an example: with the wavelength in the order of *0.1 nm*, atomic resolution can be achieved in a scattering experiment.

3.2 Coherence

In the above derivation, we assumed plane waves as initial and final states. For a real scattering experiment, this is an unphysical assumption. In the incident beam, a wave packet is produced by collimation (defining the direction of the beam) and monochromatization (defining the wavelength of the incident beam). Neither the direction $\dfrac{\boldsymbol{k}}{k}$, nor the wavelength λ have sharp values but rather have a distribution of finite width about their respective mean values. This wave packet can be described as a superposition of plane waves. As a consequence, the diffraction pattern will be a superposition of patterns for different incident wave vectors $\boldsymbol{k}$ and the question arises, which information is lost due to these non-ideal conditions. This *instrumental resolution* is intimately connected with the *coherence* of the beam. Coherence is needed, so that the interference pattern is not significantly destroyed. Coherence requires a phase relationship between the different components of the beam. Two types of coherence can be distinguished.

• *Temporal or longitudinal coherence* due to a wavelength spread.

A measure for the longitudinal coherence is given by the length, on which two components of the beam with largest wavelength difference (λ and $\lambda+\Delta\lambda$) become fully out of phase. According to the following figure, this is the case for $l_\| = n \cdot \lambda = \left(n - \dfrac{1}{2} \right)(\lambda + \Delta\lambda)$.

$$l_\| = n\lambda = \left(n - \frac{1}{2} \right)(\lambda + \Delta\lambda)$$
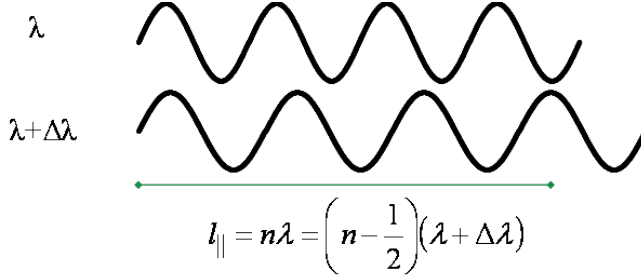
Fig. 3: *A sketch illustrating the longitudinal coherence due to a wavelength spread.*

From this, we obtain the *longitudinal coherence length* $l_\|$ as

$$l_\| = \frac{\lambda^2}{2\Delta\lambda} \tag{10}$$

• *Transversal coherence* due to source extension

Due to the extension of the source (transverse beam size), the phase relation is destroyed for large source size or large divergence. According to the following figure, a first minimum occurs for $\dfrac{\lambda}{2} = d \cdot \sin\theta \approx d \cdot \theta$.
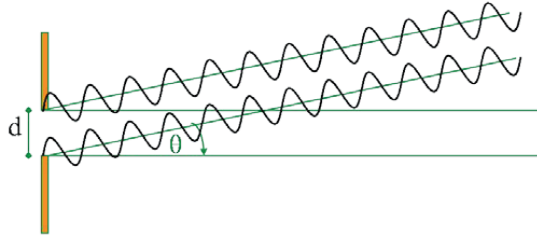
Fig. 4: *A sketch illustrating the transverse coherence due to source extension.*

From this, we obtain the *transversal coherence length* $l_\perp$ as

$$l_\perp = \frac{\lambda}{2\Delta\theta} \tag{11}$$

Here $\Delta\theta$ is the divergence of the beam. Note that $l_\perp$ can be different along different spatial directions: in many instruments, the vertical and horizontal collimations are different.

Together, the longitudinal and the two transversal coherence lengths (in two directions perpendicular to the beam propagation) define a *coherence volume*. This is a measure for a volume within the sample, in which the amplitudes of all scattered waves superimpose to produce an interference pattern. Normally, the coherence volume is significantly smaller than the sample size, typically a few *100 Å* for neutron scattering, up to *μm* for synchrotron radiation. Scattering between different coherence volumes within the sample is no longer coherent, i. e. instead of the amplitudes, the intensities of the waves contributing to the

scattering pattern have to be added. This limits the real space resolution of a scattering experiment to the extension of the coherence volume.

3.3 Pair correlation functions

After having clarified the conditions under which we can expect a coherent scattering process, let us now come back to the question, which information is accessible from the intensity distribution of a scattering experiment. From (9) we see that the phase information is lost during the measurement of the intensity. For this reason, the Fourier transform of the scattering potential is not directly accessible in most scattering experiments (note however that phase information can be obtained in certain cases).

Substituting (8) into (9) and applying variable substitution $\boldsymbol{R}=\boldsymbol{r'}\text{-}\boldsymbol{r}$, we obtain for the magnitude square of the scattering amplitude, a quantity directly accessible in a diffraction experiment:

$$
I(\boldsymbol{Q}) \sim \left| A(\boldsymbol{Q}) \right|^2 \sim \int_{V_S} d^3r' V(\boldsymbol{r}') e^{i\boldsymbol{Q}\cdot\boldsymbol{r}'} \int_{V_S} d^3r V^*(\boldsymbol{r}) e^{-i\boldsymbol{Q}\cdot\boldsymbol{r}} = \iint_{V_S} d^3r' d^3r \, V(\boldsymbol{r}) V^*(\boldsymbol{r}') e^{i\boldsymbol{Q}\cdot(\boldsymbol{r}'-\boldsymbol{r})}
$$

$$
= \iint_{V_S} d^3R \, d^3r \, V(\boldsymbol{R}+\boldsymbol{r}) V^*(\boldsymbol{r}) e^{i\boldsymbol{Q}\cdot\boldsymbol{R}}
\tag{12}
$$

This function denotes the so-called *Patterson function* in crystallography or more general the *static pair correlation function:*

$$
P(\boldsymbol{R}) = \int_{V_S} d^3r V^*(\boldsymbol{r}) V(\boldsymbol{r}+\boldsymbol{R})
\tag{13}
$$

$P(\boldsymbol{R})$ correlates the value of the scattering potential at position $\boldsymbol{r}$ with the value at the position $\boldsymbol{r}+\boldsymbol{R}$, integrated over the entire sample volume V_s. If, averaged over the sample, no correlation exists between the values of the scattering potentials at position $\boldsymbol{r}$ and $\boldsymbol{r}+\boldsymbol{R}$, then the Patterson function $P(\boldsymbol{R})$ vanishes. If, however, a periodic arrangement of a pair of atoms exists in the sample with a difference vector $\boldsymbol{R}$ between the positions, then the Patterson function will have an extremum for this vector $\boldsymbol{R}$. Thus, in a periodic arrangement the Patterson function reproduces all the vectors connecting one atom with another atom.

As will be shown in detail in lecture A4, pair correlation functions are being determined quite generally in a scattering experiment. In a coherent inelastic scattering experiment, we measure a cross section proportional to the *scattering law S(Q,ω)*, which is the Fourier transform with respect to space and time of the spatial and temporal pair correlation function:

$$
\frac{d^2\sigma}{d\omega d\Omega} \sim S(\boldsymbol{Q},\omega) = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} dt \, e^{-i\omega t} \int_{V_S} d^3r e^{i\boldsymbol{Q}r} G(r,t)
\tag{14}
$$

While the proportionality factor between the double differential cross section and the scattering law depends on the type of radiation and its specific interaction potential with the system studied, the spatial and temporal pair correlation function is only a property of the system studied and independent of the probe used:

$$
G(r,t) = \frac{1}{N} \sum_{ij} \int_{V_S} d^3r' \left\langle \delta\left(\boldsymbol{r}' - \boldsymbol{r}_j(0)\right) \cdot \delta\left(\boldsymbol{r}' + \boldsymbol{r} - \boldsymbol{r}_i(t)\right) \right\rangle = \frac{1}{N} \int_{V_S} d^3r' \left\langle \rho(\boldsymbol{r}',0) \rho(\boldsymbol{r}'+\boldsymbol{r},t) \right\rangle
\tag{15}
$$

Here, the pair correlation function is once expressed as a correlation between the position of N point-like particles (expressed by the delta functions) and once by the correlation between the densities at different positions in the sample for different times. In a magnetic system,

radiation is scattered from the atomic magnetic moments, which are vector quantities. Therefore, the scattering law becomes a tensor - the Fourier transform of the *spin pair correlations*:

$$S^{\alpha\beta}(\boldsymbol{Q},\omega) = \frac{1}{2\pi}\sum_l \int dt\, e^{i[\boldsymbol{Q}(\boldsymbol{R}_l - \boldsymbol{R}_0) - \omega t]} \left\langle S_0^\alpha(0) S_l^\beta(t) \right\rangle \tag{16}$$

α, ß denote the Cartesian coordinates *x, y, z;* $\boldsymbol{R}_0$ and $\boldsymbol{R}_l$ are the spatial coordinates of a reference spin *0* and a spin *l* in the system.

3.4 Scattering from a periodic lattice in three dimensions

We now are ready to understand the famous first diffraction experiment by Laue et al. As an example for the application of (8) and (9), we will now discuss the scattering from a three dimensional lattice of point-like scatterers. As we will see later, this situation corresponds to the scattering of thermal neutrons from a single crystal. More precisely, we will restrict ourselves to the case of a Bravais lattice with one atom at the origin of the unit cell. To each atom we attribute a "scattering length b" (see interaction potential of neutrons below). The single crystal is finite with *N, M* and *P* periods along the basis vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. The scattering potential, which we have to use in (8) is a sum over δ-functions for all scattering centers:

$$V(\boldsymbol{r}) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sum_{p=0}^{P-1} b\cdot\delta\left(\boldsymbol{r} - (n\cdot\boldsymbol{a} + m\cdot\boldsymbol{b} + p\cdot\boldsymbol{c})\right) \tag{17}$$

The scattering amplitude is calculated as a Fourier transform:

$$A(\boldsymbol{Q}) \sim b\sum_{n=0}^{N-1} e^{in\boldsymbol{Q}\cdot\boldsymbol{a}} \sum_{m=0}^{M-1} e^{im\boldsymbol{Q}\cdot\boldsymbol{b}} \sum_{p=0}^{P-1} e^{ip\boldsymbol{Q}\cdot\boldsymbol{c}} \tag{18}$$

Summing up the geometrical series, we obtain for the scattered intensity:

$$I(\boldsymbol{Q}) \sim \left| A(\boldsymbol{Q}) \right|^2 = \left| b \right|^2 \cdot \frac{\sin^2\frac{1}{2}N\boldsymbol{Q}\cdot\boldsymbol{a}}{\sin^2\frac{1}{2}\boldsymbol{Q}\cdot\boldsymbol{a}} \cdot \frac{\sin^2\frac{1}{2}M\boldsymbol{Q}\cdot\boldsymbol{b}}{\sin^2\frac{1}{2}\boldsymbol{Q}\cdot\boldsymbol{b}} \cdot \frac{\sin^2\frac{1}{2}P\boldsymbol{Q}\cdot\boldsymbol{c}}{\sin^2\frac{1}{2}\boldsymbol{Q}\cdot\boldsymbol{c}} \tag{19}$$

The dependence on the scattering vector $\boldsymbol{Q}$ is given by the so-called *Laue function* (19), which factorizes according to the three directions in space. One factor along one lattice direction $\boldsymbol{a}$ is plotted in Figure 5.
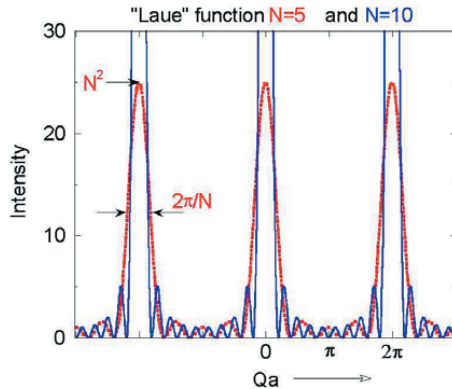
Fig. 5: *Laue function along the lattice direction $\boldsymbol{a}$ for a lattice with five and ten periods, respectively.*

The main maxima occur at the positions $Q = n \cdot 2\pi/a$. The maximum intensity scales with the square of the number of periods N^2, the half width is given approximately by $\Delta Q = 2\pi/(N \cdot a)$. The more periods contribute to coherent scattering, the sharper and higher are the main peaks. Between the main peaks, there are N-2 side maxima. With increasing number of periods N, their intensity becomes rapidly negligible compared to the intensity of the main peaks. The main peaks are of course the well-known *Bragg reflections*, which we obtain for scattering from a crystal lattice. From the position of these Bragg peaks in momentum space, the metric of the unit cell can be deduced (lattice constants *a, b, c* and unit cell angles α, β, γ). The width of the Bragg peaks is determined by the size of the coherently scattering volume (parameters *N, M,* and *P*) - and some other factors for real experiments (resolution, mosaic distribution, internal strains, ...).

Via the so-called Ewald construction, it can be shown that the Laue conditions for interference maxima to occur $\mathbf{Q} \cdot \mathbf{a} = n \cdot 2\pi$ etc. are equivalent to the *Bragg equation* for scattering from lattice planes (hkl) with interplanar spacings d_{hkl}:

$$2d_{hkl} \sin\theta_{hkl} = \lambda \tag{20}$$

4   X-rays and Neutrons

Since the first scattering experiments, some standard probes for condensed matter research have emerged, which optimally fulfil the requirements for a suitable type of radiation.

First of all, electromagnetic radiation governed by the Maxwell equations can be used. Depending on the resolution requirements, X-rays with wavelength λ about 0.1 nm are being used to achieve atomic resolution, or visible light ($\lambda \sim$ 350 - 700 nm) is employed to investigate e. g. colloidal particles in solution. Besides electromagnetic radiation, particle waves can be utilized. It turns out that thermal neutrons with a wavelength $\lambda \sim$ 0.1 nm are particularly well adapted to scattering experiments in condensed matter research. Neutrons are governed by the Schrödinger equation of quantum mechanics. An alternative is to use electrons, which for energies of around 100 keV have wavelengths in the order of 0.005 nm. As relativistic particles, they are governed by the Dirac equation. The big drawback of electrons is the strong Coulomb interaction with the electrons in the sample. Therefore neither absorption, nor multiple scattering effects can be neglected. However, the abundance of free electrons and the relative ease to produce optical elements makes them very suitable for imaging purposes (electron microscopy). Electrons, but likewise atomic beams, are also very powerful tools for surface science: due to their strong interaction with matter, both types of radiation are very surface sensitive. Low Energy Electron Diffraction LEED and Reflection High Energy Electron Diffraction RHEED are both used for in-situ studies of the crystalline structure during thin film growth, e.g. with Molecular Beam Epitaxy MBE. In what follows we will concentrate on the two probes, which are best suited for bulk studies on an atomic scale: x-rays and neutrons. We will touch upon the radiation sources, briefly discuss the main interaction processes and finally give a comparison of these probes.

4.1 X-ray Sources

Since the early days of Conrad Röntgen X-rays are being produced in the laboratory in sealed vacuum tubes, where electrons from a cathode are accelerated towards the anode. There characteristic- and/or bremsstrahlungsradiation is produced. Radiation emitted from such x-ray tubes has been widely used for structural studies in condensed matter science. However,

in 1947 a new type of radiation was discovered in a General Electric synchrotron accelerator [12]. It soon turned out that this so-called *synchrotron radiation* has superb properties, see figure 6.
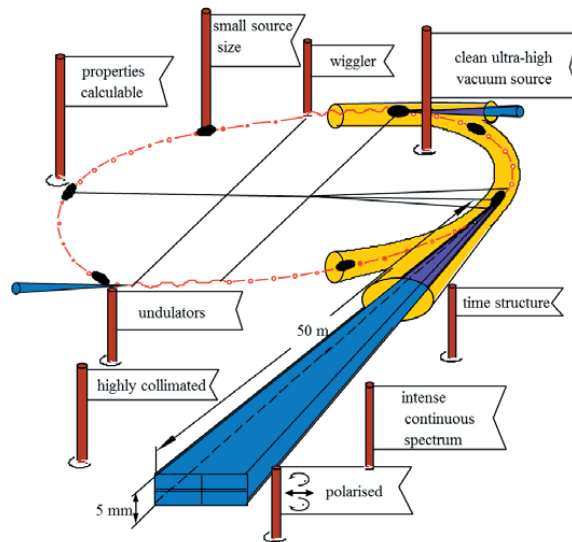
Fig. 6: *Sketch of a synchrotron radiation source indicating the properties of synchrotron radiation.*

Synchrotron radiation is emitted when relativistic charged particles (electrons or positrons) are being accelerated perpendicular to their direction of motion by an appropriate magnetic field. This happens is so-called bending magnets within circular accelerators and this type of radiation has originally been used by solid state physicists in a parasitic mode at particle physics facilities (first generation of synchrotron radiation sources). Second generation synchrotron radiation sources were dedicated to the production of synchrotron radiation, mainly from such bending magnets. However, even more intense radiation can be produced in straight sections of the accelerator by so-called insertion devices - wigglers and undulators - which consist of arrays of magnets with alternating field direction. Modern synchrotron radiation sources of the 3rd generation employ mainly these insertion devices as radiation sources, see lecture C3. This continuous improvement of the source parameters led to an exponential growth of the brilliance, i.e. the spectral photon flux, normalized to the size and divergence of the beam. A further increase of the peak brilliance can be achieved with *free electron lasers*. For the X-ray regime these are based on the SASE principle: *Self Amplified Spontaneous Emission*. In such facilities, an electron beam from a linear accelerator passes through an undulator structure, where synchrotron radiation is produced. The electromagnetic interaction between this radiation and the electron beam travelling in parallel leads to an amplification of the radiation, giving rise to extremely brilliant fully coherent x-ray flashes of about 100 fs duration. Close to DESY in Hamburg such a facility, the *European XFEL* has been realized [13]. Details will be presented in lecture C5 by Robert Feidenhans'l. The facility opens entirely new perspectives for research, see lectures D7 and D8 by Henry Chapman and Simone Techert, respectively.

4.2   Neutron Sources

While neutrons are everywhere - without neutrons we would not exist - they are extremely difficult to produce as free particles, not bound in nuclei. Free neutrons are produced by nuclear physics reactions, which require rather large and high-tech installations. Two main routes to produce free neutrons are being followed today (see figure 7 and lecture C2):

(1) **Fission** of the uranium 235 nuclei in a chain reaction; this process happens in *research reactors*.

(2) Bombarding heavy nuclei with high energetic protons; the nuclei are "heated up" when a proton is absorbed and typically 20 - 30 neutrons are being evaporated. This process is called **spallation** and requires a *spallation source* with a proton accelerator and a heavy metal target station.
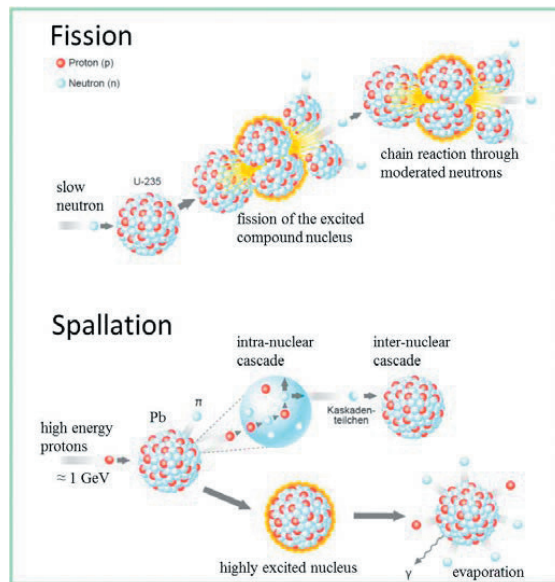
Fig. 7: *A cartoon of the processes of fission and spallation, respectively, used for the production of free neutrons (adapted from the ess-reports).*

Both processes lead to free neutrons of energies in the MeV region. These neutrons are way too fast to be useful for condensed matter studies. These so-called epithermal neutrons have to be slowed down, which is done most efficiently by collisions with light atoms - e.g. Hydrogen H or Deuterium D in light or heavy water moderators, or C in graphite as in the first reactor, the Chicago Pile 1. During the moderation process after several collisions, the neutrons thermalize and acquire the temperature of the moderator. To adjust the energy spectra to ones need, mainly three types of moderators are being employed:

Moderator	Typical temperature [k]	Neutron energy range [meV]	Neutron wavelength range [Å]
hot source (graphite block)	2500	100 - 1000	0.3 - 0.9
thermal source (H_2O / D_2O)	300	5 - 100	0.9 - 4
cold source (liquid D_2)	25	0.05 - 5	4 - 40

Note that room temperature ~ 300 K corresponds an energy of about 26 meV (1 meV $\hat{=}$ 11.6 K), which is just a typical energy of elementary excitations in a solid. Despite the effort made in these high-tech facilities, the free neutrons available for scattering studies are still extremely rare. In a high flux reactor the neutron flux i.e. the number of neutrons passing through a given area in a given time is in the order of 10^{15} neutrons/cm^2·s. If one compares this value with particle fluxes in gases, the neutron density in high flux sources corresponds to high vacuum conditions of about 10^{-6} mbar pressure. The neutrons have to be transported from the source to the experimental areas, which can either be done by simple flight tubes or so-called neutron guides. These are evacuated tubes with glass walls (often covered with metal layers to increase the performance), where neutrons are transported by total reflection from the side, top, and bottom walls in a similar manner like light in glass fibers. The neutron flux downstream at the scattering experiments is then even much lower than in the source itself and amounts to typically 10^6 - 10^8 neutrons/cm^2·s. This means that long counting times have to be taken into account to achieve reasonable statistics in the neutron detector. Just for comparison: the flux of photons of a small Helium-Neon laser with a power of 1 mW (typical for a laser pointer) amounts to some 10^{15} photons/s in a beam area well below 1 mm^2. At modern synchrotron radiation sources, a flux of some 10^{13} photon/s in a similar beamspot can be achieved.

Just as for synchrotron radiation with the XFEL, there is an European project to build the world's most powerful neutron source: the *European Spallation Source ESS*. It will outperform all existing sources by several orders of magnitude in peak flux and allow entirely new experiments to be realized [14]. This facility is currently being built in Lund, Sweden and will be presented by Andreas Schreyer as lecture C6.

Finally, one should emphasize that neutrons can also be generated through nuclear reactions by bombarding targets with protons in the MeV regime. Such facilities driven by compact accelerators will be significantly cheaper and smaller than spallation sources, while their scattering, imaging and analytical instruments promise performances comparable to nowadays medium flux research reactors or spallation sources. The "High Brilliance neutron Source" (HBS) project for such a facility at Forschungszentrum Jülich [15], will be presented by Thomas Gutberlet in lecture C7.

4.3 Interaction Processes

The principle probes for condensed matter studies, X-rays, electrons and neutrons feature different interaction processes with matter, leading to a great complementarity. The principle interaction processes are depicted schematically in fig. 8. Details will be given in lecture A3.
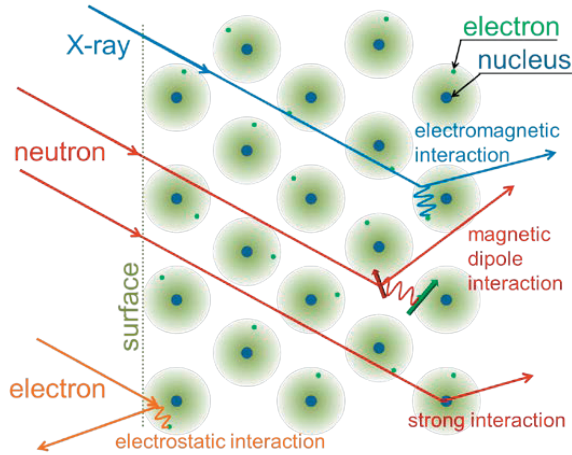
Fig. 8: *Cartoon of the scattering processes of X-rays, neutrons and electrons with atoms in a solid in a 2d representation. The most relevant interaction processes which lead to scattering events are indicated. Note that electrons are mainly scattered in a surface-near region (figure inspired by a lecture by R. Pynn, 1990).*

For X-rays, the most relevant scattering process is pure charge or *Thomson scattering* with the differential cross section for scattering from one electron of:

$$\frac{d\sigma}{d\Omega} = r_0^2 \cdot P(\theta) \tag{21}$$

where $r_0 = \frac{e}{m_0 c^2} = 2.82$ *fm* is the classical electron radius and P(θ) a factor describing the polarization dependence of Hertz' dipole radiation. The Thomson scattering process is the basis for all structural investigations with X-rays since the discovery by Max von Laue one hundred years ago. A single electron is a point-like scatterer, leading to a cross section which is independent of **Q**, apart from the polarization dependence. Scattering from the extended electron cloud of an atom, on the other hand, leads to a variation of the scattered amplitude with scattering angle described by the so-called *formfactor* - the (normalized) Fourier transform of the electron density of a single atom.

Of course, X-rays as electromagnetic radiation also interact with the spin moment of the electron. This so-called *magnetic x-ray scattering* process is a relativistic correction to charge scattering and typically six orders of magnitude weaker. At absorption edges of elements, the scattering amplitude becomes energy dependent, leading to so-called *anomalous scattering*. Anomalous scattering (see figure 9) enables contrast variation, can also be used to enhance small scattering contributions, such as scattering from charge, orbital or magnetic order, see lecture D9. For most such studies, polarization handling is required, see lecture C4.
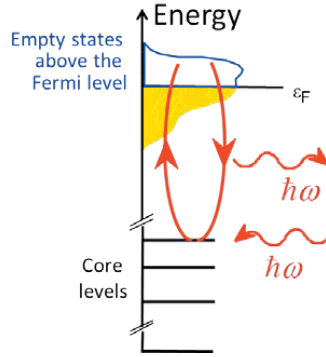
Fig. 9: *Schematic illustration of the second order perturbation process leading to anomalous scattering: core level electros are virtually excited by the incident X-rays into empty states above the Fermi level, if the photon energy is close to an absorption edge. Besides photoelectric absorption, a resonant scattering process can occur, where X-rays of the same wavelength are re-emitted.*

For neutron scattering, two main interaction processes are relevant: scattering with the nucleus due to the strong interaction (*nuclear scattering*) and scattering due to magnetic dipole-dipole interaction between the neutron's magnetic moment and the spin- or orbital moment of unpaired electrons in the solid (*magnetic scattering*).

Since the nucleus is a point-like object compared to the wavelength of thermal neutrons, the differential cross section for nuclear scattering is independent of scattering angle and given by:

$$\frac{d\sigma}{d\Omega} = |b|^2 \tag{22}$$

where b, the scattering length, is a phenomenological parameter as measure of the strength of the interaction potential. b depends not only on the atomic number, but also on the isotope and the nuclear spin orientation relative to the neutron spin.

Magnetic neutron scattering strongly depends on the polarization state of the neutron (for polarization handling, see lecture C4). The differential cross section is given by:

$$\frac{d\sigma}{d\Omega} = \left(\gamma_n r_0\right)^2 \left| -\frac{1}{2\mu_B} \langle \sigma_z{}' | \hat{\boldsymbol{\sigma}} \cdot \boldsymbol{M}_\perp(\boldsymbol{Q}) | \sigma_z \rangle \right|^2 \tag{23}$$

The pre-factor $\gamma_n r_0$ has the value $\gamma_n r_0 = 0.539 \cdot 10^{-12}\ cm = 5.39\ fm$. $\hat{\boldsymbol{\sigma}}$ denotes the spin operator, σ_z and $\sigma_z{}'$ the polarization state of the neutron before and after the scattering process, respectively. $\boldsymbol{M}_\perp(\boldsymbol{Q})$ denotes the component of the Fourier transform of the sample magnetization, which is perpendicular to the scattering vector $\boldsymbol{Q}$:

$$\boldsymbol{M}_\perp(\boldsymbol{Q}) = \hat{\boldsymbol{Q}} \times \boldsymbol{M}(\boldsymbol{Q}) \times \hat{\boldsymbol{Q}}$$
$$\boldsymbol{M}(\boldsymbol{Q}) = \int \boldsymbol{M}(\boldsymbol{r}) e^{i\boldsymbol{Q} \cdot \boldsymbol{r}} d^3 r \tag{24}$$

This tells us that with neutron scattering we are able to determine the magnetization *M(r)* in microscopic atomic spatial co-ordinates *r*, which allows one not only to determine magnetic structures, but also the magnetization distribution within a single atom.

To obtain an idea of the size of the magnetic scattering contribution relative to nuclear scattering, we can replace the matrix element in (23) for a spin ½ particle by the value of one Bohr magneton 1 μ_B. This gives an "equivalent" scattering length for a magnetic scattering of 2.696 fm for a spin ½ particle. This value corresponds quite well to the scattering length of cobalt b_{co} = 2.49 fm, which means that magnetic scattering is comparable in magnitude to nuclear scattering.

4.4   Comparison of Probes

Figure 10 shows a double logarithmic plot of the dispersion relation "wave length versus energy" for the three probes neutrons, electrons and photons. The plot demonstrates, how thermal neutrons of energy 25 meV are ideally suited to determine interatomic distances in the order of 0.1 nm, while the energy of X-rays or electrons for this wavelength is much higher. However, with modern techniques at a synchrotron radiation source, energy resolutions in the meV-region become accessible even for photons of around 10 keV corresponding to a relative energy resolution $\Delta E/E \approx 10^{-7}$ (compare lectures D4 and D6)! The graph also shows that colloids with a typical size of 100 nm are well suited for the investigation with light of energy around 2 eV. These length scales can, however, also be reached with thermal neutron scattering in the small angle region (compare lecture D1). While figure 10 thus demonstrates for which energy-wave-length combination a certain probe is particularly useful, modern experimental techniques extend the range of application by several orders of magnitude.
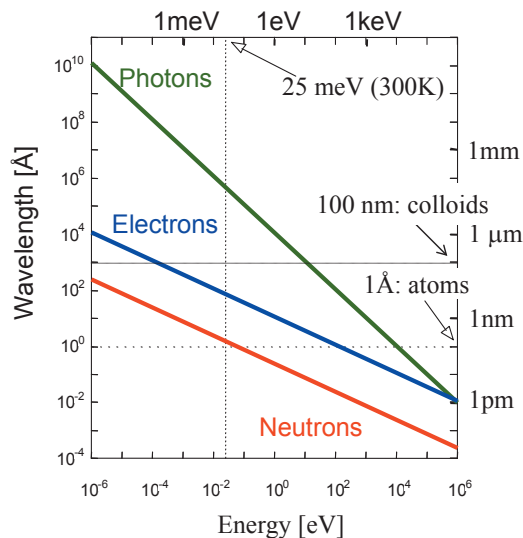
Fig. 10: *Comparison of the three probes - neutrons, electrons and photons - in a double logarithmic energy-wavelength diagram.*

It is therefore useful to compare the scattering cross sections as it is done in figure 11 for X-rays and neutrons. Note that the X-ray scattering cross sections are significantly larger as compared to the neutron scattering cross sections. This means that the signal for x-ray scattering is stronger for the same incident flux and sample size, but that caution has to be

applied that the conditions for kinematical scattering are fulfilled. For X-rays, the cross section depends on the number of electrons and thus varies in a monotonic fashion throughout the periodic table. Clearly it will be difficult to determine hydrogen positions with x-rays in the presence of heavy elements such as metal ions. Moreover, there is a very weak contrast between neighbouring elements as can be seen from the transition metals Mn, Fe and Ni in figure 11. However, this contrast can be enhanced by anomalous scattering, if the photon energy is tuned close to the absorption edge of an element (lecture D9). Moreover, anomalous scattering is sensitive to the anisotropy of the local environment of an atom. For neutrons the cross sections depend on the details of the nuclear structure and thus vary in a non-systematic fashion throughout the periodic table. As an example, there is a very high contrast between Mn and Fe. With neutrons, the hydrogen atom is clearly visible even in the presence of such heavy elements as uranium. Moreover, there is a strong contrast between the two hydrogen isotopes H and D. This fact can be exploited for soft condensed matter investigations by selectively deuterating certain molecules or functional groups and thus varying the contrast within the sample (see lectures E2, E8, E9).
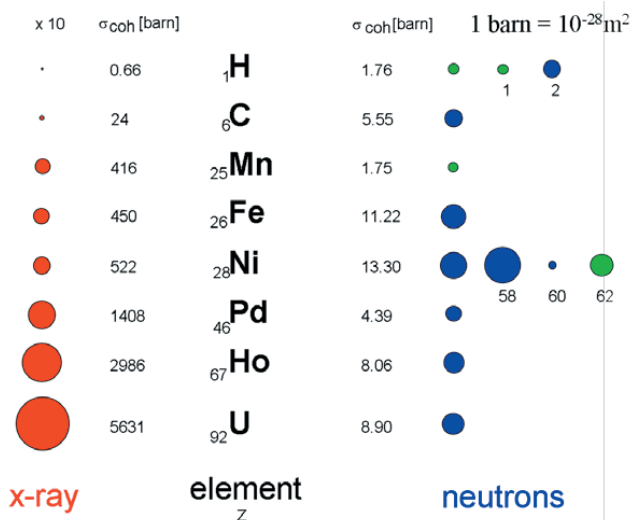
Fig. 11: *Comparison of the coherent scattering cross-sections for x-rays and neutrons for a selection of elements. The area of the coloured circles represent the scattering cross section. In the case of X-rays these areas were scaled down by a factor of 10. For neutrons, the green and blue coloured circles distinguish the cases where the scattering occurs with or without a phase shift of π.*

Finally, both neutrons and X-rays allow the investigation of magnetism on an atomic scale. Magnetic neutron scattering (lectures D4, E3, E5…) is comparable in strength to nuclear scattering, while non-resonant magnetic X-ray scattering is smaller than charge scattering by several orders of magnitude. Despite the small cross sections, non-resonant magnetic x-ray Bragg scattering from good quality single crystals yields good intensities with the brilliant beams at modern synchrotron radiation sources. While neutrons are scattered from the magnetic induction within the sample, X-rays are scattered differently from spin and orbital momentum and thus allow one to measure both form factors separately. Inelastic magnetic scattering e.g. from magnons or so called quasielastic magnetic scattering from fluctuations in disordered magnetic systems is a clear domain of neutron scattering. Finally, *resonance*

exchange scattering XRES, a variant of anomalous X-ray scattering for magnetic systems, allows one not only to get enhanced intensities, but also to study magnetism with element- and band sensitivity (lecture D9). The corresponding technique for the study of excitations is called *Resonant Inelastic X-ray Scattering* (RIXS).

With appropriate scattering methods, employing neutrons, X-rays or light, processes in condensed matter on very different time and space scales can be investigated. Which scattering method is appropriate for which region within the "scattering vector Q - energy E plane" is plotted schematically in figure 12. Via the Fourier transform, the magnitude of a scattering vector Q corresponds to a certain length scale, an energy to a certain frequency, so that the characteristic lengths and times scales for the various methods can be directly determined from the figure.
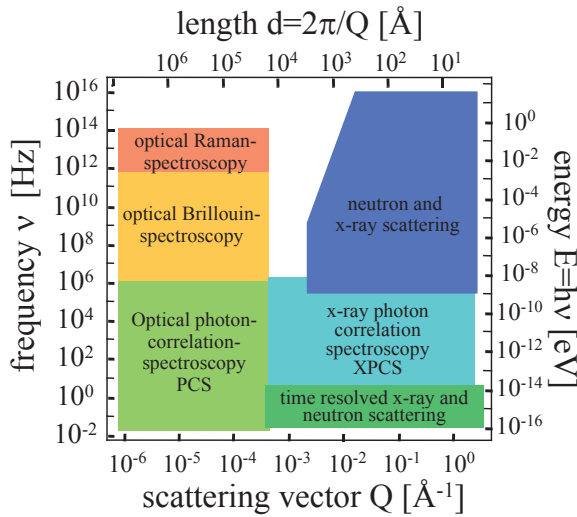
Fig. 12: *Regions in frequency v and scattering vector Q (v,Q)- or energy E and length d (E,d)-plane, which can be covered by various scattering methods.*

5   Techniques and Applications

Scattering with electromagnetic radiation (light, soft- and hard-X-rays) and neutrons cover a huge range of energy and momentum transfers (see figure 12), corresponding to an extraordinary range of length- and time scales relevant for research in condensed matter. Exemplary, this is depicted for research with neutrons in figure 13.

The very extremes of length scales - below 10^{-12} m - are the domain of nuclear and particle physics. For example, neutrons with an energy of 14 MeV, corresponding to wavelengths of 7.6 fm, are used to determine the size and shape of atomic nuclei. Furthermore, high energy neutrons are used for measurements of the charge or the electric dipole moment of the neutron to provide stringent tests of the standard model of particle physics without the need of huge and costly accelerators. On the other extreme, neutrons also provide information on length- and time scales relevant for astronomical dimensions, e. g. the decay series of radioactive isotopes produced by neutron bombardment give information on the creation of elements in the early universe. In this course, however, we are only concerned with neutrons as a probe for condensed matter research and therefore restrict ourselves to a discussion of neutron scattering. Still, the various scattering techniques cover an area in phase space from picometers pm up to meters and femtoseconds fs up to hours, an extremely impressive range!
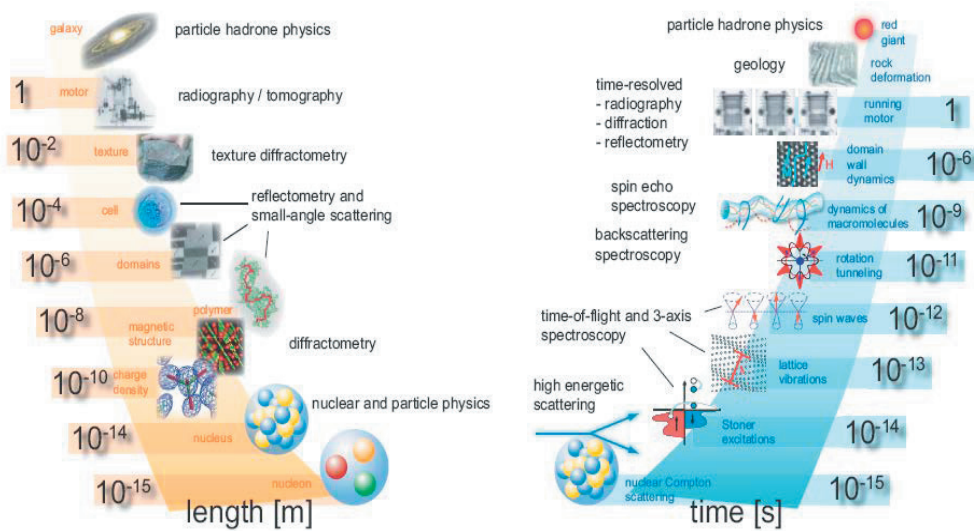
Fig. 13: *Length- and time scales covered by research with neutrons giving examples for applications and neutron techniques.*

Different specialized scattering techniques are required to obtain structural information on different length scales:

- With *wide angle diffractometry*, charge (X-rays) or magnetization (neutrons) densities can be determined within single atoms on a length scale of ca. 10 pm[1]. The position of atoms can be determined on a similar length scale, while distances between atoms lie in the 0.1 nm range (e.g. lectures B1, D3, E8).
- The sizes of large macromolecules, magnetic domains or biological cells lie in the range of nm to μm or even mm. For such studies of large-scale structures, one applies *reflectometry* (lectures D2, E4) or *small angle scattering technique* (lectures D1, E2) or *imaging* (lecture F1).
- Most materials relevant for engineering or geo-science occur neither in form of single crystals, nor in form of fine powders. Instead they have a grainy structure, often with preferred orientation of the grains. This so-called texture determines the macroscopic strength of the material along different directions. *Texture diffractometry* as a specialized technique allows one to determine this granular structure on length scales of up to mm (high energy X-ray diffraction as "3d microscope").
- Finally, for even larger structures, one uses imaging techniques, such as neutron *radiography* or *tomography* (lecture F1), which give a two dimensional projection or full 3-dimensional view into the interior of a sample due to the attenuation of the beam, the phase shift or other contrast mechanisms.

In a similar way, specialized scattering techniques are required to obtain information on the system's dynamics on different time scales:

---

[1] In this sense, X-rays and neutrons are not only nanometer nm, but even picometer pm probes!

- *Neutron Compton scattering*, where a high energy neutron in the eV energy range makes a deep inelastic collision with a nucleus in so-called impulse approximation, gives us the momentum distribution of the atoms within the solid. Interaction times are in the femtosecond fs time range.
- With *pump-probe techniques* at free electron lasers, processes in the fs to ps time range can be studied (lecture D8). For this technique one uses the time structure of the radiation and delays a "probe" pulse with respect to the "pump" pulse to study e.g. relaxation processes after excitation in real time.
- In magnetic metals, there exist single particle magnetic excitations, so-called Stoner excitations, which can be observed with inelastic scattering of high energy neutrons using the so-called *time-of-flight spectroscopy* or the *triple axis spectroscopy* technique (lecture D4). Typically, these processes range from fs to several hundred fs.
- The electronic structure of solids, including electronic relaxation processes in the fs time range, can be determined by X-ray spectroscopy techniques (lecture F2).
- Lattice vibrations (phonons) or spin waves in magnetic systems (magnons) have frequencies corresponding to periods in the picosecond ps time range (lecture B5). Again, these excitations can be observed with neutron *time-of-flight-*, neutron *triple axis spectroscopy* or at high energy resolution backscattering synchrotron beamlines (lecture D4).
- Slower processes in condensed matter are the tunneling of atoms, for example in molecular crystals or the slow dynamics of macromolecules (lectures B4, E2, E9). Characteristic time scales for these processes lie in the nanosecond ns time range. They can be observed with specialized techniques such as neutron *backscattering spectroscopy*, neutron *spin-echo spectroscopy*, light- or X-ray photon correlation spectroscopy (lecture D6).
- Even slower processes occur in condensed matter on an ever-increasing range of lengths scales. One example is the growth of domains in magnetic systems, where domain walls are pinned by impurities. These processes may occur with typical time constants of microseconds μs. Periodic processes on such time scales can be observed with *stroboscopic scattering* techniques.
- Finally, *time resolved scattering* or *imaging* techniques, where data is taken in consecutive time slots, allow one to observe processes from the millisecond ms to the hour h range.

Even within a Spring School of two weeks, it is impossible to cover all scattering techniques and applications. Some techniques will be touched briefly in the application lectures E1-E9, which cover such different fields as research on soft matter (E2) and quantum materials (E3, E1), nanoparticles (E5), engineering and energy materials (E7 & E6) and life science (E8 & E9).

6   Life at large scale facilities

Neutron and x-ray sources are rather expensive to build and to operate. Therefore, only few such installation exist world wide - especially in the field of research with neutrons, where licensing of nuclear installations is an additional aspect to be considered. Figure 14 shows the geographic distribution of the major facilities for research with neutrons. The Jülich Centre for Neutron Science JCNS is present at some of the world's best sources.

Fig. 14: *Major neutron research centers worldwide which have sources of appreciable flux and a broad instrumentation suite for condensed matter research. JCNS is present at four of the leading sources worldwide: the Heinz Maier-Leibnitz Zentrum MLZ at the research reactor FRM II in Garching, Germany, the Institute Laue-Langevin ILL in Grenoble, France, the Spallation Neutron Source SNS in Oak Ridge, USA and the Chinese Advanced Research Reactor CARR close to Beijing, China. JCNS also has a leading involvement in the European Spallation Source project, Lund, Sweden.*

The fact that there are only few sources worldwide implies that experiments at large facilities have to be organized quite different from normal lab-based experiments (see lecture C1). Efforts have to be made to use the existing sources as efficient as possible. This means (i) continuous and reliable operation of the source during a large fraction of the year; (ii) many highly performing instruments, which can run in parallel, located around every source; (iii) professional instrument operation with highly qualified staff and a stringent risk management to keep the downtime of instruments and auxiliary equipment as low as possible; (iv) and access for as many scientists as possible. While there are specialized companies which produce beamline and instrument components, there is no true commercial market for neutron or synchrotron instruments. Therefore, these instruments are being built by research centers, where usually one or a few staff scientists work closely with engineers and technicians to realize an instrument for a certain application. These highly experienced scientists will then later-on also operate the instruments. The Jülich Centre for Neutron Science JCNS has such staff scientists located at the outstations at MLZ, ILL and SNS. However, these large-scale facilities are way too expensive to be operated just for a small number of scientists. Beamtime is offered to external users from universities, research organizations (such as Max-Planck or Fraunhofer in Germany) and industry. In order for these users to obtain access to a scattering instrument, the user will obtain information from the internet on available instruments, contact the instrument scientist and discuss the planned experiments with the instrument scientist. Once a clear idea and strategy for an experiment has been worked out, the user will write a beamtime proposal where he describes in detail the scientific background, the goal of the planned experiment, the experimental strategy and the prior work. The facility issues a call for proposals in regular intervals, typically twice a year. The proposals received are distributed to members of an independent committee of

international experts, which perform a peer review of the proposals and establish a ranking. Typically overload factors between 2 to 3 occur i.e. 2 to 3 times the available beam time is being demanded by external users. Once the best experiments have been selected, the beamtime will be allocated through the facility, where the directors approve the ranking of the committee, the beamline scientist schedules the experiments on the respective instrument and the user office sends out the invitations to the external users. Many facilities will pay travel and lodging for 1 up to 2 users per experiment. It is now up to the user to prepare the experiment as well as possible. If the experiment fails because it was not well prepared, it will be very difficult to get more beamtime for the same scientific problem. Typical experiments last between 1 day and up to 2 weeks. During this time lots of data will be collected which users take home and usually spend several weeks or months to treat the data and model it.

A typical scattering facility will run about 200 days a year with a few hundred visits of user from all over the world. This is also what makes research at large facilities so attractive to young scientists: early-on in their career they will learn to work in large international collaborations, get the opportunity to work on state-of-the-art high-tech equipment and learn to organize their research as efficient as possible. You have therefore chosen well to attend this Spring School!

Conclusion
This overview was meant to give a first introduction to scattering methods, give a glimbse of the possibilities provided at current and future sources and outline the structure of the course. You can now look forward to interesting lectures, where many more details will be explained and you will learn the principles to enable you to successfully perform experiments at neutron and synchrotron radiation sources. Have lots of fun and success working with these powerful techniques!

References

[1] W. Friedrich, P. Knipping, M. von Laue, Interferenzerscheinungen bei Röntgenstrahlen, *Sitzungsberichte der Mathematisch-Physikalischen Klasse der Königlich-Bayerischen Akademie der Wissenschaften zu München* (1912) 303 - 322.

[2] W. H. Bragg, W. L. Bragg, The reflection of x-rays by crystals, *Proc. R. Soc. Lond.* **A88** (1913) 428 - 438.

[3] H. Geiger, E. Marsden, On a diffuse reflection of the alpha particles, *Proc. Roy. Soc.* **A82** (1909) 495 - 500.

[4] H. Geiger, E. Marsden, The laws of deflection of the particles through large angles, *Philosophical Magazine Series* **6** (1913) 25604 - 25623.

[5] E. Rutherford, The scattering of alpha and beta particles by matter and the structure of the atom, *Philosophical Magazine* **21** (1911) 669 - 688.

[6] J. Chadwick, Possible existence of a neutron, *Nature* **129** (1932) 312.

[7] J. Chadwick, The existence of a neutron, *Proc. Roy. Soc.* **A136** (1932) 692 - 708.

[8] http://www.atomicarchive.com/History/firstpile

[9] C. G. Shull, Early development of neutron scattering, *Nobel lecture* (1994) www.nobelprize.org.

[10] C. G. Shull, J. S. Smart, Detection of antiferromagnetism by neutron diffraction, *Physical Review* **76** (1949) 1256 - 1257.

[11] B. N. Brockhouse, Slow neutron spectroscopy and a grant enclass of the physical world, *Nobel lecture* (1994). www.nobelprize.org.

[12] R. R. Elder, A. M. Gurewitch et al., Radiation from Electrons in a Synchrotron, *Physical Review* **71** (1947) 829 - 830.

[13] http://www.XFEL.eu

[14] https://europeanspallationsource.se/

[15] http://www.fz-juelich.de/jcns/

A 2   Scattering Theory: Born series and dynamical theory

Stefan Blügel

Peter Grünberg Institut and Institute for Advanced Simulation
Forschungszentrum Jülich GmbH

Contents

1 Introduction

Since Rutherford's surprise at finding that atoms have their mass and positive charge concentrated in almost point-like nuclei, scattering methods are of extreme importance for studying the properties of condensed matter at the atomic scale. Electromagnetic waves and particle radiation are used as microscopic probes to study a rich variety of structural and dynamical properties of solids and liquids. Atomistic processes in condensed matter take place at length scales on the order of an Ångström (1Å= 10^{-10}m) and an energy scale between a meV and a few eV. Obviously, detailed information concerning atomic systems require measurements related to their behavior at very small separations. Such measurements are in general not possible unless the de Broglie wavelength ($\lambda = \frac{h}{p} = \frac{h}{mv}$) of the relative motion of the probing particle is comparable to these distances. This makes x-ray scattering and neutron scattering, in addition to electron scattering and to a certain extent also Helium scattering, to the outstanding microscopic "measurement instruments" for studying condensed matter. To push electromagnetic waves in this area one uses either x-rays with wavelengths of a few Ångströms, but in the keV energy range, or light with energies in the eV range, but wavelengths of some 1000 Å. Neutrons (and Helium atoms) make it possible to match energy and wavelengths simultaneously to the typical atomic spacings and excitation energies of solids (solid surfaces). The most important scattering properties of the probing particles are collected in Tabel. 1.

	neutrons	x-rays	electrons	LEED
energy E	10 meV	10 keV	100 keV	100 eV
wavelength λ	1 Å	1 Å	0.05 Å	1 Å
extinction length d_{ext}	10^5 Å	10^4 Å	$10^2 - 10^3$ Å	5 Å
absorption length $1/\mu_0$	10^8 Å	10^5 Å	$10^3 - 10^4$ Å	10 Å
$1/\Delta\mu$	$> 10^8$ Å	$30 \cdot 10^5$ Å	$3 \cdot (10^3 - 10^4)$ Å	10 Å

Table 1: Basic parameters for the diffraction of neutrons, electrons, x-rays and in the case low-energy electron diffraction (LEED). The extinction length, which is essentially the thickness of the crystal for which the kinematic theory breaks down and the absorption length changes by orders of magnitude between different probing particles. $\Delta\mu = \mu_0 \pm \mu_G$ is the absorption length in the case when the Bragg reflection is excited. For the case of Bragg diffraction, the absorption of the x-rays is reduced by a very large factor, known as the *Borrmann effect*, whereas the absorption of electrons is only slightly reduced.

Thus, a simultaneous spatial and temporal resolution of atomistic or magnetic processes is possible. In addition, the photon, neutron, electron and under certain conditions also Helium possess internal degrees of freedom such as a polarization vector or a spin with which the probes couple to core and valence electrons. The photon, neutron and Helium result in only week interaction with matter, which simplifies considerably the analysis and interpretation of experiments as multiple scattering processes are frequently of minor importance and can often be ignored completely, which makes an interpretation of the scattering results valid within a kinematic scattering theory.

Within the kinematic theory, it is assumed that the incoming wave of e.g. x-rays is scattered from the objects, which perform the scattering, e.g. atoms, only once. After such "once-scattering"

the intensities of the scattered waves are added taking into account the phase differences of the scattered waves in order to form the intensities of the transmitted and reflected beams. What is neglected in the kinematic theory is the interaction of the "once-scattered" waves with each other and the matter. In terms of rigorous scattering theory, the kinematic theory corresponds to the (1st) *Born approximation*, discussed in detail in the first part of the chapter lecture.

Very often, however, electrons are used as particles or the diffraction of waves and particles on large crystals of very good crystalline quality is investigated, and the intensities of the transmitted and reflected beams are measured after the waves forming them traverse large regions of space filled with atoms. In this case the interaction between the scattered waves and atoms cannot be ignored anymore, and the so-called *multiple scattering* of waves inside the crystal, which technically corresponds to going beyond the Born approximation, has to be considered in order to explain observed phenomena which cannot be understood within the kinematic theory. For centrally symmetric potentials the partial wave expansion, an angular momentum expansion of the scattered waves, offers an elegant description of the multiple scattering in terms of a partial wave scattering amplitude expressed in phase shifts, a direction we do not follow in this chapter due to lack of time. The description of multiple scattering can be essentially simplified if the perfect periodicity of the crystal is assumed. The *dynamical scattering theory* describes the interaction of waves with a regular lattice of atoms such as atomic crystal structures, or nanometer scaled multilayers, or self-arranged systems and it includes all multiple scattering effects. The variety of effects, which can be explained within the dynamical scattering theory is vast, and there are considerable differences for different types of radiation.

In this Chapter, we will provide a brief introduction to the elementary concepts and methodology of scattering theory. The focus lies on the introduction of the description of the scattering process in terms of the Hamiltonian of the scattering projectile at a finite range interaction potential of a single site target using the Lippmann-Schwinger equation, an integral equation formulation of scattering that leads then to the first Born approximation of scattering and the distorted wave Born approximation. The former is the approximation of choice if multiple scattering is unimportant and the latter is applied in the analysis of grazing-incidence small-angle scattering experiments discussed in more details chapter D2. The subject is typically part of an Advanced Quantum Mechanics curriculum and is therefore elaborated at textbooks on quantum mechanics. A selection is given as references [1–3]. The lecture continues with the discussion of the scattering on the lattice rather than a single site target resulting at the Bragg scattering [4], that will be discussed in more detail in the succeeding chapter. The foundations of the *dynamical scattering theory*, which goes beyond the kinematical scattering theory of diffraction along these lines, were set already by Darwin in 1914 [5]. On a more fundamental level the problem was treated by Ewald in 1917 [6] and later by Laue [7]. For electrons, the problem has been tackled further by Bethe [8]. Since then, many good books and reviews on the subject have been published, see e.g. the books of Zachariasen [9], Cowley [10] or Authier [11], or reviews by Slater [12], and Batterman and Cole [13]. In our manuscript, we provide a simple introduction into the dynamical scattering theory following mainly the review by Slater [12] and a beautiful review by P. H. Dederichs [14].
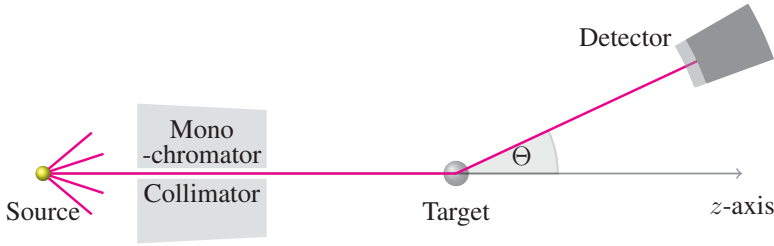
Fig. 1: A set-up of a typical diffraction experiment consisting of a particle source, a scattering target and a particle detector. The beam passes through a collimator and a monochromator with a beam opening Δr.

2 The Scattering Problem

In a scattering experiment a beam of particles is allowed to strike a target,[1] and the particles that emerge from the target area or *scattering volume*, respectively, are observed.

The Experimental Situation

A schematic representation of a standard scattering experiment appears in Fig. 1. Each scattering experiment consists of three indispensable elements: (i) The source of incident beam of particles or electromagnetic wave, to propagate with wave vector $\mathbf{k} = \frac{2\pi}{\lambda} \hat{\mathbf{k}}$ of wavelength λ along the direction $\hat{\mathbf{k}}$, which we assume without loss of generality to be the $\hat{\mathbf{z}}$-direction (the axis of the collimator). (ii) The target, that we consider stationary, a reasonable assumption in condensed matter physics and (iii) the detector, whose function is to simply count the number of particles of a particular type that arrive at its position $\mathbf{r}$ with the coordinate $\mathbf{r} = (x, y, z)$ in real space along the direction $\hat{\mathbf{r}}$ at the angle (θ, φ) with respect to the axis of the propagating incident beam. Ideally it may be set to count only particles of a given energy, spin or polarization vector, respectively. We assume throughout that the source and detector are classical objects that have a clearly defined, precisely controllable effect on the scattering process. The detector will be assumed to be 100% efficient and to have no effect on the scattered particle prior to the time it enters the detector.

The result of the scattering experiment will vary with the energy E of the incident beam. In order to simplify the analysis of the experimental data, the energy spectrum of the incident beam should be sharply peaked so that the experiment may be considered to take place at a unique energy eigenvalue E. To this end, in most experiments care is taken to achieve a monochromatic incident beam characterized by the wave vector $\mathbf{k}$. We shall assume here that the beam emerging from the collimator and monochromator is both perfectly monochromatic and perfectly collimated, as well. Of course, according to the uncertainty principle, a beam of finite cross section (of the size of the collimator opening Δr) cannot be perfectly monochromatic ($\Delta k > 0$) and perfectly collimated as well. We may, however, assume the beam to be suffi-

---

[1] In the language of elementary scattering theory one frequently refers to a target although we keep in mind that in the language of condensed matter, it is referred to as the sample.

ciently well collimated that the angular divergence may be ignored in an actual experiment. In that case we must necessarily have not a monochromatic beam represented by a plane wave, but rather a beam describable as a superposition of such waves. In this respect the collimator and monochromator can be considered the fourth indispensable elements of a scattering experiment. It shields the detector from the incoming beam to the ideal extent that no count is measured in the detector without target, and produces a small beam of monochromatic energy. We further assume that the detector has a small opening angle $d\Omega$ and is positioned at large distance from the target. Under these conditions, the scattered beam can be characterized at the position of the detector by the wave vector $\mathbf{k}'$ and energy E'. Summarizing, in a scattering experiment a wave packet of incident particles characterized by the initial state $(\mathbf{k}, E, \mathbf{e})$, denoting the polarization vector $\mathbf{e}$ of an x-ray beam as a representative of an internal degree of freedom of the particle, is scattered into the final state:

$$(E, \mathbf{k}, \mathbf{e}) \stackrel{\text{scattering}}{\longrightarrow} (E', \mathbf{k}', \mathbf{e}'). \tag{1}$$

The scattering process is characterized by the *scattering vector*

$$\mathbf{Q} = \mathbf{k}' - \mathbf{k} \tag{2}$$

and the *energy transition*

$$\hbar\omega = E' - E. \tag{3}$$

$\hbar\mathbf{Q}$ represents the momentum transfer during scattering, since according to de Broglie, the momentum of the particle corresponding to the wave with wave vector $\mathbf{k}$ is given by $\mathbf{p} = \hbar\mathbf{k}$. For *elastic scattering* (*diffraction*), it holds that $E' = E$ and $|\mathbf{k}'| = |\mathbf{k}|$ and all possible scattering vectors are located on a sphere, called *Ewald-sphere*. Structural investigations are always carried out by elastic scattering. The magnitude Q of the scattering vector can be calculated from wavelength λ and scattering angle θ as follows

$$Q = \sqrt{k^2 + k'^2 - 2kk'\cos\theta} = \sqrt{2k^2(1 - \cos\theta)} = k\sqrt{2(1 - \cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2})} = \frac{4\pi}{\lambda}\sin\frac{\theta}{2} \tag{4}$$

Description of Scattering Experiment

After the beam of particles is emitted from the collimated source, the experimenter has no control over the particles until they have reached his detector. During that time, the propagation is controlled solely by the laws of quantum mechanics and the Hamiltonian of the projectile-target system. We restrain our description to the nonrelativistic domain and may thus formulate the physical situation in terms of the solution of a Schrödinger equation using an appropriate Hamiltonian and suitable boundary conditions. The occurring phenomena can be very complex. We assume in the following that the incident particles do not interact with each other during the time of flight, rather that the particles fly one by one, and that incident particles and the target particles do not change their internal structures or states, but only scatter off each other. Internal excitations, rearrangements, charge or spin exchange are also excluded. In fact, internal degrees of freedom such as the spin or polarization vector are currently completely neglected. That means we shall consider only the purely *elastic scattering*. We further neglect the *multiple scattering* in the target and consider at first only the interaction of an incident particle with one
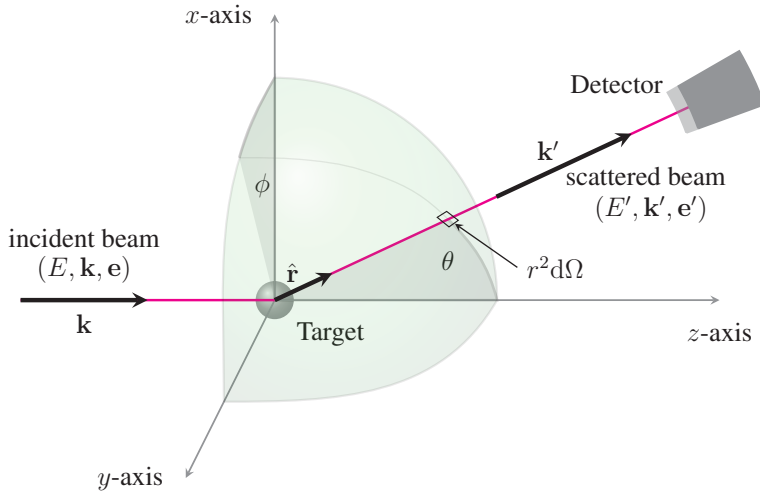
Fig. 2: The geometry of the scattering experiment.

target particle, the state of both is described by a two-particle wave function $\Psi(\mathbf{r}_P, \mathbf{r}_T)$ and the interaction is described by a potential $V(\mathbf{r})$ which depends only on the relative distance $\mathbf{r} = \mathbf{r}_T - \mathbf{r}_P$, where the subscripts T and P denote the target and the incoming particle, respectively. We shall confine ourselves in this article to scattering processes in which only short-range central forces are present. In the presence of such potentials, the particle is not under influence of the target potential when they are emitted from the source or when they enter the detector. Since for the two-body problem the motion of the center of mass $\mathbf{R}$ can be separated out, the problem reduces to the scattering of a particle with the reduced mass $\frac{1}{m} = \frac{1}{m_T} + \frac{1}{m_P}$ at the potential $V(\mathbf{r})$. Since the potential does not depend explicitly on the coordinate of the center of mass $\mathbf{R}$, the two-particle wave function can be expressed in terms of a product of single-particle wave functions $\Psi(\mathbf{r}_P, \mathbf{r}_T) = \phi(\mathbf{R})\psi(\mathbf{r})$,[2] both solutions of two separate Schrödinger equations. The elementary two-body scattering process could be intrinsically elastic, but recoil of the target particle might lead to a transfer of energy to the target. In the present context elastic scattering specifically excludes such effects. Considering a solid as target, depending on the energy of the projectile and the interaction of the constituent atoms in a solid, this can be a very good assumption, as for favorable circumstances all atoms contribute to the scattering mass of the solid (of course, in other chapters it becomes clear that atoms in a solid vibrate and a scattering event may cause inelastic excitations of phonons in the vicinity of the elastic energy). To think that the target particle or a target solid is infinitely heavy relative to the mass of the incident particle simplifies our thinking further. In this case the center of mass of the target remains stationary at the position of the target particle throughout the scattering process. Under these circumstances the relative coordinate $\mathbf{r}$ represents the actual laboratory coordinate of the light particle and the mass m is then the mass of the incident particle. These parameters enter the

---

[2] This does not hold if target and projectile are identical particles. Identical articles scattering about angle θ and $\pi - \theta$ cannot be distinguished.

time-dependent Schrödinger equation

$$i\hbar\frac{\partial \psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right]\psi \qquad \text{with} \qquad V(\mathbf{r}) = 0 \qquad \text{except } \mathbf{r} \in \text{ target region } \mathbb{T} \quad (5)$$

to be solved.

The time-dependent Schrödinger equation seems a natural starting point for the description of a scattering event as it is not a stationary process but involves individual discrete particles as projectiles, but as we see a bit later the good news is, that under reasonable assumptions that are fulfilled in typical experimental situations, the same results are obtained using a time-independent description applying the stationary Schrödinger equation.

At the vicinity of the collimator and detector, the solution of the potential-free Schrödinger equation (5) is analytically known as the free-particle wave packet:

$$\psi(\mathbf{r},t) = \frac{1}{(2\pi)^3}\int d^3k\, A(\mathbf{k})\, \psi_{\mathbf{k}}(\mathbf{r},t) \qquad \text{with} \qquad \psi_{\mathbf{k}}(\mathbf{r},t) = e^{i\mathbf{k}\mathbf{r}}e^{-i\frac{\hbar k^2}{2m}t}. \qquad (6)$$

The wave packet is expressed as a superposition of a complete set of stationary-state solutions $\psi_{\mathbf{k}}(\mathbf{r})$ of this Schrödinger equation, which are plane waves $\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}}$. The energy eigenvalue corresponding to the eigenfunction $\psi_{\mathbf{k}}$ is simply $E_{\mathbf{k}} = \frac{\hbar^2}{2m}k^2$. For convenience the target is placed at the origin of the coordinate system ($r = 0$). The coefficient $A(\mathbf{k})$ is the probability amplitude for finding the wave number $\mathbf{k}$, or momentum $\hbar\mathbf{k}$ in the initial state. We assume that the properties of the source are such that the wave packet is close to monochromatic and that the amplitude function $A(\mathbf{k})$ peaks about the average momentum $\hbar\mathbf{k} = \hbar\mathbf{k}_\circ$ with a spread in the wave number $\Delta\mathbf{k}$ that is small compared to $\mathbf{k}_\circ$ ($\Delta\mathbf{k} \ll \mathbf{k}_\circ$), related to the opening of the collimator $\Delta\mathbf{r}$ by Heisenberg's uncertainty principle ($\Delta\mathbf{k}\Delta\mathbf{r} \simeq 1$). If we finally impose the condition $\frac{(\Delta\mathbf{k})^2}{\mathbf{k}_\circ}L \ll 1$ (equivalent to the condition $\frac{\lambda_\circ}{\Delta r} \ll \frac{\Delta r}{L}$) the packet does not spread appreciably during the course of the experiment with the set-up of length L, and we can finally show that a stationary description of the scattering problem is sufficient. Under these conditions the energy-dependent phase factor of $\psi_{\mathbf{k}}(\mathbf{r},t)$ can be conveniently be approximated about the median energy $k^2 \simeq -k_\circ^2 + 2\mathbf{k}\cdot\mathbf{k}_\circ$ and Eq. (6) will become

$$\psi(\mathbf{r},t) = e^{i\frac{\hbar k_\circ^2}{2m}t}\frac{1}{(2\pi)^3}\int d^3k\, A(\mathbf{k})\, e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{v}_\circ t)} = e^{i\frac{\hbar k_\circ^2}{2m}t}\psi(\mathbf{r}-\mathbf{v}_\circ t,0), \qquad (7)$$

i.e. a wave packet $\psi(\mathbf{r},t)$ centered about the origin $r = 0$ at $t = 0$ moves at the classical velocity $\mathbf{v}_\circ = \frac{\hbar\mathbf{k}_\circ}{m}$ and the packet at time $t > 0$ will have exactly the same shape, but centered about $\mathbf{r} = \mathbf{v}_\circ t$. Thus, the initial state $\lim_{t\to-\infty}\psi(\mathbf{r},t)$ and the final state $\lim_{t\to+\infty}\psi(\mathbf{r},t)$ can be expressed by Eq. (7), but with the coefficients $A(\mathbf{k})$ in the final state having been modified compared to the ones in the initial state due to the scattering, as we shall discuss immediately.

A solution of Eq. (5) requires, however, the specification of boundary conditions imposed on the solution that reflect the physical situation in the laboratory as discussed in section 2. The proper boundary condition is a condition on the wave function when the particle and target are far apart. It can be motivated from *Huygens' principle* [15] who proposed that every point which a luminous disturbance reaches becomes a source of a spherical wave, and the sum of these secondary waves determines the form of the wave at any subsequent time. For a single target scatterer we express the wave function

$$\psi_{\mathbf{k}}(\mathbf{r}) \overset{r\to\infty}{\longrightarrow} e^{i\mathbf{k}\mathbf{r}} + \frac{1}{r}e^{ikr}f_{\mathbf{k}}(\theta,\phi) \qquad \forall\mathbf{k} \quad \text{and} \quad t > 0 \qquad (8)$$

in terms of a superposition for the incoming wave plus an outgoing scattered wave emanating from the target, removing some of the incoming particles from the incident primary beam. $f(\theta, \phi)$, $f(\hat{\mathbf{r}})$ or $f(\hat{\mathbf{k}}')$, respectively, denotes the scattering amplitude. This form of the wave function is motivated by the fact that we expect, after scattering, an outgoing spherical wave, modified by the scattering amplitude, interfering with the incoming wave; we will later show a more rigorous justification of this expression. Consistent to the lab schematics in Fig. 1, $\mathbf{k}_\circ \| \hat{\mathbf{z}}$ and the azimuthal and polar scattering angle (θ, ϕ) are given by the projection of the direction of the wavevector $\hat{\mathbf{k}}$ of the scattered wave, e.g. into the detector at direction $\hat{\mathbf{r}}$ and the $\hat{\mathbf{z}}$ direction. If we replace $\psi_\mathbf{k}(\mathbf{r}, t)$ in (6) by its asymptotic form given in (8), when the packet is far from the target, the wave function $\psi'(\mathbf{r}, t)$ (where we use a prime to denote the wave function after scattering) breaks up into two terms

$$\psi'(\mathbf{r}, t) = \psi_\mathbf{k}(\mathbf{r}, t) + \psi_{\mathrm{sc}}(\mathbf{r}, t), \qquad \text{with} \quad \psi_{\mathrm{sc}}(\mathbf{r}, t) = 0 \quad \text{for} \quad t < 0 \tag{9}$$

with the incident beam $\psi_\mathbf{k}(\mathbf{r}, t)$ identical to Eq. (7) and the scattered wave $\psi_{\mathrm{sc}}(\mathbf{r}, t)$ according to Eq. (6)

$$\psi_{\mathrm{sc}}(\mathbf{r}, t) = \frac{1}{r} \frac{1}{(2\pi)^3} \int d^3k \, f_\mathbf{k}(\hat{\mathbf{r}}) A(\mathbf{k}) \, e^{ikr} \, e^{-i\frac{\hbar k^2}{2m}t}. \tag{10}$$

If we assume now that the scattering amplitude is slowly varying over the spread of the wave numbers $\Delta \mathbf{k}$, and thus approximate $f_\mathbf{k}(\hat{\mathbf{r}}) \simeq f_{\mathbf{k}_\circ}(\hat{\mathbf{r}})$ as well as making use of the approximations

$$k \simeq \hat{\mathbf{k}}_\circ \cdot \mathbf{k} \qquad \text{and} \qquad k^2 \simeq -k_\circ^2 + 2\mathbf{k} \cdot \mathbf{k}_\circ \qquad \text{for} \qquad (\Delta k)^2 \ll k_\circ^2 \tag{11}$$

$(k = |\mathbf{k}| = |\mathbf{k}_\circ + \Delta \mathbf{k}| = [k_\circ^2 + 2\mathbf{k}_\circ \cdot \Delta \mathbf{k} + (\Delta k)^2]^{1/2} \simeq k_\circ [1 + 2\mathbf{k}_\circ \cdot \Delta \mathbf{k}/k_\circ^2]^{1/2} \simeq k_\circ + \hat{\mathbf{k}}_\circ \cdot (\mathbf{k} - \mathbf{k}_\circ))$ in which k is equal to the projection along $\mathbf{k}_\circ$ we obtain

$$\psi_{\mathrm{sc}}(\mathbf{r}, t) = e^{-i\frac{\hbar k_\circ^2}{2m}t} \frac{1}{r} f_{\mathbf{k}_\circ}(\hat{\mathbf{r}}) \frac{1}{(2\pi)^3} \int d^3k \, A(\mathbf{k}) \, e^{i\mathbf{k} \cdot (\hat{\mathbf{k}}_\circ r - \mathbf{v}_\circ t)} = e^{-i\frac{\hbar k_\circ^2}{2m}t} \frac{1}{r} f_{\mathbf{k}_\circ}(\hat{\mathbf{r}}) \, \psi(\hat{\mathbf{k}}_\circ r - \mathbf{v}_\circ t, 0) \,. \tag{12}$$

Thus, after the incident packet has passed the target a spherical scattered wave shell of thickness Δr, equal to the size of the packet, centered on the origin and having a radius $r = v_\circ t$ emerges from the target. One finds further that the incoming wave packet given in (7) and the scattered wave, Eq. (12), share absolutely the same time dependence and are the same for all k. The solution is actually a superposition of all available wave numbers according to $\int d^3k A(\mathbf{k}) \ldots$. Since there is no mode-mode coupling such as $\mathbf{k} \to \mathbf{k_1} + \mathbf{k_2}$, it is totally sufficient to solve the problem in terms of a scattering problem of $\psi_\mathbf{k}(\mathbf{r})$ on the basis of a stationary Schrödinger equation for all relevant wave vectors $\mathbf{k}$, which will be pursued during the rest of the manuscript, and keep thereby in mind that a wave packet is formed with a certain probability amplitude. This stationary problem with plane waves as incident beam simplifies the description of scattering significantly.

Coherence

The formation of a wave packet bears, however, a consequence on which we shall briefly touch upon: The scattering pattern or diffraction pattern, respectively, will be a superposition of patterns for different incident wave vectors $(\mathbf{k}, \mathbf{k} + \Delta \mathbf{k})$ and the question arises, which information is lost due to these non-ideal conditions. This "instrumental resolution" is intimately connected

with the "coherence" of the beam and the size of the *scattering volume* in comparison to the *target volume*. Coherence is needed, so that the interference pattern is not significantly destroyed. Coherence requires a phase relation between the different components of the beam. A measure for the coherence length l is given by the distance, at which two components of the beam become fully out of phase, i.e. when one wave train at position $\mathbf{r}$ exhibits a maximum, meats a wave train exhibiting a minimum, thus experiencing a phase difference of $\lambda/2$. If the coherence length $l_\parallel$ is determined by the wavelength spread, λ and $\lambda + \Delta\lambda$ one refers to the temporal or longitudinal coherence. The condition $l_\parallel = n\lambda = (n - \frac{1}{2})(\lambda + \Delta\lambda)$ translates then into

$$l_\parallel = \frac{1}{2}\frac{\lambda^2}{\Delta\lambda} \qquad \text{for longitudinal coherence}$$

and

$$l_\perp = \frac{1}{2}\frac{\lambda}{\Delta\theta} \qquad \text{for transversal coherence.}$$

Analogously one obtains the transversal coherence length $l_\perp$ shown in above equation due to the divergence of the beam $\Delta\theta$ that results from the finite transverse beam size due to the finite extension of the source. In many instruments, the vertical and horizontal collimations are different and the vertical one can even be different along different spatial directions.

Together, the longitudinal and the two transversal coherence lengths define a coherence volume. This is a measure for a volume within the sample, in which the amplitudes of all scattered waves superimpose to produce an interference pattern. Normally, the coherence volume is significantly smaller than the sample size, typically a few 100 Å for neutron scattering, up to μm for synchrotron radiation. Scattering between different coherence volumes within the sample is no longer coherent, i.e. instead of the amplitudes, the intensities of the contributions to the scattering pattern have to be added. This limits the spatial resolution of a scattering experiment to the extension of the coherence volume.

The Cross Section

A general measure of the scattered intensity $I(\Omega)$ is the *differential cross section* $(\frac{d\sigma}{d\Omega})$. It is defined by the number of particles dN counted per unit time dt scattered into a cone of solid angle $d\Omega = \sin\theta\, d\theta\, d\phi$ in the detector located at the distance r along a ray specified by the direction $\hat{\mathbf{r}}$, at angle (θ, ϕ) or the solid angle Ω, respectively, normalized to the current of the incoming particles j_{in}

$$\frac{1}{j_{\text{in}}}\frac{dN}{dt} = \left(\frac{d\sigma}{d\Omega}\right)d\Omega = \left(\frac{d\sigma}{d\Omega}\right)\frac{1}{r^2}dA. \tag{13}$$

$d\sigma$ describes a cross-sectional area with a surface normal parallel to $\mathbf{k}_\circ$, through which the number of particles dN that get scattered into the angle Ω flow per unit time. The total cross section

$$\sigma_{\text{tot}} = \int_0^{4\pi}\left(\frac{d\sigma}{d\Omega}\right)d\Omega \tag{14}$$

is the total effective geometrical cross-sectional area of the incident beam that is intercepted and the particles therein deflected by the target object.

From Eq. (13) the scattered current density is $j_{sc} = \frac{1}{r^2} j_{in} (\frac{d\sigma}{d\Omega})$. On the other hand j_{sc} can be calculated directly employing the expression of the probability current density given by

$$\mathbf{j}_{sc}(\mathbf{r}) = -i\frac{\hbar}{2m} \left[\psi_{sc}^*(\mathbf{r}) \nabla \psi_{sc}(\mathbf{r}) - \psi_{sc}(\mathbf{r}) \nabla \psi_{sc}^*(\mathbf{r})) \right] \simeq j_{in} \frac{1}{r^2} |f(\Omega)|^2 \,\hat{\mathbf{r}} + \mathcal{O}\left(\frac{1}{r^3}\right) \ldots \quad (15)$$

where as $\psi_{sc} = \frac{1}{r} e^{ikr} f(\Omega)$ is the asymptotic scattering wave, Eq. (8). We explicitly inserted here the current density $\mathbf{j}_{in} = \frac{\hbar \mathbf{k}}{m}$ to the incoming plane wave ψ_{in}. Equating the two expressions gives the relation

$$I(\Omega) \propto \left(\frac{d\sigma}{d\Omega}\right) = |f(\Omega)|^2 \quad (16)$$

for the differential cross section. This expression relates the experimental quantity, the differential cross section, to the scattering amplitude, which characterizes the wave function at large distances from the target. It is the fundamental relation between scattering theory and scattering experiments.

3   Lippmann Schwinger Equation

Having established the basic concepts for the scattering problem, we turn now to the illustration of the physical ideas that underlie the scattering analysis using integral equation methods. We recall that we are looking for the solution of the stationary Schrödinger equation

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right] \psi_{\mathbf{k}}(\mathbf{r}) = E \psi_{\mathbf{k}}(\mathbf{r}) \qquad \text{with} \qquad V(\mathbf{r}) = 0 \qquad \text{except } \mathbf{r} \in \text{ target region } \mathbb{T},$$
$$\tag{17}$$

that is consistent with the boundary condition (8) of an incident plane wave $\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}}$ and an emanating scattered wave. The energy E is determined by the energy of the incident plane wave $E_{\mathbf{k}} = \frac{\hbar^2}{2m} k^2$. By introducing the Green function $G_\circ$,

$$\left[\frac{\hbar^2}{2m} \nabla^2 + E \right] G_\circ(\mathbf{r}, \mathbf{r}'|E) = \delta(\mathbf{r} - \mathbf{r}'), \quad (18)$$

for the potential-free Schrödinger equation, the Schrödinger equation for $\psi_{\mathbf{k}}(\mathbf{r})$,

$$\left[\frac{\hbar^2}{2m} \nabla^2 + E \right] \psi_{\mathbf{k}}(\mathbf{r}) = V(\mathbf{r}) \, \psi_{\mathbf{k}}(\mathbf{r}), \quad (19)$$

can be transformed into an integral equation

$$\psi_{\mathbf{k}}'(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r}) + \int_{\mathbb{T}} d^3r' G_\circ(\mathbf{r}, \mathbf{r}'|E) \, V(\mathbf{r}') \, \psi_{\mathbf{k}}'(\mathbf{r}'), \quad (20)$$

in which the formal expression $V(\mathbf{r}) \psi_{\mathbf{k}}'(\mathbf{r})$ is conceived as inhomogeneity of the differential equation (18). This integral equation is called the *Lippmann-Schwinger equation*. Hereby, $\psi_{\mathbf{k}}(\mathbf{r})$ is the above cited plane-wave solution of the potential-free Schrödinger equation. The index $\mathbf{k}$ in ψ' expresses the fact that this state has evolved from one that in the remote past was a plane wave of the particular wavevector $\mathbf{k}$. Obviously, in the limit of zero potential, $V(\mathbf{r}) \to 0$, the scattered and the incident wave are identical, $\psi_{\mathbf{k}}'(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r})$.

The Green function $G_\circ(\mathbf{r}, \mathbf{r}'|E)$ is not uniquely determined by the Schrödinger equation (18). Also here the unique solution requires a boundary condition, which is chosen such, that the solution $\psi'_{\mathbf{k}}(\mathbf{r})$ describes outgoing scattered waves. The Green function $G_\circ(\mathbf{r}, \mathbf{r}'|E)$,

$$G_\circ(\mathbf{r}, \mathbf{r}'|E) = -\frac{2m}{\hbar^2} \frac{1}{4\pi} \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} \qquad \text{with} \qquad k = \sqrt{\frac{2m}{\hbar^2} E}, \tag{21}$$

describes then the stationary radiation of a particle of energy E, that is generated at $\mathbf{r}'$, by a spherical wave outgoing from the target. In other words, the Green function $G_\circ(\mathbf{r}, \mathbf{r}'|E)$ gives the amplitude of this wave at location $\mathbf{r}$ due to its generation by the source at $\mathbf{r}'$, under the condition that the wave is not further scattered during its propagation from $\mathbf{r}'$ to $\mathbf{r}$. By the Lippmann-Schwinger equation, the incident wave $\psi_{\mathbf{k}}(\mathbf{r})$ is superimposed with spherical waves emitted from scattering at position $\mathbf{r}'$ in the target. The amplitude of these scattered waves is proportional to the interaction potential $V(\mathbf{r}')$ and the amplitude of the total wave field $\psi'(\mathbf{r}')$ at that point.

Recalling our experimental set-up that the distance between target and detector is significantly larger than the size of the sample, for large distances between $\mathbf{r}$ and the scattering center $\mathbf{r}'$ it is useful to expand the Green function $G_\circ$ in powers of $\frac{r'}{r} \ll 1$ assuming that the extent of $\mathbf{r}'$ is restricted to the space of a small target or scattering volume, respectively, $\mathbf{r}' \in \mathbb{T}$. Approximating for $r' \ll r$

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \frac{1}{r} + \mathcal{O}\left(\frac{1}{r^2}\right) \qquad \text{and} \qquad |\mathbf{r}-\mathbf{r}'| \approx r - \hat{\mathbf{r}} \cdot \mathbf{r}' \qquad \text{with} \qquad \hat{\mathbf{r}} = \frac{\mathbf{r}}{r} \tag{22}$$

and inserting this into the relation (21) one obtains the asymptotic form, or far-field limit, respectively, of the Green function $G_\circ$,

$$G_\circ(\mathbf{r}, \mathbf{r}'|E) = -\frac{2m}{\hbar^2} \frac{1}{4\pi} \frac{e^{ikr}}{r} e^{-ik\hat{\mathbf{r}} \cdot \mathbf{r}'} + \mathcal{O}\left(\frac{1}{r^2}\right). \tag{23}$$

Inserting this expression into the Lippmann-Schwinger equation (20) one obtains the asymptotic solution of the wave function $\psi'_{\mathbf{k}}(\mathbf{r})$ for large distances $\mathbf{r}$

$$\psi'_{\mathbf{k}}(\mathbf{r}) \simeq e^{i\mathbf{k}\mathbf{r}} + \frac{1}{r} e^{ikr} f_{\mathbf{k}}(\hat{\mathbf{r}}), \tag{24}$$
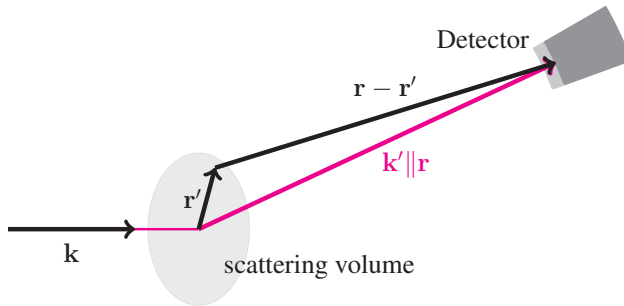

Fig. 3: Scattering geometry for the calculation of the far-field limit at the detector. In the Fraunhofer approximation, we assume that $|\mathbf{r}| \gg |\mathbf{r}'|$.

which is exactly the boundary condition (8) we conjectured from Huygens' principle, whereas the scattering amplitude $f(\hat{\mathbf{r}}) = f(\theta, \phi)$ is given by the integral,

$$f_{\mathbf{k}}(\hat{\mathbf{r}}) = -\frac{2m}{\hbar^2}\frac{1}{4\pi}\int d^3r'\, e^{-i\mathbf{k}'\mathbf{r}'}V(\mathbf{r}')\,\psi'_{\mathbf{k}}(\mathbf{r}') = -4\pi\frac{\hbar^2}{2m}\,T(\mathbf{k}',\mathbf{k}) \tag{25}$$

that can be interpreted as a transition-matrix element from the scattering state described by $\psi'_{\mathbf{k}}(\mathbf{r}')$ to the scattered state at far distances, which is a plane-wave state described by $\mathbf{k}' = k \cdot \hat{\mathbf{r}}$, the wave vector of the scattered wave in the direction of the detector, which is known in the experiment. $T(\mathbf{k}', \mathbf{k})$ is referred to as the T matrix or transition amplitude, a quantity proportional to the scattering amplitude. Due to the far-field approximation (22) the scattering pattern $f_{\mathbf{k}}(\hat{\mathbf{r}})$ is independent of the distance between target and detector, depending only on the angles to the detector from the target. In optics this is known as the Fraunhofer diffraction and in this context approximation (23) is also referred to as the Fraunhofer approximation of the Green function.

4 Born Approximation

Note that in the Lippmann-Schwinger equation (20), the wave function $\psi'(\mathbf{k})$ appears both on the left and right hand side. In a general case, there is no simple way to find exact solutions of the Lippmann-Schwinger equation. The form of the Lippmann-Schwinger equation provides a natural but approximate means that can be used for any potential, under the proper conditions, to proceed by an iterative procedure. At zeroth order in V, the scattering wave function is specified by the unperturbed incident plane wave,

$$\psi'^{(0)}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}}. \tag{26}$$

Then one can iterate the Lippmann-Schwinger equation (20) according to the rule

$$\psi'^{(n+1)}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} + \int d^3r'\, G_\circ(\mathbf{r},\mathbf{r}'|E)\,V(\mathbf{r}')\,\psi'^{(n)}_{\mathbf{k}}(\mathbf{r}') \tag{27}$$

that results in the *Born expansion* or the *Born series*, respectively, of the wave function in powers of the interaction potential V written here in a symbolic form[3]

$$\begin{aligned}
\psi'_{\mathbf{k}} &= \psi'^{(0)}_{\mathbf{k}} + \psi'^{(1)}_{\mathbf{k}} + \psi'^{(2)}_{\mathbf{k}} + \psi'^{(3)}_{\mathbf{k}} + \cdots + \tag{28}\\
&= \psi'^{(0)}_{\mathbf{k}} + G_\circ V\psi'^{(0)}_{\mathbf{k}} + G_\circ V G_\circ V\psi'^{(0)}_{\mathbf{k}} + G_\circ V G_\circ V G_\circ V\psi'^{(0)}_{\mathbf{k}} + \cdots + \tag{29}\\
&= (1 + G_\circ T)\,\psi'^{(0)}_{\mathbf{k}} \quad \text{with} \quad T = V + V G_\circ V + \cdots + = V\frac{1}{1 - G_\circ V} \tag{30}
\end{aligned}$$
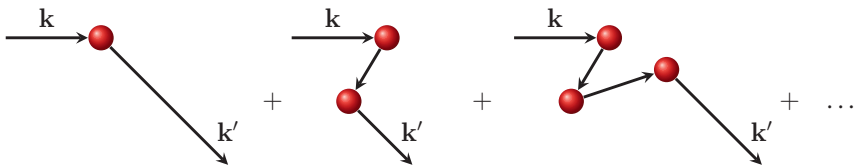
---
[3] Please note that $\psi^{(n)}_{\mathbf{k}} = (G_\circ V)^n \psi'^{(0)}_{\mathbf{k}}$. This is different from definition $\psi'^{(n)}_{\mathbf{k}}(\mathbf{r})$ in (27).

A term-by-term convergence of this series is in general not guaranteed and depends on the potential and the energy of the incident particle, even though the final expression is always

valid. Physically, an incoming particle undergoes a sequence of *multiple scattering* events from the potential. The first term in the series expansion (29) describes single scattering processes of the incident wave, while the following terms describe then scattering processes of successively higher order. Rarely are higher-order terms calculated analytically, since the complications then become so great that one might as well use a numerical method to obtain the exact solution if this is possible at all. Thus, only the first iteration of the series is taken into account, i.e. only single scattering, and the T matrix is approximated by the potential matrix $V(\mathbf{k'}, \mathbf{k})$,

$$T(\mathbf{k'}, \mathbf{k}) \simeq V(\mathbf{k'}, \mathbf{k}). \tag{31}$$

This first order term, in which the exact wave function $\psi'_{\mathbf{k}}(\mathbf{r'})$ in the integral kernel is replaced by the plane wave $e^{i\mathbf{k}\mathbf{r'}}$ is the *first Born approximation* and typically abbreviated as the *Born approximation*.[4] This approximation is most useful when calculating the scattering amplitude. In first Born approximation the general equation for the scattering amplitude (25) reads then

$$f^{(1)}_{\mathbf{k}}(\hat{\mathbf{r}}) = -\frac{2m}{\hbar^2}\frac{1}{4\pi}\int d^3r'\, e^{-i\mathbf{k'}\mathbf{r'}} V(\mathbf{r'})\, e^{i\mathbf{k}\mathbf{r'}} = -\frac{2m}{\hbar^2}\frac{1}{4\pi} V(\mathbf{Q}) \quad \text{with} \quad \mathbf{Q} = \mathbf{k} - \mathbf{k'}, \tag{32}$$

with $V(\mathbf{Q})$ denoting the Fourier transform of the potential with the momentum transfer $\mathbf{Q}$.[5] $V(\mathbf{Q})$ can be interpreted as a transition-matrix describing the transition from the incoming plane-wave of state $\mathbf{k}$ into the outgoing plane-wave state $\mathbf{k'}$ due to the action of the potential expressed in the reciprocal space at scattering angle $\mathbf{Q}$. From (16) follows then the differential cross section

$$\left(\frac{d\sigma}{d\Omega}\right)^{(1)} = \left(\frac{2m}{\hbar^2}\right)^2 \frac{\pi}{2} |V(\mathbf{Q})|^2. \tag{33}$$

The physics behind the 1st Born approximation is provided by the assumption that the incoming wave scatters only *once* inside the target potential before forming the scattered wave $\psi^{(1)}$. This is the concept behind the *kinematic theory of scattering*, that simplifies the interpretation of the scattering experiment substantially. For example, for the case of elastic scattering that we assumed all the time during the derivations, energy is conserved $|\mathbf{k}|^2 = |\mathbf{k'}|^2$, all possible scattering vectors are placed on the so-called *Ewald sphere* with radius $|\mathbf{k}|$. The length of the *scattering vector* $\mathbf{Q}$ is then given by

Fig. 4: The Ewald sphere.

$$Q(\Omega) = |\mathbf{Q}| = 2\,|\mathbf{k}|\,\sin\frac{1}{2}\Omega = \frac{4\pi}{\lambda}\sin\frac{1}{2}\Omega \qquad \text{with} \qquad \Omega = (\theta, \phi) \sphericalangle (\mathbf{k'}, \mathbf{k}). \tag{34}$$

Note, and this is the essence of the Ewald-sphere, that this shows that the differential cross section (33) does not depend on scattering angle and beam energy independently, but on a single parameter through the combination $Q = 2k\,\sin\frac{1}{2}\Omega$. By using a range of energies, $\mathbf{k}$, for

---

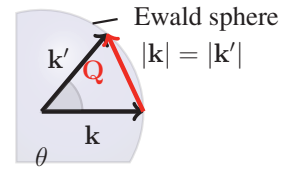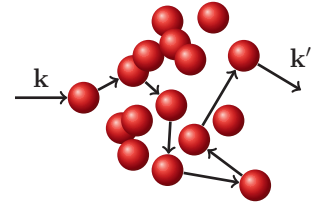[4] It should not be confused with the Born-Oppenheimer approximation.

[5] It would be mathematically more correct to denote the Fourier transformation $\mathcal{F}$ of $V(\mathbf{r})$ by a different function name e.g. $\widetilde{V}(\mathbf{Q}) = \mathcal{F}[V(r)]$. To avoid incomprehension of reading due to unduly complicated notation we replace $\widetilde{V}(\mathbf{Q})$ by $V(\mathbf{Q})$.

the incoming particles, this dependence can be used to test whether experimental data can be well described by the Born approximation. A very common use of the Born approximation is, of course, in reverse. Having found $\frac{d\sigma}{d\Omega}$, experimentally, a reverse Fourier transform can be used to obtain the form of the potential.

Example of Born Approximation: Central Potential

For a centrally symmetry potential, $V(\mathbf{r}) = V(r)$, we can make some progress with the matrix element integral (32) if we choose a polar coordinate system with $\mathbf{Q}$ along the z-axis, so that $\mathbf{Q} \cdot \mathbf{r} = Qr\cos\theta$. Then, the scattering amplitude in Born approximation $f^{(1)}(\theta)$ is written after some manipulations in the form

$$f_k^{(1)}(\theta) = -\frac{2m}{\hbar^2}\frac{1}{Q}\int_0^\infty V(r)\, r\, \sin Qr\, dr \tag{35}$$

and is seen to be independent of ϕ due to the cylindrical symmetry of the problem at hand and all scattering vectors are placed on an Ewald-circle. An example, is the *Rutherford* scattering or Coulomb scattering, respectively, where a charged particle with charge Z_1e impinges on an other charged particle with charge Z_2e under the action of a Coulomb potential, which results into the scattering amplitude

$$f_k^{(1)}(\theta) = -\frac{2m}{\hbar^2}\frac{Z_1 Z_2}{Q^2}e^2 = -\frac{1}{4}\frac{Z_1 Z_2}{\sin^2 \frac{1}{2}\theta}e^2\frac{1}{E} \tag{36}$$

and the differential cross section,

$$\left(\frac{d\sigma}{d\theta}\right)^{(1)} = \left|\frac{1}{4}\frac{Z_1 Z_2}{\sin^2 \frac{1}{2}\theta}e^2\right|^2\frac{1}{E^2} \tag{37}$$

known as the *Rutherford formula*. Due to the long-range nature of the Coulomb scattering potential, the boundary condition on the scattering wave function does not apply. We can, however, address the problem by working with the screened (Yukawa) potential, $V(r) = \frac{Z_1 Z_2}{r}e^{-\kappa r}$, leading to $f^{(1)} \propto \frac{1}{Q^2+\kappa^2}$ and taking $\kappa \to 0$, which leads then to the Rutherford formula (37). Accidentally, the first Born approximation gives the correct result of the differential cross section for the Coulomb potential.

Example of Born Approximation: Square Well Potential

Consider scattering of particles interacting via a spherical three dimensional (3D) square well potential $V(r) = V_\circ$ for $r \le R_\circ$ and zero outside ($V(r) = 0$ for $r > R_\circ$). The integral (35) for the scattering amplitude required here is then

$$f_k^{(1)}(\theta) = \frac{2m}{\hbar^2}\frac{1}{Q}\int_0^{R_\circ} V_\circ\, r\, \sin Qr\, dr = \frac{2m}{\hbar^2}\frac{1}{Q}V_\circ\left[\frac{\sin Qr - Qr\cos Qr}{Q^2}\right]_0^{R_\circ} \tag{38}$$

[6] and whence to the differential cross section

$$\left(\frac{d\sigma}{d\theta}\right)^{(1)} = \left(\frac{2m}{\hbar^2}\frac{V_\circ}{Q}\right)^2 R_\circ^2 \, j_1^2(QR_\circ) \simeq \left(\frac{2m}{\hbar^2}\frac{V_\circ}{Q}\right)^2 \begin{cases} \frac{1}{9}\left(1 - \frac{1}{5}Q^2 R_\circ^2\right) & \text{for low } E, \quad kR_\circ < 1 \\ \frac{R_\circ^2}{Q^2} & \text{for high } E, \quad kR_\circ > 1 \end{cases}.$$

(39)

From integrating over θ and ϕ the low and high energy limits for the total cross section are

$$\sigma(E \to \infty) = \pi\left(\frac{2m}{\hbar^2}\right)^2\left(\frac{V_\circ R_\circ^3}{kR_\circ}\right)^2 \qquad \sigma(E \to 0) = \sigma(E \to \infty)\frac{8}{9}\left(k^2 R_\circ^2 - \frac{2}{5}k^4 R_\circ^4 + \cdots\right).$$

(40)

The two examples illustrate some general features of scattering in the Born approximation:

(i) Born approximation is based on perturbation theory, so it works best for high energy particles.

(ii) At high energy, the scattering amplitude and the cross section are inversely proportional to the energy ($E = \hbar^2 k^2/2m$). E.g. both become smaller and the scattering weaker with increasing energy. This is a general phenomenon, if no bound states appear in the vicinity of the energy. This can be seen best by inspecting the Fourier transformed Green function $G_\circ(k|E) \propto 1/(E - \frac{\hbar^2 k^2}{2m})$ that is inverse proportional to the energy.

(iii) Scattering depends on square of the interaction potential, e.g. $V_\circ^2$, so both attractive and repulsive potentials behave the same.

(iv) The dependence on the energy of the incident beam k and scattering angle θ arises only through the combination $Q = 2k\sin\frac{\theta}{2}$. Thus as energy increases, the scattering angle θ is reduced and the scattered beam becomes more peaked in the forward direction.

(v) Angular dependence depends on the range of the potential $R_\circ$ but not on the strength $V_\circ$.

(vi) The total cross section depends on both range $R_\circ$ and depth $V_\circ$ of the potential.

Validity of first Born Approximation

This raises the practical questions (i) under which conditions the Born expansion converges and (ii) whether the first term is a good approximation. In the Born approximation the T matrix is approximated by the potential matrix V. This will not work if the denominator $|1 - VG_\circ|$ in (30) is small or zero. This is the situation at low energy, when the energy of the incoming beam coincides with bound states of the potential. Then, the Born approximation is invalid and the Born expansion will not converge. The solution to this problem is provided by the *dynamical scattering theory* discussed in section 6. According to (30) the Born approximation $T \simeq V$ is equivalent to the condition

$$\left|\iint_{\mathbb{T}} d^3r \, d^3r' \, V(\mathbf{r})G_\circ(\mathbf{r}, \mathbf{r}'|E)V(\mathbf{r}')\right| \ll 1.$$

(41)

At the same time this condition determines the radius of convergence of the Born series with respect to the strength of the potential. This condition means that the first Born approximation is valid and the Born series converges if the potential is sufficiently weak and the approximation improves as the energy is increased. Concerning the question whether the first term is itself a

---

[6] $j_0(Qr) = \sin Qr/Qr$ is the spherical Bessel function for angular momentum $\ell = 0$. Radial integration leads to Bessel function $j_1(Qr)$.

good approximation to the wave function, a convenient, although nonrigorous, criterion can be obtained by requiring that the first-order correction to the wave function be small compared to the incident wave in the region of the potential, i.e. $|\psi(1)\rangle_{\mathbf{k}}(\mathbf{r})| \ll |\psi_{\mathbf{k}}'^{(0)}(\mathbf{r})|$ which results to

$$\frac{2m}{\hbar^2}\frac{1}{4\pi}\left|\int_{\mathbb{T}} d^3r'\, \frac{e^{ikr'}}{r}V(\mathbf{r}')\,e^{i\mathbf{kr}'}\right| \ll 1. \tag{42}$$

For the above introduced spherical 3D square well potential $V(r \leq R_\circ) = V_\circ$ and $V(r > R_\circ) = 0$, this implies

$$\left|\frac{mV_\circ}{\hbar^2 k^2}\left(e^{ikR_\circ}\sin kR_\circ - kR_\circ\right)\right| \ll 1. \tag{43}$$

or

$$\frac{m}{\hbar^2}|V_\circ|R_\circ^2 \quad \ll \quad 1 \qquad \text{for low energies } kR_\circ < 1 \tag{44}$$

$$\frac{m}{\hbar^2}|V_\circ|R_\circ\frac{1}{k} \quad \ll \quad 1 \qquad \text{for high energies } kR_\circ > 1. \tag{45}$$

Since a bound state for this potential exists when $\frac{m}{\hbar^2}|V_\circ|R_\circ^2 \gtrsim 1$, as said above, the Born approximation will not be valid at low energies if the potential is so strong that it has a bound state. On the other hand criterion (45) can be satisfied for any potential by going to sufficiently high energy. When we square criterion (45) and multiply it by the geometrical cross section $\sigma_{\text{geo}} = \pi R_\circ^2$, criterion (45) reads

$$\pi\left(\frac{2m}{\hbar^2}\right)^2\left(\frac{V_\circ R_\circ^3}{kR_\circ}\right)^2 \ll \pi R_\circ^2 \qquad \Longleftrightarrow \qquad \sigma_{\text{tot}} \ll \sigma_{\text{geo}}. \tag{46}$$

and provides a hand-waving criterion when the potential is sufficiently weak so that the Born approximation gives reliable results: If the ratio of the scattering cross section and the geometrical extension of the potential is small, $u := \frac{\sigma_{\text{tot}}}{\sigma_{\text{geo}}} \ll 1$, the Born approximation can be used. For x-ray and neutron scattering, the scattering cross sections amount to a few 10^{-24} cm^2, the cross-sectional area per atom is of the order of several 10^{-16} cm^2. This results indeed in a very small potential strength of $u \sim 10^{-8} \div 10^{-7}$ for scattering on different atoms: that means, the Born approximation is justified and the easy-to-interpret kinematic interpretation of scattering results is sufficient. The arguments become invalid for the nuclear scattering of neutrons by individual nuclei as the cross-sectional area of a nucleus is eight orders of magnitude smaller and the scattering cross section and the geometrical gross section can be of similar size and the potential strength u can be even larger than 1, $u > 1$. Due to the strong Coulomb interaction potential, the probability for multiple scattering processes of electrons in solids is extremely high, making the interpretation of electron diffraction experiments very difficult. Although in neutron and x-ray scattering, the first Born approximation is almost always adequate, even for neutrons and x-rays, the kinematic scattering theory can break down, for example in the case of Bragg scattering from large nearly perfect single crystals. In this case as in the case of electron scattering the wave equation has to be solved exactly under the boundary conditions given by the crystal geometry. This is then called the dynamic scattering theory discussed in Section 6. For simple geometries, analytical solutions can be obtained. Other examples where the Born series do not converge are neutron optical phenomena like internal total reflection in a neutron guide, or grazing-incidence small-angle neutron scattering (GISANS). The same holds for x-ray scattering for example in combination with grazing-incidence small-angle x-ray scattering

(GISAXS) experiments. The *grazing-incidence small-angle scattering (GISAS)* techniques and their application will be discussed in Chapter D2. The theoretical analysis makes use of the *distorted-wave Born approximation (DWBA)*.

Distorted-Wave Born Approximation (DWBA)

In the previous Section we discussed that the Born approximation is accurate if the scattered field is small, compared to the incident field, in the scatterer. The scatterer is treated as a perturbation to free space or to a homogeneous medium, and the incident wave is a plane wave. When this smallness criteria is not met, it is often possible to generalize the idea of the Born approximation, which is frequently referred to as the distorted-wave Born approximation (DWBA). In generalization to the Born approximation, the free space zero-potential, $V_\circ(\mathbf{r}) = 0$, is replaced by a non-trivial reference potential $V_1(\mathbf{r})$ to which the scattered field $\psi'^1_{\mathbf{k}}$ is known analytically, numerically, e.g. due to the solution of the Lippmann-Schwinger equation (20),

$$\psi'^1_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} + \int_{\mathbb{T}} d^3r' G_\circ(\mathbf{r}, \mathbf{r}'|E)\, V_1(\mathbf{r}')\, \psi'^1_{\mathbf{k}}(\mathbf{r}'), \tag{47}$$

or experimentally. The interaction of interest V

$$V(\mathbf{r}) = V_1(\mathbf{r}) + \delta V(\mathbf{r}) \qquad \text{with} \qquad |\delta V| \ll |V_1| \tag{48}$$

is treated as a perturbation δV to the reference system V_1. In the distorted-wave Born approximation, the scattering field $\psi'_{\mathbf{k}}(\mathbf{r})$ due to the potential V is then determined applying the Born approximation

$$\psi'_{\mathbf{k}}(\mathbf{r}) = \psi'^1_{\mathbf{k}}(\mathbf{r}) + \int d^3r' G_1(\mathbf{r}, \mathbf{r}'|E)\, \delta V(\mathbf{r}')\, \psi'^1_{\mathbf{k}}(\mathbf{r}') \tag{49}$$

to the description of the scattering of the incident wave $\psi'^1_{\mathbf{k}}(\mathbf{r})$, the so-called "distorted" wave, due the perturbative potential $\delta V(\mathbf{r})$. The "distorted" incident wave, is the outgoing-wave solution of

$$\left[\frac{\hbar^2}{2m} \nabla^2 - V_1(\mathbf{r}) + E \right] \psi'^1_{\mathbf{k}}(\mathbf{r}) = 0, \tag{50}$$

that is supposed to be known, and $G_1(\mathbf{r}, \mathbf{r}'|E)$ is the corresponding Green function with the outgoing boundary condition for the same potential,

$$\left[\frac{\hbar^2}{2m} \nabla^2 - V_1(\mathbf{r}) + E \right] G_1(\mathbf{r}, \mathbf{r}'|E) = \delta(\mathbf{r} - \mathbf{r}'). \tag{51}$$

In analogy to the potential-free case (19), the difference to the reference system that appears in the Schrödinger equation, $\delta V(\mathbf{r})\psi'^1_{\mathbf{k}}(\mathbf{r})$, can be considered as inhomogeneity that constitutes a Lippmann-Schwinger equation with $\psi'^1_{\mathbf{k}}(\mathbf{r})$ as homogeneous solution. The Born approximation to this equation is given by Eq. (49).

To satisfy the boundary conditions we must also require that the "distorted" wave function behaves in the asymptotic limit as plane wave plus an outgoing wave

$$\psi'^1_{\mathbf{k}}(\mathbf{r}) \xrightarrow{r \to \infty} e^{i\mathbf{k}\mathbf{r}} + \frac{1}{r} e^{ikr} f^1_{\mathbf{k}}(\theta, \phi), \tag{52}$$

where, as in (25)

$$f_{\mathbf{k}}^1(\theta, \phi) = -\frac{2m}{\hbar^2} \frac{1}{4\pi} \int d^3 r'\, e^{-i\mathbf{k'r'}} V_1(\mathbf{r'})\, \psi_{\mathbf{k}}'^1(\mathbf{r'}). \tag{53}$$

This is simply the scattering amplitude for the potential $V_1(\mathbf{r})$, as if it were the only potential present, assumed to be known. The total scattering amplitude $f_{\mathbf{k}}(\theta, \phi)$ is

$$f_{\mathbf{k}}(\theta, \phi) = f_{\mathbf{k}}^1(\theta, \phi) + \delta f_{\mathbf{k}}(\theta, \phi) \tag{54}$$

where $\delta f_{\mathbf{k}}(\theta, \phi)$ is calculated in the Born approximation ($\psi_{\mathbf{k}}'(\mathbf{r}) \simeq \psi_{\mathbf{k}}'^1(\mathbf{r})$)

$$\delta f_{\mathbf{k}}(\theta, \phi) \simeq -\frac{2m}{\hbar^2} \frac{1}{4\pi} \int d^3 r'\, \psi_{\mathbf{k'}}'^{1(-)*}(\mathbf{r'}) \delta V(\mathbf{r'})\, \psi_{\mathbf{k}}'^1(\mathbf{r'}). \tag{55}$$

The scattering amplitude describes the scattering strength of an outgoing spherical wave. By inspection of Eq. (53) one finds that the first wave function of the integrand is a plane wave $e^{-i\mathbf{k'r'}}$, whose negative sign in the exponent represents an incoming plane wave. According of the standard definition of plane waves we can write $e^{-i\mathbf{k'r'}} = \psi_{\mathbf{k}}^{(-)*}$, where $(-)$ denotes the incoming boundary condition. Quite in the same way $\psi_{\mathbf{k'}}'^{1(-)*}(\mathbf{r'})$ is the known incoming wave function corresponding to the reference potential V_1.

Clearly Eq. (55) will be a good approximation if $\delta V(\mathbf{r})$ is sufficiently small, so that the additional scattering that is generated does not significantly modify the wave function. Some example in which this method is useful include scattering in which $\delta V(\mathbf{r})$ may be the spin-orbit interaction or a perturbation due to many-particle excitations, atomic scattering where $\delta V(\mathbf{r})$ may be a deviation from the Coulomb potential or from a Hartree average potential, or in case of scattering at a magnetic superlattice where $V_1(\mathbf{r})$ contains the scattering at the nuclei or electron charge distribution plus the interaction to an average magnetization, and $\delta V(\mathbf{r})$ describes the interaction to the modulated magnetic structure of the superlattice. One can think for example of a long-wave single-q or multiple-q spin-spiral structure, which is a modulation of the (collinear) ferromagnetic state, or a surface with roughness and the smooth surface as reference. The analysis of grazing-incidence small-angle scattering (GISAS) experiments to resolve the magnetic structure of superlattices [16] is an example where the DWBA is at place .

5   Scattering from a Collection of Scatterers

Finally, after having considered so far only the scattering at a single site with the target potential $V(\mathbf{r})$ placed at position $\mathbf{R}_\tau = \mathbf{0}$, prior to closing this chapter we shall relate these derivations to the scattering phenomena in solid state systems with the potential composed of an assembly of targets

$$V(\mathbf{r}) = \sum_\tau v_\tau(\mathbf{r} - \mathbf{R}_\tau) \tag{56}$$

centered at a collection of sites $\mathbf{R}_\tau$. Inserting this into the Lippmann-Schwinger equation (20), replacing the integration variable $\mathbf{r'}$ by a vector $\mathbf{r'}_\tau \in \mathbb{T}_\tau$ within the target τ and the center-of-gravity-vector $\mathbf{R}_\tau$, $\mathbf{r'}_\tau + \mathbf{R}_\tau$, and taking into consideration that the free-space Green function $G_\circ(\mathbf{r}, \mathbf{r'}|E)$ depends only on $\mathbf{r} - \mathbf{r'}$, the Lippmann-Schwinger equation for many potentials can be written as

$$\psi_{\mathbf{k}}'(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r}) + \sum_\tau \int_{\mathbb{T}_\tau} d^3 r'_\tau\, G_\circ(\mathbf{r} - \mathbf{R}_\tau, \mathbf{r'}_\tau|E)\, V(\mathbf{r'}_\tau)\, \psi_{\mathbf{k}}'(\mathbf{r'}_\tau + \mathbf{R}_\tau). \tag{57}$$

Approximating the Green function by its far-field asymptotic form (23) and considering that in the far-field solution $\psi'_{\mathbf{k}}(\mathbf{r}'_\tau + \mathbf{R}_\tau) \simeq e^{i\mathbf{k}\mathbf{R}_\tau}\psi'_{\mathbf{k}}(\mathbf{r}'_\tau)$, behind which is the Huygens' principle where the wave function generated at different sites share a phase relation, which becomes exact in the limit of the Born approximation, one obtains the asymptotic solution of the wave function $\psi'_{\mathbf{k}}(\mathbf{r})$ for large distances $\mathbf{r}$

$$\psi'_{\mathbf{k}}(\mathbf{r}) \simeq e^{i\mathbf{k}\mathbf{r}} + \sum_\tau \frac{1}{|\mathbf{r} - \mathbf{R}_\tau|} e^{ik|\mathbf{r}-\mathbf{R}_\tau|} e^{i\mathbf{k}\mathbf{R}_\tau} f_{\mathbf{k}}(\widehat{\mathbf{r} - \mathbf{R}_\tau}) \tag{58}$$

Since in the far-field approximation (22), $r \gg R_\tau$, $k|\mathbf{r} - \mathbf{R}_\tau| \simeq k(r - \hat{\mathbf{r}} \cdot \mathbf{R}_\tau) = kr - \mathbf{k}' \cdot \mathbf{R}_\tau$, we shall rewrite the asymptotic solution of the wave function for scattering at many potentials as

$$\psi'_{\mathbf{k}}(\mathbf{r}) \simeq e^{i\mathbf{k}\mathbf{r}} + \frac{1}{r}e^{ikr}F(\mathbf{Q}) \qquad \text{with} \qquad F(\mathbf{Q}) = \sum_\tau P_\tau(\mathbf{Q})\, f_{\tau\mathbf{k}}(\hat{\mathbf{r}}_\tau) \qquad \text{and} \qquad \mathbf{Q} = \mathbf{k}' - \mathbf{k}\,. \tag{59}$$

where the *static structure factor* (or *structure factor* for short) $F(\mathbf{Q})$ describes the way in which an incident beam is scattered by the atoms of a solid state system, taking into account the different scattering power of the elements through the term $f_{\tau\mathbf{k}}(\hat{\mathbf{r}}_\tau)$ also called *atomic form factor*

$$f_{\tau\mathbf{k}}(\hat{\mathbf{r}}_\tau) = -\frac{2m}{\hbar^2}\frac{1}{4\pi}\int_{\mathbb{T}_\tau} d^3 r'_\tau\, e^{-i\mathbf{k}'\mathbf{r}'_\tau} V(\mathbf{r}'_\tau)\, \psi'_{\mathbf{k}}(\mathbf{r}'_\tau) \qquad \text{with} \qquad \mathbf{r}_\tau = \mathbf{r} - \mathbf{R}_\tau \tag{60}$$

that depends only on the potential and not on the position of the atom. The atomic form factor, or scattering power, of an element depends on the type of radiation considered. Since the atoms are spatially distributed, there will be a difference in phase when considering the scattered amplitude from two atoms. This phase difference is taken into account by the phase factor

$$P_\tau(\mathbf{Q}) = e^{i\mathbf{Q}\cdot\mathbf{R}_\tau}\,, \tag{61}$$

which depends only on the position of the atoms and is completely independent of the scattering potential. So in total the structure factor separates the interference effects from the scattering within an target from the interference effects arising from scattering from different targets. Thus, the scattering intensity and the differential cross section are proportional to the square of the structure factor

$$I(\mathbf{Q}) \propto \left(\frac{d\sigma}{d\Omega}\right) = |F(\mathbf{Q})|^2 = \left|\sum_\tau P_\tau(\mathbf{Q})\, f_{\tau\mathbf{k}}(\hat{\mathbf{r}}_\tau)\right|^2\,. \tag{62}$$

If we ignore spin degrees of freedom, so that we do not have to worry whether an electron does or does not flip its spin during the scattering process, then at low energies the scattering amplitude $f(\theta)$ of particles from a cluster of atoms or a crystal becomes independent of angle (s-wave) and maybe described by the scattering length b_τ for atom τ, i.e. $f_{\tau\mathbf{k}}(\hat{\mathbf{r}}_\tau) = b_\tau$. Then, the differential cross section simplifies to

$$\left(\frac{d\sigma}{d\Omega}\right) = \left|\sum_\tau P_\tau(\mathbf{Q})\, b_\tau\right|^2\,. \tag{63}$$

If we consider scattering from a periodic crystal lattice, all atoms are same, i.e. have the same nuclear number (and we consider here also all nuclei as identical), thus $b_\tau = b$ for all atoms τ. Then, we are left with the differential cross section

$$\left(\frac{d\sigma}{d\Omega}\right) = N|b|^2 S(\mathbf{Q}) \qquad \text{with} \qquad S(\mathbf{Q}) = \frac{1}{N}\left|\sum_\tau e^{i\mathbf{Q}\mathbf{R}_\tau}\right|^2. \qquad (64)$$

$S(\mathbf{Q})$ is the form factor of the lattice, a quantity closely related to the pair density of a solid. For simplicity we consider scattering at a finite Bravais lattice. The lattice points are spanned by the lattice vector $\mathbf{R}_\tau = \mathbf{R_m} = \mathbf{A}\mathbf{m}$, where $\mathbf{A}$ is the Bravais matrix consistent with the primitive vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ of the three dimensional lattice, and $\mathbf{m} = (m_1, m_2, m_3) \in \mathbb{N}^3$ with $0 \le m_i \le (N_i - 1)$. The reciprocal lattice is defined by the matrix $\mathbf{B}$, that is orthogonal to $\mathbf{A}$, $\mathbf{B}^T\mathbf{A} = 2\pi\mathbf{1}$.
The reciprocal lattice vectors are given as $\mathbf{G_h} = \mathbf{B}\mathbf{h}$ with $\mathbf{h} = (h_1, h_2, h_3) \in \mathbb{Z}^3$. The transfer of scattering momentum vector is expressed by $\mathbf{Q} = (\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3) = \mathbf{B}\boldsymbol{\kappa}$, with $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3) \in \mathbb{R}^3$. The scattering amplitude is calculated analytically summing up the geometrical series
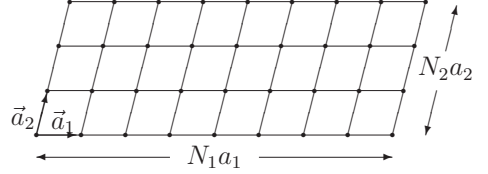
$$
\begin{aligned}
\sum_\mathbf{m} e^{i\mathbf{Q}\mathbf{R_m}} &= \sum_\mathbf{m} e^{i2\pi\boldsymbol{\kappa}\mathbf{m}} = \sum_{m_1=0}^{N_1-1}\sum_{m_2=0}^{N_2-1}\sum_{m_3=0}^{N_3-1} e^{i2\pi\kappa_1 m_1} e^{i2\pi\kappa_2 m_2} e^{i2\pi\kappa_3 m_3} \\
&= \frac{1 - e^{i2\pi\kappa_1 N_1}}{1 - e^{i2\pi\kappa_1}} \cdot \frac{1 - e^{i2\pi\kappa_2 N_2}}{1 - e^{i2\pi\kappa_2}} \cdot \frac{1 - e^{i2\pi\kappa_3 N_3}}{1 - e^{i2\pi\kappa_3}} \\
&= e^{i2\pi(\kappa_1(N_1-1)/2 + \kappa_2(N_2-1)/2 + \kappa_3(N_3-1)/2)} \cdot \frac{\sin N_1\pi\kappa_1}{\sin\pi\kappa_1} \cdot \frac{\sin N_2\pi\kappa_2}{\sin\pi\kappa_2} \cdot \frac{\sin N_3\pi\kappa_3}{\sin\pi\kappa_3}
\end{aligned}
$$

giving the scattered intensity

$$I(\mathbf{Q}) \propto \left(\frac{d\sigma}{d\Omega}\right) = |b|^2 \frac{\sin^2\frac{1}{2}N_1\mathbf{Q}_1\mathbf{a}_1}{\sin^2\frac{1}{2}\mathbf{Q}_1\mathbf{a}_1} \cdot \frac{\sin^2\frac{1}{2}N_2\mathbf{Q}_2\mathbf{a}_2}{\sin^2\frac{1}{2}\mathbf{Q}_2\mathbf{a}_2} \cdot \frac{\sin^2\frac{1}{2}N_3\mathbf{Q}_3\mathbf{a}_3}{\sin^2\frac{1}{2}\mathbf{Q}_3\mathbf{a}_3}. \qquad (65)$$

taking into account that $2\pi\kappa_i = \mathbf{Q}_i\mathbf{a}_i$, for $i = 1, 2, 3$. The dependence of the scattering intensity on the scattering vector $\mathbf{Q}$ is given by the so-called *Laue function*, which separates according to the three Bravais vectors. One factor along one lattice direction a is plotted in Fig. 5.

The main peaks are the *Bragg reflections*. They occur at integer κ_i, $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3) \in \mathbb{Z}^3$, i.e. at reciprocal lattice vectors $\mathbf{Q} = \mathbf{G_n}$. At points of Bragg reflection the coherent interference of scattering waves of all atoms add up constructively so that the maximum intensity scales with the square of the number of periods N^2. This high intensity is the reason why the Born approximation can in general not be used to describe the scattering at Bragg peaks. At small deviations $\Delta\kappa$ from the exact Bragg condition the intensity drops fast,[7] so that the total intensity integrated over a small $\Delta\kappa$ region, $\Delta\kappa \simeq 1/N$, varies only $\simeq N$. The half width is given

---

[7] For $\kappa \ll 1$ it follows that $\frac{\sin^2 N\pi\kappa}{\sin^2 \pi\kappa} \approx \frac{\sin^2 N\pi\kappa}{\pi^2\kappa^2} \implies \int_{-\infty}^{\infty} d\kappa \frac{\sin^2 N\pi\kappa}{\sin^2 \pi\kappa} \approx \frac{N}{\pi}\int_{-\infty}^{\infty} dx\frac{\sin^2 x}{x^2} = N \implies \bar{\kappa} \simeq 1/N$.
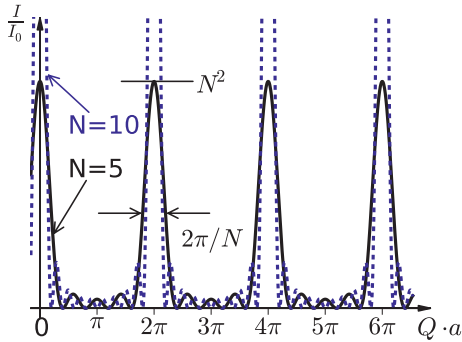
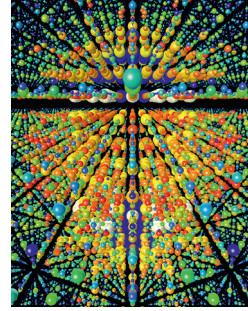**Fig. 5:** Laue function along the lattice direction a for a lattice with $N = 5$ and $N = 10$ periods.

Fig. 6: Three-dimensional rendering of x-ray diffraction data obtained from over 15 000 single nanocrystal diffraction snapshots of a protein complex (Permission of use granted by SLAC National Accelerator Laboratory, see also [17]).

approximately by $\Delta Q = \frac{2\pi}{a}\frac{1}{N}$. The more periods contribute to coherent scattering, the sharper and higher are the main peaks. Between the main peaks, there are $N - 2$ side maxima. With increasing number of periods N, their intensity becomes rapidly negligible compared to the intensity of the main peaks. From the position of these Bragg peaks in momentum space, the metric of the unit cell can be deduced (lattice constants a_1, a_2, a_3 in the three Bravais vector directions and unit cell angles α, β, γ). The width of the Bragg peaks is determined by the size of the coherently scattering volume $N = N_1 N_2 N_3$ and experimental factors (resolution) as well as details of the sample (size of crystallite, mosaic distribution, internal strains, etc.). For large N the form factor approaches a δ-function

$$\left(\frac{d\sigma}{d\Omega}\right) = N|b|^2 V_{\text{BZ}} \sum_{\mathbf{h}} \delta(\mathbf{Q} - \mathbf{G_h}), \tag{66}$$

where V_{BZ} is the Brillouin-zone volume.

6   Scattering of Electrons in Crystal

Scattering of Electronic Wave by a Periodic Perturbation

We start with a simple picture of a wave of electrons, which is to be scattered, and which we can write as $e^{i(\omega_0 t - \mathbf{k}_0 \mathbf{r})}$, where $\mathbf{k}_0$ is the propagation vector and $\mathbf{r} = (x, y, z)$ is the coordinate in real space. The wave which will do the scattering in a crystal we represent in a very simple way as $e^{-i\mathbf{G}\mathbf{r}}$, where $\mathbf{G}$ is some reciprocal vector of our crystal lattice. For simplicity, we assume in this section that the periodicity of the crystal is obeyed only along one direction, which is perpendicular to the surface of the crystal, onto which the wave $e^{i(\omega_0 t - \mathbf{k}_0 \mathbf{r})}$ is incoming. Thus, $\mathbf{G} = (0, 0, G)$. This is a typical setup of the *Bragg diffraction* experiment, see Fig. 7. Generally speaking, the solution to this scattering experiment, that is, the decomposition of the
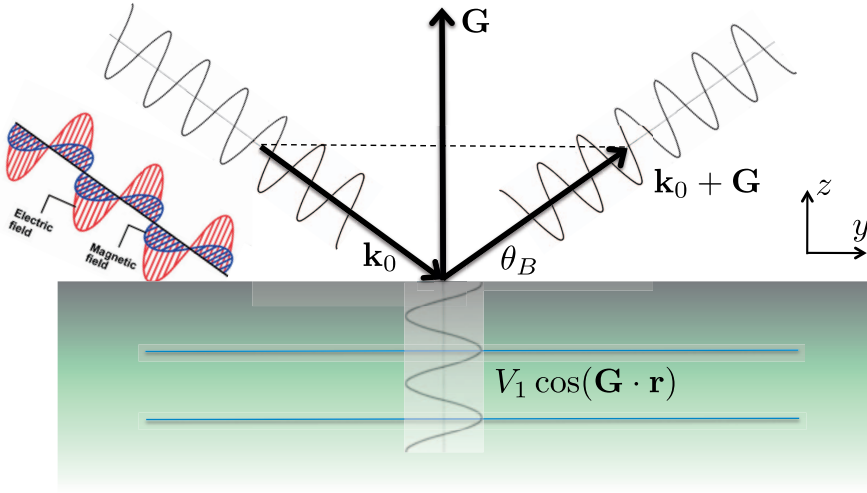
Fig. 7: A setup of a simple diffraction experiment. An incoming wave with wave vector $\mathbf{k}_0$ and energy χ_0 is scattered at a semi-infinite crystal, periodic along the z-direction. The wave on the left represents the case of x-ray radiation. In case of the Bragg scattering, the incoming wave is completely scattered into the reflected wave of the same energy and wave vector $\mathbf{k}_0 + \mathbf{G}$, where $\mathbf{G}$ is the reciprocal vector corresponding to the periodicity of the lattice.

total wavefunction in space into incident, scattered and the wavefunction inside the crystal can be obtained by finding a solution of the Schrödinger equation in $\mathbb{R}^3$, which satisfies certain boundary conditions. This solution must correspond to a constant energy $E = \chi_0$, meaning, that the wave inside the crystal, as well as incident and scattered waves have the same frequency.

Let us first find the solution to the Schrödinger equation (5) inside the crystal:

$$\frac{^2}{2m}\,^2\phi + V_1\cos(\mathbf{Gr})\phi = i\,\frac{\phi}{t} \tag{67}$$

Owing to the Bloch's theorem, valid inside periodic crystals, every solution of this equation can be characterized by a certain wave vector $\mathbf{k}$ from the Brillouin zone of the crystal (reciprocal space) [4], and the solution itself can be represented as:

$$\phi(\mathbf{r}\ t) = \sum_{n=}^{+} A_n e^{\,i[\,_0 t\ (\mathbf{k}+n\mathbf{G})\mathbf{r}]} \tag{68}$$

Substituting the Bloch wave into Eq. (5), we obtain a system of equations for the amplitudes A_n:

$$\frac{^2}{2m}(\mathbf{k}+n\mathbf{G})^2 \quad \chi_0 \quad A_n + \frac{V_1}{2}(A_{n\ 1} + A_{n+1}) = 0 \tag{69}$$

The system of equations above is completely identical to the system of equations for the coefficients of the wavefunction ϕ expanded in the complete and orthonormal basis of quantum states $\phi_n : \phi = \sum_n A_n \phi_n$. Then, if the components of the Hamiltonian matrix $H_{nm} = \gg_n H\ \phi_m \ll$
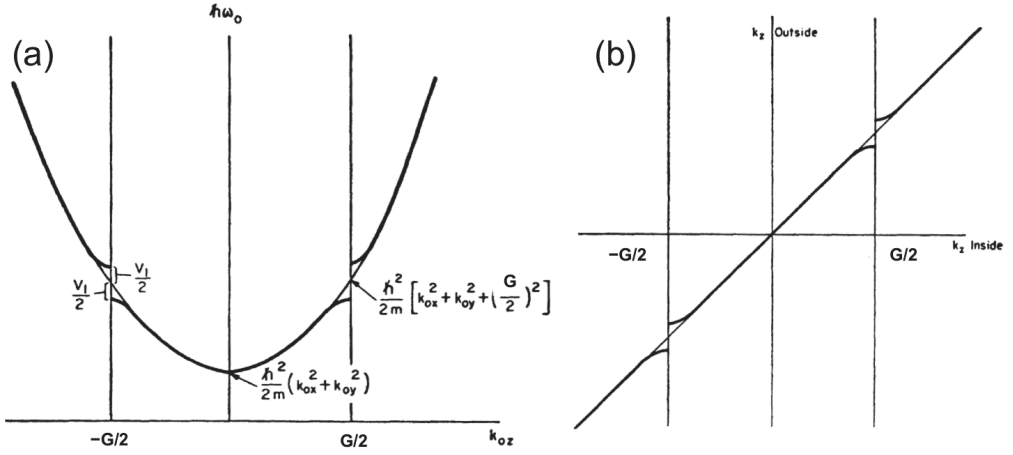
Fig. 8: (a) Energy $\hbar\omega_0$ as a function of z component of the **k**-vector for an electron wave inside the crystal. (b) The z component of the wave vector outside the crystal (k_{z0}) as a function of the z component of the wave vector inside the crystal (k_z). Taken from [12].

wavefunction ψ solves the Schrödinger equation with energy E when the following system of equations is satisfied [3]:

$$(H_{nn} - E)A_n + \sum_{m\neq n} H_{nm}A_m = 0 \tag{70}$$

As follows from Eq. (69), in our case, the H_{nn} matrix elements are given by $\frac{\hbar^2}{2m}(\mathbf{k}+n\mathbf{G})^2-\hbar\omega_0$, while the off-diagonal $H_{n,n\pm1}$ matrix elements assume the values of $V_1/2$.

In general, the system of equations Eqs. (69) and (70) has solutions only for certain values of E (or ω_0), if the Bloch vector **k** is given, which establishes the so-called dispersion relation $\omega_0(\mathbf{k})$ in a crystal. The wavefunction (68) with the vector **k** is the solution of the Schrödinger equation at energy E if the $\det(H_{nm} - E\delta_{nm}) = 0$. From this condition both $E(\mathbf{k}) = \hbar\omega_0(\mathbf{k})$ and the set of $\{A_n\}$ can be determined. The solution for $\{A_n\}$ for all n can be easily performed iteratively following Eq. (69) if, for example, A_0 and A_1 are pre-given. Effectively, the choice of A_0 and A_1 corresponds to the choice of the value of the wavefunction and its derivative when solving the second-order differential Schrödinger equation. It is known that generally, if arbitrary A_0 and A_1 are specified, the value of the wavefunction can be unbounded at infinity, however, upon an appropriate choice of ω_0 and A_1/A_0 ratio, the series (69) can be converged to provide a bounded wavefunction. Such frequencies ω_0 specify the energy spectrum of our system at **k**.

Usually, in the theory of the *x-ray diffraction* and *electron diffraction* one of the main approximations which is made is the smallness of the V_1 (or its equivalent for x-rays, see last section). This allows to use approximations which are equivalent to the perturbation theory expressions. In the drastic case of $V_1 = 0$ (free space) the solution is trivial: $A_0 = 1$, $A_n = 0, n > 1$, and $\hbar^2k^2/2m = \hbar\omega_0 = \hbar\mathbf{k_0}^2/2m$. When V_1 is sufficiently small we can derive approximate expression for the energy:

$$\hbar\omega_0 = \frac{\hbar^2\mathbf{k}^2}{2m} - \frac{V_1}{4}\left[\frac{1}{\frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{G})^2 - \hbar\omega_0} + \frac{1}{\frac{\hbar^2}{2m}(\mathbf{k}-\mathbf{G})^2 - \hbar\omega_0}\right] \tag{71}$$

and the only surviving coefficients in the first order with respect to V_1 are:

$$A_{\pm 1} = -\frac{V_1}{2} A_0 \frac{1}{\frac{\hbar^2}{2m}(\mathbf{k} \pm \mathbf{G})^2 - \hbar\omega_0} \tag{72}$$

While for most of the energies $A_0 \gg A_{\pm 1}$, in the particular case of the *Bragg scattering*, when $(\mathbf{k} \pm \mathbf{G})^2 = \mathbf{k_0}^2$, Eqs. (69) and (71) have no definite value. In this case, the degenerate perturbation theory should be used. Note that in our interpretation of the equations (69) the Bragg condition means the equality of the diagonal elements of the Hamiltonian matrix for $n = 0$ and $n = 1$. In this case, employing the degenerate perturbation theory we can show that at the point where the Bragg condition, $\mathbf{k} = \pm\mathbf{G}/2$, is met, exactly two energy solutions are possible:

$$\hbar\omega_0 = \frac{\hbar^2}{2m}\mathbf{k}^2 \pm \frac{V_1}{2}, \tag{73}$$

while there are two major participants in the Bloch wave: $A_0/A_1 = 1$. The solution $\hbar\omega_0$ as a function of $\mathbf{k}$ for the both considered above cases of the Bragg scattering and away from it, is shown in Fig. 8(a). We have to remark that without assuming that there are only two waves in the crystal in the vicinity of the Bragg point (as we shall see in the following, it is called the *two-beam approximation*), we can recover analogously the opening of the gaps in the spectrum at points $\pm n\mathbf{G}/2$, with the magnitude of the gap proportional to V_1^n. Thus, if V_1 is small, then the two-beam approximation is justified.

Lets try to interpret now the appearance of the gap in the energy spectrum of a perfect periodic crystal (through V_1 and $\mathbf{G}$) from the point of view of a simple diffraction experiment, depicted in Fig. 7. In this experiment a wave of electrons (x-rays etc.) of a certain energy $\hbar\omega_0$ and wavelength $\mathbf{k}_0$, is sent toward the (possibly even finite) film, while the intensities of the reflected (or even transmitted to the other side of the film) waves are measured. Theoretical treatment of this experiment lies in finding the solutions to the Schrödinger equation in the vacuum and in the film, which can be very complicated owing to the dynamical nature of scattering. Two necessary conditions have to be satisfied in order to uniquely solve the problem: (i) the solution of the Schödinger equation in each part of space is sought at an energy $\hbar\omega_0$, and (ii) the wavefunction and its normal space derivative have to be continuous at the boundaries of the crystal.

In our simple case, the boundary conditions, owing to the fact that we have periodicity in the crystal only along the z-axis, lead to an observation that the in-plane components of the wave vector have to be continuous: $k_x = k_{x0}, k_y = k_{y0}$, where $\mathbf{k}_0 = (k_{0x}, k_{0y}, k_{0z})$ and the wave vector of the wave inside the crystal $\mathbf{k} = (k_x, k_y, k_z)$. The z-components of the both wave vectors do not have to be the same however, and they are indeed different in the vicinity of $\pm\mathbf{G}/2$. This discrepancy is obvious looking at the energy dispersion in Fig. 8(a), in which the energy of the electrons inside the crystal as a function of k_z is given with the thick line, while the thin line stands for the energy of electrons as a function of k_{0z} in the vacuum. As we can see, in the vicinity of $\pm\mathbf{G}/2$ the constant energy line will cross the two energy profiles at two different k-points. The correspondence between the two k-vectors at a constant energy is given in Fig. 8(b).

If the energy of an incoming wave lies in the gap of the energy solutions in the crystal, the reflection of the incoming wave will be very large. In this case the waves inside the crystal will be decaying with the distance from the surface of the film, since the Bloch waves cannot be matched to the wave outside the crystal. When the energy is in the middle of the gap,
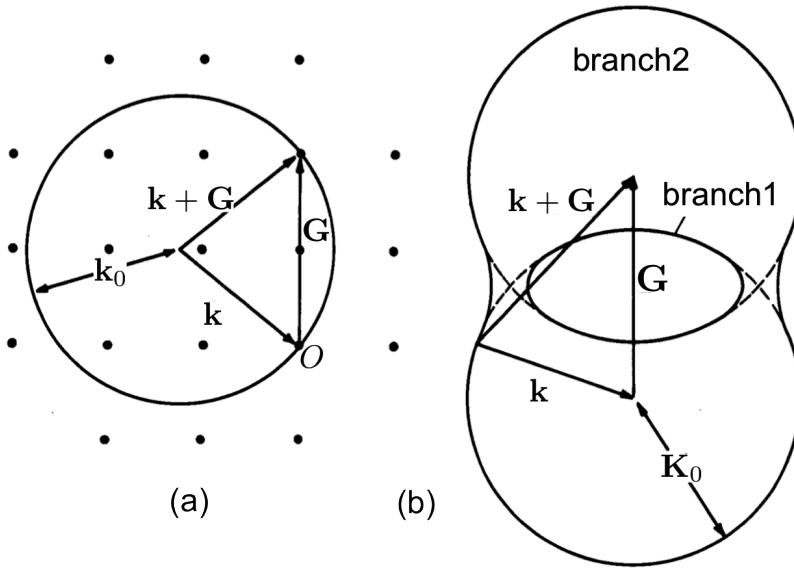
Fig. 9: (a) Ewald sphere around the Bloch vector **k** of the radius corresponding to an incoming wave $\mathbf{k}_0$. The origin of the sphere is sometimes called the Laue point. (b) Dispersion surface for the two-beam case. Taken from [14].

which occurs exactly at the Bragg angle of the incoming wave of θ_B, see Fig. 7, the reflection is complete. The range of the angles of the incident wave, θ, for which the transmission is unfavorable can be found easily:

$$\sin \theta = \left[\sin^2 \theta_B \pm \frac{V_1}{2\hbar\omega_0} \right]^{1/2}. \tag{74}$$

On the other hand, when the experiment operates away from the Bragg condition, the matching of the incoming wave to the Bloch wave inside the crystal can be perfectly done, and, as a result, the amplitude of the reflected beam is very small.

Dynamical Scattering of Electrons in Crystal

In this section we elaborate in more detail how the dynamical scattering can be studied in a way suggested in the previous section.

The motion of an electron in a solid is described by the Schrödinger equation Eq. (17). In an infinite periodic crystal the potential $V(\mathbf{r})$ is periodic with respect to translations by lattice vectors $\mathbf{R}$. The crystal potential can be split into contributions from different unit cells:

$$V(\mathbf{r}) = \sum_{\mathbf{R}} v(\mathbf{r} - \mathbf{R}) \tag{75}$$

where $v(\mathbf{r})$ is the potential of the Wigner-Seitz cell:

$$v(\mathbf{r}) = -\frac{Ze^2}{r} + \int_{V_C} d\mathbf{r}' \frac{e^2 \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \qquad \text{with} \qquad Z = \int_{V_C} d\mathbf{r}' \, \rho(\mathbf{r}'). \tag{76}$$

Here, the first term represents the attractive interaction with the nucleus of charge Z, and the second one is the repulsive interaction with the electron density ρ in the unit cell of the volume V_C. The crystal potential can be decomposed into Fourier series:

$$V(\mathbf{r}) = \sum_{\mathbf{G}} V_{\mathbf{G}} e^{i\mathbf{G}\mathbf{r}} \tag{77}$$

where the set $\{\mathbf{G}\}$ is the reciprocal lattice. The expression for the $\mathbf{G}$-components of V reads:

$$V_{\mathbf{G}} = \frac{1}{V_C} \int_{V_C} e^{-i\mathbf{G}\mathbf{r}} V(\mathbf{r}) \, d\mathbf{r} = \frac{1}{V_C} \int_{\mathbb{R}^3} e^{-i\mathbf{G}\mathbf{r}} v(\mathbf{r}) \, d\mathbf{r} \tag{78}$$

Using expression Eq. (76) we come to:

$$V_{\mathbf{G}} = -\frac{4\pi e^2}{V_C} \frac{Z - f_{\mathbf{G}}}{G^2} \tag{79}$$

with

$$f_{\mathbf{G}} = \int_{V_C} e^{-i\mathbf{G}\mathbf{r}} \rho(\mathbf{r}) \, d\mathbf{r} \tag{80}$$

The quantity $f_{\mathbf{G}}$ is known as *atomic scattering factor*, and it is normally smaller or equal than Z, ensuring that all the Fourier components of the potential are negative. For small $\mathbf{G}$'s the $V_{\mathbf{G}}$'s approach a constant, while for higher harmonics the interaction with the nucleus becomes increasingly important.

According to the Bloch's theorem valid in periodic crystals, the solution of the Schrödinger equation with a wave vector $\mathbf{k}$ can be represented as:

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} A_{\mathbf{G}}(\mathbf{k}) e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}, \tag{81}$$

where $u_{\mathbf{k}}(\mathbf{r})$ is a lattice-periodic function. The coefficients $A_{\mathbf{G}}(\mathbf{k})$ can be found by substituting the Bloch wave into the Schrödinger equation, leading thus to an infinite system of linear homogeneous equations:

$$\left\{ E_{\mathbf{k}} - \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{G})^2 \right\} A_{\mathbf{G}}(\mathbf{k}) = \sum_{\mathbf{G}'} V_{\mathbf{G}-\mathbf{G}'} A_{\mathbf{G}'}(\mathbf{k}) \tag{82}$$

By solving this system, for each $\mathbf{k}$ we can find a set of solutions $A_{\mathbf{G}}(\mathbf{k})$ which determine the Bloch wave, and the set of energies $E_{\mathbf{k}}$, with the latter one being commonly referred to as the *band structure*. This system has a solution only if the *dispersion equation* is satisfied:

$$\det \left[\left\{ E_{\mathbf{k}} - \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{G})^2 \right\} \delta_{\mathbf{G},\mathbf{G}'} - V_{\mathbf{G}-\mathbf{G}'} \right] = 0 \tag{83}$$

The effect of the *absorption* for electrons in crystals can be phenomenologically included by considering complex potential $V(\mathbf{r})$, complex $\mathbf{k}$-vectors and complex energies $E_{\mathbf{k}}$.
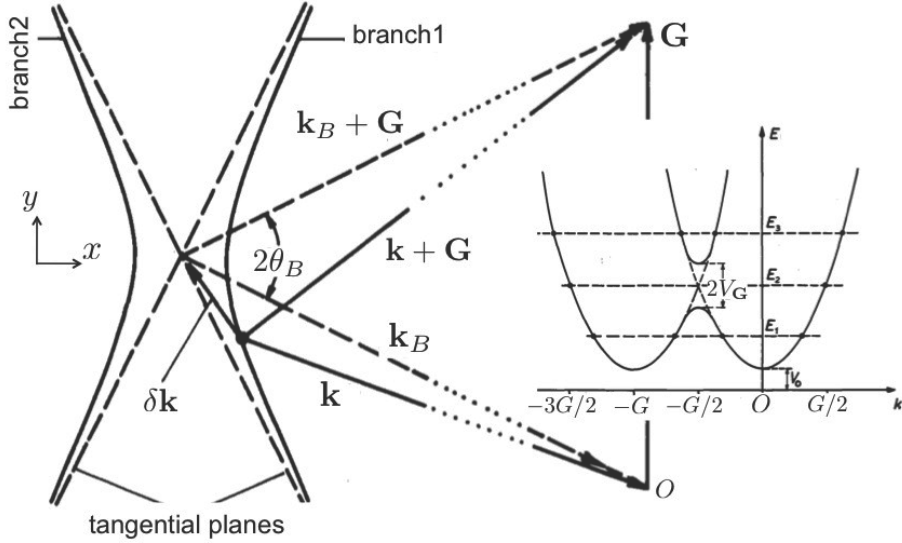
Fig. 10: Left: Dispersion surface near the Bragg spot k_B. Right: Bandstructure for the case on the right, for the direction of the Bloch vector along y. Taken from [14].

As already mentioned, for electron and x-ray diffraction the calculation of the Bloch waves and the band structures can be very much simplified since the energy is much larger than the potential, and, analogously to the previous section, we can apply the perturbation theory. Assuming first that $V(\mathbf{r}) = 0$, we get the following system of equations:

$$\left\{ \mathbf{k}_0^2 - (\mathbf{k} + \mathbf{G})^2 \right\} A_\mathbf{G} = 0, \tag{84}$$

where the energy $E_\mathbf{k}$ is given by an incoming wave with $\mathbf{k}_0$: $E_\mathbf{k} = \frac{\hbar^2}{2m} \mathbf{k}_0^2$. For given energy and given $\mathbf{k}$ the expression in brackets will in general not vanish for all $\mathbf{G}$ and there are no allowed waves. Nevertheless, for a certain $\mathbf{k}$ it may happen that $\mathbf{k}_0^2 = (\mathbf{k} + \mathbf{G})^2$ for one $\mathbf{G}$, for example $\mathbf{G} = 0$. Then $A_\mathbf{G} = \delta_{\mathbf{G},0}$ and the wave $e^{i\mathbf{kr}}$ is allowed. For $V \neq 0$ the system of equations (83) comes to:

$$\left\{ \mathbf{K}_0^2 - (\mathbf{k} + \mathbf{G})^2 \right\} A_\mathbf{G} = \sum_{\mathbf{G}' \neq \mathbf{G}} v_{\mathbf{G} - \mathbf{G}'} A_{\mathbf{G}'}, \tag{85}$$

with $v_\mathbf{G} = (2m/\hbar^2)V_\mathbf{G}$, and $\mathbf{K}_0^2 = \mathbf{k}_0^2 - v_0 = (2m/\hbar^2)(E_\mathbf{k} - V_0)$. It can be shown that for small $v_\mathbf{G}$ the secondary waves $A_\mathbf{G}$ for $\mathbf{G} \neq 0$ are small and we have only one strong beam $\psi = e^{i\mathbf{kr}}$ with a slightly renormalized $\mathbf{k}$-value.

However, if the condition $\mathbf{k}_0^2 \approx (\mathbf{k} + \mathbf{G})^2$ is fulfilled not only for the primary wave $\mathbf{G} = 0$ but also for other secondary waves $\mathbf{G} \neq 0$, then these waves may also become strong and the perturbation theory breaks down. This condition means that for certain $\mathbf{G}$ the vector $\mathbf{k} + \mathbf{G}$ lies near or at the so-called *Ewald sphere*, see Fig. 9. Because the energies $\frac{\hbar^2}{2m} \mathbf{k}^2$ and $\frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{G})^2$ are close in this case, the degenerate perturbation theory has to be applied, which takes into account on an equal basis all the excited waves. Here, we restrict ourselves to the so-called *two-beam case*, for which only two waves are prominent in the crystal: at $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$. This approximation allows to obtain main conclusions of the dynamical scattering theory in a simple

way, yet without loosing generality. We also used the two-beam approximation in the previous section. For two beams, the equations for the amplitudes of the waves read:

$$(\mathbf{K}_0^2 - \mathbf{k}^2)A_0 = v_{-\mathbf{G}}A_{\mathbf{G}}, \quad (\mathbf{K}_0^2 - (\mathbf{k}+\mathbf{G})^2)A_{\mathbf{G}} = v_{\mathbf{G}}A_0, \tag{86}$$

while the dispersion equation reads:

$$(\mathbf{K}_0^2 - \mathbf{k}^2) \cdot (\mathbf{K}_0^2 - (\mathbf{k}+\mathbf{G})^2) = v_{\mathbf{G}} \cdot v_{-\mathbf{G}}. \tag{87}$$

From the dispersion equation it is clear that in the absence of the crystal potential for a given by $\mathbf{K}_0^2$ energy the allowed $\mathbf{k}$-vectors lie on a *dispersion surface* consisting of two intersecting spheres centered around the origin, and $\mathbf{G}$, both with the radius $\mathbf{K}_0^2$, see Fig. 10. Non-zero $\mathbf{G}$-component of the potential, on the other hand, forces the intersection surface between the two spheres, at which the Bragg condition $\mathbf{k}^2 = (\mathbf{k}+\mathbf{G})^2$ is fulfilled, to split into two surfaces, of which the outer branch 2 completely surrounds the inner branch 1. While at the Bragg point $\mathbf{k} = \mathbf{k}_B$, by setting $\mathbf{k} = \mathbf{k}_B + \delta\mathbf{k}$ around the intersection, we find:

$$4(\mathbf{k}_B \cdot \delta\mathbf{k})((\mathbf{k}_B + \mathbf{G}) \cdot \delta\mathbf{k}) = v_{\mathbf{G}} \cdot v_{-\mathbf{G}}, \tag{88}$$

if we neglect the 3rd and 4th orders in $\delta\mathbf{k}$. For convenience, we decompose the $\delta\mathbf{k}$ into x and y components, see Fig. 10. Using the condition that $K_0 \sin\theta_B = G/2$ we find:

$$\delta k_x^2 - \delta k_y^2 \tan^2\theta_B = \frac{|v_{\mathbf{G}}|^2}{4K_0^2 \cos^2\theta_B}. \tag{89}$$

In this approximation, the dispersion surfaces are hyperbolas, the asymptotes of which are the tangential planes of the spheres, see Fig. 10. The smallest separation between the two branches is

$$\Delta k = \frac{|v_{\mathbf{G}}|}{K_0 \cos\theta_B} \tag{90}$$

The distance $d_{ext} = \frac{2\pi}{\Delta k}$, over which the two Bloch waves from the opposite branches get a phase difference 2π, is called the *extinction length*.

From the dispersion equation we get the energy as a function of $\mathbf{k}$ as follows:

$$\frac{2m}{\hbar^2}E(\mathbf{k}) - v_0 = K_0^2 = \frac{1}{2}(\mathbf{k}^2 + (\mathbf{k}+\mathbf{G})^2 \pm \frac{1}{2}\sqrt{(\mathbf{k}^2 - (\mathbf{k}+\mathbf{G})^2)^2 + 4|v_{\mathbf{G}}|}, \tag{91}$$

showing that for $\mathbf{k}^2 = (\mathbf{k}+\mathbf{G})^2$ there is a band gap of the width $\Delta E = 2V_{\mathbf{G}}$, as in the previous section. We have plotted the bandstructure according to the relation above along the direction of $\mathbf{G}$ in Fig. 10. When the energy E_1 is below the gap, we get four allowed k-values. For the energy in the gap E_2 we obtain only two allowed k-values, while we restore the situation for E_1 when we go above the band gap (E_3).

From the previous formulas also the expressions for the amplitudes of the waves in the crystal can be derived:

$$A_0 = \frac{1}{\sqrt{2}}\sqrt{1 \mp \frac{W}{\sqrt{1+W^2}}}, \quad A_{\mathbf{G}} = \pm\text{sign}(v_{\mathbf{G}})\frac{1}{\sqrt{2}}\sqrt{1 \pm \frac{W}{\sqrt{1+W^2}}}, \tag{92}$$
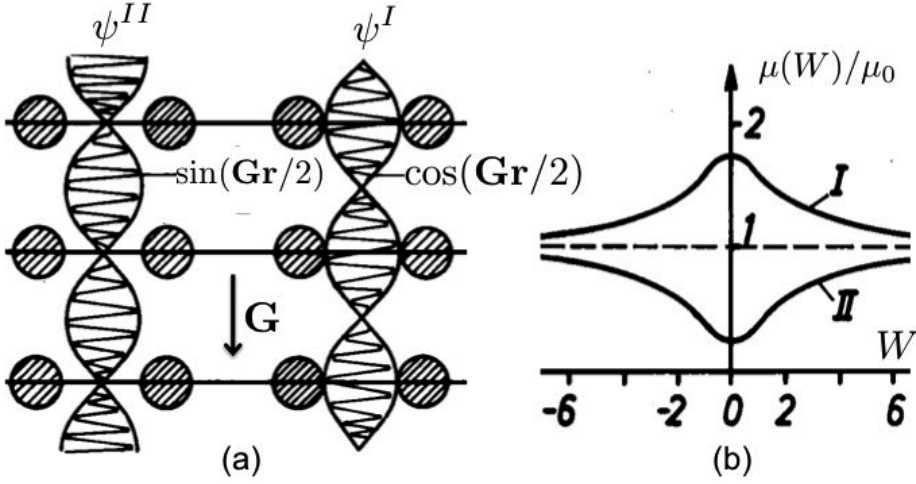
Fig. 11: (a) Modulation factors of ψ^I and ψ^{II}. (b) Absorption for Bloch wave ψ^I and ψ^{II} at $\mu_G = 3/4\mu_0$. Taken from [14].

where an important parameter W, which measures the deviation from the Bragg angle, θ_B (see Fig. 10), is given by:

$$W = \frac{\delta\mathbf{k} \cdot \mathbf{G}}{|v_\mathbf{G}|}. \tag{93}$$

The limiting case of $|W| \to \infty$ corresponds to the situation away from the Bragg spot.

For the case of exact Bragg condition $W = 0$ the two Bloch waves are:

$$\psi^I(\mathbf{r}) = \sqrt{2}e^{i(\mathbf{k}+\frac{\mathbf{G}}{2})\mathbf{r}} \cdot \cos\left(\frac{\mathbf{Gr}}{2}\right), \quad \psi^{II}(\mathbf{r}) = -i\sqrt{2}e^{i(\mathbf{k}+\frac{\mathbf{G}}{2})\mathbf{r}} \cdot \sin\left(\frac{\mathbf{Gr}}{2}\right) \tag{94}$$

In case of electron diffraction, the Fourier components of the potential are negative, and the ψ^I will be positioned on the outer branch, while ψ^{II} lies at the inner branch. This situation will be reversed for x-ray diffraction, in which case $v_\mathbf{G}$ is positive. Characteristic for both branches are the cosine and sine modulation functions. Owing to these modulation functions, ψ^I is always maximal at the atomic positions of the reflecting planes (situated along $\mathbf{G}$), while it is small in between the planes, see Fig. 11. The situation is reversed for ψ^{II}, which is maximal between the atomic planes, while it almost does not interact with atomic potential. Both waves, of course, have the same energy, however, the Bloch vector of ψ^I is larger than that of ψ^{II}, since the solutions are positioned on different branches. This manifests the larger kinetic energy of the first wave, which is compensated by larger but negative energy coming from interaction with the atoms.

For an absorbing crystal it is clear that the absorption will be very much different for ψ^I and ψ^{II}. The absorption of wave ψ^I will be higher than normal since an electron in the Bloch wave spends more time on the atoms, while wave ψ^{II} manages to avoid the atoms and has therefore a smaller than average absorption, resulting in a so-called *anomalous transmission* (or *anomalous absorption*).

7 X-ray Scattering

The theory of x-ray diffraction is quite analogous to the theory of electron diffraction with the exception that the wave fields are vector fields in nature (see Fig. 1). The interaction of x-rays with the medium arises via excitation of atomic electrons. The frequencies of motion of electrons are of the order of $\omega_0 \approx v_e/a_B$, where v_e is the velocity of electron motion around the atom, and a_B is Bohr's radius. The wavelength of x-rays is comparable to a_B while their frequencies are of the order of $2\pi c/\lambda$ with c as the speed of light, that is, much higher than those of the electrons. Thus, electrons can be treated as free. This simplification leads to the following formulation of the Maxwell's equations, which have to be solved instead of the Schrödinger equation for electrons (assuming harmonic time-dependence of the x-ray field with the frequency ω):

$$\nabla \times \mathbf{E} = i\frac{\omega}{c}\mathbf{H}, \quad \nabla \times \mathbf{H} = -i\frac{\omega}{c}\mathbf{D}, \tag{95}$$

where the electric displacement field $\mathbf{D} = \varepsilon\mathbf{E}$ and ε is the dielectric function of the medium:

$$\varepsilon(\mathbf{r}, \omega) = 1 - \frac{4\pi e^2}{m\omega^2}\rho(\mathbf{r}), \tag{96}$$

with $\rho(\mathbf{r})$ as the electronic density. Additionally,

$$\nabla \cdot \mathbf{D} = \nabla \cdot \mathbf{H} = 0, \quad \nabla \times \nabla \times \mathbf{E} = (\omega/c)^2\mathbf{D}. \tag{97}$$

Deviation of ε from unity is actually very small for x-rays and the quantity $\chi = \varepsilon - 1$ normally varies between 10^{-6} and 10^{-4}. Thus, we can write that

$$\mathbf{E} = \frac{1}{\varepsilon}\mathbf{D} \approx \mathbf{D} - \chi\mathbf{D}. \tag{98}$$

This equation is a great simplification, which, substituted into previous equations leads to an equation for $\mathbf{D}$ alone:

$$(\nabla^2 + k_0^2)\mathbf{D}(\mathbf{r}) = -\nabla \times \nabla \times [\chi(\mathbf{r})\mathbf{D}(\mathbf{r})] \tag{99}$$

In an infinite periodic crystal the charge density as well as the dielectric function are periodic and can be expanded in terms of the Fourier series in analogy to the potential for electrons Eq. (77). Analogously to electrons, the solution to Eq. (99) can be sought in the shape of a Bloch wave:

$$\mathbf{D}_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} \mathbf{D}_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}. \tag{100}$$

Since $\nabla \cdot \mathbf{D} = 0$ the vectors $\mathbf{D}_{\mathbf{G}}$ are orthogonal to $\mathbf{k} + \mathbf{G}$: $\mathbf{D}_{\mathbf{G}} \cdot (\mathbf{k} + \mathbf{G}) = 0$. It can be shown that the corresponding component $\mathbf{H}_{\mathbf{G}}$ is orthogonal both to $\mathbf{k} + \mathbf{G}$ and $\mathbf{D}_{\mathbf{G}}$, while $\mathbf{E}_{\mathbf{G}}$ lies in the plane of $\mathbf{D}_{\mathbf{G}}$ and $\mathbf{k} + \mathbf{G}$, but the difference between $\mathbf{E}_{\mathbf{G}}$ and $\mathbf{D}_{\mathbf{G}}$ is very small owing to the smallness of χ.

For each Fourier component we can introduce two polarization vectors $\mathbf{e}_{\mathbf{G}}^{\sigma} \cdot (\mathbf{k} + \mathbf{G}) = 0$ and $\mathbf{e}_{\mathbf{G}}^{\sigma} \cdot \mathbf{e}_{\mathbf{G}}^{\sigma'} = \delta_{\sigma,\sigma'}$, where σ and σ' are either "1" or "2". Thus, we can write that $\mathbf{D}_{\mathbf{G}} = \sum_{\sigma} D_{\mathbf{G}}^{\sigma}\mathbf{e}_{\mathbf{G}}^{\sigma}$. Since χ is very small, we can significantly simplify Eq. (99), arriving at the following system of equations for the scalar Fourier components of $\mathbf{D}$:

$$\left(k_0^2 - (\mathbf{k} + \mathbf{G})^2\right) D_{\mathbf{G}}^{\sigma} = \sum_{\mathbf{G}',\sigma'} \varkappa_{\mathbf{G}-\mathbf{G}'}\left(\mathbf{e}_{\mathbf{G}}^{\sigma} \cdot \mathbf{e}_{\mathbf{G}'}^{\sigma'}\right) D_{\mathbf{G}'}^{\sigma'}, \tag{101}$$
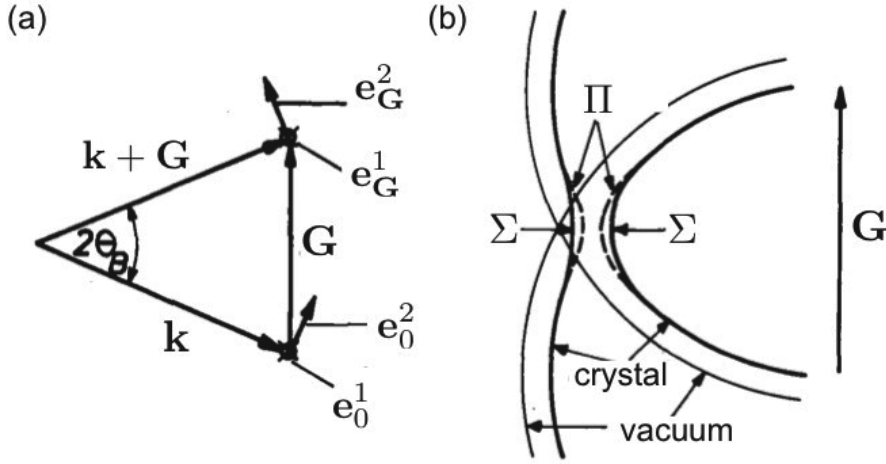
Fig. 12: (a) Direction of polarization vectors. Vectors $\mathbf{e}_0^1 = \mathbf{e}_\mathbf{G}^1$ are normal to the plane of the plot. (b) Dispersion surface for Σ and Π polarizations. Taken from [14].

where

$$\varkappa_{\mathbf{G}-\mathbf{G}'} = -k_0^2 \chi_{\mathbf{G}-\mathbf{G}'} = \frac{4\pi r_e}{V_C} f_{\mathbf{G}-\mathbf{G}'} \tag{102}$$

with $f_\mathbf{G}$ as the atomic scattering factor, Eq. (80), and r_e as the classical electron radius of 2.82×10^{-13} cm. Note that since $\mathbf{e}_\mathbf{G}^\sigma \cdot \mathbf{e}_{\mathbf{G}'}^{\sigma'}$ is in general non-zero because $\mathbf{G} \neq \mathbf{G}'$, the latter equations couple both polarizations. Also note complete similarity between Eq. (101) and Eqs. (82) and (85) derived previously for the case of electron scattering in a crystal. Thus, $\varkappa_{\mathbf{G}-\mathbf{G}'}$ is analogous to the corresponding component of the crystal potential for electrons. The eigenenergies, or, bands $\omega_\mathbf{k}$ of x-ray Bloch states in a crystal are determined from the secular equation which is very much the same as Eq. (83). Analogously to the case of electrons the effect of absorption lies in complexity of the k-vectors, eigenenergies and Fourier components of the crystal "potential" $\varkappa$.

For scattering of x-rays from a finite crystal, the boundary conditions have to be formulated. Owing to the nature of Maxwell's equations given in the beginning of the section, and smallness of χ, the boundary conditions are very simple: both normal and tangential components of $\mathbf{D}$ have to be continuous across the boundary between the crystal and the vacuum. In general one writes the fields in the vacuum before the slab (see Fig. 6) as:

$$\mathbf{D}(\mathbf{r}) = \hat{\mathbf{D}} e^{i\mathbf{k}_0 \mathbf{r}} + \sum_\mathbf{G} \mathbf{R}_\mathbf{G} e^{i\mathbf{K}_\mathbf{G}^- \mathbf{r}}, \quad z < 0, \tag{103}$$

and after the slab:

$$\mathbf{D}(\mathbf{r}) = \sum_\mathbf{G} \mathbf{T}_\mathbf{G} e^{i\mathbf{K}_\mathbf{G}^+ \mathbf{r}}, \quad z > d, \tag{104}$$

while in the crystal we seek the solution in terms of Bloch waves:

$$\mathbf{D}(\mathbf{r}) = \sum_j P_j \mathbf{D}_{\mathbf{k}_j}(\mathbf{r}), \tag{105}$$

where all the reflection ($\mathbf{R_G}$), transmission ($\mathbf{T_G}$) and P_j coefficients are found by satisfying the boundary conditions. It is noteworthy to remark, that, in analogy to electrons, the kinematic and dynamical theory of x-ray scattering can be also developed starting from the integral equation for the $\mathbf{D}$-field.

$$\mathbf{D}(\mathbf{r}) = \hat{\mathbf{D}}e^{i\mathbf{k}_0\mathbf{r}} + \nabla \times \nabla \times \int d\mathbf{r}' \frac{e^{i\mathbf{k}_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}\chi(\mathbf{r}')\mathbf{D}(\mathbf{r}'). \tag{106}$$

In analogy to electrons, if the scattering occurs away from Bragg condition, we have only one strong beam. If the Bragg reflection is excited, we have two strong beams $\mathbf{k}$ and $\mathbf{k}+\mathbf{G}$. In this case there is a natural choice for polarization vectors: (i) Σ-polarization for $\sigma = 1$: $\mathbf{e}_0^1 = \mathbf{e}_\mathbf{G}^1$ perpendicular to both $\mathbf{k}$ and $\mathbf{k}+\mathbf{G}$, and (ii) Π-polarization for $\sigma = 2$: $\mathbf{e}_0^2 = \mathbf{e}_\mathbf{G}^2$ in the plane of $\mathbf{k}$ and $\mathbf{k}+\mathbf{G}$, see Fig. 11. Via this particular choice of polarization vectors we can decouple the equations (101) for Σ and Π waves:

$$(k_0^2 - \varkappa_0 - k^2)D_0^\sigma = \varkappa_\mathbf{G}P_\sigma D_\mathbf{G}^\sigma, \tag{107}$$
$$(k_0^2 - \varkappa_0 - (\mathbf{k}+\mathbf{G})^2)D_\mathbf{G}^\sigma = \varkappa_\mathbf{G}P_\sigma D_0^\sigma, \tag{108}$$

where polarization factor $P_\sigma = \mathbf{e}_0^\sigma \cdot \mathbf{e}_\mathbf{G}^\sigma$ equals one for Σ-polarization, and $\cos 2\theta_B$ for Π-polarization. For each polarization the allowed k-vectors lie on the dispersion surface given by:

$$(k_0^2 - \varkappa_0 - k^2)(k_0^2 - \varkappa_0 - (\mathbf{k}+\mathbf{G})^2) = P_\sigma^2|\varkappa_\mathbf{G}|^2. \tag{109}$$

As we can see, far away from the Bragg condition the dispersion surfaces are spheres around $\mathbf{k} = 0$ and $\mathbf{k} = \mathbf{G}$ with the radius $\sqrt{k_0^2 - \varkappa_0}$ for both polarizations. Near the Bragg condition the degeneracy between both polarizations is lifted and the smallest separation between the outer and inner branches becomes polarization-dependent:

$$\Delta k_\sigma = \frac{2\pi}{d_{ext}} = \frac{P_\sigma \varkappa_\mathbf{G}}{k_0 \cos \theta_B}, \tag{110}$$

see Figure 12. In this figure the dispersion surface in the vacuum, degenerate for both polarizations, is shown with thin lines, while the dispersion surface in the crystal is shown with thicker lines. The expressions for the coefficients D_0^σ and $D_\mathbf{G}^\sigma$ are exactly the same as those in Eq. (92), with $v_\mathbf{G}$ replaced by $P_\sigma \varkappa_\mathbf{G}$. The resulting Σ-fields are identical to those given by the Bloch waves ψ^I and ψ^{II} from the previous section. However, for Π-polarization we do not get pure sine and cosine waves but only a combination of both, since $\mathbf{e}_0^2 \neq \mathbf{e}_\mathbf{G}^2$. Correspondingly, only Σ-waves show an anomalous transmission effect, while Π-waves never vanish at the atomic positions completely. Interestingly, for multi-beam cases, important for x-ray diffraction, the situation complicates significantly, since both polarizations cannot be anymore decoupled.

Final Remarks

I shall conclude this chapter by mentioning that in 1914 the Nobel Prize in Physics was awarded to Max von Laue "for his discovery of the diffraction of x-rays by crystals" and in 1915 the Nobel Prize in Physics was awarded to Sir William Henry Bragg and William Lawrence Bragg "for their services in the analysis of crystal structure by means of x-rays". Since the pioneering

days of von Laue, Ewald, Knipping, Friedrich, the Bragg's, Compton, etc., diffraction experiments went a long way deciphering today the atomic arrangement of noncrystalline solids such as viruses as shown in a recent experiment [17] carried out at the Linac Coherent Light Source (LCLS), at SLAC National Accelerator Laboratory in California, USA) as shown in the diffraction image Fig. 6. Physical principles established 100 years ago and subsequent theoretical and experimental methods developed, reinvented and constantly brought to perfection contribute today and will contribute in the future to the welfare of mankind.

Acknowledgements

References

[1] J.J. Sakurai, *Modern Quantum Theory*, (Addison Wesley, 1984).

[2] Wu and Ohmura, *Quantum Theory of Scattering*, (Prentice Hall, 1962).

[3] A. S. Davidov, *Quantum Mechanics*, (Pergamon Press, 1991).

[4] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, (Brooks Cole, 1976).

[5] C. G. Darwin, Phil. Mag. **27**, 315; **27**, 675 (1914).

[6] P. P. Ewald, Ann. Physik **49**, 1; **49**, 117 (1916); **54**, 519 (1917).

[7] M. v. Laue, Ergeb. Exact. Naturw. **10**, 133 (1931).

[8] H. Bethe, Ann. Physik, **87**, 55 (1928).

[9] W. H. Zachariasen, *Theory of X-Ray Diffraction in Crystals* (Dover Publications, 1945).

[10] J. M. Cowley, *Diffraction Physics* (Elsevier, 1995).

[11] A. Authier, *Dynamical Theory of X-Ray Diffraction* (Oxford University Press, 2001).

[12] J. C. Slater, *Interaction of Waves in Crystals*, Reviews of Modern Physics **30**, 197 (1958).

[13] B. W. Batterman and H. Cole, *Dynamical Diffraction of X Rays by Perfect Crystals*, Reviews of Modern Physics **36**, 681 (1964).

[14] P. H. Dederichs, *Dynamical Diffraction Theory*, Berichte der Kernforschungsanlage **797**, (1971).

[15] Chr. Huygens, Trait de la Lumiere (completed in 1678, published in Leyden in 1690).

[16] Sabrina Disch, Erik Wetterskog, Raphaël P. Hermann, German Salazar-Alvarez, Peter Busch, Thomas Brückel, Lennart Bergström, and Saeed Kamali, Nano Lett. **11**, 1651 (2011).

[17] Henry N. Chapman *et al.*, Nature **470**, 73 (2011).

A 3   Interaction of X-rays, Neutrons and Electrons with Matter

David P. DiVincenzo

PGI-2

Forschungszentrum Jülich GmbH

Contents

1   The Basics of Scattering Theory

Each of the scattering probes that we discuss here, be they particles or waves, permit, according to the tenets of quantum mechanics, a description of either sort. In fact, the wave theory is the best adapted as the unified framework that we will set up here. The incident beam will be treated as monoenergetic and unidirectional – and thus as a plane wave, with incident wave field

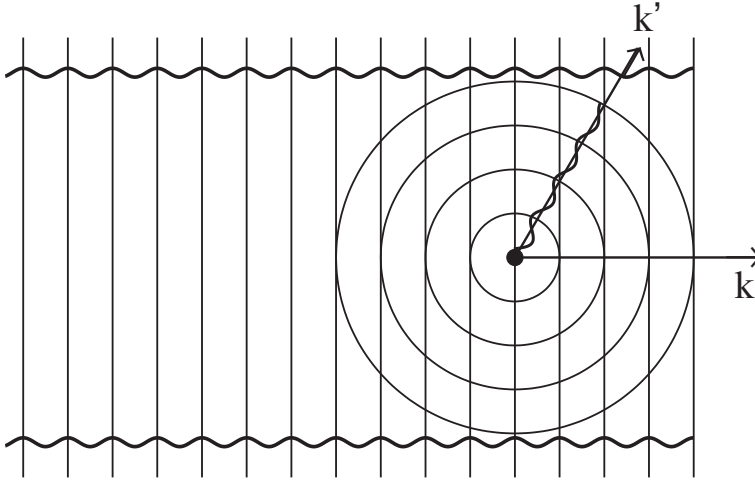$$\Psi_{inc}(\mathbf{r}) = Ae^{i\mathbf{k}\cdot\mathbf{r}} \tag{1}$$

Fig. 1: A schematic of the scattering process from an atomic target. The incident plane wave (wavy line) has wavevector k**; its constant-phase fronts are shown as straight lines. The scattered wave is an outgoing spherical wave (circles) going out in all directions, including in the wavevector direction** k'**.**

The energy of the incident scatterer is a function of the magnitude of the wavevector $|\mathbf{k}|$; for nonrelativistic electrons or neutrons, $E = \hbar^2 k^2 / 2m$, and for light, $E = \hbar c k$. The direction of propagation is of course the direction of the vector $\mathbf{k}$, which will be conveniently described in spherical polar coordinates using angles $(\theta, \varphi) = \Omega$. We assume that there is a small scattering target fixed at the origin. In the relevant wave equation, this scatterer will be described by a potential energy function $V(\vec{r})$. The "interaction region" $|\mathbf{r}| < r_0$ is assumed to be the only region in which $V(\vec{r}) \neq 0$. Outside this interaction region the wave field also contains an outgoing spherical wave of the form

$$\Psi_{scat} = Af(\Omega)\frac{e^{ikr}}{r} \tag{2}$$

We have specialized to elastic scattering (appropriate for most of the scattering experiments considered in this chapter), so that the magnitude of the incident and scattered wavevectors k are the same.

The quantity $f(\Omega)$ is the central focus of our attention, describing the amount of scattering in the direction of the solid angle Ω. Note that the complex quantity f has units of length – it is a "scattering length". It in fact directly indicates the normalized scattering flux $\Delta\sigma(\Omega)$ in a cone of solid angle $\Delta\Omega$ (see Fig. 2), in the direction Ω, for unit incident wave flux density:
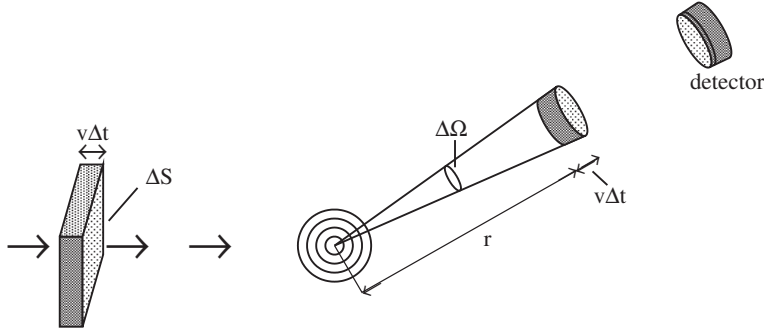

Fig. 2: Geometry of the scattering process.

$$\Delta\sigma = |f(\Omega)|^2 d\Omega \tag{3}$$

The total scattering cross section is

$$\sigma = \int |f(\Omega)|^2 d\Omega \tag{4}$$

We see that the phase of the complex scattering length $f(\Omega)$ does not appear in any of our expressions; however it is very important in the interference that occurs in scattering from two different scattering centers. This effect is beyond the scope of the present chapter.

For completeness, we note the other important quantity, the differential scattering cross section, which is simply the integrand of the quantity above:

$$\frac{d\sigma}{d\Omega} = |f(\Omega)|^2 \tag{5}$$

We end this section with a simple physical picture of the scattering cross section. Naturally, the above discussion implies that the full wave field is given by the sum of the incident and scattered waves, which is correct in the Born approximation:

$$\Psi_{tot}(\mathbf{r}) = A \left(e^{i\mathbf{k}\cdot\mathbf{r}} + f(\Omega)\frac{e^{ikr}}{r} \right) \tag{6}$$

This Born approximation expression does not take account of the fact that the flux of the incident beam is affected (and depleted) by scattering. The amount by which it is depleted is exactly the flux density through the area σ. One can have a simple picture of this result: the depleting effect

of the scatterer is exactly the same as that of a fully absorbing screen with area σ (Figure 2). The common unit for σ in scattering physics is the *barn*, which, at $10^{-28}m^2$, is actually a large unit of area in many areas of particle and nuclear physics. The term originates from the 20th century American taunt to a poor thrower, "You couldn't hit the broad side of a barn."
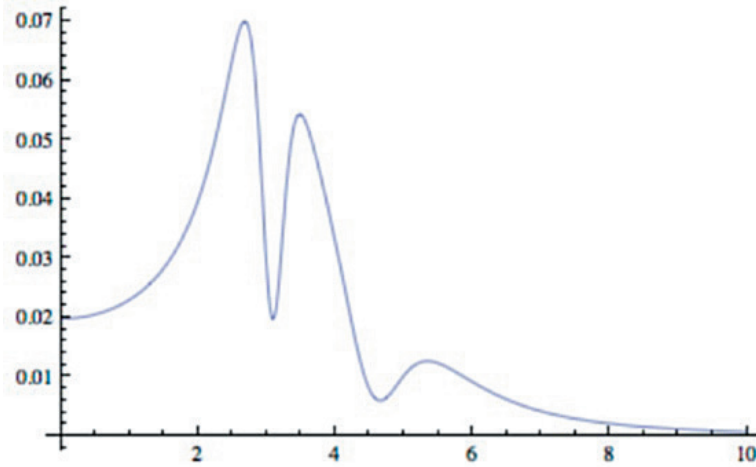

Fig. 3: A schematic of the square of ther real part of the polarizability $(\chi'(\omega))^2$ versus frequency ω. We see the low frequency, Rayleigh part (frequency independent), a complex intermediate frequency range in which anomalous dispersion and absorption occur, and then a high-frequency, Thompson part (going line $1/\omega^4$).

2   The Scattering of Electromagnetic Radiation from Atoms

While we will in this school primarily be concerned with high-energy scattering probes such as X-rays, we begin this discussion at the low-energy (that is, the low-frequency) end of the spectrum. An electromagnetic wave comprises transverse, perpendicular oscillating electric and magnetic fields. We first consider the effect of the electric fields on a target atom. At low frequencies, below that of any atomic resonances, the applied field will polarize the electrons bound to the atom, producing an electric polarization P proportional to the strength of the electric field:

$$P(\omega) = \chi(\omega)E(\omega) \tag{7}$$

At low frequencies, the electric polarizability of the atom $\chi(\omega)$ is independent of frequency ω. The resulting electric dipole oscillating with angular frequency ω, $P(\omega)e^{i\omega t}$, will radiate an outgoing spherical wave – this is the scattered wave of our general scattering theory. From classical electromagnetic theory, the efficiency with which this dipole radiates energy scales like the fourth power of the frequency; the net result for the scattering cross section is the formula for *Rayleigh scattering*:

$$\sigma_R(\omega) = \frac{8\pi}{3} \frac{\omega^4}{(4\pi\epsilon_0 c^2)^2} (\chi(0))^2 \tag{8}$$

Recall that this ω^4 dependence gives Rayleigh's explanation that the sky is blue.

Passing over the visible and ultraviolet region of the spectrum where atoms show complex resonant behavior in their scattering cross section (Figure 3), we consider a regime where the frequency is high enough that the binding of the electron to the atom is *irrelevant*; the electron oscillates as if it were in free space. In this regime a calculation of the oscillating dipole $P(\omega)$ is again straightforward, since it simply requires the calculation of the periodic displacement of a free particle subject to a sinusoidal force. The result in this regime is
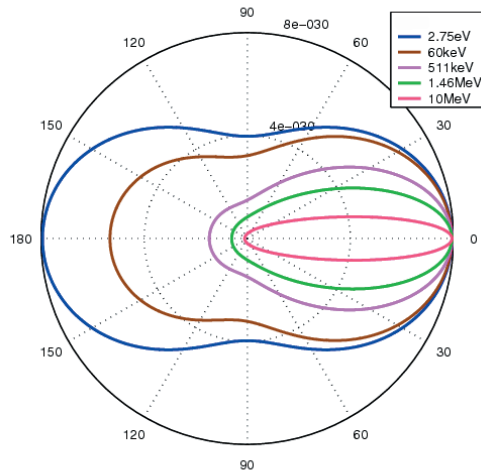

Fig. 4: Polar plot of the Klein-Nishina formula for the differential scattering cross section of X-rays by electrons. At low frequency the scattering goes to that of a classical dipole; at high frequency (the Compton regime) the cross section becomes more and more forward directed, best described as energy- and momentum-conserving photon-electron collisions. From [1].

$$\chi(\omega) = \frac{e^2}{m_e \omega^2} \tag{9}$$

The higher the frequency, the smaller the polarizability, because the electron has a shorter time in which to move. The scattering formula Eq. (8) still applies, so we get the simple, frequency-independent result for the cross section contribution per electron:

$$\sigma_T = \frac{8\pi}{3} r_e^2 \tag{10}$$

This is the regime of *Thompson scattering*. Here

$$r_e = \frac{1}{4\pi\epsilon_0} \frac{e^2}{m_e c^2} \approx 2.8 \times 10^{-13} cm \tag{11}$$

is the so-called *classical electron radius*.

Even though the binding of the electron to the atomic nucleus is irrelevant in the Thompson scattering regime, it should be understood that, in the regime of low excitation intensity, the nevertheless remains associated with the atom, so long as the distance over which the electron travels under the influence of the time-oscillatory force is much smaller than the atomic radius. In this regime, X-ray scattering is non-destructive. Naturally, if the excitation intensity is raised to the point where this oscillation distance becomes comparable to or greater than the atomic radius, we enter the regime of *high-intensity effects*, which can very realistically be achieved with strong X-ray sources such as the free-electron laser (FEL). In that case the X-ray probe is destructive, causing ionization and disruption of chemical structure, so the time available for this scattering probe to give useful information about condensed matter is limited.

Of course a more complete calculation is possible; the result for the differential cross section is

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} \left(1 + \cos^2\theta\right).$$ (12)

The angular dependence appears in Fig. 3, showing the dipolar form that is also characteristic of the Rayleigh scattering regime. This figure shows the result of a much more general calculation due to Klein and Nishina [1], who calculated this scattering taking quantum and relativistic effects into account. The *Klein-Nishina formula* for the differential cross section is
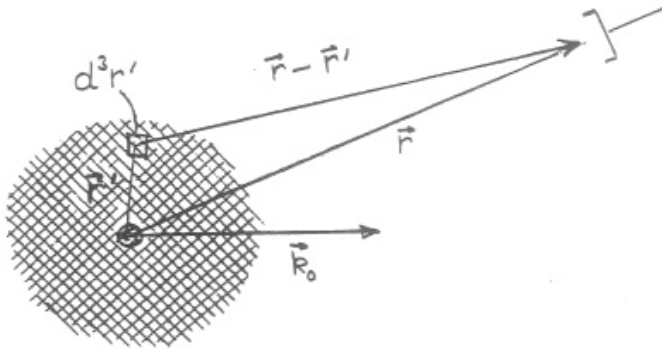

Fig. 5: Scattering geometry for discussion of atomic form factor.

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} P(\omega, \theta)^2 \left(P(\omega, \theta) + P(\omega, \theta)^{-1} - 1 + \cos^2\theta\right).$$ (13)

Here the factor

$$P(\omega, \theta) = \frac{1}{1 + (\hbar\omega/m_e c^2)(1 - \cos\theta)}$$ (14)

has a simple kinematical interpretation when we take the quantum point of view and consider the light to consist of particles (photons): it is the ratio of the photon energy after the scattering event to its original energy before scattering. Note that in the limit of small ω, $P(\omega, \theta) = 1$ and this expression reduces to the one for Thompson scattering. At high frequencies, when the photon energy $\hbar\omega$ becomes comparable to the rest energy of the electron $m_e c^2 = 511\text{keV}$, the

scattering takes on a different character, and we enter the regime of *Compton scattering*. The scattering cross section becomes much more forward-directed, as we can see from the figure; the energetic photon suffers less and less of a deflection during the scattering from the electron, the higher its energy is.
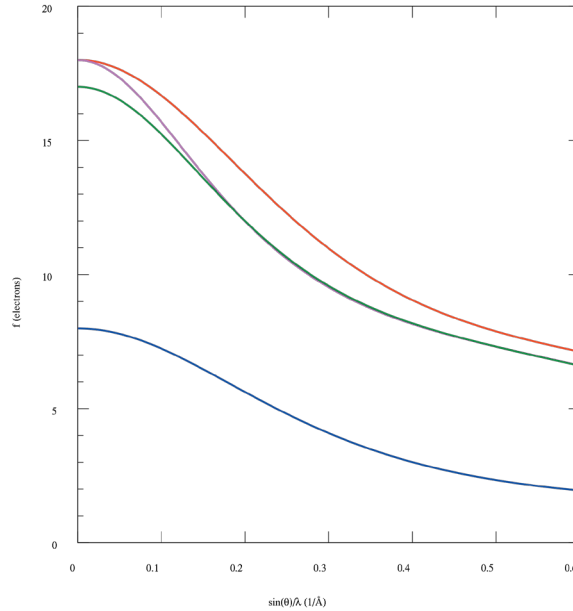

Fig. 6: The atomic form factor $f(Q)$ for several ions and elements, versus $\sin(\theta)/\lambda$. λ is wavelength, θ is scattering angle; with another 4π factor this expression is the scattering wavevector Q. Note that $f(0) = Z$ Z = total number of electrons, not nuclear charge). From top to bottom these curves are for K^+, Cl^- (note that these have the same number of electrons), Cl and O. From [2].

3 Atomic Form Factor for X-rays

The Thompson scattering formula is clearly not the whole story of X-ray scattering from an atom. Even in the (considerable) frequency range in which The scenario for Thompson scattering applies (scattering from quasi-free electrons), we need to take account of the fact that the scattering is from the cloud of electrons that is bound to the atom. This means, in short, that the scatterers are not all at the origin of the coordinate system, and we must do a calculation to sum up their contributions.

Referring to the figure, we consider each volume element d^3r' to be a source of Thompson scattering with a strength governed by the probability that an electron is found in this volume element, which is given by the electron density function according to $n(\mathbf{r}')d^3r'$. The spherical wave that is emitted from that element involves the factor
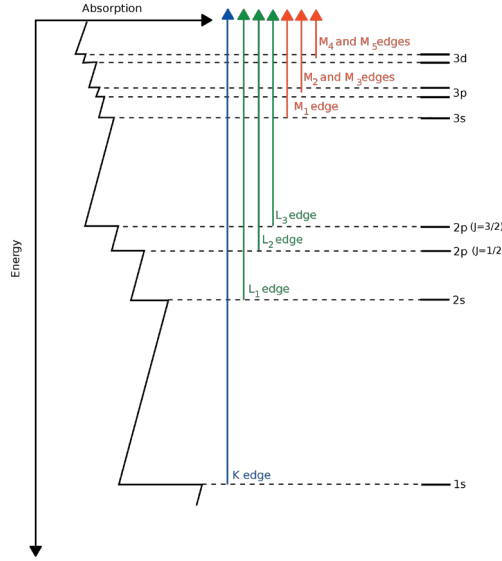
Fig. 7: Schematic of X-ray absorption structure, including many edges, for a typical atom with many inner shells. From [3].

$$n(\mathbf{r}')d^3r'\frac{e^{ik_0|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}e^{i\mathbf{k}_0\cdot\mathbf{r}'}. \tag{15}$$

Note that the final factor comes from the phase of the incident plane wave at the scattering point $\mathbf{r}'$. The overall scattering strength is given by integrating this quantity over the electron cloud:

$$\int d^3r'n(\mathbf{r}')\frac{e^{ik_0|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}e^{i\mathbf{k}_0\cdot\mathbf{r}'}. \tag{16}$$

Note that the angular factor from, e.g., the Thompson differential cross section formula will not appear inside this integral, since we will consider only the far field (i.e., $|\mathbf{r}| >> |\mathbf{r}'|$, so that the angular dependence can be put on as an overall factor once the integral is done.

Since $|\mathbf{r} - \mathbf{r}'| = r - \hat{\mathbf{r}} \cdot \mathbf{r}'$, we can approximate the integrand by

$$\frac{e^{ik_0|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}e^{i\mathbf{k}_0\cdot\mathbf{r}'} \approx \frac{1}{r}e^{ik_0r}e^{i(\mathbf{k}_0-\mathbf{k}_1)\cdot\mathbf{r}'}. \tag{17}$$

Here $\mathbf{k}_1 = \hat{\mathbf{r}}k_0$. We see here appearing the *scattering wavevector*

$$\mathbf{Q} = \mathbf{k}_0 - \mathbf{k}_1 \tag{18}$$

With this we write our scattering amplitude

$$\int d^3r'n(\mathbf{r}')e^{i\mathbf{Q}\cdot\mathbf{r}'}\frac{e^{ik_0r}}{r} = f_a(\mathbf{Q})\frac{e^{ik_0r}}{r}. \tag{19}$$

We identify the Fourier transform of the atomic electron density,

$$f_a(\mathbf{Q}) = \int d^3r' n(\mathbf{r}') e^{i\mathbf{Q}\cdot\mathbf{r}'}, \tag{20}$$

as the *atomic form factor* for X-ray scattering. It is a factor that must be accounted for in other applications of the scattering theory (e.g., for Bragg scattering). For example, it appears this way in the Thompson scattering differential cross section for an atom,

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} |f_a(\mathbf{Q})|^2 \left(1 + \cos^2\theta\right) \tag{21}$$

We can see several features of the atomic form factor from Fig. 6. It is of course isotropic for atoms, so this it depends only on $|\mathbf{Q}|$. Its value at zero is very simple:

$$f_a(0) = Z \tag{22}$$

Z being the total electron number of the atom or ion. The figure shows two cases for which this number is the same, namely for the ions K^+ and Cl^-. The extension of these functions is the reciprocal of the extent of the atomic electron cloud in real space; thus we can observe that the Cl^- ion is considerably more extended than K^+.
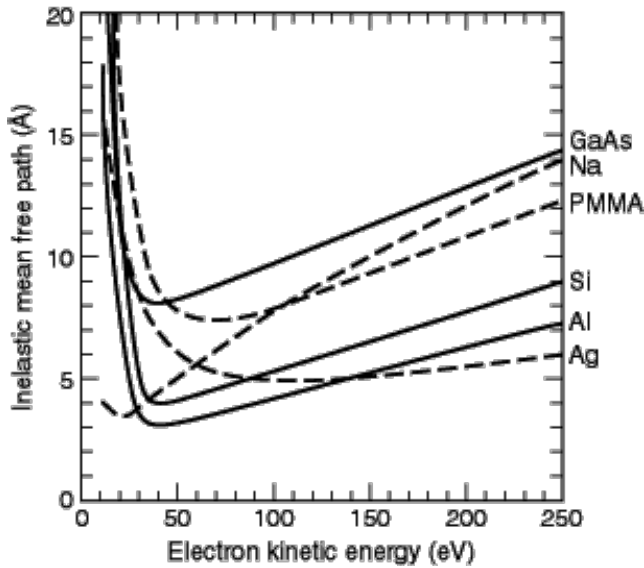

Fig. 8: Electron inelastic mean free path versus incident electron energy, for a range of materials. After [4].

4   X-ray Absorption and Dispersion

Our discussion above has so far ignored the phenomenon of absorption of radiation. We can trace this to our implicit assumption in Eq. (7) that the polarization vector is in-phase with the

applied electric field, so that the polarizability function $\chi(\omega)$ is real. In fact the polarization has an out of phase component as well; elementary electromagnetic theory shows that a polarization oscillating out of phase with the electric field results in absorption of energy. Thus, we write the polarizability function as the sum of the real and an imaginary part:

$$\chi(\omega) = \chi'(\omega) + i\chi''(\omega) \tag{23}$$

For atoms, $\chi''(\omega)$ in the X-ray regime is fairly featureless, except for sharp *X-ray edges* that appear when the radiation can eject electrons from the inner electronic shells of the atom. Figure 7 shows the occurrence of these edges, and how they are interpreted in the shell model.
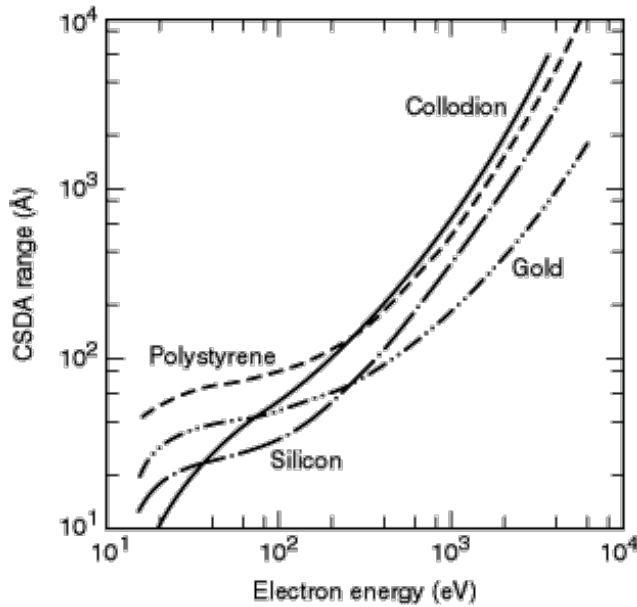

Fig. 9: Continuous Stopping Distance Approximation (CSDA) range of electrons vs. electron energy. After [4].

The occurrence of *dispersion*, significant variation in the lossless response $\chi'(\omega)$, is intimately tied to the appearance of structure in the lossy $\chi''(\omega)$ as exemplified by edges. This connection is embodied in the *Kramers-Kronig relations*. These relations for the $\chi'(\omega)$ function are

$$\chi'(\omega) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{\chi''(\omega)}{\omega' - \omega} d\omega', \tag{24}$$

$$\chi''(\omega) = -\frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{\chi'(\omega)}{\omega' - \omega} d\omega'. \tag{25}$$

The derivation of these relations is usually presented as an exercise involving Cauchy's theorem from complex analysis. I will take a moment to review a less rigorous but more physically informative demonstration involving only the elementary features of the Fourier transform. This derivation makes it more clear that the one and only one premise on which the Kramers-Kronig relations are based is the *causality* of the response of the system during scattering. Imagine

that the scattering wave impinges as a wave packet on the scatterer, so that the time dependent electric field $E(t)$ becomes non-zero only after $t = 0$. The temporal polarization response is given using the Fourier transform of the polarizability function:

$$P(t) = \int_{-\infty}^{\infty} \chi(t - t')E(t')dt' \tag{26}$$

But because such a physical response is causal, $P(t) = 0$ if $t < 0$; the response cannot begin before the excitation has arrived. But for χ, this implies that

$$\chi(t) = 0 \text{ for } t < 0. \tag{27}$$

Now, we write $\chi(t) = \chi_e(t) + \chi_o(t)$, that is, we decompose χ into a sum of an even function of time and an odd function of time. Because of the causality condition these two functions are related:

$$\chi_e(t) = sgn(t)\chi_o(t) \tag{28}$$

Here the "sign function" $sgn(t)$ is $+1$ for $t \geq 0$ and -1 for $t < 0$. Fourier transforming this equation immediately gives the first Kramers-Kronig relation: The Fourier transform of χ_e is purely real and is in fact the real part of $\chi(\omega)$, $\chi'(\omega)$. The Fourier transform of the product is a convolution, the Fourier transform of $\chi_o(t)$ is $i\chi''(\omega)$, and the transform of $sgn(t)$ is $-i/(\pi\omega)$. The other relation is obtained similarly.

Hopefully this little discussion takes some of the mystery out of these relations. What do they have to do with X-ray scattering? We can see the connection by looking at the model of H. A. Lorentz for absorption based on a model of a damped resonator with resonant frequency ω_0. This is a good model for an electron bound in an atom; it captures only qualitatively the X-ray absorption edges, which involve not just the oscillation of the electron but also the ejection of the electrons into a continuum. But the "Lorentzian lineshape" for the absorption in Lorentz's model is very simple:

$$\chi''(\omega) = \frac{e^2}{m_e} \frac{\Gamma\omega}{(\omega_0^2 - \omega^2)^2 + \Gamma^2\omega^2} \tag{29}$$

Here Γ is a linewidth or damping parameter. The Kramers-Kronig relation above requires that this absorbtion function be accompanied by the following frequency-dependent in-phase polarizability:

$$\chi'(\omega) = \frac{e^2}{m_e} \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \Gamma^2\omega^2}. \tag{30}$$

Note that this expression intepolates between the two low-absorption regimes that we have discussed above: Rayleigh scattering (for $\omega \ll \omega_0$, and Thompson scattering for $\omega \gg \omega_0$. The Kramers-Kronig constraints say that there *must* be a regime of high loss in between, and that the in-phase polarizibility much also rise to a much higher value than in either of the two limits (in fact, $\chi'(\omega)_{max} = e^2/m(2\Gamma\omega_0+\Gamma^2) \approx e^2/(2m\Gamma\omega_0)$ for $\Gamma \ll \omega_0$). This strong enhancement near an absorption feature of the real part of the polarizability, and therefore of the scattering cross section, and its strong frequency dependence, is known as *anomalous dispersion*. As you will learn elsewhere in this course, this phenomenon is used to enhance the contrast of one atomic element relative to another in X-ray scattering.

I conclude this section with a brief discussion of *magnetic X-ray scattering*. I have so far described the X-ray scattering process as involving only the electric field of the incident wave.

Naturally, the electromagnetic wave also has a magnetic component, normal to the direction of propagation and also normal to the electric field. This magnetic field also induces a response, and causes an additional contribution to the scattered spherical wave. Most importantly, this scattering is sensitive to the magnetic state of the target – the scattering from an atom will be different when its spin is up or down. Thus, such contributions to the scattering can distinguish the magnetic state (ferromagnetic, antiferromagnetic, etc.) of a material. Generally, this magnetic contribution to the scattering is weak; the scattering amplitude has a prefactor $\hbar\omega/m_e c^2$, so that this scattering is generically suppressed for X-ray photon energies below 511keV. Wise use of magnetically-dependent anomalous dispersion can enhance the magnetic signal.

5 Electron Scattering

The basic physics of the scattering of electrons from matter is the same as that for X-ray photons: in the quantum theory the electrons have a wave description, and the basic scenario of scattering, in which there is an incident wave on the target, and an outgoing scattered spherical wave. The important qualitative distinction between electron scattering and X-ray scattering is that the strength of electron scattering is much greater than that of X-rays. Electrons will not penetrate a large thickness of material as X-rays will.

In fact, the scattering cross section for electrons can be deduced directly from the cross section for electromagnetic waves, already discussed above. In the electron wave equation (the Schroedinger equation), the scattering intensity from point $\mathbf{r}$ is determined by the potential function at that point $V(\mathbf{r})$. From a calculation of the scattering problem using this equation, the scattering form factor is given by the expression

$$f^e(Q) = \frac{2m_e e}{\hbar^2} \int_0^\infty \frac{V(r)\sin(qr)r^2 dr}{qr} \tag{31}$$

Note that by convention the form factor for electron scattering also contains the scattering length; this means that it has units of meters, rather than being dimensionless as the X-ray form factor is taken to be.

One further step permits f^e to be related directly to the X-ray form factor, since we can relate the scattering potential $V(\mathbf{r})$ to the electron density $n(r)$ whose fourier transform determines the $f(Q)$ for X-rays. This relation is via the Poisson equation, $\nabla^2 V(\mathbf{r}) = -\frac{e}{\epsilon_0} n(\mathbf{r})$. Fourier transforming this equation and substituting into Eq. (31) gives the *Mott-Bethe formula* for the electron form factor for an atom with atomic number Z:

$$f^e(Q, Z) = \frac{m_e e^2}{2\pi\hbar^2\epsilon_0} \left(\frac{Z - f(Q, Z)}{Q^2} \right) \tag{32}$$

This equation also includes the form factor Z/Q^2 for the atomic nucleus.

The principal item of practical interest that I will cover here is the theory of the stopping range of low-energy electrons in solid matter. For electrons with an incident energy in the range of 5 keV, the basic picture is that electrons slow down by a large sequence of scatterings in the material, each of which leads to a small loss of energy. We speak of the *continuous slowing down approximation* (CSDA) in calculating the electron range. This calculation again involves

the polarizability of constituents $\chi(\omega)$. When summed over a large number of constituents, this response is called the *dielectric function* $\epsilon(Q, \omega)$; this expression singles out polarization leading to a scattering wavevector $\mathbf{Q}$. Then for an electron traveling with energy E, the probability of energy loss ω over a unit of distance is given by the expression

$$p(E, \omega) = \frac{m_e e^2}{\pi \hbar^2 E} \int_{q_-}^{q_+} \mathrm{Im}\left(\frac{-1}{\epsilon(Q, \omega)}\right) \frac{dQ}{Q} \tag{33}$$

Here $\hbar q_\pm = \sqrt{2mE} \pm \sqrt{2m(E - \hbar\omega)}$. $p(E, \omega)$ is known as the differential inverse mean free path. The stopping power $S(E)$, which is the energy loss per unit distance travelled along the electron path, is given by

$$S(E) = \int dE \hbar\omega p(E, \hbar\omega) \tag{34}$$

Finally, the distance over which the electron is stopped (actually, brought to a nominal kinetic energy of 10eV) is the CSDA range $R_0(E)$, given by

$$R_0(E) = \int_{10eV}^{E} \frac{dE'}{S(E')} \tag{35}$$

In Fig. 8 we show the inelastic mean free path for electrons in a wide variety of solid matierials, over the range of incident energies from 10 eV to 250 eV. This quantity continues to grow almost linearly above this energy, up to 2keV. There are two important things to note about this quantity: it has a minimum at a few tens of eV. Electrons are more penetrating at energies both above and below this. Second, the scale of this mean free path is very small, being a fraction of a nanometer over much of this energy range. Just a couple of atomic layers are effective at blocking the passage of a large fraction of electrons in this energy range.
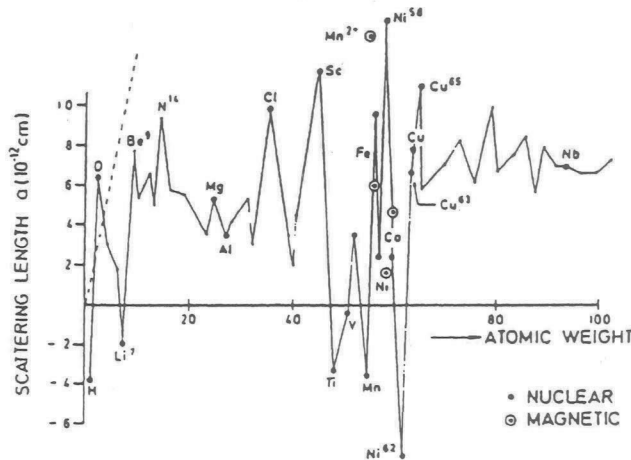

Fig. 10: Scattering length b for the atomic elements, showing the non-monotone dependence of b on atomic weight, even for isotopes of the same element. See [5].

Finally, Fig. 9 shows a sampling of the total travel range for incident electrons (note the much greater range of energies than in the previous figure) for several types of solid materials. Note that the total penetration range never exceeds 1 micrometer for any of the cases shown.

6 Neutron Scattering

The neutron, a particle with no charge, a mass very close to that of the proton, and a spin magnetic moment about 1000 times smaller than that of the electron, is a very useful scattering probe. Neutron beams of high intensity and sharply defined energy and direction can be produced and directed at targets; the neutron's lack of charge permits low-energy neutrons to penetrate deeply into matter. The free neutron is unstable; while its decay is fast compared with many radio-nuclei, at about 10 minutes this time is very long compared with that of the scattering process and detection, so that this decay can be ignored in discussions of neutron scattering. The energy range of the scattering neutrons is usually "thermal", meaning that the kinetic energy of the neutrons is reduced by moderation (passage through a non-absorbing material) into the range of $k_B T$ with $T \approx 300$K. (Moderation to lower energies is possible.) Monochromators pick off well defined energies from this moderated collection of neutrons.

Lacking an electric change, the neutron still has two means of interacting significantly with matter: first, its magnetic moment makes it sensitive to the magnetic scatterers in the target. Second, the strong nuclear force causes there to be a significant scattering cross section from each magnetic nucleus. It turns out that these contributions are roughly of the same order of magnitude, both are very important in the application of neutron scattering.

We will deal first with the scattering arising from the nuclear force. The strong interaction of the neutron with a many-nucleon atomic nucleus is very complex. However, the description we need of scattering is very much simpler, because at thermal energies, the wavelength of the quantum-mechanical (de Broglie) neutron wave is in the vicinity of 0.1 nm, comparable, in fact, to the internuclear spacings in molecules or solids. This wavelength is very long compared with the range of the strong nuclear force (about 10^{-6} nm). Thus, the neutron-nucleus interaction may be accurately represented as a delta-function at the origin; this is called the Fermi pseudopotential. Fermi writes

$$V(\mathbf{r}) = \frac{2\pi \hbar^2}{m_N} b \delta(\mathbf{r}) \tag{36}$$

Here m_N is the neutron mass. b has dimensions of a length, and is in fact the s-wave scattering length; it is also equal to the neutron form factor, since the delta-function form of the potential means that the form factor has no dependence on the scattering wavevector $\mathbf{Q}$.

In the simplest view b is just a simple scalar number. We must be a little bit more sophisticated, for several reasons. First, if the target nucleus possesses a non-zero nuclear spin quantum number I, then the scattering depends on the relative angles of the neutron spin vector $\mathbf{s}$ and the nucleus angular momentum vector $\mathbf{I}$; in general this spin dependence is quite strong. This effect is included by writing b as

$$b = b_c + \frac{2b_i}{\sqrt{I(I+1)}} \mathbf{s} \cdot \mathbf{I} \tag{37}$$

Thus, the scattering process takes two parameters to describe; for historical reasons, these parameters are called the *coherent* cross section b_c and the *incoherent* cross section b_i. In fact both

parameters describe perfectly coherent wave scattering phenomena. However, it is typical in scattering experiments to have no control over the spin state of the target nucleus (there are now many exceptions to this); thus it has been traditional to consider the nuclear spin state to be random, causing the resulting scattering to be incoherent. We will state shortly the consequences of this for scattering from atomic crystals.

The second fact about b that we wish to note is that it also possesses an imaginary part ib". As with χ'' above, this constant describes the absorption of neutrons,due to nuclear reactions, during the scattering. Finally, the b "constants" can also be functions of energy. Generally b'' has a linear energy dependence, so that in tabulations the scattering energy must be stated. The real part is in most cases energy independent at thermal energies, although it should be noted that for a small minority of the atomic nuclei, there are already resonances, with anomalous dispersion and enhanced absorption, already at low energies.

All these parameters are accurately measured and can be found tabulated, typically with a couple of digits of accuracy but sometimes much more, for all the isotopes of the periodic table of elements. Unlike the X-ray and electron scattering lengths, which increase monotonically as one moves down the periodic table, the b parameters, which depend on complex details of nuclear physics, are already large for the lightest nuclei, and vary tremendously from one element to the next, and vary to the same degree even for isotopes of the same element. We show this variation in Fig. 10. So, a crystal of pure He has a perfect periodic structure as seen either by X-ray scattering or neutron scattering (producing "Bragg peaks", see Chap ???), because pure helium consists almost entirely of one spinless isotope, He-4. But a crystal of pure selenium with equal amounts of Se-74 and Se-76 (these are both natural isotopes of Se, but these are not the natural abundances) looks highly disordered from the point of view of neutron scattering, producing a large component of non-Bragg diffuse scattering, because b_c for Se-74 and Se-76 are very different (0.8 and 12.2 barns, resp.). On the other hand, a pure crystal of arsenic looks disordered for the other reason; while there is only one stable isotope As-75, the four different permitted spin states of the I=3/2 As-75 nucleus scatter with considerably different strengths (because $b_i = -0.7$ barns).

7 Magnetic Neutron Scattering

The neutron is a chargeless particle, but it has a magnetic moment, which is about the same in magnitude as the protons, and about 100 times smaller than that of the electron. We analyse the scattering of the neutron from the field arising from the spin of an electron at position $\mathbf{r}'$:

$$\mathbf{B}_S = \frac{\mu_0}{4\pi} \nabla_{\mathbf{r}} \times \left(\boldsymbol{\mu}_e \times \nabla_{\mathbf{r}} \frac{1}{|\mathbf{r} - \mathbf{r}'|} \right), \tag{38}$$

here $\boldsymbol{\mu}_e = g_e \mu_B \mathbf{s}_e$ is the electron spin magnetic moment; the Bohr magneton is μ_B and the electron spin operator is $\mathbf{s}_e$. The field arising from electron orbital motion is

$$\mathbf{B}_L = -\frac{\mu_0 e}{4\pi} \frac{\mathbf{v}_e \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \tag{39}$$

(we use the Biot-Savart law for a particle with charge $-e$ and velocity $\mathbf{v}_e$). Here we further consider only the spin field, yielding the potential

$$V(\mathbf{r}) = -\boldsymbol{\mu}_n \mathbf{B}_S = -\boldsymbol{\mu}_n \frac{\mu_0 g_e \mu_B}{4\pi} \nabla_\mathbf{r} \times \int d\mathbf{r}' \mathbf{s}_e(\mathbf{r}') \times \nabla_\mathbf{r} \frac{1}{|\mathbf{r} - \mathbf{r}'|}, \tag{40}$$

$\boldsymbol{\mu}_n$ is the neutron magnetic moment, and $\mathbf{s}_e(\mathbf{r}')$ is the electronic spin density. The scattering amplitude requires a calculation of a double integral

$$I = \int d\mathbf{r} e^{-i\mathbf{Q}\mathbf{r}} \nabla_\mathbf{r} \times \int d\mathbf{r}' \mathbf{s}_e(\mathbf{r}') \times \nabla_\mathbf{r} \frac{1}{|\mathbf{r} - \mathbf{r}'|} \tag{41}$$

We perform this evaluation in the Furier domain

$$\frac{1}{r} = \frac{1}{2\pi^2} \int d\mathbf{q} \frac{e^{i\mathbf{q}\mathbf{r}}}{q^2}.$$

One obtains

$$\begin{aligned} I &= -\frac{1}{2\pi^2} \int d\mathbf{r}' \int d\mathbf{r} e^{-i\mathbf{Q}\mathbf{r}} \int d\mathbf{q} \hat{\mathbf{q}} \times \mathbf{s}_e(\mathbf{r}') \times \hat{\mathbf{q}} e^{i\mathbf{q}(\mathbf{r}-\mathbf{r}')} \\ &= -4\pi \hat{\mathbf{Q}} \times \int d\mathbf{r}' \mathbf{s}_e(\mathbf{r}') \times \hat{\mathbf{Q}} e^{-i\mathbf{Q}\mathbf{r}'}. \end{aligned} \tag{42}$$

Note that the application of $\nabla_\mathbf{r}$s to the exponential term $e^{i\mathbf{q}(\mathbf{r}-\mathbf{r}')}$ results in the simple factors $i\mathbf{q} = iq\hat{\mathbf{q}}$. The integration over $\mathbf{r}$ gives the delta function $\delta(\mathbf{Q} - \mathbf{q})$, permitting the integration over $\mathbf{q}$ to be completed. Specializing to the case of constant spin direction $\mathbf{s}_e(\mathbf{r}') = s_e(\mathbf{r}')\hat{\mathbf{s}}$ we get a scattering amplitude

$$f(\mathbf{Q}) = -\boldsymbol{\mu}_n \frac{2m_n}{\hbar^2} \frac{\mu_0 g_e \mu_B}{4\pi} \hat{\mathbf{Q}} \times \hat{\mathbf{s}} \times \hat{\mathbf{Q}} F_{magn}(\mathbf{Q}), \tag{43}$$

here we see the *magnetic form factor*

$$F^{magn}(\mathbf{Q}) = \int d\mathbf{r}' e^{i\mathbf{Q}\mathbf{r}'} s_e(\mathbf{r}'). \tag{44}$$

Using the usual cross-product identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$ one gets $\hat{\mathbf{Q}} \times \hat{\mathbf{s}} \times \hat{\mathbf{Q}} = \hat{\mathbf{s}} - (\hat{\mathbf{s}} \cdot \hat{\mathbf{Q}})\hat{\mathbf{Q}}$; note that this is the component of $\hat{\mathbf{s}}$ perpendicular to $\hat{\mathbf{Q}}$. Thus the scattering amplitude (44) is related to the Fourier transform of the spin density component perpendicular to the scattering vector $\mathbf{Q}$. So, magnetic neutron scattering allows a determination of both the size and the direction of the magnetisation in a material of interest. Originally neutron scattering was the only practical probe for the determination of the magnetic structure of solids. In the present time magnetic X-ray scattering with X-rays produced using synchrotron radiation sources can also deliver such information.

Let us estimate the magnitude of the magnetic scattering length (43). The neutron magnetic moment is $\boldsymbol{\mu}_n = \frac{1}{2} g_n \mu_N \boldsymbol{\sigma}$, where g_n is the neutron g-factor, μ_N the nuclear magneton and $\boldsymbol{\sigma} = 2\mathbf{s}_n$ is the Pauli spin operator. We can estimate the prefactor in (43) to be

$$\frac{2m_n}{\hbar^2} \frac{\mu_0}{4\pi} g_n \frac{e\hbar}{2m_p} g_e \frac{e\hbar}{2m_e} \approx 4 \frac{e^4}{(4\pi\epsilon_0 \hbar c)^2} \frac{4\pi\epsilon_0 \hbar^2}{m_e e^2} = 4\alpha^2 a_0,$$

where $g_n \approx -4$, $g_e \approx -2$, $m_n \approx m_p$, $\mu_0 = 1/\epsilon_0 c^2$, the Bohr magneton $\mu_B = e\hbar/2m_e$, the nuclear magneton $\mu_N = e\hbar/2m_p$, the fine structure constant $\alpha = e^2/4\pi\epsilon_0\hbar c$ and the Bohr radius $a_0 = 4\pi\epsilon_0\hbar^2/m_e e^2$ have been used. Note that $\alpha^2 a_0$ is the classical electron radius r_e; this happens to be in the same range as nuclear scattering lengths b. Thus the nuclear and magnetic scattering are of competitive size (very much unlike the X-ray case); this means that interference between the two forms of scattering can in practice occur.

References

[1] Wikipedia article, "Klein-Nishina Formula".

[2] Wikipedia article, "Atomic Form Factor".

[3] Wikipedia article, "X-ray Absorption Spectroscopy"

[4] Piero Pianetta, *X-ray data booklet*, http://xdb.lbl.gov/Section3/Sec_3-2.pdf

[5] Varley Sears, *Neutron Cross Sections*, Neutron News **3** (3), 26 (1992).

A 4 Correlation Functions Measured by Scattering Experiments

R. Zorn

Jülich Centre for Neutron Science

Forschungszentrum Jülich GmbH

Contents

1   Introduction

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

$$|\mathcal{F}[f(x)]|^2 = \mathcal{F}[\langle f(0)f(x)\rangle] \tag{1}$$

($\mathcal{F}$ denoting the Fourier transform) 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 distribution itself is not known.
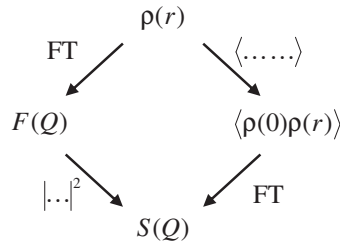

Fig. 1: The two ways to calculate the scattering law from the microscopic density, left: as the absolute-squared Fourier transform of the density, right: as the Fourier transform of the correlation function.

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 second 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.

Textbooks on scattering theory usually restrict themselves to a single probe. For neutron scattering refs. 1–3 can be recommended, for light scattering refs. 4–8, and for x-ray scattering refs. 9–11.

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 $d\Omega$ normalised to the intensity of the incident beam:

$$\frac{d\sigma}{d\Omega} = \left\langle \left| \sum_{j=1}^{N} b_j \exp(i\mathbf{Q} \cdot \mathbf{r}_j) \right|^2 \right\rangle \,. \tag{3}$$

b_j is a measure of the 'scattering power' of the particle. From the dimensions 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 the direction of the propagation of the wave. For elastic scattering $k' = k$, and the definition of $\mathbf{Q}$ 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. Quantum-mechanically Q corresponds to the momentum transfer due to the scattering process: $\hbar\mathbf{Q} = \hbar\mathbf{k}' - \hbar\mathbf{k} = \mathbf{p}' - \mathbf{p}$.

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 $d\sigma/d\Omega$ but an assembly of so-called speckles. For two reasons this situation is exceptional and the observed scattering is usually an average:
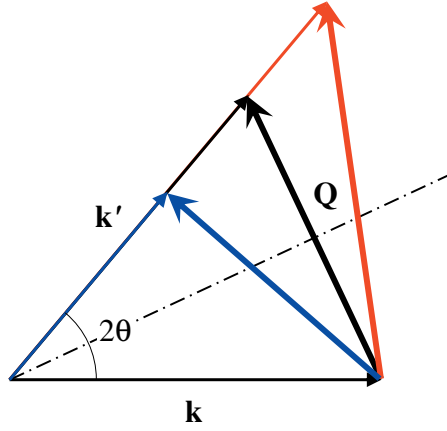
Fig. 2: Definition of the scattering vector $\mathbf{Q}$ in terms of the incident and final wave vectors $\mathbf{k}$ and $\mathbf{k}'$. The black (isosceles) triangle corresponds to elastic scattering. The blue and red ones correspond to inelastic scattering with energy loss or gain of the scattered radiation, respectively.

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 scatters. $S(\mathbf{Q})$ 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):

$$S(\mathbf{Q}) = \frac{1}{N}\left\langle\left|\int_V \mathrm{d}^3r\exp(\mathrm{i}\mathbf{Q}\cdot\mathbf{r})\rho(\mathbf{r})\right|^2\right\rangle\,. \tag{7}$$

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. In this case 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

$$\frac{d\sigma}{d\Omega} = \left\langle \left| \int_V d^3r \, \exp(i\mathbf{Q} \cdot \mathbf{r}) \rho_b(\mathbf{r}) \right|^2 \right\rangle \,. \tag{9}$$

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):

$$\begin{aligned} S(\mathbf{Q}) &= \frac{1}{N} \left\langle \left(\sum_{j=1}^{N} \exp(-i\mathbf{Q} \cdot \mathbf{r}_j) \right) \left(\sum_{k=1}^{N} \exp(i\mathbf{Q} \cdot \mathbf{r}_k) \right) \right\rangle \\ &= \frac{1}{N} \sum_{j,k=1}^{N} \left\langle \exp(i\mathbf{Q} \cdot (\mathbf{r}_k - \mathbf{r}_j)) \right\rangle \,. \end{aligned} \tag{10}$$

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.

In order to proceed in a similar way as before, we introduce the two-particle density

$$\rho(\mathbf{r}_1)\rho(\mathbf{r}_2) = \sum_{j,k=1}^{N} \delta(\mathbf{r}_1 - \mathbf{r}_j)\delta(\mathbf{r}_2 - \mathbf{r}_k) \tag{11}$$

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:

$$\langle \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle \neq \langle \rho(\mathbf{r}_1)\rangle\langle \rho(\mathbf{r}_2)\rangle = \rho_0{}^2 \tag{12}$$

($\rho_0 = N/V$). 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, $0 < |\mathbf{r}_2 - \mathbf{r}_1| < 2R$. Nevertheless, in a translationally invariant system one of the positions can be chosen arbitrarily, especially as the origin, so that

$$\langle \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle = \langle \rho(\mathbf{0})\rho(\mathbf{r}_2 - \mathbf{r}_1)\rangle = \rho_0 \left\langle \sum_{j,k=1}^{N} \delta(\mathbf{r}_j - \mathbf{r}_k + \mathbf{r}_2 - \mathbf{r}_1) \right\rangle . \tag{13}$$

For a system of identical scatterers the two-particle density (11) can now be used to express the structure factor:

$$\begin{aligned} S(\mathbf{Q}) &= \frac{1}{N}\left\langle \int_V \mathrm{d}^3 r_1 \int_V \mathrm{d}^3 r_2 \exp(i\mathbf{Q}\cdot(\mathbf{r}_2 - \mathbf{r}_1))\rho(\mathbf{r}_1)\rho(\mathbf{r}_2) \right\rangle \\ &= \frac{1}{\rho_0} \int_{V_d} \mathrm{d}^3 r \exp(i\mathbf{Q}\cdot\mathbf{r})\langle \rho(\mathbf{0})\rho(\mathbf{r})\rangle . \end{aligned} \tag{14}$$

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 liquids and colloids is

$$g(\mathbf{r}) = \frac{\langle \rho(\mathbf{0})\rho(\mathbf{r})\rangle}{\rho_0{}^2} - \frac{\delta(\mathbf{r})}{\rho_0} . \tag{15}$$

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 (11). With this pair correlation function the structure factor can be written as

$$S(\mathbf{Q}) = 1 + \rho_0 \int_{V_d} \mathrm{d}^3 r \exp(i\mathbf{Q}\cdot\mathbf{r})(g(\mathbf{r}) - 1) . \tag{16}$$

Here, the $1+$ compensates the delta function term subtracted in (15). 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 $Q = |\mathbf{Q}|$ and the volume integral reduces to a one-dimensional:

$$S(Q) = 1 + \frac{4\pi\rho_0}{Q} \int_0^\infty (g(r) - 1)\sin(Qr)r\mathrm{d}r . \tag{17}$$
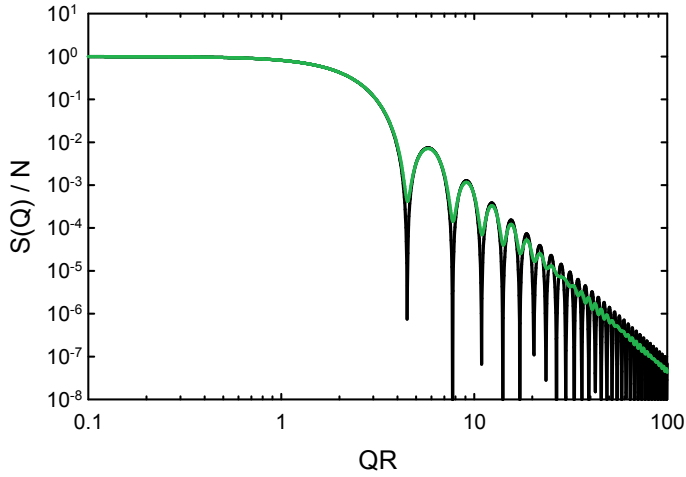
Fig. 3: Form factor of a sphere (black), average for spheres distributed equally over $R = 0.95\ldots 1.05$ (green).

Another definition, more often used for crystalline structures, is the Patterson function:

$$P(\mathbf{r}) = \frac{\langle \rho(\mathbf{0})\rho(\mathbf{r})\rangle}{\rho_0} \tag{18}$$

which Fourier-transforms into the structure factor without further prefactors:

$$S(\mathbf{Q}) = \int_{V_d} d^3r \exp(i\mathbf{Q}\cdot\mathbf{r})P(\mathbf{r})\,. \tag{19}$$

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

$$\rho(r) = \begin{cases} \rho_0 & \text{for } r < R \\ 0 & \text{for } r > R \end{cases}. \tag{20}$$

Because $\rho(r)$ only depends on the magnitude of $\mathbf{r}$ the volume integral in expression (7) can be simplified:

$$\int_V d^3r \exp(i\mathbf{Q}\cdot\mathbf{r})\rho(\mathbf{r}) = \frac{4\pi}{Q}\int_0^\infty \rho(r)\sin(Qr)r\,dr \tag{21}$$

The one-dimensional integral in this expression can be calculated as

$$
\begin{aligned}
\int_0^R \rho_0 \sin(Qr)r\,dr \quad &\overset{\substack{x = Qr \\ \downarrow}}{=} \quad \frac{\rho_0}{Q^2}\int_0^{QR} x\sin x\,dx \\
&= \quad \frac{\rho_0}{Q^2}\left(-x\cos x\Big|_0^{QR} + \int_0^{QR}\cos x\,dx\right) \\
&= \quad \frac{\rho_0}{Q^2}\left(\sin(QR) - QR\cos(QR)\right).
\end{aligned}
$$

With this result and $V = 4\pi R^3/3$ one gets:

$$P(Q) = S(Q)/N = \frac{9}{Q^6 R^6} \left(\sin(QR) - QR\cos(QR)\right)^2 . \tag{22}$$

Fig. 3 shows the form factor[1] of the sphere. The zeros predicted by (22) are usually not found in the experiment but only more-or-less pronounced minima. The reason is that experimental samples usually consist 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 green curve of Fig. 3.

2.4   Example: liquid structure factor

This second example is not an exact calculation but rather 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 inter-particle potential $V(r)$ there is no exact way to derive the pair correlation function $g(r)$. There are only approximative analytical methods [12] 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 (17) the structure factor can be calculated. Although again an exact result cannot be given, several general features can be stated: For $Q \to \infty$, $\exp\left(\mathrm{i}\mathbf{Q} \cdot \mathbf{r}\right)$ becomes a rapidly oscillating function and the integral vanishes. Then one has

$$\lim_{Q\to\infty} S(Q) = 1 . \tag{23}$$

For $Q \to 0$, $S(Q)$ measures only the overall density fluctuation, i.e. the fluctuation of the particle number:

$$\lim_{Q\to 0} S(Q) = \frac{V^2\langle\delta\rho^2\rangle}{N} = \frac{\langle N^2\rangle - \langle N\rangle^2}{\langle N\rangle} = \rho_0 k_{\mathrm{B}} T \kappa_T . \tag{24}$$

Here, k_{B} denotes the Boltzmann factor, 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 $Q_{\max} \approx 2\pi/r_{\mathrm{nn}}$.

As an example of a static neutron scattering experiment on a liquid, the results for Argon at 85 K are shown in Fig. 5 [13]. One can see that the actual neutron scattering data are restricted

---

[1] For this lecture it is sufficient to consider the form factor $P(Q)$ just as a more convenient normalisation of the structure factor which does not contain the number of scatterers anymore. But there is also a conceptual difference between form- and structure factor which will be discussed in later lectures (B3, D3...).
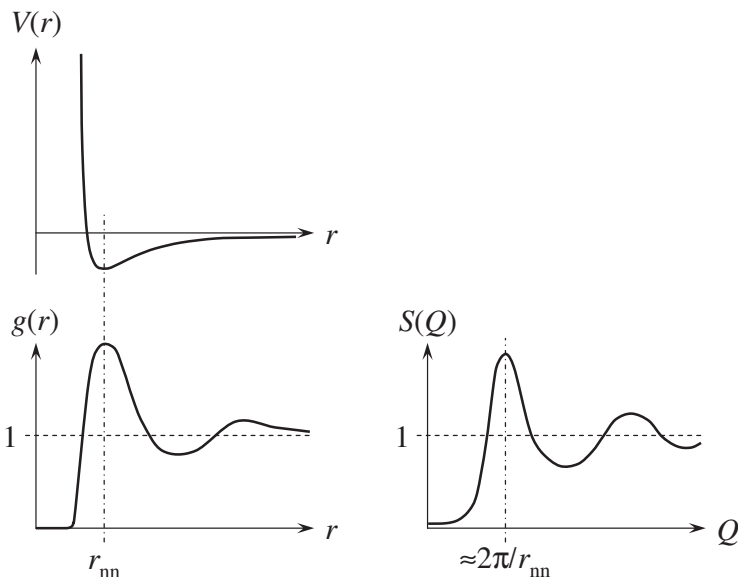
Fig. 4: Schematic representation of interaction potential $V(r)$, pair correlation function $g(r)$, and structure factor $S(Q)$.
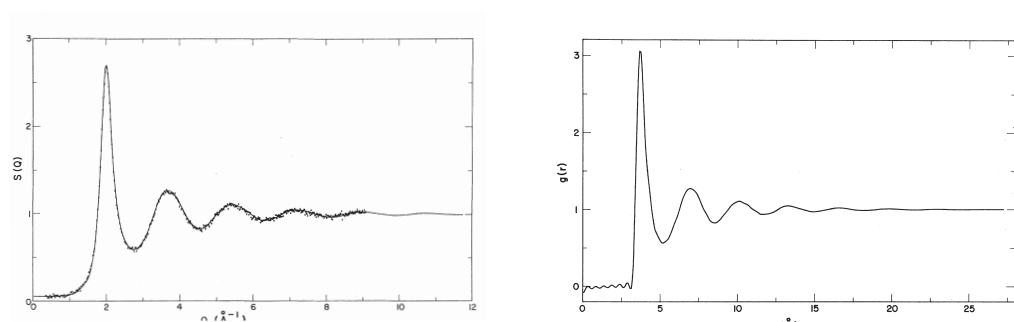

Fig. 5: Left: Structure factor $S(Q)$ of liquid Argon at 85 K. The points are resulting from a neutron scattering experiment. The line represents a smoothing and extrapolation explained in the original publication. Right: Pair correlation function $g(r)$ calculated from the smoothed and extrapolated $S(Q)$. [13]

to the range $Q = 0.4 \ldots 9$ Å^{-1}. Inverse Fourier transforming this data directly would lead to strong oscillations in the pair correlation function $g(r)$. Therefore, $S(Q)$ was smoothed and extrapolated by a combination of molecular dynamics simulation and plausibility restrictions for $g(r)$. Nevertheless, there are still oscillations visible in $g(r)$ below 3 Å which are due to the upper Q limit of the instrument used. State-of-the-art neutron diffractometers using 'hot' neutrons can reach about 15 Å^{-1}, significantly improving the quality of $g(r)$ [2].

3 Scattering from dynamic systems

Here, the more realistic situation will be considered in which the particles of the sample are moving. For moving particles, energy may be transferred to or from the scattered particle, or in the wave picture the frequency is changed by the Doppler effect. Thus an energy transfer

$$\Delta E = E' - E \equiv \hbar\omega \tag{25}$$

occurs, the scattering is in general *inelastic*, and $k' \neq k$. 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:

$$Q = \sqrt{k^2 + k'^2 - 2kk'\cos(2\theta)} \tag{26}$$

$$= \sqrt{\frac{8\pi^2}{\lambda^2} + \frac{2m\omega}{\hbar} - \frac{4\pi}{\lambda}\sqrt{\frac{4\pi^2}{\lambda^2} + \frac{2m\omega}{\hbar}}\cos(2\theta)} \tag{27}$$

for non-relativistic particles with rest mass $m > 0$, e.g. neutrons. 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. This complication may be unnecessary to consider in cases where $|\hbar\omega| \ll E$, as for x-ray scattering, Brillouin light scattering, and photon correlation spectroscopy. But it will be important for methods where $\hbar\omega$ and the incident energy of the particles E are comparable, as Raman light scattering or neutron scattering. Fig. 6 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.

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 $\hbar\omega \ldots \hbar(\omega + \mathrm{d}\omega)$. A quantum-mechanical calculation based on Fermi's Golden Rule yields [1–3]:

$$\frac{\partial^2\sigma}{\partial\Omega\partial E'} \propto \frac{k'}{k}\sum_{\lambda,\lambda'}P_\lambda\left|\langle\lambda'|\langle\mathbf{k}'|\hat{V}|\mathbf{k}\rangle|\lambda\rangle\right|^2\delta\left(\hbar\omega + E_\lambda - E_{\lambda'}\right). \tag{28}$$

Here, $|\lambda\rangle$ and $|\lambda'\rangle$ are the states of the system before and after scattering, $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$ those of the scattered particle (plane waves). P_λ is the probability of the initial state, $\hat{V}$ is the interaction between scattered particle and system. Finally $\delta(\hbar\omega + E_\lambda - E_{\lambda'})$ expresses energy conservation.

---

[2] The reader may ask why the example here and in Fig. 7 are taken from more than 40 years old literature. The reason is that such simple system as monatomic liquids are not studied anymore and there is no incentive by the scientific reward system to repeat the experiments using the currently available better experimental conditions.
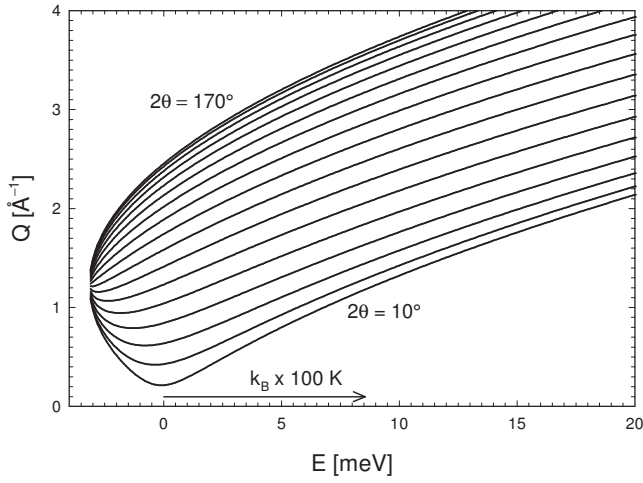
Fig. 6: Scattering vectors Q accessed by a neutron scattering experiment with the detector at scattering angles $2\theta = 10\ldots170°$ vs. the energy transfer $\hbar\omega$ (incident wavelength $\lambda = 0.51$ nm). For comparison the thermal energy $k_\mathrm{B}T$ corresponding to 100 K is indicated by an arrow.

For different probes (neutrons, light, electrons etc.) and interaction mechanisms (e.g. nuclear, magnetic) $\hat{V}$ will be different and there may be additional variables characterising the state of the scattered particle (spin, polarisation). Surprisingly, the most simple result emerges from nuclear neutron scattering mainly because the short-ranged nuclear potential can be replaced by the *Fermi pseudopotential*

$$\hat{V} \propto \sum_{j=1}^{N} b_j \delta(\mathbf{r} - \hat{\mathbf{r}}_j) \tag{29}$$

which is obviously the quantum-mechanical equivalent of the scattering length density (8). With this potential the inner matrix element can easily be calculated with the plane wave expressions for $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$:

$$\frac{\partial^2\sigma}{\partial\Omega\partial E'} = \frac{k'}{k} \sum_\lambda P_\lambda \sum_{\lambda'} \left| \sum_j \langle\lambda'|b_j \exp(\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_j)|\lambda\rangle \right|^2 \delta\left(\hbar\omega + E_\lambda - E_{\lambda'}\right). \tag{30}$$

Note that this expression neglects the spin of the neutron as well as that of the scattering system. Therefore, it is only valid for the situation of a polarised neutron beam in combination with spinless (or spin-polarised) sample nuclei.

For comparison, the expression for inelastic x-ray scattering is [14]:

$$\left.\frac{\partial^2\sigma}{\partial\Omega\partial E'}\right|_{\alpha\to\beta} \propto \frac{k'}{k} |\boldsymbol{\epsilon}_\alpha{}^* \cdot \boldsymbol{\epsilon}_\beta|^2 \sum_\lambda P_\lambda \sum_{\lambda'} \left| \sum_j \langle\lambda'|b_j \exp(\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_j)|\lambda\rangle \right|^2 \delta\left(\hbar\omega + E_\lambda - E_{\lambda'}\right).$$
$$\tag{31}$$

It can be seen that the formula is complicated by the polarisation-dependent term $|\boldsymbol{\epsilon}_\alpha{}^* \cdot \boldsymbol{\epsilon}_\beta|^2$ but otherwise contains the same 'core terms' as (30). The same is true for all scattering probes.

Therefore, we will continue with the simplest formulation in nuclear neutron scattering from spin zero nuclei.

3.1 Scattering functions

In order to relate expression (28) to a correlation function one starts with an integral representation of the delta function:

$$\delta\left(\hbar\omega + E_\lambda - E_{\lambda'}\right) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dt \exp\left(-i\left(\omega + \frac{E_\lambda - E_{\lambda'}}{\hbar}\right)t\right) \tag{32}$$

which results from the fact that the delta function is the Fourier transform of a constant one. With this expression the matrix element in equation (28) can be written as a Fourier transform in time:

$$\left|\sum_j \langle\lambda'|b_j\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_j)|\lambda\rangle\right|^2 \delta\left(\hbar\omega + E_\lambda - E_{\lambda'}\right)$$

$$= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dt \exp(-i\omega t)\exp\left(-i\frac{E_\lambda}{\hbar}t\right)\exp\left(-i\frac{E_{\lambda'}}{\hbar}t\right)$$

$$\sum_j b_j\langle\lambda'|\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_j)|\lambda\rangle\sum_k b_k^*\langle\lambda|\exp(-i\mathbf{Q}\cdot\hat{\mathbf{r}}_k)|\lambda'\rangle$$

$$= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dt \exp(-i\omega t)\sum_{j,k} b_j b_k^*\langle\lambda|\exp(-i\mathbf{Q}\cdot\hat{\mathbf{r}}_k)|\lambda'\rangle$$

$$\langle\lambda'|\exp(iE_{\lambda'}t/\hbar)\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_j)\exp(-iE_\lambda t/\hbar)|\lambda\rangle \tag{33}$$

If $\hat{\mathcal{H}}$ is the Hamiltonian of the scattering system, the fact that $|\lambda\rangle$ are energy eigenstates is expressed by

$$\hat{\mathcal{H}}|\lambda\rangle = E_\lambda|\lambda\rangle. \tag{34}$$

Iterating this equation n times yields:

$$\hat{\mathcal{H}}^n|\lambda\rangle = E_\lambda{}^n|\lambda\rangle. \tag{35}$$

By expanding the exponential into a power series one finally obtains from this relation

$$\exp(i\hat{\mathcal{H}}t/\hbar)|\lambda\rangle = \exp(iE_\lambda t/\hbar)|\lambda\rangle. \tag{36}$$

With this result and the analogous one for λ' it is possible to replace the eigenvalues E_λ in (33) by the Hamiltonian $\hat{\mathcal{H}}$:

$$\ldots \langle\lambda'|\exp(i\hat{\mathcal{H}}t/\hbar)\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_i)\exp(-i\hat{\mathcal{H}}t/\hbar)|\lambda\rangle. \tag{37}$$

In the picture of time dependent Heisenberg operators the application of the operator $\exp(i\hat{\mathcal{H}}t/\hbar)$ and its conjugate just means a propagation by time t:

$$\exp\left(i\mathbf{Q}\cdot\hat{\mathbf{r}}_i(t)\right) = \exp(i\hat{\mathcal{H}}t/\hbar)\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_i(0))\exp(-i\hat{\mathcal{H}}t/\hbar) \tag{38}$$

where we can arbitrarily set $\hat{\mathbf{r}}_i = \hat{\mathbf{r}}_i(0)$ because of time-translation invariance. Using this result the final expression for the double differential cross section is obtained:

$$\frac{\partial^2\sigma}{\partial\Omega\partial E'} = \frac{k'}{k}\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dt\exp(-i\omega t)$$

$$\sum_{\lambda}P_{\lambda}\sum_{j,k}b_jb_k^*\langle\lambda\,|\exp(-i\mathbf{Q}\cdot\hat{\mathbf{r}}_k(0))\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_j(t))|\,\lambda\rangle\ . \qquad (39)$$

(Here, the sum over λ' vanishes because of the completeness relation $\sum_{\lambda'}|\lambda'\rangle\langle\lambda'| = 1$.) In addition the initial states of the scattering system are averaged weighted with the probability of their occurrence P_λ. The latter is given by the Boltzmann distribution

$$P_\lambda = \frac{1}{Z}\exp\left(-E_\lambda/k_{\mathrm{B}}T\right)\ \text{with}\ Z = \sum_{\lambda}\exp\left(-E_\lambda/k_{\mathrm{B}}T\right)\ . \qquad (40)$$

As usual the thermal average is denoted by angular brackets, $\langle\ldots\rangle$. We now assume that the scatterers are identical with respect to the interactions ('chemically' identical) but may have different scattering lengths b_j randomly distributed over the scatterers. In this case we have to do another averaging over all distributions of scattering lengths which is written as overline: $\overline{\ldots}$. Keeping in mind that for equal indices $b_ib_i^* = |b_i|^2$ has to be averaged while for unequal indices the scattering lengths itself will be averaged we end up with the usual separation into incoherent and coherent part:

$$\frac{\partial^2\sigma}{\partial\Omega\partial E'} = \sum_{\lambda}P_\lambda$$

$$\frac{k'}{k}\frac{\overline{|b|^2}-|\bar{b}|^2}{2\pi\hbar}\int_{-\infty}^{\infty}dt\exp(-i\omega t)\sum_{j}\langle\lambda\,|\exp(-i\mathbf{Q}\cdot\hat{\mathbf{r}}_j(0))\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_j(t))|\,\lambda\rangle$$

$$+\frac{k'}{k}\frac{|\bar{b}|^2}{2\pi\hbar}\int_{-\infty}^{\infty}dt\exp(-i\omega t)\sum_{j,k}\langle\lambda\,|\exp(-i\mathbf{Q}\cdot\hat{\mathbf{r}}_j(0))\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_k(t))|\,\lambda\rangle\ . \qquad (41)$$

The first term is the incoherent scattering. It involves the coordinate vector operators of the *same* atom at different times. The second, the coherent term correlates also *different* atoms at different times. The material dependent parts are now defined as the *scattering functions*[3]

$$S_{\mathrm{coh}}(Q,\omega)\ \equiv\ \frac{1}{2\pi N}\int_{-\infty}^{\infty}dt\exp(-i\omega t)\sum_{j,k=1}^{N}\langle\exp(-i\mathbf{Q}\cdot\hat{\mathbf{r}}_j(0))\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_k(t))\rangle,\quad (42)$$

$$S_{\mathrm{inc}}(Q,\omega)\ \equiv\ \frac{1}{2\pi N}\int_{-\infty}^{\infty}dt\exp(-i\omega t)\sum_{j=1}^{N}\langle\exp(-i\mathbf{Q}\cdot\hat{\mathbf{r}}_j(0))\exp(i\mathbf{Q}\cdot\hat{\mathbf{r}}_j(t))\rangle.\quad (43)$$

The former is called the *coherent* and the latter the *incoherent* scattering function. In terms of the scattering functions the double differential cross section can be written as

$$\frac{\partial^2\sigma}{\partial\Omega\partial E'} = \frac{1}{\hbar}\frac{k'}{k}N\left(\left(\overline{|b|^2}-|\bar{b}|^2\right)S_{\mathrm{inc}}(\mathbf{Q},\omega)+|\bar{b}|^2S_{\mathrm{coh}}(\mathbf{Q},\omega)\right)\ . \qquad (44)$$

---

[3] The definition of the scattering functions in literature usually differs by a factor $1/\hbar$ which is here included in equation (44). The difference is that here $S(Q,\omega)$ is a density in frequency (with unit [time]) while in the literature it is taken as a density in energy. As pointed out by J. Wuttke, the literature definition leads to an unnecessary complication of the sum rules in subsection 3.2.

One can see that the incoherent term is weighted by the variance of scattering lengths. It did not show up in the considerations of section 2 because the assumption of equal scattering length excluded it. A treatment of the static case allowing a variance of the scattering length results in

$$\frac{d\sigma}{d\Omega} = N|\bar{b}|^2 S(\mathbf{Q}) + N\left(\overline{|b|^2} - |\bar{b}|^2\right).$$

(45)

Thus, the incoherent term exists also in scattering from a static system but it only constitutes a flat background.

The most common situation where incoherent scattering arises is that of neutron scattering. Because chemically identical atoms may be different isotopes a random variation of the scattering length is more the rule than the exception. It is also possible to 'smuggle in' the neglected dependence of neutron scattering on the spin orientation at that point. For neutron scattering with an unpolarised beam and without polarisation analysis, the random orientation of spins leads to a de-facto randomness of the scattering lengths having the same effect as a variation of isotopes.

There is often a discussion whether incoherent scattering may occur in light scattering. This is often demonstrated from polydisperse colloids. Because the form factor of a colloidal particle depends on its size, this can be treated like a variation of the scattering length for nuclei in neutron scattering. But in the opinion of the author this does not constitute a case of incoherent scattering[4] because particles of different size also have different interaction, they are 'chemically different'. The only case where one can for certain speak of incoherent light scattering is that of particles which are made distinguishable by cores with different index of refraction [15].

As an example, Fig. 7 shows the coherent and incoherent scattering function of liquid Argon at 85 K. For the coherent part one can immediately see the similarity of the envelope in the $\hbar\omega = 0$ plane to $S(Q)$ shown in Fig. 5. This is a consequence of equation (49) although strictly speaking not the peak height but the are under the peak represents $S(Q)$.

3.2   Intermediate scattering functions

In some cases it is interesting to consider the part of expression (42) before the time-frequency Fourier transform, called *coherent intermediate scattering function*:

$$I_{\text{coh}}(\mathbf{Q}, t) = \frac{1}{N} \sum_{j,k=1}^{N} \langle \exp(-i\mathbf{Q} \cdot \hat{\mathbf{r}}_j(0)) \exp(i\mathbf{Q} \cdot \hat{\mathbf{r}}_k(t)) \rangle.$$

(46)

Its value for $t = 0$ expresses the correlation between atoms *at equal times*. On one hand the integral theorem of Fourier transform tells that this is identical to the integral of the scattering function over all energy transfers:

$$I_{\text{coh}}(\mathbf{Q}, 0) = \int_{-\infty}^{\infty} S_{\text{coh}}(\mathbf{Q}, \omega) d\omega.$$

(47)

---

[4] . . . in the sense of this lecture. Often the term "incoherent scattering" is used in the light scattering community as is "diffuse scattering" by neutron scatterers. In that sense, incoherent scattering would trivially exist in light scattering, except for ordered systems as opal or butterfly wings.
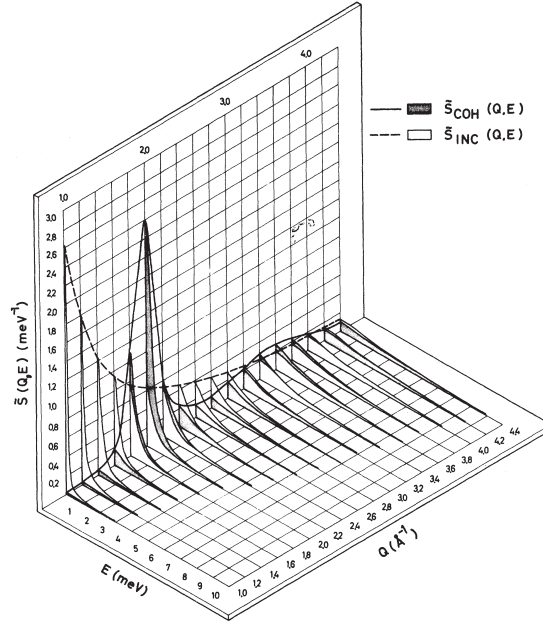
Fig. 7: Coherent and incoherent scattering function of liquid Argon at 85 K [16]. The definition of the scattering function used by the authors is slightly different. In terms of this text, the plotted quantity is $S(Q, E/\hbar)/\hbar$.

On the other hand this is just the (static) structure factor (10) because at equal t the position operators commute and the exponentials can be merged:

$$I_{\text{coh}}(\mathbf{Q}, 0) = \frac{1}{N} \sum_{j,k=1}^{N} \left\langle e^{i\mathbf{Q}\cdot(\mathbf{r}_k - \mathbf{r}_j)} \right\rangle = S(\mathbf{Q}) \,. \tag{48}$$

Combining the two expressions for the intermediate coherent scattering function yields:

$$S(\mathbf{Q}) = \int_{-\infty}^{\infty} S_{\text{coh}}(\mathbf{Q}, \omega) \mathrm{d}\omega \,. \tag{49}$$

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.

---

[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 (49): (1) On the instrument the integral is taken along a curve of constant 2θ in Fig. 6 while constant Q would correspond to a horizontal line. (2) The double differential cross-section (44) contains a factor k'/k which depends on ω via (26). (3) The detector may have an efficiency depending on wavelength which will introduce another ω-dependent weight in the experimental integration. These effects can be taken into account in the so-called Placzek corrections [13, 17].

Similarly the *incoherent intermediate scattering function* is

$$I_{\text{inc}}(\mathbf{Q}, t) = \frac{1}{N} \sum_{j=1}^{N} \langle \exp(-i\mathbf{Q} \cdot \hat{\mathbf{r}}_j(0)) \exp(i\mathbf{Q} \cdot \hat{\mathbf{r}}_j(t)) \rangle \tag{50}$$

with the sum rule

$$I_{\text{inc}}(\mathbf{Q}, 0) = \frac{1}{N} \sum_{j=1}^{N} \langle e^{i\mathbf{Q} \cdot (\mathbf{r}_j - \mathbf{r}_j)} \rangle = 1 = \int_{-\infty}^{\infty} S_{\text{inc}}(\mathbf{Q}, \omega) d\omega. \tag{51}$$

Thus, the *incoherent* intermediate scattering function is normalised to one for each Q. Note that this result is independent of the actual structure of the sample. Also it is the same result as the high Q limit $S(Q) \to 1$ (23). This is a consequence of the more general fact that coherent and incoherent scattering become indistinguishable for large Q.

3.3 Van Hove correlation functions

As in the static situation, the scattering law can be traced back to distance distribution functions, the *van Hove correlation functions* [18]. These can be derived by (inverse) Fourier transform of the intermediate scattering functions back into real space. In the coherent case:

$$G(\mathbf{r}, t) = \left(\frac{1}{2\pi}\right)^3 \int d^3Q \exp\left(-i\mathbf{Q} \cdot \mathbf{r}\right) \frac{1}{N} \sum_{j,k=1}^{N} \langle \exp\left(-i\mathbf{Q} \cdot \hat{\mathbf{r}}_j(0)\right) \exp\left(i\mathbf{Q} \cdot \hat{\mathbf{r}}_k(t)\right) \rangle. \tag{52}$$

The derivation of the relation between the coherent dynamical structure factor $S_{\text{coh}}(Q, \omega)$ and the generalized pair correlation function requires a strict quantum mechanical calculation. This problem results from the fact that the coordinate vector operators commute only at identical times. Therefore, in all algebraic manipulations the order of $\hat{\mathbf{r}}_j(0)$ and $\hat{\mathbf{r}}_j(t)$ must not be interchanged.

To begin, one writes the operator $\exp\left(-i\mathbf{Q} \cdot \hat{\mathbf{r}}_i(0)\right)$ as the Fourier transform of the delta function:

$$\exp\left(-i\mathbf{Q} \cdot \hat{\mathbf{r}}_i(0)\right) = \int d^3r' \delta\left(\mathbf{r}' - \hat{\mathbf{r}}_i(0)\right) \exp(-i\mathbf{Q} \cdot \mathbf{r}'). \tag{53}$$

Using this expression equation (52) can be rewritten as

$$\begin{aligned}
G(\mathbf{r}, t) &= \left(\frac{1}{2\pi}\right)^3 \frac{1}{N} \sum_{j,k=1}^{N} \left\langle \int d^3r' \delta\left(\mathbf{r}' - \hat{\mathbf{r}}_j(0)\right) \right. \\
&\qquad\qquad \left. \underbrace{\int d^3Q \exp\left(-i\mathbf{Q} \cdot \mathbf{r} - i\mathbf{Q} \cdot \mathbf{r}' + i\mathbf{Q} \cdot \hat{\mathbf{r}}_k(t)\right)}_{= (2\pi)^3 \delta\left(\mathbf{r} + \mathbf{r}' - \hat{\mathbf{r}}_k(t)\right)} \right\rangle \\
&= \frac{1}{N} \sum_{j,k=1}^{N} \int d^3r' \langle \delta\left(\mathbf{r} - \mathbf{r}' + \hat{\mathbf{r}}_j(0)\right) \delta\left(\mathbf{r}' - \hat{\mathbf{r}}_k(t)\right) \rangle
\end{aligned} \tag{54}$$

without changing the order of the operators at different times.

Now the particle density operator is introduced as a sum over delta functions at the particle position operators:

$$\hat{\rho}(\mathbf{r}, t) \equiv \sum_{j=1}^{N} \delta\left(\mathbf{r} - \hat{\mathbf{r}}_j(t)\right) . \tag{55}$$

With this definition the pair correlation function can be written as time-dependent density-density correlation function:

$$G(\mathbf{r}, t) = \frac{1}{N} \int \mathrm{d}^3 r' \left\langle \hat{\rho}(\mathbf{r}' - \mathbf{r}, 0) \hat{\rho}(\mathbf{r}', t) \right\rangle . \tag{56}$$

With this form of the dynamic pair correlation function the dynamical structure factor can be—analogously to equation (16)—written as the double Fourier transform of the correlator of the particle density:

$$S_{\mathrm{coh}}(\mathbf{Q}, \omega) = \frac{1}{2\pi N} \int_{-\infty}^{\infty} \mathrm{d}t \exp(-\mathrm{i}\omega t) \tag{57}$$

$$\int \mathrm{d}^3 r \int \mathrm{d}^3 r' \exp\left(\mathrm{i}\mathbf{Q} \cdot \mathbf{r}\right) \left\langle \hat{\rho}(\mathbf{r}' - \mathbf{r}, 0) \hat{\rho}(\mathbf{r}', t) \right\rangle \tag{58}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}t \exp(-\mathrm{i}\omega t) \int \mathrm{d}^3 r \exp\left(\mathrm{i}\mathbf{Q} \cdot \mathbf{r}\right) G(\mathbf{r}, t) . \tag{59}$$

Thus the scattering function is the double Fourier transform (in space and time) of the van Hove correlation function.

We now define the density operator in reciprocal space as the Fourier transform of (55):

$$\hat{\rho}_{\mathbf{Q}}(t) \equiv \sum_{j=1}^{N} \exp\left(\mathrm{i}\mathbf{Q} \cdot \hat{\mathbf{r}}_j(t)\right) \tag{60}$$

and obtain for the dynamic structure factor

$$S_{\mathrm{coh}}(\mathbf{Q}, \omega) = \frac{1}{2\pi N} \int_{-\infty}^{\infty} \mathrm{d}t \exp(-\mathrm{i}\omega t) \left\langle \hat{\rho}_{\mathbf{Q}}(0) \hat{\rho}_{-\mathbf{Q}}(t) \right\rangle . \tag{61}$$

Correspondingly, the intermediate scattering function is

$$I_{\mathrm{coh}}(\mathbf{Q}, t) = \frac{1}{N} \left\langle \hat{\rho}_{\mathbf{Q}}(0) \hat{\rho}_{-\mathbf{Q}}(t) \right\rangle \tag{62}$$

which after insertion of (60) turns out to be equivalent to (46).

Analogously, one can define the van Hove *self* correlation function by setting $k = j$ in the preceding equations, leading to

$$G_{\mathrm{s}}(\mathbf{r}, t) = \frac{1}{N} \sum_{j=1}^{N} \int \mathrm{d}^3 r' \left\langle \delta\left(\mathbf{r} - \mathbf{r}' + \hat{\mathbf{r}}_j(0)\right) \delta\left(\mathbf{r}' - \hat{\mathbf{r}}_j(t)\right) \right\rangle \tag{63}$$

as the equivalent of (54). Note that there is no equivalent of equations (61) and (62) because the definition of the density already 'mixes' all particles and the product of the density with itself inevitably contains all pair correlations.

The pair correlation function has some general properties:

1. For spatially homogeneous systems the integrand in (56) is independent of $\mathbf{r}'$ which can be arbitrarily set to the origin $\mathbf{0}$:

$$G(\mathbf{r}, t) = \frac{V}{N} \langle \hat{\rho}(-\mathbf{r}, 0)\hat{\rho}(\mathbf{0}, t) \rangle = \frac{1}{\rho_0} \langle \hat{\rho}(\mathbf{0}, 0)\hat{\rho}(\mathbf{r}, t) \rangle \,. \tag{64}$$

2. The pair correlation function has the following asymptotic behaviour: For fixed distance and $t \to \infty$ or fixed time and $r \to \infty$ the averages in equation (56) can be executed separately and in consequence

$$G(\mathbf{r}, t) \to \frac{1}{N} \int \mathrm{d}^3 r' \, \langle \hat{\rho}(\mathbf{r}' - \mathbf{r}, 0) \rangle \, \langle \hat{\rho}(\mathbf{r}', t) \rangle = \rho_0 \,. \tag{65}$$

3. For $t = 0$ the operators commute and the convolution integral of equation (56) can be carried out:

$$G(\mathbf{r}, 0) = \frac{1}{N} \sum_{j,k=1}^{N} \langle \delta \left(\mathbf{r} + \hat{\mathbf{r}}_j(0) - \hat{\mathbf{r}}_k(0) \right) \rangle \,. \tag{66}$$

Comparison of this equation with (13) or putting $t = 0$ in (64) yields the relation to the static pair correlation function:

$$G(\mathbf{r}, 0) = \frac{1}{\rho_0} \langle \rho(\mathbf{0})\rho(\mathbf{r}) \rangle \tag{67}$$

or in terms of the commonly defined $g(r)$:

$$G(\mathbf{r}, 0) = \delta(\mathbf{r}) + \rho_0 g(\mathbf{r}) \,. \tag{68}$$

This equation expresses again the fact that the diffraction experiment ($g(r)$) gives an average snapshot picture ($G(r, 0)$) of the sample.

3.4   Approximation by trajectories

In an often-used approximation the operators in the preceding derivation are replaced by variables. Especially the position operators reduce to trajectories of particles, $\mathbf{r}_j(t)$ [19]. Then the integrals of equations (54) and (63) can be carried out and yield

$$\tilde{G}(\mathbf{r}, t) \;=\; \frac{1}{N} \sum_{j,k=1}^{N} \delta(\mathbf{r} - \mathbf{r}_k(t) + \mathbf{r}_j(0)) \text{ and} \tag{69}$$

$$\tilde{G}_{\mathrm{s}}(\mathbf{r}, t) \;=\; \frac{1}{N} \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{r}_j(t) + \mathbf{r}_j(0)) \,, \tag{70}$$

respectively. The former equation expresses the probability to find *any* particle at a time t in a distance $\mathbf{r}$ from another at time 0. The latter equation denotes this probability for the *same* particle. It therefore depends only on the particle's displacement during a time interval $\Delta \mathbf{r}_j(t) = \mathbf{r}_j(t) - \mathbf{r}_j(0)$ leading to a simple expression for the intermediate incoherent scattering function:

$$\tilde{I}_{\mathrm{inc}}(\mathbf{Q}, t) = \frac{1}{N} \sum_{j=1}^{N} \langle \exp\left(-\mathrm{i}\mathbf{Q} \cdot \Delta \mathbf{r}_j(t)\right) \rangle \,. \tag{71}$$

In certain cases (if $\tilde{G}_s(\mathbf{r}, t)$ is a Gaussian in space) this expression can be further simplified using the "Gaussian approximation":

$$I_{\text{inc}}^{\text{Gauss}}(Q, t) = \exp\left(-\frac{1}{6}Q^2\langle\Delta r^2(t)\rangle\right). \tag{72}$$

Here $\langle\Delta r^2(t)\rangle$ is the average mean squared displacement which often follows simple laws, e.g. $\langle\Delta r^2(t)\rangle = 6Dt$ for simple diffusion. Because one of the prerequisites of the Gaussian approximation is that all particles move statistically in the same way (dynamic homogeneity) the particle average and the index i vanish. An analogous expression can be derived for the coherent scattering.

The most prominent difference between the full quantum-mechanical treatment and the approximation based on classical trajectories is that the real scattering functions are asymmetric with respect to the energy transfer $\hbar\omega$,

$$S_{[\text{coh}|\text{inc}]}(\mathbf{Q}, -\omega) = \exp\left(\frac{\hbar\omega}{k_{\text{B}}T}\right) S_{[\text{coh}|\text{inc}]}(\mathbf{Q}, \omega), \tag{73}$$

while those from the trajectory-based approximation are symmetric:

$$\tilde{S}_{[\text{coh}|\text{inc}]}(\mathbf{Q}, -\omega) = \tilde{S}_{[\text{coh}|\text{inc}]}(\mathbf{Q}, \omega). \tag{74}$$

Equation (73) expresses the fact that the probability for a neutron to be scattered with energy loss is always higher than the probability to be scattered with energy gain (Fig. 8). This can be understood as a *detailed balance* factor: In equilibrium, the probability for the scattering system to be in the lower energy state is higher by the factor $\exp(\hbar\omega/k_{\text{B}}T)$. Therefore the probability of scattering into a state with higher energy is more probable by the same factor than scattering into the lower energy state. The effect of the asymmetry will be noticeable for low temperatures and high energy transfers unless $T \gg \hbar\omega/k_{\text{B}}$ [6]. Considering that for room temperature $k_{\text{B}} \cdot 300\,\text{K} \approx 26\,\text{meV}$ it is clear that the condition for a classical treatment is often not fulfilled even in neutron scattering.

From the asymmetry of $S(\mathbf{Q}, \omega)$ follows that the intermediate scattering function $I(\mathbf{Q}, t)$ and the van Hove correlation function $G(\mathbf{r}, t)$ are *complex*. This is surprising but not unphysical because they are no observable quantities[7] as $S(\mathbf{Q}, \omega)$ which still has to be real. By inversion of the Fourier transform in time it follows for the intermediate scattering functions and the van Hove correlation functions:

$$I_{[\text{coh}|\text{inc}]}(\mathbf{Q}, -t) = I_{[\text{coh}|\text{inc}]}\left(\mathbf{Q}, t - \frac{i\hbar}{k_{\text{B}}T}\right) \tag{75}$$

$$G_{[\text{s}]}(\mathbf{r}, -t) = G_{[\text{s}]}\left(\mathbf{r}, t - \frac{i\hbar}{k_{\text{B}}T}\right) \tag{76}$$

---

[6] There is another condition for the validity of the approximation concerning the momentum transfer: $Q \ll \sqrt{2Mk_{\text{B}}T}/\hbar$, where M is the mass of the scattering particle. The meaning of this condition is that the De Broglie wavelength of the scatterer should be sufficiently small compared to the length scale of the scattering experiment $1/Q$.

[7] Note that the neutron spin echo spectrometer, which is said to measure the intermediate scattering function, performs an inverse *cosine* Fourier transform instead of an exponential one. Therefore it actually measures only the real part of the intermediate scattering function.
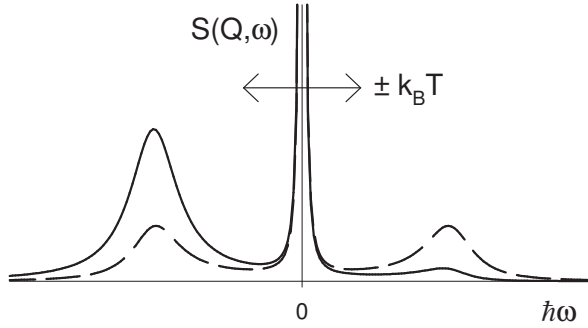
Fig. 8: Example of the asymmetry of the scattering function due to the detailed balance factor. The (neutron) energy loss side is enhanced compared to the energy gain side for the correct quantum-mechanical result (continuous curve). The classical result (dashed curve) in contrast is mirror symmetric. The arrow indicates for comparison the thermal energy $\pm k_{\mathrm{B}}T$.

The latter relations follow from the quantum-mechanical peculiarity that time-dependent operators in a correlation function may not be interchanged but $\langle \hat{A}(0)\hat{B}(t)\rangle = \langle \hat{B}(t-\mathrm{i}\hbar/k_{\mathrm{B}}T)\hat{A}(0)\rangle$ and $\langle \hat{A}(0)\hat{B}(t)\rangle^* = \langle \hat{B}^\dagger(t)\hat{A}^\dagger(0)\rangle$.

In many cases, the classical particle trajectories are much easier to derive (e.g. by molecular dynamics simulation) than their quantum-mechanical counterparts. If a full quantum-mechanical result is not available, at least a better approximation to the correct scattering functions can be derived [20] which fulfills the detailed balance relation (73). It is based on simply multiplying the square root of the prefactor in equation (73) to the neutron energy loss side and dividing it out of the energy gain side:

$$S_{[\mathrm{coh|inc}]}(\mathbf{Q},\omega) \approx \exp\left(-\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right)\tilde{S}_{[\mathrm{coh|inc}]}(\mathbf{Q},\omega)\,. \tag{77}$$

3.5   Example: ideal gas

Although this is possibly the simplest system one can imagine, the calculation is already rather intricate. The simplicity of the example is mainly based on the definition of an ideal gas, that particles do not interact. This implies that there are no correlations between different particles and the $j \neq k$ terms vanish in expressions as (42). Therefore, the incoherent and coherent quantities are equal for the ideal gas:

$$S_{\mathrm{coh}}(Q,\omega) = S_{\mathrm{inc}}(Q,\omega)\,,\ I_{\mathrm{coh}}(Q,t) = I_{\mathrm{inc}}(Q,t)\,,\ G(r,t) = G_{\mathrm{s}}(r,t)\,. \tag{78}$$

(Because the ideal gas is isotropic, r and Q are scalars.) In addition, all particles behave statistically in the same way and therefore the averages over all particles can be replaced by a single representative particle:

$$\begin{aligned}
I(Q,t) &= \frac{1}{N}\sum_{j=1}^{N}\langle\exp\left(-\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_j(0)\right)\exp\left(\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_j(t)\right)\rangle \\
&= \langle\exp\left(-\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_1(0)\right)\exp\left(\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_1(t)\right)\rangle\,.
\end{aligned} \tag{79}$$

Writing out the Heisenberg operator $\hat{\mathbf{r}}_1(t)$ in its explicit form one obtains

$$I(Q,t) = \left\langle \exp\left(-\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_1(0)\right) \exp\left(\mathrm{i}\hat{\mathcal{H}}_1 t/\hbar\right) \exp\left(\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_1(t)\right) \exp\left(-\mathrm{i}\hat{\mathcal{H}}_1 t/\hbar\right) \right\rangle \tag{80}$$

where $\hat{\mathcal{H}}_1$ is the Hamiltonian of the representative particle which is simply the square of the momentum operator divided by twice the scattering particle's mass:

$$\hat{\mathcal{H}}_1 = \frac{1}{2M}\hat{\mathbf{p}}^2\,. \tag{81}$$

Taking into account that the operator $\exp\left(-\mathrm{i}\mathbf{q}\cdot\hat{\mathbf{r}}_1\right)$ shifts the momentum

$$\exp\left(-\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_1\right)\hat{\mathbf{p}}\exp\left(\mathrm{i}\mathbf{Q}\cdot\hat{\mathbf{r}}_1\right) = \hat{\mathbf{p}} + \hbar\mathbf{Q} \tag{82}$$

one gets

$$I(Q,t) = \left\langle \exp\left(\mathrm{i}\hat{\mathcal{H}}_1' t/\hbar\right) \exp\left(-\mathrm{i}\hat{\mathcal{H}}_1 t/\hbar\right) \right\rangle \tag{83}$$

where $\hat{\mathcal{H}}_1'$ denotes the single-particle Hamiltonian with shifted momentum:

$$\hat{\mathcal{H}}_1' = \frac{1}{2M}\left(\hat{\mathbf{p}} + \hbar\mathbf{Q}\right)^2 = \hat{\mathcal{H}}_1 + \frac{\hbar}{M}\mathbf{Q}\cdot\hat{\mathbf{p}} + \frac{\hbar^2 Q^2}{2M}\,. \tag{84}$$

Insertion of (84) into (83) yields:

$$I(Q,t) = \exp\left(\frac{\mathrm{i}\hbar t Q^2}{2M}\right)\left\langle \exp\left(\frac{\mathrm{i}t\mathbf{Q}\cdot\hat{\mathbf{p}}}{M}\right)\right\rangle \tag{85}$$

The thermodynamic average in this expression can be calculated with the equilibrium distribution of momenta. Here, a Boltzmann distribution is assumed:

$$\left\langle \exp\left(\mathrm{i}t\mathbf{Q}\cdot\hat{\mathbf{p}}\right)\right\rangle = \frac{\int \mathrm{d}^3p\,\exp\left(-\frac{p^2}{2Mk_{\mathrm{B}}T}\right)\exp\left(\mathrm{i}t\mathbf{Q}\cdot\mathbf{p}\right)}{\int \mathrm{d}^3p\,\exp\left(-\frac{p^2}{2Mk_{\mathrm{B}}T}\right)} = \exp\left(-Q^2 t^2 k_{\mathrm{B}}T/2M\right)\,. \tag{86}$$

From (85) and (86) the intermediate scattering function is finally obtained:

$$I(Q,t) = \exp\left(-\frac{Q^2}{2M}\left(k_{\mathrm{B}}Tt^2 + \mathrm{i}\hbar t\right)\right)\,. \tag{87}$$

The Fourier transform of (87) can be calculated yielding

$$S(Q,\omega) = \sqrt{\frac{M}{2\pi k_{\mathrm{B}}TQ^2}}\exp\left(-\frac{M}{2\pi k_{\mathrm{B}}TQ^2}\left(\omega + \frac{\hbar Q^2}{2M}\right)^2\right)\,. \tag{88}$$

The scattering function is a Gaussian distribution around $-\hbar Q^2/2M$ (Fig. 9) showing that on average the neutrons lose the 'recoil energy' $E_{\mathrm{r}} = \hbar^2 Q^2/2M$ during the scattering event. The width of the Gaussian, $\sqrt{k_{\mathrm{B}}T/M}\hbar Q$ increases with temperature and scattering 'vector' Q.

The van Hove correlation function can be calculated immediately by inverse Fourier transform from (87) because $I(Q,t)$ is also a Gaussian in Q:

$$G(r,t) = \left(\frac{M}{2\pi k_{\mathrm{B}}Tt(t + \mathrm{i}\hbar/k_{\mathrm{B}}T)}\right)^{3/2}\exp\left(-\frac{Mr^2}{2k_{\mathrm{B}}Tt(t + \mathrm{i}\hbar/k_{\mathrm{B}}T)}\right)\,. \tag{89}$$
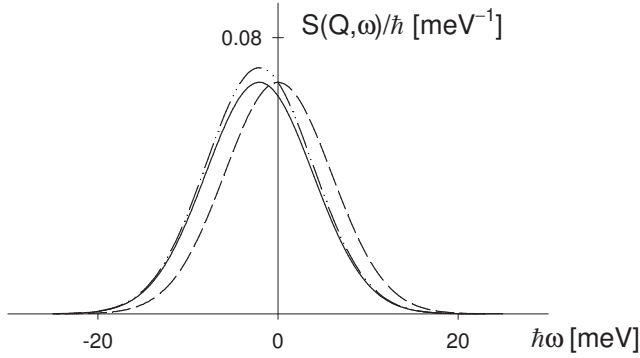
Fig. 9: Scattering from an ideal gas calculated with the parameters of helium at 100 K for $Q = 2$ Å^{-1}. The continuous curve shows the correct quantum-mechanical result, the dashed curve that of a classical calculation. The dot-dashed curve represents the approximation resulting from applying (77) to the classical result.

To demonstrate the differences arising from a classical treatment of the ideal gas, the intermediate scattering function in that approximation will be derived too. Because of the identity of the particles (50) reduces to

$$\tilde{I}(Q,t) = \langle \exp\left(i\mathbf{Q} \cdot (\mathbf{r}_1(t) - \mathbf{r}_1(0))\right) \rangle \ . \tag{90}$$

In an ideal gas the trajectory of a particle is $\mathbf{r}_1(t) = \mathbf{r}_1(0) + \mathbf{v}t$ yielding:

$$\tilde{I}(Q,t) = \langle \exp\left(i\mathbf{Q} \cdot \mathbf{v}t\right) \rangle \ . \tag{91}$$

The thermodynamic average can be calculated using the Maxwell distribution of velocities:

$$P(v) = \sqrt{\frac{2\pi k_{\mathrm{B}}T}{M}} \exp\left(-\frac{Mv^2}{2k_{\mathrm{B}}T}\right) \tag{92}$$

resulting in

$$\tilde{I}(Q,t) = \exp\left(-\frac{k_{\mathrm{B}}TQ^2t^2}{2M}\right) \ . \tag{93}$$

In contrast to the quantum-mechanical $I(Q,t)$ (87), this quantity is real and (accidentally for this example) $= |I(Q,t)|$. Because $\tilde{I}(Q,t)$ is a Gaussian in both Q and t, the Fourier transforms can be easily calculated and the scattering function is

$$\tilde{S}(Q,\omega) = \sqrt{\frac{M}{2\pi k_{\mathrm{B}}TQ^2}} \exp\left(-\frac{M\omega^2}{2\pi k_{\mathrm{B}}TQ^2}\right) \ , \tag{94}$$

a Gaussian of the same width as the exact result (88) but centred around zero energy transfer (Fig. 9). Thus, classically neutrons are scattered with no average energy transfer. The classical van Hove correlation function

$$\tilde{G}(r,t) = \left(\frac{m_{\mathrm{sc}}}{2\pi k_{\mathrm{B}}Tt^2}\right)^{3/2} \exp\left(-\frac{m_{\mathrm{sc}}r^2}{2k_{\mathrm{B}}Tt^2}\right) \tag{95}$$

conveys the meaning that at any time the distribution of distances travelled by particles of an ideal gas is Gaussian with a width increasing linearly in time.

Finally, the improved approximation of the correct results from the classical using (77) is

$$S(Q,\omega) \approx \sqrt{\frac{M}{2\pi k_B T Q^2}} \exp\left(-\frac{M}{2\pi k_B T Q^2}\left(\omega^2 + \frac{\hbar Q^2}{M}\omega\right)\right) . \tag{96}$$

It can be seen (Fig. 9) that the approximation captures the shift by the recoil energy correctly but the normalisation is wrong: $\int \tilde{S}(Q,\omega)\mathrm{d}\omega = \exp(\hbar^2 Q^2/8 M k_B T) \neq 1$. Nevertheless, equation (96) differs from (88) only in the order $\hbar^2$, which makes it better approximations than the purely classical result deviating already in $\hbar^1$ terms.

References

[1] G. E. Bacon, *Neutron Diffraction* (Clarendon Press, Oxford, 1975).

[2] G. L. Squires, *Introduction to the theory of thermal neutron scattering* (Cambridge University Press, Cambridge, 1978).

[3] S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter* (Clarendon Press, Oxford, 1984).

[4] W. Brown, *Light Scattering: Principles and Development* (Oxford University Press, Oxford, 1996).

[5] B. Chu, *Laser Light Scattering* (Academic Press, 1974).

[6] H. Z. Cummins, E. R. Pike (eds.), *Photon Correlation Spectroscopy and Velocimetry* (Plenum Press, New York, 1977).

[7] B. J. Berne, R. Pecora, *Dynamic Light Scattering* (Wiley, New York, 1976).

[8] R. Pecora, *Dynamic Light Scattering* (Plenum Press, New York, 1985).

[9] R. Hosemann, A. N. Bagchi, *Direct analysis of diffraction by matter* (North-Holland, Amsterdam, 1962).

[10] L. V. Azaroff, R. Kaplow, N. Kato, R. J. Weiss, A. J. C. Wilson, R. A. Young, *X-ray diffraction* (McGraw-Hill, 1974).

[11] O. Kratky, O. Glatter (eds.), *Small Angle X-Ray Scattering* (Academic Press, London, 1982).

[12] R. Klein, *Interacting Colloidal Suspensions*, in P. Lindner, T. Zemb (eds.), *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, pp. 351–380 (Elsevier, Amsterdam, 2002).

[13] J. L. Yarnell, M. J. Katz, R. G. Wenzel, S. H. Koenig, Phys. Rev. A **7**, 2130 (1973).

[14] S. K. Sinha, J. Phys.: Condens. Matter **13**, 7511 (2001).

[15] P. N. Pusey, H. M. Fijnaut, A. Vrij, J. Chem. Phys **77**, 4270 (1982).

[16] K. Sköld, J. M. Rowe, G. Ostrowski, P. D. Randolph, Phys. Rev. A **6**, 1107 (1972).

[17] G. Placzek, Phys. Rev. **86**, 377 (1952).

[18] L. van Hove, Phys. Rev. **95**, 249 (1954).

[19] G. H. Vineyard, Phys. Rev. **110**, 999 (1958).

[20] P. Schofield, Phys. Rev. Lett. **4**, 239 (1960).

B 1   Crystal Structures and Symmetries

G. Roth

Institute of Crystallography

RWTH Aachen University

Contents

Introduction

The term "crystal" derives from the Greek κρύσταλλος , which was first used as description of ice and later - in a more general meaning - for transparent minerals with regular morphology (regular crystal faces and edges).

Crystalline solids are thermodynamically stable in contrast to amorphous solids and are characterised by a **regular three-dimensional periodic arrangement of atoms** (ions, molecules) in space. This periodic arrangement makes it possible to determine their structure (atomic positions in 3D space) by **diffraction methods**, using the crystal lattice as a three-dimensional diffraction grating.
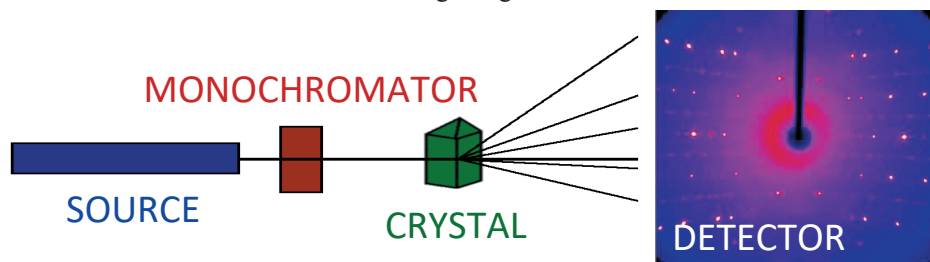
Fig. 1.1: *Sketch of a typical constant wavelength single crystal diffraction experiment. The first such experiment has been conducted by Laue et al. in 1912 (Nobel Prize in Physics 1914).*

The purpose of this chapter is to give a brief introduction into the symmetry concept underlying the description of the crystalline state.

1.1   Crystal lattices

The three-dimensional periodicity of crystals can be represented by the so-called crystal lattice. The repeat unit in form of a parallelepiped - known as the **unit cell** – is defined by 3 non-coplanar basis vectors a_1, a_2, and a_3, whose directions form the reference axes of the corresponding right-handed crystallographic coordinate system. The 6 **lattice parameters** are given as the lengths of the basis vectors $a = |a_1|$, $b = |a_2|$, $c = |a_3|$ and the angles between the basis vectors: angle $(a_1,a_2) = \gamma$, angle $(a_2,a_3) = \alpha$, angle $(a_3,a_1) = \beta$. The faces of the unit cell are named as face $(a_1,a_2) = $ C**,** face $(a_2,a_3) = $ A**,** face $(a_3,a_1) = $ B.

If the vertices of all repeat units (unit cells) are replaced by points, the result is the crystal lattice in the form of a **point lattice**. Each lattice point is given by a vector $a = ua_1+va_2+wa_3$, with u, v, w being integers. As a symmetry operation of parallel displacement, a – also known as **translation vector** – maps the atomic arrangement of the crystal (crystal structure) onto itself.
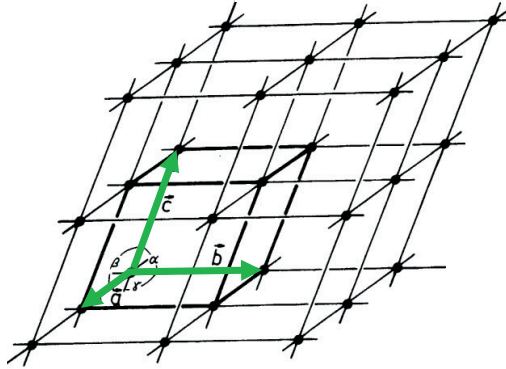
Fig. 1.2: *Notation for a unit cell (basis vectors a_1, a_2, a_3, or a, b, c) and a point lattice.*

A **lattice point** is labelled "*uvw*", according to the coefficients (integers) of the translation vector

$$a = u\,a_1 + v\,a_2 + w\,a_3 \qquad\qquad 1.1$$

from the origin to the lattice point. A **lattice direction** - given by the symbol [*uvw*] - is defined by the direction of the corresponding translation vector.

A plane passing through three lattice points is known as a **lattice plane**. Since all lattice points are equivalent (by translation symmetry) there will be infinitely many parallel planes passing through all the other points of the lattice. Such a set of equally spaced planes is known as a **set of lattice planes**. If the first plane from the origin of a set of lattice planes makes intercepts a_1/h, a_2/k, a_3/l on the axes, where h, k, l are integers, then the **Miller indices** of this set of lattice planes are (*hkl*), the three coefficients h, k, l are conventionally enclosed in parentheses.

The equation of lattice planes can be written in intercept form as

$$(hx/a_1) + (ky/a_2) + (lz/a_3) = \text{n}, \qquad\qquad 1.2$$

where n is an integer. If n = 0 the lattice plane passes through the origin; if n = 1 the plane makes intercepts a_1/h, a_2/k, a_3/l on the axes; if n = 2 the intercepts are $2a_1/h$, $2a_2/k$, $2a_3/l$; and so on.

Complementary to the crystal lattice, the so-called reciprocal lattice may be constructed, which is a useful tool for understanding the geometry of diffraction experiments. The reciprocal lattice can be thought of as the result of diffraction (of X-rays, neutrons, electrons etc.) from the crystal lattice ('direct lattice'). The points on the diffraction pattern in Fig. 1.1 (right) are actually points of the reciprocal lattice recorded during the diffraction experiment. Their nodes are indexed by the Miller-indices hkl in the same way as the nodes of the direct lattice are indexed by uvw:

$$\tau = h\,\tau_1 + k\,\tau_2 + l\,\tau_3. \qquad\qquad 1.3$$

The basis vectors τ of the reciprocal lattice can be calculated from those of the direct cell by:

$$\boldsymbol{\tau}_i = (\mathbf{a_j} \times \mathbf{a_k})/V_c, \qquad\qquad 1.4$$

where $\times$ means the cross product, and $V_c = \mathbf{a_1} \cdot (\mathbf{a_2} \times \mathbf{a_3})$ is the volume of the unit cell.

Here is a compilation of some properties of the reciprocal lattice:

• Each reciprocal lattice vector is perpendicular to two real space vectors: $\boldsymbol{\tau}_i \perp \mathbf{a_j}$ and $\mathbf{a_k}$ (for i ≠ j, k)

• The lengths of the reciprocal lattice vectors are $|\boldsymbol{\tau}_i| = 1/V_c \cdot |\mathbf{a_j}| \cdot |\mathbf{a_k}| \cdot \sin\angle(\mathbf{a_j}, \mathbf{a_k})$.

• Each point *hkl* in the reciprocal lattice refers to a set of planes (*hkl*) in real space.

• The direction of the reciprocal lattice vector $\boldsymbol{\tau}$ is normal to the (*hkl*) planes and its length is reciprocal to the interplanar spacing d_{hkl}: $|\boldsymbol{\tau}| = 1/d_{hkl}$.

• Duality principle: The reciprocal lattice of the reciprocal lattice is the direct lattice.
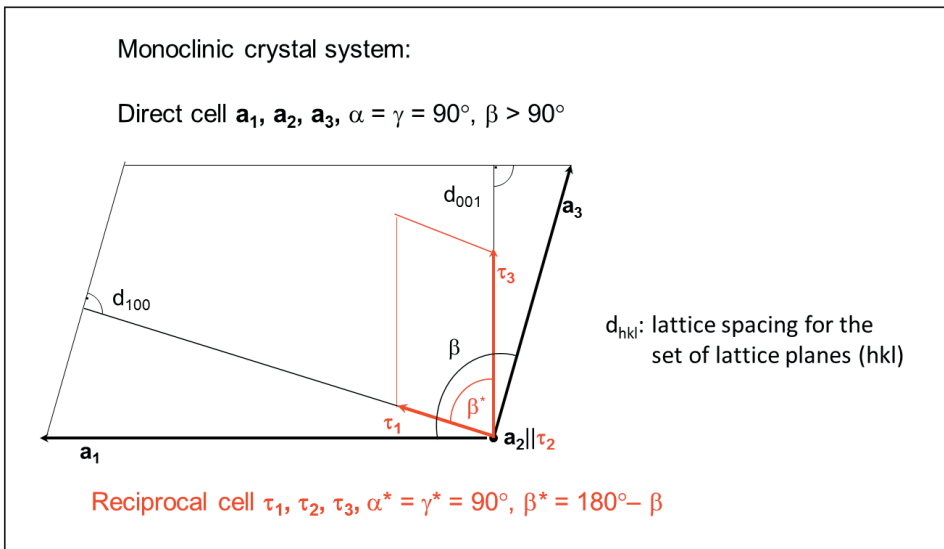
Fig. 1.3: Direct and corresponding reciprocal unit cell.

1.2 Crystallographic coordinate systems

The description of a crystal structure consists first of the choice of a unit cell as the smallest repeat unit of the crystal with its basis vectors. In this way a crystal-specific coordinate system is defined which is used to localize all the atoms in the unit cell. While - in physics and chemistry - Cartesian coordinate systems are frequently used, crystallographers often use non-orthogonal and non-orthonormal coordinate systems.

The conventional crystallographic coordinate systems are based on the symmetry of the crystals. In three dimensions there are **7** different **crystal systems** and hence 7 crystallographic coordinate systems:

Crystal system	Minimum symmetry	Conventional unit cell
triclinic	1 or $\bar{1}$	$a \neq b \neq c; \ \alpha \neq \beta \neq \gamma$
monoclinic (unique axis b)	<u>one</u> diad – 2 or m ($\parallel Y$)	$a \neq b \neq c; \ \alpha = \gamma = 90°, \ \beta > 90°$
orthorhombic	<u>three</u> mutually perpendicular diads – 2 or m ($\parallel X, Y$ and Z)	$a \neq b \neq c; \ \alpha = \beta = \gamma = 90°$
tetragonal	<u>one</u> tetrad – 4 or $\bar{4}$ ($\parallel Z$)	$a = b \neq c; \ \alpha = \beta = \gamma = 90°$
trigonal (hexagonal cell)	<u>one</u> triad – 3 or $\bar{3}$ ($\parallel Z$)	$a = b \neq c; \ \alpha = \beta = 90°, \ \gamma = 120°$
hexagonal	<u>one</u> hexad – 6 or $\bar{6}$ ($\parallel Z$)	$a = b \neq c; \ \alpha = \beta = 90°, \ \gamma = 120°$
cubic	<u>four</u> triads – 3 or $\bar{3}$ ($\parallel$ space diagonals of cube)	$a = b = c; \ \alpha = \beta = \gamma = 90°$

The choice of the origin of the coordinate system is free in principle, but for convenience it is usually chosen at a centre of symmetry (inversion centre), if present, otherwise in a point of high symmetry.

In order to complete the symmetry conventions of the coordinate systems it is necessary to add to the 7 so-called primitive unit cells of the crystal systems (primitive lattice types with only one lattice point per unit cell) 7 centred unit cells with two, three or four lattice points per unit cell (centred lattice types). These centred unit cells are consequently two, three or four times larger than the smallest repeat units of the crystals. The resulting **14 Bravais lattice types** with their centering conditions are collected in Fig. 1.4.

triclinic P

monoclinic P
monoclinic axis ∥c

monoclinic A
(0,0,0 + 0, ½, ½)

orthorhombic P

orthorhombic I
(0,0,0 + ½, ½, ½)

orthorhombic C
(0,0,0 + ½, ½,0)

orthorhombic F
(0,0,0 + ½, ½,0
½,0, ½ + 0, ½, ½)

tetragonal P

tetragonal I

hexagonal P

hexagonal/
rhombohedral R

cubic P

cubic I

cubic F
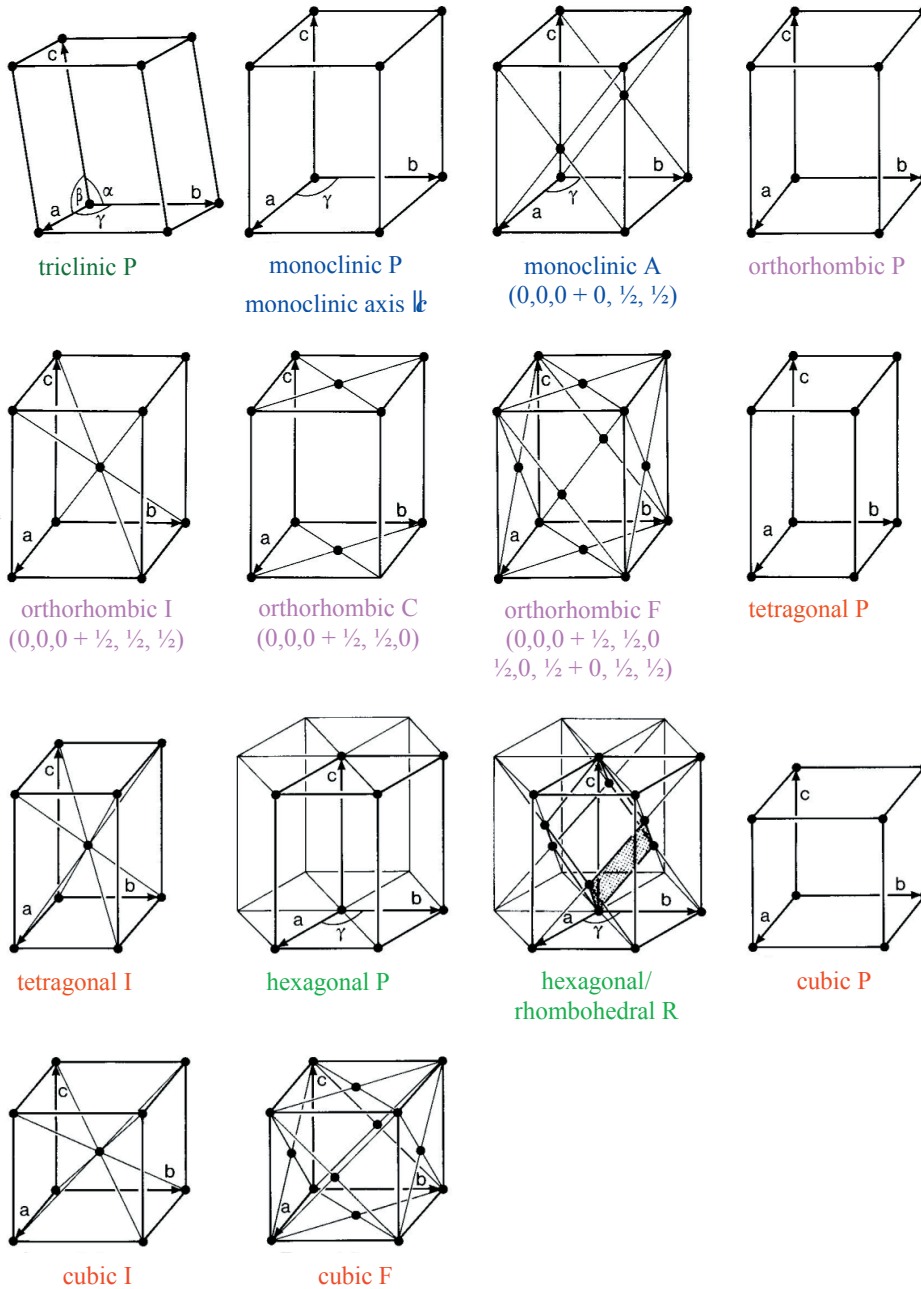
Fig. 1.4: *The 14 Bravais lattices consisting of the 7 primitive lattices P for the 7 crystal systems with only one lattice point per unit cell + the 7 centred (multiple) lattices A, B, C, I, R and F with 2, 3 and 4 lattice points per unit cell.*

1.3 Symmetry-operations and -elements

The **symmetry operations** of a crystal are isometric transformations or motions, i.e. mappings which preserve distances and, hence, also angles and volumes. An object and its transformed object superimpose in a perfect manner, they are indistinguishable.

The simplest crystallographic symmetry operation is the **translation**, which is a parallel displacement of the crystal by a translation vector a (see chapter 1.1). There is no fixed point, the entire lattice is shifted and therefore, theoretically, the crystal lattice is considered to be infinite.

Crystallographic **rotations** n around an axis by an angle $\varphi = 360°/n$ (n-fold rotations) and **rotoinversions** (combination of rotations and inversions) $\bar{n}$ are called point symmetry operations because they leave at least one point of space invariant (at least one fixed point). An important fact of crystallographic symmetry is the restriction of the rotation angles by the three-dimensional crystal lattice to $\varphi = 360°$ (n = 1), 180° (n = 2), 120° (n = 3), 90° (n = 4), 60° (n = 6). Only for these crystallographic rotations the space can be covered completely without gaps and overlaps. The rotoinversion $\bar{n} = \bar{1}$ is an **inversion** in a point, $\bar{n} = \bar{2} \equiv m$ (mirror) describes a **reflection** across a plane.

The combination of n-fold rotations with (m/n)·a translation components (m < n) $\parallel$ to the rotation axis leads to the so-called **screw rotations** n_m, e.g. 2_1, 3_2, 4_2, 6_5. These symmetry operations have no fixed points.

The combination of a reflection through a plane (glide plane) with translation components (glide vectors) of $a_1/2$, $a_2/2$, $a_3/2$, $(a_1+a_2)/2$, … $\parallel$ to this plane are known as **glide reflections** a, b, c, n, …, d. Again no fixed points exist for these symmetry operations.

In addition to the symmetry operations which represent isometric motions of an object, symmetry can also be described in (static) geometrical terms by **symmetry elements**. They form the geometrical locus, oriented in space, on which a symmetry operation is performed (line for a rotation, plane for a reflection, and point for an inversion) together with a description of this operation. Symmetry elements are mirror planes, glide planes, rotation axes, screw axes, rotoinversion axes and inversion centres. The geometrical descriptions of the crystallographic symmetry operations are illustrated in Figs. 1.5-1.7.

A symmetry operation transforms a point X with coordinates x, y, z (according to a position vector $X = xa_1 + ya_2 + za_3$) into a symmetrically equivalent point X' with coordinates x', y', z' mathematically by the linear equations

$$x' = W_{11}x + W_{12}y + W_{13}z + w_1$$
$$y' = W_{21}x + W_{22}y + W_{23}z + w_2 \qquad\qquad 1.5$$
$$z' = W_{31}x + W_{32}y + W_{33}z + w_3$$

with w_1, w_2, w_3 constituting the translational part of the symmetry operation.

Point symmetry operations

rotations	rotoinversions

1=identity

2-fold = 180°-rotation
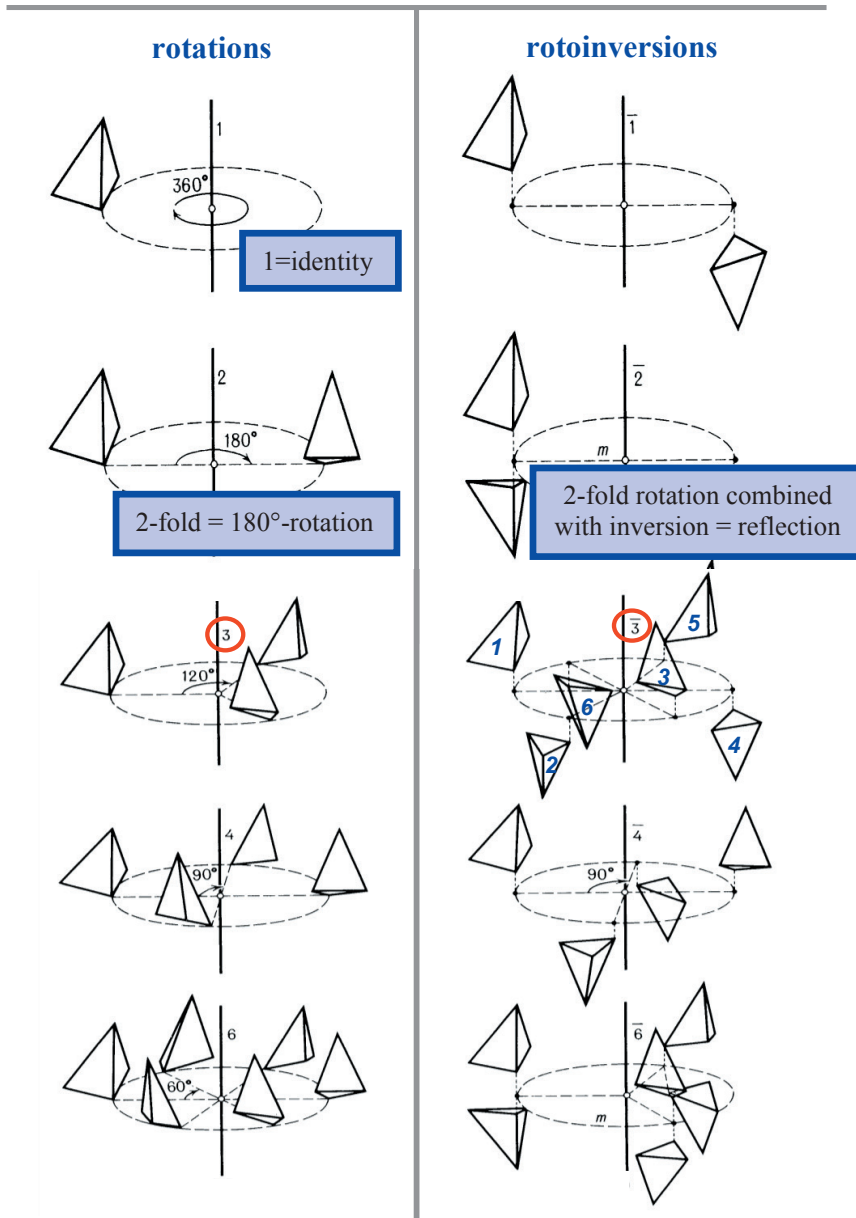
2-fold rotation combined with inversion = reflection

Fig. 1.5: *Rotations: n=1 (identity), n=2 (rot. angle 180°), n=3 (120°), n=4 (90°), n=6 (60°). Rotoinversions: $\bar{1}$ (inversion), $\bar{2} \equiv m$ (reflection), $\bar{3}=3+\bar{1}$, $\bar{4}$, $\bar{6} = 3/m$.*
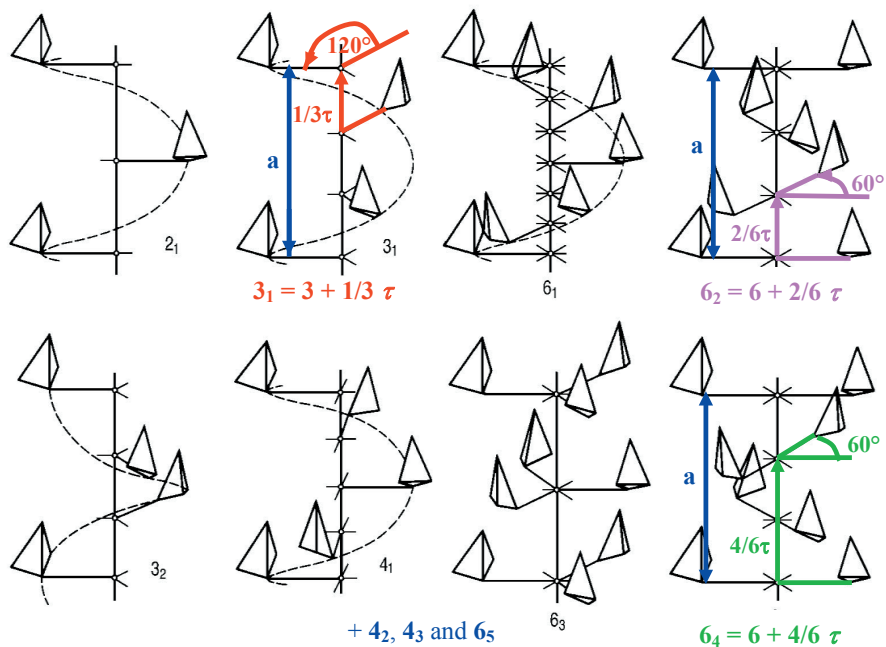
Fig. 1.6: *Screw rotations n_m: combination of rotations n and translation components $(m/n) \cdot a \parallel$ to the rotation axis.*

reflection: mirror plane m ⊥ image plane (plane of the paper)

glide reflection: glide plane a ⊥ with glide vector *a*/2
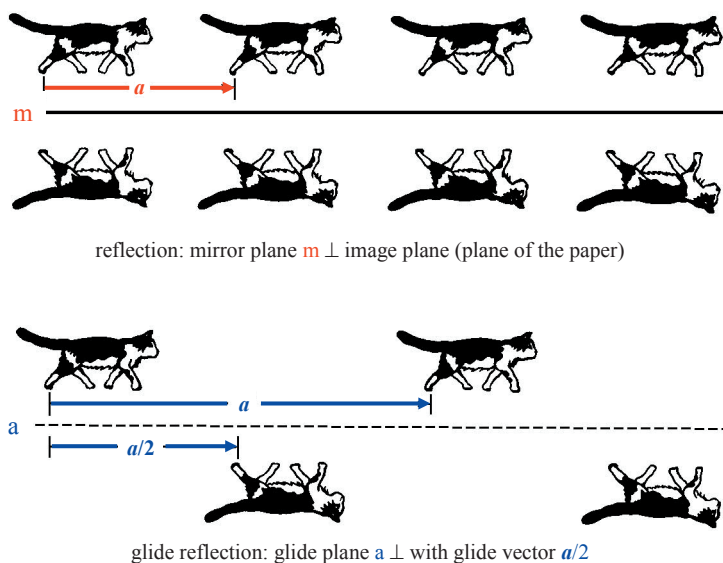
Fig. 1.7: *Examples of reflections and glide reflections.*

The above equation, re-written in matrix notation:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{pmatrix} \circ \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}; \quad X' = \mathbf{W} \circ X + w = (\mathbf{W}, \mathbf{w}) \circ X \quad 1.6$$

The (3×3) matrix **W** is the rotational part and the (3×1) column matrix *w* the translational part of the symmetry operation. The two parts **W** and *w* can be assembled into an augmented (4×4) matrix *W* according to

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} & W_{13} & w_1 \\ W_{21} & W_{22} & W_{23} & w_2 \\ W_{31} & W_{32} & W_{33} & w_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = W \circ X \quad\quad 1.7$$

Since every symmetry transformation is a "rigid-body" motion, the determinant of all matrices **W** and *W* is det **W** = det *W* = ± 1 (+ 1: preservation of handedness; - 1: change of handedness of the object).

The sequence of two symmetry operations (successive application) is given by the product of their matrices W_1 and W_2:

$$W_3 = W_1 \circ W_2 \quad\quad\quad 1.8$$

where W_3 is again a symmetry operation.

1.4 Crystallographic point groups and space groups

The symmetry of a crystal and of its crystal structure can be described by mathematical group theory. The symmetry operations are the group elements of a crystallographic group G and the combination of group elements is the successive execution of symmetry operations. All possible combinations of crystallographic point-symmetry operations in three-dimensional space lead to exactly 32 crystallographic point groups (≡ crystal classes) which all are of finite order (the maximum order is 48 for the cubic crystal class m$\overline{3}$m). For the different crystal systems they are represented by stereographic projections in Fig. 1.8. There are two types of group symbols in use: For each crystal class the corresponding Schoenflies symbol is given at the bottom left and the Hermann-Mauguin (international) symbol at the bottom right. A maximum of 3 independent main symmetry directions ("Blickrichtungen") is sufficient to describe the complete symmetry of a crystal. These symmetry directions are specifically defined for the 7 crystal systems (Hermann-Mauguin symbols). As an example, the symmetry directions of the cubic system are shown in Fig. 1.9.
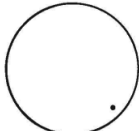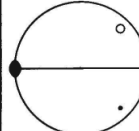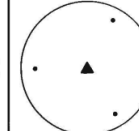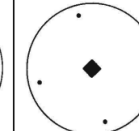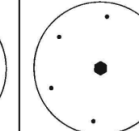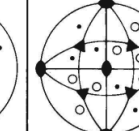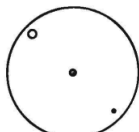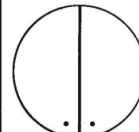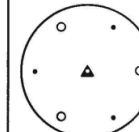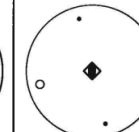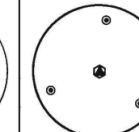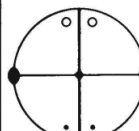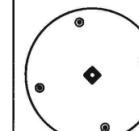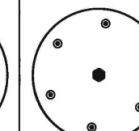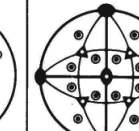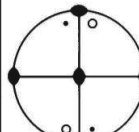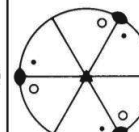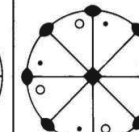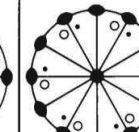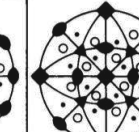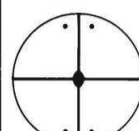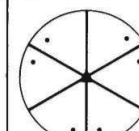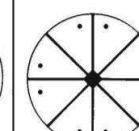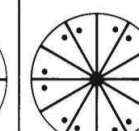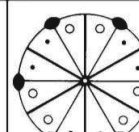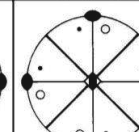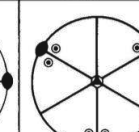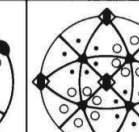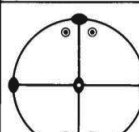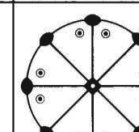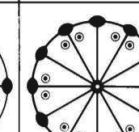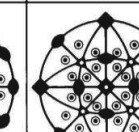
Fig. 1.8: *The 32 crystallographic point groups (crystal classes) in three-dimensional space represented by their stereographic projections. The group symbols are given according to Schoenflies (bottom left) and to Hermann-Mauguin (bottom right).*

$[100]$
$\dfrac{4}{m}$

$[111]$
$\bar{3}$

$[110]$
$\dfrac{2}{m}$
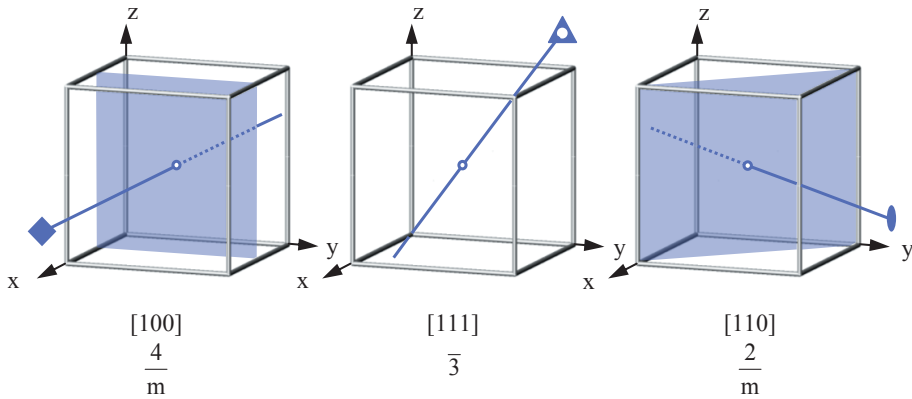
Fig. 1.9: *Symmetry directions ("Blickrichtungen") of the cubic lattice (a=b=c, α=β=γ=90°). Along [100]: 4/m, along [111]: $\bar{3}$, along [110]: 2/m.*

In three dimensions all possible combinations of the point symmetries of the 32 crystallographic point groups with the lattice translations of the 14 Bravais lattices lead to exactly 230 space groups, all of infinite order. As already mentioned, the addition of translations to the point symmetries results in new symmetry operations: Screw rotations and glide reflections. The conventional graphical symbols for the symmetry elements according to the International Tables for Crystallography Vol. A (2002) [1] are shown in Fig. 1.10.
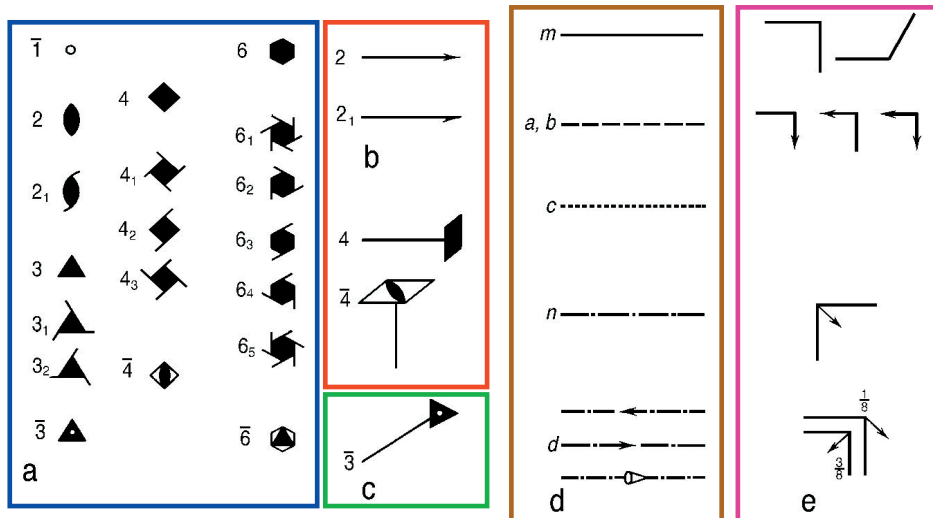
Fig. 1.10: *Conventional graphical symbols for symmetry elements:*
- symmetry axes: (a) perpendicular, (b) parallel, and (c) inclined to the plane;
- symmetry planes: (d) perpendicular and (e) parallel to the image plane.

1.5   Quasicrystals

Since the pioneering work of Shechtman et al [2] published in 1984 and honoured by the 2011 Nobel-Prize in Physics it is accepted that the crystalline state with its 3D periodic arrangement of atoms in a lattice is not the only long-range ordered ground state of matter. This quasi crystalline state also follows strict construction rules and exhibits long range order, but the rules are no longer based on the lattice concept.
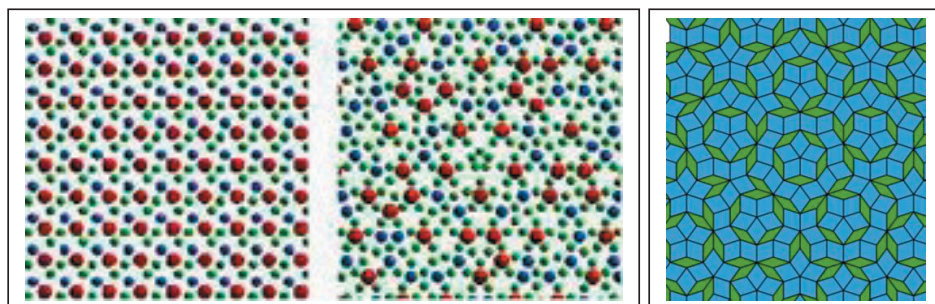
Fig. 1.11: 2D-analogues of a crystalline (left) and a quasi-crystalline structure (center) [3], Penrose tiling of a plane by two different rhombs (right) [4].

The description of quasicrystals is closely related to the so called Penrose-tilings which are a way to cover the plane completely and without overlap by a long range ordered, non-periodic arrangement of (in the case shown in Fig. 1.11) two different geometric shapes (here: rhombs). As a result of the lack of translation symmetry, the "crystallographically forbidden" rotation axes (5-fold, 8-fold, 10-fold etc., more precisely: forbidden as part of a 3D-space group symmetry) may occur in quasi crystals and also show up as symmetries of the outer shape (Fig. 1.12 left) and diffraction patterns of quasi crystals (center).
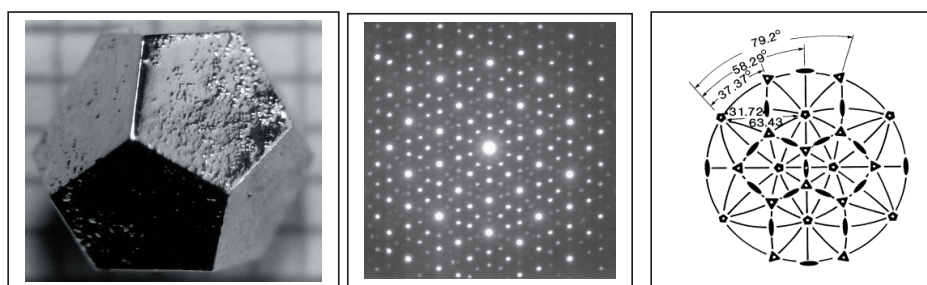
Fig. 1.12: Icosahedral quasi crystal HoMgZn (left) [5], electron diffraction pattern taken along the -5 rotoinversion axis (center) [5] and stereographic projection of the icosahedral point symmetry group m-3-5 [2].

1.6   Application: Structure description of $YBa_2Cu_3O_{7-\delta}$

The crystal structure determination with atomic resolution is achieved by diffraction experiments with X-rays, electron or neutron radiation. As an example, the results of a structure analysis by neutron diffraction on a single crystal of the ceramic high-T_C superconductor $YBa_2Cu_3O_{7-\delta}$ with $T_C = 92$ K are presented [6]. The atomic arrangement of the orthorhombic structure, space group P m m m, and the temperature-dependent electrical resistivity are shown in Fig. 1.13.
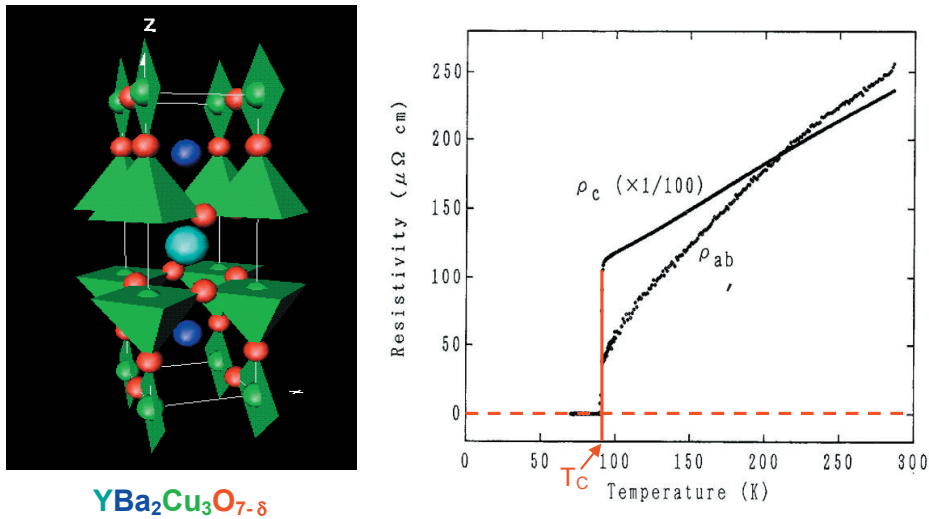
$YBa_2Cu_3O_{7-\delta}$

Fig. 1.13: *Crystal structure (unit cell) of $YBa_2Cu_3O_{7-\delta}$ with the CuO_x-polyhedra (left) and the electrical resistivity as a function of temperature $\parallel$ and $\perp$ to the [001] direction (right).*

The crystal structure contains two different Cu-O polyhedra (green): CuO_5-tetragonal pyramids and CuO_4-squares. The pyramids share corners in 2D and form double layers, the charge carriers responsible for superconductivity are supposed to be located in these double layers.

Information from the international tables on the relative locations and orientations of the symmetry elements (symmetry operations 1, 2_z, 2_y, 2_x, $\bar{1}$, m_z, m_y, m_x) of the orthorhombic space group P m m m, together with the choice of the origin (in an inversion centre), is shown in Fig. 1.14. The general position (site symmetry 1) of multiplicity 8 and all special positions with their site symmetries are listed in Fig. 1.15. There are no special reflection conditions for this space group.

$Pmmm$ D_{2h}^{1} mmm Orthorhombic

No. 47 $P\,2/m\,2/m\,2/m$ Patterson symmetry $Pmmm$
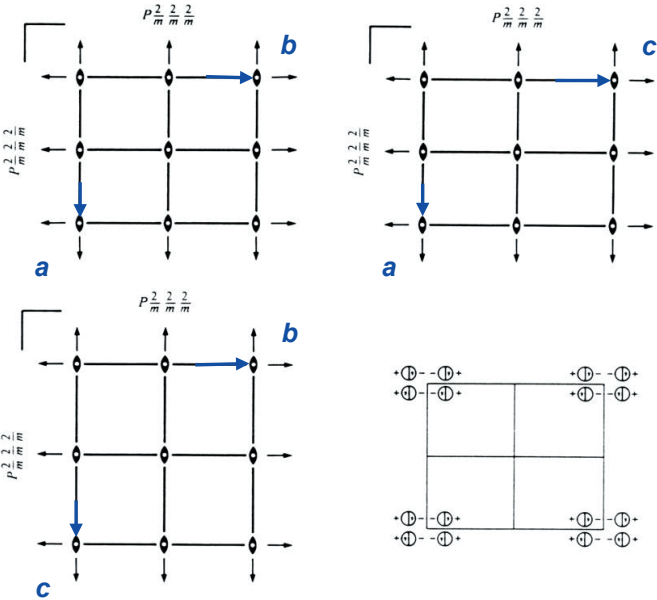
Origin at centre (mmm)

Asymmetric unit $0 \le x \le \frac{1}{2}$; $0 \le y \le \frac{1}{2}$; $0 \le z \le \frac{1}{2}$

Symmetry operations

(1) 1   (2) 2   0,0,z   (3) 2   0,y,0   (4) 2   x,0,0
(5) $\bar{1}$ 0,0,0   (6) m   x,y,0   (7) m   x,0,z   (8) m   0,y,z

Fig. 1.14:   *Description of the orthorhombic space group P m m m in [1].*

CONTINUED   **YBa$_2$Cu$_3$O$_{7-\delta}$**   No. 47   *Pmmm*

Generators selected (1); $t(1,0,0)$; $t(0,1,0)$; $t(0,0,1)$; (2); (3); (5)

Positions

Multiplicity, Wyckoff letter, Site symmetry		Coordinates				Reflection conditions
8	α	1	(1) x,y,z (2) $\bar{x},\bar{y},z$ (3) $\bar{x},y,\bar{z}$ (4) $x,\bar{y},\bar{z}$			General: no conditions
			(5) $\bar{x},\bar{y},\bar{z}$ (6) $x,y,\bar{z}$ (7) $x,\bar{y},z$ (8) $\bar{x},y,z$			

General: no conditions

Special: no extra conditions

4	z	$..m$	$x,y,\tfrac{1}{2}$	$\bar{x},\bar{y},\tfrac{1}{2}$	$\bar{x},y,\tfrac{1}{2}$	$x,\bar{y},\tfrac{1}{2}$
4	y	$..m$	$x,y,0$	$\bar{x},\bar{y},0$	$\bar{x},y,0$	$x,\bar{y},0$
4	x	$.m.$	$x,\tfrac{1}{2},z$	$\bar{x},\tfrac{1}{2},z$	$\bar{x},\tfrac{1}{2},\bar{z}$	$x,\tfrac{1}{2},\bar{z}$
4	w	$.m.$	$x,0,z$	$\bar{x},0,z$	$\bar{x},0,\bar{z}$	$x,0,\bar{z}$
4	v	$m..$	$\tfrac{1}{2},y,z$	$\tfrac{1}{2},\bar{y},z$	$\tfrac{1}{2},y,\bar{z}$	$\tfrac{1}{2},\bar{y},\bar{z}$
4	u	$m..$	$0,y,z$	$0,\bar{y},z$	$0,y,\bar{z}$	$0,\bar{y},\bar{z}$
2	t	$mm2$	$\tfrac{1}{2},\tfrac{1}{2},z$	$\tfrac{1}{2},\tfrac{1}{2},\bar{z}$		
2	s	$mm2$	$\tfrac{1}{2},0,z$	$\tfrac{1}{2},0,\bar{z}$		
2	r	$mm2$	$0,\tfrac{1}{2},z$	$0,\tfrac{1}{2},\bar{z}$		
2	q	$mm2$	$0,0,z$	$0,0,\bar{z}$		
2	p	$m2m$	$\tfrac{1}{2},y,\tfrac{1}{2}$	$\tfrac{1}{2},\bar{y},\tfrac{1}{2}$		
2	o	$m2m$	$\tfrac{1}{2},y,0$	$\tfrac{1}{2},\bar{y},0$		
2	n	$m2m$	$0,y,\tfrac{1}{2}$	$0,\bar{y},\tfrac{1}{2}$		
2	m	$m2m$	$0,y,0$	$0,\bar{y},0$		
2	l	$2mm$	$x,\tfrac{1}{2},\tfrac{1}{2}$	$\bar{x},\tfrac{1}{2},\tfrac{1}{2}$		
2	k	$2mm$	$x,\tfrac{1}{2},0$	$\bar{x},\tfrac{1}{2},0$		
2	j	$2mm$	$x,0,\tfrac{1}{2}$	$\bar{x},0,\tfrac{1}{2}$		
2	i	$2mm$	$x,0,0$	$\bar{x},0,0$		

1	h	mmm	$\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}$
1	g	mmm	$0,\tfrac{1}{2},\tfrac{1}{2}$
1	f	mmm	$\tfrac{1}{2},\tfrac{1}{2},0$
1	e	mmm	$0,\tfrac{1}{2},0$
1	d	mmm	$\tfrac{1}{2},0,\tfrac{1}{2}$
1	c	mmm	$0,0,\tfrac{1}{2}$
1	b	mmm	$\tfrac{1}{2},0,0$
1	a	mmm	$0,0,0$

Fig. 1.15:   *General and special positions (coordinates of all symmetrically equivalent positions) of space group P m m m with their site symmetries and multiplicities [1]. The special positions occupied by atoms of the YBa$_2$Cu$_3$O$_{7-\delta}$ structure are highlighted by frames.*

The atomic parameters of the structure refinement of $YBa_2Cu_3O_{6.96}$ at room temperature [6] are given in the following Table:

Atomic positions of $YBa_2Cu_3O_{6.96}$ orthorhombic, space group type P *2/m 2/m 2/m* a = 3.858 Å, b = 3.846 Å, c = 11.680 Å (at room temperature)					
atom/ion	multiplicity	site symmetry	x	y	z
$Cu1/Cu^{2+}$	1	*2/m 2/m 2/m*	0	0	0
$Cu2/Cu^{2+}$	2	*m m 2*	0	0	0.35513(4)
Y/Y^{3+}	1	*2/m 2/m 2/m*	½	½	½
Ba/Ba^{2+}	2	*m m 2*	½	½	0.18420(6)
$O1/O^{2-}$	2	*m m 2*	0	0	0.15863(5)
$O2/O^{2-}$	2	*m m 2*	0	½	0.37831(2)
$O3/O^{2-}$	2	*m m 2*	½	0	0.37631(2)
$O4/O^{2-}$	1	*2/m 2/m 2/m*	0	½	0

1.7 References

[1] International Tables for Crystallography Vol. A, *Space-group Symmetry*, edited by Th. Hahn, Dordrecht: Kluwer Academic Publishers (5. Edition, 2002)

[2] D. Shechtman, I. Blech, D. Gratias, J. Cahn: "*Metallic Phase with Long-Range Orientational Order and No Translational Symmetry*". Physical Review Letters **53**, 1951 (1984)

[3] S. Curtarolo, Duke University, http://today.duke.edu/2005/09/quasicrystals.html

[4] Image: Inductiveload/Wikimedia Commons http://en.wikipedia.org/wiki/Penrose_tiling

[5] I.R. Fisher, Z. Islam, A.F. Panchula, K.O. Cheon, M.J. Kramer, P.C. Canfield, A.I. Goldman: Philos. Mag. **B 77** 1601–1615 (1998) Images: http://en.wikipedia.org/wiki/File:Zn-Mg-HoDiffraction.JPG

[6] P. Schweiss, W. Reichardt, M. Braden, G. Collin, G. Heger, H. Claus, A. Erb: Phys. Rev. **B 49**, 1387 – 1396 (1994)

B 2   Quasicrystals and modulated structures

M. Dulle

Jülich Centre for Neutron Science

Forschungszentrum Jülich GmbH

Contents

Introduction

Already shortly after the discovery that crystals are periodic arrangements of atoms or molecules scientist discovered materials that showed similar properties in that these materials also gave distinct scattering patterns with well-defined peaks and therefore had to possess long range order. But as it turned out these structures were not periodic so the term aperiodic crystals was introduced and nowadays is used to describe composite crystals, incommensurately modulated crystals and quasicrystals. In this chapter a short introduction to the latter two is given with an emphasis on a qualitative understanding of what sets them apart from regular periodic crystals and how to obtain meaningful depictions in real space of these extraordinary structures.

1   Modulated structures

Already shortly after the discovery that crystals have a 3D periodic structure by Max von Laue using X-rays[1] and the successful description of their structures by Bragg [2] materials were found that gave spots in the diffractograms that could not be indexed using three integer indices [3]. These extra peaks are called satellites [4] and arise when a three dimensional crystal structure is modulated with respect to its unit cell. The position and direction of these satellites depends on the modulation of the crystal structure and can be used to extract that information from the scattering pattern. But a slightly different route as for regular crystals has to be taken to be able to give a quantitative description of these structures namely higher-dimensional crystallography[5,6]. By extending the description into higher dimensions it is possible again to look at these crystals as periodic structures and with that a lot of the tools that are used on 3D periodic structures can be used again in a familiar way.

1.1   Types of modulation

To keep things simple in this chapter we will only consider two types of modulation namely displacive and occupational modulation. There are also mixed forms of the latter and as a separate class composite crystals, which are crystals that have different unit cells within one crystal. However this would go well beyond an introductory lecture.

Displacive modulation can be simply visualized as the displacement of one or more elements (Atoms, Molecules, Proteins…) of the unit cell with respect to the position in the conventional crystal. This displacement of elements follows a modulation function for instance a sine as can be seen in Fig.1. Note however that the modulation function can have a lot of different forms like the superposition of several harmonic waves or a saw tooth or triangular to name just a few.

The other common type of modulation is the occupational modulation in the crystal structure. Here all the elements of the unit cell are on their expected position but their occupation is governed by a probability function. In nature also a combination of several different types of modulation can be present at the same time making the correct interpretation of diffraction data all but a simple task.

The modulation function is characterized by its wavelength(λ) and its amplitude, when looking at the wavelength of the modulation function one can differentiate between two cases, either the wavelength and the unit cell length along the modulation direction are commensurable or incommensurable. Any two real numbers are commensurable if their ratio

can be expressed as the ratio of two integers (a rational number). This leads to a periodic crystal with a larger unit cell than in the non-modulated case but it is still periodic. If however the ratio of λ and the unit cell length are incommensurable the modulated crystal is no longer periodic in its real space dimensions as no unit cell can ever be found along the modulation direction. This is visualized in Fig.1 were starting from the top the structure of an unmodulated 2D crystal is shown with the unit cell as red rectangle (top left).

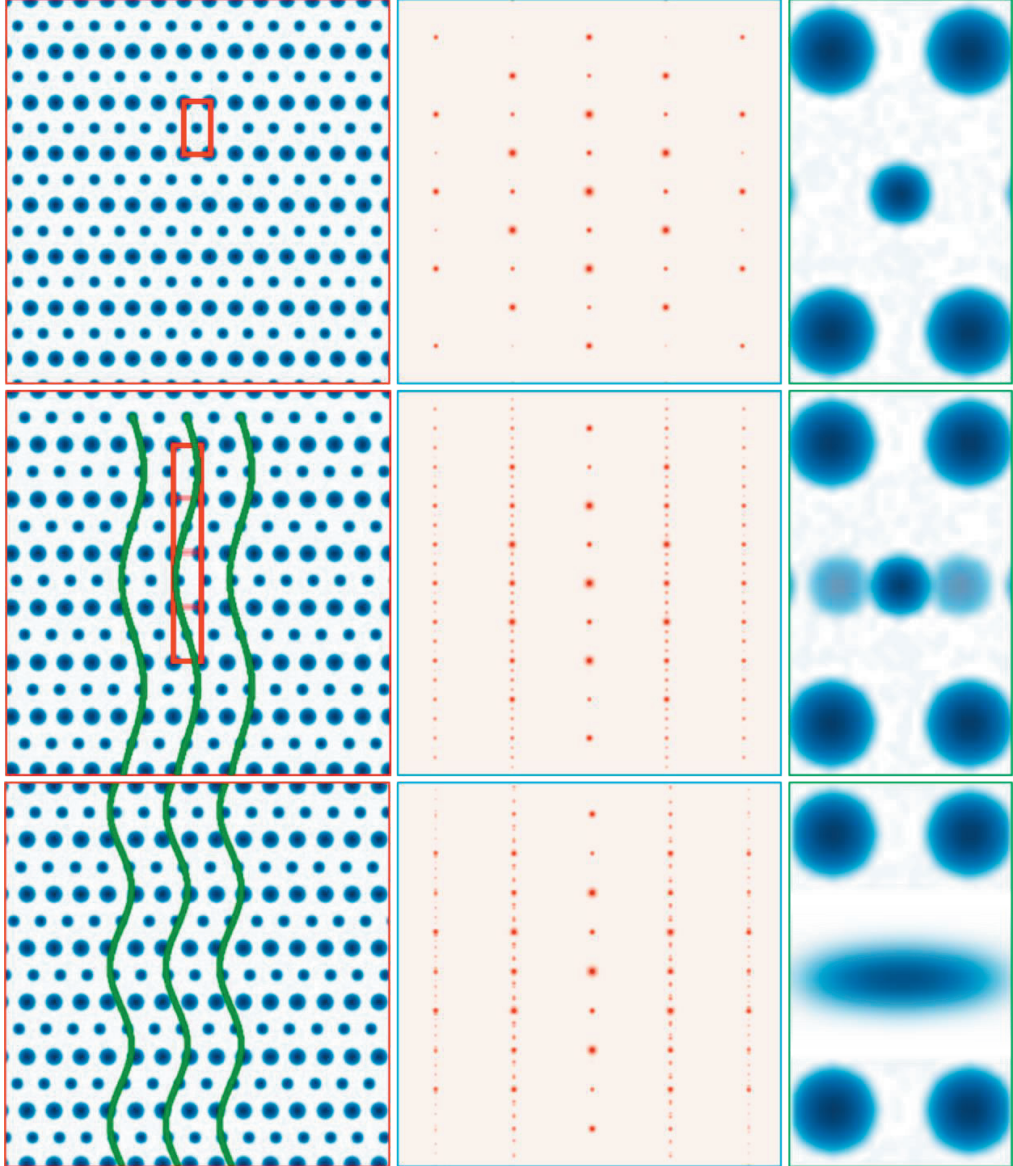**Fig. 1:** Top: unmodulated 2D crystal (left) with its scattering pattern(middle) and density map (right) Middle: commensurably modulated 2D crystal (left) with the enlarged unit cell, its scattering pattern showing main and satellite reflections (middle) and the still discrete density map. Bottom: incommensurably modulated 2D crystal, its scattering pattern(middle) and the corresponding density map

Next to it in the middle of the top row the scattering pattern of the 2D structure is shown and to the right the corresponding density map. In the middle a commensurate modulated structure of the same crystal with a λ of $4a$ is depicted. One can easily see that the modulated structure is still periodic but with a larger unit cell. When looking at the scattering pattern we still have all the reflections from the unmodulated structure, which are called the main reflections and a lot of satellites which are a result of the modulation. The density map on the right shows now three spots in the middle instead of only one. Note however that the map is still discrete as the modulated element only occupies these three positions. The bottom row shows the same crystal with an incommensurate modulation with a λ of τa (with τ being 1.618… the golden mean), hence there is no more periodicity along this direction but still sharp diffraction peaks as well as satellite reflections. The density map on the other hand is not discrete anymore as all possible positions along the modulation direction are realized. In the scattering pattern this loss of periodicity yields also satellite reflections that cannot be indexed using two indices as for all the main reflections. The same principle holds also in three dimensions and one needs a different set of indices to describe the whole scattering pattern.

1.2   Superspace and the recovery of periodicity

Just as a reminder for all the considerations laid out in this section, an incommensurately modulated crystal although not periodic in three dimensions has still perfect long range order and shows therefore distinct and sharp reflections in the scattering pattern. When looking at the diffraction peaks of a structure with long-range order each peak position can be written as

$$\mathbf{k} = \sum_{i=1}^{n} h_i \boldsymbol{a}_i^* \quad \text{integers } h_i \tag{1}$$

this is called a vector module[1] and is a more general description of a crystal encompassing both periodic but also aperiodic crystals. In case of a normal periodic crystal the dimensions of the crystal d equal the rank of the vector module such that $d = n = 3$. If we use the standard notation from crystallography $\boldsymbol{a}_1^*, \boldsymbol{a}_2^*, \boldsymbol{a}_3^*$ become $\boldsymbol{a}^*, \boldsymbol{b}^*, \boldsymbol{c}^*$ and h_1, h_2, h_3 are h, k, l. This leads to the well-known relation

$$\mathbf{k} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^* \tag{2}$$

Now when considering a modulated crystal the rank of the vector module is not equal to the dimensions of the crystal but larger in the simplest case with only a one-dimensional modulation the rank(n) of the vector module would be 4. For simplicities sake we will now only consider this case but the concept also works for two and three dimensional modulations with m equal to 5 or 6 respectively.
So the peak positions for the vector module of rank four are

$$\mathbf{k} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^* + m\boldsymbol{q} \tag{3}$$

The main reflections of the scattering pattern can be described with the indices $(h,k,l,0)$ but the satellite reflections need the fourth index m with $\mathbf{q}$ being the wave vector of the modulation. This fourth index is necessary to get a good and meaningful structure from the

---

[1] In this chapter the notation for the reciprocal vectors is different than in the rest of the book because τ is used as the symbol for the golden mean

scattering data. If only the main reflections of the scattering pattern are used for the structure determination in real space one gets the so called average structure. This structure is an average over a large number of different unit cells due to the modulation and shows very large anisotropic displacement parameters as well as unrealistic bond angles and lengths. Another approach to solve the structure is to drop the distinction of main and satellite reflections and use all reflections for indexing with three indices (*h,k,l*). This will result in a smaller reciprocal unit cell and a lot of reciprocal lattice points that are empty because their intensity is too low. The resulting structure in real space accordingly has a much larger unit cell and is called the superstructure (see Fig1 middle left).   In case of a commensurate modulation this approach is valid and works quite well in case of an incommensurate modulation however the satellite reflections will never perfectly fit the positions of the new reciprocal lattice which leads to poor agreement factors and large atomic displacement factors ect.. The only way to obtain a good incommensurable structure in real space is to keep the distinction of main and satellite reflections. In a first step the main reflections are used to determine the basic reciprocal unit cell. Then one can use the fact that the satellite reflections in reciprocal space are not arbitrarily placed. All the satellites have fixed distances from their main reflection and from each other this allows use to describe their position using the modulation vector q. Each satellite is $mq(m = \pm1, \pm2 \dots)$ away from its main reflection and can now be indexed (Fig. 2 left). The vector q can be expressed as a linear combination of fractions of the three reciprocal base vectors a^*, b^*, c^* which describe the basic cell like this.

$$q = \alpha a^* + \beta b^* + \gamma c^* \tag{4}$$

So now we have defined the vector q with respect to the reciprocal lattice of the average structure and also have now another means to differentiate between commensurate and incommensurate structures in reciprocal space. If all the components (α, β, γ) of q are rational we have a commensurate structure and if at least one is irrational it is incommensurate.

We have gone now into the fourth dimension by adding this q into the description of the scattering pattern but how can this be used to solve the structure in three dimensions? As this is only an introductory lecture only the principle of how this is done will be shown how to actually do it can be found in the references. The scattering pattern that is measured is still in 3 dimensions but as we saw we need a fourth dimension to describe it and this stems from the fact this scattering pattern in 3 dimensions is a projection of a 4 dimensional scattering pattern which is periodic. In this periodic superspace many of the concepts that are used with regular crystals can again be used because they require periodicity to work. In Fig.3 the concept of projection from a 4 dimensional reciprocal superpace onto the 3 dimensional reciprocal space (the scattering pattern) is shown (right side) as well as the indexing of a hypothetical scattering pattern using four indices (left side).
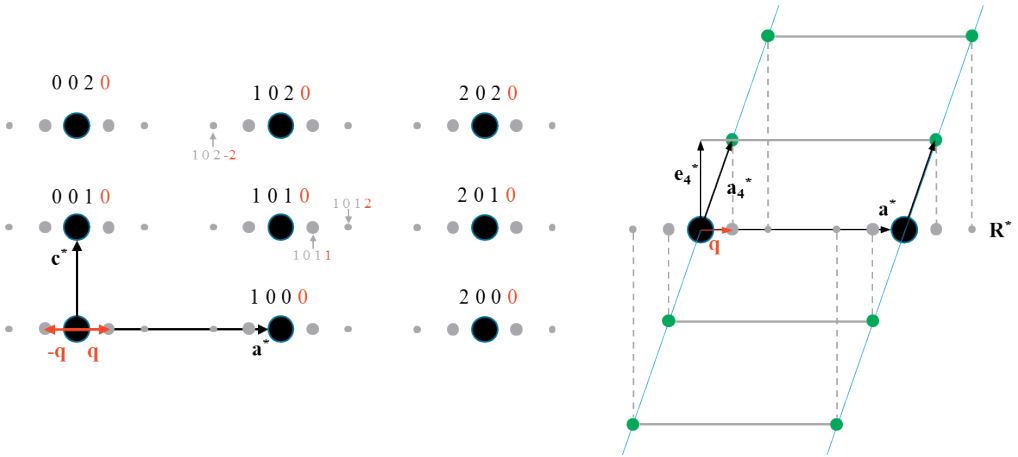
Fig. 2: The indexing of a hypothetical scattering pattern of a modulated crystal using 4 indices looking along the b^* direction (left) and the concept of projection from a 4 dimensional reciprocal superspace onto the 3 dimensional reciprocal space R^* which is the measured scattering pattern (right).

When looking at the vector **q** alone it is difficult to envision the extra dimension we introduced before. The relationship of this vector in three dimensional reciprocal space to the fourth dimension is illustrated on the right hand side of Fig.3. Here we chose **q** to be parallel to a^* (meaning that only the α component of **q** is non zero). The fourth dimension is defined by e_4^* which is perpendicular to a^*, b^* and c^* (of which only a^* is shown). The vector a_4^* is the reciprocal lattice vector of the fourth dimension and is the linear combination of e_4^* and **q**. The projection of this reciprocal lattice defined by a_4^* onto the reciprocal lattice vector a^* (dotted lines) which lies in the three dimensional reciprocal space R^* results in **q** and gives us all the satellite reflections.

1.3   Where are the atoms?

In the literature one can often read that the aperiodic structure in real space is interpreted as a cut through the (3+d) dimensional superspace, but what exactly does that mean? Before we try to shed some light on this the following convention will be introduced: the real space coordinates x, y, z will now be called x_1, x_2, x_3 and the vectors a, b, c shall be a_1, a_2, a_3. This way all the axis and coordinates in n-dimensional superspace can be denoted in a consistent fashion so that for our four dimensional case we get $a_{s1}, a_{s2}, a_{s3}, a_{s4}$ as the four perpendicular vectors of superspace, a_1, a_2, a_3 as vectors for three dimensional real space and x_1, x_2, x_3, x_4 for the atomic coordinates.

As it is impossible to draw a four dimensional structure in three dimensions we have to use sections (real space) or projections (reciprocal space) of the four dimensional structure similar to drawing a three dimensional structure as a two dimensional projection on a sheet of paper. To do this one takes one of the principal axis (a_{s1}, a_{s2} or a_{s3}) and a_{s4} and draws a two dimensional cut though the four dimensional space(Fig.4). This way we have defined the periodic superspace structure with the two vectors a_{s1} and a_{s4}. The real space vector a_1 is perpendicular to a_{s4} and at an angle to a_{s1} defined by the **q** vector component α. So in this (3+1) superspace atoms can no longer be interpreted as points but instead have to be envisioned as one dimensional objects along the fourth dimension so called atomic surfaces. These atomic surfaces are represented as a curve along the fourth dimension and this curve is periodic along a_{s4}. Please note that the two dimensional cut in Fig.4 is periodic and the same

holds for the two dimensional cuts along the other principal directions $(\boldsymbol{a}_{s2}, \boldsymbol{a}_{s4})$ and $(\boldsymbol{a}_{s3}, \boldsymbol{a}_{s4})$. This now shows nicely that by introducing this extra dimension we have gained translational symmetry again and that the aperiodic structure in three dimensions is periodic in four dimensions.
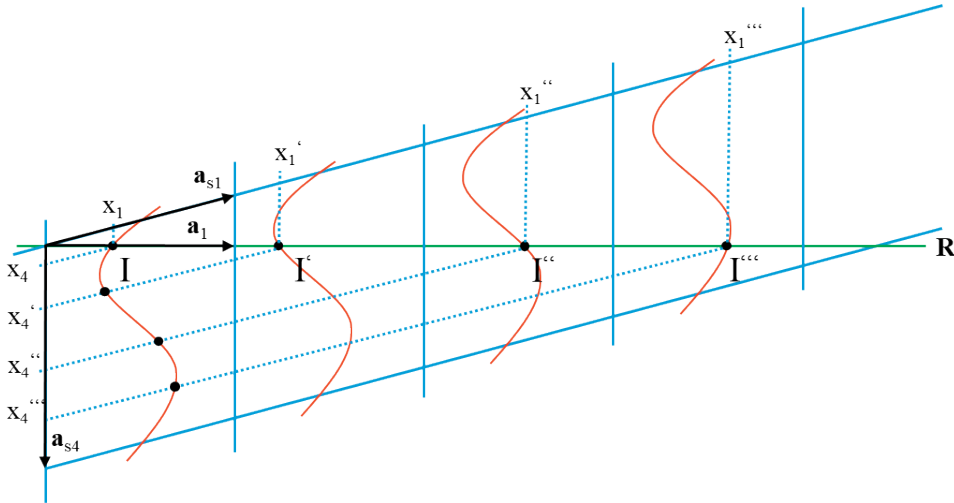

Fig. 3: The two-dimensional cut $(\boldsymbol{a}_{s1}, \boldsymbol{a}_{s4})$ is shown with the angle between $\boldsymbol{a}_{s1}$ and $\boldsymbol{a}_1$ being the $\boldsymbol{\alpha}$ component of $\mathbf{q}$. The modulated positions of the atoms are depicted as red lines. The positions of the atoms in 3D (I,I´,I´´ and I´´´) can be obtained by the intersection of the three dimensional space (line $\boldsymbol{R}$) with the modulation function. Vice versa the modulation function can be determined by shifting the points (I,I´,I´´ and I´´´) along $\boldsymbol{a}_{s4}$ into a single unit cell.

The aperiodic structure in real space can now be obtained by a cut $\boldsymbol{R}$ parallel to $\boldsymbol{a}_1$ through this two dimensional representation of superspace. When doing this for all possible two-dimensional representations the real space coordinates for all the atoms are obtained. In order to be able to do this one has to deduce the atomic modulation function describing the atomic domains. This can be done shifting all points along a cut $\boldsymbol{R}$ (I,I´,I´´ and I´´´) into one unit cell in the plot $\boldsymbol{a}_{s1}, \boldsymbol{a}_{s4}$ by translational symmetry(dotted lines in Fig.4) and this means that the atomic modulation function is depending on the three dimensional structure and vice versa. This fact is what enables us to solve the structure and get a meaningful interpretation of the scattering of modulated structures in real space.

2   Quasicrystals

When looking at modulated structures one always can determine or at least envision the basic cell, which would be the unit cell of the unmodulated structures. This fact is what enables us to differentiate between main and satellite reflections in the scattering pattern. For quasicrystals this is no longer the case. Here there is no periodic basic cell as already the basic structure is incommensurate so one cannot make any distinction between main and satellite reflections anymore. What also sets them apart from the other aperiodic crystals is that they also have non crystallographic symmetry elements like 12, 8 or 5-fold rotational axis. Since their discovery by Shechtman [7] several hundred quasicrystals have been found and characterized. In this part of the chapter we will explore how to envision these structures and on which principles they are built and the basic ideas to solve their structure using scattering methods.

2.1   The Fibonacci chain a one dimensional quasicrystal

As it has been stated in the previous section a loss of periodicity does not necessarily mean a loss of long range order. The Fibonacci chain is such an example and can be understood as a one-dimensional quasicrystal. To create this sequence we start with a finite sequence of two segments *L* and *S* and apply the following rule. $S \rightarrow L$ and $L \rightarrow LS$ to build successive strings. When starting with *L* and one applies these rules we get the following sequence of strings.

L
LS……………………………………*1*
LSL……………………………..…*2*
LSLLS……………………................*1.5*
LSLLSLSL……………………….….*1.6667*
LSLLSLSLLSLLS……………….….*1.6*
LSLLSLSLLSLLSLSLLSLSL……..…*1.625*
.
.
.
.
.
etc.

The numbers are the ratio of *L/S* and when the chain is growing ad infinitum this ratio will become $\tau = 1.618034 \ldots$ the golden mean. This golden mean will also occur again later one when dealing with icosahedral symmetries. The fourier transform of this sequence will show sharp diffraction peaks which cannot be indexed using only one integer but two. Which in turn means to describe the structure   we can use similar principles as we did with the incommensurately modulated crystals, namely the embedding into a higher dimensional space. Furthermore the chain shows self-similarity meaning that any part of the chain will show up again within the length of the whole chain, which is also a property all quasicrystals share. We will return to the Fibonacci chain when dealing with higher dimensional space and show how to use it to obtain quasicrystalline structures from scattering.

2.2 Tilings and quasicrystals

Regular Crystals belonging to any of the 230 possible space groups only show 2,3,4 and 6-fold rotational symmetry. This can be understood when looking at Fig.4.
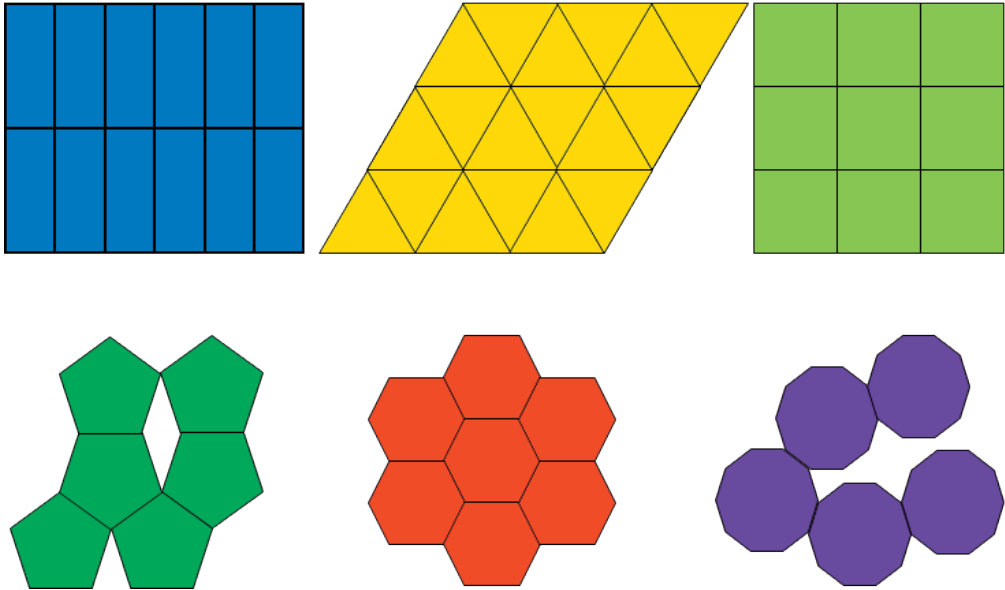

Fig. 4: Tiles showing different rotational symmetries from top left to bottom right: 2,3,4,5,6 and 7 notice how the tiles with 5 and 7 fold symmetry cannot be placed to fill the plane without gaps

Here two dimensional tilings with the aforementioned rotational symmetries are shown and one can see that they all can cover the plane without leaving any gaps. If however we use a tile with a five, seven fold or higher rotational symmetry it is not possible to cover the whole plane without gaps. There is a way to cover the plane with a set of two tiles so that the tiling has 5-fold rotational symmetry. This tiling is called a Penrose tiling [8]. At the heart of this tiling lies again the golden mean τ. There is one big difference between this tiling and the others. The Penrose tiling has no translational symmetry and is therefore not periodic but it still has as mentioned above the 5 fold rotational symmetry. There are several ways to construct a Penrose tiling. One of the most popular ones is by using a skinny and a fat rhombus as shown in Fig.5. If the edge length for both tiles is set to one the length of the long diagonal of the fat rhombus is τ and the length of the short diagonal of the skinny rhomb is $1/\tau$ and the angles in the skinny rhomb are 36°,144°,36°,144° and in the fat rhomb 72°,108°,72°,108°. If one would start to assemble these tiles without any specific matching rules you can easily end up with a periodic regular tiling again(Fig 4..) or very soon create holes in it. These matching rules can visualized by differentiating the edges of the two tiles with two colors. Only pairings of the same color are allowed and two tiles shall also not be combined to form a parallelogram. An easy way to build such a tiling is by using its self-similarity. This method is called inflation and is exactly what we did with the Fibonacci chain in 2.1 but now in 2D (Fig.5 right).
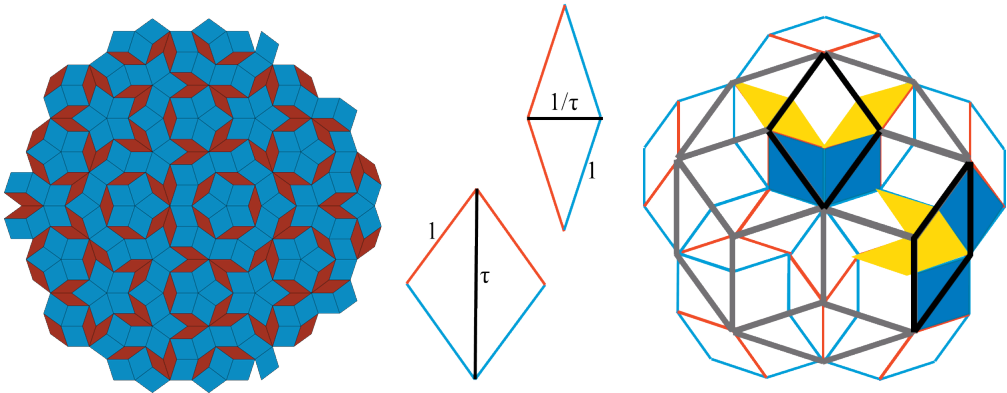
Fig. 5: Left: a Penrose tiling made of rhombs Middle:The two tiles needed for this tiling with the matching rules shown by the edge colors (only edges of the same color may be put together) Right: Tiling constructed showing the matching rules and on top of this tiling the former generation before inflation. The inflation rules for the fat and skinny rhombus are depicted by the blue and yellow rhombs inside the black edged larger rhombus the scaling from the grey/black tiling to the blue/red one is τ

One might ask what does that have to do with quasicrystals? This aperiodic tiling and others were found about a decade before quasicrystals were discovered but as it turned out both had a lot in common, they lack translational symmetry are self similar and posses rotational symmetries not allowed for periodic crystals. So the study of these tilings helped a lot to understand and interpret the quasicrystals that were found in nature. One can build also different 3D tilings of which only one is of importance when looking at icosahedral quasicrystals. The so called 3D Penrose or Ammann tiling. This tiling consists of two kinds of 3D unit tiles a prolate and an oblate rhombohedron. If these two tiles are used together with a special set of matching rules an icosahedral quasicrystal can be built. Tilings in either 2D or 3D have helped a lot to understand the properties of quasicrystals especially from a mathematical point of view and are to this day the subject of many mathematicians' studies. However there are problems when dealing with actual real quasicrystals and only thinking in terms of tiles and their decoration. One of them is that when building for instance a penrose tiling not by inflation but by following the matching rules starting from a single tile, there will be a lot of paths that lead to dead ends were a further positioning of a tile is no longer possible. The number of possible dead ends increases rapidly as the tiling grows. Another problem is that a tiling can possess the correct symmetry elements as the real structure we want to solve but what the tiling does not tell us is where the atoms sit (the decoration of the tiling). This problem together with the fact that any tiling is only one infinitely many leads us directly to a higher dimensional approach as it already did for the incommensurately modulated structures. The merits of this approach are that it restores hidden correlations (the decoration for instance) and symmetries of quasiperiodic structures because it is rooted in reciprocal space which is directly accessible to us via scattering methods.

2.3 Higher-Dimensional Crystallography (Superspace revisited)

What we have done with incommensurately modulated structures in section 1.2 and 1.3 can be done in a very similar manner with quasicrystals. Namely embedding a reciprocal structure(scattering pattern) in a higherdimensional superspace and by cutting and projecting this superspace into real space obtaining an atomic or molecular structure. The main difference being that for quasicrystals the positions of atoms are no longer described by a modulation function in superspace but by discrete atomic surfaces. The cut of real space with these atomic surfaces in the orthogonal superspace gives the positions of atoms in real space. The concept is shown in Fig.6 note that we use the same definitions of axis and coordinates as in 1.3. The red strips in the in the plane spanned by a_{s1} and a_{s2} are the atomic surfaces. In reciprocal space it is very similar as for the incommensurate structures and the details can be found in the literature.[9]

When dealing with real icosahedral quasicrystals the superspace has 6 dimensions and the atomic surfaces become 3-dimensional and have the form of complex polyhedra(see examples in 2.4). The atomic surfaces contain not only positional information but also information on the atom type as well as its surroundings. So different atomic species in a ternary intermetallic icosahedral quasicrystal will have differently sized and shaped atomic surfaces.
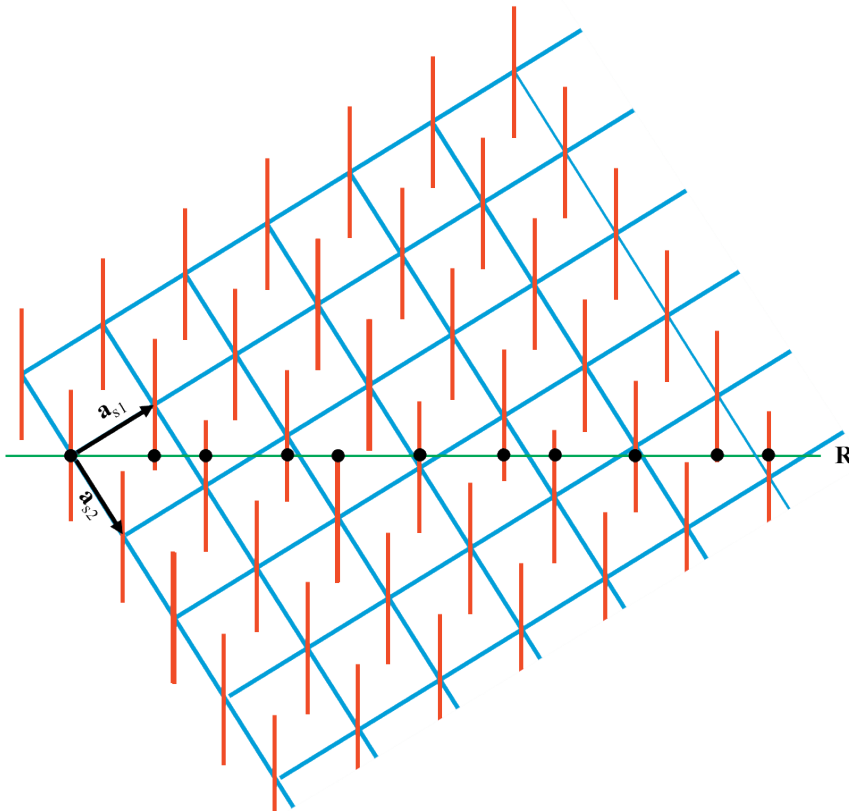

Fig. 6: Example of direct-space embedding of a 2D hyper crystal resulting in the Fibonacci sequence with the cut and project method. The red stripes are the discrete atomic surfaces of the 1D Quasicrystal in 2D hyperspace

What should become clear with this simple example is that the higher-dimensional approach to quasicrystals gives us periodic structure in a higher dimension and that this in turn allows for the use of well-established methods from crystallography to solve the structure. Still they are different from modulated structures in that they do not exhibit a periodic base structure of any kind in physical space.

2.4 Example Structures

12-fold axial quasicrystal

Axial quasicrystals that are aperiodic in only two dimensions and periodic in one perpendicular to it. The example presented here in Fig. 7 is the result of an MD simulation using the well-known Dzugotov pair potential [10]. On the left of the figure the QC is oriented with the twelvefold rotational axis in the plane of the paper below it the 2D scattering pattern along this direction is shown. Notice the twelve spots with 30° angles between each other showing nicely the 12-fold rotational symmetry. When the centers of the "tubes" along this direction are connected with lines an aperiodic tiling with 12 fold rotational symmetry is revealed. Structures very similar to this were found in block copolymer[11] and starlike micellar systems[12] as well as in binary VNi[13] alloys.
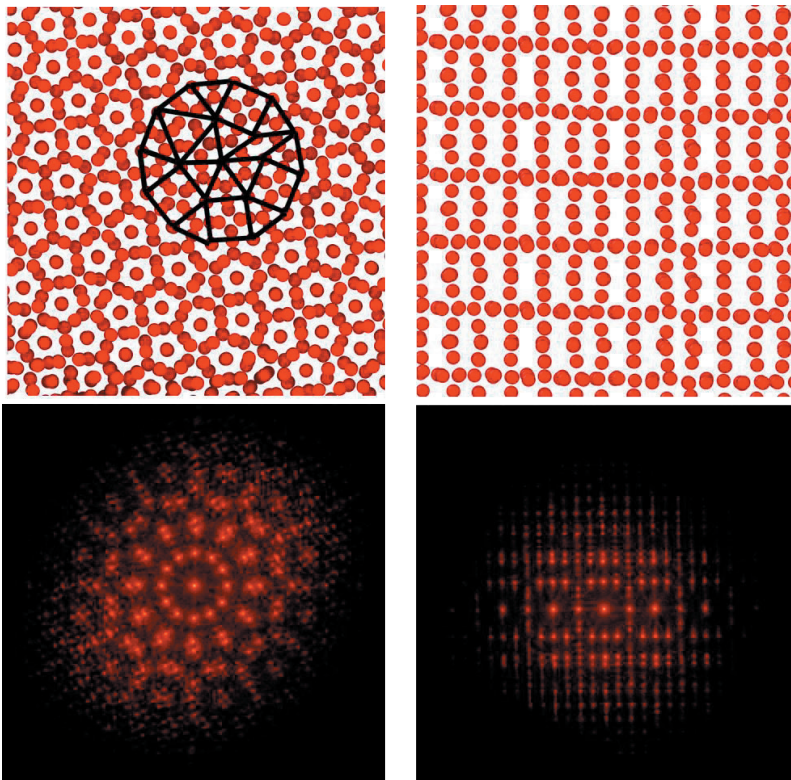
Fig. 7: Left: 12-fold rotational Quasicrystal viewed along the 12-fold axis and its 2D scattering pattern below Right: The same structure as left but viewed perpendicular to the 12 fold axis and the respective scattering pattern below. Notice that along this direction the structure is periodic

Icosahedral quasicrystal

This is an example of a one-component icosahedral quasicrystal. Also from MD simulation this time using a three well potential [14] to mimic the complex forces present in naturally occurring ternary intermetallic phases that form such icosahedral quasicrystals.
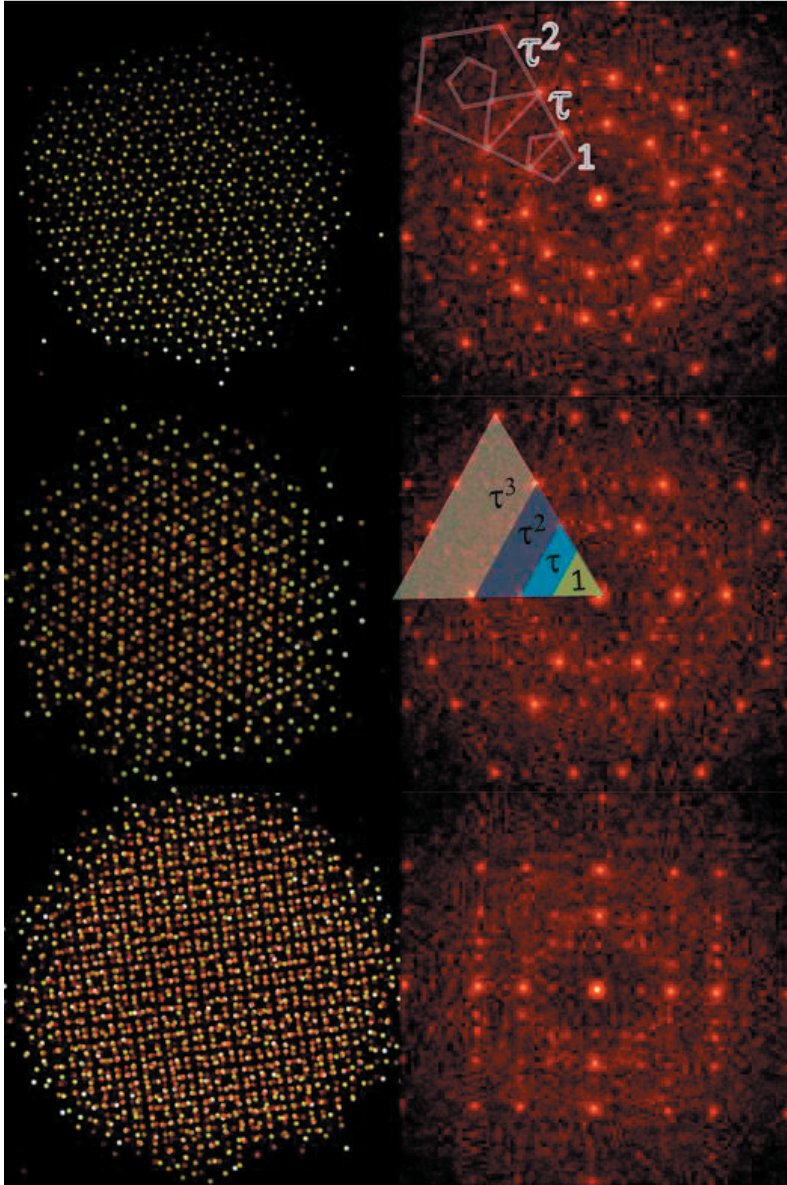
Fig. 8: Icosahedral Quasicrystal from molecular dynamics simulation top: look along the 5 fold rotational axis with the corresponding scattering right next to it. Middle: look along the 3 fold rotational axis with the scattering to the right. Bottom: look along the 2 fold rotational axis with the scattering right next to it One can nicely see the scaling with the golden mean in the scattering, but it naturally also shows up in real space but is not shown explicitly.

References

[1] W. Friedrich, P. Knipping and M. Laue, Annalen der Physik, **346**, 971-988 (1913)

[2] W. H. Bragg and W. L. Bragg, Proceedings of the Royal Society of London. Series A, **88**, 428-438 (1913)

[3] Donnay, J. D. H. Ann. Soc. Geol. Belg. B**55**, 222–230 (1935)

[4] Daniel, V. & Lipson, H. Proc. R. Soc. London, **181**, 368–377 (1943)

[5] Wolff, P. M. de Acta Cryst. A**30**, 777–785 (1974)

[6] Janner, A. & Janssen, T. Phys. Rev. B, **15**, 643–658 (1977)

[7] D. Shechtman, I. Blech, D. Gratias and J. W. Cahn, Physical Review Letters, **53**, 1951-1953 (1984)

[8] Penrose, R., Bull. Inst. Maths. & itsApplns. 10, **No. 7/8** 266-271 (1974)

[9] S. Walter and S. Deloudi, *Crystallography of quasicrystals: concepts, methods and structures*, Springer Science & Business Media, 2009

[10] M. Dzugutov, Physical Review Letters, **70**, 2924-2927 (1993)

[11] S. Fischer, A. Exner, K. Zielske, J. Perlich, S. Deloudi, W. Steurer, P. Lindner and S. Förster, Proceedings of the National Academy of Sciences, **108**, 1810-1814 (2011)

[12] X. Zeng, G. Ungar, Y. Liu, V. Percec, A. E. Dulcey and J. K. Hobbs, Nature, **428**, 157-160 (2004)

[13] H. Chen, D. X. Li and K. H. Kuo, Physical Review Letters, **60**, 1645-1648 (1988)

[14] M. Engel, P. F. Damasceno, C. L. Phillips and S. C. Glotzer, Nat Mater, **14**, 109-116 (2015).

B 3   The structure of complex fluids

H. Frielinghaus
Jülich Centre for Neutron Science
at Heinz Maier-Leibnitz Zentrum
Forschungszentrum Jülich GmbH

Contents

1   Introduction

Complex fluids are often defined by their mechanical response to external forces. So they are different from classical solids, which often only show elastic response, i.e. a Hookean spring-like behavior. And they are different from simple Newtonian fluids that are characterized by the viscosity only, a simple measure for the loss of energy during constant shear. A severe combination of both effects is the reality for complex fluids that display the so called visco-elasticity.

One prominent example for a complex fluid is ketchup. At slow shear rates it displays a high resistance against the deformation. After a threshold shear rate the fluid suddenly reduces the resistance and flows very well. This phenomenon is called shear thinning. At small shear rates the internal structure provides a locking mechanism that inhibits the flow, while at high shear rates this mechanism is overcome.

The reason for visco-elasticity is the nano-scale structure that differs from simple fluids only consisting of small molecules. The crystal-like order of classical solids is not reached to this high degree by the complex fluids. So, essential questions of complex fluids are: What is the nano-scale structure, i.e. what are the building blocks at this meso-scale? How do these building blocks arrange themselves on larger length scales, if at all? For these questions, scattering experiments are highly valuable because they provide direct insight. Hierarchically, the building blocks and the super-structures can be resolved in an ideal way.

Coming back to the example of ketchup, the locking mechanism could emerge from spherical colloidal structures that are tightly packed. Only after stronger shear the packing of the colloids can break up and give way to facilitated flow. For synthetic colloids this phenomenon is well explained while in food industry the exact mechanisms often remain a little unclear due to the many ingredients.

Generally, examples of complex fluids are polymer blends, rubber, colloidal suspensions, aqueous surfactant systems and microemulsions. In this lecture the focus of examples lies on amphiphilic systems. So molecules with a combined hydrophilic and hydrophobic property are involved. They tend to self-assembly, i.e. an arrangement of several molecules on the nano-scale. Representatives are surfactants and amphiphilic polymers that both form micelles if dissolved in water. As a third example, microemulsions as mixtures of oil, water and surfactant display similar structures as the binary examples that allows for similar methods in the analysis. The topic of pure polymer systems (i.e. blends and rubber) is covered by Chapter E2, and the understanding of colloidal suspensions is already included in the amphiphilic systems. The subject of classical solid particles is touched at the very end of this lecture.

In the lecture the concepts of aqueous surfactant systems and microemulsions are presented. Then the characterization methods are described with the emphasis on scattering methods. Finally, an example for the interaction of self-assembling systems with solid walls is presented with an outlook and possible applications.
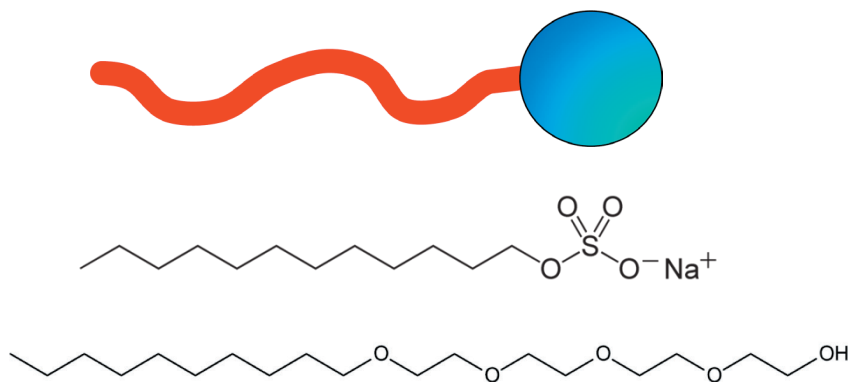
Fig. 1: Top: Conceptual drawing of a surfactant molecule. The hydrophilic head is blue while the hydrophobic tail is red. Middle: Chemical structure of sodium dodecyl sulfate (SDS). The head and tail groups are just below the conceptual drawing. Bottom: Chemical structure of tetraethyleneglycolmonodecylether ($C_{10}E_4$). The head group is four glycol groups long.

2 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 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 surfactants 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 suppress crystalline order of the hydrophobic tails [1]. 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 [2].

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

---

[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
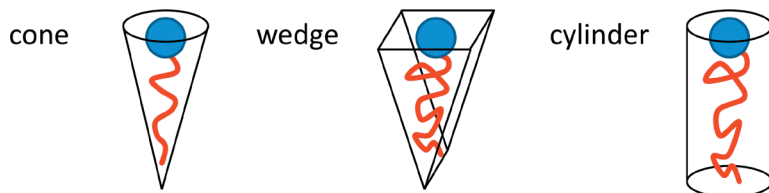
Fig. 2: Molecule geometries for different packing parameters.

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 [3] and so very precise values for the CMC can be given. The phenomenon 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 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 [4], 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 dimensionless 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. It shall be mentioned that this effect is called self-assembly already. 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 [5,6]. For the purpose of this lecture we restrict the considerations of P to a maximum of 1; otherwise reversed structures form (see below).

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
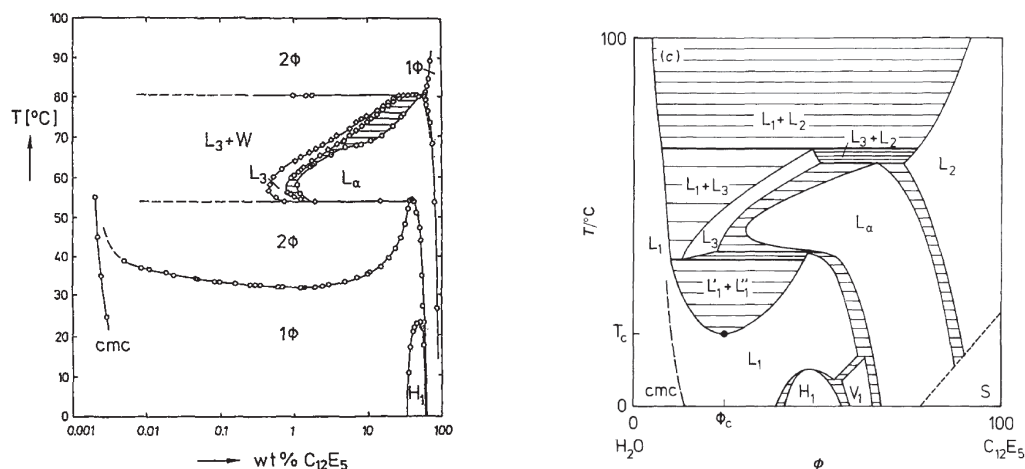
Fig. 3: An experimental phase diagram of the non-ionic surfactant $C_{12}E_5$ in water [7, 8]. Following a horizontal line at ca. 20°C from low to higher concentrations one finds the CMC at around 0.005% and the one-phase (1Φ) region. Below $\sim$1% the micelles are spherical and uncorrelated. Between 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 [7]. The abbreviations for the different phases are discussed in the text below (see also Table 2).

become cylindrical. Going up in temperature now, their length grows until the phase boundary at $\sim$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 concentrations (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 Coulomb-like interaction. Astonishingly, more ordered phases appear. The H_1-phase contains cylindrical micelles (as in 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
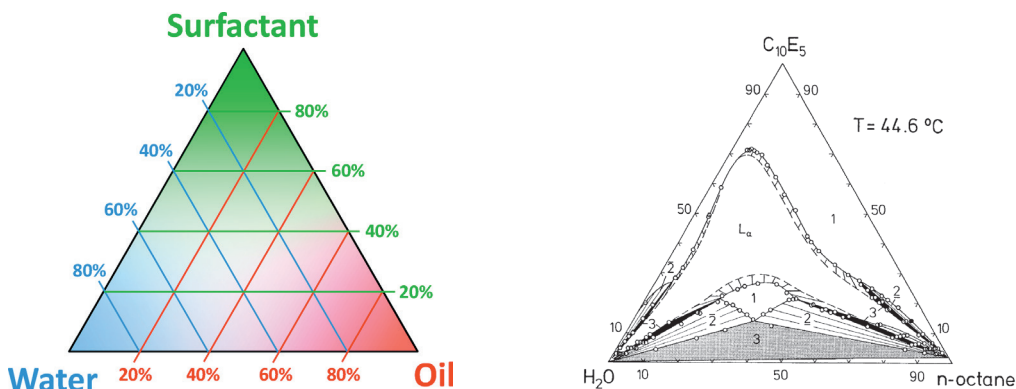
Fig. 4: Left: A scheme how to read off the compositions from the Gibbs Phase Triangle. The axis to the left of the corresponding component in the corner shows the scale of the fraction. The colored lines indicate constant composition. The whole diagram is completely symmetric. Right: A rather simple phase diagram for the system: H_2O, n-octane, $C_{10}E_4$ [9]. The numbers inside indicate the number of phases, and L_α indicates the lamellar phase. © Deutsche Bunsengesellschaft.

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-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 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. Furthermore, they are thermo-dynamically stable, which corresponds to a spontaneous formation.

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 bi-continuous 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 concen-trations 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 compo-nent inside. In principle, the reversed micelles are possible as well. As 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 [1] 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 [10]. 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:

$$F = \int dS \left(\gamma + \frac{1}{2}\kappa(c_1 + c_2 - 2c_0)^2 + \bar{\kappa}c_1c_2 \right) \tag{2}$$

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
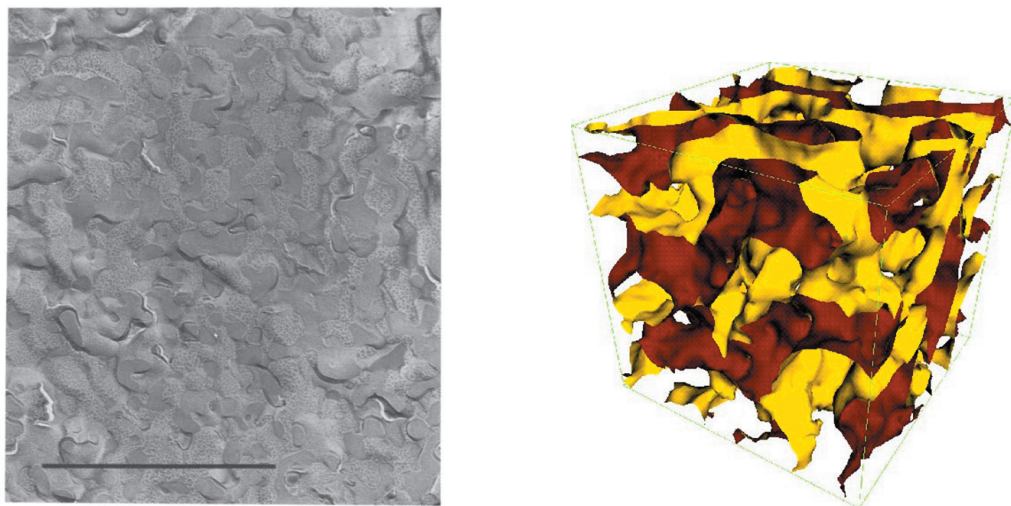
Fig. 5: Left: A transmission electron micrographof the microemulsion containing water, octane, and $C_{12}E_5$. The surfactant content was 7 wt%. The indicated bar shows a scale of 1μm. Reprinted from Ref. 11, Physica A: Statistical Mechanics and its Applications, Vol. 157, H.T. Davis, J.F. Bodet, L.E. Scriven, W.G. Miller, Microstructure and transport in midrange microemulsions, Pages 470-481, Copyright (1989), with permission from Elsevier. Right: A real space picture of the bicontinuous microemulsion according to computer simulations [12]. Actually the surfactant film is shown with the surface color being red for oil facing surface and yellow for water facing surface.
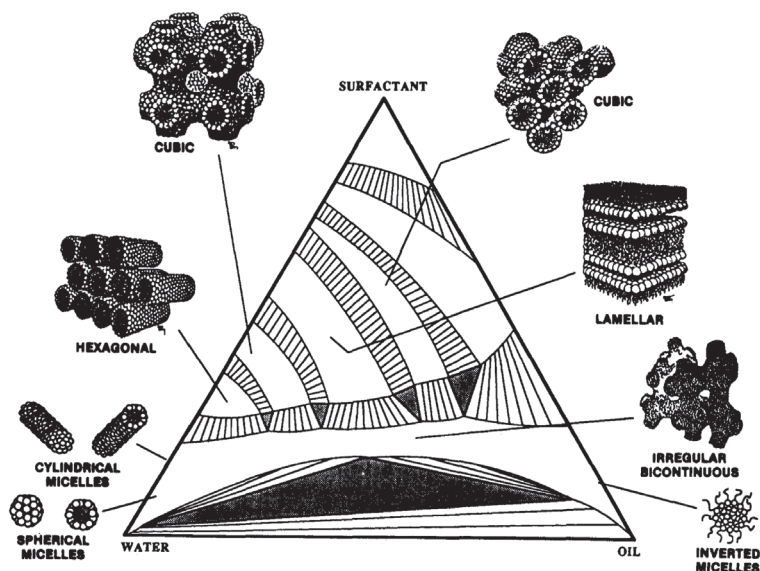
Fig. 6: A more schematic phase triangle with most typical lyotropic phases [13]. At the bottom there is the three-phase coexistence region (black) surrounded by two-phase coexistence regions. Above there is the large L_1-region with droplets, cylinders and the bicontinuous phase. In the upper half there 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. Reprinted from Ref. 13, with the permission of AIP Publishing.

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 $\kappa \approx 1..10 k_B T$ 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 [14] 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. He found that the three different shapes are separated by distinct phases. The same problem was described by Blokhuis [15] 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. $\omega = V_{\text{tot}}/S_{\text{tot}}$, 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 (called Helfrich interaction) 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 [16, 17] 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 stadion 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:

$$F_{\text{steric}} \propto c_0^{-2} \left(\frac{k_B T}{\kappa} \right)^2 \frac{\phi_{\text{oil}}^3}{(1 - \phi_{\text{oil}})^2} \tag{3}$$

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).

Another approach 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

Table 2: A survey about symbols for the different phases in aqueous surfactant systems and microemulsions. A first attempt was introduced by Tiddy [1]. Especially more exotic examples are given there. The second column gives symbols for polymeric systems [18], which will be discussed in a later section.

symbol used here	symbol (polymers)	alternative symbols	explanation
$1, 1\Phi$		L_1 or L_2	micelles & fluctuating bicontinuous phase
$2, 2\Phi$		$L_1' + L_1'', ...$	2 coexisting phases (need to be specified)
$\bar{2}, \underline{2}$			2 coexisting phases at high/low temperatures
$3, 3\Phi$		$L_1 + L_2 + L_3, ...$	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$_{\text{fcc,bcc}}$ with spherical micelles
I_2	C_2	Q_{II}	cubic phase$_{\text{fcc,bcc}}$ 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
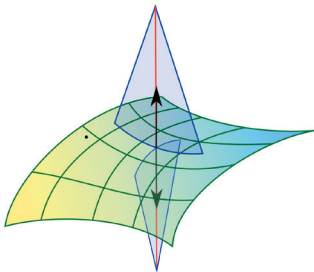
Fig. 7: An example of a surface with the two principal radii indicated. This construction can be done for any point of the surface.
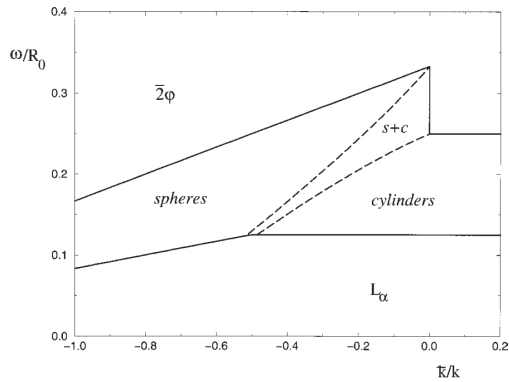
Fig. 8: Microemulsion phase diagram. The parameter ω/R_0 is given by the ratio of the total volume and the membrane surface and the equilibrium curvature, i.e. $\omega/R_0 = V_{\text{tot}}/S_{\text{tot}} \cdot c_0$. The x-axis shows the ratio of the two moduli $\bar{\kappa}/\kappa$. Reprinted from Ref. 15, with the permission of AIP Publishing.

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 [19]. 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 [19], and it reads:

$$F[\Phi] = \int dV \left(-\frac{\chi}{2}\Phi^2 + \frac{1-\Phi}{2}\ln\frac{1-\Phi}{2} + \frac{1+\Phi}{2}\ln\frac{1+\Phi}{2} - \frac{1}{2}(\nabla\Phi)^2 + \frac{1}{2}(\nabla^2\Phi)^2 - \mu\Phi \right) \tag{4}$$

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 [20]. 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. 4). 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 $\Phi = -1 + 2\phi_{oil}$. 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 4 is quite oversimplified to describe the complex behavior of microemulsions. So there exist more detailed approximations (see Refs. 21, 22), 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. 22 treats Coulomb interactions quite explicitly.
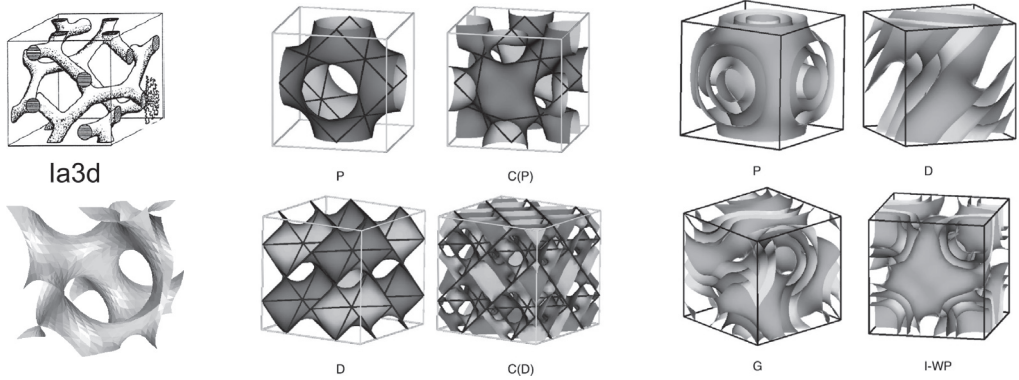
Fig. 9: Left: The most prominent structure: The gyroid phase G (G_1). Middle: The next most prominent structures in a cubic phase. Especially the D and P (V_1) are realized. Right: Double frequency structures. These structures are not realized often. Left figure reprinted with permission from Ref. 17, U.S. Schwarz, G. Gompper, Physical Review Letters, 85, 1472-1475, 2000. Copyright (2000) by the American Physical Society. Middle and right figures reprinted with permission from Ref. 16, U.S. Schwarz, G. Gompper, Physical Review E, 59, 5528-5541, 1999. Copyright (1999) by the American Physical Society.
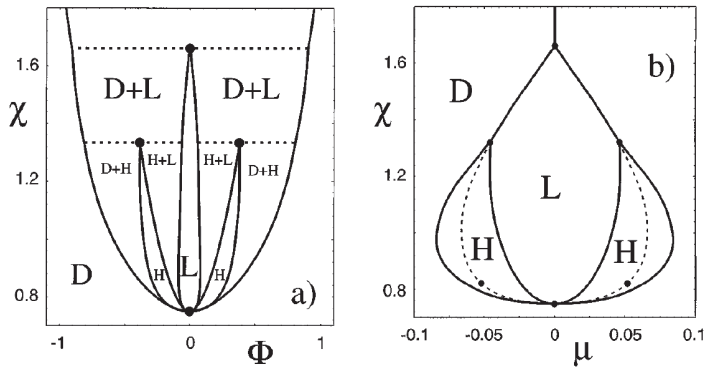
Fig. 10: Two-dimensional bulk phase diagram, showing disordered (D), lamellar (L), and hexagonal (H) phases, as a function of the interaction parameter χ. The x-axis is spanned by (a) the average order parameter $\Phi = -1 + 2\phi_{\text{oil}}$, and (b) the chemical potential μ. Dashed lines in (a) denote triple lines and dashed lines in (b) denote the (metastable) L-D transitions, which exhibit tricritical points (denoted by solid circles). Figures reprinted with permission from Ref. 19, R.R. Netz, D. Andelman, M. Schick, Physical Review Letters, 79, 1058-1061, 1997. Copyright (1997) by the American Physical Society.

4   Phase Diagram and Structure Determination

Optical measurements base on different observations: Turbidity, meniscus, and depolarization. Turbidity indicates a two- or three-phase coexistence, which is a quickly obtained criterion. The observation of the meniscus after long waiting times allows to determine the ratio of the volumes precisely. This is for instance helpful to determine the three-phase coexistence boundary (see Fig. 4). By changing the composition (or temperature) and following the minority phase to vanish the boundary to the two-phase coexistence is extrapolated. When observing a one-phase sample with crossed polarizers one finds some phases, which depolarize the light while most of the phases keep the polarization. The reason for depolarization is elongated domains, which give rise to different propagation for different polarizations. Because most of the samples are 'powder-like', i.e. with many grains, the light is simply depolarized. Elongated domains are found for the hexagonal phases H_1 and H_2, and for the lamellar phase. Please note that cubic phases are isotropic and do not give rise to depolarization. Another, not really optic criterion is also the viscosity. By shaking the sample tube, one gets rough information about the viscosity and thus can speculate about the ordered phases.

The real scientific field behind viscosity is *Rheology*. Usually, the sample is placed between two plates. One of the plates is oscillating either elongationally or rotationally with relatively small excitations. For rotations the plates do not need to be parallel. One plate is often conical to have the same shear for all radii. The still plate takes up the force. There the amplitude and phase shift is measured with respect to the moving plate. In most cases the amplitude is used to characterize the different phases. This method is rather indirect because the absolute values often do not tell the difference. But it is plausible that lamellar phases and hexagonal cylinder phases allow the domains to slide with small losses while cubic phases are rather rigid. The L_1 and L_3 phase usually show the lowest viscosity. Low temperatures can lead to rather rigid phases in parallel simply by raising the viscosity of oil and water.

Another method often used is *nuclear magnetic resonance* (NMR). Mostly the isotropic phases are distinguished from the phases with elongated domains [23]. The anisotropic phases show a quadrupolar splitting of an otherwise single line. Some more examples for NMR measurements are given in Ref. 1. Please note that the CMC can also be determined by NMR [3] as mentioned before.

4.1   Scattering Methods

By small angle neutron and x-ray scattering one can distinguish ordered phases quite well (see also Chapter D1). With neutrons limited resolution often does not allow to distinguish all higher order peaks. This is why x-rays are most often preferred. The principle behind this method is powder diffraction because usually the domains are not oriented and so many grains with different orientation coexist. Especially, cubic phases cannot be ordered by external fields. So, Debye-Scherrer rings are observed on a 2-dimensional detector. After averaging the intensities on concentric circles the intensity is simply shown as a function of the modulus of the scattering vector q. The observed peaks are indexed according to multiples of the q-value with respect to the first peak with the lowest q. These are either square roots of natural numbers or of simple fractions. In Soft Matter research these ratios are simply used as labels for the peaks instead of the full information of the Miller indices (h, k, l). Nonetheless, the experimentalist needs to

Table 3: Most important ordered structures. The reciprocal space structure consists always of delta-like peaks and can be taken as a structure factor; then the real space structure is assumed to be infinitely thin. The lattice points of reciprocal space are given by Miller indices and are sorted by vector length (as given in the third column). Colored dots give connection to Miller indices; the origin is always indicated by a black dot.
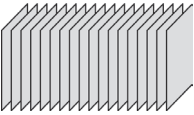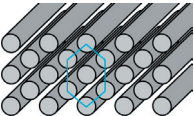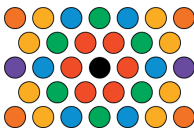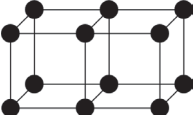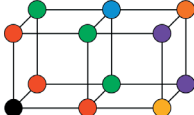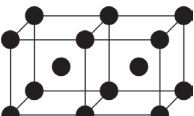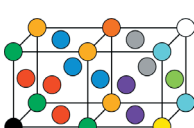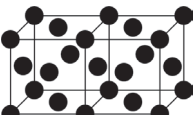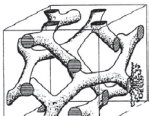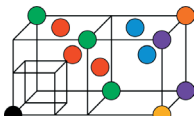
real space	reciprocal space	lattice points of reciprocal space (Miller indices)
lamellar	(lamellar)	$(h,0,0)$ ● $(1,0,0)$ 1 ● $(5,0,0)$ 5 $(9,0,0)$ 9 ● $(2,0,0)$ 2 $(6,0,0)$ 6 $(10,0,0)$ 10 ● $(3,0,0)$ 3 $(7,0,0)$ 7 $(11,0,0)$ 11 ● $(4,0,0)$ 4 $(8,0,0)$ 8 ...
hexagonal	hexagonal	$(h,k,0)$ ● $(1,0,0)$ 1 ● $(2,2,0)$ $\sqrt{12}$ $(3,3,0)$ $\sqrt{21}$ ● $(1,1,0)$ $\sqrt{3}$ $(3,1,0)$ $\sqrt{13}$ $(5,0,0)$ $\sqrt{25}$ ● $(2,0,0)$ $\sqrt{4}$ $(4,0,0)$ $\sqrt{16}$ $(4,2,0)$ $\sqrt{28}$ ● $(2,1,0)$ $\sqrt{7}$ $(3,2,0)$ $\sqrt{19}$ $(5,1,0)$ $\sqrt{31}$ ● $(3,0,0)$ $\sqrt{9}$ $(4,1,0)$ $\sqrt{21}$ $(4,3,0)$ $\sqrt{31}$
sc	sc	(h,k,l) ● $(1,0,0)$ 1 ● $(2,1,1)$ $\sqrt{6}$ $(3,1,1)$ $\sqrt{11}$ ● $(1,1,0)$ $\sqrt{2}$ $(2,2,0)$ $\sqrt{8}$ $(2,2,2)$ $\sqrt{12}$ ● $(1,1,1)$ $\sqrt{3}$ $(2,2,1)$ $\sqrt{9}$ $(3,2,0)$ $\sqrt{13}$ ● $(2,0,0)$ $\sqrt{4}$ $(3,0,0)$ $\sqrt{9}$ ... ● $(2,1,0)$ $\sqrt{5}$ $(3,1,0)$ $\sqrt{10}$
bcc	fcc	(h,k,l) and $(h+\frac{1}{2},k+\frac{1}{2},l)$ ● $(\frac{1}{2},\frac{1}{2},0)$ $\sqrt{\frac{1}{2}}$ ● $(1,1,1)$ $\sqrt{3}$ ● $(2,1,0)$ $\sqrt{5}$ ● $(1,0,0)$ $\sqrt{1}$ ● $(\frac{3}{2},1,\frac{1}{2})$ $\sqrt{\frac{7}{2}}$ $(2,\frac{3}{2},\frac{1}{2})$ $\sqrt{\frac{13}{2}}$ ● $(1,\frac{1}{2},\frac{1}{2})$ $\sqrt{\frac{3}{2}}$ ● $(2,0,0)$ $\sqrt{4}$ $(\frac{5}{2},\frac{1}{2},0)$ $\sqrt{\frac{13}{2}}$ ● $(1,1,0)$ $\sqrt{2}$ ● $(2,\frac{1}{2},\frac{1}{2})$ $\sqrt{\frac{9}{2}}$ ... ● $(\frac{3}{2},\frac{1}{2},0)$ $\sqrt{\frac{5}{2}}$ $(\frac{3}{2},\frac{3}{2},0)$ $\sqrt{\frac{9}{2}}$
fcc	bcc	(h,k,l) and $(h+\frac{1}{2},k+\frac{1}{2},l+\frac{1}{2})$ ● $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ $\sqrt{\frac{3}{4}}$ ● $(2,0,0)$ $\sqrt{4}$ $(\frac{3}{2},\frac{3}{2},\frac{3}{2})$ $\sqrt{\frac{27}{4}}$ ● $(1,0,0)$ $\sqrt{1}$ $(\frac{3}{2},\frac{3}{2},\frac{1}{2})$ $\sqrt{\frac{19}{4}}$ $(2,2,0)$ $\sqrt{8}$ ● $(1,1,0)$ $\sqrt{2}$ ● $(2,1,0)$ $\sqrt{5}$ $(\frac{5}{2},\frac{3}{2},\frac{1}{2})$ $\sqrt{\frac{35}{4}}$ ● $(\frac{3}{2},\frac{1}{2},\frac{1}{2})$ $\sqrt{\frac{11}{4}}$ ● $(2,1,1)$ $\sqrt{6}$ ... ● $(1,1,1)$ $\sqrt{3}$ $(\frac{5}{2},\frac{1}{2},\frac{1}{2})$ $\sqrt{\frac{27}{4}}$
gyroid Ia$\bar{3}$d		$(h,k,0)$ with $h+k+l=2n$ and further restrictions ● $(2,1,1)$ $\sqrt{6}$ $(3,3,2)$ $\sqrt{22}$ $(6,1,1)$ $\sqrt{38}$ ● $(2,2,0)$ $\sqrt{8}$ ● $(4,2,2)$ $\sqrt{24}$ $(5,3,2)$ $\sqrt{38}$ ● $(3,2,1)$ $\sqrt{14}$ $(4,3,1)$ $\sqrt{26}$ $(6,2,0)$ $\sqrt{40}$ ● $(4,0,0)$ $\sqrt{16}$ $(5,2,1)$ $\sqrt{30}$ $(5,4,1)$ $\sqrt{42}$ ● $(4,2,0)$ $\sqrt{20}$ $(4,4,0)$ $\sqrt{32}$ ...

Table 4: Form factors in connection with the liquid crystalline structures discussed in Table 3. The lamellar structure factor arises from a 1-dim, the cylindrical from a 2-dim, and the spherical from a 3-dimensional calculation. J_1 is the Bessel function of first kind and first order. d indicates the lamellar domain thickness, and r the radius of the cylinder and the sphere.

lamellar	$F(q) = \left(\dfrac{\sin(qd/2)}{qd/2} \right)^2$
cylindrical	$F(q) = \left(2\dfrac{J_1(qr)}{qr} \right)^2$
spherical	$F(q) = \left(3\dfrac{\sin(qr) - qr\cos(qr)}{(qr)^3} \right)^2$
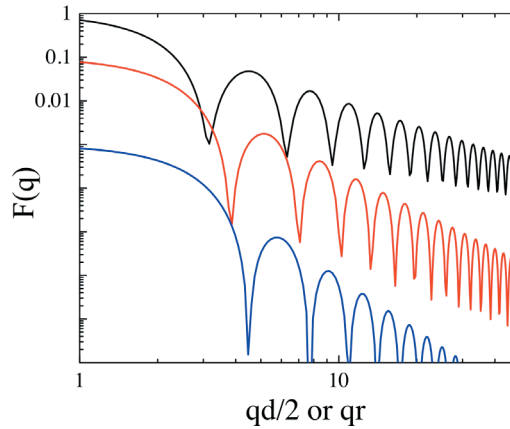
Fig. 11: Double logarithmic plots of the form factors $F(q)$ for lamellae (black), cylinders (red), and spheres (blue). For a clear representation the curves are shifted down by a factor of 1, 0.1 and 0.001. The original functions are normalized to 1 for $q \to 0$.

know the crystallography [24] behind in order to exclude some ratios, which do not exist. In Table 3 many important examples are shown.

The lamellar phase expands in two dimensions and the periodicity usually extends in z-direction (with a distance d). The reciprocal lattice is periodic in z as well. The distances are given by $q = 2\pi/d$. In reciprocal space the lattice is formed by delta-function-like peaks. In x, y-direction it arises from the infinite dimensions, and in z the strict periodicity is the reason.

The hexagonal phase consists of infinitely long cylinders (or rods) in z-direction, while in the x, y-plane the structure is periodic. The reciprocal structure is periodic in x, y (note: the lattice is turned by 60°) while the delta-like peaks do not expand in z-direction. The number of peaks with the same distance from the origin can be counted according to $n_{h,k,l} = 12$ for $h > k > 0$ and $n_{h,k,l} = 6$ otherwise. This frequency has to be taken into account for the intensities of the Debye-Scherrer rings.

For the cubic structures the following shall be mentioned: The simple cubic (sc) lattice stays simple cubic in reciprocal space. The body centered cubic (bcc) lattice becomes a face centered cubic (fcc) lattice and vice versa. The gyroid phase is presented as a more exotic phase [25]. The number of peaks of the cubic lattices can be counted according to $n_{h,k,l} = 2^s \cdot c$, where s counts the number of h, k, l different from zero (so here the sign matters), and c is a combinatorial factor being 1 for $h = k = l$, 3 for $h \neq k = l$, and 6 for $h \neq k \neq l$.

The nomenclature I_1 for the cubic phase is often quite sloppy. The name actually arises only from the word isotropic but the actual structure behind can be simple cubic, body or face centered cubic. One publication [26] demonstrates this problem for even more complicated cubic structures. Furthermore, Brownian motion – as for microemulsions – can destroy the long range order and so only fluid-like near-order is achieved. This is usually expressed in missing higher order peaks, and no crystal-like order can be determined. Contrarily, the gyroid phase is usually well distinguished from the cubic phase with globular domains, but this phase only appears for polymer based systems.

For indexing the Debye-Scherrer rings Table 3 gives a good basis. So there are many examples in the literature, which argue mainly on the basis of the appearing peaks [18]. For a quick reference this technique might be sufficient. Nonetheless, the analysis can go one step beyond if a model is assumed for the 3-dimensional structure of the domains. This means, that the lamellae have finite thickness, the cylinders have a finite diameter, and the spherical domains on the cubic lattice have a finite diameter. For such structures one then splits the problem in two parts: The periodic structure is given by Table 3, while the single domain is described by a form factor.

$$I(q_{h,k,l}) \propto S(q_{h,k,l}) \cdot F(q_{h,k,l}) \tag{5}$$

In this equation the structure factor $S(q)$ arises from the delta-function peaks of the reciprocal lattice. The form factor $F(q)$ is connected to the 3-dimensional structure of the single domain. Very often, equation 5 can be simplified to an orientationally averaged form according to:

$$I_{\mathrm{av}}(q) \propto \frac{1}{q^2} \cdot \sum_{q_{h,k,l}=q} n_{h,k,l} \cdot F(q) \tag{6}$$

Table 4 gives a list of form factors for differently shaped domains. The functions are plotted in Fig. 11. It becomes clear now that zeros of the formfactor might result in further eliminations of Debye-Scherrer rings. So the full list of Bragg peaks gives the maximum set of peaks, and the exact form factor might cause further eliminations, which can be interpreted as a higher degree of symmetry.

A very good example for peak indexing is discussed now in context with Fig. 12. The system is a polybutadiene-polyethyleneoxide (PB(1,4)-PEO) diblock copolymer in a melt. The phase separation leads to spherical domains of PEO hosted in the PB matrix. At high temperatures (203°C) the spherical domains show only a near order. The repulsive interactions are modeled by a Perkus-Yevic structure factor (excluded volume interaction). Upon lowering the temperature the liquid crystalline order becomes more prominent, which is indicated by a larger number of higher order peaks. Finally, at 100°C the best order is achieved, and the highest order visible is found at a relative q value of $\sqrt{9}$.
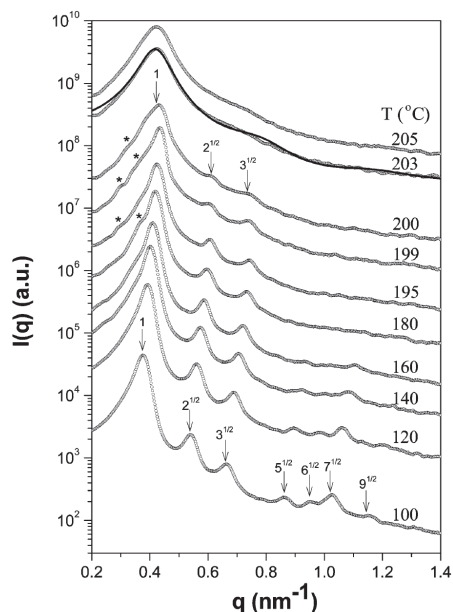
Fig. 12: In this example small angle x-ray scattering curves $I(q)$ are shown as a function of temperature [27]. For different temperatures different factors scale the curves down for better visibility. At the lowest temperature (100°C) the order is developed best, and many peaks indicate the face centered cubic order of spherical entities. The numbers give the ratios of the peak q-values. Reprinted with permission from Ref. 27. Copyright (2007) American Chemical Society.
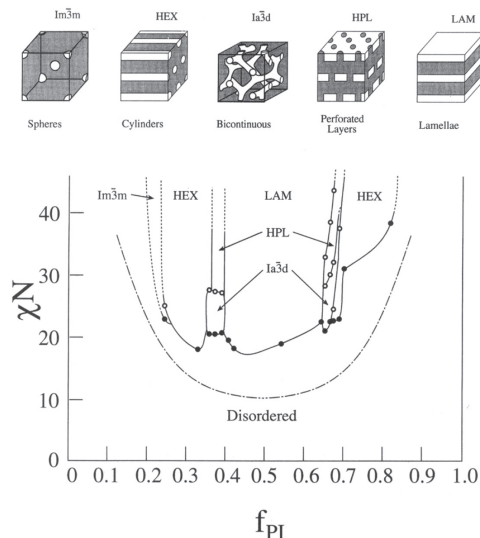
Fig. 13: Experimental phase diagram of a diblock copolymer melt polystyrene-polyisoprene (PS-PI). The scaled interaction parameter χ is plotted against the chain length ratio f_{PI}. The ordered phases are indicated as LAM for lamellar (L_α), Ia$\bar{3}$d for gyroid (G_1 and G_2), HEX for hexagonal (H_1 and H_2), and Im$\bar{3}$m for the body centered cubic phase (I_1 and I_2). The hexagonally perforated lamellar structure (HPL) is not found consistently for all polymers. Reprinted with permission from Ref. 28. Copyright (1995) American Chemical Society.

It shall be mentioned that at this stage of the theory all thermal fluctuations are neglected. Fluctuations will lead to undulations of the domain surface. This can lead to suppression of higher order Bragg peaks. For instance for the lamellar order of a surfactant film or bilayer the theory of Caillé describes the scattering. A good review can be found in Ref. 29 and further developments in Ref. 30. A simpler approximation for undulations of a domain surface can be achieved by a multiplication of a Gaussian to equation 6. Another appearance of disorder is obtained by non-perfect repetitions of the crystalline order. Then the paracrystal model [31] needs to be taken into account. The result is broader Bragg peaks for the higher orders with lower peak intensities. So any kind of disorder leads to weaker higher order Bragg peaks. It shall be mentioned that the Caillé theory actually includes both ways of disorder. Please note that the described structures are also found for diblock copolymer blends (see Fig. 13) and more complicated polymer blends with more components or blocks. The principal phase diagram (see Fig. 10) compares well with the experimental one.

If crystallographical violations appear in terms of peak positions or even more drastically in the allowed symmetry (for instance 5-fold symmetry) then the underlying structure is called a quasi-crystal (see Chapter B2). The realized structure is then connected to a projection of a crystal in more dimensions to the real 3 dimensions.

4.2   Scattering of the Bicontinuous Phase

The scattering of ordered phases is unique and supports the identification of the different phases. If the long-range order is lost higher order peaks are suppressed. One ultimate example is the bicontinuous microemulsion which develops only one peak in the scattering (Fig. 14). The scattering of the bicontinuous microemulsion has been observed quite early while the existence of real *bi*-continuous structures was still under debate. The bare scattering experiment only supported the arguments for the bicontinuous structure, and conductivity measurements and transmission electron micrographs (Fig. 5) needed to follow in concert with theoretical explanations. However, the bicontinuous structure is accepted today and scattering experiments extract the most essential structural parameters. The pronounced peak at a scattering vector q^* is connected to 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. Thus, the forward scattering is 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 $S_{\text{tot}}/V_{\text{tot}}$, and, therefore, is proportional to the membrane volume content ψ. The overall scattering function is well described by the following formula:

$$I(q) = \frac{d\Sigma}{d\Omega}(q) = \left(\frac{8\pi\langle\nu^2\rangle/\xi}{q^4 - 2(k_0^2 - \xi^{-2})q^2 + (k_0^2 + \xi^{-2})^2} \right.$$
$$\left. + \frac{G \cdot \text{erf}^{12}(1.06qR_g/\sqrt{6})}{1.5q^4R_g^4} \right) \exp(-\sigma^2q^2) + b_{\text{backgr}} \quad (7)$$

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 7 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 [32]. The term arises from a Landau description of the order parameter, similar to eq. 4. 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 [34]. 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
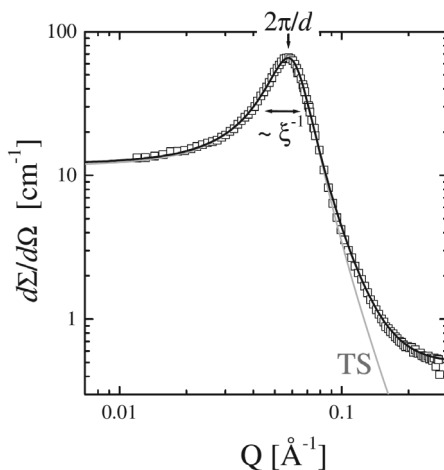
Fig. 14: A typical scattering pattern of a bicontinuous microemulsion (intensity vs. q) in bulk contrast, i.e. with D_2O, hydrogenated oil and hydrogenated surfactant. From the peak position the domain spacing is derived, while the peak width indicates the correlation length. The grey curve shows the fit with the Teubner-Strey theory. The solid line is a fit with the extended theory of equation 7. Reprinted with permission from Ref. 33. Copyright (2007) American Chemical Society.

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 fraction in equation 7 describes additional surface [33], 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 [35]. 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 structural smearing. 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. Apart from the bulk contrast (Fig. 14) with water contrasted against oil, the film contrast scattering does not display a peak [34] anymore. The analysis is more subtle and will not be discussed here.

4.3   Scattering of Polymeric and Worm-Like Micelles

While a scattering peak provides simple and unique information about the structure, for diluted phases this high level of uniqueness is not reached, but usually the dimensions of the micelles
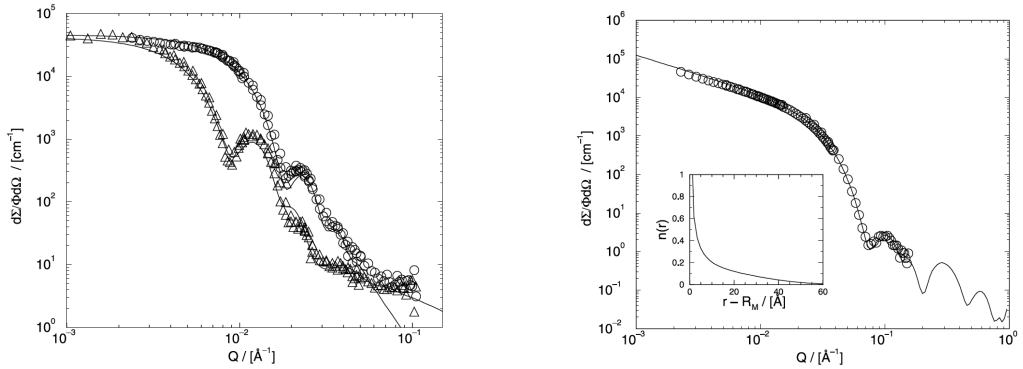
Fig. 15: Left: Scattering of spherical polymer micelles [36] from PEP_{10}-PEO_{10} and PEP_{22}-PEO_{22}. The different curves arise from full contrast (downturn at lower/higher q for higher/lower molecular mass). Note: The forward scattering can be obtained since the curves are constant for low q. Right: Scattering of cylindrical polymer micelles from PEP_1-PEO_1. The power law q^{-1} at low q is characteristic, which indicates the 'infinitely' long domains. The inset presents the monomer density as a function of the radius. Reprinted by permission from: Springer Nature, Applied Physics A: Materials Science & Processing, Micellization of symmetric PEP-PEO block copolymers in water: molecular weight dependence, H. Kaya, L. Willner, J. Allgaier, J. Stellbrink, D. Richter, Ref. 36, Copyright (2002).

can be extracted. So some examples will be discussed in the following.

The polymer PEP-PEO can be dissolved in water [36]. Then the hydrophobic PEP forms a compact core while the water-soluble PEO forms a dilute corona in the aqueous phase. As we know from simple amphiphiles there is a phase transition between spherical and cylindrical micelles. For the high molar masses spherical domains are obtained. This can be judged by the forward scattering, which is reached within our q-window. Detailed modeling of course supports the finding. For the low molar masses cylindrical domains are found. A characteristic q^{-1}-power law is found in the q-range 0.002 to 0.02Å^{-1}. This indicates relatively rigid cylindrical structures. The forward scattering is not reached in the experimental q-range, but in principle at some point the curve would reach a plateau with a finite forward scattering. So the presented experiments cannot judge about the length of the cylinders. The inset of Fig. 15b gives a radial distribution of the monomer density of the corona. In principle it is quite dilute and spans a relatively wide range to large distances. This finding supports the picture of a compact core and a dilute corona.

The best way to get access to the core and shell structure is small angle neutron scattering. The core polymer block is deuterated while the corona block is protonated. By using different compositions of H_2O and D_2O each structure can be made visible. In the full contrast (H_2O) basically the whole polymer is visible. For the core, intermediate and shell contrast, increasing amounts of D_2O were used. Results of such a study are shown in Fig. 16. In the left figure, almost spherical micelles are found because the scattering curves are nearly flat at small q. In the right figure, cylindrical micelles are found, which is indicated by the q^{-1}-power law at small scattering vectors q. The size of the compact core and the dilute corona can be directly seen from such figures. The downturn to the first minimum appears at smaller q for the core than for the corona. Furthermore, the interesting case of nearly zero average contrast is observed. Then, the forward scattering is ideally zero, while the scattering at finite q indicates alternating
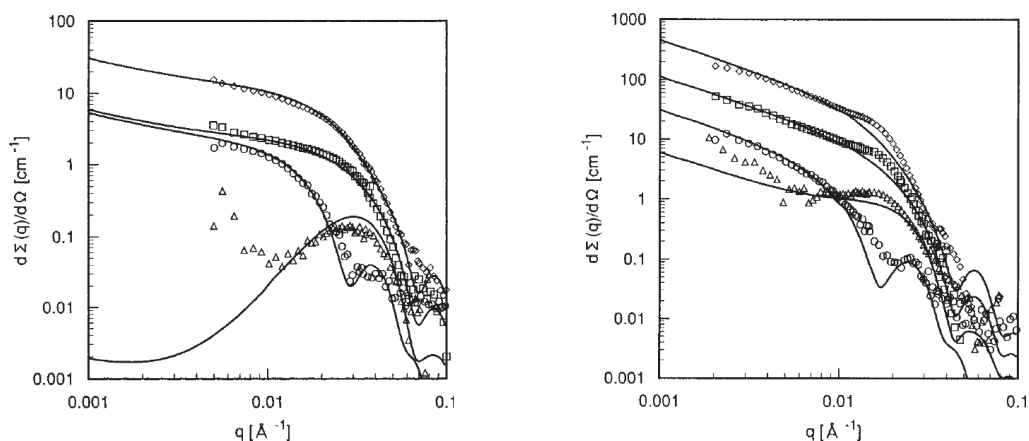
Fig. 16: Left: Scattering of almost spherical micelles of the polymer Poly(HOVE-b-NBVE) with molar masses of 7.5 and 3 kg/mol of the core and corona forming blocks. The slight inclination at small q indicates a small fraction of cylindrical micelles. The different symbols $\diamond$, $\square$, $\triangle$, and $\circ$ indicate the aqueous solvent being H_2O to D_2O. The corresponding contrasts are called full, core, intermediate and shell contrast. Interestingly, the intermediate contrast reaches nearly the zero-average contrast condition, which is indicated by a very low forward scattering. Right: Scattering of cylindrical micelles of the polymer Poly(HOVE-b-NBVE) with molar masses of 7.5 and 6 kg/mol of the core and corona forming blocks. The contrast conditions are the same as before. Reprinted with permission from Ref. 37. Copyright (1999) American Chemical Society.

structures from protonated-deuterated-protonated regions (according to the corona-core-corona structure). Macroscopically, the degree of deuteration in the whole sample is the same while locally the micelle structure appears. From the scattering contrast variation experiment and the simultaneous description of the scattering curves one obtains more precisely the different structures of the core and corona. In the simplest case one finds two radii. Especially for the corona one usually needs to assume a rather wide distribution of the segments, and so the model usually carries more information than a simple radius.

A theoretical model for polymeric micelles was formulated by Halperin [38]. It bases on three contributions to the free energy. The first contribution considers the corona. Close to the core the polymers are anchored densely enough such that a polymer gel is formed. This gel has a typical mesh size, which increases to the outer regions. Each polymer part inside such a mesh is called a blob and contributes with $1\,k_BT$ to the free energy. The second contribution comes from the polymer stretching inside the core. Here, the equilibrium chain dimension of a melt has to be compared with the polymer size in the real core. The polymer is seen as an entropic spring. The relative size gives rise to an entropic contribution to the free energy. The last contribution is the surface tension of the core with respect to the aqueous region. Here, simply the core surface times a classical surface tension yields the enthalpic contribution to the free energy. From such a model the free energy is minimized to yield the aggregation number N_{agg}, from which all dimensions and the shape of the micelle can be determined.

So far we have only described rigid cylindrical micelles. Locally, very often the micelles stay rigid while on larger length scales the fluctuations lead to the loss of correlations. In the extreme case, then the micelle follows a random walk. This concept is also known from polymers
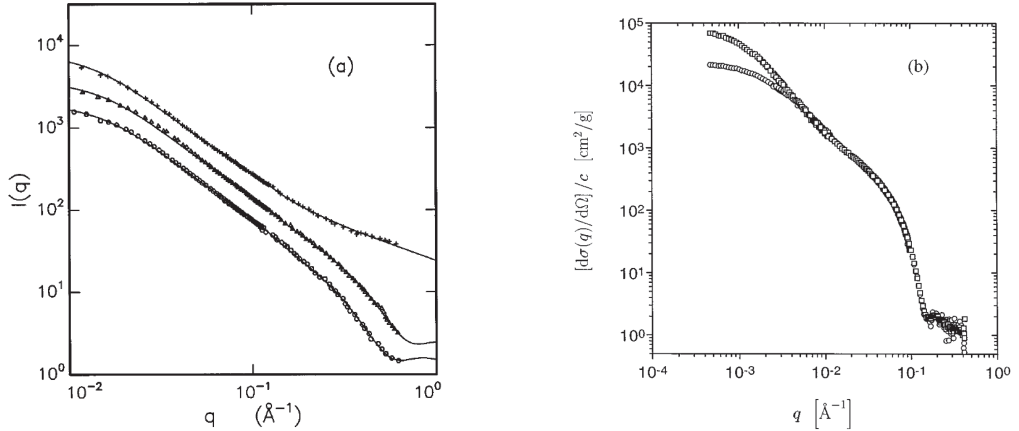
Fig. 17: Left: The polymer polystyrene in the good solvent CS_2. The three curves are obtained for deuterated backbone (**+**), deuterated phenyl ring ($\triangle$), and fully deuterated chain ($\circ$). The polymer is a worm-like object and resembles worm-like micelles with a small cross section. At low $q < 0.02\text{Å}^{-1}$ the Guinier scattering observes the whole polymer. At intermediate q up to 0.2Å^{-1} the wormlike power law with $q^{-1.67}$ is observed. At high q the finite diameter of the polymer becomes visible. Reprinted with permission from Ref. 39. Copyright (1996) American Chemical Society. Right: The system water/d-isooctane/lecithin with a molar water-to-lecithin ratio of 2.5 and lecithin concentrations of 0.9 mg/ml ($\circ$) and 3.5 mg/ml ($\square$) [5, 6]. Wormlike micelles are formed. At low $q < 0.001\text{Å}^{-1}$ the Guinier scattering observes the whole micelle. At intermediate q up to 0.02Å^{-1} the wormlike power law with $q^{-1.67}$ is observed. At slightly higher q up to 0.05Å^{-1} the local rigidity is indicated by a q^{-1}. At high q the finite diameter of the micelle is observed. Reprinted with permission from Ref. 6 as follows: G. Jerke, J.S. Pedersen, S.U. Egelhaaf, P. Schurtenberger, Physical Review E, 56, 5772-2788, 1997. Copyright (1997) by the American Physical Society.

(and diffusion processes). A polymer in a solvent with zero exchange interaction (theta solvent) shows the fractal structure of the polymer with a q^{-2} dependence. Here the self-similarity of sub-sections with the whole polymer is expressed. The exponent -2 arises from the following: A random walk lacks correlations between individual steps. For the ensemble averaged squared end-to-end distance R_{ee}^2 one obtains the proportionality to the number of steps N (see also Chapter E2). Thus, the segment number as a function of the length scale carries the exponent 2. The scattering intensity is proportional to the number of scatterers while the length scale is connected to the reciprocal q vector. Fig. 17a shows the example of a polymer in a good solvent. Accordingly, the power law changes from q^{-2} to $q^{-1.67}$ for the Flory chain. The excluded volume interactions are responsible for this power law, which means the intersections of the chain appear in the free energy as an important contribution. The other example of Fig. 17b shows the scattering experiment from an inverse micelle formed by water/d-isooctane/lecithin. Again, a clear power-law for the worm-like micelle with excluded volume interactions is observed. Apart from that, a small q-range at slightly higher q indicates the still locally rigid micelle. Both examples are very good, because the whole object is observed in the Guinier range at lowest q. From the forward scattering $I(q \rightarrow 0)$ the connection to the concentration ϕ and the volume of the particle (i.e. polymer or micelle) can be made. In the simplest case one finds $I(q \rightarrow 0) \propto \phi V_{\text{particle}}$. In references [5, 6] a more complicated renormalization theory is
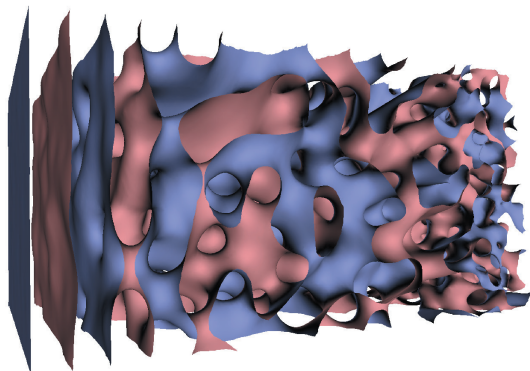
Fig. 18: A real space image of a microemulsion near a planar hydrophilic wall from a computer simulation. The surfactant layer is depicted with blue and red facing the water and oil domains. Close to the surface a lamellar order is formed while in the volume the microemulsion is bicontinuous. Reprinted with permission from Ref. 40 as follows: M. Kerscher, P. Busch, S. Mattauch, H. Frielinghaus, D. Richter, M. Belushkin, G. Gompper, Physical Review E, 83, 030401(R), 2011. Copyright (2011) by the American Physical Society.

applied. The other parameter of the Guinier-range is the overall size of the worm-like object, i.e. $R_{\mathrm{ee}} = \sqrt{6} \cdot R_{\mathrm{g}}$. Scattering scientists usually prefer the radius of gyration R_{g} instead of the end-to-end distance. Furthermore, at high q the small scale structure of the polymer or the micelle is indicated. For the micelle it is a simple circular cross-section. For the polymer a more atomistic model has to be taken into account. The overall example of worm-like objects shows that many structural parameters can be obtained from a single scattering experiment. The power laws give clear hints to the kind of structure formed. A good model is highly recommended to extract the structural parameters to the highest precision possible.

5   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. In soil remediation poisonous substances need to be washed off from the solid soil by pushing aqueous surfactant solutions or microemulsions through the volume. 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 [40]. This question was addressed by computer simulations [40]. A real space picture is shown in Fig. 18. One can see the lamellar order near the surface and the bicontinuous microemulsion in the volume. Experimentally, a microemulsion was studied by grazing incidence small angle neutron scattering (GISANS) and reflectometry. The GISANS experi-
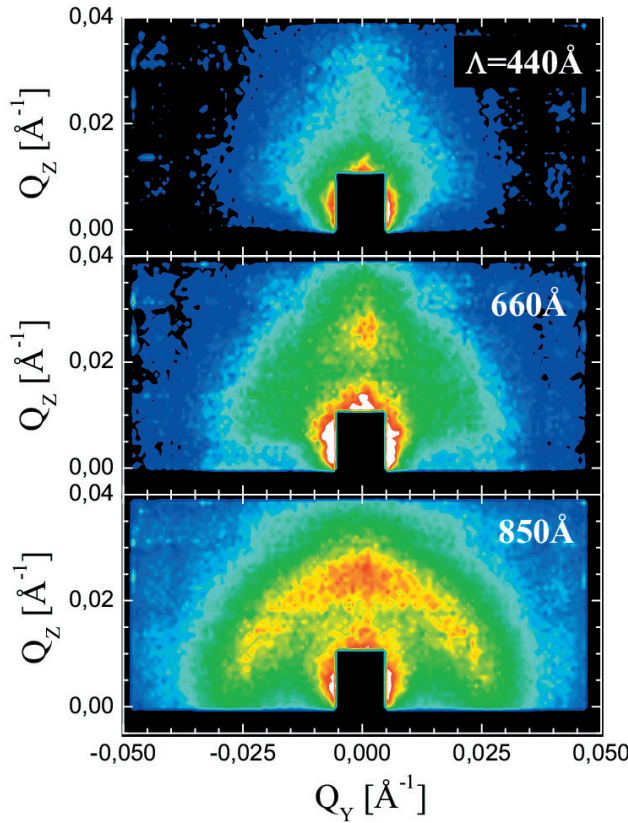
Fig. 19: GISANS patterns at different penetration depths Λ. For 440Å there is a rather strong surface scattering background in the center and a lamellar peak is only slightly indicated in the middle top. For 660Å the lamellar peak becomes stronger. For 850Å both the lamellar peak and the bicontinuous Debye-Scherrer ring are visible. Reprinted with permission from Ref. 40 as follows: M. Kerscher, P. Busch, S. Mattauch, H. Frielinghaus, D. Richter, M. Belushkin, G. Gompper, Physical Review E, 83, 030401(R), 2011. Copyright (2011) by the American Physical Society.

ments are sensitive to the lateral structures and so there were contributions from the lamellar and the bicontinuous regions (Fig. 19). Small angle scattering with grazing incidence leads to an evanescent (tunneling) wave in the sample (see also Chapter D2) and the sample is illuminated with a variable depth Λ. 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 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 isotropic bicontinuous microemulsion becomes visible as well.

The near surface ordering of the otherwise bicontinuous microemulsion is highly important because the lamellar structure gives rise to a facilitated flow along the solid surface: The lubrication effect describes a facilitated flow of the ordered parallel domains that otherwise needed to

be deformed. This effect can also be monitored macroscopically: When adding planar colloidal clay particles of approx. 1% to a microemulsion [41, 42] there exist enough lamellar domains that macroscopically allow for facilitated flow. Apart from low clay particle concentrations, the higher concentrations display another important phenomenon [43]: When the solid surfaces come close enough, the induced lamellar ordering suddenly prevails in the whole volume. This flipping from an equilibrium (bicontinuous) structure to a completely induced ordered structure is called *capillary condensation*. All in all, the additive clay is interesting and may highly likely lead to new applications in the future.

6   Summary

We have seen that there exist single molecules with chemical groups, which dislike each other. This property is called amphiphilicity. Surfactants are the most widely used substance with hydrophilic heads and hydrophobic tails. There exist ionic and non-ionic surfactants. Surfactants need to cross the critical micelle concentration to form micelles with various shapes. These shapes already have a high degree of symmetry because the surfactant molecules are all identical. At high concentrations interactions between the micelles come into play and liquid crystalline ordering takes place. These structures have again a high degree of symmetry, again because of identical building blocks.

Microemulsions involve two kinds of solvents and enlarge the variety of phases (or shapes) found in the phase diagram. Domains of water and oil form with a surfactant layer in between. The theoretical concept of Helfrich treats the surfactant layer as an elastic membrane with negligible thickness. Very important are bicontinuous phases where domains expand over macroscopic distances.

Polymers can also be treated with the concept of amphiphilicity even when the interactions are much weaker. The large number of segments leads to effectively strong enough interactions in comparison to the entropy of mixing. Micelles of amphiphilic polymers can be obtained when dissolving in water. The water soluble parts form a hairy corona.

Microemulsions tend to form a lamellar order near planar walls. Clay platelets are colloidal additives, which induce the same ordering at the solid-liquid interface. From such additives future industrial applications are to be expected.

Scattering experiments support the structure identification. They give spatial parameters in a direct way while averaging over a large sample volume. The connection to microscopic theories reveals the mechanisms of amphiphilic molecules. This understanding is supporting the next generation of applications when simple formulation experiments do not succeed anymore.

References

[1] G.J.T. Tiddy, Physics Reports **57**, 1 (1980).

[2] J.S. Keiper, R. Simhan, J.M. DeSimone, G.D. Wignall, Y.B. Melnichenko, H. Frielinghaus, J. Am. Chem. Soc. **124**, 1834 (2002).

[3] A.-G. Fournial, Y. Zhu, V. Molinier, G. Vermeersch, J.M. Aubry, N. Azarouali, Langmuir **23**, 11443 (2007).

[4] J.N. Israelachvili, D.J. Mitchell, B.W. Ninham, J. Chem. Soc. Faraday Trans. I **72**, 1525 (1976).

[5] J.S. Pedersen, S.U. Egelhaaf, P. Schurtenberger, J. Phys. Chem. **99**, 1299 (1995).

[6] G. Jerke, J.S. Pedersen, S.U. Egelhaaf, P. Schurtenberger, Phys. Rev. E **56**, 5772 (1997).

[7] R. Strey, R. Schomäcker, D. Roux, F. Nallet, U. Olsson, J. Chem. Soc. Faraday Trans. **86**, 2253 (1990).

[8] E. Jahns, H. Finkelmann, Colloid & Polymer Sci. **265**, 304 (1987).

[9] R. Strey, Phys. Chem. Chem. Phys. **97**, 742 (1993).

[10] W. Helfrich, Z. Naturforsch., A: Phys. Sci. **33**, 305 (1978).

[11] H.T. Davis, J.F. Bodet, L.E. Scriven, W.G. Miller, Physica A **157**, 470 (1989).

[12] 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).

[13] R.G. Larson, J. Chem. Phys. **91**, 2479 (1989).

[14] S.A. Safran, L.A. Turkevich, P. Pincus, J. Physique, Lettres **45**, L-69 (1984).

[15] E.M. Blokhuis, W.F.C. Sager, J. Chem. Phys. **115**, 1073 (2001).

[16] U.S. Schwarz, G. Gompper, Phys. Rev. E **59**, 5528 (1999).

[17] U.S. Schwarz, G. Gompper, Phys. Rev. Lett. **85**, 1472 (2000).

[18] N.R. Washburn, T.P. Lodge, F.S. Bates, J. Phys. Chem. B **104**, 6987 (2000).

[19] R.R. Netz, D. Andelman, M. Schick, Phys. Rev. Lett. **79**, 1058 (1997).

[20] F. Schwabl, *Statistical Mechanics* (Springer, Berlin, Heidelberg, New York, 2002).

[21] W.B. Lee, R. Mezzenga, G.H. Fredrickson, Phys. Rev. Lett. **99**, 187801 (2007).

[22] X.J. Li, M. Schick, Biophys. J. **78**, 34 (2000).

[23] C. Stubenrauch, C. Frank, R. Strey, D. Burgemeister, C. Schmidt, Langmuir **18**, 5027 (2002).

[24] T. Hahn, Ed., *International Tables for Crystallography, 5th ed.* (Kulver, Dordrecht, The Netherlands, 2002).

[25] S. Förster, A.K. Khandapur, J. Zhao, F.S. Bates, I.W. Hamley, A.J. Ryan, W. Bras, Macromolecules **27**, 6922 (1994).

[26] S. Abbas, T.P. Lodge, Langmuir **24**, 6247 (2008).

[27] Y.Y. Huang, J.Y. Hsu, H.L. Chen, T. Hashimoto, Macromolecules **40**, 406 (2007).

[28] A.K. Khandpur, S. Förster, F.S. Bates, I.W. Hamley, A.J. Ryan, W. Bras, K. Almdal, K. Mortensen, Macromolecules **28**, 8796 (1995).

[29] K.D. Göcking, M. Monkenbusch, Europhys. Lett. **43**, 135 (1998).

[30] F. Castro-Roman, L. Porcar, G. Porte, C. Ligoure, Eur. Phys. J. E **18**, 259 (2005).

[31] R. Hosemann, S.N. Bagchi, *Direct Analysis of Diffraction by Matter* (North-Holland, Amsterdam, 1962).

[32] M. Teubner, R. Strey, J. Chem. Phys. **87**, 3195 (1987).

[33] C. Frank, H. Frielinghaus, J. Allgaier, H. Prast, Langmuir **23**, 6526 (2007).

[34] 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).

[35] G. Beaucage, J. Appl. Cryst. **29**, 134 (1996).

[36] H. Kaya, L. Willner, J. Allgaier, J. Stellbrink, D. Richter, Appl. Phys. A **74**, S499 (2002).

[37] M. Nakano, H. Matsuoka, H. Yamaoka, A. Poppe, D. Richter, Macromolecules **32**, 697 (1999).

[38] A. Halperin, Macromolecules **20**, 2943 (1987).

[39] J.S. Pedersen, P. Schurtenberger, Macromolecules **29**, 7602 (1996).

[40] M. Kerscher, P. Busch, S. Mattauch, H. Frielinghaus, D. Richter, M. Belushkin, G. Gompper, Phys. Rev. E **83**, 030401(R) (2011).

[41] M. Gvaramia, G. Mangiapia, V. Pipich, M.S. Appavou, S. Jaksch, O. Holderer, M.D. Rukhadze, H. Frielinghaus, Sci. Rep. (2019) accepted.

[42] M. Gvaramia, G. Mangiapia, V. Pipich, M.S. Appavou, G. Gompper, S. Jaksch, O. Holderer, M.D. Rukhadze, H. Frielinghaus, arXiv:1709.05198 (2018).

[43] M. Gvaramia, G. Mangiapia, P. Falus, M. Ohl, O. Holderer, H. Frielinghaus, J. Coll. Interf. Sci. **525**, 161 (2018).

B 4 Dynamics of large molecules and membranes

O. Holderer

Jülich Centre for Neutron Science (JCNS) at
Heinz Maier-Leibnitz Zentrum (MLZ)
Forschungszentrum Jülich GmbH

Contents

1   Introduction

The dynamics of large molecules and membranes is a broad field where scattering techniques and in particular neutron scattering play a prominent role. Large molecules such as proteins or polymer chains are relevant in biology, materials engineering, drug design and chemistry and are of fundamental interest in chemical physics and biology. Phospholipid membranes are part of every cell and the elastic properties as well as transport through biological membranes are of utmost importance for its function.

Scattering helps to reveal fundamental properties of molecules and membranes and contributes significantly to the general knowledge about complex systems. Membrane elasticity measurements helped for example to understand the enzyme activity in DFPase/microemulsion systems [1].

The contrast in scattering experiments depends on the probe used. X-ray scattering is sensitive to the variation in electron density, light scattering to refractive index changes, whereas one peculiarity of neutron scattering is the dependence on the scattering length density, which is rather irregular across the periodic table and can be strongly different for different isotopes, e.g. H and D, which allows to vary the contrast without varying the electronic structure and hence the chemistry. With contrast variation, it is therefore possible to highlight different parts of the sample while matching out others [2]. Fig. 1 depicts the scattering length density for typical materials relevant for soft condensed matter studies, and the varying contrast of one of the most common solvents, water, while replacing part of it with deuterated water.

2   Large molecules and mesoscopic structures

The starting point for measuring and describing the dynamics of large molecules, membranes and mesoscopic structures are scattering experiments on the experimental side and correlation functions as introduced in Lecture A4. Typical examples of systems which can be studied in this way are given in Fig. 2. Domain motions of proteins are described in more detail in Lecture E8, here just the example of an Immunuglobuline protein is given Fig. 2 (top left), a Y shaped protein, where the spring constant of the joint can be obtained from scattering experiments [3]. Polymer segmental dynamics as depicted on the right by a bead spring model is shown on the right. Thermal fluctuations in polymer melts with different levels of complexity [4–6]. Membrane fluctuations play an important role in chemical systems in the form of surfactant membranes, where the elasticity and bending rigidity can be determined [7]. Biologically relevant are phospholipid membranes as model systems for cell membranes, where similarly the elasticity is key for its stability and deformability and can be investigated with scattering techniques, including effects of additives such as ibuprofen [8,9].

2.1   Scattering techniques for energy resolved studies

A typical scattering experiment is in principal relatively simple. A particle beam of defined energy (photons, neutrons, electrons) hits the sample, after the scattering event, the scattering angle and outgoing energy is measured. The technical details can get very complex are
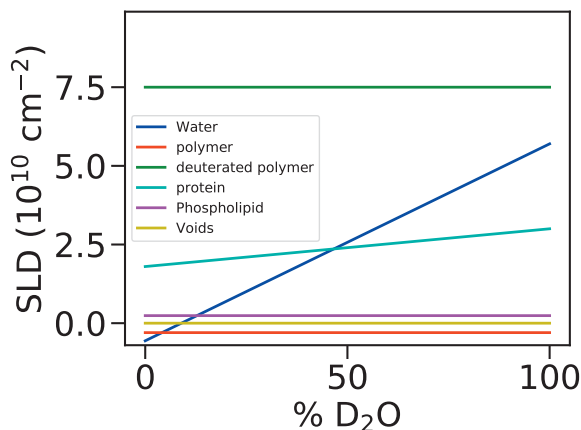
Fig. 1: The strength of neutron scattering is the ability to change the contrast in the sample by partly deuterating some components, e.g. the solvent (here water).

presented in different lectures (D1-10). For studying the dynamics of large molecules, some suitable techniques are put into the focus.

2.1.1 Neutron scattering

Neutron scattering provides length- and energy-scales suitable for large molecule and membrane dynamics. One prominent technique is neutron spin echo (NSE) spectroscopy [4, 10] (see also Lecture D6), where velocity changes during scattering are encoded in a number of spin precessions in large solenoids . The polarization at the detector allows to measure the intermediate scattering function, which is related to the correlation function in real space by a Fourier transformation.

2.1.2 DLS and XPCS

The equivalent technique to NSE with photons is dynamic light scattering (DLS) or x-ray photon correlation spectroscopy for different length scales (depending on the wavelength of the photon) and time scales of the order of μs and above.

2.1.3 Computer simulations

Molecular dynamics (MD) simulations are a virtual experiment, where the trajectories of atoms or particles in general can be followed [11]. A possible free software for such MD simulations is gromacs [12]. 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:

$$m_i \ddot{\mathbf{r}}_\mathbf{i} = \mathbf{F}_\mathbf{i} \tag{1}$$
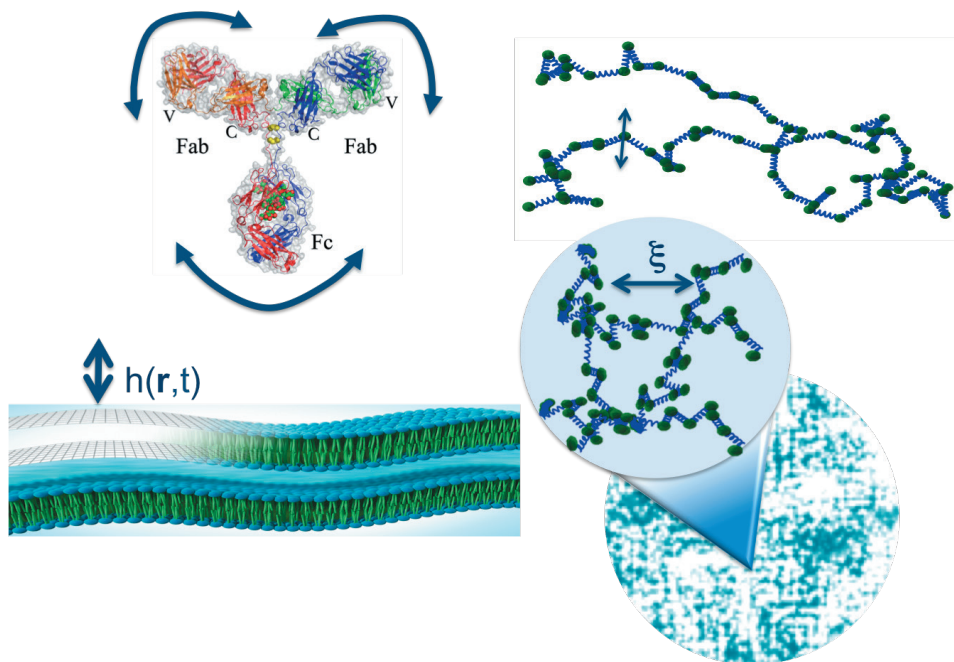
Fig. 2: Large scale motions involve protein domain motions (top left, Immunoglobuline from Ref [3]. Phospholipid membranes (bottom left) play an important role in chemistry and biology. They can serve as model systems for biological membranes, surfactant membranes are relevant in different products. Polymers in solution (modelled as a chain of beads connected with springs, top right) and microgels (crosslinked polymers on local scale, inhomogeneous particles on large scales, bottom right) will be presented here.

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 [13]. 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. Polymeric particles such as microgels including the surrounding solvent can be simulated in this way, Fig. 3 shows a visualization of such particles [14].

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.

2.2   Description of molecule dynamics

The dynamics of macromolecules can be described in terms of correlation functions , which have been introduced in the chapters A4 and D5. They describe the probability of finding a certain state of the system at time t if the state at time 0 is known. The Fourier transform of this correlation function from real space into reciprocal space leads to the intermediate scattering function $I(Q,t)$ as measured e.g. with neutron spin echo spectroscopy, dynamic light scattering or x-ray photocorrelation spectroscopy on different length- and timescales. Further Fourier transforming the intermediate scattering function from the time-domain to the energy domain results then in the scattering function $S(Q,\omega)$ measured in neutron scattering with backscattering spectroscopy, time-of-flight- or three-axis spectroscopy. In real space, the effect of a loss in correlation described with the correlation function is visualized in Fig. 4. A coarse grained polymer chain represented as beads connected by springs is taken as an example. The knowledge about the actual position of the molecule or parts of the molecule, depending on the length scale we are investigating, is expressed in terms of the correlation function, starting at 1 and decaying to zero (complete loss of knowledge about the system).

2.3   Center of mass diffusion

As an example on how to relate mathematical models with experiment, a diffusing particle will be considered. For objects of the order of 1-100 nm this is whithin the time scales of DLS, NSE and XPCS (although the observation length and time scale differs strongly between these techniques).

One possible starting point is the Langevin equation (here only in x direction), which describes the motion of a Brownian particle at position x on which a random force $f(t)$ is acting coming
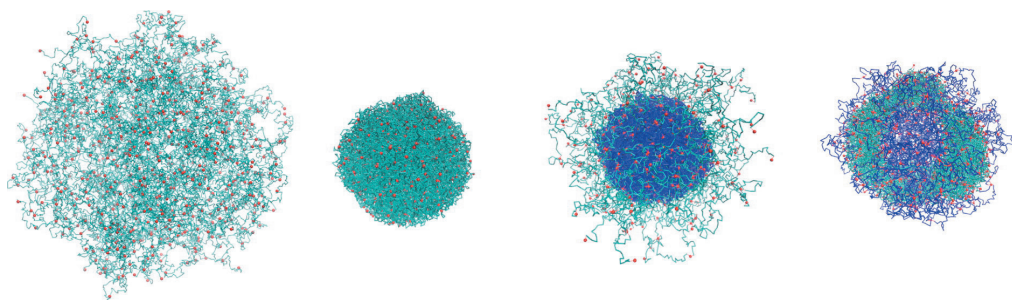
Fig. 3: Microgel particle from computer simulations. From Ref. [15]
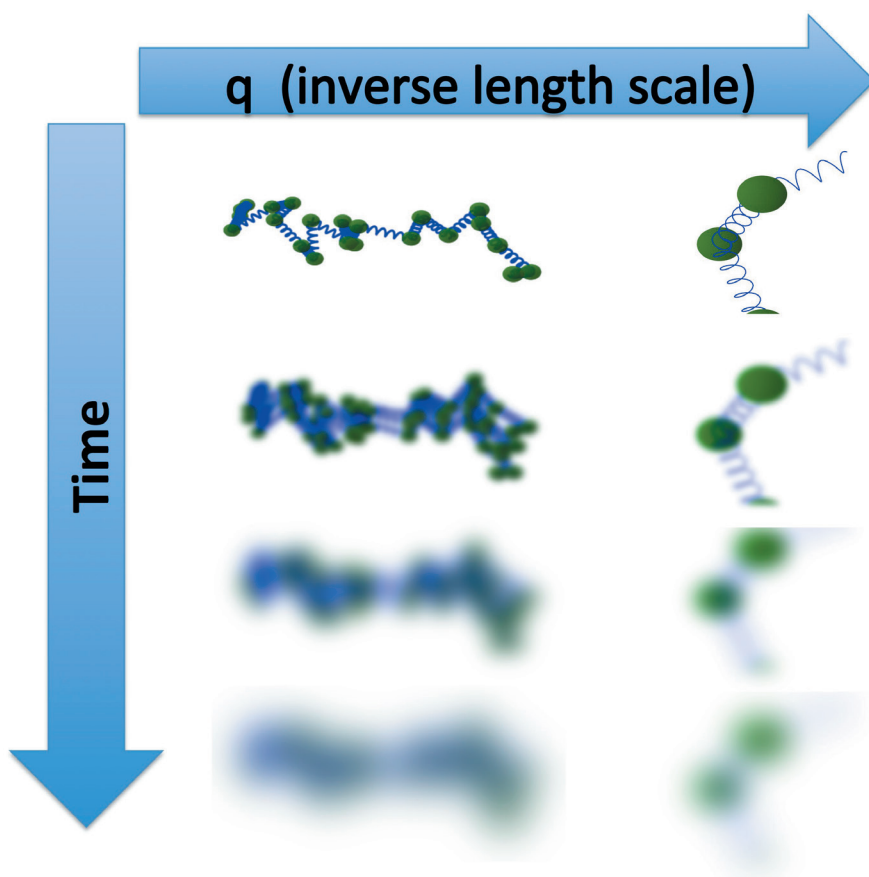
Fig. 4: Loss of correlation for a polymer chain with time on different length scales. The strength of scattering techniques is the possibility of changing the length scale of investigation. The loss of correlation in reciprocal space is the intermediate scattering function measured with DLS and XPCS (large length scales, mainly diffusion) and NSE.

from collisions with the solvent molecules:

$$\eta \frac{\partial x}{\partial t} + \frac{\partial U}{\partial x} = f(t) \tag{2}$$

where η is the viscosity and the first term the friction force and U is an external potential which exerts a force $F = -\partial U/\partial x$ on the particle.

Assuming the random force obeying a Gaussian distribution also the distribution of the position of the particle is Gaussian and can be described by

$$\Psi(x,t) = \frac{1}{4\pi Dt} \exp -\frac{x^2}{4Dt} \tag{3}$$

This probability distribution of a particles position provided it was at the origin at $t = 0$ is a solution of the Langevin equation.

The mean squared displacement (MSD) is the second moment of $\Psi(x,t)$:

$$MSD = <x^2> = \int dx x^2 \Psi(x,t) = 2nDt \tag{4}$$

with the diffusion constant D (in n dimensions) which is a quantity accessible in scattering experiments or computer simulations.

It can be shown that the Langevin equation is also equivalent to the Smoluchowski equation, where the same result is obtained starting from Ficks law of diffusion. See Ref. [16] for a more extensive derivation.

The movement of a particle can be described as a random walk. Fig. 5 illustrates the evolution of $\Psi(x,t)$ with time (left), i.e. a broadening of the particle distribution starting from a point where the position is well defined (sharp peak), which corresponds to a random walk of the particle (right).

Fourier transformation of $\Psi(x,t)$ from real into reciprocal space gives the intermediate scattering function, as measured with NSE (the factor 6 for the 3D case):

$$I(Q,t)/I(Q,0) = \exp\left(-1/6 <r^2(t)> Q^2\right) = \exp\left(-D_{CM}Q^2 t\right) \tag{5}$$

A simple diffusion therefore has a quadratic dependence on the the scattering vector Q.

2.4   Polymer dynamics: Rouse model

In the Rouse model the Gaussian polymer chain is described as beads connected by springs as depicted in Fig. 6. The springs correspond to the entropic forces between the beads and the distance between the beads corresponds to the segment length of the polymer. The polymer chain is in a heat bath from the surrounding chains. The Rouse model describes the movement of the single chain segments of such a polymer chain as Brownian movement. Again, the Langevin equation can be used as a starting point, with thermally activated fluctuations (by the
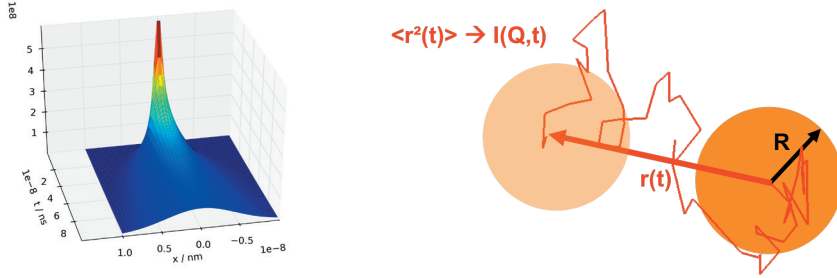
Fig. 5: Left: At time t=0 the state of the system (all atomic positions) is known, all atoms are at their initial position. With evolving time, the atoms diffuse further and further away, they are distributed over a region with width σ from their initial positions. Right: If a particle moves away from the origin in a random walk, its distance covered at time t, $r(t)$, obeys the $\langle \Delta r^2(t) \rangle = \langle r(t)^2 \rangle = 2Dt$.
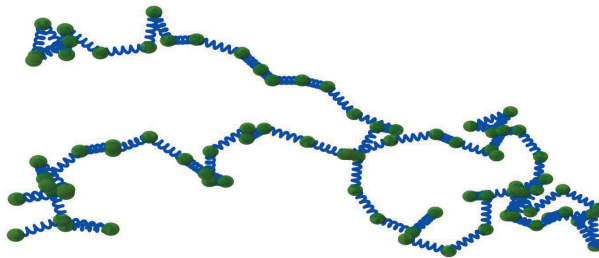
Fig. 6: Schematic representation of the polymer chain in the Rouse model [4] as a Gaussian chain with beads connected by springs. If surrounded by solvent, also hydrodynamic interactions come into play (Zimm model), i.e. the flow field of the solvent acts on neighbouring beads.

stochastic force $\mathbf{f}_n(t)$ with $< \mathbf{f}_n(t) >= 0$), friction force (with friction coefficient ζ) and the entropic force determine the relaxation of polymer chains.

The movement of the chain segments is described by a Langevin equation:

$$\zeta\frac{d\mathbf{R}_n}{dt} + \frac{3k_BT}{l^2}(\mathbf{R}_{n+1} - 2\mathbf{R}_n + \mathbf{R}_{n-1}) = \mathbf{f}_n(t) \tag{6}$$

The Langevin equation can be solved (this is worked out e.g. in Refs [4, 16]) and one can calculate again the intermediate scattering function:

$$I(Q,t) = \exp(-Q^2Dt)I_{intern}(Q,t) \tag{7}$$

with a diffusive part with a relaxation rate proportional to Q^2 and the part describing the internal relaxation, with the contributions of all modes of the polymer chain:

$$I_{intern}(Q,t) = \frac{1}{N}\sum_{n,m}^{N}\exp\left(-\frac{1}{6}Q^2B(n,m,t)\right) \tag{8}$$

with

$$B(n,m,t) = (n-m)l^2 + \frac{4R_e^2}{\pi^2}\sum_{p=0}^{pmax}\frac{1}{p^2}\cos\left(\frac{\pi pn}{N}\right)\cos\left(\frac{\pi pm}{N}\right)(1 - \exp(-tp^2/\tau_R)) \tag{9}$$

with the Rouse relaxation time τ_R, the number of modes p and R_e the end-to-end distance of the polymer chain.

Replacing the sums with integrals and applying some simplifications, $I_{intern}(Q,t)$ can be written for $QR_G >> 1$ as follows:

$$I_{intern}(Q,t) = \frac{12}{Q^2l^2}\int_0^\infty du\exp(-u - \sqrt{(\Gamma_Q t)}h(u/\sqrt{(\Gamma_Q t)})) \tag{10}$$

with the relaxation rate

$$\Gamma_Q = \frac{k_BT}{12\zeta}Q^4l^2 \tag{11}$$

and

$$h(u) = \frac{2}{\pi}\int dx\cos(xu)(1 - e^{-x^2})/x^2 \tag{12}$$

Note that the local relaxation rate depends on Q^4. When $I(Q,t)/I(Q,0)$ is plotted against the Rouse variable $\sqrt{\Gamma_Q t}$, all curves collapse onto a master curve if the Rouse model holds. This scaling behaviour is a powerful way of investigating polymer melts.

With this model, for example, the dynamic of short polymer chains in the melt can be described. With increasing molecular weight some other effects like the constraints imposed by mutual entanglements of the polymer chains become important, which are described in the reptation model by de Gennes (Nobel prize 1991).

2.5   Zimm dynamics

Polymers in solution can be described by the Zimm model , where hydrodynamic interaction between the chain segments mediated by the solvent are dominant. Moving chain segments exert forces on other segments due to the flow of the surrounding solvent. Within some approximations the system can be described again by a Langevin equation analogous to that of the Rouse model. The main modification is the inclusion of the hydrodynamic interaction represented by an Oseen tensor H_{nm}

$$\zeta_0 \frac{d\mathbf{R}_n}{dt} = \sum_m H_{nm} \frac{k_B T}{l^2} \frac{\partial^2 R_m}{\partial m^2} + \mathbf{f}_n(t) \tag{13}$$

More details can be found in literature, e.g. in Ref. [4].

Fourier transforming the corresponding real space correlation function leads again to an expression for the intermediate scattering function, which is in the discrete picture a sum over all modes in the bead-spring chain:

$$\frac{I(Q,t)}{I(Q,0)} = \frac{1}{N} \exp(-Q^2 Dt) \sum_{n,m}^N \exp\left(-\frac{1}{6} Q^2 B(n,m,t)\right) \tag{14}$$

$$B(n,m,t) = (n-m)^{2\nu} l^2 + \frac{4R_e^2}{\pi^2} \sum_{p=0}^{pmax} \frac{1}{p^{2\nu+1}} \cos\left(\frac{\pi pn}{N}\right) \cos\left(\frac{\pi pm}{N}\right) (1 - \exp(-1/\tau_p)) \tag{15}$$

with the number N of beads in the chain with end-to-end distance R_e, the modes p which contribute to the dynamics, $\nu = 0.5$ for a Gaussian chain. The relaxation time τ_p depends on the viscosity η:

$$\tau_p = \frac{\eta R_e^3}{\sqrt{(3\pi)} k_B T} p^{-3\nu} \tag{16}$$

An approximation to the intermediate scattering function of the Zimm model is a stretched exponential function which reads

$$\frac{I(Q,t)}{I(Q,0)} = \exp\left(-\left(\frac{k_B T Q^3 t}{6\pi \eta b}\right)^\beta\right) \tag{17}$$

with $b \simeq 1.354$ and $\beta \simeq 0.85$. The relaxation rate of a polymer chain in this model, $\Gamma = k_B T Q^3 t/(6\pi \eta)$, is mainly determined by the viscosity of the solvent. Internal dynamics is dominant at higher scattering vectors Q, where also the typical Q^3 dependence of the relaxation rate can be observed. At smaller scattering vectors the contribution from the center of mass diffusion is more prominent so that rather a Q^2 dependence of the relaxation rate is expected.

Fig. 7 shows a polymer in an apolar solvent (PEP in deuterated decane) together with the prediction from the Zimm model as lines (left) and in Zimm scaling, where the data from all Q values fall onto one line.
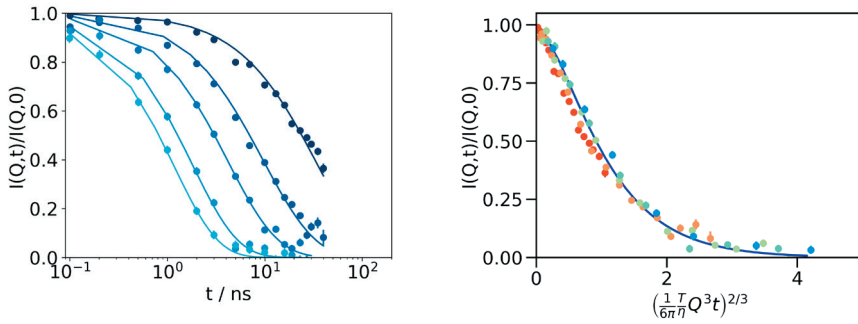
Fig. 7: Left: Intermediate scattering function I(q,t) for a polymer in solution (polyethylene-propylene in deuterated decane) together with the Zimm prediciton (lines) for Q-values from 0.05 to 0.18 Å^{-1}. Right: Zimm scaling of the same sample.

A typical approach for the interpretation of the chain dynamics is to fit $I(Q,t)$ with Eq. 17 and plot the rate Γ/Q^2 vs Q. Diffusive behaviour leads to a constant, while Zimm dynamics would result in a linear increase. The transition between thouse two gives an indication on the length scales of density fluctuations. Cross-linking of polymer chains (deviation from a Gaussian chain) or very crowded environment may change the apparent viscosity, internal structures of the polymer chain might require to modify the Zimm model, e.g. with internal friction contributions (ZIF, Zimm model with Internal Friction [17])

2.6   Membrane dynamics

The height correlation function of membrane fluctuations can be deduced again from the Langevin equation, where the Helfrich hamiltonian [18] describes the energy of the membrane. The expression for the intermediate scattering function is calculated by Zilman and Granek [19], which for stiffer membranes reduces to a stretched exponential function. For softer membranes a numerical evaluation is required which also yields the bending rigidity of the membranen [20].

In a first approximation, the intermediate scattering function can be described therefore as follows:

$$\frac{I(Q,t)}{I(Q,0)} = \exp\left(-(\Gamma t)^\beta\right) \tag{18}$$

with

$$\Gamma = 0.025\gamma_\kappa\left(\frac{k_BT}{\kappa}\right)^{1/2}\frac{k_BT}{\eta}\,Q^3 \tag{19}$$

where $\gamma_\kappa \simeq 1$ for not too soft membranes, κ is the bending rigidity of the membrane (related to the energy required to bend the membrane). The stretching exponent of $\beta = 2/3$ is characteristic of membrane dynamics together with the Q^3-dependence of the relaxation rate. This is also valid for soft membranes with $\kappa \simeq k_BT$, but for the determination of the actual value of κ a numerical integration of the height correlation function instead of the approximation in Equation 18 is required.

Since the elasticity of membranes is an important parameter for the membrane stability, fusion,
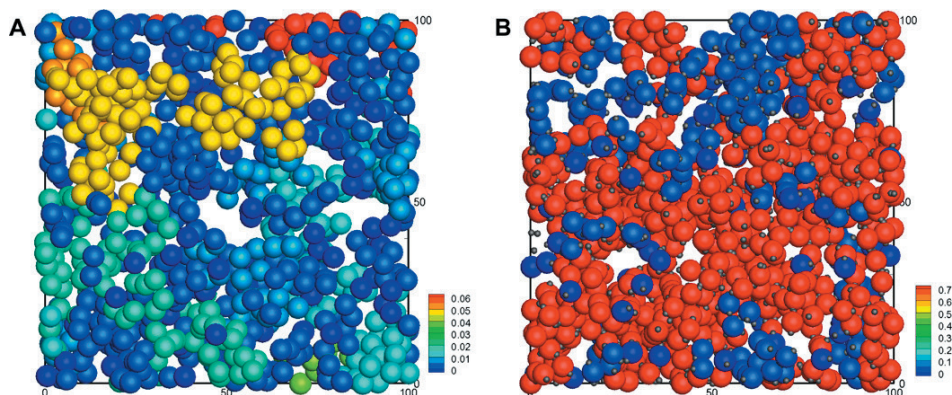
Fig. 8: Crowded protein solution: centrosymmetric interactions lead to rather small clusters (left), while the introduction of patchy interactions faciliate the formation of large clusters (right). From Ref. [21]

transport through membranes etc., studying the influence of additives such as co-polymers or membrane proteins is an important step for the understanding of the function of membranes.

2.7   Adding complexity: confinement

Fluctuations of macromolecules and membranes are affected by the presence of interfaces, e.g. solid walls, tubes, fillers, or crowding with similar molecules. Starting from the models for pure polymer melts or Gaussian coils in solution, effects of confinements or constraints have been attracted much attention in the last years, since in real life interfaces are ubiquitous.

Still, the starting point of understanding are the fundamental models starting with the Langevin equation of the system. Either one can modify the model in order to include effects of interfaces, e.g. a changed membrane fluctuation spectrum when measuring confined membranes [22–24]. Gels or microgels might impose constraints to the dynamics of polymer chains such as frozen inhomogeneities [25] or charge induced interactions by counterions in microgels [26]. Core-Shell structures in microgel particles have been analyzed together with computer simulations [14], where different parts of the particle can be separately collapsed or swollen as shown in Fig. 3. Often, details of structural properties in solution are difficult to detect directly, e.g. with small angle scattering, but leave an imprint in the dynamics and can be analyzed in this way. A last example, which is also highly relevant for biological applications, is the dynamics in a crowded protein solution (Fig. 8), where small changes in the interaction of eye lense proteins result in a massive change in viscosity, which might be highly relevant for the elasticiy of the eye lens. Adding a patchy interaction between the proteins results in the formation of large clusters (Fig. 8, right).

2.8   Final remarks

The dynamics of large molecules and membranes on length scales of some nanometers to 100s of nanometers is relevant in many applications in biology, chemistry or process engineering. Scattering techniques provide a window to this world and allow to study thermally driven molecular fluctuations and the influence of the surrounding of the molecules on the dynamics. Correlation functions are the main tool to describe these motions and allow to relate scattering experiments in a natural way to computer simulations, where tracking of particles in the "virtual" sample give access to the same quantity.

References

[1] S. Wellert, B. Tiersch, J. Koetz, A. Richardt, A. Lapp, O. Holderer, J. Gäb, M.-M. Blum, C. Schulreich, R. Stehle, T. Hellweg, European Biophysical Journal **40**, 761 (2011).

[2] N. R. Zaccai, C. W. Sandlin, J. T. Hoopes, J. E. Curtis, P. J. Fleming, K. G. Fleming, S. Krueger, in Z. Kelman (ed.), *Isotope Labeling of Biomolecules - Applications*, *Methods in Enzymology*, volume 566, pp. 159 – 210 (Academic Press, 2016).

[3] L. R. Stingaciu, O. Ivanova, M. Ohl, R. Biehl, D. Richter, Scientific reports **6**, 22148 (2016).

[4] D. Richter, M. Monkenbusch, A. Arbe, J. Colmenero, in *Neutron Spin Echo in Polymer Systems*, pp. 1–221 (Springer, 2005).

[5] C. Mark, O. Holderer, J. Allgaier, E. Hübner, W. Pyckhout-Hintzen, M. Zamponi, A. Radulescu, A. Feoktystov, M. Monkenbusch, N. Jalarvo, D. Richter, Physical review letters **119**, 047801 (2017).

[6] M. Krutyeva, A. Wischnewski, M. Monkenbusch, L. Willner, J. Maiz, C. Mijangos, A. Arbe, J. Colmenero, A. Radulescu, O. Holderer, et al., Physical review letters **110**, 108303 (2013).

[7] O. Holderer, H. Frielinghaus, M. Monkenbusch, M. Klostermann, T. Sottmann, D. Richter, Soft Matter **9**, 2308 (2013).

[8] S. Jaksch, O. Holderer, M. Gvaramia, M. Ohl, M. Monkenbusch, H. Frielinghaus, Scientific Reports **7**, 4417 (2017).

[9] S. Jaksch, F. Lipfert, A. Koutsioubas, S. Mattauch, O. Holderer, O. Ivanova, H. Frielinghaus, S. Hertrich, S. F. Fischer, B. Nickel, Physical Review E **91**, 022716 (2015).

[10] F. Mezei, in *Neutron spin echo*, pp. 1–26 (Springer, 1980).

[11] S. Pahari, C. K. Choudhury, P. R. Pandey, M. More, A. Venkatnathan, S. Roy, The Journal of Physical Chemistry B **116**, 7357 (2012).

[12] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, Journal of chemical theory and computation **4**, 435 (2008).

[13] K.-D. Kreuer, S. J. Paddison, E. Spohr, M. Schuster, Chemical reviews **104**, 4637 (2004).

[14] S. Maccarrone, A. Ghavami, O. Holderer, C. Scherzinger, P. Lindner, W. Richtering, D. Richter, R. G. Winkler, Macromolecules **49**, 3608 (2016).

[15] S. Maccarrone, A. Ghavami, O. Holderer, C. Scherzinger, P. Lindner, W. Richtering, D. Richter, R. G. Winkler, Macromolecules **49**, 3608 (2016).

[16] S. Edwards, M. Doi, *The theory of polymer dynamics. International series of monographs on physics* (1988).

[17] A. M. Stadler, L. Stingaciu, A. Radulescu, O. Holderer, M. Monkenbusch, R. Biehl, D. Richter, Journal of the American Chemical Society **136**, 6987 (2014).

[18] W. Helfrich, Z. Naturforsch. **33 a**, 305 (1978).

[19] A. Zilman, R. Granek, Phys. Rev. Lett. **77**, 4788 (1996).

[20] M. Mihailescu, M. Monkenbusch, H. Endo, J. Allgaier, G. Gompper, J. Stellbrink, D. Richter, B. Jakobs, T. Sottmann, B. Farago, The Journal of Chemical Physics **115**, 9563 (2001).

[21] S. Bucciarelli, J. S. Myung, B. Farago, S. Das, G. Vliegenthart, O. Holderer, R. G. Winkler, P. Schurtenberger, G. Gompper, A. Stradner, Science advances **2**, e1601432 (2016).

[22] H. Frielinghaus, M. Kerscher, O. Holderer, M. Monkenbusch, D. Richter, Physical Review E **85**, 041408 (2012).

[23] F. Lipfert, H. Frielinghaus, O. Holderer, S. Mattauch, M. Monkenbusch, N. Arend, D. Richter, Physical Review E **89**, 042303 (2014).

[24] O. Holderer, H. Frielinghaus, S. Wellert, F. Lipfert, M. Monkenbusch, R. Von Klitzing, D. Richter, Journal of Physics: Conference Series **528**, 012025 (2014).

[25] S. Koizumi, M. Monkenbusch, D. Richter, D. Schwahn, B. Farago, The Journal of chemical physics **121**, 12721 (2004).

[26] S. Maccarrone, O. Mergel, F. A. Plamper, O. Holderer, D. Richter, Macromolecules **49**, 1911–1917 (2016).

B 5 Elementary excitations in crystalline matter: phonons and magnons

Nicolò Violini

Jülich Centre for Neutron Science

Forschungszentrum Jülich GmbH

Contents

1   Introduction

Understanding of the thermal, mechanical and magnetic properties of solid state materials brought to the formulation of complex problems involving many strongly interacting particles. In order to solve these problems, theorists introduced approximations and assumptions that enabled to tract the problems and find suited solutions for them. Thanks to their efforts, a deep knowledge of many properties of materials has been gained and many phenomena have been understood such as the low-temperature specific heat and the deviation of the spontaneous magnetization of a ferromagnet from its saturation level. In both these cases, the theoretical solution was found in approximating the many-body system to a set of independent harmonic oscillators. In the harmonic oscillator, the energy levels are equally spaced with spacing $\hbar$ times the frequency of the oscillator. The quantum of energy to bring the system of oscillators from one excited state to a contiguous one takes the name of an elementary excitation, being called phonon or magnon, if the described system is made of nuclei or spins in the crystal, respectively. Even though the system under consideration is characterised by perfect periodicity, the latter is not necessary to prove the existence of elementary excitations in solids, which is a rather more general result. Moreover, here we provide an introduction to phonons and magnons only, but it should be noticed that other elementary excitations exist in crystals, such as plasmons and excitons. The paper is organized as follows: in section 2 we provide an introduction to the Born-Oppenheimer approximation and the relative consequences on the treatment of the many-body problems; sections 3 and 4 are devoted to introduce the concepts of phonons and magnons in crystals, respectively, and provide the basics on the methods to treat the Hamiltonian problems of nuclei and spins lattice.

2   Born-Oppenheimer approximation

The Born-Oppenheimer approximation is the assumption that the motion of atomic nuclei can be treated independently from the motion of their electrons. It represents a very fundamental step toward into the description of a complex problem, since one can separate the Hamiltonian in kinetic and potential energy with all nuclear and electronic coordinates and momenta:

$$H = H_{\text{ion}} + H_{\text{el}} + H_{\text{int}}. \tag{1}$$

In this approximation the Coulomb interaction between nuclei and electrons is taken into account by the term H_{int}. The relevant consequence of this approximation is that it allows to break the wavefunction into its nuclear and electronic components and find a description of them, i.e. a solution of the Hamiltonian problem.

It is also known as adiabatic approximation, because nuclei and electrons are seen as independent moving objects, therefore one can imagine the light electrons to adjust instantaneously (i.e. adiabatically) to the slower motion of the heavy ionic cores. It is reasonable to assume this because of the small ratio of the electron mass to the ion one.

It may seem a trivial assumption, but it is indeed a very relevant simplification, since now the problem is broken in two smaller problems for which we can find solutions.

3   Phonons

The first type of elementary excitations we describe in this lecture are called phonons. The commonly used definition of phonon is: "the quantum of energy that separates two contiguous excited states of the lattice vibrations". In order to justify this definition we will describe the Hamiltonian problem for the lattice dynamics and describe how this can be solved with the help of the harmonic approximation, which enables to define the potential energy as a positive-definite function. It should be stressed that the definition of phonons is a direct consequence of this characteristic, even in absence of perfect periodicity. On the other hand, only taking into account the periodic conditions (Born-von Karmann conditions), a mathematical description of the dispersion curves of phonons can be derived.

3.1   Harmonic approximation

In order to describe the lattice dynamics, we observe that the atomic mass is mainly in its point-like nucleus and that the forces between nuclei are transmitted through much lighter electrons. Therefore the Hamiltonian problem can be formulated as if the atomic ions moved in a potential V, made up of the potential energy of the ion configuration and of the electrons' energy in this configuration:

$$H_{\text{ion}} = \sum_l \frac{\mathbf{p}_l^2}{2m_l} + V. \tag{2}$$

The instantaneous position (at time t) of the $l-$th ion placed at position $\mathbf{r}_l$ inside the $k-$th unit cell placed at position $\mathbf{r}_k$ is:

$$\mathbf{R}_{kl} = \mathbf{r}_k + \mathbf{r}_l + \mathbf{u}_l(t), \tag{3}$$

where it is implied that the ion can vibrate around the equilibrium position with displacement $\mathbf{u}_l(t)$. Assuming small atomic displacements out of their equilibrium positions, the potential energy V can be written as a Taylor series in these atomic displacements:

$$V = V_0 + \sum_{l,\alpha} \frac{\partial V}{\partial u_{l\alpha}} u_{l\alpha} + \frac{1}{2} \sum_{l,\alpha} \sum_{l'\beta} \frac{\partial^2 V}{\partial u_{l\alpha} \partial u_{l'\beta}} u_{l\alpha} u_{l'\beta} + o(u^3) \,, \tag{4}$$

where V_0 is the static potential energy in the equilibrium configuration, i.e. when all the displacements are equal to zero. Being a constant, it is negligible. The first order term must also be equal to zero, because the sum of the forces on each atom at the equilibrium position must vanish:

$$\sum_{l,\alpha} \frac{\partial V}{\partial u_{l\alpha}} = 0 \,. \tag{5}$$

The second-order term contains the force-constants that describe the force acting on the atom l in direction α when the atom l' is displaced in direction β:

$$\Phi_{\alpha\beta}(ll') = \frac{\partial^2 V}{\partial u_{l\alpha} \partial u_{l'\beta}} \,. \tag{6}$$

$\Phi_{\alpha\beta}(ll')$ should be seen as the element $\alpha\beta$ of a matrix called dynamic matrix, whose determinant gives the eigen-frequencies of ionic motions. In equation 4 we described all higher orders

derivatives with $o(u^3)$, which by definition contains anharmonic terms in the crystal potential energy. As a result of all these considerations, the Hamiltonian can be re-written as the one of an assembly of independent harmonic oscillators:

$$H_{\text{ion,harm}} = \sum_l \frac{\mathbf{p}_l^2}{2m_l} + \frac{1}{2} \sum_{l,\alpha} \sum_{l',\beta} \Phi_{\alpha\beta}(ll') u_{l\alpha} u_{l'\beta} \tag{7}$$

for which both classical and quantum mechanical solutions exist. In other terms, the harmonic approximation is the assumption that all anharmonic terms in the Taylor series can be neglected so that the Hamiltonian is similar to the one of a system of harmonic oscillators. This enables to find solutions and eventually extend them to the case where also higher order terms are considered, using a perturbative approach, that means adding small external perturbation to the simpler problem. What does all this mean in practice? Through the adiabatic approximation, a complex problem has been broken down into smaller tractable problems for which a solution exists. For the ionic part, the harmonic approximation lead to find this solutions (more on this in the following sections). We know that viewing a crystal as an assembly of independent harmonic oscillators is an oversimplification of the problem. For instance, these lattice vibrations without coupling have infinite lifetime, which seems to be unrealistic, while for a more realistic description of the real observations that excitations have a finite lifetime, one would have to consider the higher-order terms neglected in the harmonic approximation. In the next section we will explain the steps necessary to solve the Hamiltonian problem for $H_{\text{ion,harm}}$ and how they bring us to define collective vibrations of the ions in the lattice.

3.2   Exercise: How to solve the Hamiltonian problem.

In order to write the equation of motion for the ion l-th in the unit cell k-th, we have to notice that its conjugate variables are its momentum $p_{k,l}$ and its displacement from equilibrium position $u_{l\alpha}$ in direction α:

$$\frac{dp_l}{dt} = -\frac{\partial H_{ion,harm}}{\partial u_{l\alpha}} \ , \tag{8}$$

with:

$$\frac{dp_l}{dt} = \frac{d}{dt} \frac{d(m_l u_{l\alpha})}{dt} = m_l \frac{d^2 u_{l\alpha}}{dt^2} \ , \tag{9}$$

and:

$$\frac{\partial H_{ion,harm}}{\partial u_{l\alpha}} = \frac{\partial}{\partial u_{l\alpha}} \left(\frac{1}{2} \sum_{l,\alpha} \sum_{l',\beta} \Phi_{\alpha\beta}(ll') u_{l\alpha} u_{l'\beta} \right) = \sum_{l',\beta} \Phi_{\alpha\beta}(ll') u_{l'\beta}, \tag{10}$$

one gets[1]

$$m_l \frac{\partial^2 u_{l\alpha}}{\partial t^2} = -\sum_{l',\beta} \Phi_{\alpha\beta}(ll') u_{l'\beta}. \tag{12}$$

---

[1] Omitting the dependency on k, k' and l, l' and assuming $\Phi_{\alpha\beta} = \Phi_{\beta\alpha}$ we demonstrate here the last passage in two dimensions, i.e. when $\alpha = x, y$ and $\beta = x, y$:

$$\frac{\partial}{\partial u_x} \left(\frac{1}{2} \sum_{\alpha=x,y} \sum_{\beta=x,y} \Phi_{\alpha\beta} u_\alpha u_\beta \right) = \frac{\partial}{\partial u_x} \left(\frac{1}{2} \Phi_{xx} u_x^2 + 2\Phi_{xy} u_x u_y + \Phi_{yy} u_y^2 \right) = \Phi_{xx} u_x + \Phi_{xy} u_y = \sum_\beta \Phi_{\alpha\beta} u_\beta.$$

$$\tag{11}$$

It can be verified that a plane wave-like function is a solution for this second order differential equation:

$$\mathbf{u}_{l\alpha} = \frac{b_{l\alpha}}{\sqrt{m_l}} \exp(-i\omega t), \tag{13}$$

where $b_{l\alpha}$ is the time independent amplitude of the wave, while the time dependency is contained only in the phase. With this *ansatz* solution, one gets the result:

$$\sum_{l',\beta} \left(\Phi_{\alpha\beta}(ll') - \omega^2 m_l \delta_{ll'} \delta_{\alpha\beta} \right) u_{l'\beta} = 0. \tag{14}$$

The sum runs over all atoms in the system, that means if there are N primitives unit cells in the crystal and τ atoms in a primitive unit cell, this problem has $3\tau N$ solutions, which we now label with the index q. Each of these solutions describes a collective motion, where all atoms undergo a vibration around their equilibrium position with the same frequency ω_q. All these vibrations are independent from each other.

3.3   The concept of phonon as a quantum of energy

In the last section we showed that the wave function of an harmonic oscillator is a solution of the Hamiltonian problem. In order to describe the phonon as a quantum of energy, we have to re-write the Hamiltonian in a way that simplifies it, through a change of variables, which is also called coordinate transformation. By promoting momenta and displacements to quantum operators and introducing the creation and annihilation operators:

$$b_q^+ = \left(\frac{1}{2\hbar} \right)^{1/2} \sum_{l,\alpha} b_{l,\alpha} \left((m_l)^{1/2} \omega_q - i \frac{p_{l\alpha}}{(m_l\omega_q)^{1/2}} \right) \tag{15}$$

and

$$b_q = \left(\frac{1}{2\hbar} \right)^{1/2} \sum_{l,\alpha} b_{l,\alpha} \left((m_l)^{1/2} \omega_q + i \frac{p_{l\alpha}}{(m_l\omega_q)^{1/2}} \right), \tag{16}$$

respectively, one can write the Hamiltonian $H_{\mathrm{ion,harm}}$ defined in the previous section as:

$$H_{\mathrm{ion,harm}} = \sum_{q=1}^{3\tau N} \hbar\omega_q \left(b_q^+ b_q + \frac{1}{2} \right). \tag{17}$$

The eigenstates of this are the wave function of harmonic oscillators of frequency ω_q. The difference in energy between two contiguous excited states ($\hbar\omega_q$) is called *phonon* . In other terms, the phonon is the quantum of energy necessary to raise one of these harmonic oscillators from its original energy level to the next higher or lower one.

3.4   Born-von Karman boundary conditions

In the previous discussion, we have not assumed that the crystal is perfectly periodic, therefore the results obtained are valid also in the case that this condition is not fulfilled, like for example

when the crystal contains impurity or deviations from the perfect periodicity or even when the solid under consideration is not a crystal. The existence of elementary excitations like phonons, is therefore not only characteristic of a perfect periodic lattice, but it is a rather more general feature of systems for which the kinetic energy and the potential energy are positive-definite quadratic forms, and hence can be simultaneously diagonalized. The same observation is valid for magnons, or other elementary excitations in solids. Nevertheless, when the periodicity conditions are applied to displacements amplitudes, the wavevector of the excitations can only assume discrete values in the first Brillouin zone of the crystal. These characteristics are usually called Born-von Karman boundary conditions:

$$\mathbf{u}_{l\alpha} = \frac{b_{l\alpha}}{\sqrt{m_l}} \exp(-i\omega t) = \frac{b_{l\alpha}}{\sqrt{m_l}} \exp(i\mathbf{k} \cdot \mathbf{R_{kl}} - i\omega t), \tag{18}$$

and help to calculate the function $\omega_{\mathbf{q}}$ as a function of $\mathbf{q}$, which is in the limit of small $\mathbf{q}$:

$$\omega_j(\mathbf{q}) = c_j \left(\frac{\mathbf{q}}{q} \right) q. \tag{19}$$

The label j indicates that $\omega_{\mathbf{q}}$ is a multivalued function of $\mathbf{q}$. In a system of $3N\tau$ atoms, there are 3 acoustic and $3N\tau - 3$ optic branches of this function, which are distinguished according to their different behaviour at small $\mathbf{q}$: only the acoustic branches have a vanishing value for $\mathbf{q} \to 0$. Indeed they describe a different kind of motion: the optical branches describe motions in which the sublattices vibrate in counter-phase against each other. The last concept we introduce here is dispersion curve: with this name we address the relation between the frequency $\omega_j(\mathbf{q})$ and the wavevector $\mathbf{q}$ of a normal mode of vibration in the crystal. Dispersion curves of phonons in crystals can be measured deploying inelastic scattering techniques, both with X-rays and neutrons, how it will be described in another dedicated lecture.

4   Magnons

The second type of excitations we introduce in this lecture are called magnons. In a way similar to the definition of phonons, magnons are usually addressed as: "the quantum of energy that separates two contiguous states of the **spin**-lattice vibrations". Again, as already done for the phonons in the previous section, we will describe the Hamiltonian of the spin-lattice system (Heisenberg model) and how the harmonic approximation leads again to approximate it to a system of harmonic oscillators. Again, by applying the periodicity conditions one can derive a description for the magnons dispersion curves.

4.1   Heisenberg model Hamiltonian

The most important term in the H_{el} in Equation (1) derives from spins localized on the lattice sites of primitive cells in an external magnetic field $\mathbf{B}$. This situation is described in the Heisenberg-model that describes the Hamiltonian of the spin system. Conventionally the external field is oriented along the z-axis and the Hamiltonian can be written in the following form:

$$H_{el} = \gamma \mu_B \mathbf{B} \cdot \sum_i \mathbf{S}_i - \sum_{i \neq j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j. \tag{20}$$

In this expression S_i is the spin vector of the i-th atom measured in units of $\hbar$, the first term is the potential energy of the spins in the external magnetic field $\mathbf{B}$, γ is the Landé factor or gyromagnetic ratio ($\gamma \approx 2.00232$ for a single electron) and μ_B is the Bohr magneton. The second term is the exchange energy between the i-th and the j-th atom. The exchange integral J_{ij} decays fast with increasing distance between the atoms i and j, and in practice the interaction is limited to next neighbours and it is assumed to be symmetric, i.e. $J_{ij} = J_{ji}$.

In a way similar to the case of phonons, in order to describe the magnon as a quantum of energy, we have to change variables (Holstein-Primakoff transformation), promote the spin to a quantum mechanical spin operator and define the creation and annihilation operators:

$$a = \frac{1}{(2S)^{1/2}}(S_x + iS_y) \tag{21}$$

$$a^+ = \frac{1}{(2S)^{1/2}}(S_x - iS_y), \tag{22}$$

and re-write the Hamiltonian as:

$$H_{el} = E_0 - \gamma\mu_B\mathbf{B}\sum_i a_i^+ a_i - 2S\sum_{i\neq j} J_{ij}a_i^+ a_j - 2S\sum_i J_{ii}a_i^+ a_i. \tag{23}$$

Before showing the solutions of the Hamiltonian problem, we first discuss a few aspects related to this transformation in the next section and therefore provide its justification. Moreover, we notice that the term E_0 in the Hamiltonian is the energy of a completely saturated state in which all the spins are parallel to the external field, and it is defined as:

$$E_0 = -\gamma\mu_B BS - S^2\sum_{i\neq j} J_{ij}. \tag{24}$$

The applied approximations assume small deviations of the spins from their maximum S and thus only small excitation energies. This would be rigorously true only for small temperatures and the approximation has proven to be reasonable in many cases. For example, at low temperature high energy excitations do not contribute significantly to the thermodynamic properties of ferromagnets, for the reason that high energy states are less populated than low energy ones, because of their statistical distribution function (Boltzmann-like function).

4.2   Holstein-Primakoff transformation of the Heisenberg model.

In the previous section we introduced the spin operator for a spin state $|m\rangle$ along the z direction:

$$S_z|m\rangle = m|m\rangle \tag{25}$$

that satisfies the commutation relations with the spin operators along the other orthogonal direction:

$$[S_x, S_y] = iS_z. \tag{26}$$

The Holstein-Primakoff transformation, which we applied in the previous section, makes use of the commutation relations and introduces the spin deviation quantum number, that is the deviation of S_z from the maximum value S:

$$n = S - S_z. \tag{27}$$

In the representation where n is diagonal, the spin operators S_x, S_y, S_z have defined matrix elements which can be calculated, therefore it is a useful representation to define the annihilation and creation operators a and a^+ respectively, and $n = a^+ a$. a and a^+ connect two states that differ by one quantum spin number, while n is the spin counterpart of the phonon operator. The spin wave approximation defines the spin operators as:

$$S_x = \left(\frac{S}{2}\right)^{1/2}(a + a^+) \tag{28}$$

$$S_y = \frac{1}{i}\left(\frac{S}{2}\right)^{1/2}(a - a^+) \tag{29}$$

$$S_x = (S - a^+ a), \tag{30}$$

which led to their inverse transformation in Eq.(21) and Eq.(22). This definition is rigorously valid only for values of the deviation n much smaller than the maximum spin value S, nevertheless this limit is ignored. Moreover, the Hamiltonian can contain further contributions, coming for instance from crystalline anisotropies coupling the spins to certain crystallographic directions, or the dipole-dipole interaction of the spins. These contributions are usually smaller than the terms written down and are neglected in first approximation. As already mentioned for the phonons, we can treat the simplified problem first and then include additional effect using a perturbative approach.

4.3   Normal coordinate transformation

In order to describe the magnon as a quantum of energy, we need to apply the coordinate transformation:

$$a_i = \sqrt{\frac{1}{N}}\sum_{\mathbf{q}} a_{\mathbf{q}}e^{-i\mathbf{q}\cdot\mathbf{R}_i}, \tag{31}$$

$$a_i^+ = \sqrt{\frac{1}{N}}\sum_{\mathbf{q}} a_{\mathbf{q}}^+ e^{i\mathbf{q}\cdot\mathbf{R}_i}. \tag{32}$$

Here R_i defines the position of the i-th spin in the lattice. The periodic boundary conditions determine the number of allowed values for the wavevector $\mathbf{q}$.

With this new definition of coordinates, the Hamiltonian takes the form:

$$H_{el,harm} = E_0 + \sum_{\mathbf{q}} \hbar\Omega_{\mathbf{q}} a_{\mathbf{q}}^+ a_{\mathbf{q}}, \tag{33}$$

which is similar to the $H_{ion,harm}$ obtained for phonons in a crystal. Therefore, plane-wave solutions for this problem exist and the wavevector $\mathbf{q}$ is the wavevector of the related harmonic oscillator. The energy levels are separated by the quantum of energy $\hbar\Omega_{\mathbf{q}}$ which is called *magnon*. The result is valid with the assumption to neglect magnon-magnon interaction.

To find a description of $\hbar\Omega_{\mathbf{q}}$ as a function of $\mathbf{q}$ we need to assume that the sum in Equation (33) is limited to first neighbour atoms in a primitive cell. If z is the number of first neighbours in a
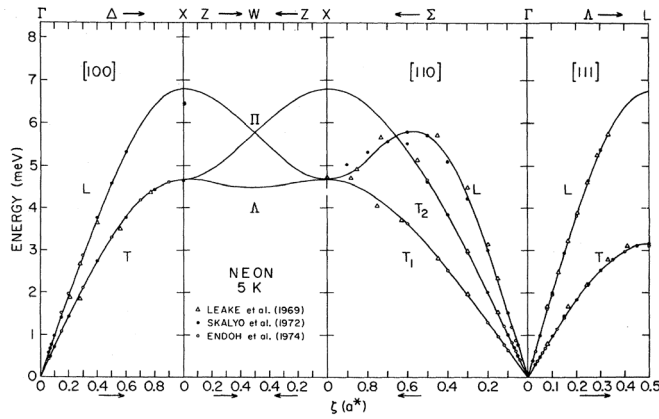
Fig. 1: Phonon dispersion curve in Neon at 5 K [4], shows a typical pattern for a monoatomic crystal. Various acoustic branches are visible, with different behaviour at large values of **q**.

cell and every atom is a center of inversion symmetry, it is possible to show that, in the limit of small **q**:

$$\hbar\omega(\mathbf{q}) = B\gamma\mu_B + 2SJ(z - \sum_l \cos(\mathbf{q} \cdot \mathbf{R_l}) \approx \gamma\mu_B B + SJ \sum_l (\mathbf{q} \cdot \mathbf{R_l})^2 + O(\mathbf{q}^4). \quad (34)$$

This shows that the dispersion $\Omega(\mathbf{q})$ is quadratic in **q** for small **q**, i.e. for long wavelengths, unlike the case of phonons where $\omega(\mathbf{q})$ is linear in q for long wavelengths.

Similar to a phonon dispersion the magnon dispersion can have several branches, e.g., different atoms in the magnetic unit cell or different sublattices can lead to not only one acoustic branch but additionally to one or several optic branches.

5   Examples of dispersion curves

In this section we report three different examples of measurements of dispersion curves in crystals by means of inelastic neutron scattering techniques. Providing an insight into this topic is beyond the scope of this lecture, and it will be given in another one. The first example is the case of solid Neon measured at 5K [4], the second is NaCl at 80K [5], the third is Fe at room temperature [6]. More details about samples and experimental methods can be found in the references. Here these examples are simply mentioned for didactic purpose, in order to show that the theoretical treatment presented in this lecture shows an experimental counterpart, which is in some cases very close to the simplified picture of the crystal as a system of harmonic oscillators.
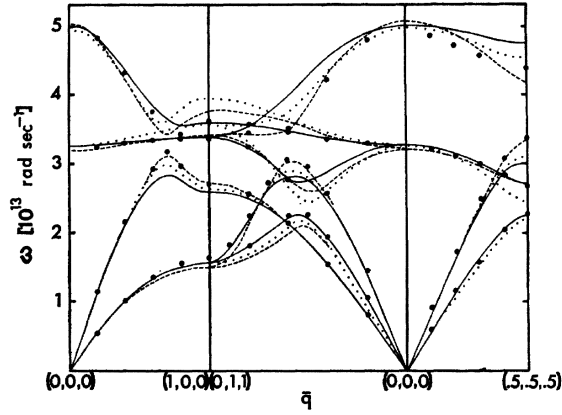
Fig. 2: Phonon dispersion curves in NaCl at 80 K [5] obtained from a neutron scattering experiment. NaCl has two atoms per unit cell, therefore 3 acoustic and 3 optic branches. All of them show different behaviour as function of q.
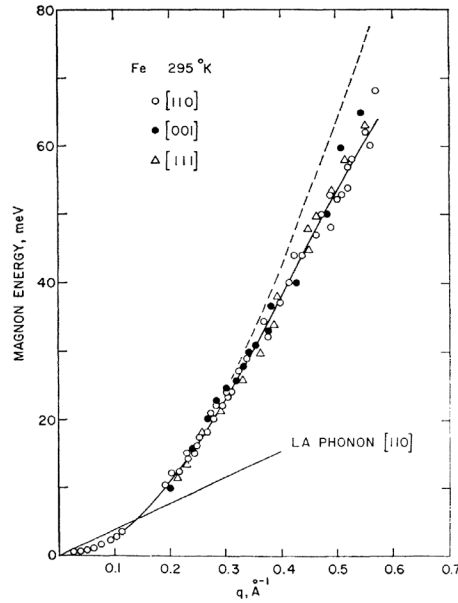
Fig. 3: Magnon dispersion curve in iron at room temperature [6]. The dashed curve represents the predicted dispersion curve for a nearest-neighbour Heisenberg Hamiltonian along [110]. The model is such as the one discussed in the previous section.

References

[1] A.A. Maradudin, *Introduction to elementary excitations in solids* (The Cortina Lectures - 1966, Plenum Press, 1969).

[2] G. L. Squires, *Introduction to the theory of thermal neutron scattering* (Cambridge University Press, Cambridge, 1978).

[3] S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter* (Clarendon Press, Oxford, 1984).

[4] Y.Endoh, G. Shirane, J. Skalyo Jr., Phys. Rev. B **11**, 1681 (1974).

[5] G. Raunio, L. Almqvist, R. Stedman, Phys. Rev. **178**, 1496 (1968).

[6] M.F. Collins, V.J. Minkiewicz, R. Nathans, L. Passell, G. Shirane, Phys. Rev. **179**, 417 (1968).

B 6   Crystal-field theory

Eva Pavarini

Institute for Advanced Simulation

Forschungszentrum Jülich GmbH

Contents

1   Introduction

The central equation of solid-state physics is the eigenvalue problem $\hat{H}\Psi = E\Psi$, defined (in the non-relativistic limit) by the many-body Hamiltonian

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i\neq i'}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|} - \sum_{i,\alpha}\frac{Z_\alpha}{|\boldsymbol{r}_i - \boldsymbol{R}_\alpha|} - \sum_\alpha \frac{1}{2M_\alpha}\nabla_\alpha^2 + \frac{1}{2}\sum_{\alpha\neq\alpha'}\frac{Z_\alpha Z_{\alpha'}}{|\boldsymbol{R}_\alpha - \boldsymbol{R}_{\alpha'}|},$$

where $\{\boldsymbol{r}_i\}$ are the coordinates of the N_e electrons, $\{\boldsymbol{R}_\alpha\}$ those of the N_n nuclei, Z_α the atomic numbers, and M_α the nuclear masses.[1] The Born-Oppenheimer Ansatz

$$\Psi(\{\boldsymbol{r}_i\}, \{\boldsymbol{R}_\alpha\}) = \psi(\{\boldsymbol{r}_i\}; \{\boldsymbol{R}_\alpha\})\Phi(\{\boldsymbol{R}_\alpha\}), \tag{1}$$

splits the Schrödinger equation $\hat{H}\Psi = E\Psi$ into the system

$$\begin{cases} \hat{H}_e\psi(\{\boldsymbol{r}_i\}; \{\boldsymbol{R}_\alpha\}) & = \varepsilon(\{\boldsymbol{R}_\alpha\})\psi(\{\boldsymbol{r}_i\}; \{\boldsymbol{R}_\alpha\}), \\[2mm] \hat{H}_n\Phi(\{\boldsymbol{R}_\alpha\}) & = E\Phi(\{\boldsymbol{R}_\alpha\}), \end{cases} \tag{2}$$

where the Hamilton operator for the electrons ($\hat{H}_e$) and that for the lattice ($\hat{H}_n$) are

$$\begin{aligned} \hat{H}_e &= -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i\neq i'}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|} - \sum_{i,\alpha}\frac{Z_\alpha}{|\boldsymbol{r}_i - \boldsymbol{R}_\alpha|} + \frac{1}{2}\sum_{\alpha\neq\alpha'}\frac{Z_\alpha Z_{\alpha'}}{|\boldsymbol{R}_\alpha - \boldsymbol{R}_{\alpha'}|} \\ &= \hat{T}_e + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn}, \end{aligned} \tag{3}$$

$$\begin{aligned} \hat{H}_n &= -\sum_\alpha \frac{1}{2M_\alpha}\nabla_\alpha^2 + \varepsilon(\{\boldsymbol{R}_\alpha\}) \\ &= \hat{T}_n + \hat{U}_n, \end{aligned} \tag{4}$$

and where in (4) we neglect non-adiabatic corrections.[2] The electronic eigenvalue $\varepsilon(\{\boldsymbol{R}_\alpha\})$ acts as potential for the nuclei and defines a Born-Oppenheimer energy surface. While (3) describes the electronic structure, (4) yields the equilibrium crystal structure of the system and the phonon modes. If the equilibrium structure $\{\boldsymbol{R}_\alpha^0\}$ is known, for example experimentally, we can focus on (3). Because $\hat{V}_{ee}$ is not separable, with increasing N_e, finding the eigenvalues and eigenvectors of (3) becomes quickly an unfeasible task, even for a single atom. The modern approach to such many-body problems consists in building, starting from (3), minimal but material specific low-energy many-body models, which retain the essential physics [1–3].

The first step in model building consists in performing density-functional theory (DFT) calculations. DFT is based on the Hohenberg-Kohn theorem, which states that the ground-state energy of the many-body Hamiltonian (3) is a functional $E[n]$ of the electron density, minimized by the ground-state density. In the Kohn-Sham DFT scheme, the ground-state energy of (3) can be obtained by solving an auxiliary Schrödinger equation $\hat{h}_e\psi = \varepsilon\psi$, with

$$\hat{h}_e = \sum_i \left[-\frac{1}{2}\nabla_i^2 + v_R(\boldsymbol{r}_i) \right] = \sum_i \hat{h}_e(\boldsymbol{r}_i). \tag{5}$$

---

[1] In this lecture we use atomic units (see Appendix A).

[2] The neglected term is $\hat{\Lambda}_n = -\sum_\alpha \frac{1}{M_\alpha}\left[\frac{1}{2}\langle\psi|\nabla_\alpha^2\psi\rangle + \langle\psi|\nabla_\alpha\psi\rangle \cdot \nabla_\alpha\right]$.

The auxiliary Hamiltonian describes N_e non-interacting electrons in an external potential, $v_R(\boldsymbol{r})$, chosen such that the ground-state electron density $n_0(\boldsymbol{r})$ of the auxiliary model equals $n(\boldsymbol{r})$, the ground-state electron density of the original interacting system. This potential can be written as

$$v_R(\boldsymbol{r}) = -\sum_\alpha \frac{Z_\alpha}{|\boldsymbol{r} - \boldsymbol{R}_\alpha|} + \int d\boldsymbol{r}' \frac{n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} + \frac{\delta E_{\mathrm{xc}}[n]}{\delta n} = v_{en}(\boldsymbol{r}) + v_H(\boldsymbol{r}) + v_{xc}(\boldsymbol{r}), \qquad (6)$$

where $v_H(\boldsymbol{r})$ is the long-range Hartree term and $E_{\mathrm{xc}}[n]$ is the so-called exchange-correlation functional. The main difficulty of DFT is that $E_{\mathrm{xc}}[n]$ is not know, and it is therefore necessary to find good approximations for it. Most common are the local-density approximation (LDA) and its extensions; they work remarkably well for several classes of materials and properties.

For *strongly correlated materials*, however, many-body effects *beyond* the LDA play a crucial role, leading to cooperative emergent phenomena; examples are transition-metal oxides with partially filled d-shells, Mott insulators, Kondo systems, and heavy fermions. For these systems simple approximations to $E_{\mathrm{xc}}[n]$ fail, even qualitatively.

Thus, the second step consists in using DFT to construct a *localized* one-electron basis; this is usually achieved building from the Bloch functions $\psi_{n\boldsymbol{k}\sigma}(\boldsymbol{r})$, obtained by solving (5) for a given crystal, material-specific Wannier functions

$$\psi_{in\sigma}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{-i\boldsymbol{R}_i \cdot \boldsymbol{k}} \, \psi_{n\boldsymbol{k}\sigma}(\boldsymbol{r}).$$

Localized Wannier functions can be constructed using different procedures: the downfolding approach, discussed in [4], the maximally-localized Wannier functions algorithm of Marzari and Vanderbilt [5], and the projectors technique.

The third step consists in writing the Hamiltonian (3) in second quantization using such localized Wannier functions as one-electron basis. The resulting many-body Hamiltonian is the sum of an LDA term $\hat{H}^{\mathrm{LDA}}$, a Coulomb term $\hat{U}$, and a double-counting correction $\hat{H}_{\mathrm{DC}}$

$$\hat{H}_e = \hat{H}^{\mathrm{LDA}} + \hat{U} - \hat{H}_{\mathrm{DC}}. \qquad (7)$$

The LDA part of the Hamiltonian is given by

$$\hat{H}^{\mathrm{LDA}} = -\sum_\sigma \sum_{in,i'n'} t_{n,n'}^{i,i'} c_{in\sigma}^\dagger c_{i'n'\sigma}, \qquad (8)$$

where $c_{in\sigma}^\dagger$ ($c_{in\sigma}$) creates (annihilates) an electron of spin σ in orbital n at site i, and

$$t_{n,n'}^{i,i'} = -\int d\boldsymbol{r} \, \overline{\psi_{in\sigma}}(\boldsymbol{r}) [-\frac{1}{2}\nabla^2 + v_{\mathrm{R}}(\boldsymbol{r})] \psi_{i'n'\sigma}(\boldsymbol{r}). \qquad (9)$$

The $i \neq i'$ contributions are the hopping integrals, while the on-site ($i = i'$) term yields the crystal-field matrix

$$\varepsilon_{n,n'}^{i,i} = -t_{n,n'}^{i,i} = \int d\boldsymbol{r} \, \overline{\psi_{in\sigma}}(\boldsymbol{r}) \left[-\frac{1}{2}\nabla^2 + v_{\mathrm{R}}(\boldsymbol{r}) \right] \psi_{in'\sigma}(\boldsymbol{r}). \qquad (10)$$

The Coulomb interaction $\hat{U}$ is given by

$$\hat{U} = \frac{1}{2} \sum_{ii'jj'} \sum_{\sigma\sigma'} \sum_{nn'pp'} U^{ij\,i'j'}_{np\,n'p'} c^\dagger_{in\sigma} c^\dagger_{jp\sigma'} c_{j'p'\sigma'} c_{i'n'\sigma},$$

with

$$U^{iji'j'}_{np\,n'p'} = \int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2 \frac{\overline{\psi_{in\sigma}(\boldsymbol{r}_1)}\,\overline{\psi_{jp\sigma'}(\boldsymbol{r}_2)}\,\psi_{j'p'\sigma'}(\boldsymbol{r}_2)\,\psi_{i'n'\sigma}(\boldsymbol{r}_1)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}. \tag{11}$$

The Coulomb tensor (11) is discussed in [6]. The double counting term $\hat{H}_{\mathrm{DC}}$ cancels the part of the electron-electron interaction contained and already well accounted for in $\hat{H}^{\mathrm{LDA}}$, such as the mean-field part of the exchange-correlation interaction and the long-range Hartree term; the difference $\hat{H}_U = \hat{U} - \hat{H}_{\mathrm{DC}}$ is therefore a short-range many-body correction to the LDA [6].

The Hamiltonian (7) still describes the full many-body problem; further approximations are necessary to make progress. Typically electrons are divided into two types, correlated or heavy electrons (e.g., d or f open shells) and uncorrelated or light electrons. For the correlated electrons the LDA fails qualitatively, and $\hat{H}_U$ has to be accounted for explicitly; for the light electrons we can instead assume that LDA is a good approximations and no correction $\hat{H}_U$ is needed.

The main effect of the light electrons is assumed to be a renormalization of the Coulomb parameters (*screening*), which, as a consequence, cannot be calculated any more as in (11); since the exact screening is not known, approximated schemes such as the constrained LDA or the constrained random-phase approximation are commonly used. The separation of electron in light and heavy is the most delicate aspect of model building, as only in few cases the distinction is really clear cut. In most cases we can only make a reasonable guess, that has to be tested *a posteriori*, e.g., comparing with experiments, or better, when doable, extending the basis of heavy electrons to include, e.g., other states close to the Fermi level.

In the last step, the minimal material-specific many-body model is solved using many-body methods. If the solution method adopted is the dynamical mean-field theory (DMFT) approach, the procedure described above defines the LDA+DMFT method [1–3].

While strong-correlation effects arise from the Coulomb matrix (11), chemistry enters mostly through the hopping integrals (9) and the crystal-field matrix (10). The purpose of this lecture is to explain the physical origin of these parameters, and the role they can play. To do this we will use some basic results of group theory. A short introduction to group theory can be found in Ref. [7], together with a longer version of this lecture. For simplicity, in most derivations we will use atomic hydrogen-like orbitals as a basis; the generalization to Wannier functions is however straightforward.

The lecture is organized as follows. First we analyze how and why, in a crystal or a molecule, the atomic l-shells split, becoming the crystal-field levels. Using group theory, we discuss the effects of symmetry reduction, from O(3) to the site symmetry at a molecular or crystal site. We then show how to calculate the splitting due to the electric field generated at a given site by the surrounding ions. Finally, we analyze covalency effects, which lead to the formations of bonds and bands (hopping integrals), and also contribute to the splitting of atomic levels.
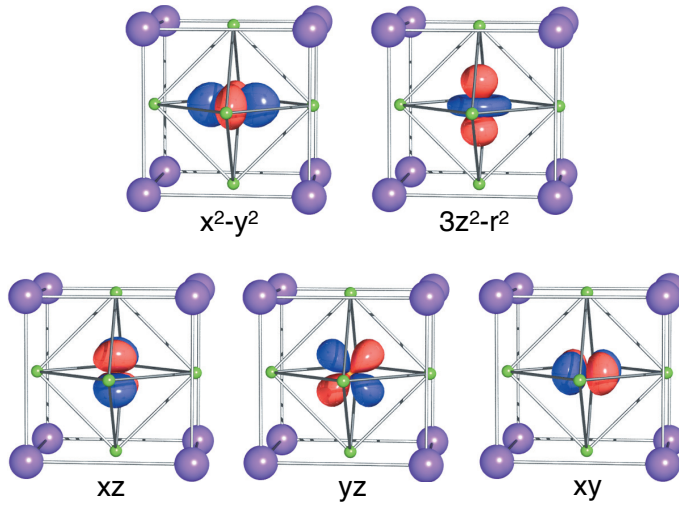
Fig. 1: The Cu e_g and t_{2g} Wannier orbitals for the cubic perovskite $KCuF_3$, obtained from first principles calculations, using a Wannier basis that spans all bands.

2 Crystal-field theory

In an atom, the potential $v_R(\boldsymbol{r})$ which determines the one-electron energies (10) is central and has (at least) all the symmetries of $O(3)$. In a molecule or a solid, $v_R(\boldsymbol{r})$ has in general lower symmetry, the symmetry of a finite point group. Thus electronic states that are degenerate in an atom can split in a solid or a molecule. The symmetry reduction arises from the *crystal field*; the latter has two components, the Coulomb potential generated by the surrounding ions and the *ligand field* due to the bonding neighbors.

2.1 Point-charge model of the crystal field

Let us assume that the crystal is ionic and the ions can be treated as point charges q_α, and let us neglect $v_H(\boldsymbol{r})$ and $v_{xc}(\boldsymbol{r})$ in (6). Then, the one-electron potential can be written as

$$v_R(\boldsymbol{r}) = \sum_\alpha \frac{q_\alpha}{|\boldsymbol{R}_\alpha - \boldsymbol{r}|} = v_0(r) + \sum_{\alpha \neq 0} \frac{q_\alpha}{|\boldsymbol{R}_\alpha - \boldsymbol{r}|} = v_0(r) + v_c(\boldsymbol{r}), \tag{12}$$

where $\boldsymbol{R}_\alpha$ are the positions of the ions and q_α their charges. The term $v_0(r)$ is the ionic central potential at site $\boldsymbol{R}_0$, and has spherical symmetry. The term $v_c(\boldsymbol{r})$ is the electric field generated at a given site $\boldsymbol{R}_0$ by all the surrounding ions in the crystal and it is called *crystal-field potential*.

Let us consider a crystal with the perovskite structure ABC_3, shown in Fig. 1 for the specific case of $KCuF_3$. We want to calculate the crystal-field potential at the site of the transition metal, B. Let us first assume that only the contribution of nearest neighbors (the negative C ions, usually oxygens or fluorine) is relevant. The C ions are located at positions $(\pm a/2, 0, 0), (0, \pm a/2, 0), (0, 0, \pm a/2)$, where a is the lattice constant, and have all the same

charge q_C. Expanding around $\boldsymbol{r} = 0$, we find that the first contribution to $v_c(\boldsymbol{r})$ with less than spherical symmetry is

$$v_{\text{oct}}(\boldsymbol{r}) = \frac{35}{4}\frac{q_C}{a^5}\left(x^4 + y^4 + z^4 - \frac{3}{5}r^4\right) = D\left(x^4 + y^4 + z^4 - \frac{3}{5}r^4\right).$$

We can rewrite this potential as

$$v_{\text{oct}}(\boldsymbol{r}) = \frac{7}{6}\sqrt{4\pi}\frac{q_C}{a^5}r^4\left[Y_0^4(\theta,\phi) + \sqrt{\frac{5}{14}}\left(Y_4^4(\theta,\phi) + Y_{-4}^4(\theta,\phi)\right)\right], \tag{13}$$

where

$$Y_0^4(\theta,\phi) = \frac{3}{16}\frac{1}{\sqrt{\pi}}(35\cos^4\theta - 30\cos^2\theta + 3),$$

$$Y_{\pm 4}^4(\theta,\phi) = \frac{3}{16}\frac{\sqrt{35}}{\sqrt{2\pi}}\sin^4\theta e^{\pm 4i\phi}.$$

Let us now calculate the crystal field due to the cubic cage of cations A (with charge q_A), shown in Fig. 1. One can show that

$$v_{\text{cube}}(\boldsymbol{r}) = -\frac{8}{9}\frac{q_A}{q_C}v_{\text{oct}}(\boldsymbol{r}),$$

i.e., $v_{\text{cube}}(\boldsymbol{r})$ has the same form as $v_{\text{oct}}(\boldsymbol{r})$; this happens because a cube and an octahedron are dual polyhedra[3] and have therefore the same symmetry properties. If $q_A/q_C > 0$, $v_{\text{cube}}(\boldsymbol{r})$ has opposite sign than $v_{\text{oct}}(\boldsymbol{r})$; however, in the case of a perovskite, cations are positive ions; thus the crystal field due to the A cage has the same sign of the field generated by the B octahedron.

2.2   Group-theory approach

The crystal-field potential $v_c(\boldsymbol{r})$ can split the $(2l+1)$-fold degeneracy of the atomic levels. To calculate how the l manifold splits, we use first group theory. We assume for simplicity that the symmetry is only O (group of the proper rotations which leave a cube invariant); using the full symmetry group of the cube, $O_h = O \otimes C_i$, does not change the result, because the spherical harmonics have fixed parity. The character table of group O is

O	E	$8C_3$	$3C_2$	$6C_2$	$6C_4$
$(x^2+y^2+z^2)$ A_1	1	1	1	1	1
A_2	1	1	1	−1	−1
$(x^2-y^2, 3z^2-r^2)$ E	2	−1	2	0	0
(x,y,z) T_1	3	0	−1	−1	1
(xy,xz,yz) T_2	3	0	−1	1	−1

$$\tag{14}$$

The characters of the reducible matrix representation Γ^l constructed using spherical harmonics with quantum numbers lm as a basis, for a rotation of an angle α, are given by

$$\chi^l(\alpha) = \frac{\sin(l+1/2)\alpha}{\sin\alpha/2}$$

---

[3] Every polyhedron has a dual which can be obtained by exchanging the location of faces and vertices.

Hence

$$\chi^l(C_2) = (-1)^l$$

$$\chi^l(C_3) = \begin{cases} 1 & l = 0, 3, \dots \\ 0 & l = 1, 4, \dots \\ -1 & l = 2, 5, \dots \end{cases}$$

$$\chi^l(C_4) = \begin{cases} 1 & l = 0, 1, 4, 5, \dots \\ -1 & l = 2, 3, 6, 7, \dots \end{cases}$$

For the s, p, d, f shells we can therefore write for representations Γ^l

O	E	$8C_3$	$3C_2$	$6C_2$	$6C_4$
Γ^s	1	1	1	1	1
Γ^p	3	0	-1	-1	1
Γ^d	5	-1	1	1	-1
Γ^f	7	1	-1	-1	-1

We can now determine how the reducible representations Γ^l splits using the decomposition formula

$$\Gamma = a_1 \Gamma_1 \oplus a_2 \Gamma_2 \oplus \cdots = \bigoplus_j a_j \Gamma_j$$

$$a_j = \frac{1}{h} \sum_k N_k [\chi_j(\mathcal{C}_k)]^* \chi(\mathcal{C}_k).$$

Hereafter the symmetry representations of electronic terms are written in lower case. We find

$$\Gamma^s = a_1$$

$$\Gamma^p = t_1$$

$$\Gamma^d = e \oplus t_2$$

$$\Gamma^f = a_2 \oplus t_1 \oplus t_2$$

Thus, the s- and the p-functions do not split, because the a_1 irreducible representation is one-dimensional and the t_1 irreducible representation is 3-dimensional. However, d-functions split into a doublet and a triplet, while f-functions into a singlet and two triplets. To calculate which functions belong to which representation we can, e.g., use projectors [7]. For d-electrons, relevant for the case of a transition-metal ion, we find that the d-shell splits into e ($x^2 - y^2, 3z^2 - r^2$) and t_2 (xy, xz, yz). The partner functions for the representations of group O are given in the first column of the character table (14), on the left.

As discussed, the full symmetry of the B site is O_h. The group O_h can be obtained as direct product, $O_h = O \otimes C_i$; with respect to O, the group O_h has twice the number of elements and classes, and thus twice the number of irreducible representations. The latter split into even ($a_{1g}, a_{2g}, e_g, t_{1g}, t_{2g}$) and odd ($a_{1u}, a_{2u}, e_u, t_{1u}, t_{2u}$) representations. The d-functions are even, and therefore $x^2 - y^2$ and $3z^2 - r^2$ are partners functions for the e_g irreducibile representation, while xy, xz, yz are partner functions for the t_{2g} irreducible representation. The p-orbitals are odd, and are partners functions for the t_{1u} representation.

2.3 Ionic effects

Group theory tells us *if* the degenerate $2l + 1$ levels split at a given site in a lattice, but not of *how much* they do split, and which orbitals are higher in energy. For the perovskite structure, we can however calculate approximately the crystal-field splitting at site B, using (13). The central atom B is a transition-metal ion in a $3d^1$ configuration (e.g., Ti^{3+} or V^{4+}), which has degeneracy $2l + 1 = 5$. The octahedral potential $v_{\mathrm{oct}}(\boldsymbol{r})$ yields the following element of matrix between states in the d^1 manifold

$$
\begin{aligned}
\langle \psi_{n20}\ |\hat{v}_{\mathrm{oct}}|\psi_{n20}\ \rangle &= +6Dq \\
\langle \psi_{n2\pm1}|\hat{v}_{\mathrm{oct}}|\psi_{n2\pm1}\rangle &= -4Dq \\
\langle \psi_{n2\pm2}|\hat{v}_{\mathrm{oct}}|\psi_{n2\pm2}\rangle &= +\ Dq \\
\langle \psi_{n2\pm2}|\hat{v}_{\mathrm{oct}}|\psi_{n2\mp2}\rangle &= +5Dq
\end{aligned}
$$

where $Dq = -q_C\langle r^4\rangle/6a^5$. The crystal-field splitting between e_g and t_{2g}-states can be the obtained by diagonalizing the crystal-field matrix

$$
H_{\mathrm{CF}} = \begin{pmatrix}
Dq & 0 & 0 & 0 & 5Dq \\
0 & -4Dq & 0 & 0 & 0 \\
0 & 0 & 6Dq & 0 & 0 \\
0 & 0 & 0 & -4Dq & 0 \\
5Dq & 0 & 0 & 0 & Dq
\end{pmatrix}.
$$

We find two degenerate e_g eigenvectors with energy $6Dq$

$$
\begin{aligned}
|\psi_{n20}\rangle &= |3z^2 - r^2\rangle, \\
\frac{1}{\sqrt{2}}\left[|\psi_{n22}\rangle + |\psi_{n2-2}\rangle\right] &= |x^2 - y^2\rangle,
\end{aligned}
$$

and three degenerate t_{2g} eigenvectors with energy $-4Dq$

$$
\begin{aligned}
\frac{i}{\sqrt{2}}\left[|\psi_{n22}\rangle - |\psi_{n2-2}\rangle\right] &= |xy\rangle, \\
\frac{1}{\sqrt{2}}\left[|\psi_{n21}\rangle - |\psi_{n2-1}\rangle\right] &= |xz\rangle, \\
\frac{i}{\sqrt{2}}\left[|\psi_{n21}\rangle + |\psi_{n2-1}\rangle\right] &= |yz\rangle.
\end{aligned}
$$

The splitting is

$$
\Delta = E_{e_g} - E_{t_{2g}} = 10Dq.
$$

Thus the e_g-states are higher in energy than the t_{2g}-states. This happens because e_g electrons point towards the negative C ions (see Fig. 1), and will therefore feel a larger Coulomb repulsion than t_{2g} electrons, which point between the negative C ions.

For a generic lattice, we can expand the crystal-field potential (12) in spherical harmonics using

$$
\frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} = \sum_{k=0}^{\infty} \frac{r_<^k}{r_>^{k+1}} \frac{4\pi}{2k+1} \sum_{q=-k}^{k} Y_q^k(\theta_2, \phi_2)\overline{Y}_q^k(\theta_1, \phi_1),
$$

where $r_<$ ($r_>$) is the smaller (larger) of r_1 and r_2. The crystal-field potential can then be written as

$$v_c(\boldsymbol{r}) = \sum_{k=0}^{\infty} \sum_{q=-k}^{k} B_q^k Y_q^k, \tag{15}$$

where $B_q^k = (-1)^q \bar{B}_{-q}^k$. Although the series in (15) is in principle infinite, one can terminate it by specifying the wavefunctions, since

$$\langle Y_m^l | Y_q^k | Y_{m'}^l \rangle = 0 \quad \text{if} \quad k > 2l.$$

For example, for p electrons $k \leq 2$, for d-electrons, $k \leq 4$, and f electrons $k \leq 6$. Thus, for d-electrons and O_h symmetry, the terms that appear in the potential (13) are actually also the only ones to be taken into account.

The derivation of (13) and (15) presented here might let us think that the first nearest neighbors are those that determine the crystal field. However, this is often not the case, because Coulomb repulsion is a long-range interaction; for example, in some systems the first nearest neighbors yield cubic symmetry at a given site but further neighbors lower the symmetry.[4]

The point charge model discussed in this section is useful to explain the relation between crystal field and site symmetry, however yields unsatisfactory results for the crystal-field splitting in real materials. Corrections beyond the point-charge approximation turn out to be important. In addition, as we will see in the next section, in many systems the crystal field has a large, sometimes dominant, covalent contribution, the ligand field. The modern approach to calculate crystal-field splittings including the ligand-field contribution is based on material-specific DFT potentials and DFT localized Wannier functions as one-electron basis. We will discuss this approach at the end of the next section.

Let us now analyze the splitting of energy levels in a many-electron $3d^n$ manifold. Apart from the crystal field (15), in calculating the energies of states in such manifold, we have also to take into account the electron-electron Coulomb repulsion. Here we briefly discuss some simple examples: $3d^1$, $3d^9$ and $3d^2$. We have seen that for a d-electron surrounded by an octahedron of negative ions, $\Delta = 10Dq$; the energy difference between the electronic configuration e_g^1 and electronic configuration t_{2g}^1 is therefore Δ. In the case of a single hole in the d-shell ($3d^9$ ion, e.g., Cu^{2+}), the energy difference between $t_{2g}^6 e_g^3$ and $t_{2g}^5 e_g^4$, is then just $-\Delta$, because of electron-hole symmetry. The d crystal-field orbitals (Wannier functions) for the $3d^9$ perovskite $KCuF_3$ (cubic structure) are shown in Fig. 1. For a generic $3d^n$ configuration we can consider two limit cases, strong or weak crystal field. If the crystal field is *strong*, one can treat Coulomb electron-electron interaction as a perturbation, and classify the atomic states according to the crystal field. Let us consider the case of a perovskite in which the central ion has electronic configuration $3t_{2g}^2$ (e.g., V^{3+}); if we neglect the electron-electron repulsion, the excited states are $t_{2g}^1 e_g^1$, with energy Δ, and e_g^2, with energy 2Δ. We can obtain a representation of the group O_h in the basis of two-electron states from the direct product of the representations in the basis

---

[4] This means that O_h is not the site symmetry.

of single-electron states. By using the decomposition formula we have

$$t_{2g} \otimes t_{2g} = a_{1g} \oplus e_g \oplus t_{1g} \oplus t_{2g}$$

$$e_g \otimes t_{2g} = t_{1g} \oplus t_{2g}$$

$$e_g \otimes e_g = a_{1g} \oplus a_{2g} \oplus e_g$$

The Coulomb repulsion acts as a perturbation and can split degenerate states belonging to different irreducible representations. In particular, the manifold t_{2g}^2 splits into $^1a_{1g}$, ^1e_g, $^1t_{2g}$, and $^3t_{1g}$ (ground state), where $^{(2S+1)}$ indicates the spin degeneracy of the state.

If the crystal field is *weak*, the opposite approach can be used; the crystal field is treated as a perturbation of the atomic Coulomb multiplets, labeled as ^{2S+1}L. In this case the two-electron ground state is the triplet ^3F and the O_h crystal field splits it into $^3t_{1g}$, $^3t_{2g}$, and $^3a_{2g}$.

Up to here we have neglected the spin-orbit interaction. The latter plays an important role, e.g., in $5d$- or f-systems. In the case in which the crystal field is weak with respect to the spin-orbit coupling, as it happens in many f-electron compounds, the total angular momentum J is a good quantum number. It is therefore useful to construct a reducible representation of the point group, Γ^J, in the basis of the eigenvectors of total angular momentum. The characters are

$$\chi^J(\alpha) = \frac{\sin(J + 1/2)\alpha}{\sin \alpha/2},$$

For half-integral values of J (odd number of electrons), $\chi^J(\alpha)$ has the property

$$\chi^J(\alpha + 2\pi) = -\chi^J(\alpha).$$

We therefore expand the original point group to include a new element, R, which represents the rotation by 2π. The new group has twice the number of elements of the original group and is known as *double group*. In the case of the group O the double group is labeled with O' and its character table is

O'	E	$8C_3$	$3C_2 + 3RC_2$	$6C_2 + 6RC_2$	$6C_4$	R	$8RC_3$	$6RC_4$
Γ_1	1	1	1	1	1	1	1	1
Γ_2	1	1	1	-1	-1	1	1	-1
Γ_3	2	-1	2	0	0	2	-1	0
Γ_4	3	0	-1	-1	1	3	0	1
Γ_5	3	0	-1	1	-1	3	0	-1
Γ_6	2	1	0	0	$\sqrt{2}$	-2	-1	$-\sqrt{2}$
Γ_7	2	1	0	0	$-\sqrt{2}$	-2	-1	$\sqrt{2}$
Γ_8	4	-1	0	0	0	-4	1	0

To determine if the atomic levels in a given J manifold split we use the same procedure adopted

for the l-shell. First we calculate the characters of all elements in the group

$$\chi^J(E) = 2J+1$$
$$\chi^J(R) = -(2J+1)$$
$$\chi^J(C_2) = 0$$
$$\chi^J(RC_2) = 0$$

$$\chi^J(C_3) = \begin{cases} 1 & J = 1/2, 7/2, \ldots \\ -1 & J = 3/2, 9/2, \ldots \\ 0 & J = 5/2, 11/2 \ldots \end{cases}$$

$$\chi^J(RC_3) = \begin{cases} -1 & J = 1/2, 7/2, \ldots \\ 1 & J = 3/2, 9/2, \ldots \\ 0 & J = 5/2, 11/2 \ldots \end{cases}$$

$$\chi^J(C_4) = \begin{cases} \sqrt{2} & J = 1/2, 9/2 \ldots \\ 0 & J = 3/2, 7/2, \ldots \\ -\sqrt{2} & J = 5/2, 13/2, \ldots \end{cases}$$

$$\chi^J(RC_4) = \begin{cases} -\sqrt{2} & J = 1/2, 9/2 \ldots \\ 0 & J = 3/2, 7/2, \ldots \\ +\sqrt{2} & J = 5/2, 13/2, \ldots \end{cases}$$

Next we use the decomposition formula to find how the reducible representation Γ^J is decomposed in irreducible ones. One can show that

$$\Gamma^{\frac{1}{2}} = \Gamma_6$$
$$\Gamma^{\frac{3}{2}} = \Gamma_8$$
$$\Gamma^{\frac{5}{2}} = \Gamma_7 \oplus \Gamma_8$$
$$\Gamma^{\frac{7}{2}} = \Gamma_6 \oplus \Gamma_7 \oplus \Gamma_8$$
$$\Gamma^{\frac{9}{2}} = \Gamma_6 \oplus 2\Gamma_8$$

Since $\Gamma_6, \Gamma_7, \Gamma_8$ have dimensionality $d \geq 2$, all levels remain at least two-fold degenerate. This is an example of *Kramers degeneracy*. Kramers theorem states that, in the presence of (only) electric fields, the energy levels of a system with odd number of fermions are at least two-fold degenerate. Kramers degeneracy is a consequence of time-reversal symmetry.

2.4   Tight-binding theory and covalency effects

In solids, electrons delocalize to form bonds and bands. In the Hamiltonian (8), these arise from the elements of matrix (9), the hopping integrals. But what is the specific form of the Hamiltonian (8) for a given system? Which parameters are large? Which are zero? The simplest way to answer these questions is to use the tight-binding method, which consists in expanding the crystal wavefunctions in the basis of functions centered at each atomic site; here we use as a basis atomic orbitals,[5] $\{\psi_{nlm}(\boldsymbol{r})\}$. Let us first consider a simple example, a homonuclear molecular ion formed by two hydrogen nuclei, located at $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, and one electron. The electronic Hamiltonian for such an H_2^+ molecular ion is

$$\hat{h}_e(\boldsymbol{r}) = -\frac{1}{2}\nabla^2 - \frac{1}{|\boldsymbol{r}-\boldsymbol{R}_1|} - \frac{1}{|\boldsymbol{r}-\boldsymbol{R}_2|} = -\frac{1}{2}\nabla^2 + v(\boldsymbol{r}-\boldsymbol{R}_1) + v(\boldsymbol{r}-\boldsymbol{R}_2) = -\frac{1}{2}\nabla^2 + v_R(\boldsymbol{r}).$$

We take as atomic basis the ground state $1s$ atomic orbitals, $\psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_1)$ and $\psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_2)$; in the free hydrogen atom they have energy ε_{1s}^0. In this basis, the Hamiltonian and the overlap matrix have the form

$$H = \varepsilon_{1s}^0\, O + \begin{pmatrix} \Delta\varepsilon_{1s} & V_{ss\sigma} \\ V_{ss\sigma} & \Delta\varepsilon_{1s} \end{pmatrix} \qquad O = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}$$

where

$$\Delta\varepsilon_{1s} = \int d\boldsymbol{r}\, \psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_\alpha)\left[v_R(\boldsymbol{r}) - v(\boldsymbol{r}-\boldsymbol{R}_\alpha)\right]\psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_\alpha), \quad \alpha = 1, 2$$

$$V_{ss\sigma} = \int d\boldsymbol{r}\, \psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_\alpha)v(\boldsymbol{r}-\boldsymbol{R}_\alpha)\psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_{\alpha'}), \quad \alpha \neq \alpha'$$

$$S = \int d\boldsymbol{r}\, \psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_\alpha)\psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_{\alpha'}), \quad \alpha \neq \alpha'.$$

The hopping integral $t = -V_{ss\sigma} > 0$ is a Slater-Koster two-center integral (Appendix B).

The ground state of the molecular ion is the *bonding* linear combination

$$\phi_{1s}^B(\boldsymbol{r}) = \left[\psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_1) + \psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_2)\right]/\sqrt{2(1+S)},$$

and has energy

$$E_B = \varepsilon_{1s}^0 + \frac{\Delta\varepsilon_{1s} + V_{ss\sigma}}{1+S}.$$

The label σ in $V_{ss\sigma}$ indicates that the bonding state is symmetrical with respect to rotations about the bond axis (see Fig. 2). The excited state is the *antibonding* state

$$\phi_{1s}^A(\boldsymbol{r}) = \left[\psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_1) - \psi_{1s}(\boldsymbol{r}-\boldsymbol{R}_2)\right]/\sqrt{2(1-S)},$$

and has energy

$$E_A = \varepsilon_{1s}^0 + \frac{\Delta\varepsilon_{1s} - V_{ss\sigma}}{1-S}.$$

---

[5] Linear Combination of Atomic Orbitals (LCAO) approach.
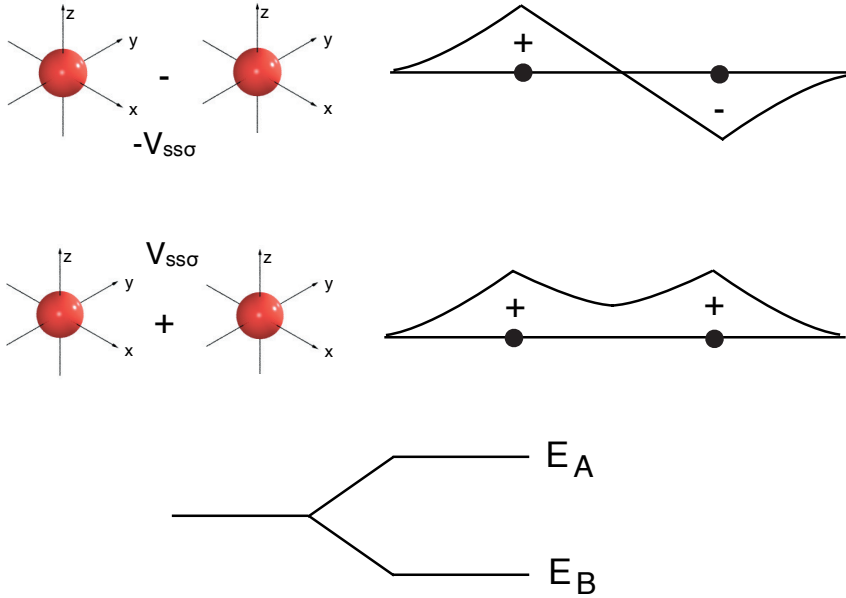
Fig. 2: Pictorial view of the antibonding (top) and bonding (bonding) state of H_2^+.

Let us now consider a crystal. If we neglect $v_{\mathrm{H}}(\boldsymbol{r})$ and $v_{\mathrm{xc}}(\boldsymbol{r})$ in (6), the one-electron Hamiltonian $\hat{h}_e(\boldsymbol{r})$ in (5) becomes

$$\hat{h}_e(\boldsymbol{r}) = -\frac{1}{2}\nabla^2 - \sum_{i,\alpha}\frac{Z_{i,\alpha}}{|\boldsymbol{r}-\boldsymbol{T}_i-\boldsymbol{R}_\alpha|} = -\frac{1}{2}\nabla^2 + \sum_{i,\alpha}v(\boldsymbol{r}-\boldsymbol{T}_i-\boldsymbol{R}_\alpha) = -\frac{1}{2}\nabla^2 + v_R(\boldsymbol{r}),$$

where $\boldsymbol{R}_\alpha$ are the positions of the basis $\{\alpha\}$ atoms in the unit cell and $\boldsymbol{T}_i$ lattice vectors. For each atomic orbital with quantum numbers lm we construct a Bloch state

$$\psi_{lm}^\alpha(\boldsymbol{k},\boldsymbol{r}) = \frac{1}{\sqrt{N}}\sum_i e^{i\boldsymbol{T}_i\cdot\boldsymbol{k}}\,\psi_{lm}(\boldsymbol{r}-\boldsymbol{T}_i-\boldsymbol{R}_\alpha). \tag{16}$$

In the Bloch basis (16), the Hamiltonian and the overlap matrix are given by

$$\begin{aligned}
H_{lm,l'm'}^{\alpha,\alpha'}(\boldsymbol{k}) &= \langle\psi_{lm}^\alpha(\boldsymbol{k})|\hat{h}_e|\psi_{l'm'}^{\alpha'}(\boldsymbol{k})\rangle, \\
O_{lm,l'm'}^{\alpha,\alpha'}(\boldsymbol{k}) &= \langle\psi_{lm}^\alpha(\boldsymbol{k})|\psi_{l'm'}^{\alpha'}(\boldsymbol{k})\rangle.
\end{aligned}$$

They define a generalized eigenvalue problem, the solution of which yields the band structure. The Hamiltonian matrix is given by

$$H_{lm,l'm'}^{\alpha,\alpha'}(\boldsymbol{k}) = \varepsilon_{l'\alpha'}^0 O_{lm,l'm'}^{\alpha,\alpha'}(\boldsymbol{k}) + \Delta\varepsilon_{lm,l'm'}^\alpha\delta_{\alpha,\alpha'} - \frac{1}{N}\sum_{i\alpha\neq i'\alpha'} e^{i(\boldsymbol{T}_{i'}-\boldsymbol{T}_i)\cdot\boldsymbol{k}}\,t_{lm,l'm'}^{i\alpha,i'\alpha'}.$$

Here $\varepsilon_{l\alpha}^0$ are atomic levels, and $\Delta\varepsilon_{lm,l'm'}^\alpha$ the crystal-field matrix

$$\Delta\varepsilon_{lm,l'm'}^\alpha = \int d\boldsymbol{r}\,\overline{\psi_{lm}}(\boldsymbol{r}-\boldsymbol{R}_\alpha)[v_R(\boldsymbol{r})-v(\boldsymbol{r}-\boldsymbol{R}_\alpha)]\psi_{l'm'}(\boldsymbol{r}-\boldsymbol{R}_\alpha), \tag{17}$$
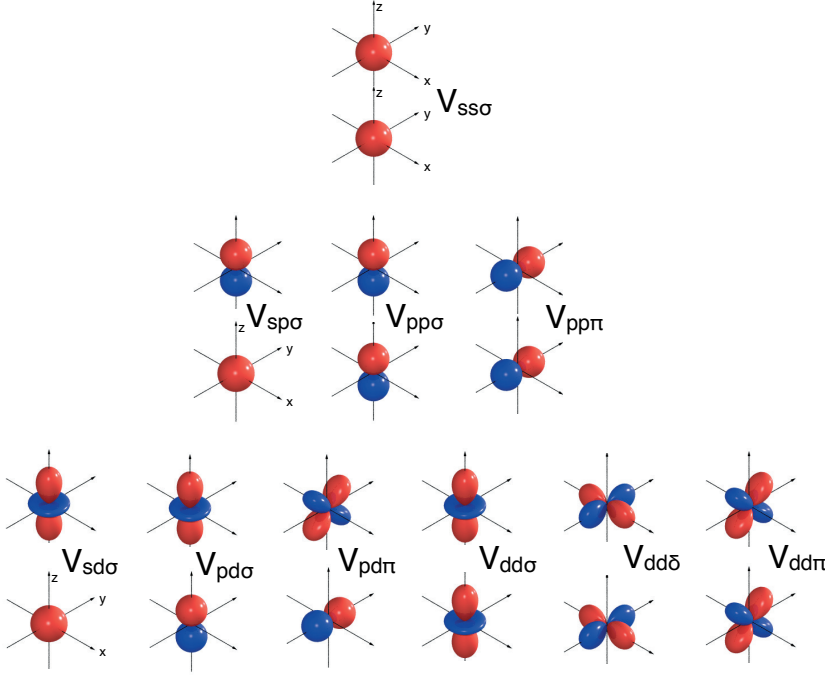
Fig. 3: Independent Slater-Koster two-center integrals for s, p and d atomic orbitals (see Appendix B). The label σ indicates that the bonding state is symmetrical with respect to rotations about the bond axis; the label π that the bond axis lies in a nodal plane; the label δ that the bond axis lies in two nodal planes.

which, as in the case of the H_2^+ ion, is a two-center integral. Finally

$$t^{i\alpha,i'\alpha'}_{lm,l'm'} = -\int d\boldsymbol{r}\, \overline{\psi}_{lm}(\boldsymbol{r} - \boldsymbol{R}_\alpha - \boldsymbol{T}_i)[v_R(\boldsymbol{r}) - v(\boldsymbol{r} - \boldsymbol{R}_{\alpha'} - \boldsymbol{T}_{i'})]\psi_{l'm'}(\boldsymbol{r} - \boldsymbol{R}_{\alpha'} - \boldsymbol{T}_{i'}).$$

The hopping integrals above contain two-center and three-center terms; if the basis is localized, we can neglect the three-center contributions and assume that $t^{i\alpha,i'\alpha'}_{lm,l'm'} \sim -V^{i\alpha,i'\alpha'}_{lm,l'm'}$, where

$$V^{i\alpha,i'\alpha'}_{lm,l'm'} = \int d\boldsymbol{r}\, \overline{\psi}_{lm}(\boldsymbol{r} - \boldsymbol{R}_\alpha - \boldsymbol{T}_i)v(\boldsymbol{r} - \boldsymbol{R}_\alpha - \boldsymbol{T}_i)\psi_{l'm'}(\boldsymbol{r} - \boldsymbol{R}_{\alpha'} - \boldsymbol{T}_{i'})$$

is a two-center integral. A general Slater-Koster two-center integral can be expressed as a function of few independent two-center integrals, shown in Fig. 3 for s, p, and d-functions. A part from the σ bond, which is the strongest, other bonds are possible; the π bonds are made of orbitals which share a nodal plane to which the bond axis belongs, and the δ bond, which has two nodal planes which contain the bond axis and the two ions; furthermore, if the ions on the two sites are different, the bond is *polar*. Fig. 4 shows how to obtain a generic two-center integral involving p and s orbitals.

Let us now consider as an example the e_g and t_{2g} bands of $KCuF_3$; we assume for simplicity that the system is an ideal cubic perovskite (point group O_h). Let us use as a basis only Cu d and F p atomic orbitals, and as matrix elements only on-site terms and pd hopping integrals.
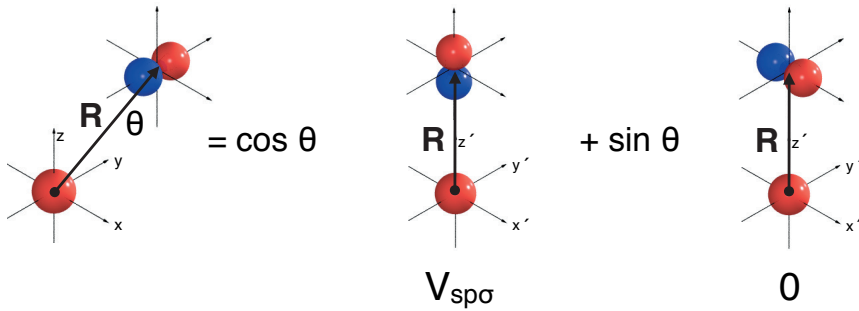
Fig. 4: Illustration of the decomposition of a general s-p two-center integral in terms of $V_{sp\sigma}$.

We label the p-orbitals on different F sites as μ^ν, where $\nu = a, b, c$ identifies the direction of the unit cell axis along which the F atom lies and $\mu = x, y, z$ the orbital; we then construct the corresponding Bloch states $|\boldsymbol{k}\ \mu^\nu\rangle$, as well as the Cu e_g Bloch states $|\boldsymbol{k}\ \mu\rangle$, $\mu = 3z^2 - r^2, x^2 - y^2$. We neglect the overlap matrix for simplicity.

The tight-binding Hamiltonian in this basis has then the form

$H_{e_g}^{\mathrm{TB}}$	$\lvert\boldsymbol{k}\ z^c\rangle$	$\lvert\boldsymbol{k}\ x^a\rangle$	$\lvert\boldsymbol{k}\ y^b\rangle$	$\lvert\boldsymbol{k}\ 3z^2 - r^2\rangle$	$\lvert\boldsymbol{k}\ x^2 - y^2\rangle$
$\lvert\boldsymbol{k}\ z^c\rangle$	ε_p	0	0	$-2V_{pd\sigma}s_z$	0
$\lvert\boldsymbol{k}\ x^a\rangle$	0	ε_p	0	$V_{pd\sigma}s_x$	$-\sqrt{3}V_{pd\sigma}s_x$
$\lvert\boldsymbol{k}\ y^b\rangle$	0	0	ε_p	$V_{pd\sigma}s_y$	$\sqrt{3}V_{pd\sigma}s_y$
$\lvert\boldsymbol{k}\ 3z^2 - r^2\rangle$	$-2V_{pd\sigma}\bar{s}_z$	$V_{pd\sigma}\bar{s}_x$	$V_{pd\sigma}\bar{s}_y$	ε_d	0
$\lvert\boldsymbol{k}\ x^2 - y^2\rangle$	0	$-\sqrt{3}V_{pd\sigma}\bar{s}_x$	$\sqrt{3}V_{pd\sigma}\bar{s}_y$	0	ε_d

where $s_\alpha = ie^{-ik_\alpha a/2}\sin k_\alpha a/2$, $\alpha = x, y, z$, $\varepsilon_p < \varepsilon_d = \varepsilon_p + \Delta_{pd}$, and $V_{pd\sigma} < 0$. If $|V_{pd\sigma}|/\Delta_{pd}$ is small, the occupied bonding-like bands have mostly F p character, while the partially filled antibonding-like bands have mostly Cu e_g character. The energies ε_d and ε_p include the crystal-field term (17). We now calculate the e_g-like bands along high-symmetry lines.[6] Along the Γ-X direction we find the dispersion relations for the e_g-like bands

$$\varepsilon_2(\boldsymbol{k}) = \varepsilon_d$$

$$\varepsilon_1(\boldsymbol{k}) = \varepsilon_p + \frac{\Delta_{pd}}{2} + \frac{\sqrt{\Delta_{pd}^2 + 16V_{pd\sigma}^2|s_x|^2}}{2}$$
$$\sim \varepsilon_d + 2t - 2t\cos k_x a \qquad (18)$$

where $t = V_{pd\sigma}^2/\Delta_{pd}$; in the last step (18) we have assumed that $|V_{pd\sigma}|/\Delta_{pd}$ is small. We can repeat the calculation for the t_{2g} bands. In this case the simplest tight-binding Hamiltonian is

$H_{t_{2g}}^{\mathrm{TB}}$	$\lvert\boldsymbol{k}\ y^a\rangle$	$\lvert\boldsymbol{k}\ x^b\rangle$	$\lvert\boldsymbol{k}\ xy\rangle$
$\lvert\boldsymbol{k}\ y^a\rangle$	ε_p	0	$2V_{pd\pi}s_x$
$\lvert\boldsymbol{k}\ x^b\rangle$	0	ε_p	$2V_{pd\pi}s_y$
$\lvert\boldsymbol{k}\ xy\rangle$	$2V_{pd\pi}\bar{s}_x$	$2V_{pd\pi}\bar{s}_y$	ε_d

---

[6] Special points: $\Gamma = (0,0,0)$, Z$= (0,0,\pi/a)$, X$= (\pi/a,0,0)$, M$= (\pi/a,\pi/a,0)$, R$= (\pi/a,\pi/a,\pi/a)$.
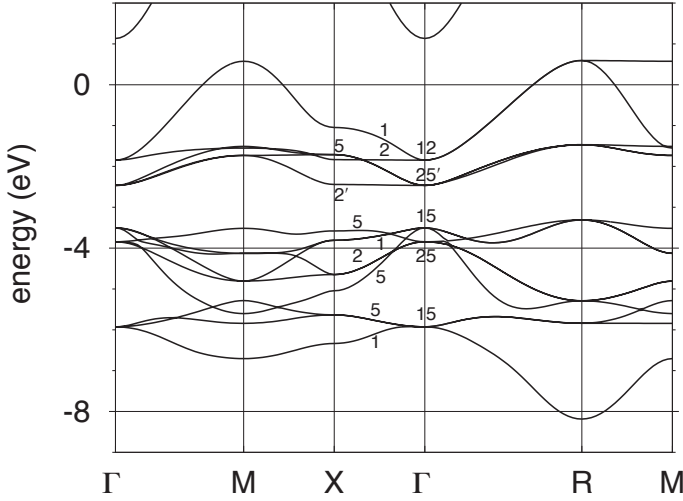
Fig. 5: LDA band structure of cubic KCuF$_3$. Labels along the direction X-Γ indicate the corresponding irreducible representations for the e_g bands.

and cyclic permutations of x, y, z. In the Γ-X direction we find

$$
\varepsilon_{2'}(\boldsymbol{k}) = \varepsilon_d
$$

$$
\begin{aligned}
\varepsilon_5(\boldsymbol{k}) &= \varepsilon_p + \frac{\Delta_{pd}}{2} + \frac{\sqrt{\Delta_{pd}^2 + 16V_{pd\pi}^2|s_x|^2}}{2} \\
&\sim \varepsilon_d + 2t - 2t\cos k_x a
\end{aligned}
$$

where $t = V_{pd\pi}^2/\Delta_{pd}$. The tight-binding model we have used so far is oversimplified, but it already qualitatively describes the e_g and t_{2g} bands in Fig. 5. A more accurate description can be obtained including other Slater-Koster integrals, such as the hopping to apical F s states, or between neighboring F p-states. With increasing number of parameters, it becomes progressively harder to estimate them, e.g., from comparison with experiments; furthermore a large number of fitting parameters makes it impossible to put a theory to a test. However, modern techniques allow us to calculate hopping integrals and crystal-field splittings *ab-initio*, using localized Wannier functions as basis instead of atomic orbitals, and the DFT potential $v_R(\boldsymbol{r})$ as one electron potential; because Wannier functions are orthogonal, the corresponding overlap matrix is diagonal. This leads to the expression (8) for the Hamiltonian, with hopping and crystal-field integrals defined as in (9) and (10).

Let us now return to the crystal-field splitting. In the point charge model discussed in the previous section, the neighboring sites are viewed as ions, and their nature and tendency towards covalent bondings are ignored. In the tight-binding approach described in this section, this corresponds to calculate the terms $\Delta\varepsilon_{lm,l'm'}$ in a basis of atomic orbitals; in the simple tight-binding model considered, this gives the splitting of e_g and t_{2g} bands at the Γ point. However,
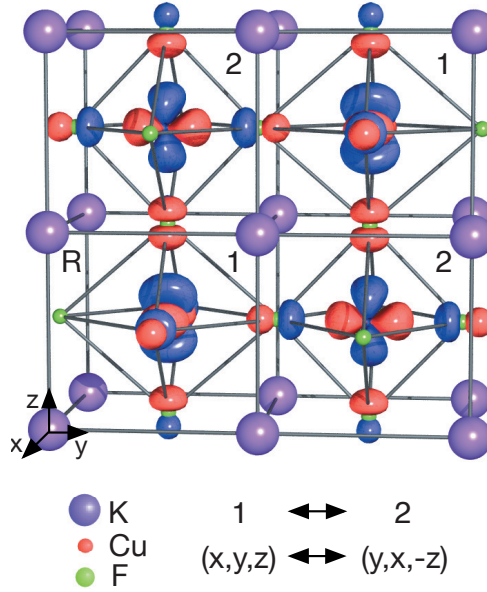
Fig. 6: Cooperative Jahn-Teller distortion and ordering of the e_g hole orbitals in KCuF$_3$. Adapted from Ref. [8]. The Wannier function of the hole orbitals is obtained by downfolding all states but the Cu e_g; thus, differently than the orbitals in Fig. 1, it has p-tails at F sites.

the ligands do matter, because they can form bonding and antibonding states with the central atom. In the case of a cubic perovskite, the t_{2g} and e_g bands are antibonding-like bands; because $V_{pd\sigma}$ (σ bond), relevant for the e_g bands, is larger than $V_{pd\pi}$ (π bond), relevant for the t_{2g} bands, the latter are lower in energy, in agreement with the results of the point-charge model. This ligand field, differently from the crystal field in the point-charge model, is mostly determined by the first shells of neighbors, because the hopping integrals decay fast with distance (Appendix B). We can understand better the effect of the ligands by considering the e_g and t_{2g} tight-binding Hamiltonians $H_{e_g}^{\mathrm{TB}}$ and $H_{t_{2g}}^{\mathrm{TB}}$ at the $\boldsymbol{k} = M = (\pi/a, \pi/a, 0)$ point. The d-like states that diagonalize the Hamiltonians are antibonding combinations of ligand p-functions and transition-metal d-functions. Two of such states can be written as

$$
\begin{aligned}
|M\psi_{x^2-y^2}\rangle &= c_{1d}|M\ x^2-y^2\rangle + c_{1p}\left[|M\ x^a\rangle - |M\ y^b\rangle\right], \\
|M\psi_{xy}\rangle &= c_{2d}|M\ xy\rangle - c_{2p}\left[|M\ y^a\rangle + |M\ x^b\rangle\right],
\end{aligned}
$$

where c_{id}, c_{ip} define the mixing, $i = 1, 2$ and $c_{id}^2 + c_{ip}^2 = 1$. If the atomic xy and $x^2 - y^2$ orbitals are degenerate, the difference in the energy of the two states depends only on the degree of mixing and the Slater-Koster integrals $V_{pd\sigma}$ and $V_{pd\pi}$. For the simple tight-binding models presented for KCuF$_3$, the additional e_g-t_{2g} splitting due to the ligands can thus be estimated as $(W_{e_g} - W_{t_{2g}})/2$, where W_{e_g} and $W_{t_{2g}}$ are the e_g and t_{2g} band width, respectively.

As previously discussed, the modern approach to tight-binding theory consists in using localized Wannier functions, instead of atomic orbitals, as a basis. In this case, one can build Wannier functions which span the e_g and t_{2g} bands only, and which have therefore the effects of the

ligands built-in. This can be seen, e.g., in Fig. 6 for the empty orbital of $KCuF_3$; the Wannier function, obtained by downfolding all states but e_g, has p tails on the neighboring F sites. In the basis of such Wannier functions, the crystal-field splitting, including ligand-field effects, can be obtained directly from the on-site elements (10) of the Hamiltonian.

3   Conclusions

The parameters of the one-electron Hamiltonian are essential ingredients of many-body models. The crystal-field splittings and the hopping integrals carry the information on the lattice and the covalency, and determine to a large extent what makes a system different from the others. The color of a transition-metal complex is for example often determined by the e_g-t_{2g} crystal-field splitting. For a given system, the hopping integrals determine the band structure and the shape of the Fermi surface; the crystal-field splitting plays a crucial role for the local properties, such as the local magnetic moments or spin states, competing with spin-orbit interaction and Coulomb repulsion. In strongly correlated systems, the competition between hopping integrals and Coulomb interaction decides if the system is a metal or a Mott insulator; the crystal-field splitting can however help the formation of a Mott insulating state by reducing the degeneracy of the relevant many-body states [4].

In this lecture we have discussed simple approaches to determine the one-electron parameters for a given system. Such approaches are based on atomic orbitals and symmetries. They are easier to use for high-symmetry systems, in which the number of parameters to determine are small; once the model is constructed, the parameters can be obtained, e.g., by fitting to experiment. In the age of massively parallel supercomputers and standard *ab-initio* codes, it might seem anachronistic to study approximate methods to calculate one-electron parameters. However, these approaches are very useful for understanding qualitatively the behavior of a given system, and the results of complex calculations. It is indeed astonishing how far we can often go in understanding a system with these methods alone. One of the reasons of the successes of tight-binding and crystal-field theory is that symmetries are fully accounted for. In developing approximations to describe numerically complex many-body effects, we should always remember that symmetries are crucial, and taking them into account is essential to understand the properties of a given material.

The modern approach to calculate one-electron parameters is based on *ab-initio* localized Wannier functions; they are built from DFT calculations (e.g., in the LDA approximation), and used as a one-electron basis to construct material-specific many-body models [1–3]. The choice of LDA Wannier functions as a basis relies on the success of the LDA in describing the properties of weakly correlated systems. These successes let us hope that the long-range and the mean-field part of the electron-electron interaction are already well accounted for by the LDA. Thanks to *ab-initio* Wannier functions it is possible to build many-body models even for low-symmetry materials, accounting, e.g., for the effects of small distortions that split the t_{2g} levels [4], a very hard task with semiempirical tight binding. When using Wannier functions as a one-electron basis to build many-body models, we should however never forget what are the assumptions behind; simple models and symmetry considerations remind us where all comes from.

Appendices

A   Constants and units

In this lecture, formulas are given in atomic units. The unit of mass m_0 is the electron mass ($m_0 = m_e$), the unit of charge e_0 is the electron charge ($e_0 = e$), the unit of length a_0 is the Bohr radius ($a_0 = a_B \sim 0.52918$ Å), and the unit of time is $t_0 = 4\pi\varepsilon_0\hbar a_0/e^2$. In these units, m_e, a_B, e and $1/4\pi\varepsilon_0$ have the numerical value 1, the speed of light is $c = 1/\alpha \sim 137$, and the unit of energy is $1\text{Ha} = e^2/4\pi\varepsilon_0 a_0 \sim 27.211$ eV.

B   Atomic orbitals

B.1   Radial functions

The nlm hydrogen-like atomic orbital is given by

$$\psi_{nlm}(\rho, \theta, \phi) = R_{nl}(\rho)Y_l^m(\theta, \phi),$$

where $R_{nl}(\rho)$ is the radial function and $Y_m^l(\theta, \phi)$ a spherical harmonic, $\rho = Zr$ and Z the atomic number. In atomic units, the radial functions are

$$R_{nl}(\rho) = \sqrt{\left(\frac{2Z}{n}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}}\, e^{-\rho/n} \left(\frac{2\rho}{n}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2\rho}{n}\right),$$

where L_{n-l-1}^{2l+1} are generalized Laguerre polynomials of degree $n - l - 1$.

The radial function for $n = 1, 2, 3$ are

$$
\begin{aligned}
R_{1s}(\rho) &= & 2 \quad & Z^{3/2} & & e^{-\rho} \\
R_{2s}(\rho) &= & \tfrac{1}{2\sqrt{2}} \quad & Z^{3/2} & & (2 - \rho)\, e^{-\rho/2} \\
R_{2p}(\rho) &= & \tfrac{1}{2\sqrt{6}} \quad & Z^{3/2} & & \rho\, e^{-\rho/2} \\
R_{3s}(\rho) &= & \tfrac{2}{3\sqrt{3}} \quad & Z^{3/2} & & (1 - 2\rho/3 + 2\rho^2/27)\, e^{-\rho/3} \\
R_{3p}(\rho) &= & \tfrac{4\sqrt{2}}{9\sqrt{3}} \quad & Z^{3/2} & & \rho(1 - \rho/6)\, e^{-\rho/3} \\
R_{3d}(\rho) &= & \tfrac{2\sqrt{2}}{81\sqrt{15}} \quad & Z^{3/2} & & \rho^2\, e^{-\rho/3}
\end{aligned}
$$

where we used the standard notation s for $l = 0$, p for $l = 1$ and d for $l = 2$.

B.2   Real harmonics

To study solids, it is usually convenient to work in the basis of real harmonics. The latter are defined in terms of the spherical harmonics as $y_{l0} = Y_0^l$, $\quad y_{lm} = \frac{1}{\sqrt{2}}(Y_{-m}^l + (-1)^m Y_m^l)$, $\quad y_{l-m} =$
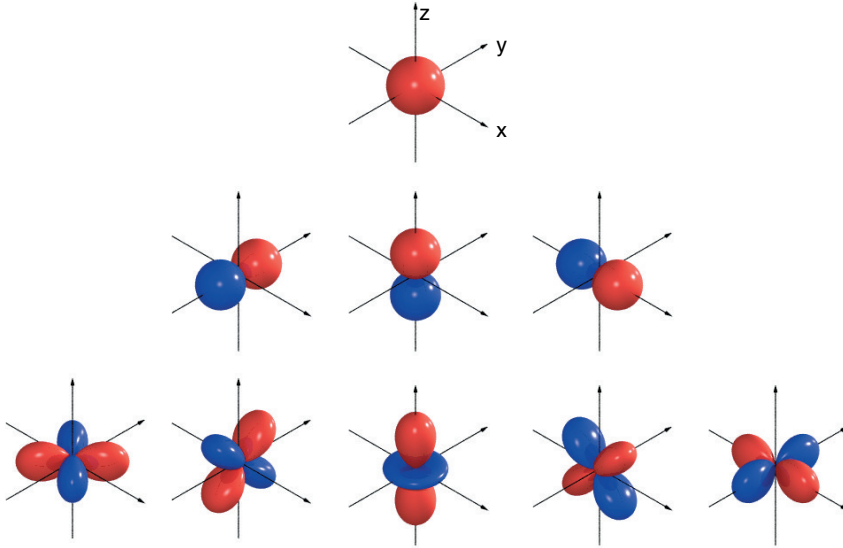
Fig. 7: The s (first row), p_y, p_z, p_x (second row), and d_{xy}, d_{yz}, $d_{3z^2-r^2}$, d_{xz}, $d_{x^2-y^2}$ (last row) real harmonics.

$\frac{i}{\sqrt{2}}(Y^l_{-m} - (-1)^m Y^l_m)$, $\quad m > 0$. Using the definitions $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$, we can express the $l = 0, 1, 2$ real harmonics (Fig. 7) as

$$s \quad\quad = y_{00} \;= Y^0_0 \quad\quad\quad\quad\quad = \sqrt{\tfrac{1}{4\pi}}$$

$$p_y \quad = y_{1-1} = \tfrac{i}{\sqrt{2}}(Y^1_{-1} + Y^1_1) = \sqrt{\tfrac{3}{4\pi}} \quad\quad y/r$$

$$p_z \quad = y_{10} \;= Y^0_2 \quad\quad\quad\quad\quad\quad = \sqrt{\tfrac{3}{4\pi}} \quad\quad z/r$$

$$p_x \quad = y_{11} \;= \tfrac{1}{\sqrt{2}}(Y^1_{-1} - Y^1_1) = \sqrt{\tfrac{3}{4\pi}} \quad\quad x/r$$

$$d_{xy} \quad = y_{2-2} = \tfrac{i}{\sqrt{2}}(Y^2_{-2} - Y^2_2) = \sqrt{\tfrac{15}{4\pi}} \quad\quad xy/r^2$$

$$d_{yz} \quad = y_{2-1} = \tfrac{i}{\sqrt{2}}(Y^2_{-1} + Y^2_1) = \sqrt{\tfrac{15}{4\pi}} \quad\quad yz/r^2$$

$$d_{3z^2-r^2} = y_{20} \;= Y^0_2 \quad\quad\quad\quad\quad = \sqrt{\tfrac{15}{4\pi}}\tfrac{1}{2\sqrt{3}} \; (3z^2 - r^2)/r^2$$

$$d_{xz} \quad = y_{21} \;= \tfrac{1}{\sqrt{2}}(Y^2_{-1} - Y^2_1) = \sqrt{\tfrac{15}{4\pi}} \quad\quad xz/r^2$$

$$d_{x^2-y^2} \;= y_{22} \;= \tfrac{1}{\sqrt{2}}(Y^2_{-2} + Y^2_2) = \sqrt{\tfrac{15}{4\pi}}\tfrac{1}{2} \; (x^2 - y^2)/r^2$$

B.3   Slater-Koster integrals

The interatomic Slater-Koster two-center integrals are defined as

$$E_{lm,l'm'} = \int d\boldsymbol{r}\, \overline{\psi}_{lm}(\boldsymbol{r}-\boldsymbol{d})V(\boldsymbol{r}-\boldsymbol{d})\psi_{l'm'}(\boldsymbol{r}).$$

They can be expressed as a function of radial integrals $V_{ll'\alpha}$, which scale with the distance d roughly as $d^{-(l+l'+1)}$ [9], and direction cosines, defined as

$$l = \boldsymbol{d}\cdot\hat{x}/d, \quad m = \boldsymbol{d}\cdot\hat{y}/d, \quad n = \boldsymbol{d}\cdot\hat{z}/d.$$

The Slater-Koster integrals for s-, p-, and d-orbitals [9] are listed below.

$E_{s,s}$	$=$	$V_{ss\sigma}$		
$E_{s,x}$	$=$	$lV_{sp\sigma}$		
$E_{x,x}$	$=$	$l^2 V_{pp\sigma}$	$+(1-l^2)V_{pp\pi}$	
$E_{x,y}$	$=$	$lm V_{pp\sigma}$	$-lm V_{pp\pi}$	
$E_{x,z}$	$=$	$ln V_{pp\sigma}$	$-ln V_{pp\pi}$	
$E_{s,xy}$	$=$	$\sqrt{3}lm V_{sd\sigma}$		
E_{s,x^2-y^2}	$=$	$\frac{1}{2}\sqrt{3}(l^2-m^2)V_{sd\sigma}$		
$E_{s,3z^2-r^2}$	$=$	$[n^2-\frac{1}{2}(l^2+m^2)]V_{sd\sigma}$		
$E_{x,xy}$	$=$	$\sqrt{3}l^2m V_{pd\sigma}$	$+m(1-2l^2)V_{pd\pi}$	
$E_{x,yz}$	$=$	$\sqrt{3}lmn V_{pd\sigma}$	$-2lmn V_{pd\pi}$	
$E_{x,zx}$	$=$	$\sqrt{3}l^2n V_{pd\sigma}$	$+n(1-2l^2)V_{pd\pi}$	
E_{x,x^2-y^2}	$=$	$\frac{\sqrt{3}}{2}l[(l^2-m^2)]V_{pd\sigma}$	$+l(1-l^2+m^2)V_{pd\pi}$	
E_{y,x^2-y^2}	$=$	$\frac{\sqrt{3}}{2}m[(l^2-m^2)]V_{pd\sigma}$	$-m(1+l^2-m^2)V_{pd\pi}$	
E_{z,x^2-y^2}	$=$	$\frac{\sqrt{3}}{2}n[(l^2-m^2)]V_{pd\sigma}$	$-n(l^2-m^2)V_{pd\pi}$	
$E_{x,3z^2-r^2}$	$=$	$l[n^2-\frac{1}{2}(l^2+m^2)]V_{pd\sigma}$	$-\sqrt{3}ln^2 V_{pd\pi}$	
$E_{y,3z^2-r^2}$	$=$	$m[n^2-\frac{1}{2}(l^2+m^2)]V_{pd\sigma}$	$-\sqrt{3}mn^2 V_{pd\pi}$	
$E_{z,3z^2-r^2}$	$=$	$n[n^2-\frac{1}{2}(l^2+m^2)]V_{pd\sigma}$	$+\sqrt{3}n(l^2+m^2)V_{pd\pi}$	
$E_{xy,xy}$	$=$	$3l^2m^2 V_{dd\sigma}$	$+(l^2+m^2-4l^2m^2)V_{dd\pi}$	$+(n^2+l^2m^2)V_{dd\delta}$
$E_{xy,yz}$	$=$	$3lm^2n V_{dd\sigma}$	$+ln(1-4m^2)V_{dd\pi}$	$+ln(m^2-1)V_{dd\delta}$
$E_{xy,zx}$	$=$	$3l^2mn V_{dd\sigma}$	$+mn(1-4l^2)V_{dd\pi}$	$+mn(l^2-1)V_{dd\delta}$
E_{xy,x^2-y^2}	$=$	$\frac{3}{2}lm(l^2-m^2)V_{dd\sigma}$	$2lm(m^2-l^2)V_{dd\pi}$	$\frac{1}{2}lm(l^2-m^2)V_{dd\delta}$
E_{yz,x^2-y^2}	$=$	$\frac{3}{2}mn(l^2-m^2)V_{dd\sigma}$	$-mn[1+2(l^2-m^2)]V_{dd\pi}$	$+mn[1+\frac{1}{2}(l^2-m^2)]V_{dd\delta}$
E_{zx,x^2-y^2}	$=$	$\frac{3}{2}nl(l^2-m^2)V_{dd\sigma}$	$+nl[1-2(l^2-m^2)]V_{dd\pi}$	$-nl[1-\frac{1}{2}(l^2-m^2)]V_{dd\delta}$
$E_{xy,3z^2-r^2}$	$=$	$\sqrt{3}lm[n^2-\frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$-2\sqrt{3}lmn^2 V_{dd\pi}$	$\frac{\sqrt{3}}{2}lm(1+n^2)V_{dd\delta}$
$E_{yz,3z^2-r^2}$	$=$	$\sqrt{3}mn[n^2-\frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$+\sqrt{3}mn(l^2+m^2-n^2)V_{dd\pi}$	$-\frac{\sqrt{3}}{2}mn(l^2+m^2)V_{dd\delta}$
$E_{zx,3z^2-r^2}$	$=$	$\sqrt{3}ln[n^2-\frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$+\sqrt{3}ln(l^2+m^2-n^2)V_{dd\pi}$	$-\frac{\sqrt{3}}{2}ln(l^2+m^2)V_{dd\delta}$
$E_{x^2-y^2,x^2-y^2}$	$=$	$\frac{3}{4}(l^2-m^2)^2 V_{dd\sigma}$	$+[l^2+m^2-(l^2-m^2)^2]V_{dd\pi}$	$+[n^2+\frac{1}{4}(l^2-m^2)^2]V_{dd\delta}$
$E_{x^2-y^2,3z^2-r^2}$	$=$	$\frac{\sqrt{3}}{2}(l^2-m^2)[n^2-\frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$+\sqrt{3}n^2(m^2-l^2)V_{dd\pi}$	$+\frac{1}{4}\sqrt{3}(1+n^2)(l^2-m^2)V_{dd\delta}$
$E_{3z^2-r^2,3z^2-r^2}$	$=$	$[n^2-\frac{1}{2}(l^2+m^2)]^2 V_{dd\sigma}$	$+3n^2(l^2+m^2)V_{dd\pi}$	$\frac{3}{4}(l^2+m^2)^2 V_{dd\delta}$

References

[1] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (eds.)
The LDA+DMFT approach to strongly correlated materials,
Reihe Modeling and Simulation, Vol. 1 (Forschungszentrum Jülich, 2011)
`http://www.cond-mat.de/events/correl11`.

[2] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (eds.)
DMFT at 25: Infinite Dimensions,
Reihe Modeling and Simulation, Vol. 4 (Forschungszentrum Jülich, 2014)
`http://www.cond-mat.de/events/correl14`.

[3] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (eds.)
DMFT: From Infinite Dimensions to Real Materials,
Reihe Modeling and Simulation, Vol. 8 (Forschungszentrum Jülich, 2018)
`http://www.cond-mat.de/events/correl18`.

[4] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, O.K. Andersen,
Phys. Rev. Lett. **92**, 176403 (2004).
E. Pavarini A. Yamasaki, J. Nuss and O.K. Andersen, New J. Phys. **7**, 188 (2005).

[5] J. Kunes: *Wannier functions and construction of model Hamiltonians*, in [1].

[6] E. Pavarini: *The LDA+DMFT Approach*, in [1].

[7] See my chapter in E. Pavarini, E. Koch, F. Anders, and M. Jarrell (eds.)
Correlated Electrons: From Models to Materials,
Reihe Modeling and Simulation, Vol. 2 (Forschungszentrum Jülich, 2012)
`http://www.cond-mat.de/events/correl12`

[8] E. Pavarini, E. Koch, and A.I. Lichtenstein, Phys. Rev. Lett. **101**, 266405 (2008).

[9] W.A. Harrison: *Electronic Structure and The Properties of Solids* (Dover, 1989).

C 1 How to write a good proposal

F. Carsughi

MLZ User Office

Forschungszentrum Jülich GmbH

Contents

1   Large-scale facilities

A large-scale facility is a research facility of a considerable dimension to conduct cutting edge basic and applied research and to promote innovation in the field. A large-scale facility needs a considerable budget for its construction, operation and for continuous upgrading the infrastructure to be at state-of-the art in its field. Initially, large-scale facilities were built and operated by a single institution for its own staff and access to such facilities was only possible through a personal agreement with the persons in charge of the available instrumentation.

Due to price increases and budget cuts, it became slowly more and more difficult for a single institution to build and operate a large-scale facility. In the middle 60s, the staff members of the research reactors DIDO and PLUTO at Harwell, UK, with Peter Egelstaff playing a leading role, started to involve the UK scientific and industrial communities in using neutrons for their own research, while at the same time the Atomic Energy Research Establishment (AERE), which operated the reactors, faced budget cuts [1]. The collaborations with the UK Universities led in 1966 to the formal establishment of the so-called "Joint programme" between AERE and the Scientific Research Council (SRC) ensuring additional financial support that allowed the reactors' operation at full capacity [2].

To my knowledge, this was the first time that a large-scale facility opened officially to external users, moving the focus of a large-scale facility from a "private" view to a user service approach. This introduced a new approach in the management of the available instrumentation: when a large-scale facility uses funds made available by external funding agencies, access to its infrastructure shall be allocated through a transparent procedure, to avoid any conflict of interest and to allow the same chance to all users. This lead to the proposal system widely used nowadays at the large-scale facilities worldwide. Unfortunately, I have not been able to find any documentary evidence when the first proposals were written for the Harwell reactors. However, I do know that when the Institute Laue-Langevin (ILL) in Grenoble was founded in 1967 by an agreement between France and Germany, it was established as a "User facility", and when it started to operate in 1973 proposals were necessary for beamtime.

The oldest documented evidence of a proposal system relates to the astronomy community and, in particular, to the European Southern Observatory (ESO) with the proposal submission deadline on September 1st, 1968, for the period N.1 [3]. Likely, the user facility concept of large-scale facilities developed in parallel in different communities, and I cannot exclude interactions between them.

Nowadays, there are large-scale facilities in many different fields, in natural sciences, as well as humanities and social sciences. As large-scale facilities are generally sponsored by public funding agencies, it is mandatory that all interested users have the same chance to access the available infrastructure without preferences or bias. Academic users, who aim to publish scientific papers, have free access. Moreover, many large-scale facilities offer also a financial support for covering their travel, accommodation and subsistence costs. On the other hand,

the access to the available instrumentation for proprietary research is possible upon the payment of an access fee.

2   Proposal system

Large-scale facilities offer the available infrastructure via a proposal system, where external users are invited to submit proposals to explain the scientific and technical details of the proposed experiment. Therefore the proposal system is addressed to non-proprietary research, namely to academic users.

The proposal system starts with the proposal preparation and follows the work through the experiment up to the scientific publication, and consists of the following steps:

a) Proposal preparation: proposals begin with a detailed analysis of the scientific problem to understand how neutrons or synchrotron X-rays may help; as these experiments are very expensive it is first worth to understand if the results can be obtained also with other conventional techniques. Once it is clear that neutrons and synchrotron X-rays provide unique results, the experimental technique shall be selected. One single technique may be available in multiple large-scale facilities and the selection of the instrument is related to many different aspects, namely technical, personal and financial ones. For example, when non-standard sample conditions are needed, the proper ancillary equipment should be available at the selected instrument. A correct definition of the requested beam time must clearly show that the proposal has been carefully prepared. Last but not least, it is always important, where applicable, to check if the experimental reports of all previous experiments have been properly submitted, as this represents an important step on the proposal preparation (experimental report s are described later at point g). Further an in depth discussion with the instrument scientist is extremely important, especially if the user is not an expert one, and sometimes even mandatory; nevertheless it is always encouraged that the proposal text is discussed with the selected instrument scientists prior submission.

b) Online submission: nowadays the proposal submission procedure is done using the facility's web pages. The requested procedures and forms are all similar and user friendly, although they differ from each other from one facility to another.

c) Proposal submission deadline: be aware of the deadline in terms of date and time and submit the proposal on time.

d) Review process: the scientific merit of each submitted proposal is assessed by at least two reviewers of the review panel and the final results are notified to the users in about 2-3 months.

e) Planning the experiment: once the proposal is accepted, please get in touch with the assigned instrument local contact for planning the experiment.

f) Experiment: each user shall register in order to access the experimental area, please do not forget this important step: many users reach the facility and can not enter the experimental area because they neglected these administrative aspects! In most of the experiments it is expected that one or more users travel to the large-scale facility to perform the experiment.

g) <u>Experimental report</u>: after the experiment many, but not all, large-scale facilities require the users to submit an experimental report with preliminary results of the experiments. This document is intended to bridge the gap between the experiment and the publication. The experimental reports are very useful during the review process as the Review Panel has the possibility to check prior experiments performed by a user.

h) <u>Publication(s)</u>: the ultimate goal of a non-proprietary investigation is to publish one or more manuscripts. The number of publications arising through use of its own instruments represents a performance indicator for a large-scale facility. With this last step, the long history of a proposal comes to an end.

2.1 Proposal types

There are many types of proposals usually accepted by large-scale facilities; the following ones are those currently used in the neutron and synchrotron radiation facilities. Proposal names may slightly vary although their meaning remains unchanged.

i) <u>Standard proposals</u>

There are usually 2 proposal rounds per year and the submitted proposals are assessed for their scientific merit by a review panel formed by external members to avoid conflict of interests and ensure a impartial review. The proposals may be accepted, rejected or put on a waiting list. Typical waiting time between the proposal submission and the experiment may vary from 3 up to 12 months.

Pros: external reviewers assess the proposals; users can perform the experiments; travel, accommodation and subsistence costs may be paid; large number of access time to allocate

Cons: strong competition due to the oversubscription of the requested instrument(s); relatively long waiting time between the proposal submission and the experiment

ii) <u>Rapid/Fast Access proposals</u>

Depending on the different facilities, programmes for a quick access, namely Rapid and/or Fast Access proposal, may slightly differ from each other, but all tend to shorten the time between the proposal submission and the experiment by a factor of 3-4 compared to the standard proposal type. They may address both short feasibility tests and also standard measurements depending on the large-scale facility.

Pros: short waiting time between the proposal submission and the experiment
Cons: usually the experiment is performed by the local contact

iii) Long Term proposals

Some facilities also accept proposals not only for one single experiment, but also for a full research programme, as may be the case for a PhD work. The proposal describes a full working plan with more experiments. In this case it is possible to proceed with the next experiment only when the step before is over and the correspondent experimental report is available and contains satisfactory results.

Pros: complex experimental plan secured in one step; reduced waiting time for the experiment
Cons: less flexibility for the experiment to perform

iv) Directors' Discretion time proposals

Proposals on very hot topics that require an urgent experiment may be assessed directly by the facilities' Director. In this case a proposal accepted for Directors' Discretion time has a very short assessment time and one can perform the experiment with a very short notice.

Pros: very short waiting time between the proposal submission and the experiment

Cons: Directors' Discretion time is allocated only in special cases

v) Internal proposals

Some large-scale facilities reserve a fraction of the available beam time for scientific experiments for their scientific staff. Proposals may be accepted either through an internal proposal round or directly by the instrument scientists.

Pros: less competition than in the case of standard proposals; more flexibility

Cons: -

An important piece of additional information requested by all the large-scale facilities is the status of the proposal. A newly submitted proposal may be a fully new proposal, a resubmission of a previously rejected proposal or a continuation of an experiment of a previously accepted proposal. In the last two cases, the review panel looks to the available prior documents during the review process.

If it is a resubmission, the user should take care to address all comments received by the Review Panel when the proposal was rejected, and if it is a continuation the user should submit the experimental report of the previous experiment, and this represents an added value to the proposal.

3   A good proposal

A proposal is a short document where the user has to describe the proposed experiment and impress the reader to get beam time. There are no easy recipes to write a good proposal as it mainly depends on the writing skills of each user, however the following indications may help you to write good proposal.

A proposal consists of many parts, such as the abstract, the technical part, the scientific and technical description and references.

The *abstract* is an efficient short summary that plays an important role, as it opens the proposals and gives the first impression to the reader. Therefore, it should be about 10-20 lines long and very carefully written.

The *technical part* is mainly formed by all technical information provided during the online submission and usually contains no text.

The *scientific and technical description* is the core of the proposal; the user should report in about two A4 pages all the necessary information. The use of a few figures is highly recommended, as they describe previous results and other important information in an efficient way, paying great attention to avoid micro- or mini-plots with no added value to the document. Font size and type described in the layout instruction shall be used. Any attempt to create additional space by using smaller fonts and figures/tables shall be avoided; moreover, duplication of information, such as the abstract for example, only reduces the available space.

As general guidelines, the scientific and technical description may include the following paragraphs:

1) <u>Introduction and state of science</u>: the state-of-the-art of the proposed research in terms of results available in the scientific literature
2) <u>Previous results</u>: the results of previous experiments and sample pre-characterization, not necessarily by using neutrons and synchrotron X-rays, performed by the user
3) <u>Aim of the proposed work</u>: the detailed description of the proposed research in terms of scientific goals and expected results; it should be clearly explained why neutrons and synchrotron X-rays are necessary for the investigation and, moreover, also why the selected experimental technique is requested
4) <u>Proposed experiment</u>: the detailed description of the proposed experiment in terms of technical configuration of the instrument, justification of the requested beam time and use of sample environment.

References includes your publication records as well as those quoted in the text. By no means should the user rely on external publications to describe the scientific and/or the technical part of the proposed experiments. All major information should be well visible in the scientific and technical description.

Last but not least, if it is not already requested on the online forms, it is worthy to mention if the proposed experiments are part of students' work. This is usually considered to be an added value during the assessment of the proposal. Many new users will quickly learn that the preparation of a proposal will become of a great help for the preparation of the experiment. The application submitted by Prof. Purcell in 1950 for a financial contribution is a very good example of a well written proposal, it came as close to proposal perfection as one could hope [4]. He wrote the two-page proposal in a very coincise and persuasive way, clearly indicating the goal, the significance, the method, the state-of-the-art and the request of the proposed investigation. And thanks to that successful application, Prof. Purcell was able to launch hydrogen-line radio astronomy, which represents now a major branch of the modern astronomy. In conclusion, the competition to grant access at large-scale facilities for scientific experiments does not differ at all from other competitive processes, such as, for example, business strategy or grant application. The art of writing efficient texts for any of these competitive processes is similar and teaching writing skills lies outside the goal of the present contribution, where guidelines and suggestions on how to improve the content of the proposal are given.

References

[1] G. Lander, private communication.
[2] M. Hutchings,
     https://neutronsources.org/files/the_end_of_the_reactors_at_harwell_____1990.pdf .
[3] ESO Observatory, Bullettin N. 4, July 1968.
[4] D.Kleppner and P.Horowitz, Physics Today, Vol. 69/1 (2016) 48.

C 2   Neutron sources and instrumentation

U. Rücker

Jülich Centre of Neutron Science

Forschungszentrum Jülich GmbH

Contents

Introduction

The neutron is a baryon without electric charge. In contrast to the charged proton, it doesn't exist as a stable free particle, but only in bound states together with protons in atomic nuclei. The free neutron has been observed first in 1930 by W. Bothe and H. Becker. They found that if energetic α particles emitted from polonium impinge on certain light elements as e.g. beryllium, boron, or lithium, an unusually penetrating radiation was produced [1].

In 1932 J. Chadwick proved that this radiation consistes of neutral particles with about the same mass as the proton. This particle had earlier been proposed as a part of atomic nuclei by E. Rutherford. Since then, it was named neutron. Chadwick has been awarded the nobel prize in physics in 1935 for the discovery of the neutron [2].

In contrast to the bound neutron, the free neutron experiences a β decay

$$n \rightarrow p^+ + e^- + \bar{v}_e \tag{1}$$

into proton, electron, and electron antineutrino with a lifetime of almost 15 minutes. This lifetime is long enough to be able to use the neutron as a probe for the investigation of other objects as e.g. a solid material.

The mass m_n of the free neutron (compare: m_p of the free proton) is

$$m_n = 1.6749 * 10^{-27} \, \text{kg} = 1.008665 \, \text{u}$$
$$(m_p = 1.6726 * 10^{-27} \, \text{kg} = 1.007277 \, \text{u}) \tag{2}$$

If the energy E of the neutron is taken into account, we always mean the kinetic energy.

$$E = \frac{1}{2} m_n \mathbf{v}^2 = \frac{\mathbf{p}^2}{2m_n} \tag{3}$$

with the velocity vector $\mathbf{v}$ and the kinetic momentum $\mathbf{p}$. A temperature equivalent T of the kinetic energy can be defined using the Boltzmann constant k_B.

$$E = k_B T \tag{4}$$

The neutron's wavelength λ is obtained from the de Broglie relation

$$\lambda = \frac{h}{m_n|\mathbf{v}|} = \frac{h}{|\mathbf{p}|} = \frac{h}{\sqrt{2m_nE}} \qquad (5)$$

and the wave vector $\mathbf{k}$ describes the propagation of the neutron wave

$$\hbar\mathbf{k} = \mathbf{p} \qquad \text{with} \qquad k = |\mathbf{k}| = \frac{2\pi}{\lambda} \quad . \qquad (6)$$

A useful unit for the wavelength λ is the Ångstrøm: 1 Å = 0.1 nm = 10^{-10} m and for the energy E it's the electron Volt: 1 eV = e * 1 V = $1.6022 * 10^{-19}$ J with the electron's charge e.

Some useful conversions are

$$\lambda\,[\text{Å}] = \frac{3956}{v\,[\text{m/s}]} \qquad \text{and} \qquad (7)$$

$$E\,[\text{meV}] = \frac{81.8}{\lambda^2\,[\text{Å}]} \quad . \qquad (8)$$

The following table 1 shows the relations between energy, velocity, wavelength, and temperature of the neutron for a few examples which are commonly used. From these numbers, it immediately becomes obvious that neutrons with a temperature around room temperature have a wavelength comparable to interatomic distances, so they are useful to investigate the atomic order of condensed matter.

Energy E	Velocity v	Wavelength λ	Temperature T
1 MeV	$1.38 * 10^7$ m/s	$2.86 * 10^{-4}$ Å	$1.16 * 10^{10}$ K
1 eV	13800 m/s	0.286 Å	11600 K
230 meV	6600 m/s	0.6 Å	2700 K
82 meV	4000 m/s	1 Å	950 K
26 meV	2230 m/s	1.77 Å	300 K
10 meV	1380 m/s	2.86 Å	116 K
1.7 meV	570 m/s	6.9 Å	20 K

Table 1: Relation between energy, velocity, wavelength, and temperature of the neutron.

1   Neutron production

Stable neutrons only exist in a bound state in atomic nuclei. A nuclear reaction is always needed to release a neutron from a nucleus. In most cases, energy is needed to release a free neutron, in all cases an activation barrier has to be overcome. Today, many reaction paths are known that yield free neutrons which are presented in the following.

1.1   Radioactive neutron sources: (α,n) reaction

The production of neutrons with α radiation has historically led to the discovery of the neutron [1], [2]. The reaction

$$^9\mathrm{Be} + \alpha \rightarrow {}^{13}\mathrm{C}^* \rightarrow {}^{12}\mathrm{C} + \mathrm{n} + \gamma \tag{9}$$

is still used today for small neutron sources used for detector calibration or constant activation purposes, e.g. in smoke detectors. Today's most common alpha emitter used in neutron sources is ^{241}Am (432 years half-life), artificially produced by Pu breeding in a nuclear reactor. Also ^{238}Pu or natural ^{226}Ra can be used together with Be to produce small neutron source.

Fig. 1:   AmBe neutron source used for detector calibration. The source's α activity is 350 MBq, yielding a neutron production of 20000 n/s. The measured dose rate emitted by this source is about 10 µSv/h

The neutron yield of a AmBe neutron source is about $6 * 10^4$ neutrons per GBq α activity. These neutron sources are hermetically sealed, so that the material can be handles safely and the α radiation is properly shielded. The strongly penetrating neutron and γ radiations are emitted.

Fig. 1 shows an AmBe neutron source used in our laboratory to calibrate neutron detectors for the radiation safety interlock system. The white "pig" on the picture is a neutron detector encapsulated in a polyethylene thermal moderator making it possible to measure the intensity of thermal and fast neutron fields (see section 2.1) for radiation safety purposes.

1.2   Photoneutrons: (γ,n) reactions

The absorption of highly energetic γ photons can be used to excite a lot of nuclei to release a neutron. This works as well with light nuclei as Be as with heavy nuclei as W or Pb. A combination of ^{123}Sb (activated to ^{124}Sb as γ source) and Be (as neutron source) can be activated in a the neutron field of a nuclear reactor and emits monoenergetic neutrons with an energie of 24 keV [3]. The half-life of this source is 60 days. Because of its low neutron energy it is used to keep a residual neutron field in a nuclear reactor during shutdown.

Another application of (γ,n) reactions is a compact accelerator based neutron source based on an electron accelerator and a thick heavy metal target. With an electron beam of some 100 MeV energy impinging on the target, a lot of high energetic bremsstrahlung is produced whose energy is sufficient to release neutrons from the same target material. An example of this type of neutron source is the HUNS source at Hokkaido university, Japan [4]. The drawback of this kind of accelerator driven neutron source is the fact that the γ radiation produced by the electrons has an energy such high that is difficult to shield and that also in the shielding material neutrons are released by (γ,n) processes.

1.3   (p,n) and (d,n) reactions

Accelerated protons or deuterons impinging on a metal target can be used to induce neutron production in a very efficient manner. Light ions can be accelerated in an electrostatic or a RF linear accelerator to energies in the range of a few MeV to a few 10 MeV. For most target materials the first neutron production channels open well below 20 MeV activation energy with additional contributions up to 100 MeV and higher. In case of deuteron beams, also the stripping of the deuteron's neutron increases the neutron production rate. Fig. 2 shows the energy dependence of the probability of different neutron production reaction channels for deuterons impinging on Be. For the deuteron stripping, there is no energy threshold, the neutron emission from Be starts at about 2 MeV deuteron energy [5].

Due to the flexibility of the light ion accelerators, this kind of neutron source can be operated in pulsed mode at high intensities. The low energy protons or deuterons are stopped after a short penetration length in the target material. Therefore, the volume where primary neutrons are produced is small, so that the neutrons can efficiently be coupled into a thermal moderator. Based on this principle is the design of the Compact Accelerator based Neutron Sources (CANS), as they are presented e.g. in lecture **C7** describing the HBS project for a modern CANS that we would like to build in Jülich.
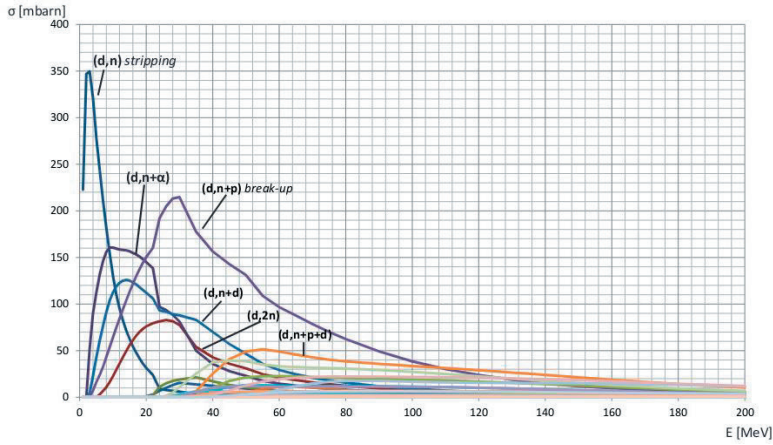
Fig. 2: Calculated energy dependence of the cross sections of different neutron emitting reactions for deuterons impinging on Be based on the TENDL nuclear data library [5].

1.4   Nuclear fusion: d+d or d+t reactions

The neutron production by nuclear fusion of deuterium (d or ^{2}H) and tritium (t or ^{3}H) can be induced at a low ion energy (below 100 keV) that can easily be achieved with a small accelerator. The two reactions available yield monochromatic neutrons:

$$^2\text{H} + {}^3\text{H} \rightarrow {}^4\text{He} + \text{n} \quad (14\,\text{MeV}) \tag{10}$$

$$^2\text{H} + {}^2\text{H} \rightarrow {}^3\text{He} + \text{n} \quad (2.5\,\text{MeV}) \tag{11}$$

Fig. 3: A fusion neutron generator GENIE 16 GT commercially available from EADS Sodern. The tritium loaded metal hydride neutron production target is located 15 cm from the end of the long cylinder.

Technologically demanding is the fact that the target material is a gas. This is solved by using a metallic hydride as target. From time to time the target is burnt up and must be refreshed.

Fig. 3 shows a commercial neutron generator based on the reaction (10). It can produce up to 10^8 n/s in continuous or pulsed operation.

1.5 Nuclear fission reactors

Nuclear fission reactors are high-intensity neutron sources available since the 1940s. If ^{235}U is irradiated with thermal neutrons, the nucleus absorbs the neutron and breaks into typically two fission fragments. In this process neutrons are emitted with an average of 2.3 neutrons per fission process.

In a nuclear reactor, these fission neutrons are used to sustain a chain reaction. A sufficient amount of ^{235}U (more than the critical mass) together with a suitable thermal moderator (see section 2.1) that slows down the originally fast fission neutrons makes it possible that at least 1 of the 2.3 fission neutrons can again be absorbed by a ^{235}U nucleus. The remaining neutrons either escape from the core region (e.g. into a neutron beamline towards a neutron scattering experiment), or they are absorbed by the moderator or any structure material. The excess neutrons need to be absorbed by a reactor control system that can tune the chain reaction to a stable level or a slight deviation for an increase or decrease of the reaction speed, i.e. the reactor power.
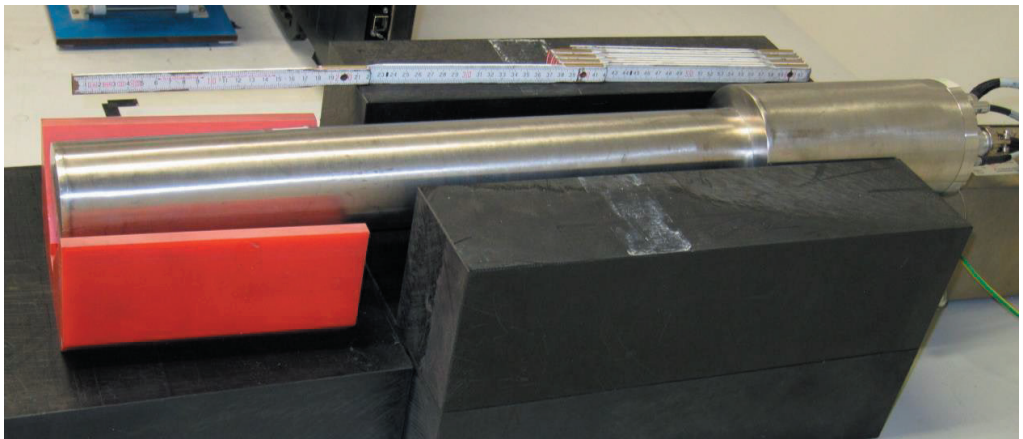
For the reactor control it is important that a small fraction of the neutrons (0.65% in the case of ^{235}U are not emitted immediately after the fission, but about 10 to 20 s later. If the chain reaction is leveled so that the prompt neutrons are not sufficient to sustain the chain reaction, the delayed neutrons can be used to control a slow change in reactor power. The reactor must never be operated in a state where the prompt neutrons alone sustain the chain reaction ("prompt critical reactor") [6]. In this case, the reactor power rises exponentially in an uncontrolable manner. This happened for example at the Chernobyl accident.

Fig. 4 shows a model of the fuel element of the FRM II research reactor in Garching near Munich. The reactor core consists of this single fuel element loaded with highly enriched ^{235}U, a light water cooling circuit passing vertically through the fuel element and a heavy water moderator and reflector that brings back thermal neutrons to the core to sustain the chain reaction and that feeds tangential beam tubes to extract neutrons from the reactor to use them in neutron beam lines outside the reactor vessel. The reactor power is controlled by a neutron absorber rod located in the centre of the fuel element.

Historically and still today, nuclear reactors are the main source of high-intensity neutron beams for research applications. As most research reactors have been built in the years 1950 – 1970, they are aging and more and more are shut down. Their role as a supply of neutron beams for research is gradually taken over by CANS and by spallation neutron sources.

Fig. 4: A model of the fuel element of the FRM II research reactor. The inclined plates
contain the fuel, the cooling water flows in between the fuel plates. The central
channel is the position of the control rod. The core of FRM-2 is the most compact
^{235}U reactor core ever built.
(The copyright for these photos is held by W. Schürmann, TU Munich)

1.6   Spallation neutron sources

Spallation is the interaction of heavy nuclei with high-energy protons, i.e. protons with energy
in the GeV range. In the collision of a nucleus with a proton at such high energy, the nucleus
is completely disintegrated. At energies above 150 MeV the de Broglie wavelength of the
proton is small enough that the proton interacts with the individual nucleons in the target
nucleus. The nucleus is highly excited, up to 30 neutrons are evaporated from a single nucleus
and several small nuclei are formed [7].

Spallation is most energy-efficient process of neutron production, but a lot of the energy is
transferred onto the neutrons, so that neutrons with energies up to 500 MeV need a high effort
to design a suitable shielding for the operation of such a source. Due to the large stopping
length of the high energy protons, the neutron production zone is elongated to several 10 cm.

Today's strongest pulsed neutron sources are spallation sources. The European Spallation
Source which will be the strongest neutron source worldwide is currently being built in Lund
(Sweden) and will be presented in lecture **C6**.

2 Manipulation of travelling neutrons

2.1 Moderation

Primary free neutrons that are produced in a nuclear reaction typically have a energy of a few MeV, in case of spallation it might be a lot higher. To make these neutrons useful to investigate atomic matter, this energy has to be reduced below 1 eV according to table 1. The useful neutron energies correspond to about room temperature, for crystallography maybe higher, for investigations of complex matter (e.g. macromolecules or layered structures) even lower. This neutron energy can be reached by bringing the neutron field into thermal equilibrium with something that has the proper temperature and that interacts strongly with the neutrons.

One important candidate to do so is hydrogen in any compound that has a high density. The hydrogen atom has approximately the same weight as the neutron, so collisions between a neutron and a hydrogen nucleus (i.e. a proton) can optimally transfer momentum from the neutron to the proton. In addition, the inelastic scattering cross section that describes the probability of an energy transfer between hydrogen nuclei and neutrons is high.

So, for neutron moderation to thermal temperatures water H_2O or heavy water D_2O are ideal. H_2O has the advantage that it is easily available and does not produce any radioactive waste, but the disadvantage of a non-negligible neutron absorption. D_2O has the advantage that it hardly absorbs neutrons, but if it does, radioactive tritium is produced. In addition, the interaction between neutrons and D_2O is weaker compared to H_2O, so that the moderator volume needs to be bigger and the neutron field is diluted. If the thermal load is low, also solid polyethylene is a good thermal neutron moderator, but it cannot be cooled as efficiently as liquid water. Another useful thermal moderator is Be which is solid but with a high thermal conductivity and has almost no neutron absorption. A thermal moderator efficiently delivers neutrons in the wavelength range between 0.7 and 2.5 Å.

If neutrons with longer wavelengths are desired, a "cold neutron source" is used to moderate thermal neutrons to even lower temperatures. Liquid H_2, liquid D_2, solid methane CH_4 or solid mesitylene C_9H_{12} are the most frequently used substances in a cold neutron source. A cold neutron source efficiently delivers neutrons in the wavelength range between 2 and 8 Å. Depending on the realization, sometimes neutrons with wavelengths up to 20 Å can be used for special applications.

In some research reactors specialized beamlines for crystallography and high-energy spectroscopy are fed by a hot neutron source. A hot neutron source is a graphite moderator heated (by the γ radiation close to the fuel element) to temperatures between 2500 and 3000 K. This moderator efficiently delivers neutrons with wavelengths down to 0.2 Å corresponding to an energy up to 2 eV.

2.2   Monochromatization

In a neutron scattering experiment it is in most cases necessary to know the neutron's wavelength or energy when it interacts with the sample. There are three methods of monochromatization commonly used:

2.2.1   Time-Of-Flight (TOF)

The TOF method uses the energy dependent velocity of the neutrons (eq. (5) and (7)). The neutron beam is structured in time by a rotating chopper that is transparent only for a short fraction or it is emitted from a pulsed source. When the neutron beam travels the distance between source / chopper to the detection system, the neutrons with shorter wavelength arrive earlier at the detector. With the control of the timing between source / chopper and the detector system, the wavelength of every neutron can be reconstructed from the arrival time.

2.2.2   Velocity selector

A velocity selector is a rotating turbine wheel with neutron absorbing channel walls, as can be seen in fig. 5. The neutron beam impinges in a direction parallel to the turbine axis. If the speed of the neutrons matches the rotational motion of the channel, the neutrons of this speed can pass the velocity selector. This technique is used for instruments that need a monochromatic beam with moderate wavelength resolution (10-20%).

Fig. 5:   Turbine wheel of the velocity selector used at KWS-3. The selector blades consist of a MgLi alloy with neutron absorbing ^{6}Li.

2.2.3   Monochromator crystals

Bragg reflections on a single crystal happen only, if wavelength and angle of incidence on a suitable lattice plane match. Therefore, a neutron beam with well-defined direction can be monochromatized by Bragg reflection from a monochromator crystal. Depending on the quality of the crystal, a wavelength resolution between 0.05% and 1.5% can be achieved. Sometimes "too good" crystals are treated mechanically to disturb the perfetion of the lattice to be able to reflect a wider wavelength band.

Fig. 6 shows a double focussing Cu monochromator. It is equipped with 105 single crystals that have been hammered to increase the reflectivity. This monochromator can be focused vertically to be able to illuminate small samples more strongly and it can be focused horizontally to be able to transport a larger wavelength band (due to different Bragg angles of the individual columns of the monochromator) to the sample on the expense of a higher beam divergence.
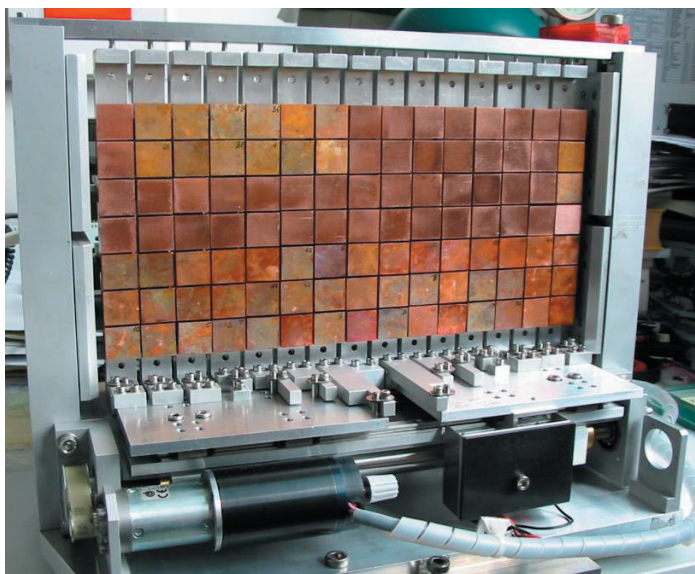

Fig. 6:   Double focussing monochromator equipped with 105 Cu single crystals. The monochromator can be bent individually in vertical and horizontal direction to be able to adapt the focussing of a large incident beam onto samples with variable size.

2.3   Neutron transport

2.3.1 Neutron guides

Originally, thermal or cold neutrons are emitted isotropically from the moderator. Instruments directly located at a beamtube of a nuclear reactor can use the entire solid angle that is illuminated through the opening of the reactor shielding. But, as space around the shielding is very limited, a way to bring neutrons efficiently to instruments at a larger distance is necessary.

Neutron guides based on the principle of total reflection (see lecture **D2**) are used to transport neutron beams with a limited divergence but with low losses over distances of up to 200 m. The walls of neutron guides are polished glass plates with a metal coating. A simple Ni coating offers total reflection up to the critical angle

$$\theta_c = \frac{0.1°}{Å} * \lambda \tag{12}$$

for neutrons with the wavelength λ. These neutrons are transported without losses. Today's supermirror technology has achieved to increase the reflection angles up to m=6 times θ_c with reflectivities between 95% (for m=2) down to 60% (for the highest angles with m=6). From eq. (12) it is obvious that neutron guides are more efficient for cold neutrons with long wavelengths than for thermal neutrons. Therefore, most cold neutron instruments are located at increased distance from the reactor surface while only thermal instruments are located directly at the beamtubes.

Fig. 7 shows the neutron guides serving all cold neutron instruments of the neutron guide hall at the former research reactor in Jülich. Straight and slightly curved guides allow to supply neutrons to instruments that have sufficient space to be operated independently.

Fig. 7:   Neutron guides serving the cold neutron instruments at the former FRJ-2 research reactor in Jülich.

2.3.2   Collimation

To be able to define the divergence of a neutron beam, a collimating device is needed. This device typically absorbs all neutrons that travel in any unwanted direction. Either this is realized by a pair of adjustable slits (typically 4 electronically movable blades defining the left, right, top, and bottom edge of the opening) at a distance of typically a few metres, or by a collimator which is an array of tunnels for neutrons with absorbing walls, so that only the neutrons that can pass straight through one of the tunnels are transmitted.

2.3.3   Focussing

The neutron flux at the sample (or detector) position of a neutron instrument can be increased by focussing the beam onto the sample position on the expense of an increased divergence, if this can be tolerated by the experimental conditions.
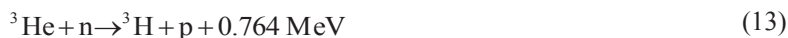
In sec. 2.2.3 already a focussing crystal monochromator has been introduced. A second option for instrument at a sufficient distance from the source is an elliptically shaped neutron guide. This kind of neutron guide can transport a larger divergence of neutron beam from one focal point (where the source is located) to the second focal point where the beam is focused on. Ideally (with a point source), every neutron would be reflected only once along the entire neutron guide. Even with the extended sources available the elliptical neutron guide is a good technique to deliver a beam with higher divergence to a distant instrument with small sample size, because the number of reflections on the walls is strongly reduced compared to straight neutron guides.

For very cold neutrons ($\lambda > 10$ Å) refractive lenses are feasible to achieve highest spatial resolution. These lenses are made of MgF_2. Due to the low index of refraction typically several 10 lenses are used in series, so that the transmission losses and the shape errors become an important parameter.

2.4   Neutron detectors

Neutrons cannot be detected electronically because they carry no charge. Again, a nuclear reaction is necessary to produce charges that can be treated with electronic devices. The capture of the neutron in a suitable converter material releases charged secondary particles with high energy which are able to ionize detector material [8].

The most important nuclear capture reactions for neutron detection are

$$^3\mathrm{He} + \mathrm{n} \rightarrow {}^3\mathrm{H} + \mathrm{p} + 0.764\,\mathrm{MeV} \tag{13}$$

$$^6\mathrm{Li} + \mathrm{n} \rightarrow {}^3\mathrm{H} + \alpha + 4.78\,\mathrm{MeV} \tag{14}$$

$$^{10}\mathrm{B} + \mathrm{n} \rightarrow {}^7\mathrm{Li} + \alpha + 2.31\,\mathrm{MeV} + \gamma\,(0.48\,\mathrm{MeV}) \tag{15}$$

Gas detectors use the ionization of gas by particle radiation within an electrical field. The ions and electrons are moving to the electrodes and induce a charge signal that can be detected by electronics. Typical detector gases for the detection of thermal or slow neutrons are $^3\mathrm{He}$ and $^{10}\mathrm{BF_3}$.

In Scintillation detectors, the neutron converter is embedded in a transparent solid matrix and the energy of the secondary particles is converted into excited states of the matrix that decay with the emission of photons. These photons in a second step can be converted into electronic signals in a photomultiplier. A typical neutron scintillator material is glass with embedded ^{6}LiF.

Both detector techniques can be used to build a position sensitive detector, where the charges are detected in a spatially resolved way. 1D or 2D spatial resolution can be achieved down to 50 μm.

3   Neutron instrumentation

For most of the neutron scattering experiments, the experimentalist is interested in the relation between the wave vectors $\mathbf{k}_i$ and $\mathbf{k}_f$ of the neutron before and after the encounter with the sample. The job of the instrument is now to define $\mathbf{k}_i$ precisely enough and to detect enough information about the scattered neutrons to be able to evaluate $\mathbf{k}_f$.

For the definition of the incident neutron state, the angles of incidence α_i (in the scattering plane) and φ_i (perpendicular to the scattering plane), the neutron wavelength λ_i or energy E_i, and the place of the interaction (x,y,z) need to be known. After the interaction α_f, φ_f and E_f need to be measured.

One important exception is spin-echo spectroscopy, where the precessing neutron spin is used as an internal clock for every single neutron, so that E_f-E_i can be measured without precise knowledge of E_i.

As the intensity of a neutron source is limited, it is not sensible to determine all parameters with high precision. Depending on the information that the experimentalist would like to gain, it is necessary to find out which value needs to be defined with what precision. The relaxation of some parameters always leads to important intensity gains. For this reason, a lot of different instrument types have been developed to have suitable instruments for different scientific questions.

3.1   Elastic neutron scattering instruments

The neutrons that probe the structure of matter do not change their energy during the interaction with the sample. As the probability for elastic neutron scattering, i.e. $E_f = E_i$, is always much higher than the probability for inelastic scattering, the detection of E_f is omitted in all elastic scattering instruments.

3.1.1   Single crystal diffractometer

A single crystal diffractometer is used to determine the crystal structure of a high-quality single crystal in the centre of the Eulerian cradle (3 motorized drives for the angles ω, χ, and φ) that makes it possible to align the sample in every orientation with respect to the incoming beam. Together with the 2θ motor that defines the angle between detector and incoming beam, all mutual orientations can be realized.

Fig. 8 shows a standard single crystal diffractometer as it is realized at many places. It requires a precise definition of the wavelength λ and a proper adjustment of the centre of the Eulerian cradle to the beam. The divergence of the angles α_i and φ_i of the incoming beam can be relaxed because the selection rules in the single crystal diffraction induce additional resolution.
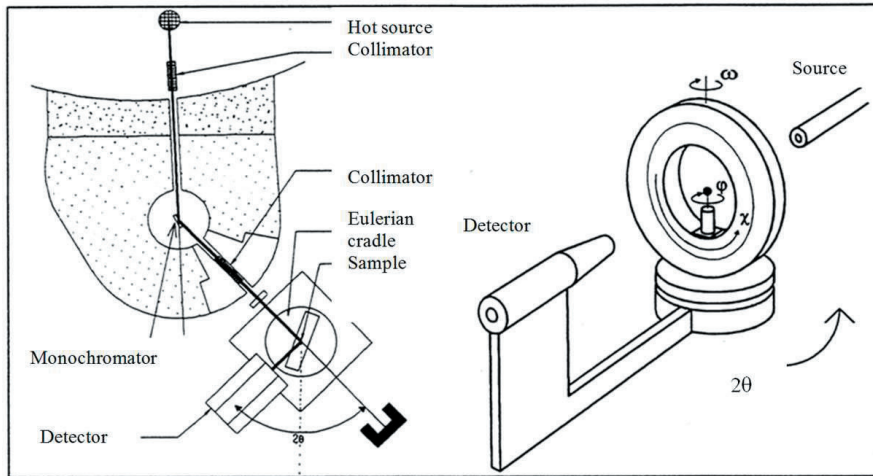
Fig. 8: Principle components of a constant wavelength 4-circle single crystal diffractometer (from [4]).

3.1.2   TOF reflectometer

A TOF reflectometer at a pulsed source is used to determine thicknesses, densities and composition of the individual layers in a layered structure. As the sample is laterally homogeneous, no in-plane resolution is necessary.
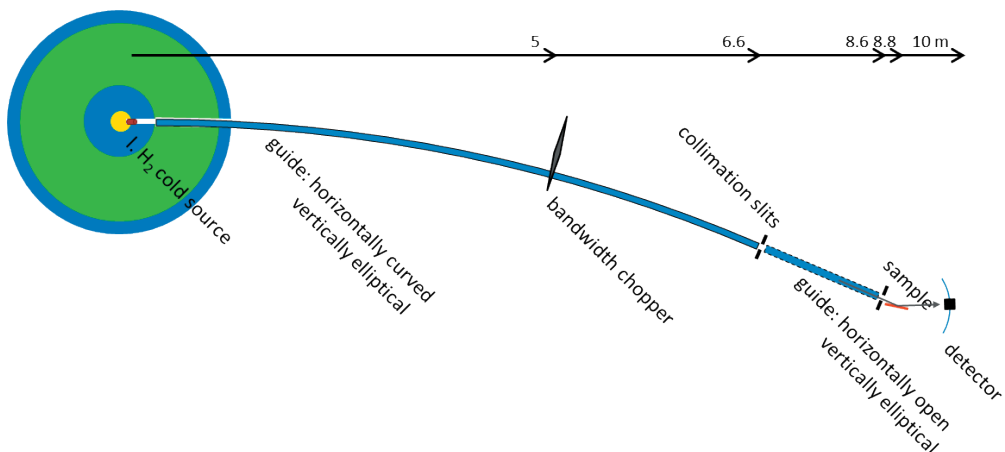

Fig. 9: Generic reflectometer to be installed at a pulsed compact neutron source [10].

Fig. 9 shows a design suitable for a reflectometer at a pulsed source with 48 Hz repetition rate and 4% duty cycle [10]. The neutron pulse emitted from the liquid H_2 cold moderator is dispersed over the 10 m flight distance so that it has a $\Delta\lambda = 0.34$ Å wavelength resolution at the detector position, which is below the necessary resolution limit of 20% for all wavelengths used. A bandwidth chopper cuts the wavelength band to 7.5 Å width (e.g. from 2 Å to 9.5 Å), so that no overlap between the fastest neutrons of one pulse and the slowest neutrons of the previous pulse occurs.

In horizontal direction, the incident angle α_i is defined by 2 collimation slits at 2 m distance from each other. In vertical direction, an elliptical neutron guide focuses the beam onto the sample with highest divergence available, because in this direction no resolution is required. The collimation slits are widely open in the vertical direction. Due to this focussing the intensity at the sample position can be increased by more than a factor 3 compared to a neutron guide with fixed height.

3.2 Inelastic neutron scattering instruments

3.2.1 TOF spectrometer

A Time-Of-Flight spectrometer is used to determine the energy transfer between neutron and sample. This is used to measure e.g. phonon excitation energies is crystalline matter. Fig. 10 shows the main components of a TOF spectrometer that can be installed at a pused neutron source. Typically, for intensity reasons, the angular resolution is relaxed, so that focussing neutron optics can be used. A big spherical detector bank covers most of the solid angle around the sample so that the neutrons scattered to almost all directions can be detected simultaneously.

The white neutron pulses emitted from the source are dispersed by Time-Of-Flight on the flight path towards the sample. Somewhere along this flight path a narrow bandwidth filter selects the energy of the incident neutrons. This may be a bandwidth chopper with a narrow opening or a focussing crystal monochromator. (Of course, in this case the instrument is not straight!) The chopper immediately in front of the sample defines the point of time when the neutrons with well-known energy E_i reach the sample. The neutrons scattered by the sample then travel the flight path towards the detectors. The time they need for this path then allows to determine the energy E_f of the neutron after the interaction with the sample and with that the energy transfer at the sample. The position of the detected neutrons determines α_f and φ_f, so that $\mathbf{k}_f$ is fully known.
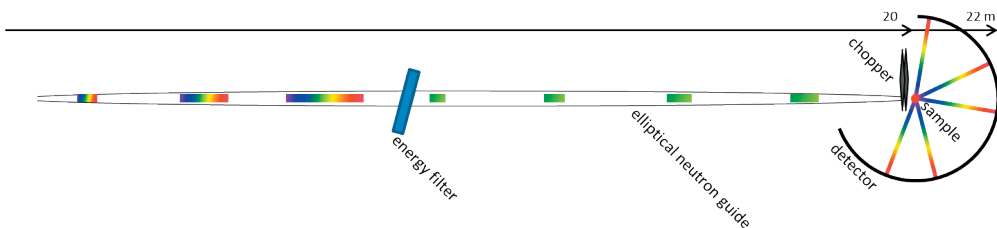

Fig. 10: Main components of a TOF spectrometer operated at a pulsed neutron source.

3.2.2   Triple axis spectrometer

In a triple axis spectrometer, the parameters desribing the incident and the scattered neutrons are determined one after the other. Fig. 11 shows a picture of the IN12 spectrometer operated in a neutron guide hall at the ILL. As a first step, the energy E_i of the incident neutrons is defined by the monochromator crystal (1st axis). The first collimator defines the angle of incidence α_i of the neutron beam arriving at the sample position. The sample can be oriented freely on the sample table, the 2nd axis is the angle α_f, at which the neutrons leaving the sample towards the analyzer are observed. The second collimator ensures that only the neutrons within the desired angular resolution can travel to the analyzer crystal. The rotation of the analyzer against the beam coming from the sample (3rd axis) defines the energy E_f of the neutrons that are finally allowed to go to the detector.
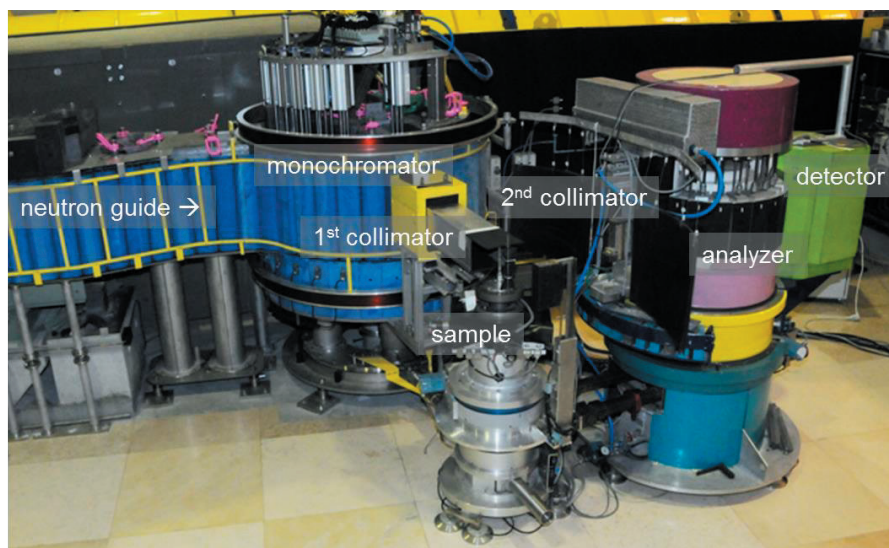
Fig. 11: Triple axis spectrometer IN12 of the JCNS operated at the ILL.

In contrast to the TOF spectrometer, the triple axis spectrometer can only measure at one single $\mathbf{k}_f$ vector at a time. This makes the measurement much slower compared to the TOF spectrometer, but the unique opportunity to define the resolution of every parameter individually allows a measurement strategy that is very much adapted to the problem under investigation. Of course, this requires a good basic knowledge of the physics that is expected in the sample.

3.2.3   Backscattering spectrometer

The backscattering spectrometer is the highest resolution spectrometer that uses the classical approach to determine E_i and E_f individually. In this lecture, I present only the secondary spectrometer measuring $\mathbf{k}_f$ of the neutrons scattered from the sample.

Fig. 12 shows the energy analyser of such a backscattering spectrometer. The energy of the scattered neutrons is determined by backscattering from Si (111) single crystals, the most perfect single crystals available. For this reason, the resolution in E_f is as good as 1.5 µeV. The amount of neutrons within this very narrow energy window is very low, so to achieve a reasonable count rate, the angular resolution of the instrument is extremely relaxed. Each of the analyser plates, which are segments of a sphere and which are covered with Si single crystal wafers bent to the spherical shape, covers a range of scattering angles between 5 and 20° and focuses the reflected neutrons on a single detector. So the entire area you can see in Fig. 12 is collected within 7 angular channels to be able to achieve a sufficient count rate well above the background level in every detector.

Fig. 12: Energy analyser bank of the backscattering spectrometer at FRJ-2. Every analyser section is a part of a sphere and is covered with perfectly aligned Si single crystals.

References

[1] https://en.wikipedia.org/wiki/Neutron

[2] https://www.nobelprize.org/prizes/physics/1935/chadwick/facts/

[3] https://www.nrc.gov/docs/ML1122/ML11229A704.pdf

[4] http://phi.phys.nagoya-u.ac.jp/JCANS/jcans.html

[5] U. Rücker et al., The Jülich high-brilliance neutron source project, Eur. Phys. J. Plus **131** (2016), 19

[6] M. Borlein, Kerntechnik, Vogel Buchverlag, Erlangen (2009), chapter 5

[7] D. Filges, F. Goldenbaum, Handbook of Spallation Research, Wiley-VCH, Weinheim (2009), chapter 1.3

[8] G. Kemmerling, lecture C5 in: 43rd IFF Spring School: Scattering Methods for Condensed Matter research, Forschungszentrum Jülich, Series Key Technologies, Vol. **33** (2012)

[9] G. Roth, lecture 4 in: Neutron Scattering, Forschungszentrum Jülich, Series Key Technologies, Vol. **172** (2018)

[10] E. Mauerhofer et al., Conceptual Design Report NOVA ERA, Forschungszentrum Jülich, Series General, Vol. **7** (2017)

C 3   Synchrotron radiation sources and instrumentation

U. Klemradt

II. Physikalisches Institut B

RWTH Aachen University

Contents

1   Introduction

Discharge tubes were made possible in the second half of the 19th century by advances in vacuum technology, leading to the discovery of both the electron in 1897 by J.J. Thomson and X-rays in 1895 by W.C. Röntgen . For the first two decades, the dominant scientific use of X-rays was in the field of medical imaging. However, the first successful recording of a diffraction pattern in 1912 by Friedrich and Knipping and the subsequently developed theory by von Laue and the Braggs opened up experimental access to the atomic scale of materials. The experiment proved not only that X-rays are capable of interference and thus possess wave characteristics, but also that crystals are periodic structures on the atomic scale, as long suspected by chemists. The now famous 'Laue experiment' is frequently considered to be the starting point of modern solid state physics and chemistry, linking for the first time (macroscopic) properties to the (atomic) structure. The subsequent progress in structure determination is probably best illustrated by the advent of protein crystallography (see talk E8), where the complexity has been increased to more than 10,000 atoms per unit cell, achieved nowadays routinely at large storage ring facilities usually referred to as synchrotrons.

However, synchrotron radiation can be used for far more than crystallography. Its spectrum extends from the infrared to the γ-region of electromagnetic waves, and typical applications include scattering, imaging and spectroscopy, performed on samples from practically all fields of science. In this overview chapter we will discuss elementary properties of synchrotron radiation and its generation in storage ring facilities, as well as the basic layout of beamlines and their instrumentation. We will use a 'downstream' approach, following the photon from the source via the mononchromator to the sample. The focus will be on hard X-rays, although reference to soft X-rays is occasionally made.

2   Properties of Synchrotron Radiation

2.1   Fundamental aspects and overview

X-rays are generated by transitions in the atomic shell structure, or by sufficiently accelerated charged particles. We start with a look at X-ray generation in tubes. Despite their enormous technological evolution over the decades, the basic design still implies a) high X-ray intensity only at characteristic wavelengths given by the target material, b) rather isotropic photon emission, c) a photon flux limited by the heat tolerance of the anode, and d) a spectrum arising from multiple scattering of electrons inside the target, which cannot be calculated analytically.

By contrast, synchrotron radiation is produced by highly energetic, charged particles deflected by magnetic fields. Synchrotron radiation occurs naturally in space and was on earth at first observed at circular accelerators built for particle physics. It was initially an unwanted by-product, since it limited the maximum energy of the accelerator. However, soon after the first observation in 1947 it was realized that the emitted light had extraordinary properties useful in atomic, molecular and condensed matter physics. The outstanding features are largely a direct consequence of the following two aspects:

1. the charged particles (typically electrons) move at highly relativistic speeds,

2. the particles are accelerated in a vacuum.

It is instructive to compare the generation of X-rays in a tube and in a synchrotron with the above list a) - d) in mind. The relativistic speed brings in the Lorentz transformation naturally, with remarkable results for the experimentalists and their samples being in the laboratory frame.

First, the usual dipole radiation pattern in the moving electron frame is observed in Lorentz-contracted form in the laboratory frame. Together with an additional boost forward, this results in a strong, natural collimation of the emitted light cone in the laboratory frame, sometimes compared to a sweeping search light (cf. Fig. 1). The natural collimation contributes significantly to the enormous photon flux available at synchrotron beamlines, since a well-defined angle of incidence - typically required in X-ray experiments - is usually realized by slits. By comparison, simple collimator systems discard most of the photons generated by an isotropic emitter as in b).

Second, the intensity radiated is not restricted to the fundamental frequency of revolution, which is in the MHz range. The relativistic velocities together with the point-like charges give rise to harmonics up to very high order. Since the individual harmonic lines are smeared out, a continuous, broad spectrum results which extends from the infrared into the X-ray range. Therefore, a synchrotron bending magnet radiates a broad, intense spectrum, in contrast to point a) above. This allowed to develop special synchrotron-based techniques which require tunable X-ray wavelengths, for example absorption spectroscopy (cf. talk F2).

Third, the acceleration of relativistic electrons by radio frequency cavities in a vacuum allows to form electron bunches of very small size (down to the micron range), resulting in an exceedingly small source size that comes very close to the ideal of a point source. This is of great importance for many experimental techniques and is described by the physical quantity 'brilliance' explained in section 2.3.

Forth, it is intuitively clear that the acceleration of charges in a vacuum is a well-defined process compared to point d) above. It leads to a clean spectrum than can be treated theoretically in full detail [2, 3]. From a practical point of view, this makes it much easier to design more elaborate equipment, like wigglers and undulators (see section 2.4). Also the fundamental SASE (self amplified stimulated emission) process in free electron lasers (cf. talk C5), which is based on undulators, was predicted quantitatively before the first machines were built.
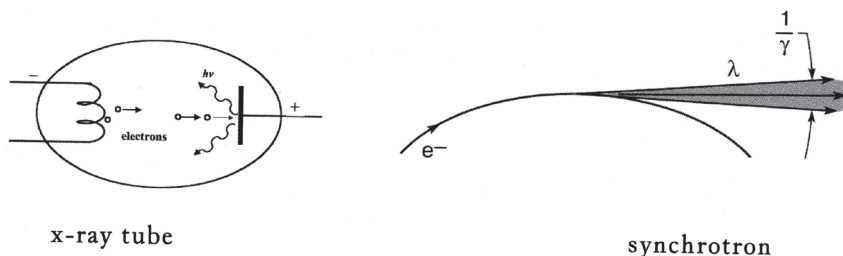

Fig. 1: X-rays generated in a wide solid angle by a tube compared to the narrow emission angle of synchrotron radiation [1]. The angle $1/\gamma$ represents a characteristic scale for the emission in the orbit plane as well as perpendicular to it.

However, it should be mentioned that the access to synchrotron radiation is somewhat restricted, since it usually involves the submission of a proposal, with several months easily passing before an experiment can be performed (cf. talk C1). Therefore, the greatest advantage of X-ray tubes today lies in the availability of a laboratory source.

2.2   Continuous spectral distribution

Of foremost importance is the energy E of the electrons circulating in the storage ring, typical values range from 0.7 GeV (SSLS, Singapore) to 8 GeV (SPring 8, Japan). The energy of the electrons then is almost entirely kinetic energy, since their rest mass m_0 is equivalent to only 511 keV, so their speed is only a few m/s below the speed of light according to

$$E = m\,c^2 = \gamma\,m_0\,c^2 \tag{1}$$

with

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}. \tag{2}$$

Typical values of γ are several thousand for a synchrotron in the GeV range, for example $\gamma = 11742$ for a 6 GeV synchrotron. The characteristic angle $\frac{1}{\gamma}$ of the emission cone then reduces to 0.085 mrad or 0.0049° (cf. Fig. 1).

One of the most important features of synchrotron radiation is the universality of the emitted continuous spectrum, shown in Fig. 2. It depends only on a few specific source parameters such as electron beam energy and the strength of the magnetic field which forces the electrons on a curved path (also expressed as the bending radius R). The spectrum scales with the critical frequency

$$\omega_c = \frac{3}{2}\frac{c}{R}\gamma^3 = \frac{3}{2}\frac{c}{R}\left(\frac{E}{m_0\,c^2}\right)^3, \tag{3}$$

which is defined as the frequency up to which exactly 50 % of the total power is radiated. The peak of the spectrum is located at about $\omega_c/3$, with analytical approximations indicated in Fig. 2. The rather featureless spectrum can be calculated via integrals over Bessel functions. For details, the interested reader is referred to the X-ray Data Booklet, which is freely available
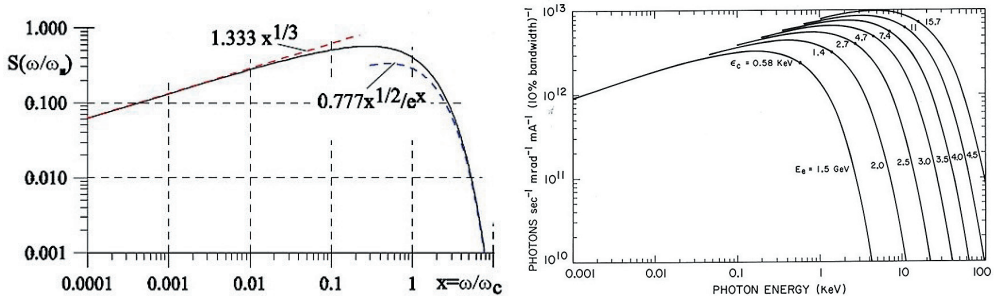

Fig. 2: Left: Spectral distribution of synchrotron radiation at fixed electron beam energy; right: hardening of the spectrum with increasing electron beam energy [4].

on the internet [5], and references given therein. The electron energy determines the hardness of the resulting photon spectrum (cf. Fig. 2 (right)). The intensity falls off quickly beyond the characteristic frequency (or equivalently, the characteristic energy $E_c := \hbar\omega_c$), whereas the part of the spectrum well below the characteristic frequency is essentially independent from the electron energy. The characteristic energy can be expressed in practical units as

$$\frac{E_c}{\text{keV}} = 0.665 \left(\frac{E}{\text{GeV}}\right)^2 \frac{B}{\text{Tesla}}. \tag{4}$$

The power radiated by a single electron is given by

$$P = \frac{e^2 c}{6\pi\epsilon_0} \frac{1}{R^2} \left(\frac{E}{m_0 c^2}\right)^4, \tag{5}$$

which underlines why only charged particles with small rest mass m_0 (e.g., electrons or positrons) are used in synchrotron radiation facilities.

2.3 Other properties: collimation, brilliance, time structure and coherence

Synchrotron radiation is extremely collimated in the plane of the accelerator. This is a relativistic effect and illustrated in Fig. 3. In the rest frame of the electron, the emitted radiation exhibits the characteristic donut-shaped dipole emission pattern. However, since the electron moves almost at the speed of light, the emission pattern appears drastically changed to an observer in the laboratory frame. The coordinate in the direction of motion is transformed via a Lorentz transformation, whereas the two other spatial coordinates remain unaffected. Accordingly, the emission is peaked very strongly in a narrow forward cone as indicated on the right hand side of Fig. 3. This radiation pattern is observable in tangential directions around the ring from all points of the electron trajectory where the electron is accelerated.

We want to define some useful quantities characterizing light sources. The brightness

$$B = \frac{\Delta P}{\Delta\Omega \cdot \Delta A} \quad ; \quad [B] = \frac{\text{photons/s}}{\text{mrad}^2 \cdot \text{mm}^2} \tag{6}$$
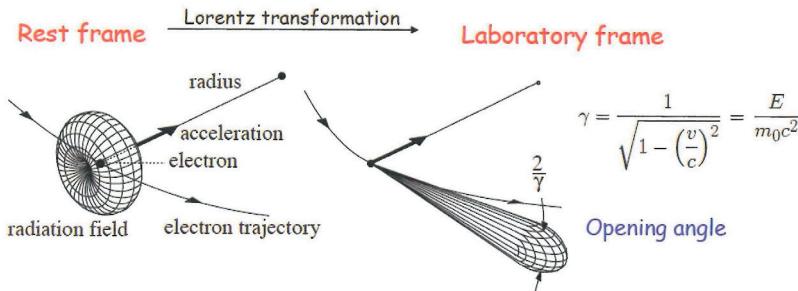

Fig. 3: Emission pattern of dipole radiation from a relativistic charged particle, viewed within its rest frame (left) and by an observer in the laboratory (right) [6].
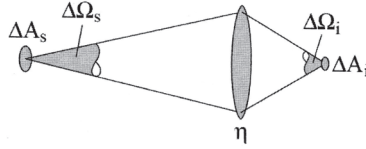
Fig. 4: A source of finite size ΔA_S emits into a solid angle $\Delta \Omega_S$ and is imaged by an optical system with efficiency η. Under ideal conditions, the product $\Delta \Omega \cdot \triangle A$ is conserved, usually expressed via the brilliance (Eq. (7)) [1].

is the power ΔP radiated by a source of area ΔA into a solid angle $\Delta \Omega$, with practical units as indicated. By contrast, the spectrally resolved brightness

$$B_{Spectral}(E) = \frac{\Delta P(E, \Delta E)}{\Delta \Omega \cdot \Delta A} \quad ; \quad [B_{Spectral}] = \frac{\text{photons/s}}{\text{mrad}^2 \cdot \text{mm}^2 \cdot 0.1\% \,\text{bandwidth}} \tag{7}$$

is called brilliance and refers only to photons within a band pass ΔE centered at the energy E. Since relative bandwidths for crystal monochromators are on the order of $\Delta E / E = 0.1\%$ (see section 3.2), the brilliance is usually given in the practical units of Eq. (7).

Both quantities take into account the area of the light emitting source, which is here the cross section of the electron beam. They are fundamental figures of merit, as the product $\triangle \Omega \cdot \Delta A$ represents an invariant not affected by focusing optics (Fig. 4). It is therefore possible to produce from the same source by some suitable optics a small, divergent beam as well as a large, nearly parallel beam according to experimental needs; however the figure of merit represents a limit to what can be achieved.

Fig. 5 shows the exponential growth of the brilliance of synchrotron X-ray sources during the last decades. The change of technology is manifest as kinks in the plot, both for the transition from tubes to synchrotrons and the advent of free electron lasers, resulting in an additional increase of peak brilliance by another 10 orders of magnitude. The reduction of the effective source size ΔA by advances in electron beam focussing has contributed significantly to the increase in brilliance.

X-ray generation in storage rings is associated with a definite time structure , analogous to a stroboscopic light source, since the electrons are not randomly distributed along the circumference of the ring. The reason lies in the bunching of the injected electrons, as they have to pass the radio frequency cavities at a certain phase in order to be accelerated. Typical electron bunch lengths are a few cm, separated from each other by several 10 cm. Due to the electron speed being close to the speed of light and the high collimation (cf. Fig. 1), an observer at a beamline registers X-ray pulses of several 10 ps duration, repeated every few nanoseconds. The maximum number of electron bunches is given by the circumference of the storage ring, divided by the number of 'buckets' for such bunches (typically several hundred). Not all possible 'buckets' are necessarily filled during synchrotron user operations, since the large variety of experiments requires different filling modes for optimal performance. Hence, the filling mode is changed regularly (for example, on a weekly basis), and the user experiments are grouped accordingly for beamtime allocation.

The angular distribution of both polarization components is different with respect to the accelerator plane. While the parallel polarization component is strongest in the plane of the ring,
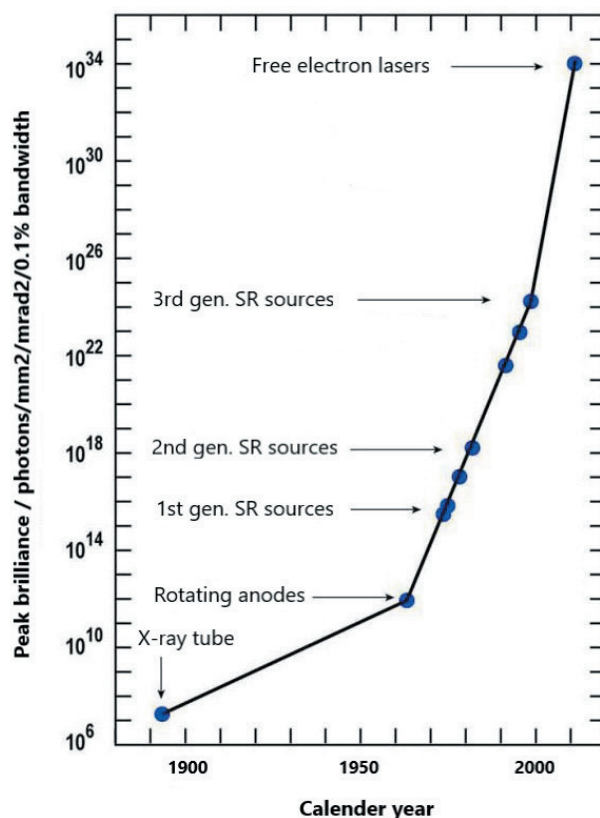
Fig. 5: Evolution of the peak brilliance in the history of X-rays from tubes to synchrotron radiation sources and free electron lasers (note the logarithmic scale).

the vertical component vanishes in this plane, with a wavelength-dependent maximum above and below. Therefore, in general an elliptical polarization results, with opposite helicities above and below the accelerator plane. Consequently, synchrotron radiation is linearly polarized in the accelerator plane. It should be pointed out that the angular distribution of the polarization components can be expressed in a universal way by Bessel functions, as was already the case for the spectral distribution, independent of any particular storage ring.

The natural polarization is one of the most important features of synchrotron radiation. Details on polarization handling are given in talk C4.

The degree of coherence is an important property of all light sources. Whereas coherent sources of visible light - such as lasers - have been available for decades, coherent X-rays with sufficient intensity have only become recently available from high-brilliance sources. Loosely speaking, the coherence of a light source describes the ability of its emitted wavefronts to form extended, stable interference patterns needed for example in holography or lensless imaging. As depicted in Fig. 6, any light source can be made coherent by appropriate experimental measures; how-
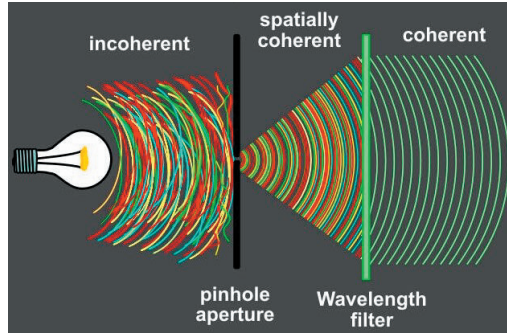
Fig. 6: Improving the spatial and temporal coherence of electromagnetic waves from an incoherent source using a pinhole and monochromatization [7].

ever, the question is how much intensity is left afterwards. Placing an aperture in front of the source increases the so-called spatial or transverse coherence. A frequent choice are pinholes, which approach the ideal of a point source. By contrast, temporal or longitudinal coherence refers to the wavelength purity of the radiation.

For a quantitative treatment of coherence, the reader is referred to talk D7.

2.4 Wigglers and undulators

Modern storage rings are more polygons than circles, with bending magnets at the corners of the polygon and straight sections between the corners. These sections are used for so-called insertion devices, where electrons are periodically deflected from sequences of small but strong (mostly permanent) magnets. Although the emission cone is intrinsically narrow for radiation from a bending magnet, the photons are constantly emitted while the electrons follow their circular path, resulting in a large horizontal fan (e.g., in the orbit plane). By contrast, radiation from insertion devices is concentrated in a small angular range in forward direction.

The most important insertion devices are wigglers and undulators. Their basic structure is identical, depicted in Fig. 7. The periodic arrangement of magnets forces the electrons onto a sinusoidal path in the horizontal plane. An important parameter to distinguish wigglers and undulators is the deflection parameter K:

$$K = 0.934 \frac{\lambda_U}{\text{cm}} \frac{B}{\text{Tesla}}, \tag{8}$$

where λ_U is the period and B is the magnetic field experienced by the electrons.

In a wiggler, the N pairs of magnets create a spectrum that is equivalent to the intensity superposition of N bending magnets. However, strong magnets separated by a small gap (cm range) yield an equivalent bending magnet that is significantly stronger than the bending magnets at the corners of the storage ring polygon. Therefore the wiggler spectrum is not only far more intense, but also shifted to higher photon energies. The maximum deflection angle from the x-axis is given by $\alpha = K/\gamma$, leading to a corresponding horizontal spread. Wigglers are characterized by high magnetic fields and long periods, resulting in $K \gg 1$.

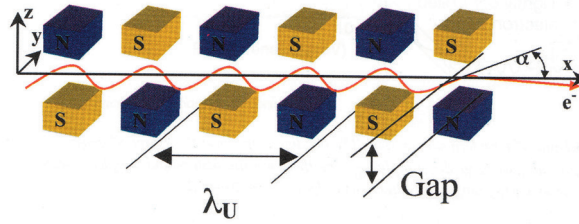Undulators are similar to wigglers except for much smaller K-values (typically $K < 1$). Their

Fig. 7: Schematic structure of wigglers and undulators. The period λ_U is typically a few cm; the strength of the magnetic field on the x-axis can be changed by varying the vertical gap between the two rows of magnets [1].

spectral and spatial radiation distributions are dominated by strong interference effects, hence undulators are not continuum sources like wigglers and bending magnets. The spectrum consists of sharp peaks at wavelengths

$$\lambda = \frac{1}{M}\frac{\lambda_U}{2\gamma^2}\left(1 + \frac{1}{2}K^2 + \gamma^2\theta^2\right),\tag{9}$$

where θ is the observation angle with respect to the undulator axis and $M = 1, 2, 3, \ldots$ indicate that also higher harmonics are generated.[1] In essence, the generation of photons with a wavelength λ on the order of 10^{-10} m in the laboratory frame is brought about by the Lorentz contraction of the undulator period λ_U (about 10^{-2} m) and the relativistic Doppler effect, yielding *two* times a γ-factor on the order of 10^4.

Since the properties of undulator radiation are based on a coherent superposition of the radiation emitted in the N periods of the device, the intensity is proportional to the square of the amplitude and hence is proportional to N^2. Undulators are frequently used at beamlines where the experiments profit most from a very intense, strongly collimated beam and need not be performed at a very specific wavelength, or when the coherence properties of the radiation are important. The tuning of the wavelength at an undulator beamline is considerably more complicated than at a wiggler beamline. Whereas at a continuum source only the monochromator needs to be changed, an undulator requires according to Eq. (9) to vary the K parameter, which is accomplished by varying the magnetic field. This is usually done by increasing the gap mechanically for shorter wavelengths and vice versa.

3   Instrumentation

3.1   Typical beamline layout

The typical beamline layout found at 3rd generation storage rings is depicted in Fig. 8. All beamlines are arranged tangentially to the storage ring. From experience gathered at 2nd generation sources, the following layout has evolved:

---

[1] In an ideal undulator, only odd harmonics can be observed on the undulator axis.
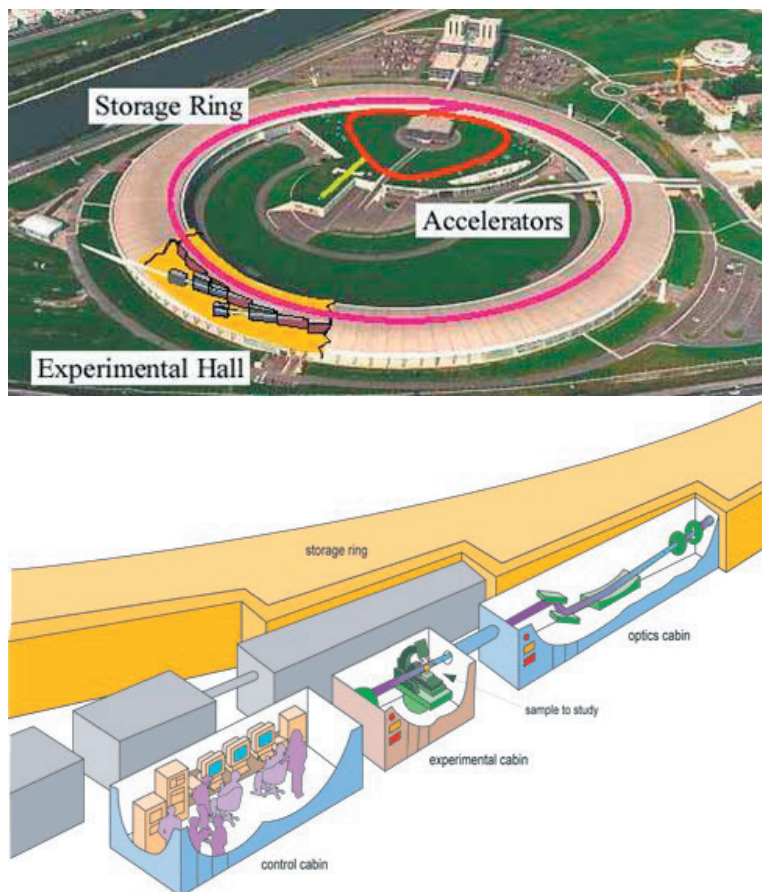
Fig. 8: The European Synchrotron Radiation Facility (ESRF) , Grenoble, with a sketch of a typical beamline layout consisting of separate hutches for the optical elements, the sample interaction and the researchers controlling the experiment (Source: ESRF).

- The optics hutch receives the incoming X-ray beam, which is modified according to the desired experimental characteristics. For example, the polychromatic beam is monochromatized here; also focusing optics are housed in this hutch.

- The experimental hutch contains the sample to be studied and a diffractometer (or other positioning devices). In case of in-situ studies, the sample environment is mounted onto the diffractometer as well, allowing to control external parameters like temperature, magnetic field etc. One or more detectors record the scattering which results from the interaction of the X-ray beam with the sample. Owing to the high level of ionizing radiation, no access during the experiments is possible, requiring remote control for each apparatus inside the experimental hutch.

- The control hutch allows the researchers to control their experiment, collect, store and transfer the data. Frequently, some basic evaluations are already performed here during

the beamtime.

It should be added that hutches which contain the X-ray beam are heavily shielded in cases of storage rings producing hard X-rays. Typically, they are called lead hutches because their shielding consists of Al/Pb sandwich plates. In case of storage rings designed for soft radiation (UV and soft X-rays), lead hutches are only found at those beamlines where hard X-rays are generated by special insertion devices. The transition from soft to hard X-rays is not sharply defined, occurring at about 2 - 4 keV photon energy. At the other end of the spectrum, the soft X-ray range overlaps with the ultraviolet range, which is also called vacuum-ultraviolet (VUV) - reminding us that the radiation is so soft that it is absorbed even by air after a short distance. In these cases, the steel vessels used to transport the beam and to perform the experiments are more than sufficient to provide shielding, and hutches do not exist.

Under ideal conditions, the brilliance is conserved for a stream of photons emanating from a source and propagating down the beamline (see Fig. 4). If an experiment with a small sample requires an intense and highly parallel photon beam, a high-brilliance source like a $3^{\rm rd}$ generation synchrotron is therefore required. Also the usable photon flux for experiments which require spatial coherence is determined by the source brilliance.

3.2   Monochromatization

Before synchrotron radiation can be profitably applied in experiments, a monochromatization is required in many cases. There are generally two classes of X-ray monochromators:

- diffraction gratings for soft X-rays, which work well for wavelengths down to about 6 Å,

- crystal monochromators for hard X-rays.

A diffraction grating is a surface with a periodic array of geometrical features of equal shape, for example grooves ('lines'). Since a grating is a dispersive element, the combination with a slit can be used for monochromatization. As shown in elementary diffraction theory, the condition for constructive interference is

$$d \left(\sin\theta_i + \sin\theta_d\right) = m\lambda, \tag{10}$$

where d is the grating period, θ_i and θ_d are the angles of the incident and the diffracted beam, respectively, λ is the wavelength and m a positive integer.
Monochromatization filters out a narrow band of wavelengths from a continuous distribution, with the width $\Delta\lambda$ of the band called the 'absolute resolution' of the monochromator. A high-quality monochromator provides only a small band, at the expense of the available intensity. Frequently, the parameter $\lambda/\Delta\lambda$, called 'resolving power', is used to quantify the monochromator performance:

$$\frac{\lambda}{\Delta\lambda} = m\, N_L \left(\sin\theta_i + \sin\theta_d\right). \tag{11}$$

Since N_L is the number of illuminated lines of the grating, many lines therefore improve the resolving power. In practical setups, the grating is combined with slits and mirrors to obtain the desired performance, which is always a trade-off between high resolution and high intensity.

In the hard X-ray range, crystals are the dominant type of monochromators. The reason can be seen by the extreme case of Eq. (10) when both angles θ_i and θ_d approach 90°. In this case, the grating period d has to be on the order of the wavelength according to $d \approx m\lambda/2$, since the period cannot exceed the wavelength significantly even for high orders. For short wavelengths on the Å scale, it is technically difficult to fabricate such gratings.[2] However, the long-range atomic order of a crystal provides a natural grating with an extremely short period. The basic law of crystal diffraction is the Bragg law

$$n\lambda = 2\,d\sin\theta, \tag{12}$$

where n is the order of the diffraction, λ is the wavelength and d is the lattice spacing. Frequently, the order n is combined with the lattice spacing d to give the actual spacing probed by the X-ray wave fields at the (h k l) reflection. With d replaced by d_{hkl}, the Bragg equation now reads

$$\lambda = 2\,d_{hkl}\sin\theta, \tag{13}$$

with common factors now allowed in the values for (h k l). For the rest of this chapter, we will use this notation.

A polychromatic X-ray beam impinging on a crystal will produce a characteristic set of Bragg reflections (Laue pattern). In principle, a receiving slit, together with shielding material, placed after the monochromatizing crystal can thus block unwanted wavelengths and select only the desired reflection. Suitable crystals for hard X-ray monochromators are based on a number of criteria. Of foremost importance are the 2d-values, since these should match the wavelengths of interest to give practical Bragg angles. The crystal quality is also an important parameter, since excellent single crystals have narrow reflection widths in contrast to mosaic crystals. Both types find applications in high-resolution and high-flux monochromators, respectively. Additional factors that have to be considered are thermal stability and resistance to radiation damage, especially for the first crystal encountered by the polychromatic synchrotron beam, which receives a high heat load.

The so-called 'rocking curve', a plot of the Bragg-reflected intensity as a function of the incident angle θ, is characteristic for each monochromator. The Bragg equation implies a δ-function, but in reality even the rocking curves for perfect crystals possess a small but finite width of a few thousands of a degree. The physics behind the finite width is that the incident X-ray wave field can only reach a finite, although large number of Bragg planes because of the finite intensity taken out of the beam by each reflecting plane (extinction of the incident beam). The width of a Bragg reflection indexed (h k l) can be calculated for a perfect crystal in the dynamical theory of X-ray diffraction to be

$$\delta\theta_B = \frac{2\,C_{Pol}\,|\,\chi_{hkl}\,|}{\sqrt{|\,b\,|}\,\sin\,(2\theta_B)}, \tag{14}$$

where C_{Pol} is the linear polarization factor, θ_B is the Bragg angle, and χ_{hkl} the corresponding Fourier coefficient of the crystal polarizability $\chi\,(\mathbf{r})$. The asymmetry factor b is defined as

---

[2] The deposition of multilayer thin films provides a possibility to create artificial one-dimensional periodic structures with a short period, which are sometimes used as low-resolution monochromators.
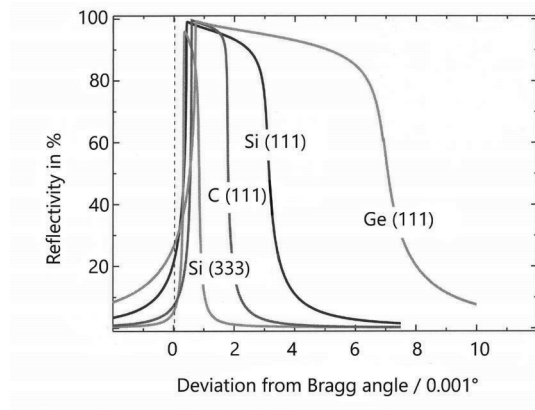
Fig. 9: Normalized rocking curves for various [111] oriented crystals. Note the shrinking of the Darwin width for Si (333) compared to the Si (111), typical for highly indexed reflections (modified after [8]).

$$b = -\frac{\sin(\theta_B + \phi)}{\sin(\theta_B - \phi)} \tag{15}$$

with ϕ being the angle between the diffracting lattice planes and the crystal surface. Typical widths of reflections are shown in Fig. 9.

There is a great variety in the design of crystal monochromators. An effective way to improve the resolving power is to use Bragg reflections from two different crystals (see Fig. 10). The highly dispersive mode is also called the (+, +) setting, since the crystals have to rotate in the same direction when changing the wavelength; consequently, the antidispersive mode is called the (+, -) setting. The latter configuration is the standard double crystal monochromator at most synchrotron beamlines, simply called DCM. It possesses the advantage that the exit beam can be kept spatially fixed when changing the wavelength by modifying the distance between the two crystals accordingly. The four-bounce monochromator depicted on the right hand side in Fig. 10 combines highest resolution with simple wavelength tuning. The multiple reflections help to suppress the tails in the single crystal reflection curves (Fig. 9), resulting in an improved resolution function.

The resolution obtained from a single reflection is discussed in the Appendix. In essence, differentiating the Bragg equation yields:

$$\frac{\Delta E}{E} = \frac{\Delta \lambda}{\lambda} = \cot \theta_B \, \Delta\theta, \tag{16}$$

with $\Delta\theta$ being the relevant angular spread. For high resolution monochromators, this will include the divergence of the incident beam as well as the intrinsic width $\delta\theta_B$ of the reflection. Typical resolutions for good single crystals are on the order of 10^{-4} to 10^{-3}. Eq. (16) makes immediately clear that backscattering arrangements with θ_B approaching $90°$ will provide the best resolution.
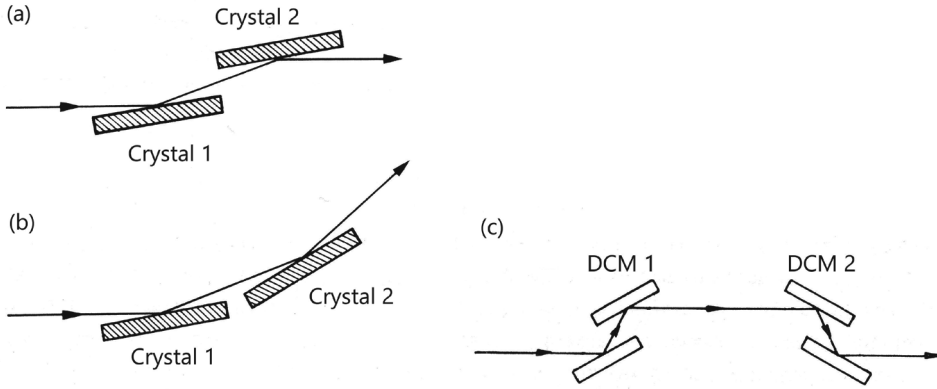
Fig. 10: A double crystal monochromator (DCM) in the (a) parallel (antidispersive) and (b) highly dispersive setting; (c) high-resolution monochromator with four reflections and zero beam deviation, built as a combination of two DCMs.

DuMond diagrams are helpful to visualize the band selected from a continuous spectrum due to the finite angular acceptance of a particular Bragg diffraction. Whereas the Bragg equation couples λ and θ to each other, the idea here is to view them as independent because of the finite reflection width given by Eq. (14), and hence display the wavelenghts transmitted by a monochromator in the λ-θ-plane. For simplicity, the transmission function is approximated by a rectangular function, so that the basic plot in a DuMond diagram consists of a graph of the sin-function from Eq. (12), but modified to have a finite width (see Fig. 13) [9]. The different resolutions resulting from the (+, +) and (+, -) settings for two crystals can easily be understood using DuMond diagrams. In the dispersive (+, +) setting, the transmission curve of the second crystal must be plotted in the opposite direction.

The small overlap area indicates that only a small range of wavelengths leaves the monochromator. The excellent resolution of this arrangement can be used for laboratory tubes to even
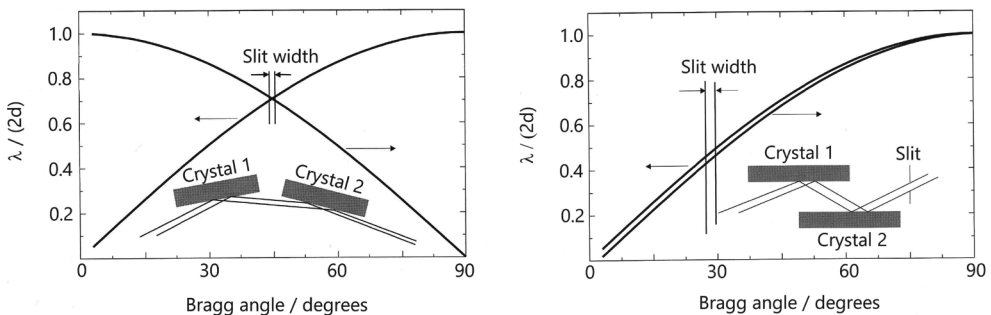

Fig. 11: DuMond diagrams and corresponding crystal arrangements for double crystal monochromators in the antidispersive and dispersive setting (modified after [8]).

separate the $K\alpha_1$ and $K\alpha_2$ lines. By contrast, the parallel (+, -) setting implies two parallel graphs with a large overlap, i.e. all wavelengths diffracted by the first crystal are diffracted by the second crystal as well, which results in far greater flux after the monochromator. Of course, in a strict sense the (+, -) setting is non-dispersive only if both crystals are equal and the same Bragg reflection is used.

It should be noted that the broad spectrum of polychromatic synchrotron radiation leads generally to the unwanted admixture of higher-order wavelengths like $\lambda/2$, $\lambda/3$ etc. in the monochromatized beam if no special measures are taken. When Si or Ge (111) reflections are used for monochromatization, the important second-order harmonics are already suppressed because of the structure factor of the diamond lattice (e.g., (222) is a forbidden reflection with zero intensity); unfortunately, this does not provide a general method for higher-order rejection. Owing to refraction effects, the higher-order harmonics possess absolute Bragg angles slightly but systematically shifted away from the fundamental reflection. This can be used to suppress higher harmonics by slightly rotating the second crystal of the DCM (on the order of a thousand of a degree), thereby moving the center of the band pass away from the harmonics. Another popular means of suppressing harmonics is the reflection from mirrors using an appropriate angle of incidence. The wavelength-dependent total reflection of X-rays under grazing incidence up to a critical angle (see talk D2) makes mirrors effective low passes (Fig. 12); in addition they can also serve to achieve focusing in one or two planes with an appropriately curved shape.
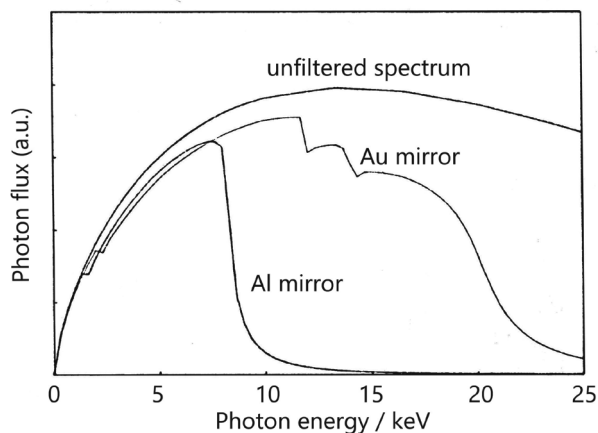
Fig. 12: The total reflection of mirrors under grazing incidence provides an effective low pass frequently used for the suppression of higher harmonics. The topmost curve is an unfiltered bending magnet spectrum. The structures in the curve of the Au mirror result from absorption edges (modified after [10]).

3.3 Sample interaction and analysis of the scattered beam

Standard instruments for angle-dispersive measurements at fixed wavelength are so-called four- or six-circle diffractometers with sufficient degrees of freedom to scan or map those parts of reciprocal space which are of particular interest in the problem to be investigated. In the simplest case, encountered in many laboratory setups, only two circles are present: the sample rotation

(usually called θ or ω), and the detector rotation (usually called 2θ, although this degree of freedom is independent from the sample rotation and is only twice the detector angle in certain scan types). Both rotations are restriced to the plane of incidence.

Four- and six-circle diffractometers possess additional degrees of freedom for sample positioning, which open up more possibilities in accessing reciprocal space. For example, the six-circle diffractometer combines the features of a four-circle and a so-called z-axis diffractometer [11]. It is useful in surface sensitive scattering, where a grazing angle of incidence (the control of which is important for depth resolution) must be kept constant during a scan and independent from the Bragg angle of the reflection to be measured. The degrees of freedom for a six-circle diffractometer are δ (detector arm rotation), ω (rotation of the sample circles), χ (sample tilt), φ (sample rotation), α (rotation of the entire diffractometer), and γ (out-of-plane detector rotation). There are also special setups for energy-dispersive measurements performed at fixed angles, usually either because there is no space for large angular variations (for example in high-pressure diffraction with anvil cells) or because of time gains (parallel measurements of a large range of q-values). The interested reader is referred to the literature for details [12].

The scattered beam can be analyzed in a variety of ways with respect to the required angular or energetic resolution. In the simplest case of a scintillation or gas counter used as a point detector, the angular divergence of the detected radiation is typically reduced by a slit or a collimator system (for example Soller slits), thus providing angular resolution in scans. If reciprocal space maps are measured, a good resolution is required in two directions [8]. High-resolution measurements frequently involve so-called analyzer crystals . Quite analogous to monochromator crystals, one or more analyzer crystals can be positioned behind the sample, deflecting the scattered beam into the detector. They can be viewed as an extreme form of slits with a width given by Eq. (14). However, they are more than that, since the scattered radiation has to fulfil the Bragg equation, resulting in a very effective suppression of inelastic background (for example, fluorescence or Compton scattering). Analyzer crystals are sometimes bent to achieve higher intensity if a relaxed angular resolution is acceptable.

The advent of very bright undulator radiation has also brought possibilities for the measurement of extremely weak inelastic scattering, which provides experimental access to elementary excitations in solids using X-rays. The magnitude of the problem becomes clearer if one realizes that, for example, in such experiments phonon energies in the 1 - 100 meV range are to be probed with 20 keV photons. Extreme resolutions in the meV or sub-meV regime can, according to Eq. (16), only be reached with analyzer crystals in backscattering geometry, using highly indexed reflections like Si (9 9 9) etc. At such high resolutions, energy scans are not performed any more by rotations, but by varying the relative temperature of the monochromator and analyzer crystals with mK precision, since the lattice constant enters the reflection width according to Eq. (23). For more details on inelastic X-ray scattering and the various spectrometer designs, the reader is referred to the literature [13, 14].

A modern synchrotron facility hosts many beamlines, all of them tailored to specific tasks and receiving radiation from insertion devices which exactly match their purposes. The following list, being certainly not exhaustive, is intended to give an impression of the degree of specialization:

- crystallography (powder diffraction, protein crystallography, high pressure),

- topography,

- small angle scattering (SAXS, GISAXS),

- surface diffraction and grazing incidence,

- absorption spectroscopy (EXAFS, XANES, dichroism),

- imaging and microscopy,

- inelastic scattering (Compton, high resolution, nuclear resonance),

- time-resolved scattering,

- coherence applications,

- photoelectron spectroscopy.

Further details of many of these specialized techniques will be presented in the talks D1 - D9 of this Spring School.

Acknowledgement

Appendices

A   Flux and Resolution in a DuMond Diagram

The relation between energy and wavelength is for photons

$$E = \frac{hc}{\lambda} \tag{17}$$

or in practical units

$$E/keV = \frac{12.398}{\lambda/(\text{Å})}. \tag{18}$$

Differentiating Eq. (17), one obtains

$$dE = -\frac{h\,c}{\lambda^2}\,d\lambda, \tag{19}$$

where the minus sign signifies that an increase dE in the energy is related to a decrease $d\lambda$ in the wavelength. For a finite but small energy interval ΔE, one obtains therefore for the relative resolution

$$\frac{\Delta E}{E} = \frac{\Delta\lambda}{\lambda}, \tag{20}$$

where the minus sign has been dropped since $\Delta\lambda$ is usually also defined to be a positive quantity.

To discuss flux and resolution in a DuMond diagram, we start with the Bragg equation in the form of Eq. (13). In a continuum of polychromatic X-rays, any small angular spread $d\theta$ is related to a corresponding spread in wavelength $d\lambda$ obtained by taking the differential of Eq. (13):

$$d\lambda = 2\,d_{hkl}\cos\theta\,d\theta, \tag{21}$$

resulting in the spectral resolution

$$\frac{d\lambda}{\lambda} = \cot\theta\,d\theta \tag{22}$$

by dividing both expressions. These are our basic equations, which are now applied to two different situations.

First, we look at the intrinsic wavelength spread $\delta\lambda$ associated with the intrinsic width $\delta\theta_B$ of a Bragg reflection indexed (h k l). Inserting Eq. (14) in Eq. (21) and using the identity $\sin(2x) = 2\sin x\,\cos x$ yields

$$\delta\lambda = 2\,d_{hkl}\cos\theta_B\,\delta\theta_B = d_{hkl}\,\frac{2\,C_{Pol}|\chi_{hkl}|}{\sqrt{b}\,\sin\theta_B}. \tag{23}$$

Second, we look at the flux resulting from polychromatic radiation with an angular divergence $\Delta\theta$. The flux is proportional to the shaded area in Fig. 13:

$$F \sim \delta\lambda \cdot \triangle\theta. \tag{24}$$

We now want to calculate the spectral resolution for a situation where both the angular divergence $\Delta\theta$ and the intrinsinc spread $\delta\theta_B$ contribute significantly to the total wavelength spread $\Delta\lambda$ using Eq. (22):

$$\frac{\Delta\lambda}{\lambda} = \cot\theta_B\,(\Delta\theta + \delta\theta_B). \tag{25}$$

With the help of Eq. (23), we obtain the final result

$$\frac{\Delta\lambda}{\lambda} = \cot\theta_B\,\Delta\theta + \frac{C_{Pol}|\chi_{hkl}|}{\sqrt{b}\,\sin^2\theta_B}.$$
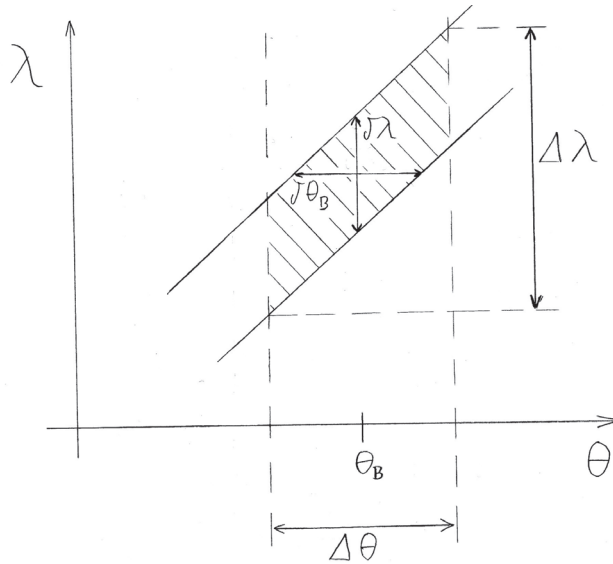
(26)

Fig. 13: DuMond diagram for a single crystal monochromator, showing only a small part of the $\sin$ -curve. The photon flux through the monochromator is proportional to the shaded area.

References

[1] S. Cramm, *Synchrotron Radiation Sources and Beamlines: Overview*, in K. Urban, C. Schneider, T. Brückel, S. Blügel, K. Tillmann, W. Schweika, M. Lentzen, L. Baumgarten (eds.), *38th IFF Spring School: Probing the Nanoworld* (Forschungszentrum Jülich, 2007), chapter B 2.

[2] D. D. Ivanenko, I. Pomeranchuk, Phys. Rev. **65**, 343 (1944).

[3] J. Schwinger, Phys. Rev. **70**, 798 (1946).

[4] *Frequency distributions of radiated energy and synchrotron radiation emission as a function of the beam energy*, https://commons.wikimedia.org/wiki/File: Frequency_distributions_of_radiated_energy.png, https: //commons.wikimedia.org/wiki/File:Syncrotron_radiation_ emission_as_a_function_of_the_beam_energy.png, [CC BY 2.5 (https://creativecommons.org/licenses/by/2.5)], via Wikimedia Commons.

[5] *X-Ray Data Booklet*, `http://xdb.lbl.gov/xdb-new.pdf`, Center for X-ray Optics and Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, USA (2009).

[6] W. Eberhardt, *Synchrotron Radiation Sources*, in M. Angst, T. Brückel, D. Richter, R. Zorn (eds.), *43rd IFF Spring School: Scattering Methods for Condensed Matter Research: Towards Novel Applications at Future Sources* (Forschungszentrum Jülich, 2012), chapter C 2.

[7] W. Beaty, *Lasers: What is Coherent Light?*, `http://amasci.com/miscon/coherenc.html` (2004).

[8] U. Pietsch, V. Holý, T. Baumbach, *High Resolution X-Ray Scattering: From Thin Films to Lateral Nanostructures* (Springer Verlag, Berlin, 2004), chapter 1.

[9] J. DuMond, Phys. Rev. **52**, 872 (1937).

[10] S. Joksch, *Röntgenoptik für Synchrotronstrahlung hoher Leistung*, in R. Hölzle (ed.), *23. IFF-Ferienkurs: Synchrotronstrahlung zur Erforschung kondensierter Materie* (Forschungszentrum Jülich, 1992), chapter 20.

[11] M. Lohmeier, E. Vlieg, J. Appl. Cryst. **26**, 706 (1993).

[12] D. Mills (ed.), *Third Generation Hard X-Ray Synchrotron Radiation Sources: Source Properties, Optics, and Experimental Techniques* (Wiley, New York, 2002).

[13] Y. Shvyd'ko, *X-Ray Optics: High-Energy-Resolution Applications* (Springer Verlag, Berlin, 2004), chapter 1.

[14] H. Sinn, J. Phys.: Condens. Matter **13**, 7525 (2001).

[15] U. Klemradt, *Synchrotron X-ray Beamlines*, in M. Angst, T. Brückel, D. Richter, R. Zorn (eds.), *43rd IFF Spring School: Scattering Methods for Condensed Matter Research: Towards Novel Applications at Future Sources* (Forschungszentrum Jülich, 2012), chapter C 4.

C 4   Polarisation Handling

J. Voigt

Jülich Centre for Neutron Science

Forschungszentrum Jülich GmbH

Contents

1   Introduction

When you lookup the expression 'Polarization' in Wikipedia, you find a large number of definitions depending on the context. E.g. in the social sciences several ambiguous definitions exist. And in particular in the physical sciences the expression 'polarization' describes different properties of a system of interest. In this lecture we deal with the polarization of a neutron or a x-ray beam and how it is affected by the interaction with a sample in a scattering experiment.

As you have learned already, the total energy and momentum of the sample and the probe particle are conserved in a scattering experiment. Therefore we can measure the momentum and energy exchanged with the sample by determination of the energy and momentum of the probe particle before and after scattering. It is then an obvious question, how the conservation of angular momentum can be used to explore the spins and orbital momentum of the sample.

The polarization of a neutron or x-ray beam must therefore be related to the angular momentum of the individual particles in the beam, namely the neutron or the photon. At the end of this lecture we will adress this point again. But first we need to understand the how the polarization of a beam is defined and how we can measure it. We address the x-ray case in section 2 and the neutron case in section 3, before we return to the common formalism used for the description, which is due to the same dimensionality of the Hilbert space in both cases.

2   Polarized x-ray beam

Most of you might have experience with polarized light, e.g. using sunglasses with polarization filters or you even might have used polarized light microscopes for your own research. For these applications we use the linear polarization of light, i.e. electric field vector of the electro-magnetic wave is oriented only in one direction. Since electro-magnetic waves are transverse, the polarisation state of any wave in vacuum can be expressed by two components. Therefore a polarization microscope uses crossed polarizers/analyzers to describe the birefringent properties of the sample. As x-rays are just electric-magnetic wave with a short wavelength, the entire formalism as you know it perhaps for optical light applies. If we describe the x-ray field in the particle picture, the polarization is related to the expectation value for the spin quantum number. The photon is a spin 1 particle with angular momentum $l = 1$. But because the the photon has no mass, the magnetic quantum number can have only values $m = \pm 1$ and the angular momentum is either parallel or antiparallel to the wave vector of the photon. These states correspond to left or right circular polarized wave, i.e. the electric field vector at one point in space lies on a circle and rotates either clock-wise or counter clock-wise. Historically the polarization state of the x-ray beam is expressed in the basis of the orthogonal linearly polarized beams. It can be transformed to a basis of left and right circularly polarized photons by superposition of the linear states with complex phases. For the reminder of the lecture we will use the former basis as it is typically used to describe the anisotropic scattering of x-ray e.g. by magnetic moments.

As you might know the polarization of light can also be parametrized by the Poincare- Stokes

parameters.

$$P_1 = \frac{\langle |E_\sigma + iE_\pi|^2 - |E_\sigma - iE_\pi|^2 \rangle}{2\langle E_\sigma^2 + E_\pi^2 \rangle} \tag{1}$$

$$P_2 = \frac{\langle |E_\sigma + E_\pi|^2 - |E_\sigma - E_\pi|^2 \rangle}{2\langle E_\sigma^2 + E_\pi^2 \rangle} \tag{2}$$

$$P_3 = \frac{\langle E_\sigma^2 - E_\pi^2 \rangle}{\langle E_\sigma^2 + E_\pi^2 \rangle}, \tag{3}$$

where E_σ, E_π denote the amplitudes of the electric field vector. P_3 is a measure for the power the wave field in σ ($P_3 = 1$) or π direction ($P_3 = -1$), P_2 corresponds to the power of the components with linear polarization at an angle $45°$ to the $\sigma\pi$ plane and P_1 to the power of the circular polarized fraction of the beam. We depart here from the convention by exchanging P_1 and P_2 in order to apply the same expressions as in the discussion of the neutron beam polarization. The Stokes parameter allows also the description of a partial polarization of the beam. At the end of the lecture we will come back to this parameters, when we discus the common formalism to describe the neutron and x-ray polarisation.

2.1 The coordinate system

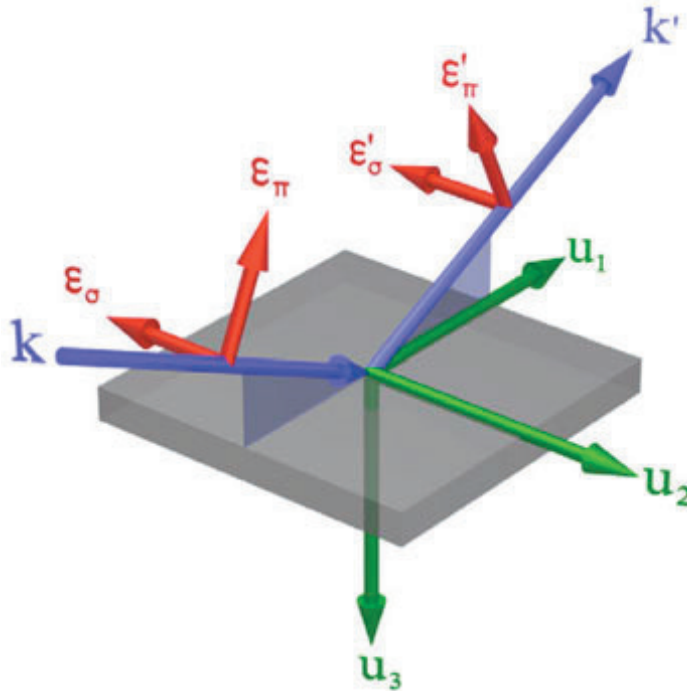**Fig. 1:** Coordinate system for scattering of a wave with wave vector **k** towards wave vector **k**′. From [1]

In a scattering experiment we illuminate the sample by a beam with wave vector **k** and detect, how much of it has been scattered into a direction of the wave vector **k**′, so that the wave vector

transfer $\mathbf{Q} = \mathbf{k}' - \mathbf{k}$ has been exchanged with the sample, see Fig 1. Based on that we use the coordinate system [1]:

$$\hat{\mathbf{u}}_1 = \frac{1}{\cos\theta}\left(\hat{\mathbf{k}} + \hat{\mathbf{k}}'\right) \tag{4}$$

$$\hat{\mathbf{u}}_2 = \frac{1}{\sin 2\theta}\left(\hat{\mathbf{k}} \times \hat{\mathbf{k}}'\right) \tag{5}$$

$$\hat{\mathbf{u}}_3 = \frac{1}{2\sin\theta}\left(\hat{\mathbf{k}} - \hat{\mathbf{k}}'\right) = -\hat{\mathbf{Q}}. \tag{6}$$

This coordinate system is also used to distinct the two possible directions of the linear polarization. ϵ_σ (from german 'Senkrecht') is used to denote the polarization, for which the electric field vector is perpendicular to the scattering plane, i.e. $\mathbf{E}_\sigma \parallel \hat{\mathbf{u}}_2$. ϵ_π (from german 'Parallel') refers to electric field vectors within the scattering plane.

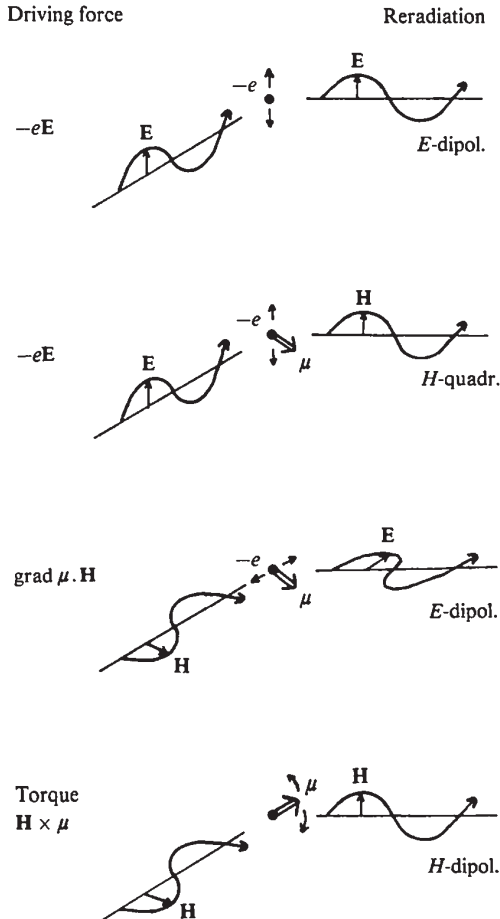## 2.2   Polarization dependence of the x-ray scattering cross section

Fig. 2: Classical picture of the scattering of x-rays by charged particle with a magnetic moment. From [2]

As the x-ray is an electromagnetic wave, it seems obvious, that it does interact with the charges and the magnetic moments in a sample, see Fig. 2, which sketches the scattering process as the acceleration of the charge/magnetic moment by the incoming wave and a reemission of the outgoing wave. The dominant process is clearly the classical Thomson scattering, where the reradiation is electrical dipole emission. As the dipole does not emit in the direction of the oscillating charge, it follows that the illumination with π polarized light depends on the scattering angle. For a scattering angle $2\theta = 90°$ the dipole oscillates exactly in the direction of the outgoing wave vector and hence does not re-emit in this direction. For σ polarized light the emission is perpendicular to the dipole axis and has hence full intensity. If we describe the effect on the polarization in the form of a matrix using as a basis σ polarized light as (1,0) and π polarized light as (0,1) we can express the action on the polarization in form of a matrix:

$$\begin{pmatrix} 1 & 0 \\ 0 & \cos 2\theta \end{pmatrix}. \tag{7}$$

In words: Illuminating a sample with σ polarized light does not affect the scattered intensity, while illuminating a sample with π polarized light results in an intensity modulation. This modulation is accounted for in the polarization part of the well known Lorentz-Polarization factor.

It turns out, that the other processes depicted in Fig. 2 are mainly relativistic corrections to this dominating process, which are at least 6 orders of magnitude weaker in the non-resonant case. Therefore higher multipole scattering effects remained rather a curiosity until the advent of dedicated synchrotron x-ray sources with a tremendous increase of brilliance, which allowed the detection of very weak scattering signals. Nowadays beamlines dedicated to magnetic or multipole scattering exist at all synchrotron facilities. In particular the resonant enhancements of this processes close to absorption edges of the elements has boosted this field and will be subject of a following lecture. Here we will concentrate on the technical aspects to control and analyze the x-ray polarization.

2.3   The synchrotron as a polarized x-ray source

In the seminal paper by de Bergevin and Brunel [2] they used the expression from eq. 7 to polarize and monochromatize the x-ray beam from an laboratory source with the respective loss in intensity. The x-ray emission from a synchrotron is instead already polarized to a very high degree. Modern x-ray beamlines at third generation synchrotron radiation sources feature a polarization very close to 100 %. The high degree of polarization is achieved as the electrons in the storage ring are basically only accelerated in the plane of the orbit. If we consider the emission within this plane (Fig. 3), the projected motion of the charge resembles an oscillating dipole in the orbit, yielding nearly perfect linear polarization in the orbit plane. Looking out of plane, the motion looks like an ellipse and hence the polarization becomes elliptical. In the lecture about synchrotron sources you will get the additional information, why synchrotron radiation is so brilliant which can also be understood from the sketch in Fig. 3.
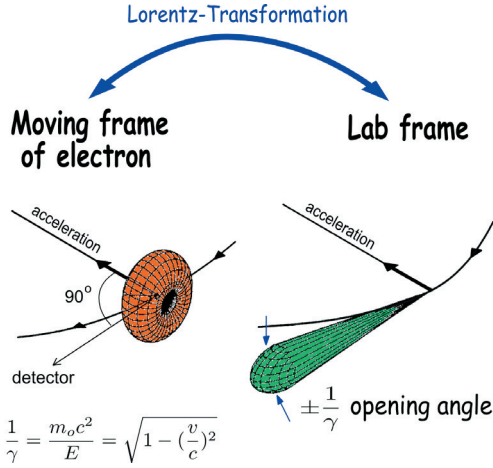
Fig. 3: Sketch of the x-ray emission from a synchrotron. The projected motion of the electron resembles a harmonic oscillation in the plane of the orbit. Performing the Lorentz transformation into the laboratory coordinate system, the emission is concentrated in the tangential forward direction. [3]

2.4 Manipulation of the beam polarization

To vary the polarization of the incident x-ray beam, we use the birefringence near a Bragg reflection, i.e. the difference in the index of refraction for σ light and π light. The phase shift is proportional to the path through the material d and the deviation from the Bragg angle $\Delta\theta$ [4]:

$$\Phi = -\frac{\pi}{2} \frac{r_e^2 \lambda^3 \Re(F_{hkl} F_{\bar{h}\bar{k}\bar{l}}) \sin 2\theta}{\pi V^2} \frac{d}{\Delta\theta} \tag{8}$$

with the classical electron radius r_e, the structure factors for the Friedel pair $(hkl), (\bar{h}\bar{k}\bar{l})$, the Bragg angle θ and V the volume of the unit cell. The Poincare-Stokes parameters achieved by the wave plate exposed to a beam with σ light are then given by [5]:

$$P_1' = -\sin\Theta \sin 2\chi \tag{9}$$

$$P_2' = \sin^2\frac{\Theta}{2} \sin 4\chi \tag{10}$$

$$P_3' = 1 + (\cos\Theta - 1)\sin^2 2\chi. \tag{11}$$

Here χ denotes the angle of the diffraction plane of the wave plate. For a suitable choice of Θ and χ on can achieve circular light from the incoming σ polarized light, i.e. by a quarter wave plate (QWP) one transforms $\mathbf{P} = (0, 0, 1)$ into $\mathbf{P}' = (1, 0, 0)$ while a half wave plate (HWP) turns $\mathbf{P} = (0, 0, 1)$ into $\mathbf{P}' = (0, 0, -1)$, π polarized light.

2.5 Analysis of the scattered beam polarization

To analyze the linear polarization of a scattered x-ray beam one uses again the expression eq. 7. So one looks for analyzer crystals, which feature a scattering angle $2\theta \approx 90°$, so that only σ polarized light can reach the detector. As you will learn in the lecture D-9 the resonant enhancement of the magnetic scattering request the use of a x-ray energy close to the absorption edge of a magnetic element. Therefor a variety of analyzer crystals exist to match absorption
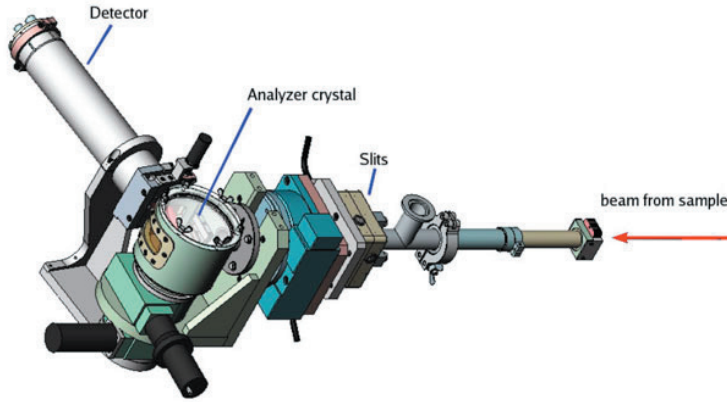
Fig. 4: 3d Model or a polarization analyzer. I contains three rotation axis: (i) the detector rotation, (ii) and analyzer crystal rotation to fulfill the Bragg condition and (iii) a rotation arond the axis defined by the beam from the sample to rotate the scattering plane of the analyzer arbitrarily. From [1]

edges. Examples for absorption edges of the elements, which are in particular important for magnetism, are given in tab. 1.

The polarization analyzer sketched in Fig. 4 has two rotation axes to fulfill the Bragg condition for the chosen analyzer crystal and x-ray energy. A third rotation axes allows the rotation of the scattering plane around the beam coming from the sample. Hence one can measure the linear polarization of the scattered beam for all directions. In particular one can determine the Stokes parameters P_2, P_3 and then via the relation $P_1 \leq 1 - P_3^2 - P_2^2$ also give a lower bound for the degree of circular polarization.

3 Polarization of neutron beams

The neutron is a spin $\frac{1}{2}$ particle with the magnetic moment

$$\mu = g_n \mathbf{S} \mu_N, \mu = -1.913 \cdot \mu_N, \mu_N = e\hbar/2m_P$$

The beam polarization is defined as the expectation value of the Pauli operators, i.e. orientation of the magnetic moment within a coordination system defined by a quantization axis parallel to the z-direction.

$$\mathbf{P} = \langle \sigma \rangle \tag{12}$$

with the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{13}$$

To determine the polarization for a specific direction, we measure the intensity of neutrons with spin aligned parallel, $I_\uparrow$, to the quantization axis $i = x, y, z$ and the intensity of the neutrons with spin aligned antiparallel to this quantization axis $I_\downarrow$. The polarization for this direction is then

$$P_i = \frac{I_\uparrow - I_\downarrow}{I_\uparrow + I_\downarrow}. \tag{14}$$

Crystal	(H K L)	d	$E(90°)$	Edges		
		[Å]	[keV]	$5f$ $M_{4,5}$	$4f$ $L_{1,2,3}$	$3d$ K
Au	(1 1 1)	2.355	3.72	U, Np		
Pt	(1 1 1)	2.266	3.87			
Cu	(2 0 0)	1.807	4.85			Ti
Graphite	(0 0 4)	1.677	5.22			Ti, V
Mo	(2 0 0)	1.574	5.57		La, Ce	V
Al	(2 2 0)	1.432	6.12		La, Ce, Pr, Nd	Cr, Mn
Cu	(2 2 0)	1.276	6.86		Sm, Nd, Eu	Mn, Fe
Au	(2 2 2)	1.177	7.44		Sm, Eu, Gd	Fe, Co
Al	(2 2 2)	1.169	7.49		Sm, Eu, Gd	Fe, Co
Pt	(2 2 0)	1.133	7.74		Sm, Eu, Gd, Tb, Dy	Co
Graphite	(0 0 6)	1.118	7.84		Eu, Gd, Tb, Dy, Ho	Co
Cu	(2 2 2)	1.042	8.41		Gd, Tb, Dy, Ho, Er, Tm	Ni, Cu
Pt	(4 0 0)	0.981	8.94		Tb, Dy, Ho, Yb	Cu
Pd	(4 0 0)	0.973	9.01		Tb, Dy, Ho, Yb	Cu
Graphite	(0 0 8)	0.839	10.48		Tm, Yb, Lu	Zn, Ga
Au	(3 3 3)	0.785	11.16			Ge

Table 1: Analyzer crystals that can be used in the hard x-ray range with the absorption edges for 'common' magnetic elements. From [1].

A parameter often used to qualify the beam polarization is the flipping ratio.

$$\text{FR} = \frac{I_\uparrow}{I_\downarrow}. \tag{15}$$

3.1   Polarization dependence of the neutron scattering cross section

In lecture A3 the interaction of neutrons with matter has been introduced. It was shown, that the neutron interacts with the atoms via nuclear and magnetic forces. The former is comprised in the nuclear scattering length, omitting details on the length scale of the nucleus. However, if the nucleus carries a spin momentum, the cross section for the neutrons depends on the orientation of the neutron spin with respect to the nuclear spin. As the nuclear spins are usually not correlated this gives rise to the nuclear spin incoherent scattering. If neutron spin and nuclear spin are parallel, the neutron spin is not changed, while for the two transverse directions the neutron spin is flipped. The polarization of the beam is therefor decreased to 1/3 and 2/3 of the intensity goes into the spin flip channel. The most prominent example of a spin incoherent scatterer is the proton with its huge neutron cross section. It gives rise to a large incoherent signal, which can be an annoying background, if one is interested in ordering phenomena in samples with a large hydrogen content, but it can also be useful to study local motions of hydrogen.

The field created by the magnetic moments in a sample exerts a force onto the neutron via its magnetic moment. For the interaction only the component of the magnetic moment, which is perpendicular to the momentum transfer or scattering vector, gives rise to scattered intensity, see Fig. 5. Considering again the coordinate system introduced in Fig. 1 it is clear, that the
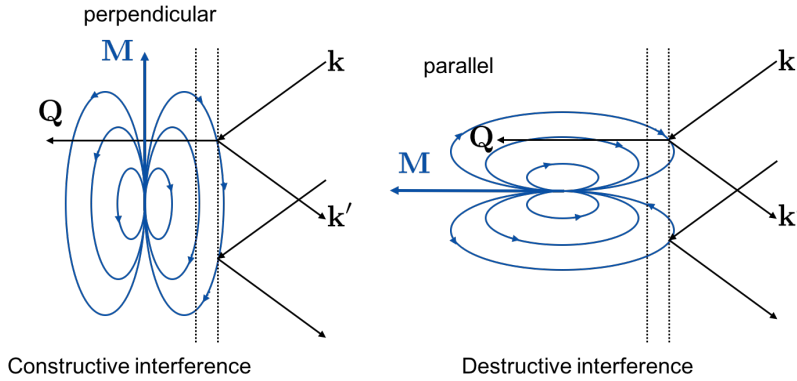
perpendicular

M

Q

k

parallel

k

M

Q

k′

k′

Constructive interference   Destructive interference

Fig. 5: Only the component of magnetic moment perpendicular to the momentum transfer give rise to constructive interference in scattering from wave vector $\mathbf{k}$ to $\mathbf{k}'$.

magnetic moments $\parallel \hat{\mathbf{u}}_3$ do not scatter the neutron. We may call this the first selection rule of neutron scattering, which relates the directions of the scattering vector and the magnetic moment in the sample. The so called magnetic interaction vector $\mathbf{M}_\perp(\mathbf{Q})$ contains only the components along $\hat{\mathbf{u}}_{1,2}$.

If we now take also into account the polarization of the neutron beam, we get another simple rule for the scattered polarization. This rule relates the directions of the interaction vector and the quantization axis for the neutron beam, which is given by an applied guide field. The matrix for the magnetic scattering is

$$A(\mathbf{Q}) = \langle S'_z | \sigma \cdot \mathbf{M}_\perp | S_z \rangle \Rightarrow \begin{pmatrix} M_{\perp,z} & M_{\perp,x} - iM_{\perp,y} \\ M_{\perp,x} + iM_{\perp,y} & -M_{\perp,z} \end{pmatrix} \tag{16}$$

Expressed in words: A magnetic moment parallel to the guide field or quantization axis does not flip the spin of the neutron. Any magnetic moment perpendicular to the quantization axis flips the neutron spin.

To become acquainted with this rules, I discuss a few examples: We start with an one dimensional case. If we consider a reflectivity experiment with a ferromagnetic sample, the scattering vector is parallel to the surface normal. Hence we probe the component of the magnetization within the surface. If the guide field is strong enough to saturate the sample, all moments are parallel to the quantization axis and the entire reflectivity does not flip the spin. So from the spin flip intensity we can draw conclusions about the anisotropy in the sample or how the magnetization rotates upon field reversal.

In Longitudinal Polarization Analysis you measure the spin dependent cross sections for the three directions $\hat{\mathbf{u}}_i$, resulting in 12 measurements. Often only 6 measurements are performed, namely the $\langle \uparrow' |, | \uparrow \rangle$ and $\langle \downarrow' |, | \uparrow \rangle$ cross sections, called Non-Spin-Flip (NSF) and Spin-Flip (SF) cross sections. In the seminal paper from Moon, Riste and Koehler [6] you find a very comprehensive and accessible presentation of the potential of this method.

Finally you might study the rotation of the spin polarization from any polarization along the directions $\hat{\mathbf{u}}_i$ into the direction $\hat{\mathbf{u}}_j$. This method is called spherical polarization analysis. As it

measures all 9 components of the polarization matrix, it requires the measurement of 36 (often only 18) cross sections. This allows then basically to reconstruct the full vector information of the magnetic interaction vector, which is often the only way to distinguish competing models of complex magnetic structures. Again we refer to the seminal papers from Blume [7] and Maleev [8] for more details.

3.2   How do we polarize a neutron beam?

The neutrons, which are released from a moderator are unpolarized. We use polarization dependent scattering and absorption processes to select a neutron spin state. In the former case, the scattering potential vanishes for one spin direction and hence only neutrons with the opposite spin state are deflected, while the momentum of the other spin state remains unchanged. In the subsequent paragraphs we describe devices made by artificial heterostructures and single crystals

The second class of polarizers uses the spin dependence of the nuclear reactions to select a spin state. We restrict ourselves to the case of ^3He, for which the absorption cross section varies by orders of magnitude, if the spin state of the neutron is parallel or antiparallel to the spin of the ^3He nucleus.

Supermirror devices: For slow neutrons one uses devices made of so called supermirrors. A supermirror features (nearly) total reflection of neutrons, when the neutron impinges on the surfaces from the vacuum or air side. They are made up from artificial superlattices with atomically flat inter faces. For a specific superlattice periodicity D, neutrons that fullfill the Bragg condition

$$\lambda = 2D \sin \theta \tag{17}$$

are strongly reflected. By combination of a range of different periodicities, one can accordingly reflect a certain range of angles θ and wavelength λ. Originally Ni coated glasses were used to build mirrors for neutrons, as it is the element with the largest neutron scattering length and hence has the largest critical angle for total reflection according to the index of refraction

$$\theta_c = \lambda \times \sqrt{\frac{Nb}{\pi}} \tag{18}$$

The reflectivity of such a supermirror Bragg peak depends on the contrast between the constituent layers. To become sensitive on the neutron polarization we use the fact, that the magnetic scattering length density depends on the relative orientation of the neutron spin and the magnetization of the layer. If the spin is parallel to the layer magnetization, the nuclear and magnetic scattering length densities add up, whereas for an antiparallel orientation they are subtracted. Choosing the right materials, the contrast between the layers in the superlattice vanishes for one spin direction, and for the respective spin direction the beam is not refracted at this interface. Therefore the reflected beam is polarized up, whereas the transmitted beam is polarized down.

A variety of different devices is build from polarizing supermirrors. For a low divergent beam, polarizing cavities are used. As the beam impinges on the plane mirror within a narrow angular
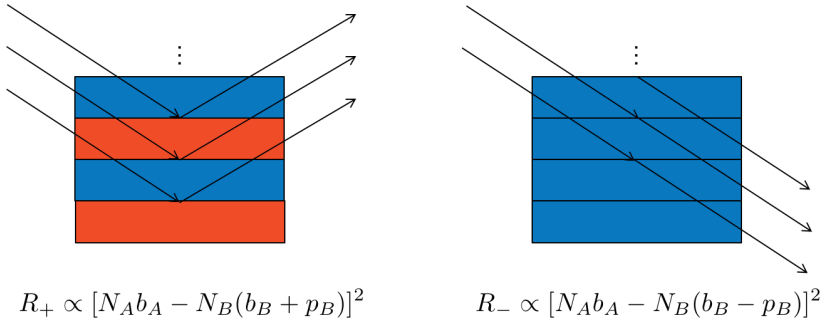
$$R_+ \propto [N_A b_A - N_B(b_B + p_B)]^2 \qquad\qquad R_- \propto [N_A b_A - N_B(b_B - p_B)]^2$$

Fig. 6: The reflectivity of a supermirror with FM and non-magnetic layers. For the $|\uparrow\rangle$ states the contrast is increased, while for the $|\downarrow\rangle$ states the contrast vanishes and hence these neutrons are not reflected.

distribution, the reflectivity for the entire beam is very similar and hence a transmission device provides both a high transmission and polarization in the transmitted beam, while a reflection of the wrong spin state only reduces the intensity but not the polarization. If the incoming beam has a broad collimation, the situation is the opposite. The momentum transfer averaged over the beam is rather high and the probability for a wrong spin state to end up in the reflected beam is very low. Typical applications for these devices are benders installed after the sample to analyze the polarization within a large solid angle or devices between a focusing monochromator and the sample.

Ferromagnetic single crystals   Following a similar principle, ferromagnetic monochromator crystals can also be used to polarise a neutron beam. In this case the structure factor for a specific Bragg reflection vanishes for one spin direction and becomes ideally large for the other spin direction. Hence only this spin direction is diffracted. Several Heusler alloys feature this property, see tab. 2. Such devices combine the functions of monochromatizing and polarizing. The peak reflectivity of these crystals is however low compared to the non-polarizing monochromator crystals also shown in the table.

Both types of devices require saturation of the ferromagnets, as any domain structure will lead to an incomplete polarization. Therefore the are enclosed in strong magnetic fields, which increases the complexity of the device.

^{3}He spin filter cells   The isotope ^{3}He features strongly different absorption cross sections for thermal neutrons, if the spin of the neutron and the spin of the ^{3}He nucleus are aligned parallel or antiparallel:

$$\sigma_{\uparrow\downarrow} = 6000\,\text{barn}, \sigma_{\uparrow\uparrow} = 5\,\text{barn} \qquad (19)$$

If the neutron beam travels through a gas volume with a large enough number of polarized ^{3}He nuclei the transmitted beam is mainly polarized in the direction of the ^{3}He polarization. To achieve a sufficient neutron polarization either the neutron path through the volume must be long enough or the pressure of the polarized ^{3}He must be sufficiently high. Since also the

polarising	refl.	d (Å)	d (Å)	refl.	non-polarising
Cu_2MnAl "Heusler"	111	3.43	3.35	002	graphite
Fe_3Si	111	3.27	3.35	002	graphite
$^{57}Fe:Fe$	110	2.03	2.08	111	copper
$Co_{0.92}Fe_{0.08}$	200	1.76	1.80	200	copper
$HoFe_2$	620	1.16	1.28	220	copper

Table 2: Lattice spacing of polarising and non-polarising Bragg reflection used for monochromators.

absorption cross section varies linearly with the neutron wavelength λ, the spin filter cells are qualified by the product $p_{^3He} \cdot l \cdot \lambda$, called opacity.

While the principle of such a device is rather simple, the technical realisation of a polarized ^{3}He gas volume is technically challenging. The ^{3}He can be polarized via two processes.

Metastability Exchange spin Optical Pumping (MEOP):   A small fraction of the ^{3}He gas is electronically polarized by staged discharge and optical pumping processes. Via metastability exchange collisions the polarization is transferred to the nuclei, that remained in the $J = 0$ electronic state. These processes must be performed at a pressure $p \approx 1$ mbar. The polarized gas is then compressed by special non-magnetic compressors to fill the filter cell with the required opacity. This process is performed offline and the filled cells are brought manually to the neutron beamlines. As the ^{3}He polarization decays with time, the cells must be changed regularly, typically once per day or two days.

Spin Exchange Optical Pumping (SEOP):   Here the cell is filled with K and Rb vapor. Rb atoms are polarized by optical pumping using a circularly polarized laser. The electronic spin is exchanged to the K atoms, which couple to the ^{3}He atoms. These processes can be performed at the working pressure of the filter cell and therefor cells have been developed, where the polarization is maintained continuously. These requires a special glas, GE 180, which is transparent for the infrared laser light and doesn't disturb the ^{3}He polarization by wall collisions. This glass is difficult to process and the production of good cells is an art and a bit of alchemy. The continuous pumping results in high degree of ^{3}He polarization, in particular for the time average. So if the space for the optical system and the oven is available, such a system is very attracting.

Polarizing the neutron beam via absorption processes does not interfere with the beam divergence. Therefore it is in particular suitable for an application in the scattered neutron beam, e.g. on reflectometers or small angle scattering instruments. Also for neutrons in the wavelength range < 2 Å ^{3}He filter cells are often the method of choice, as the critical angle for total

reflection becomes very small.

3.3   Manipulation of the neutron beam polarization

Due to its magnetic moment the neutron is subject to magnetic fields. Classically the Bloch equation of motion describes the time evolution of the magnetic moment.

$$\dot{\mu} = \gamma\mu \times \mathbf{B}, \tag{20}$$

with the gyromagnetic ratio $\gamma = 2\mu_n/\hbar$. In a constant magnetic field the moment precesses around the field direction with the Larmor frequency, as the torque is always perpendicular to the moment:

$$\omega_L = -\gamma B_z \approx 3\mathrm{kHz}\,\mathrm{Oe}^{-1} \cdot B_z \tag{21}$$

Pictorially we may imagine this as spinning top, which is stable, when the rotation frequency is high enough. Therefore we need a minimum value of a guide field B_z to maintain the beam polarization. Typically guide fields have a magnetic induction on the order of a few mT. Otherwise any weak field, e.g. the earth field will exert a torque onto spin or in our picture of the spinning top it will start to tumble.

Moving in spatially varying fields, the field acting on the neutron becomes time dependent in the neutron frame of reference. If we assume a field configuration that changes direction from vertical to horizontal transverse to the beam propagation, one can use the Bloch equation of motion to calculate the field rotation frequency

$$\omega_B = \frac{\partial\theta}{\partial x}v_x \tag{22}$$

for a field rotation perpendicular to the propagation direction x. θ denotes here the angle between the quantization axis and the actual field direction. The relation between these frequencies is used to distinct two different cases to manipulate the neutron spin. For slow field variations, the field rotation frequency is much smaller than the Larmor frequency

$$\omega_B \ll \omega_L. \tag{23}$$

The field change is then called adiabatic, as the potential energy of neutron stays always close to the initial state. Adiabatic field changes are typically realized, when the neutron travels from the vertical guide field regions into a field region created by Helmholtz coil pairs. To make sure, that the neutron moment follows the field direction, $\omega_L/\omega_B > 10$ must be always fulfilled. As the field rotation frequency scales linearly with the neutron velocity, thermal neutrons need stronger guide fields or slower rotations as cold neutrons

Flipping the spin state is essential for the measurement of the beam polarization. It is equivalent to rotate the polarization through an angle π. If a very sudden field change is realized, e.g. separating two regions of opposed magnetic fields by a superconducting Meissner sheet or a current sheet, the neutron cannot follow the field rotation. Therefore it remains in the higher energy state or in other words a $|\uparrow\rangle$ state has become a $|\downarrow\rangle$ state and vice versa. So with respect to the guide field, traversing through the sheet had the effect of a spin flipper. In a radio-frequency flipper you superimpose a static magnetic field of variable size and a field oscillating with the

Larmor frequency defined for the static field in the center of the device. In a reference frame rotating with the Larmor frequency the static field apparently vanishes in the center as the spin does not precess, while the oscillating field becomes a static longitudinal guide field. Outside the center the transverse field changes the sign and therefore in the rotating frame of reference the spin rotates adiabatically from $|\uparrow\rangle$ to $|\downarrow\rangle$.

In a Mezei flipper, the field on the inside of the flipper is perpendicular to the guide field outside. Within the flipper the neutron spin starts to precess according to the time-of-flight inside:

$$\phi = \omega_L \frac{d}{v} \tag{24}$$

Adjusting the field of a coil, the neutrons precess by an angle ϕ by travelling through the field region. As the time-of-flight is different depending on the velocity of the neutron, a π flip is reached only for a narrow velocity distribution. Therefore such a device is very well suited for a monochromatic instrument, but not for wide band instruments typically used at pulsed neutron sources. Such a device can also be used to rotate the polarization only by an angle $\frac{\pi}{2}$. This is needed in spin-echo methods where the precessing angle of the neutron spin in a magnetic field before and after interaction with the sample provides ultimate resolution information about dynamics or structure (see lectures B4 and D6).

4 Common formalism to treat x-ray and neutron polarization

While the beam polarization for neutrons and x-ray photon describe different physical properties, the formalism for the theoretical treatment is similar. For both cases a density matrix formalism has been derived to describe the polarization of the beam, the action of the different devices and the interaction of the beam with the sample [2,7–11]. In both cases the density matrix ρ has dimensions 2×2 as the Hilbert space of the spin states is 2 dimensional and also the Electric field vector has only 2 independent components. It turns out, that the relation between the density matrix and the polarization vector/Poincare-Stokes parameters is the same:

$$\rho = \frac{1}{2}(1 + \mathbf{P} \cdot \sigma). \tag{25}$$

It should however be noted, that this identity is only formal, as the Poincare-Stokes parameters do not transform as an axial vector like the neutron polarization. However, the formalism to treat the polarization in a scattering experiment is similar, which is also evident from the fact that two seminal papers on the topic of x-ray and neutron polarization are both written by M. Blume [7, 10].

The general properties of the density matrix allow to calculate the polarization before and after scattering by the simple expressions

$$\mathbf{P} = \mathrm{tr}(\sigma\rho) \tag{26}$$

$$\rho' = M\rho M^\dagger \tag{27}$$

$$\frac{d\sigma}{d\Omega} \propto \mathrm{tr}(\rho') \tag{28}$$

$$\mathbf{P}' = \frac{\mathrm{tr}(\sigma\rho')}{\mathrm{tr}(\rho')} \tag{29}$$

The matrix elements of M describe the (forward) scattering of any optical element or sample for the different polarization channels. In particular for the interaction of the beam with the sample they contain the different scattering cross sections. For a more detailed discussion of the formalism, please consult [1, 7, 8, 10].

5   Conclusion

Analysing the polarization state of the scattering probes enhances the information content of an experiment. On one side the conservation of angular momentum yields information about the angular momenta inside the sample and hence about the full vector properties of the moment distribution inside the sample. Another crucial advantage comes from the selectivity in the different channels, which allows the measurement of a weak signal next to much stronger scattering processes. Examples include the magnetic x-ray scattering in general, but also magnetic scattering from thin films, which are often dominated by the background scattering from the non-magnetic substrate. The cost you have to pay in terms of longer measurement times might be more then compensated by the improved signal to noise ratio.

In recent years, devices and/or sources have been improved to provide nowadays much stronger polarized beams than in the past. Back then users were often afraid to loose to much intensity to make succesful experiments with polarization analysis. I hope, I could convince you that this is not the case any more and that the very solid theoretical foundation of the method can allow to draw more stringent conclusions from your experiment in particular when you are concerned with magnetic moments or other multipoles in your sample.

References

[1] C. Detlefs, M. Sanchez del Rio, C. Mazzoli, *X-ray polarization: General formalism and polarization analysis*, The European Physical Journal Special Topics **208**, 359 (2012).

[2] F. de Bergevin, M. Brunel, *Diffraction of X-rays by magnetic materials. I. General formulae and measurements on ferro- and ferrimagnetic compounds*, Acta Crystallographica Section A **37**, 314 (1981).

[3] http://photon-science.desy.de/research/students_teaching/primers/synchrotron_radiation/index_eng.html (6.12.2018).

[4] C. Giles, C. Vettier, F. de Bergevin, C. Malgrange, G. Grübel, F. Grossi, *X-ray polarimetry with phase plates*, Review of Scientific Instruments **66**, 1518 (1995).

[5] L. Bouchenoire, S. D. Brown, P. Thompson, J. A. Duffy, J. W. Taylor, M. J. Cooper, *Performance of phase plates on the XMaS beamline at the ESRF*, Journal of Synchrotron Radiation **10**, 172 (2003).

[6] R. M. Moon, T. Riste, W. C. Koehler, *Polarization Analysis of Thermal-Neutron Scattering*, Phys. Rev. **181**, 920 (1969).

[7] M. Blume, *Polarization Effects in the Magnetic Elastic Scattering of Slow Neutrons*, Phys. Rev. **130**, 1670 (1963).

[8] S. V. Maleev, V. G. Baryakhtar, R. A. Suris, *The Scattering of Slow Neutrons by Complex Magnetic Structures*, Sov. Phys.: Sol. State **4**, 2533 (1963).

[9] M. Brunel, F. de Bergevin, *Diffraction of X-rays by magnetic materials. II. Measurements on antiferromagnetic Fe_2O_3*, Acta Crystallographica Section A **37**, 324 (1981).

[10] M. Blume, D. Gibbs, *Polarization dependence of magnetic x-ray scattering*, Phys. Rev. B **37**, 1779 (1988).

[11] J. P. Hill, D. F. McMorrow, *Resonant Exchange Scattering: Polarization Dependence and Correlation Function*, Acta Crystallographica Section A **52**, 236 (1996).

C 5   European XFEL

Robert Feidenhans'l
Managing Director
European XFEL
Schenefeld/Hamburg

Extended Abstract

Extended abstract

X-ray science has over the last decades undergone a tremendous revolution in particular due to the development in X-ray sources. In the last ten years hard X-ray lasers have had a major impact in the field of X-ray Science and this will accelerate in the coming years as even more intense sources like the European X-ray will become available to a broader user community.

The European XFEL started first user operation September 2017. It is the world's first hard X-ray laser facility based on superconducting accelerator technology and will deliver an unprecedented X-ray beam to the user community with a high repetition rate. The outline of the facility is shown below. More details can be found in ref. [1].
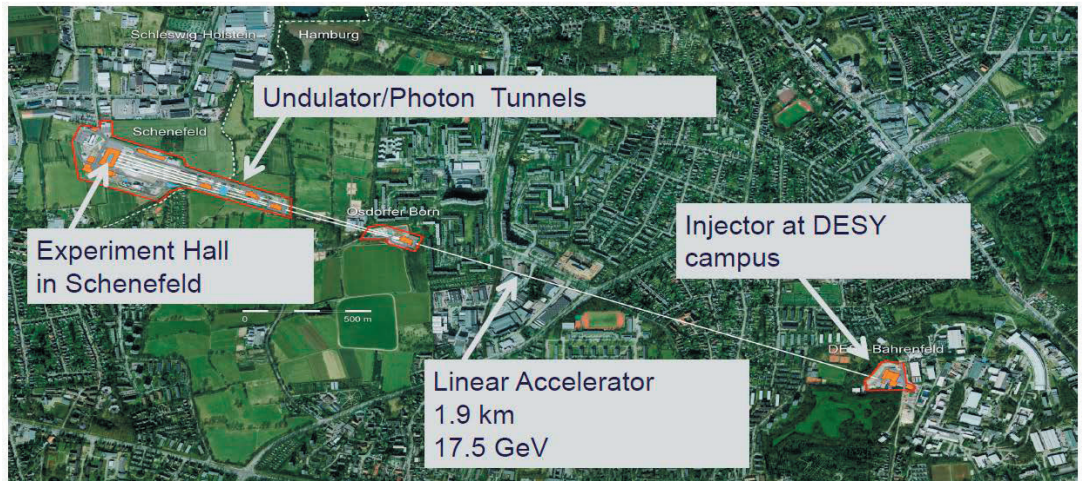

Figure 1. Outline of European XFEL

In contrast to synchrotron radiation sources X-ray Free Electron Lasers are based on linear accelerators that accelerate the electrons up to 17.5 GeV in the case of European XFEL. The electron injector is based at the DESY Campus. The linear accelerator is 1.9 km long and based on superconducting technology. This is important in order to achieve a high repetition rate of the X-ray pulses, which again allows a high data collection rate. The linear accelerator is placed in a long tunnel below ground that connects the DESY campus with the campus at Schenefeld over a distance of 3.5 km.

The X-ray pulses are generated by three undulators based on the principle of Self-Amplified Spontaneous Emission (SASE) and the undulators are hence called SASE1, SASE2 and SASE 3. In comparison to synchrotron radiation sources the X-ray pulses from X-ray free electron lasers are much more intense, much shorter duration and less abundant as shown in the sketch below. The pulse duration is so short, of the order 10-100 femtosecond, that atoms

are hardly moving on that time scale and hence 'still picture' of the atoms in principle can be taken. Of course this is only possible if the pulses are sufficient intense, but as illustrated in the sketch the pulses contain about 10^4 more photons in a single pulse than from a synchrotron radiation source. On the other hand, the pulses from the X-ray free electron laser are much less frequent than from a synchrotron radiation source. Free electron lasers sources based on normal conducting accelerator technology can only produce about 60-120 pulses per second in comparison to a synchrotron radiation sources with about 10^6 pulses per second. The superconducting technology of the accelerator at European XFEL will be able to deliver up to 27000 pulses pr sec in a so-called burst mode with 2700 pulses in a train of total duration 600 µsec with trains 10 times a second. This gives a possibility of a much higher data collection rate that at the normal conducting facilities.
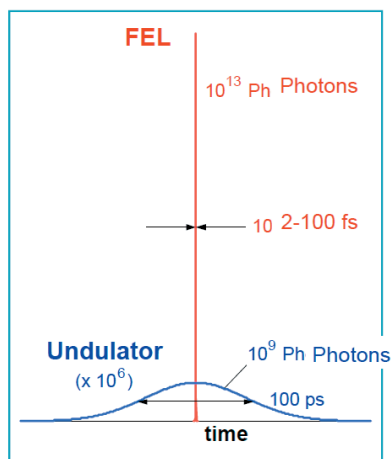

Figure 2. Sketch of the difference between X-ray pulses from a X-ray free electron laser and a synchrotron radiation source

The x-ray pulses generated by the undulators are guided to the experimental hall through long photon transport systems onto the scientific instruments. A layout of the experimental is shown below. At Osdorfer Born (see Fig. 1) the tunnel is branched in into two tunnels leading to the undulators SASE1 and SASE2. After SASE1 there is a further branch leading to the SASE3 undulator. Furthermore there are two empty tunnels that have not been filled with instrumentation yet.
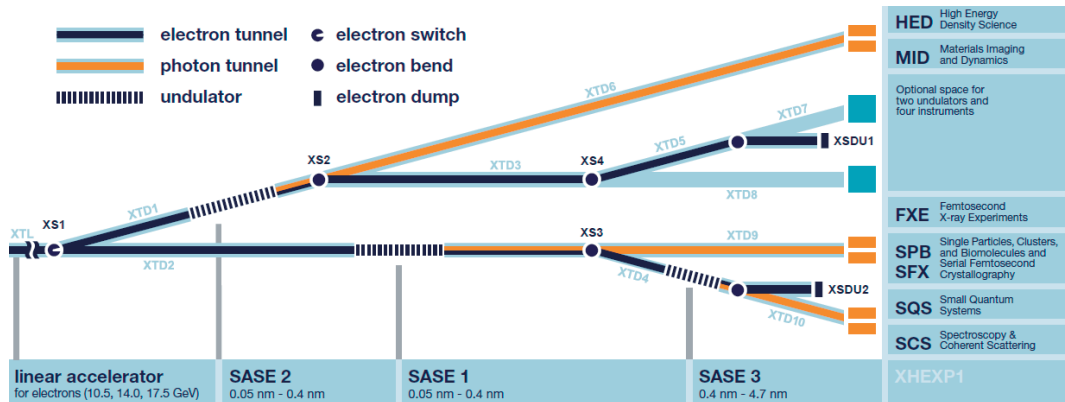
Figure 3. Layout of the experimental hall at European XFEL

The first instruments that became available to users in September 2017 were the FXE (Femtosecond X-ray Experiments) instrument for ultra-fast x-ray spectroscopy and x-ray scattering with a primary focus on photo-induced chemical reaction in liquids and the SPB/SFX (Single Particle, Biomolecules and Serial Crystallography) instrument for diffractive three dimensional imaging and structural determination for single particles, clusters and biomolecules. Both the instruments are based on the SASE1 radiation source and they share the radiation source daily on a 12h basis. Since September 2017 26 user experiments have in total been conducted on the two instruments.

In late 2018 two more instruments were taken into user operation covering namely the SQS (Small Quantum Systems) instrument dedicated to investigations of light-matter interactions in the soft x-ray regime and the SCS (Spectroscopy and Coherent Scattering) instrument dedicated to electronic, spin and structural properties in materials. Only very few experiments were conducted so far and first examples of the results might be presented in the talk.

Finally the last two instruments will be taken into operation in the spring of 2019. This are the MID (Materials, Imaging and Dynamics) with the focus to perform ultrafast imaging and study dynamic of materials with a particular focus on the application of coherent X-ray scattering and diffraction. The last instrument is the HED (High Energy Density) instruments with aims at providing capabilities to study matter at extremes states of temperature, pressure and densities including conditions in solar or extra-solar planets.

In the talk I will give examples of the some of the first experiments that have been conducted at European XFEL

References

[1]   *Appl. Sci.* **2017**, *7*(6), 592; doi:10.3390/app7060592

C 6 ESS

A. Schreyer

European Spallation Source ESS ERIC

Lund, Sweden

Please refer to:

www.esss.se

- The Organisers -

C 7   Accelerator Based High Brilliance Neutron Sources

T. Gutberlet, U. Rücker
Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

Introduction

Neutrons are an essential tool in science and innovation for probing the structure and dynamics of matter from the mesoscale to the nanoscale and from seconds to nanoseconds. In Europe science and research benefit from a globally unique landscape of various neutron sources with the flagship facilities ILL in Grenoble, France, MLZ in Garching, Germany, and ESS in Lund, Sweden, which is currently under construction and will represent the world's most powerful neutron facility. The unique capabilities of neutrons and the existing European infrastructure were highlighted recently in the report by the ESFRI Neutron Landscape Group [1].

Neutrons can be produced by fission in nuclear reactors, by spallation using high-power proton accelerators, and by nuclear reactions with low-energy proton accelerators. While the first two techniques can offer the highest total neutron current production, modern state-of-the-art accelerator technology offers the opportunity for a new landscape of novel and unique high-brilliance neutron sources based on low-energy proton accelerators. This option has not yet been exploited in neutron science in Europe. In Japan a few small compact accelerator based neutron sources exist [2,3], as well as the LENS facility in the US [4] and the IREN source in Russia [5].

Most of the small and medium-sized regional neutron facilities in Europe were established in the 1960s to 1970s in the course of research on the peaceful use of atomic energy. A number of these facilities have been closed in Denmark, Sweden or Germany in recent years or will face closure in the near future such as the BER II reactor in Berlin or the Orphée reactor in Saclay, France. Against this backdrop in neutron sources a novel concept is required to avoid a loss of access to neutrons and to maintain competitiveness in Europe [6]. Accelerator based neutron sources provide a modern, versatile and efficient opportunity to improve and spread neutron access in Europe and also a new route for supplying neutrons with leading-edge, but still affordable, research infrastructures directly to science and industry.

1   Bringing neutrons to the users

With the construction of the ESS, the European neutron user community is eagerly awaiting the commissioning of the brightest neutron source worldwide. At the same time, there is, however, the ongoing development of neutron science being undertaken at a dwindling number of neutron facilities worldwide and particularly in Europe [1]. These flagships of neutron science serve the needs of a limited number of experienced researchers, but the smaller or medium flux sources used for method development, user recruitment, education or proof-of-principle experiments are slowly vanishing. Responding to this situation compact accelerator-driven high-brilliance neutron sources as an efficient and cost effective alternative to current low- and medium-flux reactor and spallation sources are considered. Such compact

sources offer to science and industry access to neutrons as local national or regional medium-flux, but high-brilliance neutron facilities.

Based on high current accelerators with pulsed proton or deuteron beams compact neutron sources allow optimized neutron instruments to be realized, with unprecedented brilliance amplified by the target-moderator unit as part of the specific instrument design. Compact quasi one-dimensional, finger-like thermal and cold moderator systems are used to boost the cold neutron flux and offer the highest level of brilliance. Neutronic calculations show the possibility of competitive neutron fluxes of such high brilliance neutron sources. The proton beam can be distributed onto various target stations allowing the different requirements in pulse width and frequency of dedicated neutron instruments to be met.

By shaping the experiment from the source to the instrument detector, a dedicated holistic neutron experiment is set-up, and can fulfil many varied scientific requirements on demand in a flexible and efficient way. It is possible to develop unique instrumental concepts at such compact sources.

The full design allows for the scalable operation of the compact source, capable of meeting the needs of a small laboratory in a university or in industry, right up to those of a large research facility. This will offer a significant increase in access to neutrons, even after some reactor sources have been shut down, and will greatly improve the capability to exploit neutron applications in a multitude of research areas as well as in applied science.

Several projects have been launched in France [7], Spain [8], Italy [9] and Germany [10] to establish and build state-of-the-art accelerator based neutron sources. The most ambitious project is conducted at Forschungszentrum Jülich for a high brilliance accelerator driven neutron source as a large scale facility on the level of a medium size neutron source [10].

2   Accelerator based neutron sources

Early work at accelerator-driven sources was based on cyclotrons and pulsed electron-linac-driven bremsstrahlung photo-neutron sources. In these sources the gamma rays produced in the bremsstrahlung process go on to produce neutrons by photo-neutron reactions. As in the early 1970s proton beam power exceeded the 100-MeV energy level, accelerator-driven spallation neutron sources have since come to the fore. Spallation is the process induced by high energy particles delivered by accelerators (around GeV protons) striking nuclei in a massive target, which promptly emits neutrons (and other radiation) to cool off. Pulsed operation alleviates the heat load averaged over time yet the instantaneous power and neutron flux are very high.

The pulses from accelerator-driven short-pulse spallation sources (SPSSs) are short enough to differentiate the neutron flight times of events corresponding to most neutron scattering requirements. Today's high-power SPSSs include the MW-level sources, i.e., the Spallation

Neutron Source (SNS) in the US and the Japan Spallation Neutron Source (JSNS of J-PARC), and the sub-MW-level sources, i.e., ISIS in the UK. Additionally, a MW level steady-state spallation source, SINQ, has been in operation for more than twenty years in Switzerland and the new European Spallation Source (ESS), is currently under construction in Sweden.

Recent progress in accelerator technology and neutronic design has made it possible to construct small-scale accelerator-driven neutron facilities that could play a significant role in neutron technology and science [11]. Various neutron applications using compact accelerator-driven neutron sources (CANS) are becoming more and more important as they can contribute to various fields such as material science, engineering, nuclear physics, cancer therapy, electronics and so on.

Until now only a few small facilities have actually been built and are operating scattering instruments. Facilities such as HUNS in Japan [2] or the Bariloche LINAC in Argentina [12] started operation during the 1970'. While the HUNS facility is still in operation and has been upgraded recently, the facility in Bariloche has stopped operation in 2017. Additional facilities were built with LENS in the US [4] and RANS in Japan [3]. These CANS are all designed to run at power levels of some kW with user programs tailored to specialized or multipurpose applications.

The challenge to establish a competitive high brilliance neutron source is to design and operate a high current proton accelerator with beam current up to 100 mA and proton energies up to 100 MeV.

2.1 Low energy neutron processes

Nuclear reactions can produce neutrons by bombardment of a target material with low to medium energy proton (p) or deuterium (d) beams in the range of some MeV up to 100 MeV. Most efficient target materials used are light elements as lithium (Li) or beryllium (Be) or various metal targets as vanadium (V), tantalum (Ta), or tungsten (W).

The nuclear process on bombarding a light element as Be with an accelerated high energy proton beams knocks out one neutron out of the Be atom (5 neutrons, 4 protons) and produces a boron (B) isotope (4 neutrons, 5 protons) and release of energy as γ-rays. The ^{9}Be(p,n)^{9}B charge-exchange reaction is given as

$$p + {}^{9}_{4}Be = {}^{9}_{5}B + n + \gamma$$
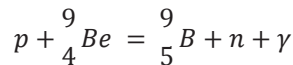
Their efficiency is around 10^{-2} n/p (or per d) with a heat release of around 1000 MeV per neutron. A similar production of neutrons is achieved using metal targets as in the ^{181}Ta(p,n)^{181}W reaction (an overview on neutron production for these elements is given in

Tab.1 and Fig. 1). For the notation of nuclear reactions see en.wikipedia.org/wiki/
Nuclear_reaction.

Nuclear process	E [MeV]	n/ion	n/(s mA)	n/(s kW)
p → Be	50	2.70%	1.68E+14	3.37E+12
d → Be	50	5.90%	3.69E+14	7.38E+12
p → Li	20	0.33%	2.08E+13	1.04E+12
p → V	50	5.08%	3.18E+14	6.35E+12
p →Ta	50	6.40%	4.00E+14	8.01E+12
p → W	50	6.95%	4.35E+14	8.70E+12

Table 1. Comparison of the primary neutron production efficiencies of different incident ions
and target materials. Analytic calculations with protons and deuterons are performed using the
TENDL 2017 data library. The neutron production is normalized to neutron/second and
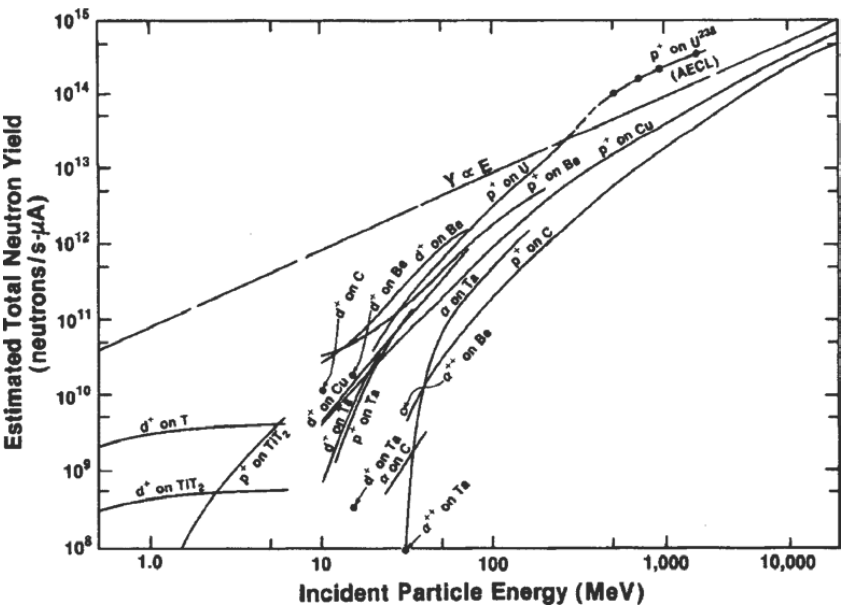milliamp and neutron/second and kilowatt.

Fig. 1: Estimated global neutron yields from nuclear reactions of particles incident on thick
targets of various materials [J.M. Carpenter, C.-K. Loong, Elements of Slow-Neutron
Scattering, Cambridge University Press, 2015].

While much less efficient than spallation, these processes allow the construction and
operation of compact neutron sources. With a volume of only a few cm^3 a very small

interaction volume can be exploited which allows a compact access to the neutron source field. By optimizing the entire setup from ion source to neutron detector, these sources become very competitive to reactor based or spallation medium flux neutron sources.

Accordingly (obtained on the basis of the TENDL nuclear data library and MCNPx) the most efficient nuclear interaction for neutron emission is the nuclear stripping effect with low energy deuterons (d, n) when using Be as target material. The break-up reaction (d, n+p) also contributes significantly to the neutron release with the maximum yield at 30 MeV deuteron energy.
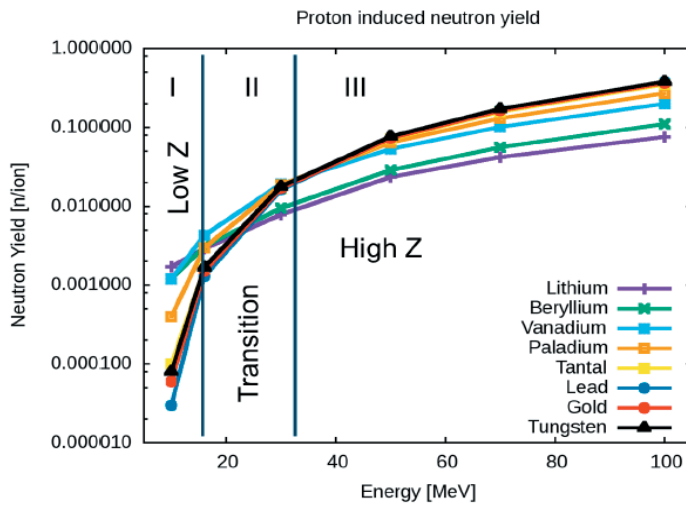

Fig. 2: Neutron yield (n/ion) for different materials for a proton beam as a function of the particle energy.

Figure 2 shows the neutron yield (normalized to the number of neutrons per ion) as a function of proton beam energy for different target materials. Heavy elements, especially gold or tungsten, are obviously more suitable as target materials. Beyond a certain beam energy (approx. 35 MeV) the yield starts to become saturated.

2.2 What is brilliance?

The brilliance of single neutron beams (B) can be defined as:

$$B = \frac{neutrons}{s \cdot cm^2 \cdot mrad^2 \cdot 1\%\frac{\Delta\lambda}{\lambda}}$$

It is given by the neutron current (number of neutrons emitted per second), normalized to the source area (cm^2), the solid angle ($mrad^2$) and the wavelength spread at which the neutrons are emitted.

The best brilliance for a specific experiment is achieved by an optimized setup starting from a dedicated target-moderator unit for each individual neutron instrument. Thus, the number of "useful" neutrons for the experiment in question is maximized, while minimizing the number of neutrons which have to be stopped in shielding. High brilliance compact accelerator driven neutron sources aim at maximizing this number. In particular they will minimize the effective source area and the solid angle at which neutrons are emitted and transported to the experiment as shown schematically in Fig. 3. By this a very small interaction volume with spherical diameter of some ten centimeters allows direct access to the neutron source field. Such an optimised setup offers a unique and flexible template for individual experimental stations and local facilities with high brilliant neutron beams.
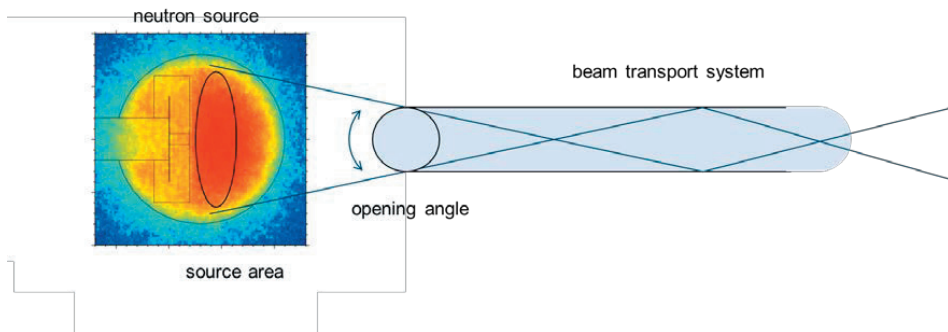

Fig. 3:   Schematic drawing of setup of neutron source and beam transport system to optimize and achieve high brilliant neutron beams.

Due to the low energy proton beam (<100 MeV) and the small interaction volume a very compact moderator and reflector unit is possible which allows to build a neutron target system with dimensions of only a few m^3 including the biological shielding.

3   High brilliance neutron source

3.1   Facility outline

One main advantage of this new type of compact neutron sources is the fact that they are scaleable: from smaller sources with a few dedicated applications, all the way to competitive medium flux sources scientifically comparable to our present day facilities at research reactors serving a full instrument suite. Using a multiplexing system the accelerated pulsed

proton beam is distributed to a number of target stations. Each target station can provide hot, thermal or cold neutrons by a number of channels to several neutron instruments. In particular different source repetition rates can be realized, tailored to the needs of individual instruments. Depending on the organisation of the target stations and the corresponding instruments a variable number of experimental setups can be offered (Fig. 4).

Main components of the HBS are
- a dedicated proton or deuteron accelerator with optimized energy (10 to 100 MeV) and current (up to 100 mA) that is multiplexed to feed several target stations that are operated at different frequencies,
- a target-moderator combination that offers a pulsed neutron beam at optimal frequency, pulse duration and neutron spectrum to fulfil the needs of a single instrument,
- optimized moderators according to the spectral needs of individual instruments,
- a number of instruments at each target station (typically up to 6, in order not to dilute the thermal flux maximum).
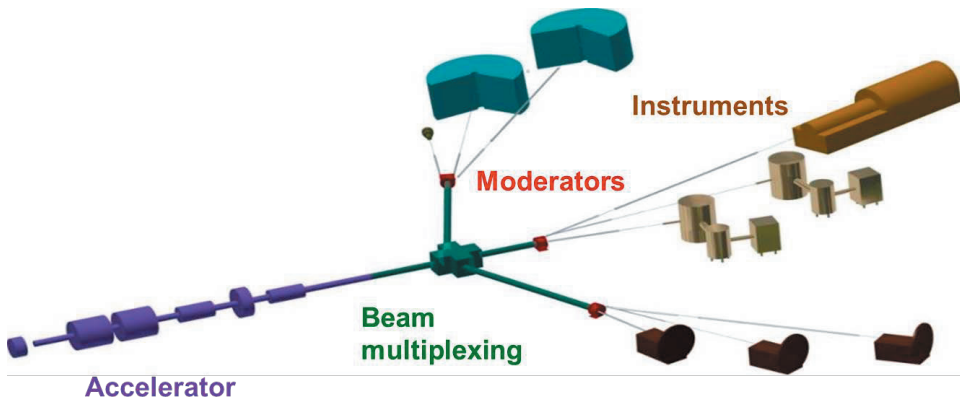

Fig. 4:   Schematic drawing of setup of high brilliance neutron source facility.

The instrumentation at such a compact HBS source can be fully competitive to the instruments operated today. In particular the possibility to realize smaller and affordable local neutron sources will change research with neutrons fundamentally by bringing neutrons to the users instead of bringing users to neutrons. Such a small size neutron source can be designed suitable to operate elastic neutron scattering instrumentation as well as neutron imaging and activation analysis on a laboratory level, e.g. at a larger university or at a company where neutron characterization on-line with the production process can improve the quality level of the products. With suffient neutron flux available also inelastic neutron scatttering instruments can be operated [13].

3.2 Accelerator

From the accelerator point of view well established technology provides nowadays a robust and stable operating accelerator system based on a high current 100 mA ion source [14], corresponding low and medium energy beam transfer systems (LEBT, MEBT), radio-frequency quadrupole (RFQ) systems to increase the initial proton energy, and drift tube linac (DTL) systems to achieve the final energy. The final energy (~100 MeV) can be accomplished adding appropriate DTL cavities (see Fig. 5).

The proton beam will be pulsed to match the requirements for optimized neutron instrumentation with requested pulse length between 100-600 µsec for optimized time-of-flight experiments. This pulse length relates to the time necessary for high energy neutrons to loose their energy on their way through appropriate thermal moderator materials from some 10 MeV to some 100 meV energy. The repetition rate of the pulses will be in the range of 10 Hz to 400 Hz to match the required energy resolution of a single neutron pulse by the instruments in time-of-flight mode and the length from the moderator. The expected duty cycle of the proton beam will be around a few percent reaching a time average power of 100 kW of the proton beam at the target.
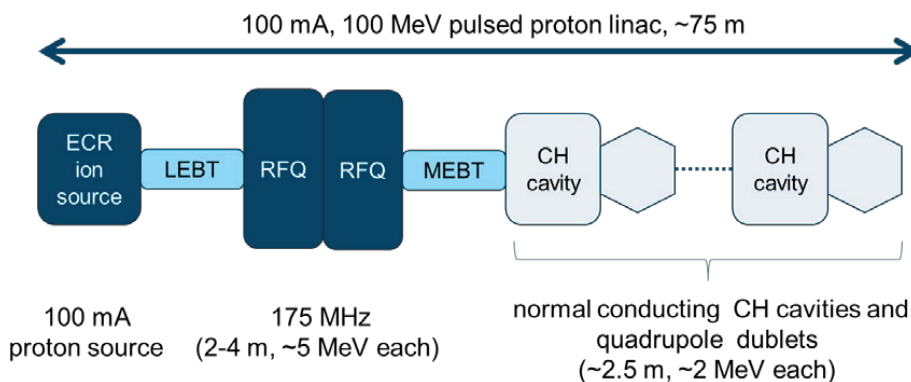

Fig. 5: Basic layout of HBS accelerator system.

3.3 Target-Moderator-Reflector Unit

The concept of the neutron source is based on the production of fast neutrons from the interaction of a 10-100 MeV pulsed proton beam with a water-cooled target of appropriate thickness. The geometry of the target and moderator-reflector assembly is shown in Fig.6.

The fast neutrons produced are slowed down to thermal energy in a dimension optimized moderator and reflector assembly. A thermal moderator like polyethylene (PE) with a thickness of ~10 cm is moderating the fast neutrons with MeV energy to thermal energies between 10 meV and 500 meV. A reflector like beryllium, lead or molybdenum with a thickness of ~20 cm increases the thermal neutron flux inside the moderator due to backscattering. Everything is surrounded by the shielding consisting of borated PE and lead. A thermal flux of 1.4×10^{11} cm^{-2}s^{-1}mA^{-1} neutrons is calculated for such a target unit. Up to six beam guides can be placed within the moderator core around the target.
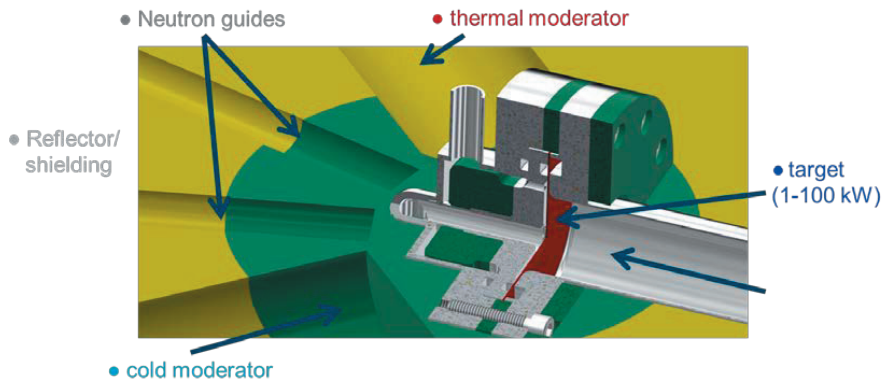

Fig. 6: Geometry of the target and moderator-reflector assembly. The proton beam hits the target from the right.

The most efficient cold neutron moderator materials for neutron scattering purposes are liquid ortho/para H_2 and solid CH_4 at cryogenic temperatures. Cold neutrons are produced in a cold finger moderator cryostat [15] containing an optimally sized mesitylene or ortho/para hydrogen moderator located at the point of maximum thermal neutron flux. With a diameter of up to 5 cm and a length of 10 cm in case of the liquid ortho/para hydrogen moderator such a cryogenic system can be inserted into the opening of the thermal moderator-reflector. The dimensions are optimized according to the principle of low dimensional moderators proposed for the ESS [16] to be used with liquid para hydrogen.

Thermal and cold neutrons are extracted using suitable advanced and well tested neutron optics. With this concept, fast, thermal and cold pulsed neutron beams can be delivered with properties (energy, time structure, divergence, energy resolution) in line with the specifications and requirements of the individual instruments and experiments respectively.

In the HBS concept there is no strict difference between a classical thermal moderator and reflector, since the materials in the simulation found to be most efficient (graphite, D_2O and beryllium) have high scattering and low absorption cross sections in the reference energy range, thus fulfilling both tasks at the same time. For radioprotection, the moderator and

reflector assembly is surrounded by a layered biological shielding structure consisting of borated polyethylene and lead.

3.4 Instrumentation and science

For brilliance optimized medium flux neutron sources and compact accelerator driven neutron sources the gain factor provided by an optimized neutron moderation and extraction system is the most efficient method to enhance the performance and compensate for the low value of the total number of neutrons produced compared to research reactors. In terms of neutron scattering experiments, an optimized set-up, starting with a dedicated target-moderator unit for each individual neutron instrument achieves the best brilliance for a specific experiment. The approach including neutron optical elements (such as filters, guides and lenses) placed as close to the source as possible along with an optimized cold source design, can be tailored specifically to produce an optimized phase space volume for certain instrument requirements. It will allow neutron spectrometers to be competitive in flux to those at medium flux reactor and spallation neutron sources [13, 17].

Based on the possibility to reach optimized neutron flux for small samples, options for scientific work at such novel sources can be considered as highly attractive in particular assuming easy and short time access close to individual labs and research areas. Hence, the instrumentation at a HBS neutron source will cover the most relevant neutron methods as powder and single crystal diffraction, small angle scattering, reflectometry, neutron spectroscopy, imaging and neutron analytics and shall depend strongly on the character of the experiments to be performed. Most of these methods will be operated in time-of-flight mode which provides an improved signal-to-noise ratio and an efficient neutron exploitation of a pulsed neutron source. We only present some short sketches of a few instruments that can be operated at a HBS neutron facility in the following.

SANS

SANS scattering instruments are workhouse instruments for neutron scattering in soft matter and metallurgy. The implementation of a SANS instrument on a HBS source benefits from the fact that SANS instruments are low resolution instruments which can nicely be operated with a wavelength resolution $\Delta\lambda/\lambda$ ranging from 10 up to 20%. For SANS measurements, a wide wavelength range is desirable to cover a Q range as large as possible. Hence a source with long pulses and a slow repetition rate is the most efficient way to operate. With a source operating with pulses of length 2 ms and with an operation frequency of 20 Hz (4% duty cycle) a versatile SANS instrument would be available, comparable to existing SANS instruments at medium flux neutron sources.
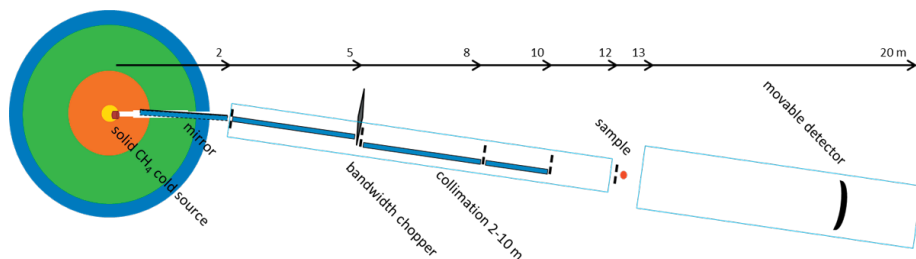
Fig. 7: Schematic layout of a time-of-flight small angle scattering instrument at the HBS.

Diffraction

Power diffraction instruments are also workhorses in neutron diffraction. Diffraction experiments yield more and more exact information than X-ray scattering when analysing complex low-symmetric crystal structures of many minerals because of the higher penetration depth of neutrons.

High resolution instruments are used for structure determination. High flux instruments are used to study phase transitions. Efficient powder diffraction in time-of-flight mode requires a large coverage of the space with detectors which turns into a rather expensive instrument (contrary to monochromatic instruments). The same comment applies to single crystal diffraction, an efficient spectrometer can be built only with a large space coverage ($\sim 2\pi$ steradians).

The powder diffractometer will be built at a thermal extraction channel. A fast rotating double chopper defines the neutron pulse length which will be around 250 msec and accordingly the wavelength resolution $\Delta\lambda/\lambda\sim1\%$. Between the chopper and the sample position an elliptical neutron guide is used to focus the neutron beam onto the sample.

Reflectometry

Neutron reflectivity is a technique which is especially suited to Time-of-Flight measurements because reflectivity signals follow a dependence $R \sim 1 / Q^4 \sim \lambda^4$ while the neutron intensity in a Maxwellian distribution follows a $I(\lambda)\sim1/\lambda^4$ dependence. This has made neutron reflectivity a technique which can be very efficiently implemented on pulsed sources. Operating with pulses of length 2 ms and with an operation frequency of 20 Hz this corresponds to a duty cycle of 4%. On a pulsed source the wavelength resolution degrades very quickly for short wavelengths.

Spectroscopy

Inelastic scattering is one of the great strength of neutron scattering thanks to the very low kinetic energy of neutrons (in the meV range). This allows for the study of excitations with energies ranging from neV (with spin-echo techniques) up to 0.1 eV using three axes techniques. However, spectroscopic measurements are usually more flux demanding than elastic studies.

At HBS time-of-flight spectrometers can be realized which provide spectroscopic measurements in the range from 20 μeV up to 1meV. The moderator pulse lengths should be long enough to allow the relaxation of the resolution by the pulsed neutron beam within a range defined by a specific instrument design. Such instruments can be built in a direct (e.g. MERLIN@ISIS) or inverse geometry (e.g. OSIRIS@ISIS). In the direct geometry, the sample is illuminated by a short monochromatic pulse. A broad range of final neutron energies arrives at different times at the detector and the neutron energy resolution is then determined by the length of the illuminating pulse. The indirect spectrometer filters a narrow spectral range of final neutron energies by means of e.g. a monochromatic analyzer or a Be transmission filter. The initial neutron beam, which has a broad spectral range, is pulsed at a rather large distance from the sample, so that the initial neutron energy can be measured by the arrival time at the detector. A rough estimate assuming an incident vertical and horizontal divergence of 1° on the sample and a 100 μeV energy resolution gives a flux of 2.5×10^{7} n/cm²/s which is on par with the flux quoted for the OSIRIS instrument at ISIS (Tab. 2). Such instruments seem perfectly viable on a compact neutron source.

	Backscattering	Cold ToF	Thermal ToF
$E_{i,f}$ (meV)	1.84	5	45
$\Delta E_i/E_i$ (%)	1	2	5
$\Delta\theta$ (°)	4	2	0.75
Δt (μs)	120	50	18
Rep. rate (Hz)	200	100	400
Flux (cm^{-2}s^{-1})	2.5×10^{7}	1.3×10^{5}	1×10^{5}
Reference instrument	OSIRIS	LET	MERLIN
Flux reference (cm^{-2}s^{-1})	2.7×10^{7}	5×10^{4}	6×10^{4}

Table 1. Parameters and performance for selected spectrometer types compared to reference instruments [13].

In the case of the Spin-Echo technique, very low wavelengths resolutions can be used (up to 20%) so that a very short instrument can be build. An instrument of 12 m on a 2 ms pulse length source with collimations suitable for spin-echo (~0.5°) would achieve a polarized brilliance of about 5×10^{6} n/cm²/s.

Radiography

Neutron radiography / tomography is widely used in research, industry or medical fields to investigate the internal structure of materials or to study dynamical processes. It is based on the measurement of the neutron attenuation properties of an object using, for example, a CCD camera or a flat-panel silicon detector, after conversion of the neutron energy into light with an appropriate scintillator.

Standard neutron radiography is a technique which is intrinsically handicapped on a pulsed source since it cannot benefit from the time structure of the source, since most of the current experiments are performed using a white beam. One can consider an instrument which would have a length of 4 meters and a pinhole of size 5x5 cm² (preceded by a 1 meter neutron guide extracting the neutron from the moderator). With such a geometry, the flux at the detector position is 2×10^7 n/s/cm². The illuminated area is rather homogeneous over an area of about 25x25 cm² which is perfectly suitable for industrial radiography using films or image plates. The L/D of this instrument is 80. Energy resolved experiments can also be performed by using either velocity selectors or crystal monochromators.

For ToF instruments, one has to set the aimed energy resolution and the useful wavelength range. Here the useful wavelength range is defined by the highest Bragg edge of the materials to be studied which is below 5Å for metals. The useful bandwidth thus ranges between 2 to 5Å. An energy resolution of 1% can be achieved which is smaller than typical Bragg edge widths (~2%).

PGNAA

Prompt gamma neutron activation analysis (PGNAA) using cold or thermal neutron beams is a powerful non-destructive analytical technique to determine the elemental or isotopic composition of small samples of various origins (geological, environmental, biological, pharmaceutical, cultural and industrial). It is based on the measurement of prompt gamma rays emitted during the de-excitation of compound nuclei formed from neutron capture in the sample. Both prompt and delayed gamma radiations are characteristic, i.e. their energy identifies the element and the isotope, while their intensities are proportional to their quantities. Thus, their detection can be used for the non-destructive analysis of a large variety of materials.

The pulsed HBS neutron source will allow to perform prompt and delayed gamma neutron activation analysis with neutron time-of-flight (TOF-P&DGNAA) in order to determine the elemental composition of heterogeneous large samples.

References

[1] Neutron scattering facilities in Europe. Present status and future perspectives, ESFRI PSE Strategy Working Group, Neutron Landscape Group, ESFRI Scripta Vol. I, 2016.

[2] M. Furusaka et al., Physics Procedia, 60, 167, 2014

[3] Y. Ikeda et al., Nucl. Instr. Meth. A, 833, 61, 2016

[4] D.V. Baxter, Eur. Phys. J. Plus, 131, 83, 2016

[5] W. Furman et al. Intense resonance neutron source (IREN) - new pulsed source for nuclear physical and applied investigations. 11th Int. Conf. Nuclear Engineering, Tokyo, JAPAN, April 20-23, 2003, ICONE 11 - 36318

[6] T. Federer, Physics Today, 69, 25, 2016; J. Cartwright, Physics World, Oct. 2016.

[7] LLB, Annual Report 2015, SONATE Compact Neutron Source.

[8] J.P. de Vicente et al., STFC, Technical Report, RAL-TR-2013-016.

[9] L. Silvestrin et al., Eur. Phys. J. Plus, 131, 72, 2016

[10] U. Rücker et al., Eur. Phys. J. Plus, 131, 19, 2016.

[11] I.S. Anderson et al., Physics Reports, 654, 1, 2016

[12] J.R. Granada et al., Eur. Phys. J. Plus, 131, 216, 2016

[13] J. Voigt et al., Nucl. Instr. Meth. A, 884, 59, 2018

[14] R. Gobin et al., Rev. Sci. Instr. 73, 922, 2002

[15] T. Cronert et al., J Phys.: Conf. Ser., 746, 012036, 2016

[16] K. Batkov et al., Nucl. Instr. Meth. A, 729, 500, 2013

[17] X. Fabreges et al., arXiv 1612.00232

D 1 Small-Angle Scattering

S. Jaksch

Jülich Centre for Neutron Science
at Heinz Maier-Leibnitz Zentrum

Forschungszentrum Jülich GmbH

Contents

1 Introduction

Small-Angle Scattering (SAS) investigates structures in samples that generally range from approximately 0.5 nm to a few 100 nm. This can both be done for isotropic samples such as blends and liquids, as well as anisotropic samples such as quasi-crystals. In order to obtain data about that size regime scattered intensity, mostly of x-rays or neutrons, is investigated at angles from close to zero, still in the region of the primary beam up to 10°, depending on the wavelength of the incoming radiation.

The two primary sources for SAS experiments are x-ray (small-angle x-ray scattering, *SAXS*) sources and neutron (small-angle neutron scattering, *SANS*) sources, which shall be the two cases discussed here. Also scattering with electrons or other particle waves is possible, but not the main use case for the purpose of this manuscript.

For most small-angle scattering instruments, both SAXS and SANS, the science case covers the investigation of self-assembled polymeric and biological systems, multi-scale systems with large size distribution of the contained particles, solutions of (nano-)particles and soft-matter systems, protein solutions, and material science investigations. In the case of SANS this is augmented by the possibility to also investigate the spin state of the sample and hence perform investigations of the magnetic structure of the sample.

In the following sections the general setup of both SAXS and SANS instruments shall be discussed, as well as data acquisition and evaluation and preparation of the sample and the experiment in general. The information contained herein should provide sufficient information for planning and performing a SAS experiment and evaluate the gathered data.

General concept

All SAS experiments, irrespective of the setup used in any specific case, rely on the concept of pinhole cameras to work. Fig.1 illustrates the geometric concept of the interplay between pinhole cameras and SAS.

In the usual case, pinhole cameras map every point of the sample (object) to a discrete point on the screen (film or detector). The smaller the hole, the better the point-to-point mapping works, since in the ideal case only a single path between object and image is available. However, this of course comes with a penalty in intensity, since the smaller hole lets less light pass through. Due to the geometry, an image taken with a pinhole camera is always upside down. While the mathematical implications shall be discussed later on in this manuscript at this point we only want to grasp the underlying concept. The information about the object is at the beginning stored in real space. Colors (wavelength) and locations are given as points on the surface of the object. When all beams have converged to the single point that is ideally the pinhole, the information is then encoded in direction of the path (or light-beam) and the wavelength of the light. This is the change between direct and reciprocal space, locations and directions. When the light falls onto the screen the information is reversed again, to location and color of a spot on the screen, into direct space.

This concept is exploited by SAS. Since we are looking at very small objects (molecules and atoms) the determination of the location with the naked eye, or even a microscope, and encoding
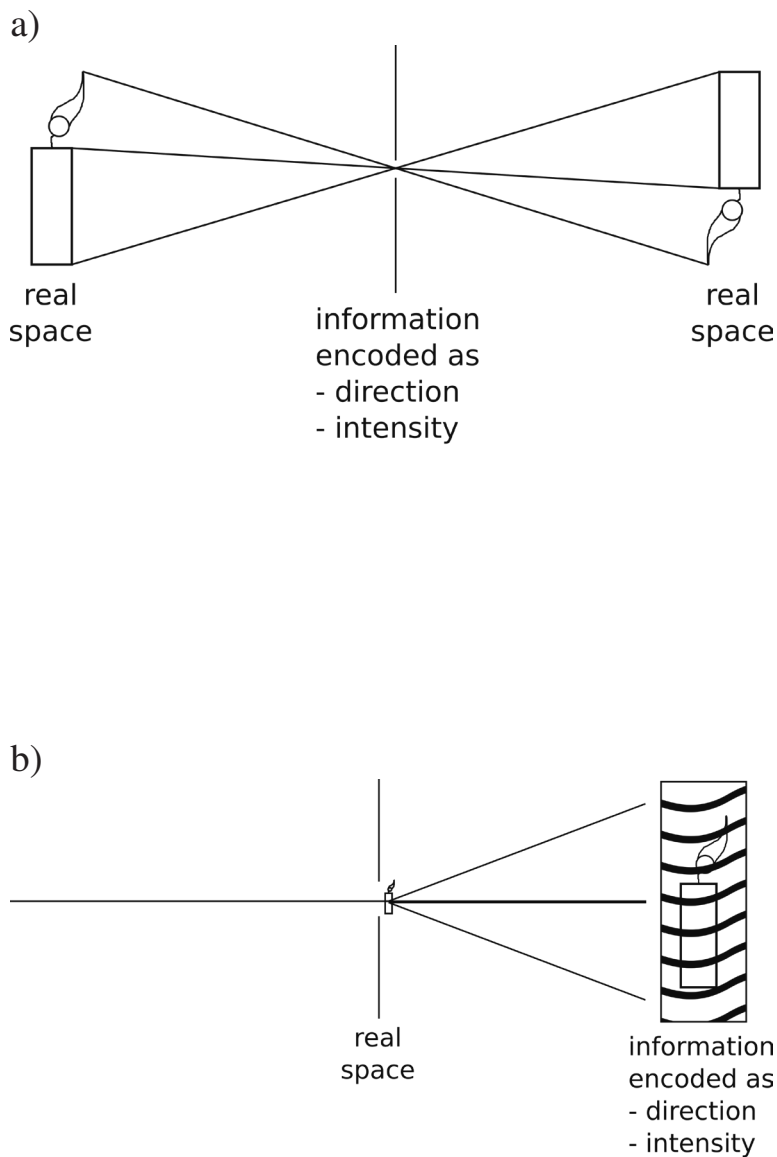
a)

real
space

information
encoded as
- direction
- intensity

real
space

b)

real
space

information
encoded as
- direction
- intensity

Fig. 1: a) Sketch of a pinhole camera and b) a simplified SAS instrument. The encoding of the real space information is in one case done inside the pinhole, in the other case the direction (and wavelength) encoded information is directly displayed on the screen (shaded area with waves). Positioning of the screen farther away improves the angular resolution and therefore the encoded information.
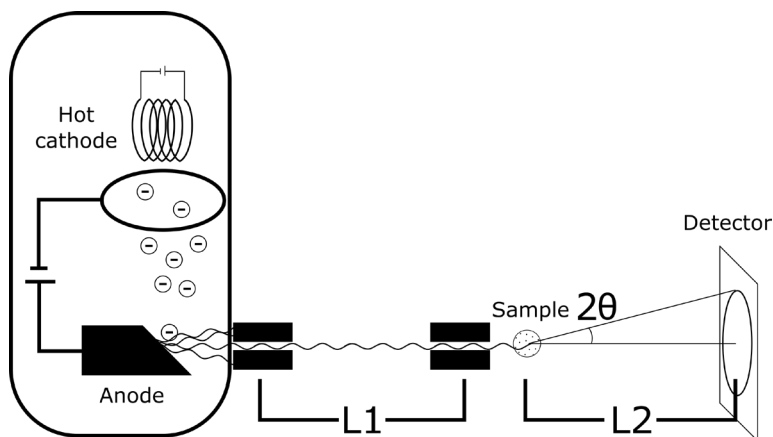
Fig. 2: Laboratory SAXS setup. The left box is a sketch of a x-ray tube, all the components are in vacuum. The flight path is also usually evacuated. L1 and L2 are the collimation and sample detector distance (SDD) respectively. In the case of laboratory setup those range usually from about 20 cm up to 1-2 m in modern setups. The collimation blocks for L1 and L2 are usually set up in both x and y direction to constrict the flight path, widely used openings are around $1\,\text{mm} \times 1\,\text{mm}$ or below. In some setups, also a slit collimation instead of a point collimation is realized to increase the intensity.

of direction is easily achievable by increasing the distances and adjusting the size of the pinhole. However, instead of using the information that has been transferred to real space again, this time the object in real space is put close to the window. This way, the information about the location of atoms and molecules in the sample is encoded into direction or reciprocal space. Since there should be no information about the light before the pinhole, the light needs to be collimated down to a small, point-like source with no angular divergence.

2   SAXS instruments

In general there are two classes of SAXS instruments. One is the laboratory type setup that can be set-up in a single laboratory with a conventional x-ray tube, or more general any metal anode setup, while the other one is a large-scale facility setup at a synchrotron that can provide higher intensities. Since the setup of both instruments differs, and also the use case is not fully identical, we shall discuss both setups separately. One thing that should be kept in mind is that the fundamental principle is identical, i.e. any experiment that can be performed at a synchrotron can also in principle be performed at a laboratory SAXS setup and is only limited in intensity. This is important for the preparation of beamtimes at a synchrotron, which in general should be thoroughly prepared in order to fully exploit all capabilities offered there.
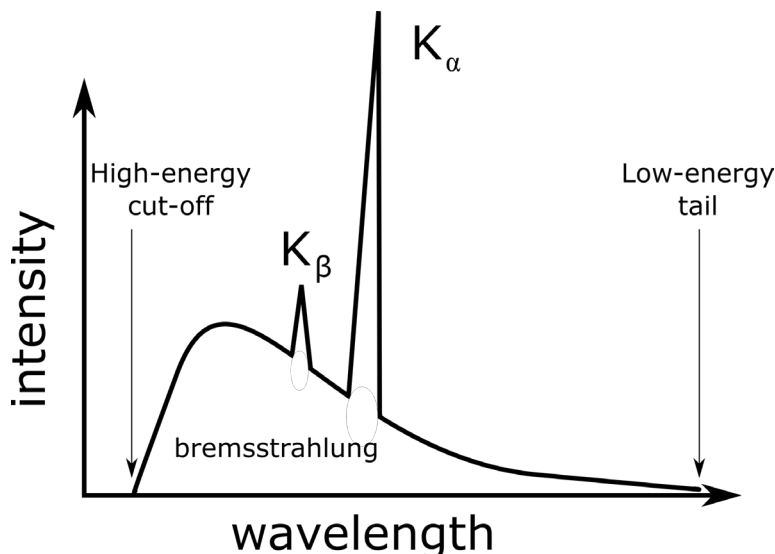
Fig. 3: Characteristic x-ray spectrum from a metal anode x-ray tube. The high-energy cut-off wavelength is given for the case that a single electron, fully accelerated by the voltage in the x-ray tube, deposits all its kinetic energy in a single photon. In an optimal setup this distribution is very narrow. Then the K_α line fully dominates the spectrum and gives a clean wavelength to perform a SAXS instrument.

Laboratory SAXS setup

Over the years a wide range of specialized SAXS instruments has become commercially available. The oldest concepts date back to the early 20th century, right after the discovery of x-rays. [1] Most of them offer specific advantages in certain use cases, such as the measurement of isotropic samples in a Kratky Camera [2], or highly adaptable sample environments. Here we shall only concentrate on the basic principle of operation. A general sketch of a SAXS instrument is shown in Fig.2. The x-rays are produced in an x-ray tube and then collimated by a set of slits. Here the collimation as such is already sufficient to obtain a coherent beam, since most of the intensity of standard x-ray tubes (and essentially all metal target x-ray sources) is concentrated into the characteristic spectral lines of the target material (see Fig.3). Common materials for the target anode are copper and molybdenum, delivering wavelengths of the most intensive K-α lines of 1.54 Å and 0.71 Å respectively. Under the assumption of a usual characteristic spectrum for the anode material the x-ray tubes can be considered monochromatic sources.

In order to achieve spatial as well as wavelength coherence most x-ray tubes work with a focused beam that is as small as technically feasible. This allows very narrow collimation slits, since it is not improving the coherence, and therefore the signal-to-noise ratio, to narrow the slit further than the initial beam spot or the pixel size of the detector, whichever be smaller. This however leads to a very high energy density, why some x-ray tube designs forgo a solid anode all together and either opt for a rotating anode, where the energy of the beam spot is distributed over a larger surface or a metal-jet anode, where the material is refluxed and can therefore not heat up beyond the point of deformation and therefore also defocussing of the beam.

Parameter	value
SDD	0.8-4 m
Pixel resolution	$172 \times 172 \mu m$
Flux	10^7 photons s^{-1}
wavelength λ	1.35 Å
Q-range	$4 \cdot 10^{-3}$-$8 \cdot 10^{-1}$ Å^{-1}

Table 1: Performance parameters for state of the art laboratory SAXS setups, in this case with a liquid metal jet anode at the GALAXI instrument. [3]

Some performance figures of current laboratory SAXS setups are given in Tab.1. It is worth noting that with the last generation of metal-jet anode setups even laboratory setups can achieve intensities comparable to what was achievable one or two decades ago at a world-class synchrotron. While this of course allows for faster measurements and smaller beam, it also means that beam damage to the sample has to be taken into account.

Synchrotron SAXS setups

While the setup in general is similar to that of a laboratory setup there are some key differences between a synchroton and a laboratory SAXS setup. Most of the differences are based on radio protection needs and are therefore immaterial to this description in terms of the SAXS measurement itself. The other main difference is in the production of the x-rays itself. Current setups at synchrotrons use undulators in order to periodically accelerate charged particles (usually electrons/positrons) perpendicular to the direction of propagation of the particle beam. This creates a very brilliant, nearly perfectly monochromatic x-ray beam along the direction of the electron beam. The monochromaticity can further be improved by a monochromator crystal. Fig.4 shows an example of an synchrotron SAXS setup. After that, the collimation is very similar to that of a laboratory SAXS setup, only the materials are chosen to be thicker in most cases to improve the absorption characteristics. Due to the monochromaticity the brilliance, coherence and signal-to-noise ratio are significantly better than that of a laboratory SAXS setup, since there is no bremsstrahlung spectrum to contribute to the background. In terms of achievable wavelength there is no limitation to use a specific K-α line of any specific material. Often common wavelengths are chosen to better correspond to laboratory measurements on identical samples. One option that is also available in some synchrotrons is the tunability of the wavelength in order to measure resonance effects in the atomic structure of the sample (anomalous SAXS, *ASAXS*) [4] or better chose the accessible Q-space. Tab.2 summarizes some of the performance figures of current synchrotron SAXS setups. For most synchrotron SAXS beamlines beam damage, especially for organic samples, is an issue and has to be taken into account when planning an experiment.

3   SANS setups

In contrast to x-rays, sufficient numbers of free neutrons can only be obtained by nuclear processes, such as fission, fusion and spallation. As large-scale facilities are needed to create the

Parameter	value
SDD	0.8-4 m
Pixel resolution	$172 \times 172 \mu m$
Flux	10^{18} photons s^{-1}
wavelength λ	0.54 - 1.38 Å

Table 2: Performance parameters for a state of the art synchrotron SAXS beamline, here P03 at DESY. [5]
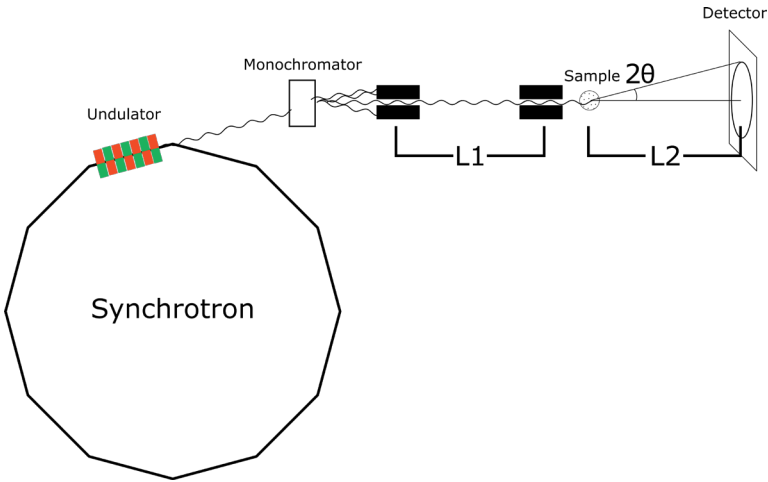

Fig. 4: Synchrotron SAXS setup. Here the radiation is produced in the storage ring of a synchrotron. In earlier designs, the x-rays were produced at the bending magnets in the ring (kinks in the ring here). This however lead to a wide spread of the produced wavelength and a high angular distribution of the radiation. An undulator from a magnet array as depicted here produces a narrow distribution of wavelength and angular divergence. The rest of the setup is comparable to the laboratory setup, albeit the intensity of the radiation is orders of magnitude higher, which allows for finer collimation slits and longer collimation distances and SDDs.

processes at a suitable rate to perform scattering experiments with them, the only facilities where neutron scattering today can be performed is at fission reactor sources and spallation sources. This of course also leads to larger efforts in terms of biological shielding.

It is an inherent feature of those reactions that the reaction products show a wide distribution of energies, with peak energies ranging up to 3 MeV kinetic energy per neutron. This leads to deBroglie wavelengths in the fermi meter region, which is unsuitable for SANS scattering experiments. Thus, in order to obtain a coherent beam it is not only necessary to collimate the neutrons but also to moderate and monochromatize them. Both processes result in losses in usable flux, since the phase space of neutrons cannot be compressed by lenses, as is the case for photons.

The moderation process is performed by collision processes in a moderator medium. The moderator is a material at temperatures around 25 K or below and the resulting neutron spectrum is a Maxwell-Boltzmann spectrum of the corresponding temperature. This results in peak wave-
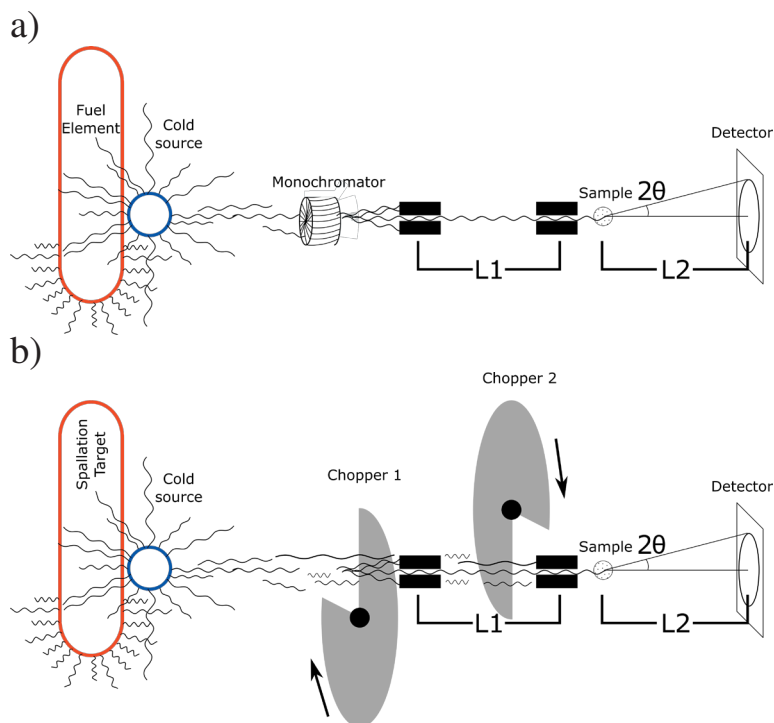
a)

b)

Fig. 5: a) Continuous source SANS setup and b) pulsed source SANS Setup. In both cases the neutron source (red) creates hot neutrons of a short wavelength. A cold source (blue) vessel (usually filled with cold ^{2}H or ^{2}D) is moderating the neutrons down to slower speeds, i.e. longer wavelengths. In both cases the collimation distance and SDD is widely adjustable for most instruments, with lengths between 1 m up to 30 m. In a SANS instrument at a continuous source a monochromator (a turbine with slightly inclined channels) selects a certain wavelength (usually between 3 and 15 Å) and afterwards the setup is very much like the one shown for SAXS setups, except that the whole instrument is larger. In case of a pulsed source choppers (rotating discs with transparent openings for neutrons) define a start and an end time for each pulse. Since neutrons, different from x-rays, are particle waves, their wavelength determines their speed. Thus, the wavelength is determined by measuring the time of arrival at the detector for each neutron. For an optimized neutron transport all components are usually evacuated.

lengths around 4 Å for the neutron beam. Neutron scattering instruments can be run both in time-of-flight mode or monochromatic mode.

A schematic of a SANS instrument is shown in Fig.5. Both cases with a monochromator and a chopper setup for time-of-flight are presented. In a continous source the neutron flux has to be interrupted for the timing of time-of-flight mode while for pulsed sources there is an inherent interruption of the neutron flux.

This moderation and collimation process in consequence means that neutrons always show an albeit small distribution of wavelenghts and therefore a lower signal to noise level than x-ray sources. Spin and isotopic incoherence add to that. Beam damage however is nigh on impossible with the weakly interacting neutrons.
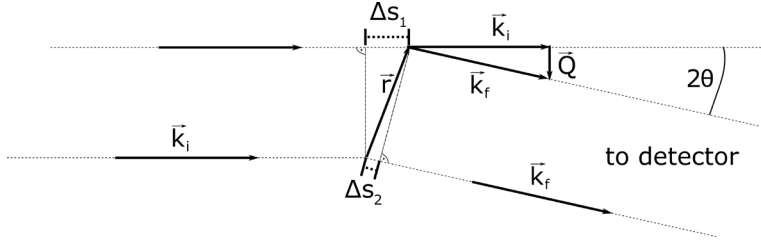
Fig. 6: Construction of **Q**. The incoming and final wavevectors $\mathbf{k}_i$ and $\mathbf{k}_f$ define both the scattering vector **Q** as well as the path length difference $\delta = \Delta s_1 - \Delta s_2$. Here it is important to note that the selection of the center of origin is arbitrary and thus can be chosen to be at the center of the construction. The calculation of the length of **Q** is then given by Eq.2.

4 Reciprocal space and Small-Angle Scattering

The need for the resolution of small angles can be directly derived from Bragg's equation

$$n\lambda = 2d \cdot \sin \delta \tag{1}$$

with n being the order of the diffraction, d being the distance between two scatterers, θ as the scattering angle and λ the wavelength of the incoming beam. In order to get interference the incoming beam has to have a wavelength that corresponds to the investigated size regime, which in both cases is on the order of a few Angstroms. Using Bragg's equation with $n = 1$, $d = 50\text{Å}$ and $\lambda = 1\text{Å}$ we arrive at $0.01 = \sin \theta \approx \theta$. Thus, the largest structures to be resolved are determined by the smallest achievable angle.

In order to allow for a setup and wavelength independent data evaluation the data is recorded in terms of **Q** or reciprocal space. The construction of that Q-space from two scattering points is shown in Fig.6. From that the magnitude of **Q**, which here for simplicity is $|\mathbf{Q}| = Q$, can be derived as

$$Q = \frac{4\pi}{\lambda} \sin \theta. \tag{2}$$

Even though **Q** is strictly speaking a vector, for most small angle problems only the absolute value Q is of interest, hence this simplification is reasonable. This is due to the isotropic scattering picture of a majority of small-angle scattering data. Another simplification that is often used is the small-angle approximation for the sine with $\sin \theta = \theta$, which is very well valid for small angles. Combining Eqs.1 and 2 also delivers a useful expression for the approximation of inter-particle distances or correlation lengths

$$d = \frac{2\pi}{Q}. \tag{3}$$

5 Resolution limits

SAS is working based on the interference of coherent radiation. That in itself imposes some limitations on the samples and properties that can be investigated.

In term of size, the object under observation has to be of the same order of magnitude as the wavelength of the incoming radiation, analogous to light interference at a double slit. Concerning the analysis in reciprocal space, also the limited size of the detector and coherence volume of the sample has to be taken into account.

The second limitation that should always be considered is that only elastic scattering renders useful results, i.e. any change in speed or wavelength of the incoming radiation will render unusable results.

Finally, multiple scattering is usually not considered for the evaluation of SAS data. This means, mostly thin samples, or those with a high transmission (usually 90% or higher), can be investigated.

6 Fourier Transform and Phase Problem

Considering the spacing of only two scattering centers as in the last section needs to be extended to an arrangement of scattering centers for evaluation of macroscopic samples, where each atom/molecule can contribute to the scattered intensity. Since the incoming wave at location $\mathbf{x}$ can be considered to be an even wave it can be described by

$$A(\mathbf{x}, t) = A_0 \exp\left(i2\pi(\nu t - \frac{\mathbf{x}}{\lambda})\right) \tag{4}$$

With a A being the amplitude as a function of position $\mathbf{x}$ and time t. A_0 is the modulus of the amplitude, ν the frequency and lambda the wavelength.

In order to calculate the correct phase shift $\Delta\phi$ after scattering from two centers as in Fig.6 we need to know the differences in travelled distance between the two waves δ. This then yields

$$\Delta\phi = \frac{2\pi\delta}{\lambda} = \mathbf{Q}\mathbf{r}, \tag{5}$$

which is equivalent to the expression $2\pi x/\lambda$ in Eq.4. Here also the relation $\mathbf{Q} = \mathbf{k}_f - \mathbf{k}_i$ was used. This then leaves us with the spherical wave scattered by the first scattering center

$$A_1(\mathbf{x}, t) = A_0 b \exp(i2\pi(\nu t - \mathbf{x}/\lambda)) \tag{6}$$

and the corresponding scattered wave from the second scattering center

$$
\begin{aligned}
A_2(\mathbf{x}, t) &= A_1(\mathbf{x}, t) \exp i\Delta\phi \tag{7} \\
&= A_0 b \exp(i2\pi(\nu t - \mathbf{x}/\lambda)) \exp i\mathbf{Q}\mathbf{r} \tag{8}
\end{aligned}
$$

This can then be combined into the full description of the amplitude with both contributions to

$$
\begin{aligned}
A(\mathbf{x}, t) &= A_1(\mathbf{x}, t) + A_2(\mathbf{x}, t) && (9) \\
&= A_0 b \exp(i2\pi(\nu t - \mathbf{x}/\lambda))(1 + \exp i\mathbf{Q}\mathbf{r}) && (10)
\end{aligned}
$$

here an arbitrary scattering efficiency b for each scattering center has been introduced, which will later be discussed for both x-rays and neutrons.

Since only intensity can be observed at the detector, we need to consider the square, calculated with the complex conjugate of the expression itself

$$
\begin{aligned}
I(\mathbf{Q}) &= A(\mathbf{x}, t) A^*(\mathbf{x}, t) && (11) \\
&= A_0^2 b^2 (1 + \exp(i\mathbf{Q}\mathbf{r}))(1 + \exp(-i\mathbf{Q}\mathbf{r})). && (12)
\end{aligned}
$$

Here the time and absolute location dependencies in Eq.10 have cancelled each other out, so we can neglect them and are left with a function that solely depends on the scattering vector $\mathbf{Q}$ and the location of the particles $\mathbf{r}$. Neglecting those dependencies allows us to generalize Eq. 10 to the case of N identical scattering centers with

$$
A(\mathbf{Q}) = A_0 b \sum_{i=1}^{N} \exp(i\mathbf{Q}\mathbf{r}_i). \tag{13}
$$

The $\mathbf{r}_i$ here signify the relative locations of all scattering centers in the sample, relative to either simply the first scattering center or any arbitrary center chosen. Indeed all arrangements are mathematically identical. Replacing the sum by a weighed integral allows also the calculation for the case of a (quasi)continuous sample with number density $\rho(\mathbf{r})$:

$$
A(\mathbf{Q}) = A_0 b \int_V \rho(\mathbf{r}) \exp i\mathbf{Q}\mathbf{r} d\mathbf{r} \tag{14}
$$

This is the Fourier transform of the number density of scattering centers with scattering efficiency b, it can also be applied for numerous scattering efficiencies.

However, since the phase information got lost while obtaining the intensity as an absolute square of the amplitudes, there is no direct analytic way of performing an inverse Fourier transform. This is why this is called the phase problem. Also, as described above, in a wide range of cases it is enough to investigate the modulus of $\mathbf{Q}$, neglecting its vector nature.

7 Scattering Efficiency

Since the physical scattering event is very dissimilar for x-rays and neutrons they shall be discussed separately here. However, it should be noted, that the nature of the scattering process does not impact on the method of data evaluation in general. Only in very specific cases, such as contrast matching or polarized scattering there is any discernible difference.

Scattering with x-rays

X-rays, as photons, interact with the sample via electromagnetic interaction. For the purpose of this manuscript it is sufficient to note that the vast majority only interact with the electron shell around the atoms and thus effectively map the electron density within the sample. Interactions with the nucleus would only occur at very high energies, which are not usually used in elastic scattering. In a rough approximation the strength of the electromagnetic interaction scales with Z^2, meaning that heavy elements, such as a wide range of common metals, scatter considerably stronger than light ones, like hydrocarbon compounds. For element analyses there is also the possibility of resonance scattering, where the chosen x-ray energies are close to the resonance gaps in the absorption spectrum of specific elements (ASAXS). [4]

Based on Thomson scattering the scattered intensity at angle 2θ is

$$I(2\theta) = I_0 \left(\frac{e^2}{mc^2} \right) \frac{1 + \cos^2 2\theta}{2} \tag{15}$$

$$\frac{I}{I_0} = \left(\frac{d\sigma}{d\Omega} \right)_2 = r_e^2 \frac{1 + \cos^2 2\theta}{2} \tag{16}$$

Here we also introduced the differential scattering cross section $\frac{d\sigma}{d\Omega}$ for a single electron and r_e being the radius of an electron. This means that the total probability for a scattering event to occur into a solid angle $d\Omega$ is exactly that value for a single, isolated electron. This probability is in units of an area. Thus, the scattering length for a single electron b_e is defined as the square root of that:

$$b_e = r_e \sqrt{\frac{1 + \cos^2 2\theta}{2}} \tag{17}$$

With those previous equations it is again important to note that small-angle scattering is mainly concerned with small angles, thus that $\cos 2\theta \approx 1$ is a very good approximation. This is also, together with backscattering, the location of the highest intensity and negligible polarization effects. The numeric values for the constants used here are $r_e = 2.818 \times 10^{-15}$ m and the scattering cross section for a single electron $\sigma_e = 6.65 \times 10^{-29}$ m^2 $= 0.665$ barn after integration over the full solid angle. As apparent with integration over the full solid angle, the relation is $\sigma = 4\pi b_e^2$.

Since usually the goal is to find the distribution of scattering centers in a volume, the density of scattering length per unit volume is of interest. This is the scattering length density (SLD)

$$\rho(\mathbf{r}) = \frac{b_e(\mathbf{r})}{V}. \tag{18}$$

A very common way of expressing scattering efficiency is using electron units. As can be seen in Eq.13 the scattering amplitude is only determined by the SLD of a single electron apart from the Fourier transform of the local density. This means the scattering intensity in electronic units can be expressed as

$$I_{eu}(Q) = \frac{I(Q)}{I_0 b_e^2} \tag{19}$$

This means, with appropriate calibration, if there is an intensity of $I_{eu} = 200\, b_e^2$ at a certain Q, that the size scale corresponding to that Q vector has 200 electrons per unit volume.

Since photons interact mainly with the electron shell, there is also an angle dependency accounting for the time averaged location probability of the electrons in the shell, which may or may not be spherical, depending on the electronic configuration of that specific atom. This would then lead to a SLD in terms of $b_e(Q) = b_e f_s(Q)$ with f_s being the atomic scattering factor for any specific element. This important to take note of, when there is a structure or form factor on the same size scale as a single atomic distance $Q = \frac{2\pi}{1.54\,\text{Å}} = 4.08\,\text{Å}^{-1}$. This is usually not in the regime of interest for small-angle scattering and will mostly vanish in the incoherent background.

Another incoherent background effect is Compton scattering, where inelastic processes change the wavelength during the scattering process. This is however again strongly suppressed at small angles. The wavelength shift occurring based on Compton scattering is following this expression

$$\Delta\lambda = \frac{h}{mc} 2\sin^2\theta \tag{20}$$

The prefactor is $\frac{h}{mc} = 0.02426\,\text{Å}$. It is also obvious that at large angles $2\theta = 180°$ the energy transfer is maximal. Since we are always investigating angles close to $\theta = 0$ the wavelength shift and hence the incoherent background is negligible compared to other experimental factors, such as slits and windows scattering.

Scattering with neutrons

Neutrons interact with the nuclei directly, which results in the atomic form factor being always spherically symmetric (billiard balls) and them being sensitive to different isotopes and spin-spin coupling. In contrast to x-rays, there is no simple expression for scattering strength as a function of isotope or atomic number. Directly neighboring elements and isotopes may have vastly different cross sections.

Based on that we have to rely on tabulated values for the cross sections and scattering lengths of different elements and isotopes (see Tab.3) and can only write the cross section and scattering length relation as

$$\frac{d\sigma}{d\Omega} = b^2 \tag{21}$$

That said, only coherent scattering can form interference patterns, i.e. no change of the nature of the radiation can take place during the scattering process. However, since the neutron can change its spin orientation through spin-spin coupling during the scattering process that may

Element	scattering length $b_{coh}/10^{-14}m$
^{1}H	-0.374
^{2}D	0.667
C	0.665
N	0.936
O	0.580
Si	0.415
Br	0.680

Table 3: Coherent scattering length of several elements and isotopes.

happen, depending on the spin orientation of the sample nuclei. Those are completely statistical processes.

As neutrons are fermions, which have spin $1/2$ the possible outcomes after a scattering process with a nucleus of spin i are $i + 1/2$ and $i - 1/2$, and the associated possible spin states are

$$\text{number of states } i + 1/2 \quad : \quad 2(i + 1/2) + 1 = 2i + 2 \tag{22}$$
$$\text{number of states } i - 1/2 \quad : \quad 2(i - 1/2) + 1 = 2i \tag{23}$$
$$\text{total number of states} \quad : \quad 4i + 2. \tag{24}$$

This immediately shows, that only for the case $i = 0$ there can be only two states. Since it is impossible to know the spin state of non-zero spin nuclei under ambient conditions, the differential cross section becomes a two-body problem of the form:

$$\frac{d\sigma}{d\Omega} = \sum_{i,j} \langle b_i b_j \rangle \exp -i\mathbf{Q}(\mathbf{r_i} - \mathbf{r_j}) \tag{25}$$

Here $\langle b_i b_j \rangle$ is the expectation value of the SLD for each $b_i b_j$ combination possible given isotope and spin variability. For this there is only one coherent outcome, where $b_i = b_j$, which then results in

$$\langle b_i b_i \rangle = \langle b_i^2 \rangle = \langle b^2 \rangle . \tag{26}$$

All other cases result in $b_i \neq b_j$ and therefore

$$\langle b_i b_j \rangle_{i \neq j} = \langle b_i \rangle \langle b_j \rangle = \langle b \rangle^2 . \tag{27}$$

This then results in

$$\frac{d\sigma}{d\Omega} = \langle b^2 \rangle \cdot \sum_{j,k} \exp\left(-i\mathbf{Q}(\mathbf{r_i} - \mathbf{r_j})\right) + N(\langle b^2 \rangle - \langle b \rangle^2). \tag{28}$$
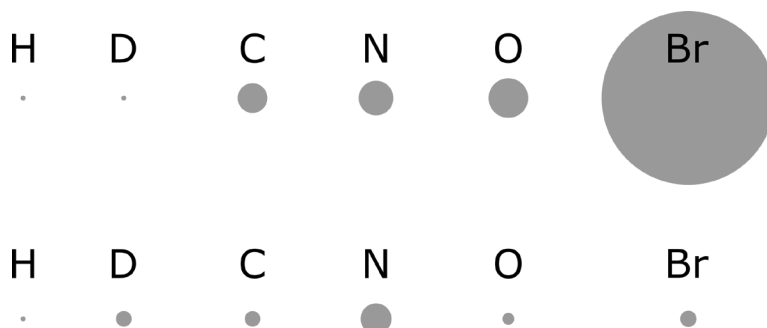
Fig. 7: Coherent cross-sections for selected elements for x-rays (top) and neutrons (bottom). The coherent scattering cross section scales linearly with the diameter of the circles. It is apparent, that the Z^2 dependency strongly emphasizes heavy elements in x-ray scattering, whereas for neutrons even single isotopes can be distinguished. However, for neutrons there is no simple analytic expression for the scattering cross-sections.

Here $\sqrt{\langle b^2 \rangle} = b_{coh}$ signifies the coherent scattering length density, since it contains information about the structure of the sample via $\mathbf{r_{ij}}$ and $\sqrt{\langle b^2 \rangle - \langle b \rangle^2} = b_{inc}$ is the incoherent cross section not containing any information about the sample structure. This cannot be suppressed instrumentally, therefore often isotopes with low incoherent scattering length are chosen in neutron scattering to suppress the incoherent background. Both coherent and incoherent scattering lengths can separately used together with Eq.18 to obtain the corresponding scattering length densities.

Scattering Cross Section and Contrast Matching

As described above there is a Z^2 dependency of the cross section of atoms in case of x-rays and the cross section values for neutrons have to be tabulated since there is no simple algebraic expression for that. The resulting differences in cross section are illustrated in Fig.7. Because different isotopes have very different cross sections for neutron scattering, in some cases it is possible to replace certain isotopes in order to arrive at desired contrast conditions.

One of the most important examples for that technique, called contrast matching, is replacing hydrogen by deuterium. This leaves the chemical composition of the sample unchanged, and hydrogen is extremely abundant in most organic compounds. The concept can in some cases be extended to be used as the Babinet principle, in order to suppress background scattering, since it is extremely preferable to have a solvent with a low background and a solute with a higher background than vice versa. A sketch of the concept is shown in Fig.8.

This method allows highlighting otherwise hidden features of the sample or suppressing dominant scattering in order to better determine a structure with a lower volume fraction and therefore less scattering contribution. Examples for that application are highlighting the shell of a sphere, by matching the core or vice versa. Also for protein samples certain structures can be matched, so that only distinct features are visible.

In order to apply contrast matching, mostly the solvent is changed. In some rare cases also the
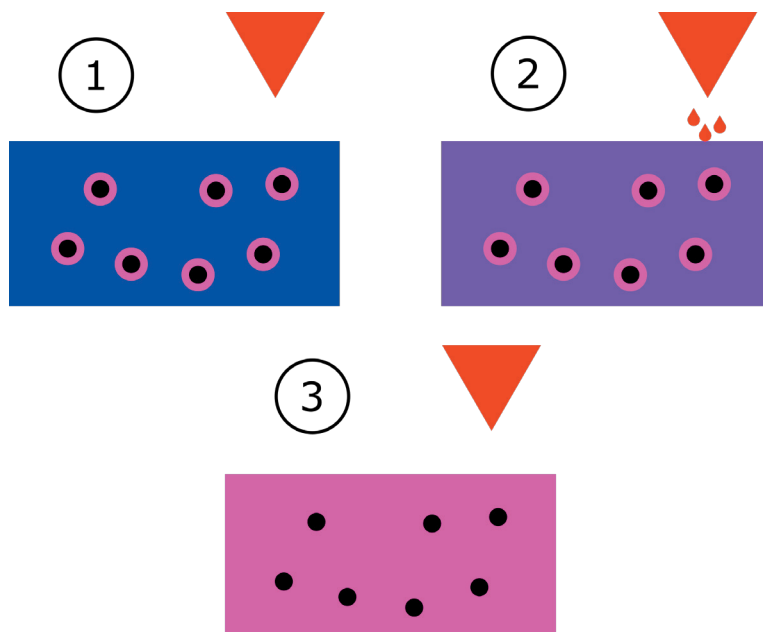
Fig. 8: Illustration of the concept of contrast matching. In step ① there are micelles with a corona (pink) dissolved in a solution (blue). The scattering length density of the corona is between the SLD of the solvent and its deuterated counterpart (red). In step ② the deuterated solvent is added to the solution, which changes the contrast conditions. Finally, in step ③ a sufficient amount of deuterated solvent has been added, so the contrast between the corona and the solvent has vanished. Now the micellar cores can be measured directly.
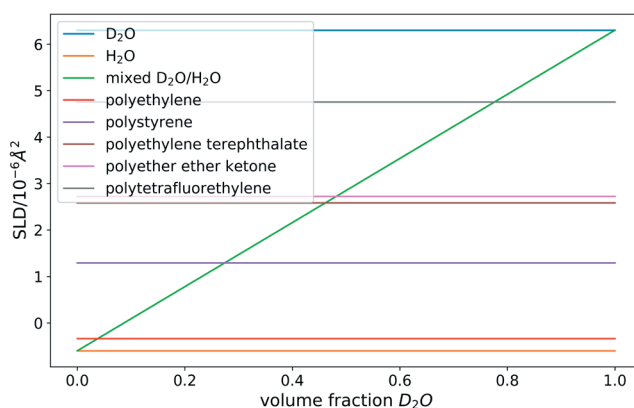
Fig. 9: Semi-analytic way to determine the necessary solvent deuteration for contrast matching. The concentration at the matching point, where the solvent has the same SLD as the polymer particles, is determined by the crossing of the mixed D_2O/H_2O SLD line and the SLD line of the respective polymer. For the calculation the scattering length density of water is calculated to $-0.6 \cdot 10^{-6} \text{Å}^2$ and the SLD of heavy water is calculated to $6.3 \cdot 10^{-6} \text{Å}^2$.

polymer or other sample is synthesized with a different isotope composition. Here the finding of the correct H/D fraction of the solvent shall be shown. Fig.9 gives an example of how to find the correct H/D fraction in a semi-analytic way. The underlying principle is expressed by

$$SLD_{sample} = SLD_{H_2O} \times H + SLD_{D_2O} \times D \tag{29}$$
$$H \equiv 1 \tag{30}$$
$$D = \frac{SLD_{sample} - SLD_{H_2O}}{SLD_{D_2O}}. \tag{31}$$

This way the volume of heavy water for each unit volume of protonated (usual) water can be calculated. It is also apparent from that calculation that only mixtures with a scattering length density between water and heavy water can be matched, and that the equations above only cover the non-trivial cases, where pure water or heavy water is not suitable. The actual volumes can then be calculated with $V_{\text{water}} = \frac{H}{H+D}$ and $V_{\text{heavy water}} = \frac{D}{H+D}$.

A prominent example for contrast matching is the matching out of the shell or core of a micelle. The contrast behavior and the resulting scattering curves are shown in Fig.10. Essentially contrast matching can improve the fitting procedure, if well known parts of the structure are matched out or emphasized by the contrast matching. This then delivers two or more different data sets that all should return comparable results. Another option is the reconstruction of embedded particles in a larger structure. Also here, the overall fitting procedure can profit from two fits with mutually corroborating results.

One concept that shall also be mentioned here is magnetic (spin-) contrast. In this context Fig.8 can be understood to be particles with a magnetic shell. As long as the spins are not aligned there is no contrast between the shell and the solvent (step ③). When an external magnetic field aligns the spins in the shell, a contrast between the shell and the solvent emerges (①). Several other possibilities with and without polarization analysis are possible, however that is beyond the scope of this manuscript.

8   Form factors

As described above, the phase problem usually prevents an analytic reconstruction of the structure from the scattered intensity by an inverse Fourier transform. There are approaches attempting the direct reconstruction of direct space information [6] or reconstruction from bead model annealing / Monte Carlo simulation [7, 8]. All these approaches have in common that a direct analytic expression for the scattering is not foreseen, and can therefore not be used as a starting point of the analysis. In the past, the model based analysis has been the most applied approach for the analysis of small-angle scattering data. Here, predetermined structures undergo a Fourier transform, whose result is then used to calculate a scattering pattern. This results in the most cases in analytic expressions that can be directly fitted to the data and are often used in a catalog-like manner in order to determine the structure of the sample. As most geometric forms can be approximated either as a sphere, a disk or a rod (see Fig.11) these are the forms that are going to be discussed here. More elaborate structures are available and can in
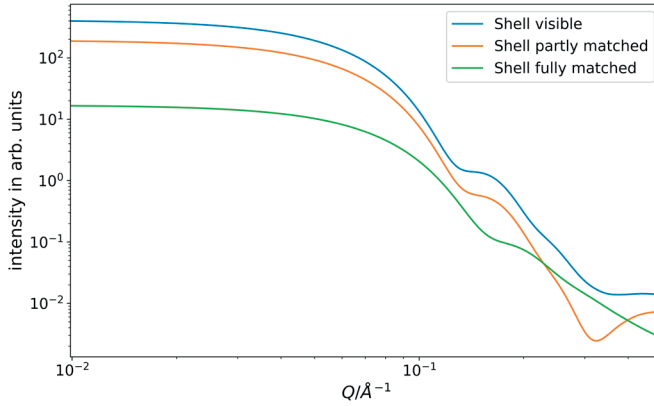
Fig. 10: Scattering curves for micelles with unmatched, partially matched and completely matched corona. The curves correspond to the scenarios ①, ② and ③ in Fig.8. Here two effects can be observed. The corona is only 50% of the radius of the core, hence it influences the scattered intensity at higher angles than the core itself, the scattering feature at $Q=0.15\text{Å}^{-1}$ corresponding to the micellar core is therefore quite stable, while the intensity at higher Q changes drastically. Considering the forward scattering the dependence of the scattering contrast between solvent and core is directly visible. The matched out corona shows the least contrast, and therefore the lowest forward scattering intensity, while the unmatched corona has the highest contrast and the highest intensity. This approach is also used, when an analytic approach to find the matching D_2O/H_2O concentration cannot be found. Several concentrations are tested and where a minimum in the scattered intensity is found, the contrast can be assumed to be matched.

principle be calculated for any structure where the form can be described by an analytic expression. A short, and by no means complete, list of programmes for the evaluation of SAS data is SasView (https://www.sasview.org), SasFit (https://kur.web.psi.ch/sans1/SANSSoft/sasfit.html) and Scatter (http://www.esrf.eu/UsersAndScience/Experiments/ CRG/BM26/SaxsWaxs/DataAnalysis/Scatter#).

Sphere

The analytic expression for the scattering created by a sphere of radius R is

$$I(Q) = N \left[3V\rho_0 \cdot \frac{\sin(QR) - QR\cos(QR)}{(QR)^3} \right]^2 \tag{32}$$

with N being the number of the scattering particles, V being the volume of a single sphere and ρ_0 being the SLD contrast between the sphere and the solvent.

This expression can be reached by using a SLD description like a step function as depicted in Fig.12. As a sphere is already spherically symmetric this can be directly put into the Fourier transform
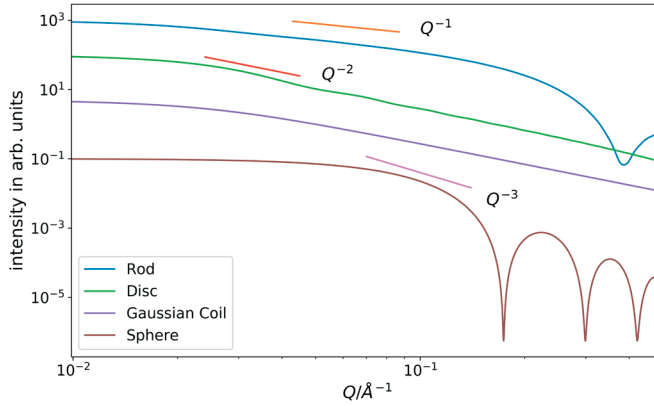
Fig. 11: Form factors for several scattering geometries. The slopes at the onset of the form factor after the plateau are shown, which is mostly determined by the fractal dimension of the scattering object. Here it also becomes apparent that solely relying on that slope may lead to misinterpretation between similarly scaling objects, here Gaussian coils and discs.

$$A(\mathbf{Q}) = \mathcal{F}(\rho(\mathbf{r})) \cdot (2\pi)^3 \tag{33}$$

$$= \int_V \rho(\mathbf{r}) \exp\left(-i\mathbf{Q}\mathbf{r}\right) dV \tag{34}$$

$$= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r=0}^{R} \rho(\mathbf{r}) \exp\left(-i\mathbf{Q}\mathbf{r}\right) r^2 \sin\theta \, dr \, d\theta \, d\phi \tag{35}$$

$$= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r=0}^{R} \rho(r) \exp\left(-iQr\cos\theta\right) r^2 \sin\theta \, dr \, d\theta \, d\phi \tag{36}$$

$$= \int_{\phi=0}^{2\pi} \int_{u=-1}^{1} \int_{r=0}^{R} \rho(r) \exp\left(-iQru\right) r^2 \, dr \, du \, d\phi \tag{37}$$

$$= 4\pi \int_{r=0}^{R} \rho(r) \left(\frac{\exp(iQru) - \exp(-iQru)}{iQr}\right) r^2 dr \tag{38}$$

$$= 4\pi \int_{r=0}^{R} \rho(r) \left(\frac{\sin Qr}{Qr}\right) r^2 dr \tag{39}$$

$$= 4\pi\rho_0 \int_{r=0}^{R} \frac{\sin Qr}{Qr} r^2 dr \tag{40}$$

$$= 4\pi\rho_0 \frac{sinQR - QR\cos QR}{Q^3} \tag{41}$$

$$= 4\pi\rho_0 \frac{sinQR - QR\cos QR}{Q^3} \tag{42}$$

$$= V\rho_0 \frac{3sinQR - QR\cos QR}{R^3 Q^3} \tag{43}$$

Here Eq.36 used the identity of $\mathbf{Q}\mathbf{r} = Qr\cos\theta$ with theta being the enclosed angle and in Eq. 37
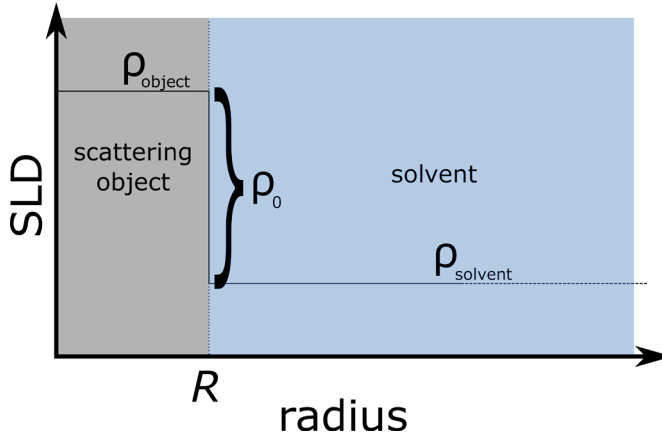
Fig. 12: Depiction of the SLD distribution along the radius of a sphere. ρ_0 is the SLD contrast, i.e. the SLD difference between the scattering particle and the solvent. R is the radius of the sphere.

$\cos\theta$ was replaced by u. In addition, spherical symmetry was exploited for the integration over the solid angle. The factor $(2\pi)^3$ is to correct for scaling differences during the Fourier transform.

This corresponds exactly to the squared term in Eq.32 which is nothing else than the squared amplitude that we calculated here. As this is only the scattering for a single, isolated sphere, the number density needs to be included to reflect the absolute scattered intensity. In case of neutron scattering this is the case for most of the instruments. X-ray instruments are often not calibrated to absolute scattering intensities and therefore need an arbitrary scaling factor. Similar approaches can be used for other analytic representations of form factors.

Thin Rod

The scattered intensity by a dilute solution of thin rods of length L is given by

$$I(Q) \quad = \quad \rho_0^2 v^2 \left(\frac{2}{QL\cos\theta} \right) \sin^2 \left(\frac{QL}{2}\cos\theta \right) \tag{44}$$

$$\rightarrow \quad \rho_0^2 v^2 \frac{2}{QL} \left(\text{Si}(QL) - \frac{1-\cos QL}{QL} \right). \tag{45}$$

Here v is the volume of the particle and the average over all orientations has been performed in the second step. The substitution $\text{Si}(QL) = \int_0^{QL} \frac{\sin u}{u} du$ was used.

Circular Disc

An infinitely thin circular disk of radius R scatters the incoming intensity as follows:

$$I(Q) = \rho_0 v^2 \frac{2}{Q^2 R^2} \left(1 - \frac{J_1(2QR)}{QR} \right) \tag{46}$$

J_1 here is the first order Bessel function.

Non-particulate scattering from a flexible chain

A flexible chain in solution cannot be described by a simple analytic form, since one needs to integrate over all possible conformations of the chain. Nevertheless, an analytic expression, the Debye scattering, can be found:

$$I(Q) = \rho_0^2 v^2 \frac{2(\exp(-Q^2 R_g^2) + Q^2 R_g^2 - 1)}{Q^2 R_g^2} \tag{47}$$

Here $R_g = \frac{1}{V} \int_V \mathbf{r}^2 \rho_0 d\mathbf{r}$ is the radius of gyration (in this case for constant SLD). A very important aspect of that scattering curve is, that it essentially scales with Q^2.

For better comparison the radius of gyration for a solid sphere of radius R is $R_g = \sqrt{\frac{3}{5}} R$, the one for a thin rod of length L is $R_g = \frac{1}{\sqrt{12}} L$ and the one for a very thin circular disc with radius R is $R_g = \frac{1}{\sqrt{2}} R$

Polydispersity

All analytic form factors, that deliver the scattered intensity, are determining the scattered intensity for particles of one exact size. In real systems, however, there are mostly distributions of different sizes. This leads to a superposition of scattering from different particle sizes. Since most particle sizes follow a Gaussian distribution, this is also a good way to fold in the particle size distribution analytically. For extremely long, or very polydisperse, particles then Schulz-Zimm distribution is used, which looks very similar to the Gaussian distribution, however has a cut-off at zero to prevent negative sizes of the particles. For specialized problems also other distributions, such as La-Place, multi-modal or other size distribution functions can be used.

The general idea is that the scattered intensity $I(Q, r)$ is folded with the size distribution function $f(r)$

$$I_{\text{real}}(Q, r) = I_{\text{ideal}}(Q, r) * f(r). \tag{48}$$

Here the subscripts real and ideal identify the real measured intensity or the ideal intensity for any calculated particulate size and form.

The effects of the convolution can be seen in Fig.13. Most notably, the minima are smeared out, and in some cases vanish completely, so they can only be estimated. Another important effect is that the slopes of inclinations cannot be completely reproduced anymore, which is especially important to distinguish scattering from different contributions. The magnitude of
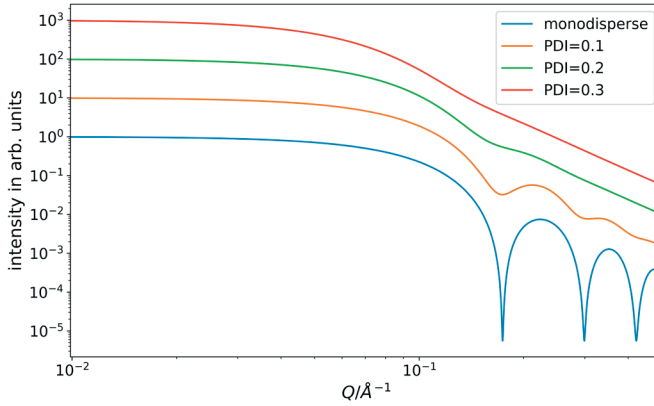
Fig. 13: Effect of polydispersity. While the positions of the minima can still be found at higher polydispersity, the higher order undulations of the form factor vanish.

the polydispersity is described by the polydispersity index $PDI = \sigma(f(r))/\mu(f(r))$ where $\sigma(f(r))$ is the standard deviation of the size distribution function and $\mu(f(r))$ is the mean of the size distribution function. Values of $PDI \geq 0.3$ are usually discarded during fitting, as then the results become unreliable in such a polydisperse sample.

In addition to this, the usual polydispersity (approximated by a Gaussian distribution) is by its very nature similar to a resolution smearing of the instrument itself. Therefore, it can easily happen to overestimate the polydispersity. If the resolution function of the instrument is known, it should be used for deconvolution before performing the fits.

9   Structure Factors

Structure factors in general describe the scattered intensity due to the arrangement of single particles. This can be because the solution is becoming to dense, and therefore the particles arrange following a nearest neighbor alignment or because the particles are attractive to each other and form aggregates. Thus, more generally a structure factor $S(Q)$ is a measure of interaction between the single particles in the solution and connected with the correlation function $c(r)$ (the probability to find a particle at a certain distance) with the relation

$$S(Q) = \frac{1}{1 - nc(Q)}. \tag{49}$$

Since the structure factor and the form factor need to be convolved in real space, in reciprocal space this converts to a multiplication, following the convolution theorem. Therefore the scattered intensity, described by form factor $F(Q)$ and structure factor $S(Q)$
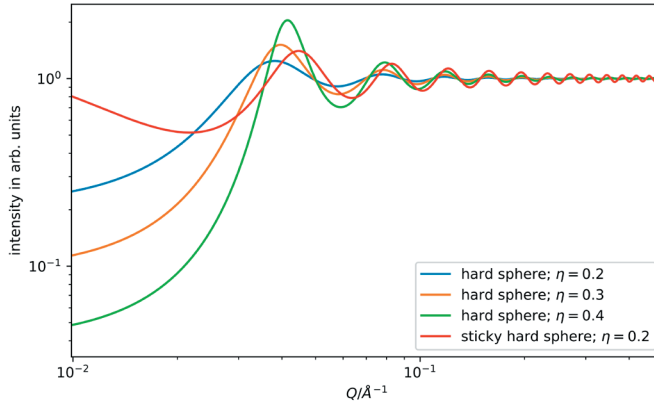
$$I(Q) = F(Q) \cdot S(Q). \tag{50}$$

Fig. 14: Examples for structure factors. The intensity of the peaks roughly scales with the volume fraction η of the particles. Also the position of the peaks is slightly dependent on that volume fraction, which makes a direct calculation of $R = \frac{2\pi}{Q_{max}}$ invalid (The hard sphere radius used here was 60 Å). A distinct difference can be noted at low Q. Here, in general, attractive interaction (sticky hard spheres) leads to an increase in scattering, while repulsive interaction leads to a decrease in intensity.

From this equation it also follows, that for a system of uncorrelated, identical particles the structure factor must be $S(Q) = 1$. Since the correlation between particles usually leads to either an aggregation or repulsion of particles over long length scales the contribution of the structure factor is most prominent at low Q-values. Also, this means that for large distances the structure factor has to level out to unity, to preserve the fact that at large Q only the inner structure of the particle is visible, not its arrangement in space. A few instructive examples for the structure factor are shown in Fig.14.

Hard Sphere Structure Factor

The hard sphere structure factor assumes an infinitely high potential below a radius R and a zero potential at higher radii. This can be described by

$$V(r) = \begin{cases} \infty \text{ for } r \leq R \\ 0 \;\; \text{for } r > R. \end{cases} \tag{51}$$

Using Eq.49 this can be rewritten as

$$S(Q) = \frac{1}{1 + 24\eta_{HS}G(2QR)/2QR}. \tag{52}$$

Here $G(x)$ is defined as

$$G(x) = \alpha \frac{(\sin(x) - x\cos(x))}{x^2} + \tag{53}$$

$$= \beta \frac{(2x\sin(x) + (2 - x^2)\cos(x - 2))}{x^3} + \tag{54}$$

$$= \gamma \frac{(-x^4\cos(x) + 4[(3x^2 - 6)\cos(x) + (x^3 - 6x)\sin(x) + 6])}{x^5} \tag{55}$$

with these definitions for α, β and γ:

$$\alpha = \frac{(1 + 2\eta_{HS})^2}{(1 - \eta_{HS})^4} \quad ; \quad \beta = \frac{6\eta_{HS}(1 + \eta_{HS}/2)^2}{(1 - \eta_{HS})^4} \quad ; \quad \gamma = \frac{\eta_{HS}/2(1 + 2\eta_{HS})^2}{(1 - \eta_{HS})^4}. \tag{56}$$

In all equations the volume fraction that is occupied by hard spheres of radius R is designated η_{HS}.

10   Reading a curve

In an experimental environment it can be useful to determine the fundamental features in a preliminary fashion without computer aided data evaluation, also known as fitting. In addition, this helps determining good starting parameters for fits. In order to do so, we are going to look at the curves shown in Fig.15. There we can determine different regions of the scattered intensity (forward scattering, Guinier regime, Debye regime and Porod regime) and determine several properties of the sample from that intensity. When applying the described techniques for directly reading a curve it has to be kept in mind that most of them are either restricted in their validity concerning the Q-space or are very general and rough descriptions of the sample.

Forward scattering

As pointed out in the discussion of the structure factor, large aggregates mostly show their presence by an increased scattering intensity at low Q. This also becomes apparent when taking Eq.3 into account. This means, in general, an increased scattering at low Q is indicative of large aggregates being present in the sample. This also correlates with an attractive potential between the single particles.

Another possibility is strongly suppressed scattering at low Q. This can be the case for strongly repulsive interaction potentials between the particles, close to what is described for the hard sphere factor above.

A leveling out of the intensity at low Q is indicative of an either dilute solution or a very weak potential between the particles. Then there is no influence at low Q and only the structure factor of the single particles is visible.
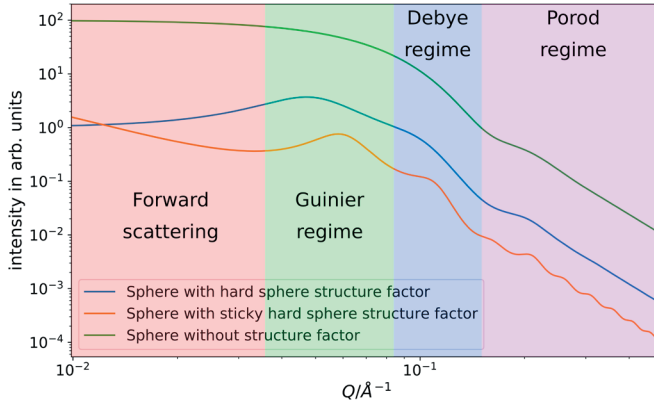
Fig. 15: Diverse scattering curves from identical spherical form factor and different structure factors.

Guinier regime

The Guinier regime is usually the crossover region, where the forward scattering is not dominant anymore and the slope of the scattering curve changes to the scattered intensity of the form factor. In this regime the overall size of the particle can be examined. This is similar to seeing something from far away: One may be able to discern the size of the particle but the distinct form remains hidden. Imagine a football and a pumpkin seen from 100 m away. They are close in size, you can properly judge it to be approximately 20 cm in diameter, but the exact form (ridges, stem of the pumpkin) remains hidden. A description that is only taking into account the scattered density of the particles as a whole, valid in that scattering regime is the Guinier Law:

$$I(Q) = \rho_0 v^2 \exp(-\frac{Q^2 R_g^2}{3}) \tag{57}$$

For details of derivation, which include a Taylor series expansion around zero of the scattered amplitude (Eq.35) and an averaging over all directions, please refer to the literature. [9, 10] Another option is to develop a series expansion for the Debye Law (Eq.47) at low Q.

In order to evaluate the data using the Guinier Law, the data needs to be plotted as shown in Fig.16. The log-log representation and plotting versus Q^2 allow to directly read the inclination of the system, multiply by 3 and use the square root in order to retrieve the particle radius.

Debye regime

In contrast to the Guinier regime, where the data can be evaluated by the Guinier law, the Debye regime signifies the area, where the particulate form manifests in the scattering, which in general cannot be fitted by the Debye law. The Debye law is only valid for the scattering from Gaussian chains. As can be seen in the form factors section 8, there is a direct correlation to the dimen-
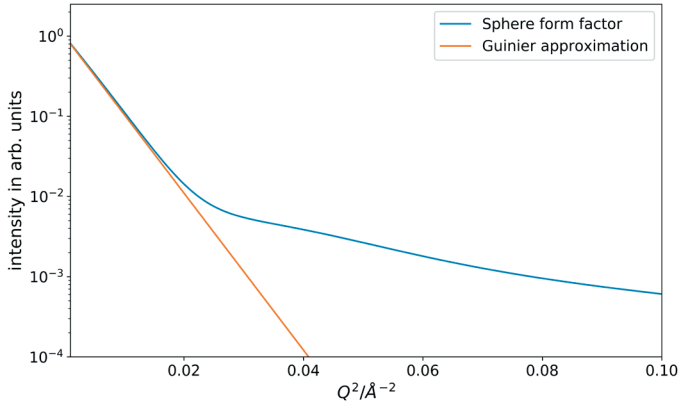
Fig. 16: Sphere form factor and Guinier approximation from Eq.57 in a Guinier plot. The radius of gyration is 25.8 Å. The estimated slope by eye was $m = -185$. With $R_g = \sqrt{-3 \cdot m} = 23.5$Å the error is within 10%, which is suitable for a naked eye approximation.

sionality of the scattering particle (sphere, disc, rod) and the slope in log-log plot, since the scattering scales with $I(Q) \sim Q^{-D}$, where D is the dimensionality of the scattering object (sphere: $D = 3$; disc: $D = 2$; rod: $D = 1$). Also the scattering from fractal objects is possible, which then results in non-integer numbers for the slope. It should be noted that this is an approximation that is only valid for the case when $1/$particle radius $\ll Q \ll 1/$fundamental building block. The fundamental building block in this case can be for example atoms or single monomers of a chain.

Porod regime

The Porod regime, is the regime where the interface between the particle and the solvent dominates the scattered intensity. It is valid for large Q (before leveling out into the incoherent background) and therefore a good approach is extrapolating the sphere form factor to large Q. The decisive property of the scattered intensity is the scaling of $I(Q) \sim Q^{-4}$. This behavior can be derived from an extrapolation of the sphere form factor (Eq.32) to very large Q:

$$I(Q) \ \propto \ \left(\frac{4}{3}\pi R^3\right)^2 \frac{9(\sin QR - QR\cos QR)^2}{Q^6 R^6} \tag{58}$$

$$= \ 8\pi^2 \left(\frac{R^2(1 + \cos 2QR)}{Q^4} - \frac{2R\sin 2QR}{Q^5} + \frac{1 - \cos 2QR}{Q^6}\right) \tag{59}$$

The higher order terms vanish at large Q delivering the characteristic Q^{-4} behavior of the scattered intensity. Here only proportionality is claimed, which is strictly true in this case. If the scattered intensity is recorded in absolute intensities, here also information about the surface of the particles can be obtained. This then follows the form

$$\lim_{Q \to \infty} I(Q) = \frac{8\pi \Delta \rho S}{Q^4}.$$

(60)

$\Delta \rho$ is here the SLD difference between the particle and the surrounding medium and S the interfacial area of the complete sample between particles and medium. This means, the absolute intensity of the Porod regime allows to determine the complete amount of surface in the sample.

Estimation of particle and feature Size

As described previously for low Q in most cases it is a good approximation to assume all particles in the sample have spherical symmetry (Section 10). The roots of the expression for the spherical form factor are in the locations $\tan(QR) = QR$, which is true for $QR \approx 4.49, 7.73, 10.90.....$ In many cases anyway only the first minimum of the form factor will be visible. This allows a fast approximation of the radius with $R \approx 4.5/Q_{min}$. Here it needs to be noted, that this is the rotational average of the particle, neglecting any structure of the particle whatsoever.

Another approach of determining the size or correlation of features is using Eq.3:

$$d = \frac{2\pi}{Q}.$$

Although this is in general only strictly true for lamellar systems and the corresponding correlations, it is still a good approximation for a summary data examination during the experiment. With that restriction in mind it can be used for virtually any feature in the scattering curve and the size of the corresponding feature in the sample.

11 Further Reading

Most of the concepts shown in this manuscript are based on previous publications. The following selection of textbooks gives the reader a good overview of the principles of SAS.

A. Guinier: X-ray diffraction in crystals, imperfect crystals, and amorphous bodies

This early textbook concentrates on SAXS, as neutron scattering at the time of writing was still in its infancy. While some of the terminology may have changed slightly over time, in many aspects this book still gives a good fundamental overview of what can be done with small-angle scattering, and how to perform a solid data analysis. In addition, this is literally the book on the Guinier Law, and where some of the basic ideas of reading scattering curves were first collected.

R.J. Roe: Methods of x-ray and neutron scattering in polymer science

Here the author nicely manages to emphasize the commonalities and differences between x-ray and neutron scattering. An overview of the methods and technologies is given, as well as a helpful mathematical appendix, reiterating some of the concepts used in the book.

G. Strobl: The physics of polymers

For soft-matter researchers this book, even though not being focused on scattering as such, gives a good overview of applicable concepts for scattering with soft-matter samples. A wide range of helpful examples highlight in which particular area any evaluation concept of the data is applicable and useful.

References

[1] A. Guinier, P. Lorrain, D. S.-M. Lorrain, J. Gillis, *X-ray diffraction in crystals, imperfect crystals, and amorphous bodies*, Physics Today **17**, 70 (1964).

[2] H. Stabinger, O. Kratky, *A new technique for the measurement of the absolute intensity of x-ray small angle scattering. The moving slit method*, Die Makromolekulare Chemie **179**, 1655.

[3] E. Kentzinger, M. Krutyeva, U. Rücker, *GALAXI: Gallium anode low-angle x-ray instrument*, Journal of large-scale research facilities JLSRF **2**, 61 (2016).

[4] H. Haubold, P. Hiller, H. Jungbluth, T. Vad, *Characterization of Electrocatalysts by in situ SAXS and XAS Investigations*, JAPANESE JOURNAL OF APPLIED PHYSICS-SUPPLEMENT- **38**, 36 (1999).

[5] S. Roth, G. Herzog, V. Körstgens, A. Buffet, M. Schwartzkopf, J. Perlich, M. A. Kashem, R. Döhrmann, R. Gehrke, A. Rothkirch, et al., *In situ observation of cluster formation during nanoparticle solution casting on a colloidal film*, Journal of Physics: Condensed Matter **23**, 254208 (2011).

[6] B. Weyerich, J. Brunner-Popela, O. Glatter, *Small-angle scattering of interacting particles. II. Generalized indirect Fourier transformation under consideration of the effective structure factor for polydisperse systems*, Journal of Applied Crystallography **32**, 197 (1999).

[7] A. Koutsioubas, S. Jaksch, J. Pérez, *DENFERT version 2: extension of ab initio structural modelling of hydrated biomolecules to the case of small-angle neutron scattering data*, Journal of applied crystallography **49**, 690 (2016).

[8] T. D. Grant, *Ab initio electron density determination directly from solution scattering data*, Nature methods **15**, 191 (2018).

[9] A. Guinier, *x-ray diffraction in crystals, imperfect crystals and amorphous bodies*.

[10] R. Roe, *Methods of x-ray and neutron scattering in polymer science*.

D 2   Scattering under grazing incidence

E. Kentzinger

Quantum Materials and Collective Phenomena
Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

1   Introduction

Neutron reflectometry is a very efficient tool to determine the nuclear and magnetic density profiles along the depth of nanometric thin films. It has been used a lot to solve soft matter problems like the self organization of diblock copolymers, the structure of liquid-liquid interfaces or the structure of biomembranes [1]. Those studies benefit a lot from the possibility of contrast variaiton, i.e. the exchange of hydrogen by deuterium.

In the mid 1980's a new field of application of neutron reflectometry emerged. Following the discoveries of interlayer exchange coupling and giant magnetoresistance effect in magnetic multilayers [2], there has been an interest to determine, depth-resolved, the magnetic profile (see lecture 10 of this book).

More recently, the interest evolved towards the determination of the correlations of in-plane fluctuations in thin films. Those fluctuations can be nuclear or magnetic, in the bulk of the layers or at their interfaces, or nanometric objects deposited on a surface. The breaking of in-plane invariance introduced by those fluctuations produce scattering of radiation out of the specular direction, called grazing incidence small angle scattering (GISAS).

In this lecture, we will concentrate on neutron and x-ray reflectometry and GISAS for the determination of nuclear and chemical profiles. Section 2 shows the calculation of specular reflection at flat and homogeneous surfaces, introducing the concepts of scattering length density, index of refraction and total external reflection. It then describes the reflectivity from various types of layered structures and the effect of interfacial roughness and interdiffusion. The two types of reflectometers one can encounter and the practical aspects of a reflectometry experiment are discussed in section 3. Finally, an example of the application of grazing incidence small angle x-ray scattering (GISAXS) for the depth-resolved investigation of the lateral arrangement of nanoparticles is depicted (section 4).

2   Description of specular reflection[1]

A monochromatic, well collimated beam impinges under a well defined, small angle $\alpha_i = \theta$ (in most cases $\theta \ll 5^o$) onto the surface of the sample. It is then partly reflected specularly from the surface, i.e. the outgoing angle $\alpha_f = \theta$ as well, and partly refracted into the material (See Fig. 1). As we will derive below, the reflection from a laterally homogeneous medium can be treated according to classical optics. Only the proper index of refraction n has to be used.

For most material, the index of refraction for neutrons is slightly smaller than 1, leading to total external reflection for small angles of incidence $\theta < \theta_c$, where θ_c depends on the material.

In the case of a single layer on the substrate, reflection and refraction take place at both the surface and the interface (Fig. 2). Then, the reflected beams from the different interfaces interfere with each other. Maximum intensity is received, when the path length difference between the two reflected beams is an integer multiple of the wavelength.

For the case of perfectly smooth surface and interfaces, an exact description of the reflected

---

[1] A large part of this section is taken from Ref. [3–6].

and transmitted intensity can be deduced from quantum theory, as will be shown in the next subsections.

When the in-plane invariance of the layers or interfaces is broken, some diffuse signal can be observed out of the specular direction (Fig. 3). This is grazing incidence small angle scattering (GISAS). Its theoretical description goes beyond the scope of this lecture [7, 8].
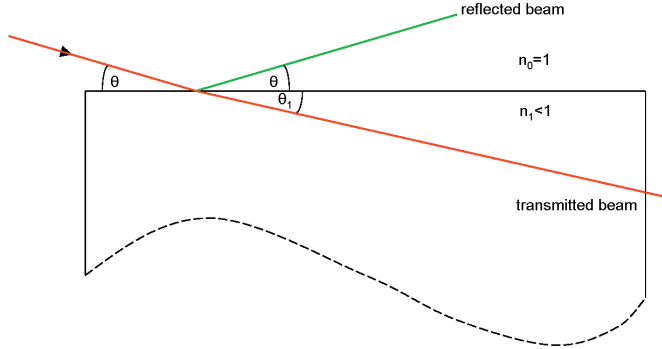

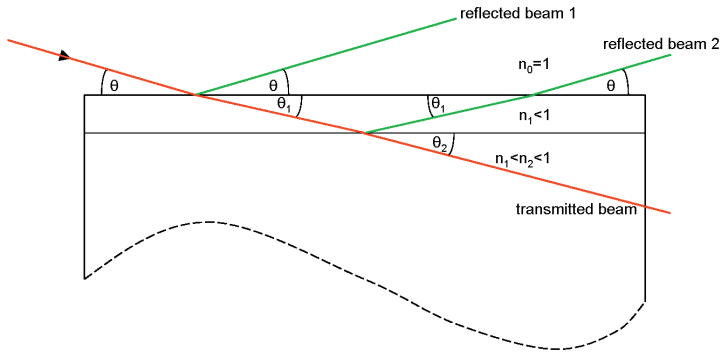**Fig. 1:** Reflection and refraction from a free surface

Fig. 2: Reflection and refraction from a single layer on a substrate

Wave equation in homogeneous medium. Optical index

The starting point is the Schrödinger equation for the wave function of the neutron:

$$\left[-\frac{\hbar^2}{2m}\Delta + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1}$$

The kinetic energy of the neutron is given by $E = \hbar^2 k^2/(2m)$ with the modulus $k = 2\pi/\lambda$ of the wave vector $\mathbf{k}$.
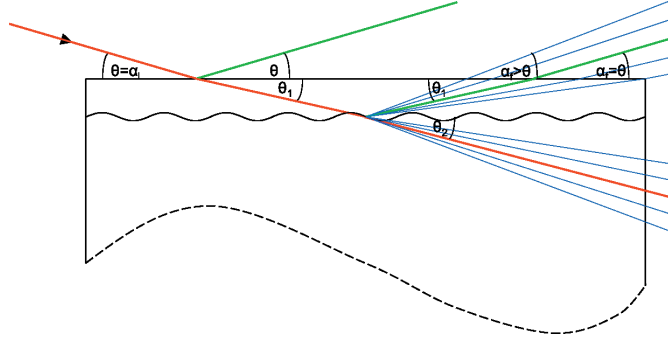
Fig. 3: Reflection, refraction and grazing incidence small angle scattering (GISAS) from a single layer on a substrate showing broken in-plane invariance

Due to the small $|\mathbf{Q}|$ values that are probed, a reflectometry experiment does not resolve the atomic structure of the sample in any of the three directions. Therefore, it is a valid approximation to describe the potential V_1 of the homogeneous material as

$$V_1 = \frac{2\pi\hbar^2}{m}\rho \qquad (2)$$

where ρ if the scattering length density (SLD) defined by

$$\rho = \sum_j N_j b_j \qquad (3)$$

where N_j is the number of nuclei per unit volume and b_j is the coherent scattering length of nucleus j. With that we receive

$$\left[\Delta + \left(k^2 - 4\pi\rho\right)\right]\psi(\mathbf{r}) = \left[\Delta + k^2\left(1 - \frac{\lambda^2}{\pi}\rho\right)\right]\psi(\mathbf{r}) = \left[\Delta + k_1^2\right]\psi(\mathbf{r}) = 0 \qquad (4)$$

with the wave vector k_1 inside the medium. From this equation, it is justified to introduce the index of refraction in the material

$$n = \frac{k_1}{k} \qquad\qquad n \simeq 1 - \frac{\lambda^2}{2\pi}\rho \qquad (5)$$

It is a number very close to 1 for thermal and cold neutrons. The quantity $1 - n$ is of the order of 10^{-6} to 10^{-5}. For most materials it is positive (because the coherent scattering length b_j is positive for most isotopes), so that n is smaller than 1. This means that the transmitted beam is refracted towards the sample surface, which is opposite to the daily experience with light refracted at a glass or liquid surface.

Solution for a sharp surface. Fresnel's formulas

In analogy to classical optics, we can derive e.g. Fresnel's formulas. For the solution of the wave equation at a sharp surface between air and a semi-infinite medium, we assume the surface of the sample to be at $z = 0$. The potential is then

$$V(z) = \begin{cases} 0 & for \quad z > 0 \\ V_1 & for \quad z \leq 0 \end{cases} \tag{6}$$

As the potential V is independent of the in-plane coordinates x and y, the wave function in the Schrödinger equation (4) is of the form

$$\psi(\mathbf{r}) = e^{i(k_x x + k_y y)} \psi_z(z) \tag{7}$$

with the in plane components k_x and k_y of $\mathbf{k}$ independent of z. The Schrödinger equation then reduces to the one dimensional equation

$$\frac{d^2 \psi_z(z)}{dz^2} + k_z^2(z)\psi_z(z) = 0 \tag{8}$$

with $k_z(z)$ depending on the medium. The general solution is given by

$$\psi_{zl}(z) = t_l e^{ik_{zl}z} + r_l e^{-ik_{zl}z}, \tag{9}$$

where the index l distinguishes between vacuum (l=0) and medium (l=1). The unique solution is determined by the boundary conditions. The incoming wave in the vacuum before interaction with the sample is a plane wave of norm 1, i.e. t_0 is equal to 1. In a half-infinite medium, there is no reflected wave, because there is nothing to reflect from, i.e. r_1 vanishes. In addition, the wave function and its first derivative must be continuous at the interface. So we receive the following boundary conditions:

$$t_0 = 1 \;\; ; \;\; r_1 = 0 \;\; ; \;\; \psi_{z0}(z = 0) = \psi_{z1}(z = 0) \;\; ; \;\; \frac{d\psi_{z0}}{dz}(z = 0) = \frac{d\psi_{z1}}{dz}(z = 0). \tag{10}$$

When we insert (9) into (10) we receive the continuity equations for the wave function:

$$1 + r_0 = t_1 \;\; ; \;\; k_{z0}(1 - r_0) = k_{z1}t_1. \tag{11}$$

t_1 is the amplitude of the transmitted wave and r_0 is the amplitude of the reflected wave. The reflectivity R is defined as the modulus squared of the ratio of the amplitudes or reflected and incoming waves, the transmissivity T is defined as the modulus squared of the ratio of the amplitudes or transmitted and incoming waves.

$$R = |r_0|^2 \;\; ; \;\; T = |t_1|^2 \tag{12}$$

In conclusion, we arrive at the Fresnel's formulas for the reflection and the refraction at a flat interface

$$\text{Reflectivity}: \qquad R = \left| \frac{k_{z0} - k_{z1}}{k_{z0} + k_{z1}} \right|^2 \tag{13}$$

$$\text{Transmissivity}: \qquad T = \left| \frac{2k_{z0}}{k_{z0} + k_{z1}} \right|^2 \tag{14}$$

Snell's law of refraction. Total external reflection

Taking into account the continuity relation for the wave vector component tangential to the surface

$$k_{x0} = k_{x1} \qquad k_{y0} = k_{y1} \tag{15}$$

together with $k_1 = k_0 n_1$ (Eq. 5), Snell's law for refraction follows from trigonometry:

$$\frac{\cos \theta}{\cos \theta_1} = \frac{k_1}{k_0} = n_1 \tag{16}$$

The fact that in most cases the index of refraction is $n_1 < 1$ means that the transmitted beam is refracted towards the sample surface ($\theta_1 < \theta$ in Fig. 1). For angles of incidence θ below the so called critical angle θ_c with

$$n_1 = \cos \theta_c \qquad \theta_c \simeq \lambda \sqrt{\frac{\rho}{\pi}} \tag{17}$$

total reflection is observed, i.e. all intensity is reflected and no wave propagating in z-direction exists in the sample. Only an evanescent wave in the z-direction with propagation parallel to the surface is induced. For angle of incidence above θ_c, the beam can partially penetrate the sample and is only partly reflected.

From Snell's law (Eq. 17) and the definition of the index of refraction in Eq. (4) one can relate the normal components of the incoming and refracted wave vectors

$$k_{z1}^2 = k_{z0}^2 - k_{z0,c}^2 \quad \text{with} \quad k_{z0,c} = \frac{2\pi}{\lambda} \sin \theta_c = \sqrt{4\pi\rho}. \tag{18}$$

This confirms that, for angles of incidence θ below θ_c, k_{z1} becomes purely imaginary and the refracted wave is an evanescent wave in the z-direction.

The last relation allows to express the Fresnel coefficients (Eq. 13 and 14) as a function of one variable only. In general the measured reflectivity is represented as a function of θ or the magnitude of the scattering wave vector $Q = 2k_{z0}$:

$$R = \left| \frac{Q - \sqrt{Q^2 - Q_c^2}}{Q + \sqrt{Q^2 - Q_c^2}} \right|^2 \tag{19}$$

When $Q \gg Q_c$, the preceding equation reduces to:

$$R \simeq \frac{1}{16} \frac{Q_c^4}{Q^4} \tag{20}$$

which is the formula for the reflectivity within the Born approximation [9]. This shows that the reflectivity above the critical angle decreases sharply with Q.

Once again, coming back to the wave function inside the surface, one finds using Eq. (18) that, when $\theta < \theta_c$:

$$\psi_{z1}(z) = t_1 e^{i\left(k_{z0}^2 - k_{z0,c}^2\right)^{1/2} z} = t_1 e^{-\frac{1}{2}\left(Q_c^2 - Q^2\right)^{1/2} z}. \tag{21}$$

This result is very important, because it shows that when the energy of the particle normal to the surface is smaller than the potential barrier, the wave still can penetrate the medium on a characteristic depth of $2/\sqrt{Q_c^2 - Q^2}$. This evanescent wave propagates itself along the surface with a wave vector equal to (k_x, k_y) and then leaves the volume in the specular direction. For example for Ni ($\rho = 9.41 \times 10^{-6}$ Å^{-2}), the penetration depth is of the order of 200 Å at $Q = 0$; if one neglects absorption, it raises rapidly to infinity at $Q = Q_c$. No conservation rule is broken: the reflectivity equals 1 because this wave represent no transmitted flux in the medium.

Fig. 4 represents, on a linear scale, the reflectivity and the transmissivity of a substrate as a function of the angle of incidence θ. The reflectivity equals 1 for angles smaller than the critical angle θ_c and decreases rapidly above this value (Eq. 20). The transmissivity increases monotonously up to a value of 4 at θ_c and decreases to 1 at large angles. This result might look very surprising at first sight. The value of 4 for the transmissivity comes from the fact that the incident and the reflected waves in vacuum superpose to form a stationary wave of amplitude exactly equal to 2 at the interface with the medium. For the intensity, we obtain a factor of 4.
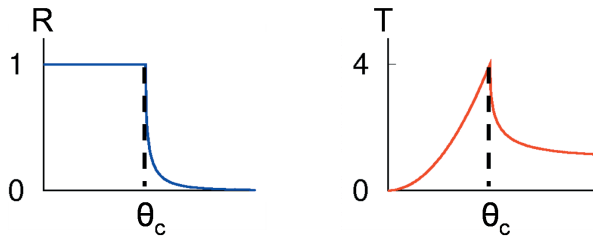

Fig. 4: Reflectivity and transmissivity of a substrate as a function of the angle of incidence

Reflectivity from layered systems

In a layered system, the same Ansatz as in Eq. (9) can be written in each layer l. The coefficients of reflection r_l and transmission t_l can be deduced recursively from the continuity relations of the wave function and its derivative at each interface. If N is the number of layers, and considering the vacuum on top of the multilayer and the substrate below, 2(N+2) coefficients have to be calculated. The number of interfaces being N+1, the continuity relations lead to 2(N+1) equations. Two other equations are obtained considering that the transmission into the vacuum is equal to one ($t_0 = 1$) and that, in the substrate, there is no reflected wave ($r_{N+1} = 0$), leading in total to a number of equations equal to the number of coefficients to determine. The calculation of the coefficients of reflection and transmission in each layer and, in particular, the calculation of the reflectivity in air are therefore possible [10].

Here we just want to demonstrate with very simple arguments how interference effects from layered structures arise and how the intensity modulations in Q-space are related to real space length scales.

Fig. (2) shows how interference can occur in a system composed of a single layer of thickness d deposited on a substrate. Interference occurs between beams reflected from the surface and those first transmitted in the layer, reflected from the interface between layer and substrate and then leaving the layer into vacuum. To a good approximation, refraction at the top surface can be neglected for incident angles twice the critical angle or total reflection. In this case $\theta = \theta_1$ in Fig. (2) holds. Since the index of refraction of the neutrons is very close to one, this approximation is valid even for rather small angles of incidence. Then the optical path length difference between the two beams is:

$$\Delta = 2d \sin \theta \qquad (22)$$

We can now determine the distance between interference maxima from the condition that the path length difference has to differ by one wavelength: $\lambda = 2d \cdot \delta(\sin \theta) \simeq 2d \cdot \delta\theta$. With $Q = \frac{4\pi}{\lambda} \sin \theta \simeq \frac{4\pi}{\lambda}\theta$ we final obtain:

$$\delta Q \simeq \frac{2\pi}{d} \qquad (23)$$

We can see that the interference phenomena in Q-space are connected with real space length scales in a reciprocal way. (23) tells us that there will be a number of interference maxima at a distance in Q of $\frac{2\pi}{d}$. These interference phenomena are called "Kiessig fringes". Fig. 5 shows calculations of the reflectivity of a Ni layer deposited on a Si substrate. One observes that the reflectivities above the critical angle for total reflection decrease rapidly, therefore the ordinate is on a logarithmic scale. The oscillations of the reflectivity due to the above described interference effect can be observed. At small angles, due to the effect of refraction, the interference maxima are a bit denser distributed than at higher angles where formula (23) can be used to determine the layer thickness from the distance between the interference maxima. The thinner layer corresponds to an interference scheme with a bigger period. In both cases the minima of the interference scheme lay on the reflectivity of the Si substrate.

Note that for a 100 Å thick layer of Ni, that has a scattering length density (SLD) approximately
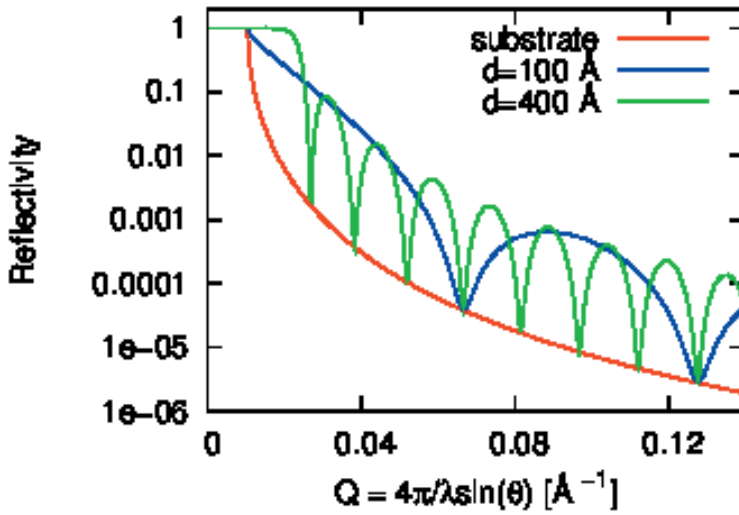
Fig. 5: Reflectivity of a Si substrate and reflectivity of a Ni layer ($\rho = 9.41 \times 10^{-6}$ Å^{-2}) on Si substrate ($\rho = 2.15 \times 10^{-6}$ Å^{-2}). Simulations are performed for two layer thicknesses d.

4 times larger than the one of Si, the critical angle of total reflection is determined by the SLD of Si and not by the one of Ni. This comes from the penetration depth of the neutrons that is bigger than 100 Å. For a 400 Å thick Ni layer, the θ_c approaches the one of Ni and the total reflection plateau is somewhat rounded.
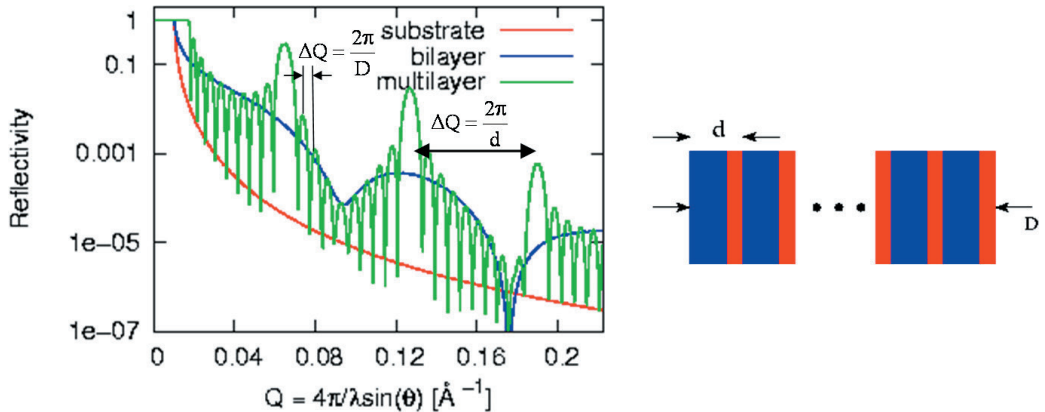
Fig. 6: Reflectivities of a Ni/Ti bilayer and of a Ni/Ti multilayer on Si substrate. Simulations are performed for Ni and Ti thicknesses of 70 and 30 Å respectively.

Fig. 6 shows the simulation of the neutron reflectivity from a multilayer on a Si substrate. This multilayer is composed of 10 double layers of 70 Å Ni and 30 Å Ti. On can clearly see the

pronounced maxima due to the periodicity of the Ni/Ti double layer of thickness 100 Å. In between, one observes many weaker oscillation (be attentive to the logarithmic scale) with a period given by the total thickness of the multilayer.

Roughness and interdiffusion

Until now we assumed perfectly flat interfaces. A real interface will, however, always show a certain roughness at the atomic level, as shown in Fig. 7. The height profile of the interface is completely described by the parametrization $z(x, y)$. Such a detailed information is not at all interesting. Much more interesting are parameters that statistically describe the interface, such as the mean squared deviation from an ideally flat interface, or the lateral correlation length. Those parameters can be determined from reflectometry and scattering under grazing incidence [7].
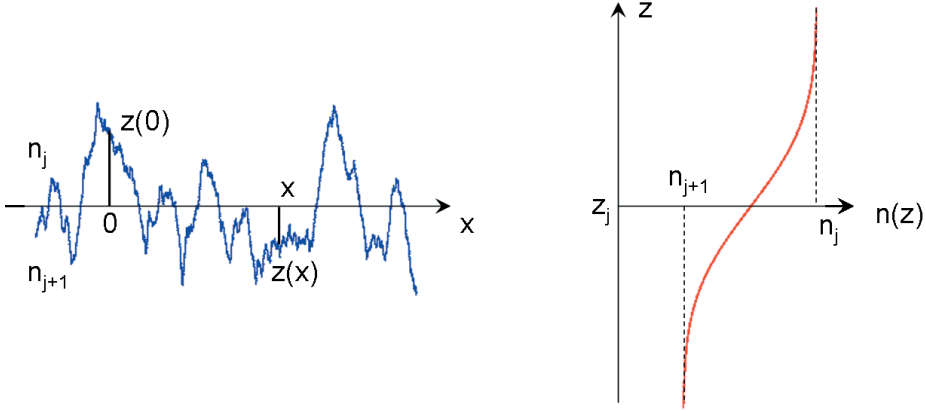
Fig. 7: Roughness of a real interface, characterized by the parametrization $z(x, y)$ and dependency of the refractive index on z.

As simplest model, we assume that the height coordinate z follows a random distribution of values around the nominal value z_j of the flat interface. The random distribution being described by a Gaussian function

$$P(\Delta z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\Delta z^2}{2\sigma^2}\right), \tag{24}$$

the profile of index of refraction between layers j and $j+1$ takes the form:

$$n(z) = \frac{n_j + n_{j+1}}{2} - \frac{n_j - n_{j+1}}{2}\text{erf}\left(\frac{z - z_j}{\sqrt{2}\sigma_j}\right) \tag{25}$$

with the "Error" function:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt. \tag{26}$$

The reflectivity from such a rough interface is obtained from the average of the reflectivities from a sequence of layers that describe the profile of refraction index. This average is performed in detail in Ref. [11]. As a result one obtains that the Fresnel coefficient for an ideally flat interface has to be modified by an exponential damping factor in the following way:

$$R_{rough} = R_{flat} \cdot \exp\left(-4\sigma_j^2 k_{zj} k_{zj+1}\right). \tag{27}$$

In this equation, σ_j is the root mean squared deviation from the nominal position of the flat interface.

The effects of interfacial roughness on the neutron reflectivity from a Si substrate and from a Ni layer on Si substrate have been simulated in Fig. 8. On the left side of Fig. 8 one can observe that the effect of roughness is to decrease the reflectivity at large wave vector transfers. The effect of roughness will be seen, if the value of the scattering wave vector gets bigger than $1/\sigma$. Therefore, if one wants to determine very small roughness amplitudes, one has to measure the reflectivity till very large reflection angles and over a large dynamical range.

The right side of Fig. 8 shows the effect of the roughness of a single layer. The simulations have been performed for ideally flat interfaces, for a rough surface of the layer, for a rough interface between layer and substrate and for the case where both interfaces are rough. One can see that the four cases can be well differentiated. When only one of the two interfaces is rough, the interference pattern due to the reflection on the top and bottom interfaces is suppressed at large wave vectors. If both interfaces are rough, a faster decrease of the averaged reflectivity takes place.
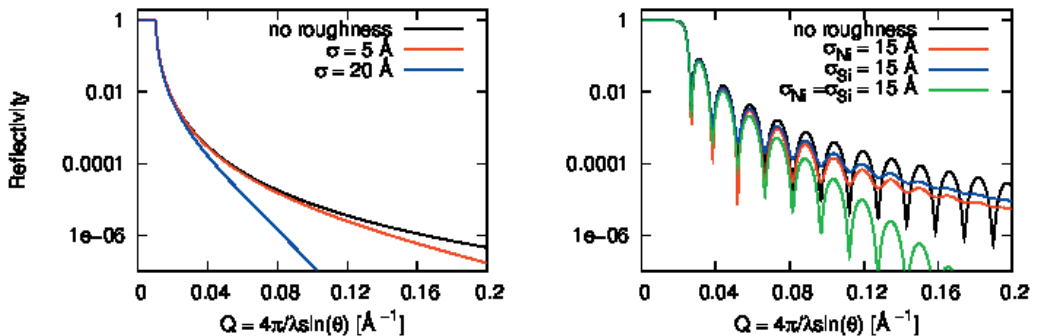

Fig. 8: Left: Neutron reflectivity at the interface between vacuum and Si. Right: Neutron reflectivity from a 400 Å thick Ni layer on Si substrate. Effect of interfacial roughness.

Finally, one should point out that a specular reflectivity measurement can only describe the profile of scattering length density normal to the interface. This means that a reflectivity measurement can not differentiate between interfacial roughness and interdiffusion, as interdiffusion

will induce the same profile of refraction index as in Fig. 7. But what happens to the intensity loss described by the exponential factor of Eq. (27)? In the case of a diffuse interface, this intensity goes into the transmitted beam because there is no potential gradient in a direction different than the one normal to the interface. On the other hand, in the case of a rough interface, the intensity loss comes from scattering by lateral fluctuations of the potential, leading to intensities that can be observed in directions other than the specular direction: this is off-specular diffuse scattering. A statistical function like the height-height pair correlation function can be determined from the measurement of off-specular scattering [7].

3   Measurement of neutron reflectivity

The principal components of a reflectivity experiment are (i) a radiation source, (ii) a wavelength selector (monochromator, choppers), (iii) a collimation system, (iv) the sample and (v) a detection system.

The aim of a neutron specular reflectivity experiment is to measure the reflectivity as a function of the scattering wave vector Q perpendicular to the sample surface:

$$Q = \frac{4\pi}{\lambda} \sin \theta \tag{28}$$

The measurement can be done by changing either the angle of incidence θ on the sample or the wavelength λ, or both.

Monochromatic instruments

At a nuclear reactor source, the measurements are usually performed at a fixed value of λ, using θ-2θ scans (2θ being the detector angle). The wavelength selection can be obtained by Bragg scattering on a monochromator crystal or by using a velocity selector. Fig. 9 describes such an instrument. This is the MARIA reflectometer of the JCNS located at the FRM-II source in Garching [12]. The neutrons are brought from the cold source to the instrument using a supermirror coated guide (see lecture 2 of this book). A certain wavelength with a spread of 10 % is chosen by adjusting the rotation speed of a velocity selector. The wavelength spread can be reduced by using a Fermi chopper and time-of-flight detection. The neutron beam is then collimated by a pair of slits in order to define the angle of incidence of the neutrons relative to the sample surface with a certain precision. The neutrons are then detected on a two dimensional position sensitive detector. Such a detector allows to record at the same time not only the specular reflectivity signal but also the signals of off-specular scattering and grazing incidence small angle scattering. The projection of the spin of the neutron on a quantization axis can be selected before interaction with the sample by using a polarizer and after interaction with the sample by using a polarization analyzer, allowing to retrieve information about the norm and angle of the layer magnetizations in a magnetic sample (see lecture 10). The polarizer uses magnetic supermirrors and the analyzer uses a nuclear polarized [3]He gas to select the spin projection.
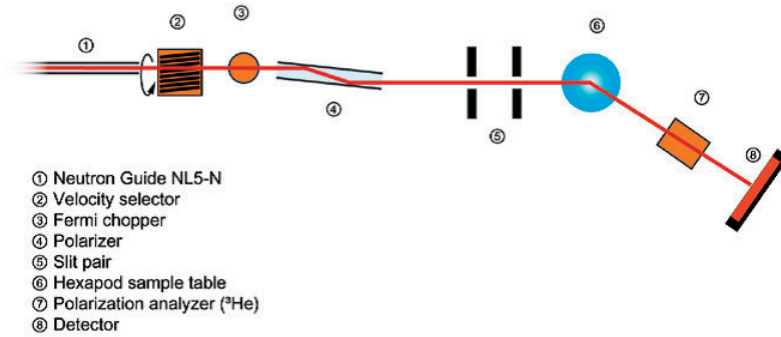
① Neutron Guide NL5-N
② Velocity selector
③ Fermi chopper
④ Polarizer
⑤ Slit pair
⑥ Hexapod sample table
⑦ Polarization analyzer (³He)
⑧ Detector

Fig. 9: A monochromatic instrument: MARIA of the JCNS at MLZ [12].

Time-of-flight instruments

At a spallation source, the measurements are performed at fixed values of θ and as a function of λ. This is the time-of-flight technique, that consists in sending a pulsed white beam on the sample. Since the speed of the neutron varies as the inverse of the wavelength, the latter is directly related to the time t taken by the neutron to travel from the pulsed source to the detector (over the distance L) by:

$$\lambda = \frac{h}{mL}t. \tag{29}$$

For a reflectivity measurement, the angle is fixed and the reflectivity curve is obtained by measuring the reflectivity signal for each wavelength of the available spectrum, each wavelength corresponding to a different scattering wave-vector magnitude. Sometimes it is necessary to use several angles of incidence because the Q range is not large enough.

An example of time-of-flight reflectometer is presented in Fig. 10. This is the magnetism reflectometer of the Spallation Neutron Source (SNS) in Oak Ridge, USA [13]. Neutrons coming from the moderator are first deflected by 2.5^o using a channel beam bender, composed of a stack of supermirrors, in order to achieve enough separation with the neighbour instrument (a liquid reflectometer) and in order to deliver to the sample a "clean" neutron beam, essentially free of fast neutrons and γ radiation. As much useful neutrons as possible are transported to the sample by using a supermirror coated tapered neutron guide that focuses the beam horizontally and vertically to a size comparable to usual sample sizes, i.e. several cm^2. The bandwidth choppers are used to select a wavelength width (λ from 2 to 5 Å), in order to avoid frame overlap. A chopper is a rotating disk with windows transparent to neutrons. When two choppers are mounted at a certain distance one with respect to the other, the delay between the window openings and the width of the windows can be chosen to achieve a transmission of only those neutrons having speeds contained in a certain range. The phenomenon of frame overlap happens when the slow neutrons of a pulse are overtaken by the fast neutrons of the next pulse. A time-of-flight detection cannot differentiate between those neutrons. Therefore, frame overlap has to be avoided. The function of the second of the three choppers is to absorb the very slow neutrons. This instrument has also collimation slits, a position sensitive detector and polarizing and analyzing

devices whose functions are the same as the ones explained in the preceding section.
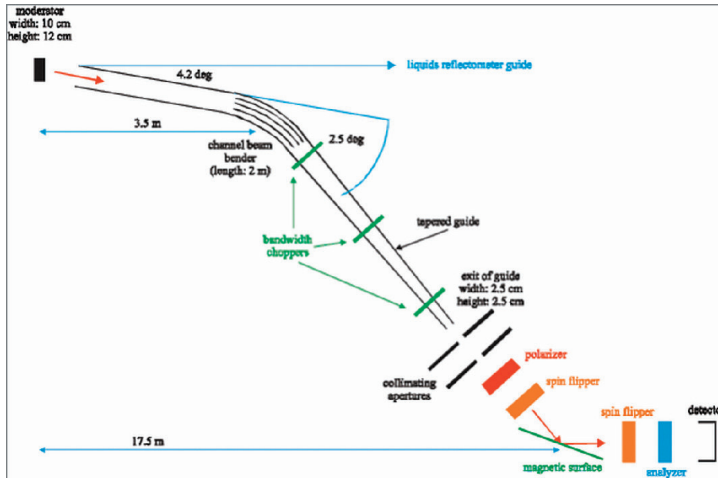

Fig. 10: A time-of-flight instrument: The magnetism reflectometer of the SNS [13].

Resolution

The reflectivity signal decreases very rapidly above the critical angle of total reflection when Q increases (see Eq. (20), $R \propto 1/Q^4$). In order to win some intensity, either the collimation slits can be opened or the wavelength spread $\delta\lambda$ can be increased, at the price of a loss in resolution in scattering wave vector. The dispersion in Q is given by (for $\theta \ll 1$ rad.):

$$\delta Q \simeq \sqrt{\left(\frac{4\pi}{\lambda}\frac{\delta\lambda}{\lambda}\theta\right)^2 + \left(\frac{4\pi}{\lambda}\delta\theta\right)^2} \tag{30}$$

where $\delta\theta$ is the beam angular divergence. The divergence of the incident beam is usually determined by the two collimation slits if the beam is smaller than the effective width of the sample seen by the neutron beam, or by the first slit and the sample itself if the sample is small enough to be totally illuminated by the neutron beam. The experimental reflectivity is then the calculated reflectivity convoluted by a resolution function whose width is given by δQ. Experience shows that a Gaussian function works well to reproduce the resolution effects. In Fig. 11 the reflectivity is calculated for a perfect instrument and by taking into account the effects of angular divergence and wavelength spread. As can be inferred from Eq. (30), angular divergence induces a loss of resolution independent of θ, and wavelength spread degrades the resolution as θ increases. This example shows that, when preparing a reflectometry experiment and depending on the sample under study, a good compromise between intensity and resolution has to be found.
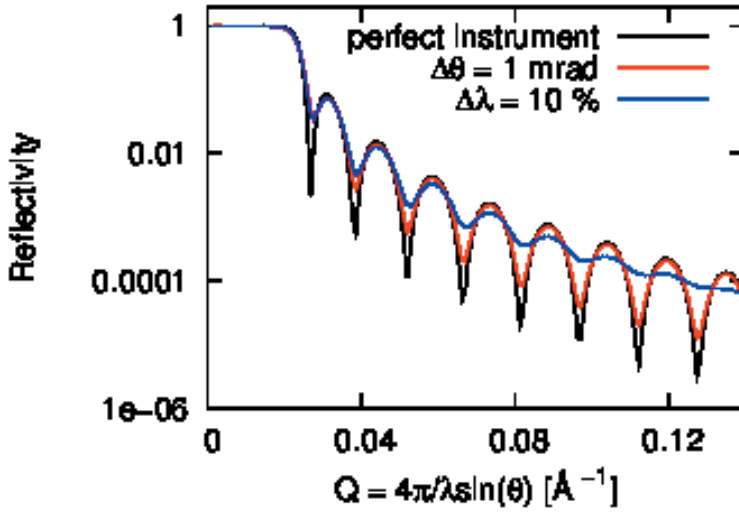
Fig. 11: Effect of $\delta\theta$ and $\delta\lambda$. Comparison between a perfect instrument, an instrumental $\delta\theta$, and a $\delta\lambda$ for a measurement on a 400 Å thick Ni layer on Si substrate

4 Crystallography at the nanoscale: GISAXS from a nanoparticle assembly

The prime aim of this section is to emphasize on the added information provided by Grazing Incidence Small Angle Scattering (GISAS) with respect to other surface characterization techniques like Atomic Force Microscopy (AFM) and Scanning Electron Microscopy (SEM). AFM and SEM give information on the in-plane fluctuations of the order parameter, while GISAS allows a full 3 dimensionnal investigation, i.e. gives depth-resolved infomation on those in-plane fluctuations.

The study reported here [14] concerns the investigation of the ordering in an assembly of magnetic nanoparticles deposited on a surface. Those nanoparticles, of truncated cubic shape, were deposited under an applied magnetic field, leading to the formation of mesocrystals, i.e. columns (see Fig. 12), several hundreds nanometres high and several micrometres diameter, composed of a single crystalline arrangement of nanoparticles (see insert of Fig. 13). The whole assembly is a 2 dimensional orientational average of such mesocrystals.

The in-plane arrangement of the nanoparticles in each mesocrystal has been determined by SEM and consists of a square lattice (see insert of Fig. 13) of lattice parameter 13.1 nm. Only five different cubic and tetragonal Bravais lattices are compatible with this 2 dimensional arrangement: simple cubic (sc), simple tetragonal (st), body centred cubic (bcc), body centred tetragonal (bct) and face centred cubic (fcc). Two (bcc and fcc) of those five lattices can be excluded from packing and geometrical conditions.

The actual 3 dimensional Bravais lattice has been determined by GISAXS, the geometry of the

experiment being depicted in Fig. 12: a beam of x-rays, well collimated in both directions perpendicular to $\mathbf{k_i}$, impinges on the surface under an angle of incidence α_i and the scattered x-rays are collected on a position sensitive detector. Each detector pixel is defined by the values of the angles θ and α_f. Taking into acount the smallness of those three angles, the components of the scattering wave vector along the three axes depicted in Fig. 12 are given by:

$$Q_x = k \left(\alpha_i^2 - \alpha_f^2 - \theta^2 \right)/2, \quad Q_y = k\theta, \quad Q_z = k \left(\alpha_i + \alpha_f \right), \quad \text{with} \quad k = \frac{2\pi}{\lambda}. \quad (31)$$
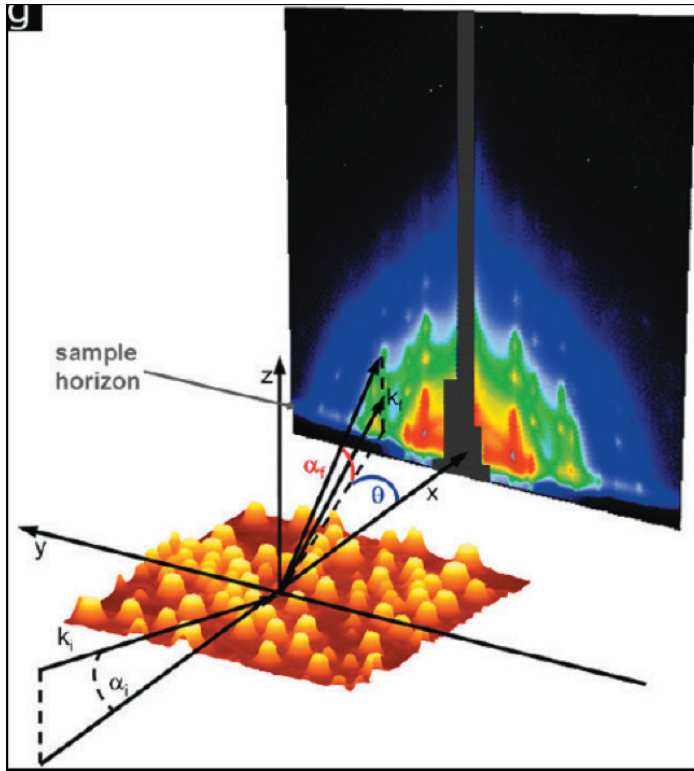
Fig. 12: Atomic Force Microscopy (AFM) image of the assembly of magnetic nanoparticles and geometry of the GISAXS experiment. GISAXS signal is collected on a 2 dimensional position sensitive detector. Taken from [14].

The thus obtained GISAXS pattern at a certain angle of incidence α_i of the incoming beam close to the critical angle of total reflection is given in Fig. 13. This pattern shows a whole bunch of local intensity maxima at positions in Q_y and Q_z that are characteristic of the crystalline stacking respectively in-plane and out-of-plane. A relation connecting all the Q_z coordinates of the local maxima to their Q_y coordinates is obtained by a combination of Snell's law and Bragg's law leading to an extinction rule and an out-of-plane lattice parameter (17.8 nm) characteristic of a bct packing of the nanoparticle lattice.
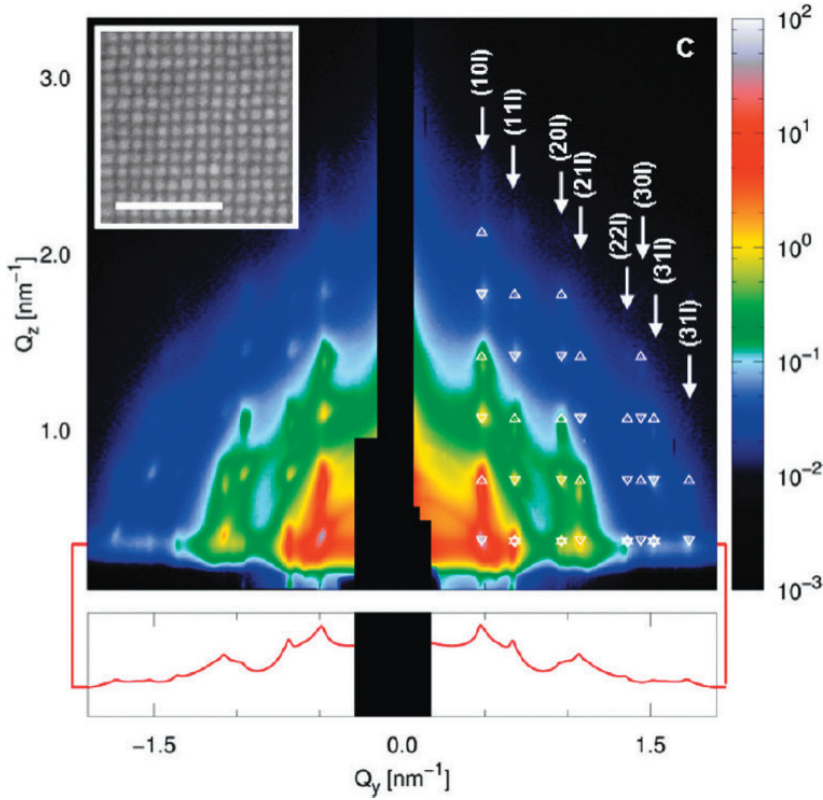
Fig. 13: GISAXS pattern of the assembly of magnetic nanoparticles. Insert: Scanning Electron Microscopy (SEM) image of the top of a mesocrystal of nanoparticles; scale bar represents 100 nm. Taken from [14].

5   Conclusion

This chapter has given an overview of reflectometry and GISAS as a tool for the investigation of thin films, their interfaces or mesoscopic objects deposited on a surface. We have presented a formalism which makes it possible to describe the specular reflectivity on non-magnetic systems. The formalism of neutron reflectometry for the investigation of the magnetic moment orientations in magnetic multilayers is presented in the next chapter of this book, together with several application examples.

6   Take-Home Message

Reflectometry of X-rays and neutrons from thin films with thicknesses in the mesoscopic range gives access to the scattering length density profile along their depth. The information is an average over the in-plane coordinates. With GISAS, in-plane correlations of the scattering

length density can be investigated. Moreover, depth resolution is accessed by interpreting the GISAS signal as a function of α_i or α_f.

References

[1] A special issue of *Langmuir* covers a broad range of applications of this technique for the characterization of surfaces and interfaces in the fields of soft matter and biology, Lamgmuir **25(7)**, (2009).

[2] P. Grünberg, J. Phys. Condens. Matter **13**, 7691 (2001).

[3] U.Rücker and E. Kentzinger, in, *Probing the Nanoworld, 38th IFF Springschool* (Schriften des Forschungszentrum Jülich, Materie und Material, Band 34, 2007). Lecture D4

[4] Th. Brückel and E. Kentzinger in, *Magnetische Schichtsysteme, 30. Ferienkurs des IFF* (Schriften des Forschungszentrum Jülich, Materie und Material, Band 2, 1999). Kapitel B3

[5] Th. Brückel in, *Laboratory course, Neutron Scattering* (Schriften des Forschungszentrum Jülich, Reihe Schlüsseltechnologien, Band 39, 2012). Lecture 14

[6] G. Fragneto, Collection SFN **11**, 89 (2010).

[7] S.K. Sinha, E.B. Sirota, S. Garoff, H.B. Stanley, Phys. Rev. B **38:4**, 2297 (1988).

[8] J. Burle, C. Durniak, J. M. Fisher, M. Ganeva, G. Pospelov, W. Van Herck, J. Wuttke (2017), BornAgain - Software for simulating and fitting X-ray and neutron small-angle scattering at grazing incidence, http://www.bornagainproject.org

[9] M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Oxford, 1989).

[10] L.G. Parratt, Phys. Rev. **95**, 359 (1954).

[11] L. Névot, P. Croce, Rev. de Phys. Appl. **15**, 761 (1980).

[12] Heinz Maier-Leibnitz Zentrum. MARIA: Magnetic reflectometer with high incident angle, Journal of large-scale research facilities **1**, A8 (2015). http://dx.doi.org/10.17815/jlsrf-1-29

[13] Courtesy of Valeria Lauter (Oak Ridge National Laboratory)

[14] S. Disch, E. Wetterskog, R.P. Hermann, G. Salazar-Alvarez, P. Busch, Th. Brückel, L. Bergström, S. Kamali, Nano Letters **11**, 1651 (2011). dx.doi.org/10.1021/nl200126v

D 3   Powder and Single Crystal Diffractometry: Chemical and Magnetic Structures

M. Meven

Institut für Kristallographie

RWTH Aachen

Contents

1   Introduction

Many mechanical, thermal, optical, electrical and magnetic properties of solid matter depend significantly on its atomic structure. Therefore, a good understanding of the physical properties needs not only the knowledge about the particles inside (atoms, ions, molecules) but also about their spatial arrangement. For most cases diffraction is *the* tool to answer questions about the atomic and/or magnetic structure of a system. Beyond this, neutron diffraction allows to answer questions where other techniques fail.

1.1   Crystallographic Basics

In the ideal case a complete solid matter consists of small identical units (same content, same size, same orientation like sugar pieces in a box). These units are called unit cells. A solid matter made of these cells is called a single crystal. The shape of a unit cell is equivalent to a parallelepiped that is defined by its base vectors $\mathbf{a}_1$, $\mathbf{a}_2$ und $\mathbf{a}_3$ and that can be described by its lattice constants a, b, c; α, β and γ (fig. 1). Typical lengths of the edges of such cells are between a few and a few ten Ångström (1 Å = 10^{-10} m). The combination of various restrictions of the lattice constants between $a \neq b \neq c$; $\alpha \neq \beta \neq \gamma \neq 90°$ (triclinic) and $a = b = c$; $\alpha = \beta = \gamma = 90°$ (cubic) yields seven crystal systems. The request to choose the system with the highest symmetry to describe the crystal structure yields fourteen Bravais lattices, seven primitive and seven centered lattices.
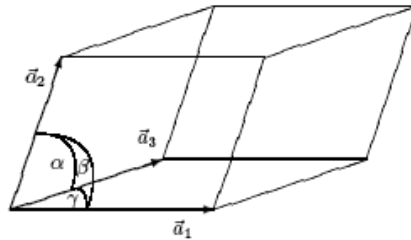

Fig. 1:   Unit cell with $|\mathbf{a}_1|=a$, $|\mathbf{a}_2|=b$, $|\mathbf{a}_3|=c$, α, β, γ.

Each unit cell contains one or more particles i. The referring atomic positions $\mathbf{x}_i = x_i\mathbf{a}_1 + y_i\mathbf{a}_2 + z_i\mathbf{a}_3$ are described in relative coordinates $0 \leq x_i$; y_i; $z_i < 1$. The application of different symmetry operations (mirrors, rotations, glide mirrors, screw axes) on the atoms in one cell yield the 230 different space groups (see [1]).

The description of a crystal using identical unit cells allows the representation as a three-dimensional lattice network. Each lattice point can be described as lattice vector $\mathbf{t} = u\mathbf{a}_1 + v\mathbf{a}_2 + w\mathbf{a}_3$; $u, v, w \in \mathbf{Z}$. From this picture we get the central word for diffraction in crystals; the *lattice plane* or *diffraction plane*. The orientations of these planes in the crystal are described by the so called *Miller indices* h, k and l with $h, k, l \in \mathbf{Z}$ (see fig. 2). The reciprocal base

vectors $\mathbf{a}_1{}^*$, $\mathbf{a}_2{}^*$, $\mathbf{a}_3{}^*$ create the reciprocal space with: $\mathbf{a}_i{}^* \, \mathbf{a}_j = 2\pi \, \delta_{ij}$ with $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$. Please note: The commonly used notation in crystallography is $\mathbf{a}_i{}^* \, \mathbf{a}_j = \delta_{ij}$, which will not be considered in the following text. Each point $\mathbf{Q} = h{*}\mathbf{a}_1{}^* + k{*}\mathbf{a}_2{}^* + l{*}\mathbf{a}_3{}^*$ represents the normal vector of a (hkl) Plane. Each plane cuts the crystal lattice along its base vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ at $(1/h)\mathbf{a}_1$, $(1/k)\mathbf{a}_2$ and $(1/l)\mathbf{a}_3$. A Miller index of zero means that the referring axis will be cut in infinity. Thus, the lattice plane is parallel to this axis.
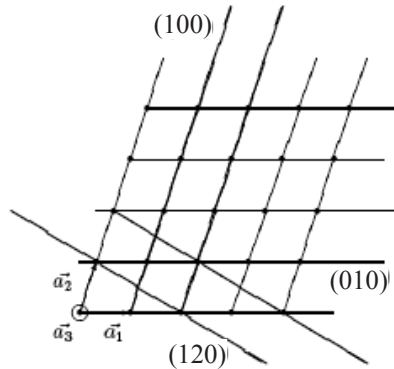
Fig. 2:   Different lattice planes in a crystal lattice ($\mathbf{a}_3$ = viewing direction).

The atoms in a unit cell are not rigidly fixed at their positions. They oscillate around their positions (e.g. thermal excitation). A simple description for this is the model of coupled springs. In this model atoms are connected via springs whose forces describe the binding forces between the atoms (e.g. van der Waals, Coulomb, valence). The back driving forces of the springs are proportional to the deviation x_i of the atoms from their mean positions and to the force constant D, thus the force is $F = -D \, \Delta x$ (harmonic approximation).

Therefore, the atoms oscillate with $x_i = A_i \sin(vt)$ around their mean positions with the frequency v and the amplitude A_i. Both, v and A_i are influenced by the force constant D_j of the springs and the atomic masses m_i of the neighbouring atoms. The resulting lattice oscillations are called phonons in reference to the photons (light particles) in optics, which as well transport energy in dependence of their frequency. A more complex and detailed description of phonons in dependence on the lattice structure and the atomic interaction effects is given in lattice dynamics. In the harmonic approximation the displacements of an atom can be described with an oscillation ellipsoid. This ellipsoid describes the preferred volume in which the atom is placed. Its so called mean square displacements (MSD) $U_{jk}{}^i$ represent the different sizes of the ellipsoid along the different main directions j, k in the crystal. The simplest case is a sphere with isotropic MSD B_i. In the next paragraph MSD are discussed from the point of view of diffraction analysis.

A full description of a single crystal contains information about lattice class, lattice constants and unit cell, space group and all atomic positions and their MSD. If an atomic site i is not fully occupied by the same element (occ_i = 100%) or by more than one element (crystal with vacancies or mixed crystals), this site occupancy factor can and has to be determined in addition. Especially the isotope specific interaction of neutrons can be a useful tool to resolve structural details, e.g. via. deuteration.

1.2 Structure Determination with Diffraction

Diffraction means coherent elastic scattering of a wave on a crystal. Because of the quantum mechanical wave/particle dualism X-rays as well as neutron beams offer the requested wave properties:

X-rays: $E = h\nu$; $\lambda = c/\nu$
Neutrons: $E_{kin} = \frac{1}{2} m_n v^2 = h\nu = p^2/2m_n$; $\lambda = h/p$; $p \sim \sqrt{(m_n k_B T)}$

h: Planck's constant; ν: oscillation frequency; λ: wavelength; c: light speed; v: particle velocity; p: impact; m_n: neutron mass; k_B: Boltzmann constant; T: temperature

Only the scattering cross section partners are different (X-rays: scattering on the electron shell of the atoms; neutrons: core (and magnetic) scattering) as explained in detail below. In scattering experiments the information about structural properties is hidden in the scattering intensities I.

In the following pages we will discuss only elastic scattering ($\lambda_{in} = \lambda_{out}$). The scattering cross section of the radiation with the crystal lattice can be described as following:

Parallel waves of the incoming radiation with constant λ are reflected by lattice planes which are ordered parallel with a constant distance of d. This is very similar to a light beam reflected by a mirror. The angle of the diffracted beam is equal to the angle of the incoming beam, thus the total angle between incoming and outgoing beam is 2θ (see fig. 3).
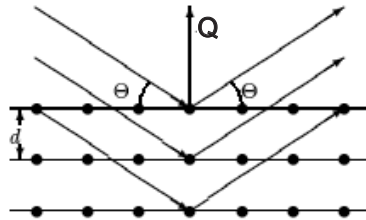
Fig. 3: Scattering on lattice planes.

The overlap of all beams diffracted by a single lattice plane results in constructive interference only if the combination of the angle θ, lattice plane distance d and wavelength λ meets Bragg's law:

$$2d \sin \theta = \lambda$$

The largest distance $d_{hkl} = 2\pi/Q$ of neighboured parallel lattice planes in a crystal is never larger than the largest lattice constant $d_{hkl} \leq \max(a; b; c)$. Therefore, it can only be a few Å or less. For a cubic unit cell ($a = b = c$; $\alpha = \beta = \gamma = 90°$) this yields:
$d_{hkl} = a/\sqrt{(h^2+k^2+l^2)}$

With increasing scattering angles also the indices (*hkl*) increase while the lattice plane distances shrink with a lower limit of $d_{min} = \lambda/2$. Therefore, scattering experiments need wavelengths λ in the same order of magnitude of the lattice constants or below. This correpsonds to X-ray energies of about 10 keV or neutron energies about 25 meV (thermal neutrons).

The following scheme (fig. 4) represents a diffraction experiment using the *angular dispersive* setup. The beam from a "white" source is diffracted by a monochromator with a fixed *d* spacing and a fixed Bragg angle $2\theta_M$ to select a constant wavelength λ.
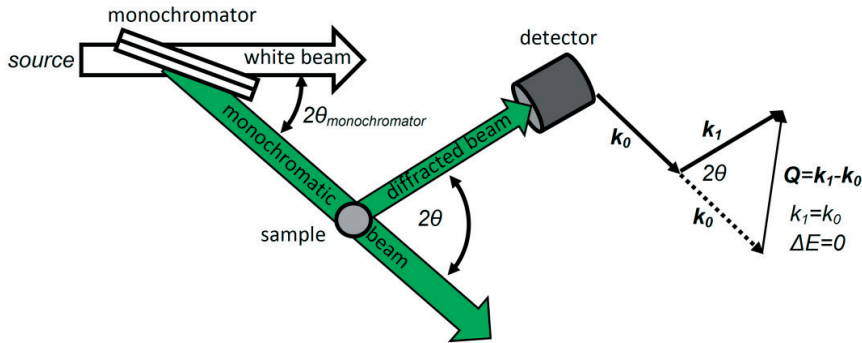
Fig. 4:   Schematic representation of a constant wavelength diffractometer [21].

Ewald Construction: In reciprocal space each Bragg reflection (*hkl*) is represented by a point $\mathbf{Q} = h\mathbf{a_1}^* + k\mathbf{a_2}^* + l\mathbf{a_3}^*$. A scattered beam with the wave vector $\mathbf{k}$ fulfills Braggs law if the relationship $\mathbf{k} = \mathbf{k_0} + \mathbf{Q}$, $|\mathbf{k}|=|\mathbf{k_0}|=2\pi/\lambda$ is true, as shown in fig. 5. During an experiment the available reciprocal space can be described by an Ewald sphere with a diameter of $4\pi/\lambda$ and the (000)-point as cross point of $\mathbf{k_0}$ direction and the centre of the diameter of the sphere. The rotation of the crystal lattice during the diffraction experiment is equal to a synchronous movement of the reciprocal lattice around the (000) point. If Bragg's law is fulfilled, one point (*hkl*) of the reciprocal lattices lies exactly on the Ewald sphere. The angle between the $\mathbf{k}$ vector and the $\mathbf{k_0}$ vektor is 2θ. The limited radius of $2\pi/\lambda$ of the Ewald sphere limits also the visibility of (*hkl*) reflections to $|\mathbf{Q}| < 4\pi/\lambda$.
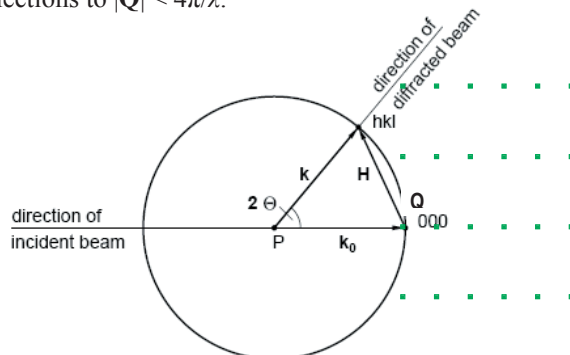

Fig. 5:   Ewald construction in reciprocal space, green dots = reciprocal lattice vectors [18].

Determination of the unit cell: Following Braggs law the scattering angle 2θ varies (for λ=const.) according to the lattice distance d_{hkl}. Thus for a given λ and known scattering angles 2θ one can calculate the different d values of the different layers in the lattice of a crystal. With this knowledge is possible to determine the lattice system and the lattice constants of the unit cell (although not always unambigously!).

Atomic Positions in the unit cell: The outer shape of a unit cell does not tell anything about the atomic positions $\mathbf{x}_i = (x_i\, y_i\, z_i)$ of the atoms in this cell. For their determination one has to measure also the quantities of the different reflection intensities of a crystal. This works because of the relationship between the intensities of Bragg reflections and the specific cross section of the selected radiation with each element in a unit cell. Generally one can use the following formula for the intensity of a Bragg reflection (hkl) with $\mathbf{Q}$ (kinetic scattering theory):

$$I_{hkl} \sim |F_{hkl}|^2 \text{ with } F_{hkl} = \sum_{i=1}^{n} s_i(\mathbf{Q})\, \exp(2\pi\iota(hx_i+ky_i+lz_i))$$

The **structure factor** F is a complex function describing the overlap of the scattering waves of each atom i (n per unit cell). $s_i(\mathbf{Q})$ describes the scattering strength of the i-th atom on its position $\mathbf{x}_i$ in dependence of the scattering vector $\mathbf{Q}$, which depends on the character of cross section as described below.

In this context one remark concerning statistics: For measurements of radiation the pure statistical error σ is the square root of the number of measured events, e.g. X-ray or neutron particles. Thus, 100 events yield an error of 10% while 10,000 events yield an error of only 1%!

Mean Square Displacements (MSD): Thermal movements of atoms around their average positions reduce Bragg intensities in a diffraction experiment. In the simplest (=isotropic) description the parameter B_i is used to define a shell of electron or nucleus density around the average atomic position where the atom oscillates harmonically. The reduced probability to find an electron/nucleus at the average atomic position attenuates also the scattering probability. For higher temperatures (above a few Kelvin) the MSD B_i of the atoms increase linearly to the temperature T, this means $B \sim T$. Near a temperature of 0 K the MSD become constant with values larger than zero (zero point oscillation of the quantum mechanical harmonic oscillator). In the structure factor the true scattering power s_i of the i-th atom has to be corrected by an angle-dependent factor (the so called Debye-Waller factor):

$$s_i(\mathbf{Q}) \rightarrow s_i(\mathbf{Q}) * \exp(-B_i(\sin(\theta_{\mathbf{Q}})/\lambda)^2)$$

This Debye-Waller factor decreases with increasing temperature and yields an attenuation of the Bragg reflection intensities. At the same time this factor becomes significantly smaller with larger $\sin(\theta)/\lambda \sim |\mathbf{Q}|$. Therefore, especially reflections with large indices become weaker. An improved description of probability density with anisotropic MSD U_{ij} contains the following exponential function:

$$s_i(\mathbf{Q}) \rightarrow s_i(\mathbf{Q}) \exp(-\tfrac{1}{2}\,(U_{11}{}^i\,h^2 a^{*2} \quad + \quad U_{22}{}^i\,k^2 b^{*2} \quad + \quad U_{33}{}^i\,l^2 c^{*2} + $$
$$+ \; 2U_{13}{}^i\,hl\,a^*c^* + 2U_{12}{}^i\,hk\,a^*b^* + 2U_{23}{}^i\,kl\,b^*c^*))$$

Here the U_{ij} describe the dimensions of an ellipsoid instead of a shell. The transformation between B and U_{eq} (isotropic MSD calculated from the anisotropic U_{ij} with identical volume the corresponding shell and ellipsoid) is: $B = 8\pi^2 U_{eq}$

For some structures the experimentally determined MSD are significantly larger than from the harmonic calculations of the thermal movement only. Static local deformations, point defects, mixed compounds, anharmonic oscillations or double well potentials (regions with two energetically equal atomic positions very near to each other where an atom has a 50%/50% chance to occupy one position or the other) can cause this additional contribution to the pure thermal Debye-Waller factor. In the following text only the term MSD will be used to avoid misunderstandings.

1.3 Comparison of X-ray and Neutron Radiation

The different nature of interaction with matter of X-rays and neutrons explains why more comprehensive studies use both techniques. The following table (fig. 6) and pictures show the main similarities and differences between the two radiation types.

Properties	X-ray/γ	Neutrons
Mass [kg]	0	$1.673*10^{-27}$
Energy [eV]	10^3 - 10^6	10^{-3} – 100 0.025 (thermal)
magn. Moment	no	yes
Wave length λ [Å]	0.3 - 3 1.5 (Cu-$K\alpha$)	0.3 – 20 1.8 (thermal)
typ. speed [m/s]	$3*10^8$	2500 (thermal)
Interaction with	e-shell Z specific	cores/isotopes spin

Fig. 6: Properties of X-rays vs. neutrons.

X-ray Radiation interacts as electromagnetic radiation only with the electron density in a crystal. This contains all electrons whether they contribute to a chemical bond or not. The electronic scattering power – the so called atomic form factor $f(\sin\theta/\lambda)$ or shorter f_Z - of an atom depends on the number Z of its shell electrons ($f(\sin(\theta=0)/\lambda) = Z$). To be exact, $f(\sin\theta/\lambda)$ is the Fourier transform of the radial electron density distribution $n_e(r)$: $f(\sin\theta/\lambda) = \int_0^\infty 4\pi^2 n_e(r)\,\sin(\mu r)/\mu r\,dr$ with $\mu = 4\pi\sin\theta/\lambda$. Heavy atoms with many electrons contribute much stronger to reflection intensities ($I \sim Z^2$) than light atoms with less electrons. The reason for the $\sin\theta/\lambda$-dependence of f_Z is the diameter of the electron shell. It has the same order of magnitude as the wavelength λ and cannot be described as point like scattering centre. Thus, for large

scattering angles the atomic form factors vanish as well as the reflection intensities relying on them. The atomic form factors are derived from theoretical spherical electron density functions (e. g. Hartree-Fock). The resulting $f(\sin\theta/\lambda)$ curves of all elements (separated for free atoms and ions) are listed in the international tables. Their analytical approximation is described by seven coefficients (c, a_i, b_i; $1 \leq i \leq 3$), see [1].

Neutron Radiation interacts with the cores and the magnetic moments of atoms. The neutron analogon to the X-ray form factor - the scattering length b_c - is therefore not only dependent on the element but the isotope. At the same time b values of elements neighboured in the periodic table can differ significantly. Nevertheless, the scattering lengths do not differ around several orders of magnitude like in the case of the atomic form factors f_Z. Therefore, in a compound with light and heavy atoms the heavy atoms do not dominate necessarily the Bragg intensities. Furthermore, the core potential with a diameter of about 10^{-15}Å is a point like scattering centre and thus the scattering lengths b_c are independent of the Bragg angles and $\sin\theta/\lambda$ respectively. This results in large intensities even at large scattering angles. The magnetic scattering lengths b_m can generate magnetic Bragg intensities comparable in their order of magnitude to the intensities of core scattering. On the other hand side the magnetic scattering lengths depend strongly on the $\sin\theta/\lambda$ value due to the large spacial distribution of magnetic fields in a crystal. Similarly to the X-ray form factors the $b_m(\sin\theta/\lambda)$ curves are listed in the International Tables [1]. Therefore, one can easily measure magnetic structures with neutrons and separate them from the atomic structure.
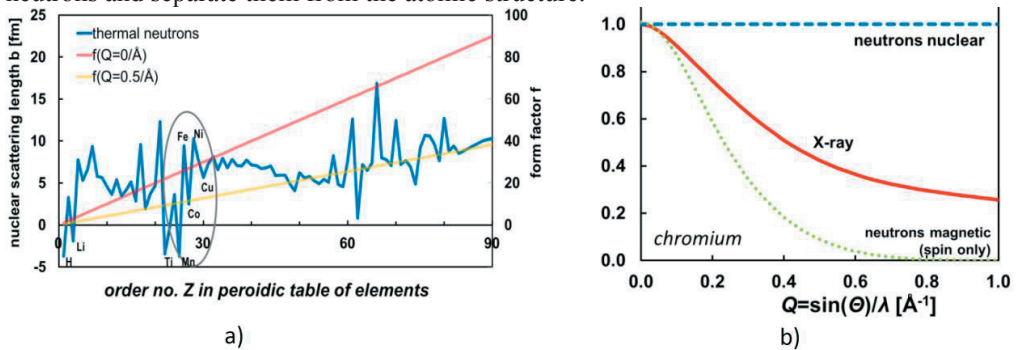
Fig. 7:   a) Scattering lengths and X-ray form factors as a function of the order number Z in the table of elements. Note the **Q** dependence of f and the strong variation of b within the 3d metal group. b) normalized form factors of chromium (absolute value of the scattering amplitude) as a function of the scattering angle for X-ray and nuclear and magnetic neutron scattering [21].

Comparison: In summary in the same diffraction experiment the different character of X-ray and neutron radiation yield different pieces of information that can be combined. X-rays yield electron densities in a crystal while neutron scattering reveals the exact atomic positions. This fact is important because for polarised atoms the core position and the centre of gravity of electron densities are not identical any more. In compounds with light and heavy atoms structural changes driven by light elements need additional diffraction experiments with neutrons to reveal their influence and accurate atomic positions respectively. There is also

another aspect; one has to take into account that for X-rays the intensities depend in two ways on $\sin\theta/\lambda$: once by the atomic form factor f_Z, and twice by the temperature dependent Debye-Waller factor (see above). The first dependence vanishes if using neutron diffraction with b_c=const. and decouples the structure factors from the influence of the MSD. In general this yields much more accurate MSD U_{ij} especially for the light atoms and might be helpful to reveal double well potentials in (partially) disordered compounds.
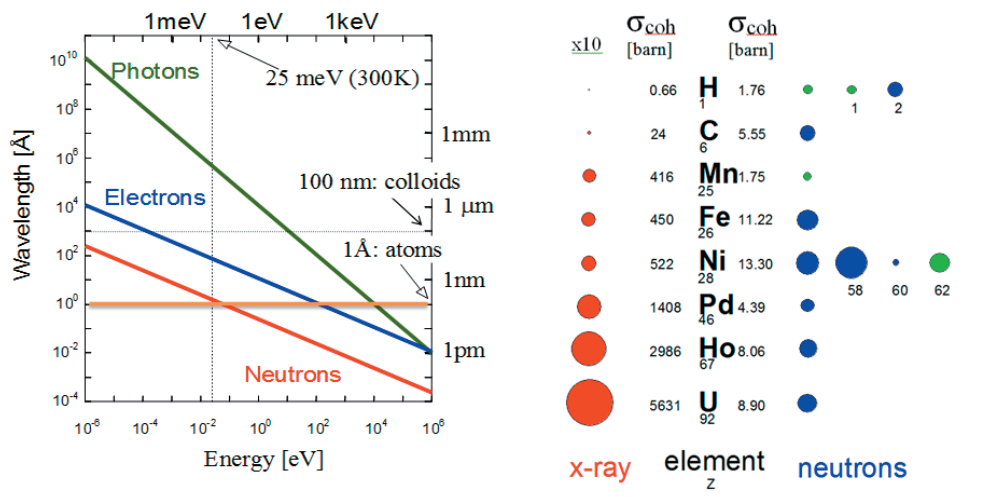
Fig. 8: left: energy-wavelength diagram for photons (X-rays), neutrons and electrons, right: scattering strengths (coherent cross-section, 1 barn = 10^{-28} m²) of selected atoms for X-rays (red) and neutrons for wavelengths at 1 Å (blue > 0, green < 0) [20].

1.4 Special Effects and Aspects

From the relation $I \sim |F|^2$ one can derive that the scattering intensities of a homogenous illuminated sample increases with its volume. But there are other effects than MSD that can attenuate intensities. These effects can be absorption, extinction, polarization and the Lorentz factor:

Absorption can be described by the Lambert-Beer law:

$I = I_0 \exp(-\mu x)$, μ/cm^{-1} = linear absorption coefficient, x/cm = mean path through sample

The linear absorption coefficient is an isotropic property of matter and depends on the wavelength and kind of radiation. For X-rays penetration depths are only a few millimetre or below (e.g. for silicon with $\mu_{MoK\alpha}$=1.546 mm^{-1}, $\mu_{CuK\beta}$=14.84 mm^{-1} with penetration depths of 3 mm and 0.3 mm respectively). This limits transmission experiments to sample diameter of typically below 0.3 mm. To correct a bias of intensities from different scattering path lengths through the sample one has to measure accurately the sample size along all directions. Even

for spherical samples the mean path lengths depend on 2θ. In addition the sample environment must have extraordinary small absorption.

Thermal neutrons show for most elements a penetration depth of several centimeters. Thus, sample diameters of some millimeters together with large and complex sample environments (furnaces, magnets, etc.) can be used easily. On the other hand side the growth of sufficiently large samples for neutron single crystal diffraction can be a significant challenge.

Extinction also reduces scattered intensities although its character is completely different form that of absorption as it emerges from multiple scattering. The principle of the *extinction effect* (not to be mixed up with crystallographic *extinction rules*!) can be explained easily taking into account that each diffracted beam can be seen as a new primary beam for the neighbouring lattice planes, see fig. 9, left. Therefore, the diffracted beam becomes partially backscattered towards the direction of the very first primary beam (in fact this means to switch from the simplified kinematic to the dynamic scattering theory!). Condition for this effect is a merely perfect crystal. Especially for very strong reflections this effect can reduce intensities dramatically (by more than 50%).

Theoretical models which include a quantitative description of the extinction effect were developed from Zachariasen (1962) and Becker and Coppens [2, 3, 4, 5, 6]. These models assume either that the crystal consists of different mosaic blocks of a *typical size* or that there is a *typical mosaicity* between the different perfect grains or both to describe the strength of the extinction effect. It is also possible to take into account anisotropic extinction effect if the crystal quality is also anisotropic. Nowaday, in most refinement programs [7] extinction correction is included. In general extinction is a problem of sample quality and size and therefore more commonly a problem for neutron diffraction and not so often for X-ray diffraction with much smaller samples and larger absorption. Sometimes shorter wavelengths where extinction effects become weaker can be used as solution.

Umweg excitation or **Renninger effect** occurs if the diffracted beam of the first lattice plane $(h_1k_1l_1)$ works as the primary beam for a second non identical set of lattice planes $(h_2k_2l_2)$ that by accident also fulfils Bragg's law (fig. 9 right). The result is a diffracted beam virtually generated by a third lattice plane $(h_3k_3l_3)$. If the structure factors of the two first planes, F_{h1k1l1}, F_{h2k2l2} are strong, the measured intensity I_{h3k3l3} might be larger than the true one generated from F_{h3k3l3}. This second type of multiple scattering is especially problematic in cases, where $(h_3k_3l_3)$ is forbidden because of the extinction rules of the (true) space group.
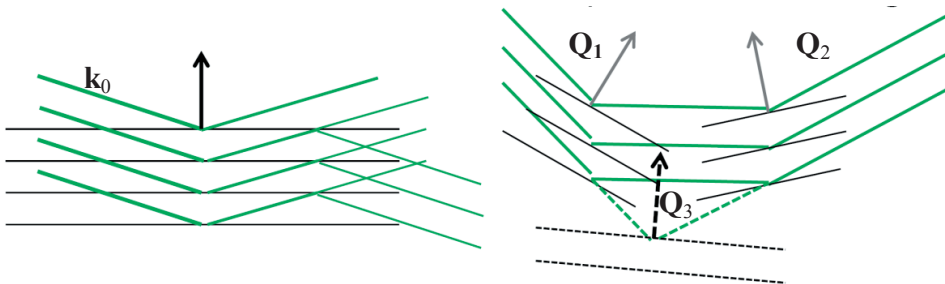

Fig. 9:   Multiple scattering; left: extinction effect; right: Renninger effect [21].

Polarization: X-ray radiation means electromagnetic radiation. Therefore, the primary beam of an X-ray tube is not polarized. The radiation hits the sample under an diffraction angle of θ where it can be separated into two waves of same intensity, firstly with an electrical field vector parallel $E_{||}$ and secondly perpendicular $E_\perp$ towards the θ-axis. Whilst the radiation with $E_{||}$ will not be attenuated the radiation with $E_\perp$ undergoes an attenuation with $E_\perp \rightarrow \cos(2\theta)$ $E_\perp$. The polarization factor P for the attenuation has then the following formula ($I \sim E^2$):

$$P = (1+\cos(2\theta)^2)/2$$

Additional optical components like monochromator crystals also have an impact on the polarization and have to be taken into account accordingly.

Lorentz factor: The Lorentz factor L is a purely geometrical factor. It results from the effectively elongated stay of the sample near the reflection position during an ω- and $\omega/2\theta$-scan respectively towards higher 2θ values for the same angular speed $\Delta\omega/\Delta t$ (see Ewald construction, fig. 5):

$$L = 1/\sin(2\theta)$$

This has to be taken into account for any kind of radiation in a diffraction experiment.

Technical limits: The different interactions of X-ray and neutron radiation with the atoms in a crystal make neutrons apparently the better choice for diffraction experiments which focus on atomic positions, mean square displacements and magnetism. X-rays are preferable for studies on electron densities and chemical bonds. But one has to take into account the available flux of X-rays and neutrons respectively. Diffraction with a monochromatic beam needs a sharp band of energies/wavelengths in the order of $\Delta\lambda/\lambda < 10^{-2} - 10^{-3}$ or even smaller. For such a small bandwidth the flux of neutrons is several orders of magnitude smaller than the flux of X-rays of a corresponding synchrotron source or X-ray tube in the laboratory. The reason for this is that in an X-ray tube most X-rays are generated in a small energy band, the characteristic lines of the tube target (K_α, K_β, etc.). Additional metal foils used as filter allow to cut off unwanted characteristic lines which yields quasi monochromatic radiation of a single wavelength. Neutrons generated by fission in a research reactor distribute to a broad spectrum of wavelengths. To reduce the bandwith one has to use a monochromator crystal. This reduces significantly the number of available neutrons for the diffraction experiment. Thus, the weak flux of neutrons and the weak cross section of neutrons with matter have to be compensated with large sample sizes of several millimeters. For the same reason the monochromatization of the neutrons is normally chosen to be not too sharp (resolution about $\Delta\lambda/\lambda \approx 10^{-2}$ for neutrons, $\Delta\lambda/\lambda \approx 10^{-4} - 10^{-5}$ for synchrotron).

1.5   From Measurement to Model – Structure Refinement

Goal of powder diffraction and single crystal diffraction is to compare experimental diffraction data with several structure models and to find the one that fits best. The differences between both methods are explained in the following sections. For instance, powder diffraction experiments yield $I/2\theta$ diagrams while single crystal diffraction experiments yield typically a list of corresponding intensities and error bars of a set of Bragg reflections and therefore slightly different structure refinement procedures are used to find a structure solution. Nevertheless the general approach described for single crystal data analysis in this section is very similar to the one for powder diffraction which will not be discussed here.

To get a structural model from the experimentally collected integral Bragg intensities one needs several steps in advance. Firstly one has to make sure that all reflections are measured properly (no shading, no $\lambda/2$-contamination, no multiple scattering). Damaged reflections have to be excluded from further treatment.
During data refinement not only the quantities of the relative intensities but also their errors are taken into account. The total statistical error σ of an integral intensity I_{obs} of a single reflection is calculated as following:

$$\sigma^2 = I_{obs} + I_{background} + (k\, I_{total})^2$$

The part $\sigma_0^2 = I_{total}$, $I_{total} = I_{obs} + I_{background}$ refers to the error caused by counting statistics. It contains as well the effective intensity I_{obs} as the contribution of the background. But other effects also influence the reproducibility of a measurement (and thus the total error), e.g. specific the instrumental errors. Those errors are collected in the so called *McCandlish-Factor* k and increase to the total error. Obviously, the total error cannot drop below the instrumental limit of the experiment and thus the impact of strong reflections does not become exaggerated in the refinement. The determination of k is done by measuring the same set of reflections (so called standard reflections) several times during a data collection. The mean variation of the averaged value represents k. In addition, the repeated measurement of standard reflections offers the opportunity to notice unwanted changes during experiment like structural changes or release from the sample holder.
To make sure the comparability of all reflections with each other, all intensities and errors are normalized to the same time of measurement (or monitor count rate) and undergo the Lorentz and (in the X-ray case) polarization correction.
Finally in advance of the data refinement there can be done a numerical (e.g. with DataP, [8]) or an empirical absorption correction if necessary. The quality of a measurement is checked in advance of the data refinement by comparing symmetry equivalent reflections and systematic extinctions to confirm the Laue group and space group symmetry. The result is written as internal R-value:

$$R_{int} = \left(\sum_{k=1}^{m}\left(\sum_{j=1}^{n_k} (<I_k> - I_j)^2\right)\right) / \left(\sum_{k=1}\sum_{j=1}^{n_k}(I_j^2)_k\right)$$

R_{int} represents the mean error of a single reflection j of a group k of n_k symmetry equivalent reflections, corresponding to its group and the total number m of all symmetrically independent groups. Therefore R_{int} is also a good mark to check the absorption correction. After these preliminary steps one can start the final data refinement.

At the beginning one has to develop a structural model. The problem with that is that we measure only the absolute values $|F_{hkl}|$ and not the complete structure factor $F_{hkl} = |F_{hkl}|\exp(\iota\phi)$ including its phase ϕ. Therefore, generally the direct Fourier transform of the reflection information F_{hkl} from reciprocal space into the density information ρ in the direct space (electron density for X-rays, probability density of atomic cores for neutrons) with

$$\rho(\mathbf{x}) \sim \Sigma_h \Sigma_k \Sigma_l \, F_{hkl} \exp(-2\pi(hx+ky+lz))$$

is not possible. This can be done only by direct methods like Patterson, heavy atom method or anomal dispersion for X-rays.

In the so called refinement program a given structural model (space group, lattice constants, atomic form factors, MSD, etc.) are compared with the experimental data and fitted. In a least squares routine those programs try to optimize (typically over several cycles) the free parameters to reduce the difference between the calculated structure factors F_{calc} and intensities $|F_{calc}|^2$ respectively and the experimentally found F_{obs} and $|F_{obs}|^2$ respectively. To quantify the quality of a measurement there are several values in use:

1. *unweighted R-value*: $R_u = \Sigma_{hkl} |F_{obs}^2 - F_{calc}^2| / \Sigma_{hkl} F_{obs}^2$
This value gives the alignment of the whole number of reflections without their specific errors.

2. weighted *R-value*: $R_w = (\Sigma_{hkl} w \, (F_{obs}^2 - F_{calc}^2)^2) / \Sigma_{hkl} w \, F_{obs}^4$
This value represents the alignment of the whole number of reflections including their specific errors or weights ($w \sim 1/\sigma^2$). Sometimes weights are adopted in a way to suppress unwanted influence of the refinement algorithm by weak or badly defined reflections. Be aware that such corrections have to be done extremely carefully because otherwise the refinement adopts the data to the selected structural model and not the model to the experimental data!

3. *Goodness of Fit S*: $S^2 = (\Sigma_{hkl} w \, (F_{obs}^2 - F_{calc}^2)/(n_{hkl\text{-reflections}} - n_{\text{free parameters}})$

Ideally, S should have a value near one if the weighting scheme and the structure model fit to the experimental data set.

2   Powder Diffraction

2.1   Method

Crystalline powder consists of a large number of randomly oriented microscopic single crystals, so called crystallites. Using a monochromatic beam a powder sample in a small capillary or on a flat table generates cones of diffracted radiation, the so called Debye-Scherrer cones. Following Bragg's law the discrete angles at which intensity can be found by a photo film, single or area detector one can calculate the crystallographic unit cell and its fundamental symmetry (trigonal, monoclinic, hexagonal, orthorhombic, tetragonal, cubic). The projection of the integrated intensities from the detector area yields a diffractogram that presents the intensity distribution against 2θ and Q respectively. In a Rietveld procedure the full reflection pattern (profiles and intensities) is analysed.
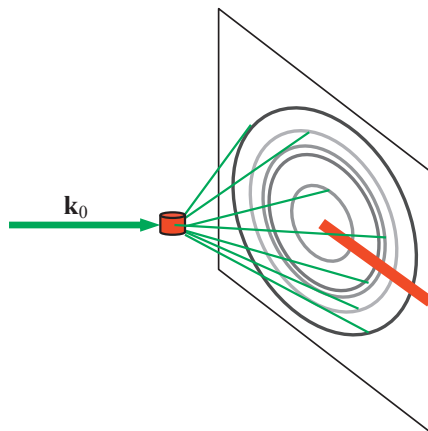
Fig. 10: Sketch of powder diffraction towards a flat 2D detector (red line: 1D detector).

The main advantage of powder diffraction is the speed of data collections. At high flux sources a complete diffractogram can be collected within minutes. This allows not only to perform data analysis routinely and for many samples in a short time but can also be used to do in situ experiments where for instance a chemical reaction can be visualised by changes in a sequence of diffractograms measured one after each other. The other advantage with powder samples is the fact, that it is much easier to prepare a sample with a huge number of small crystals instead of one large single crystal. The main disadvantage of powder diffraction is that lattice planes that are crystallographically different will occur at the same position in the diffractogram due to the projection of diffraction intensities towards the diffraction angle. One problem of sample preparation can occur from crystallites of anisotropic shape (needles, plates). In powder samples these crystallites might order not fully randomly. As a result, the Debye-Scherrer cones are not homogeneously illuminated which yields diffractograms with misleading intensity distributions. This effect is called texture.
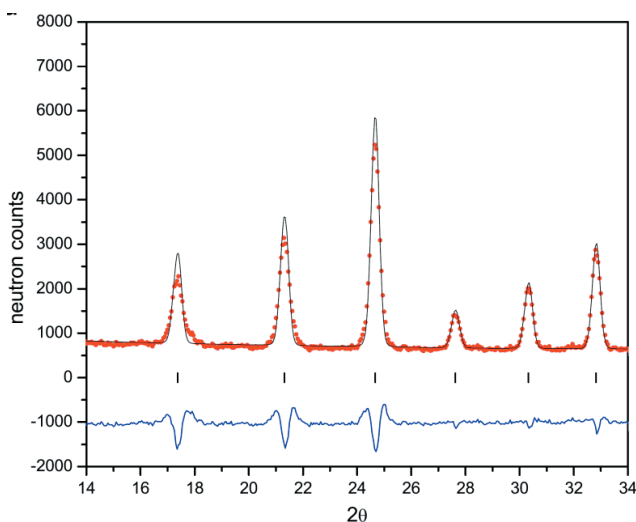
Fig. 11: Diffractogram with experimental data (dots), fit (line) and difference plot (blue) [19].

2.2 Instrumentation

Laboratory X-ray powder diffractometers use characteristic K_α radiation from a Cu, Mo, Ag or W targets in an X-ray tube. These instruments have a high flux and good resolution. Powder diffractometers at a synchrotron source offer additionally variable wavelengths, very high photon flux up to 10^{12} photons/s and very high resolution.

The following figure shows on the left a laboratory X-ray diffractometer. Its beam tube generates mainly characteristic K_α and K_β radiation while the amount of "Bremsstrahlung" is rather small.

As the K_β radiation can be cut out instantly by a metal foil filter (e.g. Zn for Cu radiation) behind the beam tube no additional measures have to be taken to use a monochromatic primary beam. In this case the sample stage stays fixed at its position. The source moves clockwise while the detector does the same counter clockwise yielding at any time an angle between primary beam and sample of θ and 2θ between primary and secondary beam to fulfil Bragg's law.

An example for a thermal neutron powder diffractometer with fixed wavelengths of 1.11 Å, 1.55 Å or 2.54 Å is the instrument SPODI in the experimental hall of the neutron research source Heinz Maier-Leibnitz (FRM II) on the right. As the neutron source generates a broad spectrum of thermal neutrons this "white beam" needs to be monochromatised by a monochromator crystal first (e.g. Cu-220, Ge-311, etc.). The resulting monochromatic neutron flux is several orders of magnitude smaller than the photon flux at a synchrotron.

To overcome this problem the detector consists of a set of ³He detectors covering about 160° in 2θ and counting simultaneously to measure several diffractograms per day.
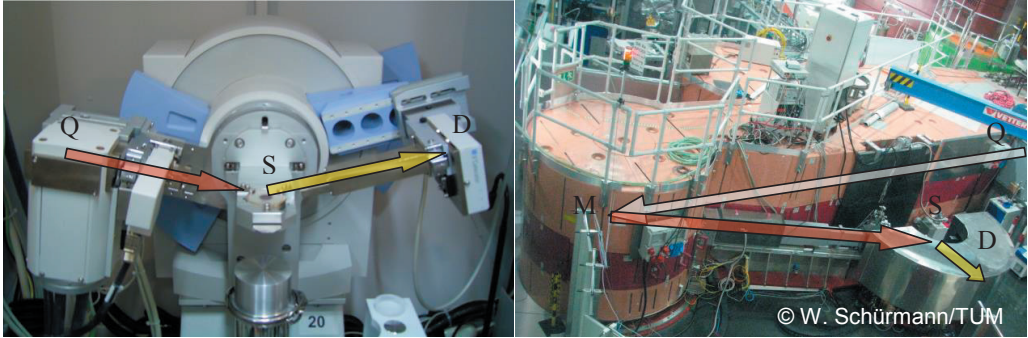
Fig. 12: Powder diffractometer: left: Laboratory X-ray diffractometer with $\theta/2\theta$ geometry (Bragg-Brentano), right: Neutron diffractometer SPODI (Q: source, S: sample, D: detector, M: monochromator).

2.3   Examples

The method of powder diffraction warranties that no Braggs reflections inside the observed 2θ range are missed. This is highly advantageous if samples of uncertain composition and/or structure have to be studied and can also be used to observe multiple phases simultaneously or behind each other, e.g. in temperature dependent experiments. As data collection in powder diffraction is rather fast these in situ experiments allow not only to study the results of a structural or chemical change but also the intermediate states.

In Situ Experiments on Zirconia Zirconia play an important role in many industrial applications like electrolytes, ion exchangers, catalysators and piezo-electrics. Therefore, the detailed understanding of its structural properties and synthesis are very important. A typical synthesis route is shown in fig. 13.
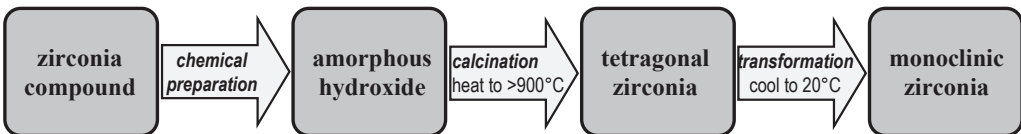
Fig. 13: Wet chemical/furnacing route [9].

Without going into details the following key aspects can be shown: The zirconium compound at the beginning of synthesis – $Zr(OH)_2SO_4 \cdot 3H_2O$ – contains a huge amount of water/OH groups and shows amorphous behaviour. During the heating process a loss of water/OH groups is generated that results in a crystallisation in the tetragonal zirconia phase. After cooling down the structural phase transition from the tetragonal to monoclinic phase occurs during the cooling cycle but depends strongly on the dwelling time (for details see [9] and references within).
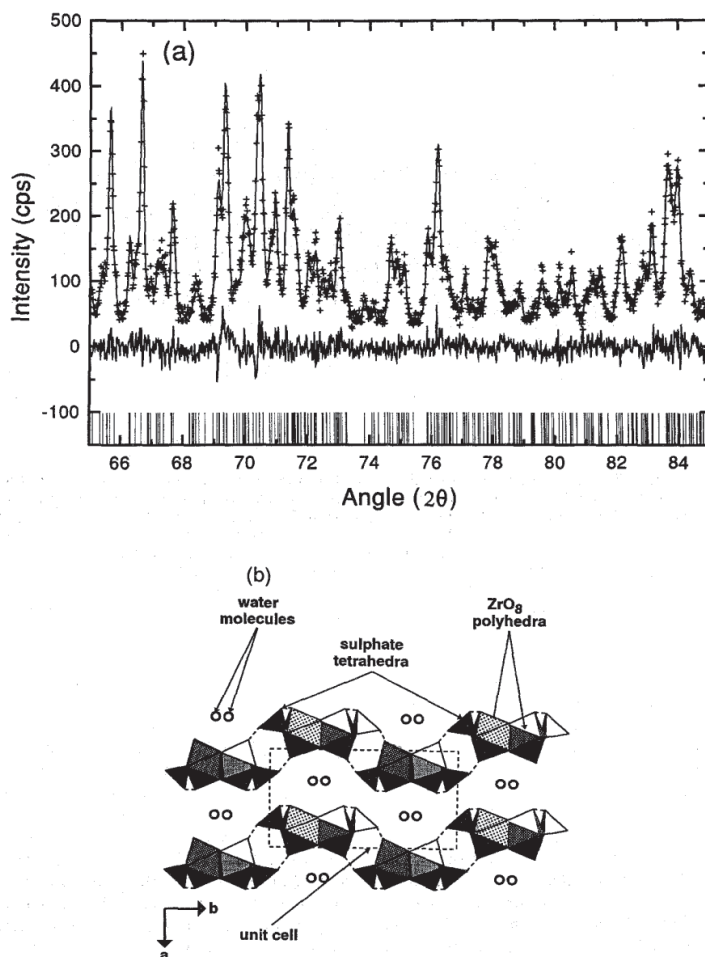
Fig. 14: T dependent laboratory powder diffractogram of Zirconia [9].

(a) Part (65-85°C) of the laboratory X-ray powder diffraction pattern obtained on $Zr(OH)_2SO_4 \cdot 3H_2O$ is shown to illustrate the detail and high quality that can be achieved using a fully monochromatized laboratory X-ray source. The data are displayed as a *Rietveld plot* in which the upper trace shows the observed data as dots and the calculated pattern as a solid line; the middle trace gives the difference between the observed and calculated data; and the lower set of vertical bars identifies each reflection. (b) A representation of the $Zr(OH)_2SO_4 \cdot 3H_2O$ structure (after Gascoigne et al. 1994) illustrating zigzag chains, viewed end on down the **c**-direction, which are hydrogen bonded together (broken lines) to form curraged sheets, parallel to **a** and **b**, which themselves are hydrogen bonded together via interlayer water molecules [9].

The figure above shows a part of a laboratory powder diffractogram and the corresponding Rietveld plot (perpendicular dashes below). The second half of this figure shows the tetragonal structure. The following figure shows the corresponding angular dispersive powder diffractograms for laboratory and synchrotron sources. The superior resolution of the synchrotron source is obvious.
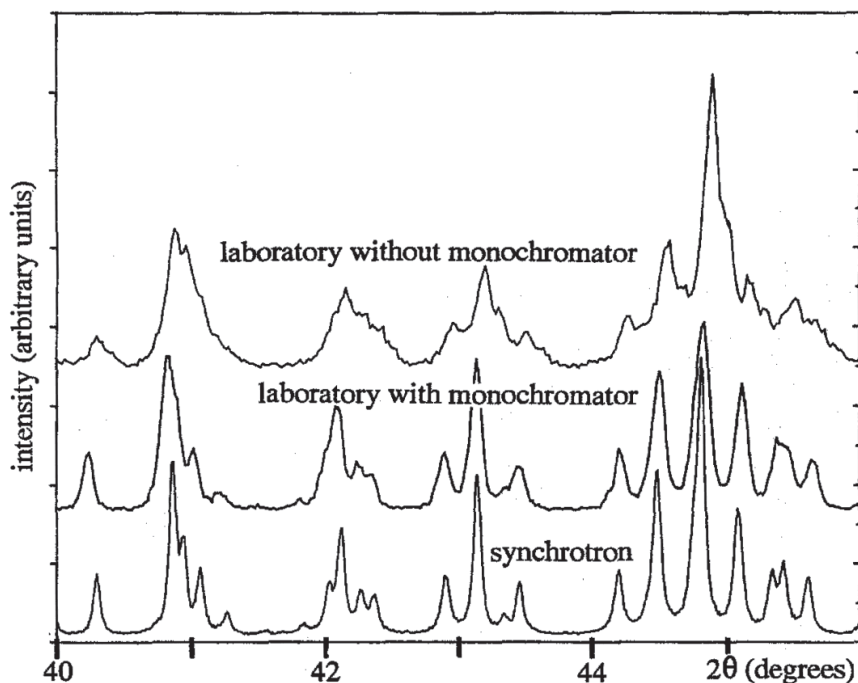
Fig. 15: Comparison of laboratory and synchrotron powder diffractograms [9].
  Portions (considerably expanded, between 40° and 46°) of powder diffraction patterns for $Zr(OH)_2$ $SO_4 \cdot 3H_2O$ obtained using three different collection modes: the upper trace is with a conventional (Bragg-Brentano) laboratory X-ray powder diffractometer utilizing the full $K_{\alpha 1 \alpha 2}$-doublet from a Cu X-ray source (average wavelength $\lambda = 1.5418$ Å); the middle trace is again with a laboratory X-ray source but with the $K_{\alpha 2}$ radiation removed by means of a germanium (100) monochromator [resultant $\lambda(K_{\alpha 1}) = 1.5406$ Å]; the lower trace is from a synchrotron (station 2.3 on the Daresbury source) powder diffractometer monochromated to $\lambda = 1.51603$ Å. the progressive improvements in pattern quality are self-evident, particularly with the closely-spaced multiple peaks around 41°, 42° and 45°.

The final figure (fig. 16) of this section presents the T dependent synchrotron energy dispersive powder diffractogram (EDD) which shows clearly the evolution and devolution of Bragg peaks during the different stages covering the range from the amorphous pattern of the hydroxide to the monoclinic oxide (The patterns were collected on station 9.7 of the Daresbury synchrotron source using a diffraction angle of $2\theta = 7.8°$ and a collection of 60s per pattern [9]).
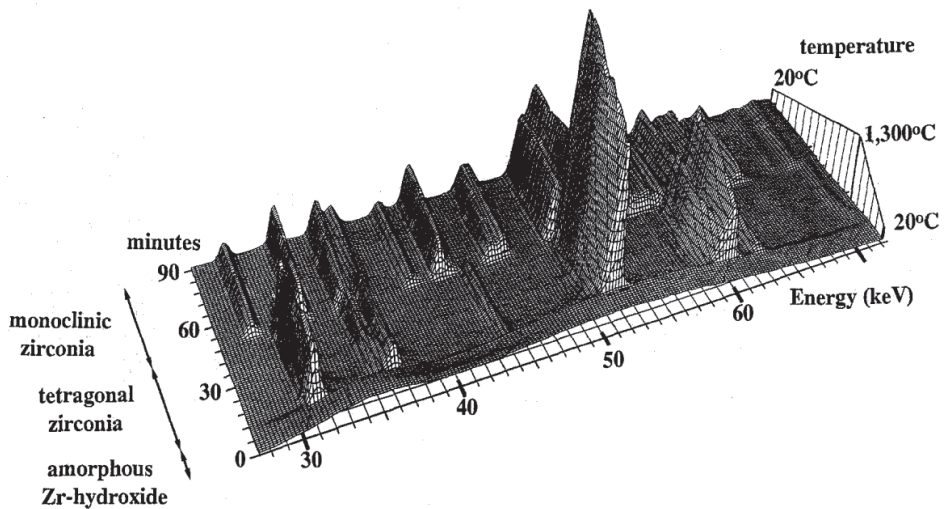
Fig. 16: *T* dependent synchrotron powder diffractogram of Zirconia [9].

3   Single Crystal Diffraction

3.1   Method

Classical angle dispersive single crystal diffractometers use a fixed constant wavelength λ. As the sample is one large single crystal, a lattice plane fulfilling Bragg's law will generate not a Debye-Scherrer cone but a single diffracted beam. The schematic drawing in fig. 17 shows a single crystal (red dot) in the centre of a Eulerian cradle with three axes ω, χ, φ. These three axes allow the crystallographic coordinate system of the sample to be oriented relatively to the coordinate system of the diffractometer towards any direction. Therefore any lattice plane can be oriented in a way that its diffracted beam hits the detector in the diffraction plane defined by the 2θ axis. By measuring each lattice plane separately the problem of reflection overlaps as described above for powder diffraction can be overcome. Also certain anisotropic effects can be detected that stay hidden in powder diffractograms. But the not simultaneous measurement of Bragg intensities is very time consuming. This problem can be healed partially by using image plate detectors.
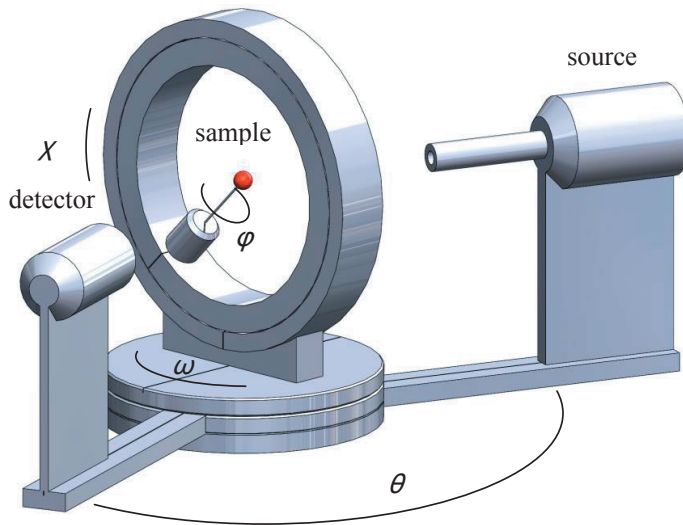
Fig. 17: Scheme of a single crystal diffractometer with single detector and Eulerian cradle.

3.2 Instrumentation

An example for a classical neutron single crystal diffractometer is HEiDi from JCNS at the hot source at FRM II (fig. 18). The instrument offers discrete wavelengths between 1.2 Å and 0.4 Å and a large Q range.

The use of single detectors makes diffraction of single crystals a very time consuming task. Especially for X-ray diffraction analysis the development of new large area detectors, so called image plates with a laser read out at the end of last century helped to drop the necessary beam time for an experiment by up to one order of magnitude [10]. Comparable to the rotating crystal technique the sample is only rotated around one fixed axis φ. Special software transforms the collected Bragg intensities into an image of the intensity peak distribution in the reciprocal space.
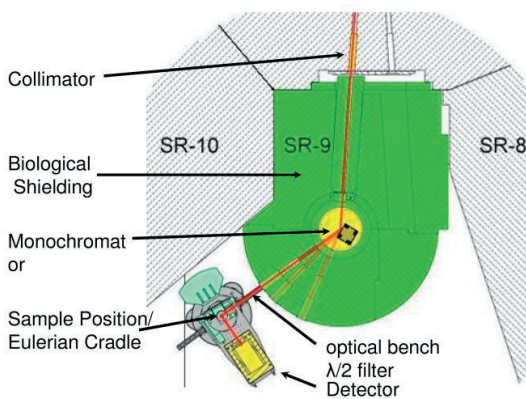
Fig. 18: Classical neutron diffactometer HEiDi with Eulerian cradle and mounted cryostat.
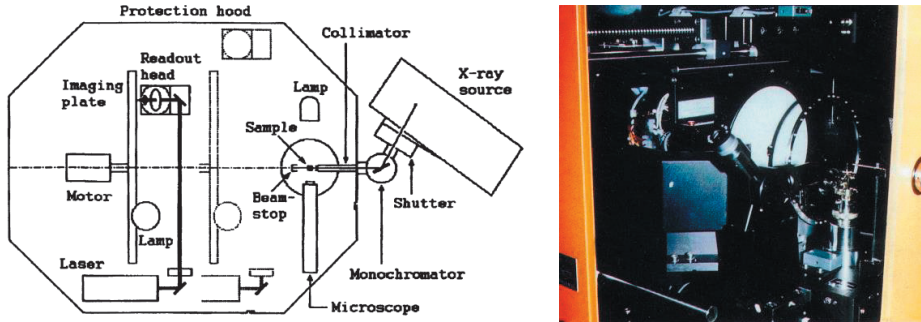
Fig. 19: X-ray single crystal image plate diffractometer (IPDS STOE, [10]).

3.3   Examples

Mixed Crystals - ZrAs$_x$Se$_y$: This metallic compound shows an anomaly in its electric resistivity below 16 Kelvin [22]. The chemical structure of this compound at room temperature is tetragonal (*P4/nmm*, a = 3.78 Å, c = 8.14 Å). To investigate the question, whether the electric resistivity is related to a possible order/disorder or redistribution of the As and Se atoms on crystallographic *2a* and/or *2c* positions taking into account the possibility of vacancies requires the combination of X-ray and neutron diffraction [11].
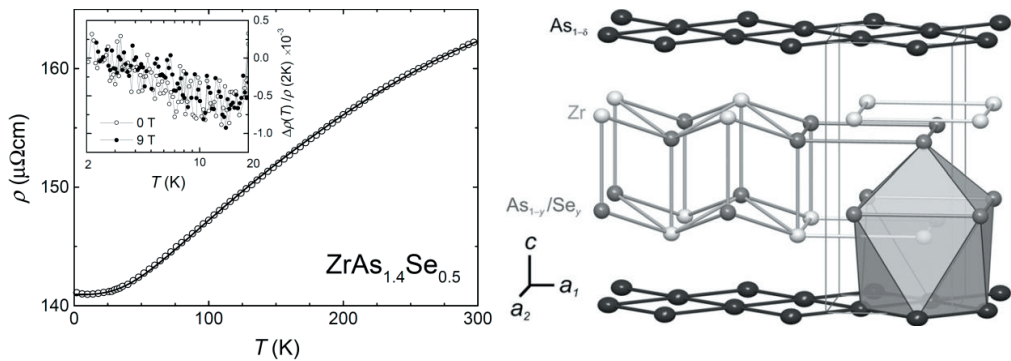
Fig. 20: Electrical anomaly of ρ below 16 K [22] and As/Se distribution in ZrAs$_x$Se$_y$ [11].

X-ray diffraction cannot distinguish between As and Se because of Z_{As} = 33, Z_{Se} = 34 are too similar. Nevertheless, X-ray diffraction can be used to determine the number of vacancies at the possible lattice positions of both elements. Chemical analysis is used to check the stochiometry of the samples. These constraints were used in a single crystal diffraction neutron experiment to determine the As and Se distribution in this compound taking into account the different neutron scattering lengths of As and Se with b_{As} = 6.58 fm, b_{Se} = 7.97 fm. The data shown in the table below were taken at three different temperatures, two above and one below the anomaly. Please notice that at room temperature a Bragg data set was taken

up to $Q \approx 10$ Å^{-1} while the two low temperature Bragg data sets were taken up to $Q \approx 14$ Å^{-1}. The increase in Q range by about 40% increases not only the number of available reflections by a factor of about 4 but drops the error bars of the atomic parameters by one order of magnitude!

The atomic positions of the elements in the unit cell of this compound are not randomly chosen. According to the space group of the unit cell of this compound there are symmetry restrictions. For instance, the first As position (3/4 1/4 0) stays fixed during the refinement process. Only the z positions of the other atoms and the anisotropic mean square displacements and occupancies of two of the three atomic positions can be set as free parameters. In combination with two additional parameters (scaling parameter for the normalisation of the sample volume and extinction parameter for eliminating extinction effects) the maximum number of free parameters is limited to 12.

Specimen		Crystal I		Crystal II[a]	
Chemical composition (WDXS)		$Zr_{1.000(3)}As_{1.595(3)}Se_{0.393(1)}$		$Zr_{1.000(2)}As_{1.420(3)}Se_{0.560(2)}$	
Crystal dimensions/mm		$2 \times 2 \times 2$		$1.5 \times 1.5 \times 1.5$	
Temperature/K		300		25.0(1)	2.3(1)
Radiation		Neutron		Neutron	
Diffractometer		RESI		HEiDi	
Wavelength/pm		104.9(1)		55.2(1)	
Space group		P4/*nmm*		P4/*nmm*	
Z		2		2	
Cell dimensions	a	378.30(7)		375.0(3)	
Neutron/pm	c	813.6(1)		803.7(1)	
$V/10^6$ pm^3		116.44		113.02	
Cell dimensions	a	375.76(2)		375.86(2)	
X-ray/pm	c	807.80(5)		805.45(7)	
$2\theta_{max}$/deg		110.8		75.0	
hkl range		± 4	$-6-+8$		$-6-+7$
		$-4-+3$	$-3-+8$		$-3-+7$
		± 12	$-17-+16$		± 16
No. reflections		355	1679		1518
No. parameter		12	12		12
GOF		1.26	1.09		1.14
$R1/wR2$		0.051/0.132	0.027/0.057		0.023/0.047
Largest difference peak		0.74	0.97		0.96
Refined composition		$ZrAs_{1.48(10)}Se_{0.48}$	$ZrAs_{1.36(2)}Se_{0.62}$		$ZrAs_{1.33(2)}Se_{0.65}$

Fig. 21: Table of Bragg data statistics with neutron single crystal diffraction [11].

Atom	Site	x	y	z	U_{11}	U_{33}	U_{eq}	occ.
Zr	2c	$^1/_4$	$^1/_4$	0.2666(2)	0.0062(6)	0.0072(7)	0.0065(5)	1
				0.26505(3)	**0.00247(7)**	**0.00181(7)**	**0.00225(6)**	
				0.26506(2)	*0.00237(7)*	*0.00176(7)*	*0.00217(6)*	
As	2a	$^3/_4$	$^1/_4$	0	0.0089(8)	0.0072(8)	0.0083(7)	0.96(2)
					0.00671(9)	**0.00204(8)**	**0.00515(7)**	**0.979(4)**
					0.00644(8)	*0.00193(8)*	*0.00494(7)*	*0.975(4)*
As/Se	2c	$^1/_4$	$^1/_4$	0.6205(2)	0.0058(6)	0.0077(7)	0.0065(5)	0.52(8)/0.48
				0.62157(2)	**0.00153(6)**	**0.00233(6)**	**0.00180(5)**	**0.38(2)/0.62**
				0.62152(2)	*0.00162(6)*	*0.00238(6)*	*0.00187(5)*	*0.35(2)/0.65*

Fig. 22: Results of the refinements [11].

The high accuracy of the data concerning atomic positions, MSDs and occupancies and the combination of X-ray and neutron diffraction show unambigously that there are

- only As and vacancies on the *2a* position,
- no vacancies but As and Se on the *2c* position,
- no incalation of Zr in interstitial positions and
- no structural change down to 2.5 Kelvin.

From this result, a structural change can be excluded as cause for the observed anomaly of the electric resistivity.

HT Superconductors: $La_{2-x}Sr_xCuO_4$ is one of the cuprate superconductors with K_2NiF_4-structure (layered perovskite) for whose discovery the Noble prize was granted in 1988 (Bednorz and Müller [12]). Pure La_2CuO_4 is an isolator while doping with earth alcali metals (Ca^{2+}, Sr^{2+}, Ba^{2+}) on the La^{3+} lattice positions generates depending on the degree of doping superconductivity with a maximum T_c of 38 K for Sr doping of $x = 0.15$ (see fig. 23).
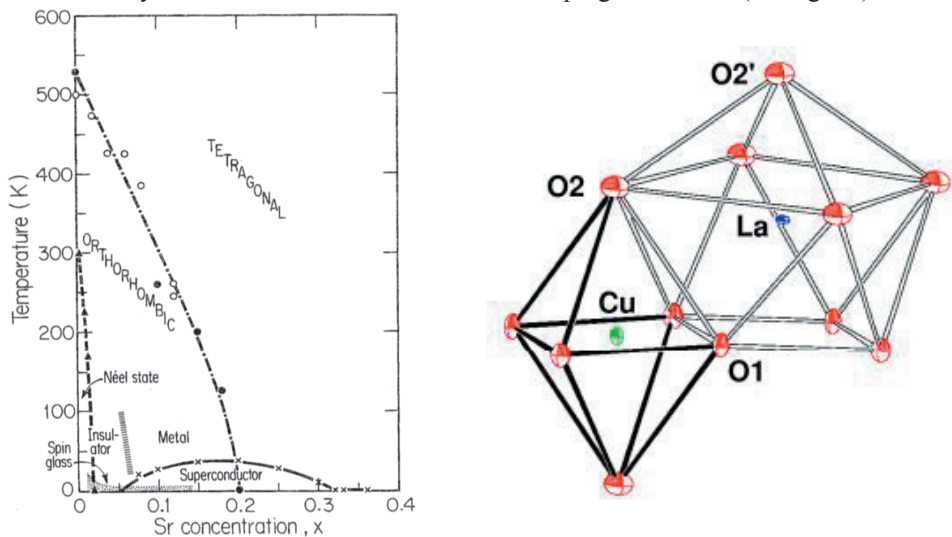
Fig. 23: Phase diagram [13] and structural units, e.g. CuO_6 octahedra and LaO_9 polihedra.

Pure La_2CuO_4 undergoes at $T_{t-o} = 530$ K a structural phase transition from the tetragonal high temperature phase (HTT) *F4/mmm* ($a = b = 5.384$ Å, $c = 13.204$ Å, $\alpha = \beta = \gamma = 90°$ at $T_{t-o} = 540$ K) to the orthorhombic low temperature phase (LTO) *Abma* ($a = 5.409$ Å, $b = 5.357$ Å, $c = 13.144$ Å, $\alpha = \beta = \gamma = 90°$ at room temperature). For $La_{2-x}Sr_xCuO_4$ the phase transition temperature T_{t-o} drops with increased doping and disappears above $x = 0.2$.

The following aspects of structure analysis can be learned from this example:
Twinning: The structural phase transition yields a symmetry reduction that makes the single crystal separate into domains of identical structure with different but well defined orientations to each other. This effect is called *twinning*. As can be seen in fig. 24 the transition into the low temperature phase tilts the CuO_6 octahedrons around their [010] axis. The two axes of identical length in the HTT phase, $\mathbf{a}_1$ and $\mathbf{a}_2$, are not equal in the LTO phase anymore. Instead, the longer one becomes the new **a** axis, the shorter one becomes the **b** axis. Whether $\mathbf{a}_1$ or $\mathbf{a}_2$ becomes the new *a* axis depends only on the real structure of the crystal, for instance grain boundaries or point defects. Two equivalent crystallographic space groups describe the LTO phase:

Abma ($\mathbf{a}_1 \to \mathbf{a}$, $\mathbf{a}_2 \to \mathbf{b}$) and Bmab ($\mathbf{a}_1 \to \mathbf{b}$, $\mathbf{a}_2 \to \mathbf{a}$)

For the structure factors in the LTO is valid: $F_{Abma}(hkl) = F_{Bmab}(khl)$
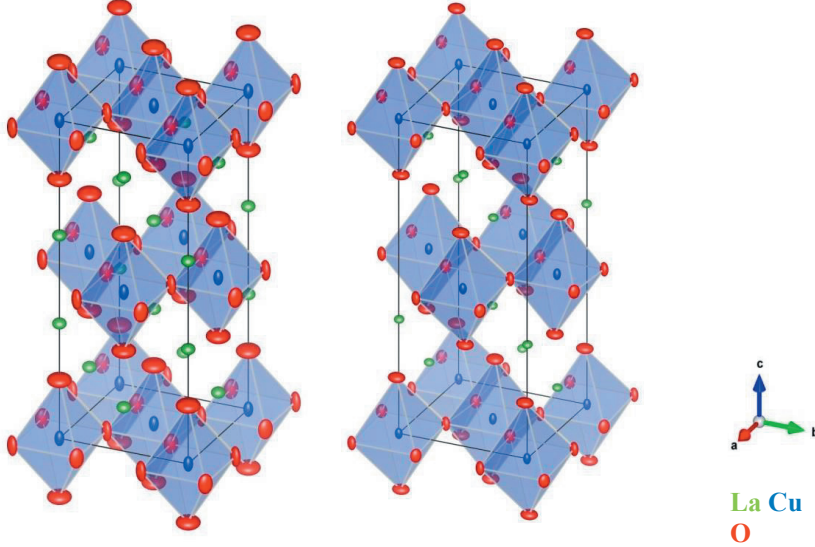

Fig. 24: Tetragonal HTT phase (left) and orthorhombic LTO phase (right) [21].

In the real structure of the crystal there exist four domain types in total which are separated in pairs with $Abma_1/Bmab_1$ (I/II) with the (1-10) mirror plane as grain boundary and $Abma_2/Bmab_2$ (III/IV) with the (110) mirror plane as grain boundary, see. fig. 25.

The most accurate observation of this effect is possible only by single crystal diffraction if the diffraction plane and the *ab* plane of the sample are parallel. Be aware that the *d* spacing for the {220} reflections of all 4 domains remains identical in the orthorhombic phase and no reflection splitting is observable with powder diffraction. The {400} reflections separate into {400} and {040} reflections with slightly different *d* spacing and can be revealed by powder diffraction only if resolution is sufficiently high.

An equal distribution of the volumetric portion of each single domain yields a ratio of intensities of 1:2:1 for the triple splitting. The distance $\Delta\omega$ between the single peaks of a (hkl) reflex gives because of $(a + b)/2 = a_{1/2}$ an information about the orthorhombic *a/b* splitting. For the triple splitting of a (hh0) reflex is valid: $\Delta\omega = 90° - 2\arctan(b/a)$
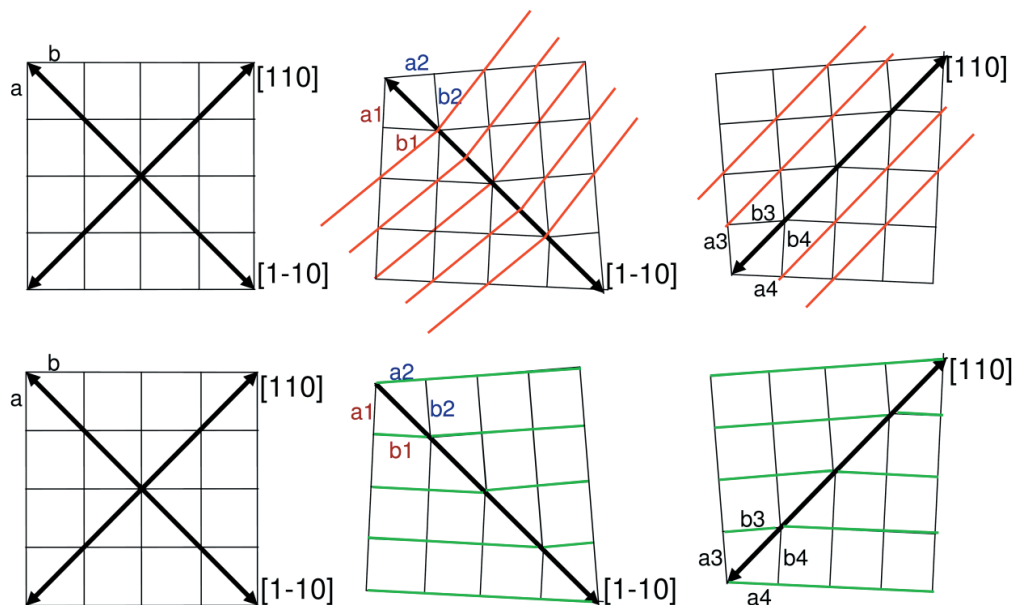
Fig. 25: HT phase (left) and domain pair (right) with (220) planes (red) and (400) planes (green), **c** perpendicular.
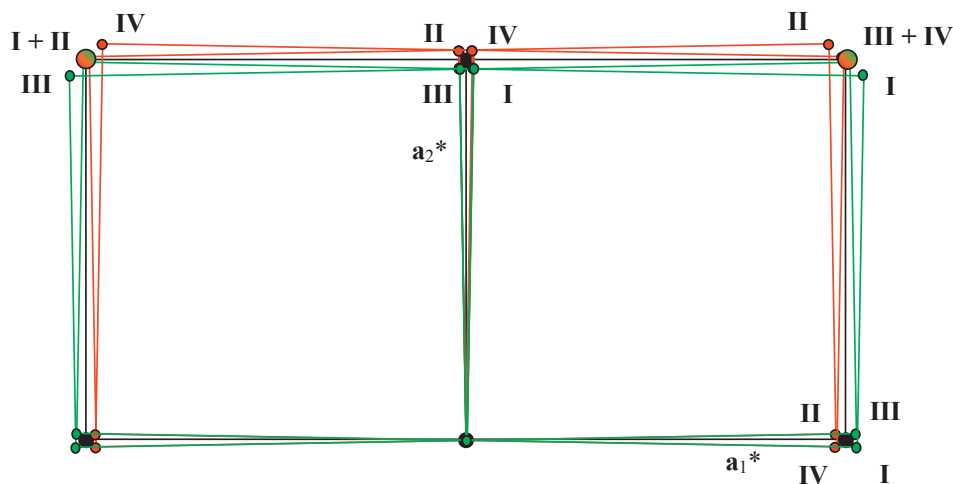
Fig. 26: Reciprocal space of twinned La$_2$CuO$_4$ (**a**$_1$* and **a**$_2$* from tetragonal space group, **c*** perpendicular).

Due to the face (=F) centering in the HTT phase only reflections with h, k, l of equal parity (g for even, u for odd) are allowed - (uuu) and (ggg). They are called in the following main structure reflections. The loss of symmetry in the LTO phase generates additional reflections, called super structure reflections (e = even index, u = odd index): In the Abma structure (ugg),

$l \neq 0$ and (*guu*), in the B*mab* structure (*gug*), $l \neq 0$ and (*ugu*). Forbidden remain in both the HTT and the LTO phase (*uuu*), (*ggu*), (*ug0*) and (*gu0*). There is no overlap between the superstructures from the A*bma* and the B*mab* domains. Therefore, although the real crystal is twinned, one can quantify the orthorhombic distortion.
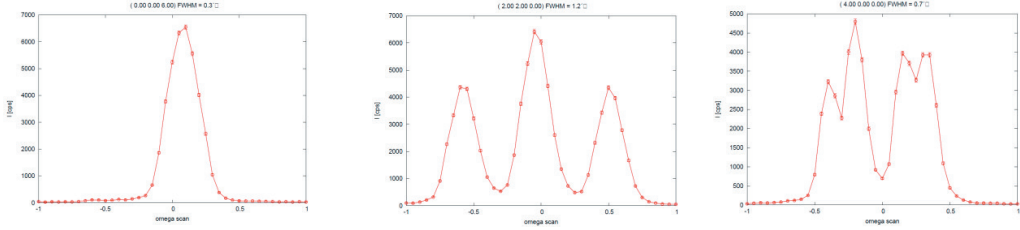
Fig. 27: Intensity distributions of (006), (220) and (400) reflection of twinned La$_2$CuO$_4$.

The intensity contribution of the single domains corresponding to the whole intensity of a reflection can be described (taking into account the incoherent overlap of single intensities and the volumetric portions V_{A1} to V_{B2} of the domains) as follows:

$I_{obs}(hkl)$ $\quad = I_{Abma1}(hkl) + I_{Bmab1}(hkl) + I_{Abma2}(hkl) + I_{Bmab2}(khl)$ or

$V_{total}|F_{obs}(hkl)|^2 = V_{A1}|F_{Abma1}(hkl)|^2 + V_{B1}|F_{Bmab1}(hkl)|^2 + V_{A2}|F_{Abma2}(hkl)|^2 + V_{B2}|F_{Bmab2}(hkl)|^2$

$\qquad\qquad = (V_{A1} + V_{A2})|F_{Abma1}(hkl)|^2 + (V_{B1} + V_{B2})|F_{Bmab1}(hkl)|^2$

$\qquad\qquad = V_{total} \{\alpha|F_{Abma}(hkl)|^2 + (1 - \alpha)|F_{Abma}(khl)|^2\}$

with α being the relative portion of the volume of *Abma* domains to the crystal.

Because of the extinction rules in the LTO phase for the super structure reflections is valid: $I_{obs}(hkl) \sim \alpha|F_{Abma}(hkl)|^2$ for *Abma* and $I_{obs}(hkl) \sim (1 - \alpha|F_{Abma}(khl)|^2$ for *Bmab*. Thus, one can classify directly intensities to the volumetric portions of the domain types *Abma* and *Bmab* respectively. Therefore, by using one single additional parameter α to describe the relation between the twins in the structure one can determine the orthorhombic single crystal structure! This holds true although the Bragg reflections contain contributions of up to four different domains.

Light elements: The phase transition is driven by a displacement of the oxygen atoms (see fig. 24). As the oxygen atoms are much lighter than any other element in this compound, the accurate observation of these displacements depends strongly on the chosen radiation. The atomic positions of Cu and La do not change significantly between the HTT and LTO phase and the structure factor for the superstructure reflections can be written as

$F(hkl) \sim \sum_i s_i \exp(-2\pi\iota(hx_i + ky_i + lz_i) = F(hkl)_{\text{apex oxygen}} + F(hkl)_{\text{in plane oxygen}} + F(hkl)_{\text{Remains}}$
$\qquad\qquad\qquad\qquad \rightarrow F(hkl)_{\text{apex oxygen}} + F(hkl)_{\text{in plane oxygen}}$

As the apex oxygen moves away from the $z = 0$ position to $(x\ 0\ z)$ the corresponding superstructure reflection for h uneven is

$F(hkl)_{apex\ oxygen} = \sin(2\pi h x)\cos(2\pi l z)$ for h odd

In the case of X-rays the weak form factor of the oxygen ($Z = 8$) against the Cu ($Z = 29$) and La ($Z = 57$) atoms make this intensity contribution almost invisible ($\ll$ 1% of main reflections). In the case of neutrons the scattering lengths of all atoms are in the same order of magnitude ($b_O = 5.803$ fm, $b_{Cu} = 7.718$ fm, $b_{La} = 8.24$ fm) and therefore also the superstructure reflections yield easily measureable intensities significantly larger than 1% of the strongest main structure reflections.

Mean square displacements: Pure La_2CuO_4 shows a purely linear behaviour of the mean square displacements with temperature. Deviations from this harmonic behaviour of the Debye-Waller-factors can be a hint for a disturbance like an order-disorder phase transition. As the $La_{1.85}Sr_{0.15}CuO_4$ compound shows the highest $T_c = 38$ K it was discussed whether an order/disorder phase transition could be related to superconductivity. Bragg data sets taken with neutron single crystal diffraction at three temperatures above and below the structural phase transition ($T_{t-o} = 186$ K) and the superconducting state ($T_c = 38$ K) show no anomaly for all atoms including the two oxygens O1 (in-plane) and O2 (apex) (fig. 28). The only anomaly found there is the increase of $U_{33}(O1)$ and $U_{11}(O2)$ for *all* temperatures compared to the undoped La_2CuO_4. Harmonic lattice dynamical calculations from experimentally determined phonon dispersion curves taking into account the Sr doping were in good agreement with this observation. Thus, the random distribution of Sr atoms on La sites introduces static disorder into the structure [14].
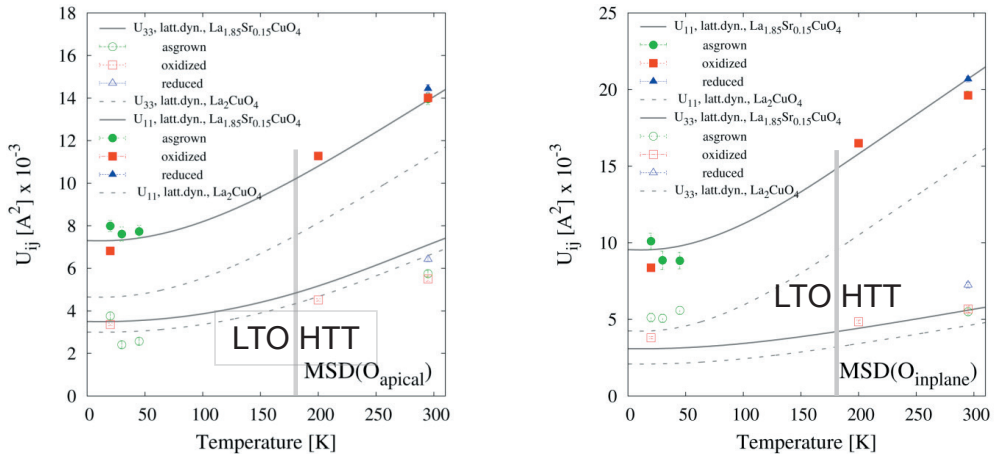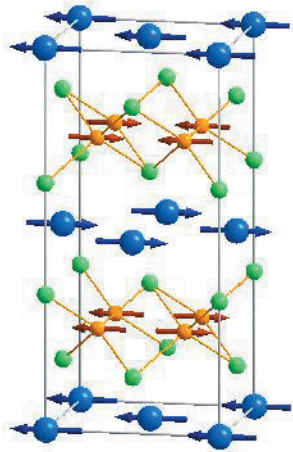

Fig. 28: MSD of O_{apical} (left) and $O_{inplane}$ (right) for $La_{1.85}Sr_{0.15}CuO_4$ (data from [14], grey vertical line = T_{t-o}). The dotted lines in the middle of the diagrams for O1 and O2 are $U_{33}(O1)$ and $U_{11}(O2)$ of the undoped parent compound La_2CuO_4.

Magnetic compounds: Some years ago FeAs compounds were found to show – like the cuprates – superconductivity if doped. It was also found that the structural properties are similar to the cuprates, e.g. a layered structure and a phase transition from a tetragonal high temperature phase ($I4/mmm$) to an orthorhomic low temperature phase (*Fmmm*, fig. 29).

An interesting member of these compounds is $EuFe_2As_2$, where doping with potassium generates superconductivity with $T_c(Eu_{0.5}K_{0.5}Fe_2As_2)$ = 31 K. The undoped compound shows antiferromagnetic (=AF) ordering of the Fe^{2+} and Eu^{2+} atoms but at different temperatures, 190 K and 19 K respectively. Neutron single crystal diffraction was used to study the nuclear *and* magnetic structures in detail as the magnetic moment of the neutron is sensitive to magnetic order (fig. 30) [15].

Fig. 29: *Fmmm* phase of $EuFe_2As_2$ [15].
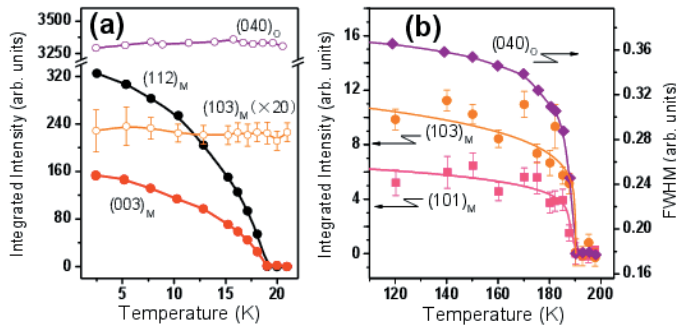
Fig. 30: T dependent measurements of magnetic Bragg reflections (Eu^{2+} left, Fe^{2+} right)[15].

Like for the example of La_2CuO_4 the orthorhombic structure is twinned. Careful profile analysis and Bragg data collections were used to reveal the details of the orthorhombic structure [15]. As can be seen in fig. 32 the nuclear and magnetic reflections are well separated. Additionally, the comparison of nuclear and magnetic reflections between measurement (F_{obs}) and model (F_{calc}) show good agreement for both (fig. 31).
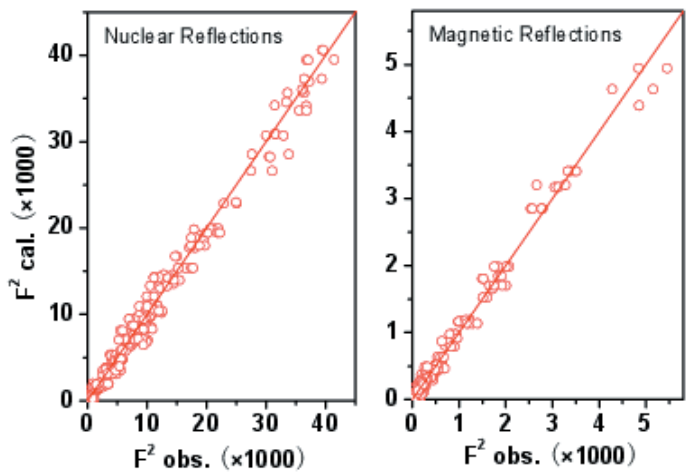
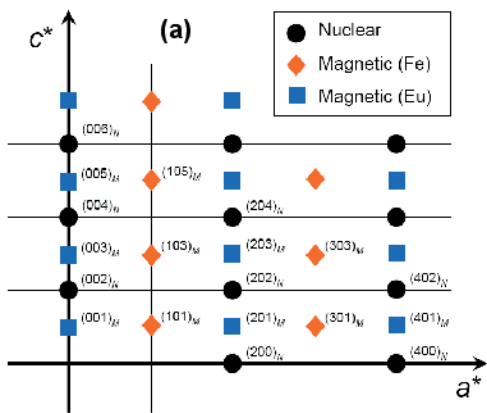**Fig. 31:** Correlation diagrams of nuclear and magnetic reflections [15].

Fig. 32: Distribution of magnetic and nuclear peaks in reciprocal space along **b*** [15].

4   Summary

Powder and single crystal diffractometry are two of the most versatile tools for detailed studies on chemical and magnetic structures. The different interactions of X-ray and neutron radiation with matter allows these techniques to contribute important informations to almost every scientific area of solid state physics, chemistry, biology and material sciences. This script can offer only a short overview of the methods and their applications. Additional information can be found in the attached list of literature.

The following table presents a compact guideline which radiation and technique might be best suited to answer a specific scientific question:

	Powder	Single Crystal
X-rays	sample mass µg	sample size µm
	very fast, in situ experiments	fast (1/2 - 2 days)
	(intrinsic & resolution dep.) reflection	very high angular resolution, anisotropic
	intensity damping by temperature factor *and* structure factor!	
	weak interaction with light elements (H, N, O, etc.)	
	no discrimination of neighbouring elements (e.g. Co, Fe, Cu)	
	large absorption effects, polarization effects	
	limited sample environment (T, H)	
	e- density maps - characterization of chemical bonds	
Neutrons	sample mass mg	sample size mm
	fast, in situ experiments	medium (3 days -7 days)
	reflection overlapp	good angular resolution
	intensity damping *only* by temperature factor	
	strong interaction with light elements, isotope specific!	
	discrimination of neighbouring elements and isotopes (H/D)	
	weak absorption effects, no polarization effects (unless pol. neutrons)	
	core and spin density maps, magnetic ordering!	

5   References

[1] Th. Hahn (ed.), Space-group symmetry, International Tables for Crystallography Vol. A, Kluver Academic Publishers (1995).
[2] W.H. Zachariasen, Acta Cryst. 18 703 (1965).
[3] W.H. Zachariasen, Acta Cryst. 18 705 (1965).
[4] P. Coppens and W.C. Hamilton, Acta Cryst. A 26 71-83 (1970).
[5] P.J. Becker and P. Coppens, Acta Cryst. A 30 129-147 (1974).
[6] P.J. Becker and P. Coppens, Acta Cryst. A 30 148-153 (1974).
[7] U.H. Zucker, E. Perrenthaler, W.F. Kuhs, R. Bachmann and H. Schulz J. of Appl. Crystallogr., 16 358 (1983).
[8] P. Coppens, W.C. Hamilton, S. Wilkins, M.S. Lehmann and Savariault, Datap, http://www.ill.fr/data treat/diftreat.html#single (1999).
[9] X. Turrillas, P. Barnes, D. Gascoigne, J.Z. Turner, S.L. Jones, C.J. Norman, C.F. Pygall and A.J. Dent; Radiat. Phys. Chem. Vol. 45, No. 3, pp. 491-508 (1995)
[10] M. Ermrich, F. Hahn and E.R. Wölfel, Textures and Microstructures 29, 89-101 (1997)
[11] R. Niewa, A. Czulucki, M. Schmidt, G. Auffermann, T. Cichorek, M. Meven, B. Pedersen, F. Steglich, R. Kniep, *J. Solid State Chem.* 183, 1309 (2010).
[12] J. Bednorz and K. Müller, Z. Phys. B 64, 189 (1986)
[13] R.J. Birgeneau and G. Shirane, Physical Properties of High Temperature Superconductors I, Editor D.M. Ginsberg, World Scientific (1989).
[14] M. Braden, M. Meven, W. Reichardt, L. Pintschovius, M.T. Fernandez-Diaz, G. Heger, F. Nakamura, and T. Fujita; Phys. Rev. B 63, 140510 (2001).
[15] Y. Xiao, Y. Su, M. Meven, R. Mittal, C. M. N. Kumar, T. Chatterji, S. Price, J. Persson, N. Kumar, S. K. Dhar, A. Thamizhavel, and Th. Brueckel, Phys. Rev. B 80, 174424 (2009).
[16] J. Strempfer, Th. Brückel, W. Caliebe, A. Vernes, H. Ebert, W. Prandl, J.R. Schneider; Eur. Phys. J. B 14, 63-72 (2000).
[17] J. Strempfer, Th. Brückel, G.J. McIntyre, F. Tasset, Th. Zeiske, K. Burger, W. Prandl; Physica B 267-268, 56-59 (1999).
[18] Th. Brückel, G. Heger, D. Richter and R. Zorn (Eds.), Lectures of the JCNS Laboratory Course, Schriften des Forschungszentrums Jülich (2007).
[19] M. Hoelzel, A. Senyshyn, N. Juenke, H. Boysen, W. Schmahl, H. Fuess; Nuclear Instruments and Methods in Physics Research A 667, 32–37 (2012).
[20] T. Brückel, A Neutron Primer in Neutron Scattering, Lectures of the JCNS Laboratory course, ed. T. Brückel, G. Heger, D. Richter, G. Roth and R. Zorn, Schriften des Forschungszentrums Jülich, Reihe Schlüsseltechnologien Bd. 15, pp 2:1-32 (2010).
[21] Meven, M. and Roth, G.. Neutron Diffraction. In Handbook of Solid State Chemistry (eds R. Dronskowski, S. Kikkawa and A. Stein), Vol. 3, Wiley-VCH Verlag GmbH & Co. KGaA, 77-108 (2017).
[22] M. Schmidt, T. Cichorek, R. Niewa, A. Schlechte, Y. Prots, F. Steglich and R. Kniep; arXiv:cond-mat/0504485v1 [cond-mat.str-el] (2005)

Literature

N.W. Ashcroft and N.D. Mermin, Festkörperphysik, Oldenbourg 2001.

H. Ibach and H. Lüth, Festkörperphysik, Einführung in die Grundlagen, 6. Ed. Springer 2002.

C. Kittel, Einführung in die Festkörperphysik, 10. Edition, Oldenbourg 1993.

W. Borchardt-Ott, Kristallographie. Eine Einführung für Naturwissenschaftler, 6. Auflage Springer 2002.

W. Kleber, Einführung in die Kristallographie, Oldenbourg 1998.

H. Dachs, Neutron Diffraction, Springer (1978).

D.J. Dyson, X-ray and Electron Diffraction Studies in Material Science, Maney Pub 2004.

C. Giacovazzo, Fundamentals of Crystallography, 2nd Ed., Oxford University Press 2002.

L.A. Aslanov, Crystallographic Instrumentation, Oxford University Press 1998.

M.T. Dove, Structure and Dynamics. An Atomic View of Materials, Oxford University Press 2003.

W. Clegg, Crystal Structure Analysis. Principles and Practice, Oxford University Press 2001.

D4 Inelastic scattering: lattice and magnetic excitations

I. Radelytskyi

Jülich Centre for Neutron Science

Forschungszentrum Jülich GmbH

Contents

1 Aim of inelastic measurements

The theoretical explanation of lattice vibrations (phonons) and magnetic interactions (magnons) in the solid state materials was presented in lecture B5. The given course depicts the common experimental equipments which can be used for getting the above-mentioned information.

The physical properties of materials depend on their microstructure and the interactions between elementary particles. The atoms in a solid oscillate near their equilibrium positions with energies depending on temperature. The lattice vibrations have an effect on the thermal/electrical conductivity, specific heat, optical and dielectric properties, diffusion mechanism, magnetic properties, phase change phenomena, etc. Hence; the understanding of magnons and phonons plays an important role. The knowledge of atoms positions in the crystal lattice and mechanism of atoms interactions allows to find and investigate new materials, and to understand deeply their physical properties.

Lattice vibrations and magnetic excitations can be studied by inelastic scattering investigations. The energy transfer notifies about the excitation energy while the momentum transfer informs about the (an)isotropy of the interactions. The detailed information of atom movements in the crystal lattice can be described by an experimentally determined dispersion. Theoretical calculations are very powerful tools for the description of above mentioned correlations, but it is still difficult to predict all experimental results.

It is worth mentioning the usefulness of the X-ray methods for structure determination. In our days, it is also possible to investigate the dynamics by X-ray inelastic scattering methods (e.g. on ID28 at the ESRF, BL12XU at the SPring-8). Nevertheless, the standard neutron investigations of dynamic properties are still most useful because of X-ray properties: low penetration depth, slight sensitiveness to the light elements and non - interaction with spins in comparison with neutrons. This lecture will focus on inelastic neutron scattering.

1.1 Dispersion

The dispersion relation is introduced by Nicolo Violini in the chapter B5. Its example of dispersion for the gold sample is presented in Fig.1.
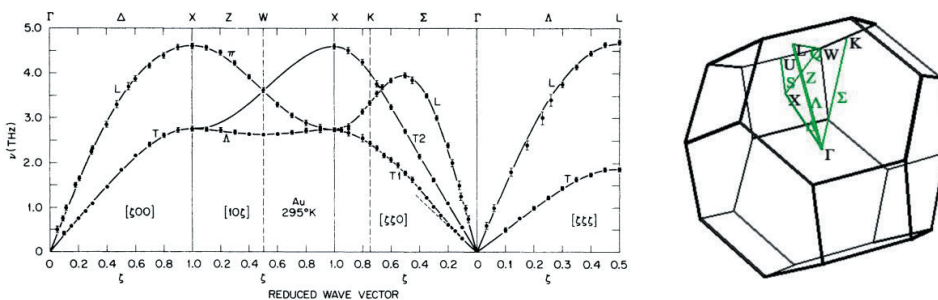

Fig. 1: Dispersion relation for fcc gold. The measurements were performed along different main symmetry directions (figure taken from ref. [1]). Right: Brillouin zone of fcc lattice with the points of symmetry.

Here we point onto the connection with the experimental investigations. The atoms have bonded position in the solid state lattice. Their displacements cause a change of their neighbors position due to the bonding forces and the related energies. Therefore, the structural symmetry and the kind of atomic bonding decide their displacement in amplitude and direction. These movements are correlated and therefore to be measured as coherent waves. In other words, we call such collective lattice excitations **phonons**. Transverse and longitudinal phonon modes illustrate the displacement of atoms away from their equilibrium position perpendicular and parallel to the propagation vector of the wave, respectively. The dispersion relation describes the energy of these excitations depending on their symmetry.

Since phonons can be presented by their reciprocal lattices vector in one Brillouin zone because of the elementary cell repetition and appropriate translation symmetry in the crystal, dispersions are measured at suited q-points along all interesting directions as seen in Fig. 1.

The spin collective excitations (magnon) in magnetic materials can be described by dispersion relations, too. More details are given in the subchapter 4.2. Summarizing, the collective movement of atoms and spins are described by phonons and magnons studies, respectively. To achieve this, inelastic measurements depending on energy and momentum transfer are performed. These data provide the total experimental information for the next data analysis. Sec. 2 depicts the momentum and energy transfer principles in reciprocal space. The experimental Sec. 3 describes the widely used equipments for inelastic measurements. The basic useful formulas for data analysis are presented in the Sec. 4.

2 Kinematics

Elastic scattering. Figure 2 depicts the elastic and inelastic neutron scattering processes. The reciprocal space for the cubic structure is illustrated. The scattering plane is perpendicular to the direction [100].
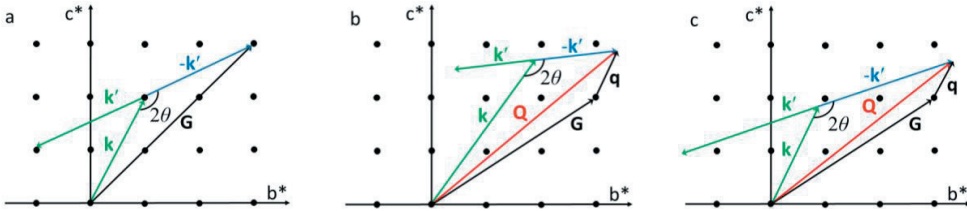

Fig. 2:   a) Scattering triangle in elastic case $|\mathbf{k}| = |\mathbf{k'}|$ (Bragg reflection). The $\mathbf{k}$ and $\mathbf{k'}$ are incident (before interaction with sample) and final (after interaction) wave vectors. b) Scattering triangle for inelastic case and energy loss $\mathbf{k} > \mathbf{k'}$. c) Scattering triangle for inelastic case and energy gain $\mathbf{k} < \mathbf{k'}$.

The energy of incident neutrons has to be similar to the energy of the outgoing neutrons in case of elastic scattering (Fig. 2a). The Bragg condition is shown in the following equation:

$$n\lambda = 2d_{hkl}sin\theta \tag{1}$$

$$|\mathbf{Q}| = |\mathbf{G}_{hkl}| = \frac{2\pi}{d_{hkl}} = 2ksin\theta = \frac{4\pi}{\lambda}sin\theta \tag{2}$$

where **Q** is momentum transfer (scattering vector), G_{hkl} is a reciprocal lattice vector, d_{hkl} represents the interplanar distance of the lattice planes, hkl denotes the Miller indices, 2θ the angle between k and k', n is a positive integer. Any point in the reciprocal space could be reached by setting the appropriate incident wave vector, application of scattering angle 2θ and sample rotation.

Inelastic scattering. In case of inelastic scattering, the incident and final neutrons have different energies. Let us consider the visualization for inelastic scattering (Fig. 2b and 2c). The example of energy loss $E > E'$ is presented in Fig. 2b in contrast to the energy gain $E < E'$ in Fig. 2c.

If neutrons with a wavevector k interact with a single crystal and scatter at an angle 2θ with a wavevector k', the momentum transfer **Q** is:

$$Q = k - k' = G_{hkl} + q \tag{3}$$

where q represents the momentum transfer related to the reciprocal lattice vector.

There are kinematic and physical limits of the accessible energy range and momentum transfers for the inelastic neutrons scattering. Measurements are possible when the scattering triangle is closed, this is called kinematic constraints. An example of the available (**Q**,E) space for different wavelengths is given in Fig.3 for a typical time-of-flight-spectrometer:
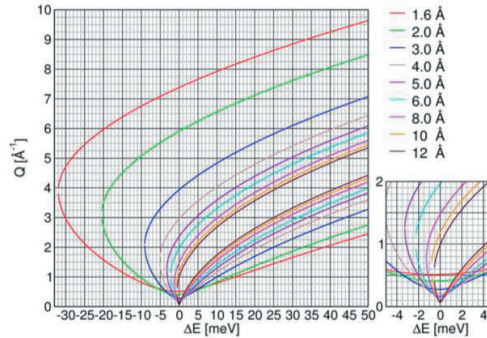

Fig. 3: The accessible (**Q**,E) space for the time-of-flight instrument, shown here for TOFTOF at MLZ (figure taken from ref. [2]).

The energy transfer is expressed by the following formula $\hbar\omega = E - E'$. The relation between **Q**, k and k' can be calculated on the basis of scattering triangle:

$$Q^2 = k^2 + k'^2 - 2kk'cos2\theta. \tag{4}$$

Using the energy transfer formula, the fundamental limit of experimental dynamic range can be described by the eq. (5):

$$\frac{\hbar^2 Q^2}{2m_n} = 2E - \hbar\omega - 2\sqrt{E(E - \hbar\omega)}cos2\theta. \tag{5}$$

It means that during the measurements we get the limit of accessible **Q** and E points even if it is physically possible to perform experiment for more broad (**Q**,E) range by using appropriate

incident and scattered neutrons energies E, E' and scattering angle 2θ. In addition, there are physical limits connected to the available space due to the spectrometer itself or the mounted sample environment. The E, E' should be chosen carefully to close the scattering triangle.

3 Methods

There are different techniques used for inelastic neutron scattering which can cover more than six order of magnitude in energy transfer. Such wide energy range and dedicated sample environment make it impossible to build one universal instrument. It is important to know the energy of excitation in the sample. The neutron energy ranges from few meV to 100 meV, which match the energy scale of the investigated excitations. Appropriately, the instruments are used by one or two methods of the neutron beam monochromatisation such as Bragg reflection and time-of-flight methods. Incident neutron energy is usually selected either with a monochromator making use of the Bragg law (TAS instruments) or a velocity selector (TOF instruments). These two techniques will be described in the following chapter.

3.1 The neutron three-axis spectrometer

The three-axes spectrometer (TAS) are used for single crystal investigation. This concept was invented by Bertram Brockhouse in 1961, and he was awarded for it by the Nobel Prize in 1994. The name TAS comes from the three main rotating axes of the instrument: monochromator, sample table and analyzer. (see Fig. 4).
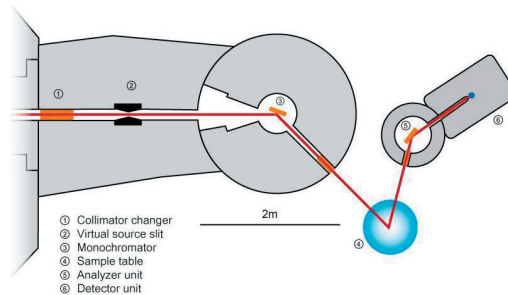

① Collimator changer
② Virtual source slit
③ Monochromator
④ Sample table
⑤ Analyzer unit
⑥ Detector unit

Fig. 4:   The schematic construction of cold-neutron three-axes spectrometer PANDA at MLZ (figure taken from ref. [3]).

The first axis is the monochromator. From the white incoming neutron beam the needed neutron energy is choosen by changing the scattering angle at the monochromator (crystals) following the Bragg law (2). The monochromatic neutron beam of energy E is scattered onto the sample by rotating the sample table and the secondary spectrometer – analyzer and detector – around the monochromator. To fulfill the (in)elastic condition on the sample, every change in scattering angle needs a change in sample rotation – the sample is the second axis of rotation. The analyzer is positioned after the sample. To realize the requested energy transfer ΔE, the neutron final energy E' is selected at the analyzer by Bragg scattering – third axis of rotation. The neutrons with the correct final energy are scattered towards the detector.

In addition, there are options to modify the properties of the equipment. Instead of collimation, the virtual source concept and focusing techniques can be used to increase the intensities from monochromator and analyzer. While collimation improves the **Q** resolution on cost of intensity, the focused mode (curving monochomator/analyzer) increases the intensity drastically on cost of **Q** resolution, retaining the energy resolution.

TAS measurements are limited to the 2D scattering plane, for measurements of the third direction the sample needs to be re-mounted. The second disadvantages of the method is the fact of point-to-point measurement. Its gain is the high efficiency of the single-point measuring, and the flexible, if needed energy and momentum high resolution.

Depending on the stiffness of the dispersion $(\mathbf{Q},\omega)$, it is useful to perform either const.-energy or const.q-scans (see Fig. 5). In the first case an energy transfer is fixed and a measurement of momentum transfer is performed (Fig.5a). On the other hand, the constant **Q**-scan can be conducted by fixing the momentum transfer: the scattering intensity is measured as a function of energy (Fig.5b).
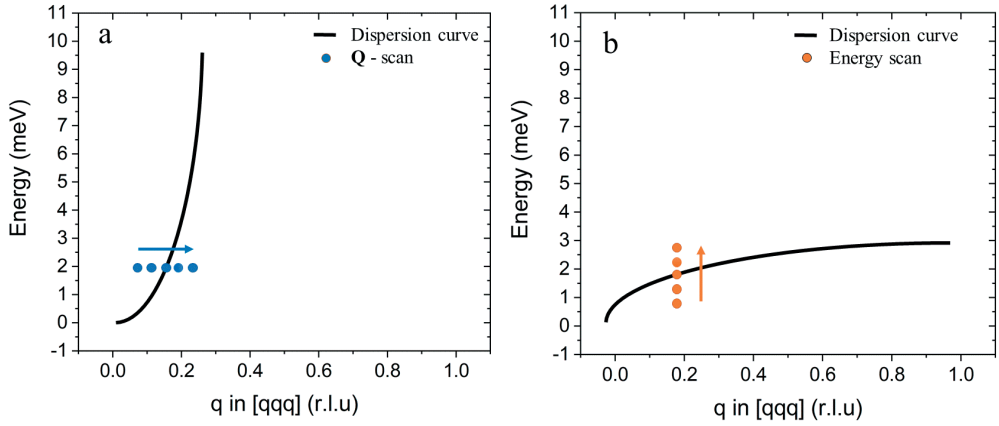

Fig. 5:   Q, E scans of dispersion curve.

The energy and momentum transfer scans experiments are common for lattice vibration and magnetic excitation investigations. Measuring with fixed final energy does not require any correction for the variation of the resolution and the analyzer reflectivity with final energy and scattering angle. Even if the incident energy is changing during the experiment – which may request additional effort to consider special constraints – the experimental range is large and the data can be interpreted directly.

The precision of a three-axes spectrometer is determined by its **resolution**. The analytical explanation of resolution function can be found in refs. [4,5,6]. The resolution of incident and final wave vector depends on the crystal properties of monochromator and analyzer in the following way:

$$\frac{\Delta \chi}{\chi} = \frac{\Delta d_{hkl}}{d_{hkl}} + cot\theta \Delta \theta .$$ (6)

The beam resolution after scattering from the crystal is thoroughly visualized on the Fig. 6.

We can imagine an experiment with not collimated beam scattering from the perfect crystal (Fig.6a). As a result, the neutron beam will spread with Δk^{col} in perpendicular direction to the Bragg reflection wave vector d_{hkl}. The Fig. 6b represents the parallel beam interaction with a non-ideal crystal. The mosaicity of the sample leads to a spread of the scattered wave vector Δk^{mos} in the direction parallel to the scattering vector. In the reality we have a combination of non- perfectly collimated beam and crystals with mosaicity (Fig. 6c). This case will provide to extent of the spread of wave vector to ellipsoidal shape. The idea of explanation was taken from [7, 8].
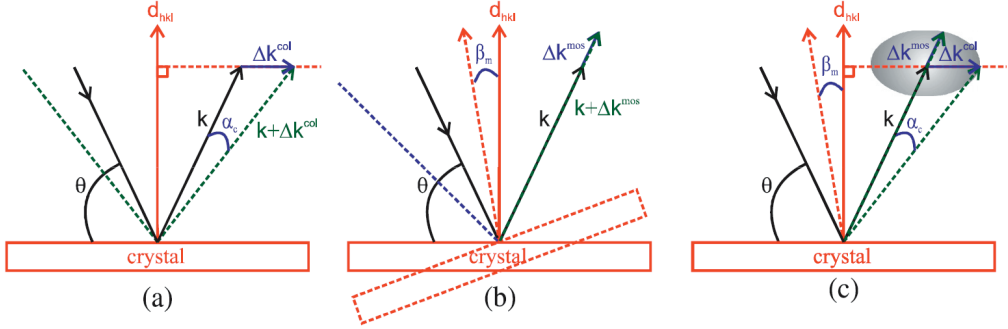
Fig. 6: Illustration of the effect of mosaicity and beam divergence on the resolution. (figure taken from ref. [7]): a) the spread in wave vector caused by a diverging beam and crystal with no mosaicity; b) perfect beam collimation, spread of beam caused only by crystal with non-zero mosaicity; c) the combined effect of not perfect beam collimation and crystal mosaicity. The α_c is collimator acceptance and β_m divergence of monochromator mosaicity.

The total spread in wavevector can be calculated as $\Delta k = \Delta k^{mos} + \Delta k^{col}$. Δk^{mos} and Δk^{col} are the spread of k from the mosaicity of the crystal and from divergence of the beam. The spread of wave vectors can be calculated by following formulas for amplitudes of incident wave vector perpendicular to Q (7) and along k (8) [8].

$$|\Delta k^{mos}| = \beta_m k \cot\theta \tag{7}$$

$$|\Delta k^{col}| = \alpha_c k \operatorname{cosec}\theta \tag{8}$$

The divergence of incident neutrons which are going through a neutron guide to the monochromator depends on the beam collimation. The increasing resolution by using high quality monochromator single crystals decreases the number of neutrons which are scattered by the monochromator. To increase the number of scattered neutrons, crystal with a larger mosaicity can be used.

Sometimes, high resolution is needed for studying sharp peaks. To achieve this, collimator can be introduced between all axis of TAS instrument. The price of high resolution is low intensity. Depending on the sample properties, it is important to choose a right combination of above-mentioned instrument parameters to perform a successful and efficient experiment.

During the investigation of dispersion curves the **focusing condition** should be used for increasing intensity of scattered neutrons from the sample. The visual explanation focusing and **defocusing condition** is shown in Fig. 7. The resolution has as usual ellipsoidal shape according to relation $\Delta k^{mos}/\Delta k^{col}$ equal to about 0.2 to 0.4.
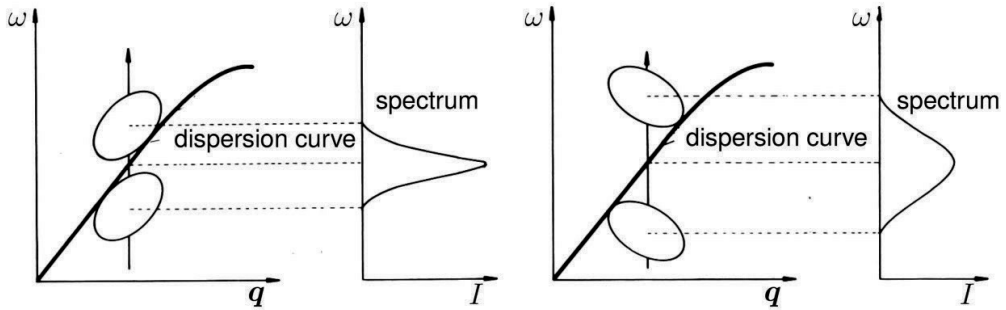

Fig. 7:   Focused and unfocused measurement (figure taken from ref. [9]).

The focusing condition strongly depends on the elongation of resolution ellipsoid. The widths of the observed peak also changes with the orientation of the ellipsoid regarding the slope of the dispersion curve. In the focused condition, the main axis of the resolution ellipsoid should be parallel to a dispersion curve and a sharper peak with higher intensity will be observed. In the other case, the peak will be broader with smaller intensity.

3.2   Multiplexing TAS

In the previous section, there are shown the main advantages of three-axes spectrometers such as very broad range of momentum and energy transfer, optimization of resolution according to experiment request. Nevertheless, by one instrument configuration we measure only one experimental "point" in ($\mathbf{Q}$, ω). Extended dispersions, possible continua and overview studies would profit from wider scans with many points measured at the same time. To improve the efficiency of mapping on the basis of the three-axes spectrometer, multiplexing analyzer-detector systems are constructed. In contrast to the classical analyzer presented on the Fig. 4 the multiplexing construction contains several channels which conduct a measurement at different 2θ angle or/and energies at the same time. The first example, Flatcone at ILL, has been successfully used for a long time already. It consists of one analyzer at fixed final energy for cold and thermal neutrons each for every scattering channel [10,11]. A construction of a tunable analyzer array covering a wide angular scattering range with the possibility to change the final energy is performed on UFO (JCNS at ILL) and the PUMA (MLZ). Both options will be in user operation soon. Another type of inelastic neutron multiplexed analyzer-detector system is realized for Multi-FLEXX (HZB), CAMEA (PSI) and the BAMBUS option for PANDA (MLZ) - see Fig. 8. The concept follows the idea to scatter out of plane at several consecutive positioned analyzers, using the transmitted neutron intensities.
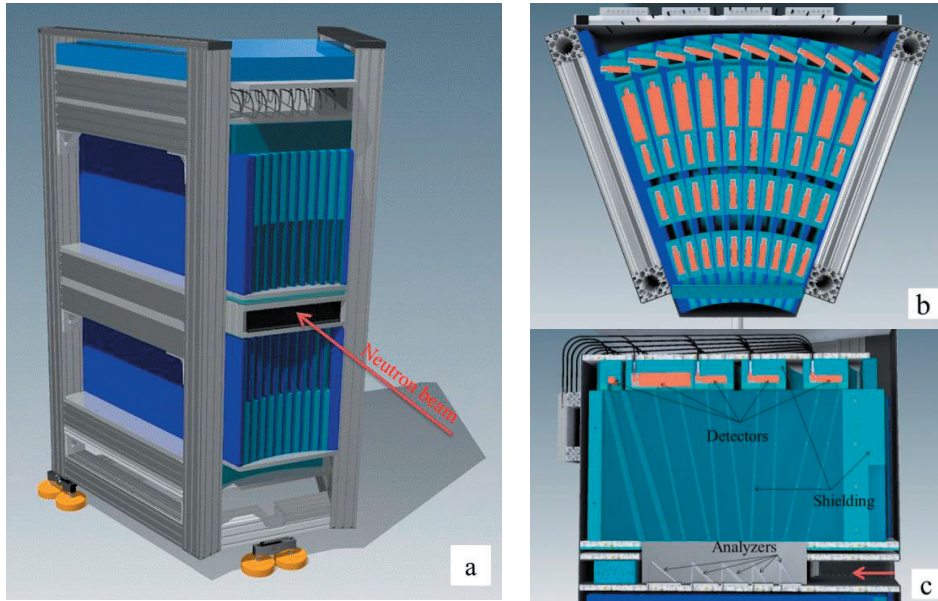
Fig. 8: The prototype of BAMBUS: a new inelastic neutron multiplexed analyzer for cold neutrons three-axes spectrometer PANDA at MLZ. a) Model of BAMBUS b) Construction of the cassettes (top view) c) Construction of the cassette (view from the side).

By one "shot" one can measure broad *2θ* range with few different energy transfers, see Fig. 9.
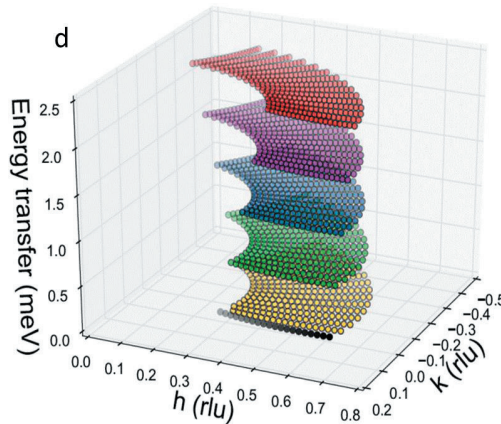

Fig. 9: The example of scan simulation using BAMBUS multiplexed analyzer for cold neutrons three-axes spectrometer PANDA at MLZ.

The complementary options are developed due to different scientific cases (PUMA – for time dependent and stroboscopic measurements) and by experimental requests (CAMEA – single

instrument, BAMBUS – to be exchanged against classical analyzer within less than 1h to continue with better resolution and higher scatter intensity for precise measurements). It is expected that multiplexing TAS will improve the efficient use of TAS measurements.

3.3 Time-of-flight spectroscopy

The first energy-dependent neutron investigations the time-of-flight technique (TOF) was used about 84 years ago [12].The name TOF is given because of measuring time of scattered neutron flight from the sample to detector when the distance between them is well known. Let us consider the main physical basis before explaining the working scheme of the instrument. The typical energy of excitations like magnons and phonons is in the energy range from few μeV to 10^3 meV. The classical experimental speed of neutrons is about $v_n = 395.6$ m/s ($\lambda = 10$ Å) and $v_n = 3956$ m/s ($\lambda = 1$Å) fits properly to these excitations: mass m_n and velocity v_n of neutrons allow to calculate the energy by:

$$E_n = \frac{1}{2} m_n v_n^2. \tag{9}$$

Respectively to wavelengths 10 Å and 1 Å, the energies of neutrons are 0.818 meV and 81.8 meV. These physical properties are used in time-of-flight spectrometer constructions, for an example see Fig.11. By the use of choppers on one hand a time structure of the neutron can be created and on the other hand functionalities like suppression of higher orders, frame overlap and background can be realized. Different kinds of choppers are used, the distance between the choppers is known and the speed of neutrons can be tuned by changing the disc rotation (Fig. 10).
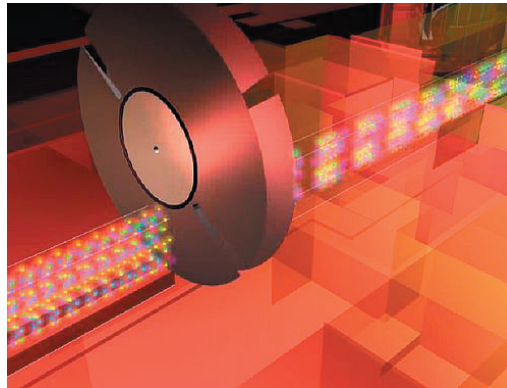

Fig. 10: The example of disc chopper IN5B at the ILL (figure taken from ref. [13]).

In the example (In4C – ILL, Fig.11), the white beam is partially monochromatised by performing counter-rotation of disc choppers. The resulting „neutron packet" is scattered by the monochromator. By setting the incident angle of the monochromator according to the Bragg law a precise neutron energy selection is performed. The scattered monochromatic beam passes through the Fermi chopper, for better time resolution. The short parallel monochromatic neutron pulses interact with the sample. Finally, the neutron energy loss or gain is analysed by recording the arrival time of the neutrons at the detector, in addition to their scattering 2θ angle.

Using a large detector array, multiple information about structure and dynamics of the sample can be observed.

The advantage of TOF instrument is to measure the complete energy and momentum transfers using large solid angle detectors covering. On the other hand we can use neutrons during short burst time of about 20 μs and wait until all neutrons reach the detector to record the full energy spectrum. Only when all these processes are completed the next pulse can be used. As a result, during the TOF instrument experiment at continuous source we are using only 1-2% of the neutron beam.

Spallation sources already provide a short pulse structure of the neutron beam and therefore they are well suited for TOF spectroscopy. Intrinsic limitations occur due to the limited solid angle $d\Omega$ by scattering into the detectors, and due to the **Q** resolution because of resolution limits of detection. The time-of-flight method is a progressive technique for receiving broad information about **Q** and E in reciprocal space. In other words, it is perfect for mapping. Nevertheless, recent developments e.g. at IN5 (ILL), NEAT (HZB) or LET (ISIS) allow detailed high q-resolution measurements with TOF technique by the use of position-sensitive detectors. New instruments – like TOPAS at MLZ or BIFROST at ESS – will allow polarization analysis, too.
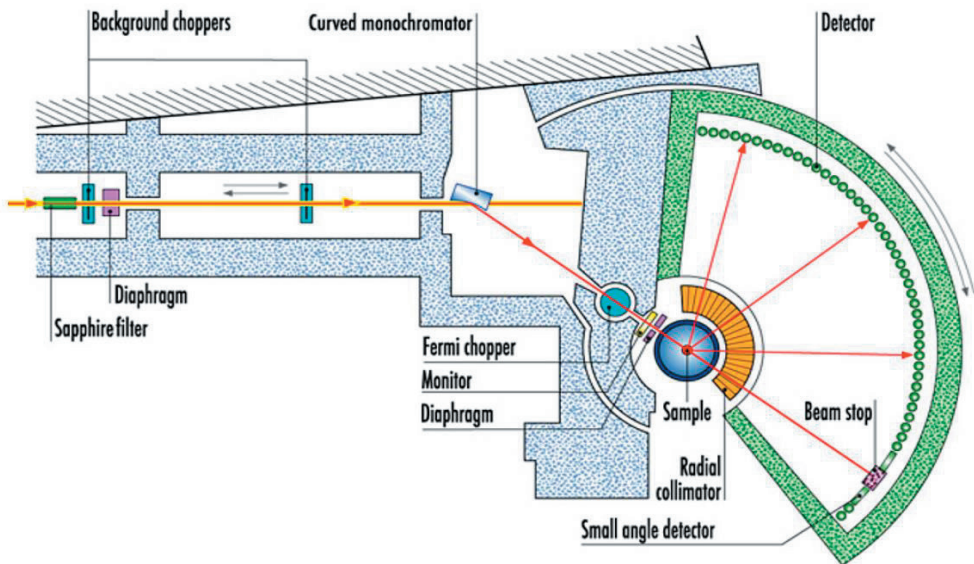

Fig. 11: The high-flux time-of-flight spectrometer IN4C at the ILL (figure taken from ref. [14]).

4 Response

Usually, the goal of inelastic scattering is measurement of neutron momentum and energy exchange with the sample. The data about dynamic properties of a system can be studied from dispersion curve investigations. The scattered beam including information about the sample is

described by the double differential cross-section. The Fermi golden rule [15,16,17] is presented below:

$$\frac{d^2\sigma}{d\Omega dE'} = \left(\frac{m}{2\pi\hbar^2}\right)^2 \frac{k'}{k} \sum_{n_0} p(n_0) \sum_{n_1} \left|\left\langle k'\sigma'n_1 |V| k\sigma n_0 \right\rangle\right|^2 \delta(E_1 - E_0 - \hbar\omega) \tag{10}$$

where E_s is the energy of sample, n_0 and n_1 are the initial and final state of the system of energy E_0 and E_1 respectivly, V is potential of interaction (neutron-unpaired electron, neutron-nuclear spin interactions, neutron-nucleus interaction, etc). The factor k'/k shows that the cross-section is proportional to the ratio of the outgoing and incoming neutron beam. The first sum goes through the all initial states of the system with probability $p(n_0) = [(e^{-E_0 n_0/k_b T})/(\Sigma_{n_0} e^{-E_0 n_0/k_b T})]$ at a temperature T.

According to the literature [18] the eq. (10) can be presented for unpolarized neutrons:

$$\frac{d^2\sigma}{d\Omega d\omega} = A(\mathbf{Q})S(\mathbf{Q},\omega). \tag{11}$$

In the formula (11) the factor $A(\mathbf{Q})$ represents the properties of scattered particles and also the interaction potential with the sample. The information about dynamics properties of the sample is contained in the scattering function $S(\mathbf{Q},\omega)$. It does not depend on actual incident and final wave vector values. The scattering function includes information about positions and motions of atoms or spins in the sample. According to the dissipation-fluctuation theorem, it can be presented by eq. (12):

$$S(\mathbf{Q},\omega) = \frac{\chi''(\mathbf{Q},\omega)}{1-\exp(-\hbar\omega)/(k_B T)}. \tag{12}$$

The $S(\mathbf{Q},\omega)$ is proportional to imaginary part of the dynamical susceptibility.

The experimental intensity is convolution product of instrument resolution function and scattering function in $(\mathbf{Q}, E)$ space:

$$\mathrm{I}(\mathbf{Q},\omega) \propto \iint \mathrm{R}((\mathbf{Q}'-\mathbf{Q}),(\omega'-\omega))\mathrm{S}(\mathbf{Q},\omega)\mathrm{d}\mathbf{Q}'\mathrm{d}\omega'. \tag{13}$$

The influence of resolution on experimental results is discussed in the section 3.1. The resolution function $R(\mathbf{Q},\omega)$ depends on a probability of distribution for the initial and final wave vectors:

$$R(\mathbf{Q},\omega) = \iint p(\mathbf{k})p(\mathbf{k}')\delta(\mathbf{Q}'-(\mathbf{k}'-\mathbf{k}))\delta(\omega'-\frac{\hbar}{2m}(k^2-k'^2))dkdk'. \tag{14}$$

According to the experiment expectations, the probability of distribution can be changed before (collimator mounted in the incident beam) and after (the collimator located between the sample and detector) the sample using beam collimation.

4.1   Single phonon scattering

This section will show the basic formulas for the description of neutron scattering with coherent single phonons. The double differential cross section will be used for its analysis. It is

containing more complete information [19] about the scattered system than the dispersion relation connected with delta function eq.(19).

$$\left(\frac{d^2\sigma}{d\Omega d\omega}\right)^{\pm}_{coh} = \frac{k'}{k}\frac{(2\pi)^3}{\upsilon_0}\sum_{\tau}\sum_{j,q} \tag{15}$$

$$\cdot\left|\sum_i \frac{b_i}{\sqrt{m_i}}\exp\left[-W_i(\mathbf{Q})+i\,\mathbf{Q}\cdot\mathbf{R}_i\right]\left(\mathbf{Q}\cdot\mathbf{e}_i^j\right)\right|^2 \tag{16}$$

$$\cdot\omega_j^{-1} \tag{17}$$

$$\cdot\left(n\left(\omega_J(\mathbf{q})\right)+\frac{1}{2}\pm\frac{1}{2}\right) \tag{18}$$

$$\delta\left(\omega\mp\omega_j(\boldsymbol{q})\right)\delta(\boldsymbol{Q}\mp\boldsymbol{q}-\boldsymbol{\tau}). \tag{19}$$

The + sign describes a phonon creation and – sign shows annihilation.

Let us consider all parts of the double differential cross section equation. The dynamic structure factor is described by the equation (16). It is proportional to nuclear coherent scattering length b_i. The formula contains the Debye-Waller factor for the displacement $\boldsymbol{u}$ of the atoms away from their equilibrium position, $W = <\mathbf{Q}\cdot\boldsymbol{u}>^2$. The scattered intensity is proportional to $|\mathbf{Q}\cdot\boldsymbol{e}^j_i|^2$ where $\boldsymbol{e}^j_i$ is the polarization vector of a phonon mode. It reflects the reason to favor phonon measurements at large $\mathbf{Q}$. In contrast to lattice vibrations, the magnetic excitation is more preferable to investigate with small momentum transfer (see below – 4.2). This helps us understand what kind of excitation we get during this experiment. It is possible to check the type of inelastic excitation using measurements in different Brillouin zones. The intensity of phonon modes should increase with increasing $\mathbf{Q}$ and disappear for magnetic excitations. It is important to mark that the contribution to the structure factor is only provided by the parallel displacement component to the scattering vector $\mathbf{Q}$ according to scalar product in $|\mathbf{Q}\cdot\boldsymbol{e}^j_i|^2$. The v_0 is the volume of the unit cell.

Eq. (17) shows the inverse proportionality between the scattered intensity and the phonon energy. This points to the difficulty of measuring phonons at high energy transfer.

In case of inelastic scattering the neutrons interact with phonons with the possibilities to either "loose" or "gain" energy, i.e. to create or annihilate the phonons. Moreover, it is worth mentioning that phonons behave like bosons. The probability of the temperature dependence of energy level occupation can be expressed by population factor, also known as a Bose factor. Eq. (18) shows the occupation of phonons and in our case $n(\omega_j(\boldsymbol{q}))$ can be expressed in the following way:

$$n(\omega_j(\mathbf{q})) = \left[\exp(\hbar\omega_j(\mathbf{q})/k_BT)-1\right]^{-1}. \tag{20}$$

The incoherent excitation also can play an important role during the experiment. It describes a situation when there is no interference between the scattered neutron waves from the different atoms in the sample. Using the incoherent inelastic scattering we can study the single atom dynamics in contrast to coherent one which represents the information about collective motion

of atoms and their positions. The incoherent scattering can be described by the following incoherent double differential cross-section [19]:

$$\left(\frac{d^2\sigma}{d\Omega d\omega}\right)_{inc}^{\pm} = \frac{\mathbf{k}'}{\mathbf{k}}\frac{(2\pi)^3}{\upsilon_0} \tag{21}$$

$$\sum_i \frac{1}{2M_i}\left(\overline{b_i^2} - \overline{b_i}^2\right)\exp\left(-2W_i(\mathbf{Q})\right) \tag{22}$$

$$\times\sum_j \frac{\left|\mathbf{Q}\cdot\mathbf{e}_{ij}\right|^2}{\omega_j}\cdot\left(n(\omega_j(\mathbf{q})) + \frac{1}{2}\pm\frac{1}{2}\right)\delta(\omega\mp\omega_j) \tag{23}$$

where $b_{inc\,i}^2 \equiv \overline{b^2}_i - \overline{b}_i^{\,2}$ is the nuclear incoherent scattering length, where i labels an atom and j phonon branch.

4.2 Magnetic excitations

The magnetic interaction is possible to investigate due to the neutron property of having a magnetic moment (S=1/2) which interacts with the orbital and spin angular moments of unpaired electrons via the dipole-dipole interaction. As the result, the inelastic neutron scattering is a common method for spin excitation studies.

This section describes the double differential cross-section for single magnon scattering. The formula is presented in a comparable way to single phonon scattering expressions:

$$\left(\frac{d^2\sigma}{d\Omega d\omega}\right)_{mag}^{\pm} = \frac{\mathbf{k}'}{\mathbf{k}}\frac{(2\pi)^3}{\upsilon_0} \tag{24}$$

$$\frac{1}{2}S\left(\frac{\gamma r_0}{2}gF(\mathbf{Q})\right)^2 \tag{25}$$

$$\exp\left(-2W(\mathbf{Q})\right) \tag{26}$$

$$\left(1+\hat{\mathbf{Q}}_z^{\,2}\right) \tag{27}$$

$$\sum_{\tau,q}\left(n(\omega(\mathbf{q})) + \frac{1}{2}\pm\frac{1}{2}\right) \tag{28}$$

$$\delta\left(\omega\mp\omega(\mathbf{q})\right)\delta\left(\mathbf{Q}\mp\mathbf{q}-\boldsymbol{\tau}\right). \tag{29}$$

The expression (25) provides the strength of the dipolar neutron-electron interaction and has the same role as the scattering length b for neutron-nuclear scattering [20].Nevertheless, b is constant due to strong interaction between two point-like objects, the magnetic form factor is a consequence of the spacial distribution of the electron system interacting with neutron spin. Where γ is the gyromagnetic factor of the neutron, r_o is the classical electron radius, the Landé factor and magnetic form factor are presented by g and $F(\mathbf{Q})$, respectively. In other words, the magnetic form factor it is the Fourier transform of the unpaired electron distribution around the nuclei. The Debye-Waller factor (26) describes the movement of the nuclei in the crystal lattice. The eq.(27) shows that only magnetization (or fluctuations) which are perpendicular to the wave

vector transfer **Q** are observed. They can be expressed by: $\mathbf{M}_\perp = \hat{\boldsymbol{Q}} \times (\mathbf{M} \times \boldsymbol{Q})$, were $\hat{\boldsymbol{Q}}$ is unit vector. The eq. (27) depict one magnon scattering in the quantization direction z.

Let us consider the scattering function or dynamic structure factor which describes the correlation in space and time of the magnetic moments in the system. The connection of $S_{mag}(Q,\omega)$ with dynamic magnetic susceptibility can be expressed by eq.:

$$S_{mag}(\mathbf{Q},\omega) = \left(\frac{1}{2}gF(\mathbf{Q})\right)^2 \exp(-2W(\mathbf{Q}))\frac{1}{\pi}\frac{1}{1-\exp\left(\dfrac{-\hbar\omega}{k_B T}\right)} \cdot \sum_{\alpha\beta}\left(\delta_{\alpha\beta} - \hat{Q}_\alpha \hat{Q}_\beta\right)\chi''_{\alpha\beta}(\mathbf{Q},\omega) \tag{30}$$

where the sum over the Cartesian coordinates α and β shows that only components perpendicular to momentum transfer **Q** can be observed.

Using the following expression, the connection of total momentum of the electron in the paramagnetic system with scattering function is shown by [17]:

$$S_{mag}(\mathbf{Q},\omega) = \left(\frac{1}{2}gF(\mathbf{Q})\right)^2 \frac{2}{3}\mathbf{J}(\mathbf{J}+1) \tag{31}$$

where the total moment is presented as $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$.

5 Conclusion

The interaction between neutrons and solid state materials contains unique information about structure and dynamic properties of atoms as well as magnetic properties of materials such as magnetic excitations, magnetization density, and magnetic order. The inelastic neutron scattering instruments presented – special techniques like neutron spin echo or backscattering are not discussed here in detail – provide opportunities to make these information achievable. The measurement should start with some strategy of instruments application to reach the highest quality results in a short time. Often the classical complex instruments cooperation will be a starting time-of-flight mapping followed by the investigation of special dispersion curves by three-axes spectrometers. Latest developments on TOF instruments increase the energy and **Q**-resolution, in contrast advanced modelling allows TAS experiments without prior TOF studies. Nevertheless, the basic physical properties of the sample such as precise structure determination, magnetic properties, resistivity, specific heat etc. need to be studied below and above the phase transition temperature/s beforehand inelastic scattering investigations

To summarize the three-axes spectrometer is perfect for dispersive excitation investigations for single crystals sample due to **Q** and E measurements. The instrument can only perform one point per measure with high precision fraction of momentum transfer and receive important information about atoms position. One of the strong advantages of TAS is the possibility to polarize both the incoming and outgoing beams to and from the sample. Additionally, high neutron flux and capability to use the multianalyzer-multodetector systems make a three-axes spectrometer the powerful tool for inelastic scattering investigation.

In contrast, the time-of-flying spectrometer is very useful for isotropic sample study and it is not always suitable for single crystal experiments. Nevertheless, TOF instrument is ideal to

make a broad overview of dispersive excitation in **Q**, E space with high energy resolution. It is an irreplaceable equipment for mapping especially after proving flux intensity by using short pulse spallation source. The combination of TOF and TAS experiments is perfect for receiving the complete and high precision results.

References

[1] J. W. Lynn, H. G. Smith, and R. M. Nicklow, Phys.Rev. B, **8**, 3493 (1973).

[2] https://www.mlz-garching.de/toftof.

[3] https://mlz-garching.de/instrumente-und-labore/spektroskopie/panda.html.

[4] M.J.Cooper and R. Nathans. Act. Cryst., **23**, 357 (1967).

[5] M. Popovici. Act. Cryst.A, **31**, 507, (1975).

[6] G. Eckold and O. Sobolev, Nuclear Instr. Methods A 752 54 (2014).

[7] Z. Yamani, Z. Tun, D.H. Ryan , NRC Research, 771 (2010).

[8] G. Dolling, *In Dynamical properties of solids, Edited by* G.K. Horton and A.A. Maradudin. Vol. 1, p. 541. North-Holland Publishers, Amsterdam. 1974; C. Stassis in Ref. 9, Vol.A, p. 369.

[9] PANDA - Three Axis Spectrometer P. Link A. Schneidewind and P. Cermak in "Laboratory Course Neutron Scattering Experiment Manual" T. Brückel S. Förster and G. Roth (eds.) Key Technologies, Volume 148, ISBN 978-3-95806-244-3 (2017).

[10] https://www.ill.eu/users/instruments/instruments-list/flatcone/description/instrument-layout/.

[11] M. Kempa, B. Janousová, J. Saroun, P. Flores, M. Boehm, F. Demmel, J. Kulda, Physica B, **385,** 1080 (2006).

[12] J. R. Dunning, G. B. Pegram, G. A. Fink, D. P. Mitchell, Phys.Rev. , **48,** 265 (1935).

[13] https://www.ill.eu/fileadmin/user_upload/ILL/3_Users/Instruments/Instruments_list/00_-_SPECTROSCOPY/00_-_TIME_OF_FLIGHT/IN5/html5/IN5B-principle/IN5.html.

[14] https://www.ill.eu/users/instruments/instruments-list/in4c/description/instrument-layout/.

[15] G.L. Squires, *Introduction to the theory of thermal neutron scattering*, (Cambridge University Press,1984).

[16] S.W. Lovesey, *Theory of neutron scattering from condensed matter*, (Clarendon Press, Oxford, 1984).

[17] D.L.Price & K. Sköld, *Neutron scattering*, K. Sköld & D.L.Price eds, *Methods of Experimental Physics* 23A, (Academic Press, New York, 1986).

[18] Léon Van Hove, Phys. Rev., **95**(1): 249, 1954.

[19] G.L. Squires, *Introduction to the theory of thermal neutron scattering*, (Cambridge University Press, 1978).

[20] F. Hippert, E.Geisler, J.L. Hodeau, E. Lelievre-Berna, J.R. Regnard, *Neutron and X-ray spectroscopy* (Springer, Netherlands, 2006).

D 5   Quasielastic Scattering

Joachim Wuttke

Jülich Centre for Neutron Science

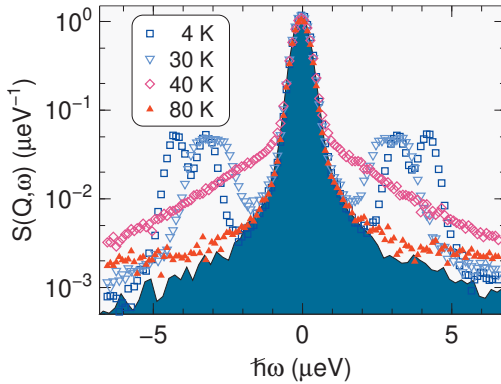Forschungszentrum Jülich GmbH

Contents

Fig. 1: Neutron backscattering spectra of an organic ionic crystal (2:1 picric acid and tetramethylpyrazine) [1], measured on SPHERES. The inelastic features are due to methyl group rotation. With increasing temperature, they soften and broaden and finally merge into a quasielastic central peak, due to the crossover from tunneling to thermally activated jumps. The dark area marks the resolution function, measured using an elastic standard scatterer.

1   Introduction

Quasielastic neutron scattering (QENS) is inelastic scattering, measured with fine energy resolution, and used for the study of motions that are much slower than typical vibration modes. In this chapter, we will discuss different kinds of *slow motion*. But first, we should explicate the term *quasielastic*.

The term *quasielastic scattering* comes from nuclear physics; it designates the limiting case of inelastic scattering, close to elastic scattering, where the energy transfer is much smaller than the incident energy of the scattered particles.[1]

The term *quasielastic peak* (or *line*) designates a spectral distribution that is centered around $\hbar\omega \simeq 0$. It can be seen as a broadened elastic peak, in the same way as a Gaussian or a Lorentzian can be seen as a broadened delta function.

These three criteria (slow motion, small energy transfer, broadened elastic peak) are fuzzy, and only weakly correlated. In practice, whatever can be subsumed under one of them may be designated as QENS. Any stricter terminology would collide with the manyfold borderline and crossover cases of which examples are given in Figs. 1 and 2. In both examples, a meaningful analysis of the quasielastic high-temperature peaks is only possible if their relation to some inelastic low-temperature modes is taken into consideration.

Fig. 1 shows backscattering spectra of an organic crystals. With an energy window $|\hbar\omega| < 7$ μeV and a fixed final neutron energy $E_{\mathrm{f}} = 2.08$ meV, this is clearly quasielastic scattering in the sense $|\hbar\omega| \ll E_{\mathrm{f}}$. All scattering intensity beyond the resolution-broadened elastic line is due to methyl group rotation. At low temperature, there are two pairs of inelastic lines. With increasing temperature, they soften, broaden, and merge into one quasielastic peak, revealing the crossover from quantum tunneling to thermally activated jumps (Sects. 4.3–4.4).

Fig. 2 shows time-of-flight spectra of an ionic solution in the glassy and liquid state. The energy

---

[1] In light scattering, the terminology is particularly confused. The light scattering analogue of quasielastic neutron scattering employs a grating spectrometer or a Fabry-Perot interferometer to analyse the energy of scattered photons; these techniques are commonly called *high-resolution inelastic light scattering* or *Rayleigh-Brillouin scattering*. In contrast, in *quasielastic light scattering*, also called *dynamic light scattering* or (much more to the point) *photon correlation spectroscopy*, scattered photons are counted regardless of their energy; photon counts are then correlated by some real-time circuitry.
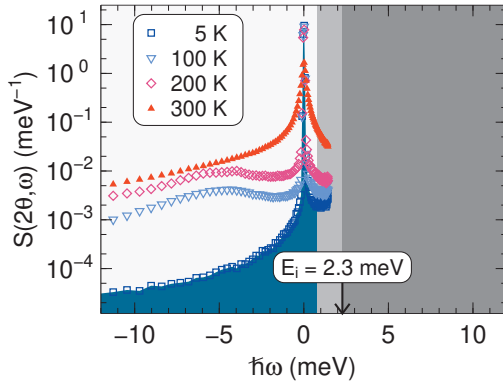
Fig. 2: Neutron time-of-flight spectra of a LiCl:H_2O solution [2], measured on TOFTOF. They dark gray area on the right is kinematically inaccessible. In the light gray stripe with $\hbar\omega \lesssim E_i$, data analysis is not possible because scattered neutrons are so slow that they are overtaken by the next pulse (*frame overlap*). The broad maximum around 5 meV is typical for glasses (*boson peak*); in the liquid state (at 300 K), there is a broad quasielastic peak.

window extends to a multiple of the incident neutron energy $E_i = 2.3$ meV. This is clearly not quasielastic in the sense $|\hbar\omega| \ll E_f$. However, at 300 K there is a quasielastic peak: centered around $\hbar\omega \simeq 0$, and substantially broader than the resolution (measured using the same sample at 5 K). It is due to structural relaxation (Sect. 4.6), and with decreasing temperature becomes so narrow that it is no longer resolved by TOFTOF, leaving only an inelastic phonon spectrum.

Also in the following, all examples will come from time-of-flight or backscattering spectrometers. The spin-echo method, which is special in several respects (implicit Fourier transform, preference for coherent scattering and small Q), shall be left to another chapter (D6).

Much, but not all of the following, is covered in depth in the monographies [3–5]. I have not yet seen the forthcoming book [6].

2 Measuring the double differential cross section

When textbooks say that inelastic neutron scattering measures the double differential cross section $\partial^2\sigma/\partial\Omega/\partial\omega$, then this is a typical pedagogical simplification. Experimentalists need to understand the dependence of recorded neutron counts on the sample properties in more detail so that they can assess possible distortions and apply appropriate corrections. Here we briefly present the two biggest concerns, multiple scattering and instrumental resolution.

2.1 Attenuation and multiple scattering

Scattering from a small sample is described by the double-differential cross section, which gives the number of neutrons per time and per incident flux that are scattered into a solid angle $d\Omega$ and a frequency interval $d\omega$. In real samples, however, some neutrons are scattered more than once. In single crystals, coherent multiple scattering of waves leads to a variety of effects (D3) that are studied by dynamical diffraction theory. In powders and in disordered systems, and also in inelastic scattering from single crystals, these interference effects play no role because the coherence length of scattered radiation is much smaller than the mean free path. Therefore we shall discuss multiple scattering in terms not of wave functions, but of particle currents.

This is the domain of *transport theory*. The neutron distribution in phase space, $f(\mathbf{r}, \mathbf{k}, t)$, obeys

a Boltzmann equation. We request a solution in form of a multiple-scattering expansion $f = \sum_{j=0} f_j$, where f_j is the distribution of neutrons that have been scattered j times. The stationary Boltzmann equation, brought into recursive form, is solved by [7, 8]

$$
\begin{aligned}
f_0(\mathbf{r}, \mathbf{k}) &= f_0(\mathbf{r} - L_{\mathbf{r},\hat{\mathbf{k}}}\hat{\mathbf{k}}, \mathbf{k})\, e^{-\Sigma(k)L_{\mathbf{r},\hat{\mathbf{k}}}}, \\
f_j(\mathbf{r}, \mathbf{k}) &= \int d^3 k'\, \Gamma(\mathbf{k}', \mathbf{k}) \int_0^{L_{\mathbf{r},\hat{\mathbf{k}}}} d\xi\, e^{-\Sigma(k)\xi}\, f_{j-1}(\mathbf{r} - \xi\hat{\mathbf{k}}, \mathbf{k}'),
\end{aligned}
\tag{1}
$$

where $j \geq 1$, $\mathbf{r}$ is inside the sample, $L_{\mathbf{r},\hat{\mathbf{k}}}$ is the distance from $\mathbf{r}$ in direction $\hat{\mathbf{k}}$ to the sample surface, $\Sigma = \Sigma_a + \Sigma_s$ is the probability per unit length for loss by absorption or scattering, and $\Gamma(\mathbf{k}', \mathbf{k})d^3 k$ is the probability per unit length for scattering from wavevector $\mathbf{k}'$ towards an element $d^3 k$ around $\mathbf{k}$, given by the transference function

$$
\Gamma(\mathbf{k}', \mathbf{k}) = \frac{\hbar}{m} \frac{k'}{k^2} \frac{1}{V} \frac{\partial^2 \sigma}{\partial\Omega\partial\omega},
\tag{2}
$$

where V is the volume associated with the double differential cross section.

The single-scattering intensity, given by f_1, is proportional to $\partial^2\sigma/\partial\Omega\partial\omega$. It differs from the thin-sample limit by exponential attenuation factors. For a simple sample geometry and a homogenous incident beam, the ξ integral in (1) can be carried out analytically to yield a single attenuation factor $\mathcal{A}(\Sigma(k_{\text{in}}), \Sigma(k_{\text{out}}))$. The multiple-scattering terms for $j \geq 2$ contain self-convolutions of the transference function and increasingly complicated attenuation factors.

In inelastic scattering, as long as one is only interested in the central frequency, width, and approximate intensity of excitation lines, attenuation and multiple scattering are of little concern. Multiple scattering by dispersionless excitations would generate peaks at overtone and combination frequencies, but in practice the excitation frequency *does* depend on Q. Since multiple scattering involves a convolution in Q, it averages over peaks at different frequencies, and results in a smooth background against which the single-scattering peaks stand out. In contrast, when it comes to the width, intensity, and lineshape of quasielastic peaks, then multiple scattering can cause severe distortions and spurious results [9], and therefore ought to be examined more often than in current QENS practice.

Inverting the dependence of f on the double differential cross section is an ill-posed problem. At best, multiple scattering is treated as a small correction, estimated from sample-specific models. With or without such correction effort, the first and most important measure is using *thin* samples that keep as weak as reasonably possible. Thinner samples require of course longer measuring times, and result in a smaller signal-to-noise ratio. The typical compromise sample scatters about 5 to 10 % of the incident neutrons. For hydrogen-rich materials, this means a sub-mm thickness, which makes powder samples difficult to prepare.

2.2   Instrumental resolution

Every spectrometer has a finite *resolution*. A resolution function is the conditional probability $R(\omega|\omega')$ that a scattering event with energy transfer ω' is registered in the channel ω. Accordingly, a true, 'theoretical' spectrum $S^{\text{th}}(\omega')$ gives rise to an observed, 'experimental' spectrum

$$
S^{\text{ex}}(\omega) = \int d\omega'\, R(\omega|\omega')S^{\text{th}}(\omega').
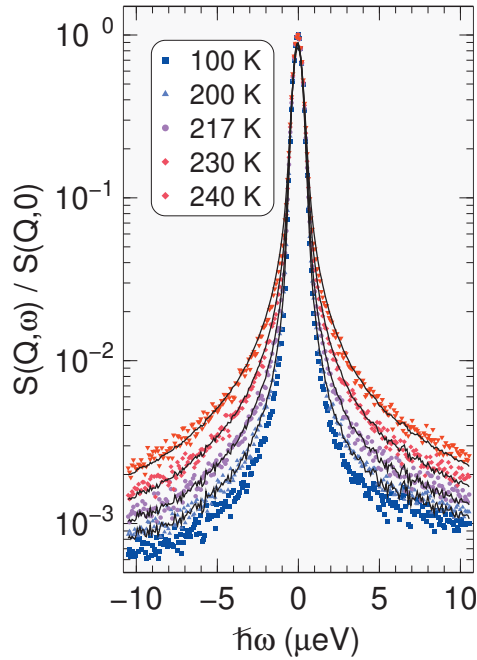\tag{3}
$$

Fig. 3: Spectra of hydration water in deuterated c-phycocyanin protein powder, measured on the backscattering spectrometer SPHERES of JCNS [10]. Solid lines are fit with a Kohlrausch-Williams-Watts function ($\beta = 0.5$, see Eq. (40) below), numerically convolved with the resolution measured at 100 K. As in many other QENS experiments, quasielastic scattering first appears deep in the wings of the resolution function, whereas no broadening can be seen at half maximum.

To make R tractable, one usually assumes $R(\omega|\omega') \simeq R(\omega - \omega')$, which makes of (3) a convolution integral, $S^{\text{ex}} \simeq R \otimes S^{\text{th}}$, and allows for an experimental determination of R by measuring the spectrum of an elastic scatterer.[2]

The resolution functions of time-of-flight and backscattering spectrometers are in a very first approximation Gaussian. Typical scattering functions (Lorentz or Kohlrausch-Williams-Watts functions, see below) are qualitatively different in shape; when scaled for equal maximum and equal width at half maximum, they have much broader wings than a Gaussian. In consequence, the onset of quasielastic scattering is typically detected as additional scattering deep in the wings of the resolution function before any broadening is observed in the width at half maximum (Fig. 3). For this reason, in high-resolution neutron scattering the signal-to-noise ratio is a more important figure of merit than the nominal resolution width.

In principle, resolution effects can be removed from experimental data by Fourier deconvolution:

$$I^{\text{th}}(t) = I^{\text{ex}}(t)/\tilde{R}(t). \tag{4}$$

The number of independent t points is limited by the Nyquist sampling theorem. For most of these t, (4) results in the division of two small, noisy numbers. Therefore one must introduce a cut-off time, restricting I^{th} to a relatively small number of short-time data points. This loss of information is normally not acceptable; instead of deconvoluting experimental data, it is preferable to fit the measured data S^{ex} with a theoretical function S^{th} that has been numerically convolved with the measured resolution R (or a smoothened model thereof). However, explicit

[2] $S^{\text{th}} = \delta$ implies $S^{\text{ex}} = R$. Phonon scattering can be tolerated since it mainly involves energies far outside the resolution peak. Usually, the resolution measurement is done either with vanadium (a perfectly incoherent scatterer so that it can also be used for detector calibration) or with the sample at low temperature (which minimizes variations of sample geometry and environment, and also can be used for intensity normalization).
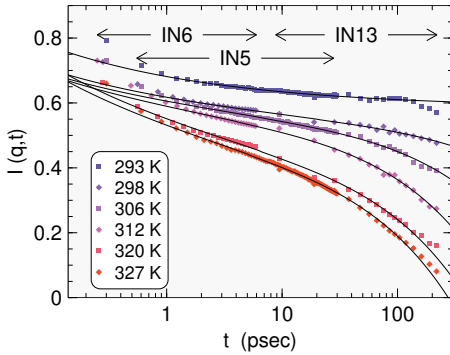
Fig. 4: Intermediate scattering function of the glass-forming liquid ortho-terphenyl, measured on three different spectrometers of the ILL, and combined after Fourier deconvolution [11]. Solid lines are fits with a mode-coupling scaling function.

Fourier deconvolution is attractive for combining spectral measurements from different spectrometers (Fig. 4) or for comparing neutron scattering with molecular dynamics simulations.

3   Interpreting the scattering function

Neutron spectra are usually analysed and presented in form of the scattering function $S(Q, \omega)$, and so for good reasons: $S(Q, \omega)$ depends neither on cross sections nor on the neutron wavelength or any other details of the scattering experiment; it just contains the sample physics the scattering experiment is meant to reveal. Microscopic expressions for $S(Q, \omega)$ or its Fourier transforms $I(Q, t)$ and $G(r, t)$ provide the link from the scattering experiment to microscopic theory or to atomistic simulations or to other information obtained from experiments. This section introduces some basic concepts that are helpful for the interpretation of $S(Q, \omega)$.

3.1   Rebinning from angle to scattering wavenumber

Backscattering or time-of-flight spectrometers have detectors at fixed scattering angles. This facilitates the conversion of raw neutron counts into the double differential cross section, which is a function of 2θ and ω.

For further conversion into the scattering function $S(Q, \omega)$, one needs to replace the dependence on 2θ by one on Q. This is trivial for backscattering where $\hbar|\omega| \ll E_i$ ensures $Q \simeq 2k_i \sin\theta$. But in general, Q is a function of 2θ and ω (Fig. 5). Therefore, the experimental data need to be rebinned. Furthermore, for any given Q, only a restricted ω range is accessible. To avoid these difficulties, some experiments are interpreted in terms of $S(2\theta, \omega)$... and in some papers, spectra are incorrectly labelled $S(Q, \omega)$ although they clearly belong to constant θ.

3.2   The scattering function as an idealization

In first Born approximation, the double differential cross section is given by

$$\frac{\partial^2 \sigma}{\partial\Omega\partial\omega} = \frac{k_{\text{out}}}{k_{\text{in}}} \int \frac{dt}{2\pi} e^{i\omega t} \frac{1}{N} \sum_{j}^{N} \sum_{l}^{N} \left\langle b_j e^{i\mathbf{Q}\mathbf{R}_j(0)} b_l e^{-i\mathbf{Q}\mathbf{R}_l(t)} \right\rangle, \tag{5}$$
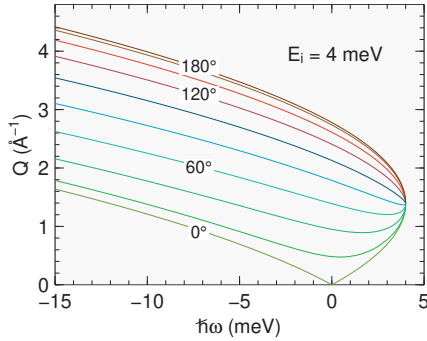
Fig. 5: Dynamic range of a multi-detector time-of-flight spectrometer: At given detector angle 2θ, the scattering wavenumber Q is a function of the energy transfer $\hbar\omega$.

where the scattering lengths b_j, dependent on nuclear spins, and the nuclear positions $\mathbf{R}_j$ are operators, and $\langle\ldots\rangle$ is a quantum-statistical average. Except for nuclear magnetism, spins and positions are uncorrelated so that $\langle b_j b_l\rangle$ can be averaged independently from the positional function $\left\langle e^{i\mathbf{QR}_j(0)}e^{-i\mathbf{QR}_l(t)}\right\rangle$. At this point, to evaluate the averages, one must distinguish whether or not j equals l. This motivates the distinction of *incoherent* and *coherent scattering*.

In textbook-like oversimplification, one would write

$$\frac{\partial^2\sigma}{\partial\Omega\partial\omega} = \frac{k_{\text{out}}}{k_{\text{in}}}\frac{4\pi}{N}\left\{\sigma_{\text{inc}}S_{\text{inc}}(\mathbf{Q},\omega) + \sigma_{\text{coh}}S_{\text{coh}}(\mathbf{Q},\omega)\right\}, \tag{6}$$

with $\sigma_{\text{inc}} = \langle b^2\rangle - \langle b\rangle^2$ and $\sigma_{\text{coh}} = \langle b\rangle^2$, and with *scattering functions* S_{inc}, S_{coh} that only depend on position operators. However, most samples contain more than one chemical element, and atoms belonging to different elements move differently. This can be formalized in two different ways: Either one extends (6) by summing over different element-specific scattering functions, or one keeps (6) intact and redefines S_{inc} and S_{coh} as scattering-length weighted functions. The latter choice is prevalent.

Often, one chemical element dominates the scattering cross section so that all other contributions can be neglected in the interpretation of the scattering-length weighted scattering function. This is especially true for samples that contain hydrogen: The incoherent cross section of the isotope ^{1}H is so big that all other elements can be neglected, and the scattering can be entirely attributed to the self-correlation of the hydrogen nuclei. This is the case for most examples in the present chapter.

Two more simplifications facilitate the interpretation of the scattering function: Most QENS experiments address either disordered materials or ordered matter in powder form. In both cases one can take the *powder average*, so that only a scalar Q dependence is left in the scattering function $S(Q,\omega)$. And the *detailed balance* correction

$$\tilde{S}(Q,\omega) := e^{\hbar\omega/2k_{\text{B}}T}S(Q,\omega) \tag{7}$$

compensates for the different probabilities of up- and downscattering. In the following, we take (7) for granted, and omit the tilde from S.
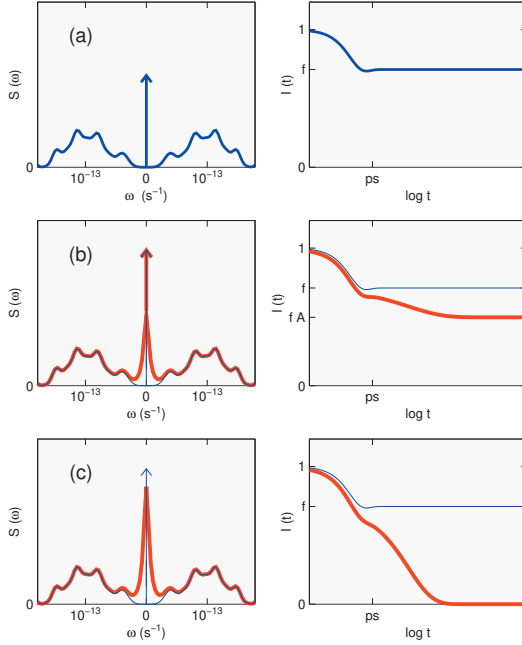
Fig. 6: Hand-drawn $S(Q,\omega)$ and $I(Q,t)$ for some schematic models: (a) An elastic delta line and an inelastic mode distribution. (b) At higher temperatures (red), there is an additional quasileastic component. (c) Here, the central peak of the red spectrum is entirely quasielastic; it has no delta component.

3.3   Frozen, localized, and diffusive dynamics

For brevity, let us specialize to incoherent scattering and drop the subscript 'inc'. For the qualitative physical interpretation of QENS spectra, it is of central importance to switch forth and back between the scattering function

$$S(\mathbf{Q},\omega) = \frac{1}{2\pi} \int dt\, e^{-i\omega t}\, I(\mathbf{Q},t), \tag{8}$$

and its Fourier transform in time, the *intermediate scattering function*

$$I(\mathbf{Q},t) = \frac{1}{N} \sum_j \left\langle e^{-i\mathbf{Q}\hat{\mathbf{r}}_j(0)} e^{i\mathbf{Q}\hat{\mathbf{r}}_j(t)} \right\rangle, \tag{9}$$

which captures *self correlations* of *tagged particles*.

By construction, I has the initial value $I(\mathbf{Q},0) = 1$, from which we obtain the *sum rule*

$$\int_{-\infty}^{+\infty} d\omega\, S(Q,\omega) = 1. \tag{10}$$

No such rule holds for $S(2\theta,\omega)$ or for $\partial^2\sigma/\partial\Omega/\partial\omega$. Also note that experiments only cover restricted ω ranges (Fig. 5). Therefore, integration of experimental scattering functions will usually yield less than 1, except in absence inelastic scattering. This is the rationale for the *normalization* of experimental spectra to a low-temperature measurement.

Fig. 6 shows Fourier transform pairs $S(Q,\omega)$, $I(Q,t)$ for three idealized situations. In (a), there is a phonon spectrum but no quasielastic scattering. Accordingly, the central peak at $\omega = 0$ is a

delta line. The Fourier transform of a delta function is a constant. Therefore $I(Q,t)$ decays from its initial value 1 towards a long-time limit $f_Q > 0$. This f_Q is just the amplitude of the delta component of $S(Q,\omega)$, and is called the *Debye-Waller factor*, or, specifically for incoherent scattering, the *Lamb-Mössbauer factor*.

In (b), there is *some* quasielastic scattering, but on top of it there still an elastic delta line of amplitude $f_Q A_Q$. In the Fourier transform, scattering at small $|\omega|$ corresponds to relaxation on long time scales, therefore $I(Q,t)$ decays in two steps that can be more or less pronounced. Since correlations remain finite, the second step is ascribed to *localized motion*. The factor A_Q is called the *elastic incoherent structure factor* (EISF).

In (c), the entire central peak is quasielastic; there is no delta component. Accordingly, $I(Q,t)$ decays to 0. This implies long-ranged diffusive motion.

To summarize these three cases, we write the scattering function as

$$S(Q,\omega) = f_Q S_{\mathrm{slow}}(Q,\omega) + (1 - f_Q)S_{\mathrm{fast}}(Q,\omega). \tag{11}$$

The fast component is typically the phonon spectrum; it lies almost entirely outside the dynamic range of a backscattering spectrometer. The slow component

$$S_{\mathrm{slow}}(Q,\omega) = A_Q\delta(\omega) + S_{\mathrm{QENS}}(Q,\omega) \tag{12}$$

consists of the elastic delta line, with amplitude given by the EISF, and of a quasielastic spectrum. In case (a), $A_Q = 1$ and $S_{\mathrm{QENS}} = 0$; in case (c), $A_Q = 0$. So only in case (b) all three components are present.

4   Sample physics models

For certain types of samples, there exist idealized models that can be solved analytically, yielding closed expressions for the scattering function. These models include harmonic vibrations, intramolecular rotation, diffusion, and structural relaxation, briefly described in the following subsections. Even where these models do not apply literally they provide an indispensable reference for discussing crossover scenarios, correction terms, or refined models.

4.1   Harmonic vibrations and the mean squared displacement

For harmonic vibrations, S_{fast} and f_Q can be derived in closed form. While the phonon spectrum is out of scope here, the result for f_Q is of the outmost importance for the analysis of QENS data. Vibrations are described in terms of displacements $\mathbf{u}_j(t)$ from equilibrium positions $\mathbf{R}_j$,

$$\mathbf{r}_j(t) = \mathbf{R}_j + \mathbf{u}_j(t). \tag{13}$$

If there are only harmonic forces, then the Bloch theorem [13] reduces (9) to

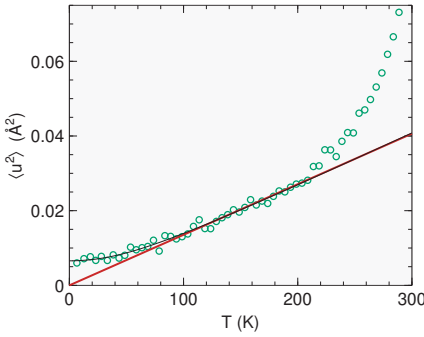$$I(\mathbf{Q},t) = \frac{1}{N}\sum_j \mathrm{e}^{-2W_j(\mathbf{Q},\mathbf{0})}\mathrm{e}^{2W_j(\mathbf{Q},t)} \tag{14}$$

Fig. 7: Temperature dependence of the mean squared displacement in glycerol $C_3H_5(OD)_3$, obtained from elastic thermal neutron backscattering [12]. The black line is a fit with the Debye model. The red line is the linear high-temperature asymptote of that model. Deviations from the fit above the glass transition ($T_g = 185$ K) are due to quasielastic broadening.

with

$$2W_j(\mathbf{Q}, t) := \langle (\mathbf{Q}\mathbf{u}_j(0))(\mathbf{Q}\mathbf{u}_j(t)) \rangle . \tag{15}$$

In isotropic systems, an orientational average gives

$$2W_j(Q, t) = \frac{Q^2}{3} \langle \mathbf{u}_j(0)\mathbf{u}_j(t) \rangle . \tag{16}$$

Each particle partakes in a huge number of oscillatory modes, which quickly run out of phase. Therefore, within little more than one typical phonon period, $2W_j(\mathbf{Q}, t)$ approaches 0.

Assuming that there is just one kind of scatterers,

$$f_Q \equiv I(\mathbf{Q}, t \to \infty) = e^{-Q^2 \langle \mathbf{u}^2 \rangle /3}. \tag{17}$$

So the Lamb-Mössbauer factor reveals the *mean squared displacement* (MSD).[3] Experimentally, one may obtain the MSD from the slope of $-\ln f_Q$ versus Q^2. However, one must expect serious distortions from multiple scattering [9, 15].

Harmonic theory provides a closed expression for the MSD in terms of the *vibrational density of states* $g(\omega)$,

$$\langle u_x^2 \rangle = \frac{\hbar}{6m} \int d\omega \, \frac{g(\omega)}{\omega} \coth \frac{\hbar\omega}{2k_B T}. \tag{18}$$

Specially for the *Debye model*, one finds that the MSD as function of temperature crosses over from a constant plateau, due to zero-point oscillations, to linear growth. At even higher temperatures, deviations from the linearity $\langle x^2 \rangle \propto T$ may then reveal anharmonicity or quasielastic broadening (Fig. 7).

4.2 Two-site jumps

Probably the simplest model of anharmonic localized motion consists of a proton that jumps between two positions $\mathbf{r}_1$ and $\mathbf{r}_2$. In practice, this motion may combine with confined diffusion so that one should rather talk of jumps between two cages. Fig. 8 shows that this actually

---

[3] There is widespread confusion in notation and about factors 2 and 3 [14]: If one refers to just one Cartesian component $x \equiv u_x$ of the displacement, then one may write $\langle x^2 \rangle = \langle \mathbf{u}^2 \rangle /3$. And the relative displacement of two independent scatterers is $\langle \mathbf{r}^2 \rangle = 2 \langle \mathbf{u}^2 \rangle$.
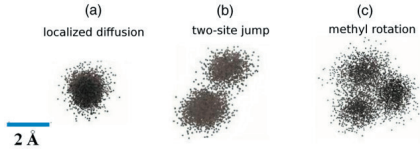
Fig. 8: Scatter plots of time-sampled position of three different hydrogen atoms in Green Fluorescent Protein, from a molecular dynamics simulation. Reproduced with kind permission from [16].

happens for some hydrogen atoms in a protein, while other H atoms are confined to *one* cage, or partake in methyl group rotation, to be discussed in the next subsection.

The probability $p(\mathbf{r}, t)$ of finding the proton at time t at site $\mathbf{r}$ obeys the rate equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} p(\mathbf{r}_1, t) \\ p(\mathbf{r}_2, t) \end{pmatrix} = - \begin{pmatrix} \lambda_1 & -\lambda_2 \\ -\lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} p(\mathbf{r}_1, t) \\ p(\mathbf{r}_2, t) \end{pmatrix} \tag{19}$$

with transition rates λ_n. The matrix has the eigenvalues 0 and $\Gamma := \lambda_1 + \lambda_2$. The rate equation is solved by $p(\mathbf{r}_n, t) = a_n + b_n \exp(-\Gamma t)$. In the limit $t \to \infty$, the occupation ratio must be $p_1/p_2 = \lambda_2/\lambda_1$ to satisfy $\mathrm{d}p/\mathrm{d}t = 0$. Combined with the normalization condition $\sum_n p(\mathbf{r}_n, t) = 1$, we find

$$a_1 = p(\mathbf{r}_1, \infty) = \lambda_2/\Gamma, \quad a_2 = p(\mathbf{r}_2, \infty) = \lambda_1/\Gamma. \tag{20}$$

Making the probabilities conditional upon the initial condition $p(\mathbf{r}_1, 0) = 1$, we compute

$$\begin{aligned} p(\mathbf{r}_1, t | \mathbf{r}_1, 0) &= a_1 + a_2 \exp(-\Gamma t), \\ p(\mathbf{r}_2, t | \mathbf{r}_1, 0) &= a_2 \left(1 - \exp(-\Gamma t)\right), \end{aligned} \tag{21}$$

and similarly for $p(\mathbf{r}_2, 0) = 1$. Using the equilibrium occupation probabilities (20) we obtain the intermediate self correlation function

$$\begin{aligned} I(\mathbf{Q}, t) &= \langle \mathrm{e}^{\mathrm{i}\mathbf{Q}\mathbf{r}(t)} \mathrm{e}^{-\mathrm{i}\mathbf{Q}\mathbf{r}(0)} \rangle \\ &= p(\mathbf{r}_1, \infty) \left[p(\mathbf{r}_1, t | \mathbf{r}_1, 0) + p(\mathbf{r}_2, t | \mathbf{r}_1, 0) \mathrm{e}^{\mathrm{i}\mathbf{Q}\mathbf{d}} \right] + \\ &\quad p(\mathbf{r}_2, \infty) \left[p(\mathbf{r}_2, t | \mathbf{r}_1, 0) + p(\mathbf{r}_1, t | \mathbf{r}_1, 0) \mathrm{e}^{-\mathrm{i}\mathbf{Q}\mathbf{d}} \right] \end{aligned} \tag{22}$$

with the jump vector $\mathbf{d} := \mathbf{r}_2 - \mathbf{r}_1$. Regrouping terms, abbreviating

$$\begin{aligned} A_0(\mathbf{Q}) &:= a_1^2 + a_2^2 + 2a_1 a_2 \cos \mathbf{Q}\mathbf{d}, \\ A_1(\mathbf{Q}) &:= 2a_1 a_2 (1 - \cos \mathbf{Q}\mathbf{d}), \end{aligned} \tag{23}$$

and evaluating the Fourier transform of $\exp(-\Gamma t)$, we get the incoherent scattering function

$$S(\mathbf{Q}, \omega) = A_0(\mathbf{Q})\delta(\omega) + A_1(\mathbf{Q})\mathcal{L}(\omega; \Gamma), \tag{24}$$

which consists of an elastic line and a quasielastic component. The latter is *Lorentzian*,

$$\mathcal{L}(\omega; \Gamma) := \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + \omega^2}. \tag{25}$$

As discussed above, the presence of an elastic line is characteristic for *localized motion*: Since the jumping proton is confined to a finite region in space, its self correlation function never decays to zero.

For powder samples, we average over the orientations of $\mathbf{d}$,

$$\overline{\cos \mathbf{Qd}} = \frac{1}{4\pi} \int_0^\pi \mathrm{d}\vartheta \, 2\pi \sin \vartheta \, \cos(Qd \cos \vartheta) = \frac{\sin Qd}{Qd} = j_0(Qd) \qquad (26)$$

with the spherical Bessel function j_0. In the simplest case, for jumps between two equivalent positions, we have $\lambda_1 = \lambda_2$ and $a_1 = a_2 = 1/2$, so that

$$A_0(Q) = (1 + j_0(Qd))/2,$$
$$A_1(Q) = (1 - j_0(Qd))/2. \qquad (27)$$

4.3   Rotational jump diffusion

Rotation of molecules or molecular sidegroups is a strong and often dominant source of quasi-elastic scattering. As the simplest and most important example, we consider the rotation of a methyl group. In polymers and proteins, this degree of freedom has important effects upon structure and mechanical properties.

We consider the group R–CH$_3$ as stiff (CH bond length $d = 1.097 \pm 0.004$ Å, HCH angle $\theta = 106.5 \pm 1.5°$). The only degree of freedom is the rotation around the R–C bond. The moment of inertia is

$$I = \sum m d_\perp^2 = 2md^2(1 - \cos \theta). \qquad (28)$$

The rotational motion can be described by a wave function ψ that depends on one single coordinate, the rotation angle ϕ. The Schrödinger equation is

$$\left\{ B \frac{\partial^2}{\partial \phi^2} - V(\phi) + E \right\} \psi(\phi) = 0 \qquad (29)$$

with the *rotational constant*

$$B := \frac{\hbar^2}{2I} = 670 \text{ μeV}. \qquad (30)$$

For free rotation ($V = 0$), solutions that possess the requested periodicity are sine and cosine functions of argument $J\phi$, with integer J. Accordingly, the energy levels are $E = BJ^2$.

In condensed matter, however, the potential V caused by the local environment cannot be neglected. Due to the symmetry of the CH$_3$ group, the Fourier expansion of $V(\phi)$ contains only sine and cosine functions with argument $3m\phi$, with integer m. In most applications, it is sufficient to retain only one term,

$$V(\phi) \doteq V_3 \cos(3\phi). \qquad (31)$$

The strength of the potential can then be expressed by the dimensionless number V_3/B. In the following we specialize to the case of a *strong potential*, $V_3/B \gg 10$, which is by far the most frequent one.

In a strong potential of form (31), the CH$_3$ group has three preferential orientations (Fig. 8c), separated by potential walls. The motion of the CH$_3$ group consists mainly of small excursions from the preferred orientations, called *libration*. Quantum-mechanically, they are zero-point oscillations in an approximately harmonic potential. Occasionally though, there are *thermally*
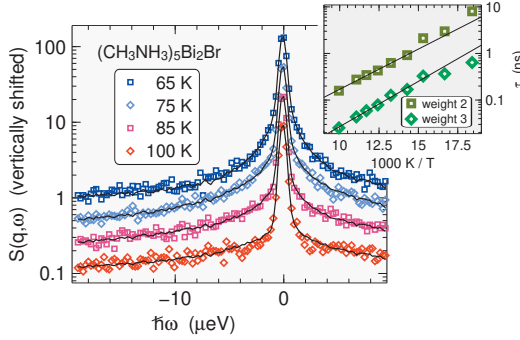
Fig. 9: Backscattering spectra of $(CH_3NH_3)_5Bi_2Br$, measured on SPHERES [17]. The five methylammonium cations fall into two different categories: at room temperature, two of them are ordered, three are disordered [18]. Therefore, we fitted the spectra with two Lorentzians with an amplitude ratio of 2:3. The resulting relaxation times have an Arrhenius temperature dependence, shown in the inset.

activated jumps between the three cages. Modelling this *jump diffusion* requires only a little extension of the two-site jump model introduced above. The transition matrix in the rate equation takes the form

$$\begin{pmatrix} 2\lambda & -\lambda & -\lambda \\ -\lambda & 2\lambda & -\lambda \\ -\lambda & -\lambda & 2\lambda \end{pmatrix}, \tag{32}$$

which has the eigenvalues 0, 3λ, 3λ. Thanks to the degeneracy of the nonzero eigenvalue, the scattering law retains the simple form (24), with $\Gamma = 3\lambda$, and with amplitudes

$$A_0(Q) = (1 + 2j_0(Qr\sqrt{3}))/3,$$
$$A_1(Q) = (2 - 2j_0(Qd\sqrt{3}))/3, \tag{33}$$

where r is the radius of the circle on which the rest positions are located. This model has proven successful in a huge number of experiments; Fig. 9 shows an arbitrarily chosen recent example.

If the rotational potential has a C_2 symmetry, then there are six equivalent equilibrium positions, connected by $60°$ jumps, so that the transition matrix is of rank 6. After some computation it is found to have three different non-zero eigenvalues Γ_μ. In such a situation, the inelastic part of scattering law no longer factorises into a Q dependent and a ω dependent function. Instead, one has a sum of Lorentzians of different widths:

$$S(\mathbf{Q}, \omega) = A_0(\mathbf{Q})\delta(\omega) + \sum_\mu A_\mu(\mathbf{Q})\mathcal{L}(\omega; \Gamma_\mu). \tag{34}$$

This equation holds quite generally for systems described by a rate equation of the form (19) with an arbitrary, symmetric transition matrix. In particular, it holds for rotational jump diffusion of molecules that have more than one axis of rotation [4].

4.4 Rotational tunneling

At low temperatures, almost exclusively the vibrational ground state is occupied. Yet reorientational motion beyond librations is possible by means of quantum mechanical tunneling: The wave functions of the three localised *pocket states* ψ_m ($m = 1, 2, 3$) have nonzero overlap. Therefore, the eigenstates are a linear combination of pocket states.[4] Periodicity and threefold

---

[4] This is an extremely simplified outline of the theory. In a serious treatment, to get all symmetry requirements right, one must also take into account the nuclear spins of the H atoms [3].
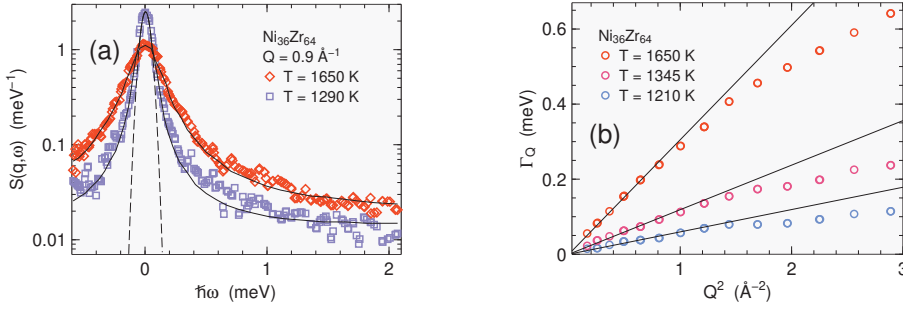
Fig. 10: Atomic dynamics of liquid Zr-Ni, investigated with the time-of-flight spectrometer TOFTOF [19]. (a) Selected spectra with Lorentzian fits (39). The dashed Gaussian with a FWHM of 95 μeV approximates the instrumental resolution. (b) Linewidths Γ_Q, from Lorentzian fits to the spectra. For small Q, scattering is dominated by incoherent contributions from Ni so that the initial slope of Γ_Q vs. Q^2 yields the self-diffusion coefficient of Ni.

symmetry allow three such combinations: a plain additive one

$$\psi_1 + \psi_2 + \psi_3, \tag{35}$$

and two superpositions with phase rotations

$$\psi_1 + \mathrm{e}^{\pm i 2\pi/3}\psi_2 + \mathrm{e}^{\pm i 4\pi/3}\psi_3. \tag{36}$$

In the language of group theory, state (35) has symmetry A, the degenerate states (36) are labelled E^a, E^b. It is found that A is the ground state. The *tunneling splitting* $\hbar\Omega_\mathrm{t}$ between the states A and E is determined by the overlap integral $\langle\psi_m|V|\psi_n\rangle$ $(m \neq n)$, which depends exponentially on the height of the potential wall. Experiments that detect tunneling transitions provide therefore a very sensitive probe of the rotational potential; conversely, if the potential is not accurately known, it is almost impossible to predict whether a tunneling transition will show up in a given experimental energy range.

In neutron scattering, a tunneling transition appears as a pair of inelastic peaks at $\pm\hbar\Omega_\mathrm{t}$. The spectral shape of these peaks is well described by Lorentzians $\mathcal{L}(\omega\pm\Omega_\mathrm{t};\Gamma)$. With rising temperatures, the occupancy of excited vibrational levels increase. This facilitates transitions between A and E sublevels and results in a decrease of $\hbar\Omega_\mathrm{t}$ and an increase of the line width Γ. Upon further temperature increase, thermal motion of neighbouring molecules causes so strong potential fluctuations that the picture of quantum tunneling is no longer applicable. Instead, the motion between different pocket states must be described as thermally activated jump diffusion, as exposed in the previous subsection. An experimental example has already been given in Fig. 1.

4.5   Diffusion

The simplest model for *self-diffusion* is Brownian motion. It can be described as a memory-less *random walk*. Particle trajectories $\mathbf{r}(t)$ are subject to the *Langevin equation*. Equivalently, the *diffusion equation* is applied to the Van Hove self-correlation function, resulting in the solution

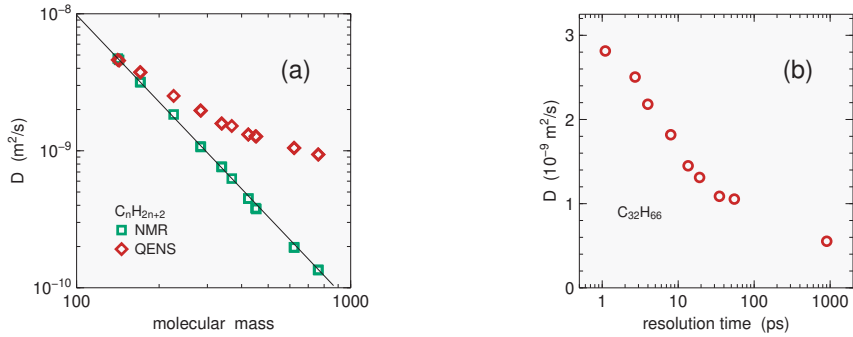$$G_\mathrm{s}(\mathbf{r},t) = (4\pi Dt)^{-3/2}\mathrm{e}^{-r^2/4Dt}. \tag{37}$$

Fig. 11: Hydrogen dynamics in n-alkanes. (a) Diffusion coefficient D, determined by pulsed-field gradient NMR and by neutron scattering. The data agree only for the shortest molecule ($n = 8$). In longer chains, QENS deviates from NMR because it measures not only center-of-mass molecular translation, but also rotation and intramolecular motion [20]. (b) In an exemplary series of TOFTOF measurements with different instrumental resolutions naive fits result in a dependence of the apparent diffusion coefficients on the *resolution time* [21].

By Fourier transform, we find the intermediate scattering function

$$I(Q,t) = \exp(-DQ^2 t) \tag{38}$$

and the scattering function

$$S(Q,\omega) = \mathcal{L}(\omega; DQ^2). \tag{39}$$

On a time-of-flight spectrometer, with experimental scales of the order $Q \sim \text{Å}^{-1}$ and $\hbar\omega \sim 0.1 \ldots 10$ meV, one can resolve diffusion coefficients D of the order $10^{-10} \ldots 10^{-8}$ m²/s.

As anticipated in Sect. 3.3 and Fig. 6, there is no elastic scattering component, in contrast to the localized jump models of Sects. 4.2–4.4. In practice, however, there may be some elastic scattering from the sample container.[5] And in an important class of applications, one investigates diffusion in a solid matrix, e. g. hydrogen diffusion in metals [5]. Therefore, absence of elastic scattering is not a reliable indicator of long-ranged diffusion. The key indicator is rather the pronounced Q dependence of the Lorentzian width $\Gamma = DQ^2$, in contrast to the Q-independent width of the jump models. Conversely, (39) has a fixed amplitude 1, whereas the Lorentzians of the jump models oscillate with Q.

The straightforward determination of D from Lorentzian fits (39) works best in simple atomic systems. Recent examples are provided by metallic melts, which can be studied under very clean experimental conditions using electromagnetic levitation (Fig. 10). Results improve significantly upon macroscopic laboratory measurements that suffer from convective contributions.

In molecular liquids the applicability of (39) is not ascertained a priori because the atomic motion seen by neutron scattering is a superposition of translations, rotations, vibrations and rearrangements. This has been demonstrated very clearly in a systematic study of alkanes C_nH_{2n+2} (Fig.11) [20, 21].

---

[5] It is debatable whether container scattering should be subtracted from the raw data, or taken into account as part of the fit model. Either way of correcting is only approximative because of sample-container multiple scattering.

4.6 Structural relaxation

Atoms or molecules that constitute a *liquid* also undergo long-range diffusion. On short time and length scales, however, their trajectories are *not* memory-less random walks. Particles rather rattle in transient cages, occasionally probe a step out of the cage, and more often than not do the next step in the backward direction. This delays the decay of correlations, and is well described by Kohlrausch's *stretched exponential function*

$$\exp(-(t/\tau)^\beta) \tag{40}$$

that is also known from other experimental probes of *structural relaxation* [22]. Typical values of β are between 0.4 and 0.8.

To use (40) in spectral fits, its Fourier transform, the *Kohlrausch-Williams-Watts function*, must be computed numerically [23]. Fig. 3 shows an application to the motion of supercooled hydration water in a protein powder.

Empirical fit functions like (40) do not capture the short-time dynamics of rattling particles. For this, one needs a microscopic theory, as proposed by the *mode-coupling theory* of structural relaxation [24, 25]. This theory starts from a closed equation of motion for density correlations, and leads to scaling functions like those used in the fits of Fig. 4.

References

[1] G. Bator *et al.*, Chem. Phys. **410**, 55 (2013).

[2] T. Unruh and J. Wuttke, *MLZ Experimental Report 2512* (2009).

[3] W. Press, *Single-particle rotation in molecular crystals* (Springer tracts in modern physics 92), Springer: Berlin (1981).

[4] M. Bée, *Quasielastic Neutron Scattering*, Adam Hilger: Bristol (1988).

[5] R. Hempelmann, *Quasielastic Neutron Scattering and Solid State Diffusion*, Clarendon Press: Oxford (2000).

[6] M. T. F. Telling and V. Garcia Sakai, *A Practical Guide to Quasi-Elastic Neutron Scattering*, Royal Society of Chemistry (2019).

[7] G. H. Vineyard, Phys. Rev. **96**, 93 (1954).

[8] V. P. Sears, Adv. Phys. **24**, 2 (1975).

[9] J. Wuttke, Phys. Rev. E **62**, 6531 (2000).

[10] W. Doster *et al.*, Phys. Rev. Lett. **104**, 098101 (2010).

[11] J. Wuttke *et al.*, Z. Phys. B **91**, 357 (1993).

[12] J. Wuttke, W. Petry, G. Coddens and F. Fujara, Phys. Rev. E **52**, 4026 (1995).

[13] N. W. Mermin, J. Math. Phys. **7**, 1038 (1966).

[14] W. Doster, Eur. Biophys. J. **37**, 591 (2008).

[15] R. Zorn, Nucl. Inst. Meth. A **572**, 874 (2007).

[16] L. Hong *et al.*, Phys. Rev. Lett. **110**, 028104 (2013).

[17] A. Piecha, J. Wuttke, R. Jakubas and G. Bator, *MLZ experimental report 3961* (2010).

[18] J. Matuszewski, R. Jakubas, L. Sobczyk and T. Głowiak, Acta Cryst. C **46**, 1385 (1990).

[19] D. Holland-Moritz *et al.*, Phys. Rev. B **79**, 064204 (2009).

[20] C. Smuda, S. Busch, G. Gemmecker and T. Unruh, J. Chem. Phys. **129**, 014513 (2008).

[21] T. Unruh *et al.*, J. Chem. Phys. **129**, 121106 (2008).

[22] R. Böhmer, K. L. Ngai, C. A. Angell and D. J. Plazek, J. Chem. Phys. **99**, 4201 (1993).

[23] J. Wuttke, Algorithms **5**, 604 (2012).

[24] T. Voigtmann, in *Soft Matter. From Synthetic to Biological Materials*, edited by J. K. G. Dhont *et al.* (Lecture Notes of the 39th Spring School), Forschungszentrum Jülich: Jülich (2008).

[25] W. Götze, *Complex Dynamics of Glass-Forming Liquids. A Mode-Coupling Theory*, Oxford University Press: Oxford (2009).

D 6 High-resolution spectroscopies: spin-echo and speckle spectroscopy

S. Pasini

Instrumentation and Methods
Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

1   Introduction

This chapter is dedicated to the scattering methods in neutron and x-ray spectroscopy that can
extend either the time/energy or the Q-ranges beyond those of techniques like e.g. Time of
Flight (ToF), Backscattering (BS) or dynamic light scattering (DLS). Here we will concentrate
on two relevant techniques: The Neutron Spin-Echo (NSE) and the X- ray photon correlation
spectroscopy (XPCS). NSE and XCPS go under the name of high resolution spectroscopy be-
cause they push the measurable capabilities towards longer time (lower energies) in neutron
scattering, as is the case for NSE, or towards larger Q's in x-ray spectroscopy, as for XCPS.
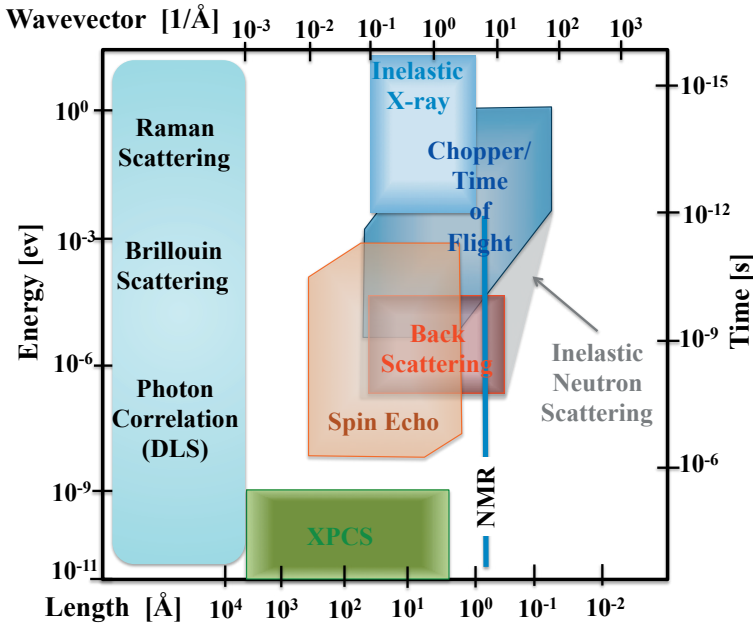
Fig. 1: Qualitative dynamical ranges of energy and time-scale for different spectroscopic tech-
niques vs Q and and length-scale.

Figure 1 provides a comparison between different techniques in spectroscopy and it shows the
domain of applicability of NSE and XPCS and their complementarity. Similarly to dynamic
light scattering, XPCS analyses the intensity correlations of scattered electromagnetic radiation
while enlarging the accessible Q-range of DLS of three order of magnitude. Differently from
DLS, X- ray photon correlation spectroscopy uses the coherence of a X-ray beam while the con-
trast is given by differences in the electron density and not in the index of refraction. Compared
to NSE the time domain of XPCS suffers more on the side of very short correlation times.

Introduced and developed by F. Mezei in 1972 [1], the neutron spin-echo technique makes use
of the direction of the neutron spin as a stopwatch to measure small differences in the neutron
velocity before and after the scattering. The Larmor precession of a spin in a magnetic field is
the pillar concept of all spin-echo techniques among which the NSE combines the polarization
of the neutron beam with a direct geometry, in a type of time-of-flight setup. The correlation
time ranges from some picoseconds up to several hundreds of nanoseconds. Similar to XPCS,

the typical length scale probed by NSE is in the nanometer range while the time-range extends to several hundred of ns. For example, a time range of e.g. 100 ns corresponds to effective molecular motion velocities of 1 nm/100 ns = 1 cm/s. If one compares this value to the typical velocity of 200-1000 m/s of the incoming neutrons, it becomes clear why this technique goes under the name of high resolution spectroscopy.

The scattering function

$$S_{i,j}(\mathbf{Q}, \omega) = \frac{1}{2\pi\hbar} \int dt \int d\mathbf{r} \, e^{-i\omega t - i\mathbf{Q}\cdot\mathbf{r}} \, \langle \rho_i(\mathbf{r}' - \mathbf{r}, t)\rho_j(\mathbf{r}', t') \rangle \tag{1}$$

is the common denominator not only of NSE and XCPS but in general of many scattering techniques. In the last equation $\rho_i(\mathbf{r}, t)$ is the density of species i at position $\mathbf{r}$ and time t and it is the coarse grained version of $\sum_l \delta(\mathbf{r} - \mathbf{R}_l^i(t))$, where $\mathbf{R}_l^i$ denotes the positions of the atoms, l, belonging to species i. The Fourier transform of $S(\mathbf{Q}, \omega)$ is the intermediate scattering function

$$I_{i,j}(\mathbf{Q}, t) = \hbar \int d\omega e^{i\omega t} S_{i,j}(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int dt \int d\mathbf{r} \, e^{i\mathbf{Q}\cdot\mathbf{r}} \, \langle \rho_i(\mathbf{r}' - \mathbf{r}, t)\rho_j(\mathbf{r}', t') \rangle. \tag{2}$$

2   The Principles of Neutron Spin-Echo Spectroscopy

Besides trivial factors like the sample amount, transmission and incoming beam intensity, the scattering intensity observed in a neutron scattering experiment depends on the double differential cross section via the scattering function $S(\mathbf{Q}, \omega)$:

$$\frac{d^2\sigma}{d\Omega dE'} = C \, \frac{k'}{k} \sum_{i,j} b_i \, b_j \, S_{i,j}(\mathbf{Q}, \omega). \tag{3}$$

Here k', k are the modulus of the scattered and incoming neutrons respectively. C is proportional to the amount of sample and b_i is the scattering length of atoms of type i and the momentum transfer is $\mathbf{Q} = \mathbf{k}' - \mathbf{k}$.

In neutron spectroscopy one is interested in the energy transferred ($\hbar\omega$) to or from the neutron during the scattering. The wave vector $\mathbf{k}$ relates linearly to the velocity of the neutron $\mathbf{v}$ (momentum $\mathbf{p} = m\mathbf{v}$) as to 1/wavelength

$$\mathbf{v}m = \hbar\mathbf{k}, \quad \frac{2\pi}{\lambda} = k, \tag{4}$$

and the energy transfer can be thus expressed as the difference in kinetic energy between the incoming and the scattered neutron

$$\hbar\omega = (\hbar^2/2m)(k^2 - k'^2). \tag{5}$$

In inelastic scattering the difference of the velocity $\Delta v = v' - v$ is large enough (usually $\Delta v/v > 1\%$) to be measured directly by comparing the average velocity of the incoming beam with the velocity distribution of the scattered neutrons. However, if very small differences have to be detected a sufficiently narrow velocity distribution of the incoming beam (from the continuous velocity spectrum emitted by the moderator of the neutron source) is needed. But this
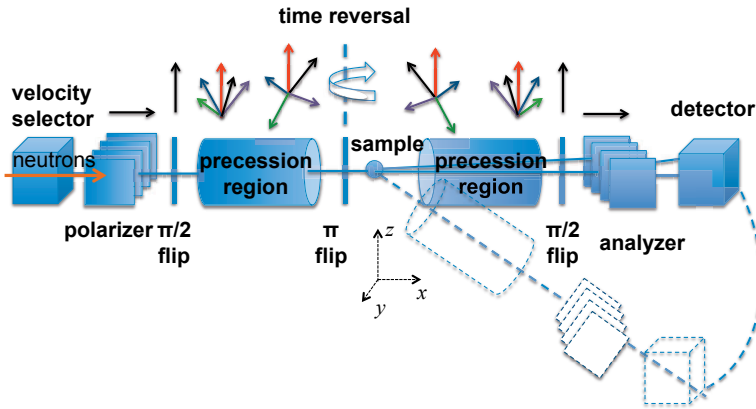
Fig. 2: Schematic setup of a NSE spectrometer. The neutrons enter the instrument from the left. For a description of the single components see the text. The arrows on the top provide a coarse representation of the spin polarization, which fans out in the first arm and it is then recovered in the second arm, because of the time inversion at the -flipper.

would leave only a very small number of neutrons in the incoming beam. As an example let's consider a typical instrument of inelastic scattering, the neutron Time-of-Flight (ToF) spectrometer. The ToF spectrometer is a time-resolved technique where the difference between the known incoming time and velocity of the pulsed neutrons and the velocity and position at the detector after the scattering is measured [2]. In order to improve the resolution of a ToF e.g. of 10% a factor 0.1 must be applied both to the monochromatization as well as to the chopper opening time[1], which increases the resolution at the price of reducing the intensity by a factor $0 1_{monoch} \quad 0 1_{chopper} = 10 ^2$. Thus, the attempt to reach high resolution beyond the $E E \quad 1\%$ in conventional time-of-flight methods leads inevitably to an unacceptable loss of intensity (as expressed in terms of detector count rate).

This hardship can be biased if each neutron could use a kind of individual "stopwatch" in such a way that time of flight before and after the scattering of each neutron path is encoded in an extra parameter (e.g. the direction of the neutron spin) and the difference is obtained at detection. If this stopwatch had a sufficient time resolution it would be possible to observe very small velocity changes even if a beam with a wide range of initial neutron velocities is used. This allows to escape the intensity trap.

NSE uses a polarized incoming beam and a sequence of spin rotations; the reading is performed by the cosine type transmission function of an analyzer and yields only ensemble averages and not individual rotation angles. The intensity at the detector is modulated accordingly. The encountered v values are distributed according to $S(Q \quad k \quad v)$. The detector signal is then proportional to the integral of precession-angle-cosine modulated intensity contributions with weight according to the distribution of v implied by $S(Q)$. I.e. it is proportional to the cosine-Fourier transform of $S(Q)$, the intermediate scattering function $I(Q t)$.

We go know into the details of the NSE technique. Figure 2 shows the schematic setup of a stan-

---

[1] e.g. the interval of time the device lets neutrons pass through.

dard (IN11-type [2]) NSE-spectrometer [3–5]. The neutrons, arriving from the left side, pass first through a velocity selector and they become monochromatized within a specific wavelength-band $\Delta\lambda$, typically between 10% and 20%. Then the beam is polarized longitudinally (i.e. parallel to the direction of flight) before hitting a so-called $\pi/2$-flipper. This device rotates adiabatically the spin direction of $90°$. At this moment the stopwatch is started: The beam enters a region (usually delimited by a cylindrical magnet) with a magnetic field parallel to the direction of motion (the x-axis in the picture). The spin-clockwork is then effected by the (Larmor) precession of the neutron spins around the axis defined by the orientation of the magnetic field $\mathbf{B}$. Extra correction coils inside the precession area guarantee that all neutrons experience the same magnetic field within a certain divergence. The Larmor precession frequency is proportional to the magnitude of the magnetic field, $\omega_L = \gamma B$, where $\gamma = 2\pi\ 2913.06598\ 10^4$ s/Tesla is the gyromagnetic ratio of the neutrons and B is the modulus of the magnetic induction along the path l. If v is the velocity of the neutron, the angle Ψ accumulated by the neutron spin before reaching the sample is proportional to the time spent inside the magnetic field along its trajectory l

$$\Psi = \frac{\gamma}{v} B\, l = \alpha + n2\pi \tag{6}$$

with n an integer. But because each neutron has a slightly different velocity and it follows a different path inside the magnetic field, Ψ changes from particle to particle. Thus the spin vectors - that were initially prepared all aligned perpendicularly to B at the $\pi/2$-flipper - become distributed (ideally) on a disc orthogonal to the field direction at the exit of the magnetic region.

The sample position defines the symmetric point between the first and the second arm of the spectrometer. Close to (ideally at) the sample position a π-flipper is located. As the name suggests, this flipper rotates the spin of $180°$ around a vertical axis. Thus $\Psi \to \Psi_1 = -\alpha + n2\pi$. The second arm is symmetric to the first one, i.e. the precession field and the path length are identical. First we consider the case where no velocity change occurs upon scattering (elastic scattering). Then the angle accumulated in the second arm is simply $\Psi_2 = \alpha + n2\pi$ and the total angle precessed by the neutron spin, before this passes through the second $\pi/2$-flipper, is the sum $\Psi_1 + \Psi_2 = 2n2\pi$. This means that, in the case of an elastic scattering, the beam ends up with the same polarization as before the first $\pi/2$-flipper.

At the second $\pi/2$-flipper, just before the analyzer, the stopwatch is stopped. The direction of the spin is rotated back and the beam is polarized longitudinally again (at least in the case of elastic scattering). Similarly to the polarizer, the analyzer lets only one component of the spin through, i.e. the longitudinal one.

More generally, after ensemble averaging over all neutrons, the count rate at the detector must range between a minimum of zero (all spins are oriented along the negative x direction) up to a maximum of 1 (all spins are oriented along the positive x direction). This means that the measured intensity will be proportional to a *transmission function* of the kind: $(1 + \cos\langle\Psi\rangle)/2$, with $\langle\Psi\rangle$ the expectation value of the angle between spin and axial direction ($\langle\Psi\rangle = 0$ for elastic scattering).

Now we consider the case of quasi-elastic scattering , i.e. $v' = v + \Delta v$ with $\Delta v/v \ll 1$. The total angle accumulated in the two arms will be

$$\Psi_1 + \Psi_2 = \frac{\gamma J_1}{v} - \frac{\gamma J_2}{v + \Delta v} + 4n\pi, \tag{7}$$

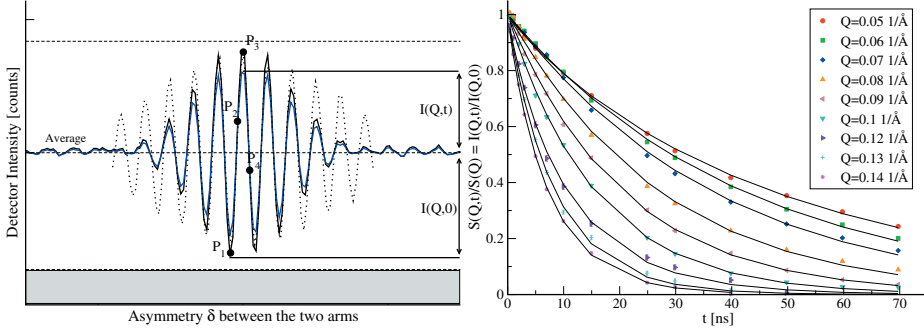[2] https://www.ill.eu/users/instruments/instruments-list/in11/

Fig. 3: The left panel shows typical echoes of the intensity at the detector, obtained by varying the symmetry of the two precession fields. At least three points are necessary to determine the amplitude, the phase and the average but more than three points are normally used. The dashed lines represents the possible maximum and the minimum amplitude obtainable in an experiment, which are used to normalize the echo amplitude. The solid black line is the case of elastic scattering while a quasi-elastic is the blue solid line. The dotted curve shows the effect of a 10% monochromatization on the echo width compared to a 20% band (sold blue and black lines). The right panel shows typical experimental NSE curves for a solution of micelle (sodium dodecyl sulfate (SDS)) in heavy water and salt. Each Fourier-time (t) corresponds to a different magnetic field in the precession coils. Each curve stands for a different scattering angle (at a given wavelength). The intermediate scattering function $I(Q, t)$ is normalized with respect to a reference sample ($S(Q)$) that scatters elastically. The curves are a guide for the eye.

where $J_{1,2}$ stand for the field integrals along all neutron paths in the primary and secondary arm of precession

$$J_1 = \int_{l(\pi/2)_1}^{l(\pi)} B dl, \quad J_2 = \int_{l(\pi)}^{l(\pi/2)_2} B dl, \tag{8}$$

here $l(\pi)$ and $l(\pi/2)_{1,2})$ represent the positions of the flippers. The minus in Eq.7 is because the π flipper inverts the sign of Ψ.

Equation 7 can be expanded in $\Delta v/v$ to obtain $\Psi_1 + \Psi_2 \approx \frac{\gamma J \hbar \omega}{m v^3}$, where we used $J_1 = J_2 = J$ (that holds under the condition of symmetry between the two arms) and the relation $\hbar \omega = m v^2/2 - m v'^2/2 = m v \Delta v + O((\Delta v)^2)$. This expression relates the neutron velocity changes with the total precession angle accumulated by the spins, in the approximation of quasi-elastic scattering. Observing the linear dependence of v as $1/\lambda$ (Eq. 4) it becomes evident that cold neutrons are indispensable to be able to measure tiny changes in the neutron velocity, $\hbar \omega \propto (\Psi_1 + \Psi_2)/\lambda^3$.

The intensity I at the detector is then the transmission function weighted by the distribution function of the incoming wavelengths w_λ, within the wavelength-band $\Delta \lambda$, and by the spectrum of the sample w_ω as given by its scattering function $S(Q, \omega)$

$$I = \eta \iint \frac{1}{2} \left[1 \pm \cos\left(-\frac{\gamma J_1}{(h/m)\lambda^{-1}} + \frac{\gamma J_2}{(h/m)\lambda^{-1} + \lambda\omega/2\pi}\right) \right] S(Q, \omega) w_\lambda(\lambda) \, d\omega \, d\lambda. \tag{9}$$

In writing Eq. 9 we used Eqs. 5 and 4 which allow to substitute Δv with $\lambda\omega/2\pi$. η is a calibration factor of the instrument.

After some more mathematics, observing the distribution of incoming wavelengths and the distribution of deviations of the field integrals from their nominal value (inhomogeneity), the following approximate expression for the detected intensity is obtained:

$$I_{Det} \propto \frac{1}{2}\left[S(Q) \pm \frac{\exp(-\Psi^2)}{\mathcal{N}} \ \cos(\delta\gamma\frac{m}{h}\lambda) \int \cos(J\lambda^3\gamma\frac{m^2}{2\pi h^2}\omega) \ S(Q,\omega)d\omega \right]. \qquad (10)$$

In the last equation we introduced δ as the asymmetry of magnetic field integral between both spectrometer arms and then we used $J_1 = J_2 + \delta$. The configuration $\delta = 0$ is also called the symmetry point. In a standard NSE experiment the symmetry is perturbed by means of a small, auxiliary magnetic coil (usually called phase coil) in one of the two arms. The asymmetry parameter δ is then proportional to the current in the phase coil. Under the assumption that the incoming wavelength distribution and the distribution of field integrals around their nominal values are Gaussians, the following explicit relations hold: $\mathcal{N}^2 = 1 + 4\Sigma^2\Lambda^2\gamma^2m^2/h^2 \simeq 1$ and $\Psi^2 = [\Sigma^2\lambda_0^2 + \Lambda^2\delta^2]\gamma^2m^2/h^2/\mathcal{N}^2$. $\Sigma^2 = \langle\Delta J^2\rangle/2$ relates to the field integral inhomogeneity and $\Lambda = \Delta_{FWHM}\lambda/4/\sqrt{\ln 2}$ to the full width at half maximum (FWHM) of the incoming wavelength distribution. For 10% wavelength width, $\lambda_0 = 1\,\mathrm{nm}$ and $\Sigma = 2 \times 10^{-6}\,\mathrm{Tm}$ the resolution factor is $R = \exp(-\Psi^2) = 0.43$.

In an experiment, see figure 3, the asymmetry δ is varied to determine the amplitude, A, of the first cosine factor in Eq. 10:

$$A = \pm\underbrace{\exp(-\Psi^2)/\mathcal{N}}_{=R} \int \cos(\underbrace{J\lambda^3\gamma\frac{m^2}{2\pi h^2}}_{=t} \ \omega) \ S(Q,\omega)d\omega. \qquad (11)$$

As the cosine-Fourier transform of the scattering function the amplitude A is, up to the resolution R, proportional to intermediate scattering function $I(Q,t)$. The resolution factor R is determined by measuring a reference sample with purely elastic scattering. Thus the output of a NSE experiment is always normalized to the intensity of this reference sample. The normalization cancels the $S(Q)$ term in Eq. 10 and allows a representation of the data as curves with intercept at 1 for $t \to 0$ (according to sample typology sometimes a background needs to be measured and subtracted).

3 The Principles of X-Ray Photon Correlation Spectroscopy

X-ray photon correlation spectroscopy (XPCS) measures the intermediate scattering function on a mesoscopic scale. Similarly to photon correlation spectroscopy with visible light or quasi-elastic light spectroscopy, i.e. DLS, XCPS provides the intensity-intensity correlation function of a speckle-pattern in the scattered radiation. But instead of implementing visible light from a laser, XPCS exploit the coherence of a X-ray beam. A coherent illumination is indispensable to obtain an interference pattern which is used to enhance the sensitivity of, e.g., the dynamical features in the scattering experiment. A speckle pattern is a pattern with random intensity modulation which can be observed when a spatially and temporally coherent light is scattered by a sample such that random fluctuations of the optical path are introduced which are comparable in length-scale with the wavelength. The speckles are related to the exact spatial arrangement of the disorder. Such information is not accessible with incoherent light because the diffraction
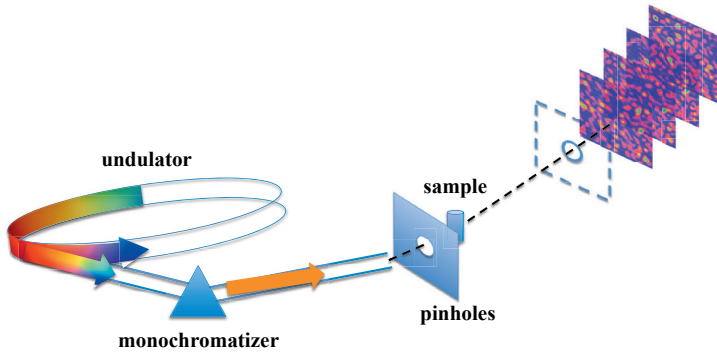
Fig. 4: Schematic setup of a XPCS instrument and of a snapshots at different instants of speckle patterns at the detector.

pattern observed in an ordinary diffraction experiment is typically an ensemble average containing only information on the average correlations in the sample. If the spatial arrangement changes with time, the corresponding speckle pattern also changes. The intensity fluctuations of the speckles thus provide information about the underlying dynamics.

The intensity-intensity correlation is a second order correlation. The correlation of the intensity within a single speckle provides the information about the dynamics of the fluctuations at a given Q. Because the intensity is proportional to the square modulus of the electric field ($\mathcal{E}^2$), we can write the second order correlation function g_2 as

$$g_2(Q,t) = \frac{\langle \mathcal{E}^2(Q,t')\mathcal{E}^2(Q,t'+t)\rangle}{\langle \mathcal{E}^2(Q,t')\rangle^2} = 1 + A(Q)[I(Q,t)/S(Q)]^2 \tag{12}$$

where the optical contrast factor $A(Q)$, with $0 < A(Q) < 1$, is determined by the experimental setup [6] and by the coherence of the beam. In order to derive the r.h.s. of Eq. 12 one needs to remember that $\langle \mathcal{E}^2(Q,t')\rangle = \langle \mathcal{E}^2(Q,t'+t)\rangle = \langle \mathcal{E}^2(Q,t)\rangle$, that $\langle \mathcal{E}(Q,t')\mathcal{E}(Q,t'+t)\rangle$ is zero and that at last $\langle \mathcal{E}(Q,t')\mathcal{E}^*(Q,t'+t)\rangle$ is proportional to $\langle \mathcal{E}^2(Q,t)\rangle$ times the Fourier-transform of the power spectrum (and similarly for the complex conjugate). This is known as the the Siegert relation [7] which relates the field correlation function g_1 to g_2 as $g_2 = 1 + |g_1|^2$, with g_1 being proportional to the second order correlation function of the electric field intensity. The normalized electric field correlation function, in turn, yields the normalized dynamic structure factor, also known as the intermediate scattering function $I(Q,t)/S(Q)$.

XPCS extends the domain of DLS towards longer times and smaller length scales, thus it deserves to be mentioned among the high-resolution techniques in spectroscopy. Compared to NSE the time domain of XPCS is more restricted in the short-time regimes.

The instrumental setup is shown in figure 4. The coherence properties of the X-ray beam as well as the size of the illuminated sample volume enter in the time dependent correlation function, while the size of the illuminated region on the sample determines the size of the speckles, i.e. the solid angle $\Delta\Omega$, which a peak or a minimum of intensity covers in the random speckle texture of scattered intensity. The detector element size needs cover at most $\Delta\Omega$. For example would a sample of $30\,\mu$m height (width) be observed with a wavelength of $\lambda = 1.5$ Å it allows for a detector at 5 m distance with an opening height (width) of $25\,\mu$m. If the detection area is enlarged the sample volume must be reduced or, vice-versa, a larger sample volume requires

smaller detection elements. A way out of this intensity limiting trap is the use of area detectors (CCD cameras) which have small size individual pixels [6].

Besides the detector pixel size also the ratio between the size of the source ($\sigma_{x,y}$) and the distance R determines the coherence of the beam and it is one of the parameters entering the contrast A. This property is defined as the transverse coherence length $\xi_{x,y} = \lambda R/(2\pi\sigma_{x,y})$ which is to be compared to the sample size. On the other hand the longitudinal coherence length of the beam is proportional to the degree of monochromatization of the beam itself $\Lambda_c = \lambda(\Delta E/E)/\pi$. Up to $Q < 10^4 \times 2\pi/\Delta x$ the transverse coherence lengths dominate for a monochromatization of $\Delta E/E \simeq 10^{-4}$.

Only modern synchrotron radiation facilities provide the required source characteristics and the necessary intensity. As far as the considerations on coherence are concerned the question is how good the resolution in Q has to be in order to be able to resolve modulations which are the size of the scattering volume.

As mentioned before, in XPCS the scattering contrast , e.g. in polymer samples, is given by differences in electron density, that in turn depends on the monomer type. In cases where the contrast is small and the intensity low (as is for pure hydrocarbons with oxygen or hydrogen atoms) the area detection using a CCD camera is very important but unfortunately it limits the shortest time to several ms [6, 8]. Different are the cases with (pseudo) Bragg peaks in specular reflection, e.g. in smectic membranes undulations [9, 10], which deliver a much higher intensity at the detector and a fast single avalanche photodiode can be used to reach a resolution time down to a few nano-second.

The XPCS technique is an emerging technique [11–14] that still suffers from the lacking of a detector with fast time resolution, combined with the intensity advantage of multi-speckle detection as obtained with a CCD camera. Other limitations of the XPCS techniques pertain possible damages to the sample: The high flux can cause heating, or modifications of the sample, e.g. the radiation can generate cross linking in polymers and charging of colloids. In same cases a poor signal to noise can also be a problem. The effects from radiation damage may be mitigated if new sample volumes are exposed after each short period of irradiation.

4   Examples and applications

Because of its wide dynamical ranges high resolution spectroscopy (HRS) is well established in many fields of investigation both in soft and in hard matter. The list of examples we provide here is meant for the reader to get a first impression of the potential of these techniques but it far from being exhaustive.

The trends in the investigation of soft- and bio-materials go more and more in the direction of nano- and mesoscopic structures. This means that the relevant length scale extends from the single molecular size of 0.5-1 nm over typical macromolecular or protein dimensions (5-20 nm) up to larger length scales of e.g. bio-material or living matter. The time scale depends more on the spatial resolution and on the size-dependent diffusivity of the systems under investigation. Relaxation times of many 100 ns can be observed even for length scales of only few nm.

The dynamics of polymers, and in general of organic molecules like microemulsions and bio-

logical systems is widely studied by means of high resolution spectroscopy, and especially in experiment with neutrons the possibility of manipulating the contrast by an appropriate H/D-replacement is particularly useful. In functional and responsive systems (like e.g. the so-called self-healing materials) the structure affecting the function is often nano- or macromolecular-sized and it contributes to the scattering function at low Q where the time scale can range up to many thousand of nano-seconds.

The study of energy materials by means of high resolution spectroscopy has also been an emerging topic in the last years. The need for investigation at large Q of the slow diffusion due proton "jumps" in solid electrolytes and similar enhances again the requirement for high resolution at small wavelengths. I.e. $Q = 2$ Å^{-1} at a time scale of circa 40-50 ns is only accessible with high resolution spectroscopy.

Extending the limits of investigation toward higher time-scales is not the only advantage of HRS. Having optimized the technique for the high-resolution realm allows at the same time to run experiments in a regime of lower-resolution with higher performance. This characteristic makes it possible to reduce the measuring time of experiments that would otherwise take too long with other techniques, thus allowing systematic investigations of e.g. fast-perishable samples.

Among the hard-matter systems, magnetic materials are also studied by means of high resolution methods. These can be used to obtain a direct and detailed view into magnetic structures and their dynamics. Neutron spin-echo and X-ray photon correlation spectroscopy are suited to analyze spin/magnetic fluctuations over a large dynamical range, e.g., in systems with frustration, in spin glasses or in superconductors.

Between the two methods described here the neutron spin-echo is the one that is maybe more restrictive for magnetic materials. The main limitations are depolarization effects on the beam due to (large) magnetic domains in the sample. Usually ferromagnetic scattering requires an extra setup. The case of paramagnetic scattering is different because it can be performed without any invasive modifications of the NSE-setup: Due to the magnetic interaction at the sample an "intrinsic" spin-flip takes already place which makes the implementation of the π-flipper unnecessary. On the other hand paramagnetic scattering at NSE profits from the fact that the NSE signal comes only from (para)magnetic scattering and it can be distinctively separated from nuclear density fluctuations.

The choice between NSE and XPCS is not only dictated by the specific Q- and time-needs, which in turn are influenced by the characteristics of sample, but also by some advantages that one method has over the other. The "contrast matching" surely plays an important role in neutron scattering and it allows, if the percentage of protonated (H) and deuterated (D) material is accurately calibrated, to put in evidence specific (internal) motions in the sample. At last XPCS does not suffer from loss of signal due to beam depolarization as is the case of NSE, especially when either ferromagnetic samples or magnetic fields need to be implemented.

4.1   Brownian motion and collective diffusion

The most simple example one can consider is that of a colloidal system, e.g., of a suspension of spheres in a simple liquid. We start with the one-dimensional motion of a spherical particle

of radius R, mass m, velocity v in a fluid with viscosity η. Experience teaches us that the force on the particles is given by the friction force fv, which is proportional to the velocity of the particles. The friction coefficient of a sphere with radius R is given by the Einstein-Stokes formula $f = 6\pi\eta R$.

Brownian motion is the result of random collisions of the atoms of the embedding fluid ("heat bath"). Thus, in addition to the friction force a randomly fluctuating force must be introduced to account for these random collisions

$$\langle \zeta(t) \rangle = 0 \qquad \text{and} \qquad \langle \zeta(t)\zeta(t') \rangle = A\delta(t - t'). \tag{13}$$

Then the equation of motion of the Brownian particles can be written as the Langevin equation

$$m\frac{dv}{dt} + f\,v = \zeta(t) \tag{14}$$

of which a general solution is given by

$$v(t) = e^{-ft/m}v(0) + \int_0^t dt'e^{-f(t-t')/m}\zeta(t')/m. \tag{15}$$

It is important to stress the role of the second term in Eq. 15: Without it (i.e. if $\zeta(t) = 0$) the velocity would decay to zero for long times, thus contradicting the fact that at thermal equilibrium the mean squared velocity must be $\langle v_{eq}^2 \rangle = k_B T/m$. Here $v(0)$ stands for the initial velocity and k_B for the Boltzmann constant. The random force provides indeed the 'extra' velocity that fulfills the condition at equilibrium. The mean velocity is simply given by $\langle v(t) \rangle = e^{-ft/m}v(0)$ because $\langle \zeta(t) \rangle = 0$. Three terms contribute to the mean squared velocity $\langle v(t)^2 \rangle$: the $v(0)^2$ and the cross term

$$e^{-2ft/m}v(0)^2, \qquad \frac{2}{m}e^{-ft/m}v(0)\int_0^t dt'e^{-f(t-t')/m}\zeta(t') \tag{16}$$

and the second order term in the random force

$$\frac{1}{m^2}\int_0^t dt'e^{-f(t-t')/m}\zeta(t')\int_0^t dt''e^{-f(t-t'')/m}\zeta(t''). \tag{17}$$

On averaging over the noise the cross term vanishes while the second order term can be written as the double integral $\int \int dt'dt''\delta(t' - t'')$. The delta function removes one of the integrals, the remaining one can be easily calculated. The result is

$$\langle v(t)^2 \rangle = e^{-2ft/m}v(0)^2 + \frac{A}{fm}(1 - e^{-2ft/m}). \tag{18}$$

If $\langle v(t)^2 \rangle \to \frac{A}{fm}$ for $t \to \infty$, then it must be $A = fk_B T$.

The next step is to calculate the mean square displacement $\langle \Delta x(t)^2 \rangle := \langle [x(t) - \langle x(t) \rangle]^2 \rangle$, that yields direct information about the diffusion of the particles. Given that the displacement of the particle is $x(t) = \int_0^t dt'v(t')$, we first need to calculate the velocity correlation function $\langle v(t)v(t') \rangle$. Starting from Eq. 15 and taking into account that terms linear in $\zeta(t)$ vanish on average, one can derive the following relation

$$\langle v(t)v(t') \rangle = \frac{k_B T}{m}e^{-f|t-t'|/m}, \tag{19}$$

where $\langle \zeta(t)\zeta(t') \rangle = A\delta(t - t')$ has been used.

The 'free-particle' diffusion coefficient is given in general by $D_0 = \int_0^\infty dt \langle v(t)v(0) \rangle$ in one dimension, for the three dimensional case one must take the scalar product of the velocity and introduce a factor $1/3$ before the integral. If one inserts Eq. 19 into the integral before, one can obtain the known expression for the diffusion constant

$$D_0 = \frac{k_B T}{f}. \tag{20}$$

From Eq. 19 we can integrate in the time and find the mean square displacement

$$\langle \Delta x(t)^2 \rangle = \frac{2k_B T}{f}[|t| - \tau(1 - e^{t/\tau})] \tag{21}$$

with $\tau := m/f$. Expanding the previous result to 3 dimensions a factor 3 would multiply in front of the above expression.

Given Eq. 20 the term $2k_B T|t|/f$ in Eq. 21 corresponds to the free diffusion. At short time one finds that $e^{t/\tau} \simeq 1 - t/\tau + (t/\tau)^2/2 - \cdots$. The motion of the particles before their initial velocity is damped out by friction is then $\langle \Delta x(t)^2 \rangle = (k_B T/m)t^2$, i.e. $\Delta x \propto t$. This inertial force may be ignored in most colloidal system because $\tau \ll t$, where t is the time-scale of interest.

We try to derive now an expression for the intermediate scattering function; in three dimension it can be written as:

$$I(Q, t) = \frac{1}{N} \sum_{i,j} \langle e^{i\mathbf{Q}\cdot(\mathbf{r}_i(0) - \mathbf{r}_j(t))} \rangle = \frac{1}{N} \sum_{i,j} \langle e^{i\mathbf{Q}\cdot(\mathbf{r}_i(0) - \mathbf{r}_j(0))} e^{-i\mathbf{Q}\cdot\Delta\mathbf{r}_j(t)} \rangle \tag{22}$$

where N is the number of particles in the system and "0" stands for the initial time (zero). For an accurate derivation of what follows please refer to Pusey and Tough in Ref. [15].

To an equilibrium - where no preferential direction of drift is expected - the arrangement of the particles $\{r_i(0)\}$ will give rise on average to the static structure factor $S(Q) = S(Q, t = 0)$.

For short times the particle motion can be thought of Brownian diffusion around the equilibrium position if the total arrangement does not change significantly. As long as the initial velocities are not correlated, the scattering contributions will be self-contributions from N single scatterers that exhibit the diffusion constant $D_0 = k_B T/f$. Therefore we can assume a Gaussian distribution of $\{\Delta r_j(t)\}$ and write $\langle e^{-i\mathbf{Q}\cdot\Delta\mathbf{r}_j(t)} \rangle = e^{\langle -i\mathbf{Q}\cdot\Delta\mathbf{r}_j(t) \rangle^2}$. With the relation derived before for $\langle \Delta\mathbf{r}_j(t)^2 \rangle$ at short times we can write

$$I(Q, t) = S(Q)e^{-6D_0 Q^2 t}. \tag{23}$$

For longer times effects of confinement might slow down the self diffusion if we imagine that in a dense system the path of diffusion can be blocked by neighbouring particles. It must be also noticed that collective diffusion rather probes the relaxation of density gradients which does not necessarily imply the diffusion of single particles over larger distances.
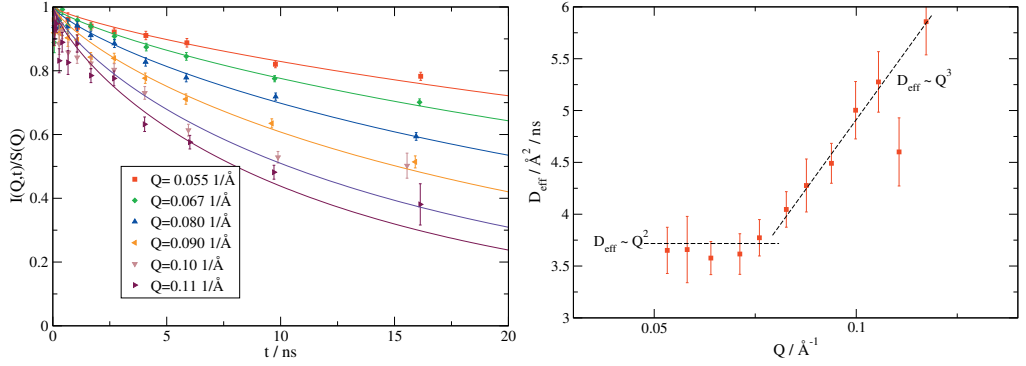
Fig. 5: Left the NSE data for a microgel (PNIPAM) in D_2O at different Q's measured at 20°C. Right the fitted diffusion constant (see text).

If also hydrodynamic interactions are considered, e.g. between the particles and the solvent, then the situation becomes more complicated. In this case the previous considerations do not hold anymore and the effective diffusion scales with Q with higher powers then Q^2, $D_{eff}(Q) = D_0 H(Q)$, where the hydrodynamic factor $H(Q)$ is a function of the pair correlation function.

The quantity $D_0 H(Q)$ depends on the hydrodynamic radius r_h and on the fluid viscosity η and it can be determined with a combination of diffraction experiment and spectroscopy measurement.

Figure 5 shows the experimental curves for a a protonated PNIPAM microgel in heavy water measured with NSE. Microgels are cross-linked polymers that have a volume phase transition into a collapsed state above 32°C [16].

The curves can be fitted with a simple stretched exponential $\frac{I(Q,t)}{S(Q)} = \exp\left[-(t/\tau)^\beta\right]$, where $\beta = 0.8$ is typical for spin-echo [17] and τ is a fitting parameter. We are interested here in the center-of-mass diffusion of the microgel at short times. From Eq. 23 we know that the intermediate scattering function is related to the diffusion constant. Through the mean value $\langle \tau \rangle$ we can define an effective diffusion constant as

$$D_{\text{eff}}(Q) = \frac{\langle \tau \rangle}{Q^2} = \frac{1}{(\tau/\beta)\Gamma(1/\beta)\,Q^2}, \tag{24}$$

where $\Gamma(x)$ is the Gamma function. In the right panel in figure 5 the coefficient is plotted versus Q. Two regimes are visible: For small Q a pure diffusive behaviour with D scaling as Q^2 and a second regime at higher Q that deviates from the expected Q-square behaviour. At this length-scale the interaction with the solvent becomes visible and the so-called Zimm dynamics is dominant with its typical scaling $D \propto Q^3$.

4.2 Microrheology

Microrheology [18, 19] is an other application of experiments that follow the particle diffusion in terms of time dependent mean squared displacements. Here the diffusing particles are microscopic probes that respond to the viscoelastic properties of the surrounding fluid.

First attempts to use XPCS experiments in microrheology was made using silica particles of
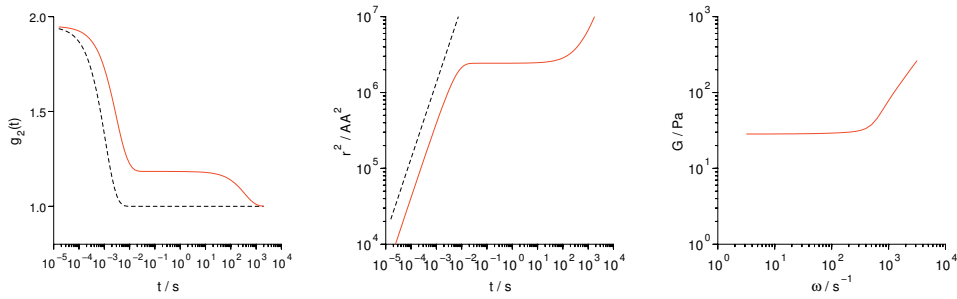
Fig. 6: Schematic representation of a part of the results obtained for gellan in aqueous solution (5%) adapted from Ref. [18]. Dashed lines correspond to probe particles in pure water. The sequence shows the generic appearance of the raw XPCS results in terms of $g_2(t)$ via the evaluation step yielding the msq-displacement and finally leading to the frequency dependent shear modulus $G(\omega)$. The figure displays the real part.

$222\,\mathrm{nm}$ radius as probes and polystyrene sulfonate polymers with comb architecture or gellan, a polysaccharide in aqueous solution as viscoelastic fluid [18].

Figure 6 shows a calculated g_2 curve as it would be obtained from a XPCS experiment. While the first step from $g_2(t) \rightarrow g_1(t) \rightarrow \langle \Delta r^2(t) \rangle$ was explained in the previous sections, the passages to the shear modulus $G(\omega)$ is more complicated. We refer to [19] for the details but basically one has to use first the Laplace transform with $-i\omega \rightarrow s$. The Laplace transform of $\langle \Delta r^2(t) \rangle$ is then evaluated numerically. Then a model expression is used to fit the results for $G(s) \simeq \sum_{i=1,n} a_i s^{\nu_i}$, at last $G(\omega)$ is obtained by analytic continuation ($s \rightarrow -i\omega$).

References

[1] F. Mezei, Zeitschrift Fur Physik **255**, 146 (1972).

[2] N. Violini, J. Voigt, S. Pasini, T. Brückel, Nuclear Instruments and Methods in Physics Research A **736**, 31 (2014).

[3] F. Mezei (ed.), *Neutron Spin Echo*, number 128 in Lecture Notes in Physics Vol. 128 (Springer, Berlin, Heidelberg, New York, 1980).

[4] F. Mezei, C. Pappas, T. Gutberlet (eds.), *Neutron Spin Echo Spectroscopy*, number 601 in Lecture Notes in Physics (Springer, Berlin, Heidelberg, New York, 2003).

[5] M. Monkenbusch, R. Schatzler, D. Richter, Nuclear Instruments & Methods in Physics Research Section A-Accelerators Spectrometers Detectors and Associated Equipment **399**, 301 (1997).

[6] H. Kim, K. Lim, Bulletin Of The Korean Chemical Society **25**, 382 (2004).

[7] H. Voigt, S. Hess, Physica A **202**, 145 (1994).

[8] C. Li, T. Koga, J. Jiang, S. Sharma, S. Narayanan, L. Lurio, Y. Hu, X. Jiao, S. Sinha, S. Billet, D. Sosnowik, H. Kim, J. Sokolov, M. Rafailovich, Macromolecules **38**, 5144 (2005).

[9] I. Sikharulidze, W. de Jeu, Physical Review E **72**, 011704 (2005).

[10] I. Sikharulidze, I. Dolbnya, A. Madsen, W. de Jeu, Optics Communications **247**, 111 (2005).

[11] P. Falus, L. Lurio, S. Mochrie, Journal Of Synchrotron Radiation **13**, 253 (2006).

[12] P. Falus, M. Borthwick, S. Mochrie, Review Of Scientific Instruments **75**, 4383 (2004).

[13] D. Lumma, M. Borthwick, P. Falus, L. Lurio, S. Mochrie, Physical Review Letters **86**, 2042 (2001).

[14] G. Grubel, F. Zontone, Journal Of Alloys And Compounds **362**, 3 (2004).

[15] P. Pusey, R. Tough, Journal of Physics A: Math. Gen. **15**, 1291 (1982).

[16] S. Maccarrone, O. Mergel, F. Plamper, O. Holderer, R. Richter, Macromolecules **49**, 1911–1917 (2016).

[17] D. Richter, M. Monkenbusch, A. Arbe, J. Colmenero, Neutron Spin Echo In Polymer Systems **174**, 1 (2005).

[18] A. Papagiannopoulos, T. Waigh, A. Fluerasu, C. Fernyhough, A. Madsen, Journal Of Physics-Condensed Matter **17**, L279 (2005).

[19] T. Mason, D. Weitz, Physical Review Letters **74**, 1250 (1995).

D 7   Coherent Imaging with X-ray Free-Electron Lasers

H. N. Chapman

Center for Free-Electron Laser Science

DESY, Hamburg

Contents

1   Introduction   2

2   Coherent Scattering from Isolated Objects   3

   2.1   The Phase Problem   6

   2.2   Iterative Phase Retrieval Algorithms   10

   2.3   Coherence and Detector Requirements   13

3   Outrunning Radiation Damage   14

   3.1   Diffraction Before Destruction   14

   3.2   Diffraction Termination   15

4   Diffraction Data from Identical Particles   19

   4.1   Serial Nanocrystallography   19

   4.2   Single-Particle Diffraction   22

   4.3   Diffraction from Crystals with Translational Disorder   23

5   Outlook   25

Lecture Notes of the 50th IFF Spring School "Scattering! Soft, Functional and Quantum Materials" This is an Open Access publication distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (Forschungszentrum Jülich, 2019)

1   Introduction

X-ray free-electron lasers (FELs) such as the Linac Coherent Light Source (LCLS) [1], SACLA [2], and the European XFEL [3], produce laser-like pulses of X-rays of less than 10 fs to 300 fs duration and up to about 10^{13} photons per pulse. Just as with the introduction of the synchrotron, X-ray FELs are bringing new capabilities that are having great impact on many scientific disciplines by opening up the study of materials at the length scale of interatomic distances and at the corresponding time scales of atomic motion [4]. In condensed matter science, for example, materials properties can be drastically altered by inducing transient structures using ultrafast light pulses [5]. Similarly, matter at extreme limits of temperature and pressure, similar to conditions in the cores of stars and planets, can be created transiently in the laboratory using intense optical pulses. The extremely intense and brief pulses delivered by X-ray FELs open up the direct imaging of processes in these systems, which previously could only be studied by spectroscopic means. In the biological sciences, time-resolved X-ray crystallography has elucidated the atomic motions that occur on stimulation of light-sensitive proteins, with femtosecond temporal resolution [6]. The extreme irradiance[1] of the pulses may let us shrink crystal sizes all the way down to single molecules, giving three-dimensional movies of conformational dynamics and chemical reactions, and allowing the imaging of macromolecules that cannot be easily crystallized.

X-ray crystallography is indeed the inspiration for the methods of imaging at X-ray FELs [7]. Crystallography can obtain atomic-resolution images of protein macromolecules without the use of a lens, based on the interpretation of the coherent scattering pattern detected in the far field. Since this method is lensless, there are no technological limitations to the achievable resolution caused by the perfection of objective lenses that can be made. Instead, the image is synthesized from the measured diffraction intensities. Each intensity sample is related to the strength of each spatial frequency component of the synthesized image. The shift of each spatial frequency component in real space is given by the corresponding phase of the diffracted wavefield. Crucially, this information cannot be obtained by measurement. The correct image, formed by the sum of the correctly positioned periodic components, cannot be synthesized without this information. This so-called "phase problem" has been largely solved in protein crystallography by constraining the phases through additional information about the structure such its atomicity (if scattering to high enough resolution was obtained), protein sequence, or known similar structure. Anomalous diffraction methods can be used to determine the locations of heavy atoms, which in turn can act as holographic references to obtain the remaining image. This general imaging strategy works even better for non-crystalline materials, since the information content of a continuous diffraction pattern of a non-periodic object is large enough to completely constrain the phases, allowing *ab initio* reconstruction of both two- and three-dimensional images [8, 9]. The phase retrieval algorithms in this case were first developed in the context of electron microscopy [10] and optical remote sensing [11].

Both crystallography and diffraction measurements from non-periodic objects requires sufficient spatial and temporal coherence of the illumination wavefield so that the scattered waves from extreme points in the object interfere to create the fringe pattern on the detector that encodes their separation. The lack of crystallinity does come at a severe cost of diffraction signal,

---

[1] Pulse irradiance is given in units of photons per unit area and time. It is often referred to as "intensity", although in crystallography that word commonly means "diffracted counts" which is the meaning employed in this chapter.

since the integrated Bragg intensity is proportional to the number of units in the crystal [12] and the number of repeats in even small protein crystals can number in the billions. Unfortunately the lower signal cannot simply be compensated by longer exposure, since biological materials can only withstand a limited dose before they are destroyed by the very beam that is used to obtain the image [13]. But here the ultrafast pulses of X-ray FELs come to the rescue, by diffracting from the sample before the effects of radiation damage have set in [14]. This concept of "diffraction before destruction" holds to atomic resolution, at doses over 1000 times higher than can be tolerated by slow exposures [15, 16]. The development of this imaging technique is following two tracks: application to unique structures, such as cells or soot particles, that can only be imaged in a single shot; and to reproducible objects, such as viruses, macromolecular complexes, and protein nanocrystals, and where signals can be accumulated over many copies in a rapid series of diffraction measurements. These are extremely active fields of research, and progress has been accelerating as more hard X-ray FEL facilties have been brought on-line.

In this lecture, brief descriptions of coherent scattering and iterative phasing techniques are given, showing how 2D and 3D images of finite-sized non-periodic objects can be recovered from diffraction data. Examples of the application of this technique are presented from experiments carried out at the FLASH soft X-ray FEL in Hamburg and the LCLS, and the research of single-particle diffractive imaging and "diffraction before destruction" is summarized, including computational and experimental investigations of the rate of perturbation of a structure under intense X-ray illumination. A new method of the crystallography of disordered macromolecular crystals is introduced, whereby "single molecule" diffraction is obtained by considering the crystal as an ensemble of aligned molecules.

2 Coherent Scattering from Isolated Objects

The interaction of x-rays with matter can be described by the inhomogeneous Helmholtz equation,

$$(\nabla^2 + k^2)\psi(\mathbf{x}) = \Phi(\mathbf{x})\psi(\mathbf{x}) \tag{1}$$

for an incident plane wave of wavelength λ or wavenumber $k = 2\pi/\lambda$. The scattering potential of a static and non-magnetic object may be expressed as

$$\Phi(\mathbf{x}) = k^2(1 - n^2(\mathbf{x})) = 4\pi r_e \rho(\mathbf{x}) \tag{2}$$

where n is the refractive index, r_e the classical radius of the electron, and $\rho(\mathbf{x})$ the electron density of the object. In the Born approximation the solution to (1) is of the form

$$\psi(\mathbf{x}) = e^{i\mathbf{k} \cdot \mathbf{x}} + \frac{e^{ikr}}{r} f(\mathbf{Q}) \tag{3}$$

where

$$f(\mathbf{Q}) = -r_e \int \rho(\mathbf{x}) \exp(i\mathbf{Q} \cdot \mathbf{x}) \, d\mathbf{x}. \tag{4}$$

and $\mathbf{Q} = \mathbf{k}' - \mathbf{k}$ is the photon momentum transfer. The physical picture that explains the form of the scattering factor $f(\mathbf{Q})$ is shown in Fig. 1 (a). A ray scattered in a direction $\mathbf{k}'$ from a point $\mathbf{x}_1$ will acquire a path difference of $\ell_1 = (\mathbf{x}_1 \cdot \hat{\mathbf{k}}' - \mathbf{x}_1 \cdot \hat{\mathbf{k}})$ relative to a ray scattering from the origin $\mathbf{O}$, where $\hat{\mathbf{k}}$ are unit vectors. This is the difference of the lengths of the thick
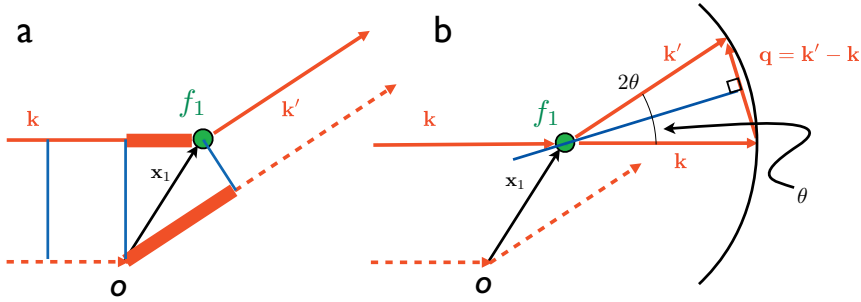
Fig. 1: The far-field scattering geometry and the Ewald-sphere construction.

lines in Fig. 1 (a). The accumulated phase will therefore be $\phi_1 = (2\pi/\lambda)\ell_1 = \mathbf{x}_1 \cdot \mathbf{Q}$. The point scatterer itself may cause a modification to the wave by the complex constant f_1, giving a scattering $f_1 \exp(i\phi_1) = f_1 \exp(i\mathbf{x}_1 \cdot \mathbf{Q})$. Equation (4) is simply the integration over all scatterers in the object. The object's electron density is assumed to be continuous, although it can certainly be atomistic and may possess translational periodicity (i.e. be a crystal).

Equation (4) states that the scattering amplitude f is given by the Fourier transform of the electron density. That is, the strength of diffraction in the direction $\mathbf{k}_{\text{out}}$ only depends on the Fourier component $\tilde{\rho}(\mathbf{q})$, where we define the Fourier transform (in three dimensions) as

$$\tilde{g}(\mathbf{q}) \equiv \mathcal{F}\{g(\mathbf{x})\} \equiv \int g(\mathbf{x}) \exp(i\mathbf{x} \cdot \mathbf{q})\,d\mathbf{x}. \tag{5}$$

This component is a particular spatial frequency in the object, which may be thought of as a volume grating of a particular wavenumber $|\mathbf{q}|$ and direction $\hat{\mathbf{q}}$. From Fig. 1 (b), it is seen that the magnitude of the momentum transfer $\mathbf{Q}$ is given simply by

$$|\mathbf{Q}| = 2|\mathbf{k}| \sin\theta = \frac{4\pi}{\lambda} \sin\theta \tag{6}$$

for a scattering angle 2θ, and that due to the conservation of $\mathbf{k}$ (that is, elastic scattering) the vector $\mathbf{Q}$ lies on the surface of a sphere (called the Ewald sphere). We see from the diagram in Fig. 1 (b) that the scattered ray appears to reflect at an angle θ from a plane normal to $\mathbf{Q}$. That is, the ray reflects from the volume grating which is tilted at the angle θ relative to the incoming wave-vector. The ray only reflects if the period of the volume grating, $d = 2\pi/|\mathbf{Q}|$ satisfies Eqn. (6), which is to say $d = \lambda/(2\sin\theta)$ which is well recognized as Bragg's law. We stress that although Bragg's law and the Ewald sphere construction are well known concepts in crystallography, there is no requirement of periodicity of the object in the derivation or application of these concepts.

A typical coherent diffraction experiment illuminates a sample of electron density $\rho(\mathbf{x})$ (or refractive index $n(\mathbf{x})$) with a quasi-monochromatic plane wave and measures the diffraction pattern in the far field with a planar detector (such as a bare CCD or pixel-array detector). This is a measure of the intensity of the wavefield, given by

$$I(\mathbf{Q}) = I_0\,\Omega_p\,P\,r_e^2\,|\tilde{\rho}(\mathbf{Q})|^2, \tag{7}$$
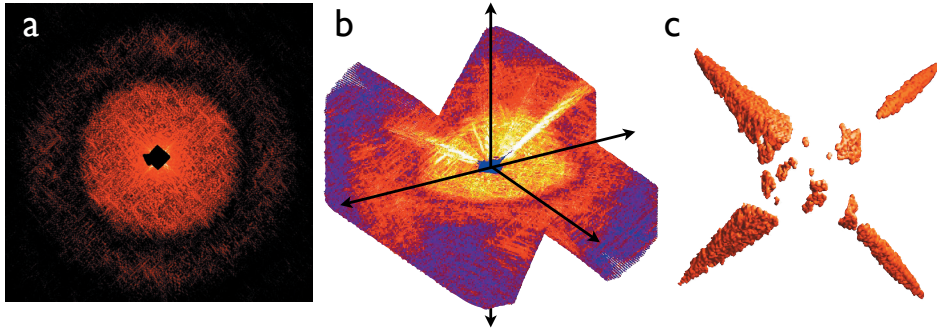
Fig. 2: Diffraction data collected from a 3D test object, showing (a) a diffraction pattern recorded at a single orientation, (b) 3D diffraction intensities collected at orientations from $-70°$ to $+70°$, and (c) the reconstructed volume image [9].

for pixels of solid angle Ω_p, and where P is the polarization factor. The mapping from pixel coordinate to $\mathbf{Q}$ is easily obtained by the geometry shown in Fig. 1 where $\mathbf{k}'$ points in the direction of the pixel. Note that a single exposure in this geometry records only information about the object for spatial frequencies that lie on the Ewald sphere. Other frequencies are missing in the measurement and the only way to record the full 3D information is to record additional diffraction patterns for different orientations of the object relative to the incident beam. Rotating the sample rotates its Fourier spectrum such that it sweeps across the Ewald sphere. From the point of view of the object, the measurements at different orientations can be thought of as orienting the Ewald sphere at different rotations (about the point $\mathbf{q} = 0$) so that it samples the 3D intensities $|\tilde{\rho}(\mathbf{q})|^2$. An example of 3D diffraction data collected at a synchrotron beamline is shown in Fig. 2.

Since the sample is completely destroyed by the interaction with a single high-irradiance FEL pulse, one object can only give rise to a single measurement (or possibly multiple simultaneous measurements). Full 3D information therefore requires combining many diffraction patterns, each from a different orientation of an identical copy of the object. Combining data from many patterns is also needed to increase the overall signal. A calculated single molecule diffraction pattern is shown in Fig. 3 for an incident pulse fluence of 10^{12} photons focused to a 0.3-μm diameter spot at 8 keV photon energy (10^6 J/cm^2), corresponding to an irradiance of 3×10^{19} W/cm^2 with a pulse duration of 30 fs. The pattern is noisy because of low photon counts (most detector pixel values are zero or one photon count). Therefore, even with the extremely high irradiances from an X-ray FEL, some averaging of the signal from many particles is required to increase the signal to noise ratio even at a single orientation. The photon count per pixel of solid-angle Ω_p per shot, averaged over shells of Q, for biological material can be estimated by

$$\langle I(Q) \rangle_Q = I_0 \, \Omega_p \, P \, r_e^2 \, \langle |\tilde{\rho}(Q)|^2 \rangle_Q = I_0 \, \Omega_p \, P \, r_e^2 \, N_{\text{atom}} \, |f(Q)|^2, \tag{8}$$

where N_{atom} are the number of atoms in the molecule, and f is an average atomic scattering factor (e.g. close to that of carbon). Full diffraction information requires sampling at the Shannon rate Δq_S as described in Sec. 2.1 on p. 7.9. In this case we have $\Omega_p \approx (\lambda/4\pi)^2 \Delta q_S^2 = (\lambda/4w)^2$
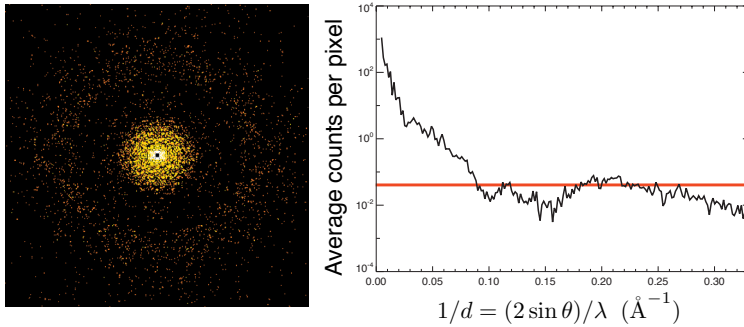
Fig. 3: Calculated diffraction pattern of a single cow pea mosaic virus (CPMV) particle for 10^{12} incident photons focused to a 0.3-μm diameter spot, at $8\,\text{keV}$ photon energy.

for a particle of width w, or

$$\langle I(Q)\rangle_Q = I_0\,P\,r_e^2\,\frac{N_{\text{atom}}\lambda^2}{16w^2}\,|f(Q)|^2. \tag{9}$$

Equations (8) and (9) assume that the positions of atoms are completely uncorrelated, which is approximately true at resolutions approaching the atomic scale. They give an estimate of average counts per pixel per particle for a single shot. For the CPMV particle, of width $w = 31\,\text{nm}$, we obtain $\langle I(Q)\rangle = 0.04$ photons per Shannon pixel, indicated by the red line in the plot of Fig. 3. We see from Eqn. (9) that there is a large incentive to perform measurements at as long a wavelength as possible that can support the desired resolution, and that the scattering signal is not strongly dependent on the size of the object, since $N_{\text{atom}} \propto w^3$. With this model we find that biological particles of around 10 nm diameter require a pulse fluence of 10^{13} photons/$(0.1\mu\text{m})^2$ for an average of 0.1 photon per Shannon pixel at a resolution of $3\,\text{Å}$ and a photon energy of $8\,\text{keV}$. For a pulse duration of 30 fs this requires the exceptionally high irradiance of $4 \times 10^{21}\,\text{W/cm}^2$, the effect of which is examined in Sec. 3. At $3\,\text{keV}$ photon energy, the required irradiance is reduced to $2 \times 10^{20}\,\text{W/cm}^2$, although the increased photoabsorption cross section at this photon energy leads to faster destruction of the object (requiring even shorter pulses, below 10 fs). The signal level of 0.1 photon per Shannon pixel is higher than required by the averaging and assembly procedures discussed in Sec. 4.

2.1   The Phase Problem

The reconstruction of a 2D or 3D image of the electron density $\rho(\mathbf{x})$ requires the inversion of Eqn. (7). While the modulus of the Fourier amplitudes can be obtained from $\sqrt{I(\mathbf{q})}$, the phases are missing. That is, we know the strength $|\tilde{\rho}|$ of each volume grating (of a specified period and direction given by $\mathbf{q}$) in the image, but not how these gratings are shifted with respect to each other. If the phases are known then the synthesized image is a coherent sum of these properly-positioned frequency components, which is simply the inverse Fourier transform of $\tilde{\rho}(\mathbf{q})$:

$$i(\mathbf{x}) = \mathcal{F}^{-1}\{\tilde{\rho}(\mathbf{q})\} \equiv \int \tilde{\rho}(\mathbf{q})\exp(-i\mathbf{q}\cdot\mathbf{x})\,d\mathbf{q}. \tag{10}$$

In the case of reconstructing an image from a single diffraction pattern, this image synthesis is the same as would be carried out by a perfect lens. The lens of course avoids the need to retrieve phases, since it acts upon the far-field wavefield incident on its pupil, and in a microscope the image intensity is directly measured, not the far-field pattern. The resolution of the lens-based image is equivalent to that retrieved from the diffraction pattern if the acceptance of the lens is the same as the angular extent of the recorded diffraction pattern. For example, a diffraction pattern that extends to a scattering angle of 2θ will resolve spatial periods in the object as small as $d = \lambda/(2\sin\theta)$. Similarly the lens-based image (of a coherently illuminated object) has a resolution commonly expressed as $d = \lambda/\mathrm{NA} = \lambda/\sin(2\theta)$, where NA is the numerical aperture of the lens.

Note that just having a detector of large angular extent does not necessarily guarantee a high-resolution image. The high-angle diffraction data of course must be detected above noise. The same is indeed true for a lens-based microscope, where the resolution of the image might not fulfill that expected by the numerical aperture of the lens. An estimation of the resolution in an X-ray or electron microscope is obtained by examining the Fourier spectrum of the recorded image. For our "lensless" imaging method we can look directly at the recorded diffraction pattern, although the resolution of the image will depend upon the reliability of the phases obtained. Methods of estimating this from the reproducibility of retrieved phases have been developed [17], but this does not necessarily mean the phases are the correct ones. Perhaps the most satisfactory way of estimating resolution is based on comparing images (or diffraction phases) retrieved in separate measurements, as is common in crystallography and single-particle electron microscopy.

The coherent X-ray diffraction data in Fig. 2 contains $512 \times 512 \times 512$ pixels, or measurements of Fourier amplitudes, $\sqrt{I(\mathbf{q})}$. Inversion to an image in this case requires the recovery of over 10^8 phases. The retrieval of these phases depends upon the observation that the number of independent measurements may exceed the number of degrees of freedom describing the image. A much simpler example is the diffraction pattern of a pair of point scatterers, separated by a distance w, and measured to a resolution $Q_{\max} = 2\pi/d$. In one dimension there are only four values to describe this object: the distance between the points, the moduli of the scattering strengths of both points, and the relative phase between them. From Eqns. (4) and (7) the diffraction pattern is a fringe pattern

$$I(\mathbf{Q}) \propto |\tilde{\rho}(\mathbf{Q})|^2 = \left|a_1 e^{i\phi_1} + a_2 e^{i\phi_2} \exp(i\mathbf{w}\cdot\mathbf{Q})\right|^2 = a_1^2 + a_2^2 + 2a_1 a_2 \cos(wQ - \phi_1 + \phi_2), \quad (11)$$

which has a period $Q_p = 2\pi/w$ and a contrast $(I_{\max} - I_{\min})/(I_{\max} + I_{\min}) = 2a_1 a_2/(a_1^2 + a_2^2)$. The fringe pattern is also shifted from the origin by the phase difference $\phi_1 - \phi_2$. The measurements of these quantities, and the overall strength of the diffraction pattern is enough to retrieve the amplitudes of the scatterers although there is an ambiguity in the sign of the phase difference and in which point has amplitude a_1. In this example, the number of measurements matches the number of degrees of freedom in the image (ignoring the trivial differences just mentioned). Consider now an object consisting of three points in a row, each separated by w. This will give diffraction fringes spaced by $2\pi/w$ due to interference between the scattering from the extreme points, as for the two-point case, and also at half that period, π/w, from interference of scattering from neighboring points. In this case it is not possible to uniquely assign the strengths and phases of the points. For example, the two objects $O_1 = \{1, 4, 4\}$ and $O_2 = \{2, 5, 2\}$ have equal diffraction patterns, where the three values are the scattering amplitudes of the three equally-spaced points (see Fig.4 (a)). In fact, the 3-point problem generally has two solutions, since
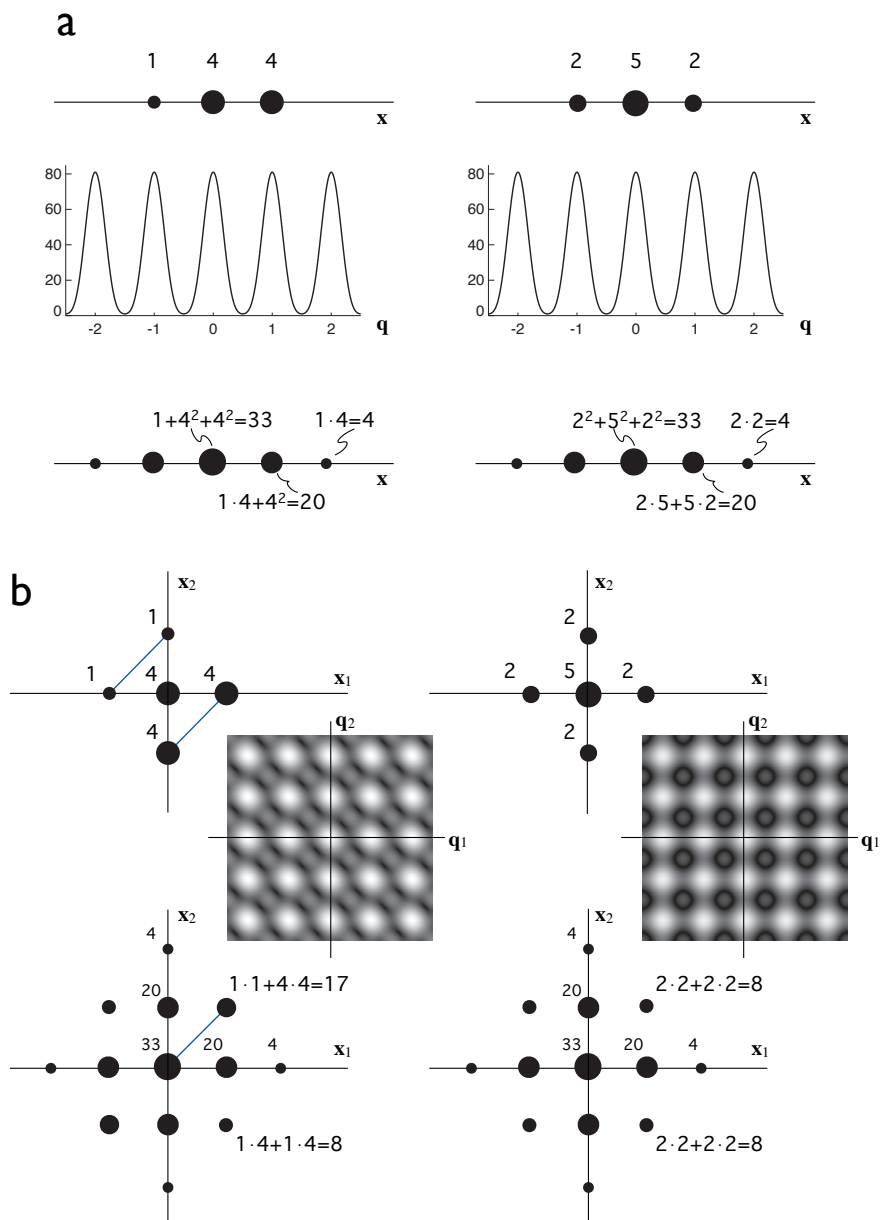
Fig. 4: (a) Two different 3-point objects (top) have the same 1D diffraction pattern (center) and spectrum of diffraction intensities (or autocorrelation) (bottom). (b) This equivalence is broken by interference between objects in different directions, introduced in two or more dimensions.

$\{1, a + b, ab\}$ and $\{a, 1 + ab, b\}$ generate the same pattern. Therefore, image reconstruction is not generally unique in one dimension. In two dimensions it has been shown that the solution is most likely unique, except for rare and special cases [18]. This is essentially due to the fact that for each line in a 2D image there are orthogonal line images that constrain the image values. For example 5-point 2D objects constructed from the sequences O_1 or O_2 can be distinguished from each other through the difference in strength of the diagonal fringes in the patterns as shown in Fig. 4 (b). While it may be possible to construct two or more 5-point 2D objects with the same pattern, these are unusual cases. Image reconstruction is even more robust in three dimensions.

The frequencies in the intensity pattern are represented by *its* Fourier spectrum $\tilde{I}(\mathbf{x})$. This real-space function is proportional to $\mathcal{F}^{-1}\{|\tilde{\rho}(\mathbf{q})|^2\} = \rho(\mathbf{x}) \otimes \rho^*(-\mathbf{x})$, the autocorrelation of the electron density (as obtained from object spatial frequencies on the Ewald sphere in the case of a single coherent diffraction pattern). We saw in the case of the two-point object that the diffraction pattern consists of just two frequencies: the zero frequency (a dc term) and one of period $2\pi/w$. $\tilde{I}(\mathbf{x})$ gives the phase and amplitude of these frequencies, as well as the negative frequency which adds no further information since $\tilde{I}(\mathbf{x})$ is Hermitian, $\tilde{I}^*(-\mathbf{x}) = \tilde{I}(\mathbf{x})$. The image retrieval problem is equivalent to finding a compact function $\rho(\mathbf{x})$ from its autocorrelation function. Each sampled $\tilde{I}(\mathbf{x})$ is dependent on all pairs of image points whose positions differ by the vector $\mathbf{x}$.

The simple examples of the point objects illustrates some of the requirements in measuring diffraction data and hints at how the missing phases are encoded in the diffraction pattern. An object of maximum width w will give rise to fringes of finest period $2\pi/w$ in the diffraction pattern. The diffraction pattern is therefore band-limited, and from Shannon's sampling theorem the complete intensity field is determined from greater than two equally-spaced samples for every $2\pi/w$ period: $\Delta q > \Delta q_S = \pi/w$ (where we refer to Δq_S as the Shannon sampling interval) . Collecting samples at finer intervals than π/w does not provide any more information (although in practice there are reasons to do so, discussed below). For a resolution $q_{max} = 2\pi/d$, there are $N_S = 2q_{max}/\Delta q_S = 4w/d$ Shannon samples (if the diffraction pattern is measured from $-q_{max}$ to $+q_{max}$). The Shannon sampling interval in real space is $\Delta x_s = d/2$, and so the minimum real-space field of view that can be retrieved from adequately-sampled diffraction data is $N_S \Delta x_s = 2w$, twice the width of the compact object. The diffraction pattern from a general 1D complex-valued object of width w will consist of $N_S = 4w/d$ independent measurements, for a resolution of d. The number of degrees of freedom in the complex-valued image is $2w/\Delta x_s$ (a real and imaginary value per each of the $w/\Delta x_s$ real-space samples). This is exactly equal to the number of Shannon samples in the pattern, which is a necessary condition to retrieve a unique image (although, as demonstrated above, not a sufficient condition). In 2D, an object of square support of width w gives rise to N_S^2 independent measurements in its diffraction pattern, needed to recover $2(w/\Delta x_s)^2 = N_S^2/2$ image coefficients. In 3D there are four times as many independent measurements as unknowns. Another way of viewing this is that the region in real space that is constrained becomes a larger fraction of the sampled space as the number of dimensions increases: outside the support we know that the image amplitude is zero.

The ratio Ω, of the number of independent measurements to the number of independent image coefficients, appears to be a good indicator of the ability to reconstruct the image [19]. The shape of the object's support can increase this overdetermination ratio, with non-convex shapes and those without centrosymmetry leading to higher values . For example a triangle support

gives $\Omega = 3$ in 2D, compared with $\Omega = 2$ for a square. Objects with well-separated components are generally easier to reconstruct. A 2D object support consisting of two squares of width w with their center points separated by $> 2w$ has $\Omega = 3$ since the autocorrelation support consists of three squares of width $2w$. The ratio Ω is increased further when when one of the components is smaller than the other, approaching the case of Fourier transform holography as one of the components approaches a delta function (see [20]). The autocorrelation vectors of magnitude greater than the width of either of the components must be due to pairs of image points in separated components, which further constrains the reconstruction.

Note that a crystal consisting of identical objects arranged in a cubic lattice of width w gives rise to Bragg peaks separated by $\Delta q_B = 2\pi/w$, which is twice the Shannon sampling rate Δq_S (the underlying molecular transform is under sampled in each dimension by a factor of 2 relative to the Shannon rate). In this case the number of unknowns to be retrieved is twice the number of independent measurements in 3D, $\Omega = 1/2$, which explains why solving the phase problem for crystals is more difficult than for non-periodic objects. Non-crystallographic symmetry (where two or more copies of a molecule occur in the crystal which are not related by the symmetry of the lattice) and changing the unit cell parameters by swelling are two methods to increase the number of independent measurements to improve phasing efforts. In many cases the molecule does not fill the entire volume of the unit cell and the smooth solvent region reduces the number of coefficients to recover. In large complexes this solvent volume can reach as high as 80%, which would increase Ω from $1/2$ to 1.25 (if the support of the molecule was known). In other instances, some proportion of the molecular structure may be known (e.g. an unknown drug fragment bound to a known target), which may significantly change the balance of known parameters to retrieved phases.

2.2   Iterative Phase Retrieval Algorithms

Reconstructing an image from a coherent diffraction pattern of an isolated object requires retrieving the diffraction phases, utilizing the constraint that the real-space image $\rho(\mathbf{x})$ is zero outside its support, plus any other constraints that can be applied. Iterative algorithms successively apply operators with the aim to converge to the image which is contained within the support and whose Fourier spectrum matches the measurement.

The simplest iterative scheme is Fienup's Error Reduction (ER) algorithm [11], which begins from the square root of the measured diffraction intensities and a random guess of the phases: $\tilde{\rho}'_1(\mathbf{q}) = \sqrt{I}(\mathbf{q})\exp\{i\phi(\mathbf{q})\}$. The image $\rho'_1(\mathbf{x})$ is then formed by an inverse Fourier transformation. This image will not be zero everywhere outside the actual support S of the object. The estimate $\rho_2(\mathbf{x})$ is obtained by setting $\rho'_1(\mathbf{x})$ to zero outside S. This is transformed to $\tilde{\rho}_2(\mathbf{q})$ which will no longer be in agreement with the measured Fourier amplitudes. An update is formed by setting $\tilde{\rho}'_2(\mathbf{q}) = \sqrt{I}(\mathbf{q})\exp\{i\phi_2(\mathbf{q})\}$, where $\phi_2(\mathbf{q})$ are the phases from $\tilde{\rho}_2(\mathbf{q})$. This procedure is then iterated, ideally until both sets of constraints are satisfied.

The image can be represented by a vector in an N-dimensional vector space, where N is the number of pixels (or voxels) in the image. The complex amplitude at each pixel gives the value of the vector along each corresponding dimension. In this representation the ER algorithm can be written as

$$\rho_{n+1}(\mathbf{x}) = P_S\,P_M\,\rho_n(\mathbf{x}) \tag{12}$$

where the projection operators are given by

$$P_S \, \rho(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}) & \text{if } \mathbf{x} \in S \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

and

$$\tilde{P}_M \, \tilde{\rho}(\mathbf{q}) = \sqrt{\frac{I(\mathbf{q})}{|\tilde{\rho}(\mathbf{q})|^2}} \, \tilde{\rho}(\mathbf{q}) \tag{14}$$

with

$$P_M = \mathcal{F}^{-1} \, \tilde{P}_M \, \mathcal{F}. \tag{15}$$

Projection operators have the property that $P^2 = P$. Orthogonal projection operators furthermore project a vector ρ to the closest point of its subspace, such that the Euclidean distance $\|P\rho - \rho\|$ is minimised, where $\| \cdot \|$ is defined as the sum of the square moduli of the vector components. Error metrics describing how well the modulus and support constraints are satisfied can be expressed as distances $\epsilon_S = \|P_S\rho - \rho\|$ and $\epsilon_M = \|P_M\rho - \rho\|$, respectively. It is clear from these definitions that an iteration of $P_S \, P_M$ will always decrease the errors ϵ_S and ϵ_M, which is why Fienup called this the Error Reduction algorithm. However, this algorithm has no method to escape from local minima, which would require a step that increases these metrics. Numerous algorithms have been proposed to overcome the stagnation problem inherent in the ER algorithm, of which the most popular is the hybrid input-output (HIO) algorithm [11], which can be expressed as

$$\rho_{n+1}(\mathbf{x}) = (P_S \, P_M + (I - P_S)(I - \beta P_M))\rho_n(\mathbf{x}), \tag{16}$$

in the case when a only a support constraint is applied in real space. Here I is the identity operator. Inside the support, where $(I - P_S)\rho = 0$, the modulus constraint is applied as in the ER algorithm. Outside the support, instead of setting ρ to zero to exactly satisfy the support constraint, the iterate ρ_{n+1} is formed by subtracting $\beta P_M \, \rho_n$ from ρ_n, where β is a constant usually in the range 0 to 1. The inspiration for this algorithm comes from control theory, with the idea to provide a negative feedback to the operation of applying the modulus constraint. The input to P_M is compensated at those points where the support constraint is violated. This allows the algorithm to escape local minima. Often several iterations of ER are inter-dispersed between HIO steps. Other algorithms improve the convergence rate by taking bigger steps in image space or altering the search strategy, as explained and reviewed by Marchesini [21]. Reconstruction problems become more feasible when the number of independent measurements can be increased for a given object at a particular imaging resolution (described by a particular number of unknowns). Likewise, improved reconstructions can be obtained by transforming the object into a representation that requires fewer unknowns to describe it [22]. At atomic resolution, a representation of a molecule as a collection of atoms, rather than an array of electron density, may require less parameters to recover, for example.

Iterations proceed until the error metrics converge. The final iterate is not necessarily equal to the solution, $\bar{\rho}$, which is the intersection of the two sets, $\bar{\rho} = P_S\bar{\rho} = P_M\bar{\rho}$ [23]. The solution can be found from $\bar{\rho}(\mathbf{x}) = P_M \, \rho_n(\mathbf{x})$. Due to measurement noise a single true solution cannot be distinguished from a family of images that satisfy all constraints to within the the errors. An average solution can be determined by continuing the iterations and generating solutions say every 100 iterations, or by rerunning the algorithm from random phases [24]. This average is
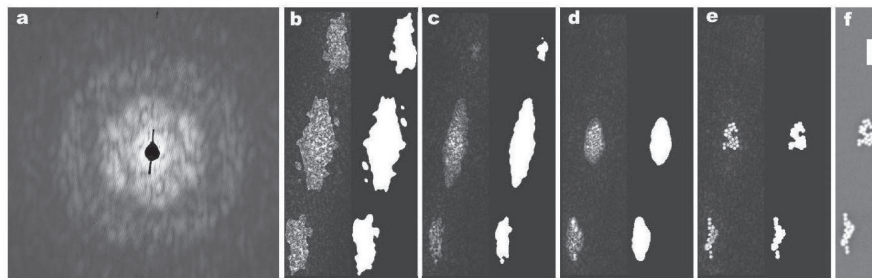
Fig. 5: Image reconstruction from the coherent diffraction pattern in (a). The process starts from zero phases, giving the autocorrelation of the image, in (b). The first support is obtained by a threshold of this image. As phase-retrieval iterations proceed the support is re-estimated, as shown in (c) to (e). An SEM image of the object is in (f). From [8].

unique, and is perhaps the best estimate of the true image that could be obtained from the data. The procedure also allows us to determine the reliability of the retrieved phases. Diffraction phases that always reconstruct to the same value will add complex amplitudes constructively, whereas the sum of amplitudes with random phases will tend to zero. By comparing the modulus of this average to the square root of the measured diffraction intensities we obtain a value at each diffraction pixel that is < 1 and only equal to 1 when phases are exactly consistent. This value tends to decrease with increasing resolution due to the fact that diffraction amplitudes decrease with resolution and are more influenced by noise. This ratio indicates how well the spatial frequencies of the image are represented in the phasing process and is accordingly referred to as the phase retrieval transfer function (PRTF). The average solution can be thought of as the true solution imaged through an optical system with a transfer function given by the PRTF. When performing the average the constant phase term must be normalized to the same value for each solution, otherwise the PRTF will be < 1 for the zero frequency (representing an attenuating optical system). Other low-order aberrations including wavefront tilt (displacement of the image), defocus, and Seidel aberrations such as astigmatism and coma, can also be removed from each solution before determining $\langle \bar{\rho} \rangle$. Since these types of aberrations only cause blurring or shifting of the image, they are not fully constrained by a loose support, unless a positivity constraint is applied (the out of focus image of a positive-density object will have negative amplitudes) [25].

When imaging at hard X-ray wavelengths away from atomic resonances, or imaging binary objects at any wavelength, a positivity constraint can be applied to the real-space image values. With soft X-rays (photon energies below about 1 keV) most interesting objects are complex valued. In this case, a successful reconstruction requires a support constraint that closely matches the actual boundary of the object, since with a loose support an out-of-focus image would be a valid solution. However, it is unlikely that the support is known to such accuracy. Marchesini noticed that the tighter the support constraint the better the reconstruction, leading him to propose the Shrinkwrap method [8]. As illustrated in Fig. 5 this adds a dynamic support constraint to an iterative transform algorithm, where the support is updated occasionally based on the current reconstruction. Image pixels below a certain threshold of a blurred version of the current reconstruction are treated as being outside the object. Starting from the support of

the image autocorrelation, the support tends to gradually shrink to the boundary of the object, improving the reconstruction as it does so (which then gives an improved estimate of the support). A stopping criterion must be added otherwise the method tends to over shrink; when this happens the error ϵ_S increases abruptly, allowing this transition to be determined.

In many experimental situations, especially single-shot imaging at X-ray free-electron lasers, there are missing data at low scattering angles near the zero-order beam. Unless the sample has been designed to be mostly absorbing (as can be achieved in imaging magnetic domains in thin films, for example) the zero-order beam will be extremely intense, and is usually managed with a beam stop or a hole in the detector. If this missing region is as small as a single speckle, only the overall scattering strength of the object is lost. As this region increases more information is lost, often causing iterative algorithms to fail. When only a support constraint is applied, there will be particular modes that are neither constrained by the diffraction measurement nor the support [24]. However, it is possible to reconstruct images missing these low frequencies, just as a lens with an annular pupil will give rise to high-pass filtered images [26].

2.3 Coherence and Detector Requirements

Reconstructing images of non-periodic objects from their X-ray diffraction patterns has often been found to be more problematic than experienced in computer simulations or visible light experiments. This may be due to the fact that most X-ray experiments have been carried out with partially-coherent beams. The van Cittert-Zernike theorem states that the mutual coherence function of a beam propagating from an incoherent source is given by the Fourier transform of the source intensity distribution. Application of this theorem to an object illuminated by this partially-coherent beam shows that its far-field diffraction pattern is given by

$$I(\mathbf{q}) = |\tilde{\rho}(\mathbf{q})|^2 \otimes p(\mathbf{q}) \tag{17}$$

where p is the spatial distribution of the source, such that its angular extent is given by $p(4\pi\theta/\lambda)$. More generally, in the Gaussian-Schell model the source can described in terms of mutually incoherent modes and the diffraction pattern is then the sum of intensity patterns arising from each mode. The image reconstruction as described in the previous section assumes full coherence, with $p(\mathbf{q}) = \delta(\mathbf{q})$. As can be seen from Eqn. (17) the effect of the source is to reduce the contrast of the diffraction pattern, and in particular causing the zeroes in the intensity pattern to have positive values. This causes problems in assigning phases to the values $\sqrt{I(\mathbf{q})}$ since there will be a discontinuity from negative to positive diffraction amplitudes.

The effect of the finite width of detector pixels is also a convolution, given by Eqn. (17), where $p(\mathbf{q})$ is now the pixel sensitivity. The decrease in fringe contrast in the diffraction pattern causes a modulation of the autocorrelation by the function $\tilde{p}(\mathbf{x})$, the modulation transfer function of the detector or the mutual coherence function of the source . The width of this function (i.e. the transverse coherence length) must be significantly wider than the width of the object autocorrelation, or twice the largest diameter of the object. For example, a square-pixel detector that is fully sensitive across its area (which is approximately the case for CCD detectors) will not detect the finest fringes in the pattern when the pixel spacing matches the Shannon sampling rate. That is because a pixel integrates over half a period of these finest fringes, which will be exactly equal to the integration over the other half period in its neighboring pixel. Improved estimates of the Shannon samples are obtained by increasing the pixel density.

Often the function p can be estimated or measured so that the diffraction pattern contrast can be corrected. This deconvolution is carried out by dividing the autocorrelation $\tilde{I}(\mathbf{q})$ by the MTF $\tilde{p}(\mathbf{q})$, a high-pass filtering operation. Since this deconvolution procedure can amplify noise, following Whitehead *et al.* [27] we could propose to *convolve* the current iterate with p before comparing with the measurement to update the modulus of $\tilde{\rho}$, by replacing Eqn. (14) with

$$\tilde{P}_M \, \tilde{\rho}(\mathbf{q}) = \sqrt{\frac{I(\mathbf{q})}{|\tilde{\rho}(\mathbf{q})|^2 \otimes p(\mathbf{q})}} \, \tilde{\rho}(\mathbf{q}). \tag{18}$$

In this way the moduli are updated by first computing the partially-coherent diffraction pattern that would arise from the current estimate of $\rho(\mathbf{x})$, and comparing this with the measurement $I(\mathbf{q})$. Equation (18) is not a projection operator, since repeated application may continue to change ρ. The modal method for reconstruction of images from partially-coherent diffraction patterns derived by Whitehead *et al.* [27] has nevertheless been found to substantially improve the success of image reconstructions using experimental data, even for X-ray FEL experiments with beams that are almost fully spatially coherent. When the width of p is small enough that the highest period fringes discerned in the pattern are due to the most distant inter-object separation, then even rough estimates of p lead to improved reconstructions.

The assumption of quasi-monochromaticity inherent in the discussion up until now requires a bandwidth spread $\Delta\lambda/\lambda < 1/N_S$. This condition can be relaxed if the spectrum is known such that the polychromatic diffraction pattern due to the current estimate of $\rho(\mathbf{x})$ can be calculated [28], in a similar treatment to partial spatial coherence. Additionally, the framework of partial coherence can be applied to changes in the sample itself, either due to X-ray induced damage [29] (and see below) or sample motion [30].

3 Outrunning Radiation Damage

3.1 Diffraction Before Destruction

As mentioned in the Introduction, recording diffraction from non-crystalline material requires a vastly greater exposure than required for crystalline material. The limitation to acquiring high-resolution diffraction information is due to radiation damage: the very radiation used to determine structure impairs that structure. The dose (energy per unit mass) that cryogenically cooled biological material can withstand is about $30\,\mathrm{MGy}$ before changes are observed in the molecular constituents of the sample. By using pulses shorter than the timescale of radiation damage processes, we can literally outrun damage and increase the dose by many orders of magnitude [14] . With short-pulse X-ray FEL radiation at high irradiance (10^{18} to 10^{21} W/cm^2 or doses up to $100\,\mathrm{GGy}$ at $8\,\mathrm{keV}$ photon energy) the high degree of ionization of every atom in the sample will cause an isolated object to Coulomb explode, yet the inertia of atoms provides on the order of $10\,\mathrm{fs}$ before atomic displacements will significantly alter the diffraction pattern at atomic resolutions. The electron density of ions is of course lower than neutral atoms, and this ionization will also give rise to modification to the pattern even before atomic motion occurs.

This method of "diffraction before destruction" was first demonstrated at the FLASH FEL at soft X-ray wavelengths [31], and has been verified to hold at atomic resolution [32]. For samples composed primarily of light elements of similar atomic weight, photoabsorption will occur

largely randomly throughout the sample. This random sequence of ionization will on average cause an uncorrelated (that is, Q-independent—at least at low resolution) addition of diffuse diffraction and reduce the overall contrast. However, when heavier atoms, such as iron, are present in the sample photoabsorption will not be isotropically distributed over the sample. In this case the non-uniformity of this absorption of energy can lead to motions and ionisation of particular atoms that are coordinated across many unit cells of a crystal, or are repeatable from shot to shot. Diffraction measurements with pulses of 80 fs at photon energies both above and below the Fe absorption edge of ferredoxin crystals at intensities of about about 10^{19} W/cm^2, indeed revealed such changes [33]. If large enough, the ability to make intensity-dependent changes to specific parts of the structure (like heavy atoms) gives a new strategy to phasing crystal diffraction [34, 35], similar to the traditional method of isomorphic replacement. Calculations also indicate that phase shifts induced in waves scattering from heavy atoms near atomic resonances still may occur under highly-ionising conditions. Despite large degrees of ionisation, opportunities exist for phasing using an extension of the method of anomalous diffraction [34].

3.2 Diffraction Termination

Molecular dynamics calculations can give detailed simulations of the X-ray induced explosion that take into account the initial molecular structure of samples. These calculations are computationally expensive, and so are limited to small samples such as single macromolecules or small crystals. A continuum approach applies methods developed to model dense plasmas such as in stellar interiors (somewhat larger than a molecule), but makes the assumption that in small regions the dynamics of the sample are isotropic. A code called Cretin was used to simulate the structural changes at atomic resolution in small protein crystals of about 1 μm diameter, embedded in a water tamper [15]. The calculations make discrete time steps in which the atomic populations and electron and photon distributions are calculated utilizing known transition rates and opacities. These provide electron and ion temperatures, ionization states and ion collisions, from which mean atomic displacements are computed, as shown in Fig. 6. The RMS displacements are obtained from a diffusion equation where the diffusion coefficient is calculated from the ion collision frequencies and temperatures. As seen in the figure the RMS atomic displacement σ increases approximately as $t^{3/2}$ over the time t of a constant-irradiance pulse. The explosion occurs faster for higher pulse irradiance, with σ approximately proportional to the square root of pulse irradiance (photons/unit area/unit time) for the range of irradiance considered here. At 8 keV photon energies, this dependence continues beyond an irradiance of 10^{21} W/cm^2 which corresponds to focusing 10^{13} photons in 10 fs to a spot of 0.4 μm diameter. At this photon energy and irradiance, an RMS displacement of 1 Å is reached in about 10 fs.

Each X-ray induced explosion of an identical object will be different due to the random sequence of ionization and atomic displacements. The diffraction signal obtained by averaging over many instances of these explosions can be obtained through the derivation, found in several text books (see e.g. [36]), for describing small random displacements in crystals due to thermal motion. It is interesting to note that this common derivation makes no explicit assumption as to whether the almost-identical objects are arranged and exposed together (as in a crystal) or exposed in a serial fashion and diffraction intensities then summed. This is the case if there is no spatial correlation to the displacements and hence no correlation of these displacements from unit cell to unit cell (if considering a crystal), and that the RMS displacement within
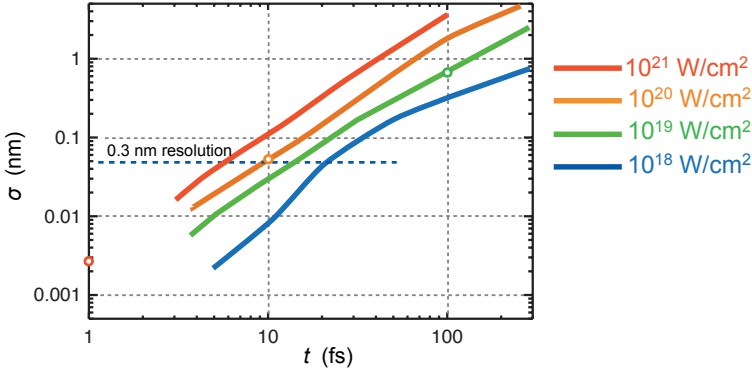
Fig. 6: Plot of the RMS isotropic atomic displacement σ as a function of time t during pulses of constant irradiance as shown in the legend, for a photon energy of 8 keV. The open circles indicate the time required to reach a fluence of 10^6 J/cm^2 as in the simulation of Fig. 3, showing that higher pulse power (photons per unit time) gives less damage for a given total scattering signal. Lower resolution can tolerate longer (and therefore less intense) pulses. From [15].

a single object is the same as the RMS displacement in the whole crystal. We assume that diffraction patterns are properly oriented prior to averaging, by the methods of Sec. 4. The effect of isotropic displacements is to multiply the diffraction signal at a momentum transfer Q by a factor of $\exp(-Q^2\sigma^2)$. Here σ is the RMS of the component of the displacement in the direction of $\mathbf{Q}$, equal to the 1D component of the mean displacements for isotropic displacements, as plotted in Fig. 6. In our accelerating explosion, σ increases during the exposure. The instantaneous diffraction signal at a resolution d is reduced by $1/e$ when σ reaches a value of $d/(2\pi)$. Displacements of only 0.5 Å terminate the accumulation of intensities at 3 Å resolution. The summed diffraction signal from recorded patterns is the pulse-integrated signal. Following Eqn. (7) this diffraction signal for a constant irradiance pulse of duration T is given by

$$I(\mathbf{Q}) = I_0\, T\, \Omega_p\, P\, r_e^2\, \left\{ |\tilde{\rho}(\mathbf{Q})|^2\, g(Q;T) + N_{\text{atom}} |f|^2 (1 - g(Q;T)) \right\}, \tag{19}$$

$$g(Q;T) = \frac{1}{T} \int_0^T \exp\{-Q^2\sigma^2(t)\}\, dt. \tag{20}$$

Here we have omitted another multiplicative Wilson factor of $\exp(-BQ^2/8\pi^2)$ to describe the structure variation inherent in the unexposed samples, and note that the total incident fluence is given by $I_0\, T$ (in units of photons per unit area).

Equation (19) consists of two terms. The first is the undamaged diffraction pattern $|\tilde{\rho}(\mathbf{Q})|^2$ modified by the dynamic disorder factor $g(Q;T)$, a function that decreases monotonically with increasing Q. The second term is a slowly-varying background of scattered counts that is proportional to the overall scattering strength of the sample and which increases monotonically with increasing Q. This term is 0 at $Q = 0$. Thus, in general, the disordering takes scattered counts out of the information-containing Shannon samples into a background that diminishes the contrast. This occurs first at high Q and works its way towards lower resolution with time. Applying the empirical time dependence observed in Fig. 6 of $\sigma(t) = \sigma_T \cdot (t/T)^{3/2}$, with σ_T

the RMS displacement at the end of the pulse, we find

$$g(Q;T) = \int_0^1 \exp(-Q^2\sigma_T^2 t'^3)dt' = \frac{\Gamma(4/3)}{(Q^2\sigma_T^2)^{1/3}} - \frac{1}{3}E_{2/3}(Q^2\sigma_T^2) \tag{21}$$

where $E_{2/3}(x)$ is the exponential integral function of order $2/3$, and which tends to zero as the argument x increases, and $\Gamma(4/3) = 0.89$. That is, for high resolution $g(Q;T) \approx (Q\,\sigma_T)^{-2/3}$.

The behavior of g can perhaps be better understood from Eqn. (20). At early times in the pulse $\sigma(t)$ is small and the instantaneous disorder factor (the integrand) is close to unity and the diffraction pattern, proportional to $|\tilde\rho(\mathbf{Q})|^2$, continues to accumulate. When σ reaches a value of $1/Q$ this accumulation stops and the background then accumulates. This happens at a time t_{off} such that $1/Q = \sigma_T(t_{\text{off}}/T)^{3/2}$, or $t_{\text{off}} = (Q\,\sigma_T)^{-2/3}T$. The proportion of the pulse that contributes to the cooperative diffraction is given by t_{off}/T, in agreement with the limiting value of the integration of Eqn. (21). Pulses longer than t_{off} do not contribute any more information at resolutions of Q or better, and simply accumulate background for the rest of the pulse.

Heavy atoms, such as metals, in the molecule tend to have higher photoabsorption cross sections than the majority lighter elements, and hence lead to a greater degree of ionisation than most of the molecule. This situation is by no means random and isotropic as has been assumed in the discussion of this section. One may initially conclude that the heavy atoms are ionised more quickly than the surroundings and hence their scattering strength vanishes more quickly than the surroundings. Experiments suggest another scenario, which is that the heavily ionized atom draws in charge from the surrounding atoms, recharging that atom so that it can continue to efficiently ionise and yet remain at a high enough electron density that it continues to efficiently scatter [37]. In this case the structure surrounding the heavy atom might melt away during the pulse, leaving the heavy atom to scatter disproportionally into the diffraction pattern. In either case (atoms disappearing more quickly than the surroundings or lasting longer than the surroundings) there is, in addition to the disorder caused by the explosion, a correlated change in the structure over the course of the pulse. This leads to a modification of Eqn. 19, by the modification of the average electron density $\rho(\mathbf{r})$ used to compute its Fourier transform $\tilde\rho(\mathbf{Q})$, an incoherent addition of the diffraction of the density excess (depending on the time that it persists), and an incoherent sum of a cross-term consisting of the interference between excess density and the average unmodified diffraction. Such a sum of diffraction intensities can not be represented as a Fourier transform of a single structure, nor can it be accurately modelled by assuming a change in occupancy of particular atoms. Thus, using conventional structure refinement procedures, this will give errors in in the recovered electron density.

Our analysis suggests that the source metric to maximize signal acquisition and accuracy is not the total pulse fluence, but the pulse irradiance (or source power, considering that the spot size is essentially dictated by focusing optics). If we simply increase the pulse energy (number of photons) by proportionally increasing the pulse length, then we only increase the background without improving the signal. Increasing the irradiance increases the rate of signal photons arriving on the detector but does shorten the time t_{off} that these photons accumulate. However, given the linear dependence of σ_T^2 on pulse irradiance observed in Fig. 6 we find that the total signal scales as $I_0\,t_{\text{off}} \propto I_0^{2/3}$. However, if irradiance is scaled higher than $10^{21}\,\text{W/cm}^2$ (at $8\,\text{keV}$ photon energy) we can expect almost every electron to be stripped from every atom in the sample. In this case the turn-off time of diffraction will not be necessarily limited by atomic motion, but the variability in the atomic scattering factors at undisplaced atomic positions. This has no spatial correlation, and leads to a diffuse background that is independent of
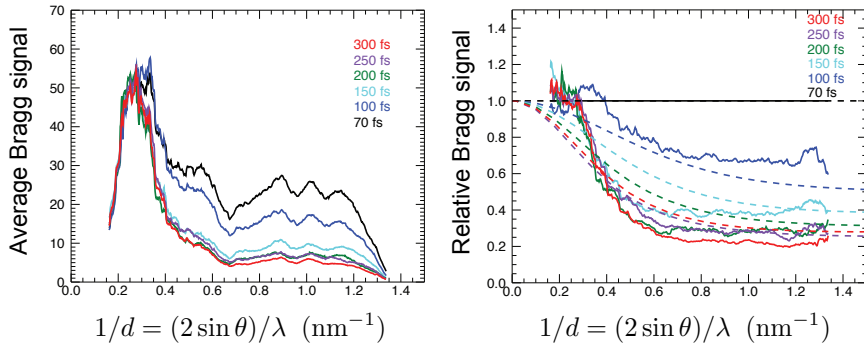
Fig. 7: (left) Average Bragg signal of Photosystem I micro crystals as a function of resolution for different pulse durations for pulse irradiance up to $10^{17}\,\mathrm{W/cm^2}$ at a photon energy of 2 keV. (right) The ratio of the average Bragg signal to that at the shortest pulse duration, and comparison to ratios of the dynamic disorder factor $g(Q;T)$ obtained from continuum modeling of the evolution of the RMS atomic displacements as shown in Fig. 6. From [15].

Q and a corresponding uniform decrease in the signal. In this case, our Q-independent turn-off time will depend on the atomic relaxation processes, which are not much longer than several femtoseconds.

Among the first experiments to be carried out at the LCLS, the world's first hard X-ray FEL [1], was to measure the effect of pulse duration and fluence on coherent diffraction patterns. The samples chosen for confirming "diffraction before destruction" and mapping out the behavior of diffraction termination were protein nanocrystals flowing in a water jet. These are biological material and are confined in a water medium, as per the system modeled in Fig. 6. The crystallinity gives a great advantage that the diffraction signal $|\tilde{\rho}(\mathbf{Q})|^2$ is confined to Bragg peaks, whereas the background due to the explosive disordering varies very slowly across the detector. The presence of discernible Bragg peaks indicates that structure at a particular length scale persisted for some fraction of the pulse. Indeed, the measurement of the Bragg intensities as a function of pulse duration allowed us to determine this fraction, and hence quantify t_{off} and σ_T. For even small protein nanocrystals, and pulses not very much longer than t_{off}, the disorder background is much weaker than the Bragg peaks. Therefore, for nanocrystallography it is not necessary to limit pulse durations to t_{off}. The Bragg peaks terminate when the disorder reaches a level of $\sigma = d/(2\pi)$ for a resolution length d. The effect of the explosion is a filtering of the Bragg values by the dynamic disorder factor $g(Q;T)$, which can be corrected by dividing the Bragg signal by an estimate of this function based on modeling of the explosion or from measurements at several pulse durations. In addition, the method of nanocrystallography opens up protein structure determination to samples that cannot be grown into crystals of sufficient size or quality for standard synchrotron radiation measurements or are particularly sensitive to radiation damage, making this a very attractive and important method for biological structure determination [32].

Figure 7 shows measurements of the termination of Bragg peaks as a function of resolution for different pulse durations, carried out at a photon energy of 2 keV (6 Å wavelength) on nano and

microcrystals of Photosystem I. This is a large membrane complex involved in photosynthesis, and chosen for those experiments primarily because of its large unit cell and because it had proven to be a very difficult sample for conventional synchrotron measurements which required crystals several hundred microns large. As can be seen from Fig. 7 the diffraction efficiency was very much lower at high resolution for longer pulses compared with the shorter pulses, simply because only the first part of the pulse contributed to the diffraction. Even so, the diffraction patterns with 300-fs pulses gave Bragg peaks, at 1-nm resolution, that were easily measured and well above background. The model predictions contained no fitted parameters, and the agreement with the experimental results gives confidence of the predictions for atomic-resolution imaging.

The large proportion of light elements in the Photosystem I complex may be responsible for the observed isotropic disorder (of ionisation and atomic displacements). However, crystals of smaller molecules, and specifically those with heavy atoms have shown different behaviour. Ferredoxin molecules, mentioned above, contain two clusters that each consist of a structure of four Fe and four S atoms. Experiments with high-intensity X-ray FEL pulses of 80 fs duration, reveal changes to the refined electron density of these clusters [33]. Molecular dynamics simulations of the experiment predict a coordinated and reproducible motion of the S atoms away from the heavier Fe atoms during the pulse [38]. Such effects will cause modifications to Eqn. 19 as discussed above in Sec. 3.2. These experiments show that while such effects seem strong enough to offer ways to help phase crystal diffraction [34], more experiments and modelling are required to gain a complete understanding of the process and use this to achieve more accurate molecular structures by the technique of diffraction before destruction.

4 Diffraction Data from Identical Particles

As we have seen, the intense and extremely brief pulses from X-ray FELs may overcome problems of radiation damage, but a single object only survives a single pulse. In the scattering geometries discussed here, this limits information collected in a single pulse to a two-dimensional slice (on the Ewald sphere) through the three-dimensional diffraction intensities $|\tilde{\rho}(\mathbf{x})|^2$. The acquisition of full three-dimensional information requires combining data from many copies of identical objects in different orientations. Summing data from many objects also increases the signal beyond that available in a single-shot pattern, which is limited by the number of incident photons per pulse as estimated by Eqn. 9. There are several ways in principle to acquire data from multiple particles so that scattering information can be combined into a single 3D dataset. We consider here methods in which the orientation of the particle to the frame of the laboratory is not known *a priori*, referred to by Elser as *cryptotomography* [39].

4.1 Serial Nanocrystallography

The most familiar and most successful method for increasing the diffraction signal of macromolecules and their assemblies is by forming crystals of these samples . Every copy in the crystal is oriented with respect to all others, and these are arranged in a 3D lattice which gives constructive addition of the scattering in the particular locations of Bragg peaks. Since the real-space electron density of the crystal is a convolution of the repeated motif with a lattice, the
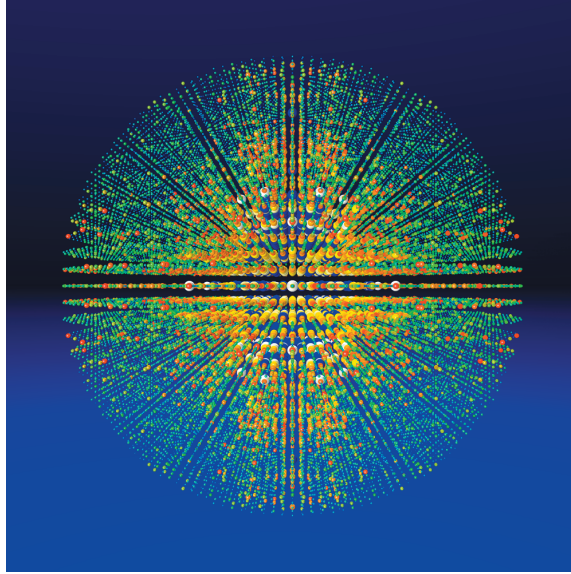
Fig. 8: The assembled diffraction data from 56,000 nanocrystal diffraction patterns of Photosystem I collected at LCLS at a photon energy of 2 keV. Figure by Thomas White, CFEL DESY.

Fourier representation $\tilde{\rho}(\mathbf{q})$ is the transform of the motif modulated by the Fourier transform of the lattice. The total integrated signal in each Bragg peak is N_{mol} times greater than that of the single particle, where N_{mol} is the total number of repeats in the crystal, and furthermore this signal is concentrated into a Fourier-space volume much smaller than the Shannon volume. For short X-ray FEL pulses, the crystalline diffraction (from structure that is correlated from unit cell to unit cell) will always give much higher counts per pixel than the disordered background. This background (which is also proportional to the total number of atoms in the crystal—see Eqn. (19)) uniformly fills the detector area between Bragg peaks whereas the peaks themselves are concentrated into approximately $1/N_{mol}^{2/3}$ of the area between peaks. That is, a $10 \times 10 \times 10$ crystal will have peaks about 100 times above background when $g = 1/2$. The well-known disadvantage of crystals, mentioned in Sec. 2.1, is that the Bragg-peak spacing Δq_B is twice the Shannon spacing Δq_S for an object with the same width as the real-space lattice period, leading to fewer independent measurements than image coefficients.

In serial crystallography [32], crystals of a few micrometer or smaller are delivered to the X-ray beam in a jet of a water. A crystal is hit by chance and is situated in a random and unknown orientation. The determination of this orientation can be readily made from the pattern itself by indexing the Bragg peaks [40]. Several so-called auto-indexing programs have been developed for crystallography which search for a repeating lattice in the measured diffraction pattern, knowing the mapping of that pattern onto the Ewald sphere. In the simple example of a monoclinic crystal with different unit cell spacings in each dimension, the reciprocal lattice basis vectors can be identified quite readily from the observed reciprocal lattice spacings and arrangement. Each Bragg peak in the pattern can thereby be indexed by its 3D Miller index $\mathbf{h}$ and thereby properly accumulated in a running sum of $|\tilde{\rho}(\mathbf{h})|^2$. In this way the diffraction
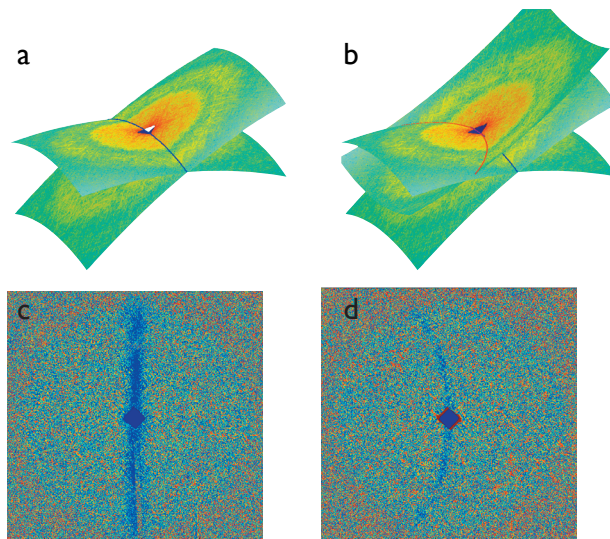
Fig. 9: Identification of common arcs of intersection in experimental X-ray diffraction patterns taken from the same 3D nano object at two orientations. The object and diffraction data are the same as in Fig. 2. (c) shows the difference of diffraction intensities, $I_1(\mathbf{Q}) - I_2(\mathbf{Q})$ and (d) shows $I_1(\mathbf{Q}) - I_2(-\mathbf{Q})$. Figure by Gösta Huldt, Uppsala University.

data are built up, averaging over all crystal shapes and sizes, just as in the case for X-ray powder diffraction. The difference in this case is that the summation is done in three dimensions instead of the single magnitude of Q, giving us complete 3D information as if collecting data from a rotating single undamaged crystal. An example of this 3D assembly is shown in Fig. 8. One popular software pipeline for serial crystallography, developed over the last decade, utilises the program OnDa [41] for on-line data monitoring and analysis, Cheetah [42] for processing data after collection (for example to extract the detector frames that actually contain crystal diffraction), and CrystFEL [43, 44] to aggregate the data into a set of structure factors.

Serial crystallography was originally developed as a way to investigate radiation damage (as discussed in Sec. 3.2) and as a path to single-molecule diffraction (Sec. 4.2) by reducing the size of crystals [45]. Along this path there are many fascinating opportunities. Serial crystallography has enabled time-resolved measurements of protein structures at atomic resolution and timescales below 1 ps, to reveal the isomerisation that occurs to the chromophore of a photoactive protein after absorption of a visible-light photon [6]. Single-fibre diffraction, in which single fibres are measured in a serial fashion, allows 3D structure factors to be obtained by determining the orientation of each fibre from its weak diffraction pattern [46]. Likewise, structures have be obtained from two-dimensional macromolecular crystals, containing about 1000 molecules per exposure [47].
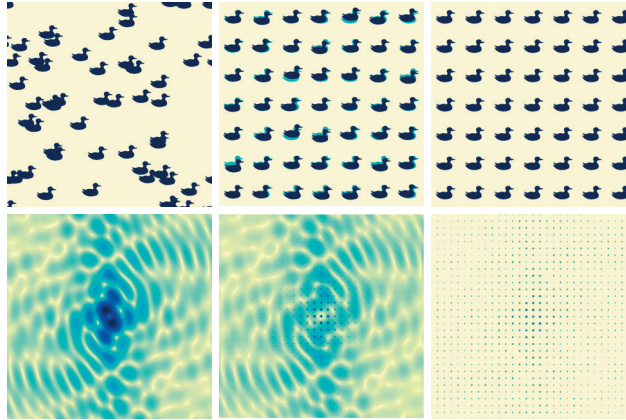
Fig. 10: Diffraction from an ensemble of similarly oriented objects depends on correlations between their positions. (left) A random arrangement of objects gives rise to the incoherent sum of the continuous diffraction pattern of each object. (middle) A crystal with a degree of translational disorder consists of Bragg peaks formed from the coherent sum of diffraction from all objects, modulated by a DebyeWaller factor that describes the suppression of Bragg peaks at resolutions greater than the disorder length divided by 2π. At those resolutions the incoherent sum of single-object diffraction occurs. (right) A perfect crystal produces solely the coherent sum of diffraction from the periodic array of scatterers.

4.2   Single-Particle Diffraction

Single-particle X-ray diffraction aims to carry out the same acquisition and analysis steps as for serial nanocrystallography, without the advantage of a lattice to guide us to the orientation, and without the addition of scattering from many unit cells that is concentrated into narrow Bragg peaks. It is the special case of the single-unit cell crystal. It is of particular interest because it could be applied to any series of identical objects that cannot be crystallized and allows measurement of all independent Shannon samples for the application of the phase-retrieval methods described in Sec. 2.2. The orientation analysis must be carried out on patterns with much lower signal than a single photon per pixel, as is the case for pulse energies that can be generated at X-ray FELs, such as shown in Fig. 3.

One method for determining the relative orientation of patterns is by searching for common arcs of intersection. Since every diffraction pattern samples information on an Ewald sphere that passes through the origin of Fourier space, any two patterns will always intersect along an arc that also passes through the origin. If $\rho(\mathbf{x})$ is real valued, as it is away from atomic resonances and for hard X-ray wavelengths, then $|\tilde{\rho}(\mathbf{q})|^2$ is centrosymmetric, and there will be another arc of intersection when forming the centrosymmetric inversion of one of the patterns. (Note that the diffraction pattern itself will not have centrosymmetry because the Ewald sphere is not centrosymmetric.) These two intersections allow the identification of the relative 3D orientation of the patterns. The identification of the two arcs does work even if the object density is not strictly real-valued, as shown in Fig. 9. Identifying the arcs of intersection requires a minimum signal to noise level in the patterns, which could be built up by first classifying patterns into

groups of similar enough orientation that they can be summed together without knowing their relationship to other patterns. This is the approach common in the similar method in cryo-electron microscopy, where noisy real-space images are first classified into groups of similar images and then the class averages are oriented. It has been found that when working directly with diffraction amplitudes, however, that the process of correlation and orientation can be merged into one process. Recent computational experiments [48] suggest that the correlations between pairs of patterns can be made on arcs, from which a consistent set of orientations for all patterns can then be found. After arbitrarily fixing the orientation of one pattern, $N-1$ estimates of the orientation of each of the remaining $N-1$ patterns can be obtained. This procedure makes use of all correlations between pairs of patterns.

Other analysis methods for cryptotomography include the expansion-maximization-compression (EMC) framework and topological mapping [49]. The EMC method seeks to build up a model $W(\mathbf{q})$ of $|\tilde{\rho}(\mathbf{q})|^2$ by placing the two dimensional diffraction pattern measurements over a distribution of their most likely orientations. The method is iterative, starting from unknown orientation assignments. In each iteration, the model W is expanded into diffraction patterns that would arise from such an object over a finite set of predetermined uniformly distributed orientations. The maximum likelihood step follows in which the model is changed to best agree with each of the model tomograms. In the final compression step the model tomograms are made consistent with a single updated model $W'(\mathbf{q})$. This method has successfully been applied to reconstruct the relatively simple 3D diffraction data of ellipsoidal iron oxide nano-particles measured at the FLASH FEL [50]. Elser has determined the lowest signal limits required to carry out such a procedure and finds that objects can be recovered at counts considerably less than a single photon per pixel per pattern [39]. A method that appears conceptually different (but is related [51]) is a topographic mapping approach. Utilizing the concept described in Sec. 2.2 of a diffraction pattern as a vector in a finite-dimensional vector space, it is noted that the diffraction patterns (or images) obtained at different orientations must map out a continuous 3D manifold in that higher-dimensional space. Using generative topographic mapping methods, in which this manifold is mapped out from noisy measurements, Fung et al. [52] recovered the structure of a molecule from simulated patterns with less than 0.04 counts per pixel.

4.3   Diffraction from Crystals with Translational Disorder

As compared with single-molecule diffraction, crystallisation provides a huge amplification of the molecular diffraction at the cost of loss of information. However, there are other ways to design experiments that may give sufficiently sampled diffraction stronger than the faint signals obtainable from a single molecule. For example, a random, rather than periodic, arrangement of molecules will give rise to an incoherent sum of the continuous single-molecule diffraction, or an amplification equal to the number of illuminated molecules. This requires that the molecules are oriented in one or very few directions, since otherwise only an orientational average will be obtained. Alignment of molecules in the gas phase can be achieved using a laser pulse [53], as demonstrated in aligned-molecule diffraction experiments at the LCLS [54]. Since the alignment axis is fixed in the laboratory frame, and a molecular beam refreshes the sample on every pulse, data can be accumulated over many pulses until the required signal level is reached, before changing the angle to obtain 3D diffraction data in a rotation series.

As it turns out, a crystal is actually a very good starting point to achieve an ensemble of aligned
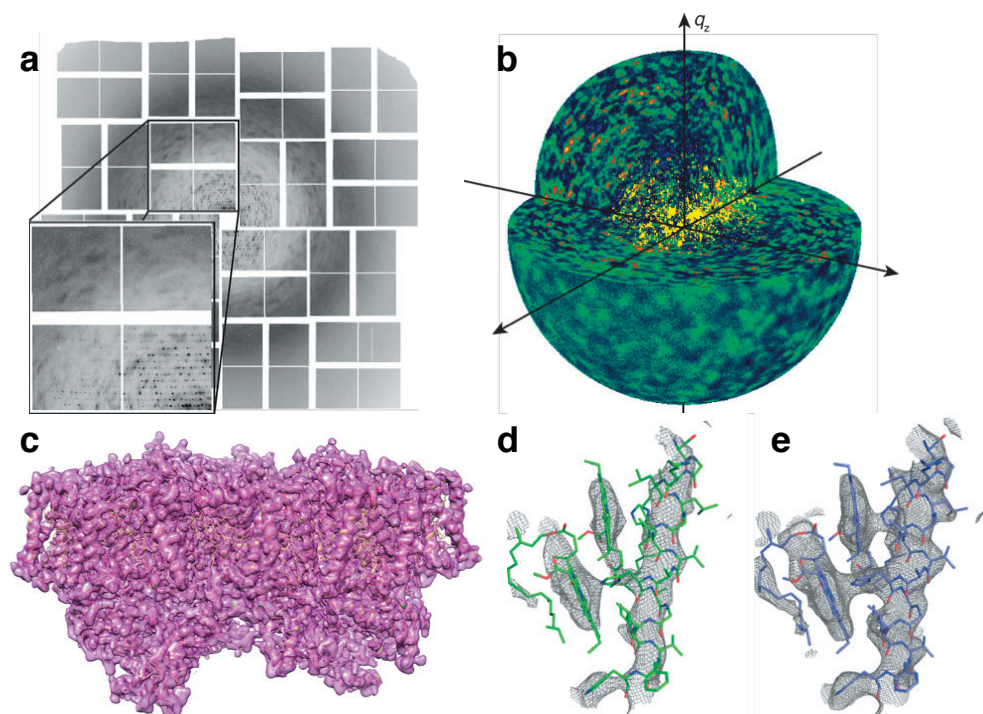
Fig. 11: Weak continuous diffraction (a) was observed in individual snapshot diffraction patterns recorded from Photosystem II crystals at the LCLS [55]. A merge of 2885 patterns in 3D reciprocal space, shows continuous diffraction beyond the Bragg peaks (b). (c) The continuous diffraction was phased using an iterative phasing algorithm to obtain a 3D image of the electron density of the Photosystem II dimer. A detail of two chlorophylls of the dimer shows the improvement obtained from performing a structural refinement using the Bragg data only (to a resolution of 4.5 Å) (d) as compared with using together the Bragg and continuous diffraction (to a resolution of 3.5 Å) (e). Reprinted from ref. [55]

molecules, in a more natural environment than the gas phase. As discussed in Sec. 3.2, random displacements of rigid objects from ideal lattice positions, with a mean square displacement σ^2 in each direction, will give rise to the modulation of Bragg intensities by the DebyeWaller factor $\exp(-Q^2\sigma^2)$. Now, as opposed to the situation of Eqn. 19, it is the entire macromolecule that moves as a rigid unit instead of a single atom. Therefore, the scattering that occurs at scattering angles beyond the Bragg peaks, given by f in Eqn. 19, is the transform of the entire rigid object—the macromolecule—as depicted in Fig. 10. Such disorder is not unexpected in protein crystals. Large macromolecules form crystals with only tenuous intermolecular contacts with solvent filling the interstitial spaces. Disorder may therefore be dominated by random translations as would be the case in the orderdisorder phase transition from a crystal to a nematic liquid crystal. Recently, continuous diffraction was observed from crystals of Photosystem II extending in resolution to 3.5 Å, beyond the 4.5 Å resolution of observable Bragg peaks [55] (see Fig. 11). It was found that the Fourier transform of this continuous diffraction yielded an autocorrelation map consistent with that of the Photosystem II dimer complex, supporting the hypothesis of the dimer as the rigid unit in a translationally-disordered lattice. The continuous diffraction could be phased by an iterative transform algorithm, using a fixed support volume as a real-space constraint, created by blurring out an initial electron density map obtained by refining a model using the Bragg intensities. The 3D image obtained by the phasing the continuous diffraction showed much clearer definition of structural elements such as helices, even though the phasing algorithm had no knowledge of such structures. A potentially fruitful approach to crystallography may be to develop methods to induce translational disorder in otherwise well-ordered crystals.

5   Outlook

There are now five hard-X-ray FELs operating in the world, giving many opportunities to explore the possibilities and methodologies of imaging with brilliant coherent X-ray pulses. There is still much development needed before single-particle structure determination can be carried out routinely, but we now have strong foundations for these endeavors, as described here. All the steps for structure determination have been elucidated theoretically with experimental confirmation. The compelling idea that the extreme irradiance of X-ray FEL pulses would allow us to obtain more information from a sample than possible at conventional sources, by virtue of outrunning the explosion caused by that high irradiance, has been examined in detail over a range of photon wavelengths, pulse durations, and fluences. These measurements give confidence that "diffraction before destruction" is valid to atomic resolution. We have found that nanocrystal diffraction is particularly impervious to the effect of the X-ray induced explosion since the diffraction from the periodic undamaged component of the structure is easily discriminated from the scattering of the disordered component. Our model of diffraction termination shows that the metric to optimize sources for imaging is pulse power. Today, the LCLS produces pulses above 100 GW and designs have been proposed that could produce pulses beyond 1 TW. [56, 57].

It is straightforward to calculate the diffraction pattern of a molecular assembly (from Eqn. (7)), showing that signals at atomic resolution will invariably be at extremely low photon counts and approaching as many photons in the diffraction pattern as there are atoms in the sample. Theoretical tools have been developed to combine data from serial measurements of identical

objects to build up the three-dimensional Fourier spectrum of the average object. The experimental challenges lay in being able to actually record such weak diffraction patterns without the introduction of extraneous noise sources. Is it really possible to illuminate a macromolecular object with 10^{13} photons and yet ensure that the single photons measured in the detector pixels have scattered from that object, and are not due to scattering from optical elements, gas molecules, or other means? Once the data assembly has been accomplished, the reconstruction of a three-dimensional object is ensured by the phase retrieval algorithms that have been adequately demonstrated in many contexts. Real experimental effects such as wavelength bandwidth, spatial coherence, and detector response can all be characterized and accounted for. The phase retrieval of diffraction data sample at or beyond the Shannon rate gives an overdetermined dataset in two and three dimensions, and *ab initio* image retrieval is possible. That is, the phase problem is more easily solved than in the case of crystals, where over 50 years of insights and breakthroughs can additionally be drawn upon or adapted.

These ideas and underpinnings have already given us a new route to obtain high-resolution room-temperature structural information from protein nanocrystals that are too small to be analyzed at conventional sources, and a way to obtain high-resolution electron density maps from the diffraction of disordered crystals. The short pulse duration from X-ray FELs gives inherently high temporal resolution, which can be exploited in pump-probe experiments where a sample is stimulated by a short laser pulse at a precise time before the arrival of the X-ray pulse. The femtosecond pulses are over 1000 times shorter than synchrotron pulses, giving access to a full exploration of the motions involved in chemical reactions, molecular vibrations, and laser-matter interactions. The serial method can be applied to measurement of irreversible reactions. Since the sample is destroyed by the pulse anyway, there is no requirement to bring it back to the ground state as is the case for stroboscopic measurements. These experimental techniques are continuing to evolve as the availability of X-ray FEL beamlines and instrumentation increases, providing a very bright future for X-ray imaging, in more ways than one.

References

[1] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F. J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H. D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky, J. Galayda, *First lasing and operation of an ångstrom-wavelength free-electron laser*, Nature Photon. **4**, 641 (2010).

[2] M. Yabashi, H. Tanaka, T. Ishikawa, *Overview of the SACLA facility*, J. Synchr. Rad. **22**, 477 (2015).

[3] M. Altarelli, R. Brinkmann, M. Chergui, W. Decking, B. Dobson, S. Düsterer, G. Grübel, W. Graeff, H. Graafsma, J. Hajdu, J. Marangos, J. Pflüger, H. Redlin, D. Riley, I. Robinson, J. Rossbach, A. Schwarz, K. Tiedtke, T. Tschentscher, I. Vartaniants, H. Wabnitz, H. Weise, R. Wichmann, K. Witte, A. Wolf, M. Wulff, M. Yurkov, *The European X-Ray Free-Electron Laser Technical design report*, Technical report, DESY XFEL Project Team (2007).

[4] E. A. Seddon, J. A. Clarke, D. J. Dunning, C. Masciovecchio, C. J. Milne, F. Parmigiani, D. Rugg, J. C. H. Spence, N. R. Thompson, K. Ueda, S. M. Vinko, J. S. Wark, W. Wurth, *Short-wavelength free-electron laser sources and science: a review*, Rep. Prog. Phys. **80**, 115901 (2017).

[5] R. Mankowsky, A. Subedi, M. Först, S. O. Mariager, M. Chollet, H. T. Lemke, J. S. Robinson, J. M. Glownia, M. P. Minitti, A. Frano, M. Fechner, N. A. Spaldin, T. Loew, B. Keimer, A. Georges, A. Cavalleri, *Nonlinear lattice dynamics as a basis for enhanced superconductivity in $YBa_2Cu_3O_{6.5}$*, Nature **516**, 71 EP (2014).

[6] K. Pande, C. D. M. Hutchison, G. Groenhof, A. Aquila, J. S. Robinson, J. Tenboer, S. Basu, S. Boutet, D. P. DePonte, M. Liang, T. A. White, N. A. Zatsepin, O. Yefanov, D. Morozov, D. Oberthuer, C. Gati, G. Subramanian, D. James, Y. Zhao, J. Koralek, J. Brayshaw, C. Kupitz, C. Conrad, S. Roy-Chowdhury, J. D. Coe, M. Metz, P. L. Xavier, T. D. Grant, J. E. Koglin, G. Ketawala, R. Fromme, V. Šrajer, R. Henning, J. C. H. Spence, A. Ourmazd, P. Schwander, U. Weierstall, M. Frank, P. Fromme, A. Barty, H. N. Chapman, K. Moffat, J. J. van Thor, M. Schmidt, *Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein*, Science **352**, 725 (2016).

[7] D. Sayre, H. N. Chapman, J. Miao, *On the Extendibility of X-ray Crystallography to Non-crystals*, Acta Cryst. A **54**, 232 (1998).

[8] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, J. C. H. Spence, *X-ray image reconstruction from a diffraction pattern alone*, Phys. Rev. B **68**, 140101 (2003).

[9] H. N. Chapman, A. Barty, S. Marchesini, A. Noy, S. P. Hau-Riege, C. Cui, M. R. Howells, R. Rosen, H. He, J. C. H. Spence, U. Weierstall, T. Beetz, C. Jacobsen, D. Shapiro, *High-resolution ab initio three-dimensional x-ray diffraction microscopy*, J. Opt. Soc. Am. A **23**, 1179 (2006).

[10] R. W. Gerchberg, O. Saxton, *Practical algorithm for determination of phase from image and diffraction plane pictures*, Optik **35**, 237 (1972).

[11] J. R. Fienup, *Reconstruction of an object from the modulus of its Fourier transform*, Opt. Lett. **3**, 27 (1978).

[12] J. M. Holton, *A beginner's guide to radiation damage*, J. Synchr. Rad. **16**, 133 (2009).

[13] M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. M. Holton, C. J. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D. A. Shapiro, J. C. H. Spence, D. Starodub, *An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy*, J. Electron Spectrosc. Rel. Phenom. **170**, 4 (2009).

[14] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, *Potential for biomolecular imaging with femtosecond X-ray pulses*, Nature **406**, 753 (2000).

[15] A. Barty, C. Caleman, A. Aquila, N. Timneanu, L. Lomb, T. A. White, J. Andreasson, D. Arnlund, S. Bajt, T. R. M. Barends, M. Barthelmess, M. J. Bogan, C. Bostedt, J. D. Bozek, R. Coffee, N. Coppola, J. Davidsson, D. P. DePonte, R. B. Doak, T. Ekeberg, V. Elser, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, P. Fromme, H. Graafsma, L. Gumprecht, J. Hajdu, C. Y. Hampton, R. Hartmann, A. Hartmann, G. Hauser, H. Hirsemann, P. Holl, M. S. Hunter, L. Johansson, S. Kassemeyer, N. Kimmel, R. A. Kirian, M. Liang, F. R. N. C. Maia, E. Malmerberg, S. Marchesini, A. V. Martin, K. Nass, R. Neutze, C. Reich, D. Rolles, B. Rudek, A. Rudenko, H. Scott, I. Schlichting, J. Schulz, M. M. Seibert, R. L. Shoeman, R. G. Sierra, H. Soltau, J. C. H. Spence, F. Stellato, S. Stern, L. Struder, J. Ullrich, X. Wang, G. Weidenspointner, U. Weierstall, C. B. Wunderer, H. N. Chapman, *Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements*, Nature Photon. **6**, 35 (2012).

[16] H. N. Chapman, C. Caleman, N. Timneanu, *Diffraction before destruction*, Phil. Trans. R. Soc. Lond. B **369**, 20130313 (2014).

[17] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neimann, D. Sayre, *Biological imaging by soft x-ray diffraction microscopy*, Proc. Nat. Acad. Sci. **102**, 15343 (2005).

[18] R. H. T. Bates, *Fourier phase problems are uniquely solvable in more than one dimension: 1. Underlying theory*, Optik **61**, 247 (1982).

[19] V. Elser, R. P. Millane, *Reconstruction of an object from its symmetry-averaged diffraction pattern*, Acta Cryst. A **64**, 273 (2008).

[20] I. McNulty, J. Kirz, C. Jacobsen, E. H. Anderson, M. R. Howells, D. P. Kern, *High-Resolution Imaging by Fourier Transform X-ray Holography*, Science **256**, 1009 (1992).

[21] S. Marchesini, *A unified evaluation of iterative projection algorithms for phase retrieval*, Rev. Sci. Instrum. **78**, 011301 (2007).

[22] Y. Shechtman, Y. Eldar, O. Cohen, H. Chapman, J. Miao, M. Segev, *Phase Retrieval with Application to Optical Imaging: A contemporary overview*, Signal Proc. Mag., IEEE **32**, 87 (2015).

[23] V. Elser, *Phase retrieval by iterated projections*, J. Opt. Soc. Am. A **20**, 40 (2003).

[24] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, D. Sayre, *Reconstruction of a yeast cell from X-ray diffraction data*, Acta Cryst. A **62**, 248 (2006).

[25] S. Marchesini, H. N. Chapman, A. Barty, C. Cui, M. R. Howells, J. C. H. Spence, U. Weierstall, A. M. Minor, *Phase Aberrations in Diffraction Microscopy*, IPAP Conf. Ser. **7**, 380 (2006).

[26] A. Martin, N. Loh, C. Hampton, R. Sierra, F. Wang, A. Aquila, S. Bajt, M. Barthelmess, C. Bostedt, J. Bozek, N. Coppola, S. Epp, B. Erk, H. Fleckenstein, L. Foucar, M. Frank, H. Graafsma, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Liang, L. Lomb, F. Maia, S. Marchesini, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, J. Schulz, R. Shoeman, H. Soltau, D. Starodub, J. Steinbrener, F. Stellato, L. Strüder, J. Ullrich, G. Weidenspointner, T. White, C. Wunderer, A. Barty, I. Schlichting, M. Bogan, H. Chapman, *Femtosecond dark-field imaging with an X-ray free electron laser*, Opt. Express **20**, 13501 (2012).

[27] L. W. Whitehead, G. J. Williams, H. M. Quiney, D. J. Vine, R. A. Dilanian, S. Flewett, K. A. Nugent, A. G. Peele, E. Balaur, I. McNulty, *Diffractive Imaging Using Partially Coherent X Rays*, Phys. Rev. Lett. **103**, 243902 (2009).

[28] B. Abbey, L. W. Whitehead, H. M. Quiney, D. J. Vine, G. A. Cadenazzi, C. A. Henderson, K. A. Nugent, E. Balaur, C. T. Putkunz, A. G. Peele, W. J., McNultyI., *Lensless imaging using broadband X-ray sources*, Nature Photon. **5**, 420 (2011).

[29] H. M. Quiney, K. A. Nugent, *Biomolecular imaging and electronic damage using X-ray free-electron lasers*, Nature Phys. **7**, 142 (2011).

[30] J. N. Clark, C. T. Putkunz, E. K. Curwood, D. J. Vine, R. Scholten, I. McNulty, K. A. Nugent, A. G. Peele, *Dynamic sample imaging in coherent diffractive imaging*, Opt. Lett. **36**, 1954 (2011).

[31] H. N. Chapman, A. Barty, M. J. Bogan, S. Boutet, M. Frank, S. P. Hau-Riege, S. Marchesini, B. W. Woods, S. Bajt, W. H. Benner, R. A. London, E. Plonjes, M. Kuhlmann, R. Treusch, S. Dusterer, T. Tschentscher, J. R. Schneider, E. Spiller, T. Moller, C. Bostedt, M. Hoener, D. A. Shapiro, K. O. Hodgson, D. van der Spoel, F. Burmeister, M. Bergh, C. Caleman, G. Huldt, M. M. Seibert, F. R. N. C. Maia, R. W. Lee, A. Szoke, N. Timneanu, J. Hajdu, *Femtosecond diffractive imaging with a soft-X-ray free-electron laser*, Nature Phys. **2**, 839 (2006).

[32] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Homke, C. Reich, D. Pietschner, L. Struder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kuhnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y.

Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jonsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.-D. Schroter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, J. C. H. Spence, *Femtosecond X-ray protein nanocrystallography*, Nature **470**, 73 (2011).

[33] K. Nass, L. Foucar, T. R. M. Barends, E. Hartmann, S. Botha, R. L. Shoeman, R. B. Doak, R. Alonso-Mori, A. Aquila, S. Bajt, A. Barty, R. Bean, K. R. Beyerlein, M. Bublitz, N. Drachmann, J. Gregersen, H. O. Jönsson, W. Kabsch, S. Kassemeyer, J. E. Koglin, M. Krumrey, D. Mattle, M. Messerschmidt, P. Nissen, L. Reinhard, O. Sitsel, D. Sokaras, G. J. Williams, S. Hau-Riege, N. Timneanu, C. Caleman, H. N. Chapman, S. Boutet, I. Schlichting, *Indications of radiation damage in ferredoxin microcrystals using high-intensity X-FEL beams*, J. Synchr. Rad. **22**, 225 (2015).

[34] S.-K. Son, H. N. Chapman, R. Santra, *Multiwavelength Anomalous Diffraction at High X-Ray Intensity*, Phys. Rev. Lett. **107**, 218102 (2011).

[35] S.-K. Son, H. N. Chapman, R. Santra, *Determination of multiwavelength anomalous diffraction coefficients at high x-ray intensity*, J. Phys. B: At. Mol. Opt. Phys. **46**, 164015 (2013).

[36] B. E. Warren, *X-ray diffraction* (Dover, 1990).

[37] A. Rudenko, L. Inhester, K. Hanasaki, X. Li, S. J. Robatjazi, B. Erk, R. Boll, K. Toyota, Y. Hao, O. Vendrell, C. Bomme, E. Savelyev, B. Rudek, L. Foucar, S. H. Southworth, C. S. Lehmann, B. Kraessig, T. Marchenko, M. Simon, K. Ueda, K. R. Ferguson, M. Bucher, T. Gorkhover, S. Carron, R. Alonso-Mori, J. E. Koglin, J. Correa, G. J. Williams, S. Boutet, L. Young, C. Bostedt, S. K. Son, R. Santra, D. Rolles, *Femtosecond response of polyatomic molecules to ultra-intense hard X-rays*, Nature **546**, 129 (2017).

[38] S. P. Hau-Riege, B. J. Bennion, *Reproducible radiation-damage processes in proteins irradiated by intense x-ray pulses*, Phys. Rev. E **91**, 022705 (2015).

[39] V. Elser, *Noise limits in the assembly of diffraction data*, arXiv p. 0709.3858 (2007).

[40] R. A. Kirian, T. A. White, J. M. Holton, H. N. Chapman, P. Fromme, A. Barty, L. Lomb, A. Aquila, F. R. N. C. Maia, A. V. Martin, R. Fromme, X. Wang, M. S. Hunter, K. E. Schmidt, J. C. H. Spence, *Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals*, Acta Cryst. A **67**, 131 (2011).

[41] V. Mariani, A. Morgan, C. H. Yoon, T. J. Lane, T. A. White, C. O'Grady, M. Kuhn, S. Aplin, J. Koglin, A. Barty, H. N. Chapman, *OnDA: online data analysis and feedback for serial X-ray imaging*, J. Appl. Cryst. **49**, 1073 (2016).

[42] A. Barty, R. A. Kirian, F. R. N. C. Maia, M. Hantke, C. H. Yoon, T. A. White, H. Chapman, *Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data*, J. Appl. Cryst. **47**, 1118 (2014).

[43] T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty, H. N. Chapman, *CrystFEL: a software suite for snapshot serial crystallography*, J. Appl. Cryst. **45**, 335 (2012).

[44] T. A. White, V. Mariani, W. Brehm, O. Yefanov, A. Barty, K. R. Beyerlein, F. Chervinskii, L. Galli, C. Gati, T. Nakane, A. Tolstikova, K. Yamashita, C. H. Yoon, K. Diederichs, H. N. Chapman, *Recent developments in CrystFEL*, J. Appl. Cryst. **49**, 680 (2016).

[45] H. N. Chapman, *X-Ray Free-Electron Lasers for the Structure and Dynamics of Macromolecules*, Ann. Rev. Biochem. **88**, pMID: 30601681 (2019).

[46] C. Seuring, K. Ayyer, E. Filippaki, M. Barthelmess, J.-N. Longchamp, P. Ringler, T. Pardini, D. H. Wojtas, M. A. Coleman, K. Dörner, S. Fuglerud, G. Hammarin, B. Habenstein, A. E. Langkilde, A. Loquet, A. Meents, R. Riek, H. Stahlberg, S. Boutet, M. S. Hunter, J. Koglin, M. Liang, H. M. Ginn, R. P. Millane, M. Frank, A. Barty, H. N. Chapman, *Femtosecond X-ray coherent diffraction of aligned amyloid fibrils on low background graphene*, Nature Comm. **9**, 1836 (2018).

[47] C. M. Casadei, K. Nass, A. Barty, M. S. Hunter, C. Padeste, C.-J. Tsai, S. Boutet, M. Messerschmidt, L. Sala, G. J. Williams, D. Ozerov, M. Coleman, X.-D. Li, M. Frank, B. Pedrini, *Structure-factor amplitude reconstruction from serial femtosecond crystallography of two-dimensional membrane-protein crystals*, IUCrJ **6** (2019).

[48] G. Bortel, M. Tegze, *Common arc method for diffraction pattern orientation*, Acta Cryst. A **67**, 533 (2011).

[49] K. Ayyer, T.-Y. Lan, V. Elser, N. D. Loh, *Dragonfly: an implementation of the expand–maximize–compress algorithm for single-particle imaging*, J. Appl. Cryst. **49**, 1320 (2016).

[50] N. D. Loh, M. J. Bogan, V. Elser, A. Barty, S. Boutet, S. Bajt, J. Hajdu, T. Ekeberg, F. R. N. C. Maia, J. Schulz, M. M. Seibert, B. Iwan, N. Timneanu, S. Marchesini, I. Schlichting, R. L. Shoeman, L. Lomb, M. Frank, M. Liang, H. N. Chapman, *Cryptotomography: Reconstructing 3D Fourier Intensities from Randomly Oriented Single-Shot Diffraction Patterns*, Phys. Rev. Lett. **104**, 225501 (2010).

[51] B. Moths, A. Ourmazd, *Bayesian algorithms for recovering structure from single-particle diffraction snapshots of unknown orientation: a comparison*, Acta Cryst. A **67**, 481 (2011).

[52] R. Fung, V. Shneerson, D. K. Saldin, A. Ourmazd, *Structure from fleeting illumination of faint spinning objects in flight*, Nature Phys. **5**, 64 (2009).

[53] H. Stapelfeldt, T. Seideman, *Colloquium: Aligning molecules with strong laser pulses*, Rev. Mod. Phys. **75**, 543 (2003).

[54] J. Küpper, S. Stern, L. Holmegaard, F. Filsinger, A. Rouzée, A. Rudenko, P. Johnsson, A. V. Martin, M. Adolph, A. Aquila, S. Bajt, A. Barty, C. Bostedt, J. Bozek, C. Caleman, R. Coffee, N. Coppola, T. Delmas, S. Epp, B. Erk, L. Foucar, T. Gorkhover, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, P. Holl, A. Hömke, N. Kimmel, F. Krasniqi, K.-U. Kühnel, J. Maurer, M. Messerschmidt, R. Moshammer, C. Reich, B. Rudek, R. Santra,

I. Schlichting, C. Schmidt, S. Schorb, J. Schulz, H. Soltau, H. Spence, John C. D. Starodub, L. Strüder, J. Thøgersen, J. Vrakking, Marc J. G. Weidenspointner, T. A. White, C. Wunderer, G. Meijer, J. Ullrich, H. Stapelfeldt, D. Rolles, H. N. Chapman, *X-Ray Diffraction from Isolated and Strongly Aligned Gas-Phase Molecules with a Free-Electron Laser*, Phys. Rev. Lett. **112**, 083002 (2014).

[55] K. Ayyer, O. M. Yefanov, D. Oberthür, S. Roy-Chowdhury, L. Galli, V. Mariani, S. Basu, J. Coe, C. E. Conrad, R. Fromme, A. Schaffer, K. Dörner, D. James, C. Kupitz, M. Metz, G. Nelson, P. L. Xavier, K. R. Beyerlein, M. Schmidt, I. Sarrou, J. C. H. Spence, U. Weierstall, T. A. White, J.-H. Yang, Y. Zhao, M. Liang, A. Aquila, M. S. Hunter, J. S. Robinson, J. E. Koglin, S. Boutet, P. Fromme, A. Barty, H. N. Chapman, *Macromolecular diffractive imaging using imperfect crystals*, Nature **530**, 202 (2016).

[56] G. Geloni, V. Kocharyan, E. Saldin, *A novel self-seeding scheme for hard X-ray FELs*, J. Mod. Opt. **58**, 1391 (2011).

[57] W. Fawley, J. Frisch, Z. Huang, Y. Jiao, H.-D. Nuhn, C. Pellegrini, S. Reiche, J. Wu, *Toward TW-level, hard X-ray pulses at LCLS*, Technical Report TN-11-3, LCLS (2011).

D 8   Time-resolved Techniques (in the X-ray Regime)

S. Techert

FS-Strukturdynamik (bio)chemischer Systeme

Deutsches Elektronen-Synchrotron DESY

Contents

The tutorial is based on an article published in:

Introduction

In a chemical reaction, typical time scales of atomic or molecular motion start from femtoseconds, meaning the millionth of a billionth of a second. Life-relevant motions, however, can be as slow as seconds or even up to minutes' or hours' time scales. The origin of these time scale differences is based on the complexity of the coordination space of a proceeding reaction [1-6].
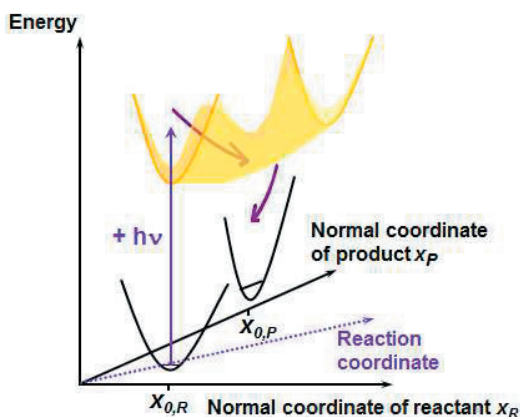
Figure 1: Potential energy hyper-surface of a chemical reaction. The reaction coordinate (abscissa) of a chemical reaction is defined as a one-dimensional projection of the normal coordinates spanning the reactant's potential energy hypersurface versus the normal coordinates spanning the product's potential energy hypersurface, and their corresponding transitions [2,3].

In a classical kinetic scheme, the gradients on the potential energy hypersurface define the molecular dynamics. Statistically population-weighted, they compose to the kinetics of a chemical reaction. The coordinates describing the dynamics and kinetics of a chemical reaction are the reaction coordinate and the energy. The reaction coordinate is defined as a one-dimensional projection of the reactant's and product's normal coordinates, which span the potential energy hypersurface of reactant and product and the potential energy hypersurfaces of their transitions (Figure 1). The energy gradient along a reaction coordinate is defined as reaction dynamics, the energy gradient along the normal coordinates as molecular dynamics.

Commonly, the potential energy is shown in a kinetic curve under the graph, and the axis description of the ordinate presents the sum of the potential energy of the nuclei involved in the chemical reaction and their kinetic energy (Fig. 2). Potential and kinetic energy of molecules can be detangled through the projection of the potential energy onto the total energy axis. The activated complex and transition state (according to Eyring) includes an imaginary mode or an imaginary Eigen value [4-6].

From a chemical physicist's point of view, one would like to understand which elementary chemical processes happen at which time scales, and how these time scales are inter-connected. To what extend do structural motifs "freeze in" time and dynamics information of chemical reactions? Which type of apparatus needs to be built and which kind of methods need to be developed for investigating the created femtosecond "time stamps" in the structure of complex matter during a chemical or biochemical reaction?
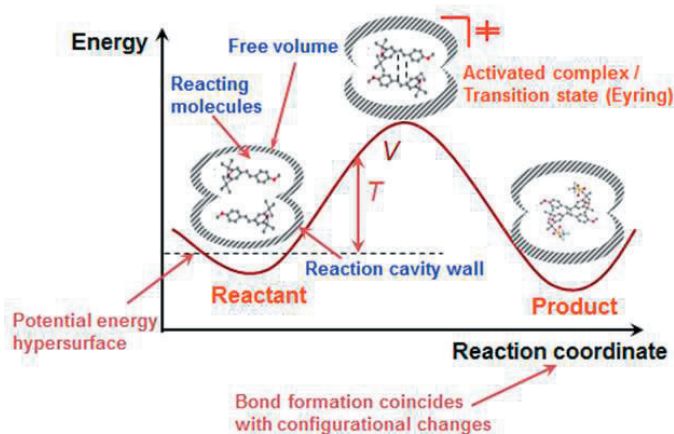
Figure 2: Total energy of a chemical system as the sum of potential energy (V) and kinetic energy (T) of the molecules (ordinate). Inside the graph, only the contribution of the potential energy is plotted; its projection onto the ordinate allows determining the kinetic energy.

Ultrafast X-ray methods bear the potential of determining the complexity of chemical reactions – *during their reactions*, in particular in the bulk, with techniques which utilize the specific characteristics of X-ray/matter interaction: in well-ordered systems, X-ray crystallography as a Thomson scattering process allows for element-specific determinations such as electron densities (redox states) and high-precision spatial resolution determination of atoms in lattices from which hydrogen bonding, chemical bonding or van der Waals stacks can be derived. In less-ordered and disordered systems, X-rays deliver element-specific information of the investigated molecules utilizing, for example, X-ray spectroscopy, X-ray absorption, photo-induced electron cascades, or X-ray and electron emission properties. Site-specific information is obtained by elastic and inelastic scattering processes such as multidimensional X-ray spectroscopy, X-ray diffraction, or X-ray scattering. The chemical consequences of X-ray/matter interactions leading to fragmentation can be characterized by X-ray mass spectrometry.

Characteristic for the X-ray photons generated in synchrotrons and free-electron lasers are their

(i) energy tunability (allowing for excitation-energy-sensitive methods like X-ray spectroscopy or advanced X-ray diffraction methods)
(ii) pulsed structure (allowing for in-situ and time-resolved X-ray methods)
(iii) defined polarization (allowing for advances in X-ray spectroscopy)
(iv) coherence (allowing for X-ray imaging or correlation spectroscopy methods)
(v) high flux (allowing for high-resolution X-ray experiments in all experimental domains)

Fourth generation accelerator-based light sources (free-electron lasers, FELs) in the VUV or X-ray regime deliver ultra-brilliant coherent radiation in very short pulses (10^{12}–10^{13} photons/bunch/10–100 fs). In order to fully exploit their unique photon capabilities, novel instrumentation is required based on single-shot (collection) schemes. Moreover, hundreds up to trillions of fragment particles, ions, electrons or scattered photons can emerge when a single light flash impinges on matter with intensities up to (predicted for the XFEL) 10^{22} W/cm^2. In order to meet these challenges, in the starting time of FLASH (Free-Electron Laser

in Hamburg [6a]) and the LCLS (Linac Coherent Light Source [6b]), various experimental chambers and endstations have been designed.

Starting from basic principles [1-6], we will in the following summarize the FEL methods developed so far, including ultrafast X-ray diffraction and crystallography for condensed state chemistry studies and applications to organic electronics [7-24], for soft condensed matter and protein applications [25-41]; ultrafast soft X-ray spectroscopy and ultrafast 2-dimensional X-ray spectroscopy for bimolecular reaction studies in the liquid phase [42-53]; ultrafast X-ray scattering and ultrafast X-ray emission spectroscopy for the study of photo catalysis [54-64]; and ultrafast photoelectron diffraction and Coulomb explosion schemes for the study of gas-phase reactions and fundamentals of chemical reaction dynamics [65-84].

1   The Concept: Filming Chemical Reactions in Real Time Utilizing Ultrafast High-Flux X-ray Sources

In a proof-of-principle experiment at the white beam beamline at the ESRF (The European Synchrotron Research Facility) in 2001, it has been demonstrated that high-flux, pulsed X-rays – as created with synchrotrons of the 3rd generation – can act as the "photons of choice" for studying the dynamics and kinetics of small chemical systems on their complex reaction landscape [1]. These studies have been used to define various expectation values for time-resolved experiments at free-electron lasers and saddling the ground for ultrafast X-ray experiments at these sources. Since then, also the phrase of "*recording the molecular movie*" has been born (Fig. 3) [2-6b].
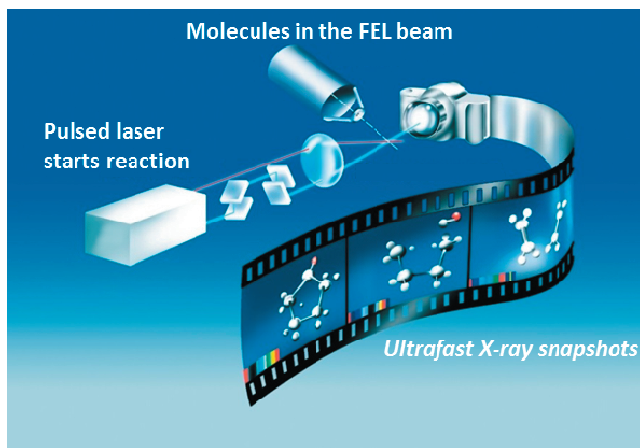
Figure 3: Principle of the "molecular movie". After the initialization of a chemical reaction by a pulsed trigger laser, ultrafast X-ray snapshots and photographs of the X-ray pulses are collected as a function of time (courtesy by DESY, MPIbpC and European XFEL).

Figure 3 summarizes the principles of such a "molecular movie" approach: after the initialization of a chemical reaction with a short laser pulse, ultrafast X-ray FEL snapshots take photographs of the X-ray spectroscopic or X-ray diffraction signal. By varying the time delay between laser pump and X-ray probe pulse, information about the structural changes as a function of time are collected.

Time-wise, the criterion for "recording the molecular movie" is given when the time resolution of the pump and probe sources meet the time scales of the *structural dynamics*

investigated. The resolution criterion for *structural dynamics* studies is fulfilled in chemistry, when the refined structure allows for determining the electron density or the charge densities (which are equal in the redox state) around a moving atom. High-resolution X-ray crystallography studies allow for the study towards such precision [7-24].

2 Crystallography with Ultrahigh Temporal and Spacial Resolution

Figure 4: Evolution of the average brilliance of synchrotron radiation light of synchrotrons of the 2nd generation (DORIS) though the ESRF towards LCLS free electron laser – characterized at the high-resolution diffraction pattern of the same molecular crystal system with same crystal quality.

Crystallography with Ultrahigh Temporal and Spatial Resolution allows to Study Photochemical Reactions beyond Conventional Quantum Chemical Approaches. Far beyond any present laboratory technique, time-resolved synchrotron (picosecond time resolution) and FEL femtosecond time resolution) experiments emphasize the uniqueness of the pulsed, ultrafast, high brilliant and coherent x-ray methods and metrology. For chemical bond breaking and bond formation, the criterion for spatial resolution is met in periodic systems (crystallography) when 0.01 to 0.001 Å resolution diffractograms yield high precision structural information [7-24].

Figure 4 reflects the changes of X-ray synchrotron beam characteristics when evolving from synchrotrons of the 2nd generation towards hard X-ray free electron lasers. The diffractograms have been collected on a molecular crystal of same crystal quality, and same orientation.

Utilizing broadband wiggler radiation in 2nd generation synchrotrons (F1/DORIS), ~~rich~~ Laue diffraction pattern have been collected. Taped undulator radiation of synchrotrons of the third generation yields in quasi Pink-Laue diffractograms (ID09/ESRF). Compared to the Pink Laue white beam at XPP as well as CXI beamline of LCLS, the FEL radiation is about one to two orders of magnitude smaller in bandwidth, allowing only the investigation of a statistical number of Bragg reflections for small molecular crystals when the crystal is rotated.

Compared to the nanocrystallographic approaches also explained in this book, small molecular crystallography at 4th generation synchrotrons could be utilized by the combination of traditional Laue crystallography and FEL specific serial crystallography techniques. Small molecule crystallography in a serial type approach is possible based on a single shot data collection strategy analogous to the time-resolved Laue diffraction.
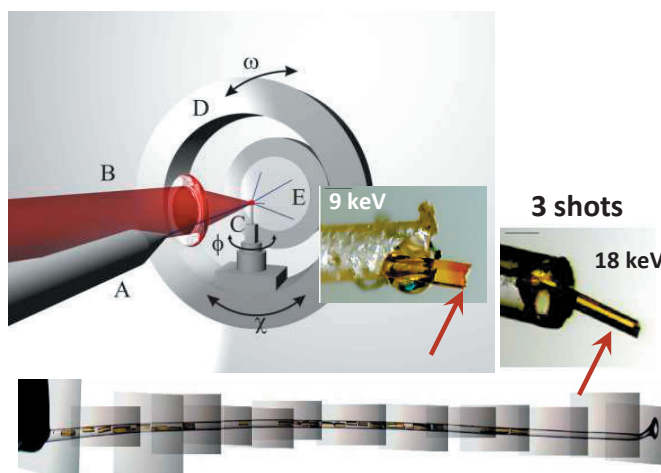
Figure 5: Small molecule crystallography at LCLS, XPP and CXI beamlines (test phase). Left side, top: Used three circle diffractometer. Right side, top: organic crystals exposed to 9 keV X-ray radiation (single shot) inducing severe radiation damage and 18 keV radiation allowing for the collection of various quasi Laue diffractograms with the Laue pink FEL beam under different orientations. Bottom: small molecule crystals lined up for a three-shot-serial type of crystallography experiment.

In contrast to conventional Laue crystallography, for normalization purposes during FEL experiments, every diffractogram is associated to an online collected X-ray spectrum. Utilizing high X-ray energies well above 15 keV (use of the third harmonic and smaller X-ray / atom cross section) with no monochromatization in pink Laue modus reduces radiation damage so that with a monotonically running spindle and randomly changing X-ray wavelength with known X-ray spectral characteristics, various orientations under defined X-ray conditions can be collected. They are sufficient for determining the orientation matrix of small molecules and hence following the indexing of the collected diffractograms.

Since the studied materials are normally compounds of small amounts, highest quality crystals are stacked behind each other in a capillary (or other type of sample target holder) and high energy X-ray radiation is utilized for collecting high resolution diffraction patterns and minimizing accumulative radiation damage. Due to the monochromaticity of FEL X-ray beams even in pink Laue mode, quasi Pink Laue diffraction pattern will be recorded for various orientations – allowing a precise determination of the orientation matrix of the small

molecule crystals. Additionally, on a single shot base, the Bragg peak intensities are wavelength and X-ray intensity normalized.
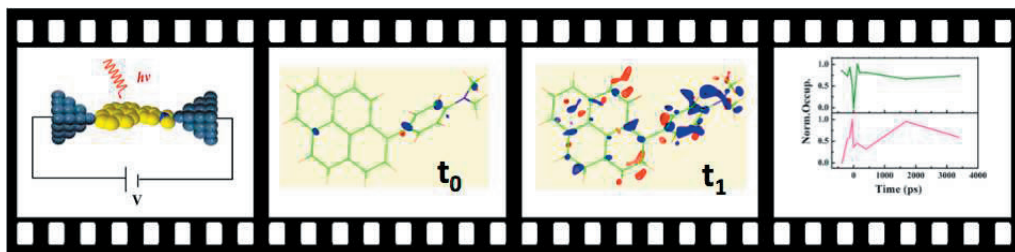
Figure 6: Molecular movie of a non-conventional molecular diode which cannot be described within the Born Oppenheimer approximation. Within femtoseconds and nearly immediately after optical photo-absorption the electrons are rearranged towards a conducting state meanwhile the structure knocks into a tilted configuration.

Figure 6 depicts the refined result of such an experiment- the femtosecond structural dynamics or "molecular movie" of a molecular crystal, which consist of only light elements (carbon, nitrogen and oxygen). The patented system has the most efficient optical light / electron transfer rate possible (100%), by utilizing quantum effects such as electron and structural dynamics pathways which cannot be described through the conventional Born Oppenheimer approximation.

Understanding this "beyond Born Oppenheimer" behaviour and combining the properties of this type of system with smart semiconducting plastic types of compounds, it is possible to build fully flexible solid type of solar cells and organic light emitting diodes with very high efficiency (Figure 7).
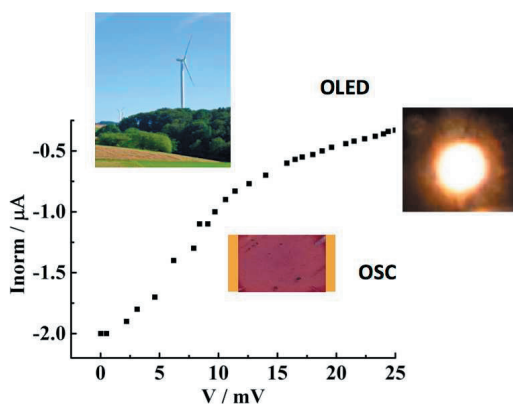
Figure 7: Once structural relaxation processes are understood the derived structural and dynamical properties can be utilized to improve functional dynamics performance allowing developing new classes of organic solar cells (OSC) (see current-voltage (I-V)curve) or organic light emitting diodes (OLED) based on plastic (see inset).

3   Applications in Energy Research

Both as a consequence of successive technical developments, and based on the chemical rules and time laws derived during the various method development process, it has become possible to optimize functional performances of solar cell organic materials and devices (Figure 7).

By moving the fundamental method developments into the application regime (again as a proof for the methods' developments), the circle of promises closes that 3rd and 4th generation X-ray sources may help to optimize strategies for modern material performances. In order to test whether the structural dynamics laws derived through the "molecular movie" approach can have direct consequences for applications of the molecular system, electronic devices such as organic solar cells and organic light emitting diodes have been built and Figure 7 presents such efforts. As can be seen on the current voltage (I-V) curve, when combining with semiconducting plastic material (which by itself is not photo active), efficient organic solar cells with high mechanic flexibility can be designed.

Since the time-resolved X-ray methods allow for detangling local to global and global to local structural responses, desired functional actions of a device like energy storage can be distinguished from "energy-eating" processes based on non-desired heating and energy quenching processes. In Figure 7, the performance of an optimized all-over organic solar cell is presented. Small atomic changes on the light-absorbing chromophore unit lead to a complete switching of its functional dynamics – from light absorbing solar cell devices [20-24] to a light emitting organic diode [22]. In another example the understanding of the crystallization processes of organic material out of time-resolved X-ray diffraction (TRXRD) studies has been influenced by the optimization of the recycling process of molten PET bottles to ultra-hard polyethylene [23]. Such ultra-hard plastic material is currently used in every 2nd wind craft machine produced world-wide.

The current examples emphasize, however, that the real world, functional materials, pharmaceuticals, catalysts or energy converting materials are not always crystalline and far from being periodically ideally arranged as it could look like when performing model type of investigations. If the intrinsic spatial resolution of the system does not allow for such detailed investigations, a combination of ultrafast X-ray spectroscopy and ultrafast X-ray diffraction or scattering as the *"local to global approach"* deliver configuration and charge information of the molecules studied [25-64]. This approach will be described in the following chapter.

4   "From Local to Global": Ultrafast X-ray Spectroscopy and Diffraction Shake Their Hands

"From Local to Global": as ultrafast X-ray spectroscopy and diffraction shake hands, they open up new ways to study complex chemical reactions. As common or photon based research, X-ray laser science allows coupling X-ray techniques coming from complementary pools of methods. Such a merge resembles the "local to global" approach when combining ultrafast X-ray diffraction with high-spatial resolution with high-energy resolution X-ray spectroscopy. Both methods are technically demanding by their own: ultraprecise structure determination requires the use of very hard X-ray radiation (starting from 18 keV X-ray

energy) and very high angular momentum collection, on one hand. On the other hand, X-ray spectroscopy with ultra-high energy resolution requires highest spectrometer grating resolution or the implementation of 2- dimensional X-ray laser spectroscopy techniques, and all of that on the ultrafast time scale (and combined).

X-ray spectroscopy [25,26] and diffraction [27,28] are complementary techniques. X-ray spectroscopy allows probing the electronic properties in an element-, orbital- and site-specific way, for example bonding or oxidation state changes [29-32]. Here we get local information for the system under investigation hence the method is referenced as the *local approach*. In X-ray diffraction the structural changes of the whole bulk are probed which is termed as the *global approach*. With these approaches it is possible to get the overall structural properties of the target system. Furthermore, by applying both methods more complete information can be obtained. However, to follow whole reaction pathways or reaction intermediates between the start and end of a reaction, the experimental approaches have to be extended using the time-resolved method, i.e. the pump-probe scheme, as described in the previous section and shown in Figure 8.
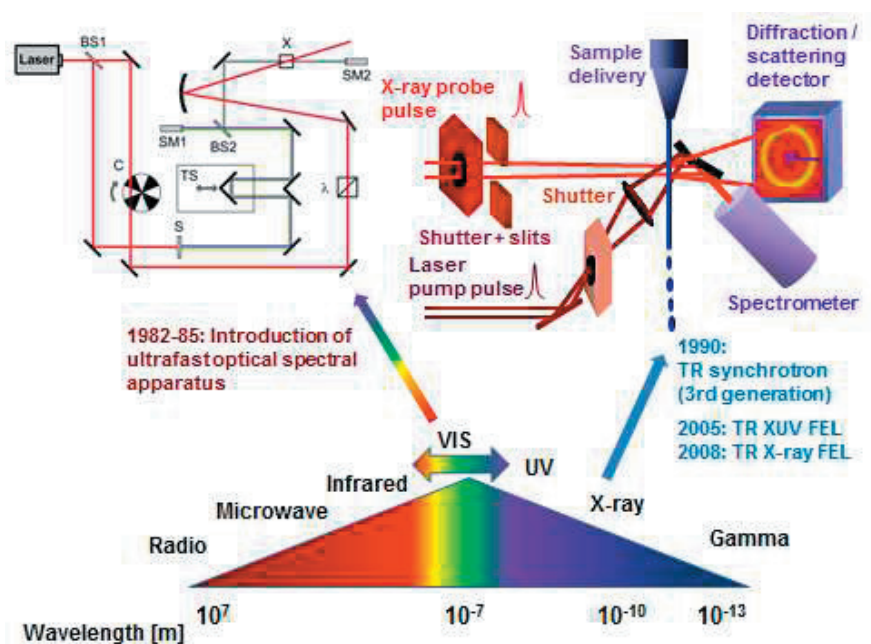
Figure 8: Principle of various types of pump-probe experiments applied to the optical region (left) and the corresponding type of experiments in the X-ray region (right). An optical laser pump initiates a chemical reaction and an optical (left) or X-ray laser pulse (right) probes the proceeding reaction pathways. Right: Both the X-ray spectroscopy and X-ray diffraction signal can be recorded, yielding complementary information.

With the availability of the first X-ray FEL, FLASH at DESY in Hamburg, Germany, it became possible to implement X-ray spectroscopy and diffraction for investigating ultrafast chemical reaction processes. In the following, we present two pioneering soft X-ray experiments performed at the FEL facilities FLASH and LCLS.

The first example is a time-resolved X-ray diffraction measurement on silver-behenate [16,19,33,34], the second example is a *fs* time-resolved X-ray spectroscopy experiment on iron-pentacarbonyl ($Fe(CO)_5$) [35-37]. Figure 9 (a) shows the apparatus which has been developed for such experiments at FELs. The ultrafast X-ray diffraction of a silver-containing redox system embedded in a supramolecular organic structure has been studied (Figure 9b). The experiment is a proof-of-principle, utilizing FEL radiation for ultrafast X-ray diffraction of chemical systems in real-time. By investigating the time-evolution of the Bragg reflections (Figure 9c) complex photo-induced transformation kinetics of partially photo-chemically induced and heat-propagation-influenced reaction kinetics has been found.

The decay curve depicts the propagation of the ultrafast transformation throughout the whole material. The black areas present the non-transformed material, the white areas the island of transformed material.
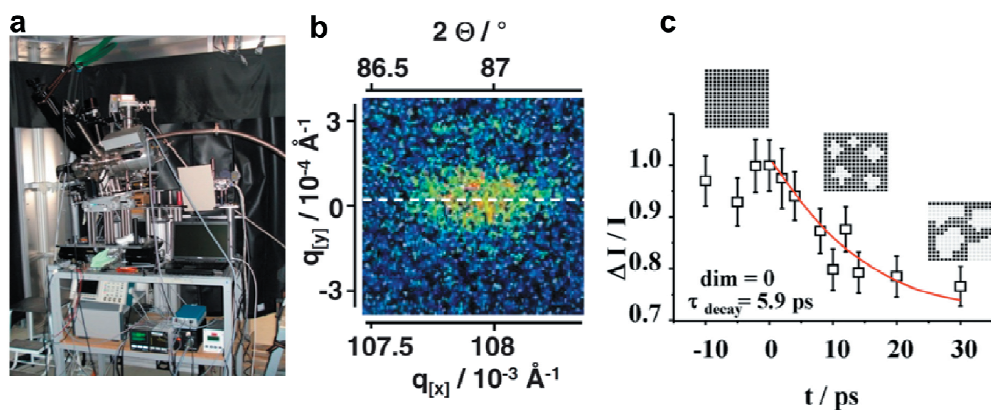
Figure 9: (a) The X-ray photon endstation at the FEL, FLASH at The modular built-up endstation has been used for the experiment presented in **(b)** and **(c)**. **(b)** Bragg diffraction peak (110) of silver-behenate studied by single-shot FEL pulses [16,19,33,34]. **(c)** The time-dependent behavior of the Bragg peak after photo-excitation.

5   Applications in Photocatalysis

Another experiment in the field of photocatalyis but utilizing soft X-ray spectroscopy has recently been worked out. For pioneering time-resolved X-ray spectroscopy measurements the FLASH end station in Figure 9 has been modified by adding a soft X-ray spectrometer in Rowland circle geometry [35]. This so called Liquid Jet (LJ) end station has then been successfully used in resonant inelastic X-ray scattering (RIXS) experiments on $Fe(CO)_5$ at the LCLS-FEL at SLAC in Stanford, CA, USA. Similar end stations have been built to perform soft X-ray spectroscopy experiments at high flux X-ray facilities at DESY [38] and HZB. In the above mentioned experiment [36], the dissociation of iron pentacarbonyl ($Fe(CO)_5$) in ethanol has been studied in real-time. After optical excitation with 266 nm photons, Figure 10(a), $Fe(CO)_5$ dissociates into iron-tetracarbonyl ($Fe(CO)_4$) and carbon monoxide (CO) under solvent-assistance. For every time-delay, the incident monochromatic FEL photon energy has been scanned over the Fe $2p$ edge, and the X-ray emission spectra have been recorded.
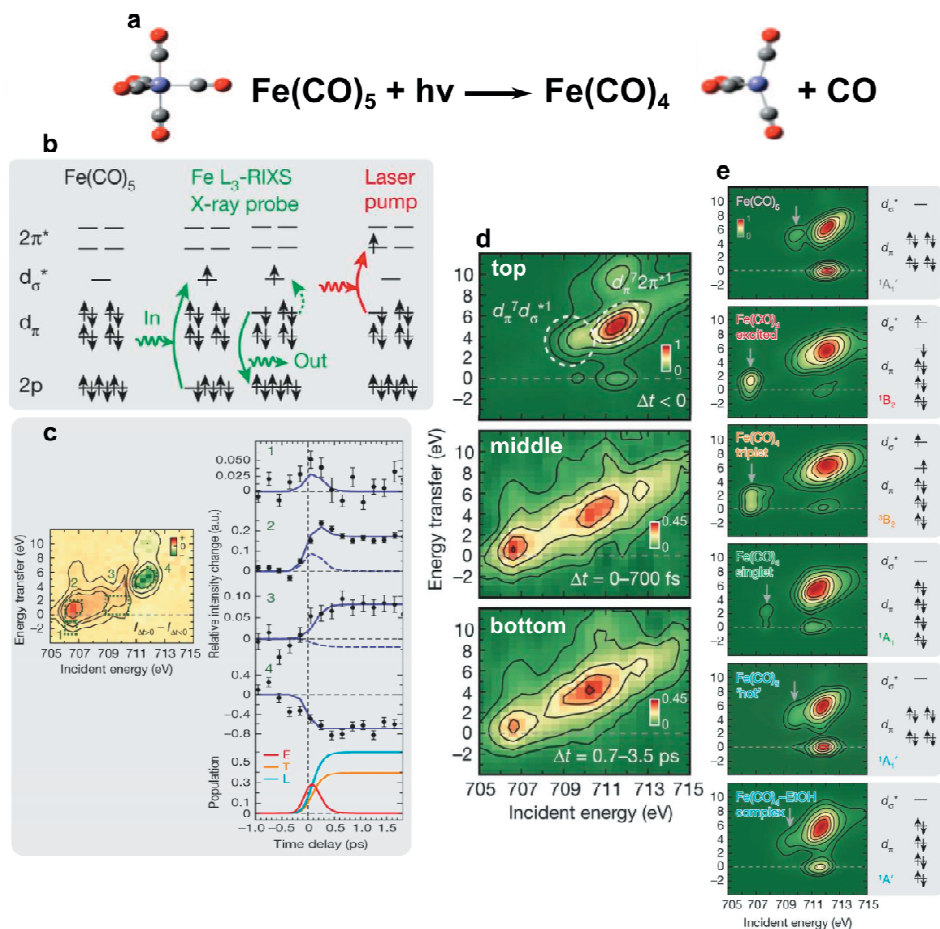
Figure 10 [36]: **(a)** Schematic diagram of Fe(CO)$_5$ photo-dissociation. **(b-e)** Evolution of the valence-electronic structure of Fe(CO)$_4$ in ethanol upon femtosecond spin crossover and complex formation. **(b)** Electron configuration of the Fe(CO)$_5$ ground state, and single-electron transitions of the laser pump - X-ray probe processes. The optical d$_\pi \rightarrow$2π* excitation triggers the dissociation and the RIXS scattering to final valence-excited states involves the probing of the d$_\pi \rightarrow$d$_\sigma$* transitions. **(c, left)** RIXS difference map of summed intensities of all spectra recorded at positive time-delays minus normalized spectra at negative time points, with the regions marked for the temporal evolution plots shown adjacent. **(c, right)** Measured intensities (1-4) versus pump-probe delay (circles with error bars reflecting twice the standard deviation) with the best global fit of a kinetic model (solid lines) with extracted populations of excited state (E), triplet state (T) and ligated complex (L), Fe(CO)$_4$. The dashed lines in 2 and 3 represent alternative models without considering triplet and complexed Fe(CO)$_4$, respectively. **(d)**: The experimental Fe L$_3$-RIXS intensities (encoded in color) versus energy transfer and incident photon energy. **(d, top)** Fe(CO)$_5$ ground state (negative delays, probe before pump, scattering to d$_\pi^7$d$_\sigma$*^{1} and d$_\pi^7$2π*^{1} marked by circles). **(d, middle and bottom)** Difference intensities for 0-700 fs and 0.7-3.5 ps time delays, respectively. **(e)** Calculated Fe L$_3$-RIXS intensities (color coded as in **(d)** and molecular-orbital diagrams of Fe(CO)$_5$ (ground state and hot), excited, triplet, and singlet Fe(CO)$_4$ (all three non-complexed) and solvent complexed singlet Fe(CO)$_4$-EtOH.

The time evolution was deduced from differences of the pumped (positive time-delay) and the un-pumped (negative time-delay) $Fe(CO)_5$ emission spectra. Figure 10 summarizes the ethanol-assisted $Fe(CO)_5$ photo-dissociation pathways and simulations are compared to the experiment. The complexity of reaction increases due to the formation and decay of a triplet state.

The involved orbitals are assigned according to the Fe $2p$ and $3d$, or ligand $2p$ characteristics, and according to the symmetry along the Fe-CO bonds. The star (*) in Figure 10 (b), marks the antibonding orbitals. The dissociation of $(Fe(CO)_5 \rightarrow Fe(CO)_4 + CO)$ is initiated by the optical $d_\pi \rightarrow 2\pi^*$ excitation. The RIXS measurements at the Fe L_3-absorption edge to final valence-excited states involves the probing of the $d_\pi \rightarrow d_\sigma^*$ transition. In Figure 10 (c, left), the difference RIXS spectra (RIXS intensity of incident photon energy versus energy transfer in eV) of the summed pumped and un-pumped sample is illustrated.

Figure 10 (c, right) displays the time structure in the regions (1-4) according to Figure (c, left), in comparison to simulated populations of the excited (E) and triplet (T) states and ligated complex (L). Figure 10 (d) shows the measured RIXS spectra for $Fe(CO)_5$ at negative time delays and the difference after delay intervals of 0 - 700 fs and 0.7 - 3.5 ps respectively. By subtracting the negative time delay with a weight of 0.9, the pumped contributions are isolated. The experimental data is compared to RIXS calculations displayed in Figure 10 (e). Simulated Fe L_3-RIXS intensities and molecular-orbital diagrams of $Fe(CO)_5$ in the ground and hot state and $Fe(CO)_4$ in various valence excited states are depicted. The $2p \rightarrow$ LUMO resonance positions and $d_\pi \rightarrow d_\sigma^*$ RIXS transitions are marked by arrows. The RIXS pattern at early time evolution can only be reproduced when a complex between the $Fe(CO)_4$ with the solvent is taken into account.

6   Applications in Liquid Phase Reaction Dynamics

In the hard X-ray regime [39-41], the "local to global" approach has been established by combining X-ray spectroscopy with X-ray scattering techniques [42, 43]. The ultrafast structural dynamics of various metal organic systems has been studied applying these techniques [44-48].

7   Applications in Biophysics

In biophysics, the ultrafast photon-in/photon-out developments utilizing high flux X-ray sources allow investigating the properties of biorelevant solvents and proteins during their structural reactions [39-52]. The literature-referenced examples include various types of small up to macromolecular model systems studied with FEL radiation.

Depending on the time scale of the system studied, it is furthermore possible to merge X-ray scattering techniques like diffuse X-ray scattering with pressure jump, temperature jump, electric field modulations, and structural freezing methods or, on the chemical modulation side, with, rapid mixing or photo-switching methods. Beyond the contributions summarized in

this book, the combination of synchrotrons or free-electron laser radiation with other techniques, such as electrospray ionization mass spectrometry, allows deriving entirely novel experimental techniques for investigating macromolecules [53-57]. E.g. a mass spectrometric study on gas-phase ubiquitin at FLASH has revealed a fast local structural response, leading to small fragments with yields increasing linearly with photon intensity [60].
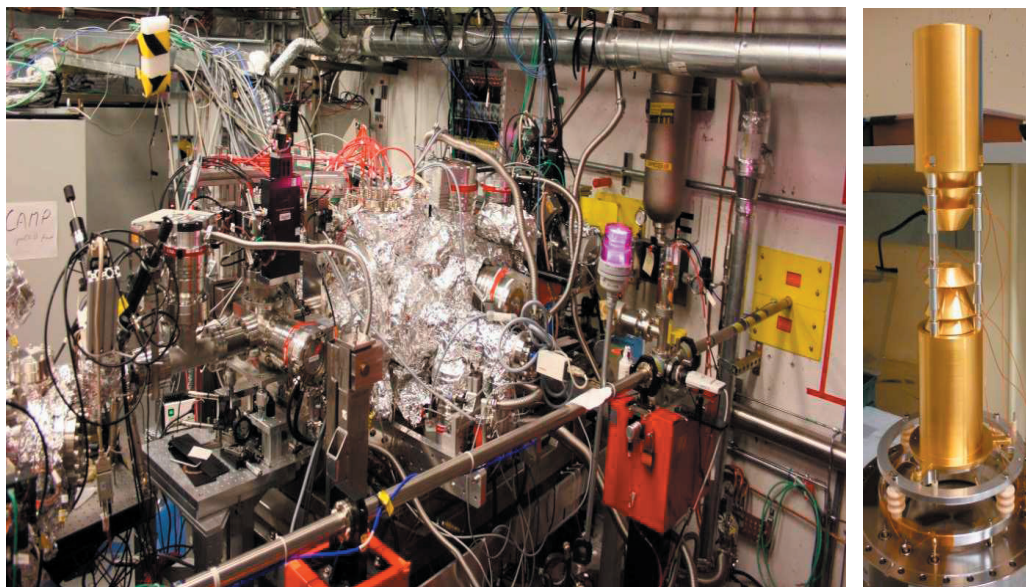
Figure 11: FLASH/CFEL-ASG MultiPurpose (CAMP) endstation [61] at the AMO beamline at the LCLS in 2012. Since 2014, CAMP is a permanent endstation at FLASH/BL1 at DESY in Hamburg. Various experimental techniques can be employed in the instrument, including novel time-resolved X-ray studies of chemical reactions in the gas phase. On the right side, CAMP's double-sided velocity-map-imaging spectrometer is shown.

8 Ultrafast Imaging of Gas-Phase Reactions Allows Investigating Fundamental Definitions in Chemistry

The advent of X-ray free-electron lasers enabled not only novel studies in the condensed phase, as described in the first parts of this chapter, but also brought forward unprecedented possibilities to study dynamical processes in the gas phase. The very short and very intense X-ray pulses made it possible, for the first time, to probe ultrafast photo-induced molecular dynamics by electron or ion momentum spectroscopy following multiple, element-specific inner-shell absorption.

These new fourth generation light sources also called for novel, dedicated instrumentation. To this end, different endstations have been developed, initially in particular for the use at the atomic, molecular, and optical physics (AMO) beamline, which was the first beamline to become operational at the LCLS [58-70] in 2009. Several of the early, pioneering experiments

from 2009 to 2012 have been performed at the AMO beamline in the CFEL-ASG MultiPurpose (CAMP) endstation [61], see Figure 11, which was developed within the Max-Planck Advanced Study Group (ASG) at the Center for Free-Electron Laser Science (CFEL) in Hamburg. Experiments in this instrument range from X-ray imaging of biomolecules, nanocrystals, and clusters [70-77] to (time-resolved) ion and electron spectroscopy on atoms and small gas-phase molecules [78-84]. Since 2014, CAMP is a permanent user endstation at FLASH/BL1 at DESY in Hamburg [76], and its successor, LAMP, has become operational at the AMO beamline at the LCLS [77].

Moreover, the high-field physics (HFP) instrument has been developed at the LCLS [78], housing an ion momentum spectrometer and several electron time-of-flight spectrometers mounted under different angles. It also offers a pulsed, supersonic molecular beam, delivering gas-phase molecules to the interaction region. The example given in the following [79], as well as multiple other pioneering gas-phase studies [89-95] have been conducted in the HFP instrument.
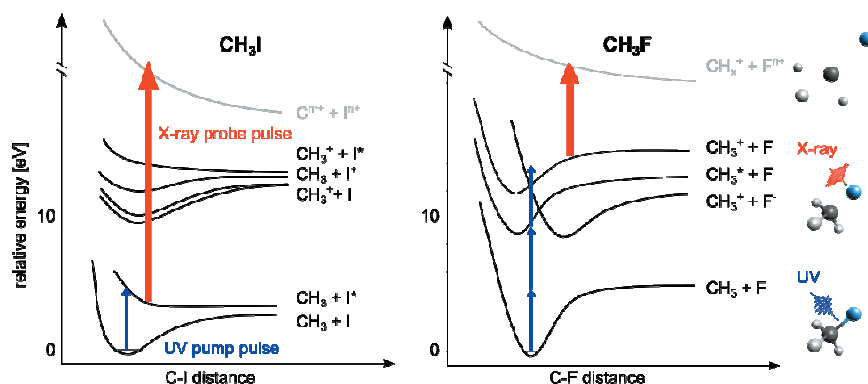
Figure 12: Schematic potential energy curves for iodomethane (left) and fluoromethane (right). Absorption of one 267 nm photon in CH_3I leads to resonant population of a repulsive neutral state, whereas multi-photon UV absorption in CH_3F populates to several higher-lying ionic states. After a given time delay, these states are probed by Coulomb explosion following inner-shell ionization of the respective halogen atom by one or several X-ray photons, probing the transition from a molecule to isolated atoms.

Gas-phase FEL experiments allow questioning fundamental definitions in chemistry, for example the investigation of processes beyond the Born-Oppenheimer approximation. As an example, a UV-pump, X-ray-probe study of two complementary halomethane molecules with different photochemistry is shown here.

In Figure 12, the schematic potential energy curves (PECs) of iodomethane (CH_3I) and fluoromethane (CH_3F) are displayed, illustrating that the different halogen species give rise to qualitatively different PECs. One reason for this is the considerably different electronegativity of iodine and fluorine, stabilizing the C-F bond in contrast to the C-I bond. Upon absorption of one 267 nm UV photon, CH_3I dissociates into two neutral fragments, $CH_3 + I$, whereas in CH_3F, no PEC is resonantly accessible at 4.6 eV. In the latter case, absorption of at least three UV photons in the same molecule populates several higher-lying ionic PECs, also resulting in dissociation of the molecules.

After a tunable time delay, an intense X-ray pulse (727 eV, 1 mJ) probes the dissociating system by ionizing predominantly the iodine (3d) or the fluorine (1s) level, respectively, because of their large absorption cross section (3.3 Mb for I and 0.4 Mb for F, compared to 0.1 Mb for CH_3), resulting in a localized positive charge on the halogen. At these very high X-ray intensities, a single molecule can absorb many photons, such that very highly charged ions up to I^{21+}/F^{4+} and C^{4+} are created. As the charge is initially created locally at the halogen atom though, the fact that highly charged carbon ions are also detected, already shows that the charge rearranges within the molecule before or during the fragmentation.

In Figure 13, the calculated electrostatic potentials of an I^{6+} + CH_3 are plotted for three different internuclear distances between the two fragments, together with the binding energy of the highest occupied orbital. In the intact molecule (a), the electrons are delocalized, but as the fragments move apart the potential barrier rises, until at a certain critical distance (b), it reaches the electron binding energy. Therefore, for larger distances (c) the electrons can classically be regarded as localized at one of the two fragments. It is this transition from a bound molecule to isolated atoms that is probed by time-resolved ion spectroscopy.
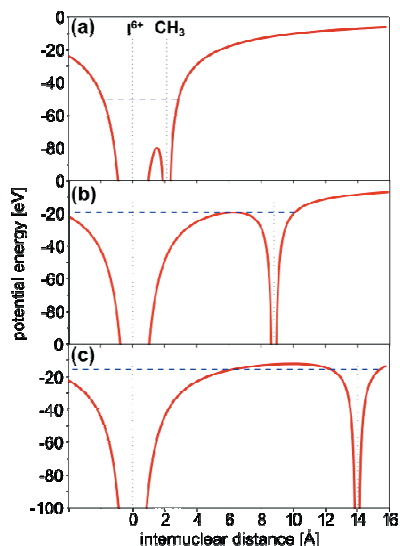
Figure 13: Calculated Coulomb potentials for an I^{6+} atom and a neutral methyl radical for (a) the equilibrium distance, (b) the cricital distance (see text) and (c) for isolated atoms (in a classical picture). The dashed blue line indicates the energy of the electron in the highest occupied orbital.

The delay-dependent time-of-flight peaks of selected ions of iodomethane and fluoromethane are shown in Figure 12. It is evident that the fragmentation patterns of the two molecules are qualitatively different. For iodine charge states $\geq I^{4+}$, the appearance of low-energy ions at positive delays is clearly visible (channel 3 in (a)). These ions originate from the pump-probe process as indicated for iodomethane in Figure 12, and can be used to extract the critical internuclear distance, up to which electron transfer from methyl to iodine is classically allowed for a given charge state.

Signatures of long-distance intramolecular electron transfer have been observed for both, CH_3I and CH_3F, and the reconstructed critical distances (up to 15 Å for I^{21+}) are in good agreement with a classical over-the-barrier model.

Two other channels can be seen in Figure 14 that correspond to Coulomb explosion of intact molecules by only the FEL (1) and to ionic dissociation induced by multi-photon UV absorption (2), as illustrated for fluoromethane in Figure 12 , which also occurs with a lower probability in CH_3I. The low-energy channel is absent in the fluorine ions.
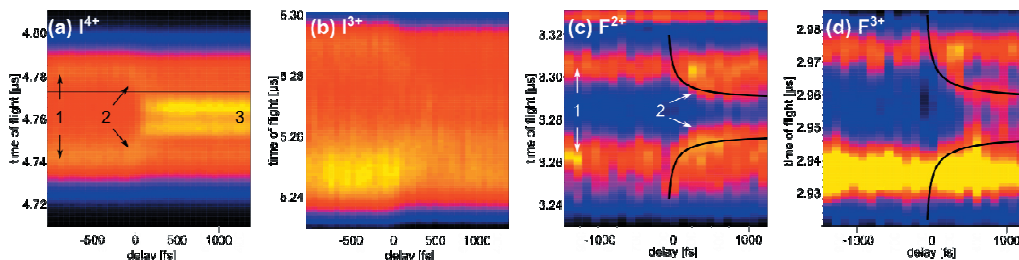
Figure 14: Time-of-flight spectra as a function of the pump-probe delay for selected fragments of iodomethane and fluoromethane. Different fragmentation channels are indicated by 1, 2, 3 (see text). Additionally, in (c) and (d), calculated delay-dependent time-of-flight curves are overlaid with the data, corresponding to an asymptotic kinetic energy of 0.4 eV. Positive delays correspond to the UV pulse arriving before the X-ray pulse.

Acknowledgements

References

The Concept: Filming Chemical Reactions in Real Time Utilizing Ultrafast High Flux X-ray Sources

[1]   S. Techert, F. Schotte, M. Wulff, *Picosecond X-ray Diffraction Probed Transient Structural Changes in Organic Solids*, Phys. Rev. Lett. 86, 2030-2034 (2001). doi: 10.1103/PhysRevLett.86.2030

[2]   a) S. Schmatz, J. Hauschildt, *Four-Mode Calculations of Resonance States of Intermediate Complexes in the S_N2 Reaction Cl^- + CH_3Cl' -> $ClCH_3$ + Cl'^-*, J. Chem. Phys. 118, 4499 (2003); b) C. Hennig, S. Schmatz, *Rotational effects in complex-forming bimolecular substitution reactions: A quantum-mechanical approach*, J. Chem. Phys. 131, 224303 (2009); c) S. Schmatz, *Approximate calculation of anharmonic densities of vibrational states for very large molecules, Chem. Phys.* 346, 198 (2008).

[3]   S. Techert, S. Schmatz, *Time-resolved X-Ray Scattering and the Observation of Intramolecular Reaction Dynamics in Liquids, Z. Phys. Chem.* 216 *(Juergen Troe Festschrift)*, 575-583 (2002). doi: 10.1515/zpch-2015-0610

[4]   S. Techert, *First, Second and Third Order Correlation Function in Time-resolved X-ray Diffraction*, J. Appl. Cryst. 37, 445-458 (2004). doi: 10.1107/S0021889804007381

[5]   A. Debnarova, S. Techert, S. Schmatz, *Ab Initio Treatment of Time-resolved X-ray Scattering: Application to the Photoisomerization of Stilbene, J. Chem. Phys.* 125, 224101-1 - 9 (2006). doi: 10.1063/1.2400231

[6]   a) FLASH; b) LCLS, c) SACLA, d) SwissFEL, e) EXFEL

Crystallography with Ultrahigh Temporal and Ultrahigh Spatial Resolution

[7]   F. Schotte, S. Techert, P. Anfinrud, V. Srajer, K. Moffat, M. Wulff, *Recent Advantages in the Generation of Pulsed Synchrotron Radiation Suitable for Picosecond Time-resolved X-Ray Studies,* Kapitel in *Third-Generation Hard X-ray Synchrotron Radiation Sources*, Ed. D. M. Mills, 345, Wiley-Interscience (2002).

[8]   S. Techert, *Analysis of Chemical Reactions at Ultra-fast Time Resolution*, TESLA - Technical Design Report I, 18-20 (2001). G. Busse, Th. Tschentscher, A. Plech, M. Wulff, B. Frederichs, S. Techert, *First Investigations on the Kinetics of the Topochemical Reaction of p-Formyl-trans-cinnamic Acid by Time-resolved X-ray Diffraction, Faraday Discuss.* 122, 105-117 (2003). doi: 10.1039/B202831J

[9]   E. Collet, M.-H. Lemée-Cailleau, M. Buron-Le Cointe, H. Cailleau, M. Wulff, T. Luty, S.-Y. Koshihara, M. Meyer, L. Toupet, P. Rabiller, S. Techert, *Laser-induced Ferroelectric Structural Order in an Organic Charge-Transfer Crystal, Science* 300, 612-614 (2003). doi: 10.1126/science.1082001

[10]   S. Techert, K. A. Zachariasse, *Structure Determination of the Intramolecular Charge Transfer State in Crystalline 4-(Diisopropylamino)benzonitrile from Picosecond X-ray Diffraction*, J. Am. Chem. Soc. 126 (17), 5593–5600 (2004). doi: 10.1021/ja0379518

[11]   J. Davaasambuu, P. Durand, S. Techert, *Experimental Requirements for Light-induced Reactions in Powders Investigated by Time-resolved X-ray Diffraction*, J. Synchrotron Rad. 11, 483-489 (2004). doi: 10.1107/S090904950402463X

[12]   A. L. Cavalieri, D. M. Fritz, S. H. Lee, P. H. Bucksbaum, D. A. Reis, J. Rudati, D. M. Mills, P. H. Fuoss, G. B. Stephenson, C. C. Kao, D. P. Siddons, D. P. Lowney, A. G.

MacPhee, D. Weinstein, R. W. Falcone, R. Pahl, J. Als-Nielsen, C. Blome, S. Düsterer, R. Ischebeck, H. Schlarb, H. Schulte-Schrepping, Th. Tschentscher, J. Schneider, O. Hignette, F. Sette, K. Sokolowski-Tinten, H. N. Chapman, R. W. Lee, T. N. Hansen, O. Synnergren, J. Larsson, S. Techert, J. Sheppard, J. S. Wark, M. Bergh, C. Caleman, G. Huldt, D. van der Spoel, N. Timneanu, J. Hajdu, R. A. Akre, E. Bong, P. Emma, P. Krejcik, J. Arthur, S. Brennan, K. J. Gaffney, A. M. Lindenberg, K. Luening, J. B. Hastings, *Clocking Femtosecond X-Rays*, *Phys. Rev. Lett.* 94, 114801-114804 (2005). doi: 10.1103/PhysRevLett.94.114801

[13] A. M. Lindenberg, J. Larsson, K. Sokolowski-Tinten, K. Gaffney, C. Blome, O. Synnergren, J. Sheppard, C. Caleman, A. G. MacPhee, D. Weinstein, D. P. Lowney, T. Allison, T. Matthews, R. W. Falcone, A. L. Cavalieri, D. M. Fritz, S. H. Lee, P. H. Bucksbaum, D. A. Reis, J. Rudati, D. M. Mills, P. H. Fuoss, G. B. Stephenson, C. C. Kao, D. P. Siddons, R. Pahl, J. Als-Nielsen, S. Dusterer, R. Ischebeck, H. Schlarb, H. Shulte-Schrepping, Th. Tschentscher, J. Schneider, O. Hignette, F. Sette, H. N. Chapman, R. W. Lee, T. N. Hansen, S. Techert, J. S. Wark, M. Bergh, G. Huldt, D. van der Spoel, M. Timneanu, J. Hajdu, D. von der Linde, R. A. Akre, E. Bong, P. Emma, P. Krejcik, J. Arthur, S. Brennan, K. Luening, J. B. Hastings, *Atomic-scale Visualization of Inertial Dynamics*, *Science* 308, 392-394 (2005). doi: 10.1126/science.1107996

[14] C. Blome, Th. Tschentscher, J. Davaasambuu, P. Durand, S. Techert, *Ultrafast Time-Resolved Powder Diffraction Using Free-Electron Laser Radiation*, *AIP Conference Proceedings* 879 (1), 1254-1257 (2007). doi: 10.1107/S0909049505026464

[15] J. Hallmann, W. Morgenroth, C. Paulmann, J. Davaasambuu, Q.-Y. Kong, M. Wulff, S. Techert, *Time-resolved X-ray Diffraction of the Photochromic a-Styrylpyrylium Trifluoromethanesulfonate Crystal Films Reveals Ultrafast Structural Switching*, *J. Am. Chem. Soc.* 131, 15018-15025 (2009). doi: 10.1021/ja905484u

[16] I. Rajkovic, G. Busse, J. Hallmann, R. More, M. Petri, W. Quevedo, F. Krasniqi, A. Rudenko, T. Tschentscher, N. Stojanovic, S. Düsterer, R. Treusch, M. Tolkiehn, S. Techert, *Diffraction Properties of Periodic Lattices under Free Electron Laser Radiation*, *Phys. Rev. Lett.* 104, 125503-6 (2010). doi: 0.1103/PhysRevLett.104.125503

[17] M. Messerschmidt, T. Tschentscher, M. Camaratta, A. Meents, C. Sager, G. Busse, J. Davaasambuu, S. Techert, *Ultrafast Potential Energy Surface Softening of One-Dimensional Organic Conductors Revealed by Picosecond Time-resolved Laue Crystallography*, *J. Phys. Chem. A* 114, 7677–7681 (2010). doi: 10.1021/jp104081b

[18] H. O. Sorensen, S. Schmidt, J. P. Wright, G. B. M. Vaughan, S. Techert, E. F. Garman, J. Oddershede, J. Davaasambu, K. S. Paithankar, C. Gundlach, H. F. Poulsen, *Multigrain Crystallography*, *Z. Kristallogr.* 227, 63–78 (2012). doi: 10.1524/zkri.2012.1438

[19] W. Quevedo, G. Busse, J. Hallmann, R. More, M. Petri, F. Krasniqi, A. Rudenko, Th. Tschentscher, N. Stojanovic, S. Düsterer, R. Treusch, M. Tolkiehn, S. Techert, I. Rajkovic, *Ultrafast Time Dynamics Studies of Periodic Lattices under Free Electron Laser Radiation*, *J. Appl. Phys.* 112, 093519 (2012). doi: 10.1063/1.4764918

Applications in Energy Research

[20] S. Thekku Veedu, D. Raiser, R. Kia, M. Scholz, S. Techert, *Ultrafast Dynamical Study of Pyrene-N,N-dimethylaniline (PyDMA) as an Organic Molecular Diode in Solid State*, *J. Phys. Chem. B* 118 (12), 3291 – 3297 (2014). doi: 10.1021/jp4121222

[21] S. Mildner, J. Hoffmann, P. E. Blöchl, S. Techert, C. Jooss, *Temperature and Doping Dependent Optical Absorption in the Small-Polaron System Pr₁₋ₓCaₓMnO₃*, *Phys. Rev. B* 92, 035145 (2015). doi: 10.1103/PhysRevB.92.035145

[22] K. R. Idzik, P. J. Cywinski, W. Kuznik, J. Frydel, T. Licha, T. Ratajczyk, *The Optical Properties and Quantum Chemical Calculations of Thienyl and Furyl Derivatives of Pyrene*, *Phys. Chem. Chem. Phys.*, 17 (35) 22758-22769 (2015). doi**:** 10.1039/C5CP03013G

[23] M. Petri, *private communication* (2016) and http://www.armacell-core-foams.com/www/armacell/INETArmacell.nsf/standard/DC2C5D13EA92BD46802576E200526C26

[24] D. Raiser, S. Mildner, B. Ifland, M. Sotoudeh, P. Blöchl, S. Techert, C. Jooss, *Evolution of Hot Polaron States with a Nanosecond Lifetime in a Manganite Peroskite*, *Adv. Energy Mat* 1, 1-9 (2017). doi: 10.1002/aenm.201602174

"From Local to Global" Approach: Ultrafast X-ray Spectroscopy and Ultrafast X-ray Diffraction Shake Their Hands

[25] I. Rajkovic, J. Hallmann, S. Grübel, R. More, W. Quevedo, M. Petri, S. Techert, *Development of a Multipurpose Vacuum Chamber for Serial Optical and Diffraction Experiments with Free Electron Laser Radiation*, *Rev. Sci. Instr.* 81, 045105-1-6 (2010). doi: 10.1063/1.3327816

[26] J. Hallmann, W. Quevedo, G. Grübel, I. Rajkovic, G. Busse, R. Moré, M. Petri, S. Techert, *First Steps towards Probing Chemical Reaction Dynamics with Free Electron Laser Radiation (Special Issue Five Years of FLASH)*, *J. Phys. B: At. Mol. Opt. Phys.* 43 194009-194016 (2010). doi: 10.1088/0953-4075/43/19/194009

[27] *P. Wernet, K. Kunnus, S. Schreck, W. Quevedo, S. Techert, F.M.F. de Groot, M. Odelius, A. Föhlisch, Dissecting Local Atomic and Intermolecular Interactions of Transition-metal Ions in Solution with Selective X-ray Spectroscopy, J. Phys. Chem. Lett.* 3 (23), 3448–3453 (2012). doi: 10.1021/jz301486u

[28] K. Kunnus, I. Rajkovic, S. Schreck, W. Quevedo, S. Eckert, M. Beye, E. Suljoti, C. Weniger, C. Kalus, S. Grübel, M. Scholz, D. Nordlund, W. Zhang, R. W. Hartsock, K. J. Gaffney, W. F. Schlotter, J. J. Turner, B. Kennedy, F. Hennies, S. Techert, P. Wernet, A. Föhlisch, *A Setup for Resonant Inelastic Soft X-ray Scattering on Liquids at Free Electron Laser Light Sources*, *Rev. Sci. Instrum.* 83, 123109-17 (2012), doi: 10.1063/1.4772685

[29] R. Mitzner, J. Rehanek, J. Kern, S. Gul, J. Hattne, T. Taguchi, R. Alonso-Mori, R. Tran, C. Weniger, H. Schröder, W. Quevedo, H. Laksmono, R. G. Sierra, G. Han, B. Lassalle-Kaiser, S. Koroidov, K. Kubicek, S. Schreck, K. Kunnus, M. Brzhezinskaya, A. Firsov, M.l P. Minitti, J. J. Turner, S. Moeller, N. K. Sauter, M. J. Bogan, D. Nordlund, W. F. Schlotter, J. Messinger, A. Borovik, S. Techert, F. M. F. de Groot, A. Föhlisch, A. Erko, U. Bergmann, V. K. Yachandra, P. Wernet, J. Yano, *Mn L-edge X-ray Absorption Spectroscopy of Dilute Systems Relevant to Photosystem II Using a X-ray Free Electron Laser*, *J. Phys. Chem. Lett.* 12, 1021 (2013). doi: 10.1021/jz401837f

[30] S. Schreck, M. Beye, J. A. Sellberg, T. McQueen, H. Laksmono, D. DePonte, B. Kennedy, S. Eckert, D. Schlesinger, D. Nordlund, H. Ogasawara, R. G. Sierra, V. H. Segtnan, K. Kubicek, W. F. Schlotter, G. Dakovski, U. Bergmann, S. Techert, L. G. M. Pettersson, J. Stoehr, P. Wernet, M. J. Bogan, Y. Harada, A. Nilsson, A. Föhlisch, *Reabsorption of Soft X-Ray Emission at High X-Ray Free-Electron Laser Fluences*, *Phys. Rev. Lett.* 113, 153002-6 (2014). doi: PhysRevLett.111.180001

Applications in Photocatalysis

[31] Ph. Wernet, K. Kunnus, I. Josefsson, I. Rajkovic, W. Quevedo, M. Beye, S. Schreck, S. Grübel, M. Scholz, D. Nordlund, W. Zhang, R. W. Hartsock, W. F. Schlotter, J. J. Turner, B. Kennedy, F. Hennies, F. M. F. de Groot, K. J. Gaffney, S. Techert, M. Odelius, A. Föhlisch, *Orbital-specific Mapping of the Ligand Exchange Dynamics of Fe(CO)5 in Solution, Nature* 520 (7545), 78-81 (2015). doi:10.1038/nature14296

[32] K. Kunnus, I. Josefsson, I. Rajkovic, S. Schreck, W. Quevedo, M. Beye, C. Weniger, S. Grübel, M. Scholz, D. Nordlund, W. Zhang, R. Hartsock, K. J. Gaffney, W. F. Schlotter, J. J. Turner, B. Kennedy, F. Hennies, F. M. F. de Groot, S. Techert, M. Odelius, P. Wernet, A. Föhlisch, *Identification of the Dominant Photochemical Pathways and Mechanistic Insights to the Ultrafast Ligand Exchange of Fe(CO)5 to Fe(CO)4EtOH, Struct. Dyn.* 3, 043204-6 (2016). doi: 10.1063/1.4941602

[33] *K. Kunnus, I. Josefsson, I. Rajkovic, S. Schreck, W. Quevedo, M. Beye, S. Grübel, M. Scholz, D. Nordlund, W. Zhang, R. Hartsock, K. J. Gaffney, W. F. Schlotter, J. J. Turner, B. Kennedy, F. Hennies, S. Techert, P. Wernet, M. Odelius, A. Föhlisch, Anti-Stokes Resonant X-ray Raman Scattering for Atom Specific and Excited State Selective Dynamics,* N. J. Phys. 18, 103011-6 (2016). doi:10.1088/1367-2630/18/10/103011

[34] Z. Yin, H. B. Peters, U. Hahn, M. Agaker, A. Hage, F. Scholz, R. Reinninger, F. Siewert, J. Nordgren, J. Viefhaus, S. Techert, *A New Compact VUV-Spectrometer for RIXS Studies at DESY, Rev. Sci. Instr.* 86, 093109,1–5 (2015). doi: A151379RR

Applications in Liquid Phase Reaction Dynamics

[35] R. Neutze, R. Wouts, S. Techert, A. Kirrander, J. Davidson, M. Kocsis, F. Schotte, M. Wulff, *Visualising Photo-Chemical Dynamics in Solution through Picosecond X-ray Scattering, Phys. Rev. Lett.* 87, 195508-195512 (2001). doi: 10.1103/PhysRevLett.86.2030

[36] S. Techert, R. Neutze, *Ultrafast Chemistry, TESLA - Technical Design Report* V, 133-139 (2001).

[37] A. Debnarova, S. Techert, S. Schmatz, *Computational Studies of the X-ray Scattering Properties of Laser Aligned Stilbene, J. Chem. Phys.* 134, 054302-1-7 (2011). doi: 10.1063/1.3523569

[38] W. Zhang, R. Alonso-Mori, U. Bergmann, C. Bressler, M. Chollet, A. Galler, W. Gawelda, R. G. Hadt, R. W. Hartsock, T. Kroll, K. S. Kjær, K. Kubicek, H. T. Lemke, H. W. Liang, D. A. Meyer, M. M. Nielsen, C. Purser, J. S. Robinson, E. I. Solomon, Z. Sun, D. Sokaras, T. B. van Driel, G. Vankó, T.-C. Weng, D. Zhu, K. J. Gaffney, *Tracking Excited State Charge and Spin Dynamics in Iron Coordination Complexes, Nature* 509 (7500), 345-8 (2014). doi: 10.1038/nature13252

[39] S. E. Canton, K. S. Kjær, G. Vankó, T. B. van Driel, S. Adachi, A. Bordage, C. Bressler, P. Chabera, M. Christensen, A. O. Dohn, A. Galler, W. Gawelda, D. Gosztola, K. Haldrup, T. Harlang, Y. Liu, K. B. Møller, Z. Németh, S. Nozawa, M. Pápai, T. Sato, K. Suarez-Alcantara, T. Togashi, K. Tono, J. Uhlig, D. A. Vithanage, K. Wärnmark, M. Yabashi, J. Zhang, V. Sundström, M. M. Nielsen, *Visualizing the Non-equilibrium Dynamics of Photoinduced Intramolecular Electron Transfer with Femtosecond X-ray Pulses, Nat. Comm.* 6(3), 6359-6362 (2015). doi:10.1038/ncomms7359

[40] K. Haldrup, W. Gawelda, R. Abela, R. Alonso-Mori, U. Bergmann, A. Bordage, M. Cammarata, S. E. Canton, A. O. Dohn, T. B. Van Driel, D. M. Fritz, A. Galler, P. Glatzel, T. Harlang, K. S. Kjær, H. T. Lemke, K. B. Møller, Z. Németh, M. Pápai, N. Sas, J. Uhlig, D. Zhu, V. Vankó Sundström, M. M. Nielsen, C. Bressler, *Observing Solvation Dynamics with Simultaneous Femtosecond X-ray Emission Spectroscopy and X-ray Scattering*, J. Phys. Chem. B 120, 6, 1158-1168 (2016). doi: 10.1021/acs.jpcb.5b12471

[41] W. Zhang, K. S. Kjær, R. Alonso-Mori, U. Bergmann, M. Chollet, L. A. Fredin, R. G. Hadt, R. W. Hartsock, T. Harlang, T. Kroll, K. Kubicek, H. T. Lemke, H. W. Liang, Y. Liu, M. M. Nielsen, P. Persson, J. S. Robinson, E. I. Solomon, Z. Sun, D. Sokaras, T. B. van Driel, T.-C. Weng, D. Zhu, K. Warnmark, V. Sundstrom, K. J. Gaffney, *Manipulating Charge Transfer Excited State Relaxation and Spin Crossover in Iron Coordination Complexes with Ligand Substitution*, Chem. Sci. 8, 515-523 (2017). online published (2016). doi: 10.1039/c6sc03070j

Application in Biophysics

[42] A. Ramos, S. Techert, *Influence of the Water Structure on the Acetylcholinesterase Efficiency*, Biophys. J. 89, 1990–2003 (2005). doi:10.1529/biophysj.104.055798

[43] W. Quevedo, M. Petri, G. Busse, S. Techert, *On the Mechanism of Photo-induced Phase Transitions in Ternary Liquid Crystal Systems Near Thermal Equilibrium*, J. Chem. Phys. 129, 024502-1-10 (2008). doi: 10.1063/1.2943200

[44] M. Petri, W. Quevedo, G. Busse, S. Techert, *Photo-Induced Phase Transitions to Liquid Crystal Phases: Influence of the Chain Length from C8E4 to C14E4*, Materials, 2 (Special Issue Liquid Crystals), 1305-1322 (2009). doi:10.3390/ma2031305

[45] M. Petri, A. Menzel, O. Bunk, G. Busse, S. Techert, *Concentration Effects on the Dynamics of Liquid Crystalline Self-Assembly: Time-Resolved X-ray Scattering Studies*, J. Phys. Chem. A 115, 2176–2183 (2011). doi: 10.1021/jp1108224

[46] A. Debnarova, S. Techert, S. Schmatz, *Limitations of High-intensity Soft X-ray Laser Fields for the Characterisation of Water Chemistry: Coulomb Explosion of the Octamer Water Cluster*, Phys. Chem. Chem. Phys. 14, 9606-9614 (2012). doi: 10.1039/C2CP40598A

[47] M. Petri, S. Frey, A. Menzel, D. Görlich, S. Techert, *Structural Characterization of Nanoscale Meshworks within a Nucleoporin FG Hydrogel*, Biomacromolec. 13, 1882-1889 (2012). doi: 10.1021/bm300412q

[48] Z. Yin et al., *Experimental setup for high resolution X-ray spectroscopy of solids and liquid samples*, Proc. SPIE 8849, X-Ray Lasers and Coherent X-Ray Sources: Development and Applications X, 88490I (2013). doi: 10.1117/12.20239

[49] R. Jain, T. Burg, M. Petri, S. Kirschbaum, H. Feindt, S. Steltenkamp, S. Sonnenkalb, S. Becker, C. Griesinger, A. Menzel, S. Techert, *Efficient Microchannel Device for Scattering Experiments with High Flux X-ray Sources and their Perspective towards Application in Neutron Scattering Science*, Eur. Phys. J. E 36, 109-118 (2013). doi: 10.1140/epje/i2013-13109-9

[50] Z. Yin, I Rajkovic, K. Kubicek, W. Quevedo, A. Pietzsch, P. Wernet, A. Föhlisch, S. Techert, *Probing the Hofmeister Effect with Ultra-fast Core-hole Spectroscopy*, J. Phys. Chem. B 118, 9398-9403 (2014). doi: 10.1021/JP504577A

[51] C. Kupitz, S. Basu, I. Grotjohann, R. Fromme, N. a. Zatsepin, K. N. Rendek, M. S. Hunter, R. L. Shoeman, T. A. White, D. Wang, D. James, J.-H. Yang, D. E. Cobb, B. Reeder, R. G. Sierra, H. Liu, A. Barty, A. L. Aquila, D. Deponte, R. A. Kirian, S.

Bari, J. J. Bergkamp, K. R. Beyerlein, M. J. Bogan, C. Caleman, T.-C. Chao, C. E. Conrad, K. M. Davis, H. Fleckenstein, L. Galli, S. P. Hau-Riege, S. Kassemeyer, H. Laksmono, M. Liang, L. Lomb, S. Marchesini, A. V. Martin, M. Messerschmidt, D. Milathianaki, K. Nass, A. Ros, S. Roy-Chowdhury, K. Schmidt, M. Seibert, J. Steinbrener, F. Stellato, L. Yan, C. Yoon, T. A. Moore, A. L. Moore, Y. Pushkar, G. J. Williams, S. Boutet, R. B. Doak, U. Weierstall, M. Frank, H. N. Chapman, J. C. H. Spence, P. Fromme, *Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser, Nature* 513, 261-265 (2014), doi:10.1038/nature13453

[52] D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. DePonte, R. L. Shoeman, D. Wang, D. James, G. Katona, S. Westenhoff, T. A. White, A. Aquila, S. Bari, P. Berntsen, M. Bogan, T. B. van Driel, R. B. Doak, K. S. Kjær, M. Frank, R. Fromme, I. Grotjohann, R. Henning, M. S. Hunter, R. A. Kirian, I. Kosheleva, C. Kupitz, M. Liang, A. V Martin, M. M. Nielsen, M. Messerschmidt, M. M. Seibert, J. Sjöhamn, F. Stellato, U. Weierstall, N. A. Zatsepin, J. C. H. Spence, P. Fromme, I. Schlichting, S. Boutet, G. Groenhof, H. N. Chapman, R. Neutze, *Visualizing a Protein Quake with Time-Resolved X-Ray Scattering at a Free-Electron Laser*, *Nat. Methods* 11, 923-931 (2014). doi: 10.1038/nmeth.3067

[53] J. A. Sellberg, T. A. McQueen, H. Laksmono, S. Schreck, M. Beye, D. P. DePonte, B. O'Kennedy, D. Nordlund, R. G. Sierra, D. Schlesinger, T. Tokushima, I. Zhovtobriukh, S. Eckert, V. H. Segtnan, H. Ogasawara, K. Kubicek, S. Techert, U. Bergmann, G. L. Dakovski, W. F. Schlotter, Y. Harada, M. J. Bogan, P. Wernet, A. Föhlisch, L. G. M. Pettersson, A. Nilsson, *X-ray Emission Spectroscopy of Bulk Liquid Water in "No-man's Land"*, *J. Chem. Phys.* 142, 044505-12 (2015). doi:A14.08.0384R

[54] S. Schreck, A. Pietzsch, B. Kennedy, W. Quevedo, C. Siate, S. Techert, V. N. Strocov, T. Schmitt, F. Hennies, J.-E. Rubensson, A. Föhlisch, *Ground State Potential Energy Surfaces around Selected Atoms from Resonant Inelastic X-ray Scattering, Nat. Sci. Rep.* 7, 20054 (2016). doi: SREP-15-08332-T

[55] R. Jain, S. Techert, *Time-resolved and In-situ X-ray Scattering Methods beyond Photoactivation: Utilizing High-flux X-ray Sources for the Study of Ubiquitious Non-photon Active Proteins, Special Issue: "Synchrotron Applications in Life Science"*, *Prot. Pept. Lett.* 23, 01(2016). doi: 10.2174/0929866523666160106153847

[56] S. Bari, O. Gonzalez-Magaña, G. Reitsma, J. Werner, S. Schippers, R. Hoekstra, T. Schlathölter, *Photodissociation of Protonated Leucine-Enkephalin in the VUV Range of 8-40 eV*, *J. Chem. Phys.* 134, 24314 (2011). doi: 10.1063/1.3515301

[57] O. González-Magaña, M. Tiemens, G. Reitsma, L. Boschman, M. Door, S. Bari, P. O. Lahaie, J. R. Wagner, M. A. Huels, R. Hoekstra, T. Schlathölter, *Fragmentation of protonated oligonucleotides by energetic photons and Cq+ ions*, *Phys. Rev. A* 87, 032702-12 (2013). doi:10.1103/PhysRevA.87.032702

[58] *A. R. Milosavljević, C. Nicolas, M. L. Ranković, F. Canon, C. Miron, A. Giuliani*, *K-Shell Excitation and Ionization of a Gas-Phase Protein: Interplay between Electronic Structure and Protein Folding*, *J. Phys. Chem. Lett.* 6, 3132-3138 (2015). doi: 10.1021/acs.jpclett.5b01288

[59] J. Schulz, S. Bari, J. Buck, C. Uetrecht, *Sample refreshment schemes for high repetition rate FEL experiments, Proc. SPIE* 8778, Advances in X-Ray Free Electron Lasers II: Instrumentation, 87780T (2013). doi:10.1117/12.2019754

[60] T. Schlathölter, G. Reitsma, D. Egorov, O. Gonzalez-Magaña, S. Bari, L. Boschman, E. Bodewits, K. Schnorr, G. Schmid, C. D. Schröter, R. Moshammer, R. Hoekstra, *Multiple Ionization of Free Ubiquitin Molecular Ions in Extreme Ultraviolet Free-*

Electron Laser Pulses, *Angew. Chem. Int. Edit.* 55, 10741-10745 (2016). doi: 10.1002/anie.201605335

Ultrafast Imaging of Gas-Phase Reactions Allows Investigating Fundamental Definitions in Chemistry

[61] Strüder, L., Epp, S., Rolles, D., Hartmann, R., Holl, P., Lutz, G., Soltau, H., Eckart, R., Reich, C., Heinzinger, K., Thamm, C., Rudenko, A., Krasniqi, F., Kühnel, K.-U., Bauer, C., Schröter, C.-D., Moshammer, R., Techert, S., Miessner, D., Porro, M., Hölker, O., Meidinger, N., Kimmel, N., Andritschke, R., Schopper, F., Weidenspointner, G., Ziegler, A., Pietschner, D., Herrmann, S., Pietsch, U., Walenta, A., Leitenberger, W., Bostedt, C., Möller, T., Rupp, D., Adolph, M., Graafsma, H., Hirsemann, H., Gärtner, K., Richter, R., Foucar, L., Shoeman, R.L., Schlichting, I., Ullrich, J., 2010. Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources. Nuclear Instruments and Methods in Physics Research A 614, 483–496. doi:10.1016/j.nima.2009.12.053

[62] Barty, A., Caleman, C., Aquila, A., Timneanu, N., Lomb, L., White, T.A., Andreasson, J., Arnlund, D., Bajt, S., Barends, T.R.M., Barthelmess, M., Bogan, M.J., Bostedt, C., Bozek, J.D., Coffee, R., Coppola, N., Davidsson, J., DePonte, D.P., Doak, R.B., Ekeberg, T., Elser, V., Epp, S.W., Erk, B., Fleckenstein, H., Foucar, L., Fromme, P., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C.Y., Hartmann, R., Hartmann, A., Hauser, G., Hirsemann, H., Holl, P., Hunter, M.S., Johansson, L., Kassemeyer, S., Kimmel, N., Kirian, R.A., Liang, M., Maia, F.R.N.C., Malmerberg, E., Marchesini, S., Martin, A.V., Nass, K., Neutze, R., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Scott, H., Schlichting, I., Schulz, J., Seibert, M.M., Shoeman, R.L., Sierra, R.G., Soltau, H., Spence, J.C.H., Stellato, F., Stern, S., Strüder, L., Ullrich, J., Wang, X., Weidenspointner, G., Weierstall, U., Wunderer, C.B., Chapman, H.N., 2011. Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. Nature Photonics 6, 35–40. doi:10.1038/nphoton.2011.297

[63] Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A., Hunter, M.S., Schulz, J., DePonte, D.P., Weierstall, U., Doak, R.B., Maia, F.R.N.C., Martin, A.V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R.L., Epp, S.W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M.J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Homke, A., Reich, C., Pietschner, D., Struder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kuhnel, K.-U., Messerschmidt, M., Bozek, J.D., Hau-Riege, S.P., Frank, M., Hampton, C.Y., Sierra, R.G., Starodub, D., Williams, G.J., Hajdu, J., Timneanu, N., Seibert, M.M., Andreasson, J., Rocker, A., Jonsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schroter, C.-D., Krasniqi, F., Bott, M., Schmidt, K.E., Wang, X., Grotjohann, I., Holton, J.M., Barends, T.R.M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., Spence, J.C.H., 2011. Femtosecond X-ray protein nanocrystallography. Nature 470, 73–77. doi:10.1038/nature09750

[64] Gomez, L.F., Ferguson, K.R., Cryan, J.P., Bacellar, C., Tanyag, R.M.P., Jones, C., Schorb, S., Anielski, D., Belkacem, A., Bernando, C., Boll, R., Bozek, J., Carron, S., Chen, G., Delmas, T., Englert, L., Epp, S.W., Erk, B., Foucar, L., Hartmann, R.,

Hexemer, A., Huth, M., Kwok, J., Leone, S.R., Ma, J.H.S., Maia, F.R.N.C., Malmerberg, E., Marchesini, S., Neumark, D.M., Poon, B., Prell, J., Rolles, D., Rudek, B., Rudenko, A., Seifrid, M., Siefermann, K.R., Sturm, F.P., Swiggers, M., Ullrich, J., Weise, F., Zwart, P., Bostedt, C., Gessner, O., Vilesov, A.F., 2014. Shapes and vorticities of superfluid helium nanodroplets. Science 345, 906–909. doi:10.1126/science.1252395

[65] Gorkhover, T., Adolph, M., Rupp, D., Schorb, S., Epp, S.W., Erk, B., Foucar, L., Hartmann, R., Kimmel, N., Kühnel, K.-U., Rolles, D., Rudek, B., Rudenko, A., Andritschke, R., Aquila, A., Bozek, J.D., Coppola, N., Erke, T., Filsinger, F., Gorke, H., Graafsma, H., Gumprecht, L., Hauser, G., Herrmann, S., Hirsemann, H., Hömke, A., Holl, P., Kaiser, C., Krasniqi, F., Meyer, J.-H., Matysek, M., Messerschmidt, M., Miessner, D., Nilsson, B., Pietschner, D., Potdevin, G., Reich, C., Schaller, G., Schmidt, C., Schopper, F., Schröter, C.D., Schulz, J., Soltau, H., Weidenspointner, G., Schlichting, I., Strüder, L., Ullrich, J., Möller, T., Bostedt, C., 2012. Nanoplasma Dynamics of Single Large Xenon Clusters Irradiated with Superintense X-Ray Pulses from the Linac Coherent Light Source Free-Electron Laser. Physical Review Letters 108, 245005–245010. doi:10.1103/PhysRevLett.108.245005

[66] Gorkhover, T., Schorb, S., Coffee, R., Adolph, M., Foucar, L., Rupp, D., Aquila, A., Bozek, J.D., Epp, S.W., Erk, B., Gumprecht, L., Holmegaard, L., Hartmann, A., Hartmann, R., Hauser, G., Holl, P., Hömke, A., Johnsson, P., Kimmel, N., Kühnel, K.-U., Messerschmidt, M., Reich, C., Rouzée, A., Rudek, B., Schmidt, C., Schulz, J., Soltau, H., Stern, S., Weidenspointner, G., White, B., Küpper, J., Strüder, L., Schlichting, I., Ullrich, J., Rolles, D., Rudenko, A., Möller, T., Bostedt, C., 2016. Femtosecond and nanometre visualization of structural dynamics in superheated nanoparticles. Nature Photonics 10, 93–97. doi:10.1038/nphoton.2015.264

[67] Hau-Riege, S.P., Graf, A., Döppner, T., London, R.A., Krzywinski, J., Fortmann, C., Glenzer, S.H., Frank, M., Sokolowski-Tinten, K., Messerschmidt, M., Bostedt, C., Schorb, S., Bradley, J.A., Lutman, A., Rolles, D., Rudenko, A., Rudek, B., 2012. Ultrafast Transitions from Solid to Liquid and Plasma States of Graphite Induced by X-Ray Free-Electron Laser Pulses. Physical Review Letters 108. doi:10.1103/PhysRevLett.108.217402

[68] Loh, N.D., Hampton, C.Y., Martin, A.V., Starodub, D., Sierra, R.G., Barty, A., Aquila, A., Schulz, J., Lomb, L., Steinbrener, J., Shoeman, R.L., Kassemeyer, S., Bostedt, C., Bozek, J., Epp, S.W., Erk, B., Hartmann, R., Rolles, D., Rudenko, A., Rudek, B., Foucar, L., Kimmel, N., Weidenspointner, G., Hauser, G., Holl, P., Pedersoli, E., Liang, M., Hunter, M.M., Gumprecht, L., Coppola, N., Wunderer, C., Graafsma, H., Maia, F.R.N.C., Ekeberg, T., Hantke, M., Fleckenstein, H., Hirsemann, H., Nass, K., White, T.A., Tobias, H.J., Farquar, G.R., Benner, W.H., Hau-Riege, S.P., Reich, C., Hartmann, A., Soltau, H., Marchesini, S., Bajt, S., Barthelmess, M., Bucksbaum, P., Hodgson, K.O., Strüder, L., Ullrich, J., Frank, M., Schlichting, I., Chapman, H.N., Bogan, M.J., 2012. Fractal morphology, imaging and mass spectrometry of single aerosol particles in flight. Nature 486, 513–517. doi:10.1038/nature11222

[69] Seibert, M.M., Ekeberg, T., Maia, F.R.N.C., Svenda, M., Andreasson, J., Jönsson, O., Odić, D., Iwan, B., Rocker, A., Westphal, D., Hantke, M., DePonte, D.P., Barty, A., Schulz, J., Gumprecht, L., Coppola, N., Aquila, A., Liang, M., White, T.A., Martin, A., Caleman, C., Stern, S., Abergel, C., Seltzer, V., Claverie, J.-M., Bostedt, C., Bozek, J.D., Boutet, S., Miahnahri, A.A., Messerschmidt, M., Krzywinski, J., Williams, G., Hodgson, K.O., Bogan, M.J., Hampton, C.Y., Sierra, R.G., Starodub, D., Andersson, I., Bajt, S., Barthelmess, M., Spence, J.C.H., Fromme, P., Weierstall, U., Kirian, R., Hunter, M., Doak, R.B., Marchesini, S., Hau-Riege, S.P., Frank, M., Shoeman, R.L.,

Lomb, L., Epp, S.W., Hartmann, R., Rolles, D., Rudenko, A., Schmidt, C., Foucar, L., Kimmel, N., Holl, P., Rudek, B., Erk, B., Hömke, A., Reich, C., Pietschner, D., Weidenspointner, G., Strüder, L., Hauser, G., Gorke, H., Ullrich, J., Schlichting, I., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K.-U., Andritschke, R., Schröter, C.-D., Krasniqi, F., Bott, M., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., Chapman, H.N., Hajdu, J., 2011. Single mimivirus particles intercepted and imaged with an X-ray laser. Nature 470, 78–81. doi:10.1038/nature09748

[70] Boll, R., Anielski, D., Bostedt, C., Bozek, J., Christensen, L., Coffee, R., De, S., Decleva, P., Epp, S., Erk, B., Foucar, L., Krasniqi, F., Küpper, J., Rouzée, A., Rudek, B., Rudenko, A., Schorb, S., Stapelfeldt, H., Stener, M., Stern, S., Techert, S., Trippel, S., Vrakking, M., Ullrich, J., Rolles, D., 2013. Femtosecond photoelectron diffraction on laser-aligned molecules: Towards time-resolved imaging of molecular structure. Physical Review A 88, 061402(R). doi:10.1103/PhysRevA.88.061402

[71] Boll, R., Rouzée, A., Adolph, M., Anielski, D., Aquila, A., Bari, S., Bomme, C., Bostedt, C., Bozek, J.D., Chapman, H.N., Christensen, L., Coffee, R., Coppola, N., De, S., Decleva, P., Epp, S.W., Erk, B., Filsinger, F., Foucar, L., Gorkhover, T., Gumprecht, L., Hömke, A., Holmegaard, L., Johnsson, P., Kienitz, J.S., Kierspel, T., Krasniqi, F., Kühnel, K.-U., Maurer, J., Messerschmidt, M., Moshammer, R., Müller, N.L.M., Rudek, B., Savelyev, E., Schlichting, I., Schmidt, C., Scholz, F., Schorb, S., Schulz, J., Seltmann, J., Stener, M., Stern, S., Techert, S., Thøgersen, J., Trippel, S., Viefhaus, J., Vrakking, M., Stapelfeldt, H., Küpper, J., Ullrich, J., Rudenko, A., Rolles, D., 2014. Imaging molecular structure through femtosecond photoelectron diffraction on aligned and oriented gas-phase molecules. Faraday Discussions 171, 57–80. doi:10.1039/C4FD00037D

[72] Erk, B., Boll, R., Trippel, S., Anielski, D., Foucar, L., Rudek, B., Epp, S.W., Coffee, R., Carron, S., Schorb, S., Ferguson, K.R., Swiggers, M., Bozek, J.D., Simon, M., Marchenko, T., Küpper, J., Schlichting, I., Ullrich, J., Bostedt, C., Rolles, D., Rudenko, A., 2014. Imaging charge transfer in iodomethane upon x-ray photoabsorption. Science 345, 288–291. doi:10.1126/science.1253607

[73] Erk, B., Rolles, D., Foucar, L., Rudek, B., Epp, S.W., Cryle, M., Bostedt, C., Schorb, S., Bozek, J., Rouzee, A., Hundertmark, A., Marchenko, T., Simon, M., Filsinger, F., Christensen, L., De, S., Trippel, S., Küpper, J., Stapelfeldt, H., Wada, S., Ueda, K., Swiggers, M., Messerschmidt, M., Schröter, C.D., Moshammer, R., Schlichting, I., Ullrich, J., Rudenko, A., 2013. Ultrafast Charge Rearrangement and Nuclear Dynamics upon Inner-Shell Multiple Ionization of Small Polyatomic Molecules. Physical Review Letters 110, 053003. doi:10.1103/PhysRevLett.110.053003

[74] Rolles, D., Boll, R., Adolph, M., Aquila, A., Bostedt, C., Bozek, J.D., Chapman, H.N., Coffee, R., Coppola, N., Decleva, P., Delmas, T., Epp, S.W., Erk, B., Filsinger, F., Foucar, L., Gumprecht, L., Hömke, A., Gorkhover, T., Holmegaard, L., Johnsson, P., Kaiser, C., Krasniqi, F., Kühnel, K.-U., Maurer, J., Messerschmidt, M., Moshammer, R., Quevedo, W., Rajkovic, I., Rouzée, A., Rudek, B., Schlichting, I., Schmidt, C., Schorb, S., Schröter, C.D., Schulz, J., Stapelfeldt, H., Stener, M., Stern, S., Techert, S., Thøgersen, J., Vrakking, M.J.J., Rudenko, A., Küpper, J., Ullrich, J., 2014. Femtosecond x-ray photoelectron diffraction on gas-phase dibromobenzene molecules. Journal of Physics B 47, 124035. doi:10.1088/0953-4075/47/12/124035

[75] Rudek, B., Son, S.-K., Foucar, L., Epp, S.W., Erk, B., Hartmann, R., Adolph, M., Andritschke, R., Aquila, A., Berrah, N., Bostedt, C., Bozek, J., Coppola, N., Filsinger, F., Gorke, H., Gorkhover, T., Graafsma, H., Gumprecht, L., Hartmann, A., Hauser, G., Herrmann, S., Hirsemann, H., Holl, P., Hömke, A., Journel, L., Kaiser, C., Kimmel, N.,

Krasniqi, F., Kühnel, K.-U., Matysek, M., Messerschmidt, M., Miesner, D., Möller, T., Moshammer, R., Nagaya, K., Nilsson, B., Potdevin, G., Pietschner, D., Reich, C., Rupp, D., Schaller, G., Schlichting, I., Schmidt, C., Schopper, F., Schorb, S., Schröter, C.-D., Schulz, J., Simon, M., Soltau, H., Strüder, L., Ueda, K., Weidenspointner, G., Santra, R., Ullrich, J., Rudenko, A., Rolles, D., 2012. Ultra-efficient ionization of heavy atoms by intense X-ray free-electron laser pulses. Nature Photonics 6, 858–865. doi:10.1038/nphoton.2012.261

[76] Erk, B., Müller, J.P., Cédric, B., Boll, R., Brenner, G., Chapman, H.N., Correa, J., Dachraoui, H., Düsterer, S., Dziarzhytski, S., Eisebitt, S., Feldhaus, J., Graafsma, H., Grunewald, S., Gumprecht, L., Hartmann, R., Hauser, G., Keitel, B., Kuhlmann, M., Müller, E., Plönjes, E., Ramm, D., Rupp, D., Rompotis, D., Sauppe, M., Savelyev, E., Schlichting, I., Strüder, L., Swiderski, A., Techert, S., Tiedtke, K., Tilp, T., Treusch, R., Ullrich, J., Moshammer, R., Möller, T., Rolles, *CAMP@FLASH: an End-station for Imaging, Electron- and Ion-Spectroscopy and Pump–Probe Experiments at the FLASH Free-Electron Laser*, *J. Synchr. Rad.* **25 (5)**, 1-12 (2018). doi: 10.1107/S1600577518008585

[77] Ferguson, K.R., Bucher, M., Bozek, J.D., Carron, S., Castagna, J.-C., Coffee, R., Curiel, G.I., Holmes, M., Krzywinski, J., Messerschmidt, M., Minitti, M., Mitra, A., Moeller, S., Noonan, P., Osipov, T., Schorb, S., Swiggers, M., Wallace, A., Yin, J., Bostedt, C., 2015. The Atomic, Molecular and Optical Science instrument at the Linac Coherent Light Source. Journal of Synchrotron Radiation 22, 492–497. doi:10.1107/S1600577515004646

[78] Bozek, J.D., 2009. AMO instrumentation for the LCLS X-ray FEL. The European Physical Journal Special Topics 169, 129–132. doi:10.1140/epjst/e2009-00982-y

[79] Boll, R., Erk, B., Coffee, R., Trippel, S., Kierspel, T., Bomme, C., Bozek, J.D., Burkett, M., Carron, S., Ferguson, K.R., Foucar, L., Küpper, J., Marchenko, T., Miron, C., Patanen, M., Osipov, T., Schorb, S., Simon, M., Swiggers, M., Techert, S., Ueda, K., Bostedt, C., Rolles, D., Rudenko, A., 2016. Charge transfer in dissociating iodomethane and fluoromethane molecules ionized by intense femtosecond X-ray pulses. Structural Dynamics 3, 043207. doi:10.1063/1.4944344

[80] Fang, L., Osipov, T., Murphy, B., Tarantelli, F., Kukk, E., Cryan, J.P., Glownia, M., Bucksbaum, P.H., Coffee, R.N., Chen, M., Buth, C., Berrah, N., 2012. Multiphoton Ionization as a clock to Reveal Molecular Dynamics with Intense Short X-ray Free Electron Laser Pulses. Physical Review Letters 109, 263001. doi:10.1103/PhysRevLett.109.263001

[81] Hoener, M., Fang, L., Kornilov, O., Gessner, O., Pratt, S.T., Gühr, M., Kanter, E.P., Blaga, C., Bostedt, C., Bozek, J.D., Bucksbaum, P.H., Buth, C., Chen, M., Coffee, R., Cryan, J., DiMauro, L., Glownia, M., Hosler, E., Kukk, E., Leone, S.R., McFarland, B., Messerschmidt, M., Murphy, B., Petrovic, V., Rolles, D., Berrah, N., 2010. Ultraintense X-Ray Induced Ionization, Dissociation, and Frustrated Absorption in Molecular Nitrogen. Physical Review Letters 104. doi:10.1103/PhysRevLett.104.253002

[82] Liekhus-Schmaltz, C.E., Tenney, I., Osipov, T., Sanchez-Gonzalez, A., Berrah, N., Boll, R., Bomme, C., Bostedt, C., Bozek, J.D., Carron, S., Coffee, R., Devin, J., Erk, B., Ferguson, K.R., Field, R.W., Foucar, L., Frasinski, L.J., Glownia, J.M., Gühr, M., Kamalov, A., Krzywinski, J., Li, H., Marangos, J.P., Martinez, T.J., McFarland, B.K., Miyabe, S., Murphy, B., Natan, A., Rolles, D., Rudenko, A., Siano, M., Simpson, E.R., Spector, L., Swiggers, M., Walke, D., Wang, S., Weber, T., Bucksbaum, P.H., Petrovic, V.S., 2015. Ultrafast isomerization initiated by X-ray core ionization. Nature Communications 6, 8199. doi:10.1038/ncomms9199

[83] Meyer, M., Radcliffe, P., Tschentscher, T., Costello, J., Cavalieri, A., Grguras, I., Maier, A., Kienberger, R., Bozek, J., Bostedt, C., Schorb, S., Coffee, R., Messerschmidt, M., Roedig, C., Sistrunk, E., Di Mauro, L., Doumy, G., Ueda, K., Wada, S., Düsterer, S., Kazansky, A., Kabachnik, N., 2012. Angle-Resolved Electron Spectroscopy of Laser-Assisted Auger Decay Induced by a Few-Femtosecond X-Ray Pulse. Physical Review Letters 108. doi:10.1103/PhysRevLett.108.063007

[84] Murphy, B.F., Osipov, T., Jurek, Z., Fang, L., Son, S.-K., Mucke, M., Eland, J.H.D., Zhaunerchyk, V., Feifel, R., Avaldi, L., Bolognesi, P., Bostedt, C., Bozek, J.D., Grilj, J., Guehr, M., Frasinski, L.J., Glownia, J., Ha, D.T., Hoffmann, K., Kukk, E., McFarland, B.K., Miron, C., Sistrunk, E., Squibb, R.J., Ueda, K., Santra, R., Berrah, N., 2014. Femtosecond X-ray-induced explosion of C60 at extreme intensity. Nature Communications 5. doi:10.1038/ncomms5281

[85] Salén, P., van der Meulen, P., Schmidt, H.T., Thomas, R.D., Larsson, M., Feifel, R., Piancastelli, M.N., Fang, L., Murphy, B., Osipov, T., Berrah, N., Kukk, E., Ueda, K., Bozek, J.D., Bostedt, C., Wada, S., Richter, R., Feyer, V., Prince, K.C., 2012. Experimental Verification of the Chemical Sensitivity of Two-Site Double Core-Hole States Formed by an X-Ray Free-Electron Laser. Physical Review Letters 108. doi:10.1103/PhysRevLett.108.153003

[86] Young, L., Kanter, E.P., Krässig, B., Li, Y., March, A.M., Pratt, S.T., Santra, R., Southworth, S.H., Rohringer, N., DiMauro, L.F., Doumy, G., Roedig, C.A., Berrah, N., Fang, L., Hoener, M., Budsbaum, P.H., Cryan, J.P., Ghimire, S., Glownia, J.M., Reis, D.A., Bozek, J.D., Bostedt, C., Messerschmidt, M., 2010. Femtosecond electronic response of atoms to ultra-intense X-rays. Nature 466, 56–61. doi:10.1038/nature09177

D 9   Resonant x-ray scattering: spin, charge, and orbital order

M. Angst

Quantum Materials and Collective Phenomena
Jülich Centre for Neutron Science
Forschungszentrum Jülich GmbH

Contents

1   Introduction

Most of the structural information known about various crystalline compounds has been obtained by x-ray diffraction. "Normal" x-ray diffraction relies on the Thomson scattering process (c.f. chapter A3) and is sensitive to the electron density, and thus mainly to heavier atoms ($f_{\text{Th}} \propto Z$). In addition to Thomson scattering there is a second scattering process of x-rays on matter, termed anomalous scattering or resonant scattering. Resonant scattering is tightly connected with x-ray absorption (c.f. chapter F2), and as such is significant for the energy of incoming photons being close to the energy of an absorption edge. Because the energy of absorption edges depend on the element, resonant scattering is automatically element-sensitive. Resonant scattering can therefore be used to determine the location of elements within a structure that might be hard to distinguish with Thomson scattering because of a similar atomic number Z. The most significant contribution by resonant diffraction to structural information is in the field of macromolecular crystallography (c.f. chapters E2 and E7): so-called MAD (multiple anomalous diffraction or multiple anomalous dispersion), based on diffraction measurements with multiple photon energies near the absorption edge of a heavy element, often inserted by isomorphic replacement, is now routinely used to solve the phase problem for crystals of complex macromolecules [1].

Apart from its use in the solution of complex crystal structures, resonant scattering is increasingly also employed to access electronic properties. Being both a coherent scattering process amenable to interference and a process involving absorption and re-emission of a photon, resonant scattering combines spectroscopy and diffraction: All the sensitivity to crystal field levels, neighboring backscatterers, and even spin polarization present in absorption spectroscopy (XANES, EXAFS, XMCD; c.f. chapter F2) affect resonant scattering as well, with periodic variations of such properties being resolvable in addition.

Such periodic variations of electronic properties are particularly relevant in correlated electron materials [2]. Electrons in these materials have localizing tendencies, causing them to acquire "atomic like" degrees of freedom, such as ion valence, occupation of atomic orbitals, and local magnetic moment. However, electrons can still hop between different atomic sites, which facilitates interactions and under the right circumstances an ordered arrangement of these electronic degrees of freedom. For example, as illustrated in Fig. 1, an initially delocalized extra electron per two sites can localize in a periodic fashion (charge order CO or ion valence order), different atomic orbitals filled by the extra electron may also be periodically arranged (orbital order), and the same applies to the atomic magnetic moments. The different electronic degrees of freedom are strongly coupled, and they are coupled with the lattice as well. Out of this "soup" of interacting degrees of freedom, many phenomena and functionalities with applications potential may arise, including for example colossal magnetoresistance in perovskite manganites.
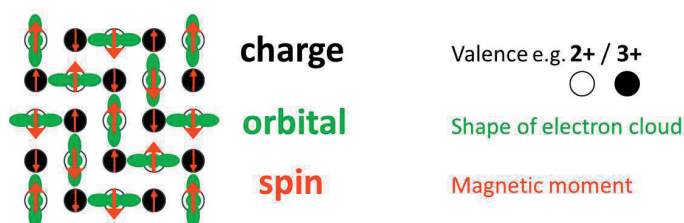
Fig. 1: Illustration of possible ordering patterns of electronic degrees of freedom in correlated electron materials.

Electronic orderings are difficult to see by Thomson scattering, because the associated relative change of the electron density is very small. Resonant x-ray scattering is directly sensitive to the electronic degrees of freedom. It has therefore developed into a key technique to elucidate the complex ordering phenomena in correlated materials, the focus of the present chapter. In the next section I will recapitalize origin and description of resonant x-ray scattering, including complex polarization dependence. The sections following this will be devoted to the use of resonant x-ray scattering to elucidate charge, orbital, and spin orderings, respectively. The chapter closes with a brief outline of resonant *inelastic* x-ray scattering (RIXS) and a summary of strengths and weaknesses, as well as an outlook to extensions not covered in detail.

2 Resonant x-ray scattering

2.1 Origin of resonant x-ray scattering

Whereas Thomson scattering can be described well within a classical picture of an electromagnetic wave forcing an electron into a harmonic oscillation, the starting point for the description of resonant scattering is necessarily the interaction between matter and x-rays in a quantum mechanical picture. For our purpose it is sufficient to consider the non-relativistic interaction of photons with a single atom (i.e. neglecting electron transfer and band structure effects, which is usually a good approximation). This is implemented by substituting in the Hamiltonian of the atom $\hat{\mathbf{p}} \rightarrow \hat{\mathbf{p}} - \frac{e}{c}\hat{\mathbf{A}}(\mathbf{r})$, yielding an interaction Hamiltonian consisting of two parts [3]:

$$\hat{\mathcal{H}}_{\text{int}} = \hat{\mathcal{H}}_{\text{int}}^{(1)} + \hat{\mathcal{H}}_{\text{int}}^{(2)} = \sum_j \frac{-e}{mc}\hat{\mathbf{A}}(\mathbf{r}_j) \cdot \hat{\mathbf{p}}_j + \sum_j \frac{e^2}{2mc^2}\hat{\mathbf{A}}(\mathbf{r}_j)^2, \tag{1}$$

with the vector potential in second quantization a sum over creation and annihilation operators of photons with a certain wavevector $\mathbf{k}$ and polarization state λ ($\hat{\varepsilon}$ is the polarization vector):

$$\hat{\mathbf{A}}(\mathbf{r}) \propto \sum_{\mathbf{k};\lambda} \frac{1}{\sqrt{k}}\hat{\varepsilon}_{\mathbf{k};\lambda}\left(\hat{a}_{\mathbf{k};\lambda}e^{-i\mathbf{k}\cdot\mathbf{r}} + \hat{a}_{\mathbf{k};\lambda}^{\dagger}e^{i\mathbf{k}\cdot\mathbf{r}}\right). \tag{2}$$

The interaction is then treated as a perturbation, and in first order (corresponding to Fermi's golden rule) $\hat{\mathcal{H}}_{\text{int}}^{(1)}$, being linear in creation and annihilation operators, can only lead to absorption[1] or emission of a photon, whereas $\hat{\mathcal{H}}_{\text{int}}^{(2)}$ can lead to scattering, annihilating a photon in state $\mathbf{k}; \lambda$ and creating one in state $\mathbf{k}'; \lambda'$: $d\sigma/d\Omega \propto |f_{\text{Th}}|^2$ and

$$f_{\text{Th}} \propto \left\langle \mathbf{k}'\lambda'; a \left| \hat{\mathcal{H}}_{\text{int}}^{(2)} \right| \mathbf{k}\lambda; a \right\rangle \propto \hat{\varepsilon}_{\mathbf{k}';\lambda'}^{\dagger} \cdot \hat{\varepsilon}_{\mathbf{k};\lambda} \left\langle a \left| e^{i\mathbf{Q}\cdot\mathbf{r}} \right| a \right\rangle = \hat{\varepsilon}_{\mathbf{k}';\lambda'}^{\dagger} \cdot \hat{\varepsilon}_{\mathbf{k};\lambda}\, \mathcal{F}\left\{|\Psi_a|^2\right\}(\mathbf{Q}), \tag{3}$$

which coincides with the classical Thomson scattering on an atom, including the polarization factor and the $\mathbf{Q} = \mathbf{k}' - \mathbf{k}$ dependent atomic form factor, the Fourier transform of the electron density (c.f. ch. A3)[2]. Perturbation in second order corresponds to transitions via intermediate states $|l\rangle$. In second order, there is a scattering contribution from $\hat{\mathcal{H}}_{\text{int}}^{(1)}$:

$$f_{\text{res}} \propto \sum_l \frac{\left\langle \mathbf{k}'\lambda'; a \left| \hat{\mathcal{H}}_{\text{int}}^{(1)\dagger} \right| 0; l \right\rangle \left\langle 0; l \left| \hat{\mathcal{H}}_{\text{int}}^{(1)} \right| \mathbf{k}\lambda; a \right\rangle}{E_l - E_a - \hbar\omega - i\Delta E_l}. \tag{4}$$

---

[1] As such, it is the relevant interaction for absorption spectroscopy described in chapter F2.

[2] Note that the relativistic version would yield very weak non-resonant magnetic scattering, also described in A3.
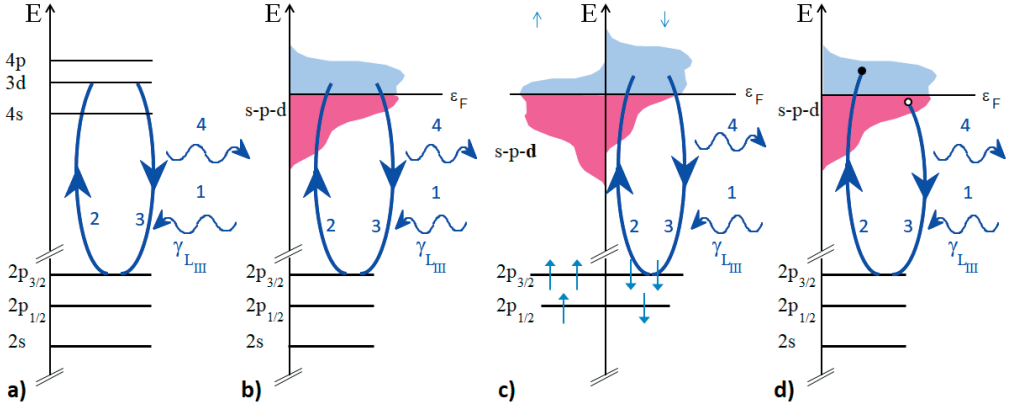
Fig. 2: a,b) Illustration resonant scattering process for the example of an L_{III} edge on an isolated atom (a) or an atom in a solid (b). An incoming photon of suitable energy (1) is absorbed by bumping (2) a $2p_{\frac{3}{2}}$ core electron to an empty 3d state. The electron then falls back (3) into its initial state, under emission of a photon of the same energy (4). c) As b), but spin-resolved. d) (direct) resonant inelastic scattering process, in which both energy and momentum is transferred to the solid.

Here, the intermediate state corresponds to the photon being absorbed by bumping up an electron from the ground state $|a\rangle$ to an excited state $|l\rangle$, which due to the possibility of spontaneous emission has a finite lifetime, corresponding to a non-zero half-width in energy ΔE_l. For an isolated atom, resonant scattering is illustrated in Fig. 2a. A contribution to resonant scattering is possible only for existing and empty intermediate states, here a 3d state filled by a $2p_{\frac{3}{2}}$ core electron. For photons of an energy corresponding to the energy difference between 3d and $2p_{\frac{3}{2}}$ the denominator in Eq. (4) becomes very small, leading to a large contribution to the scattering cross-section within a small photon energy range − hence the term resonant scattering. For atoms within a solid, intermediate state energies will be affected by crystal field effects (c.f. ch. B6), and broadened into bands by electron hopping (see Fig. 2b).

The matrix elements appearing in Eq. (4) are the same ones appearing in x-ray absorption and emission, indicating the natural connection between resonant scattering and absorption spectroscopy:

$$\left\langle 0; l \left| \hat{\mathcal{H}}_{\text{int}}^{(1)} \right| \mathbf{k}\lambda; a \right\rangle \propto \left\langle l \left| \sum_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \hat{\boldsymbol{\varepsilon}}_{\mathbf{k};\lambda} \cdot \hat{\mathbf{p}}_j \right| a \right\rangle. \tag{5}$$

Eq. (5) cannot be evaluated exactly, so instead a Taylor-expansion of the exponential is done, called multipole expansion in this context: $e^{-i\mathbf{k}\cdot\mathbf{r}_j} = 1 - i\mathbf{k} \cdot \mathbf{r}_j - \ldots$. The first contribution 1 is particularly simple and in most cases yields the by far dominant contribution to the resonant scattering, called electric dipole ($E1$) contribution. Further contributions correspondingly yield electric quadrupole, octupole, etc contributions, which despite being important in some circumstances will not be further treated in this chapter (nor will the less relevant magnetic multipole contributions that would have been obtained by using a relativistic interaction Hamiltonian). Because $|l\rangle$ and $|a\rangle$ are Eigenstates of the system Hamiltonian, the momentum operators $\hat{\mathbf{p}}_j$ can be transformed into position operators $\hat{\mathbf{r}}_j$. Introducing the electric dipole operator $\hat{\mathbf{D}} = e\sum_j \hat{\mathbf{r}}_j$,

the resonant scattering factor in dipole approximation becomes

$$f_{\text{res},E1}(\omega,\hat{\varepsilon}_{\mathbf{k'};\lambda'},\hat{\varepsilon}_{\mathbf{k};\lambda}) \propto \sum_l \frac{\left\langle a \left| \left(\hat{\varepsilon}_{\mathbf{k'};\lambda'} \cdot \hat{\mathbf{D}}\right)^{\dagger} \right| l \right\rangle \left\langle l \left| \hat{\varepsilon}_{\mathbf{k};\lambda} \cdot \hat{\mathbf{D}} \right| a \right\rangle}{E_l - E_a - \hbar\omega - \mathrm{i}\Delta E_l}, \tag{6}$$

where the dependence of the scattering factor on the photon energy $\hbar\omega$ and the polarizations $\hat{\varepsilon}$ of the incident and scattered photon has been made explicit. Note that $f_{\text{res},E1}$ does not explicitly depend on $\mathbf{k}$, $\mathbf{k'}$, or $\mathbf{Q}$, the latter in contrast to f_{Th}. This is a feature only of the dipole approximation, it is not true for resonant scattering in general.

Due to the symmetry of the dipole operator, the transition matrix elements in (6) can be non-zero only if certain *selection rules* are fulfilled between initial and intermediate state (as also treated in chapter F2):
- $\Delta\ell = \pm 1$
- $\Delta m_\ell = 0, \pm 1$ (depends on photon polarization state, c.f. Sec. 5)
- $\Delta s = \Delta m_s = 0$ (the dipole operator does not act on the spin[3])

Referring back to Fig. 2a, the indicated transitions $2\text{p}_{\frac{3}{2}} \rightarrow 3\text{d}$ are dipole-allowed, whereas $2\text{p}_{\frac{3}{2}} \rightarrow 4\text{p}$ is dipole-forbidden ($\Delta\ell = 0$) and could only happen at much lower rates via higher multipole transitions. This remains true in a solid. Of the free density of states above the Fermi-level, only the part with 3d-character is relevant for dipole resonant scattering at the L_{III} edge.

The atom is in the same state before and after the resonant scattering process, which implies that the scattering is coherent , i.e. interference between x-rays scattered resonantly on different atoms can take place. In a crystal, consequently Bragg diffraction occurs just as for Thomson scattering (and neutron scattering), just with the atomic scattering factor $f_{\text{at}} = f_{\text{Th}} + f_{\text{res}}$:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\mathbf{Q},\omega,\hat{\varepsilon}_{\mathbf{k'};\lambda'},\hat{\varepsilon}_{\mathbf{k};\lambda}) \propto |F(\mathbf{Q},\omega,\hat{\varepsilon}_{\mathbf{k'};\lambda'},\hat{\varepsilon}_{\mathbf{k};\lambda})|^2 \tag{7}$$

$$F(\mathbf{Q},\omega,\hat{\varepsilon}_{\mathbf{k'};\lambda'},\hat{\varepsilon}_{\mathbf{k};\lambda}) = \sum_j (f_{\text{Th},j}(\mathbf{Q},\hat{\varepsilon}_{\mathbf{k'};\lambda'},\hat{\varepsilon}_{\mathbf{k};\lambda}) + f_{\text{res},j}(\omega,\hat{\varepsilon}_{\mathbf{k'};\lambda'},\hat{\varepsilon}_{\mathbf{k};\lambda}))\, \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_j}. \tag{8}$$

In contrast to f_{Th}, which is real, the resonant scattering factor f_{res} is complex. In the literature this is often made explicit by writing for the atomic scattering factor $f = f_{\text{Th}} + f' + \mathrm{i}\, f''$. Because the index of refraction is linked to the forward scattering amplitude [4], $n(\omega) - 1 \propto F(\mathbf{Q} = 0, \omega)$ (dropping polarization-dependences for simplicity), the index of refraction becomes complex as well. The imaginary part corresponds to an attenuation of the x-ray beam within the material, i.e. to absorption. The absorption is obtained by carrying out the resonant process depicted in Fig. 2a,b only up to step 2, i.e. with the bumped up electron remaining in the higher level. As this is possible only if the incoming photon has sufficient energy (higher energy can go into kinetic energy of the photo electron), the absorption cross section, and therefore also the imaginary part of the refractive index, and the imaginary part of f_{res}, has a sudden increase at certain energy thresholds called absorption edges. The real and imaginary parts of the anomalous part of the dispersion, and of f_{res}, are connected by the so-called Kramers-Kronig relations based on causality (see chapter A3). As illustrated in Fig. 3, at the absorption edges, the jump in the imaginary part f_{res} (absorption) is accompanied by a downwards spike in the real part. Yttrium

---

[3] This is strict only in the absence of spin-orbit coupling. C.f. Sec. 5.
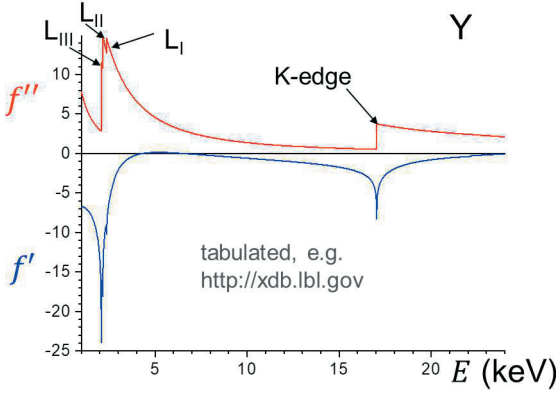
Fig. 3: Calculated energy-dependence of the anomalous scattering factors $f'' = \Im\{f_{\text{res}}\}$ (top, proportional to absorption cross section) and $f' = \Re\{f_{\text{res}}\}$ for Yttrium.

has $Z = 39$ electrons, so $f_{\text{Th}}(\mathbf{Q} = 0) = 39$: Even at forward scattering, the atomic scattering factor can experience dramatic corrections particularly in the region of the L edges. Because f_{Th} decreases with increasing $\mathbf{Q}$ whereas f_{res} does not, the modification by the latter becomes increasingly dominant at higher $\mathbf{Q}$ values. The connection between the absorption cross section of an atom and the complex scattering factor can be restated directly in the form of the *optical theorem* [4]:

$$\sigma_{\text{tot}} = 4\pi \, / \, k \, \cdot \Im\{f_{\text{at}}(\mathbf{Q} = 0)\}. \tag{9}$$

This theorem is of rather general validity, resting on basic conservation laws. For x-rays at not extremely high energies, the total cross section is dominated by the absorption cross section[4], and only f_{res} contributes an imaginary part. This is an independent confirmation of the deep connection between resonant scattering and absorption already apparent in the form of Eq. (4).

2.2 Polarization dependence: scattering factor as a tensor

For the consideration of dispersion and the optical theorem, we had neglected polarization dependences. Yet it is clear from chapter F2 that the absorption depends on the polarization of the incoming photons, and therefore so must the resonant scattering factor, as indeed made explicit in Eq. (6). The polarization dependence can be conveniently taken into account by defining the resonant scattering factor as a rank-two tensor (dropping the $E1$ dipole approximation indicator from now on):

$$\hat{\boldsymbol{\varepsilon}}_{\mathbf{k}';\lambda'}^{\dagger} \cdot \hat{f}_{\text{res}}(\omega) \cdot \hat{\boldsymbol{\varepsilon}}_{\mathbf{k};\lambda} = f_{\text{res},E1}(\omega, \hat{\boldsymbol{\varepsilon}}_{\mathbf{k}';\lambda'}, \hat{\boldsymbol{\varepsilon}}_{\mathbf{k};\lambda}). \tag{10}$$

Within this tensor formalism, the contribution to the total atomic scattering factor by Thomson scattering is simply a scalar, as the Thomson polarization dependence is $\hat{\boldsymbol{\varepsilon}}_{\mathbf{k}';\lambda'}^{\dagger} \cdot \hat{\boldsymbol{\varepsilon}}_{\mathbf{k};\lambda}$ according to (3). $\hat{f}_{\text{res}}$ depends on the symmetry of the atom (e.g. broken by a magnetic moment) and its surrounding, in particular the crystal field (c.f. ch. B6). For a free atom with no magnetic moment (magnetic moments will be considered in Sec. 5), the spherical symmetry implies that $\hat{f}_{\text{res}}$ is a scalar, i.e. the polarization dependence of the resonant scattering would be the same as for Thomson scattering. The same applies for atoms in sites with cubic site symmetry. However, for lower site symmetries, $\hat{f}_{\text{res}}$ will generally no longer be a scalar and more complex polarization dependences can be expected. Generally, for each symmetry element $\hat{S}$ of the point

---

[4] I.e. $\sigma_{\text{tot}} \simeq \sigma_{\text{abs}}$. Note that theorem (9) remains valid in tensor (Sec. 2.2) form, with $\hat{\boldsymbol{\varepsilon}}^{\dagger} \cdot \hat{\sigma}_{\text{abs}} \cdot \hat{\boldsymbol{\varepsilon}} \equiv \sigma_{\text{abs}}(\hat{\boldsymbol{\varepsilon}})$.
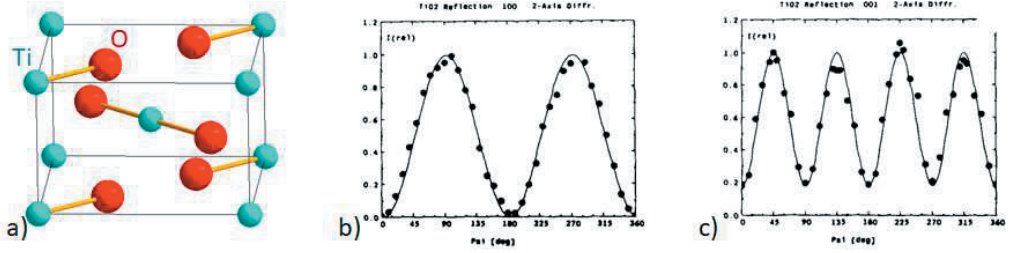
Fig. 4: a) Crystal structure of rutile TiO_2. b,c) Azimuth scans on two "forbidden" reflections [(100) in b, (001) in c], measured with the photon energy tuned to the Ti K edge. From [8]; reproduced with permission of the International Union of Crystallography.

group leaving the site invariant (c.f. ch. B1), $\hat{f}_{res} = \hat{S}^{-1}\hat{f}_{res}\hat{S}$ has to hold, i.e. $\hat{f}_{res}$ is invariant under these symmetry operations.

Consider for example an atom on a site with D_{2h} (mmm) symmetry. The above condition allows only three independent tensor elements. For a coordinate system that has the mirrors perpendicular to the x, y, z axes, the tensor is diagonal, whereas for a coordinate system with mirrors m_{001}, m_{110} and $m_{1\bar{1}0}$, as in the example below, the tensor has the form

$$\hat{f}_{res} = \begin{pmatrix} a & b & 0 \\ b & a & 0 \\ 0 & 0 & c \end{pmatrix}. \tag{11}$$

In general, there are more than one positions for the resonating atom. Within the tensor formalism, the crystallographic structure factor F introduced in Eq. (8) will be a tensor as well. The deviation from a scalar can have profound effects. A classical effect is the violation of reflection conditions resulting from glide reflections and screw rotations (see ch. B1). The violation occurs only upon resonance, and is due to the anisotropy of the tensor of susceptibility (i.e. of $\hat{f}_{res}$). It is therefore often called ATS scattering (sometimes also Templeton-Templeton scattering) [5–7]. For atom sites that are connected by a space-group element of the form $\mathbf{r}^{(2)} = \hat{S} \cdot \mathbf{r}^{(1)} + \mathbf{t}$, the tensor transforms according to $\hat{f}_{res}^{(2)} = \hat{S}^{-1}\hat{f}_{res}^{(1)}\hat{S}$. Let us consider as a concrete example rutile TiO_2. Rutile has a tetragonal structure (see Fig. 4a) with space group $P4_2/mnm$ (#136). Ti is at the Wyckoff position $2a$, with positions (000) and $(\frac{1}{2}\frac{1}{2}\frac{1}{2})$. The site symmetry is D_{2h}, with the mirrors such that the tensor at (000) has the form of Eq. (11). $(\frac{1}{2}\frac{1}{2}\frac{1}{2})$ is connected to (000) by n-glide mirrors, which transform $\hat{f}_{res}$ to the form

$$\hat{f}_{res}^{(\frac{1}{2}\frac{1}{2}\frac{1}{2})} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \hat{f}_{res}^{(000)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} a & -b & 0 \\ -b & a & 0 \\ 0 & 0 & c \end{pmatrix}. \tag{12}$$

The crystallographic structure factor (8) works out as

$$\hat{F}_{hk\ell} = \hat{f}_{res}^{(000)} + (-1)^{h+k+\ell} \cdot \hat{f}_{res}^{(\frac{1}{2}\frac{1}{2}\frac{1}{2})}, \tag{13}$$

which for $h+k+\ell$ even this evaluates to a diagonal tensor, but for $h+k+\ell$ odd an off-diagonal tensor is obtained, which is non-zero solely because of the anisotropy of the tensors $\hat{f}_{res}^{(000)}$
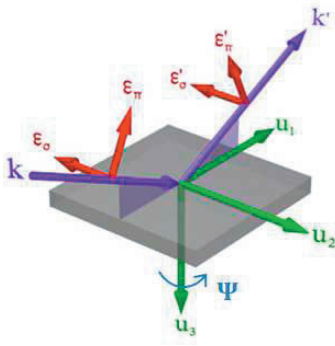
Fig. 5: Scattering geometry, adapted coordinate system and polarization vectors. Azimuth Ψ scans rotate the sample around $u_3 \| \mathbf{Q}$, which maintains the Bragg condition, but changes the direction of the polarization vectors relative to the x, y, z-coordinate system rotating with the sample. From [9], © 2012 EDP Sciences, Springer-Verlag; used under CC BY 4.0 / Added Ψ to the original.

and $\hat{f}_{\mathrm{res}}^{(\frac{1}{2}\frac{1}{2}\frac{1}{2})}$. Including the polarization dependences one obtains (abbreviating the polarization states)

$$\hat{\varepsilon}'^{\dagger} \cdot \hat{F}_{hk\ell}^{(h+k+\ell \text{ odd})} \cdot \hat{\varepsilon} = \hat{\varepsilon}'^{\dagger} \cdot 2b \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \hat{\varepsilon} = 2b \left(\hat{\varepsilon}_x \hat{\varepsilon}_y'^{\dagger} + \hat{\varepsilon}_y \hat{\varepsilon}_x'^{\dagger} \right). \tag{14}$$

To link this to actual experimental conditions, consider Fig. 5. Any linear polarization (see ch. C4 for the creation and analysis of polarized x-ray beams) of incoming and scattered beam can be decomposed into components perpendicular to the scattering plane ($\hat{\varepsilon}_\sigma$,$\hat{\varepsilon}_\sigma'$) and parallel to the scattering plane ($\hat{\varepsilon}_\pi$,$\hat{\varepsilon}_\pi'$), with components conveniently described in the adapted u_1, u_2, u_3 coordinate system. Keeping the Bragg condition, we still have the freedom to rotate the scattering plane (and u_1, u_2) around the $\mathbf{Q}$ vector (equivalent to a sample rotation around $\mathbf{Q}$)[5].

Sticking with the example of rutile, we consider the glide-plane forbidden (001) reflection, with incoming x-rays that are σ polarized and an azimuth definition $\Psi = 0$ for $u_1 \| x$. For arbitrary azimuth, the incoming polarization then has the x, y, z components $\hat{\varepsilon}_\sigma = (-\sin\Psi, -\cos\Psi, 0)$, and the scattered x-rays have $\hat{\varepsilon}_\sigma'^{\dagger} = \hat{\varepsilon}_\sigma' = \hat{\varepsilon}_\sigma$ and $\hat{\varepsilon}_\pi'^{\dagger} = \hat{\varepsilon}_\pi' = (-\sin\theta\cos\Psi, \sin\theta\sin\Psi, \cos\theta)$. Thus, with (14), we have $F_{001}^{\sigma\to\sigma'} = 2c\sin\Psi\cos\Psi = 4c\sin 2\Psi$. There is also scattering with rotation of the polarization, $\sigma \to \pi'$, which cannot happen for Thomson scattering or any scalar atomic scattering factor: $F_{001}^{\sigma\to\pi'} = 2c\sin\theta\cos 2\Psi$. Without polarization analysis, both these channels contribute equally to the intensity

$$I_{001}^{\sigma\to\sigma'+\pi'} = 4c^2 \left(\sin^2\theta + \cos^2\theta\sin^2 2\Psi \right). \tag{15}$$

This corresponds to a dependence of the resonant intensity upon the azimuth with four-fold symmetry, which might be expected given the tetragonal space group. It is indeed what Kirfel and coworkers [8] observed, when they investigated rutile with resonant scattering at the Ti K edge (see Fig. 4c, in which the symbols are experimental results and the line corresponds to Eq. (15). They also observed good agreement for other "forbidden" reflections, such as (100) (Fig. 4b), with different periodicities in Ψ. What makes the off-diagonal entry c in the resonant scattering tensor of rutile not only symmetry-allowed, but significant enough to be readily observed in the experiment is the very anisotropic environment, with oxygen atoms only in two rather than four directions. This will acts as an anisotropic crystal field (c.f. ch. B6), which splits the 4p levels, which according to the dipole selection rules (p. D9.5) are the intermediate states for the K edge.

---

[5] As an alternative to such azimuth scans, rotating $\hat{\varepsilon}$ around $\mathbf{k}$ by phase plates may be used (c.f. ch. C4).

Elements	Edge	Transition	E (keV)	Resonance	comment
3d	K	1s→4p	5-9	weak	small overlap
		1s→3d			dipole-forbidden, small overlap
3d	L_I	2s→3d	0.5-1.2	weak	dipole-forbidden, small overlap
3d	L_{II}, L_{III}	2p→3d	0.4-1.0	strong	large overlap,
					high spin-polarization of 3d
4f	K	1s→5p	40-63	weak	small overlap
4f	L_I	2s→5d	6.5-11	weak	dipole-forbidden, small overlap
4f	L_{II}, L_{III}	2p→5d	6-10	medium	5d spin-polarized by the 4f
		2p→4f			dipole-forb., 4f local moments
4f	M_I	3s→5p	1.4-2.5	weak	small overlap
4f	M_{II}, M_{III}	3p→5d	1.3-2.2	medium	
		3p→4f		to strong	dipole-forbidden
4f	M_{IV}, M_V	3d→4f	0.9-1.6	strong	large overlap,
					high spin polarization of 4f
5f	M_{IV}, M_V	3d→5f	3.3-3.9	strong	large overlap,
					high spin polarization of 5f

Table 1: Overview of absorption edges for different types of elements.

2.3   Spectroscopic aspects

From the considerations in Sec. 2.1, in particular Fig. 3, it is clear that significant resonant scattering can be expected only for photon energies close to an absorption edge. In the selection of an appropriate edge, one would first like a strong resonant enhancement, i.e. large transition matrix elements (5) to the intermediate states. This can be the case mostly only for dipole-allowed transitions, and in addition there should be a large overlap between the wavefunctions of the initial and intermediate states. The latter is generally not the case for K edges, which therefore are only weakly resonant. Furthermore, one would like to have intermediate states that are of direct interest. For transition metal elements, these are the 3d states, which is where the electronic ordering processes (c.f. Fig. 1) take place. For rare earth elements these are the 4f states. The properties of the different edges for various classes of elements are listed in table 1. As can be seen, the most interesting edges for 3d transition metal elements (L_{II}, L_{III}) as well as for rare earths (M_{IV}, M_V) are in the soft x-ray region of photon energies. This leads to particular experimental challenges: First, the long wavelength severely limits the **Q**-range that can be accessed. Of the superstructure reflections only very few (or even none in some cases) can typically be reached, and structural Bragg reflections are in most cases out of reach, making the sample alignment difficult; for the same reason crystals for polarization analysis (c.f. ch. C4) have to be replaced by graded multilayers [10]. Second, in the soft x-ray region absorption is a major issue. Even air absorbs too much, so that specially designed instruments maintaining UHV conditions have to be used. Furthermore, absorption by the sample itself implies that only a thin layer below the surface of the sample is actually probed. Nevertheless, resonant scattering with soft x-rays is a rapidly expanding field [11].

For $2p \rightarrow 3d$ transitions, strong spin-orbit coupling splits the initial state into two states with $j = 1/2$ (L_{II} edge) and $j = 3/2$ (L_{III} edge), by 10 to 20 eV. The same applies for several other tabulated transitions. Furthermore, the initial state energy is shifted by the valence (by a few eV), discussed in the next section. Even for a free atom, the intermediate states do not exactly correspond to the atomic multiplet structure, because the presence of the core hole left by the bumped up electron, which modifies the screening. Within a solid, the core hole has a localizing tendency, ameliorating charge transfer effects. An important impact of the environment on the intermediate states is made by the crystal field (see ch. B6) of the coordinating ligands, often oxygen, which splits up the intermediate state levels. While this splitting is small for the well localized 4f states, it can reach a few eV for 3d states. It can also affect 4p states, as seen in the example of rutile above. Further effects impacting the absorption fine structure and thus also the energy line shape for resonant scattering are discussed in chapter F2. The overall line shape is rather complex, yet it is crucial for the elucidation of the electronic degrees of freedom, as we will see. For simulations of spectra, either semi-empirical atomic multiplet calculations in a crystal field are done or density functional theory based cluster calculations, with codes such as FDMNES [12]. The latter, more "ab initio", approach works well for K edges, but often fails for L edges due to stronger interaction with the core hole and charge transfer effects.

3 Charge order and resonant scattering

The concept of charge order (CO) as the localization of extra electrons on some sites in a periodic fashion (as sketched in Fig. 1), driven by the Coulomb repulsion between these electrons, has great intuitive appeal. However, this picture basically assumes ionic bonding, whereas in nature bonds are never purely ionic, but e.g. in transition metal oxides (TMO) have at least partial covalent character. The nature of CO as a variation of actual charge at an atom as defined e.g. by integrating charge density within some sphere around the nucleus, is therefore questionable (see [13] for a fuller discussion). Nevertheless, the affected atoms of different valence states, do behave like atoms of the corresponding valence state in other compounds with fixed valence state. This applies to being Jahn-Teller active [2] or not, the magnetic moment associated with the atom, exchange interactions determining magnetic order, electric polarization effects, location of absorption edges (relevant for resonant scattering), and typical bond lengths to the coordinating ligands.

The last of these effects implies atom shifts with the same periodicity as the CO, and therefore superstructure reflections due to CO can also be observed with non-resonant Thomson or with neutron scattering: indeed, subsequent to a refinement (c.f. ch. D3) of the corresponding crystal structure the CO pattern can be assessed with the empirical bond-valence-sum, which deduces the valence state from the bondlengths to the ligands (see [2]). However, this deduction of CO is somewhat indirect, and in some cases where the charge modulation is only a small fraction of an electron the corresponding atom shifts may be difficult to see.

The fundamental aspect of CO in TMO is a different occupation of 3d states (this is indispensable e.g. for the change of the magnetic moment). This should affect resonant scattering at the $L_{II,III}$ edges, because only empty states are available as intermediate states. However, according to Eqs. (6) and (8), resonant scattering contrast may also arise in two other ways. First, atom shifts of the transition metal atom will lead to a resonant signal. Typically, the transition metal
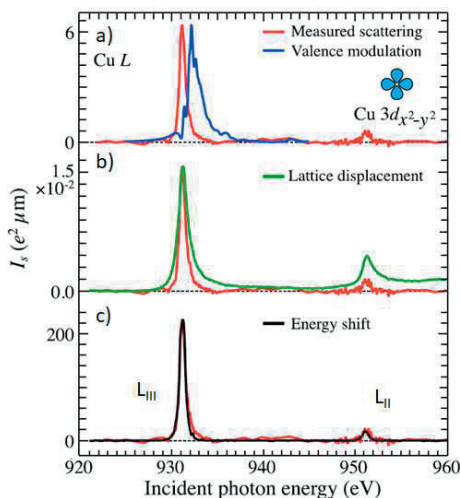
Fig. 6: Resonant scattering at the Cu L edges on the charge order reflection $(0.24, 0, 1.5)$ in $La_{1.475}Nd_{0.4}Sr_{0.125}CuO_4$. The experimental data (red lines) are compared with the calculated spectra according to 3d occupation (a), atom displacement (b), and core level energy shift (c). Reprinted with permission from [14], © 2012 American Physical Society.

atoms shift considerably less than the ligands do though. Second, a shift in the energy of the initial state, will also lead to a resonant signal. Such an energy shift does indeed occur[6]: for the higher valence state, the missing outer electron implies less effective screening of the nuclear charge potential is felt by the inner electron in the initial state, which is therefore more tightly bound. The lower energy of the initial state corresponds to an increase of the energy of the absorption edge, e.g. going from Fe^{2+} to Fe^{3+} by about $4\,eV$ for the K edge and by about $1.5\,eV$ for the L edges. Without detailed modeling, it is not clear how important each of the above effects is for the resonant scattering. Such a modeling has been carried out by Achkar *et al.* [14], who carried out resonant scattering on $La_{1.475}Nd_{0.4}Sr_{0.125}CuO_4$, which exhibits charge stripe order (like many cuprates) that might be relevant for the mechanism of high-temperature superconductivity in cuprates (c.f. ch. E1). The stripe order is incommensurate, leading to superstructure reflections at $(0.24, 0, 1.5)$, the energy spectrum through the L edges of which is shown in Fig. 6 (red lines). Comparison of the measured spectrum with spectra calculated for modulations of the 3d occupation (panel a), atom displacement (b), and core level energy shift (c) show the best fit for the model taking into account the energy shift only. According to the calculations, this is also by far the strongest contribution, with the contribution by the 3d modulation a factor of 30 smaller (and thus hard to confirm from the data directly), and the small shifts of the Cu atoms irrelevant (four orders of magnitude lower).

This indicates that the resonant scattering on charge ordered compounds is generally dominated by the energy shift of the core levels, even for absorption edges which directly probe the 3d states in which CO takes place. Hence, without very accurate modeling of the resonance spectra, the technique does not *directly* probe the 3d occupation, it rather probes the valence shift of the edges, an indirect effect, which is however generally linear with the amplitude of the charge modulation so that the latter can be estimated. A similar, though less accurate, estimation is also possible from structural information with the already mentioned bond-valence-sum method.

Given the reliance on on the core level energy shift resonant scattering for CO works as well in principle at the K edge as at the L edges. The main drawback is that the resonant signal will

---

[6] See also chapter F2 for these valence-induced edge shifts.
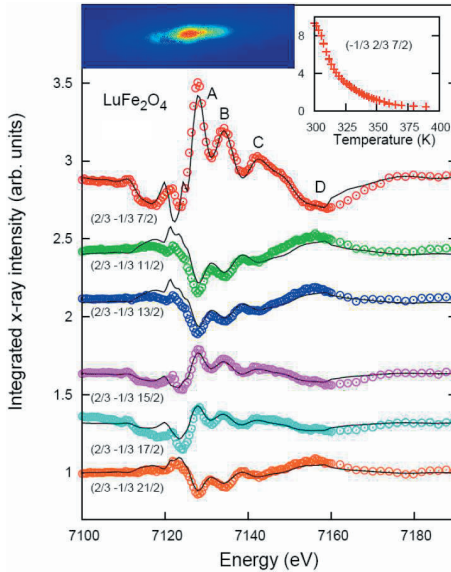
Fig. 7: Resonant scattering spectra at the Fe K edge on several charge order reflections in LuFe$_2$O$_4$, corrected for background, polarization, and attenuation by absorption. The experimental data (symbols) are compared with a model fit. Reprinted with permission from [15], © 2012 American Physical Society.

be much weaker (c.f. table 1), often smaller than the Thomson scattering, the effective intensity of which also changes going through the absorption edge due to the increasing attenuation of the primary and scattered beam, necessitating a correction prior to any modeling of the spectra. There is a major advantage, on the other hand, in that many superstructure reflections can be reached, facilitating the analysis of many spectra. This allows, with a model of the charge configuration, to fit the energy-dependence of the complex resonant scattering factors of the atoms in both valence states individually. Figure 7 shows an example of such spectra [15] measured at the Fe K edge on CO reflections of LuFe$_2$O$_4$, a compound which had attracted attention after being proposed as an example of multiferroicity originating from CO. The lines correspond to the best fits of individual resonant scattering factors for the two Fe valence states. The corresponding imaginary parts, proportional to absorption as discussed in Sec. 2.1, show a shift of the edge of $4\,\mathrm{eV}$, indicating full or nearly full Fe^{2+}/Fe^{3+} CO.

Given that CO resonant scattering is dominated by a simple energy shift, there is no reason why it would not be isotropic, i.e. have the same polarization dependence as Thomson scattering. The resonant scattering factors can thus be assumed to be scalars, which greatly simplifies the analysis. In principle, anisotropies may nevertheless be introduced by crystal field effects, which after all introduce anisotropies already for purely structural reflections as seen on the example of rutile above (Fig. 4). However, apart from reflections that may be "forbidden" by the space group describing the CO superstructure, deviations from isotropic behavior can be expected to be small, except for special circumstances like one of the valence states also exhibiting a "ferro" orbital order (see Sec. 4).

Given that the configuration of the CO and an estimation of the amplitude of the valence difference can also be obtained with non-resonant scattering by structure refinement and bond-valence-sum analysis, one might assume that the impact of resonant scattering on the understanding of materials exhibiting CO has been very limited. However, if the structural modulation is so weak that the corresponding superstructure reflections cannot or can only barely be detected, the structural route is often precluded. In these cases, the charge modulation will also
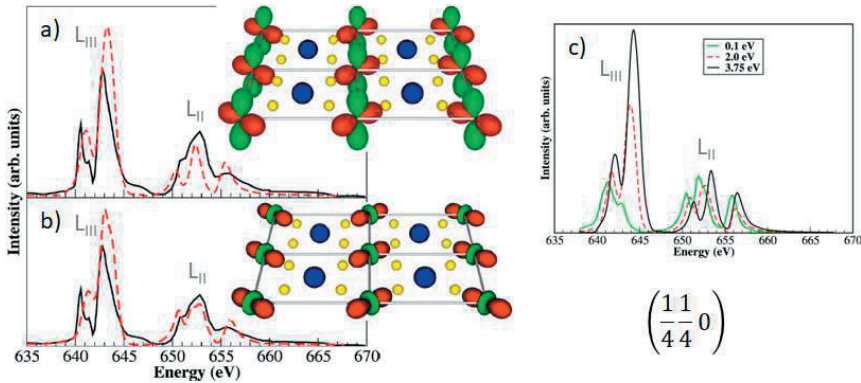
Fig. 8: Energy spectrum of resonant scattering at the Mn L edges on the orbital order reflection $(\frac{1}{4}\frac{1}{4}0)$ of La$_{0.5}$Sr$_{1.5}$MnO$_4$. a,b) The experimental data (black lines) are compared with best fits (red broken lines) of two distinct models of orbital order: $x^2 - y^2$ (a) and $3x^2 - r^2$ (b). c) Simulated effect of reducing the crystal field splitting. Reprinted with permission from [19], © 2012 American Physical Society.

be small, so that studies generally have to be conducted at the L edges, which have a stronger resonance. Many corresponding studies are reviewed in [11]. Probably the most active field is the ubiquitous CO (or charge density waves) in cuprate high-temperature superconductors [16], where the CO may be related to the exotic mechanism of superconductivity (c.f. ch. E1).

4   Orbital order and resonant scattering

Orbital order OO modulates the direction of occupied orbitals (typically 3d orbitals in transition metal oxides), as depicted in Fig. 1. The directionality of the occupied orbitals necessarily introduces anisotropy into the resonant scattering: here, deviations from Thomson-like polarization dependences should be expected and the scattering factors have to be treated as tensors. The first claims of a direct observation of orbital order was made on La$_{0.5}$Sr$_{1.5}$MnO$_4$ (LSMO) by resonant diffraction at the Mn K edge, based on the observation of a two-fold azimuthal angle dependence (very similar to the one shown in Fig. 4b) of the $(\frac{3}{4}\frac{3}{4}0)$ reflection [17]. However, resonant scattering at the K edge probes the 4p states rather than the 3d states in which OO takes place. Like for a valence change (c.f. Sec. 3), changed orbital occupation also results in shifts of the Ligand atoms in order to accommodate the filled orbital, a so-called Jahn-Teller distortion. These atom shifts make the orbital order, like charge order, indirectly visible also for non-resonant scattering [2]. Furthermore, the distorted crystal field will split the (empty) 4p leading to anisotropic scattering that can be explained entirely by the crystal structure, analogous to the ATS scattering observed on rutile (Fig. 4). In the case of LSMO, model calculations soon indicated that a contribution on the K edge resonant scattering by the direct $3d - 4p$ interaction is very small compared to the contribution by the shifted ligands [18]. Basically, K edge resonant scattering probes the Jahn-Teller distortion, which could also be obtained by non-resonant scattering.

This is different for resonant scattering at the L$_{II}$ and L$_{III}$ edges, which does probe the 3d

states [20]. The occupied orbital being unavailable *does* have a direct impact on resonant scattering. On the other hand there is also an impact by the energy shifts of the various 3d orbitals that are concomitant with the Jahn-Teller distortion. In order to disentangle these effects, experimental spectra have to be compared to model calculations. In modeling, the effect of Jahn-Teller distortion can be distinguished from the effect of orbital occupation by lowering the crystal field splitting, as exemplified for an OO reflection of LSMO (Fig. 8c): Reducing crystal field effects drastically lowers the intensity at the L_{III} edge, whereas the intensity at the L_{II} edge remains roughly the same. Fitting L edges resonant scattering spectra also allowed to distinguish between two alternative models of orbital ordering, which lead to similar displacements of the ligand atoms and thus could not be distinguished by non-resonant scattering: the model favored by simultaneous fitting of the OO reflection (Fig. 8a,b) and a magnetic reflection (not shown) has filled orbitals of $x^2 - y^2$ type (a) rather than of $3x^2 - r^2$ type (b) as previously thought [19].

Orbital order reflections occur due to a contrast between atoms with the same number of electrons, but different directions of the occupied orbital (c.f. e.g. sketch in Fig. 8a). This leaves only the anisotropic part of the scattering tensor to contribute. The situation is actually quite analogous to the "forbidden" reflections in rutile discussed in Sec. 2.2: Resonant scattering at a Ti edge does not directly take into account the oxygen position, the rutile structure (Fig. 4a) then simplifies to a *bcc* structure with reflections allowed only for $h + k + \ell = 2n$. The addition of oxygen provides for an anisotropy on Ti, which is different for the Ti at (000) and $(\frac{1}{2}\frac{1}{2}\frac{1}{2})$, analogous to a decoration by orbital order. The $h + k + \ell = 2n + 1$ reflections correspond to the superstructure reflections of this decoration by the O crystal field. The focus in the case of rutile on "forbidden" reflections is simply to avoid a large non-resonant and isotropic background by Thomson scattering from the oxygen atoms. In the case of orbital order, such a background is unavoidable because the ligand atom shifts by the Jahn-Teller distortions implies a Thomson scattering contribution (this is negligible at L edges with strong resonances, but can be significant for measurements at the K edge). Because the symmetry of the split (filled or empty) orbitals is coupled by the crystal field to the symmetry of the positions of the ligand atoms (c.f. ch. B6), the polarization dependences of the contributions due to OO (d orbital occupation) and due to the crystal field distortion are the same. If the distorted ligand positions, and thus the site symmetry, are known, the polarization dependence can be calculated in the same way as for ATS reflections (c.f. Sec. 2.2). In this case, azimuthal scans and polarization analysis will not yield new information for a *pure* OO reflection. Judicious settings of azimuth and polarization channel can, however, suppress the influence of Thomson scattering, and full analysis may help to decide whether or not there is a magnetic contribution to the intensity.

5   Spin order: x-ray resonant magnetic scattering

Spin order (SO) is in several respects different from the charge and orbital order considered in Secs. 3 and 4. First, spin order doesn't cause atom shifts of the same periodicity[7]. It is therefore not possible to determine the spin structure indirectly from Thomson scattering, and difficult to do so with non-resonant magnetic x-ray scattering because of very low intensities. It is possible to get the spin structure by magnetic neutron scattering (c.f. ch. A3), but this requires relatively large samples and generally can measure the sample as a whole only. In

---

[7] Exchange striction may lead to small atom shifts, but with half the magnetic periodicity.

the case of a presence of more than one magnetic elements, or of more than one valence state of the magnetic element, it is not directly sensitive to element or valence state. Second, the dipole operator in Eq. (6) and indeed the whole interaction Hamiltonian (1), does not operate on spins at all. It is therefore somewhat mysterious how resonant scattering can see SO. The reason is that the atomic Hamiltonian contains relativistic corrections: exchange splitting (c.f. Fig. 2c) and, particularly important, spin-orbit-coupling (SOC). SOC implies that for example with $2p_{\frac{3}{2}}$ and $2p_{\frac{1}{2}}$ as initial state split by SOC any orbital m_ℓ selectivity of the transition is partially transferred onto spin m_s selectivity (c.f. ch. F2). The orbital selectivity is there: if the dipole operator $\hat{D}$ in (6) is decomposed into components $\hat{D}_z$ and $\hat{D}_\pm = \hat{D}_x \pm i\hat{D}_y$, the m_ℓ dipole selection rule becomes $\Delta m_\ell = 0$ for $\hat{D}_z$ and $\Delta m_\ell = \pm 1$ for $\hat{D}_\pm$, which ties in with the corresponding components of the x-ray polarization ($\hat{\varepsilon}_\pm = \hat{\varepsilon}_x \pm i\hat{\varepsilon}_y$ corresponding to circular polarization). Third, the direction of the atomic spin (and thus quantization axis) $\hat{s}$ may break the site symmetry of the atom, leading to more independent entries of the atomic scattering tensor [11].

For a free atom with the spherical symmetry broken only by a magnetic moment in $\hat{s}$ direction, the resonant scattering factor has been worked out [21, 22] as

$$f_{\text{res},E1}(\omega, \hat{\varepsilon}', \hat{\varepsilon}) = (\hat{\varepsilon}'^\dagger \cdot \hat{\varepsilon})\, F^{(0)} - i(\hat{\varepsilon}'^\dagger \times \hat{\varepsilon}) \cdot \hat{s}\, F^{(1)} + (\hat{\varepsilon}'^\dagger \cdot \hat{s})(\hat{\varepsilon} \cdot \hat{s})\, F^{(2)}, \qquad (16)$$

which in tensor form corresponds to

$$\hat{f}_{\text{res},E1}(\omega) = F^{(0)} + iF^{(1)} \begin{pmatrix} 0 & -\hat{s}_z & \hat{s}_y \\ \hat{s}_z & 0 & -\hat{s}_x \\ -\hat{s}_y & \hat{s}_x & 0 \end{pmatrix} + F^{(2)} \begin{pmatrix} \hat{s}_x^2 & \hat{s}_x\hat{s}_y & \hat{s}_x\hat{s}_z \\ \hat{s}_x\hat{s}_y & \hat{s}_y^2 & \hat{s}_y\hat{s}_z \\ \hat{s}_x\hat{s}_z & \hat{s}_y\hat{s}_z & \hat{s}_z^2 \end{pmatrix}. \qquad (17)$$

Here, the complex factors $F^{(n)} = F^{(n)}(\omega)$ denote various linear combinations [21, 22] of transition probabilities to symmetry-appropriate intermediate states, weighed in particular by the probability that the corresponding intermediate states are empty. Their calculation is rather involved (c.f. Sec. 2.3). Therefore, it is generally not possible to accurately determine the absolute value of the ordered magnetic moment, in contrast to magnetic neutron scattering. Form (16) is strictly valid only for free magnetic atoms. For magnetic atoms in sites with lower than cubic symmetry (c.f. Sec. 2.2) an appropriately symmetry-adapted tensor-form has to be used [11].

The first term in Eqs. (16) and (17) does not depend on the atomic magnetic moment, it corresponds to a non-magnetic resonant scattering contribution that is isotropic, like Thomson scattering. The second term, linear in $\hat{s}$, leads to first-order satellite magnetic reflections, i.e. at the same position as reflections in magnetic neutron scattering. By the optical theorem (9) it is connected to x-ray magnetic circular dichroism in absorption (see ch. F2). The third term, quadratic in $\hat{s}$, leads to second-order satellite magnetic reflections. By the optical theorem, this term is connected to x-ray magnetic linear dichroism (ch. F2). The second order reflections are usually weak, and in commensurate magnetic structures often fall together with structural reflections.

The focus in the following will therefore be on the second term. It is convenient to express the polarization dependences in terms of σ and π polarization channels (c.f. Fig. 5). This can be done in matrix form [22], by defining (see also ch. C4):

$$\widetilde{f}_{\text{res},E1} \equiv \begin{pmatrix} \hat{\varepsilon}_\sigma^\dagger \cdot \hat{f}_{\text{res},E1} \cdot \hat{\varepsilon}_\sigma' & \hat{\varepsilon}_\sigma^\dagger \cdot \hat{f}_{\text{res},E1} \cdot \hat{\varepsilon}_\pi' \\ \hat{\varepsilon}_\pi^\dagger \cdot \hat{f}_{\text{res},E1} \cdot \hat{\varepsilon}_\sigma' & \hat{\varepsilon}_\pi^\dagger \cdot \hat{f}_{\text{res},E1} \cdot \hat{\varepsilon}_\pi' \end{pmatrix}. \qquad (18)$$
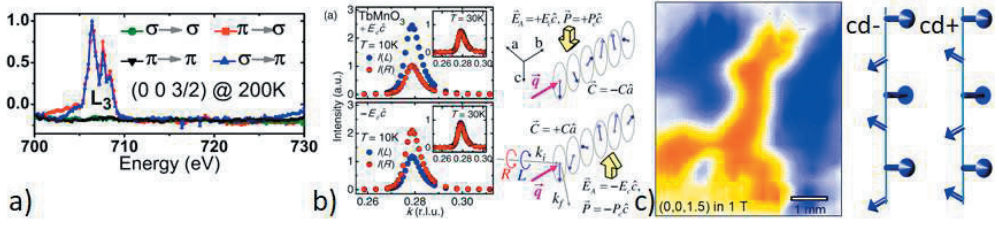
Fig. 9: a) Fe L_{III} edge resonant scattering with polarization analysis on the magnetic $(00\frac{3}{2})$ reflection of $LuFe_2O_4$. Reprinted with permission from [23], © 2012 American Physical Society. b) Scattered intensity along $(0k0)$ for circular polarized x-rays at Mn L_{II} edge on $TbMnO_3$ with opposite direction of applied electric field. Reprinted with permission from [24], © 2012 American Physical Society. c) Mapping of circular dichroic contrast $(I_+ - I_-)/(I_+ + I_-)$ of Fe L_{III} edge resonant scattering on magnetic reflection in a field-induced phase of $Ba_{0.5}Sr_{1.5}Zn_2Fe_{12}O_{22}$. Simplest possible model of chiral block spin structure leading to circular dichroism is also shown. After [25].

Eq. (17) then becomes, omitting the third term,

$$\widetilde{f}_{\text{res},E1} = F^{(0)} \begin{pmatrix} 1 & 0 \\ 0 & \cos 2\theta \end{pmatrix} - \mathrm{i}F^{(1)} \begin{pmatrix} 0 & \hat{s}_1 \cos\theta + \hat{s}_3 \sin\theta \\ \hat{s}_3 \sin\theta - \hat{s}_1 \cos\theta & -\hat{s}_2 \sin 2\theta \end{pmatrix} + \cdots \quad (19)$$

with the indices $1, 2, 3$ corresponding to the coordinate system shown in Fig. 5. Examining the $F^{(1)}$ term here, we see that $\sigma \to \sigma'$ magnetic scattering is absent, while the moment component perpendicular to the scattering plane gives rise to $\pi \to \pi'$ scattering and the other moment components give rise to $\sigma \to \pi'$ and $\pi \to \sigma'$ scattering with rotated polarization. The $F^{(0)}$ term here gives the polarization dependence of isotropic resonant scattering (equal to the one of Thomson scattering) in matrix form.

Consider the example of the Fe L edge spectrum of the $(00\frac{3}{2})$ reflection of $LuFe_2O_4$, a compound undergoing a complex charge and spin ordering, which was measured for all four polarization channels (Fig. 9a) [23]. The intensity was depressed too much by the polarization analysis [10] to discern the signal at the L_{II} edge, but the detailed form of the spectrum is not necessary to reach some conclusions. First, scattering in the $\sigma \to \sigma'$ channel is absent. This strongly indicates that the reflection is purely magnetic, as a charge order contribution would be expected to be at least approximately isotropic. Second, there is also no intensity in the $\pi \to \pi'$ channel, implying that the ordered magnetic moment has no component perpendicular to the scattering plane. Third, the intensities in the $\sigma \to \pi'$ and $\pi \to \sigma'$ channels are equal. This implies (the matrix elements will be squared to give the corresponding intensities, c.f. Eq. (7) that either $\hat{s}_1$ or $\hat{s}_3$ are zero. As the azimuth angle was chosen arbitrarily, it is much more likely that $\hat{s}_1$ is zero, i.e. we have a collinear spin structure with the moments pointing in $Q\|c$ direction. A potential contribution by orbital order can rather be excluded as well, since that would lead to complex interference phenomena [26] excluding the observed behavior except for special azimuths.

That the example taken is from L edge resonant scattering is no coincidence. Generally, resonant magnetic scattering signals at the K edge are extremely weak. This has several reasons: i) K edge resonances due to small overlap with non-s shells, ii) $1s \to 3d$ transitions are dipole-forbidden so that either quadrupole transitions have to be employed or only a possible induced

polarization in the 4p shell can be probed, iii) in contrast to 2p, 1s is not split by spin-orbit coupling so weaker SOC in the intermediate states has to be relied on. An exception is the K edge of oxygen, on which a large resonant enhancement was found and explained as polarization of the ligand O 2p states hybridized with unoccupied transition metal 3d states [27]. For transition metals, the need to use L edges limits resonant magnetic scattering to the soft x-ray region. For rare earths, significant signals can be obtained at the L edges in the hard x-ray region, mainly probing the 5d electrons, polarized by the 4f ones, or even the 4f states directly by non-negligible quadrupole transitions. In fact, the first observation of resonant magnetic scattering was on Holmium L edges, with both dipole and quadrupole contributions [28].

As pointed out above, the main term in resonant magnetic scattering is linked by the optical theorem (9) to circular dichroism in absorption. It seems natural then to wonder what effect resonant scattering with incoming circularly polarized x-rays will have. Calculating this out from Eq. (16), one obtains for the circular dichroism in the incoming x-rays (no polarization analysis) $I_+ - I_- \propto (\mathbf{s}(\mathbf{Q})^\dagger \times \mathbf{s}(\mathbf{Q})) \cdot \hat{\varepsilon}'_\pi$, i.e. this is sensitive to the vector spin chirality $\mathbf{s}_j \times \mathbf{s}_k$. This was exploited already 20 years ago, to study closure domains in ferromagnetic FePd films [29]. Spin chirality is of particular interest in a family of multiferroics, in which spin chirality induces an electric polarization, e.g. through a "spin-current" mechanism [30]. The inversion of the spin chirality then corresponds to an inversion of this polarization. In the "classic" (for this spin-induced multiferroicity) compound $TbMnO_3$, circular dichroism in resonant scattering was successfully used to show that the application of an electric field can invert spin chirality [24], as shown in Fig. 9b. In principle, the magnetic neutron scattering cross section also contains a term sensitive to the spin chirality. However, neutron fluxes are comparatively small, in effect implying a large beam so that only the average chirality within the whole sample can be probed. X-rays, in contrast, can be focused on a small spot of the sample, and with rastering the beam it is possible to map e.g. chiral domains. A recent example is shown in Fig. 9c [25]. The measurement was conducted on the $(00\frac{3}{2})$ reflection of $Ba_{0.5}Sr_{1.5}Zn_2Fe_{12}O_{22}$, at $10\,\mathrm{K}$ and $1\,\mathrm{T}$, i.e. under conditions for which field-induced ferroelectricity had been observed [31]. The large relative dichroic contrast in resonant scattering demonstrates the existence of magnetic domains of opposite chirality, which intriguingly is inverted upon inverting the magnetic field. The simplest "block" spin structure consistent with these chiralities (also shown in panel c) is consistent with ferroelectricity induced by the spin-current mechanism [30].

6 Electronic excitations: resonant inelastic x-ray scattering

So far we have looked at electronic degrees of freedom purely in terms of static ordering (Fig. 1). What about possible dynamics and excitations? Resonant scattering as depicted in Fig. 2 panels a to c, implied to be elastic, is not suitable to elucidate excitations, but there are also resonant *inelastic* scattering (RIXS) processes like the process (called direct RIXS) shown in Fig. 2d: an atomic core electron is bumped up to a (more delocalized) state of the appropriate symmetry above the Fermi level, and the core hole is then filled from another state below the Fermi energy, implying that the outgoing photon is of a bit lower energy than the incoming photon. On p. 5, it was pointed out that the scattering process was coherent because the state of the atom was unchanged. RIXS works as a coherent scattering process, because the intermediate state(s) are not localized on the atom. As an example consider a RIXS process bumping a core electron into an empty CEF-split 3d orbital and filling the core hole from a filled 3d orbital. As illustrated
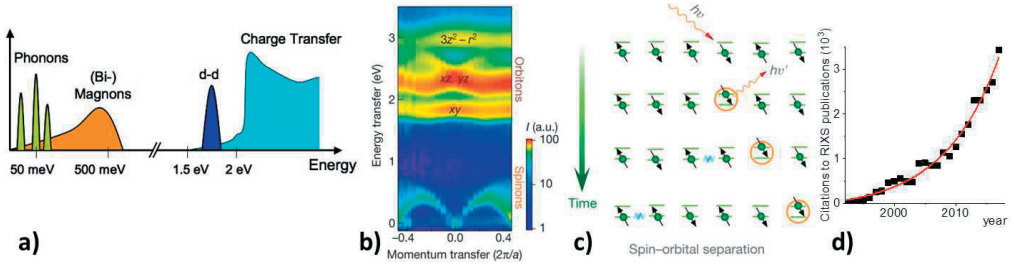
Fig. 10: a) Different elementary excitations in correlated materials, which RIXS is sensitive to, with approximate energy scales. Reprinted with permission from [33], © 2011 American Physical Society. b) RIXS (at Cu L_{III}) intensity map on Sr_2CuO_3 and c) time-evolution of orbital and spin degrees of freedom after excitation by RIXS. b,c) reprinted by permission from Nature Publishing Group [34], © 2012. d) Evolution of annual number of citations to RIXS publications, with exponential fit.

in Fig. 10c, this corresponds to a combined orbital (crystal field) and spin excitation, which subsequently propagate independently from site to site. I.e. these excitations are collective excitations with some crystal momentum. Fig. 10a shows various collective excitations that are amenable to RIXS investigations. At high energies beyond approximately $2\,eV$, charge transfer processes occur, i.e. transfer of a hole on the transition metal to the ligands, which then becomes mobile. At slightly lower energy $d-d$ (crystal field, orbital) excitations occur, usually with small dispersion, while spin excitations occur well below $1\,eV$, and lattice excitations (phonons[8]) below $\sim 100\,meV$.

An example of a mapping of the dispersion of $d-d$ and spin excitations is shown in Fig. 10b. In the investigated [34] Sr_2CuO_3, the ground state consists of one Cu 3d hole in the x^2-y^2 orbital, with antiferromagnetic order of its spin. Dispersive excitations between about 1.5 and $3\,eV$ can be assigned to excitations to different orbitals split by the crystal field, while highly dispersive excitations at lower energy correspond to spin excitations. In order for the hole excited to a higher orbital to become mobile, the AF alignment of the spins has to be broken, a so-called spinon, which subsequently moves independent of the orbital excitation.

RIXS, which within the scope of this introductory lecture cannot be treated more comprehensively (see [33] for a review), has made a large and recently exponentially increasing impact on the understanding of correlated electron materials (see Fig. 10d) for the evolution of the annual number of citations to RIXS publications. RIXS is particularly useful at edges in the soft x-ray region that directly probe the shells of interest: in contrast to resonant elastic scattering, the limited **Q** range is a less severe issue as there is always a significant range of the Brillouin zone around (000) that is reachable. The recently increasing importance of this technique is driven primarily by improvements in the (transfer) energy resolution. Having reached $\sim 1\,eV$ around 2001 and $\sim 100\,meV$ around 2008 [33], the resolution is now pushing $10\,meV$ and beyond [35, 36], on the way to reaching the resolutions of typical thermal neutron triple-axis spectrometers (c.f. ch. C2) and able to resolve more and more of e.g. the magnetic excitations (c.f. ch. B5 and D4).

---

[8] As RIXS is directly sensitive to electronic degrees of freedom, it offers the prospect of obtaining element and momentum-resolved electron-phonon coupling [32].

7 Summary and outlook

Resonant x-ray scattering has become an important technique for the study of ordering processes of electronic degrees of freedom, as well as of the corresponding excitations. For charge and orbital order, indirect information could also be obtained by normal (Thomson) x-ray scattering and refinements of the crystal structure (provided that the atom shifts are sufficiently large). The information from resonant scattering is also indirect, particularly charge order and for orbital order investigated at K edges, and usually does not provide direct evidence of the electronic degree of freedom being the driving force, as pointed out e.g. in [37]. However, resonant scattering is much more sensitive to weak modulations, and the L edge resonant scattering energy spectrum, while difficult to accurately model, *does* contain detailed electronic information not obtainable from the atomic positions alone.

For the spin degree of freedom, i.e. magnetism, resonant x-ray scattering mainly competes with magnetic neutron scattering. In this, resonant scattering has a few inherent disadvantages. Owing to the complex scattering process, it is generally not feasible to determine the absolute ordered magnetic moment. Furthermore, for the moments of transition metal atoms, feasible resonant signals can only be obtained in the soft x-ray region, which severely restricts the number of magnetic reflections that can be reached. Unknown complex magnetic structures can therefore not be solved by resonant scattering alone. On the other hand, there are inherent advantages as well. One of those is the element (and even valence) specifity, which is unique to the method. Another is the much better $\mathbf{Q}$ resolution, compared to neutron scattering, which e.g. allows to settle the question of whether a magnetic structure is exactly commensurate or not. A large advantage is furthermore that much smaller samples (thin films or crystals) can be investigates with x-rays than with neutrons; of new materials of interest, often only small crystals can be obtained, or small crystals are of significantly better quality than larger ones. Furthermore, samples containing elements with a large neutron absorption cross-section can also be alternatively studied with resonant x-ray scattering. Finally, the x-ray beam can be focused on a small spot and rastered over the sample, obtaining maps of (e.g. chiral) magnetic domains.

Within the scope of this introduction to resonant x-ray scattering, a number of exciting further possibilities had to be omitted. We restricted ourselves to transitions within the dipole approximation. Within resonant scattering research, quadrupole (and even higher) transitions are employed as well (for magnetic resonant scattering, [22] contains also the terms for quadrupole scattering, with generally much more complex polarization dependences). Taking into account up to quadrupole transitions, additional ordering phenomena can be probed, such as the ordered arrangement of octupoles [38] or toroidal moments [39]. As examples we mostly used single crystal diffraction (c.f. ch. D3), only briefly outlining inelastic (c.f. ch. D4) scattering (RIXS), which has recently become more and more important. Resonant scattering contrast not only works for crystals, but for most of the techniques outlined within this book: One can investigate for example magnetic nanoparticles with resonant small-angle scattering (c.f. ch. D1), or interface-effects in heterostructures with resonant grazing incidence scattering (c.f. ch. D2), image magnetic nanostructures with resonant coherent diffraction (c.f. ch. D7), or ultra-fast processes involving electronic degrees of freedom with pump-(resonant)probe experiments (c.f. ch. D8).

References

[1] K. R. Legate, http://doi.org/10.1038/nature13367 (2014).

[2] M. Angst, ch. 12 "Correlated electrons", in "Laboratory Course Neutron Scattering Lectures", Th. Brückel, S. Förster, G. Roth, R. Zorn (Eds.), key technologies vol. 172 (Forschungszentrum Jülich GmbH, 2018). http://hdl.handle.net/2128/19595.

[3] J. J. Sakurai, Advanced Quantum Mechanics (Series in Advanced Physics) (Reading, MA: Addison-Wesley,1976).

[4] R. G. Newton, *Optical theorem and beyond*, Am. J. Phys. **44**, 639 (1976).

[5] D. H. Templeton, L. K. Templeton, Acta Cryst. A **36**, 237 (1980).

[6] V. E. Dmitrienko, Acta Cryst. A **39**, 29 (1983).

[7] V. E. Dmitrienko, K. Ishida, A. Kirfel, E. N. Ovchinnikova, Acta Cryst. A **61**, 481 (2005).

[8] A. Kirfel, A. Petcov, K. Eichhorn, Acta Cryst. A **47**, 180 (1991).

[9] C. Detlefs, M. Sanchez del Rio, C. Mazzoli, Eur. Phys. J. Spec. Top. **208**, 359 (2012).

[10] U. Staub, V. Scagnoli, Y. Bodenthin, M. García-Fernández, R. Wetter, A. M. Mulders, H. Grimmer, M. Horisberger, J. Synch. Rad. **15**, 469 (2008).

[11] J. Fink, E. Schierle, E. Weschke, J. Geck, Rep. Prog. Phys. **76**, 056502 (2013).

[12] Y. Joly, Phys. Rev. B **63**, 125120 (2001).

[13] M. Angst, ch. 3 in 'Multiferroic Materials - Properties, Techniques, and Applications', J. Wang (Ed.), pp. 67-97 (CRC Press, Taylor & Francis Group, Boca Raton, 2016).

[14] A. J. Achkar, F. He, R. Sutarto, J. Geck, H. Zhang, Y.-J. Kim, D. G. Hawthorn, Phys. Rev. Lett. **110**, 017001 (2013).

[15] A. M. Mulders, S. M. Lawrence, U. Staub, M. Garcia-Fernandez, V. Scagnoli, C. Mazzoli, E. Pomjakushina, K. Conder, Y. Wang, Phys. Rev. Lett. **103**, 077602 (2009).

[16] R. Comin, A. Damascelli, Ann. Rev. Cond. Mat. Phys. **7**, 369 (2016).

[17] Y. Murakami, H. Kawada, H. Kawata, M. Tanaka, T. Arima, Y. Moritomo, Y. Tokura, Phys. Rev. Lett. **80**, 1932 (1998).

[18] M. Benfatto, Y. Joly, C. R. Natoli, Phys. Rev. Lett. **83**, 636 (1999).

[19] S. B. Wilkins, N. Stojić, T. A. W. Beale, N. Binggeli, C. W. M. Castleton, P. Bencok, D. Prabhakaran, A. T. Boothroyd, P. D. Hatton, M. Altarelli, Phys. Rev. B **71** (2005).

[20] C. W. M. Castleton, M. Altarelli, Phys. Rev. B **62**, 1033 (2000).

[21] J. P. Hannon, G. T. Trammell, M. Blume, D. Gibbs, Phys. Rev. Lett. **61**, 1245 (1988).

[22] J. P. Hill, D. F. McMorrow, Acta Cryst. A **52**, 236 (1996).

[23] J. de Groot, K. Marty, M. D. Lumsden, A. D. Christianson, S. E. Nagler, S. Adiga, W. J. H. Borghols, K. Schmalzl, Z. Yamani, S. R. Bland, R. de Souza, U. Staub, W. Schweika, Y. Su, M. Angst, Phys. Rev. Lett. **108**, 037206 (2012).

[24] H. Jang, J.-S. Lee, K.-T. Ko, W.-S. Noh, T. Y. Koo, J.-Y. Kim, K.-B. Lee, J.-H. Park, C. L. Zhang, S. B. Kim, S.-W. Cheong, Phys. Rev. Lett. **106**, 047203 (2011).

[25] P. Thakuria, Ph.D. thesis, RWTH Aachen University (2017).

[26] S. Ji, C. Song, J. Koo, K.-B. Lee, Y. J. Park, J. Y. Kim, J.-H. Park, H. J. Shin, J. S. Rhyee, B. H. Oh, B. K. Cho, Phys. Rev. Lett. **91**, 257205 (2003).

[27] T. A. W. Beale, S. B. Wilkins, R. D. Johnson, S. R. Bland, Y. Joly, T. R. Forrest, D. F. McMorrow, F. Yakhou, D. Prabhakaran, A. T. Boothroyd, P. D. Hatton, Phys. Rev. Lett. **105**, 087203 (2010).

[28] D. Gibbs, D. R. Harshman, E. D. Isaacs, D. B. McWhan, D. Mills, C. Vettier, Phys. Rev. Lett. **61**, 1241 (1988).

[29] H. A. Dürr, E. Dudzik, S. S. Dhesi, J. B. Goedkoop, G. van der Laan, M. Belakhovsky, C. Mocuta, A. Marty, Y. Samson, Science **284**, 2166 (1999).

[30] H. Katsura, N. Nagaosa, A. Balatsky, Phys. Rev. Lett. **95**, 057205 (2005).

[31] T. Kimura, G. Lawes, A. P. Ramirez, Phys. Rev. Lett. **94**, 137201 (2005).

[32] L. J. P. Ament, M. van Veenendaal, J. van den Brink, EPL **95**, 27008 (2011).

[33] L. J. P. Ament, M. van Veenendaal, T. P. Devereaux, J. P. Hill, J. van den Brink, Rev. Mod. Phys. **83**, 705 (2011).

[34] J. Schlappa, K. Wohlfeld, K. J. Zhou, M. Mourigal, M. W. Haverkort, V. N. Strocov, L. Hozoi, C. Monney, S. Nishimoto, S. Singh, A. Revcolevschi, J.-S. Caux, L. Patthey, H. M. Rønnow, J. van den Brink, T. Schmitt, Nature **485**, 82 (2012).

[35] J. Dvorak, I. Jarrige, V. Bisogni, S. Coburn, W. Leonhardt, Rev. Sci. Instrum. **87**, 115109 (2016).

[36] J. Kim, D. Casa, A. Said, R. Krakora, B. J. Kim, E. Kasman, X. Huang, T. Gog, Sci. Rept. **8**, 1958 (2018).

[37] S. D. Matteo, J. Phys.: Conf. Series **190**, 012008 (2009).

[38] T. Nagao, J. ichi Igarashi, Phys. Rev. B **74**, 104404 (2006).

[39] T. Arima, J.-H. Jung, M. Matsubara, M. Kubota, J.-P. He, Y. Kaneko, Y. Tokura, J. Phys. Soc. Jpn. **74**, 1419 (2005).

E 1 Superconductivity

Y. Xiao

Shenzhen Graduate School

Peking University, China

Contents

1 Introduction

Superconductivity is an exotic state of matter that has fascinated generations of scientists ever since its discovery in mercury in 1911 [1]. After more than 100 years, there are whole classes of superconducting materials that we still do not fully understand. In particular, understanding the mechanism of high temperature superconductivity in cuprates discovered in 1986 by Bednorz and Müller [2], and in iron-based superconductors [3] has remained as one of the hardest tasks in condensed matter physics.

In this lecture note, the fundamentals of superconductivity, as commonly covered in most textbooks on solid state physics, will be given first. Detailed descriptions on the phenomenon of superconductivity, phenomenological theories i.e. the London theory and Ginzburg-Landau theory and the landmark microscopic BCS theory can be found in this section. A brief overview of large classes of the known superconducting materials, including both conventional and unconventional superconductors will be given as well. The high temperature superconductivity is a vast and always rapidly moving field, it is neither in our intention nor possible to present a comprehensive review on the current status as well as a thorough description of the underlying physics. Therefore, only some important aspects and basic understanding of the high-T_c superconductors will be discussed here. The choice of the covered topic is more or less based on our own research activities in this field with the main focus on the neutron scattering studies.

2 Fundamentals of Superconductivity

2.1 Zero resistivity

In order to explain the electronic property of metals, the Drude theory was developed by Drude in 1900. In the framework of the Drude theory, electrons are treated as classical particles. When electrons are moving through the solid, they will collide with scattering center and change their direction and velocity. In a metal, the conductivity can be defined by the constitutive equation as the proportionality between electrical current density $\mathbf{J}$ and electric field $\mathbf{E}$: $\mathbf{J}=\sigma\mathbf{E}$. The electrical conductivity σ is given by Drude theory as:

$$\sigma = \frac{ne^2\tau}{m} \tag{1}$$

where τ is the mean life time and m is the effective mass of the conduction electrons. The resistivity ρ is the reciprocal of the conductivity. i.e. $\rho = 1/\sigma$ and $\rho \propto \tau^{-1}$. The resistivity is the sum of the contributions from different scattering processes. The scattering mechanisms can be impurity scattering, electron-electron scattering and electron-phonon scattering. Furthermore, these scattering processes act independently and they have different mean life times. The total resistivity reads:

$$\rho = \frac{m}{ne^2}(\tau_i^{-1} + \tau_{e-e}^{-1} + \tau_{e-p}^{-1}) \tag{2}$$

The temperature dependencies of these mean life times are also different. The impurity s-cattering life time is independent of temperature, while the lifetimes of electron-electron and electron-phonon scattering will exhibit temperature variation. At low temperatures, electron-electron scattering and electron-phonon scattering are negligible. Therefore, one expects a constant value for resistivity at zero temperature as the residual resistivity.
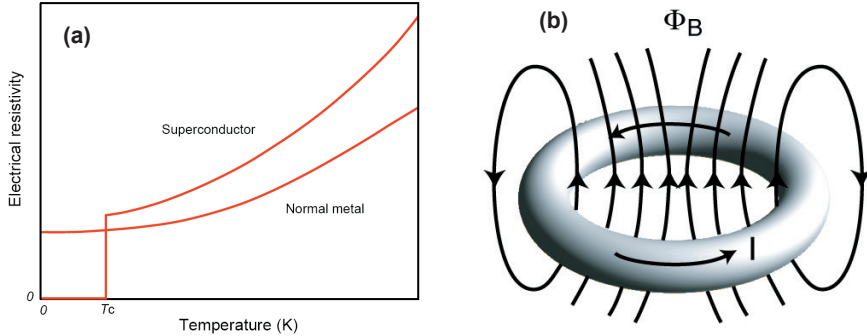
Fig. 1: (a): Temperature dependence of the electrical resistivity for superconductors and normal metals. (b): The superconducting current maintains through the superconducting ring [4,5].

In 1911, Heike Kamerlingh Onnes performed an experiment to test the validity of the Drude theory by measuring the resistivity of mercury at low temperature. Surprisingly, he found the electrical resistance dropped sharply to zero below 4.2 K [1]. Thus the superconductivity was discovered and it represents a new state of matter [4,5].

As shown in Fig. 1(a), the resistivity of superconductor drops to zero when temperature is below the critical temperature T_c. However, due to the experimental difficulty, we are not be able to measure the zero resistivity. The existence of persistent current in a closed loop of superconducting wire can be considered as the strong evidence of superconductivity. As shown in Fig. 1(b), a circulating current I can be introduced in the superconducting ring. If the superconductor has the zero resistivity, the energy stored in the ring will keep constant and the current will continue flowing in the ring. Experimentally, it was found that almost no detectable decay of the current in superconducting ring for years.

2.2   Meissner effect

Suppose we take a superconductor and place it under magnetic field, the external magnetic field will penetrate into the superconductor if the superconductor is in its normal state, *i.e.* $T > T_c$. Then we will get almost same value of magnetic field inside and outside of superconductor as indicated in Fig. 2(a). When we cool the superconducting sample below T_c in the presence of the same field, the magnetic field will be expelled from the sample. This phenomenon was discovered by Meissner in 1933 and named as Meissner effect [6].

As we known, the superconductor exhibits zero resistivity. By $\mathbf{E} = \rho\mathbf{J}$, we will have $\mathbf{E} = 0$ inside of superconductor. By using the Maxwell equation $\nabla\times\mathbf{E} = -\partial\mathbf{B}/\partial t$, we got $\partial\mathbf{B}/\partial t = 0$. It is also known that magnetic flux density $\mathbf{B}$ is related with magnetic field $\mathbf{H}$ and magnetization
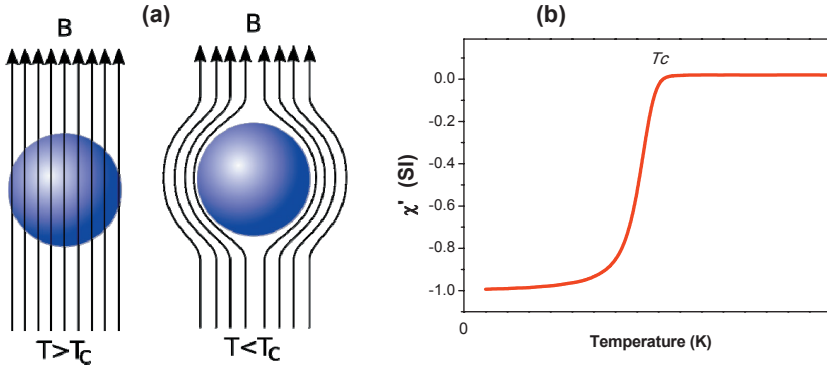
Fig. 2: (a): The Meissner effect in superconductors. (b): Temperature dependence of suscepti-bility (χ) of bulk superconductor [4, 5].

of sample **M** by **B** = μ_0(**H**+**M**). Eventually, we find that $\partial \mathbf{M}/\partial \mathbf{H} = \chi = -1$ for superconductor. Susceptibility $\chi = -1$ indicated that the superconductors possess perfect diamagnetism. To study the magnetic susceptibility as a function of temperature, we will be able to characterize the superconducting sample. As shown in Fig. 2(b), we will obtain $\chi = -1$ for bulk superconductor below T_c, which is the solid evidence for Meissner effect.

2.3 London theory

In 1935, London brothers developed the first theory to explain the magnetic properties of su-perconductors [7]. By applying the two-fluid model, the 1st London equation can be obtained, which relate the superconducting current density **J** with the electric field **E**:

$$\mathbf{E} = \mu_0 \lambda_L^2 \frac{\partial \mathbf{J}}{\partial t} \tag{3}$$

Combination of Eq. (3) with Maxwell equation $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$, London equation can also be rewritten in terms of magnetic field **B** and superconducting current density **J**, which is called 2nd London equation:

$$\mathbf{B} = -\mu_0 \lambda_L^2 \nabla \times \mathbf{J} \tag{4}$$

In both Eq. (3) and Eq. (4), λ_L is the London penetration depth with the dimension of length,

$$\lambda_L = \left(\frac{m_e}{\mu_0 n_s e^2}\right)^{1/2} \tag{5}$$

The London equations provide a simple phenomenological model to explain the Meissner effect. It also implies that the magnetic field will only penetrate the surface layer of depth λ_L and the field equals to zero inside the bulk superconductor.

2.4 Ginzburg-Landau theory

In 1930, Landau had developed a theory for second-order phase transition. Many second-order phase transitions can be characterized by an appropriate order parameter, the order parameter shows different value at high temperature disordered state and low temperature order state. For example, the magnetic order parameter is always used to describe the magnetic phase transition from ferromagnetism to paramagnetism. In 1950, Ginzburg and Landau proposed to describe the superconducting state with a more complex order parameter Ψ, here Ψ is spatially varied, and $|\Psi|^2$ is proportional to the density of super electrons, *i.e.* $|\Psi|^2 = n_s(r)$. Ψ is nonzero in the superconducting state, while it equals zero in normal state above critical temperature T_c [8].

Since superconducting state is a thermal equilibrium state, its thermal dynamic property can be described with free energy density f_s. For temperature close to critical temperature, free energy can be expanded as a function of order parameter $|\Psi|$,

$$f_s(T) = f_n(T) + a(T)|\Psi|^2 + \frac{b(T)}{2}|\Psi|^4 + \cdots \cdot \tag{6}$$

where $f_s(T)$ is the free energy density of the normal state, a and b are the temperature dependent parameters, In order to get minimum for f_s, $b(T)$ has to be positive, while $a(T)$ can be either positive or negative, corresponding to $T > T_c$ or $T < T_c$, respectively. If we plot the difference of free energy density as a function of Ψ, we will get two different curves for $a(T) > 0$ and $a(T) < 0$. These two curves have different minimum: at $\Psi = 0$ for $T > T_c$ and at $|\Psi|^2 = -a(T)/b(T)$ for $T < T_c$.
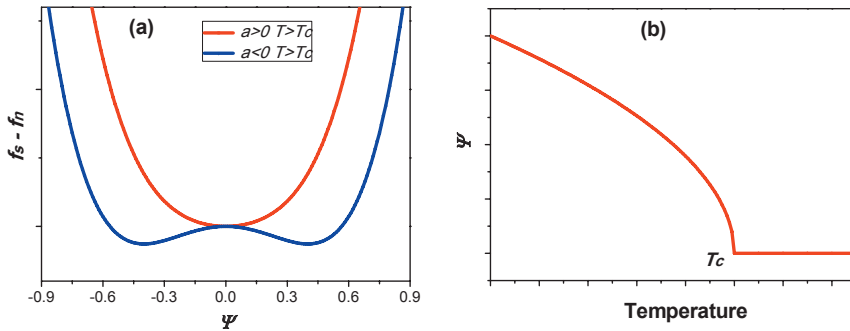
Fig. 3: (a): Difference between the free energy in the normal and superconducting state as a function of order parameter Ψ. (b): Temperature dependence of order parameter Ψ in superconductor.

In a spatially inhomogeneous superconductor, the order parameter depends on position. A new term depending on the gradient of $\Psi(r)$ should be included in the free energy. If we also consider the effect of magnetic field $\mathbf{B} = \mu_0\mathbf{H}$, another additional term should also be included. Therefore the free energy of superconductor in the magnetic field is:

$$f_s(T) = f_n(T) + a|\Psi|^2 + \frac{b}{2}|\Psi|^4 + \frac{1}{2m_s}|(-i\hbar\nabla - 2eA)\Psi|^2 + \mu_0\frac{|H|^2}{2} \tag{7}$$

By minimizing the free energy of the system, we can get two Ginzburg-Landau equations:

$$a\Psi + b|\Psi|^2\Psi + \frac{1}{2m_s}(-i\hbar\nabla - 2eA)^2\Psi = 0 \tag{8}$$

$$\mathbf{J}_s = -\frac{2e\hbar i}{i2m_s}(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*) - \frac{(2e)^2}{m_s}|\Psi|^2 A \tag{9}$$

Suppose there exists an interface between normal state and superconducting state. By assuming that $\Psi(r)$ is continuous and the boundary condition at $\Psi(0) = 0$, we can solve the first Ginzburg-Landau equation and get $\Psi(r)$:

$$\Psi(r) = \Psi(0)tanh(\frac{r}{\sqrt{2}\xi(T)}) \tag{10}$$

where $\Psi(0)$ is the order parameter far from the interface in the superconducting and $\xi(T)$ is called Ginzburg-Landau coherence length:

$$\xi(T) = \sqrt{\frac{\hbar^2}{2m_s|a(T)|}} \tag{11}$$

Beside of the London penetration depth λ_L, Ginzburg-Landau coherence length $\xi(T)$ is another fundamental length scale associated with superconductivity. The ratio between these two length scales is denoted as Ginzburg-Landau parameter, which is independent of temperature,

$$\kappa = \frac{\lambda(T)}{\xi(T)} \tag{12}$$

Usually, the ratio $\kappa = 1/\sqrt{2}$ is adopted as the criterion to define the type-I and type-II superconductors:

$$\kappa \leq 1/\sqrt{2} \qquad (Type - I) \tag{13}$$

$$\kappa \geq 1/\sqrt{2} \qquad (Type - II) \tag{14}$$

For type-I superconductor, the field inside is zero due to the Meissner effect, when external field is larger than critical field H_c, the superconductivity is destroyed suddenly. While, there are two different critical fields in type-II superconductor: the lower critical field H_{C1} and upper critical field H_{C2}. If external field is smaller than H_{C1}, the sample is perfect diamagnet. If external field exceeds H_{C1} but below H_{C2}, the superconductor enters the so called Shubnikov phase, in which the magnetic flux penetrates the superconductor in the form of vortices [9]. If external field increases further, the vortex cores are getting closer and almost overlap when external field reached upper critical field H_{C2}. The superconductivity will be totally destroyed once the field exceeds H_{C2}.

2.5 Electron-pairing and the BCS theory

As a phenomenological theory, the Ginzburg-Landau theory was quite successful in explaining many physical properties of superconductor. However, it can not explain the microscopic origins of superconductivity. For instance, the physical meaning of the Ginzburg-Landau order parameter was still no clear. In 1957, Bardeen, Cooper, and Schrieffer (BCS) proposed a microscopic theory which can provide the physical interpretation of the nature of order parameter and describe the macroscopic wavefunction of conduction electrons [10]. The key idea of BCS theory is that the crystal lattice phonons can act as the exchange bosons and give an attractive interaction between the conduction electrons near the Fermi surface. Thus, the pair bound state of electrons is formed and the paired conduction electrons are called "Cooper pair".
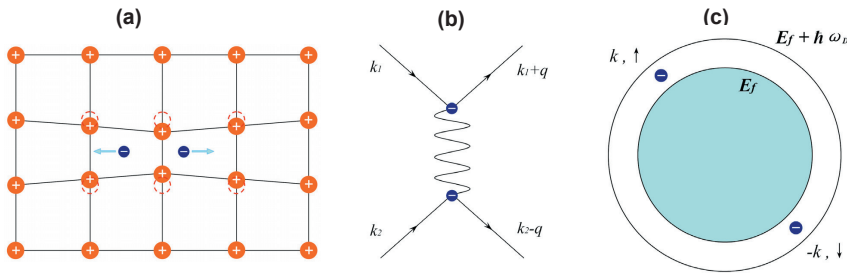
Fig. 4: (a) Two electrons paired when moving through crystal lattice due to the electron phonon coupling. (b) Interaction of electrons via exchange of boson (crystal lattice phonon). (c) Attractive interaction between two electrons close to the Fermi surface.

It is well known that bare electrons will repel each other due to the strong electrostatic Coulomb repulsion. However, if we consider the electrons in a medium, say, surrounded by charged ions in a crystal lattice [Fig. 4(a)], the Coulomb interaction will be largely reduced by the screening. Furthermore, the attractive electron-electron interaction is generated via the exchange of virtual exchange bosons such as phonons. As presented in Fig. 4(b), the Feynman diagram illustrated the interaction of electrons via exchange phonons. Because the total wave vector is conserved, we have $k_1+k_2 = k_1+q+k_2-q = k'_1+k'_2$. Considering a situation that only two additional electrons located outside of the spherical Fermi surface at $T = 0$. The interaction between two additional electrons will take place within the range $\hbar\omega_D$ of Fermi surface, *i.e.* $E_f < E_k < E_f+\hbar\omega_D$. To ensure the momentum conservation and to minimize the energy, two additional electrons will pair up as Cooper pair with no center of mass motion, as demonstrated in Fig. 4(c).

The coupling of the electron spins of the Cooper pairs will result in two different total spin, $S = 0$ or $S = 1$. Thus the spin wave function can be:

$$\phi = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) \qquad (S = 0, singlet) \qquad (15)$$

$$or \quad \phi = \begin{cases} |\uparrow\uparrow\rangle \\ \dfrac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \qquad (S = 1, triplet) \\ |\downarrow\downarrow\rangle \end{cases} \tag{16}$$

Furthermore, based on the distribution of pairing amplitude in k space, the pairing symmetry can be classified as s, p, d, f...waves.

By using the language of second quantization, the pairing hamiltonian of singlet superconductor can be given as:

$$\mathcal{H} = \sum_{k,\sigma} \epsilon(k) n_{k\sigma} + \sum_{k,k'} V_{k,k'} c_{k\uparrow}^{\dagger} c_{-k\downarrow}^{\dagger} c_{-k'\downarrow} c_{k'\uparrow} \tag{17}$$

By defining $b_k = \langle c_{-k\downarrow} c_{k\uparrow} \rangle$ and $\triangle_k = -\sum_{k'} V_{k,k'} \langle c_{-k\downarrow} c_{k\uparrow} \rangle$, the model Hamiltonian is expressed as:

$$\mathcal{H} = \mu N + \sum_{k,\sigma} \xi_k c_{k\sigma}^{\dagger} c_{k\sigma} - \sum_{k} (\triangle_k c_{k\uparrow}^{\dagger} c_{-k\downarrow}^{\dagger} + \triangle_k^* c_{-k\downarrow}^{\dagger} c_{k\uparrow}^{\dagger}) \tag{18}$$

The above Hamiltonian can be diagonalized by Bogoliubov-Valatin transformation:

$$c_{k\uparrow} = u_k^* \gamma_{k\uparrow} + v_k \gamma_{-k\downarrow}^{\dagger} \tag{19}$$

$$c_{k\uparrow}^{\dagger} = -v_k^* \gamma_{k\uparrow} + u_k \gamma_{-k\downarrow}^{\dagger} \tag{20}$$

with $|u_k|^2 + |v_k|^2 = 1$.

If we insert these operators into the model Hamiltonian, use the relation between u_k and v_k, then properly choose u_k and v_k, the following relations will be obtained:

$$2\xi_k u_k v_k + \triangle_k^* v_k^2 - \triangle_k u_k^2 = 0 \tag{21}$$

Simplify above equation by multiple $\triangle_k^* / u_k^2$ to both two sides gives:

$$\frac{\triangle_k^* v_k}{u_k} = \sqrt{\xi_k^2 + |\triangle_k|^2} - \xi_k \tag{22}$$

Excitation energy E_k is defined as $E_k = \sqrt{\xi_k^2 + |\triangle_k|^2}$ [Fig. 5(a)], thus, v_k and u_k can be expressed in term of E_k as:

$$|v_k|^2 = 1 - |u_k|^2 = \frac{1}{2}(1 - \frac{\xi}{E_k}) \tag{23}$$

In BCS theory, Bardeen, Cooper and Schrieffer proposed the ground state as:

$$|\Psi_G\rangle = \prod_k (u_k + v_k c_{k\uparrow}^\dagger c_{-k\downarrow}^\dagger)|0\rangle \tag{24}$$

Noted that $|u_k|^2 + |v_k|^2 = 1$. This implies that the parameter u_k and v_k are the probability amplitudes. The probability of the pair $(k\uparrow, -k\downarrow)$ being occupied is $|v_k|^2$, while the probability of the pair being unoccupied is $|u_k|^2$. The relations between occupation probability $|v_k|^2$ and $|u_k|^2$ is shown in Fig. 5(b).
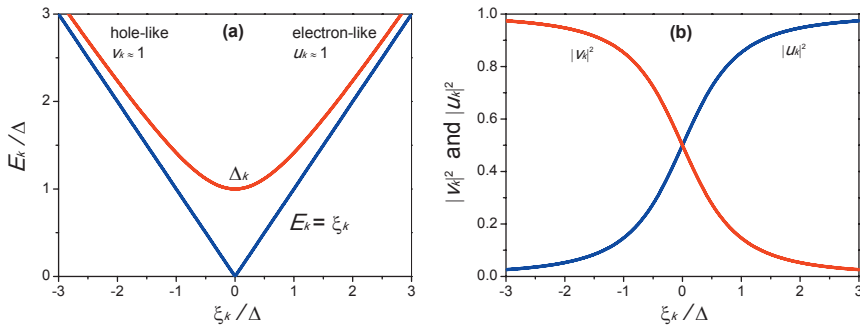
Fig. 5: (a) Excitation energy E_k as a function of ξ_k in the normal and superconducting state. (b) Occupation probability $|v_k|^2$ and $|u_k|^2$ as a function of ξ_k at T = 0 near the Fermi level.

The ground state energy can be expressed as:

$$\langle E_G \rangle = \sum_k \left(\xi_k - \frac{\xi_k^2}{E_k} \right) - \frac{\Delta^2}{V_0} \tag{25}$$

Because the occupation probability is given by the Fermi-Dirac distribution, the definition of $\triangle_k$ can be rewritten as:

$$\begin{aligned}
\triangle_k &= -\sum_{k'} V_{k,k'} \langle c_{-k\downarrow} c_{k\uparrow} \rangle \\
&= -\sum_{k'} V_{k,k'} u_{k'} v_{k'} (1 - 2f(E_{k'})) \\
&= -\sum_{k'} V_{k,k'} \frac{\triangle_{k'}}{2E_{k'}} tanh\left(\frac{E_{k'}}{2k_B T} \right)
\end{aligned} \tag{26}$$

Based on the assumption of week coupling in BCS theory, *i.e.* $V_{k,k'} = -V_0$, $\triangle_{k'} = \triangle$, the summation of above equation can be converted into an integration over energy, then we arrive at the BCS gap equation:

$$1 = \lambda \int_0^{\hbar\omega_D} \frac{1}{E_k} tanh\left(\frac{E_k}{2k_B T} \right) d\xi_k \tag{27}$$

where $\lambda = V_0 D(\mu)$ is the dimensionless electro-phonon coupling constant, $D(\mu)$ is the density of states. The BCS gap equation gives the temperature dependence of gap energy, in particular, it gives not only the energy gap at zero temperature but also the ordering temperature T_c.

For temperature approaching to T_c, we have $\triangle \rightarrow 0$. Then we can get the equation for the critical temperature:

$$k_B T_c = 1.13 \hbar \omega_D e^{-1/\lambda} \tag{28}$$

Also at low temperature the ratio between $\triangle$ and T_c can be determined as

$$\frac{\triangle(0)}{k_B T_c} = 1.764 \tag{29}$$

Above relation is one of the most important result deduced from the BCS theory, indeed, it was obeyed by all classical metallic superconductors as an universal amplitude ratio.

3   A survey of the superconducting materials

After superconductivity was found in Hg with T_c = 4.2 K, great efforts are made to search for new superconductors. Till now, thousands of superconducting materials have been found [11]. However, superconductivity is still a low temperature phenomenon. To find a room-temperature superconductor seems still a long way to go. In the following parts, the main superconducting materials are classified.

• **Superconducting element**
After Hg, Superconductivity is also found in some other elements, such as Sn, Pb and La. A-mong all elements, Pb possesses the highest T_c of 9.2 K in the ambient pressure. Although some elements are non-superconductor at very low temperature, the superconductivity can e-merge when they are subjected to high pressure or fabricated as thin films.

• **Superconducting alloys and compounds**
As a solid solution of different kinds of atoms, some alloys also exhibit superconductivity, such as NbTi (T_c = 9.5 K) and NbTa (T_c = 6.0 K). Higher T_c are also found in other Nb-content A_3B compounds, e.g. T_c = 18 K for Nb_3Sn and T_c = 23.2 K for Nb_3Ge. Besides, the superconductivity with T_c up to 40 K was discover in MgB_2 [12].

• **Organic superconductors**
Usually, the Organic compounds are insulators, but it was found that some organic compounds are superconductors. The Critical temperatures of organic superconductors are still in the range of classical superconductors [13].

• **Heavy-fermion superconductors**
In Heavy-fermion system, the electrons have large effective mass of about 200 times the free electron mass. Superconductivity was found in some heavy-fermion system such as $CeCu_2Si_2$. It was believed that heavy Fermion superconductors belong to the unconventional superconductor, however, the mechanism of this kind of superconductor is still not clear [14].

● **Superconducting fullerene compounds**
Fullerene, with the formula C_{60} was discovered as the third form of carbon in 1980s. After intercalating with exotic atoms, the doped C_{60} molecular crystals will exhibit the superconductivity with T_c up to 40 K [15].

● **Superconducting sulfur hydride**
In 2015, it was observed that hydrogen sulfide (H_2S) becomes superconductive at below 203 K when it is placed under extremely high pressure of 155 GPa [16]. H_2S set the new record for superconducting critical temperature and highlighted a potential way to achieve the room-temperature superconductivity in the compressed hydrogen-rich materials.

● **Superconducting graphene superlattices**
In 2018, superconductivity was discovered in a graphene bilayer where one layer was offset by a "magic angle" of 1.1° relative to the other [17]. Although the zero-resistance state in system is observed at low temperature of 1.7 K, the results suggest that graphene can be act as a new platform for investigating unconventional superconductivity.
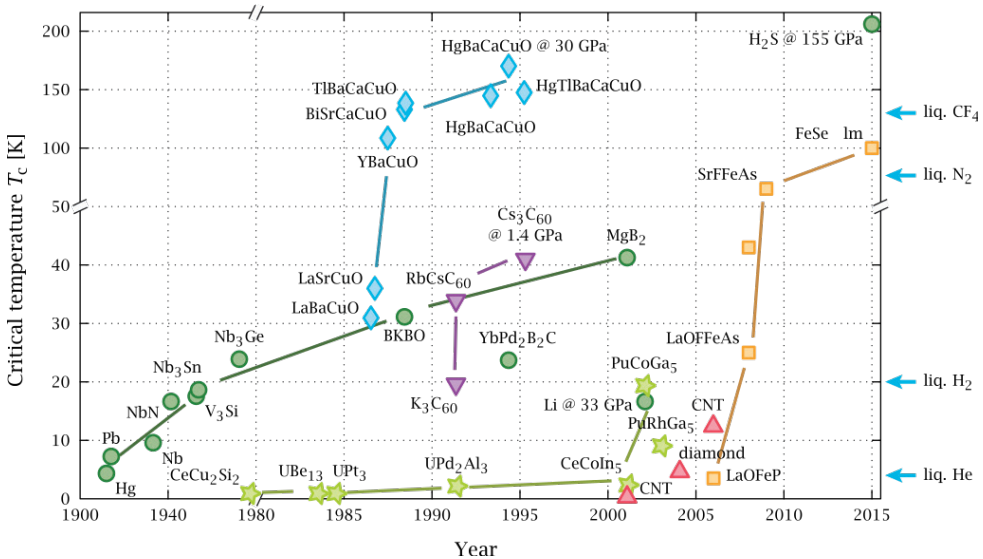
Fig. 6: Timeline of the discovery of superconductors [18].

● **High-T_c cuprate superconductors and Iron-based high-T_c superconductors**

The superconducting materials possessing unusually high critical temperature are labeled as high-temperature (high-T_c) superconductors. In contrast to ordinary superconductors with critical temperature below 30 K, the high-T_c cuprate superconductors and Iron-based high-T_c superconductors behave as superconductor at high temperature above 130 K and 50K at ambient pressure,respectively. Extensive attention are drawn on high-T_c superconductors since they provide a new opportunity to investigate the mechanism of unconventional superconductivity. Introductions on high-T_c cuprate superconductors and high-T_c iron-based superconductors are given in the next two sections.

4   High-T_c cuprate superconductors

In 1986, the first cuprate superconductor $La_{2-x}Ba_xCuO_4$ with $T_c \approx 30$ K was discovered by Bednorz and Müller [2]. It was quite surprising that high T_c is shown in cuprate since the oxidize compounds are always insulators or poor conductors. The researchers are immediately motivated by this breakthrough and start searching for new cuprate compounds with higher T_c. Soon after, the superconductivity is observed in $YBa_2Cu_3O_{6+x}$ with $T_c \approx 90$ K, which is well above the temperature of liquid nitrogen. So far the highest T_c of 133 K at ambient pressure was found in Hg-based cuprate $HgBa_2Ca_2Cu_3O_8$ [19]. Obviously, the Cuprate with high T_c are unconventional superconductors and it can not be explained in the framework of BCS theory. Therefore, new theory is required to explain the physics of the cuprate superconductors. However, the fundamental mechanism of high-temperature superconductivity is still unclear and it was still considered as the topic of the frontier of condensed matter physics.

From the point of view of crystal structure it was found that all high T_c cuprate have a layered structure. All Cuprate have one or more layers of copper oxide (CuO_2) and the CuO_2 layer are spaced by layers containing elements such as lanthanum, barium or yttrium. The schematic view of crystal structure of $YBa_2Cu_3O_7$ is shown in Fig. 7(a). Based on the large number of experimental results the universal phase diagram relating the critical temperature to doping level can be draw Fig. 7(b).
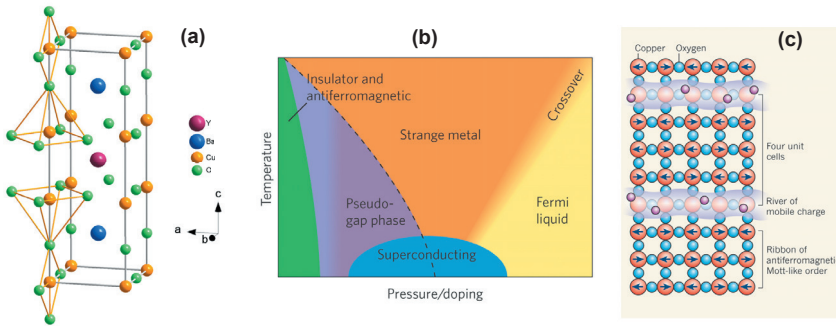
Fig. 7: (a) Crystal structure of $YBa_2Cu_3O_7$. (b) Schematic phase diagram of hole doped cuprate high T_c cuprate superconductor [20]. (c) Stripe-like electronic order in the cuprates [21].

Considering the phase diagram with increasing doping, the material is still an antiferromagnetic insulator at the lowest doping level. With increasing doping, the material enters the superconducting phase and the critical temperature T_c exhibits a dome-like dependence on doping with further increases in the doping level. At the over doped level, the material again become non-superconducting. Between the antiferromagnetic and superconducting phase zone, there is phase named pseudogap phase, in which physical properties show behavior of the existence of an energy gap. It is still controversial whether the pseudogap arisen from competing orders or it is the precursor of superconductivity.

As shown in Fig. 7(b), the undoped Cuprate compounds are Mott insulator with long-range antiferromagnetic order of Cu. Upon doping with holes, the stripes of spin and charge order formed. Both charge and spins are periodically modulated in the stripe phase. As illustrated

in Fig. 7(c), inhomogeneity arises due to the hole doping process. The antiferromagnetic spin order in the spin-part of stripe is similar as in the undoped antiferromagnetic Mott insulator, whereas the charge can conduct between the spin-part of stripe. Experimentally, the strip phase is directly detected by neutron scattering study. Information on the period of both charge and spin density modulations can be obtained. Theoretically, it was thought that the formation of stripe phase was attributed to the competition between the kinetic energy of the electrons, the antiferromagnetic interaction among spins, and the Coulomb interaction between charges. It was believed that the existence of the strip phase might give rise to superconductivity. Although qualitative understanding of the nature of the superconducting state itself has been achieved to some extent, profound unresolved issues concerning the astonishing complexity of the phase diagram and so on are still need to be explored in effort to unravel the physics of cuprates [22].

5 Iron-based high-T_c superconductors

In 2008, a new family of high-T_c superconductors, iron-based superconductors were discovered [3]. There are a number of homologous families of iron-based superconductors discovered so far, which are short-named after the stoichiometries of their parent compounds (as shown in Fig. 8). Iron-based superconductors are usually termed as iron pnictides or iron chalcogenides depending on their compositions. All iron-based superconductors have a common layer of iron atoms which are tetrahedrally coordinated by pnictogen or chalcogen atoms [23]. They differ only in the details of the buffer layers. Experiments and theory now agree that the superconducting electrons in all the iron-based superconductors flow in the planes that contain Fe. Despite of the compositional variety, the Fe-containing planes have the same structure from material to material.
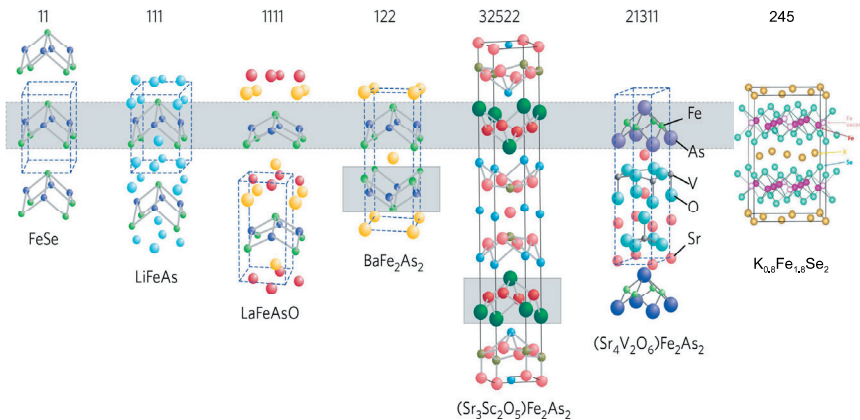
Fig. 8: Several homologous families of iron-based superconductors [23].

Starting from non-superconducting parent compounds, superconductivity can be achieved either by doping, or by the application of pressure in some materials. Up to now, the highest T_c attained is 57.4 K in the electron-doped compound $Ca_{0.4}Nd_{0.6}FeAsF$. Since the parent compounds are already metallic, the effect of the doping can not be solely related to the introduction

of free charge carriers. It has been suggested that the modifications of the Fermi surface, which are similar under pressure and chemical doping, are important for inducing superconductivity in the iron-based compounds. The measurements of resistivity, magnetic susceptibility as well as more bulk-sensitive techniques, such as thermal expansion, heat capacity and neutron and x-ray diffraction etc. would allow for an accurate determination of the phase diagrams of the iron-based superconductors. The phase diagrams electron and hole-doped $BaFe_2As_2$ is shown in Fig. 9. One of the most fascinating phenomena is the apparent coexistence and competition between superconductivity and the spin density wave phase in the underdoped regime of $Ba(Fe_{1-x}Ni_x)_2As_2$. The exact nature of this phenomenon remains to be established.
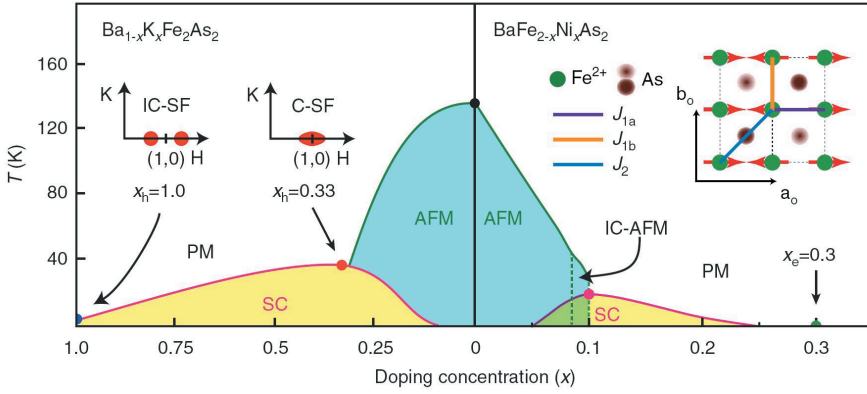
Fig. 9: The electronic phase diagram of electron and hole-doped $BaFe_2As_2$ [24].

The nature of the magnetic ordering and spin fluctuations in superconductors has had a rich and interesting history, and has been a topic of special interest ever since the parent compounds of the high-T_c cuprates were found to be antiferromagnetic Mott insulators that exhibit huge exchange energies within the Cu-O planes. These strongly correlated spin fluctuations persist into the superconducting regime, often developing a spin resonance mode whose energy scales with T_c and whose intensity exhibits a superconducting order-parameter-like behavior. Iron-based superconductors represent another remarkable example in which superconductivity is in close proximity to magnetism [25]. Although neutrons do not couple directly to the superconducting order parameter, neutron scattering methods have nevertheless played a decisive role in the understanding of the interplay between superconductivity and magnetism, as demonstrated by the determination of magnetic ordering in the parent compound and the observations of the spin resonant mode in the superconducting counterparts.

6   Summary

While superconductivity remains one of the biggest challenges in condensed matter physics, the understanding of its mechanism has advanced tremendously over the past years or so. It is expected that the research on superconductivity and discovery of new high temperature superconducting system in future will lead to a wide range of technical application.

References

[1] H. K. Onnes, Leiden Commun. **120b, 122b, 124c**, (1998).

[2] J. G. Bednorz and K. A. Mller, Zeitschrift fr Physik B Condensed Matter **64**, 189 (1986).

[3] Y. Kamihara et al., J. Am. Chem. Soc. **130**, 3296 (2008).

[4] C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, Inc., New York, 2004).

[5] P. Hofmann, *Solid State Physics* (Wiley-VCH, Berlin, 2008).

[6] W. Meissner and R. Ochsenfeld, Naturwissenschaften **21**, 787 (1933).

[7] F. London and H. London, Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences **149**, 71 (1935).

[8] V. L. Ginzburg and L. D. Landau, Zh. Eksp. Teor. Fiz **20**, 1064 (1950).

[9] A. A. Abrikosov, Sov. Phys. JETP **5**, 1174 (1957).

[10] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

[11] Frank J. Owens and Charles P.Poole, *The new superconductors* (Plenum Press, New York, 1996).

[12] J. Nagamatsu et al., Nature **410**, 63 (2001).

[13] T. Ishiguro, K. Yamaji, and G. Saito, *Organic Superconductors* (Springer, berlin and Heidelberg, 1998).

[14] C. Pfleiderer, Rev. Mod. Phys. **81**, 1551 (2009).

[15] A. Y. Ganin, Nature Materials **7**, 367 (2008).

[16] A. Drozdov et al., Nature **525**, 73 (2015).

[17] Y. Cao et al., Nature **556**, 43 (2018).

[18] Link:, http://en.wikipedia.org/wiki/Superconductivity , (2018).

[19] A. Schilling, M. Cantoni, J. D. Guo, and H. R. Ott, Nature **363**, 56 (1993).

[20] P. C. Canfield, Nature Materials **10**, 259 (2011).

[21] K. A. Moler, Nature **468**, 643 (2010).

[22] B. Keimer et al., Nature **518**, 179 (2015).

[23] D. C. Johnston, Advances in Physics **59**, 803 (2010).

[24] M. Wang et al., Nature Communication **4**, 2874 (2013).

[25] P. Dai, Reviews of Modern Physics **87**, 855 (2015).

E 2   Polymer Structures and Dynamics

M. Kruteva

Jülich Centre for Neutron Science

Institute of Complex Systems

Forschungszentrum Jülich GmbH

Contents

Introduction

Since last decade polymers surround us in everyday life. Nowadays can not imagine a house, a car or working place without modern equipment and components made of polymer-based composite materials. Polymers are now produced in great quantity and variety. Polymers are used as film packaging, solid molded forms for automobile body parts and TV cabinets, composites for golf clubs and aircraft parts (airframe as well as interior), foams for coffee cups and refrigerator insulation, fibers for clothing and carpets, adhesives for attaching anything to anything, rubber for tires and tubing, paints and other coatings to beautify and prolong the life of other materials, and a myriad of other uses. It would be impossible to conceive of our modern world without the ubiquitous presence of polymeric materials. Polymers have become an integral part of our society, serving sophisticated functions that improve the quality of our life.

Polymers are one of the most important products of chemical industry. Production of plastic increased from 1950 to 2016 by a factor more than 300. 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.

This lecture aims to identify general principles of polymer chain structure and motion on a molecular scale which sustain macroscopic properties. In particular, recent concepts and experimental results on these structural peculiarities and motional mechanisms will be presented. The general models of structure and dynamics of polymer chain will be derived. Finally, a few examples will present (i) how structure and dynamics of the polymer chain depends on their topology (ring *vs* linear), (ii) what is influence of the confinements on the polymer structure and dynamics and (iii) association properties of supramolecular polymers in a melt state.

Fig. 1:   Turnover along products (Europe 2012)

1   Macroscopic properties of polymers

Dynamic processes in polymers occur over a wide range of length and time scales. Fig. 2 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.
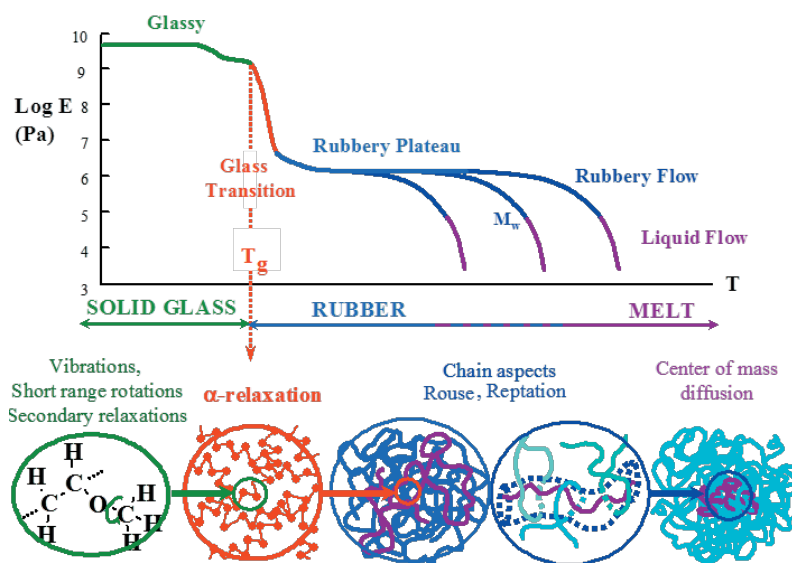
Fig.2: Schematic presentation of the modulus of a polymer melt as a function of temperature. The sketches below symbolise the types of motions that give rise to the macroscopic behaviour.

Fig. 3 quantifies bebebe 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 behavior of polymer melt are universal power laws in the molecular weight dependence of viscosity and diffusion.

Fig. 4 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.
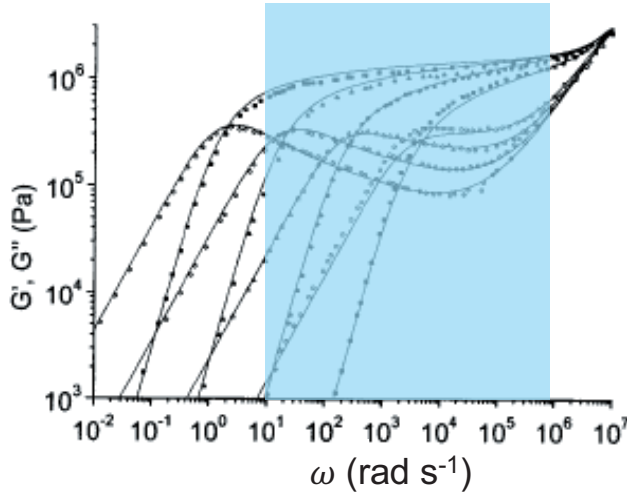
Fig.3: Real and imaginary part G' and G'' of the dynamic modulus for polymer melts of different molecular weight. The blue area displays the region where the polymer melt responses elastically.

This is shown in Fig. 5, where the inverse diffusion coefficient is displayed as a function of molecular weight in a double logarithmic form. Again, at low molecular weight M_w we realize 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 revealing a power law $M_w^{-2.3}$ instead of M_w^{-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 behavior of polymer melts. We will go through a hierarchy of models that will let us understand, why long chain molecules exhibit the shown dynamical features.
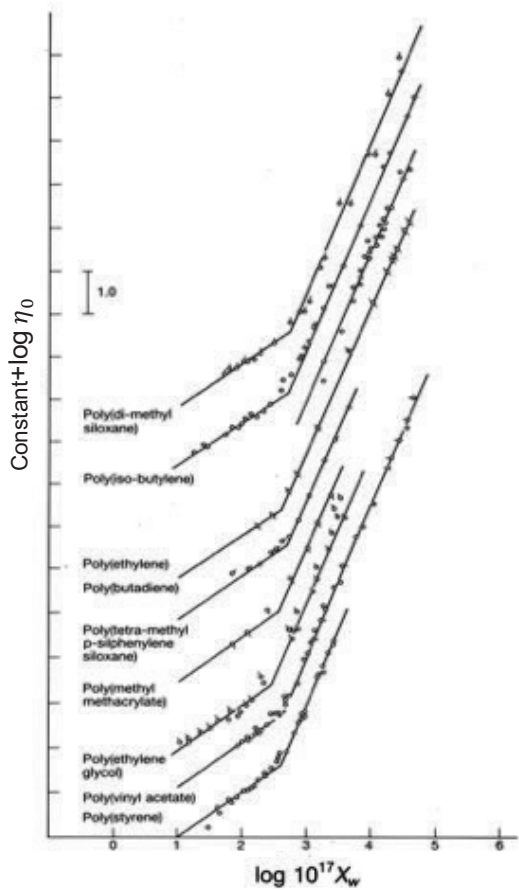
Fig.4: Polymer melt viscosity for various polymers as a function of molecular weight (parameter χ_w is proportional to M_w). The curves are shifted by a constant. The data are characterized by crossover between two power laws. [1]
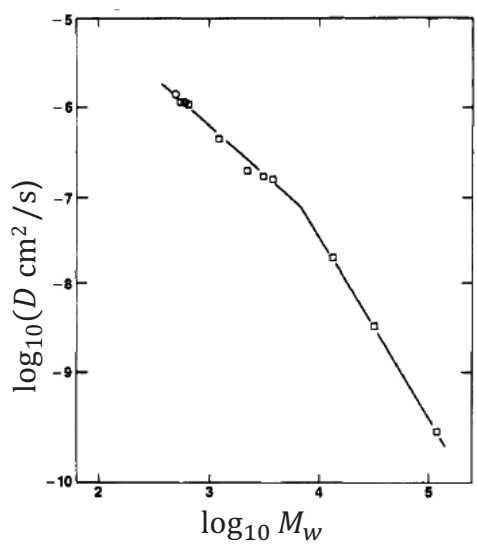
Fig.5: Translational diffusion coefficient in a polymer melt measured for polyethylene (PE) as a function of molecular weight: Again, two different power laws are visible.

2   Structure

Polymers have a very rich behaviour (see Fig.2) and are ideal for investigations by means of neutron scattering techniques. As reported in the appropriate *basic scattering* lectures, this is due to the natural difference between a proton and a deuteron. Although chemically virtually no other properties are induced if some hydrogens (H) are replaced by some deuteriums (D), this labelling allows to study e.g. the effect of different environments on the structure of polymers, on parts of the chains and so on. This is a strong advantage over scattering by X-rays which is limited to systems which differ considerably in electron density. The H/D labelling of polymers does not change the X-ray scattering patterns. On the other hand, the study of polymers by X-ray is exciting and rewarding because of the wide variation in organization that can be observed. Therefore X-ray is very often used as a valuable complementary technique, especially in multiphase systems. In particular, the long-lived methastable phases formed by block copolymers were observed using small angle X-ray scattering (SAXS) [2]. Another complex system consisted of diblock copolymers decorated by supramoplecular groups were investigated to clarify details of microscopic structure and changes in the association behavior as a function of temperature. [3]

Despite of unsensitivity of X-ray scattering pattern to the H/D labelling, tunability of X-ray wavelength at a synchrotron source makes it possible to vary absorbtion of the constituent elements of the sample (e.g. Pt, Au or some other heavy-atoms). In this contex it is worth to mention the Multiple wavelength Anomalous Diffraction (MAD) method allowing to solve the substructure of the anomalously diffracting atoms and hence the structure of whole molecule [4].

2.1   Conformation of polymer chain

A polymer is a chain of several polyatomic units called monomers covalently bonded together. Since virtual all kinds of molecules can act as a monomeric unit, thereby only differing in the ways in which they can be bound together, a wealth of synthetic and naturally occurring polymers with enormous diversity in properties is nowadays known. We mention e.g proteins, DNA, glass, thermoplasts and rubber. They all belong to the class of polymers. Given their importance, an adequate description of model polymers has become a prerequisite. This brief summary serves to introduce the reader to some of the basic models with respect to scattering. The simplest descriptions of single-molecule models can then be expanded to cyclic and branched polymers with some minor changes.

Polymers can often be imagined as spaghetti-like or coiled molecules. Their stiffness varies from very flexible to rigid in the case of rods. Their configuration i.e their spatial distribution of segments changes all the time as the result of brownian dynamics which therefore leads to an enormous amount of possible configurations itself. The motion and dynamics of polymers will be discussed at full length in the session on polymer dynamics below. Here, we will stick to the determination and static investigation of their time-averaged structures.

Basic models for linear chains to describe their statistical properties were developed. With different levels of sophistication and approximations these models are able to take into account typical conformations of linear polymer chains. In this chapter we will summarize these models

on scattering of polymer chains in equilibrium and introduce at this stage also the corresponding length scales.

The simplest model to deal with is the *freely jointed chain* (FJC). An example for such a configuration is shown in Fig. 6 and illustrates the idea that polymer chains are performing random walks. With this, the ensemble of systems can be characterized by average quantities. One is the so-called end-to-end vector

$$\vec{R}_{ee} = \sum_{i=1}^{N} \vec{r}_i \tag{1}$$

and its square

$$R_{ee}^2 = \sum_{i,j=1}^{N} \vec{r}_i \vec{r}_j \tag{2}$$

which is the scalar quantity. N is a number of segments in polymer chain. The bond vector length (or step) $|\vec{r}_i|$, connecting 2 atoms is l.

A further important parameter which has its counterpart in classical mechanics is the radius of gyration R_g:

$$R_g^2 = \frac{1}{N+1} \sum_{i=0}^{N} \vec{s}_i^2 \tag{3}$$

where the vectors $\vec{s}$ are the distance vectors with respect to the center of mass, $\vec{R}_{c.o.m}$ with

$$\vec{R}_{c.o.m} = \frac{1}{N} \sum_{i=0}^{N} \vec{r}_i \tag{4}$$
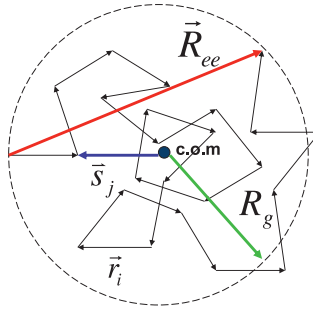
Fig.6: A freely joined chain and definitions: end-to-end distance $\vec{R}_{ee}$, gyration radii R_g, segment vector $\vec{r}_i$ and $\vec{s}_j$ is a vector between the segment and centre-of-mass (c.o.m.).

Averaging over all conformations in formulas (1-4) leads for the case of a fully freely jointed chain to a mean square end-to-end distance of

$$\langle R_{ee}^2 \rangle = \sum_{i,j=1}^{N} \langle \vec{r}_i \vec{r}_j \rangle = Nl^2 + 2 \sum_{i<j}^{N} \langle \vec{r}_i \vec{r}_j \rangle \tag{5}$$

whereas the average end-to-end distance $\langle R_{ee} \rangle = 0$ due to the ensemble average. With the help of Largange theorem the average radius of gyration (Eqs.(3)-(4)) can be rewritten

$$\langle R_g^2 \rangle = \frac{1}{2(N+1)^2} \sum_{i,j}^{N} \langle (\vec{r}_i - \vec{r}_j)^2 \rangle = \frac{1}{2(N+1)^2} \sum_{i,j}^{N} \langle r_{ij}^2 \rangle \tag{6}$$

where $\langle (\vec{r}_i - \vec{r}_j)^2 \rangle = \langle r_{ij}^2 \rangle$ is a square distance between two segments.

Now the calculation of the end-to-end distance which characterizes the size of a chain and constitutes the diameter of a hypothetical sphere including the full chain, can be performed for different chain models. The FJC model assumes an equal probability in all directions of all bond vectors, random bond rotation angles while keeping the bond length constant. The orientation of each segment or bond appears in the Eq.(5) as an angle between two segments:

$$\langle \vec{r}_i \vec{r}_j \rangle = \langle \cos \theta_{ij} \rangle \tag{7}$$

In the FJC model the orientation of each segment is therefore independent of all others and as a consequence the second term which is the scalar product in Eq.(5) averages out to zero. The mean radius of gyration can then be obtained from evaluating the double sum in Eq. 5 and using $r_{ij}^2 = (j-i)l^2$ - a result of random walk statistics -becomes

$$\langle R_g^2 \rangle = \frac{1}{6} N l^2 \tag{8}$$

From both end-to-end distance and radius of gyration an important statement which applies to polymer chain is already observed: the size depends on the square root of the number of basic steps (number of segments), $\sqrt{N}$. This is a result which has its equivalent in the random walk statistics where now the position of the segment is replaced by the trajectory of a randomly-diffusing particle and so the variable becomes the time, $\sqrt{t}$. Further we identify l with the size of basic step or segment length in corresponding model.

Compared to ideal chain, real polymer chains are not connected in this freely jointed way. Instead, bond angles θ_{ij} (Eq.(7)) assume well-defined values and correlations between bonds within the chain as well as from the environment will affect the size. Intuitively it can be understood that these correlations will vanish with increasing separation distance $|i - j|$, i.e. $\lim_{|i-j| \to \infty} \langle \cos \theta_{ij} \rangle = 0$. Also, the flexibility of polymer chains is restricted by the fact that rotations with the so-called bond rotation angle are enabled within a certain range only due to steric reasons.

If we take the end-to-end distance determined by FJC model as the reference (Eq.(5)) then the difference to real polymers will be quantified in terms of a parameter which is denoted C_∞. This is defined by following equation

$$\langle R_{ee}^2 \rangle = C_\infty N l^2 \tag{9}$$

In the case of FJC $C_\infty = 1$ per definition.

A first step to the real chain conformation is the *freely rotating chain* model (FRC) in which now valence angles are restricted additionally but still torsional angles are left free to rotate. Assuming also that all bond lengths and angles are the same, the flexibility parameter after some calculations becomes

$$C_\infty = \frac{1 + \langle \cos\theta \rangle}{1 - \langle \cos\theta \rangle} \tag{10}$$

E.g. for saturated carbon chains $\theta = 68°$ so $C_\infty \approx 2$.

A further extension is the chain model with *hindered rotation* (HRC) which is better known as the *rotational isomeric state* model by Flory (RIS) [5]. This adds restrictions now also on the torsional degree of freedom. Instead, using the rotational barriers, a good estimate for allowed, energetically and sterically different configurations can be readily obtained. For C_∞, values considerably larger than 3 are found. Experimentally, the parameter as determined from small angle neutron scattering investigations is situated in the range between 3 and 10. We can summarize this by stating that highy-coiled polymers are characterized by low C_∞ values whereas extended or loosely-coiled chains show, on the contrary, high C_∞.

The distribution of the end-to-end vector is treated in many text books on statistical physics and will therefore not be repeated here. It shows that the distribution function $p(R)$ of any intramolecular distance corresponding to N segments in a single polymer chain follows a Gaussian distribution with

$$p(R, N) = \left(\frac{3}{2\pi\, Nl^2} \right)^{3/2} \exp\left(-\frac{3\, R^2}{2\, Nl^2} \right) \tag{11}$$

2.2   Scattering of single polymer chain

The size of polymer chain can be measured by various scattering experiments (light scatering, SAXS, SANS). The principles and practic details of such experiments are given in corresponding lecture notes of this School. In thin chapter the consequences of a linear chain conformation to the neutron scattering intensity will be considered.

A scattering volume contains many polymer chains, each with N scatterers i.e. monomeric units or segments here and with a coherent scattering length b. Their density is not constant and shows random fluctuations around an average value. This density $n_i(r) = 1$ if the monomer of type i sits at r_i. We define average density as $\langle n(r) \rangle = \langle n \rangle = N/V_s$ where V_s is a sample volume. The density fluctuation is then $\Delta n(r) = n_i(r) - \langle n \rangle$. The static structure factor $S(q)$ is defined as the density-density correlation function in reciprocal space $\langle n(-q)n(+q) \rangle$. Thus, for in the system of n_p chains, which all have N monomers with a scattering length b, the coherent scattering intensity can be calculated using:

$$I(q) = \frac{b^2}{V_s} \sum_{\alpha,\beta=1}^{n_p} \sum_{i,j=1}^{N} \left\langle \exp\left(i\vec{q}(\vec{r}_{\alpha,i} - \vec{r}_{\beta,j}) \right) \right\rangle = \frac{b^2}{V_s} S(q) \tag{12}$$

The coherent scattering intensity (Eq.(12)) is proportional to the macroscopic differential cross section per unit volume, $\dfrac{d\Sigma}{d\Omega}(q) = \dfrac{1}{V_s}\dfrac{d\sigma}{d\Omega}(q)$. It has dimension [cm^{-1}] and is proportional to the scattering length density (SLD) that will be discussed below.

In the double summations the indices i and j are the monomer numbers and symbol α and β are different chains. $r_{\alpha,i}$ is thus the position of the i-th monomer on chain α. This can be rewritten into 2 parts: i.e. the intra-chain scattering is due to the contribution from 2 monomers on the same chain and an inter-chain part arises in the case that both monomers are situated on 2 different chains. For the simplest case of a dilute dispersion the inter-chain contribution to the scattering intensity is negligible and we can further assume

$$\langle \exp\left(i\vec{q}\left(\vec{r}_{\alpha,i} - \vec{r}_{\beta,j}\right)\right)\rangle = 0$$
$$\sum_{\alpha,\beta} \cdots \rightarrow \sum_{\alpha} = n_p \tag{13}$$

because the chains are sufficiently well separated in space and the phase factor from each chain will be completely uncorrelated. Then the contribution of the inter-chain term cancels out on average. This allows us to determine the single chain structure factor $S_1(q)$. We can then re-write Eq.(12) (n_p and double sum over α,β cancel) as

$$S_1(q) = \frac{1}{N} \sum_{m,l=1}^{N} \langle \exp\left(i\vec{q}(\vec{r}_i - \vec{r}_j)\right)\rangle \tag{14}$$

We will come back to this in the context of the scattering of concentrated blends where inter-chain contributions have to be included in the full description. At this point it suffices to evaluate for pedagogical reasons the small q expansion. From this the radius of gyration which was introduced in the first part of this contribution, independent of the shape or structure of polymer chain can be determined. The Taylor expansion of the structure factor in the small q region then yields

$$NS_1(q) = \sum_{i,j} \langle 1 + i\vec{q}(\vec{r}_i - \vec{r}_j) - \frac{1}{2}q^2 \sum_{i,j}(\vec{r}_i - \vec{r}_j)^2 + \cdots \rangle$$
$$= N^2 - \frac{1}{2}q^2 \sum_{i,j} \langle (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 + \cdots \rangle \tag{15}$$
$$= N^2 - \frac{1}{2}q^2 \sum_{i,j} \frac{1}{3} \langle (r_i - r_j)^2 + \cdots \rangle = N^2 \left(1 - \frac{1}{3}q^2 R_g^2\right)$$

using the former result for R_g. Eq.(15) clearly proves that the small q limit of the structure factor always yields both N and R_g.

Let us now calculate the structure factor for the Gaussian chain, i.e. where all distances along the chain obey Gaussian statstics (Eq.(11)). Then the phase factor can be transformed to

$$\langle \exp(i\vec{q}\vec{r}_{i,j})\rangle = \exp\left(-\frac{1}{2}\langle (qr_{ij})^2\rangle\right) = \exp\left(-\frac{1}{6}q^2\langle r_{ij}^2\rangle\right) \tag{16}$$

For the isotropic Gaussian chain we find and it has been used in the Eq.(15)

$$\langle (x_i - x_j)^2\rangle = \langle (y_i - y_j)^2\rangle = \langle (z_i - z_j)^2\rangle = \frac{1}{3}l^2|i - j| \tag{17}$$

This result can be used and Eq. 13 can be evaluated in the full q-range with

$$S_1(q) = \frac{1}{N} \sum_{i,j=1}^{N} \langle \exp\left(i\vec{q}(\vec{r}_i - \vec{r}_j)\right)\rangle = \frac{1}{N} \int_0^N di \int_0^N dj \exp\left(-\frac{1}{6}q^2 l^2 |i - j|\right) \tag{18}$$

Here, the discrete sum has been already replaced by the continuous integral form (for N large enough, typically $N > 80$). It is the basic result for Gaussian chains. Its strength will be shown on selected examples, covering the most important fields of current polymer investigations. This integral can be solved analytically. The form factor $P(q)$ is defined as $S_1(q)/N$ and in discrete form is

$$P(q) = \frac{1}{N^2} \sum_{i,j} \exp\left(-\frac{1}{6}q^2 l^2 |i - j|\right) \tag{19}$$

For long chains ($N \to \infty$), $P(q)$ is called the Debye function $g_D(x)$ with

$$g_D(x) = \frac{2}{x^2}(\exp(-x) - 1 + x) \tag{20}$$

and the argument $x = (qR_g)^2$. A SANS experiment with fit to the Debye function is presented in Fig. 7 in a linear scale and in the form of a 2nd moment Kratky representation achieved when $g_D(x)$ is multiplied by q^2. The Kratky representation emphasises the high q-regime. For Gaussian chains with an asymptotic q^2 behaviour the high q regime then assumes a plateau. As shown the data are in perfect agreement with the Gaussian chain results derived above.

Small and high q regimes can also be explicitly obtained from expanding Eq. (20) for $x \to 0$. For small q, we then have $g_D(x) \approx 1 - x/3 + \cdots = 1 - (1/3)q^2 R_g^2 + \cdots$. Reversely, this is the Taylor equivalent of $\exp(-x)$, which can be conveniently plotted in either the Guinier representation $\ln(g_D(x))$ vs q^2 from which then the slope contains R_g, or in a Zimm plot i.e. the inverse structure factor $1/S(q)$ vs. q^2. The disadvantage of the latter is that slope and intercept in the linearized form are coupled whereas they are independent in the logarithmic way. The low q expansion allows a fast determination of R_g. An ultra-fast single-point estimate of the chain dimension can be obtained as follows: for $q^* = 1/R_g$ it can be easily obtained then that the scattering intensity at q^* has dropped to $2/3$ of the forward scattering value at $q=0$.

Thus, in the low q regime we obtain direct information on the chain molecular weight M_w (via N) and the radius of gyration R_g. For high q, Eq.(20) on the other hand gives

$$g_D(x) \approx \frac{2}{x} = \frac{2}{q^2 R_g^2} \sim \frac{1}{C_\infty} \tag{21}$$

However, since data at high q, where the level of incoherent background plays an important role (see Lecture Notes B3 and D1), can be sensitively affected by erroneous subtraction, its use for the estimation of R_g is limited.
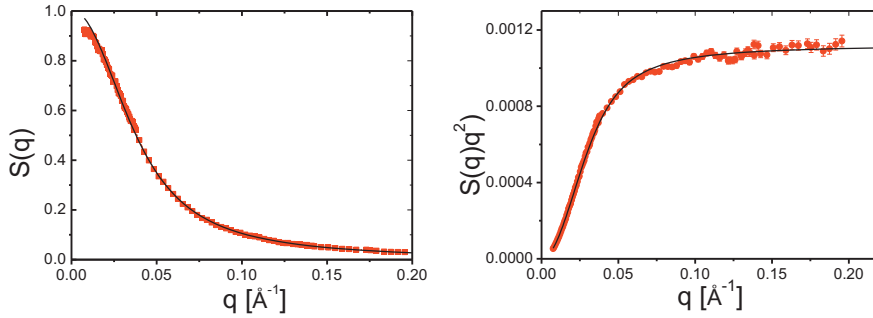
Fig.7: SANS scattering intensity for polymer chain described by the Debye function (lines). The data are normalized to S(q) on a poly-alkylene-oxide polymer, obtained at KWS2@MLZ, Munich [6]. On the left, the Debye curve fits ideally, showing the Guinier region and the high q-dependence q^{-2}. On the right, the Kratky region (plateau) is highlighted.

2.3   Contrast variation

One of the strong assets of neutron scattering is the ability to vary the scattering contrast by either selective deuterium labelling or using a mixture of hydrogenated(H)/deuterated(D) solvents. In soft matter, contrast variation schemes have allowed studies of individual components and partial structures.

The above applied to a general system in which the contrast was given by a polymer vs. a background with zero scattering length. This is not what one wants to measure nor is it generally the case. The scattering intensity is proportional to the contrast factor and can be changed by deuterium (D) labelling. Using D-labelling to increase the contrast factor is common method to study polymer systems by small angle neutron scattering (SANS) technique. It is worth to mention that the experimental proof of the random coil conformation of polymer chains in the melt or in the glassy state, as proposed in the 1950s by Flory [7] was only possible in the 1970s [8] with the development of SANS. This SANS experiment was performed in Forschungszentrum Jülich using using contrast variation and deuteration of single molecules.

We will not dwell on the numerous applications of contrast variation method in this lecture and refer to the lecture notes B3. It is worth to mention that for polymer melts a mixture of about 10% H-polymer in a matrix of D-polymer is typically used.

2.4   Scattering of polymer blend

Let us consider now the mixture of H and D polymers. The total structure factor in a dense system is obtained from Eq.(12) and split into an intra-chain and an inter-chain part.

$$S(q) = MN^2P(q) + M(M-1)N^2R(q) \tag{22}$$

If we consider a mixture of 2 polymers (H and D) with M chains, which are identical in length with N monomers and have no preferential interactions with each other, we define the form factor $P(q)$ and the inter-chain structure factor $R(q)$ which is sometimes even playing a

dominating role. $P(q)$ has been defined before and $R(q)$ is formulated similarly with double sums:

$$P(q) = \frac{1}{N^2} \sum_{m,l}^{N} < \exp(iq(r_{1,m} - r_{1,l})) >$$

$$R(q) = \frac{1}{N^2} \sum_{m,l}^{N} < \exp(iq(r_{1,m} - r_{2,l})) >$$

(23)

The total coherent scattering intensity then becomes (for $N_H = N_D$):

$$I(q) = \frac{1}{V_s} \sum_{i}^{2} \sum_{j}^{2} b_i b_j S_{ij}(q)$$

(24)

With for $i = H$ and $j = D$

$$S_{DD}(q) = M_D N_D^2 P_D(q) + M_D^2 N_D^2 R_D(q)$$

$$S_{HH}(q) = M_H N_H^2 P_H(q) + M_H^2 N_H^2 R_H(q)$$

(25)

$$S_{HD}(q) = M_D M_H N_H N_D R_{HD}(q)$$

With the incompressibility hypothesis one gets to

$$S_{DD} + S_{DH} = 0 \rightarrow S_{DD} = -S_{HD}$$

$$S_{DD} = -S_{HH}$$

(26)

We have implicitly assumed that the hydrogenous and deuterated polymers are identical and therefore exhibit the same intra-chain $P(q)$ and inter-chain structure factor $R(q)$, independent of the isotopic labelling. The inter-chain contribution can be expressed in terms of the intra-chain contributions which simplifies the scattering intensity to a rather simple expression in terms of a single chain contribution factor only.

$$I(q) = \frac{1}{V_s} (b_D^2 S_{DD}(q) + b_H^2 S_{HH}(q) + 2b_D b_H S_{HD}(q))$$

$$I(q) = \frac{(b_D - b_H)^2}{V_s} S_{DD}(q)$$

(27)

2.5   Scattering of ring polymers

Ring polymers are very interesting macromolecules in that they differ structurally and dynamically from the typical linear polymers due to the total absence of chain ends that are dominating for the dynamics in the melt. They have become of renewed interest in the last years due to the synthetic efforts in producing linear contaminant-free ring structures while minimizing concatenation reactions. As Fig. 8 shows, there are 2 different ways to define the intra-chain correlation between two monomers i and j. Both ways are equally probable and the

occurrence is taken into account by multiplying the probabilities of selecting each different path. The form factor for a polymer ring can be calculated using a multivariate Gaussian distribution approach [9]. For a Gaussian polymer ring, $P(q)$ can be calculated as follows:

$$P(q) = \frac{1}{N^2} \sum_{i,j}^{N} \exp\left(-\frac{q^2 \langle r_{ij}^2 \rangle}{6}\right) \tag{28}$$

For the mean square distance between 2 monomers and the corresponding form factor $P(q)$ we obtain then in analogy with the linear chain and including the closure relation

$$\langle r_{ij}^2 \rangle = l^2 |i - j| \left(1 - \frac{|i - j|}{N}\right)$$

$$\rightarrow P(q) = \frac{1}{N^2} N \sum_{i,j}^{N} \exp\left(-\frac{q^2 l^2 |i - j|}{6}\left(1 - \frac{|i - j|}{N}\right)\right) \tag{29}$$

The single chain structure factor differs in a sensitive way from the pure linear curve as can be seen in Fig. 8. It leads to a peaked structure in the Kratky representation. This expresses the similarity of a ring polymer with a star or branched polymer [9]. The monomers of a ring are on average located at closer distance to the center of mass than in a linear chain, so an enhancement of the compacticity of the structure can be expected. The peak arises due to the increased correlation through the closed cycle and is therefore related to R_g of the ring. Latter can be calculated to be smaller than the linear by $\sqrt{2}$.
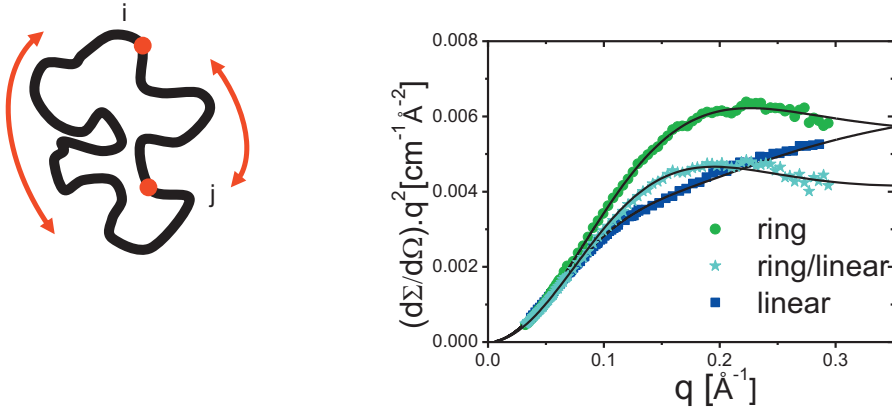
Fig. 8: Left: A polymer ring can be constructed by closing a linear chain. 2 possible ways to define the distance r_{ij}. Right: Kratky representation of scattering intensity to highlight the difference in the structure due to ring closure. In the ring/linear mixture the peak shifts to lower q and indicates a swelling of the ring conformation by penetrating linear chains. For the details see [11].

Ring polymers in the melt are significantly more compact than the Gaussian prediction. The form factor may be quantitatively described in terms of Gaussian linear chain and including the closure relation. The structural properties of the polymer ring scale with the size (N). In particular, the obtained scaling exponent ν consistently describes not only the chain statistics

but also the N dependence of the radius of gyration $R_{g,ring}^2 \approx N^{2\nu}$ as well as the N-dependence of the peaking in the Kratky plot. The exponent $\nu=0.43$ is different from the linear Gaussian chain ($\nu=0.5$) For the larger rings the evolution towards a conformation in the direction of a mass fractal is observed. [10]

2.6 Supramolecular polymers: an example for RPA

In the section 2.4 the structure factor of polymer blend consisted of deuterated and hydrogenaited polymers has been presented. Here we describe the neutron scattering experiment on the supramolecular polymer melt cosisted of the linking building blocks. Typically these polymers are formed by means of weak-to-intermediate non-covalent interactions such as e.g. hydrogen bonding. One of the main advantages of hydrogen bonding is its reversibility towards e.g. temperature or mechanical distortion.

We focus on well-defined bifunctional and differently isotope-labeled building blocks of poly(ethylene glycol) (PEG) chains. Thymine (Thy) and Diaminotriazine (DAT) hydrogen bonding groups were implanted telechelically, showing chain-like associations.

Although neutron scattering techniques are ideally suited to elucidate the underlying microscopic structure, these systems cannot be investigated in the same way as normal polymers. A conventional small angle scattering (SANS) experiment bases on the contrast between hydrogeneous and deuterated polymer specimens in the melt allows to study the form-factor of single polymer chains. In supramolecular melts, on the other hand, hydrogeneous and deuterated associated building blocks would randomize within the shortest times, e.g. driven by temperature. In this way all information about the association in the melt state is lost. The only way to measure the supramolecular association by SANS is facilitated by formation of an as ideal as possible alternating structure of hydrogeneous and deuterated heterocomplementary building blocks. This then results in a correlation peak in the scattering curve, as is the case for the covalent analogue [12]. Any competition between hetero- and homo-complementarity, however, in this living case inevitably leads to random copolymerization as well and destructive interference of correlations and that, however, can also be quantified by an evaluation of the SANS curves.

The scattering intensity or structure factor for a general multicomponent polymer system is commonly described by the random phase approximation (RPA), which is considerably simplified for binary systems A and B where A and B have different scattering length densities and lack any specific interaction. In the case presented here A is hydrogenated block together with supramolecular groups and B is a deuterated block. This theory has been reviewed in detail for most common architectures in the literature to which we refer [12–14] and here we represent the simple case of linear blocks as an example.

The classical RPA expression which is obtained, should apply for the envisaged purely alternating AB multiblock copolymer with:

$$\frac{d\Sigma}{d\Omega} \frac{1}{\Delta\rho^2} = S_{RPA} = \frac{S_{AA}^0 S_{BB}^0 - (S_{AB}^0)^2}{S_{AA}^0 + S_{BB}^0 + 2S_{AB}^0} \tag{30}$$

S^0 stand for the partial ideal structure factors of both building blocks . For simplicity, here and in the following equation the q-dependence was dropped.
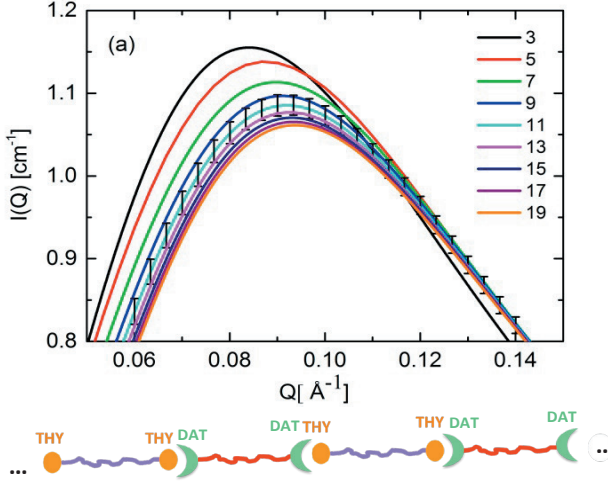
Fig. 9: Scattering intensities calculated for the supramolecular melt consisted of alternating associated building blocks (number of blocks varies from 3 to 19) using Eq.(30). Sketch on the bottom represents the structure of supramolecular melt: Thy-H-Thy are hydrogenated blocks; DAT-D-DAT represents deuterated blocks.

The partial structure factors (AA, BB and AB) which now appear in this resulting multiblock copolymer differ from the pure diblock ones due to the new correlations that arise between the blocks. Intra- and interblock correlations for the same and different scattering components have to be accounted for. In the Figure 9 calculations of the scattering intensity for the supramolecular melt consisted of alternatingly associating blocks is presented. The peak amplitude and position is the result of a delicate balance of the S_{AA} and S_{BB} versus S_{AB} terms (Eq.(30)). Since the forward scattering gives the weight average of the mass distribution, the peak height can be considered as proportional to weight-averaged aggregation number $\langle N_{agg} \rangle_w$ [15].

3   Dynamics

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 (Fig. 2). 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 2 ps to 600 ns and accesses momentum transfers q between 0.01 Å^{-1} and 3 Å^{-1}.

Coherent quasi- and inelastic neutron scattering reveals the dynamic structure factor $S(q,t)$ or its Fourier transformed counterpart $S(q,\omega)$ [16]

$$S(q,t) = \frac{1}{N} \sum_{ij} \langle \exp(-i\vec{q}\vec{r}_j(t)) \exp(-i\vec{q}\vec{r}_i(0)) \rangle \tag{31}$$

where $\vec{r}_j(t)$ and $\vec{r}_i(0)$ are the position vectors of the scatterers at time t and time $t = 0$ respectively, N is the number of scatterers and $\vec{q}/(2\pi)$ 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 $|\vec{b}|^2$.

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.(31) 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 $(\langle b^2 \rangle - |\vec{b}^2|)$.

In Gaussian approximation which is widely used for the calculation of neutron dynamic structure factors for polymer dynamics Eq.(31) is approximated by

$$S_{inc}(q, t) = \frac{1}{N} \sum \exp\left(-\frac{q^2}{6}\langle(\vec{r}_i(t) - \vec{r}_i(0))^2\rangle\right) \tag{32}$$

3.1 The standard model for polymer dynamics (Rouse)

If we want to describe the motion of a polymer, we could start with the atoms of a chain (see also chapter 2.1) 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. 10).

This is achieved with the Rouse model [17]. 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.
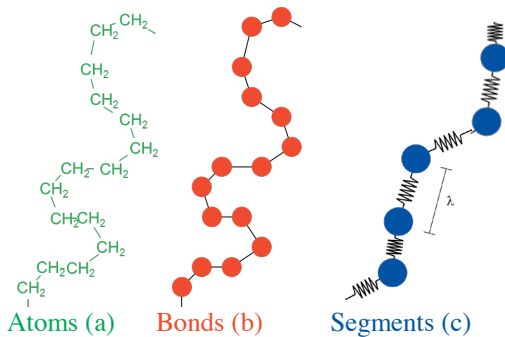
Fig. 10: Schematic representation of the coarse graining process that leads to the Rouse model

(a) Polyethylene chain
(b) Unified atom chain
(c) Chain represented by beads and springs

With that we can write down a Langevin equation for the chain segmental motion.

$$\zeta_0 \frac{dr_n}{dt} = \frac{k_B T}{l^2}(r_{n+1} - 2r_n + r_{n-1}) + f_n(t) \tag{33a}$$

$$\zeta_0 \frac{dr_n}{dt} = k \frac{\partial^2 r_n}{\partial n^2} + f_n(t) \tag{33b}$$

Here ζ_0 is the friction of the beat with the heat bath $k_B T/l^2 = \kappa$ is the entropic spring constant, 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. [18]). Eq.(33b) is the continuous version of Eq.(33a), 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

$$\eta = \frac{\zeta_0 l^2 N \rho}{36} \tag{34}$$

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.(33) leads to

$$D = \frac{k_B T}{\zeta_o N} \propto N^{-1} \tag{35}$$

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 $6D\tau = R_{ee}^2 = Nl^2$, where R_{ee} is the chain end-to-end distance.

Using Eq.(35) and equating for τ_R we get

$$\tau_R \simeq \frac{N^2 l^2 \zeta_0}{6 k_B T} \tag{36}$$

differing only slightly from the correct value. A solution of Eq.(33) yields
$\tau_R = \dfrac{N^2 l^2 \zeta_0}{3\pi^2 k_B T} = \tau_0 N^2; \tau_0 = \dfrac{l^2 \zeta_0}{3\pi^2 k_B T} = \dfrac{1}{W\pi^2}$. W is elementary Rouse rate which is related to

monomer (bead) friction ζ_0. Now, any subsection of the chain with $\dfrac{N}{p}$ segments relaxes as the

whole chain. Therefore, we have

$$\tau_p \propto \tau_0 \left(\frac{N}{p}\right)^2 \tag{37}$$

at the time τ_p the chain section with N/p monomers moves over its own distance. Thus,

$$< (r_n(\tau_p) - r_n(0))^2 > = l^2 \frac{N}{p} = l^2 \left(\frac{\tau_p}{\tau_0}\right)^{\frac{1}{2}} \tag{38}$$

the most right part of Eq. (38) follows from Eq. (37)

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 Eq. (3) with the proper boundary conditions of force free ends (see [1]). They turn out as

$$\phi_p(n) = \frac{1}{N} \cos\left(\frac{p\pi}{N} n\right) \tag{39}$$

These modes relax with the characteristic times $\tau_p = \dfrac{\tau_R}{p^2}$.

Since this Eq.(38) 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. (33), the final result for the time dependent mean square displacement is:

$$\langle r_n^2(t) \rangle = \sqrt{\frac{12 l^2 k_B T}{\pi \zeta_0} t} \tag{40}$$

The segment self correlation function that is measured with quasielastic incoherent neutron scattering directly accesses this quantity. In Gaussian approximation we have

$$S_{self}(q, t) = \exp\left[-\frac{q^2}{6}\langle r^2(t)\rangle\right] = \exp[-D_R q^2 t] * \exp\left[-\frac{2}{\sqrt{\pi}}(\Omega_R(q)t)^{\frac{1}{2}}\right];$$

$$\Omega_R(q) = \frac{k_B T l^2}{12\zeta_0} q^4 \tag{41}$$

The second part of this equation is obtained by inserting of Eq. (40). $\Omega_R(Q)$ 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 [19]. 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.
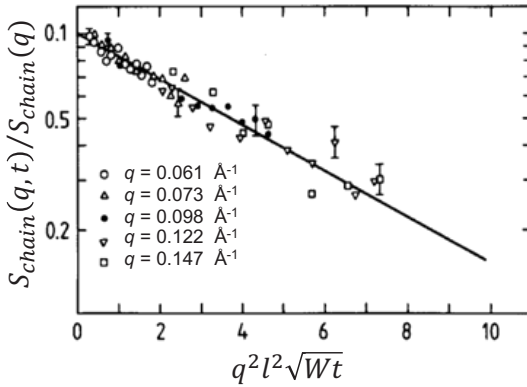
Fig. 11: Dynamic structure factor for the segmental self motion in a PDMS melt. The data are scaled with the Rouse variable (Eq.(41)). Solid line predicted $t^{\frac{1}{2}}$ relaxation by the Rouse model.

Figure 11 displays the obtained self correlation function for PDMS [20] in a presentation where the logarithm of the scattering function is plotted versus $(\Omega_R(Q)*t)^{1/2}$ the scaling variable of Eq.(41). In this way all the data collapse on one single master curve that according to Eq.(41) 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 [18]. The result may be expressed in terms of the Rouse modes (Eq.(39)) and the corresponding relaxation times $\tau_p = {}^{\tau_R}/_{p^2}$.

$$S_{chain}(q,t) = \frac{1}{N}\exp[-q^2 D_R t]\sum_{i,j}\exp\left\{-\frac{1}{6}|i-j|q^2 l^2\right\} *$$

$$\exp\left\{-\frac{2}{3}\frac{R_{ee}^2 q^2}{\pi^2}\sum_p \frac{1}{p^2}\cos\left(\frac{p\pi i}{N}\right)\cos\left(\frac{p\pi j}{N}\right)\left(1-\exp\left\{-\frac{tp^2}{\tau_R}\right\}\right)\right\} \qquad (42)$$

for small q ($qR_{ee} < 1$) the second and third terms are negligible and $S_{chain}(q,t)$ describes the centre-of-mass diffusion of the chain.

$$S_{self}(q,t) = \frac{1}{N} S_{chain}(q,t) = \exp(-D_R q^2 t) \tag{43}$$

For $qR_{ee} > 1$ and $t < \tau_R$ the internal relaxation processes 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 (Eqs.(19)-(20)). Similar to the self correlation function chain still follows the universal decay if plotted versus the Rouse variable $(\Omega_R*t)^{1/2}$.

By using a mixture of protonated and deuterated chains, NSE directly measures the coherent single chain dynamic structure factor $S_{chain}(q,t)$ that is the spatial Fourier transform of the monomer-monomer dynamic pair correlation function (see also lecture notes A4 and B4). As an example, the NSE data for the bulk PDMS melt are shown in Figure 12.
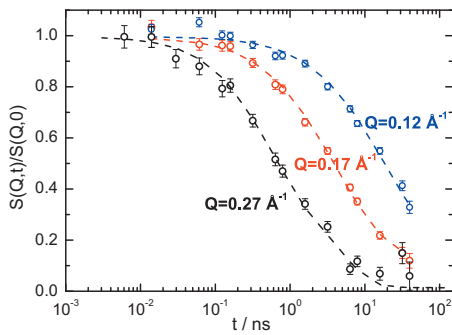
Fig. 12: NSE data for the PDMS melt measured at different values of the momentum transfer q [21]. Dashed lines present fitting curves obtained by using Eq.(42).

The experimental decay (Fig.12) can be nicely described by the Rouse model demonstrating validity of the eigenmode approach.

3.2   Role of inter-chain interactions: entanglements

Macroscopically the dynamics of long chain polymer melts is characterised by a plateau regime in the dynamic modulus (Fig.3). 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 [18] and Doi and Edwards [1]. 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.

$$d^2 = \frac{4}{5} \frac{R_{ee}^2}{M} \frac{k_B T}{G_N^0} \tag{44}$$

where R_{ee} 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.

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_{tube} = $ (contour length L)^2. Thus, for the longest relaxation time we obtain

$$\tau_\eta \approx \frac{L^2}{D_R} \approx N^3 \zeta_0 \tag{45}$$

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

$$D_{rep} = \frac{R_{ee}^2}{\tau_\eta} \approx \frac{N}{N^3 \zeta_0} \approx \frac{1}{N^2 \zeta_0} \tag{46}$$

We now consider the predictions of the reptation model for the mean square displacements of the chain segments [1,22]. For short times, when the chain segments have not yet realized the topological constraints ($r^2 < d^2$) we expect unrestricted Rouse motion $\langle r^2(t)\rangle \approx t^{\frac{1}{2}}$ (Eq.(40)). Experimentally this was the case for PDMS (Figure 11). 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.(40) 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 three-dimensional real space is $d\left\langle \left(s_n(t) - s_n(0)\right)^2\right\rangle^{1/2}$. With that we obtain

$$\langle r^2(t)\rangle = \begin{cases} 2d\left(\frac{k_B T \ ^2 t}{\zeta_0 \pi}\right)^{1/4} & \tau_e < t < \tau_R \\ \\ 2d\left(\frac{k_B T t}{N\zeta_0}\right)^{1/2} & \tau_R < t < \tau_d \end{cases} \tag{47}$$

In Figure 13 the two situations correspond to the second and the third process. 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.(45)). 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 mean square displacement. However, as Fatkullin and Kimmich have shown [23], 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 $t > \tau_e$.

$$S_{self}(q, t > \tau_e) = \exp\left(\frac{q^4 d^2}{72} \frac{\langle r^2(t)\rangle}{3}\right) \text{erfc}\left[\frac{q^2 d}{6\sqrt{2}} \sqrt{\frac{\langle r^2(t)\rangle}{3}}\right] \tag{48}$$

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. (47) 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 $t^{\frac{1}{2}}$ to a $t^{\frac{1}{4}}$ law. Indeed, the predicted cross over from $t^{\frac{1}{2}}$ to a $t^{\frac{1}{4}}$ was nicely observed by NSE [23].

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.

Figure 14 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 presented by Eq.(42) 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.
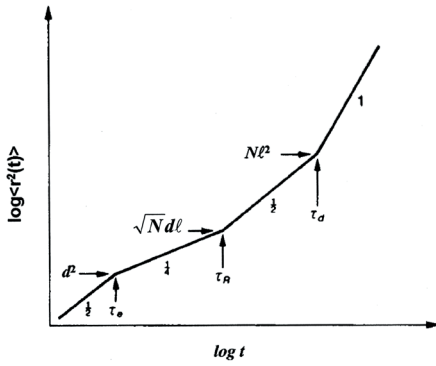
Fig. 13: Predicted time dependence of the segmental mean square displacements. Several power laws are visible. At short times we observe the Rouse regime (proportional $t^{1/2}$) then local reptation takes over (proportional $t^{1/4}$) there after reptation prevails ($t^{1/2}$) and finally translational chain diffusion occurs (proportional t).

(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.
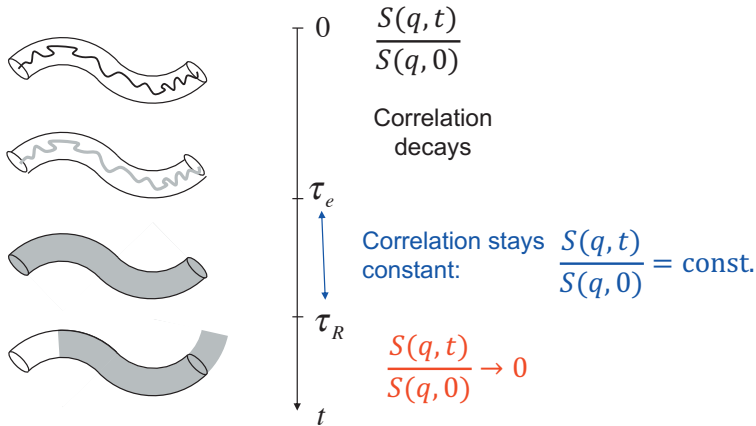
Fig.14: Schematic presentation of the various stages in the time development of the single chain dynamic structure factor. At short times unrestricted Rouse Dynamics time takes place beyond τ_e fluctuations fill the tube; For times larger than τ_R the chain creeps out of the tube.

De Gennes [24] and Doi and Edwards [1] 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.

$$\frac{S_{chain}(q,t)}{S_{chain}(q)} = \left[1 - \exp\left(-\frac{q^2 d^2}{36}\right)\right] S^{loc} + \exp\left(-\frac{q^2 d^2}{36}\right) S^{esc} \tag{49}$$

The local reptation part was calculated as

$$S^{loc}(q,t) = \exp\left(\frac{t}{\tau_0}\right) \mathrm{erfc}\left(\sqrt{\frac{t}{\tau_0}}\right) \tag{50}$$

With $\tau_0 = \frac{36}{W l^4 q^4}$ is related to the Rouse rate W.

A general expression for $S_{esc}(q,t)$ due to pure reptation was given by Doi and Edwards [1]. 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.

The confined motion of a single chain in the melt and effect of topological constraints of the tube were illustrated experimentally for different polymers (see [16] and reference inside). Figure 15 presents the dynamic structure factor from entangled polyethylene (PE) melt (M_w=36 kg/mol). Now as a function of the Rouse variable $(\Omega_R^* t)^{1/2}$ [25] other than in Figure 11, 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 $d = 46$ Å, a value, that is a lower estimate for the tube diameter, since S^{loc} is not fully relaxed. The horizontal lines in Figure 15 are the predictions from this Debye Waller factor estimate. A description with the tube dynamic structure factor (Eq.(49)) yields only the slightly higher value of $d = 48$Å.
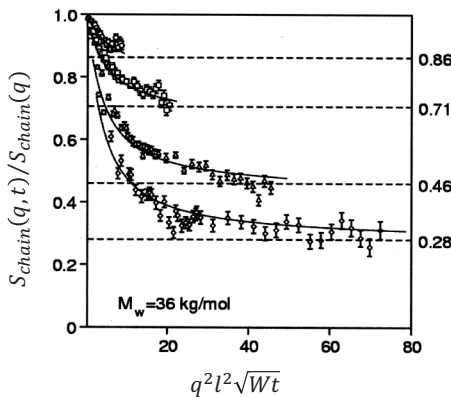
Fig. 15: Single chain dynamic structure factor from a long polyethylene melt scaled with a Rouse variable. The dashed line described the Debye Waller factor approximation for the long time plateaus (see text).

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 1 on the macroscopic properties 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). The CLF effect evolves from the participation of the chain ends in the local reptation process and is an inherent property 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. In order to clarify CLF and CR further and separate these different dynamic processes, a number of pioner experiments were designed and performed (see e.g. [26,27]).

3.3   Influence of polymer-solid surface interaction on polymer dynamics

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) [21]. Fig. 16 displays the system. Fig. 16a 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.

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 (see Fig. 12 representing bulk PDMS measured as a reference).
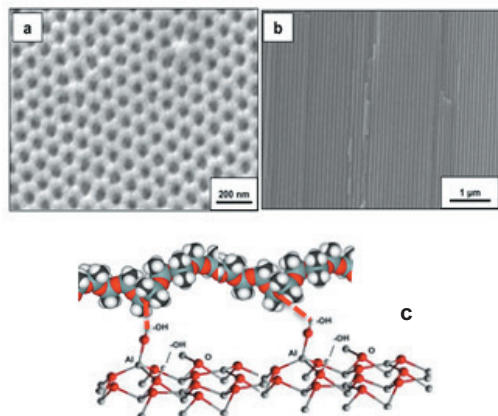
Fig.16: (a, b) Electron micrograph (SEM) of anionic aluminum oxide (AAO) nanopores, (c) model interaction of a PDMS chain with the AAO surface OH groups via H-bonds.

Figure 17 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 (compare with the bulk PDMS behaviour presented in the fig. 12).
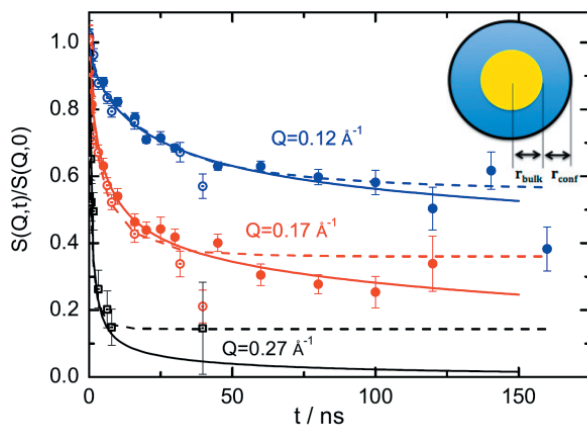
Fig.17: NSE results for confined PDMS. Dashed lines show the fitting of the data using East. Solid lines present the fitting using a continuous transition region from suppressed to free Rouse models (see text). A sketch of the two-phase model is shown in the inset.

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 $\alpha(q)$. Thus, we have

$$S_{chain}(q,t) = (1 - \alpha(q))S_{Rouse}(q,t) + \alpha(q) \qquad (51)$$

where $S_{Rouse}(q,t)$ is the dynamic structure factor originating from the Rouse model (see Eq.(42)).

With this approach we arrive at the dashed lines in Fig. 17. As already seen qualitatively the plateau $\alpha(q)$ 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

$$\alpha(q) = \exp\left(-\frac{q^2 d^2}{36}\right) \qquad (52)$$

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

$$S_{chain}(q,t) = A_{bulk}S_{bulk}(q,t) + A_{conf}S_{conf}(q,t) \qquad (53)$$

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 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 (Eq.(42) 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. 17. 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 $d = \sqrt{\left\langle R_{E,subchain}^2 \right\rangle} = \sqrt{N/p_{min}} \, l^2 = 3,4nm$. 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. 18.

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.

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.
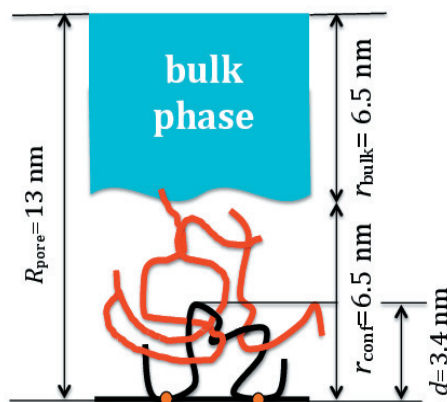
Fig.18: Schematic representation of the artificial surface-induced entanglements in the confined polymer melt. The black line represents the chain adsorbed on the surface of an AAO nanopore, and the red lines show entangled chains in the confined phase.

3.4 Dynamics of ring polymers

As it has been shown in the chapter 2.5, the ring polymers exhibit unique topology. Due to luck 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.o.m.) behavior $\langle r_{cm}^2(t) \rangle \sim t^{3/4}$ at early times, before the transition to normal translational diffusion takes place at about $\langle r_{cm}^2 \rangle = 2.5 R_g^2$. At higher molecular weights $D \sim N^{-2}$ was found. Regarding the segmental relaxation, simulations observe a significant slowing down towards $\langle r_{seg}^2(t) \rangle \sim t^{\nu}$ with $\nu = 0.25 - 0.35$. Such a power law is also characteristic for the local reptation regime where the Rouse modes relax within the stiff confining tube yielding $\langle r_{seg}^2(t) \rangle \sim t^{1/4}$. 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. [28]

Analysis of segmental dynamics of ring polymers obtained by NSE was done based on Rouse dynamic structure factor modified for rings [29]:

$$
\begin{aligned}
S(q,t) = \frac{1}{N} \sum_{i,j} \exp \Bigg[& -\frac{\langle r_{cm}^2 \rangle q^2}{6} - \frac{1}{6} q^2 l^2 |i-j|^{2\nu} \left(1 - \frac{|i-j|}{N} \right)^{2\nu} \\
& - \frac{4q^2 N^{2\nu} l^2}{6\pi^2} \sum_{p_{min,even}}^{N} \frac{1}{p^2} \cos\left(p\pi \frac{|i-j|}{N} \right) \left(1 - \exp\left(-\frac{tp^2}{\tau_R} \right) \right) \\
& - F \frac{4q^2 N^{2\nu} l^2}{6\pi^2} \sum_{p,even}^{p_{min}-2} \frac{1}{p^2} \cos\left(p\pi \frac{(i-j)}{N} \right) \left(1 - \exp\left(-\frac{tp^4}{\tau_{cross}} \right) \right) \Bigg]
\end{aligned}
\tag{54}
$$

The sum in Eq.(54) over the indices i, j runs over all monomer coordinates of the ring, the sum over p in the exponent addresses all Rouse modes (Eq.(39)) – for a ring only even modes p contributes. It turned out that 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. It is remarkable that for different ring sizes (10 kg/mol or 20 kg/mol) the number of reduced Rouse modes p_{min} refers to a ring section of about 60 monomers that can relax without topological hindrance. For the detailes see [10] .

The presented experiments allowed the first access to the dynamics of ring systems on a molecular scale. The investigated rings are placed in the crossover regime towards mass-fractal behavior with following peculiarities (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 mean square displacement $\sim t^{3/4}$ 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 $\nu = 0.32$ is found.

3.5   Association dynamics of supramolecular polymers

Supramolecular polymerization has emerged as an attractive alternative for the synthesis of functional polymer materials with tailored properties (e.g. self-healing). In particular, supramolecular polymers based on hydrogen bonds are effective, stable and allow the tuning of bonding strength through the choice of specific hydrogen bond matching pairs [30]. One of the simplest examples of associating supramolecular systems are telechelic polymers [31]. Typical building blocks are oligomers (spacers) having self-complementary binding stickers with hydrogen bonding motifs [32]. Self-complementary denotes the association of two identical end-groups, whereas in the heterocomplementary case, two different end-groups are binding together.

Combined analysis of small-angle neutron scattering (SANS) partially considered in the chapter 2.6, linear rheology and pulsed field gradient (PFG) NMR spectroscopy experiments on the supramolecular association and chain structure of well-defined telechelically modified short poly(ethylene glycol) (PEG) building blocks in the bulk was performed. PEG was functionalized with directed heterocomplementary triple hydrogen-bonding end-groups, thymine (Thy) and diaminotriazine (DAT). The polarity of the backbone polymer is comparable to the end groups and avoids clustering of the groups basing on energetic arguments. Their linear association behavior in the ideal melt state was investigated on the microscopic/molecular level as a function of temperature. By means of a selective labeling scheme, which should ideally lead to the formation of alternating hydrogeneous-deuterated building block sequences if the hydrogen bonding reaction is exclusively heterocomplementary, we showed that the Thy–DAT association is dominant and a Thy–Thy homoassociation is approximately three times less probable. Latter non-directed association gives rise to a considerable amount of random-copolymerization without affecting seriously neither the macroscopic melt viscosity nor the diffusivity of the supramolecular associates. From the q-dependence of a multiblock RPA structure factor measured by SANS, the linear association in the melt is confirmed. Furthermore, the diffusion and viscosity study reveals simple Rouse dynamics of supramolecular polymer chains with molecular weight much larger than the entanglement mass M_e (M_e of PEG is 1.9 kg/mol). The Rouse-like dynamics of long supramolecular chains indicates a short lifetime of hydrogen bonds of the end groups. Our results are in excellent agreement with the related polycondensation theory [15].

In addition, the pulsed field gradient PFG NMR data clearly shows that all experiments were performed in the fast exchange regime. With a minimum experimental observation time of $\Delta = 50$ ms, the requirement of fast exchange limits the lifetime to $\tau_H \ll \Delta \rightarrow \tau_H \ll 50$ ms. A much more stringent limit comes from the observation of Rouse dynamics for all associated chains. The chain length of the building blocks amounts to the length of one entanglement strand; i.e., all associated chains should sense topological constraints in some way. However, this does not seem to be the case. Thus, the lifetime of the hydrogen bonds must be short enough that confining tubes do not exist on the time scale of a few characteristic entanglement times τ_e which is the relaxation time of an entanglement strand of length N_e.

NSE measurements performed on the same system showed that the dynamics of the supramolecular chains is Rouse-like at the timescale up to 200 ns controlled by hydrogen bond breaking. Bond breaking changes the mode contributions but not the mode relaxation times as was assumed previously [33] Moreover, based on the previous results [15], the dynamical

NSE experiment directly reveals the lifetimes (lifetime of the H-bond and whole associate), which are one of the key parameters determining macroscopic properties of supramolecular polymer melts. In a temperature range from 353K to 450K we arrive at bond breaking times that range from about 800 ns to 30 ns. Despite the fact that the association numbers $\langle N_{agg} \rangle_w$ ~ 3-8 would imply entanglements and reptation type dynamics the viscosity and diffusion coefficients exactly correspond to the Rouse prediction for the distribution of association numbers. Entanglements obviously have no effect because of their premature release by chain breaking times in the 10-100's ns time scale [34].

4 Summary

We have presented some representative results on the structure and dynamics of macromolecules. We have discussed possibilities to investigate the *structure* of model polymeric chains in various environments and constitutions, some in more detail than others while keeping the focus on dense, amorphous and well-mixed systems with different topology. We have tried to summarize here only the most important concepts of the statistical structure of ideally monodisperse chains and we have demonstrated how the particular structure affects the scattering behaviour. This text should be a basis in order to be of help in calculating more complex structures which are built up from the here discussed linear polymer chain. It was a general aim thus to provide some useful basics which can be transferred to related sample systems. We gave an example how the linar chain approach can be extended to the case of ring and supramolecular polymers.

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 mentioned 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 supramolecular telechelic polymers 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. NSE spectra directly reveal the so far intangible H-bond lifetimes in the supramolecular melt and demonstrate that for both the microscopic and the macroscopic dynamics of the supramolecular ensemble the instantaneous average of the M_w distribution governs the system response at least as long as the Rouse picture applies.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.

References

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

[2] J. Bang and T. P. Lodge, Phys. Rev. Lett. **93**, 245701 (2004).

[3] S. Chen, T. Yan, M. Fischer, A. Mordvinkin, K. Saalwächter, T. Thurn-Albrecht, and W. H. Binder, Angew. Chemie **129**, 13196 (2017).

[4] W. A. Hendrickson, Acta Crystallogr. Sect. A Found. Crystallogr. **52**, C1 (1996).

[5] Paul J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969).

[6] C. Gerstl, G. J. Schneider, W. Pyckhout-Hintzen, J. Allgaier, S. Willbold, D. Hofmann, U. Disko, H. Frielinghaus, and D. Richter, Macromolecules **44**, 6077 (2011).

[7] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, 1953).

[8] R. G. Kirste, W. A. Kruse, and J. Schelten, Die Makromol. Chemie **162**, 299 (1972).

[9] B. Hammouda, NIST Cent. Neutron Res. 224 (2010).

[10] D. Richter, S. Gooßen, and A. Wischnewski, Soft Matter **11**, 8535 (2015).

[11] A. R. Brás, R. Pasquino, T. Koukoulas, G. Tsolou, O. Holderer, A. Radulescu, J. Allgaier, V. G. Mavrantzas, W. Pyckhout-Hintzen, A. Wischnewski, D. Vlassopoulos, and D. Richter, Soft Matter **7**, 11169 (2011).

[12] G. Hadziioannou, H. Benoît, W. Tang, K. Shull, and C. C. Han, Polymer (Guildf). **33**, 4677 (1992).

[13] H. Benoit and G. Hadziioannou, Macromolecules **21**, 1449 (1988).

[14] L. Leibler and H. Benoit, Polymer (Guildf). **22**, 195 (1981).

[15] M. Krutyeva, A. R. Brás, W. Antonius, C. H. Hövelmann, A. S. Poulos, J. Allgaier, A. Radulescu, P. Lindner, W. Pyckhout-Hintzen, A. Wischnewski, and D. Richter, Macromolecules **48**, 8933 (2015).

[16] D. Richter, M. Monkenbusch, A. Arbe, and J. Colmenero, *Neutron Spin Echo in Polymer Systems* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005).

[17] P. E. Rouse, J. Chem. Phys. **21**, 1272 (1953).

[18] P. G. de Gennes, J. Chem. Phys. **55**, 572 (1971).

[19] F. Mezei, C. Pappas, and T. Gutberlet, editors , *Neutron Spin Echo Spectroscopy* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003).

[20] D. Richter, B. Ewen, B. Farago, and T. Wagner, Phys. Rev. Lett. **62**, 2140 (1989).

[21] M. Krutyeva, A. Wischnewski, M. Monkenbusch, L. Willner, J. Maiz, C. Mijangos, A. Arbe, J. Colmenero, A. Radulescu, O. Holderer, M. Ohl, and D. Richter, Phys. Rev. Lett. **110**, 108303 (2013).

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

[23] N. Fatkullin and R. Kimmich, Phys. Rev. E **52**, 3273 (1995).

[24] P. G. De Gennes, J. Phys. **42**, 735 (1981).

[25] P. Schleger, B. Farago, C. Lartigue, A. Kollmar, and D. Richter, Phys. Rev. Lett. **81**, 124 (1998).

[26] A. Wischnewski, M. Monkenbusch, L. Willner, D. Richter, A. E. Likhtman, T. C. B. McLeish, and B. Farago, Phys. Rev. Lett. **88**, 058301 (2002).

[27] M. Zamponi, A. Wischnewski, M. Monkenbusch, L. Willner, D. Richter, A. E. Likhtman, G. Kali, and B. Farago, Phys. Rev. Lett. **96**, 238302 (2006).

[28] S. Gooßen, A. R. Brás, M. Krutyeva, M. Sharp, P. Falus, A. Feoktystov, U. Gasser, W. Pyckhout-Hintzen, A. Wischnewski, and D. Richter, Phys. Rev. Lett. **113**, 168302 (2014).

[29] G. Tsolou, N. Stratikis, C. Baig, P. S. Stephanou, and V. G. Mavrantzas,

Macromolecules **43**, 10692 (2010).

[30] T. Rossow and S. Seiffert, Adv. Polym. Sci. **268**, 1 (2015).

[31] P. G. Khalatur and A. R. Khokhlov, Macromol. Theory Simulations **5**, 877 (1996).

[32] J. Cortese, C. Soulié-Ziakovic, S. Tencé-Girault, and L. Leibler, J. Am. Chem. Soc. **134**, 3671 (2012).

[33] E. B. Stukalin and K. F. Freed, J. Chem. Phys. **125**, 184905 (2006).

[34] M. Monkenbusch, M. Krutyeva, W. Pyckhout-Hintzen, W. Antonius, C. H. Hövelmann, J. Allgaier, A. Brás, B. Farago, A. Wischnewski, and D. Richter, Phys. Rev. Lett. **117**, 147802 (2016).

E 3   Quantum Materials

Shibabrata Nandi

Jülich Centre for Neutron Science

Forschungszentrum Jülich GmbH

Contents

1 Introduction

Quantum materials are those in which microscopic quantum properties are observable at the macroscopic length scale. In the last few years, there have been a plethora of research activities on quantum materials. Such materials include superconductors, low dimensional quantum magnets, topological insulators and the latest one in this series is the Weyl semimetals. Electrons in the solid possess many degrees of freedom such as charge, spin, and orbital and characterized by the topological nature determined by the potential on the crystal lattice structure. The many degrees of freedom couple together and collectively manifest in various electrical, magnetic and optical properties of the materials as schematically shown in Fig.1 [1].

The collective electrons exhibit various macroscopic quantum phenomena, the representative example among all these is the superconductivity. One of the biggest challenges in condensed matter physics is to understand the unconventional superconductivity and to increase the superconducting transition temperature. Although significant progress has been made in the last decades, finding room temperature superconductor is still proved to be illusive. Apart from superconductivity, there are many quantum phenomena which are observable at room temperature such as magnetism and ferroelectricty. There are common features of these quantum phenomena: the order parameter which behaves like classical quantity despite the underlying quantum nature of the different phases.

The topological nature of the electronic states play a crucial role in understanding and predicting quantum properties. Topology of the band structure proved to be crucial for the understanding and developments in quantum materials. The quantum Hall effect, topological insulators and topological superconductors are all characterized by non trivial topologies in Hilbert space, the Berry phase which describes the connections and curvature of the subspace of Hilbert space plays the central role in the unified principle of topological nature.

The surface conductivity of the topological insulators depend on the topological properties of the bulk electronic wave functions and are therefore, exciting example of quantum materials. Weyl semimetals - whose quasiparticle excitations are Weyl fermions, are interesting because they show several quantum anomalies arising due to chiral properties of the Weyl fermions. Weyl semimetals show unique transport, optical and thermoelectric responses. For instance, they show unique surface states which is known as "Fermi Arcs". These Fermi arcs originates from the specific bulk band topologies and degeneracies of the energy levels near the Fermi level.

2 Quantum Materials: Weyl semimetals

As discussed in the previous section, quantum materials can be of various types. In the following sections, we like to discuss one particular class of quantum materials, namely Weyl semimetals. Weyl semimetals are those materials whose quasiparticle excitations are Weyl fermions. Weyl fermions played a crucial role in quantum field theory but never been observed as a fundamental particle in vacuum. Weyl fermions are massless spin half particle with definite chirality. Initially, it was though that neutrinos are Weyl fermions. However, with the discovery of finite neutrino mass [2, 3], there are no fundamental particle which is believed to be a
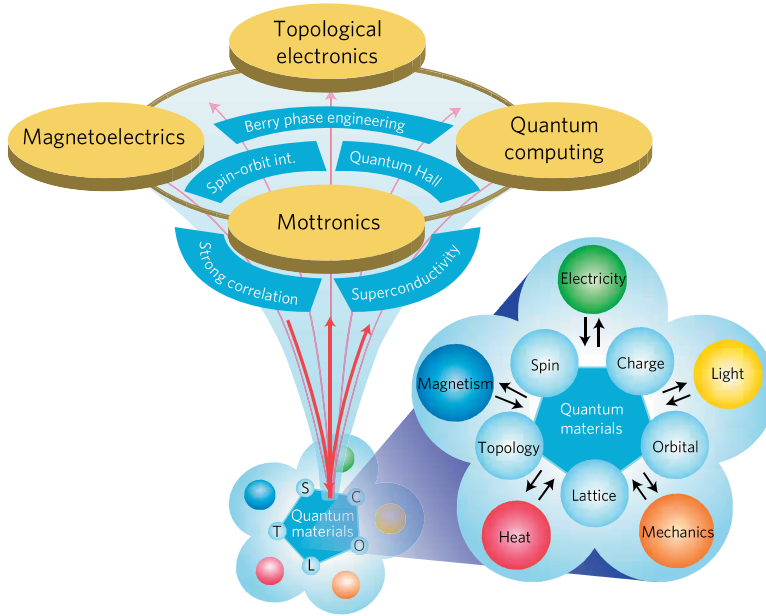
Fig. 1: The pentagon in the right hand corner shows various degrees of freedom of strongly correlated electrons in solids, which respond to external stimuli. These strong couplings lead to the emergent functions with the cross correlations among different physical observables and to developments towards the applications of emergent functions of quantum materials such as Mottronics, magnetoelectrics, topological electronics, and quantum computing. Adapted from Ref. [1].

maslesss Weyl fermion. Recently, it was theoretically predicted the existence of Weyl fermions as a quasiparticle excitation in condensed matter systems [4] and later it was experimentally discovered in various systems which are termed as Weyl semimetals.

2.1 Weyl Fermions: Basics Ideas

Here we first like to discuss two topics which will provide more insights into the existence of Weyl fermions in condensed matter physics. First, we like to discuss accidental degeneracies and then staring from basic Hamiltonian we will show the existence of Weyl fermions. We will also discuss various topological aspects of Weyl fermions which are necessary to understand basics properties manifested by the Weyl fermions.

2.2   Accidental degeneracies and band touching

During study of energy bands in solids physicist Conyers Herring [5] asked the following question: under what conditions two bands will have the same energy? Degenerecy of energy bands occur in solids at points with a high crystal symmetry. He also argued that the degeneracy might also occurs due to accident which he termed as 'accidental degeneracy'. To understand the accidental degeneracy, let us consider most general 2×2 Hamiltonian in 3D which can be in written as: [6]

$$H(\mathbf{k}) = f_0(\mathbf{k})\mathcal{I} + f_1(\mathbf{k})\sigma_x + f_2(\mathbf{k})\sigma_y + f_3(\mathbf{k})\sigma_z \tag{1}$$

where $\mathcal{I}$ is the identity matrix and σ's are the Pauli matrices. The energy splitting between the pair of energy levels is given by:

$$\Delta E = \sqrt{f_1^2 + f_2^2 + f_3^2}$$

In the absence of any symmetry, the energy levels can be brought into coincidence if all the parameters are simultaneously zero, *i.e.* $f_1 = f_2 = f_3 = 0$. Requirement of three independent parameters to be simultaneously zero indicate that we have to work in three special dimensions to make accidental degeneracy happen. Since there were no fine tuning of the parameters, we expect that the accidental degeneracy to be stable under small perturbations. The band touching points in Weyl semimetals are termed as Weyl points. In line with the previous argument, we expect Weyl points are stable under small perturbations *i.e.* we can move them in k space or even annihilate them. However, it is impossible to open a gap. Any perturbative term can be incorporated in one of the parameters in Eq. 1. Therefore, the system becomes a protected semimetal where Weyl fermions can be found [7]. This in contrast to the 2D Hamiltonian of Graphene where an external perturbation adds a mass term in the Hamiltonian and might open a gap in the energy spectrum.

2.3   Weyl Fermions from Dirac equation

The Dirac equation in d spatial dimension can be written as [6]:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0$$

Where we have taken effective speed of light $c = 1$ and the gamma matrices satisfy the anticommutation relation. $\mu = 0, 1,....d$ label time and space dimensions. When the mass term is zero, the Dirac equation reduces to the famous Weyl equation [8]:

$$i\partial_t \psi_\pm = H_\pm \psi_\pm$$

where

$$H_\pm = \mp \mathbf{p} \cdot \sigma$$

where $\psi_\pm$ are effectively two component vectors. Thus Weyl fermions propagates parallel or anti-parallel to the spin, which defines their chirality. In Weyl semimetals, momentum separated Weyl fermions with opposite chirality is present.

2.4   Topological aspects of Weyl Fermions

Various properties of Weyl semimetals such as Chiral magnetic anomaly and anomalous Hall effect can be understood from the topological properties of band structures. The Berry connection related to the topological aspects of band structure is defined as:

$$\mathcal{A}_n(\mathbf{k}) = -i \langle u_n(\mathbf{k}) | \nabla_{\mathbf{k}} | u_n(\mathbf{k}) \rangle$$

where $|u_n(\mathbf{k})\rangle$ describes Bloch wave function at the wave vector $\mathbf{k}$ and n denotes the band index. This can be alternatively defined as the surface integral of the Berry flux/ curvature. The Berry curvature $\Omega_n(\mathbf{k}) = \nabla_k \times \mathcal{A}_n(\mathbf{k})$ is the curl of Berry connection. Since the divergence of the curl is always zero, $\nabla_k \cdot \Omega_n(\mathbf{k}) = 0$. This generally holds as long as the bands are nondegenarate. However, this does not hold in a Weyl semimetal since at the Weyl node at least two degenerate bands meet. Therefore, the Berry connection and the Berry curvature are ill-defined at a Weyl node. The Berry curvature at the Weyl points of left and right chirality can be derived as: [9]

$$\Omega_R(\mathbf{k}) = \frac{1}{4\pi} \frac{\mathbf{k}}{k^3} \quad \text{and} \quad \Omega_H(\mathbf{k}) = -\frac{1}{4\pi} \frac{\mathbf{k}}{k^3}$$

Therefore, $\nabla_k \cdot \Omega_{R/L}(\mathbf{k}) = \pm\delta(\mathbf{k})$ which shows that Weyl nodes are indeed monopoles of Berry curvature. Thus, they act as a source/ sink of quantized Berry flux in the momentum space.

2.5   Detection of Weyl Fermions in Real Materials

There exist various properties of Weyl semimetals that characterize them. One is the observation of the surface Fermi Arcs which can be detected via Angle Resolved Photo-Emission Spectroscopy (ARPES). Chiral anomaly which manifests in anomalous longitudinal magnetoresistance, and anomalous Hall effect are the other two main features of Weyl semimetals. In the following sections, we like to first outline basic theories behind the above mentioned properties and give illustrative examples where they have been experimentally observed.

(a) Surface Fermi Arcs

In conventional solids, the Fermi surface on the surface forms a closed curve. However, in Weyl semimetals these loops becomes "arcs" as argued by Wan *et al*. [4] using band topology. Weyl points acts as a monopole or anti-monopole and acts as a source or sink of Berry flux as discussed earlier. The Fermi arcs terminate at the location of the bulk Weyl points as shown schematically in Fig. 2.

Angle-resolved photoemission spectroscopy (ARPES), scanning tunneling microscopy (STM), and quantum oscillations provide important insight into the band structure of the material. Particularly, ARPES becomes instrumental in determining surface states of the topological insulators due to it's intrinsic surface sensitivity. Surface Fermi arcs has been observed in various materials including NaP and TaAs [10–14]. Here we pick up TaAs as a representative example where Fermi Arcs were detected for the first time using ARPES [11].
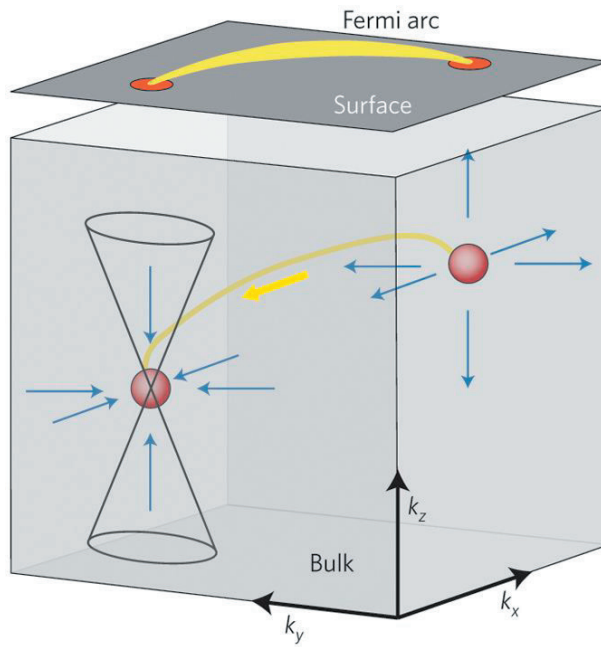
Fig. 2: Depiction of Weyl nodes and Fermi arc surface states in Weyl semimetals. The Weyl nodes acts as a source of Berry flux of monopoles and anti-monopoles. The dispersion is linear near the Weyl nodes. Adapted from Reference [7].
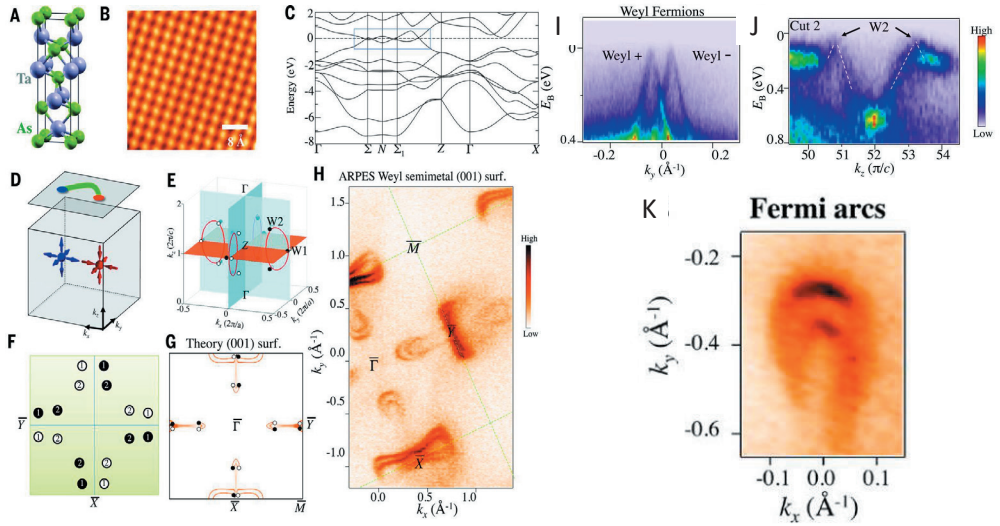
Fig. 3: (A, B) Crystal structure and STM topographic image of TaAs's (0 0 1) surface. (C) First-principles band structure calculations of TaAs without spin-orbit coupling. The blue box highlights the locations where bulk bands touch in the Brillouin zone. (D) Illustration of the simplest Weyl semimetal state that has two single Weyl nodes with the opposite (± 1) chiral charges in the bulk. (E) In the presence of spin-orbit coupling, position of the 24 Weyl nodes are shown (small black and white circles). Black and white show the opposite chiral charges of the Weyl nodes. (F) A schematic (not to-scale) showing the projected Weyl nodes and their projected chiral charges. (G) Theoretically calculated band structure of the Fermi surface on the (0 0 1) surface of TaAs. (H) The ARPES measured Fermi surface of the (0 0 1) cleaving plane of TaAs. (I) $E - k_y$ dispersion map which clearly shows the two linearly dispersive W2 Weyl cones. (J) $E - k_z$ dispersion map showing that the W2 Weyl cone also disperses linearly along the out-of-plane k_z direction. (K) Zoomed in image of the horse-shoe feature in H. Adapted from Ref. [11].

Low photon energy ARPES map for the TaAs is shown in Fig. 3 H. Three dominant features were observed: (a) a horseshoe shaped feature in the vicinity of the midpoint of each $\bar{\Gamma} - \bar{X}$ or $\bar{\Gamma} - \bar{Y}$ line, (b) a bowtielike feature centered at the $\bar{X}$ point, and (c) an extended feature centered at the $\bar{Y}$ point. The horseshoe shaped features suggest the existence of two arcs, and their termination points in k-space seem to coincide with the surface projection of the W2 Weyl nodes as shown in Fig. 3 G. Because the horseshoe feature consists of two nonclosed curves, it can either arise from two Fermi arcs or a closed contour. These two features can be distinguished from each other from the energy dependence of the these features which was termed as "bulk-boundary" correspondence by Xu *et al.* [11]. The calculated surface Fermi surface (Fig. 3G) using fine tuned parameters reproduces and explains ARPES data very well. This serves as an important cross-check that the data and interpretation are self-consistent. Specifically, the surface calculation indeed also reveals the horseshoe Fermi arcs that connect the projected W2 Weyl nodes near the midpoints of each $\bar{\Gamma} - \bar{X}$ or $\bar{\Gamma} - \bar{Y}$ line (Fig. 1G). In addition, their calculation shows the bowtie surface states centered at the $\bar{X}$ point, also consistent with the ARPES data.

Finally, bulk band structure of the TaAs was probed by soft x-ray (SX) ARPES, which reveal the existence of bulk Weyl cones and Weyl nodes as shown in Fig. 3 (I-J). This serves as an independent proof of the Weyl semimetallic state in TaAs. Xu *et al.* have also shown that the k-space locations of the surface Fermi arc terminations match with the projection of the bulk Weyl nodes on the surface BZ by combining both the surface and bulk Fermi surface data, further confirming topological nature of the Weyl semimetals (bulk-boundary correspondence).

Chiral Magnetic Anomaly

One of the main consequence of the existence of handedness of the Weyl fermions is the chiral anomaly [15] which manifests in anomalous magnetotransport properties as observed in various proposed Weyl semimetals [16, 17]. Chiral anomaly in condensed matter physics - is the non-conservation of electron quasiparticle number in the vicinity of a Weyl cone of a given chirality. The number of electron in the vicinity of each node in 3D is modified in the presence of electric (**E**) and magnetic fields (**B**) as: [6]

$$\frac{dn_{R/L}}{dt} = \pm \frac{e^2}{h^2} \mathbf{E} \cdot \mathbf{B} \tag{2}$$

Equation 2 immediately tells us that the number of electrons at an individual node is not conserved. However, chirality is balanced in a real condensed matter system, leading to axial charge current. The external magnetic field leads to Landau levels which disperse along the field direction and the chemical potential is shifted by the application of external electric field as shown in Fig. 4 . The zeroth Landau level can be shown to be chiral, *i.e.* it propagates only along or opposite to the field direction, with opposite velocities at the two chiral nodes. A current along the magnetic field proportional to the magnetic field strength is generated if a chiral chemical potential difference between the two chiral node exists. The current is given by:

$$j_c = \frac{e^2}{h^2} B \Delta\epsilon \tag{3}$$

where $\Delta\epsilon$ is the effective chemical potential difference. It can be noted that this current can not
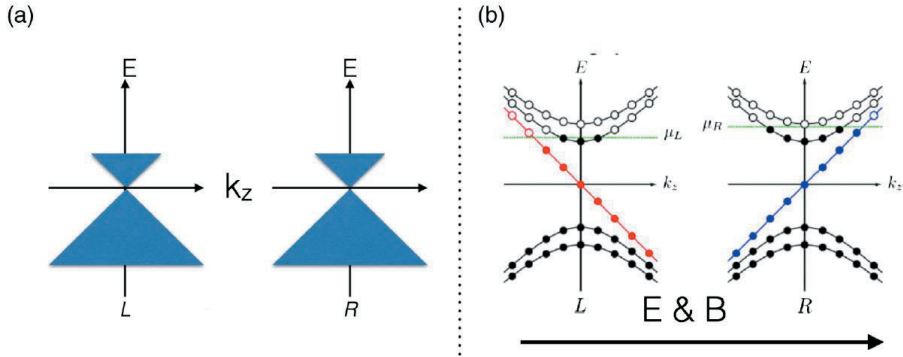
Fig. 4: Physical origin of the chiral anomaly. (a) Weyl nodes with left (L) and right (R) chiralities in the absence of external electric and magnetic fields. (b) Landau levels in the presence of external magnetic fields. The lowest landau level is chiral. In the presence of an external electric field parallel to the magnetic field creates a valley imbalace. Figure adapted from Ref. [6].

be an equilibrium dc current since no voltage is applied. However, in a nonequilibrium situation such as time-dependent electric fields can generate non-zero currents provided the frequency of the applied electric field is faster than the internode scattering time [18].

In a dc-settings, the density difference between two Weyl nodes is determined by the balance between chiral charge pumping (Eq. 2) and the inter-Weyl node scattering time ($1/\tau_a$). Therefore, from Eqs. 2 and 3, it can be seen that the chiral current $j_c \propto B(\mathbf{E}\cdot\mathbf{B})\tau_a$. Hence, magnetoconductivity (proportional to current) is quadratic in magnetic field strength and maximum along the field direction. It can be shown that the magnetoconductivity varies as [6]:

$$\sigma(B) = \sigma_0 + \frac{e^2 B^2 \tau_a}{4\pi^4 g(\epsilon_F)}$$

where $g(\varepsilon_F)$ is the density of states at the Fermi level and σ_0 is the zero field conductivity. Therefore, the magnetoconductivy is positive (magnetoresistance is negative) and varies quadratically along the field direction in a Weyl semimetal. In general, in a normal metal the magnetoresistance (MR) is usually positive since a magnetic field tends to restrict the motion of the electrons. However, following the previously mentioned argument, a negative magnetoresistance is expected in Weyl semimetal and it depends on the relative orientation of the $\mathbf{E}$ and $\mathbf{B}$ fields.

In realistic Weyl semimetals such as TaAs, NbAs, TaP, NbP and GdPtBi negative magnetoresistance has been observed [19–22] and was taken as an evidence of the chiral magnetic anomaly
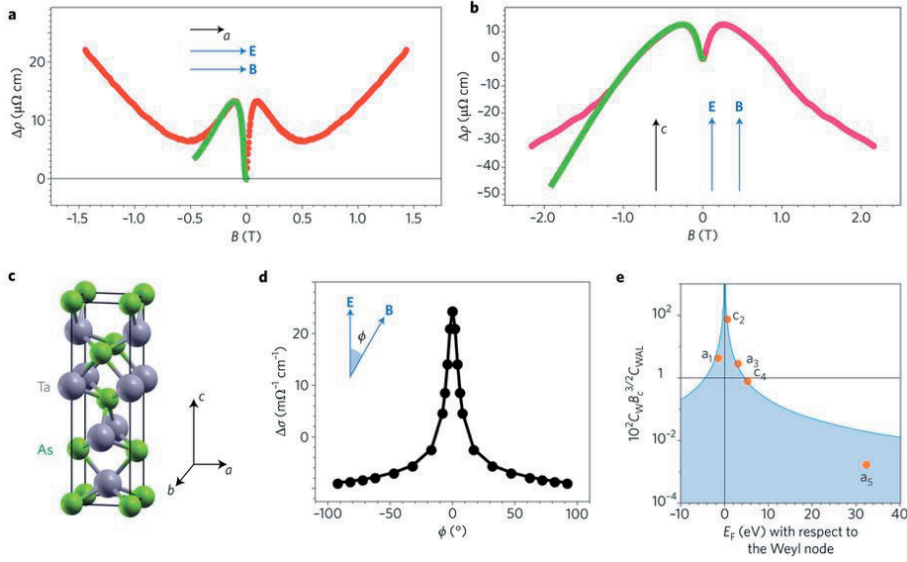
Fig. 5: (a, b) Longitudinal magnetoresistance for the TaAs sample at 2 K for the two crystallographic directions *a* and *c,* respectively. (c) Crystal structure of the TaAs (d) Magnetoresistance as a function of angle between electric and magnetic fields. (e) Chemical potential dependence of the chiral coefficient C_w. Figure adapted from [23].

as explained above. As a representative example, we show the negative magnetoresistnace for the TaAs [23]. Three regions of magnetoresistance with different characteristics can be identified from Fig. 5(a). In the low field region, there is a sharp increase in the magnetoresistance as expected from weak-antilocalization effect in Weyl semimetal [24]. In the intermediate field region, the magnetoresistance follows the dependence as expected from chiral magnetic anomaly in Weyl semimetals and is independent of the direction of the parallel field direction with respect to the crystallographic axis (compare Fig. 5(a) and (b)). In the higher field regime, magnetoresistance again increases with the magnetic field. The negative longitudinal magnetoresistance can be fitted with:

$$\Delta \sigma = N C_w B^2$$

where N is the number of Weyl nodes and C_w is the chiral coefficient. In the simplest form, chiral coefficient (C_w) can be written as $C_w = \frac{e^4 \tau_a}{4\pi^4 \hbar^4 E_F^2}$ [25]. As shown in Fig. 5(e) the fitted chiral coefficient has a strong dependence on the Fermi energy and can be regarded as a test to the Weyl nature of the electrons.

Anomalous Hall Effect

In the presence of an external magnetic field, the Hall effect (current flow perpendicular to electric field) is present in all conductors. In ferromagnetic metals like Fe, Co, and Ni, however, the Hall effect is anomalous and controlled by magnetization rather than by the Lorentz forces

[26]. Anomalous Hall effect can also be seen in Weyl semimetals. For the simplest example of magnetic Weyl semimetal, where the chemical potential is at the Weyl nodes, the expression for the Hall conductance can be written as [6]:

$$\sigma_{ab} = \epsilon_{abc}\frac{e^2}{2\pi h}q^c \tag{4}$$

assuming $\boldsymbol{q} = 2k_0\hat{z}$, the above expression reduces to $\sigma_{xy} = \frac{e^2}{2\pi h}2k_0$ where $2k_0$ is the separation between the two Weyl nodes of opposite chiralities. Microscopically, the anomalous Hall conductance is directly related to the Berry curvature as introduced earlier and can be written as [27, 28]:

$$\sigma_{ab} = \frac{e^2}{\hbar}\int\frac{d\boldsymbol{k}}{(2\pi)^3}\Omega_c(\boldsymbol{k})f(\boldsymbol{k}) \tag{5}$$

where $f(\mathbf{k})$ is the Fermi distribution function, $\Omega_c(\boldsymbol{k})$ is the c-th component of the Berry curvature for the wave vector $\mathbf{k}$.

The chiral magnetic anomaly (CME) and anomalous Hall effect (AHE) can be further understood by the axion electrodynamics of the Weyl fermions. The axion term to the electromagnetic Lagrangian reads as [6, 9]:

$$S_\theta = \frac{1}{2\pi}\frac{e^2}{h}\int dtd\boldsymbol{r}\theta(\boldsymbol{r},t)\mathbf{E}\cdot\mathbf{B}$$

where θ term is given by

$$\theta(\boldsymbol{r},t) = 2(\mathbf{k_0}\cdot\mathbf{r}-b_0 t)$$

where $\mathbf{k_0}$ is the position of the Weyl nodes and $2b_0$ is the chemical potential difference between the Weyl nodes. Minimizing the action in standard fashion leads to the following equations of motion for charge and current densities:

$$\rho = \frac{1}{2\pi}\frac{e^2}{h}2\mathbf{k_0}\cdot\mathbf{B} \tag{6}$$

and

$$\mathbf{J} = \frac{1}{2\pi}\frac{e^2}{h}(2\mathbf{k_0}\times\mathbf{E} - 2b_0\mathbf{B}) \tag{7}$$

Equation 6 and the first term in Eq. 7 represent the anomalous Hall effect that is expected to occur in magnetic Weyl semimetal with broken time reversal symmetry and is equivalent to the Hall conductance in Eq. 4. The second term in Eq. 7 is equivalent to the chiral magnetic effect as discussed in previous section where the energy difference $\Delta\epsilon = 2\hbar b_0$.

Anomalous Hall effect has been observed in various Weyl semimetals including Mn_3Sn, Mn_3Ge and Heusler compounds [22, 29–31]. As a representative example, we show the anomalous Hall effect in Mn_3Sn. Mn_3Sn orders magnetically at $T_N = 420$ K in a non-collinear magnetic structure as shown in the Fig. 6(b) [32]. In addition to the non-collinear antiferromgnetism, there is a small ferromagnetic component along the c direction (0.0002 μ_B per Mn atom). The Hall conductivity, $\sigma_H = -\rho_H/\rho^2$, for in-plane fields along both $[2\,\bar{1}\,\bar{1}\,0]$ and $[0\,1\,\bar{1}\,0]$ shows a large
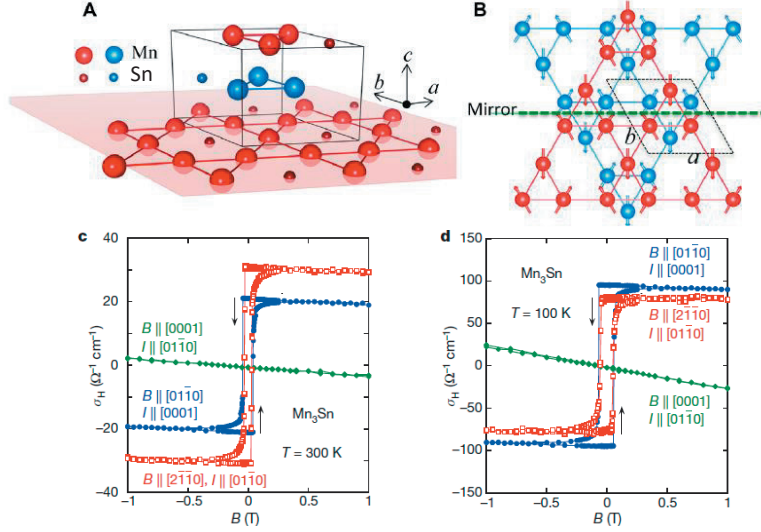
Fig. 6: (a, b) Crystal and magnetic structured of Mn_3Sn. The Hall conductivity σ_H versus B measured in B $\parallel$ $[2\,\bar{1}\,\bar{1}\,0]$ and $[0\,1\,\bar{1}\,0]$ and $[0\,0\,0\,1]$ obtained at 300 K (c) and 100 K (d). Adapted from Ref. [29, 30].

jump and narrow hysteresis (Fig. 6 c, d). For instance, with B $\parallel$ $[0\,1\,\bar{1}\,0]$, σ_H has large values near zero field, $\sim 20\ \Omega^{-1}cm^{-1}$ at 300 K and nearly $100\ \Omega^{-1}cm^{-1}$ at 100 K. This is again quite large for an AFM metal and comparable to those values found in ferromagnetic metals [26, 33]. On the other hand, the Hall conductivity for B $\parallel$ $[0\,0\,0\,1]$ (c axis) shows no hysteresis but only a linear field dependence. In the case of in-plane magnetic fields, the inverse triangular magnetic order breaks the in-plane hexagonal symmetry of the lattice, and thus may induce an AHE in the a–b plane. Indeed Kübler *et al.* [27] have theoretically found a large AHE in Mn_3Sn, calculating the anomalous Hall conductivity through the Brillouin zone integration of the Berry curvature according to Eq. 5. The authors interpreted that the non-vanishing Berry curvature arises from the chiral spin structure. Switching of the Hall conductivity may be related to the switching of the staggered moment direction of the triangular spin structure [29] as result of switching small ferromagnetic component.

3   Outlook

Weyl semimetals are fascinating materials to investigate both theoretically and experimentally. It connects microscopic quantum phenomena to macroscopic properties in a non-trivial manner. The chiral magnetic effect (CME) and anomalous Hall effect (AHE) can be observable at room temperature which might paves the way for the future spintronics and quantum comput-

ing. Combining with extremely large magnetoresistance and quantum effects of spins, Weyl semimetals are the candidates for future electronic devices known as "Weyltronics". Furthermore, Topological character of the Weyl semimetals implies that the exotic properties are robust against external perturbations - making them ideal candidate for practical applications. Due to massless quasiparticle excitations, Weyl electrons are highly mobile and could lead to much faster and low-power consumer electronic devices. Correlation and topology make the Weyl semimetals 'systems with infinite possibilities'.

References

[1] Y. Tokura, M. Kawasaki, N. Nagaosa, Nat. Phys. **13**, 1056 (2017).

[2] T. Kajita, Rev. Mod. Phys. **88**, 030501 (2016).

[3] A. B. McDonald, Rev. Mod. Phys. **88**, 030502 (2016).

[4] X. Wan, A. M. Turner, A. Vishwanath, S. Y. Savrasov, Phys. Rev. B **83**, 205101 (2011).

[5] C. Herring, Phys. Rev. **52**, 365 (1937).

[6] N. P. Armitage, E. J. Mele, A. Vishwanath, Rev. Mod. Phys. **90**, 015001 (2018).

[7] B. A. Bernevig, Nat. Phys. **11**, 698 (2015).

[8] H. Weyl, Zeitschrift für Physik **56**, 330 (1929).

[9] A. A. Burkov, Nat. Mater. **15**, 1145 (2016).

[10] B. Q. Lv, H. M. Weng, B. B. Fu, X. P. Wang, H. Miao, J. Ma, P. Richard, X. C. Huang, L. X. Zhao, G. F. Chen, Z. Fang, X. Dai, T. Qian, H. Ding, Phys. Rev. X **5**, 031013 (2015).

[11] S.-Y. Xu, I. Belopolski, N. Alidoust, M. Neupane, G. Bian, C. Zhang, R. Sankar, G. Chang, Z. Yuan, C.-C. Lee, S.-M. Huang, H. Zheng, J. Ma, D. S. Sanchez, B. Wang, A. Bansil, F. Chou, P. P. Shibayev, H. Lin, S. Jia, M. Z. Hasan, Science **349**, 613 (2015).

[12] S.-M. Huang, S.-Y. Xu, I. Belopolski, C.-C. Lee, G. Chang, B. Wang, N. Alidoust, G. Bian, M. Neupane, C. Zhang, S. Jia, A. Bansil, H. Lin, M. Z. Hasan, Nat. Commun. **6**, 7373 (2015).

[13] X. Huang, L. Zhao, Y. Long, P. Wang, D. Chen, Z. Yang, H. Liang, M. Xue, H. Weng, Z. Fang, X. Dai, G. Chen, Phys. Rev. X **5**, 011029 (2015).

[14] S.-Y. Xu, I. Belopolski, N. Alidoust, M. Neupane, G. Bian, C. Zhang, R. Sankar, G. Chang, Z. Yuan, C.-C. Lee, S.-M. Huang, H. Zheng, J. Ma, D. S. Sanchez, B. Wang, A. Bansil, F. Chou, P. P. Shibayev, H. Lin, S. Jia, M. Z. Hasan, Science **349**, 622 (2015).

[15] H. Nielsen, M. Ninomiya, Phys. Lett. B **130**, 389 (1983).

[16] A. A. Zyuzin, A. A. Burkov, Phys. Rev. B **86**, 115133 (2012).

[17] A. A. Burkov, Phys. Rev. B **91**, 245157 (2015).

[18] Z. Jian-Hui, J. Hua, N. Qian, S. Jun-Ren, Chin. Phys. Lett. **30**, 027101 (2013).

[19] X. Huang, L. Zhao, Y. Long, P. Wang, D. Chen, Z. Yang, H. Liang, M. Xue, H. Weng, Z. Fang, X. Dai, G. Chen, Phys. Rev. X **5**, 031023 (2015).

[20] F. Arnold, C. Shekhar, S.-C. Wu, Y. Sun, R. D. dos Reis, N. Kumar, M. Naumann, M. O. Ajeesh, M. Schmidt, A. G. Grushin, J. H. Bardarson, M. Baenitz, D. Sokolov, H. Borrmann, M. Nicklas, C. Felser, E. Hassinger, B. Yan, Nat. Commun. **7**, 11615 (2016).

[21] Y. Li, Z. Wang, P. Li, X. Yang, Z. Shen, F. Sheng, X. Li, Y. Lu, Y. Zheng, Z.-A. Xu, Frontiers of Physics **12**, 127205 (2017).

[22] M. Hirschberger, S. Kushwaha, HijunWang, Q. Gibson, S. Liang, C. A. Belvin, B. A. Bernevig, R. J. Cava, N. P. Ong, Nat. Mater. **15**, 1161 (2016).

[23] C.-L. Zhang, S.-Y. Xu, I. Belopolski, Z. Yuan, Z. Lin, B. Tong, G. Bian, N. Alidoust, C.-C. Lee, S.-M. Huang, T.-R. Chang, G. Chang, C.-H. Hsu, H.-T. Jeng, M. Neupane, D. S. Sanchez, H. Zheng, J. Wang, H. Lin, C. Zhang, H.-Z. Lu, S.-Q. Shen, T. Neupert, M. Z. Hasan, S. Jia, Nat. Commun. **7**, 10735 (2016).

[24] H.-Z. Lu, S.-Q. Shen, Phys. Rev. B **92**, 035203 (2015).

[25] A. A. Burkov, J. Phys. Condens. Matter **27**, 113201 (2015).

[26] N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, N. P. Ong, Rev. Mod. Phys. **82**, 1539 (2010).

[27] J. Kübler, C. Felser, Euro Phys. Lett. **108**, 67001 (2014).

[28] H. Chen, Q. Niu, A. H. MacDonald, Phys. Rev. Lett. **112**, 017205 (2014).

[29] S. Nakatsuji, N. Kiyohara, T. Higo, Nature **527**, 212 (2015).

[30] A. K. Nayak, J. E. Fischer, Y. Sun, B. Yan, J. Karel, A. C. Komarek, C. Shekhar, N. Kumar, W. Schnelle, J. Kübler, C. Felser, S. S. P. Parkin, Science Advances **2**, e1501870 (2016).

[31] C. Shekhar, N. Kumar, V. Grinenko, S. Singh, R. Sarkar, H. Luetkens, S.-C. Wu, Y. Zhang, A. C. Komarek, E. Kampert, Y. Skourski, J. Wosnitza, W. Schnelle, A. McCollam, U. Zeitler, J. Kübler, B. Yan, H.-H. Klauss, S. S. P. Parkin, C. Felser, P. Natl. Acad. Sci. USA **115**, 9140 (2018).

[32] S. Tomiyoshi, Y. Yamaguchi, J. Phys. Soc. Jpn. **51**, 247 (1982).

[33] T. Miyasato, N. Abe, T. Fujii, A. Asamitsu, S. Onoda, Y. Onose, N. Nagaosa, Y. Tokura, Phys. Rev. Lett. **99**, 086602 (2007).

E 4   Thin Film Heterostructures

Anirban Sarkar

Quantum Materials and Collective Phenomena
Jülich Centre for Neutron Science (JCNS)
Forschungszentrum Jülich GmbH

Contents

1   Introduction

Two-dimensional materials created *ab-initio* by the process of condensation of atoms, molecules, or ions, called thin films, have unique properties significantly different from the corresponding bulk materials as a result of their physical dimensions, geometry, non-equilibrium microstructure, and metallurgy. '*Thin*' is a relative term, but most deposition techniques control layer thickness from fractions of a nanometer (monolayer) to several micrometers. Thin film heterostructures are composed of layers of different materials and thicknesses in the nanometer range. The heterostructures can be formed by either depositing a single film on a solid substrate viz. a film/substrate bi-layer, or making multilayers of same/different materials stacked over the substrate, as shown in Fig. 1.

Reduced dimensions and the presence of interfaces, between different materials, lead to very interesting phenomena. Further, the characteristic features of thin film heterostructures can be drastically modified and tailored to obtain the desired and required physical characteristics. It provides a unique ability to combine materials with different or even antagonistic ground states (func-

Fig. 1: Schematic showing thickness range of thin film (top) and various thin film heterostructures (bottom) where different color represents different materials.

tionalities) – e.g., metal/insulator-oxide/semi-conductor, ferromagnetic/ferroelectric [1], ferromagnet/superconductor [2] heterostructures etc. Physical effects like the interlayer exchange coupling [3], giant magnetoresistance [4, 5], tunnel magnetoresistance [6, 7] or the current induced magnetization switching [8, 9] are based on the fact that not only the charge but also the spin of the electron can be used for information storage and processing [10]. New effects might be expected if the size of element or period of the arrangement is smaller than some characteristic length scale of the material, like the wavelength of electrons at the Fermi level, the spin diffusion length of the electrons, or the domain wall width [11, 12].

These features form the basis for development of a host of extraordinary thin film device applications in the last two decades. On the one extreme, these applications are in the sub-micron dimensions in such areas as very large scale integration (VLSI), Josephson junction quantum interference devices, magnetic bubbles, and integrated optics. On the other extreme, large-area thin films are being used as selective coatings for solar thermal conversion, solar cells for photovoltaic conversion, and protection and passivating layers.

In this chapter, we begin with few examples of most commercial and scientifically relevant thin film heterostructures which holds great promises for potential application in electronic devices. Finally, we conclude with a brief discussion on the relevance of scattering methods in elaborating such heterostructures.

2   Examples of Thin Film Heterostructures

2.1   Semiconductor heterostructures

Thin film heterostructures are extensively used in the semiconductor industry. An important component in any electronic device is the semiconductor transistor. **Heterojunction Bipolar Transistor** (HBT) is a type of bipolar junction transistor (BJT) which uses differing semiconductor materials for the emitter and base regions, creating a heterojunction [13]. The difference lies in the fact that a normal BJT is fabricated by doping different areas of a single semiconductor wafer. Fig. 2 shows a cross-sectional representation of a AlGaAs/GaAs HBT [14] where AlGaAs acts as the emitter and heavily doped p$^+$-GaAs layer acts as the transistor base. HBTs are used for digital and analog microwave applications with frequencies up to several hundred GHz e. g. power amplifiers in mobile telephones and laser drivers.

Another example of a most extensively used heterostructure is the **Metal-Oxide-Semiconductor** (MOS) structure [Fig. 3(a)]. It constitutes the main component of any MOS-field-effect transistors (MOSFETs) and is present in all integrated circuits (IC's). The traditional MOS-structure is obtained by growing a layer of silicon dioxide (SiO$_2$) on top of a silicon substrate and depositing a layer of metal or polycrystalline silicon (the latter is commonly used). As the SiO$_2$ is a dielectric material, its structure is equivalent to a planar capacitor, with one of the electrodes replaced by a semiconductor.
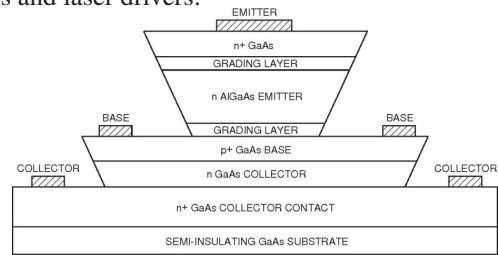
Fig. 2: Cross-sectional representation of the AlGaAs/GaAs Heterojunction Bipolar Transistor [14].
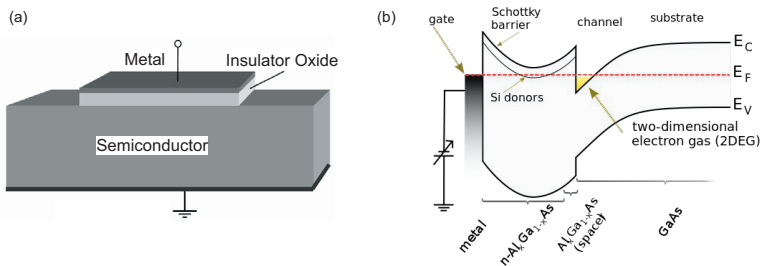
Fig. 3: (a) Basic MOS structure, and (b) band diagram of GaAs/AlGaAs heterojunction-based HEMT, at equilibrium. (image source: *wikipedia*)

A **High-electron-mobility transistor** (HEMT) [15], also known as heterostructure-FET (HFET) or modulation-doped FET (MODFET), is a field-effect transistor incorporating a junction between two materials with different band gaps (i. e., a heterojunction) as the channel instead of a doped region (as is generally the case for MOSFET). Carrier diffusion across the junction leads to accumulation of electrons inside the narrow bandgap material forming a 2 dimensional electron-gas (2DEG) resulting to the high current in the device. Fig. 3(b) shows the band diagram of GaAs/AlGaAs heterojunction-based HEMT, at equilibrium.

Applications involve – microwave and millimeter wave communications, imaging, radar, and radio astronomy – any application where high gain and low noise at high frequencies are required. Commercially available HEMT-based devices can be discrete transistors but are more usually in the form of a 'monolithic microwave integrated circuit' (MMIC). HEMTs are found in equipment ranging from cellphones and Direct Broadcasting Satellite (DBS) receivers to radar and radio astronomy. Furthermore, GaN-HEMTs on silicon substrates are used as power switching transistors for voltage converter applications because of low on-state resistances, and low switching losses due to the wide bandgap properties.

Last but not the least, are the ***Thin Film Solar Cells***. They are second generation solar cells fabricated by depositing one or more thin layers, or thin film (TF) of photovoltaic material on a substrate, such as glass, plastic or metal. The performance of a solar cell is determined by it's efficiency i.e., the portion of energy in the form of sunlight that can be converted via photovoltaics into electricity. In a recent consortium of Fraunhofer ISE, Soitec, CEA-Leti and Helmholtz Center Berlin investigated a four-junction solar cell using wafer bonding technology for terrestrial concentrator applications [16]. A new record efficiency of 44.7% (AM1.5d, 297 suns) has been achieved (see Fig. 4) which lies well above the theoretically predicted limit of 33% for normal bulk

Fig. 4: I-V-curve and performance data of four-junction concentrator GaInP/GaAs//GaInAsP/GaInAs solar cell. The measurement was conducted at the CalLab of Fraunhofer ISE. [16]

semiconductor based solar cells. Another defining term in the overall behavior of a solar cell is the fill factor (FF). This factor is a measure of quality of a solar cell. This is the available power at the maximum power point (P_m) divided by the open circuit voltage (V_{OC}) and the short circuit current (I_{SC}). From the application point of view, thin film solar cells are used to build photovoltaic module for generation of electricity from solar energy for commercial and residential use. Moreover, the main application lies in powering space vehicles such as satellites and telescopes (e.g., Hubble). They provide a very economical and reliable way of powering objects which would otherwise need expensive and cumbersome fuel sources.

2.2   Magnetoresistive element heterostructures

So far we have discussed about some of the thin film heterostructures where the device works by manipulation of the electronic charge. Magnetic thin film heterostructures however provides another intrinsic property of the electrons to be manipulated, i.e., the spin degrees of freedom [10, 17]. They are often used in magnetic sensors, actuators, and data recording/ storage. It is well known that the magnetic field can affect the transport of carriers in a material, which basically manifests the magnetoresistance (MR) of that material, i.e., the change in electrical resistance of the material due to application of magnetic field. Heterostructures built with such magnetic thin films would therefore provide an unique way to manipulate carrier transport by applying magnetic fields externally.

In 1936, F. Mott proposed a concept of spin dependent conduction and the possibility of considering the two electron spin channels as separate non-interacting conduction ways within a ferromagnetic (FM) metal [18]. This, also known as the two current model, was extensively used for explaining various magnetoresistive phenomena in the materials [19, 20]. However, Jullière was first to measure a large change in the magnetoresistance (∼14% at 4.2 K) of a **Magnetic Tunnel Junction (MTJ)** formed by Fe/GeO/Co (ferromagnet/insulator/ferromagnet - FIF) heterostructure [6]. According to Julliére's model – if the magnetizations of the ferromagnetic film (F) are in a parallel orientation it is more likely that electrons will tunnel through the insulating film than if they are in the opposite (antiparallel) orientation. Consequently, such a junction can be switched between two states of electrical resistance, one with low resistance ($R_{\uparrow\uparrow}$) and one with very high resistance ($R_{\uparrow\downarrow}$). The corresponding tunneling magnetoresistance (TMR) in an F/I/F is defined as, $TMR = \Delta R/R_M = (R_{\uparrow\downarrow} - R_{\uparrow\uparrow})/R_{\uparrow\uparrow}$. Defining spin polarization, P_i of the ferromagnetic layer (F_i) as $P_i = \frac{N_{\uparrow i} - N_{\downarrow i}}{N_{\uparrow i} + N_{\downarrow i}}$ where N's are the spin resolved DOS for majority ($\uparrow$) and minority ($\downarrow$) spins in F_i, the TMR can also be expressed as, $TMR = 2P_1 P_2/(1 - P_1 P_2)$. Replacing the insulating film with oxide tunnel barrier higher TMR (∼12%) could be achieved at room temperature [7]. Presently, the read head of the hard disk drive uses a MTJ of CoFeB/MgO/CoFeB heterostructure which shows a 600% change at room temperature [21].

Going back to Jullière's model, it directly predicts the **Spin Valve** effect [22], i.e., the resistance of such a device can be changed by a manipulation of the relative orientation of the magnetizations M1 and M2 in the ferromagnetic layers F1 and F2, respectively (from Fig. 5). Such an orientation can be sustained even in the absence of an external power. In the simplest case, a spin valve consists of a non-magnetic material sandwiched between two ferromag-

Fig. 5: Schematic illustration of F/I/F tunnel junctions.

nets, one of which is fixed (pinned) by an antiferromagnet which acts to raise its magnetic coercivity and behaves as a "hard" layer (e.g., CoFe), while the other is free (unpinned) and behaves as a "soft" layer (e.g., NiFe), see Fig. 6 [22]. Due to the difference in coercivity, the soft layer changes polarity at lower applied magnetic field strength than the hard one. Upon application of a magnetic field of appropriate strength, the soft layer switches polarity, producing two distinct states: a parallel, low-resistance state, and an anti-parallel, high-resistance state.

In such a system if a current (un-polarized) is passing into a ferromagnet whose majority spin is spin up, for example, then electrons with spin up will pass through relatively unhindered, while electrons with spin down will either 'reflect' or spin flip scatter to spin up upon encountering the ferromagnet to find an empty energy state in the new material. Thus if both the fixed and free layers are polarized in the same direction, the device has relatively low electrical resistance, whereas if the applied magnetic field is reversed and the free layer's polarity also reverses, then the device has a higher resistance due to the extra energy required for spin flip scattering. Fig. 7 shows a band diagram explaining the spin-valve effect.
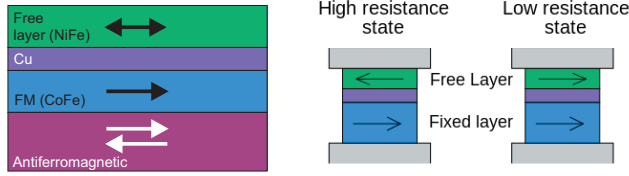
Fig. 6: Sketch of a simple spin valve structure with CoFe as the magnetic "hard" layer and NiFe as the free "soft" layer [22]. The anti-parallel (high-resistance) and parallel (low-resistance) state of the spin valve is represented on right.

An antiferromagnetic layer is required to pin one of the ferromagnetic layers (i.e., make it fixed or magnetically hard). This results from a large negative exchange coupling energy between a ferromagnet and an antiferromagnet in contact. The non-magnetic layer is required to decouple the two ferromagnetic layers so that at least one of them remains free (magnetically soft). The basic operating principles of a pseudo spin valve are identical to that of an ordinary spin valve, but instead of changing the magnetic coercivity of the different ferromagnetic layers by pinning one with an antiferromagnetic layer, the two

Fig. 7: Band diagram for explaining the spin valve effect.

layers are made of different ferromagnets with different coercivities e.g., NiFe and Co. Note that coercivities are largely an extrinsic property of materials and thus determined by processing conditions.

The difference between *TMR* and *Spin Valve* is that the former works on the principle of spin-dependent tunneling while the later on spin-dependent scattering processes. Nevertheless, this concept of the spin valve effect eventually led to the discovery of the ***Giant Magnetoresistance (GMR)*** effect [4, 5] promising application in non-volatile memory devices. Large change in the magnetoresistance value was obtained in antiferromagnetically coupled multilayers of Fe/Cr, where thin layers of magnetic metals (Fe) are separated by layers of non-magnetic metals (Cr). The magnetic layers are coupled through the non-magnetic layers in either a ferromagnetic or antiferromagnetic configuration depending upon the thickness of the non-magnetic layer. Although the details of the coupling mechanism are not known, RKKY-like mechanism predicts result which are similar to the experimental observations. Magnetoresistance effects up to $\sim$50% were observed at low temperatures. This effect was subsequently found to occur in a number of multilayer magnetic systems. Fig. 8(a) shows the change in MR ($\sim$1.5%) obtained for the Fe/Cr/Fe tri-layer heterostructure at RT by the group of P. Grünberg [5] and Fig. 8(b) shows that for the multilayer structure at 4.2 K ($\sim$80%) by the group led by A. Fert [4], for which they were awarded the Nobel Prize in Physics (2007).

From the application point of view both TMR and GMR devices are often used as magnetic field sensors and actuators. However, a greater impact was made on the development of modern day technology and data storage. Coding the high-resistive state of the device as '1' while the low-
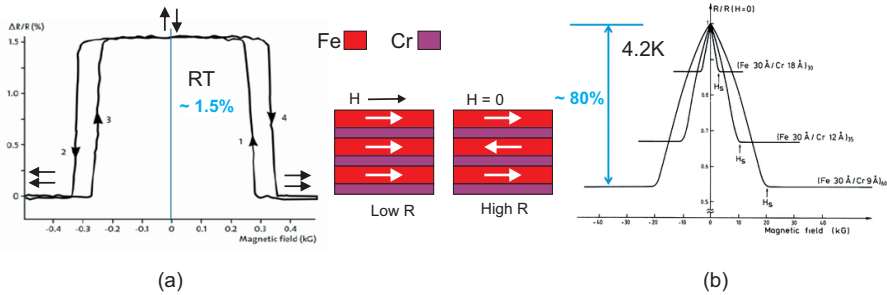
Fig. 8: Original findings about GMR that resulted in the Nobel prize for (a) P. A. Grünberg (Fe/Cr/Fe tri-layer) [5] and for (b) A. Fert (multilayered structure) [4]

resistive state as '0' a binary state could be realized and therefore can be used in computation and machine language processing. Using this idea, in 1997 IBM launched 16GB hard disk drive (HDD) where they used the GMR structure as the magnetic read head. By optimizing the heterostructure material and with the advent of modern day technology the storage capacity in modern HDD has gone through a giant leap, however the underlying physics remains the same.

Fig. 9 shows a schematic of the read and write technique used in recording and reading data stored magnetically in a HDD. The read element can be used both in current-in-plane (CIP) as well as current-perpendicular to the plane (CPP) geometry. The platter of the HDD is made of a magnetic thin film with well defined domain structure. The information here is stored in a
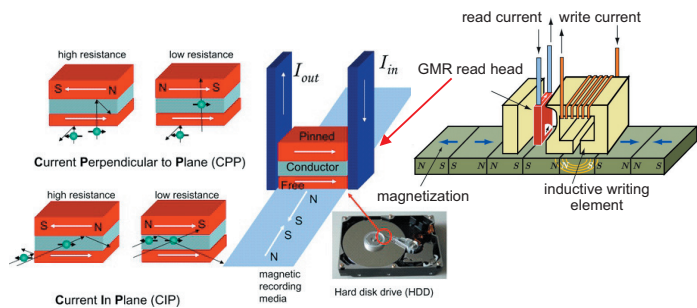
Fig. 9: GMR read head used in magnetic hard disk drives, working both in current-in-plane (CIP) and current-perpendicular to the plane (CPP) geometry.

bit (small area of hard disk) fixing the direction of collective spins (equivalent to magnetization) in the bit. While the writing is done by induction using an electromagnet, the read-out is done using a GMR head. The GMR head has a pinned magnetic layer and the resistance of the GMR element either high '1' or low '0' can be read out as bits written on the magnetic platter. On a similar characteristics note TMR element found its way for the application in magnetoresistive random access memory (MRAM) and spin-transfer-torque MRAM (STT-MRAM) with Everspin launching in 2016 the very first 256MB STT-MRAM.

2.3 Heterostructures in Research

While on one hand large number of research have been dedicated to the development of magnetoresistive elements which have proven to be commercially successful, on the other hand a number of research have been dedicated for fabricating pure spin device, the idea of which was

fueled by the theoretical model of Datta-Das Spin-FET [23], shown in Fig. 10. Such a device would depend on the transfer of spin, i.e., by manipulation of the angular momentum and involve no charge transport [24]. Therefore such devices would result in faster processing, low power consumption, low heat dissipation and can be used in non-volatile memory.
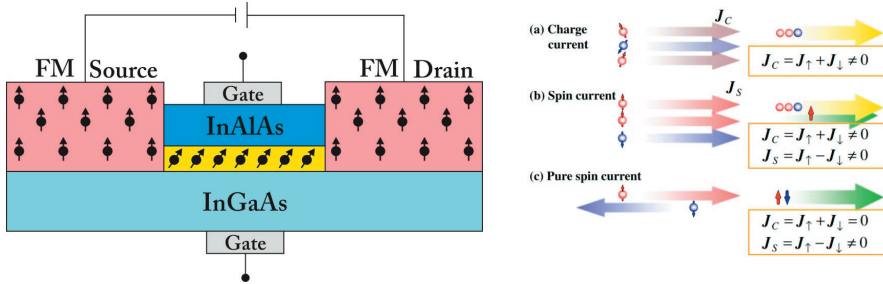
Fig. 10: Schematic representation of Datta-Das Spin-FET model (left) [23] and spin-current concept (right) where J_C and J_S is the charge and spin current density, respectively. $J_\uparrow$ and $J_\downarrow$ represents the current density due to up-spin ($\uparrow$) and down-spin ($\downarrow$), respectively.

One of the main requirement of such a device to integrate into a electronic device would be to actively convert charge current to spin-polarized current and *vice versa*. Therefore, research have been dedicated to actively manipulate the spin in the material not only via application of an external magnetic field but also using electric fields.

Since a ferromagnetic (FM) or a half-metallic (HM) material is spin-polarized (has unequal number of up and down spin due to exchange splitting), they have been extensively used in forming heterostructures and understanding their interface/ proximity effects. Some of the most studied heterostructures and their respective properties in this field are listed as follows:

1. FM/Supersonductors(S) : interface/proximity effects [25, 26], nucleation of superconductivity [27] and flux pinning [28].
2. FM/Semiconductor(SC) : efficient spin injection and manipulation within semiconductors [29].
3. FM/Antiferromagnet(AFM) : studies on interface and exchange bias effect [30].
4. FM/Heavy-metal(H) : studies on spin pumpung [31] and magnetic Skyrmions [32, 33]
5. FM/Ferroelectric(FE) or FM/Piezoelectric(PE) : artificial multiferroics [1] and voltage control of magnetism [34].

Apart from FM or HM materials, transition metal oxide (TMOs) constitute probably one of the most interesting classes of solids, exhibiting a variety of structures and properties. The phenomenal range of electronic and magnetic properties exhibited by transition metal oxides is especially noteworthy [35, 36]. One of the most important TMOs heterostructure which has been studied intensively in the past decade is the $LaAlO_3/SrTiO_3$ (LAO/STO) system [37]. The interface between lanthanum aluminate ($LaAlO_3$) and strontium titanate ($SrTiO_3$) is a notable materials interface because it exhibits properties not found in its constituent materials. Individually, $LaAlO_3$ and $SrTiO_3$ are non-magnetic insulators, yet $LaAlO_3/SrTiO_3$ interfaces can exhibit electrical conductivity [38], superconductivity [39], ferromagnetism [40], large negative in-plane magnetoresistance [41], and giant persistent photoconductivity [42].

3   Characterization using Scattering techniques

In this section we show you the power of scattering techniques in characterizing the thin film heterostructures with some examples.

3.1   *in-situ* Electron Diffration Techniques

Reflection High-Energy Electron Diffraction (RHEED) and ***Low-Energy Electron Diffraction (LEED)*** are two of the well known electron diffraction techniques used for *in-situ* monitoring of the thin film growth, structure and surface ordering of crystalline thin films.
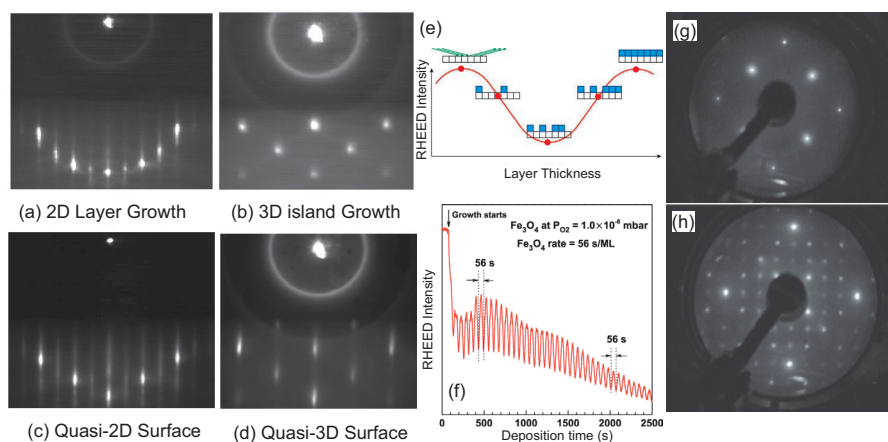
Fig. 11: RHEED patters of thin film surface representing growth of – (a) homogeneous 2D (dimensional) layer growth, (b) 3D island growth, (c) Quasi-2D growth, and (d) Quasi-3D growth often associated with rough surface. RHEED intensity oscillations (e) typically associated with the surface coverage of the growing film, and (f) 40 nm Fe_3O_4 thin film grown on MgO(001) substrate [43]. LEED pattern taken from [43] for (g) MgO(001) substrate and (h) after the growth of Fe_3O_4 thin film on MgO(001).

In RHEED, beam of electrons are accelerated with a high energy of 5-20 keV ($\lambda \approx 0.17$-0.06 Å) and are incident on the sample surface at a glancing angle, 1-3°. The elastically diffracted electron beam from the sample surface interfere constructively at specific angles and form regular patterns on the detector/ screen. Structural changes on the surface during growth are generally monitored using RHEED patterns. Fig. 11(a)-(d) shows the variation of the RHEED pattern from a uniform 2-dimensional (2D) layer growth to 3D island growths. The intensities of individual spots on the RHEED pattern fluctuate in a periodic manner as a result of the relative surface coverage of the growing thin film, see Fig. 11(e). An example of a 40 nm Fe_3O_4 thin film grown on MgO(001) is shown in Fig. 11(f) where each full period corresponds to formation of a single atomic layer thin film [43]. The oscillation period is highly dependent on the material system, electron energy and incident angle, so researchers obtain empirical data to correlate the intensity oscillations and film coverage before using RHEED for monitoring film growth.

LEED on the other hand images surface by bombardment with collimated low energy electron

beam of 20-200 eV. Contrary to RHEED the incident electron beam is projected perpendicular to the surface. Informations of the sample surface display as spots of diffracted electrons on a fluorescent screen. Fig. 11 (g) and (h) shows a LEED pattern from pure MgO(001) substrate and a corresponding 40 nm Fe_3O_4 thin film grown on it, respectively [43]. It must be noted that both techniques (RHEED and LEED) are sensitive to the roughness of the sample surface.

3.2 Orbital ordering in transition metal oxide heterostructures

In this section we will discuss about the article by Benckiser et al. [44] on oxide heterostructure consisting of a superlattice of $LaNiO_3$/$LaAlO_3$ (LNO/LAO) on a $SrTiO_3$ (STO) substrate. In their work they combines scattering and X-ray absorption spectroscopy to gain comprehensive insight into the sample properties. Within the framework of this lecture only parts of the paper are discussed.

Transition-metal atom in a crystal with perovskite structure (ABO_3) is surrounded by six oxygen ions, O^{2-}, which induces crystal field splitting of the d orbitals of the transition metal ion [35]. Wave functions pointing toward O^{2-} ions have higher energy in comparison with those pointing between them. Hence, there will be a lower lying triply degenerate t_{2g} level and a higher lying doubly degenerate e_g level. The five d orbitals are depicted in Fig. 12.

$LaNiO_3$ (LNO) has a configuration of $La^{3+}Ni^{3+}O_3^{2-}$ which means that the Ni^{3+} ($3d^7$, $S = 1/2$) has 7 electrons in the valance shell. In bulk LNO, the Ni^{3+} forms a paramagnetic metallic state [45] with no orbital preference [46] leading to a fully occupied three t_{2g} level with six electrons and a doubly degenerate e_g levels with the remaining one electron. Theoretically, it was suggested that strain and confinement could break this

Fig. 12: Five d-orbitals. In the cubic crystal field the fivefold degeneracy is lifted to two higher-lying e_g orbitals and three lower-lying t_{2g} orbitals [35].

symmetry in $LaNiO_3$/$LaAlO_3$ superlattices and thereby allow manipulation of the orbital occupancy [47]. $LaAlO_3$:$La^{3+}Al^{3+}O_3^{2-}$ (LAO) is a band insulator which is used for the confinement of electronic states in the heterostructure. Model calculations suggest, that the degeneracy of the Ni bulk e_g orbitals in LNO is lifted and that the $x^2 - y^2$ level is preferred [44] when the LNO/LAO heterostructure is exposed to a tensile strain, as for epitaxial growth on a STO substrate. This can be understood in terms that tensile strain increases the distance to the in-plane (xy-plane) oxygen ions while consequently decreasing its distance in the vertical direction (i.e., $3z^2 - r^2$ orbital). To summarize, Benckiser et al. [44] evidence this prediction experimentally.

A [4 u.c. LAO/ 4 u.c. LNO]$_8$ heterostructure was grown on STO(001) substrate by pulsed-laser deposition, where u.c. stands for unit cell. Fig. 13(a) shows the high-resolution x-ray

Fig. 13: a. High-resolution $\theta/2\theta$ scan along the specular truncation rod in vicinity of the STO(001) Bragg peak. The x-axis scale is presented in reciprocal lattice units of the STO substrate (a = 3.905 Å). b. Reciprocal space map around the STO(103) substrate Bragg peak of the same specimen. c. Atomic Force Microscopy image of STO substrate surface (2×2 μm area). Adopted from ref. 44.

diffraction pattern measured with Cu-Kα radiation around the STO(001) Bragg peak of the heterostructure. Symmetrically around the LNO-LAO(001) peak superlattice satellites peaks and thickness fringes are observed. From the period of the thickness fringes one could obtain a total thickness 247±8 Å of the heterostructure. Fig. 13(b) shows the reciprocal space mapping around the (103) Bragg peak. The layer peak position relative to the relaxation line (to origin; orange) and the fully strained position (vertical gray line) indicate that the investigated superlattice is partially, but not fully relaxed. From this measurement in- and out-of-plane lattice constants of a = 3.853(2) Å and c = 3.780(1) Å (averaged for LNO and LAO) were obtained.

X-ray reflectivity (XRR) measurements were performed at room temperature in specular geometry, that is, the momentum transfer q was parallel to the surface normal z. In Fig. 14(a), the resulting data for a hard X-ray energy of 8047 eV, which is far from resonance, is shown. The data reflect the high quality of the investigated superstructure. At energies corresponding to the Ni $L_{2,3}$ absorption edges, i. e. 854.7 eV and 872.2 eV, the reflectivity was obtained utilizing linearly σ and π polarized light of the incident X-rays (see the sketch in Fig. 14(a)), Fig. 14(b) and (c). Clearly, a dependence of the reflectivities on the polarization of the light can be seen which shows that the X-ray absorption is polarization dependent. Superlattice peaks up to the third order, denoted by SL(00l), l = 1, 2, 3 can be observed. The layer thickness and interface roughness of the superlattice (see the table below in Fig. 14) were obtained by the authors utilizing a fit program based on the Parrat formalism [48]. The solid black line in the reflectivity plot Fig. 14(a)-(c) is the best fit to the data using this algorithm. The estimated error bars of the individual thickness and roughness are approximately ±0.5Å. The resulting profiles are shown in Fig. 14(a)-(c) and the structural parameters are given in the table below the figures. It turns out that different interface roughness and thickness has to be considered for the first LNO layer on STO (acting as buffer layer) and the last LAO layer stack (acting as cap layer) for the best fit.

To gain insight into the orbital occupation X-ray absorption spectra (XAS) were measured in total-electron yield (TEY) and fluorescence-yield (FY) modes, Fig. 14(d). The TEY is interface sensitive while the FY is bulk sensitive. As both measurement modes reveal nearly identical results, only the fluorescence-yield mode is discussed in the following. The Ni L_2 and L_3 absorption edges are located around 872 eV and 855 eV, respectively. First, the white line of the Ni L_3 absorption edge has to be separated from the strong white line of La M_4 by subtracting the latter in form of an approximated Lorentzian profile. The result is shown in Fig. 14(e).
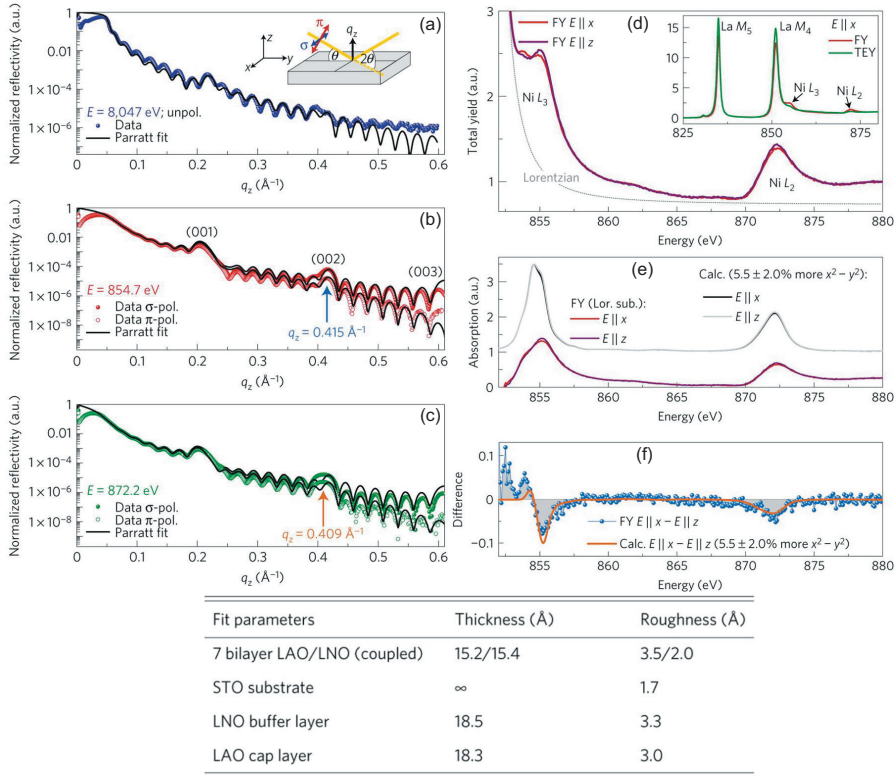
Fig. 14: XRR of [4 u.c LAO/ 4 u.c. LNO]$_8$ superlattice for – (a) E = 8047 eV (Cu-Kα), (b) E = 854.7 eV (Ni L$_3$) and (c) E = 872.2 eV (Ni L$_2$). All data have been normalized to unity at q_z=0. (d)-(f) Polarization-dependent XAS spectrum across the Ni $L_{2,3}$ for $E \parallel x$ (in-plane) and $E \parallel z$ (out-of-plane) polarization. Inset (d) TEY and FY spectra for $E \parallel x$ in the full energy range including the La $M_{4,5}$ white lines. (e) Polarized FY spectra after subtraction of a Lorentzian profile fitted to the La M_4 line shown together with results for Ni^{3+} XAS spectra with 5.5% higher $x^2 - y^2$ occupation, using the cluster calculation. (c) Difference spectra ($E \parallel x - E \parallel z$) calculated from the measured (blue points) and calculated (orange line). Adopted from ref. 44.

The dichroic difference spectrum in Fig. 14(f) clearly shows dips at the Ni L_2 and L_3 white line energies, which the authors attribute to natural linear dichroism. Remember, that the different absorption is also visible in the reflectivity data which were taken at the same energies, Fig. 14(b)-(c).

In general, natural linear dichroism reflects an anisotropy of the charge distribution around a particular ion. To obtain a quantitative estimation of the imbalance in the e_g band occupation in the heterostructure, Benckiser *et. al.* applied the sum rule for linear dichroism, which relates the total integrated intensity of the polarized spectra with different polarizations ($I_{E\parallel x,z}$) to the hole occupation $\bar{n}_{3z^2-r^2}$ and $\bar{n}_{x^2-y^2}$ in the e_g orbitals. After integration in the range of 853-877 eV of the spectra which were separated from the La M_4 contribution, a ratio of the hole occupation of $\frac{\bar{n}_{3z^2-r^2}}{\bar{n}_{x^2-y^2}} = 1.030(5)$ is obtained. From this the orbital polarization,

$$P = \frac{n_{x^2-y^2} - n_{3z^2-r^2}}{n_{x^2-y^2} + n_{3z^2-r^2}} = 5 \pm 2\% \,. \tag{1}$$

is derived with $n_{3z^2-r^2}$ and $n_{x^2-y^2}$ the numbers of electrons. Theoretical calculations find an orbital polarization of 5.5±2% more $n_{x^2-y^2}$ than $n_{3z^2-r^2}$ which means the values agree very well and the slightly higher occupation of the $x^2 - y^2$ orbital is confirmed.

In summary, by combining X-ray scattering and X-ray absorption spectroscopy it was found that the degeneracy of the Ni 3d e_g orbital is lifted along the superlattice normal z in the LNO/LAO superlattice grown on STO. Further investigation reveals that the population of the different orbitals depends also on the neighboring layer, the discussion of which is beyond the scope of this lecture. If you are interested please refer to Benckiser *et al.* [44].

3.3   Magnetization Depth Profile

Efficient spin-polarized transport across a FM/Semiconductor heterostructure presents well-known problems of – the conductivity mismatch and interdiffusion at the interface between the two materials. It is observed that the transport of spins across the FM/SC interface strongly depends on the interface properties viz. roughness, interdiffusion, alloying, and magnetic moments. Therefore one of the fundamental quantities required to understand spin-dependent transport properties in such systems is the magnetic moment at the FM/SC interface. In the following we will discuss about the investigation of interface magnetic moment of Fe/Ge multilayer using neutron reflectivity done by Singh *et. al.* [49].

A Fe/Ge multilayer represented as: $Si(100)/Ge_{10nm}$(buffer-layer)/$[Fe_{7.5nm}/Ge_{5.5nm}]_{\times 10}/Au_{3nm}$ (capping-layer) was grown by radio frequency (RF) sputtering technique. The heterostructure was characterized using both neutron reflectometry (NR) and x-ray reflectometry (XRR). Layer thickness, density profile, and interface roughness were obtained from the reflectivity measurement. Magnetic moment density profile can be obtained from Polarized Neutron Reflectometry (PNR) measurements. Details of the NR, PNR, and XRR can be obtained from Refs. [50–52] and have also been discussed extensively in the course of this spring school. Therefore we would focus mainly on the obtained results.

Fig. 15(a) and (b) shows the XRR (using Cu-K$_\alpha$) and NR (using unpolarized neutrons) measurements, respectively. The thickness and roughness parameters obtained after fitting the NR measurement using a genetic algorithm [53] are given in the table Fig. 15(d). The Bragg peak in NR at $Q = 0.05$ Å^{-1}, corresponds to bilayer thickness of 126 Å in real space ($d = 2\pi/Q$). The interlayer roughness obtained from the fits to NR and XRR are in the range of 8 to 12 Å for the film. These parameters were used unaltered for obtaining magnetic depth profile from the PNR data Fig. 15(c). The PNR profiles were obtained from the sample for spin-up (+) and spin-down (-) neutrons with respect to the sample polarization in the plane of the thin film and the corresponding reflectivities are represented by R$^+$ and R$^-$, respectively. An in-plane field of 400 G was applied on the sample, which was enough to saturate its in-plane magnetization. To model the interface better, the authors used spin asymmetry (ASYM) parameter to best fit the magnetic moment density profile, where ASYM $(Q) = [R^+(Q)\text{-}R^-(Q)]/[R^+(Q)\text{+}R^-(Q)]$. The ASYM parameter is shown in Fig. 15(e) (open circles) along with the best fit to the model of magnetic depth profile (solid line). According to their model the best fit was obtained when they considers that the magnetic moment of each Fe layer is split into three sublayers – Fe1 (at Ge/Fe interface), Fe2 (center of the Fe layer), and Fe3 (at the Fe/Ge interface) – with lower magnetic moments at the interface than at the center. The magnetic moment of 1.30±0.08 μ_B per Fe over
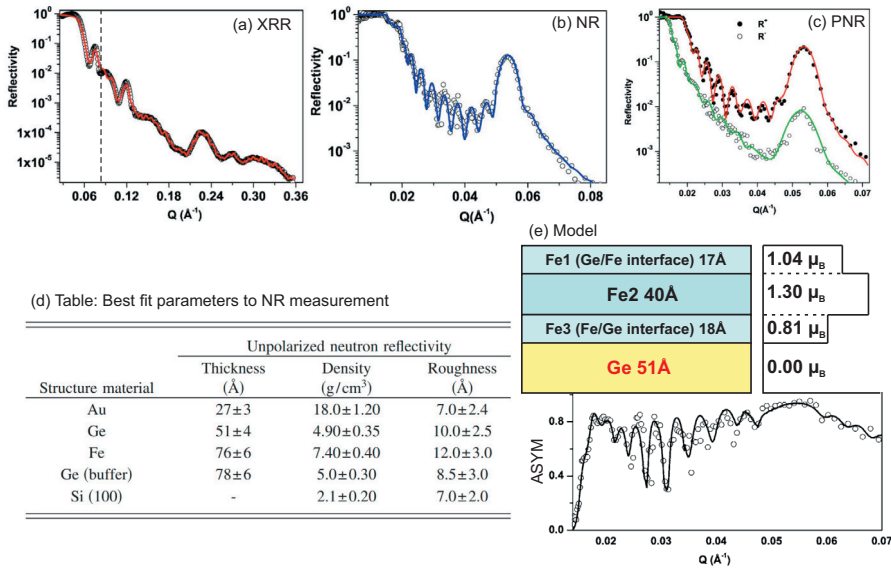
Fig. 15: (a) XRR, (b) NR with unpolarized neutron, and (c) Polarized Neutron reflectometry (PNR) measurement of the Fe/Ge multilayer structure. The open/closed circles in all the reflecivity plots represents the measured data points while the solid line represent the simulated best fit. (d) Parameters for the Fe/Ge multilayer structure obtained by best fit from the NR measurement. And (e) Schematic of the Model used for the fit of ASYM function obtained from the PNR results. Adopted from ref. 49

40 Å at the center of the Fe layer (Fe2 sublayer) away from the interfaces, 0.81 ± 0.04 μ_B over 18 Å for Fe3 sublayer (Fe on Ge) and 1.04 ± 0.05 μ_B over 17 Å for Fe1 sublayer (Ge on Fe).

This inequality in the Fe moments at the two interfaces have been attributed to cumulative effect of roughness and interdiffusion of Ge into Fe layer leading to band hybridization. According to the proposed model the moment at Fe3 interface is expected to be less than that at Fe1 because when polycrystalline Fe is deposited on amorphous Ge, there is more interdiffusion than the other case, i.e., Ge deposition on Fe. Therefore from the above example, we could see that PNR is able to yield magnetic moment density in a multilayer system as a function of depth, with high accuracy [50, 52] and has an advantage over other bulk magnetic measurement techniques in this regard.

3.4 Exchange bias of CoFe/IrMn heterostructures

Exchange bias arises in heterostructures consisting of e.g., ferromagnetic (FM) and antiferromagnetic (AFM) thin films [54,55]. This can be realized by measuring the shift in the hysteresis loop due to the exchange bias field H_{eb}, after annealing the sample above the Néel temperature (T_N, i. e., the AFM ordering temperature) and then cooling down in an external magnetic field H_{FC}. Fig. 16 shows a schematic illustration of the exchange bias effect on the hysteresis of a FM/AFM bi-layer heterostructure, above and below T_N.

To show the power of magnetization mea-
surement and polarized neutron scattering for
investigation of exchange biased systems a
$Co_{80}Fe_{20}/Ir_{20}Mn_{80}$ thin film heterostructure
has been chosen [56, 57]. Bi-layers of 3.0 nm
$Co_{80}Fe_{20}/$ 6.0 nm $Ir_{20}Mn_{80}$ are grown using
dc-magnetron sputtering. The number of bi-
layers is varied from $N = 1$ to 10. While
CoFe is a FM, IrMn is an AFM with T_N
= 520 K. Hence, by heating the sample for
60 min to 533 K and then cooling it down to
room temperature in an external field $H_{FC} =$
+130 Oe an unidirectional anisotropy is in-
duced. Fig. 17 shows the SQUID measure-
ment of the heterostructure where the loops

Fig. 16: Schematic illustration of the hystere-
sis due to exchange bias effect at the FM/AFM
interface where H_{eb} is the exchange-bias field.

reflect the magnetic behavior of the heterostructure in total. The shape of the hysteresis implies
that there are two different loops, which is most obvious for $N = 1$ in Fig. 17(a). The first loop
with low coercivity can be attributed to the hysteresis of the soft magnetic NiFe buffer. The
second one with higher coercivity relates to the hysteresis of the CoFe/IrMn heterostructure and
exhibits, as expected, a shift of the hysteresis loop to negative fields. The contribution of the
FM/AFM heterostructure to the hysteresis loop increases with the number of bilayers. For $N =$
10 the contribution of the NiFe layer is hardly visible. The bias field increases with the number
of bilayers from 330 Oe for $N = 1$ to about 900 Oe for $N = 10$ and the shape of the loop changes
from rectangular to slanted.

Fig. 17: SQUID magnetization loops of [3.0 nm $Co_{80}Fe_{20}/$ 6.0 nm $Ir_{20}Mn_{80}]_N/$ 10.0 nm
NiFe/SiO_2 heterostructures with $N =$ (a) 1, (b) 3, and (c) 10, for circled numbers in (c) re-
fer to Fig. 18. (d) AFM image of the sample surface, and (e) variation of the grain size and H_{EB}
with N. Adopted from refs. 56, 57.

Atomic force microscopy shows a decrease of the grain size with increasing number of bilayers
from 650 nm for $N = 1$ to 60 nm for $N = 10$, see Fig. 17(d) and (e). The enhancement of the
exchange bias field can be correlated with the shrinking of the grain size which is explained in
the framework of the domain-state model of the exchange bias [58, 59]. Grain boundaries are

Fig. 18: Specular reflectivities R_{++} (blue), R_{--} (red), R_{+-} (green), and R_{-+} (black) of a [3.0 nm CoFe/ 6.0 nm IrMn]$_{10}$/ 10.0 nm NiFe/SiO$_2$ at different positions along the magnetization loop (circled numbers refer to Fig. 17). Filled symbols are the data and open symbols the fits. Taken from ref. 56.
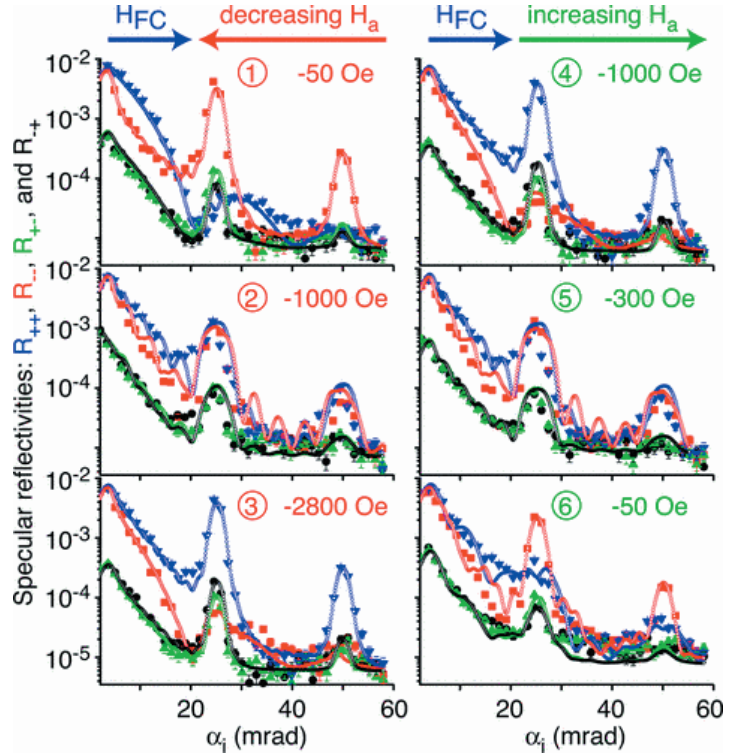
energetically preferred sites for domain walls in antiferromagnetic materials. Hence, smaller grains increase the density of domains in the antiferromagnet which leads to an increase of excess magnetization in the antiferromagnet and increases H$_{EB}$ [59]. For multilayers with $N{\sim}10$ a nontrivial magnetic reversal behavior is expected due to a gradual variation of the grain size and H$_{EB}$ along the growth direction. This will be studied in the following.

In order to study the magnetic reversal of the CoFe/IrMn heterostructure in dependence of the film depth, that means to resolve the magnetization states of the different layers, polarized neutron reflectivity (PNR) measurements with polarization analysis were performed in specular geometry at the reflectometer HADAS of FZ-Jülich. With PNR four characteristic reflectivities are measured as a function of the incident angle. Spin flip (SF) and non-spin-flip (NSF) reflectivities are distinguished with respect to the neutron polarization parallel (+) or antiparallel (-) to the external field. When the in-plane magnetic induction of the sample is parallel to the neutron polarization (i.e., the applied field direction) the neutron polarization is not changed and a NSF reflectivity is observed. In contrast, components of the magnetization orthogonal to the neutron spin can flip the polarization of the incident neutron which results in a non-zero SF reflectivity.

Generally, two reversal processes are distinguished. In case of coherent magnetization rotation the magnetization of the sample may be considered as a single magnetization vector which rotates within the sample plane from one field parallel orientation to the opposite. Hence, there will be a significant increase of the specular reflectivities in the SF channels (R_{+-} and R_{-+}) which are exclusively of magnetic origin and correspond to magnetization components perpendicular to the external field which is also the guiding field for the neutron polarization. Reversal by domain nucleation and propagation does not provide enhanced SF intensities because the

magnetization is always collinear to the external field and solely shows up in the specular NSF reflectivities.

PNR measurements of the [3.0 nm CoFe/ 6.0 nm IrMn]$_{10}$/ 10.0 nm NiFe/SiO$_2$ sample were performed in an external field of different strength. The field direction was collinear to the field used for field cooling, H$_{FC}$. Fig. 18 shows the specular reflectivities. The two peaks in the NSF channels (R_{++} and R_{--}) at 25 and 50 mrad are the first- and second-order Bragg reflections of the heterostructure. The corresponding weak peaks in the SF channels (R_{+-} and R_{-+}) are due to the limited polarization efficiencies of the setup. The peaks correspond to the bilayer thickness of the sample. The NSF and the SF reflectivities are quite similar in Fig. 18 ② and ⑤ which are taken at fields where the magnetization reversal takes place, see also the hysteresis loop in Fig. 17. This means, the reversal for both loop branches proceeds via a state with an almost vanishing magnetization component collinear to the external field and a coherent magnetization rotation can be excluded. Moreover, from off-specular measurements Paul *et al.* even find that the in-plane magnetization component perpendicular to the external field is inhomogeneous on a length scale below 1 μm which is rather attributed to fluctuations of this component than to magnetic domains. For all other fields R_{++} or R_{--} have higher intensities and reflect a net magnetization collinear to the external field, while the specular SF intensities are always much weaker.

The shape of the reflectivities contains further magnetic information which can be retrieved by applying least-square fits based on an extension of the Paratt formalism [48] to magnetic multilayers (open symbols in Fig. 18). From off-specular measurements one knows that there is no correlation of the magnetization perpendicular to the film planes [57] which means every layer may be considered to be uncoupled from the others. This information is utilized for the fitting procedure of the specular reflectivities. In order to obtain nuclear and magnetic scattering length den-

Fig. 19: Switching sequence along the heterostructure from the bottom to top and back as obtained from the fits. For some field values, each CoFe layer in the heterostructure is represented by a box and the arrows indicate a layer magnetization parallel (red) or antiparallel (green) to H$_{FC}$. Taken from ref. 57.

sities, layer thicknesses and interface roughness, fits of the specular intensities in the saturated states are performed. With these parameters fixed the reflectivities of the non-saturated states are fitted assuming different magnetization configurations, with the magnetization within each layer aligned either parallel or antiparallel to the external field. Decreasing the external field the CoFe layers switch sequentially from bottom to the top, and increasing the external field the reversal proceeds in the opposite direction, see Fig. 19. Both, the increasing exchange bias field in the hysteresis loops and the sequential switching of the layers are clearly related to the microstructure, i. e., the decrease of the grain size with increasing number of layers and from bottom to the top within the same multilayer and could be interpreted thanks to polarized neutron scattering.

4 Conclusion

In the last few years, it has been clear that the advancement of future information and device technology rely solely on magnetic thin film heterostructures and efficient manipulation of the carrier spin. Along with advanced material and device processing there is also a vast need to develop new characterization tools to properly understand the material properties and device functionalities. In this chapter, a effort have been made to present some of the most studied thin film heterostructures. We have also learned that the scattering technique provides a unique ability to view such heterostructures at the atomic length scales. However, the data interpretation relies a lot on modeling and big efforts are nowadays performed in this direction. A limiting factor for the application of those methods is that they are intensity limited, especially at neutron sources. Optimized instrumentation at present high flux reactor sources and at present the future spallation neutron sources are pushing this limit.

References

[1] C. Song, B. Cui, F. li, X. Zhou, F. Pan, Prog. Mater. Sci. **87**, 33 (2017).

[2] P. Przyslupski, I. Komissarov, W. Paszkowicz, P. Dluzewski, R. Minikayev, M. Sawicki, Phys. Rev. B **69**, 134428 (2004).

[3] X. Waintal, E. B. Myers, P. W. Brouwer, D. C. Ralph, Phys. Rev. B **62**, 12317 (2000).

[4] M. Baibich, J. Broto, A. Fert, F. N. van Dau, F. Petroff, P. Eitenne, G. Creuzet, A. Friederich, J. Chazelas, Phys. Rev. Lett. **61**, 2472 (1988).

[5] G. Binasch, P. Grünberg, F. Saurenbach, W. Zin, Phys. Rev. B **39**, 4828 (1989).

[6] M. Julliere, Phys. Lett. **54A**, 225 (1975).

[7] J. S. Moodera, L. R. Kinder, T. M. Wong, R. Meservey, Phys. Rev. Lett. **74**, 3273 (1995).

[8] E. B. Myers, D. C. Ralph, J. A. Katine, R. N. Louie, R. Buhrman, Science **285**, 867 (1999).

[9] P. Grünberg, J. Phys. Condens. Matter **13**, 7691 (2001).

[10] I. Žutic, J. Fabian, S. D. Sarma, Rev. Mod. Phys. **76**, 323 (2004).

[11] J. I. Martin, J. Nogues, K. Liu, J. L. Vincent, I. K. Schuller, J. Magn. Magn. Mater. **256**, 449 (2003).

[12] R. Skomskii, J. Phys. Condens. Matter **15**, R841 (2003).

[13] J. J. Liou, C. I. Huang, Microelectronics Journal **32**, 419 (2001).

[14] S. Kayali, G. Ponchak, R. Shaw, Jet Propulsion Laboratory (JPL) Publication **96**, 25 (1996).

[15] T. Mimura, IEEE Trans. Microw. Theory Techn. **50**, 780 (2002).

[16] A. W. Bett, *et. al.*, *28th European Photovoltaic Solar Energy Conference and Exhibition* (EPSCE, Villepinte, France, 2013).

[17] S. A. Wolff, *et. al.*, Science **294**, 1488 (2001).

[18] N. F. Mott, Proc. R. Soc. London Ser. A **156**, 368 (1936).

[19] A. Fert, I. A. Campbell, Phys. Rev. Lett. **21**, 1190 (1968).

[20] T. Valet, A. Fert, Phys. Rev. B **48**, 7099 (1993).

[21] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, H. Ohno, Appl. Phys. Lett. **93**, 082508 (2008).

[22] H. Kanai, *et. al.*, IEEE Trans. Magn. **32**, 3368 (1996).

[23] S. Datta, B. Das, Appl. Phys. Lett. **56**, 665 (1990).

[24] Y. P. Feng, *et. al.*, WIREs. Comput. Mol. Sci. **e**, 1313 (2017).

[25] J. W. Robinson, J. D. Witt, M. G. Blamire, Nature **329**, 5987 (2010).

[26] R. S. Keizer, S. T. B. Goennenwein, T. M. Klapwijk, G. Miao, G. Xiao, A. Gupta, Nature **439**, 825 (2006).

[27] A. Y. Aladyshkin, A. V. Silhanek, W. Gillijns, V. V. Moshchalkov, Supercond. Sci. Technol. **22**, 053001 (2009).

[28] V. V. M. M. Lange, M. J. V. Bael, Mod. Phys. Lett. B **17**, 519 (2003).

[29] H. Jaffrès, A. Fert, Phys. Rev. B **91**, 8111 (2002).

[30] F. Radu, H. Zabel, *Exchange Bias Effect of Ferro-/Antiferromagnetic Heterostructures. In: Zabel H., Bader S.D. (eds) Magnetic Heterostructures*, volume 227 (Springer, Berlin, Heidelberg, 2008).

[31] H. Nakayama, K. Ando, K. Harii, Y. Kajiwara, T. Yoshino, K. Uchida, E. Saitoh, IEEE Trans. on Magn. **46**, 2202 (2010).

[32] R. Wiesendanger, Nat. Rev. Mater. **1**, 16044 (2016).

[33] N. Romming, A. Kubetzka, C. Hanneken, K. von Bergmann, R. Wiesendanger, Phys. Rev. Lett. **114**, 177203 (2015).

[34] C. A. F. Vaz, U. Staub, J. Mater. Chem. C **1**, 6731 (2013).

[35] Y. Tokura, N. Nagos, Science **288**, 462 (2000).

[36] T. Hotta, Rep. Prog. Phys. **69**, 2061 (2006).

[37] H.-L. Lu, Z.-M. Liao, L. Zhang, W.-T. Yuan, Y. Wang, X.-M. Ma, D.-P. Yu, Sci. Rep. **3**, 2870 (2013).

[38] A. Ohtomo, H. Y. Hwang, Nature **427**, 423 (2004).

[39] S. Gariglio, N. Reyren, A. D. Caviglia, J.-M. Triscone, J. Phys.: Condens. Matter **21**, 164213 (2009).

[40] J. A. Bert, B. Kalisky, C. Bell, M. Kim, Y. Hikita, H. Y. Hwang, K. A. Moler, Nat. Phys. **7**, 767 (2011).

[41] M. B. Shalom, M. Sachs, D. Rakhmilevitch, A. Palevski, Y. Dagan, Phys. Rev. Lett. **104**, 126802 (2010).

[42] A. Tebano, E. Fabbri, D. Pergolesi, G. Balestrino, E. Traversa, ACS Nano. **6**, 1278 (2012).

[43] X. Liu, D. Rata, C. Chang, A. C. Komarek, L. H. Tjeng, Phys. Rev. B **90**, 125142 (2014).

[44] E. Benckiser, *et. al.*, Nat. Mat. **10**, 180 (2011).

[45] J.-S. Zhou, J. B. Goodenough, B. Dabrowski, Phys. Rev. Lett. **95**, 127204 (2005).

[46] U. Staub, *et. al.*, Phys. Rev. Lett. **88**, 126402 (2002).

[47] J. Chaloupka, G. Khaliullin, Phys. Rev. Lett. **100**, 16404 (2008).

[48] L. G. Parratt, Phys. Rev. **95**, 359 (1954).

[49] S. Singh, S. Basu, M. Gupta, M. Vedpathakz, R. H. Kodama, J. Appl. Phys. **101**, 033913 (2007).

[50] S. J. Blundell, J. A. C. Bland, Rhys. Rev. B. **46**, 3391 (1992).

[51] J. A. C. Bland, *Polarised Neutron Reflection ch. from Ultrathin Magnetic Structures*, volume I (Springer-Verlag, Berlin, Germany, 1994).

[52] G. P. Felcher, K. E. Gray, R. T. Kampwirth, M. B. Brodsky, Physica B **136**, 59 (1986).

[53] S. Singh, S. Basu, Solid State Phys. **44**, 257 (2001).

[54] W. H. Meiklejohn, C. P. Bean, Phys. Rev. **102**, 1413 (1956).

[55] J. Nogués, I. K. Schuller, J. Magn. Magn. Mater. **192**, 203 (1999).

[56] A. Paul, E. Kentzinger, U. Rücker, D. E. Bürgler, P. Grünberg, Phys. Rev. B **70**, 224410 (2004).

[57] A. Paul, E. Kentzinger, U. Rücker, D. E. Bürgler, P. Grünberg, Eur. Phys. J. B **45**, 249 (2005).

[58] P. Miltényi, M. Gierlings, J. Keller, B. Beschoten, G. Güntherodt, U. Nowak, K. D. Usadel, Phys. Rev. Lett. **84**, 4224 (2000).

[59] U. Nowak, K. D. Usadel, J. Keller, P. Miltényi, B. Beschoten, G. Güntherodt, Phys. Rev. B **66**, 014430 (2002).

E 5   Nanoparticles – Structure and Magnetism

Sabrina Disch
Department für Chemie
Universität zu Köln

Contents

1   Introduction

Nanoparticles are intensely investigated in many fields of research as the reduction of size is usually accompanied by significant effects on their physical properties as compared to the respective bulk materials. Electronically, nanoparticles establish the transition between individual atoms and continuous, bulk materials, and consequently adopt a transition state between classical and quantum phenomena. A typical example is the particle size-dependent band gap in semiconductor nanocrystals. Further size-dependent phenomena include the localized surface plasmon resonance in metal nanoparticles, superparamagnetism in magnetic nanoparticles, and enhanced catalytic activities due to the large surface-to-volume ratio of small nanoparticles. A comprehensive overview of functional nanoparticles and their applications is given in [1].

The additional requirements arising from the reduction of the crystal dimensionality make the controlled synthesis of nanoparticles more complex than the synthesis of the respective bulk materials. An important parameter is the nanoparticle morphology, including the particle size and shape. Whenever the statistical average of a large amount of nanoparticles is probed, *i. e.* for investigation of surface-related properties involving statistical methods such as scattering techniques, a narrow size distribution is desired. Fine control over the kinetics of nucleation and growth processes is thus required. Moreover, stability against precipitation may be desired, which additionally involves control of the surface chemistry. A large variety of synthesis methods has been developed in the last decades in order to fulfill these requirements and to prepare nanoparticles with tunable particle size, shape, and surface properties. Bottom-up approaches are generally preferred for enhanced sample homogeneity on the nanoscale and improved control of the particle size. Large-scale synthesis of monodisperse nanocrystals with defined shape is nowadays routinely achieved by thermal decomposition of precursors in the presence of organic stabilizing agents [2, 3]. Detailed insight into controlled nanoparticle synthesis can be found in [4–6].

The aim of this chapter is to highlight the potential of selected scattering techniques for the structural and magnetic investigation of nanoparticles. We will first concentrate on the particle morphology. Different techniques for characterization of the particle size and size distribution will be introduced, and the different accessible characteristic sizes will be discussed in section 2. The magnetic morphology of nanoparticles is accessible using polarized small-angle neutron scattering and will be addressed in section 3. Although nanoparticles might in a first approximation be regarded as nanocrystals with an atomic structure similar to the bulk, structural disorder becomes more evident with reduced particle size and has a strong effect on the materials properties. Therefore, the atomic Pair Distribution (PDF) analysis will be introduced for investigation of the atomic scale structure and disorder in section 4.

2   Nanoparticle morphology

In addition to phase composition and crystal structures, the sample morphology, including particle size and shape as well as size distribution, is highly relevant for both fundamental science and technological applications of nanomaterials. It is thus crucial to distinguish between the different characteristic sizes depending on the technique applied for their determination. In this section, a typical example on the basis of ferrite nanoparticles will be discussed in order to

illustrate the different information gained on size ranges as probed by different scattering techniques. A sample of cobalt ferrite nanoparticles with a mean diameter of ~ 10 nm is presented in Figure 1. These nanoparticles were prepared by thermal decomposition of a mixed iron-cobalt oleate precursor [7], a technique that is widely used for preparation of ferrite nanoparticles with narrow size distribution [8], and are stabilized by a ligand shell of oleic acid.
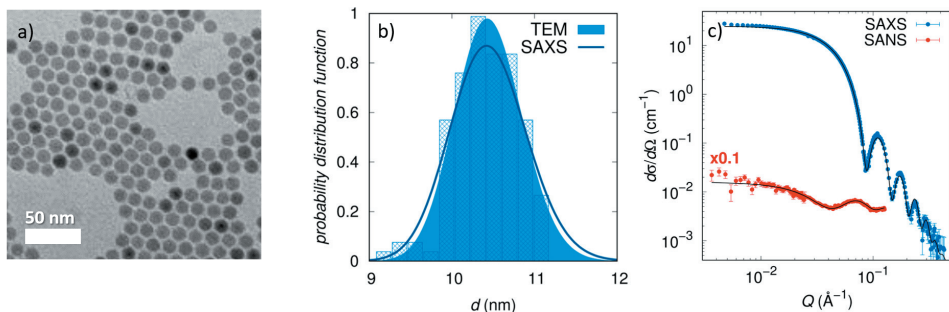
Fig. 1: a) TEM micrograph of cobalt ferrite nanoparticles. b) Lognormal probability distribution function of the particle diameter obtained from TEM (transparent filled curve) and SAXS data (solid line). c) SAXS and SANS data with form factor fit (black line). [9]

Transmission Electron Microscopy (TEM) gives a first impression of particle size, shape, and size distribution, albeit on a local scale where up to few hundreds of nanoparticles can be evaluated. Measurement of more than 100 individual particles of the sample shown in Figure 1a yields the size histogram presented in Figure 1b. The histogram is evaluated using a lognormal distribution of the particle size

$$D(d, d_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma d} \cdot \exp\left(-\frac{(\ln(d/d_0))^2}{2\sigma^2}\right), \tag{1}$$

with the mean particle diameter d_0 and the lognormal standard deviation σ. The advantage of this distribution function over a normal distribution (Gaussian) is the logarithmic weighting that ensures that even for small particle sizes and wide size distributions no negative particle size is taken into account. We obtain for our cobalt ferrite nanoparticles a mean diameter of $d = 10.4(1)$ nm with a size distribution of $\sigma = 3.9(3)\%$, a narrow size distribution that can be considered monodisperse.

Small-angle scattering (introduced in detail in chapter D1) allows for the characterization of density fluctuations in the nanometer size range. The magnitude of the scattering wave vector Q is typically probed in a range of 10^{-3} to 0.6 Å^{-1}, corresponding to real space distances of 1 nm up to several 100 nm. This size range is relevant for a variety of materials such as polymers, biological samples, microemulsions, colloids, and superconductors, and the technique therefore finds application in diverse research fields including soft matter research, biology, materials science, and solid state physics. Small-angle scattering provides statistical information on the sample on a global scale, *i.e.* with $10^{13} - 10^{14}$ particles analyzed at the same time. When applied to nanoparticles embedded in a homogeneous matrix, *i.e.* a dispersion medium, the conventional particle-matrix formalism is sufficient to describe the observed signal according

to

$$\frac{d\sigma(Q)}{d\Omega} = \frac{N}{V}\Delta\rho^2 P(Q)S(Q), \tag{2}$$

where $\frac{N}{V}$ is the particle number density, $\Delta\rho$ the scattering contrast between the particles and the matrix, $P(Q)$ the particle form factor, and $S(Q)$ the structure factor. The structure factor gives information on the interparticle interactions and the resulting interference of scattering contributions of different particles. For characterization of the nanoparticle morphology, dilute dispersions of non-interacting nanoparticles are typically investigated, such that the structure factor approaches $S(Q) = 1$ and can be neglected. The particle form factor describes the morphology of the individual particles and fulfills the condition of $P(0) = V_p^2$ with V_p the particle volume. Form factors have been derived analytically for a large number of different particle shapes, and a comprehensive list of frequently used form factors has been given by Pedersen [10].

The SAXS data of our sample of cobalt ferrite nanoparticles shown in Figure 1c has been modeled using a spherical form factor with a lognormal particle size distribution. We determine a mean particle diameter of $d = 10.4(1)$ nm with a particle size distribution of $\sigma = 4.4(4)$ %. These results are in line with the TEM results and demonstrate that the particle size determined using small-angle scattering is the *density-based* particle size. As a result of the different contrasts for X-ray and neutron scattering, the SANS data of the same sample reveals a distinct form factor. The nanoparticles with their non-deuterated oleic acid ligand shell were dispersed in deuterated solvent (here d_8-toluene), leading to an enhanced H/D contrast between shell and matrix observed using SANS. Therefore, SANS analysis is sensitive to the ligand shell thickness, which is negligible for X-ray scattering. From the SANS data presented in Figure 1c, a ligand shell thickness of $d_{OA} = 1.2(1)$ nm is derived. This is reasonable considering the maximum length of the oleic acid molecule of ~ 2 nm.
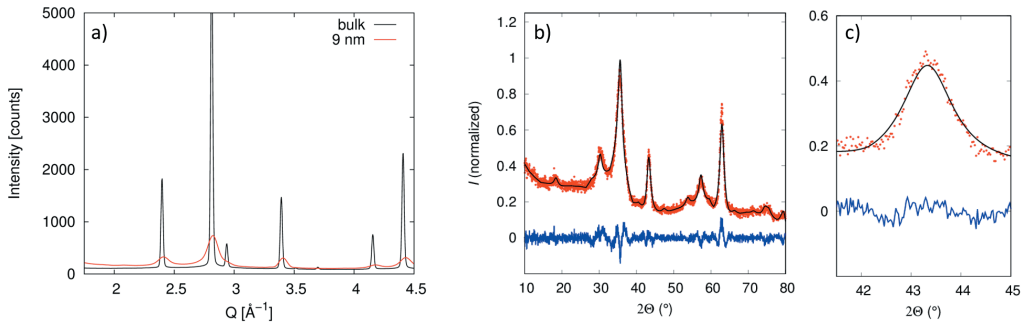
Fig. 2: X-ray diffraction (XRD) by nanoparticles. a) Comparison of XRD data for bulk magnetite and iron oxide nanoparticles. b) XRD data for cobalt ferrite nanoparticles (red) with Rietveld refinement (black line) and difference (blue). c) Detail of the (400) reflection. [9]

Powder diffraction accounts for atomic scale crystal structures, covered in detail in chapter D3. The most prominent effect of nanoscale particle sizes on the diffraction pattern is a strongly increased reflection broadening, even leading to reflection overlap as visible in Figure 2a. The broad and overlapping reflections complicate structural analysis of nanomaterials using Rietveld techniques significantly. However, the quantitative reflection broadening itself can be used as

a measure of the crystallographic grain size. An average grain size L_{hkl} in the hkl direction of the crystal structure is obtained according to the Scherrer equation

$$L_{hkl} = \frac{K\lambda}{\Delta\theta_{hkl}\cos(\theta_{hkl})},\tag{3}$$

with $\Delta\theta_{hkl}$ the full width at half maximum of the hkl reflection, corrected for instrumental resolution, and θ_{hkl} the reflection position, both in radian, λ the used wavelength, and K the phenomenological Scherrer constant.

X-ray powder diffraction data of our nanoparticle sample is presented in Figure 2b. Exemplarily, the (400) reflection is modeled in Figure 2c, yielding a $\Delta\theta_{400} = 0.97°$ at $2\theta = 43.3°$. Using the Scherrer equation and a Scherrer constant of $K = 0.9$, a mean particle size of $L_{400} = 8.8$ nm is obtained. The obtained *structurally coherent grain size* is significantly smaller than the particle size determined using TEM and SAS techniques. This is reasonable considering that reflection broadening gives information on the average structural correlation length in the nanoparticles. A smaller particle size as determined by diffraction techniques relates to a range of structural coherence smaller than the entire nanoparticle and may thus indicate reduced crystallinity due to structural disorder such as lattice strain close to the nanoparticle surface.

Dynamic light scattering (DLS, also: photon correlation spectroscopy) represents an additional technique that provides information on nanoscale particle sizes. DLS is a time-resolved technique for investigation of the dynamics of samples using light scattering. In contrast to the direct investigation of particle size and shape using static scattering techniques, the concept of DLS is the determination of the diffusion properties (Brownian motion) of the sample in solution. These are correlated with the particle size, and thus give indirect information on particle sizes and size distributions. A more detailed introduction into DLS is beyond the scope of this chapter, but can be found in [11].

The typical result of a DLS measurement is presented for our example of cobalt ferrite nanoparticles in Figure 3a. For an ideally diluted sample, the inverse relaxation time Γ accessible by DLS is related to the translational diffusion coefficient D via the Landau-Placzek relation

$$\Gamma = Dq^2.\tag{4}$$

Note that for the scattering wave vector in light scattering experiments, the refractive index of the solvent n_0 has to be taken into account according to $q = \frac{4\pi n_0}{\lambda}\sin\theta$. The translational diffusion coefficient D gives access to the *hydrodynamic radius* R_h of the scattering particles via the Stokes-Einstein relation

$$D = \frac{k_B T}{6\pi\eta_{solv}R_h},\tag{5}$$

with the Boltzmann constant k_B, the temperature T, and the viscosity of the solvent η_{solv}. The hydrodynamic radius, also referred to as Stokes radius, is defined as the radius of a hypothetic, hard sphere that has the same diffusion properties in a specific solvent as the studied particle. Evaluation of the DLS data shown in Figure 3a yields a translational diffusion coefficient of $D = 5.17\cdot10^{-11} m^2/s$. Along with the measurement temperature of $T = 298$ K and the viscosity of the solvent (toluene) $\eta_{solv} = 0.56$ mPa s, we obtain a hydrodynamic radius of $R_h = 7.5$ nm.
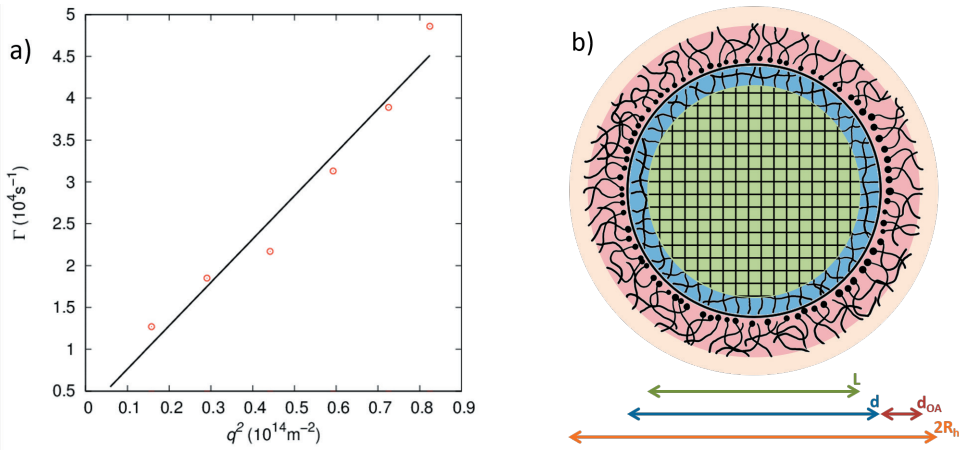
Fig. 3: a) DLS data by cobalt ferrite nanoparticles [9]. b) Characteristic nanoparticle sizes obtained using small-angle scattering (*density-related particle diameter* d and ligand shell thickness d_{OA}), X-ray diffraction (*structurally coherent grain size* L), and dynamic light scattering (*hydrodynamic radius* R_h).

Finally, Figure 3b illustrates the different characteristic sizes obtained using the presented techniques. The *density-related particle size* is obtained using small-angle scattering techniques. Using SAXS, the inorganic particle radius of our ferrite nanoparticles in toluene is obtained, whereas SANS is sensitive to the H/D contrast between the organic ligand shell and the deuterated solvent and provides additional information on the ligand shell thickness. DLS gives access to the *hydrodynamic particle size*, which is larger than the inorganic particle itself and is affected by particle-solvent interactions. X-ray diffraction is sensitive to the *structurally coherent grain size* that is often smaller than the density-related particle size, indicating structural disorder near the particle surface.

3   Magnetic nanoparticle morphology

After the focus on the nanoparticle morphology in the previous section, this section will introduce the use of polarized small-angle neutron scattering for determination of the magnetic nanoparticle morphology. Similar to the determination of the nanoparticle form factor corresponding to particle size and size distribution, the particle-matrix formalism will be applied here to the magnetic form factor sensitive to the magnetization distribution in magnetic nanoparticles. For simplicity, we will maintain the framework on non-interacting magnetic nanoparticles, where any structure factor, nuclear or magnetic, can be neglected. This situation corresponds to dilute systems of structurally and magnetically non-interacting nanoparticles.

The magnetization distribution in nanoparticles strongly depends on magnetic shape and surface anisotropies as well as structural and magnetic surface disorder. The spin structure of single-domain nanoparticles can be considered as a superspin of coupled atomic spins in the core, surrounded by a magnetically dead or disordered region towards the particle surface. This
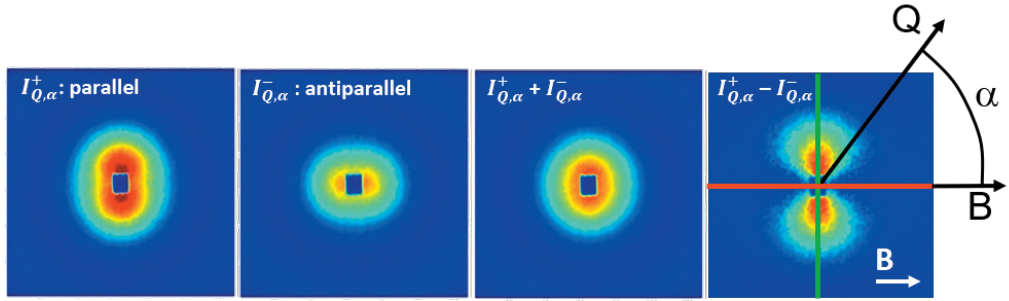
Fig. 4: Magnetic scattering anisotropy for incident neutron polarization parallel ($I_{Q,\alpha}^+$) and antiparallel($I_{Q,\alpha}^-$) to the applied field direction and unpolarized scattering intensity ($I_{Q,\alpha}^+ + I_{Q,\alpha}^-$). The nuclear-magnetic interference term $I_{Q,\alpha}^+ - I_{Q,\alpha}^-$ is maximized perpendicular (green) and vanishes parallel to the applied field (red).

generally accepted concept results mainly from theory and macroscopic magnetization [12, 13], where a lower saturation magnetization than expected from the bulk material is commonly observed and attributed to surface disorder effects [14–16]. In contrast to the macroscopic, spatially averaging techniques, polarized SANS provides the spatial resolution in the nanoscale required to resolve the magnetization distribution within the particle by separation of nuclear and magnetic scattering amplitudes. A difference between nuclear and magnetic particle sizes indicates surface spin disorder, whereas the magnetic scattering length density (SLD) is directly related to the absolute magnetization. It was recently found that the extent of surface spin disorder in noninteracting ferrite nanoparticles varies with an applied magnetic field [9].

Separation of nuclear and magnetic scattering contributions is achieved using polarized neutron scattering. For small-angle scattering, a half-polarized (SANSPOL) experiment (i. e. using a polarized incident neutron beam without polarization analysis after the scattering event) is sufficient to resolve the longitudinal magnetization distribution in the sample parallel to an applied magnetic field. Using longitudinal polarization analysis (POLARIS), both parallel and perpendicular components of the magnetization distribution are accessible [17]. For a dilute sample of non-interacting nanoparticles, the SANSPOL scattering cross section for incoming neutrons polarized either parallel ($I_{Q,\alpha}^+$) or antiparallel ($I_{Q,\alpha}^-$) to a saturating applied magnetic field is given by

$$I_{Q,\alpha}^\pm = F_N^2(Q) + [F_M^2(Q) \mp 2\xi^\pm F_N(Q)F_M(Q)]\sin^2\alpha, \qquad (6)$$

where F_N and F_M denote the nuclear and magnetic scattering amplitudes, $\xi^\pm$ corresponds to the degree of the incident polarization (in case of ξ^- including the flipper efficiency), and α corresponds to the azimuthal angle between applied magnetic field and the scattering wave vector **Q** (Figure 4). Note that eq. (6) is a simplification that does not take into account any scattering contributions arising from spin misalignment and therefore only applies to perfectly aligned magnetization, as in a fully saturated state. For a more precise derivation of the scattering cross sections, refer to [17, 18]. Due to the angular dependence of the magnetic scattering amplitude, even a non-polarized experiment ($I_{Q,\alpha}^+ + I_{Q,\alpha}^-$) in saturating magnetic field gives access to the purely nuclear form factor according to

$$I_{Q,\alpha}^+ + I_{Q,\alpha}^- = 2F_N^2(Q) + 2F_M^2(Q)\sin^2\alpha \qquad (7)$$

for $\alpha = 0$. Once the nuclear scattering amplitude is known, the magnetic scattering amplitude is conveniently accessible from the difference of the SANSPOL cross sections

$$I_{Q,\alpha}^+ - I_{Q,\alpha}^- = -2(\xi^+ + \xi^-)F_N(Q)F_M(Q)\sin^2\alpha \qquad (8)$$

for $\alpha = 90°$. This nuclear-magnetic interference term has the advantage that all background scattering contributions (e. g. from non-magnetic impurities, spin-incoherent scattering, or spin-misalignment contributions) are eliminated. The magnetic scattering amplitude is sensitive to the spatial distribution of magnetic scattering length density within the nanoparticle, similar to the nuclear scattering amplitude which corresponds to the nuclear scattering length density distribution, i. e. the particle morphology. The magnetic scattering length density scales with the magnetic moment and its orientation towards the applied magnetic field (see also Chapter C4). For magnetic nanoparticles, this typically results in a Langevin-type field dependence of the magnetic scattering amplitude for $\alpha = 90°$.
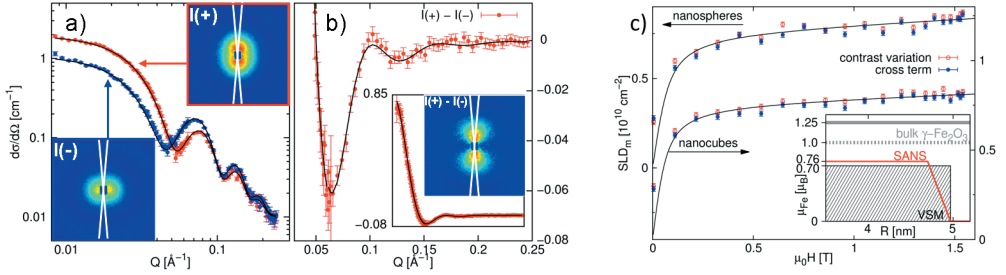
Fig. 5: Polarized SANS by iron oxide nanospheres [19]. Left: Form factor refinements at $\mu_0 H = 1.5$ T. a) Magnetic contrast variation of I_Q^+ and I_Q^-. b) Magnetic-nuclear interference term derived from $I_Q^+ - I_Q^-$. Insets: full intensity range of the interference term and $10°$ sectors used for integration. c) Field dependence of the magnetic SLD in the particle core for nanospheres and nanocubes, refined to the Langevin behavior (eq. (9)). Inset: spatial magnetization distribution in the nanospheres (SANS) compared to the macroscopic (VSM) and the theoretical bulk γ-Fe$_2$O$_3$ moments (solid: 0 K; dashed: 300 K).

The availability of monodisperse samples in large quantities through modern synthesis approaches enables a precise description of the magnetization distribution, and hence the discrimination of surface spin disorder, in magnetic nanoparticles. As an example for the application of polarized SANS, the quantitative spatial magnetization distribution in iron oxide nanocubes and nanospheres will be presented [19]. In this study, nearly monodisperse ($\sigma_{log} \leq 7\%$) samples of maghemite (γ-Fe$_2$O$_3$) nanospheres and nanocubes with 5 nm radius and 8.4 nm edge length, respectively, were investigated. The polarized SANS cross sections obtained from dilute nanoparticle dispersions in deuterated toluene at a maximum magnetic field of 1.5 T are presented in Fig 5. After determination of the nuclear scattering amplitude from both SAXS and nuclear SANS, the magnetic scattering amplitude was refined in both the individual polarized scattering cross sections (Figure 5a, according to eq. (6)) and the nuclear-magnetic interference

term (Figure 5b, according to eq. (8)). The derived spatial distribution of the magnetic scattering length density reveals a reduced magnetization in a thin surface layer that is attributed to surface spin disorder. Such surface spin disorder may result from local structural deviations and broken symmetry at the particle surface. The surface spin disorder layer is found slightly thicker for the nanocubes (0.5 nm) than for the nanospheres (0.3 nm), an effect likely related to the cubic shape anisotropy and corresponding to a larger degree of spin canting in the cube corners.

Next to the qualitative magnetization profile obtained in high magnetic field, field-dependent analysis of the magnetic scattering length density provides quantitative information on the magnetization in the particle core. The field-dependent orientation of the superparamagnetic nanoparticles is taken into account using a Langevin-type approach according to

$$SLD_m(H) = SLD_m^{sat} \cdot \mathcal{L}(\mu_R\mu_0 H/\mathrm{k_B}T) + \beta\mu_0 H, \tag{9}$$

with the Langevin function $\mathcal{L}(x) = \coth(x) - \frac{1}{x}$, where μ_R is the integral particle moment, $\mu_0 H$ the applied magnetic field, $\mathrm{k_B}$ the Boltzmann constant, T the temperature, and β a phenomenological term related to excess magnetic susceptibility observed in superparamagnetic nanoparticle systems [12]. The saturating magnetic scattering length density SLD_m^{sat} is related to the nanoparticle magnetization and averaged atomic magnetic moment according to

$$M_v = 2\mu_{\mathrm{Fe}}^{\perp}/V_m = SLD_m^{sat}/(\gamma_n r_0/2), \tag{10}$$

with V_m the molecular volume of a Fe_2O_3 formula unit, M_v the magnetization, and $\mu_{\mathrm{Fe}}^{\perp}$ the average magnetic moment per iron atom in the ferrimagnetic iron oxide perpendicular to $\mathbf{Q}$. The gyromagnetic factor of the neutron γ_n and the classical electron radius r_0 result in $\frac{1}{2}\gamma_n r_0 = 2.7$ fm. The field-dependence of the magnetic scattering length density in the maghemite nanospheres and nanocubes is presented in Figure 5c and reveals a shape-independent spontaneous magnetization of only 76% of the bulk material. The inset of Figure 5c illustrates the quantitative magnetization distribution in the nanospheres in comparison to macroscopic magnetization and the bulk material. Whereas the observed surface spin disorder layer explains a slight reduction in core magnetization found macroscopically (VSM) as compared to polarized SANS, the magnetization in the particle core is still considerably lower than expected for the bulk material. In consequence, the low magnetization observed macroscopically in nanoparticles results in part from surface spin disorder, but to a much larger extent from reduced magnetization throughout the entire nanoparticle. The reduced magnetization in iron oxide nanoparticles has recently been related to structural disorder phenomena such as antiphase boundaries as observed using HRTEM and diffraction techniques [20, 21].

4 Atomic Scale Structure

In section 2, we have discussed the investigation of nanoparticle morphology in detail. We have seen how a reduced crystalline coherent domain size affects the reflection profiles in diffraction experiments, and that structural disorder may be present within the nanoparticle. In this section, analysis of the atomic pair distribution function will be presented as an approach towards the atomic scale structure in nanoparticles and disordered, *i. e.* crystallographically challenging,

materials.[1] A more detailed introduction into total scattering and pair distribution function analysis is covered in [22–25].

Bragg scattering is based on periodic boundaries and translational symmetries and gives information on the long range average structure of a material under study. For amorphous or locally disordered materials, as well as nanoparticles, these requirements of periodic boundaries are not fulfilled anymore. This results in tremendous reflection broadening as well as diffuse scattering contributions which are usually discarded as scattering background. For analysis of the local structure, however, the diffuse scattering contribution is crucial. In consequence, traditional crystallographic techniques are not sufficient for a complete description of the atomic scale structure in nanoparticles. Based on Fourier transformation of the total scattering intensity including both Bragg and diffuse scattering contributions, analysis of the atomic Pair Distribution Function (PDF) gives access to both the local structural information and the average long range structure in real space. This approach is based on the Wiener-Khintchine theorem, stating that the absolute square of a Fourier transform equals the Fourier transform of the autocorrelation function

$$|\mathcal{F}[f(x)]|^2 = \mathcal{F}[\langle f(0)f(x)\rangle], \tag{11}$$

with the Fourier transform $\mathcal{F}$. In consequence, there are two possible ways to derive the scattering intensity of a scattering density distribution, namely as the absolute square of the Fourier transform or the Fourier transform of the autocorrelation function, *i. e.* the atomic pair distribution function (Figure 6), of the scattering density. The concept of obtaining atomic distances via Fourier transform of the scattering intensity is very similar to the Patterson function, which is used to solve the phase problem in X-ray crystallography. However, as in case of the PDF not only the Bragg intensities are considered but the total scattering function, the obtained PDF is not necessarily a periodic function resembling the unit cell symmetry. It is instead a continuous function reflecting all atomic distances within the material.

The different steps from experimental scattering data to the pair distribution function are shown exemplarily in Figure 7. In most cases, the experimental setup for measurements of the total scattering function $S(Q)$ is very similar to conventional powder diffraction experiments.

---

[1] This section is based on the Lecture Notes D5 of the 43rd IFF Spring School "Scattering Methods for Condensed Matter Research: Towards Novel Applications at Future Sources" (Forschungszentrum Jülich, 2012).
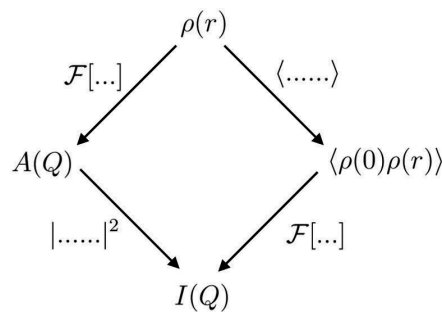
Fig. 6: Two approaches to derive the scattering intensity $I(Q)$ from a microscopic scattering density $\rho(r)$. Left: the absolute square of the Fourier transform. Right: the Fourier transform of the autocorrelation function (corresponding to the PDF).
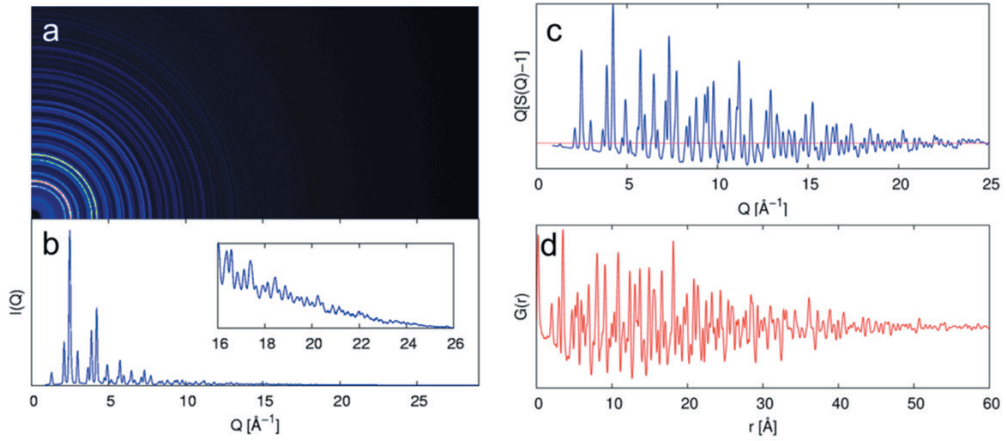
Fig. 7: Data reduction of high-energy X-ray scattering data to the PDF of magnetite: a) 2D scattering intensity. b) Radially integrated and normalized intensity $I(Q)$ after background subtraction. c) Reduced total scattering function $F(Q)$. d) Pair distribution function $G(r)$.

The most important requirements for total scattering measurements exceeding those for conventional powder diffraction are a wide accessible Q range with good statistics and an accurate separation of instrumental and sample related contributions to the $S(Q)$. As valid for all scattering techniques, a balance between good resolution in Q and high scattering intensity has to be found. Good Q resolution is mainly important for investigation of structural coherence lengths or r dependent parameters. The total scattering structure function $S(Q)$ contains both Bragg scattering and diffuse scattering contributions and can be derived from the measured scattering intensity by a number of corrections and normalizations according to

$$S(Q) = \frac{I(Q) - \sum c_i \mid f_i(Q) \mid^2}{\mid \sum c_i f_i(Q) \mid^2}, \tag{12}$$

where $I(Q)$ is the coherent scattering intensity from a powder sample after correction for background and experimental contributions and normalization to flux and number of atoms in the sample. c_i is the atomic concentration and $f_i(Q)$ the atomic form factor (in case of neutron scattering to be replaced by the neutron scattering length b_i) of the element i in the sample. The $S(Q)$ is often presented as the reduced total scattering function $F(Q)$

$$F(Q) = Q[S(Q) - 1], \tag{13}$$

which emphasizes the scattering contributions at high Q. The atomic pair distribution function (PDF) $G(r)$ is directly accessible from the $F(Q)$ through Fourier transform

$$G(r) = 2/\pi \int_0^\infty Q[S(Q) - 1] \sin(Qr) dQ. \tag{14}$$

As explained above, special care is required for decision of the range of data used for Fourier transform. In essence, a large Q_{max}, desired for minimization of termination ripples, has to be balanced against the signal to noise ratio at large Q. As the experimentally derived $S(Q)$

and $F(Q)$ include Bragg scattering as well as diffuse scattering contributions, the obtained PDF contains information on both the average and the local structure. The physical meaning of the $G(r)$ is a weighted probability of finding atomic pairs separated by r. It is defined as

$$G(r) = 4\pi r[\rho(r) - \rho_0], \tag{15}$$

where $\rho(r)$ is the microscopic pair density, ρ_0 is the average number density, and r is the radial distance. $G(r)$ approaches $-4\pi r\rho_0$ in the low r range and oscillates around zero for large r with positive peaks indicating atomic densities exceeding the average number density, whereas negative peaks can be related to distances with less than average number density (note that due to negative neutron scattering lengths, atomic pair distances may also be indicated by negative peaks in neutron PDF). An example $G(r)$ for Nickel powder is given in Figure 8a. The
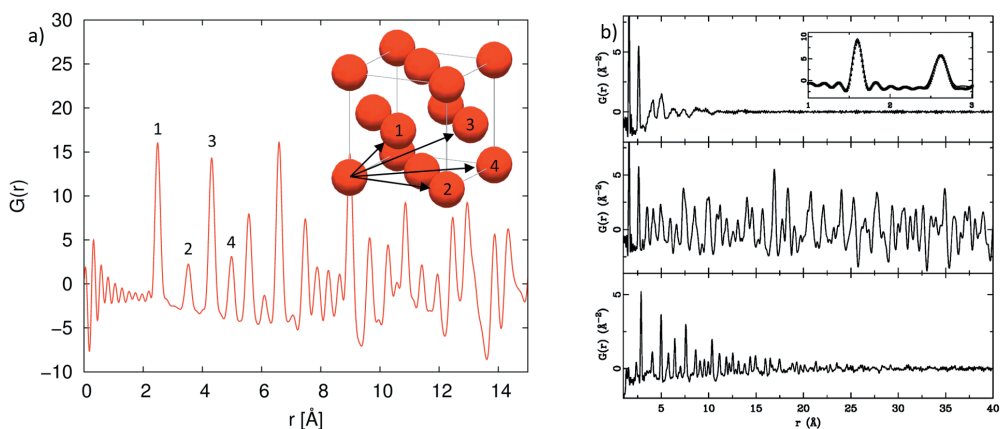
Fig. 8: a) PDF obtained for Nickel powder. The first nearest neighbor correlations are indicated in the *fcc* crystal structure. b) Experimentally obtained PDFs of silica glass (top) and crystalline quartz (center). A direct comparison of the first nearest neighbor distances of both data sets is given in the inset. Experimentally obtained PDF of ∼3.5 nm gold nanoparticles (bottom). Republished with permission of RSC, from [24].

PDF allows for a direct determination of bond lengths from the peak positions. These bond lengths represent the average distance between the positions of pairs of atoms, as opposed to the distance between the average positions of atoms accessible by traditional crystallographic analysis [25]. This difference may become important for largely disordered systems or systems with pronounced lattice dynamics. The peak shape in the PDF gives information on the bond length distribution due to lattice dynamics, correlated motion, or static disorder. A larger peak width indicates a broad distribution of bond lengths, whereas a narrow peak width corresponds to a sharp distribution of bond lengths, which is indicative for correlated motion. Furthermore, relative coordination numbers can be extracted in a quantitative analysis of the PDF from the integrated peak intensity. These may be known for crystalline materials from traditional crystallographic analysis, but in case of compositional disorder, the coordination numbers may reflect local correlations between site occupancies which are not accessible by traditional crystallography [25]. The decay of the peak intensity at large r gives information on the structural coherence of the material under study. For long range ordered crystalline materials, the $G(r)$
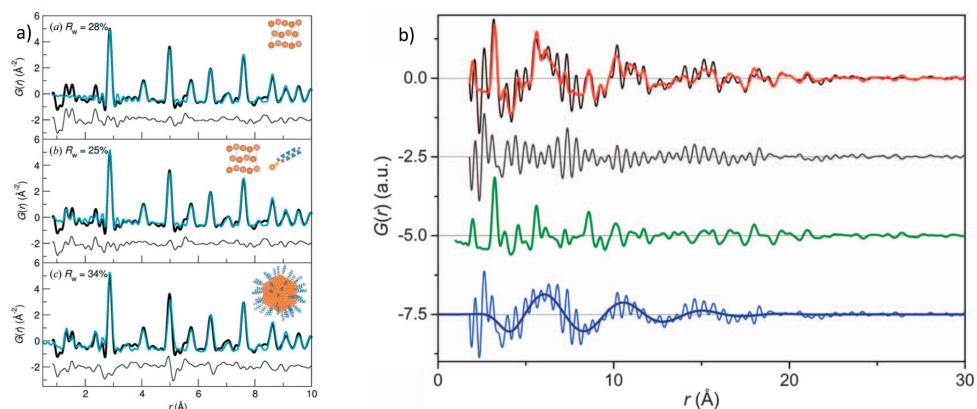
Fig. 9: a) Experimental PDF of Au nanoparticles with refinements of the nanoparticle structure as a single phase (top), the nanoparticle structure and an additional ligand phase (center), and the entire nanoparticle with ligand chains (bottom). Reproduced with permission of the International Union of Crystallography [26]. b) Fit of ZnO nanoparticles with citrate ligands in propanol. Experimental d-PDF with fit (red) and overall difference (grey), along with fit of the nanoparticle contribution to the PDF (green) and the contribution and fit of the restructured solvent (blue). From [27]. Reprinted with permission from AAAS.

should exhibit peaks in an infinite r range, whereas for purely local correlations no PDF intensity is observed beyond the real space correlation length. Similar to reciprocal space analysis, where the peak width results from convolution of particle size (and lattice strain) broadening with the instrumental resolution, the damping of the PDF at large r contains contributions from both the structural coherence length and the instrumental resolution. Examples for different structural coherence lengths are given in Figure 8b. Whereas the PDF of crystalline quartz powder exhibits sharp peaks throughout the entire r range, the PDF of a fused silica glass rod decays very quickly after the first sharp peaks. These peaks correspond to the Si-O and Si-Si nearest neighbor distances, illustrating that the local environment is the same for silica glass and crystalline quartz. The most obvious effect of nanoscale particle sizes on the measured PDF is the smaller structural coherence length as compared to long range crystalline materials, resulting in a massive damping of the PDF at large r.

There are different approaches to describe the damping of the PDF of nanosized materials. It can be described by an envelope function depending of the particle shape and size, which is multiplied with the calculated PDF for the bulk average structure and the instrumental damping factors [28]. Such envelope functions have been derived for spherical particles including particle size distributions and different particle shapes [28, 29]. However, this approach restricts the structural analysis to the average structure throughout the nanoparticle without discrimination of surface-related correlations and disorder. In a different approach, Page *et al.* have demonstrated the variation of the envelope function by different particle sizes and shapes through simulation of the PDF for entire nanoparticles [26]. Such whole-particle structure modeling allows for differentiation of disorder of the nanoparticle compositional against core-shell structures and even provides insight into the surface coverage of an organic ligand shell. Whereas the presence of a thiol ligand shell in the studied gold nanoparticle samples has been confirmed

by a two phase refinement of the nanoparticle and the ligand phases, modeling of the entire particle including the ligand shell allowed for determination of the surface coverage and the Au-S distance (Figure 9a) [26]. Knowing the molecular structure of the ligand, the orientation of the ligand with respect to the particle radial direction was defined. Further examples of the whole-particle approach to modeling nanoparticle PDF include the determination of nanoparticle structure, shape, and defect structure in II-VI semiconductor nanoparticles [30]. Modeling of the PDF revealed a high stacking fault density between wurtzite and zincblende modifications which was not accessible by Rietveld refinement. For CdSe/ZnS core shell nanoparticles PDF analysis further revealed the structurally incoherent attachment of particle core and shell.

Beyond the structural characterization of nanoparticles, differential PDF techniques have recently been employed to investigate the solvent restructuring by nanoparticles in dispersion (Figure 9b). Such investigations require a precise correction for the contributions of both the bulk solvent and the nanoparticles, which were subtracted before and after Fourier transform, respectively. The derived double-differential PDF reveals enhanced short-range order within the solvent that extends up to 2 nm beyond the nanoparticle surface [27].

References

[1] A. M. Schmidt, *46th IFF Spring School "Functional Soft Matter"*, chapter A1 - Functional Nanoparticles (Forschungszentrum Jülich, 2015).

[2] J. Park, K. An, Y. Hwang, J.-G. Park, H.-J. Noh, J.-Y. Kim, J.-H. Park, N.-M. Hwang, T. Hyeon, *Ultra-large-scale syntheses of monodisperse nanocrystals*, Nat. Mater. **3**, 891 (2004).

[3] J. Park, E. Lee, N.-M. Hwang, M. Kang, S. C. Kim, Y. Hwang, J.-G. Park, H.-J. Noh, J.-Y. Kim, J.-H. Park, T. Hyeon, *One-Nanometer-Scale Size-Controlled Synthesis of Monodisperse Magnetic Iron Oxide Nanoparticles*, Angew. Chemie - Int. Ed. **117**, 2932 (2005).

[4] A. H. Lu, E. L. Salabas, F. Schüth, *Magnetic nanoparticles: Synthesis, protection, functionalization, and application*, Angew. Chemie - Int. Ed. **46**, 1222 (2007).

[5] Y. Xia, Y. Xiong, B. Lim, S. E. Skrabalak, *Shape-controlled synthesis of metal nanocrystals: Simple chemistry meets complex physics?*, Angew. Chemie - Int. Ed. **48**, 60 (2009).

[6] J. Park, J. Joo, G. K. Soon, Y. Jang, T. Hyeon, *Synthesis of monodisperse spherical nanocrystals*, Angew. Chemie - Int. Ed. **46**, 4630 (2007).

[7] D. Zákutná, D. Honecker, S. Disch, *Phase-transfer and stabilization of highly monodisperse ferrite nanoparticles into polar solvents by ligand exchange synthesis*, J. Nanosci. Nanotechnol. Accepted (2018).

[8] E. Wetterskog, M. Agthe, A. Mayence, J. Grins, D. Wang, S. Rana, A. Ahniyaz, G. Salazar-Alvarez, L. Bergström, *Precise control over shape and size of iron oxide nanocrystals suitable for assembly into ordered particle arrays*, Sci. Technol. Adv. Mater. **15**, 055010 (2014).

[9] D. Zákutná, S. Disch, *unpublished* (2019).

[10] J. S. Pedersen, *Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting*, Adv. Colloid Interface Sci. **70**, 171 (1997).

[11] W. Schärtl, *Light Scattering from Polymer Solutions and Nanoparticle Dispersions* (Springer, 2007).

[12] R. H. Kodama, a. E. Berkowitz, E. J. McNiff Jr., S. Foner, *Surface Spin Disorder in $NiFe_2O_4$ Nanoparticles*, Phys. Rev. Lett. **77**, 394 (1996).

[13] R. Kodama, A. Berkowitz, *Atomic-scale magnetic modeling of oxide nanoparticles*, Phys. Rev. B **59**, 6321 (1999).

[14] J. Curiale, M. Granada, H. E. Troiani, R. D. Sánchez, A. G. Leyva, P. Levy, K. Samwer, *Magnetic dead layer in ferromagnetic manganite nanoparticles*, Appl. Phys. Lett. **95**, 93 (2009).

[15] A. Kovács, K. Sato, V. K. Lazarov, P. L. Galindo, T. J. Konno, Y. Hirotsu, *Direct observation of a surface induced disordering process in magnetic nanoparticles*, Phys. Rev. Lett. **103**, 1 (2009).

[16] P. Dutta, S. Pal, M. S. Seehra, N. Shah, G. P. Huffman, *Size dependence of magnetic parameters and surface disorder in magnetite nanoparticles*, J. Appl. Phys. **105**, 2007 (2009).

[17] S. Mühlbauer, D. Honecker, E. Périgo, F. Bergner, S. Disch, A. Heinemann, S. Erokhin, D. Berkov, C. Leighton, M. R. Eskildsen, A. Michels, *Magnetic small-angle neutron scattering*, Rev. Mod. Phys. Accepted (2018).

[18] A. Wiedenmann, *Polarized SANS for probing magnetic nanostructures*, Phys. B **356**, 246 (2005).

[19] S. Disch, E. Wetterskog, R. P. Hermann, A. Wiedenmann, U. Vainio, G. Salazar-Alvarez, L. Bergström, T. Brückel, *Quantitative spatial magnetization distribution in iron oxide nanocubes and nanospheres by polarized small-angle neutron scattering*, New J. Phys. **14**, 013025 (2012).

[20] Z. Nedelkoski, D. Kepaptsoglou, L. Lari, T. Wen, R. A. Booth, S. D. Oberdick, P. L. Galindo, Q. M. Ramasse, R. F. L. Evans, S. Majetich, V. K. Lazarov, *Origin of reduced magnetization and domain formation in small magnetite nanoparticles*, Sci. Rep. **7**, 45997 (2017).

[21] E. Wetterskog, C. W. Tai, J. Grins, L. Bergström, G. Salazar-Alvarez, *Anomalous magnetic properties of nanoparticles arising from defect structures: Topotaxial oxidation of $Fe_{1-x}O|Fe_{3-\delta}O_4$ core|shell nanocubes to single-phase particles*, ACS Nano **7**, 7132 (2013).

[22] T. Egami, S. Billinge, *Underneath the Bragg peaks - Structural Analysis of Complex Materials* (Pergamon, Amsterdam, 2003).

[23] S. J. L. Billinge, M. G. Kanatzidis, *Beyond crystallography: the study of disorder, nanocrystallinity and crystallographically challenged materials with pair distribution functions*, Chem. Commun. p. 749 (2004).

[24] T. Proffen, H. Kim, *Advances in total scattering analysis*, J. Mater. Chem. **19**, 5078 (2009).

[25] C. A. Young, A. L. Goodwin, *Applications of pair distribution function methods to contemporary problems in materials chemistry*, J. Mater. Chem. **21**, 6464 (2011).

[26] K. Page, T. C. Hood, T. Proffen, R. B. Neder, *Building and refining complete nanoparticle structures with total scattering data*, Journal of Applied Crystallography **44**, 327 (2011).

[27] M. Zobel, R. B. Neder, S. A. J. Kimber, *Universal solvent restructuring induced by colloidal nanoparticles*, Science **347**, 292 (2015).

[28] R. C. Howell, T. Proffen, S. D. Conradson, *Pair distribution function and structure factor of spherical particles*, Phys. Rev. B **73**, 094107 (2006).

[29] K. Kodama, S. Iikubo, T. Taguchi, S.-i. Shamoto, *Finite size effects of nanoparticles on the atomic pair distribution functions*, Acta Crystallographica Section A **62**, 444 (2006).

[30] R. B. Neder, V. I. Korsunskiy, C. Chory, G. Müller, A. Hofmann, S. Dembski, C. Graf, E. Rühl, *Structural characterization of II-VI semiconductor nanoparticles*, physica status solidi c **4**, 3221.

E 6   Energy Materials

Michael Ohl
Jülich Centre for Neutron Science
Institute of Complex Systems
Forschungszentrum Jülich GmbH

Contents

---

1   Introduction

In our modern society we have to deal with all sorts of challenges. To accomplish the needs of our growing society it becomes more and more important to raise the question of the demand of this modern society. While it can be easier expressed and defined what is needed the technical limitations often hinder us to accomplish the goals needs e.g. any form of technical solution should be sustainable. Among others like life science and nanotechnology the demand for energy application increases. Energy is a very serious question and since our societies demand on energy is rising it becomes more and more important. As such, many researchers studied, study and will study energy related questions. Many questions can be raised but the main questions lie in generation, storage and transfer of energy in different forms like motion, electricity etc. All these open questions will not find one sole answer and are subject to dynamic process in which they will be reformulated. In addition, there are too much and huge variety of devices and applications available which form a wide selection and development of innovative solutions. While some are already established others are still subject to intense research [1].

Here we will focus on materials which offer the properties to serve in generation, storage and transfer of energy in different forms. This is not an attempt to cover all available energy material related applications of neutron scattering but shall give some insight of the range of possibilities.

We are reaching a point where new materials with improved physical properties are urgently needed for energy innovation. Also one shall never forget that any solution shall be sustainable for a long term solution e.g. being environmentally friendly. As such modern research fields for example like nanotechnology interplays strongly with energy questions. In many areas of energy research, simply observing and using materials available in nature is no longer enough to drive significant changes. We need new approaches to find solutions for the challenges we face in the field of energy materials. Our ability to actively design and tailor material properties will define what is possible in our energy future.

There are a lot of energy materials reported at least what one would call energy materials. However, beside materials which are further away from any applications there are a number of applications in our daily life which use already a wide range of energy materials. A list of what we believe depends on energy material is given below:

- Photovoltaics/ Solarpower
- Hydrogen storage
- Fuel cells
- Magneto caloric effect
- Ion Batteries

Energy materials in these fields will play a major role for our future. However, most of these fields can't satisfactory be covered in this course due to time limitations so one has to focus on

a few. Energy materials even studied solely with scattering methods could easily form an own spring school.

The question remains what can be covered with scattering methods? In this short communication we will focus mainly on neutron albeit X-ray scattering is a major tool as well.

Neutron scattering is practiced at research reactors and spallation neutron sources that provide neutron radiation with varying intensities. One generally can distinguish between elastic and inelastic neutron scattering. Elastic neutron scattering is commonly realized with neutron diffraction techniques/ instruments and their main application is to analyze structures.
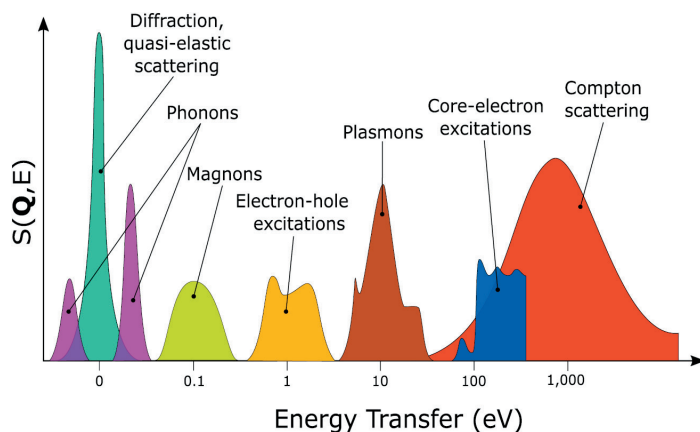
Fig. 1: Quasielastic neutron scattering (QENS) and inelastic scattering (INS) versus energy transfer for a variety of applications. X-ray scattering is shown at higher energy transfers as well.

Structure is determined usually with neutron diffraction (ND). The structure factor, S(Q) describes scattered neutron intensity in terms of the momentum transfer Q, where $Q = 4\pi \sin\theta /\lambda$, and θ is the angle of the scattered neutrons with λ being the incident-neutron wavelength. For a single crystal, the scattering will consist of Bragg peaks. In an ideal powder sample, small crystallites are randomly oriented and scattering from a particular set of lattice planes corresponds to the scattering obtained by turning a single crystal. In powder samples, Debye-Scherrer cones are obtained in place of Bragg peaks, where intensity from the cones can be determined simultaneously using large area detector arrays.

Inelastic neutron scattering is used for studying atomic vibrations and other excitations in a wide energy range (see Fig. 1). Also one distinguishes between quasielastic (QENS) and purely inelastic scattering (INS). Whereas QENS is usually understood as a broadening of an elastic peak – so it is "quasi - elastic" (see D5). The signal is usually obtained when unfolding the instrumental resolution.

In situ studies were considered as well and are of increasing importance. The most important question arise from the analysis of the study of energy materials and is of course related to the understanding of structure- and dynamic-function relations. Comprehending the working mechanism, at the atomic and molecular scale, is the key to progressing alternative and sustainable-energy technologies, and fundamental to this is the study of the materials during operation. In situ and even operando studies are commonplace and necessary in energy materials research. The in-situ technique, often applied to materials under equilibrium, has been extended in recent years to operando studies, where the materials are studied under non-equilibrium conditions whilst performing their function. The advent of new-generation reactor and spallation neutron sources, as well as associated faster instrumentation, has greatly assisted in facilitating such research [1].

Please be aware, that all accessible scientific question of energy materials can just be answered in the given time and length scale regime for this neutron scattering techniques as shown in Fig. 1. Because neutrons are electrically neutral, they penetrate more deeply into matter than electrically charged particles of comparable kinetic energy and thus are valuable as probes of bulk properties.
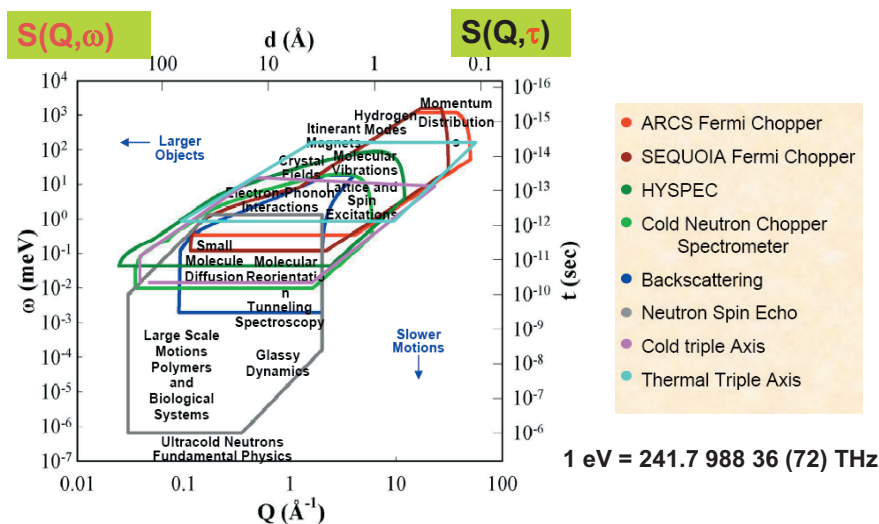
Fig. 2: $S(Q,\omega)$ range for a variety of inelastic neutron scattering instruments at the spallation neutron source in the Oak Ridge National Laboratory. Please note, the momentum and energy transfer are related which justifies the odd shape of the covered area for the different instrument types.

Neutrons interact with atomic nuclei and with magnetic fields from unpaired electrons, causing pronounced interference and energy transfer effects in neutron scattering experiments. Unlike an X-ray photon with a similar wavelength, which interacts with the electron cloud surrounding the nucleus, neutrons interact primarily with the nucleus itself, as described by Fermi pseudopotential. Neutron scattering and absorption cross sections vary widely from isotope to

isotope despite X-ray scattering where the cross section is mainly proportional to the atomic number squared (Z^2). Subsequently, often both technique have a unique advantage when studying energy materials with a decision on a by-case basis.

Neutron scattering can also be incoherent or coherent, depending on isotope. Among all isotopes, hydrogen has one of the highest scattering cross section. Important elements like carbon and oxygen are quite visible in neutron scattering—this is in marked contrast to X-ray scattering. Thus neutrons can be used to analyze materials with low atomic numbers, including proteins and surfactants. X-ray scattering can be done at synchrotron sources but very high intensities are needed, which may cause the structures to change or in general be destroyed. The nucleus provides a very short range, as isotropic potential varies randomly from isotope to isotope, which makes it feasible to tune the (scattering) contrast to suit the experiment e.g. also by replacing certain atoms like hydrogen with their deuterium counterpart.

Scattering almost always presents both elastic and inelastic components. The fraction of elastic scattering is determined by the Debye-Waller factor or the so called Mössbauer-Lamb factor.

Achieving a precise velocity, i.e. a precise energy and de Broglie wavelength, of a neutron beam is important. Such single-energy beams are termed 'monochromatic', and monochromaticity is mostly achieved either with a crystal monochromator or with a chopper system e.g. at a time-of-flight (TOF) spectrometer. With choppers in the time-of-flight technique, neutrons are sent through a sequence of at least two rotating slits such that only neutrons of a particular velocity are selected. Spallation sources have been developed that can create a rapid pulse of neutrons. The pulse contains neutrons of many different velocities or de Broglie wavelengths, but separate velocities of the scattered neutrons can be determined afterwards by measuring the time-of-flight of the neutrons between the sample and neutron detector.

We know, magnetic neutron scattering is possible at several research facilities as well. The neutron has a net electric charge of zero, but has a significant magnetic moment, although only about 0.1% of that of the electron. Nevertheless, it is large enough to scatter from local magnetic fields inside condensed matter, providing a weakly interacting and hence penetrating probe of ordered magnetic structures and electron spin fluctuations.

2   Energy Materials

The most pressing question for neutron scattering scientist are, how can we unravel information with neutron scattering which is limited in its time and length scale range ?

Also one should note, despite the timeline for new generation of energy materials and their associated applications/ devices is very short, neutron scattering often requires a much longer time. Experiments must be properly planned, successfully proposed, performed in a limited time frame and evaluated before one could make any statements. All this requires of course much more time than regular in house laboratory measurements/ experiments. However, for a

lot of energy relevant applications improvements of their relevant parameters reaches a saturation. So a more fundamental understanding of the relevant processes is needed for further improvements e.g. Li ion conductivity mechanism in batteries. And of course too often just neutrons can enlighten and answer questions for this fundamental understanding [1, 2].

For example neutron diffraction offers information about atomic positions and neutron spectroscopy gives information about atomic (or molecular) motion. Also, neutron absorption reveals information concerning material composition through radiography and tomography. The instrumentation for both photons and neutron are well known even so technology is subject to permanent research with innovative ideas and finally benefit for the applicable research. For neutrons there is an almost analogous group of techniques that together cover length scales from fractions of an Å to microns (and up to many centimetres for radiography) and timescales that cover from femtoseconds to hundreds of nanoseconds (see Fig. 2). The generic properties of neutrons lead to the recurrent use of particular neutron scattering and neutron-based analysis throughout this contribution, and this section explains the basic principles of these [5, 6].

2.1   Photovoltaics and Solarpower

Photovoltaics are best known as a method for generating electric power using solar cells to convert sunlight into a flow of electrons with the photovoltaic effect [3]. It can be used to power equipment or to recharge a battery. The first practical application of photovoltaics was to power orbiting satellites and other spacecraft, but today the majority of photovoltaic modules are used for grid connected power generation. In this case inverters are required to convert the DC to AC. There is a smaller market for off-grid power for remote dwellings, recreational vehicles, electric cars, roadside emergency telephones, remote sensing and cathodic protection of pipelines.

Photovoltaic power generation employs solar panels composed of a number of solar cells containing a photovoltaic material [4]. Copper solar cables connect modules (module cable), arrays (array cable), and sub-fields. Because of the growing demand for renewable energy sources, the manufacturing of solar cells and photovoltaic arrays has advanced considerably in recent years [5].

Solar photovoltaic power generation has long been seen as a clean energy technology which draws upon the planet's most plentiful and widely distributed renewable energy source – the sun. Cells require protection from the environment and are usually packaged tightly in solar panels. Photovoltaic power capacity is measured as maximum power output under standardized test conditions (STC) in "W_p" (watts peak) [6]. The actual power output at a particular point in time may be less than or greater than this standardized, or "rated", value, depending on geographical location, time of day, weather conditions, and other factors.

Crystalline silicon photovoltaics are only one type of photo voltaic (PV), and while they represent the majority of solar cells produced currently there are many new and promising technologies that have the potential to be scaled up to meet future energy needs. As of 2018, crystalline silicon cell technology serves as the basis for several PV module types, including monocrystalline, multicrystalline and bifacial. Regarding the scattering cross sections of the

atoms therein for neutron scattering mostly structure analysis can be done. However, the question is still not answered how especially neutron scattering can help to boost this technology.
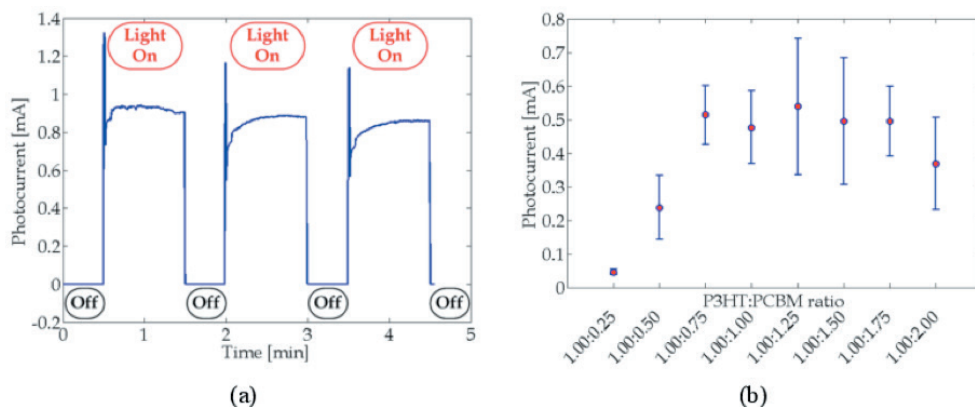
Fig. 3: Photocurrent as a function of light/ time (a) and ratio of the organic photovoltaic P3HT:PCBM (b).

The analysis of the structure e.g. a crystalline long range order or even disorder as in polymeric glasses is one way but often done easier with X-ray scattering. In addition, there is another group of PV the so called Organic photovoltaics (OPV) which fall into the thin-film category of manufacturing, and typically operate around the 12% efficiency range which is lower than the 15–21% typically seen by silicon based PVs. On the other side OPVs offer some advantages over the generic silicon based PVs. Among other advantages, they are:

- It is light in weight.
- It is semi-transparent.
- It is environmental friendly.
- OPV are 1000 times thinner compared to silicon solar cells. Hence it results into huge savings of materials and consecutively they are cheaper.
- They can be easier adapted to various odd surfaces.
- It is flexible due to use of PV modules.
- It offers short energy payback time.
- It is easy to integrate.

On the downside stands, they degrade in time faster, the efficiency is lower, moisture and air along with sunlight further degrades its performance over time as well.

Direct access with neutron scattering techniques on the process: conversion of light into a flow of electrons is not possible since this process does not lie in the range of dynamic neutron scattering (see Fig. 2). Nevertheless organic photovoltaics with their layer structure is very attractive and one of the most prominent candidate of an OPV is a blend called P3HT:PCBM (Fig. 3). In this blend one component serves as donor while the other is the acceptor. When

applying light a photocurrent is released and there is a optimal composition of donor and acceptor.

There were already several attempts to study this system with neutron scattering. One of the techniques which can be used is neutron reflectometry which is suited for thin layers in the range from 0.5 nm to 350 nm. The layer structure can be looked at. With an additional technique called depth profiling one can get information about the concentration or composition profile perpendicular to the surface or interface. The interface of P3HT and PCBM is relevant for proper charge transfer and it can be studied via neutron scattering techniques. There might be some potential but not very much is done in this field yet.

To summarize, even so PV is a modern and increasing research field, neutron scattering is playing a minor role to study its properties and reveal the relevant parameters. The main underlying process when light is converted into electricity doesn't lie in the accessible length and energy range and just interfaces of donor and acceptor are investigated in thin films.

2.2   Hydrogen storage

Hydrogen has been identified as a fuel of choice for providing clean energy for transport and other applications across the world and the development of materials to store hydrogen efficiently and safely is crucial to this endeavor.

To realize a "Hydrogen Economy" requires solutions and devices in three fundamental areas: hydrogen production, hydrogen storage, and fuel cells. We will discuss fuel cells in the next chapters. Neutron scattering is used in all these areas but the most significant impact is in hydrogen storage devices. Especially Inelastic neutron scattering (INS) is of utmost importance to study hydrogen storage because they directly excite the motions of hydrogen atoms with bonds to other atoms. Subsequently, the local structure, vibrational dynamics, diffusion processes and nature of hydrogen-bond dynamics can be addressed with neutron scattering. The large incoherent neutron-cross section for hydrogen supports studies with neutron scattering and often a labeling or replacement of hydrogen with deuterium when chemically feasible is an immense advantage [1]. Consequently, intensity of e.g. peaks in vibrational spectra can be directly assigned to hydrogen bonds.

Hydrogen can be stored locally in large amount e.g. as a cryogenic liquid but for mobile applications the requirements are higher. In cars mostly molecular hydrogen is stored as a gas of liquid. Since these technologies are subject to research for many years, reasonable effort has been achieved but the requirement for cars with a range of 500km and more is still a challenge.

The progress in this regard is strongly depending on the material synthesis in a huge amount of storage material. Often standard thermodynamic measurement techniques are used to characterize those materials. However the use especially of neutron scattering techniques provide additional essential details on the interactions of hydrogen with the host material. In addition these studies serve as a basis for further computer modelling as well as input and validation of the results. Also the search and choice of suitable hydrogen storage media is supported. It is believed that one of the most powerful experimental methods for molecular-

level studies of hydrogen in materials involve the use of neutrons because of their outstanding sensitivity to hydrogen. Neutrons provide the rotational and vibrational dynamics of the adsorbed hydrogen and additionally diffusive motions through the material. The dynamics of the hydrogen molecules are in strong interaction with its surroundings and can therefore be taken as an indirect measure for its performance as storage material. However, microscopic diffusion of hydrogen is critical for the desired rapid loading, and release, but has not been studied in great detail to date. On the other side, INS is an established technique for such studies and can be easily applied at least but not limited with so called backscattering, time-of-flight and vibrational spectroscopy instruments. Hydrogen storage material is ideal and both diffraction and spectroscopic investigations can readily be performed which additionally can be done under a wide range of experimental conditions such as pressure, temperature, hydrogen content [1].

For a lot of hydrogen storage materials the structure and thus knowledge of the positions of H – species is a necessity for further inelastic studies unraveling the nature of reorientation or diffusion processes which help to understand discharge processes etc.. It is natural that one first has to decipher the structure and in the second step studies the dynamic processes in hydrogen storage material. Without the knowledge of the structure any interpretation of the inelastic data even with computer simulation is weakened. Many of the computation methods depend on the knowledge of the long – range structure of these materials. Also, the combination of measurements with computational methods is giving even more information about those processes.

2.3   Fuel cells

A fuel cell is an electrochemical cell that converts the potential energy from a fuel into electricity through an electrochemical reaction of hydrogen fuel with oxygen or another oxidizing agent . Fuel cells can produce electricity continuously for as long as fuel and oxygen are supplied and is thus often considered as a viable alternative to a more intense usage for example to replace a combustion engines in vehicles. Hydrogen is the basic fuel, but fuel cells also require oxygen. Every fuel cell has two electrodes, the anode and cathode. In addition, they possess an electrolyte which carries electrically charged particles (Hydrogen). The reactions that produce electricity take place at the electrodes. The fuel cells possess an electrolyte which supports the motion of the electrically charged particles from one electrode to the other. In addition, a catalyst which is a chemical substance increases the rate of a reaction at the electrodes without being consumed.

One great appeal of fuel cells is that they generate electricity with very little pollution–much of the hydrogen and oxygen used in generating electricity ultimately combine to form a harmless byproduct, namely water. A detail of terminology: a single fuel cell generates a tiny amount of direct current (DC) voltage. In practice, many fuel cells are usually assembled into a stack.

The function principle of a fuel cell is shown in Fig. 5 in more detail. At the anode a catalyst causes the fuel to undergo oxidation reactions that generate protons (positively charged hydrogen ions) and electrons. The protons flow from the anode to the cathode through the electrolyte after the reaction. At the same time, electrons are drawn from the anode to the

cathode through an external circuit, producing a direct current/ electricity. At the cathode, another catalyst causes hydrogen ions, electrons, and oxygen to react, forming water. Note, between anode and cathode often a membrane or separating layer is used which acts as a barrier film separating the gases in the anode and cathode compartments of the fuel cell. In practice, many fuel cells are usually assembled into a stack. Cell or stack, the principles are the same.

There are many types of fuel cells, but they all consist of an anode, a cathode, and an electrolyte that allows positively charged hydrogen ions (protons) to move between the two sides of the fuel cell. They are all different by the type of electrolyte and startup time. One very important type of fuel cells are proton exchange membrane fuel cells (PEMFC) in which the charge transporting media are protons.

Already in the previous chapter it was mentioned how well hydrogen atoms can be detected with neutrons in a surrounding "matrix" which is a result of the high incoherent cross section of hydrogen. Also labelling or exchanging hydrogen with deuterium helps to directly assign observed scattering intensity to the hydrogen.

The transport mechanism can be considered as translational diffusion in an electrolyte [7]. Those dynamic processes can be detected with backscattering and time-of-flight neutron spectroscopy.
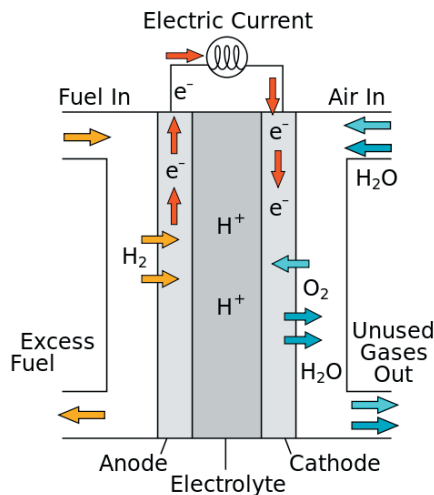
Fig. 4: Function principle of fuel cells (https://en.wikipedia.org/wiki/Fuel_cell).

One example shall be given now in which QENS contributed immense to the understanding of the proton conductivity process. The proton/ hydrogen conductor: $La_{1-x}Ba_{1+x}GaO_{4-x/2}$ is a compound with distorted tetrahedral (GaO_4) units (Fig. 5). Increasing the barium content in expense of lanthanum content leads to oxygen vacancies for charge compensation. Water can be incorporated into the oxygen vacancies as protonic defects from humid atmospheres and oxide ion conduction in dry atmospheres, and proton conduction in humid atmospheres appears [8, 9].
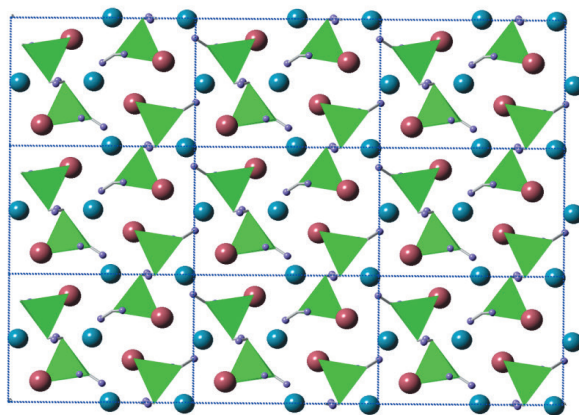
Fig. 5: Illustration of the $La_{1-x}Ba_{1+x}GaO_{4-x/2}$ orthorombic $P2_12_12_1$ structure with distorted GaO_4 tetrahedra (green in the plot). The blue spheres represent the lanthanum and the brown spheres represent the barium.

The question now remained: how does the ion conduction works ?

As a consequence, the compound $La_{0.8}Ba_{1.2}GaO_{3.9} \times 0.08 H_2O$ with water was investigated with incoherent QENS [10]. Neutron-spin echo measurements were performed (Fig. 6) and the data was fitted with an exponential function for a single diffusion process (D6 - Equation 38, upper plot of Fig. 6) resulting in a characteristic temperature dependence of the diffusion coefficient (lower plot of Fig. 7) in dependence of the temperature.

For the NSE experiment the diffusion coefficient followed a straight line and the slope of the temperature dependent diffusion coefficient was evaluated resulting in an activation energy for this single diffusion process of 0.4 eV. Comparison with literature resulted in an assignment of an intra-tetrahedra jump of the hydrogen atom from the atom O2 to O4 (see Fig. 7).
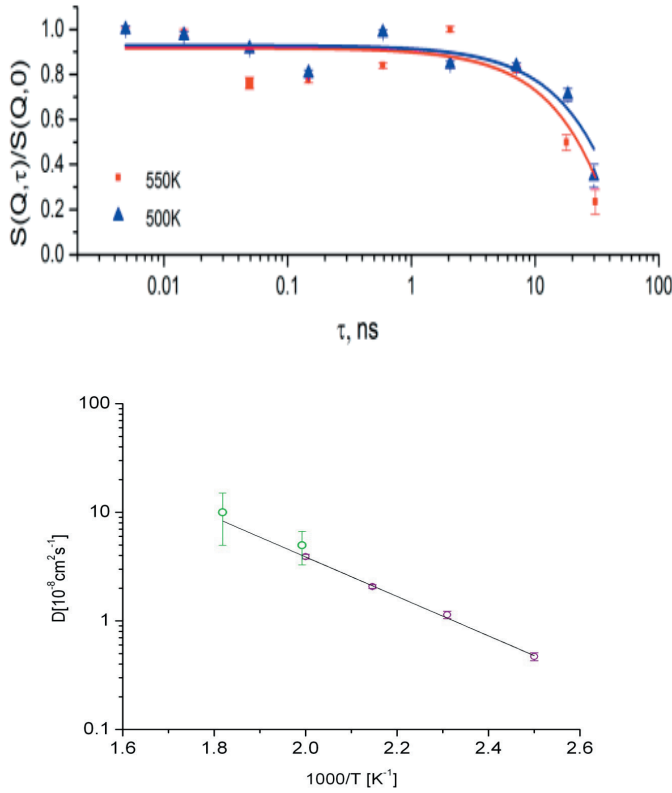
Fig. 6: Upper plot shows the normalized intermediate scattering function $S(Q,\tau)/S(Q,0)$ which was fitted with a single exponential function and lower plot shows the resulting temperature dependent diffusion coefficient (data was added from the wider component of the backscattering measurement).

The second measurement technique was backscattering spectroscopy (see D5). The measured scattering function $S(Q,\omega)$ was fitted with two Lorentzian functions from which the narrow component half width at half minimum (HWHM) was received and plotted. The plot of the HWHM showed a Q^2 dependence obeying the so called Fick's law [2] and it indicated rather long range translational diffusion. In addition, an activation energy of 0.07eV was obtained and compared to the literature [11] and the jump process of the hydrogen was assumed to be inter-tetrahedra.

To summarize, both jump processes for the hydrogen had been decoded with QENS, the inter-tetrahedra jumps (blue arrows in right plot of Fig. 7) with a rather low rate limiting activation energy of 0.07eV and the intra-tetrahedra jumps (gray arrows in right plot of Fig. 7) with an activation energy of 0.44eV.

In this case QENS, with direct access to the Hydrogen motion, disclosed the proton conduction mechanism which might pave the way to better fuel cells with an optimized conduction mechanism.
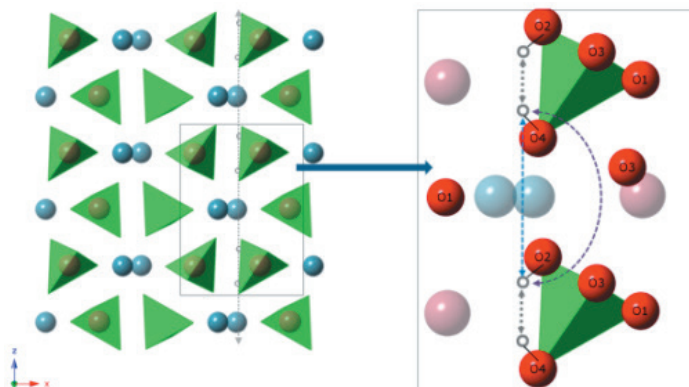
Fig. 7:   Zoom of the structure which shows on the right side the intra-tetrahedra (gray arrows) and inter-tetrahedra (blue arrows) jumps.

Another very common technique is neutron imaging [12]. The aim of these studies are on water distribution inside and near the gas diffusion layers around the membrane ("in-plane" study) but due to space limitations can't be considered here.

2.4 Magneto Calorics

The magnetocaloric effect (MCE, from magnet and calorie) is a magneto-thermodynamic phenomenon in which a temperature change of a suitable material is caused by exposing the material to a changing magnetic field. This is also known by low temperature physicists as adiabatic demagnetization. In that part of the refrigeration process, a decrease in the strength of an externally applied magnetic field allows the magnetic domains of a magnetocaloric material to become disoriented from the magnetic field by the agitating action of the thermal energy (phonons) present in the material. If the material is isolated so that no energy is allowed to (re)migrate into the material during this time, (i.e., an adiabatic process) the temperature drops as the domains absorb the thermal energy to perform their reorientation. The randomization of the domains occurs in a similar fashion to the randomization at the curie temperature of a ferromagnetic material, except that magnetic dipoles overcome a decreasing external magnetic field while energy remains constant, instead of magnetic domains being disrupted from internal ferromagnetism as energy is added.

Subsequently, the magneto caloric effect can be studied with magnetic neutron scattering. The neutron has a net electric charge of zero, but has a significant magnetic moment. It is large enough to scatter from local magnetic fields inside condensed matter, providing a weakly

interacting and hence penetrating probe of ordered magnetic structures and electron spin fluctuations [13]. It can be used to unravel the magnetic spin ordering e.g. complete disorder to an ferromagnetic order and vice versa - thermo-diffractograms with powder is a very helpful tool. It has been used especially to determine the structural phase transition which should occur close to the actual operating temperature.

Following a powder diffraction measurement, a Pair-distribution function (PDF) analysis might become increasingly important in studying these materials since it provides local structure, interatomic distances, bond-angles and coordination numbers in disordered materials such as glassy and amorphous materials (for examples see [14,15]). The essential difference between conventional ND and PDF (linked to the neutron total-scattering experiment) is that while in ND only the Bragg peaks are considered, neutron total scattering means that also the diffuse, weaker, scattering present between the Bragg peaks is analyzed where deviations from the average can be measured. It is this "extra" scattering that provides information about the structure on a local scale and is therefore of high importance for structural studies when the material is not fully periodic. The PDF or $G(r)$ is obtained from the structure factor $S(Q)$ with a Fourier transformation,

$$G(r) = \frac{2}{\pi} \int_0^\infty S(Q-1)(Qr) \, Q \, dQ \qquad (1)$$

It is the analysis of this quantity that gives information about the local structure of the material. The Fourier transform requires data over a large Q range to avoid truncation effects, so the technique typically uses rather short-wavelengths.

Applications are mainly related to magnetic refrigeration which is a cooling technology based on the magnetocaloric effect. This technique can be used to attain extremely low temperatures, as well as the ranges used in common refrigerator. Magnetic refrigeration was the first method developed for cooling below about 0.3K (a temperature attainable by He refrigeration, that is pumping on the 3He vapors). Moreover, magneto calorics play a minor role in energy related applications and is therefore not subject of extensive research compared to PV or ion - batteries.

To summarize, magnetic scattering which interacts with the spin of the neutron is an excellent tool to decipher the intrinsic properties related to structural changes of the magneto-caloric effect.

2.5   Ion Batteries

From all ion batteries the Lithium ion battery is the most common form and their performance for most applications is best because Li-ion battery technology provides the highest energy densities of commercialized battery-technologies and has found widespread use in portable electronic applications A Li-ion battery is one type of a rechargeable battery or accumulator. It stores electrical energy. The battery mainly consists of two electrodes (anode and cathode), Lithium, electrolyte and a separator (see Fig. 8).
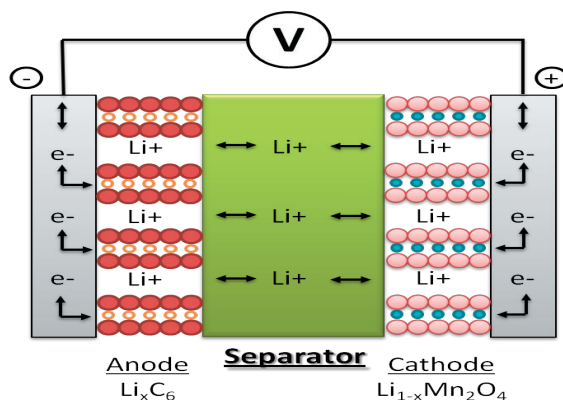
Fig. 8: Schematic drawing of the operating principle of a Li-ion battery.

When discharging, the Li-ions move from the negative electrode to the positive electrode and vice versa when charging. Li-ion batteries use often Li in a compound with a high degree of dissociation so the Li-ion can transport the charge and travel through the separator unhindered. Their main application are portable electronic devices with a high energy density, small memory effect for long lifetime and low self-discharge.

Application of Li-ion batteries in electrical vehicles and as static storage media is emerging, however, improved performance and reduced cost, combined with safety enhancements, are required. This has initiated a worldwide research effort for Li-ion electrode and electrolyte materials that combine desirable properties such as high energy and power density, low cost, high abundance of component elements, and electrochemical stability (Fig. 9). Also there is a growing interest of Li as electric charge carrier in batteries for the military, battery electric vehicles and aerospace. High power requires Li-ions to be highly mobile throughout the electrolyte [16].

Since a lot of these aspect can be solved with engineering approaches without understanding the underlying processes on a microscopic level, neutron scattering was lately considered to aid in this regard and studying this aspect could be rewarding when using neutron scattering [1].

The work and progress in the discovery, understanding and development depends on the available technique which needs to be again appropriate in time and length scale to gain insight into the role of Li-ion battery electrolytes. The role of neutron scattering is again strongly related to sensitivity to Li compared to X-rays and electrons. The coherent cross section of Li often allows the determination of the Li positions atomic displacement parameters with common diffraction techniques. In addition, the relative large coherent neutron scattering cross section of Li gives contrast that Li distributions can be studied using imaging and, in thin-films which can be studied with neutron reflectometry as well. Also the incoherent cross section allows the study of Li mobility. However, to measure Li diffusion is very difficult since the cross sections in a mixture with a possible electrolyte depend individually on the atoms therein.
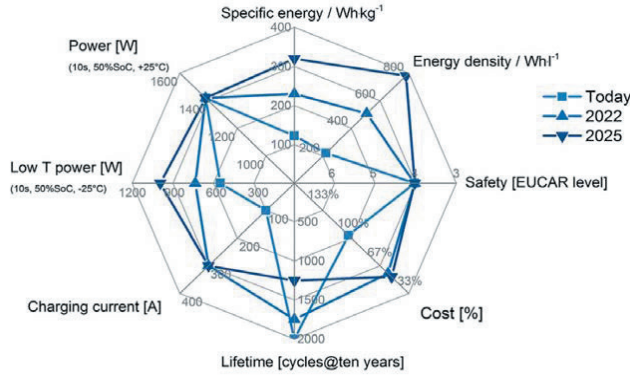
Fig. 9: Relevant parameters to build modern, powerful batteries for a wide variety of applications. Note: the different blue areas denote the status and future foreseeable progress of the specific features/ characteristics for Li-ion batteries [16].

As a detour, one is able to measure the dynamics of anions of the involved Li salt after dissociation or the hydrogen group in the electrolyte with QENS or INS. It will indirectly provide information about the Li dynamics. This new approach is the combination of the measurement of the Li dynamics and comparison of data with computer simulations. Even if the Li-ion dynamics is invisible with experimental techniques like neutron scattering the electrolyte are and those are highly responsible to boost the Li-ion mobility and enlighten the conducting mechanism. The picture underlying any computer simulation involves the motion of the Li-ions and results in a dynamic structure factor $S^{sim}(Q, \omega)$. Just when comparing to the experimentally determined dynamic structure factor $S^{exp}(Q, \omega)$ and receiving reasonable agreement the underlying microscopic picture for the simulation is confirmed and the Li-ion dynamic is resolved [17]. This on the other side is a necessity to improve Li-ion mobility in batteries for faster charging and discharging processes.

Note, naturally occurring Li is composed of 7.5 % Li[6] and 92.5 % Li[7]. The larger coherent neutron-scattering cross section and lower absorption cross-section of Li[7] make it possible to improve data quality by tailoring compositionally (isotopically) samples but depends also on the neutron-scattering technique. For example backscattering and time-of-flight spectrometer prefer incoherent and neutron-spin echo spectrometer prefer to measure coherent signals.

However, one prominent example are Li batteries (Li salt is called LiTFSi - Lithium Bis(Trifluoromethanesulfonyl)Imide) with a polymer electrolytes (Polyethylene oxid - PEO). This blend was studied combining neutron-spin echo (NSE) and dielectric spectroscopy with molecular dynamics (MD) computer simulations [18, 19].
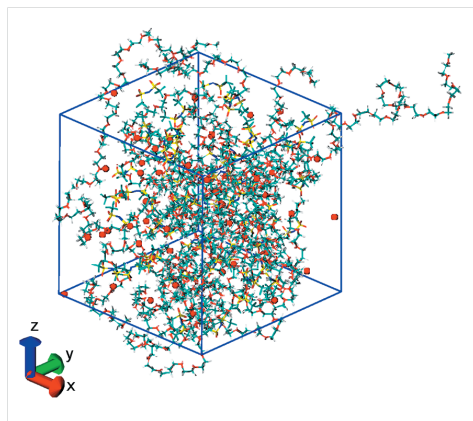
Fig. 10: Illustration of the Li-ion PEO electrolyte system. The PEO polymer chain is symbolized as greenish red chain and the Li-ions are the red dots in the box surrounding the polymer chain [19].

For the computer simulation a model was created in which the polymer electrolyte and Li ions are inside a box (Fig. 10). The box size determines the number of atoms and as a consequence the computation time. Note, before a simulation can be started the system has to be in an thermal equilibrium which can be achieved with preceding calculations at a much higher temperature when everything is thermally more activated.
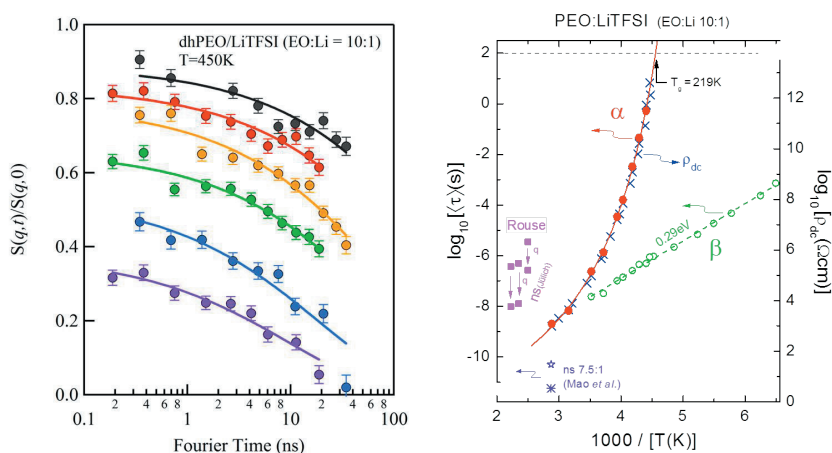
Fig. 11: Left picture: Intermediate scattering function $S(Q,\omega)$ calculated with molecular dynamic simulations (solid line) and measured with neutron-spin echo technique (circles). Right: Arrheniusdiagram of polyethylene oxid (PEO) polymer and Li-ions. Both were studied with and without LiTFSi salt [19].

The intermediate scattering function $S(Q, \tau)$ was calculated with molecular dynamic simulations and measured with neutron-spin echo technique and the results are shown in the left plot of Fig. 11. An agreement was achieved between both techniques and consequently the simulation data described the measurement well. Based on the labelling of the neutron-spin echo measurements one can assume, the backbone dynamics is observed in the experiment. Intermediate scattering function $S(Q, \omega)$ was calculated with molecular dynamic simulations and measured with neutron-spin echo technique and the results are shown in Fig. 11 [19].

The main results are summarized in a relaxation time map (right plot of Fig. 11) covering wide ranges of temperature and time. The temperature dependence of the DC conductivity and the dielectric α relaxation time is found to be identical which indicates a strong coupling between both. The relaxation times obtained from the NSE measurements at 0.05 Å^{-1} < q < 0.2 Å^{-1} are of similar magnitude as the relaxation time of Lithium predicted by MD simulation.

Furthermore the results suggest that the characteristic live times of the ions within the oxygen cages are mainly determined by the more local α relaxation that corresponds to local segmental motions of polymers, to a much lesser extent by the main chain relaxation, and not at all by the β relaxation or other faster processes. It is the first time decisive experimental evidence for a microscopic picture of the Li-ion transportation process is shown in which the PEO chain forms EO cages over several monomer units and the Li-ion ''jump'' from cage to cage. In this experiment and simulation even the characteristic jump time of about 3nm and decay times of the cages with appr. 5ns were determined. It would be difficult to achieve any comparable information with other techniques.

In this example, QENS and computer simulation were able to indirectly unravel the Li-ion transport process in this polymer matrix which in return might improve the ability to produce better Li batteries in the future.

Summary

While neutron scattering is an established technique to understand the properties in condensed matter, it became clear, it is exceptionally-well suited to study the transport of energy and charge-carrying molecules and ions which are strongly related to their dynamics in energy material applications. To study the structure of energy material is a matter of course. Species of interest in leading energy-technologies include H_2, H^+, and Li^+ which have particularly advantageous neutron-scattering properties that render these techniques of analysis ideal for such studies and consequently, neutron-based analysis is common-place for hydrogen storage, fuel-cell and battery materials etc. Research to understand the functionality of solar cell etc. materials rely on unique aspects of neutron scattering as well. Also, it is of utmost importance to tailor the neutron scattering experiments exactly to the question in demand and often a combination of several techniques unravels the full picture in which neutron scattering plays a major role.

References

[1] G. J. Kearley and V. K. *Peterson*, "*Neutron Applications* in *Materials for Energy",* Springer International Publishing (2015).

[2] S. W. Lovesey, "*Theory of neutron scattering from condensed matter*", Clarendon Press, Bd 1 (1984).

[3] Photovoltaic Effect Archived 14 July 2011 at the Wayback Machine.. Mrsolar.com. (2010).

[4] M. Z. Jacobson,"*Review of Solutions to Global Warming, Air Pollution and Energy Security*", Energy & Environmental Science. **2** (2), 148–173 (2009).

[5] "*BP Solar to Expand Its Solar Cell Plants in Spain and India"*, Archived, 26 September 2007 at the Wayback Machine.. Renewableenergyaccess.com. (2007). Retrieved on 3 June 2012.

[6] Luque, Antonio & Hegedus, Steven. *Handbook of Photovoltaic Science and Engineering*. John Wiley and Sons. (2003*).* ISBN 978-0-471-49196-5.

[7] O.-E. Haas et al., *"Surface Self-Diffusion and Mean Displacement of Hydrogen on Graphite and PEM Fuel Cell Catalyst Support"*, J. Phys. Chem. C **113**, 20281 – 20289 (2009).

[8] S. Li et al., "*$La_{1-x}Ba_{1+x}GaO_{4-x/2:}$ a novel high temperature proton conductor*", Chem. Comm. **21***,* 2694–2695 (2003).

[9] F. Schönberger et al., "*Investigation of proton conduction in $La_{1-x}Ba_{1+x}GaO_{4-x/2}$ and $La_{1-x}Sr_{2+x}GaO_{5-x/2}$*", Solid State Ion. **176**, 2951–2953 (2005).

[10] N. Jalarvo et al., "*Atomic Scale Picture of the Ion Conduction Mechanism in a Tetrahedral Network of Lanthanum Barium Gallate*", Chem. Mater. **25**, 2741 (2013).

[11] E. Kendrick et al., "*Cooperative mechanism of fast ion conduction in gallium based oxides with tetrahedral moieties*", Nat Mat. **6**, 871 (2007).

[12] P. Boillat et al., "*Neutron Imaging of Fuel Cells – Recent Trends and Future Prospects"*, *Current opinion in Electrochemistry*", **5** (1), 3-10 (2017).

[13] T. Chatterjii, Book 2006, "*Neutron Scattering from Magnetic Materials*", Elsevier Science, ISBN: 978-0-444-51050-1.

[14] O. Gourdon et al., ChemInform Abstract: "*Toward a Better Understanding of the Magnetocaloric Effect: An Experimental and Theoretical Study of $MnFe_4Si_3$"*, J. Sol. State. Chem. **216** (32) (2014).

[15] M. Gottschlich et al., "*Study of the antiferromagnetism of Mn$_5$Si$_3$: An inverse magnetocaloric effect material*", J. Mat. Chem. **22** (30), 15275-15284 (2012).

[16] *A.* Eftekhari et al., *"Lithium-Ion Batteries with High Rate Capabilities".* ACS Sustainable Chemistry & Engineering. **5** (3), 2799–2816 (2017).

[17] D. Andre, et al., "*Future Generations of cathode materials: an automotive industry perspective*", J. Mater. Chem. **A3**, 6709 (2015).

[18] C. Jafta et al., "*Ion Dynamics in Ionic-Liquid-Based Li-Ion Electrolytes Investigated by Neutron Scattering and Dielectric Spectroscopy*", ChemSusChem **11**, 3512 –3523 (2018).

[19] C. Do et al., "*Li$^+$ Transport in Poly(Ethylene Oxide) Based Electrolytes: Neutron Scattering, Dielectric Spectroscopy, and Molecular Dynamics Simulations*", Phys. Rev. Lett. **111**, 018301 (2013).

E 7   Engineering Materials

Martin Müller

Materials Physics

Institute of Materials Research

Helmholtz-Zentrum Geesthacht

Abstract

Abstract

The last two decades have seen a strong and steady increase of the use of synchrotron radiation and neutrons in engineering materials science. The application of small-angle scattering has been at the starting point of this success story, in particular using neutrons. Their specific contrast is well suited to e. g. observe precipitate formation in alloys. The non-destructive measurements of residual stresses and orientation distribution functions (texture) using diffraction have also become standard techniques. Radiography and tomography are complementary to scattering methods and are also widely used. – The recently updated 2nd edition of a topical book [1] may serve as a reference here.

Engineering materials science profits very much from the possibility to carry out *in situ* and *in operando* experiments, combining scattering with real-time investigations of materials processing. Here, particularly high-energy X-rays available at synchrotron radiation sources allow for unprecedented spatial and temporal resolution.

In my lecture, I will address the abovementioned methods in view of their specific use for the investigation of engineering materials science. Their use is then illustrated by a number of examples, ranging from residual stress measurements in aircraft components over high-temperature materials development to *in situ* laser beam welding.

Reference

[1] P. Staron, A. Schreyer, H. Clemens, S. Mayer (Eds.), *Neutrons and Synchrotron Radiation in Engineering Materials Science* (Wiley-VCH, Weinheim, 2nd Ed. 2017).

E 8 Protein Crystallography and Protein Dynamics

R. Biehl and T. E. Schrader

Jülich Centre for Neutron Science & Institute of Complex Systems

Forschungszentrum Jülich GmbH

Contents

Introduction

Apart from water proteins are the most abundant molecules in living cells. They are constantly synthesized in the cell by the transcription and translation mechanism where first the DNA is read out to produce a messenger RNA which encodes the proteins and in the subsequent translation process the protein is synthesized by the ribosomes which itself is a protein/RNA-complex. Proteins fulfil numerous functions in the cell, for example enzymatic catalysis which enhances the speed with which molecules like fatty acids are synthesized or transport and storage of important molecules like oxygen or they are involved in immunology, just to name a few [1]. In order to perform these functions, proteins adopt a unique three-dimensional structure with a carefully controlled mixture of flexibility and stiffness. For an understanding of their function knowledge of this three-dimensional structure is a prerequisite.

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 [2]. 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.

Ideally one would like to produce a movie where one can follow the functioning protein in action in slow motion with atomic resolution. In practice, there are techniques available which have a sufficient time resolution but do only provide very limited structural information as for example time resolved infrared spectroscopy. On the other hand, there are methods which provide full atomic resolution but with limited time resolution. With those methods one often stops the enzymatic process of a protein in an intermediate state by using inhibitor molecules which trap the catalytic process of the protein in a certain intermediate state.

The first part of the article will focus on the latter static, elastic scattering techniques among which x-ray protein crystallography is the most widely used one. Here, there are some promising attempts to do time resolved x-ray scattering to have both structural resolution at the atomic level and time resolution [3]. But this technique of time resolved x-ray scattering is often limited to certain systems which undergo a photo-initiated process. This part will also introduce the method of neutron protein crystallography since there are some similarities but also some differences to x-rays as probes. Finally, an example case study is discussed where these both techniques give complimentary information.

The second part of the article illustrates the power of inelastic neutron scattering for the elucidation of equilibrium motions of proteins in solution. 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 during the last 10 years the state-of-the-art involves ensemble docking. This part shows examples how protein internal dynamics can be measured.

1 Some Basics about Proteins

1.1 Amino acids as the building blocks of proteins

Proteins consist of a chain of amino acids. In that sense they are biopolymers and since they in general have charged side chains they can also be called polyelectrolytes. The first information about a protein is therefore the number of amino acid residues it contains. This number can span quite a wide range between 10 and 25 000. Typically, a protein contains 100 amino acids. The average molecular weight per amino acid is around 100 g/mol. Despite there are many different amino acids (or to be more exact: 2-amino carboxylic acids) present in living organisms not all of them are used to build up proteins. The proteinogenic amino acids are shown in Figure 1. Covalently attached to the central C-atom, the so-called C_α atom, are an amino functional group, a carboxylic group, the side chain atoms and finally one hydrogen atom. Since the C_α atom has a sp^3 hybridization sterically all four constituents point into the corners of a tetrahedron. Since (apart from the amino acid glycine) this C_α atom has four different constituents it forms a chiral center. So, two different arrangements of these four constituents are possible which lead to the L- and D-enantiomers of the corresponding amino acid, according to Fischer's convention. But in nature only the L-enantiomers are found as building blocks of proteins.
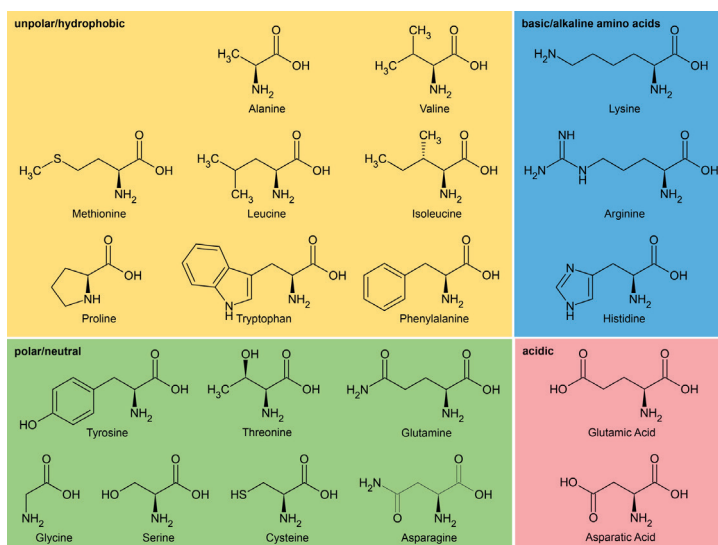
Figure 1 A compilation of all 20 amino acids found in natural proteins. The N-terminal amino group is shown in its neutral charge state pointing to the bottom of the page. Covalently attached to it is the C_α atom which carries the corresponding side chain group (Adapted from http://upload.wikimedia.org/wikipedia/commons/7/7d/Overview_proteinogenic_amino_acids-DE.sv).

The sequence of amino acids is the **primary structure** of the protein. It can be displayed as a line of text with a one letter code representing one amino acid. It is a common convention that the line of text starts at the N-terminal end of the amino acid chain i. e. with the amino acid with side chain R_1 in the example given in Figure 2 on the left.

A special feature of the peptide bond is its partial double bond character of the chemical bond between the carbon and nitrogen atom. The lone electron pair at the nitrogen atom is delocalized and has some existence probability between the atoms forming the peptide bond. This causes the bond length to shrink below the value of a single bond. As a consequence, the
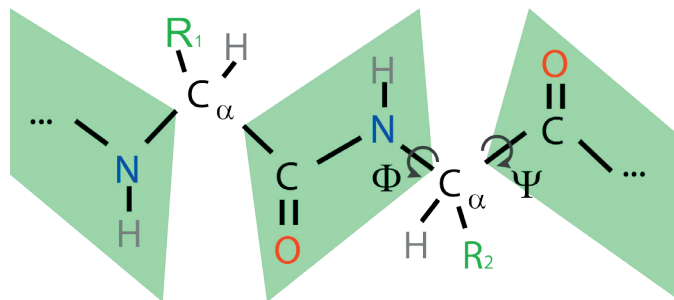
Figure 2 Due to a partial double bond character of the C-N peptide bond the rotation around it is hindered by a steep potential. This causes the four atoms OCNH to form a planar structure (marked in green).

rotation around the CN-bond is hindered and the four atoms OCNH form a planar geometry denoted by a green polygon in Figure 2. This is why one amino acid only contributes two degrees of freedom to the amino acid backbone which are denoted by the dihedral angles Φ and Ψ, for their definition see Figure 2.

1.2   The three-dimensional structure of proteins

This structure is stabilized by four different interactions. First of all, there is the possibility of establishing **hydrogen bonds** between two parts of the backbone, but also between side chains or between a side chain and a part of the backbone. Another stabilizing mechanism is a formed **salt bridge** between a negatively and a positively charged side chain, e. g. aspartic acid and lysine. The third interaction is the formation of a **hydrophobic cluster** or core. Hereby the surrounding water plays a major role which makes it a mostly entropic effect. It is more favorable for the water molecules to form hydrogen bonds with each other than to stick between some hydrophobic side chains. This is why those side chains tend to be packed together in the folding process resulting in van der Waals interactions among them. The fourth stabilizing moment of a three-dimensional fold of a protein is a **disulfide bridge** between two cysteine residues.

When the primary structure only gives the linear sequence of amino acids, the **secondary structure** of a protein denotes all arrangements of the protein backbone stabilized by a regular hydrogen bonding pattern. These hydrogen bonds are solely between different parts of the backbone. There are several structural motifs of that kind which occur frequently in proteins. Some of these motifs have been given a name e.g. α-helix or parallel β-sheet. The **tertiary structure** of a protein denotes the three-dimensional arrangement of all atoms of the protein in space, including the side chains. This information can be obtained by structural techniques as for example protein crystallography (will be discussed below), or solution NMR. Some proteins need more than one amino acid chain to be functional. The arrangement of the different amino acid chains is then referred to as the protein's **quaternary structure**.

2   Crystallography

2.1   The Physics of Protein Crystallography

The following chapter will show how most of this structural information has been obtained. For both techniques x-ray and neutron protein crystallography a single crystal of the protein of interest is required since the scattering of one protein molecule is very weak. So, in general a crystal has to be grown, especially large ones in case of neutron crystallography since the neutron luminosity of modern sources is much smaller than the x-ray flux reached by synchrotron sources.
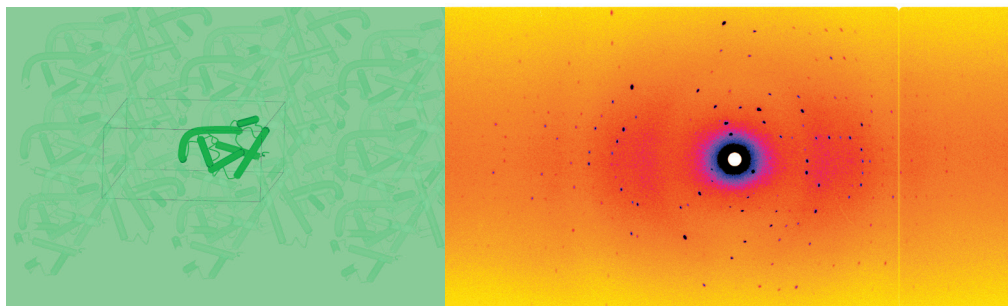
Figure 3 Real space arrangement of myoglobin molecules in a crystal of space group P2₁ (on the left) versus diffraction pattern (right) of a myoglobin crystal.

The crystal then serves as a noiseless amplifier of the diffraction signal. But the arrangement of proteins in a crystal brings in another advantage, since the orientational averaging can be avoided, which is always present in the solution phase. Figure 3, left hand side shows the regular arrangement of myoglobin molecules in a crystal lattice. The unit cell of the monoclinic lattice (space group P2₁) is indicated by black lines. It bears two myoglobin molecules in one unit cell. The picture on the right shows a diffraction pattern recorded with the instrument BioDiff on a myoglobin crystal.

An X-ray macromolecular crystallography beamline

Synchrotron beamlines provide extremely high photon flux for x-ray crystallography. Due to the high demand from the structural biology community, often more than one macromolecular crystallography beamline is operated at a synchrotron. Those beamlines are optimized for special wavelengths and focal diameters. As an example, the beamline BL14.2 has a beam size at the sample position of 150 μm x 100 μm (FWHM). Its Rayonix MX-225 detector has a pixel size of 37 μm. Without on chip binning one frame amounts to 6144 x 6144 pixels. The exposure time per frame is typically between 3 to 10 seconds. Typically, the sample crystals are kept at temperature as low as 100 Kto avoid radiation damages. To record a full data set takes about 10-30 minutes. The largest diagonal of a typical protein crystal ranges between 10 to 500 μm

A neutron macromolecular crystallography instrument

Since x-rays are scattered from the electrons in the crystal and neutrons from the nuclei, hydrogen atoms are hardly seen in x-ray crystallography experiments. Only at very high resolutions of 1 Å or less there is a chance to observe hydrogen atom positions. This resolution is often not within reach because of the quality of protein crystals. Here neutron protein crystallography must be employed to retrieve the hydrogen atom positions. Moreover, neutron scattering can distinguish between different isotopes, especially between hydrogen and deuterium. Whereas from x-ray crystallography the electron density in the unit cell of the crystal can be calculated, neutron protein crystallography yields the nuclear scattering length density, which is a signed quantity.

A major drawback of the method neutron protein crystallography is the required crystal size. Due to the much smaller neutron flux as compared to typical x-ray fluxes the crystals required for a neutron crystallography study must be much larger as compared to x-ray crystallography. Here, often crystal diagonals of 1 mm and more have to be reached.

As an example of a neutron diffraction instrument optimized for protein crystallography the instrument BioDiff at the FRM II shall be introduced. It is a collaboration between the Forschungszentrum Jülich (FZJ) and the Forschungs-Neutronenquelle Heinz Maier-Leibnitz (FRM II).
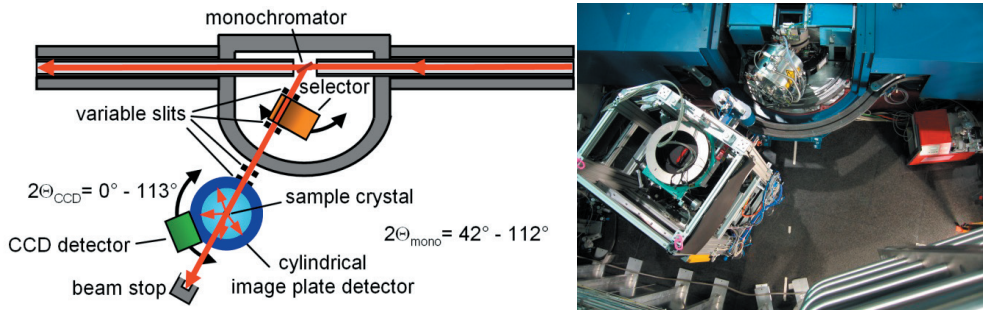
Figure 4 Schematic view of the BioDiff instrument (left) and a picture taken from a similar view point with the biological shielding removed (right).

Figure 4 shows a schematic top view of the instrument and a corresponding picture when the biological shielding has been removed. The neutron beam from the cold neutron source of the FRM II reactor enters from the right. By Bragg reflection from a pyrolytic graphite crystal (002-reflex) neutrons are taken out from the white neutron spectrum and pass a variable slit and a velocity selector. The velocity selector acts as a $\lambda/2$ filter to remove higher orders. Together with the pyrolytic graphite crystal it forms a monochromator with a $\Delta\lambda/\lambda$ of ca. 2.5 %. Behind the velocity selector the beam passes a second variable slit and the main instrument shutter. Before entering the detector drum of the image plate detector through a Zirconium window a collimator made out of boron carbide apertures with fixed diameters between 3 mm and 5 mm shape the beam to fit to the sample size. At present the sample is usually contained in a quartz glass tube. It is fixed to a standard goniometer which is mounted upside-down from the sample stage on top of the instrument. After passing the sample the main neutron beam exits the detector drum through a second Zirconium window and hits finally the beam stop. The cylindrical image plate detector can be read out with three different resolutions of 125 μm, 250 μm and 500 μm. As an alternative, one can lower the image plate detector and swing in a neutron sensitive scintillator which is imaged onto a CCD-chip. This CCD-camera set up is used for a fast alignment of the sample crystal with respect to the neutron beam.

The diffraction pattern shown in Figure 3, right hand side has been recorded with the image plate detector. In fact, a complete crystallographic data set on a typical protein crystal takes about 10 days to record and involves about 300 detector images. In between two detector images the crystal is rotated with the goniometer by 0.3° typically. Of course, the required time to record a data set is much longer as in case of x-ray diffraction.

Some mathematical aspects of diffraction by a protein crystal

Having recorded a complete data set on a crystal some data treatment is necessary in order to calculate meaningful atom positions. Here, only a brief introduction can be given. More details can be found in text books [4,5].

Assuming a number of n atoms per unit cell the structure factor of a single unit cell can be written as:

$$F(\mathbf{S}) = \sum_{j=1}^{n} f_j \exp(2\pi i \mathbf{r}_j \mathbf{S}) \tag{1}$$

Here $\mathbf{r}_j$ denote the atom position of atom j and $\mathbf{S}$ is the scattering vector perpendicular to the plane which reflects the incident beam. In the previous lectures the scattering vector was defined. It relates to $\mathbf{S}$ with the following relation: $\mathbf{Q} = 2\pi\mathbf{S}$. In crystallography it is just more convenient to use $\mathbf{S}$ instead of . f_j can be seen here either as the scattering length of atom j

in the neutron scattering case or the atomic scattering factor in case of x-ray diffraction. One can generalize this approach by switching form the summation to an integration to yield:

$$F(\mathbf{S}) = \int\limits_{unitcell} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r}\mathbf{S}) d^3\mathbf{r} \tag{2}$$

where $\rho(\mathbf{r})$ is the electron density distribution or the scattering length density respectively.

Since a crystal consists of AxBxC unit cells, the structure factor of the crystal can be composed as

$$F_{cryst.}(\mathbf{S}) = F(\mathbf{S}) \sum_{u=0}^{A} \exp(2\pi i u \mathbf{a}\mathbf{S}) \sum_{v=0}^{B} \exp(2\pi i v \mathbf{b}\mathbf{S}) \sum_{w=0}^{C} \exp(2\pi i w \mathbf{c}\mathbf{S}) \tag{3}$$

The vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ denote basis vectors of the unit cell. For an increasing number of unit cells, the sums can be represented by delta functions leading to the Laue conditions for the structure factor being non-zero:

$$\mathbf{a}\mathbf{S} = h, \mathbf{b}\mathbf{S} = k, \mathbf{c}\mathbf{S} = l \tag{4}$$

This means that one only gets constructive interference, when the scattering vector is perpendicular to planes in the crystal which can be denoted by the index vector $\mathbf{h} = hkl$. For this reason, the diffraction pattern of a single crystals shows distinct peaks, the so-called Bragg peaks. The Bragg law can be easily derived from equation 4. Figure 5 shows the Ewald sphere construction. It is a tool to construct the direction of the diffracted beam. The Ewald sphere has its origin at the position of the crystal. Its radius is the reciprocal wavelength used in the scattering experiment. The origin of the reciprocal lattice is placed at the intersection of the sphere with the incident beam direction. Whenever the orientation of the reciprocal lattice is such that another point of the reciprocal lattice lies on the Ewald sphere a diffracted beam results in the direction of line running from the center of the Ewald sphere through that point.
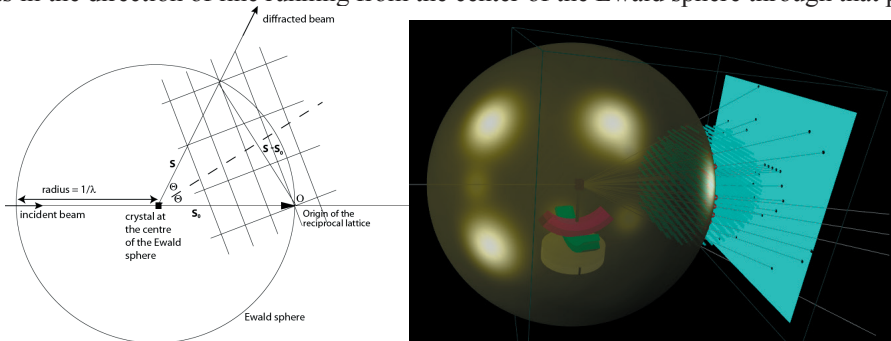
Figure 5 Ewald sphere: On the left the schematic shows how to construct the Ewald sphere. On the right an Ewald sphere (golden color) construction is shown in three dimensions. The blue square represents a flat two-dimensional detector.

When the crystal is rotated the reciprocal space rotates with it resulting in other lattice points to cause diffracted beams. In practice the incident beam is not strictly monochromatic but has a wavelength distribution which causes the Ewald sphere to be elongated to form a spherical shell of a certain thickness. This increases the number of diffracted beams observed. The beam divergence adds also to its thickness.

So, the positions of the diffracted beams on the detector only depend on the reciprocal lattice. The structure of the protein inside the unit cell is encoded in the amplitude and phase of the structure factor.

To obtain the electron density or the nuclear scattering length density one has to perform the inverse Fourier transformation:

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h}} F(\mathbf{h}) \exp(-2\pi i \mathbf{r} \mathbf{h}) \tag{5}$$

Here V is the volume of the unit cell. Unfortunately, only the modulus squared of the structure factor is measured as intensity on the detector. The phase information is lost which is known as the phase problem of crystallography.

There are several solutions to the phase problem which are predominantly applicable for the x-ray diffraction case:

a. isomorphous replacement: Several crystals of the same crystal structure have to be available for this method. First a crystallographic data set is recorded on an untreated crystal. Then crystals are soaked in at least two different heavy atom salt solutions. In the best case, the different heavy atom ions occupy different regular positions in the unit cell. From these (at least) two crystallographic data sets recorded on the heavy atom treated crystals phase information can be retrieved which is then used to determine the phases of the data set of the untreated crystal.

b. anomalous dispersion: Often it is possible to replace one distinct methionine amino acid with an artificial selenomethionine one. The selenium atom has a suitable absorption edge on which anomalous scattering can be performed by tuning the wavelength of the beamline to the anomalous regime. Crystallographic data sets are then recorded at different wavelengths from which the phase information can be calculated. In some cases, this approach can also be adopted for sulfur atoms present in naturally occurring cysteine residues.

c. molecular replacement: From the primary structure one can search the protein data bank (pdb) for proteins with a similar amino acid sequence. If one finds enough fragments which seem to be sufficiently homologous to the unknown structure one can use those fragments for the calculation of initial phases. In further refinement steps these phases can be improved. Since the number of unique structures entered in the protein data bank is growing this method is increasingly favored over other methods.

The phase problem of the neutron data sets is solved by using the x-ray structure and the molecular replacement technique.

Model building and refinement

With the data treatment one has now arrived at a contour map $\rho(\mathbf{r})$ be it either a nuclear scattering length density or an electron density. Now the information on the primary structure of the protein is used and the backbone is coarsely fitted into the contour map either manually or employing software. From this model new amplitudes and phases of the structure factor are calculated using eq. 1. The modulus squared of the structure factor is again compared with the data. One could now think of a least square based fitting procedure to find the optimum arrangement of the protein atoms in the unit cell. In practice however, maximum likelihood and simulated annealing molecular dynamics simulations are used because those are superior to the least square approach in terms of overcoming local minima. In these molecular dynamics simulations, a lot of stereochemical information is used as restraints for example known bond lengths of single bonds between two carbon atoms or bond angles. The agreement between model and observed contour map is often measured by calculating a so-called R-factor. A good agreement between model and measured data leads to an R-factor of about 0.2. R-factors of 0.5 and above are indicative for a random or no agreement between model and data.

2.2 An example case study: Fighting Meticillin resistant bacteria

This example also shows the nice interplay between x-ray and neutron crystallography. It is chosen because of the relevance it has to our everyday life. The protein involved here is the β-lactamase. It is produced by bacteria and partly secreted to their surrounding environment in order to split a certain bond in the four-membered ring (=β-lactam motif) of a series of very common antibiotics. This is one of the mechanisms which render such bacteria resistant against this type of antibiotics. The antibiotic drug called meticillin also bears such a four membered ring. Meticillin resistant *Staphylococcus aureus* is assumed to possess this mechanism of resistance against β-lactam antibiotics. This is why it would be interesting to investigate the catalytic cycle of this protein in order to block it. This would result in the bacteria being affected by the antibiotics again.

The experiment was designed in a clever way. The catalytic cycle of the β-lactamase consists of an acylation step where a covalent adduct between protein and substrate (here the antibiotics molecule) is formed (Figure 6 top part). Here, the CO double bond is split into a single bond and a tetrahedral intermediate between substrate and protein is produced. The second part is a deacylation step where the four-membered ring (the so-called β-lactam motif) is split and the product is released. The product cannot act as an antibiotic any more due to its different chemical structure.
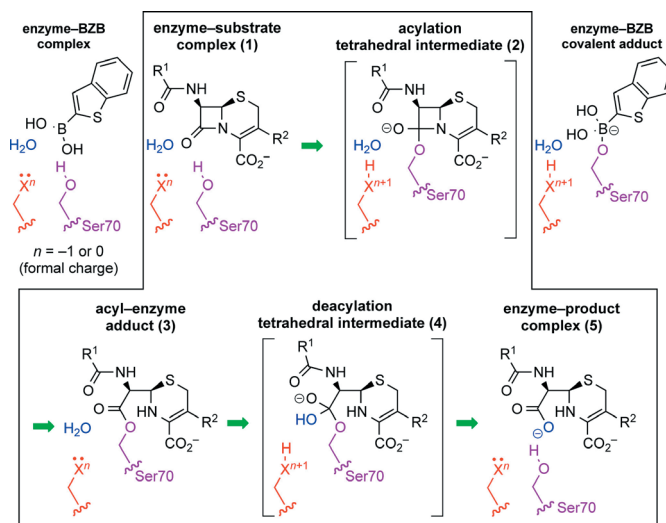
Figure 6 Catalytic cycle of a β -lactamase enzyme. (The Figure is taken from ref. [6]. This research was originally published in Journal of Biological Chemistry, Stephen J. Tomanicek et al. "Neutron and X-ray Crystal Structures of a Perdeuterated Enzyme Inhibitor Complex Reveal the Catalytic Proton Network of the Toho-1 β-Lactamase for the Acylation Reaction." © the American Society for Biochemistry and Molecular Biology.)

Due to its relevance to clinics the enzyme family of β-lactamases has been subjected to countless studies mostly by x-ray crystallography but the main question yet not addressed was the nature of the active base which takes the excess proton in the acylation step. It was not clear which part, which side chain of the protein takes over this role. But this was the key information for improving drugs to block this enzyme. Since a base is best detected when it takes a proton in the acylation step, the problem consisted of stopping the catalytic cycle in the acylation step and hunting for a proton which was not there in the ground state of the protein when no substrate was bound.
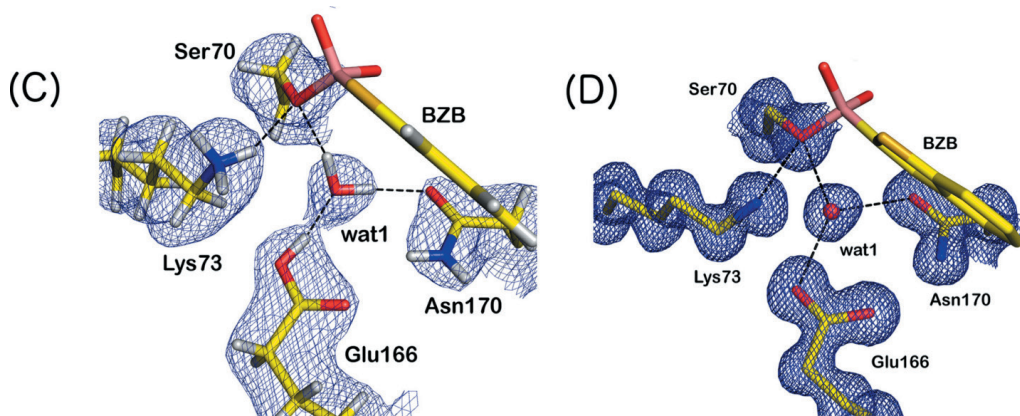
Figure 7 X-ray (right) and neutron (left) data on the BZB adduct in the catalytic cycle of a β-lactamase enzyme. (The Figure is taken from ref. [6]. This research was originally published in Journal of Biological Chemistry, Stephen J. Tomanicek et al. "Neutron and X-ray Crystal Structures of a Perdeuterated Enzyme Inhibitor Complex Reveal the Catalytic Proton Network of the Toho-1 β-Lactamase for the Acylation Reaction." © the American Society for Biochemistry and Molecular Biology.)

Here, a different ligand was used as the natural antibiotics: benzothiophene-2-boronic Acid (BZB). This ligand was known to stop the protein in the acylation step. In addition to that, this ligand was also known to mimic this covalent intermediate between substrate and protein. In addition to that, neutron protein crystallography was used to detect the proton unambiguously. For the latter technique the protein was expressed in deuterated media such that a fully deuterated protein resulted. The BZB was also synthesized using deuterium instead of hydrogen atoms. In fact, even a special Boron isotope was used which shows less neutron absorption than the natural abundant boron which is known as a good neutron absorber. Figure 7, left side shows the x-ray structure and Figure 7 right side shows the neutron structure. The proton is clearly seen in the neutron structure and therefore the side chain Glutamate 166 could be identified as the base in the acylation step. With this knowledge better drugs can be designed which bind this sidechain firmly blocking the protein's catalytic cycle in its first step. These additional drugs will break the resistance of those bacteria against common antibiotics.

3   Protein dynamics

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 [7]. 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 [8] 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. Including the dynamics of proteins as transitions in a configurational ensemble the model of "conformational selection" allows substrate binding in a specific configuration as reached due to the thermal motions.

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 [9]. The accessible configurational space can spread over a single deep minimum or a broader rugged valley where in both the local energy depth defines the occupancy of a configuration 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.

3.1   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 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.

Atomic vibration

The fastest movements with highest energy are atomic vibrations as they are common for any molecule. Figure 8a 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 [10]. 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. Figure 8c shows an example spectrum of $(PPG)_n$ an synthetic polypeptide with a similar structure to natural collagen [11]. 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 [12]. As shown in Figure 8b 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.
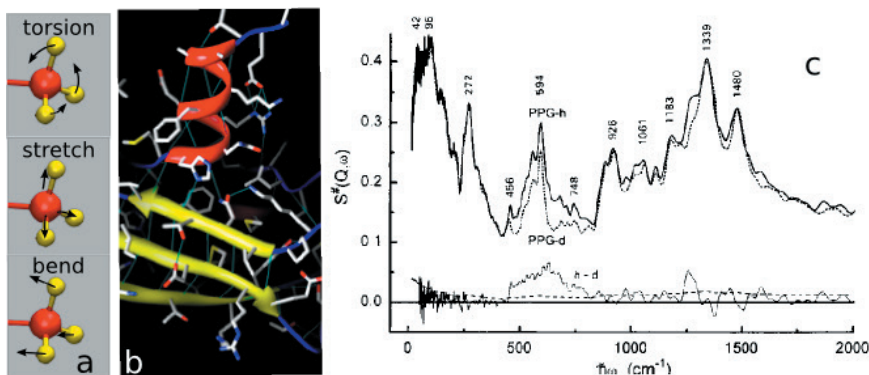
Figure 8 Example geometries of torsion, symmetric stretching or bending vibrations of a methyl group as the side group of alanine. b) Secondary structure elements β-sheets and α-helices (in yellow, red) with side chain elements. The secondary structure is stabilized by the hydrogen bonds (thin blue lines). 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. [11]

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 [13]. 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 [14]. Figure 9A 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.

Figure 9 B-E show 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.
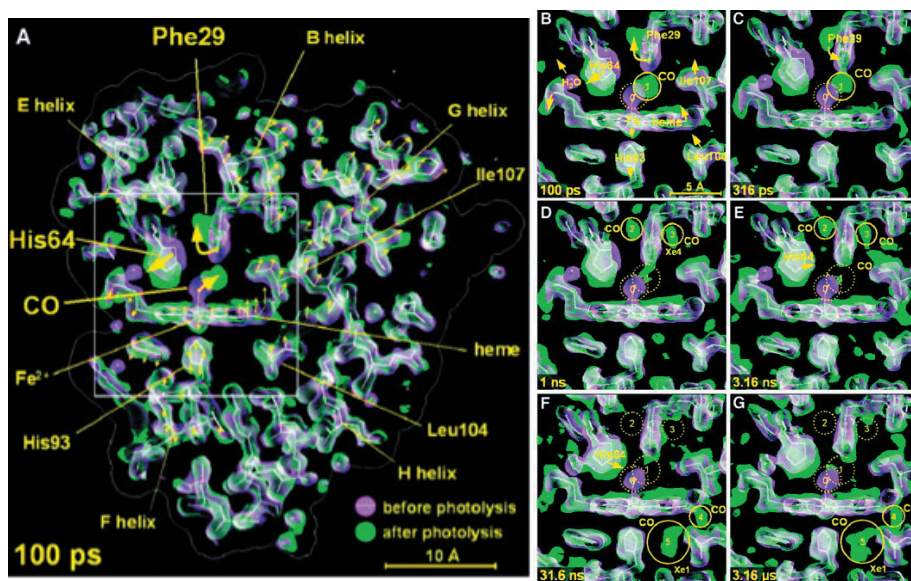
Figure 9: Electron density map of myoglobin before (magenta) and after photolysis (green). Overlapping densities are shown in white. The stick model included shows the unphotolyzed state. Large arrows indicate 3 large changes while small arrows indicate small rearrangements in the whole protein. Sequence 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 [14].

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.

3.2 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. SANS, SAXS 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 [15]. 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 [16]. 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.

Phosphoglycerate kinase as a classical hinge
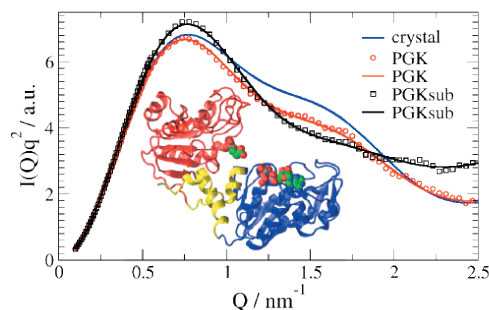
Figure 10 Form factor measured by SANS from PGK and PGKsub (Kratky plot: Q versus $Q^2 I(Q)$). Lines show the calculated form factor from the crystal structure (blue) and from structures deformed along the softest normal modes to fit the experimental data. The inset shows the protein with the hinge in yellow, the main domains in blue and red and the substrates as spheres. Figure from [20].

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 [17]. PGK is composed of two separated domains connected by a hinge region as shown in Figure 8 [18]. 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 [19]. 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.

Figure 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 [20]. 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.

NSE measurements show the relaxation of the intermediate scattering function dependent on diffusion and internal dynamics as shown in Figure 11 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.
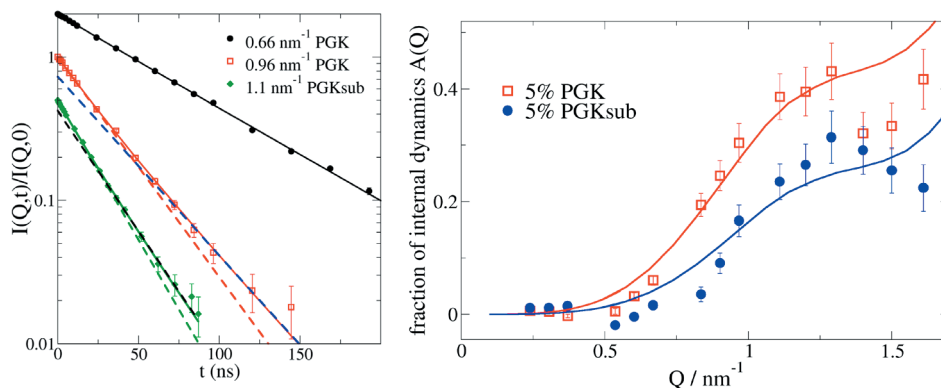
Figure 11 Left: Semi logarithmic plot of I(Q,t)/I(Q,t=0) for selected Q values (PGK, black and red; PGKsub, green; data are shifted for clarity and are equal 1 for t=0). Red and green dashed lines represent the initial slope extrapolated to long times. The blue and black dashed lines represent the long-time 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 allowing. Figure from [20].

The observed relaxation times are 60 ns for the substrate free and 45 ns for the substrate bound PGK. Figure 11 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.

Fast antibody fragment motion: flexible linkers act as entropic spring
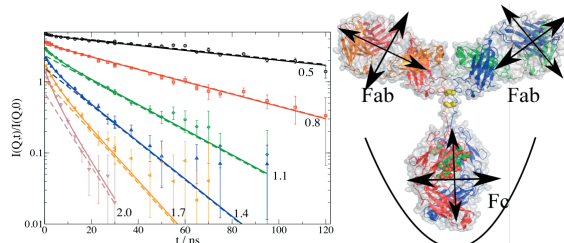
Intermediate scattering functions I(Q,t)/I(Q,0) of IgG. Dashed lines show translational and rotational diffusion and are close to a single exponential for times t > 15 ns. Solid lines include additional internal dynamics of each domain in its own harmonic potential with a 7ns relaxation time.

The fragment motion of IgG was examined by Stingaciu et al by NSE [21]. In the experiment, aside from strong diffusion contributions, a clear signature of the internal dynamics on a timescale of 6 - 7 ns was observed. The combination of translational and rotational diffusion of the rigid protein describes the observed long-time dynamics on an absolute scale, despite the fact that we have a mixed population of monomers and dimers, as found in serum.

The fragment motion shows itself in terms of a strong decay of $S(Q,t)$ at short times, well separated from the overall diffusional relaxation. The data were analyzed in terms of an Ornstein-Uhlenbeck like relaxation within a harmonic potential. Such an approach neglects the details of the complex linker region interaction on the residue level. Nevertheless, the dynamics is described in an excellent way. Even for a short flexible linker the ensemble average seems to be sufficient to produce the characteristics of a harmonic spring. The

obtained spring constant ($\approx 10 \, pN/nm$) appears to be realistic compared to an entropic spring of similar length. The resulting forces stay below the limits that would be necessary to unfold the secondary structure of the attached fragments. The observed effective friction appears to be close to the friction of a free unbound fragment in the solvent. None of the directions seems to be suppressed.

The pre-existing equilibrium hypothesis [22], with multiple local minima in the configurational energy landscape, will have long transition times between the minima but is compatible with fast 6-7 ns motions in the local minima. Consequently, NSE observes the fast dynamics in pre-existing equilibrium states. The conformational flexibility in a pre-existing equilibrium configuration might be needed to adapt faster to a specific antigen when the antigen approaches the binding site of a fragment.

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
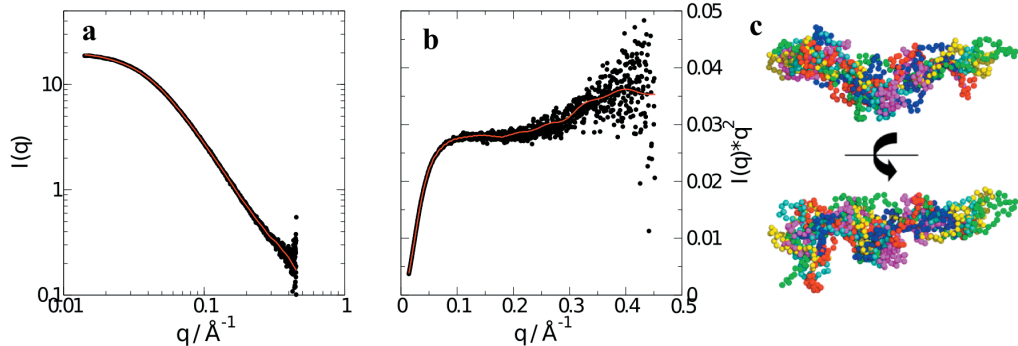
Figure 12 SAXS data from MBP at 4.5 mg/ml. a) The red solid lines is a fit with the Debye equation for a Gaussian chain. (b) Kratky plot. The line is a result of the scattering from the most probable conformational ensembles shown in c. c) Representative coarse-grained conformations of MBP as determined 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 [23].

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 [24]. 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) [25].

Figure 12 a+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 an ensemble was selected that represent the SAXS data [26]. The resulting characteristic molecular shapes are displayed in Figure 12c. The model conformations indicate an elongated structure with a relatively compact core and flexible ends on both sites.
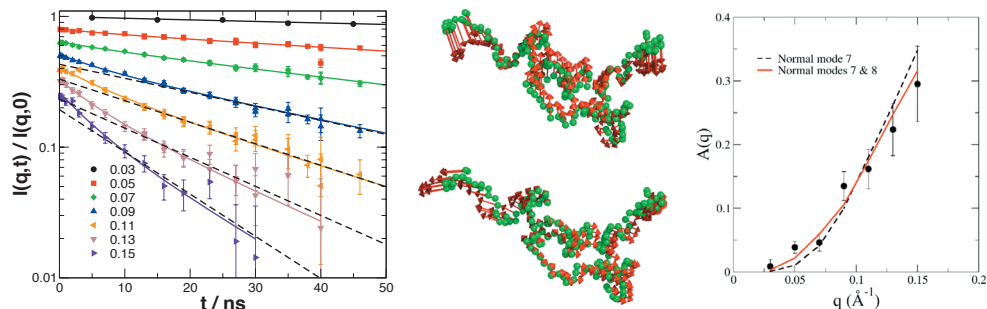
Figure 13 Left) NSE Data for MBP: All spectra start at unity but are shifted consecutively by a factor of 0.8 for clarity shown up to 50 ns. Solid lines are fits to the NSE data with the structural model. The dashed lines are exponential fits for t>20 ns to extrapolate the long-time rigid body dynamics. 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 [23].

Figure 13 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.

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 Figure 13 at the right. Normal mode analysis was used to describe the deformation of the structural models. Figure 13 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 Figure 12c 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.

Acknowledgement

References

[1] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemie* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

[2] R. D. Vale, Cell **112**, 467 (2003).

[3]   P. Fromme and J. C. Spence, Curr. Opin. Struct. Biol. **21**, 509 (2011).

[4]   J. Drenth, *Principles of Protein X-Ray Crystallography* (Springer-Verlag GmbH, 1999).

[5]   N. Niimura and A. Podjarny, *Neutron Protein Crystallography: Hydrogen, Protons, and Hydration in Bio-Macromolecules* (Oxford University Press, 2011).

[6]   S. J. Tomanicek, R. F. Standaert, K. L. Weiss, A. Ostermann, T. E. Schrader, J. D. Ng, and L. Coates, J. Biol. Chem. **288**, 4715 (2013).

[7]   E. Fischer, Berichte Der Dtsch. Chem. Gesellschaft **27**, 2985 (1894).

[8]   D. E. Koshland, Proc. Natl. Acad. Sci. U. S. A. **44**, 98 (1958).

[9]   J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).

[10]   A. Barth, Prog. Biophys. Mol. Biol. **74**, 141 (2000).

[11]   H. D. Middendorf, R. L. Hayward, S. F. Parker, J. Bradshaw, and a Miller, Biophys. J. **69**, 660 (1995).

[12]   H. Susi and D. M. Byler, *Enzyme Structure Part K* (Elsevier, 1986).

[13]   Z. Ren, D. Bourgeois, J. R. Helliwell, K. Moffat, V. Šrajer, and B. L. Stoddard, J. Synchrotron Radiat. **6**, 891 (1999).

[14]   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).

[15]   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).

[16]   R. Biehl and D. Richter, J. Phys. Condens. Matter **26**, 503103 (2014).

[17]   R. K. Scopes, Enzym. **8**, 335 (1973).

[18]   T. N. Bryant, H. C. Watson, and P. L. Wendell, Nature **247**, 14 (1974).

[19]   B. E. Bernstein, P. A. Michels, W. G. Hol, P. A. M. Micheis, and W. G. Hol, Lett. to Nat. **385**, 275 (1997).

[20]   R. Inoue, R. Biehl, T. Rosenkranz, J. Fitter, M. Monkenbusch, A. Radulescu, B. Farago, and D. Richter, Biophys. J. **99**, 2309 (2010).

[21]   L. R. Stingaciu, O. Ivanova, M. Ohl, R. Biehl, and D. Richter, Sci. Rep. **6**, 22148 (2016).

[22]   L. Pauling, J. Am. Chem. Soc. **62**, 2643 (1940).

[23]   A. M. Stadler, L. Stingaciu, A. Radulescu, O. Holderer, M. Monkenbusch, R. Biehl, and D. Richter, J. Am. Chem. Soc. **136**, 6987 (2014).

[24]   G. Harauz, N. Ishiyama, C. M. . Hill, I. R. Bates, D. S. Libich, and C. Farès, Micron **35**, 503 (2004).

[25]   E. Polverini, A. Fasano, F. Zito, P. Riccio, and P. Cavatorta, Eur. Biophys. J. **28**, 351 (1999).

[26]   P. Bernado, E. Mylonas, M. V Petoukhov, M. Blackledge, and D. I. Svergun, J Am Chem Soc **129**, 5656 (2007).

E 9   Scattering Applications in Life Sciences

Andreas Stadler & Alexandros Koutsioumpas
Jülich Centre for Neutron Science (JCNS)
Forschungszentrum Jülich GmbH

Contents

1   Introduction

Proteins and nucleic-acids are the two most basic ingredients of living organisms as they carry out most basic functions of cells including reproduction, catalysis of biochemical reaction, response to stimuli and structure formation. Determination of their three dimensional structure is of paramount importance for obtaining clues about how these molecules perform their function. In fact great scientific effort has been devoted from the mid-20th century till today for the development of experimental and theoretical tools specifically for the elucidation of biomolecular structure. As was described in previous chapters, the "workhorses" of high resolution structure determination are x-ray/neutron diffraction and NMR, with cryo-electon microscopy lastly gaining popularity.

In the current chapter we will briefly review alternative biophysical scattering tools that play a complementary role to the above-mentioned established techniques. In particular we will see how Small-angle scattering (SAS) and Reflectivity despite the lower resolution that they offer, can be used for the extraction of structural information concerning systems of biological interest and also present some illustrative examples involving proteins, nucleic acids and biological lipid membranes.

2   Biomolecular structure from Small Angle Scattering

SAS is a well-suited experimental method to investigate low-resolution structures of biological macromolecules in solution. Advantages of small-angle neutron (SANS) and X-ray scattering (SAXS) are, for instance, that proteins do not need to be crystallised, but can be studied as relatively diluted solutions with protein concentrations of a few mg/mL. It is often the case that crystal structures of proteins can be determined, but flexible regions of the protein that are necessary for biological function are not visible in the crystal structures. Here SAXS and SANS can be used to gain structural insight into the conformational properties of those flexible and unstructured regions. Concerning protein complexes, SAXS and SANS allows to determine the arrangements of the components provided that their crystallographic structures are known. By using the concept of neutron contrast variation the position and low-resolution shapes of DNA or lipid molecules in complex with a protein can be determined. SANS is particularly well-suited for the low-resolution structure determination of membrane proteins that are solubilised by detergent molecules or nanodiscs. In case that proteins exhibit large-scale structural changes as response to external stimuli, SAXS and SANS are well-suited to monitor kinetic changes in a broad time range from the millisecond to several hours.

2.1   Theoretical background

In previous chapters we have already encountered examples where SANS and SAXS provide direct information about the overall size, form and interaction between particles dissolved in a solvent. For the purposes of the present chapter we will consider the special case of monodisperse and dilute solutions of biological molecules where no interaction between different molecules takes place. We first start by recalling (chapter D1) that the scattered wave from an assembly of

atoms is given by the sum of waves scattered, so

$$A(\mathbf{Q}) = \sum_j f_j \exp(i\mathbf{Q}\mathbf{r}_j) \tag{1}$$

where $f_j, \mathbf{r}_j$ are the scattering amplitudes and position vectors respectively. This equation reflects that the scattered wave is the Fourier transform of the scattering amplitudes and positions of the atomic assembly. We note that f_j for neutrons and x-rays is different (see discussion chapter D1.6). For neutrons f_j is Q-independent due to the fact that nuclei sizes are much smaller than the typical wavelegths used in scattering experiments, so that they act as point scatterers. On the contrary, since x-rays are scattered by the atomic electron cloud which has a size comparable to the wavelength of the incident radiation, a Q-dependent atomic form factor should be taken into account.

Equation (1) is valid for a collection of atoms in vacuum. However in solution scattering experiments an assembly of atoms (particle or molecule) is studied in aqueous solutions. At the scale of $1 nm$ or more, water has no structure and acts as a homogeneous "matrix" of constant scattering density ρ_w surrounding the atoms. In order to calculate the scattered intensity from such a system we have to consider three contributions: a) the scattering from the atoms b) the solvent and c) the excluded solvent volume v_j due to the presence of the atoms. Due to the homogeneous nature of the solvent contribution (b) can be considered as zero so that the overall scattering is given by the difference between contributions (a) and (c) as given by the equation

$$A(\mathbf{Q}) = \sum_j (f_j - \rho_w v_j) \exp(i\mathbf{Q}\mathbf{r}_j) \tag{2}$$

The term $f_j - \rho_w v_j$ essentially represents the contrast amplitude of the atom j with respect to the solvent. It is important to note that due to the large difference of the scattering length of hydrogen and deuterium, the contrast of a particle in relation to water can be easily manipulated in scattering experiments by the use of different water/heavy water ratios.

As explained in chapter D5.5, in a scattering experiment only the intensity of the scattered wave $I(\mathbf{Q}) = |A(\mathbf{Q})|^2$ is measured, thus phase information is lost. Given that we examine the case of dilute monodisperse solutions where no interparticle interference is present, the scattered intensity from unit volume of solution containing N particles can be written as:

$$I_N(\mathbf{Q}) = \sum_{n=1}^{N} |\mathbf{A}_n(\mathbf{Q})|^2 \tag{3}$$

which can be further simplified by applying rotational averaging (since the atomic assemblies may assume all different orientations in respect to the beam)

$$I_N(Q) = N < |\mathbf{A}_n(\mathbf{Q})|^2 > \tag{4}$$

Peter Debye was the first [1] to calculate the rotational average in equation (4) using the result $< exp(-i\mathbf{Q}\mathbf{r}) >= \sin Qr/Qr$ and thus giving the Debye equation that writes

$$I_N(Q) = N \sum_j \sum_k (f_j - \rho_w v_j)(f_k - \rho_w v_k)\frac{\sin Qr_{jk}}{Qr_{jk}} \tag{5}$$

Before we move on into exploring the application of the Debye equation for biological macro-molecules, it is instructive to expand $\sin Qr/Qr$ in terms of a series of powers of Q close to $Q = 0$ (Maclaurin expansion) thus obtaining

$$I_N(Q) = N \sum_j \sum_k (f_j - \rho_w v_j)(f_k - \rho_w v_k)(1 - \frac{r_{jk}^2 Q^2}{3!} + ...) \tag{6}$$

In analogy to classical mechanics we may define a "scattering" radius of gyration where the role of mass is played by the contrast:

$$R_G^2 = \frac{\sum_j \sum_k (f_j - \rho_w v_j)(f_k - \rho_w v_k)r_{jk}^2}{2(\sum_j (f_j - \rho_w v_j))^2} \tag{7}$$

Combining the last two equations we get

$$I_N(Q) = N(\sum_j (f_j - \rho_w v_j)^2(1 - \frac{1}{3}R_G^2 Q^2 + ...) \tag{8}$$

We note that the expression $N(\sum_j (f_j - \rho_w v_j)^2$ is equal to the scattered intensity at $Q = 0$, so that

$$I_N(Q) = I_N(0)(1 - \frac{1}{3}R_G^2 Q^2 + ...) \tag{9}$$

The two terms in parentheses can be viewed as the first two terms of the Maclaurin expansion of $exp(-\frac{1}{3}R_G^2 Q^2)$. This result leads to the useful Guinier relationship $lnI(Q) = lnI(0) - (1/3)R_G^2 Q^2$ which dictates that in a logarithmic plot of the scattered intensity versus Q^2 the slope at small Q is equal to $-(1/3)R_G^2$, thus giving us a straightforward way to estimate the radius of gyration of the particle in solution irrespective of its contrast and shape.

2.2   SAS calculations based on atomistic models

The Debye equation gives us a convenient way to calculate the expected small-angle scattering curve for a molecule or particle dissolved in water. In practice in studies involving the use of bio-SAS we encounter one of the two following general scenarios a) an all-atom structure of a molecule or molecular complex is available by crystallographic input or by a structure prediction methodology and we want to verify that the structure in solution is compatible and b) there is none or limited information about the structure of the system and only a small-angle scattering curve is available.

For the first scenario we have to solve "direct problem" and the Debye equation gives us a straightforward way to accomplish such a task. There are many scientific software programs
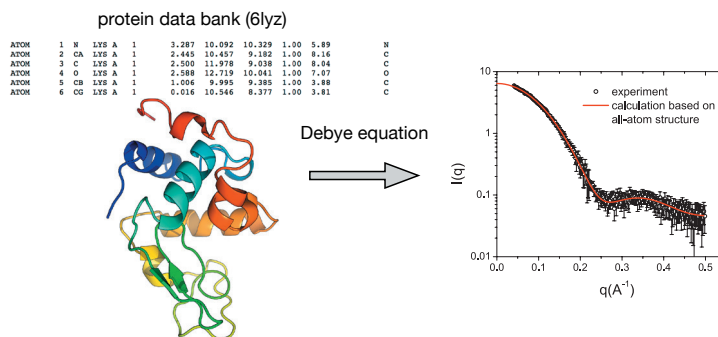
Fig. 1: General overview of solving the "direct problem" in a biological small angle experiment. Given the all-atom crystallographic structure of a protein (lysozyme: PDB code 6LYZ), the experimental SAXS curve can be compared with the theoretically calculated scattering using the Debye equation. For this particular example we additionally take into account the presence of a 0.3 nm thick hydration layer around the surface of the protein with an increased solvent density of 10%.

specialised for performing such calculations in a fast and accurate way [1]. One important aspect during such calculations is that an additional contribution to the scattering that has to be taken into account that is the related to the special organisation of water molecules around the surface of biomolecules. This so called (hydration or solvation layer) of ≈ 0.3 nm thickness around proteins, has been shown to be around 10% denser than bulk water [3].

In Fig. 1 we present a representative example of a comparison between a SAXS curve from a small ($R_g = 1.4$nm as found by the Guinier plot) 129 aminoacid hen egg lysozyme protein and the corresponding theoretically calculated curve using the crystallographic three dimensional structure (resolution 0.2 nm) available in the protein data bank (code: 6LYZ [4]). We note the agreement between the experimental and calculated curve that implies that the protein retains in solution the structure that is found by the crystallographic investigation. Of course the success of the above comparison between a model structure and the small-angle measurement depends on the quality of the sample. Radiation damage, aggregation or the presence of multimers in solution might lead to deviations from the scattering curve of the pure molecular structure that we aim to probe. There is extensive literature on how such effects can be detected and also avoided during a bio-SAS experiment [5].

The presented approach can be extended to cases where the structure of parts of a molecule or complex are known and their relative arrangement is investigated. Candidate structures can be prepared and compared against small-angle scattering data, while also additional experimental constraints from other techniques (like NMR) or energetic calculations may aid towards finding a global solution that is representative of the actual conformation. The interested reader is referred to informative reviews on the subject [6, 7].

---

[1] Probably the program CRYSOL is the most commonly used [2]

2.3   Ab initio modelling based on SAS data

The second general scenario in bio-SAS studies is related to molecules of totally unknown structure. In that case we need to solve the "inverse problem" that is extracting the maximum amount of structural information from scattering data. As we have already discussed, the low-Q region contains information about the size (radius of gyration) of the molecule (or particle) in solution. However the intermediate Q-region up to $\approx 3nm^{-1}$ contains information about the overall shape on the molecule. At such a spatial resolution $2\pi/Q \approx 2nm$, folded biological molecules like globular proteins and nucleic acids are characterised by a given contrast relative to the solvent (water), since they share common structural motifs. That means that we may approximate biomolecular structures as compact envelopes of uniform contrast relative to the solvent. Such approaches were pioneered by Svergun and Stuhrmann [8], who showed that, by incorporation of a priori constraints concerning the sld or electron density, a nonlinear reconstruction procedure can provide reliable low-resolution envelopes, representative of molecular shape.

An even more versatile way for the finding such a molecular envelope is to represent the molecule as a collection of "dummy-atoms" of given contrast in space, and search through a minimisation procedure a compact arrangement that reproduces the measured scattering properties of the molecule. Algorithms that attempt such shape reconstructions (DALAI GA [9], DAMMIN/IF [10], DENFERT [11]) belong in the class of "ab-initio" shape reconstruction methodologies for low-resolution structure recovery from SAS data. In Fig. 2 and Fig. 3 we present such a shape reconstructions from SAXS and SANS data for hen egg lysozyme and bovine bovine β-lactoglobulin proteins respectively.
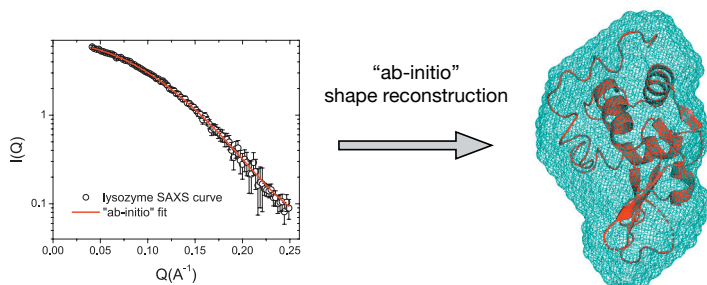
Fig. 2: General overview of solving the "inverse problem" in a biological small angle experiment. The shape reconstruction from lysozyme SAXS data was performed using the program DENFERT [11]. On the right the obtained molecular envelope (cyan) is superimposed with the crystallographic cartoon representation of the molecule, protein data bank code - 6LYZ [4]

.

From these two examples it is evident that despite the loss of phase and rotational averaging in small-angle experiments, the scattering curve is still encoding enough information that guides us to representative low-resolution molecular envelopes. In practice by performing such calculations we do not end up with a unique solution but with multiple structurally similar envelopes that give equal agreement with the measured curve. Variation between different envelopes is indicative of the stability of the solution, while tools for superimposing and averaging this pool
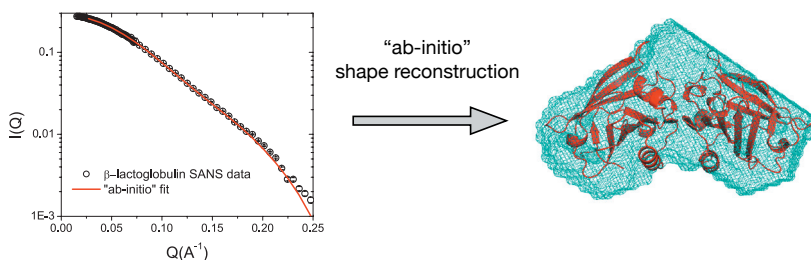
Fig. 3: Shape reconstruction from bovine β-lactoglobulin SANS data (measured at KWS1 instrument at MLZ) using the program DENFERT [11]. On the right the obtained molecular envelope (cyan) is superimposed with the crystallographic cartoon representation of the molecule's dimer, protein data bank code - 1BEB [12]

.

of structures are available.

The information that is obtained by "ab-initio" reconstructions can be used in different ways in integrative structural biology studies. Questions like the multimerization of a protein or large conformational transitions in solution may be elucidated by low-resolution models. Furthermore the natural difference in contrast between proteins and nucleic acids and also the variation of contrast (by selective deuteration of a molecule) add even more information that can be used in order to study more complicated systems like membrane proteins and biomolecular complexes. Based on the described SAS theoretical toolset, the next section gives examples of application in a system of biophysical interest.

3   Structural investigations on light-signal sensation and termination

Sensation of light by organisms is an important biological process. Dedicated light-sensing photo-receptors are able to detect photons and transmit the detected light-signal further into the cell. Receptor proteins cycle between a dark-adapted ground state and a light-activated signalling state. However, after a certain time, the active receptor proteins needs to be deactivated to terminate the signalling process. Arresting is a protein that interacts with a certain class of activated receptor proteins (so called G protein coupled receptors GPCRs). Arrestin terminates the active signalling state bringing the receptor back into its ground state and, hence, terminates the signalling process.

As a first illustrative case, we present investigations combining crystallography and SAXS on structural changes that occur upon photo-illumination in a photo-receptor protein. Here, we investigated the structural transitions of a photoreceptor protein from Pseudomonas putida - abbreviated as PpSB1-LOV - belonging to the larger light, oxygen, voltage (LOV) receptor family. Crystal structures of PpSB1-LOV in the dark- and the light-state have been determined using X-ray crystallography, see figure 4. However, due to protein flexibility and disorder

no information of residues located at t at the N- and C-terminal region could be obtained by crystallography, see figure 4. Especially residues located at the end of the long alpha-helix Jalpha are relevant for light-signal transduction.
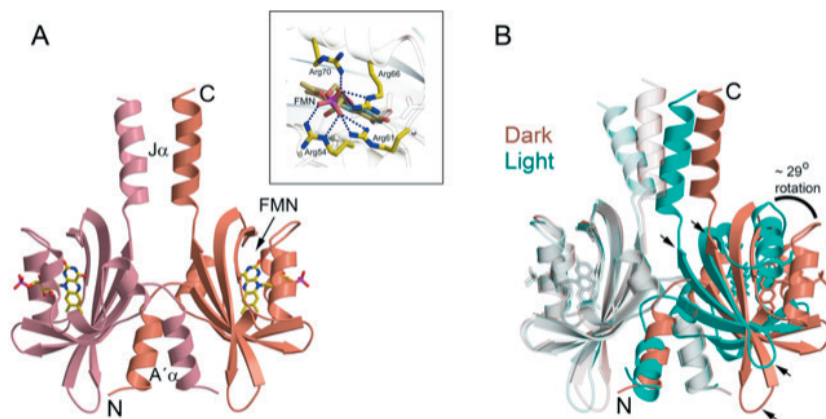
Fig. 4: Crystal structure of PpSB1-LOV. (A) Dark-state structure (PDB ID: 5J3W). (B) Superposition of dark-state (salmon, PDB ID: 5J3W) and light-state structures (cyan, PDB ID: 3SW1) states showing structural differences. A significant number of residues located at the top and bottom of the protein structures could not be determined due to disorder and large flexibility [13]

To circumvent that problem and to gain information on those flexible and unstructured regions, we have performed SAXS experiments of the light- and dark-states of PpSB1-LOV in solution. The dark adapted state was measured in the absence of light, the protein solution was illuminated with blue light shortly before the SAXS experiment to excite the fully populated light-state. Measured SAXS data of dark- and light-state PpSB1-LOV are shown in figure 5. To gain a structure based understanding of the measured SAXS data, coarse-grained modelling has been performed and the obtained models are shown in figure 5 C.

Disordered residues at the bottom of the models (see figure 5C) belong to the so-called His-tag. The His-tag has been included into the protein sequence for facilitated sample purification by affinity columns. Residues on the top of the model show light-induced structural changes in the photoreceptor. In the dark-state structure, residues are more disordered indicating of a high flexibility in that region. Upon photo-illumination, the end regions seam to form a coiled-coil element that acts as effective prolongation of the long Jalpha helix. Formation of that structural element appears to be relevant for the biological function enabling contact formation with further interaction proteins.

Let us consider now a second case study that belongs to the field of light-signal sensation and transduction. As mentioned above, signal termination is an essential component of the signal sensation mechanism. A widely studied system to investigate the molecular mechanism of signal termination is arrestin protein, which occurs in the visual signalling cascade. Visual arrestin-1 (arr-1) is proposed to exist in an inactive form that can bind to the phosphorylated and light-activated state of the photo-receptor rhodopsin. A single point-mutation of arrestin (R175E) can bind to light-activated rhodopsin independent of phosphorylation. The R175 mu-
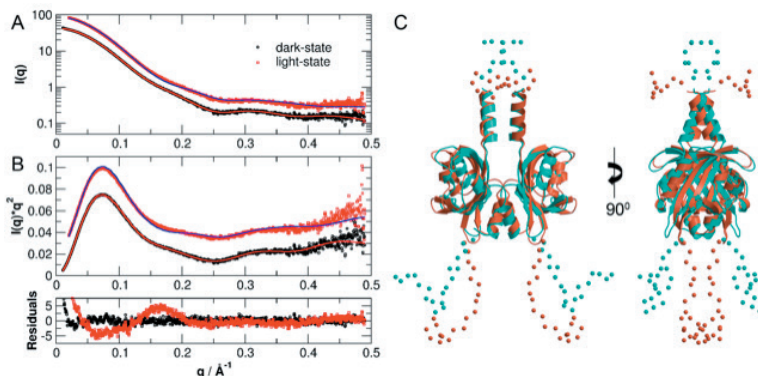
Fig. 5: Small-angle scattering data of the photoreceptor protein PpSB1-LOV-R66I in the light- and dark-states and coarse-grained structural models based on available crystal structures. (A) and (B) Experimental data and theoretical scattering curves (solid lines) of the light- and dark-state crystal structures (PDB ID's: 3SW1 and 5J3W) including flexible ends that are not seen in the crystal structures. Data are shifted for clarity. The lower panel shows the residuals of the models. (C). Coarse-grained structural models of light (cyan) and dark (salmon) states showing the structural difference between light- and dark-states [13].

tation lead to the release of the long C-tail of arrestin, which is crucial for the interaction with light-activated rhodopsin. A pre-active variant of arrentin - named p44 - has also been reported that lacks that C-tail entirely. Crystal structures of the three arrestin variants could be obtained. However, due to its high flexibility no structural information could be obtained on the important C-tail, nor of the oligomerisation state of the arrestin proteins in solution. Therefore, SAXS experiments have been performed to determine the oligomeric states and gain information on the role of the C-tail. Experimental SAXS data including theoretical cuves based on structural modelling are shown in figure 6. SAXS results confirm that the wild type arr-1 is dimeric in solution, with the best fit comprising A/D chains of the crystal structure (PDB ID 3UGX). In contrast, both mutant R175E and p44 are monomeric. In case of R175E, the SAXS data fits best to the structural model, when the C-tail is included and modelled as flexible chain. This is in agreement with ab initio models of R175E, which is more extended in the region of the flexible C-tail when compared to that of p44 (figures 6b, 6c).

From the two case studies presented above, it is obvious that SAS yields important information on protein oligomeric state and of disordered and flexible regions that are relevant for light-sensation and signal termination of photo-receptor proteins. While the two SAS studies mentioned above have been performed under static steady-state conditions, the technique can also give valuable information on structural changes that evolve as a function of time after a certain external stimulus. Concerning photoreceptor proteins, the external trigger is the excitation with blue light. The dark-adapted state of the photo-receptor is illuminated with blue light,
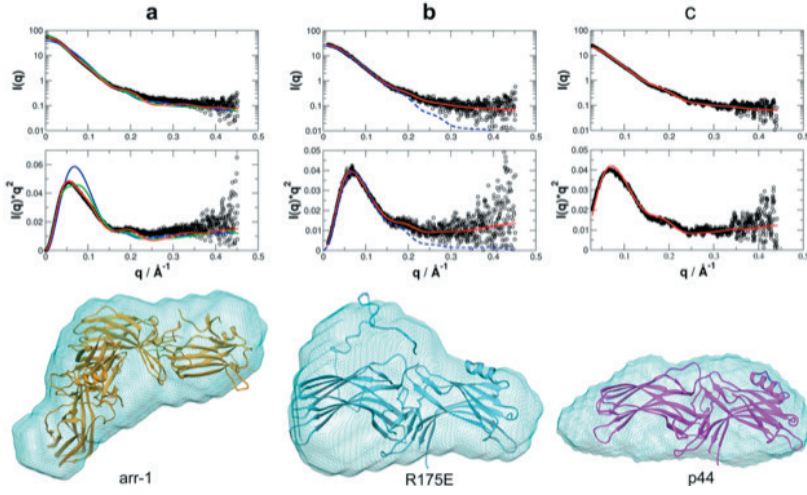
Fig. 6: Arrestin measured by SAXS. (a) arr-1, (b) the mutant R175E and (c) the mutant p44. SAXS data are shown as log-lin plots and as so-called Kratky plots (black circles). The different solid lines are theoretical curves as obtained from the crystal structures including a flexible C-tail for R175E. (a) Crystal structures considered are either the monomer (blue line: molecule A), and two different dimer models (green line: molecule A and B; red line: molecule A and D) that are found in the crystal structure (PDB ID: 3UGX). In (b) the curves of the R175E monomer including the flexible C-terminus (red line), and without the C-tail (dashed blue line) is given. In (c) the curve of the crystal structure of p44 (PDB ID: 3UGU) monomer (red line) is plotted. At the bottom: Crystal structures: Gold, dimer with molecules A and D (PDB ID: 3UGX) in (a); blue R175E in (b), and magenta p44 (PDB ID: 3UGU). Obtained ab initio models from SAXS are overlaid with the crystal structures. (Reference: Structural evidence for the role of polar core residue Arg175 in arrestin activation [14].

which brings the protein in its light-excited state. The blue light is then switched off, and the light-excited state of the photo-receptor recovers back into its dark-adapted ground state. The time-scale of the recovery process differs between various kinds of photo-receptors and also can be "tuned" by point-mutations in close vicinity around the chromophore. As one illustrative example, we consider one scientific study, where we have used time-resolved SAXS experiments to follow the light-to-dark recovery process of the photo-receptor protein PpSB1-LOV-R66I in solution.

The protein sample was illuminated at the beginning with blue light and subsequently experimental SAXS data were recorded as a function of time. Using the knowledge that the light-state of the protein recovers to the dark-state as a function of time the experimental data $I(Q, t)$ could be fitted as a simple linear combination of light- and dark-states according to

$$I(Q,t) = \phi(t)I(Q)_{light} + \big[1 - \phi(t)\big]I(Q)_{dark} \tag{10}$$

where $I(Q)_{light}$ and $I(Q)_{dark}$ are the theoretical scattering curves of the light- and dark-state models that have been obtained from SAXS experiments performed under static steady-state conditions before (see figure 5), and $\phi(t)$ is the time-dependent light-state fraction that was

used as the only free fit parameter.

Representative SAXS curves at different time points together with the theoretical fits according to equation 10 are shown in figure 7. The two-component fits yield a good representation of the time-dependent SAXS data. To illustrate changes of the SAXS pattern as a function of time the difference pattern $\Delta I(Q,t)$ of SAXS data measured at selected time points minus the SAXS data recorded at $t = 0$ min is shown in figure 7 as well with with fits according to the two-state model with the only free parameter ϕ.
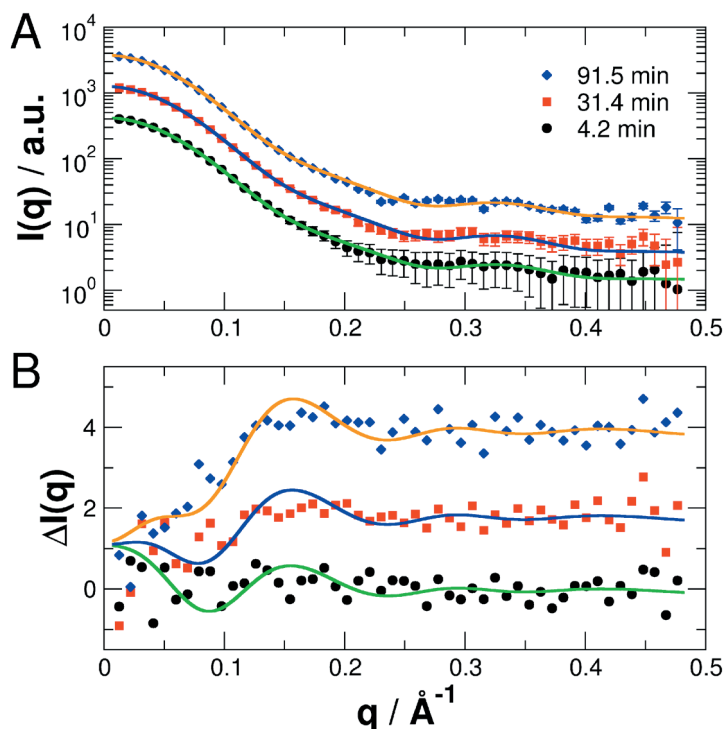
Fig. 7: (A) Experimental SAXS data measured at selected time-points after photo-illumination of the photo-receptor mutant PpSB1_LOV-R66I. Solid lines are fits to the data assuming a two-state population of known PpSB1-LOV-R66I light- and dark-state structures based on the steady-state models with the only free fit parameter ϕ that represents the population of the light-state (equation 10). Data and fits are shifted for clarity. (B) Difference SAXS patterns $\Delta I(Q,t)$ of data measured for the time points shown in (A) minus the SAXS pattern recorded at $t = 0$ min.

The structural transition of the photo-receptor protein is clearly visible in the $\Delta I(Q,t)$ patterns, see figure 7, by a loss of intensity in the q-range of $0.02 - 0.2\,\text{Å}^{-1}$. Large-scale structural re-arrangements are expected to be most visible in this q-range. Therefore, to follow the recovery process in a model-independent approach $\Delta I(Q,t))$ was integrated up to $0.2\,\text{Å}^{-1}$. The time behaviour of the normalised integral of $\Delta I(Q,t)$ and the light-state fraction $\phi(t)$ determined from the structure-based fit are shown in figure 8, and inform on the structural change from

the photo-excited light-state to the dark-adapted ground state. The decay could be described well with a single-exponential function, which demonstrated directly that no intermediate states are populated during the structural transition of the photoreceptor protein. Furthermore, we obtained a relaxation time of $\tau_{REC} = 33.5 \pm 3.6$ min and 35.3 ± 5.2 min for the normalised integral of $\Delta I(Q, t)$ and the light-state fraction $\phi(t)$, respectively. Within the statistical uncertainty both relaxation times are essentially the same.
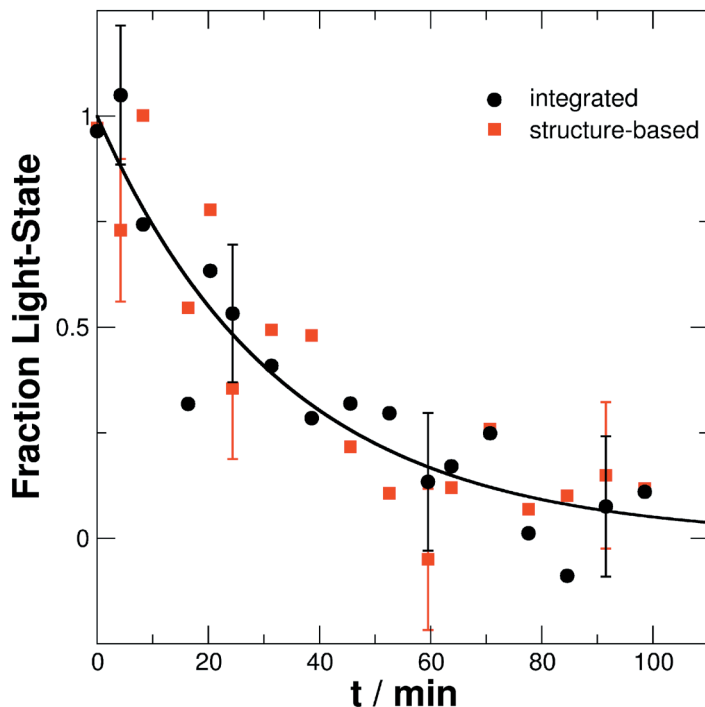
Fig. 8: Light-state fraction of PpSB1-LOV-R66I as a function of time. Determined from the integral of $\Delta I(Q, t)$ up to $0.2\,\text{Å}^{-1}$ and a structure-based fit of $I(Q, t)$ assuming a mixed population of known light- and dark-states CORAL models . The solid line is a single-exponential fit to the data points of the integral of $\Delta I(Q, t)$. For clarity error bars are shown for selected points.

Therefore, to conclude, we could show with time-resolved SAXS that the kinetics of the light-to dark-relaxation process proceeds directly from the light-excited to the dark-adapted state without the populations of any transient intermediates on the investigated time-scale. Structural intermediates would be visible by deviations from simple-exponential behaviour. The time-scale of the kinetic process is in the order of 30 min, which is due to the investigated point-mutations. Different point mutations of the photo-receptor protein result in slower or faster relaxation kinetics.

4 Biological membrane sub-nanometer structure from neutron reflectivity

So far in this chapter we have presented examples of small-angle scattering applications in the determination of the structure of biological molecules in bulk solution. Neutron and X-ray reflectivity represent another class of scattering techniques that can give information about the organisation of bio-molecules not in a bulk solution but at interfaces. In chapter D2 the basic principles of x-ray and neutron specular reflection have been presented. For x-rays and cold neutrons the index of refraction of materials is close to unity so that measurable reflection is constrained at grazing incidence angles (typically up to a few degrees). However the use of short wavelengths in comparison to the reflection of light at visible wavelengths, gives the ability to probe down to quite small length scales, since the dependence of neutron or x-ray reflected intensity on the magnitude of the scattering wavevector Q contains information about the molecular distribution at the sub-nanometer scale.

The essence of a reflectivity experiment consists of registering specular reflection as a function of the incidence angle θ or equivalently as a function of $Q = 4\pi \sin\theta/\lambda$ where λ is the wavelength of the incident radiation. Experiments can be performed at different types of interfaces like air/liquid, air/solid, liquid/liquid and solid/liquid. The experimental information can be related to the refractive index profile normal to the interface, providing information about interfacial composition. More specifically in neutron reflectivity experiments, the reflected and refracted neutron beams may be considered as plane waves, so that the solution of the time-independent Schrödinger equation is reduced to an one-dimensional problem. In fact for scattering length density profiles (sld) that can be modelled as a sequence of layers with different sld's ρ, the reflected intensity can be calculated exactly using the convenient matrix formalism [15, 16] where for each layer n, a characteristic matrix is defined as

$$c_n = \begin{bmatrix} \exp(k_n d_n) & r_n \exp(k_n d_n) \\ r_{n,n+1} \exp(-k_n d_n) & \exp(-k_n d_n) \end{bmatrix} \tag{11}$$

where $k_n = [Q^2/4 - 4\pi(\rho_n - \rho_0)]^{1/2}$ is the neutron wavevector in layer n, ρ_n, d_n and σ_n are the sld, thickness and roughness of layer n respectively and $r_{n,n+1}$ is the Fresnel reflection coefficient between n and $n+1$ given by

$$r_{n,n+1} = \frac{k_n - k_{n+1}}{k_n + k_{n+1}} \exp\left(-2k_n k_{n+1} \sigma_{n,n+1}^2\right) \tag{12}$$

The overall system matrix M is given by

$$M = \prod_{n=0}^{\text{\# of layers}} c_n \tag{13}$$

from which finally the specular reflected intensity is calculated as

$$R(Q) = |M_{11}/M_{21}|^2 \tag{14}$$

The set of equations (11) - (14) provides a convenient way for calculating the expected specular neutron reflectivity for a given sld profile [2]. Furthermore, given a set of measurements and a model for the sld profile, the parameters of each layer (thickness, sld, roughness) can be varied and the set that gives the best agreement with the experimental data can be determined. As we will see in the following, the ability in neutron reflectivity measurements to vary the contrast of the solvent or of parts of the studied system itself, can lead to a vast reduction of the ambiguity of the final obtained structural model at the interface. In order to build some intuition about the practicalities and capabilities of the technique for the study on biological systems we will present a set of examples of increasing complexity that all have to do with biomimetic lipid membranes.

Biological lipid membranes represent an important self-assembly system in living organisms, that perform the function of separation and compartmentalisation by acting as selective permeable barriers. In their most abundant form as the outer layer of cells (typically $4 - 5$ nm thick), they provide a bilayer fluid matrix for embedded membrane proteins that in turn regulate the cell's communication with the surrounding environment. Due to the high complexity of membranes in living cells, model phospholipid bilayer membranes with well-defined composition serve as relatively simpler counterparts in experimental studies.

It is a well established that small unilamellar vesicles composed of phospholipids under certain conditions can fuse on hydrophilic substrates (like silica) and produce a single planar double lipid layer (membrane) which is separated from the surface by a subnanometer thick hydration layer (see inset of Fig. 9). These supported bilayers are ideal for performing neutron reflectivity measurements at solid/liquid interfaces with the aim of extracting structural information concerning the molecular structure of membranes perpendicular to the supporting surface.
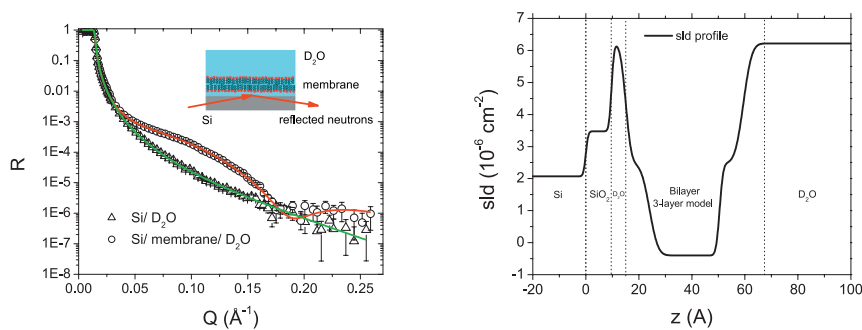
Fig. 9: (left) Neutron reflectivity (data measured at MARIA, MLZ, Garching [17]) of the silicon/D_2O interface before and after the formation of supported DPPC bilayer by vesicle fusion. Full lines represent model fits of the experimental data. (right) sld profile obtained using a three layer model (head/tails/heads) for the supported lipid membrane.

---

[2] We note that almost the same formalism can be used for the case of x-ray reflectivity with a change having to do with the replacement of sld's by electron densities. In the present chapter we will focus on neutron reflectivity since for the study of biological systems, despite it's lower resolution, presents some unique advantages related to the ability to vary contrast and also the possibilities to study "buried" interfaces due to the generally low absorption of neutron by most materials.

In Fig. 9 (left), we present two experimental reflectivity curves corresponding to a) a pure Silicon/ heavy water interface and b) to the same interface after the formation of a supported 1,2-dihexadecanoyl-*sn*-glycero-3-phosphocholine (DPPC) phospholipid membrane on the substrate. It can be seen that the Si/D$_2$O curve, expect from the small total reflection plateau at low-Q, follows a quasi Q^{-4} (Fresnel) behaviour [3]. Upon the formation of the DPPC membrane, we observe that the appearance of a characteristic "Kiessig fringe" related to the presence of a bilayer membrane "floating" above the substrate and especially to its high contrast hydrocarbon core relative to D$_2$O.

By modelling the interface using 5 layers as silicon oxide / hydration layer / hydrophilic lipid heads / hydrophobic lipids tails (hydrocarbon core) / hydrophilic lipid heads, we may use equations (11) - (14) for calculating the expected reflectivity of the model and by a proper minimisation procedure find the model parameters that provide the best agreement with the acquired experimental curves. Using such a priori information and also tabulated sld values for the different parts of the system we end with a pretty detailed sld profile of the interface (Fig. 9 (right)) which essentially reflects the structural details (thickness, roughness and surface coverage) of the supported membrane.

As an example of the sensitivity of the technique, we showcase measurements of supported lipid bilayers at two different temperatures (Fig. 10). It is well established that at a given temperature a lipid bilayer can exist either in the gel (solid) or liquid (liquid-crystalline) state. This phase behaviour is dominated by Van der Waals attractive interactions between lipid tails, and is affected by the length and degree of hydrocarbon tail saturation. The two phase states are associated with large differences in the overall fluidity and molecular packing. In the gel phase it is expected that lipid tails are more ordered and membrane thickness is larger comparing to the liquid state. The transition temperature for DPPC lipids is equal to $T_m = 40.9°$C. In Fig. 10 we clearly see that as we pass from $T = 60°$C (fluid phase) to $T = 20°$C the characteristic fringe is shifted towards lower Q. This inverse-space shift signifies a thickening of the the membrane. A fit of the curves with the above mentioned 5-layer model reveals (inset Fig. 10) a change of overall membrane thickness from 47.7 Å to 57.8 Å, in line with theoretical predictions and previous experimental studies.

Till now we have presented examples where only a single solvent contrast has been used and it has to be noted that our ability to extract such detailed structural insight is also based on the a priori information that is incorporated in the models used for the fit of the data. However in many cases such information might not be available or very subtle structure details need to be resolved. In such scenarios our ability to perform neutron reflectivity measurements using multiple solvent contrasts (or the possibility to manipulate the contrast of parts of the system by selective deuteration), provides the chance to decisively reduce the ambiguity of the determined model parameters.

In practice this approach involves the exchange of the D$_2$O/H$_2$O ratio during a measurement. In principle such a modification is not affecting the interfacial structure, the physicochemical properties of the solvent remain largely the same and only the sld values of the hydrated parts of the system are varied (in the present case this mainly concerns the small hydration layer near the surface and the hydrophilic lipid head regions). Such a set of measurements for a

---

[3] The small deviations from the Fresnel behaviour are due to the presence of a $\approx$ 1nm thick silicon oxide native layer on the surface of Silicon.
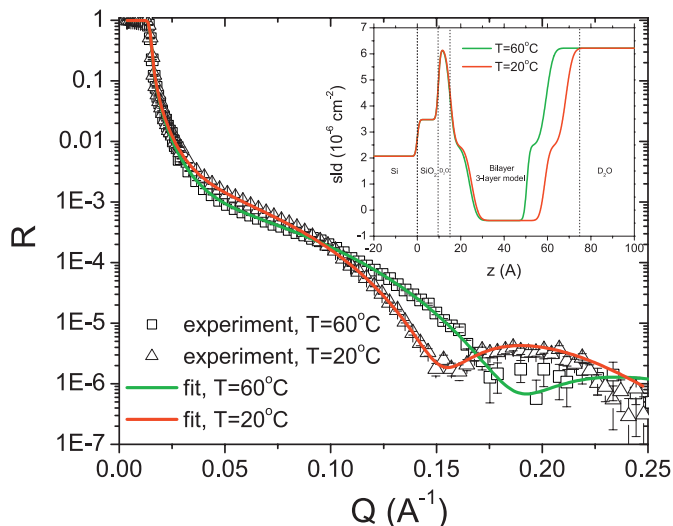
Fig. 10: Reflectivity curves and model fitting results (full lines) corresponding to the supported DPPC bilayer at two different temperatures above and below T_m. In the inset the obtained sld profiles from the fitting procedure are plotted.

supported DOPC phospholipid bilayer at three different contrasts can be seen is Fig. 11. Simultaneous fit of all three curves where structural parameters like thickness and roughness are kept the same for every contrast and only the sld of hydrated parts is left to vary depending on the sld of the bulk solvent, gives confidence about the validity of the obtained fitted parameters. This is especially true in cases where we want to probe the interaction of proteins, small peptides or nanoparticles with supported lipid membranes. By acquiring measurements at multiple contrasts we may answer questions about the penetration, additional hydration or even destabilisation of membranes due to the presence of such molecules or particles.

Single zwitterionic lipid membranes that were discussed in the subchapter, probably constitute the simplest membrane mimetic system. However there is an increasing number of published works [18] where the complexity of supported membranes in neutron reflectivity studies is augmented using multiple types of lipids, cholesterol or even total lipid extracts from living cells [19], in an effort to better mimic the actual cell membrane. All this coupled with the ability to manipulate the contrast of bio-molecules that interact with membranes like, peripheral and integral membrane proteins, antimicrobial peptides and so forth, opens the road for many exciting applications of neutron reflectivity at the interface between physics and biology.
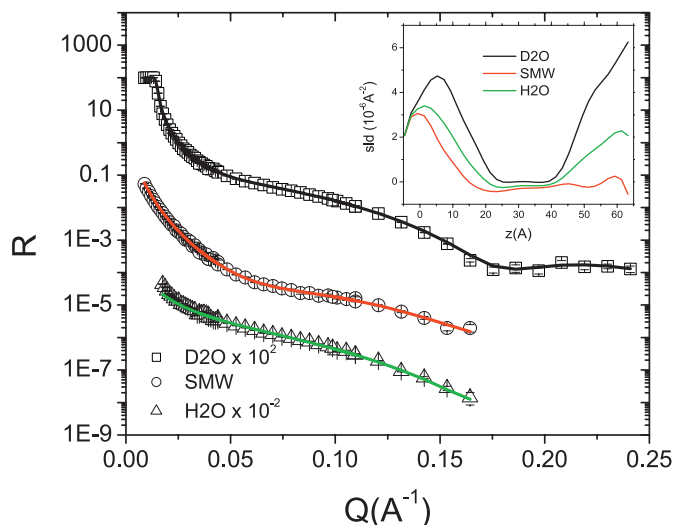
Fig. 11: Three different contrast reflectivity curves and model fitting results (full lines) corresponding to a supported DOPC bilayer. In the inset the obtained sld profiles from the fitting procedure are plotted.

References

[1] P. Debye, *Zerstreuung von Röntgenstrahlen*, Annalen der Physik **351**, 809 (1915).

[2] D. Svergun, C. Barberato, M. H. J. Koch, *CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates*, Journal of Applied Crystallography **28**, 768 (1995).

[3] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, G. Zaccai, *Protein hydration in solution: Experimental observation by x-ray and neutron scattering*, Proceedings of the National Academy of Sciences **95**, 2267 (1998).

[4] R. Diamond, *Real-space refinement of the structure of hen egg-white lysozyme*, Journal of Molecular Biology **82**, 371 (1974).

[5] D. A. Jacques, J. Trewhella, *Small-angle scattering for structural biologyExpanding the frontier while avoiding the pitfalls*, Protein Science **19**, 642 (2010).

[6] M. H. J. Koch, P. Vachette, D. I. Svergun, *Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution*, Quarterly Reviews of Biophysics **36**, 147227 (2003).

[7] C. D. Putnam, M. Hammel, G. L. Hura, J. A. Tainer, *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular struc-*

tures, conformations and assemblies in solution, Quarterly reviews of biophysics **40**, 191 (2007).

[8] D. I. Svergun, H. B. Stuhrmann, *New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations*, Acta Crystallographica Section A **47**, 736 (1991).

[9] P. Chacn, F. Morn, J. Daz, E. Pantos, J. Andreu, *Low-Resolution Structures of Proteins in Solution Retrieved from X-Ray Scattering with a Genetic Algorithm*, Biophysical Journal **74**, 2760 (1998).

[10] D. Svergun, *Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing*, Biophysical Journal **76**, 2879 (1999).

[11] A. Koutsioubas, J. Pérez, *Incorporation of a hydration layer in the 'dummy atom' ab initio structural modelling of biological macromolecules*, Journal of Applied Crystallography **46**, 1884 (2013).

[12] S. Brownlow, J. H. M. Cabral, R. Cooper, D. R. Flower, S. J. Yewdall, I. Polikarpov, A. C. North, L. Sawyer, *Bovine -lactoglobulin at 1.8 resolution still an enigmatic lipocalin*, Structure **5**, 481 (1997).

[13] K. Röllen, J. Granzin, R. Batra-Safferling, A. M. Stadler, *Small-angle X-ray scattering study of the kinetics of light-dark transition in a LOV protein*, PLOS ONE **13**, 1 (2018).

[14] J. Granzin, A. Stadler, A. Cousin, R. Schlesinger, R. Batra-Safferling, *Structural evidence for the role of polar core residue Arg175 in arrestin activation*, Scientific Reports **5**, 15808 (2015).

[15] Abelès, Florin, *La théorie générale des couches minces*, J. Phys. Radium **11**, 307 (1950).

[16] M. Born, E. Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, W. L. Wilcock, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light.* (Cambridge University Press, 1999).

[17] A. Koutsioubas, *Combined Coarse-Grained Molecular Dynamics and Neutron Reflectivity Characterization of Supported Lipid Membranes*, The Journal of Physical Chemistry B **120**, 11474 (2016).

[18] H. P. Wacklin, *Neutron reflection from supported lipid membranes*, Current Opinion in Colloid & Interface Science **15**, 445 (2010).

[19] T. K. Lind, H. Wacklin, J. Schiller, M. Moulin, M. Haertlein, T. G. Pomorski, M. Cardenas, *Formation and Characterization of Supported Lipid Bilayers Composed of Hydrogenated and Deuterated Escherichia coli Lipids*, PLOS ONE **10**, 1 (2015).

F 1   Neutron Imaging

M. Strobl

Head of Neutron Imaging and Applied Materials Group (NIAG)

Laboratory for Neutron Scattering and Imaging

Paul Scherrer Institut, Villigen, Switzerland

Contents

Introduction

In contrast to other neutron scattering methods, neutron imaging assesses structures and processes with real space resolution on macroscopic length scales [1,2]. In general neutron imaging utilizes the same neutron interactions with matter as used by scattering techniques. Hence, the same qualities of neutron radiation are exploited such as

(i) **good transmission** for many materials, in particular structural materials,

(ii) **high sensitivity to some light elements** and especially hydrogen,

(iii) **isotope sensitivity** with an outstanding example amongst others being hydrogen and deuterium, enaling particularly **contrast variation**,

and last but not least

(iv) the **match of wavelength and energy** in the thermal range with **crystallographic distances and excitations**

(v) the **magnetic moment** of the neutron providing sensitivity to magnetic fields and strcutures.

Correspondingly neutron imaging has a wide range of applications in particular for scientific problems and non-destructive testing where x-ray imaging fails and the above characteristics are required.

An outstanding example being extensive studies in material science and industrial research&development (R&D) in the first two decades of the millenium that helped enable market readiness of commercial fuel cells (Fig. 1)[3-6]. Here several of the above mentioned advantages of neutrons are exploited. Neutrons penetrate the metallic casing of fuel cells, still being very sensitive to the water distribution, which can hence be probed in operando in real operational fuel cells. In order to understand the exchange of water in continuously wet areas of the cell Deuteration has been used, exploiting the isotope contrast. Lately even the use of polarized neutrons has been considered to simultaneously assess current flows within operational cells.
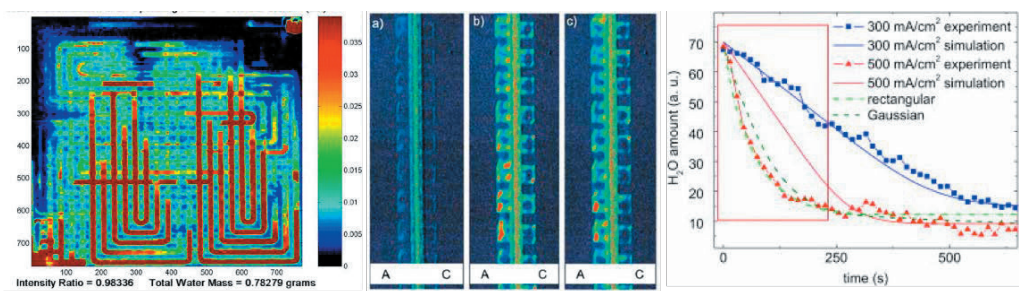
Fig. 1: In operando fuel cell studies mainly focussed on the water balance critical to the performance of fuel cells. The figure left gives an example of a through plane investigation of a full blown fuel cell in operation, displaying on the first glance the distribution of the produced water in the gas supply channels; More details with high spatial resolution are shown in the in-plane images in the middle, displaying individual channels as well as the membranes and gas diffusion layers in between; Finally on the right hand side a quantification and modelling result deduced from studies utilizing H_2O/D_2O isotope contrast. It underlines further that the images

presented rarely are the results of a study but data at an intermediate reduction state finally serving quantitative measurements and modelling. In the specific case it has been found that at higher load the water management cannot be described anymore by a simple two phase flow model, but a discrete, eruptive transport model had to be developed based on the data;

Latest since the introduction of digital detectors for neutron imaging, one has to distinguish between different fundamental imaging modes with respect to the resolved dimensions. While initially, when images where recorded on film, only

- **2-dimensional projection imaging** was available, the digital data recording enabled an extension first to
- **3 dimensions**, either
  (i) **3D tomographic volumes** or
  (ii) **time resolved imaging** combining two spatial dimensions with the time dimension. Latest developments of acquisition technology enables
- **4D imaging**, namely time-resolved tomography, with latest reports having 3 sec time resolution for full 3D tomographies, enabling to observe processes real time in 3 dimensions.
- Taking into account techniques reconstructing 3D vector fields, the dimensionality of resulting data volumes is meanwhile ranging **up to 6D and beyond**.

Currenlty best **resolution**s achieved range down to below **5 μm in 2D imaging** and around **10 μm in 3D** imaging. Best time resolutions provide movies of processes with around **1000 frames per second**, and in so-called strobo-kinetic measurements, that image repetitive processes avergaing over a number of cycles, time resolution in the **microsecond range** are reported. Time resolved tomographies have been performed with **below 10 seconds per tomography** and spatial resolutions then ranging at few 100 μm. Note that spatial resolution is not to be confused with pixel resolution. In the ideal case the spatial resolution, which is defined as the full width at half maximum (FWHM) of the pointspread or at 10% modulation transfer in the reciprocal unit of line pairs per length, is at bbest twice the actual pixel size.

1   Neutron Imaging – Spatial Resolution

A general imaging experiment requires both, spatial resolution and contrast. While we will deal with the different contrast mechanisms in the next section, here we shall focus on spatial resolution.

Contrary to the wide spread believe that the spatial resolution capability of neutron imaging builds upon a highly collimated low divergence beam, state-of-the-art neutron imaging instruments utilize highly divergent beams for efficient imaging of large objects. This misconception is largely based on the fundamental difference, that other methods view the beam as an integrated entity described by overall parameters like wavelength, wavelength spread, polarisation, divergence etc. This is justified by the nature of these measurements integrating parameters over the whole probed area, respectively volume. This is fundamentally different in imaging with instantaneous spatial resolution (in contrast to scanning methods). To provide an example of this issue, we consider the use of a velocity selector for scattering and imaging. The specification of such devices has it, that the beam divergence has an impact on the wavelength resolution provided. This is due to the geometry

of the functional components, basically absorbing radial lamella mounted prallel on a rotation axis, which itself is parallel to the beam. The lamellas are twisted such to enable passage of a specific velocity of neutrons at a certain operation frequency. Obviously the geometry for passage is optimized for a specific beam direction and hence differs for neutrons with angular deviations, which is adding to the wavelength spread of neutrons passing. While the specified wavelength resolution describes the conditions sufficiently for an integrated assessment of the beam, the divergence dependence leads to lateral shift of the central wavelength over a spatially extended beam at the detector, where the spatial extension is based on the initial divergence at the upstream velocity selector device. As a consequence, the local wavelength resolution on an imaging detector might have a better wavelength definition, but the nominal wavelength will shift laterally across the image. Hence, local beam parameters have to be evaluated on the imaging detector rather than the overall, in this case, wavelength and wavelength resolution of the whole beam.

1.1 Spatial resolution - Geometry

Without the availability of sufficient lenses for neutrons the spatial resolution capability of neutron imaging is limited by the applied pinhole camera approach (Fig. 2)[1]. In analogy to the camera obscura technique, but adapted to transmission imaging, the resolution capability is determined by a pinhole D on the radiation source side, the distance L from there to the sample and the distance l from the sample to the detector. The ratio L/D is referred to as the collimation ratio, which defines the divergence of the neutron beam at a point in the sample and hence the principle intrinsic resolution limitation, because the image blur d introduced by this geometry will depend only on the L/D ratio and the sample to detector distance l of individual measurements.

$$d = l/(L/D). \tag{1}$$
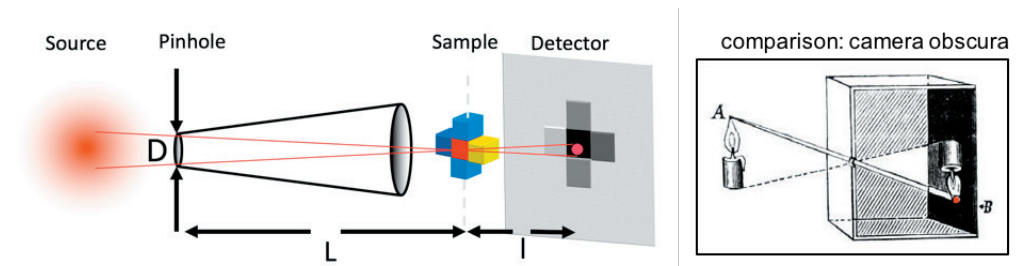
Fig. 2 Schematic of the pinnhole imaging geometry of neutron imaging (left) and the analogy to the camera obscura principle (right); (red dots denote geometrical image blur)

From this it is obvious, that the most important measure to optimize results is to place samples always as close as possible to the detector, to minnimize the introduced image blur independent of other instrument parameters. Any other measure to improve the resolution capability in this geometry is a trade with instrument performance in terms of neutron flux. The parameters left to tune are the pinhole size D and the distance L, where modern instruments offer a variation of both within reasonable limitations. However, these impact the available flux density, as both the reduction of D and the increase L, while leading to improved resolution capabilities, lead to a reciprocal flux density decrease scaling with the

square of the collimation improvement, as can be seen easily, because they both impact 2 dimensions equally (note: flux density).

To further illustrate the limitations met in this respect, let's consider in an otherwise optimized set-up the attempt to improve resolution by a factor of two (in both dimensions, indeed!). Increasing *L/D* accordingly implies a flux density loss of a factor of four as discussed above. However, also the effective pixel size in the detection has to be reduced reciprocally by a factor of two, which is another factor of four in area. Hence, to achieve the same statistical significance an improvement of the resolution by a factor of two implies an increase in the required exposure by a factor of 16(!), i.e. more than an order of magnitude, which is not yet taking into account efficiency losses in the scintillator, when increasing the resolution requires optimizing the scintillator of a state-of-the-art detector. This has to be set in relation to maximum flux densities of the order of 10^8 n/cm^2/s achieved in the best state-of-the-art instruments.

Different approaches of overcoming this situation with neutron focussing devices (lenses) are discussed continuously and recently in particular an approach utilizing so-called Wolter Optics is considered by some to have significant potential for imaging applications [7]. However, to date no significant and reliable gains could be demonstrated and achieved, which would represent a distinct and broad relevance in the field.

1.2   Spatial resolution - Detector

Spatial resolution indeed depends intrinsically also on the ability of spatial detection resolution. With the flux limitations implied by spatial resolution of imaging outlined above, it must appear obvious that not only the spatial resolution capability, but with it also the efficiency of detection plays a crucial role for detectors and detection technology (Fig. 3).

The most utilized standard detector technology for neutron imaging is the scintillator/camera detector [1]. The advantages of this system compared to other digital detectors like e.g. amorphous silicon flat panels or microchannel plate(MCP)/medipix detectors [8] are the high quantum efficiency and particularly the high flexibility in trading spatial resolution and field of view (FOV) (Fig. 3), which are related to each other with respect to the total pixel number of the utilized camera chip. The detector system generally consists of a custom designed light tight box including a window covered by a specific scintillator, a mirror deflecting the light perpendicular out of the direct neutron beam towards the light optics and camera. Light optics (mostly commercial photographic lens systems), camera and scintillator can be exchanged straightforwardly according to the specific needs of individual measurements. Fields of view can vary within a range of about 0.5 x 0.5 m^2 to below 5 x 5 mm^2, corresponding effective pixel sizes from about 500 x 500 μm^2 to 1.5 x 1.5 μm^2. Scintillators of different sizes, thicknesses and materials can be attached. In general higher resolution requires thinner scintillator screens, with correspondingly opposite effects on efficiency and resolution capability (thicker scintillator screens imply up to a certain limit higher efficiency (detection probability), but also stronger blurring of the light output).

The main scintillator screen materials used are: LiF/ZnS where LiF acts as converter and ZnS as actual scintillation material for the secondary radiation from the capture by Li. Often the ZnS is doped, e.g. by Ag, to shift the wavelength of the emitted light to the range of the highest sensitivity of the used sensor chip. Mainly for high resolution also GADOX (Gd$_2$O$_2$S

– Gadulinium Oxysulfide) is used, which provides a higher capture probability essential for the thin screens for high resolution. The lower light output per neutron capture can even be an advantage for signal quality in the limit of low count rates. For extreme spatial resolutions (best currently in standard operation < 5 μm) even scintillators of isotope enriched ^{157}Gd have been developed and are utilized [9].
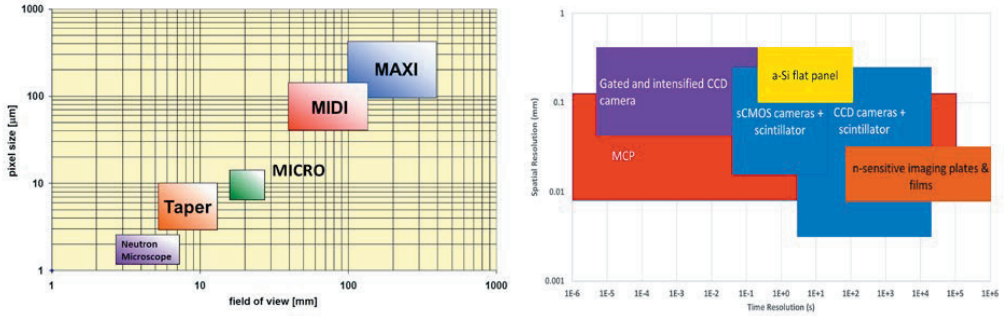
Fig. 3 Example of scintillator/camera imaging detector suite for various spatial resolution versus field-of-view requirements at NIAG at PSI, as well as an overview of utilized neutron imaging detectors in the context of spatial and temporal resolution;

Cameras utilized in neutron imaging feature CCD and CMOS technology with 1024 x 1024 up to 4096 x 4096 pixels. These are generally cooled in operation through air, water or Peltier cooling in order to enable low electronic noise. The choice of camera depends mainly on resolution, noise and exposure time consideration. These cameras are combined with state-of-the-art photographic optics, mostly focal lengths of about 50 mm and 100 mm are in use for large field of view and high resolution, respectively. For extreme resolutions more dedicated scientific optics enabling high light efficiency at high resolution are utilised [10,11].

In general the image detected by a camera detector $I_d(x,y)$ is affected by intrinsic detector noise I_{dc}, dominated by the dark current or read-out noise. An initial image correction is hennce simple background correction with an image I_{dc} recorded with closed shutters according to $I(x,z) = I_d(x,z) - I_{dc}(x,z)$.

Alternative detector solutions convey, but are not limited to, flat panel detectors, which provide digital images on relatively large fields of view, however, with limited resolution in the range of 100 μm, MCP/medipix detectors which provide resolutions down to around 10 μm, however with limited fields-of-view (FOV) of only around 10 cm^2 (Fig. 3). For some applications also film is used for image detection still, providing a unique combination of large fields of view and high spatial resolution, however, digitisation is tidious and prohibits time resolved and tomographic studies. Currently there is a large numer of imaging detector devlopments on the way, triggered mainly by the need for imaging detectors capable of time-of-flight resolution at pulsed neutron sources. In this endeavor the MCP/timepix detector can currently be regarded the benchmark, with some limitations, however, in FOV, count rate capability with respect to the last generation of spallation sources and long term stability for quantitative annalyses.

1.3   Three-dimensional spatial resolution - Tomography

Three dimensional spatial resolution of the internal structure of an object requires a tomographic approach. In general for measurements this implies that a significant number of projection images have to be recorded on an angular range of at least 180 degrees. In order to match the spatial resolution d achieved in the two dimensional transmission images, referred to as radiographies, the number of projections n required depends on the maximum radius of the object R with

$$n = (R/d)\pi/2 \tag{2}$$

The principle equation to retrieve cross section information from projections was derived by J. Radon [12] at the beginning of the 20th century. The basic transformation of a parallel beam tomography is, hence, named Radon transformation and written as

$$P(\theta,t) = \int \int f(x,y)\delta_D(x\cos\theta + y\sin\theta - t)dxdy \tag{3}$$

Here $f(x,y)$ is the two dimensional function representing a slice of the sample which is projected into a number of one-dimensional functions $I(t)$ at specific projection angles θ, with δ_D being the Dirac Delta function. Just like the equation and the term tomography implies three dimensional reconstructions are here based on stacking the reconstruction of two dimensional cross-sectional slices retrieved from the one-dimensional projection functions represented by single lines in the projection images, parametrized with t, while x and y are in the coordinate system of the sample.

To reconstruct the cross section $f(x,y)$ one can explore a large number of paths. The simplest and most straight forward analytical reconstruction is direct backprojection as depicted in Fig. 4, left. In the simple backprojection, the intensities of the measured functions are smeared back onto a pixelated cross section according to their projection angle corresponding to

$$f'(x,y) = \frac{1}{n}\sum_{i=1}^{n} P(\theta_i, x\cos\theta_i + y\sin\theta_i) \tag{4}$$

Figure 4 illustrates, how adding a growing number of projections improves the reconstruction $f'(x,y)$ of the slice. However, the right figure demonstrates more clearly the bias introduced through this simple approach particularly through the incomplete discrete sampling. The most applied reconstruction is hence a filtered backprojection algorithm. It correspondingly utilizes a filter in addition in order to even out the higher density in the center of the image as compared to the edges. The filter is applied in the frequency domain based on the Fourier slice theorem and in its most basic form has the shape $|\omega|$ where the Fourier transform reads

$$\hat{P}(\theta_i,\omega) = \int P(\theta_i,\omega)e^{-i2\pi\omega t}dt \tag{5}$$

and the filter is applied to the Fourier transformed projection $\hat{P}(\theta_i, \omega)$ as

$$P_f(\theta_i, \omega) = \int \hat{P}(\theta_i, \omega)\, e^{-i2\pi\omega t} |\omega|\, d\omega \tag{6}$$

which enables the backprojection of the filtered projection through

$$f'(x, y) = \frac{1}{n} \sum_{i=1}^{n} P_f(\theta_i, x\cos\theta_i + y\sin\theta_i) \tag{7}$$

A comparison of results of a simple shape with and without filtered back projection is shown in Fig. 4 on the right hand side and clearly underlines the requirement for filtering. In reality a number of different filters are used and are available in a wide range of reconstruction programs, commercially as well as freeware [13].
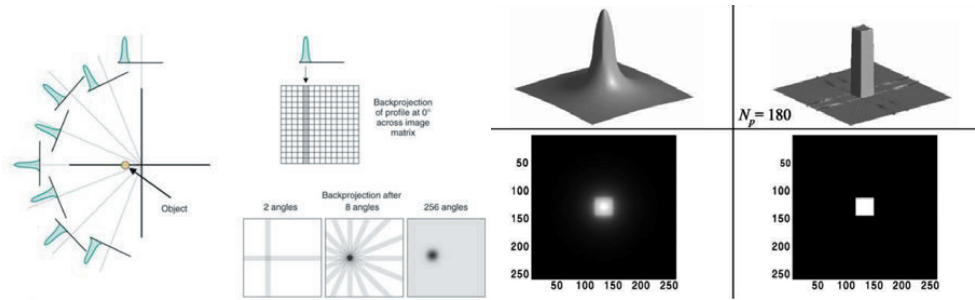
Fig 4 Schematic illustration of the simple backprojection principle and on the right hand side a comparison of a corresponding result with the result of the broadly utilized filtered backprojection reconstruction for a slice of an object with a squared cross section;

However, reconstruction can be done in many ways particularly also through iterative algorithms, which are especially useful for limited data sets and limited data quality. In addition, different beam conditions, such as cone beam geometries or the spatial resolved reconstruction of different parameters require different approaches and a departure from some of the fundamental concepts utilized here. This would be the case e.g. for the reconstruction of vectorial entities instead of scalars. In addition, strictly considered, neutron imaging experiments rather use conic beams, where the independent slice approach is not fully supported. However, within the spatial resolutions utilized normally the parallel beam proves generally a sufficient approximation.

2   Neutron Imaging - Contrast

The second indispensible ingredient for imaging, besides the spatial resolution, is contrast. Here the picture changed significantly within the past one or two decades, adding contrast modalities far beyond the conventional wavelength integrated neutron attenuation [1,14-16]. The basic concepts of contrast with neutrons will, however, be outlined along the lines of this conventional, initial and still dominating contrast modality, which lends its strengths from the most unique characteristics of neutron radiation. It has to be noted that referring here to neutron radiation shall be restricted to the range of thermal and cold neutrons. This does not mean that other regimes, especially of higher energy neutrons are not utilized for imaging. They are. And higher energies even increase the material volumes that can be penetrated,

however, in turn the resolution and particularly the detector efficiency suffer severely. Nevertheless, beyond the mainstream useful applications have also been demonstrated, most notably lately element, isotope and temperature sensitive neutron imaging utilizing resonance absorption [17].

The unique fundamental features of neutron radiation outlined in the introduction imply their value for application in radiographic imaging despite the issues with focussing, flux and spatial resolution stressed in the previous section. The specific characteristics of neutrons in the interaction with matter play a key role in the application of neutron imaging (just as much as in scattering) and most of them are key drivers for conventional attenuation contrast imaging as well, as is illustrated also in Fig. 5.
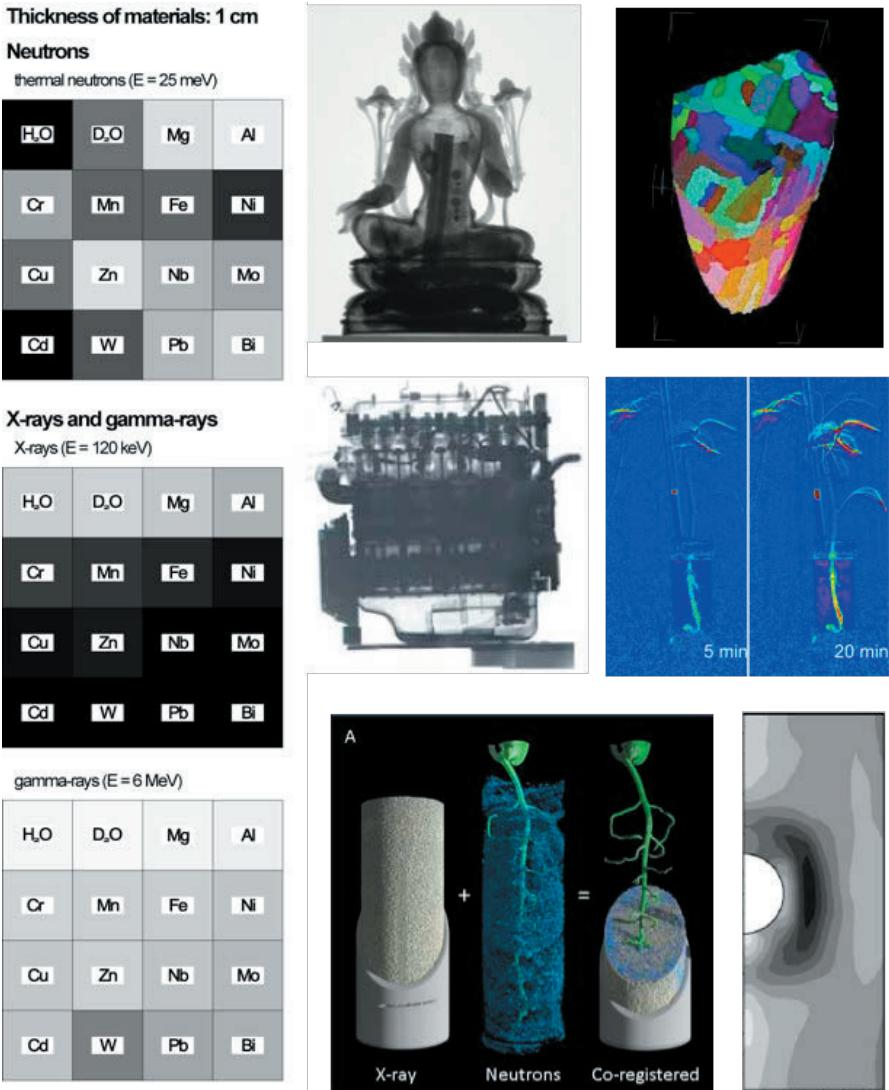
Fig. 5 left hand side: comparison of neutron attenuation contrast with x-rays and gamma rays of certain energies; the scattered contrast and significant differences between

elements close in the periodic tabble in contrast to photons is illustrated; right hand side from left to right and top to bottom: attenuation contrast neuton image of cultural heritage object utilizing the ability of neutrons to transmit the metallic sculpture and still being highly sensitive to hydrogeneous organic material inside, enabling their non-destructive study; dark-field contrast neutron tomography displaying the magnetic domain network in the bulk of an FeNi crystal, where it is not amenabble by any other technique [18]; attenuation contrast neutron image of a whole BMW motor block, with neutrons still being sennsitive to the oil distribution inside; two slices of a time resolved attenuation contrast neutron imaging study utilizing D_2O/H_2O contrast to study the wateruptake of a plant; study of the interactions in a root-soil-water system through bi-modal simultaneous neutron and x-ray tomography, where the x-rays better resolve the soil structure, while neutrons provide better contrast for water and root; diffraction contrast neutron image providing a through plane strain map around a cold expanded hole of 1 cm diameter in steel [19];

In order to correctly detect an image $I(x,z)$ based on the contrast provided by the sample the measured image $I_m(x,z)$ has to be background corrected by an image $I_{bg}(x,z)$ and normalized by the incident beam profile measured as $I_0(x,z)$, according to

$$I(x,z) = (I_m(x,z) - I_{bg}(x,z))/(I_0(x,z) - I_{bg}(x,z)) \tag{8}$$

Note that $I_0(x,z)$ does not always have to be an open beam, or so-called flat-field image, but e.g. when observing changes over time, it can be very useful to normalize with the image of the initial state at time $t = 0$ (compare images Fig. 5 mid row, right hand side). Note further, that the background image can contain a significant number of constituents, starting from the discussed detector read-out noise and might even contain biasing contributions from the sample itself. Correction strategies in particular for measurements requiring an extensive level of quantification are not limited to simple or more sophisticated measurements, but are sometimes including complex correction algorithms [20].
(In the following sections in equations, image notation will assume background not further relevant and will be simplyfied accordingly, if not locally stated differently.)

2.1 Attenuation contrast

Attenuation is often wrongly referred to as absorption in imaging. However, in particular in cold and thermal neutron imaging scattering contributions are in many cases dominating the measured attenuation contrast. The entity providing contrast and hence measured is referred to as the linear attenuation coefficient $\mu = N\sigma_{tot} = N(\sigma_a + \sigma_s)$ and correspondingly is based on the total microscopic cross section σ_{tot}, which is the sum of the microscopic absorption σ_a and scattering cross section σ_s. With N being the particle density the linear attenuation coefficeint has the unit of a length, generally given in cm^{-1}. Therefore the total contrast measured with respect to the incident beam depends also on the material thickness and the equation constituting an image is a two dimennsional version of the Beer Lambert law

$$I(x,z) = I_0(x,z)e^{-\int \mu(x,y,z)dy}$$
$$\tag{9}$$

and consequently the relevant projection image will rather read

$$P(x,z) = -\ln\frac{I(x,z)}{I_0(x,z)} = \int \mu(x,y,z)dy$$

(10)

A number of example images as well as contrast to be expected for different elements are displayed in figure 5. However, only one 3D example is shown, only four out of the six images are conventional attenuation contrast images and the data presented is more or less heavily processed. Apart from the introduced backgrounnd correction, in many cases imaging data requires further processing through filtering as it is affected e.g. by Poisson noise in the counting statistics, and e.g. tomographic reconstruction can amplify biases. Sophisticated filtering operation are hence often key for good results but operate at the edge of resolution diluation and bias. Fig. 6 nicely illustrates differennt filter attempts and a successful filter operation on a tomographic slice (Fig. 6 left).
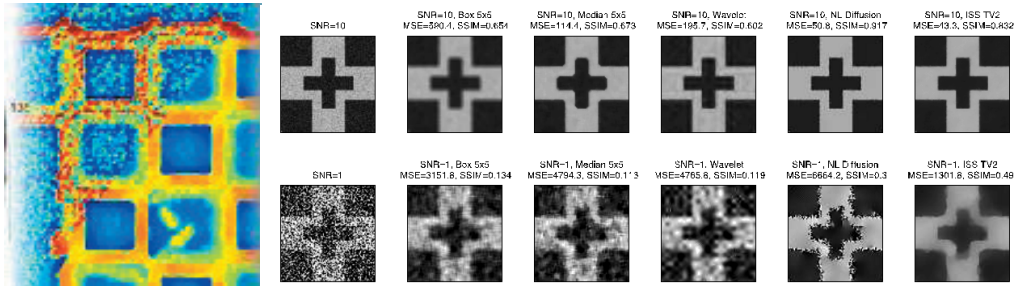
Fig. 6 Initial image quality and filtered result in a tomographic slice of a diesel particulate filter ob the left; various image filtering examples for initial images of different statistical quality;

In general the images produced and presented here are, however, still intermediate results, and processing to get these can be regarded data reduction processes, but the actual analyses is still to come. In contrast to scattering techniques, which continuously produce similar results and data sets, which are fitted with similar models, always looking at the same kind of problem, the situation is significantly different in imaging, where depending on the image and sample, the features to be extracted from conventional attenuation contrast images can be as different as the identification of a single failure in a component in non-destructive testing, to detailed structural analyses and modelling of processes in material science similar to any other material science characterisation technique, including neutron scattering (compare Fig. 1).

However, the application of a broad spectrum in favor of a high flux for high resolution imaging implies potential bias to quantification based on the attenuation coefficient. The energy dependence of $\mu(x,y,z,E)$ does not only imply that the applied energy spectrum has to be taken into account in any quantification of the attenuation coefficient distribution, but also, that significant variations with the energy can lead to effects similar to what is known as beam hardening in x-ray imaging, were penetration increases toward harder, higher energy, radiation. In imaging with neutrons in the thermal and in particular cold energy range, a very peculiar wavelength dependence (in modern neutron imaging the characterisation by wavelength replaces the characterisation through energy) characterizes the transmission spectra of crystalline materials (with high coherent scattering cross sections). Applying wavelength resolved imaging, this specific diffraction related feature can be exploited in what

is refered to as diffraction contrast [14]. Eventually, it makes imaging kind of just another scattering technique (Fig. 7 right), but operating with intirnsic real space resolution.

2.2   Diffraction contrast – Bragg edges

Recalling that the attenuation contrast is based on the sum of the absorption and scattering cross sections it is worthwhile particularly for crystalline materials to have a closer look into the contributions to the scattering cross section for cold and thermal neutrons (Fig. 7). The scattering cross section can further be broken down according to different interaction characteristics into coherent and incoherent as well as into elastic and inelastic contributions providing the options of coeherent elastic (Fig. 7 black) and coherent inelastic (Fig. 7 red) as well as incoherent elastic and incoherent inelastic scattering (Fig. 7 green and blue) contributions.
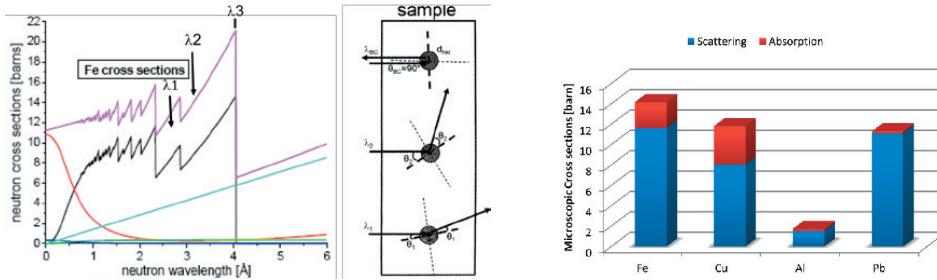
Fig. 7 Different contributions to the total neutron cross section (magenta) depending on neutron wavelength focussing on the cold and thermal energy range, using Fe as an example with a dominating part of coherent elastic (Bragg) scattering (black line); mid schematic together with left hand side graph illustrate the principle of the Bragg edges in the cross section; right hand side: scattering versus absorption contributions for four important metallic materials;

Fig. 7 left side depicts the individual contributions for iron, and it can be seen that in this case the coherent elastic cross section, through the distinct Bragg scattering, dominates the attenuation in the thermal and cold range up to around 4 Å. The turquoise line representing the absortion cross section is dominating beyond this point with a linear growing tendency with increasing wavelength (decreasing energy), while the mangenta line is the sum of all contributions. The coherent elastic Bragg scattering induces a distinct pattern characterised by well defined Bragg edges. The middle figure 7 is to illustrate the principle causing the Bragg edge pattern. At a specific hkl lattice plane family neutrons of a certain wavelength can deflect only at certain orientations of the lattice to the beam according to Bragg's law

$$\lambda = 2d_{hkl} \sin\theta \tag{11}$$

which implies that above a certain wavelength $\lambda = 2d_{hkl}$ where the Bragg angle is $\theta=\pi/2$ no more Bragg scattering can take place for wavelengths longer than two times the specific lattice distance d_{hkl}, and hence the transmission increases steeply beyond that point in the spectrum.

This, however, implies, that there is significant crystallographic information encoded in the transmission spectra, which can be exploited with spatial resolution through wavelength resolved attenuation contrast imaging, providing diffraction contrast [14].

Instrumental attempts to do so are manifold. At continuous sources velocity selctors and double crystal monochromators are utilized for low (around $d\lambda/\lambda$ 10%) and medium (around $d\lambda/\lambda$ 3%) wavelength resolution in conventional imaging instruments, respectively. The utilisation of time-of-flight for imaging has been explored at conntinuous and pulsed sources, at the latter in particular for high wavelength resolutions $d\lambda/\lambda$ below 1%.

Applications convey quantitative investigations of phase fractions and phase transformation with 2D and 3D spatial resolution as well as strain mapping with latest attempts proving the principle feasibility of strain tomography. In addition, spatial varitions of texture can be visualised and grain size analyses has been attempted. In Fig. 5 in the bottom right corner the first successfully measured strain map around a cold expanded hole (10mm diameter) in an 12mm thick ferritic steel plate is presented. The grey scale is spread from strains below 0 being white and above 900 microstrain being black [19].

Strains and phases can be evaluated particularly for powder like samples with random grain orientations. Strains are characterized by small shifts of Bragg edges. To measure such, good wavelength resolution is required. This resolution can be relaxed for phases, which can be quantified based on edge heights. Measurements of two discrete wavelengths, sometimes even a single monochromatic measurement, are hence sufficient for the latter, while the edge analyses for strain requires detailed sampling of edges. In the case of texture the distinct Bragg edge pattern characteristic of a crystal structure vanishes depending on the degree of texture as can be seen in the middel graph in Fig. 8 top. Finally very large grains, within the spatial resolution of the imaging set-up (around 100 μm), show the signature of discrete Bragg peaks rather than a Bragg edge structure. These patterns (Fig. 8 top right) can be used in tomographic approaches to reconstruct the grain network and morphology of a sample (Fig. 8 bottom right) [21].
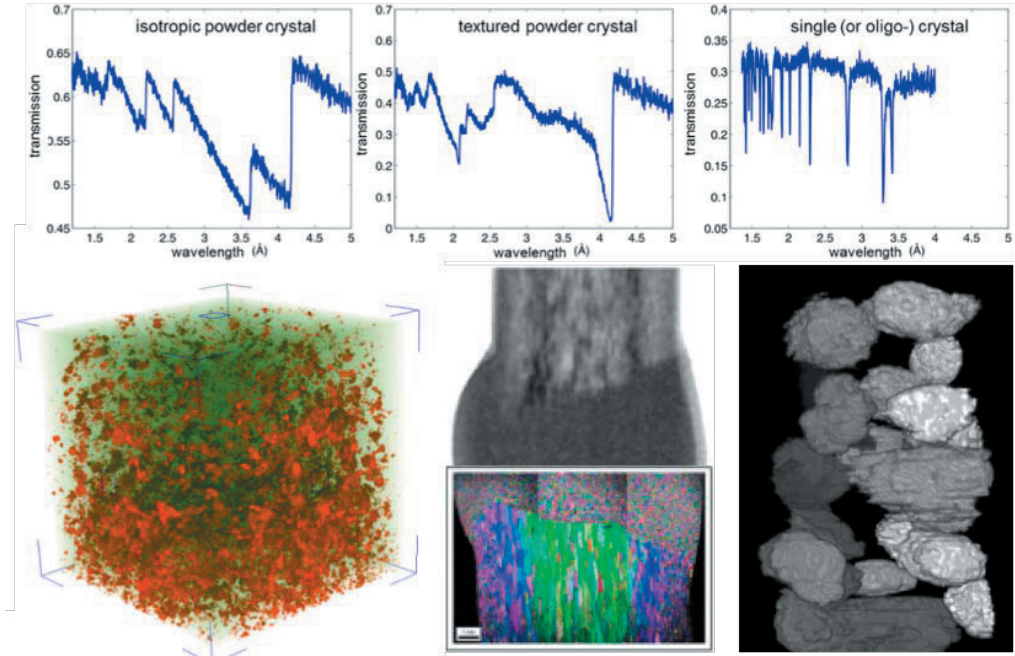
Fig. 8 Top: transmission spectra illustrating 3 distinct cases of powder like random oriented small grained crystallic samples, similar but textured samples and with grain sizes within the range of direct spatial resolution of the imaging set-up annd hence limited number of grains; Bottom: application examples convey mapping of crystalline phases (here mix of α and γ phase in steel in 3D), varying grain structure in a weld of rolled steel and single grains in an Fe sample reconstructed tomographically [21];

2.3 Dark-field contrast – Small angle scattering

Dark-field contrast can be achieved, when an imaging set-up is sensitive to small angle scattering. Nowadays this is realized through spatial modulation of the full extended beam utilized for imaging [15]. Spatially resolved small angle scattering is then detected by local losses of the modulation visibility V [22], which is defined as

$$V=(I_{max}-I_{min})/(I_{max}+I_{min}) \tag{12}$$

Different approaches exist to induce and detect spatial beam modulation and in particular its visibility [15]. The utilized modulation period p has a key influence on the scattering vector range probed and whether it is ranging in the ultra-small, very-small or small angle scattering range [15,23]. Tthe smallest modulations are produced in the micrometer range and are hence beyond the spatial resolution of the imaging set-up. Therefore such modulations have to be measured and analysed within every individual pixel. This is normally achieved with analyser gratings with the period of the modulation, and a scan of these enables the step-wise visualistation of the modulation in each pixel. If the modulation is spatially resolved, then it is superimposed to the attenuation image, and the spatial resolution of the small angle scattering information is limited by the modulation period [24].

Micrometer sized modulation periods are realized with Talbot-Lau grating interferometers, and larger periods e.g. in so-called far-field grating interferometers, using Moire fringes and even Spin-Echo devices [15,24].

Even applications not quantifying the small angle scatter signal have proven very use- and successful not only to detect structural variation but in particular magnetic domain walls, which can be directly visualized. Fig. 5 top right shows the first ever observation of bulk magnetic domains through the application of dark-field contrast neutron tomography [18].
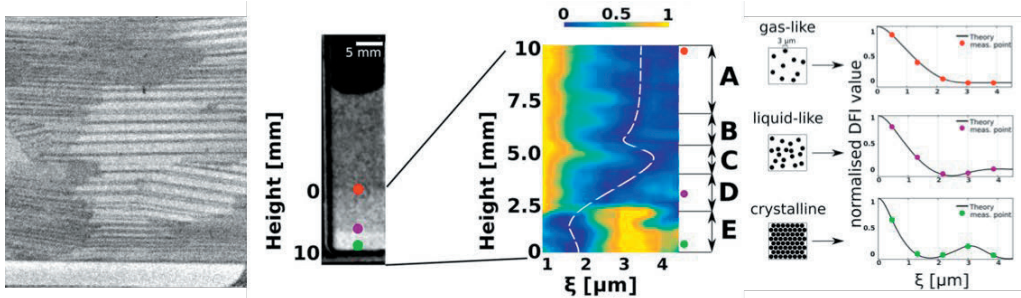
Fig. 9 left: magnetic domain walls in GOSS oriented steel used in transformers [25], mid to right: height resolved different phases in a system of mircospheres diluted in an H_2O/D_2O mixture and forming a crytal-like phase when sedimenting on the bottom of the imaged cuvette [26];

However, dark-field contrast is also suited to quantify small angle scattering and hence structures in the micro- to nanometer regime, beyond direct spatial resolution of neutron imaging. The particular correlation length ξ probed in a measurement can be described as [23]

$$\xi = \lambda L_s / p \tag{13}$$

and hence depends apart from the period p also on the utilized wavelength and the sample to detector/analyser grating distance L_s. When the period is fixed (like in grating set-ups in contrast to magnetic set-ups in spin-echo) the wavelength and to some extend the distance (note, with corresponding impact on spatial image resolution) are appropriate parameters to alter ξ in order to scan a specific correlation length range. Correlation length ranges beyond one order of magnitude have been reported within boundaries of some 10 nm up to nearly 10 µm [15].

Small angle scattering structures can pixel-wise or within the spatial resolution of the image be characterised due to the relation [23]

$$V(\xi, x, z) \Big/ V_0(\xi, x, z) = e^{\int \Sigma_s(x,y,z)(G(\xi,x,y,z)-1)dy} \tag{14}$$

Here V_0 is the visibility of the modulation without sample, Σ_s is the small angle scattering cross section based on the scattering length density contrast of the materials and the density of

the scattering structures as well as the wavelenth square. $G(\xi)$ a projected real space correlation function of the scattering structure containing the morphology information [23]. Note that this correlation function is reduced to one dimension according to the currenntly used modulation techniques being sensitive in one dimension only, comparable to slit-smeared small angle instruments. In contrast to small angle scattering instruments, here measurements remain in real space, because the modulation technique constitutes a backprojection from Fourier to real space according to [23]

$$V_s(\xi, q) = V_0(\xi) \int_{-\infty}^{\infty} dq \, S(q) \, cos(\xi q)$$

(15)

where $S(q)$ is the small angle scattering function and q the modulus of the scattering vector.

Applications are in the study of structures and structural developments which are not homogeneous throughout the sample volume as required in conventional SANS. Fig. 9 provides a few examples of dark-field contrast imaging, including one exploiting the spatial resolved quantitative SANS evaluation.

2.4   Polarisation contrast – Spin precession

The sensitivity of neutrons to magnetic fields and structures is one of the main applications in neutron science and in particular neutron scattering. However, this specific feature can also be exploited in spatially resolved polarized neutron imaging studies in the macroscopic domain [16, 27]. While examples have already been shown with respect to dark-field contrast, polarisation contrast utilizes polarisation analyses and spin precession in order to visualize and analyse magnetic structures and fields. The basic principle of polarized neutron imaging builts on the Larmor precession of the neutron spin when traversing a magnetic field.
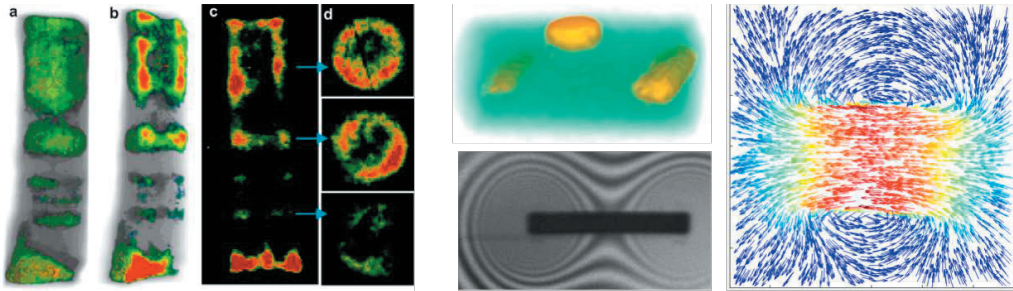
Fig. 10 left: Curie temperature mapping through depolarisation imaging, providing 3D maps of the already ferromagnetic parts of a sample at a certain temperature [16]; mid top: 3D reconstrcution of magnetic field areas trapped in a lead superconductor close to the critical temperature [27]; mid bottom: straightforward polarized neutron radiography of the field of a dipol magnet [28] and right hand side: a result of the first 3D magnetic field reconstruction from a polarimetric neutron tomography (color indicates field strength)[29];

The spin precession of the polarisation vector **P** around a magnetic field vector **B** can be expressed by

$$\frac{d\boldsymbol{P}}{dt} = \gamma \boldsymbol{P} \times \boldsymbol{B} \tag{16}$$

and hence the accumulated rotation angle

$$\varphi = \omega_L t = \frac{\gamma}{v} \int_{path} B ds = \frac{\gamma m \lambda}{h} \int_{path} B ds \tag{17}$$

where γ = -4$\pi\mu$/h = 1.832 10^8 rad/s/T being the gyromagnetic factor of the neutron, with h being the Planck constant. v denotes the velocity of the neutron and m its mass. Utilizing a polarized incident beam and a polarisation analyser between sample and detector and for illustration assuming a magnetic field perpendicular to the polarisation of the beam the signal in the resulting image can be described by

$$I(x, y) = I_0(x, y) \exp\left(-\int \Sigma(s) ds\right)(1 + \cos\varphi)/2 \tag{18}$$

and when normalised for the attenuation simply

$$I(x, y) = I_0(x, y)(1 + \cos\varphi)/2 \tag{19}$$

This enables a visualisation, but a quantification only for simple cases, sufficient a priori knowledge or significant field modelling efforts. Nevertheless the detection of magnetic field variations in the bulk of materials, where it is not amenable to any other technique has proven useful. In addition, more complex techniques have developed and meanwhile the 3D reconstruction of magnetic vector fields has become possible. Also simple implementations, not striving for reconstructing the field distribution but to detect e.g. the phase transition from paramagnetic to ferromagnetic at the Curie temperature simply probing depolarisation turned out interesting for numerous applications as well. Fig. 10 provides examples of different complexity and applications.

References

[1] M. Strobl, N. Kardjilov, A. Hilger, I. Manke, J. Banhart, Topical Review: Advances in neutron radiography and tomography, J. Phys. D 42 (2009) 243001.

[2] N. Kardjilov, I. Manke, A. Hilger, M. Strobl, J. Banhart, Neutron imaging in material science, Materials Today, 14(6) (2011) 248–256.

[3] Hickner M A, Siegel N P, Chen K S, Hussey D S, Jacobson D L and Arif M, Understanding liquid water distribution and removal phenomena in an operating PEMFC via neutron radiography J. Electrochem. Soc. 155 (2008) B427.

[4] J. Biesdorf, P. Oberholzer, F. Bernauer, A. Kaestner, P. Vontobel, E. H. Lehmann, T. J. Schmidt und P. Boillat, Dual spectrum neutron radiography: Identification of phase transitions between frozen and liquid water, Physical Review Letters 112, 24 (2014) 248301.

[5] I. Manke, Ch. Hartnig, N. Kardjilov, M. Messerschmidt, A. Hilger, M. Strobl, W. Lehnert, J. Banhart, Characterization of water exchange and two-phase flow in porous gas diffusion materials by H-D contrast neutron radiography, Appl. Phys. Lett. 92, 244101 (2008).

[6] A. Bazylak, Liquid water visualization in PEM fuel cells: a review, International Journal of Hydrogen Energy 34, 9 (2009) 3845-3857

[7] D.F.R.Mildnera, M.V.Gubarev, Wolter optics for neutron focusing, Nucl. Instr. Meth. A 634, 1 (2011) 7-11.

[8] O. H. Siegmund, J. V. Vallerga, A. S. Tremsin, J. Mcphate and B. Feller, High spatial resolution neutron sensing microchannel plate detectors Nucl. Instrum. Methods A 576 (2007)178.

[9] P. Trtik, E. Lehmann, Isotopically-enriched gadolinium-157 oxysulfide scintillator screens for the high-resolution neutron imaging, NIM-A, 788 (2015) 67-70.

[10] P.Trtik, E. Lehmann, Progress in high-resolution neutron imaging at the Paul Scherrer Institut – The Neutron Microscope project, Journal of Physics - Conference Series, 746 (2016) 012004.

[11] M. Morgano, P. Trtik, M. Meyer, E. H. Lehmann J. Hovind and M. Strobl, Unlocking high spatial resolution in neutron imaging through an add-on fibre optics taper, Optics Express 26, 2 (2018) 1809.

[12] Radon J 1917 Ber. Saechsischer Akad. Wiss. 29 262.

[13] A.P. Kaestner, MuhRec - a new tomography reconstructor, NIMA 651, 1 (2011). software: https://www.psi.ch/niag/muhrec

[14] R. Woracek, J. Santisteban, A. Fedrigo, M. Strobl, Diffraction in neutron imaging—A review, Nucl. Inst. Meth. A 878 (2018) 141–158.

[15] M. Strobl, R. P. Harti, C. Grünzweig, R. Woracek, J. Plomp, Small Angle Scattering in Neutron Imaging—A Review, J. Imaging 3 (2018) 64.

[16] M. Strobl, H. Heimonen, S. Schmidt, M. Sales, N. Kardjilov, A. Hilger, I. Manke, T. Shinohara, J. Valsecchi, Jacopo Topical review: Polarisation measurements in neutron imaging, J. Phys. D, in the press.

[17] A. S. Tremsin, J. B. McPhate, J. V. Vallerga, O. H. W. Siegmund, W. Kockelmann, E. M. Schooneveld, N. J. Rhodes, W. Bruce Feller, High Resolution Neutron Resonance Absorption Imaging at a Pulsed Neutron Beamline, IEEE Transactions on Nucl. Sci. 59, 6 (2012).

[18] I. Manke, N. Kardjilov, R. Schäfer, A. Hilger, M. Strobl, M. Dawson, C. Grünzweig, G. Behr, M. Hentschel, C. David, A. Kupsch, A. Lange, J. BanhartThree-dimensional imaging of magnetic domains, Nature Commun. 1, 125 (2010).

[19] J.R. Santisteban, L. Edwards, M.E. Fitzpatrick, A. Steuwer, P.J. Withers, M.R. Daymond, M.W. Johnson, N. Rhodes, E.M. Schooneveld, Strain imaging by Bragg edge neutron transmission, Nucl. Instr. Meth. A 481 (2002) 765–768.

[20] P. Boillat, C. Carminati, F. Schmid, C. Grünzweig, J. Hovind, A. Kaestner, D. Mannes, M. Morgano, M. Siegwart, P. Trtik, P. Vontobel, and E.H. Lehmann, Chasing quantitative biases in neutron imaging with scintillator-camera detectors: a practical method with black body grids, Optics Express 26, 12, (2018) 15769-15784.

[21] A. Cereser, M. Strobl, S. A. Hall, A. Steuwer, R. Kiyanagi, A. S. Tremsin, E. B. Knudsen, T. Shinohara, P. K. Willendrup, A. Bastos da Silva Fanta, S. Iyengar, P. M. Larsen, T. Hanashima, T. Moyoshi, P. M. Kadletz, P. Krooß, T. Niendorf, M. Sales, W. W. Schmahl, and S. Schmidt
Time-of-Flight Three Dimensional Neutron Diffraction in Transmission Mode for Mapping Crystal Grain Structures, Scientific Reports 2017; 7: 9561.

[22] M. Strobl, C. Grünzweig, A. Hilger, I. Manke, N. Kardjilov, C. David, F. Pfeiffer, Neutron dark-field tomography, Phys. Rev. Lett. 101, 123902 (2008).

[23] M. Strobl, General solution for quantitative dark-field contrast imaging with grating interferometers, Scientific Reports 4 (2014) 7243.

[24] M. Strobl, M. Sales, J. Plomp, W. G. Bouwman, A. S. Tremsin, A. Kaestner, C. Pappas, K. Habicht, Quantitative Neutron Dark-field Imaging through Spin-Echo Interferometry, Scientific Reports, 5, 16576 (2015).

[25] R. P. Harti, M. Strobl, R. Schäfer, N. Kardjilov, A. S. Tremsin & C. Grünzweig, Dynamic volume magnetic domain wall imaging in grain oriented electrical steel at power frequencies with accumulative high-frame rate neutron dark-field imaging, Scientific Reports 8, 15754 (2018).

[26] R. P. Harti, M. Strobl, B. Betz, K. Jefimovs, M. Kagias, C. Gruenzweig, Sub-pixel correlation length neutron imaging: Spatially resolved scattering information of microstructures on a macroscopic scale, Scientific Reports 7:44588 (2017).

[27] N. Kardjilov, I.Manke, M. Strobl, A. Hilger, W. Treimer, M. Meissner, T. Krist, J. Banhart
Three-dimensional imaging of magnetic fields with polarized neutrons, Nature Physics, 4 (2008).

[28] M. Strobl, N. Kardjilov, A. Hilger, E. Jericha, G. Badurek, I. Manke, Imaging with polarized neutrons, Phys. B 404 (2009) 2611–2614.

[29] M. Sales, M. Strobl, T. Shinohara, A. Tremsin, L. Theil Kuhn, W. Lionheart, N. Desai, A. Bjorholm Dahl, and S. Schmidt, Three Dimensional Polarimetric Neutron Tomography of Magnetic Fields, Scientific Reports 8, 2214 (2018).

F 2 Photon-Based Spectroscopies of Electronic and Chemical States

C.M. Schneider

Peter Grünberg Institut PGI-6

Forschungszentrum Jülich GmbH

Contents

1   Introduction

Nowadays the study of electronic and chemical states in modern condensed matter physics is largely based on a versatile portfolio of spectroscopies employing excitations with photons. A powerful toolbox of available photon sources covers a broad range of wavelengths from the infrared to the hard x-ray regime, and allows one to choose well-defined polarization states and time structures even down to the femtosecond level. We can distinguish two main classes of spectroscopies: (i) photon-in/photon-out and (ii) photon-in/electron-out approaches. The first class particularly contains photoabsorption spectroscopy, which probes the empty electronic states in a solid. The second class contains photoelectron spectroscopy, the tool of choice to address occupied electronic states. These techniques permit access to a very wide variety of materials and their electronic structure and chemical behavior. Consequently, there is an enormous wealth of results and interesting examples on different systems available, which are well worth being discussed. For good reasons, this lecture must focus on a few essential aspects and a very personal selection of examples. For an in-depth study of photon-based spectroscopies and phenomena studied by them, the reader is referred to a number of excellent textbooks and review articles covering this field [1–6].

2   Principles of Photoexcitation

The common basis of all photon-based spectroscopies is the interaction of the photon with the electrons in a solid, a molecule, or even an atom. In order to derive the details of the excitation process caused by the photon we need to look at the electronic structure first. The electronic structure of a solid can be calculated – in principle – by solving a Schrödinger (nonrelativistic approximation) or Dirac equation (relativistic interactions included) for the respective lattice structure. For most metallic systems with weak electronic correlations, i.e. being close to the limit of a homogeneous electron gas, a quite successful description has been achieved within the framework of density functional theory (DFT) using the local density (LDA) or more general local spin density approximations (LSDA) for the exchange-correlation potential [7]. In order to discuss the salient features of the photoexcitation process, we will first adopt this effective single-particle picture, although we are well aware of the fact that it has shortcomings capturing electronic correlations properly.

2.1   Ground state electronic structure

One important result of the theoretical treatment is that the quantum mechanical wave function describing an electronic state in the solid depends on the symmetries of the Hamiltonian (e.g. lattice symmetries, inversion symmetry, time reversal symmetry, etc.) and for reasons of completeness must include also the electron spin. In the single particle picture the Hamiltonian can be written as

$$\left(\left[\frac{1}{2m} \left(\mathbf{p} - \frac{e}{c} \mathbf{A} \right)^2 + eV(\mathbf{r}) \right] + i \frac{e\hbar}{4m^2 c^2} \mathbf{E} \cdot \mathbf{p} - \frac{e^2 \hbar}{2mc} \sigma \cdot \mathbf{B} - \frac{e\hbar}{4m^2 c^2} \sigma \cdot (\mathbf{E} \times \mathbf{p}) \pm eV_{exc}^{\uparrow\downarrow}(\mathbf{r}) \right) \phi = E_{nls}(\mathbf{k}) \phi. \tag{1}$$
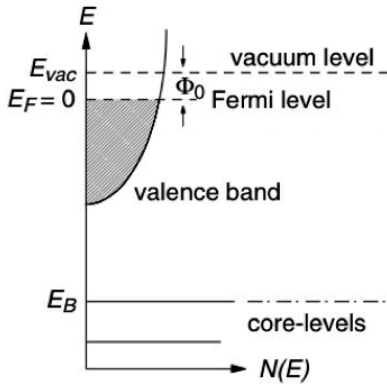
Fig. 1: Simplified picture of an electronic structure in a metallic solid in terms of the density of states $N(E)$. The electronic states are filled up to the Fermi level, which lies in the valence band. Above the valence band the electronic states are empty. At higher binding energies E_B one finds localized core electrons leading to sharp levels in $N(E)$. Electrons can only leave the crystal, if they are excited above the vacuum level, the minimum energy necessary refers to the Fermi level and is called work function Φ_0. After [3].

with ϕ and $E_{nls}(\mathbf{k})$ denoting the single electron wave function and energy eigenvalue, respectively. The terms in square brackets in eq. 2.1 represent the Hamiltonian of a system subjected to an electromagnetic field (vector potential $\mathbf{A}$). This part contains all crystalline symmetries through the potential $V(\mathbf{r})$. The Darwin term ($\sim \mathbf{E} \cdot \mathbf{p}$) may be understood as a relativistic correction to the electron energy. The fourth term contains the interaction of the spins – described by the Pauli spin matrices σ – with an external magnetic field $\mathbf{B}$. The last two terms contain the spin-dependent interactions through spin-orbit coupling and the exchange-correlation potential $V_{exc}^{\uparrow\downarrow}(\mathbf{r})$. The latter is responsible for the formation of spontaneously ordered magnetic states in solids. All spin-dependent terms in the Hamiltonian tend to reduce the symmetry of the system in one way or the other, leading to the splitting and hybridization of degenerate states. A full set of energy eigenvalues obtained from the DFT treatment forms a band or valence electronic structure $E_n(\mathbf{k})$ of the solid with the band index n and the electron wave vector $\mathbf{k}$. The wave functions are Bloch functions and are further classified by the orbital momentum quantum number ℓ and the spin quantum number s. Formally, the respective states may be written as $|n, \ell, \mathbf{k}, s\rangle$ with their energy eigenstates $E_{nls}(\mathbf{k})$. Another important quantity derived from the valence electronic structure is the density of states (DOS) $N(E)$, which is obtained by summing up all electronic states for a given binding energy E. The example in Fig. 1 displays the simplified case of a single nearly-free electron-like band, which leads to a parabolic DOS. In realistic cases the DOS is highly structured [8].

In addition to this valence electronic states comprising delocalized electrons, the full electronic structure of a solid also contains atomic-like localized core levels at higher binding energies $E_B \gtrsim 30$ eV (the binding energy is referred to the Fermi energy, i.e. $E_B = 0$ at E_F). In the ground state, the core states are completely occupied, whereas the valence states are occupied up to the Fermi energy E_F in the case of metals. Above E_F the electronic states are empty. As a consequence, the Fermi level separates the density of states $N(E)$ into an occupied and an empty part.

In semiconductors and insulators the Fermi level lies in a band gap and the intrinsic bulk states are occupied only up to the valence band edge in the undoped case. At semiconductors band bending and surface photovoltage phenomena occur, which complicate the intepretation of photoemission results from semiconductors [9].
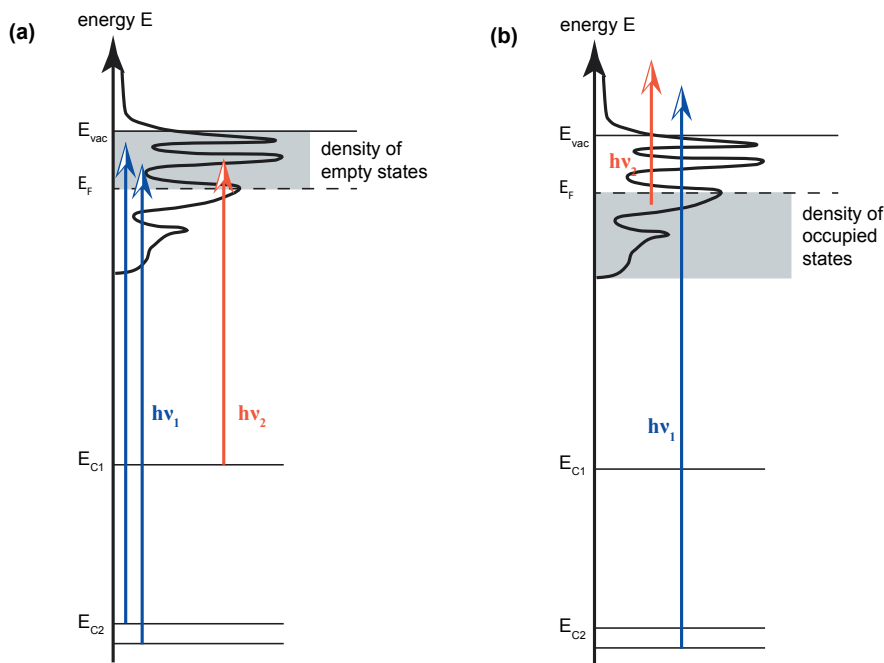
Fig. 2: (a) Schematics of the photoexcitation process determining photoabsorption spectroscopies. (b) Photoexcitation process resulting in photoelectron emission.

2.2 Electronic transitions

Irradiating such an electronic system with photons of a given energy $h\nu$ will excite (photo)-electrons from occupied into unoccupied states in the electronic structure, generally resulting in a reduction of the photons transmitted through the sample. If the unoccupied states are located *below* the vacuum level E_{vac} the excited photoelectrons will remain in the solid (Fig. 2a). However, there may be secondary processes, which generate a cascade of secondary electrons leaving the crystal. This secondary electron yield is a measure of the absorption (see Chapt. 3). Since the occupied state is usually a core level, the absorption spectra, which are recorded as a function of photon energy $h\nu$ contain chemical information. At the same time this absorption spectroscopy probes the unoccupied density of states.

If the transition takes the photoelectrons *above* E_{vac}, they can leave the crystal and can be measured by an electron spectrometer placed in front of the sample (Fig. 2b). The spectra are usually recorded for a well-defined photon energy as a function of kinetic energy E_{kin} and momentum vector **k** of the photoelectrons and yield characteristic signatures of the occupied valence electronic states and/or the core levels.

The quantum mechanical essence of the photoexcitation process is captured in Fermi's Golden Rule, which describes the transition probability between two electronic levels $|i\rangle$ and $|f\rangle$ with binding energies E_i and E_f, respectively:

$$P_{i \to f} = \frac{2\pi}{\hbar} |\langle f| \mathcal{O} |i\rangle|^2 \, \delta(E_f - E_i - h\nu) \tag{2}$$

In the simplest approach, the two levels $|i\rangle$ and $|f\rangle$ may be taken from the ground state electronic structure of the solid – which neglects the role of electronic correlations in the excitation process, as we will see below. The most important quantity in eq. (2) is the transition matrix element

$$M_{fi} = \langle f| \mathcal{O} |i\rangle \tag{3}$$

which essentially depends on the symmetries of the electronic wave functions and the photonic operator $\mathcal{O}$, whereas the delta function ensures energy conservation in the excitation process. For low photon flux densities the operator $\mathcal{O}$ can be treated within linear response theory and takes the form

$$M_{fi} = \frac{-e}{mc} \langle f| \mathbf{A}(\mathbf{r}) \cdot \mathbf{p} |i\rangle \tag{4}$$

with $\mathbf{A}(\mathbf{r})$ the vector potential of the electromagnetic field and $\mathbf{p}$ the momentum operator. It is usually assumed that the wavelength of the electromagnetic field is large compared to interatomic distances, i.e. $\mathbf{A}(\mathbf{r})$ varies only marginally in the spatial region contributing to the transition matrix element [1]. This view is commonly known as *dipole approximation* and simplifies the transition matrix element to

$$M_{fi} = \frac{-ie}{\hbar c} A_0 (E_f - E_i) \langle \psi_f| \mathbf{e} \cdot \mathbf{r} |\psi_i\rangle \tag{5}$$

with the complex amplitude of the vector potential A_0, its polarization vector $\mathbf{e}$, and the wavefunctions of the final and initial states ψ_f and ψ_i, respectively. This form of the transition matrix element is extremely valuable, as the quantity $\langle \psi_f| \mathbf{e} \cdot \mathbf{r} |\psi_i\rangle$ can be evaluated for selected symmetries of the wave functions and yields dipole selection rules, which are very useful for a qualitative interpretation of photoexcitation spectra. This can be most easily seen for atomic levels, the wavefunctions of which can be expressed in terms of a radial part and a part containing spherical harmonics $Y_{l,m}$. As the operator $\mathbf{e} \cdot \mathbf{r}$ can also be represented in terms of spherical harmonics (e.g. $Y_{1,0}$ for linearly polarized light, or $Y_{1,\pm m}$ for circularly polarized light), the matrix element $\langle \psi_f| \mathbf{e} \cdot \mathbf{r} |\psi_i\rangle$ can be fully calculated by evaluating products of spherical harmonics. The particular mathematics of spherical harmonics allows the matrix element to be nonzero only for particular relations between l_f, l_i, m_f, m_i, which is the basis of the selection rules. Although being only strictly valid for atomic systems, this approach has also been successfully extended to approximately describe the behavior of electronic states at high symmetry points in solids. A more general treatment of dipole selection rules in solids has been developed on the basis of group theory [1].

We have already mentioned above that the central aspect in the photoexcitation step concerns dipole selection rules. These rules predict allowed electronic transitions based on the symmetry

---

[1] This assumption should be revisited, if we go to photoexcitation with hard x-rays.

of the electronic wave functions involved. In an atomic picture, these selection rules take the form:

$$\Delta L = \pm 1 \tag{6}$$

$$\Delta m_L = 0, \pm 1 \tag{7}$$

The photon carries an amount of angular momentum of $|L| = 1$ with it's polarization state being determined by $m_J = 0$ (linear polarization) and $m_J = \pm 1$ (right- and left-hand circular polarization, respectively). Linearly polarized light can be represented as a superposition of right- and left-hand circularly polarized waves. To illustrate the action of these selection rules we take the example of the excitation of an atomic $2p$ level. According to Eq. (10) we will find two types of allowed transitions, which may contribute to the photoemission spectrum

$$p \to \begin{cases} d \text{ for } \Delta L = +1 \\ s \text{ for } \Delta L = -1 \end{cases} \tag{8}$$

For the evaluation of Eq. (7) it is useful to consider that atomic states are usually subject to spin-orbit coupling, which leads to a characteristic splitting of the atomic levels and leaves only the total angular momentum $J = L + S$ as a good quantum number. Consequently, our p-level splits into a $p_{3/2}$ and a $p_{1/2}$ state and with linearly polarized light, we will have allowed transitions of the type

$$\begin{aligned} p_{3/2} &\to d_{3/2} \\ p_{1/2} &\to s_{1/2} \\ p_{-1/2} &\to s_{-1/2} \\ p_{-3/2} &\to d_{-3/2} \end{aligned} \tag{9}$$

whereas circularly polarized light gives us

$$\begin{aligned} p_{3/2} &\to d_{5/2} \\ p_{1/2} &\to d_{3/2} \\ p_{-1/2} &\to s_{1/2} \\ p_{-3/2} &\to s_{-1/2} \end{aligned} \text{ for } (\Delta m_L = +1) \quad \text{and} \quad \begin{aligned} p_{3/2} &\to s_{1/2} \\ p_{1/2} &\to s_{-1/2} \\ p_{-1/2} &\to d_{-3/2} \\ p_{-3/2} &\to d_{-5/2} \end{aligned} \text{ for } (\Delta m_L = -1) \tag{10}$$

Note that the dipole operator of the light acts only on the orbital part of the electronic wave function, i.e. on the spatial symmetries, but it cannot interact with the electron spin S directly. However, because spin-orbit coupling ties the spin to specific orbitals, a selective excitation can yield spin polarized photoelectrons even from nonmagnetic materials. This phenomenon is called optical spin-orientation [10] and is also the basis of all magnetodichroic effects observed in photoabsorption and photoemission. In order to see how this works let us have a closer look at the $p \to s$ transitions described by Eq. (10) (selection rules for crystalline symmetries see Appendix). The states $s_{1/2} \equiv |\uparrow\rangle$ and $s_{-1/2} \equiv |\downarrow\rangle$ may be regarded as pure spin states.

For positive light circularity we have transitions starting at $p_{-3/2}$ and $p_{-1/2}$. Usually, the probabilities for the two transitions - which can be simply calculated from the Clebsch-Gordon coefficients [11] - differ by a factor of 3, i.e. the amount of photoelectrons excited into $s_{-1/2}$ is 3 times that for the $s_{1/2}$ state. If we assume that we have a nonmagnetic situation, the $s_{-1/2}$ and $s_{1/2}$ states will be energetically degenerate and a summation over the two photocurrent contributions will yield a spin polarization of the photoelectrons of $P = -50\%$. The same treatment for negative light circularity yields $P = 50\%$, i.e. a reversal of the circularity also reverses the sign of the photoelectron spin polarization. This optical spin-orientation effect in the $p \to s$ transitions is particularly exploited in spin-polarized GaAs photocathodes [12], but it can also be observed as a general photoemission phenomenon in basically all materials. It also governs magnetic circular and linear dichroism effects in absorption spectroscopies, which are widely used to study complex magnetic materials.

The quantity measured in the experiment is related to $I(h\nu)$, which is composed by the transitions between all possible initial (i) and final states (f) for a given photon energy $h\nu$

$$I(h\nu) \sim \sum_{i,f} |\langle f| \mathcal{O} |i\rangle|^2 \delta(\epsilon_f - \epsilon_i - h\nu) \tag{11}$$

In a photoabsorption experiment $I(h\nu)$ will be related to the absorption cross section, whereas in a photoemission experiment $I(h\nu)$ will describe the photoelectron current leaving the sample.

3 Photoabsorption Spectroscopy

In photoabsorption spectroscopy we excite the photoelectrons only up to below the vacuum level, i.e. the final states are provided by the empty electronic bands between the Fermi level E_F and the vacuum level E_{vac}. The absorption of the photons is usually determined by measuring the intensity I_t transmitted through a sample of thickness d by a photomultiplier, photodiode or CCD array and comparing it to the incident light intensity I_0 in the framework of Beer's law (Fig. 3a)

$$I_t = I_0 \exp\left(-\mu_x d\right). \tag{12}$$

The absorption coefficient μ_x in the transmission geometry depends on the photon energy via the quantity $I(h\nu)$ in eq. 11. Provided that $h\nu$ is large enough to trigger secondary processes – usually in the soft x-ray regime – there opens up a second detection channel based on emitted electrons. The total amount of photoelectrons emitted from a sample I_e is called *electron yield*. The electron yield implies an integration over all electron kinetic energies and is proportional to the absorption coefficient μ_x. It must be pointed out that due to the limited information depth of the emitted electrons (cf. Chap. 4) the electron yield I_e provides information about the surface region of the sample, whereas the transmitted intensity I_t carries information about the bulk. The electron yield can be measured either by collecting the electrons in front of the sample, for instance, by a Faraday cup or more conveniently by measuring the photocurrent from ground to the sample, which restores charge neutrality (Fig. 3b).

3.1   Absorption spectroscopy with soft x-rays

To illustrate the two types of absorption experiments we briefly discuss x-ray absorption spectroscopy (XAS) at the L-edges of the transition metal cobalt. The photoexcitation takes place from the $2p$ core levels into the density of states between E_F and E_{vac}, just as we have discussed in Chapt. 2. The $2p$ core states of cobalt are split into $2p_{3/2}$ and $2p_{1/2}$ levels with binding energies at $E_B \approx 778$ eV and $E_B \approx 793$ eV, respectively. The transmitted light intensity (Fig. 4) reveals two major depressions at photon energies corresponding to these binding energy values, indicating the electronic transitions from the $2p_{3/2}$ and $2p_{1/2}$ core levels into s- and d-type states in the empty density of states. These strong absorption features are also known as "white lines" [14].

The total electron yield spectrum (Fig. 4b) shows maximum signal intensity at the position of the absorption edges and is basically a mirror image of the spectrum in Fig. 4a. The insets in the figures detail the microscopic electronic processes. Whereas the transmission spectrum is directly related to the absorption of the photon, the total yield spectrum is the result of a more complicated three-step process. The first step – photoexcitation – creates a hole in the core level. This core hole has a finite lifetime and decays by either radiative (x-ray fluorescence) or nonradiative processes (Auger process). In the Auger process the core hole is filled by an electron from a level at lower binding energy, e.g. the $3p$, $3s$ states or even the occupied valence states. The energy is then transferred to another electron from the same or a state at lower binding energy, exciting the Auger electron well above the vacuum level. For example, the binding energy difference between the $2p_{3/2}$ and the $3p$ levels of cobalt is around $\Delta E \sim 720$ eV. If this energy is transferred to another $3p$ electron, the kinetic energy of the resulting Auger electron will be around $E_{kin} \sim 660$ eV. Inside the solid this highly energetic electron undergoes many inelastic electron-electron scattering processes, thereby efficiently loosing energy, which is subsequently transferred to a whole cascade of secondary electrons. For an electron to leave the crystal, however, its kinetic energy still has to be above the vacuum level. As a consequence of the strong electron-electron scattering, this condition is only met by Auger electrons created in the surface-near region within a depth described by the mean free path λ_e of the electrons. Typically for XAS in the soft x-ray regime up to $h\nu \lesssim 1$ keV the total yield sampling depth in metals is of the order of $\lambda_e \sim 20\text{Å}$ [15].
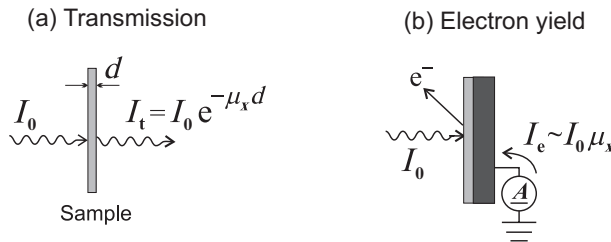
Fig. 3: Principle geometries for photoabsorption experiments. (a) Transmission geometry. Measured quantity is the transmitted light I_t. (b) Electron yield approach. Measured quantity is the total yield of emitted electrons I_e, alternatively the neutralization current from ground to sample. From [13].

(a) Transmission
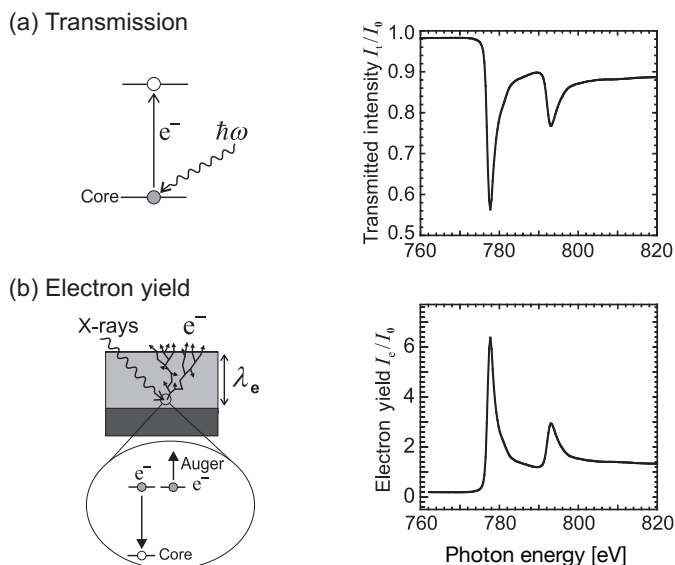
(b) Electron yield

Fig. 4: X-ray absorption spectra at the Co $L_{2,3}$ edges. (a) Transmitted intensity. (b) Total electron yield. From [13].

The details of the absorption spectrum are determined by the electronic configuration of both the core level and the empty density of states. Both are affected by the chemical environment, which causes modifications of the valence density of states through chemical bonds and chemical shifts of the core states binding energies. This is illustrated in Fig. 5, which compares the x-ray absorption spectra of the elemental metals Fe, Co and Ni with the related oxides $LaFeO_3$, CoO

Fig. 5: X-ray absorption spectra at the Fe, Co, and Ni $L_{2,3}$ edges for the pure metals and selected oxides $LaFeO_3$, CoO and NiO. From [16].

and NiO. In the clean ferromagnets, we observe two smooth absorption lines in each spectrum, which are characteristic for the metallic state. Upon oxidation, these absorption lines split into multiplet structures, which are also fingerprints of the respective oxide state. The spectral details of the multiplet splitting are related to the geometric arrangement of the atoms in the unit cell and the charge transfer from the metal due to the chemical bond to oxygen.

This example at the transition metal L-edges demonstrates that XAS can be used to obtain detailed information about the chemical state and the electronic configuration of the individual constituents by addressing the relevant absorption edges. Similar studies can be performed with hard x-rays (up to $h\nu = 10$ keV) at the K absorption edges. The hard x-rays have a much higher penetration depth and are able to provide true bulk information.

3.2 Polarization effects in XAS

So far, we have neglected the polarization character of the incident light X-ray absorption spectroscopy. As we can infer from the dipole selection rules, however, the absorption spectra should depend strongly on the polarization state of the exciting radiation, which can be conveniently selected at modern synchrotron radiation facilities. This polarization dependence is a powerful tool and is widely exploited in a broad range of X-ray absorption spectroscopies, which are often called as near-edge x-ray absorption fine structure (NEXAFS) spectroscopies [18].

The influence of the polarization character can be nicely demonstrated for oriented molecular orbitals. Figure 6 shows NEXAFS spectra from benzene molecules adsorbed on a Ag(110) surface. The molecules are lying flat on the surface. The directional bonding between carbon atoms in the benzene ring results in σ- and π-orbitals, which are oriented in the ring plane and perpendicular to it, respectively. In an x-ray absorption process at the carbon K-edge the empty orbitals $\sigma^\star$ and $\pi^\star$ are the final states in the photoexcitation from the C $1s$ level. Using linearly polarized light, the dipole matrix elements becomes large, if the electric field vector $\mathbf{E}$
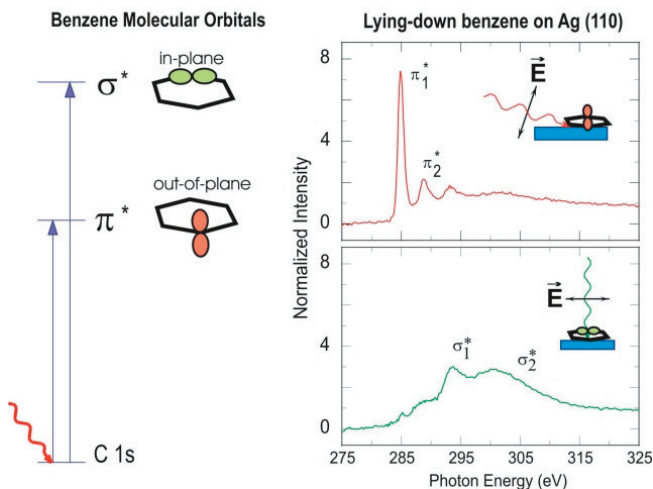
Fig. 6: Orientational dependence of x-ray absorption spectra from molecular orbitals in the system C_6H_6/Ag(110). The flat-lying molecules provide σ- and π-orbitals, which are oriented in the ring plane and perpendicular to it and can be selectively probed by aligning the electric field vector $\mathbf{E}$ of the linearly polarized light along the orbital's symmetry axis. From [17].

is oriented parallel to the orbital's symmetry axis. This means that the excitation into the $\pi^\star$ orbital is strongest for $\mathbf{E}$ perpendicular to the surface (grazing incidence), whereas $\mathbf{E}$ parallel to the surface probes the $\sigma^\star$ orbitals (normal incidence). The resulting absorption spectra differ strongly in the region of the $\pi^\star_{1,2}$ absorption lines. Note that in the grazing incidence geometry there is always also a small component of $\mathbf{E}$ along the surface. Therefore, there are also small spectral contributions from excitation into the $\sigma^\star_{1,2}$ orbitals. They can be easily identified by comparison to the normal incidence spectrum.

The orientational dependence of dipole matrix elements with respect to the electric field vector of the light is sometimes also called "search light" effect. It is not only exploited for the study of molecular structures as illustrated above, but also in solid state systems to investigate charge, orbital and spin ordering effects through natural or magnetic dichroism phenomena.

3.3   Spin effects in XAS

In Chapt. 2 we had already discussed the role of the spin and spin-orbit coupling on the transition matrix elements, leading to the phenomenon of optical spin-orientation by means of circularly polarized light. In x-ray absorption spectroscopy this phenomenon has enabled very powerful techniques to investigate magnetic systems, namely the x-ray magnetic circular dichroism (XMCD) and the x-ray magnetic linear dichroism (XMLD).

In the following we will limit ourselves to a short description of the XMCD, more details can be found in recent reviews [19, 20]. This effect was experimentally observed first at the K-edge of Fe, i.e. using circularly polarized hard x-rays [21]. The circularly polarized light is characterized by a helicity vector ζ, which is oriented parallel or antiparallel to the propagation direction of the light $\mathbf{q}$ in the case of right-handed or left-handed circular polarization, respectively. The magnitude of the circular magnetic dichroism I_{XMCD} depends on the alignment of ζ with respect to the spin quantization axis in the ferromagnet, which is given by the magnetization vector $\mathbf{M}$

$$I_{XMCD} \sim \zeta \cdot \mathbf{M}. \tag{13}$$

This means that the maximum dichroic signal is obtained when comparing parallel and anti-parallel alignments of $\mathbf{M}$ and ζ, i.e.

$$\Delta I_{XMCD} = I(\zeta \uparrow\downarrow \mathbf{M}) - I(\zeta \downarrow\uparrow \mathbf{M}). \tag{14}$$

The quantity $\Delta I_{XMCD}(h\nu)$ can be conveniently measured by reversing the magnetization for a given light helicity or by reversing the light helicity for a given magnetization direction. In addition, sometimes instead of the intensity difference $\Delta I(h\nu)$ the intensity asymmetry $A_{XMCD}(h\nu)$ is given

$$A_{XMCD} = \frac{I(\zeta \uparrow\downarrow \mathbf{M}) - I(\zeta \downarrow\uparrow \mathbf{M})}{I(\zeta \uparrow\downarrow \mathbf{M}) + I(\zeta \downarrow\uparrow \mathbf{M})}. \tag{15}$$

We recall from the discussion of the optical spin orientation in Chapt. 2 that the crucial quantity is the spin-orbit splitting of the electronic states involved in the optical excitation. In the case of
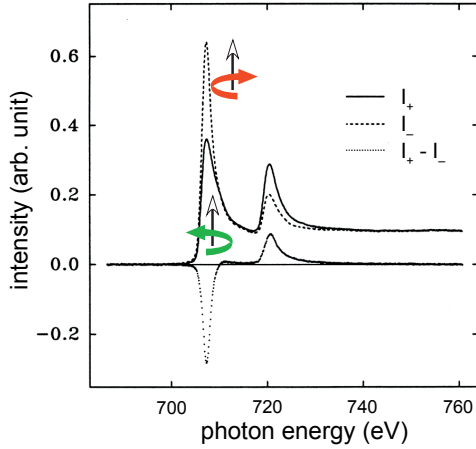
Fig. 7: X-ray magnetic circular dichroism at the Fe $L_{2,3}$-edges. The spectra I_+ and I_- represent the parallel and antiparallel configurations of light helicity and magnetization. The magnetic dichroism is represented by the difference $\Delta I = I_+ - I_-$, which exhibits a characteristic negative excursion at the L_3-edge and a smaller positive excursion at the L_2-edge.

the K-edge magnetic dichroism mentioned above the core level has angular momentum $L = 0$ and the only spin-orbit split states are contributed by the valence electronic states. The spin-orbit splitting of the valence states is quite small, of the order of a few 10 meV. As a consequence the observed Fe K-edge XMCD signal is only of the order $A_{XMCD} \sim 10^{-4}$ [21].

The situation is quite different for the L-edge XMCD. The spin-orbit splitting of the $2p$ core states is of the order of 10 eV. The resulting magnetic dichroism reaches values of $A_{XMCD} \sim 10^{-1}$ as is demonstrated for the case of Fe (Fig. 7). The spectra for the parallel and antiparallel orientation of M and ζ, I_+ and I_-, exhibit strong variations. The difference between the spectra $\Delta I = I_+ - I_-$ reveals a characteristic signature, namely a negative excursion at the L_3-edge and a smaller positive excursion at the L_2-edge. The first experiment of this type was performed at the Ni $L_{2,3}$ edge [22].

The power of XMCD comes from the fact that a careful analysis of the intensity and magnetic dichroism spectra in the framework of the "sum rule" formalism yields quantitative information about the magnetic moment, which can even be decomposed into the orbital and spin parts μ_L and μ_S [19,20]. This sum rule approach works particularly well for the late transition metals Fe, Co and Ni, whereas for the early transition metals more elaborate photoexcitation calculations are needed to describe and analyze the magnetic dichroism spectra.

An example for a more complex magnetic material is Magnetite, Fe_3O_4. Magnetite comprises several magnetic sublattices, the magnetizations of which are partially compensating each other, i.e. magnetite is a ferrimagnet. The absorption spectrum at the Fe $L_{2,3}$ edges is rather complex and exhibits the typical multiplet structures being related to Fe ions in different oxidation state and bonding configuration (Fig. 8a, top). This becomes particularly clear in the XMCD spectrum which exhibits a wealth of features, particularly the "W"-type variation of the dichroism at the Fe L_3 edge (Fig. 8a, bottom). The individual features are related to the three different iron sites occuring in magnetite. The Fe^{2+} cation is only found in an octahedral configuration (B site), whereas the Fe^{3+} ion takes both octahedral and tetrahedral (A site) configurations. For each cation and configuration calculations yield a specific contribution to the XMCD spectrum. For example, at the Fe L_3 edge the cation in tetrahedral configuration Fe_{tetr}^{3+} causes a strong positive dichroism, whereas the two B-site cations Fe_{oct}^{2+} and Fe_{oct}^{3+} each

cause a negative dichroism (Fig. 8b). These signs reflect the fact that the B-site cations couple ferromagnetically, whereas the coupling between A- and B-site cations is antiferromagnetic and hence yields an opposite sign in the dichroism.

This example demonstrates that x-ray magnetic dichroism is able to even separate the magnetic contributions of different ionic species in a complex magnetic material. As we have seen the XMCD is sensitive to the *spin orientation* in the ferromagnet. We should mention that there is also the phenomenon of x-ray magnetic linear dichroism (XMLD), which is sensitive to the *spin alignment* in a magnetic material. This is particularly relevant in antiferromagnetic materials, which do not exhibit a net magnetic moment. They can be conveniently spectroscopically investigated by XMLD [24]. The combination of XMCD and XMLD is thus a very powerful tool to study systems composed of constituents with different magnetic orderings.
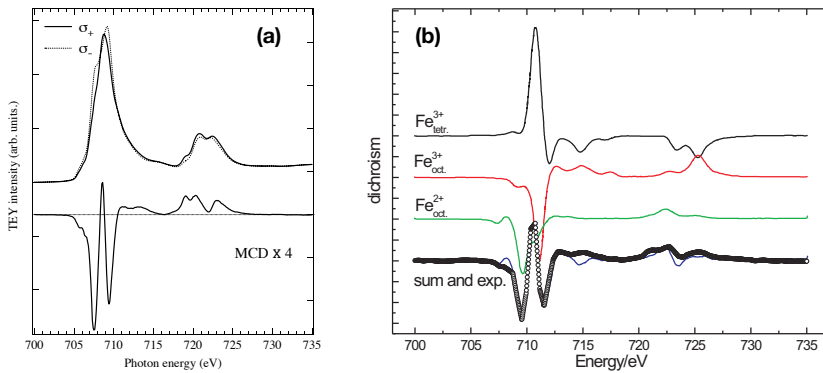
Fig. 8: X-ray absorption spectra at the Fe $L_{2,3}$ edges for the oxide Fe_3O_4. (a) Experimental data. (b) Calculations of the magnetic dichroism resulting from the various Fe ions in tetra- and octahedral configurations. From [23].

4   Photoelectron Emission Spectroscopy

We will now turn to the case when the photoelectron is excited above the vacuum level E_{vac}. Once the photoelectron has been excited into the upper state it will propagate through the solid with a kinetic energy of $E_{kin} = h\nu - E_B$ according to energy conservation. The propagation direction is determined by the electron momentum $\hbar\mathbf{k}_i$, i.e. the electron wave vector $\mathbf{k}_i$ inside the crystal. Under certain conditions, which we will discuss below, the photoelectron may leave the sample through the surface and can be detected by a photoelectron spectrometer with respect to its kinetic energy E_{kin} and wavevector $\mathbf{k}_o$ outside the crystal (Fig. 9). With additional effort also the electron spin polarization $\mathbf{P}$ may be determined. The energy analysis requires an energy dispersive element – in the example a hemispherical capacitor, and the selection of the wave vector wavevector $\mathbf{k}_o$ is achieved by a angle-resolving electron optics. Access to the photoelectron spin polarization is provided by so-called spin polarimeters. This approach is generally termed angle-resolved photoelectron spectroscopy, ARPES.

The principle scheme of the photoemission process is sketched in Fig. 10. Depending on the photon energy the spectrum contains only a signature of the valence bands or signatures from the
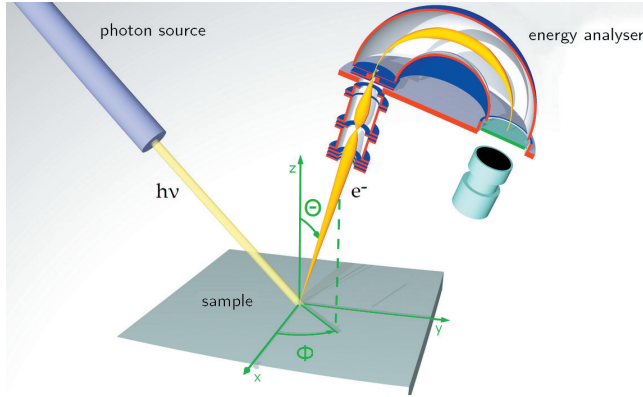
Fig. 9: Schematical picture of an angle-resolved photoemission experiment. The electron optics in front of the hemispherical energy analyser selects the photoelectrons according to the angles θ, ϕ with respect to the surface normal.

valence bands and core levels. The valence band part of the spectrum may be highly structured, depending on the details of the band structure and the dipole selection rules. The photoemission from core level states produces sharper spectral lines, which have a certain spectral width, though. Note that this is a discrepancy to our initial discussion of Fermi's golden rule, in which the delta function predicts infinitely sharp lines. The reason for this is two-fold. First, a certain spectral broadening results from the limited energy resolution of the spectrometer. Second, the single particle picture of the photoemission process is a simplification. The realistic line shape of a photoemission line is determined by the many-electron nature of the photoemission process, which will be discussed in sect. 4.2.2.

4.1   Basic photoemission considerations

Note that in this single particle picture the photoemission spectrum is represented by a series of sharp lines, which is not what is observed in the experiment. The reason of this discrepancy is the many-electron nature of the photoemission process, which will be discussed in sect. 4.2.2.

We recall that the photoexcitation is only the first part of the entire photoemission process. Once the electron has been excited into the upper level above E_{vac} it will move through the crystal. With a certain probability this highly energetic or "hot" electron will arrive at the surface. In order to pass through the surface the electron wave function has to couple to a free-electron like state in the vacuum outside the crystal. The proper quantum mechanical treatment of the entire photoemission process is the so-called *one-step model*, which – at least in principle – permits a quantitative interpretation of photoemission spectra. However, a more intuitive access to the underlying physics is provided by the simpler *three-step model* which we will briefly sketch in the following.
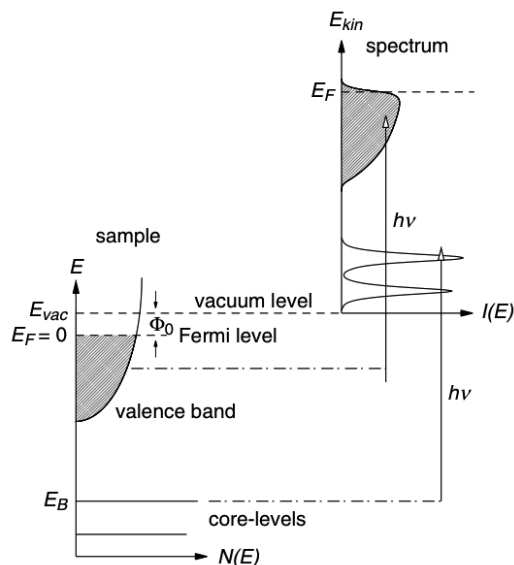
Fig. 10: Principle of the photoemission process. The electrons excited into states above the vacuum level form a photoelectron spectrum reflecting a broad valence electronic distribution (shaded area) and sharper emission lines from the core levels. From [3].

4.1.1 Three-Step Model of Photoemission

This model separates the photoemission of a single[2] electron into subsequent processes dealing with (i) the *photoexcitation*, (ii) the *transport* of the hot electron to the surface, and (iii) the *transmission* of the electron through the surface into the vacuum (Fig. 11). On a quantum mechanical level these steps have to be connected in a suitable way in order to allow the electronic wave function to propagate from one step to the next.

Photoexcitation   The photoexcitation step and the role of the dipole selection rules was already discussed in Chapt. 2. However, we have limited ourselves to atomic-like states, whereas ARPES is mostly used to investigate the valence electronic states in a crystal which have Bloch-like character. The wavefunctions for both initial and final states in ARPES are therefore determined by the crystal symmetries and take different symmetry notations, which are usually related to high-symmetry points and directions in the Brillouin zone. As a consequence the explicit form of the dipole selection rules also changes (see Appendix).

Propagation   The propagation of the hot electron is captured in the second step of the three-step model. It takes into account that due to the strong Coulomb interaction in a solid the hot electron will suffer very efficient elastic and inelastic scattering processes, which affect both the energy and angular distribution of the photoemission spectrum observed outside the crystal. The main mechanisms are scattering due to electron-electron interactions and at defects. In

---

[2] This single-electron picture is convenient, because it allows in many cases a qualitative interpretation of photoemission spectra on the grounds of electronic structure calculations within the framework of density functional theory. It neglects, however, electronic correlations in the electronic structure which can be significant in certain materials or materials classes.
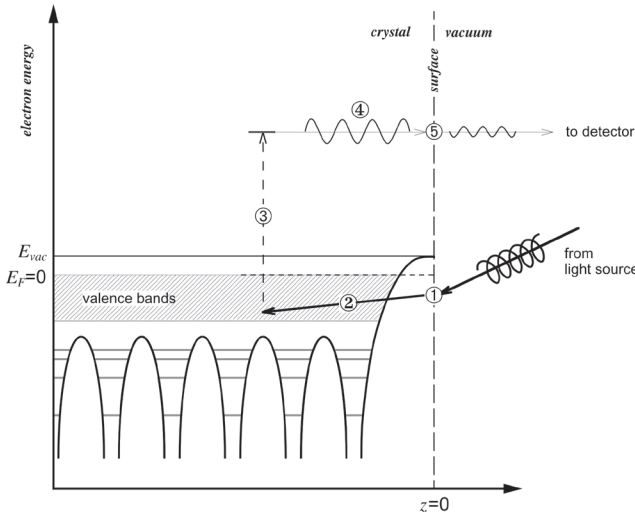
Fig. 11: Schematial representation of the three step model. The numbers denote: (1) refraction of the electromagnetic wave at the surface, (2) penetration of the photon into the solid, (3) photoexcitation, (4) propagation of the photoelectron to the surface, and (5) diffraction of the electron wave at the surface.

particular, the inelastic scattering processes lead to a relaxation of the photoelectron towards the Fermi level. The effect of the inelastic scattering is usually discussed in the framework of an exponential damping of the photoelectron intensity along the path l

$$I(l) = I_0 \exp(-\frac{l}{\lambda_{in}}) \tag{16}$$

and is generally taken into account by the quantity λ_{in}, the *inelastic mean free path* (IMFP). The concept of the IMFP is essential to describe the finite information depth in a photoemission experiment. λ_{in} describes the average distance between two subsequent inelastic scattering events. This quantity is generally believed to follow a very similar behavior in different materials, which leads us to the well-known *universal curve* of the energy dependence of λ_{in}. A closer look, however, reveals that the curve is universal with respect to the general shape only. In quantitative respects the curve depends on the electronic structure of the element or material in question (Fig. 12).

Common to the IMFP curves is a minimum of λ_{in} of only a few Ångstroms at kinetic energies of ˜100 eV and an increase towards both lower and higher energies. On this basis we can understand the high surface sensitivity of photoelectron spectroscopy at intermediate energies, which is both a virtue and a limitation. It can be – at least partly – overcome by exciting the photoelectrons with hard x-ray photons, i.e. photon energies of 6 - 10 keV (see Chapter 4.4.4). From Fig. 12 we see that at a kinetic energy of 10 keV the IMFP increases up to 10 nm. This energy-dependent variation of the information depth forms the basis for hard x-ray photoemission.

Transmission   In the final step of the three-step model the photoelectron has arrived at the surface and will leave the crystal. For this purpose, however, it has to pass through the surface potential barrier which matches the periodic potential inside the crystal to the vacuum outside. From simple quantum mechanics we know that an electron wave passing across a potential step
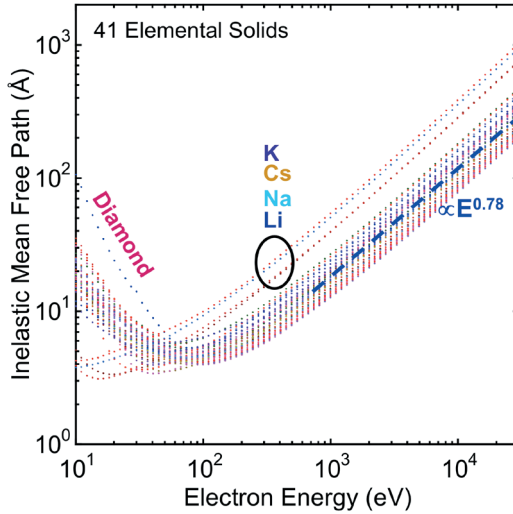
Fig. 12: IMFP values for 41 elements, calculated using the TPP-2M formula: Li, Be, three forms of carbon (graphite, diamond, glassy C), Na, Mg, Al, Si, K, Sc, Ti, V, Cr, Fe, Co, Ni, Cu, Ge, Y, Nb, Mo, Ru, Rh, Pd, Ag, In, Sn, Cs, Gd, Tb, Dy, Hf, Ta, W, Re, Os, Ir, Pt, Au, and Bi. Five "outlier" elements (diamond and the alkali metals) are included to illustrate the influence of the electronic structure characteristics. The dashed straight line for higher energies represents a variation as $\lambda_{in} \sim E_{kin}^{0.78}$, and is a reasonable first approximation to the variation for all of the elements shown. From [25].

Φ will be elastically scattered and diffracted, i.e. it will change its trajectory. In reality the surface potential barrier is not a step function, but smoothly varies as $\Phi(z)$. The details of the scattering process depend on the shape of $\Phi(z)$. The diffraction of the photoelectron in the surface potential is the reason that the wavevectors of the electrons inside the crystal $\mathbf{k}_{in}$ and outside in the vacuum $\mathbf{k}_{out}$ are not conserved. A conservation law exists only for the component parallel to the surface plane

$$k_{\parallel,in} = k_{\parallel,out} + G_{\parallel} \tag{17}$$

with $G_{\parallel}$ being a two-dimensional reciprocal lattice vector in the surface Brillouin zone. In order to relate the perpendicular component of the photoelectron wave vector $k_{\perp,out}$ to the electronic states inside the solid, more sophisticated methods for the band mapping have to be used [1]. The simplest one assumes the final electronic states to be described by a nearly-free electron parabola $E_f = \hbar^2 k_f^2 / 2m^*$ with an effective electron mass m^*. As the initial and final states are connected by vertical transitions in the band structure, k_f can be basically determined from E_f. This procedure fails, however, if the final state band structure deviates considerably from the nearly-free electron picture, which is the case particularly at hybridization regions. In this case, it is more reasonable to use final states from a band structure calculation for comparison. However, all these analyses will only yield a qualitative interpretation of the photoemission spectra. For a quantitative interpretation, a full photoemission calculation within a one-step model is needed.

4.2   Refinement of the Photoemission Picture

4.2.1   One-Step Model of Photoemission

A considerable drawback of the three-step model is its limitation to a qualitative description of the photoelectron spectra. For a more quantitative description, first of all, the transition probabilities for all electronic excitations must be calculated on the basis of a realistic band structure for a semi-infinite system, for example, derived from density functional theory. The formalism must also take into account surface states or transitions into evanescent final states.

Secondly, we must properly consider the multiple scattering events which the hot electron suffers in the final state. These are caused by the strong electron-electron interaction. This situation is more adequately described by the so-called one-step model, which considers the excitation and subsequent transport in a common framework. In particular, the multiple scattering of the electron waves in the surface-near region and in the surface potential is treated in analogy to formalisms developed for low energy electron diffraction (LEED). In LEED an electron wave enters the crystal and subsequently goes through a multiple scattering process. If we invert the order on the time scale, we retrieve our final state in the photoemission process, with the excited electron propagating towards the surface. This state is therefore also called a *time-reversed LEED state* [26].

On the basis of one-step photoemission theories one arrives at a reasonable quantitative description of the photoelectron spectra. This may even include effects due to spin-orbit coupling and the electron spin, in which case a relativistic Dirac-type formalism is involved. In this way it becomes possible to calculate magnetic dichroism and spin polarization spectra.

4.2.2   Beyond the One-Electron Model

Electronic correlations   In the above discussions we have always implicitly assumed that the electronic system under investigation can be modelled within an effective single-electron picture. This has the advantage that the features appearing in the photoemission spectra may be directly related to specific interband transitions in the electronic structure, involving band states or core levels. However, the single-particle picture may fail to capture the essential physics of a particular system, because it may underestimate the correlations in a many-electron system. This is true for the entire family of so-called highly-correlated systems, which includes transition metal oxides and other materials exhibiting phenomena such as high-temperature superconductivity, colossal magnetoresistance or multiferroicity. Such systems are usually described by theoretical approaches beyond simple LDA, for example, LDA+U or dynamical mean field theory (DMFT). The interpretation of photoemission results from such systems is more involved and must take into account the influence of the electron-electron interactions in all steps of the photoemission process.

Spectral shape of photoemission lines   There is, however, a second way through which electronic interactions enter the photoemission experiment. According to Fermi's golden rule (eq. 2) the δ-function taking care of the energy conservation implies all photoemission signatures to be infinitely sharp lines. This should hold particularly for core level photoemission lines.

The experiment reveals that the photoemission lines have a finite width, which is only partially determined by the energy resolution of the spectrometer (Fig. 10). This width is due to the multielectron character of the photoemission process itself.

Whenever a photoelectron is excited to the upper level, it leaves behind a hole in the lower level. Strictly speaking the photoemission process converts an N-electron system into an $(N-1)$-electron system, if the photoelectron has left the crystal, before the hole has been filled again. The photoelectron and the hole interact with each other through Coulomb interaction, which may lead to a renormalization of the binding energies, the appearance of spectral satellites, and a finite linewidth of the spectral line. This has two profound consequences. First, the terms ground and excited state become a different meaning, because they rather refer to an N and $(N-1)$ electronic system, respectively. Second, photoelectron spectroscopy always measures an excited state of matter rather than the electronic ground state. In a somewhat larger picture this is a nice illustration of one of the paradigms in quantum physics, according to which a measurement always affects and alters the system measured. Fortunately, modern condensed matter theory is able nowadays to handle many-electron systems both in the ground and excited state and can therefore provide a full photoemission calculation.

The role of electronic interactions in the photoexcitation spectrum is often taken into account within a Green function formalism [2]. The Green function $G(\mathbf{k}, \epsilon)$ describes the behavior of a quasiparticle, which is "dressed" by the electronic correlations and the electron-hole interaction. Their influence is globally expressed by means of the complex self-energy Σ. Without going through the details of the formalism, from the Green function one can finally calculate the *spectral density function* $A(\mathbf{k}, \epsilon)$, which may be compared to experimental results

$$A(\mathbf{k}, \epsilon) = -\frac{1}{\pi}\mathrm{Im}G(\mathbf{k}, \epsilon) = -\frac{1}{\pi}\frac{\mathrm{Im}\Sigma(\mathbf{k}, \epsilon)}{[\epsilon - \epsilon_k - \mathrm{Re}\Sigma(\mathbf{k}, \epsilon)]^2 + [\mathrm{Im}\Sigma(\mathbf{k}, \epsilon)]^2}. \tag{18}$$

A closer inspection of eq. 18 reveals that the real part of Σ introduces a renormalization of the energy ϵ_k of the spectral feature, whereas the imaginary part of Σ describes the finite lifetime of the quasiparticle state, resulting in a finite spectral width. As we will see below, the self-energy Σ can be conveniently employed to include further interactions, such as electron-phonon and electron-magnon coupling.

The spectral density function replaces the delta function in eq. 11. The total photocurrent is again determined by summing up over all dipole-allowed optical transitions between the many-electron states Φ_f and Φ_i weighted by the spectral density. We then arrive at the following description of the photocurrent [27]

$$\begin{aligned}
I(h\nu) \quad &\sim \quad \sum_{f,i} |\langle \Phi_f| \mathcal{O} |\Phi_i\rangle|^2 \, A_{ii}(\epsilon_f - h\nu) \tag{19}\\
&= \quad \frac{1}{\pi}\sum_{f,i} |\langle \Phi_f| \mathcal{O} |\Phi_i\rangle|^2 \frac{|\mathrm{Im}\Sigma(\epsilon_i)|}{[\epsilon_f - \epsilon_i - h\nu - \mathrm{Re}\Sigma(\epsilon_i)]^2 + [\mathrm{Im}\Sigma(\epsilon_i)]^2}.
\end{aligned}$$

This is the basis for modern photoemission calculations, which attempt a quantitative interpretation of the experimental data.

4.3   Techniques

4.3.1   Electron Spectrometers

During the last decade, we have seen a considerable improvement in photoelectron spectrometer technology, mainly driven by the continuous quest for improved spectral resolution. By now, commercially available energy analyzers may be able to achieve an energy resolution below 1 meV [28]. The most common type of photoelectron spectrometers nowadays are display-type energy filters which are able to efficiently acquire intensity distributions over a certain range of angles and energies within a single measurement. An example for such a hemispherical display spectrometer is given in Fig. 13. The photoelectrons moving away from the sample surface are accepted by a lens system, which defines the angular spread, i.e. the $k_\parallel$ value transmitted. The hemispherical capacitor employed to disperse the electrons according to their kinetic energy is usually operated at a fixed pass energy E_{pass} in order to keep the energy resolution constant throughout a spectrum. The slit between the lens and the hemispheres separates the angular and energy information. The lens system therefore also has the task to accelerate/decelerate the electrons to the pass energy. After being dispersed in the electrostatic field a part of the electrons leaves the analyzer through a second aperture towards the areal electron detector. This usually comprises a combination of a multichannel plate (MCP) and position-sensitive read-out. The MCP consists of an array of narrow channels each being typical several 10 μm
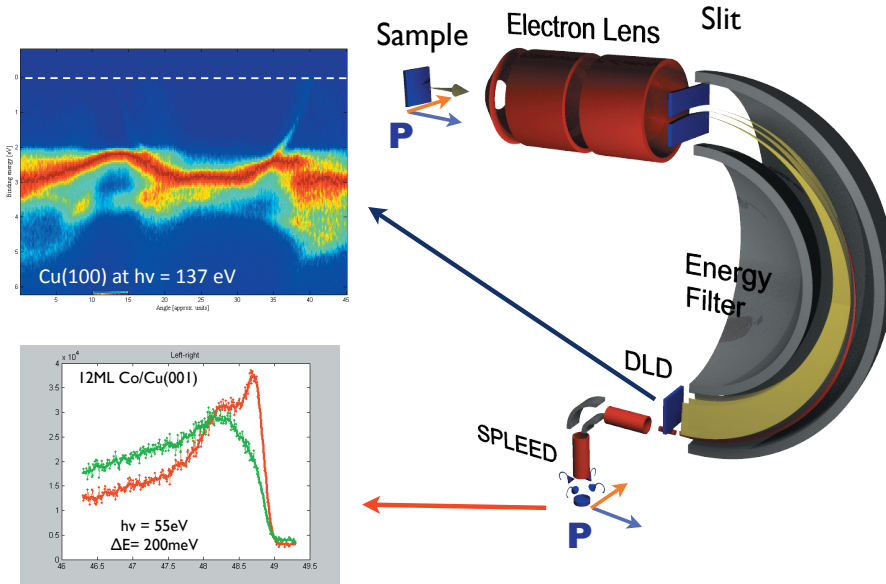
Fig. 13: Hemispherical photoelectron spectrometer with two-dimensional delayline detector (DLD) and a one-dimensional SPLEED spin polarization analyzer operated in the PGI-6 laboratory.
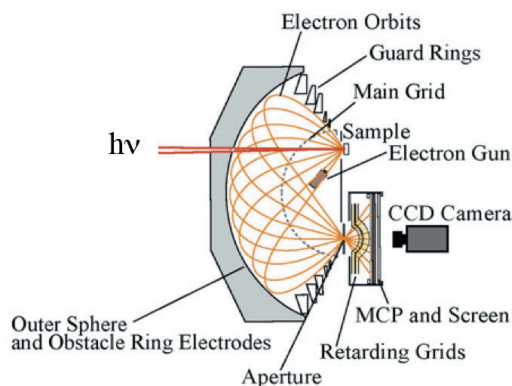
Fig. 14: Cross sectional scheme of the DIANA spherical capacitor spectrometer. The angular distribution of the photoelectrons is imaged onto a two-dimensional detector and captured by a CCD camera. From [30].

in diameter. In each channel a photoelectron is amplified by a factor of $10^4 - 10^6$. This signal is then transported into the position-sensitive readout. The read-out may be a phosphor screen observed by an intelligent camera system which sorts and counts the events into a 3-D array in a computer.

Alternatively, there are resistive anode type detectors, which directly output voltage pulses to a multichannel analyzer. The specific design in Fig. 13 in fact features such a resistive anode arrangement called a delayline detector (DLD) [29] to make room for a second detector measuring the spin polarization of the photoelectrons (see below).

For specific purposes a wide variety of specialized electron spectrometers have been developed over the years. Most of them employ the electrostatic dispersion principle or a time-of-flight approach, in which the kinetic energy of the electrons is converted into a transit time along a defined trajectory. One of the challenges is to increase the angular acceptance of the analyzer in order to be able to capture a larger part of the photoelectron angular distribution in front of the sample. A very interesting design in this respect is the display-type spherical analyzer DIANA (Fig. 14) [30]. The electron trajectories emerging from the sample surface even at large emission angles are guided with high angular fidelity into the detector. The spectrometer is capable of mapping almost the entire half-space in front of the sample. This is particulary useful, for example, for photoelectron diffraction studies.

A very contemporary approach is the *momentum microscopy*, which combines the hemispherical analyzer with an immersion lens optics [31]. An immersion lens applies a strong electrostatic field between the electron optics and the sample to collect the entire angular distribution of photoelectrons in front of the sample and image it through the energy filter onto an imaging detector. This approach can be used with laboratory sources and synchrotron radiation and is to date the most efficient way of analyzing electronic band structures. In the most recent version, momentum microscopy has also been combined with a new generation of spin polarization detectors, permitting unique spin-resolved photoemission studies [32].
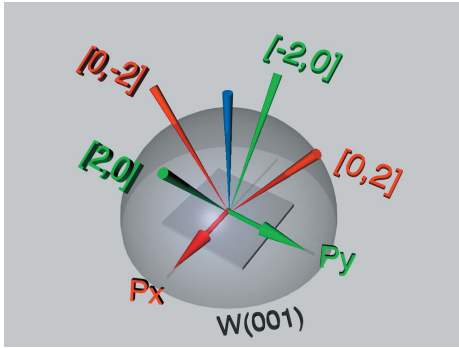
Fig. 15: Sketch of the SPLEED spin detection principle using a W(001) crystal. The spin polarization components P_x (red) and P_y (green) are measured simultaneously.

4.3.2 Spin Analysis

A "complete" photoemission experiment must analyze the photoelectron current with respect to all quantum numbers $|n, \ell, \mathbf{k}, s\rangle$. This includes also the electron spin, which carries important information about spin-dependent excitation and scattering processes in the solid. It can be shown that for an ensemble of electrons, the quantity spin can be expressed by a vector in real space, the spin polarization $\mathbf{P}$, the direction of which is defined by a quantization axis in the solid or the entire experiment [33]. Spin-dependent effects arise either through spin-orbit coupling or exchange interaction, the latter being a characteristic quantity in magnetic systems.

Several types of spin polarization analyzers have been developed over the years. Their common principle of operation is based on the spin-dependent scattering of the photoelectrons off a target. The spin-dependence in the scattering process comes about by the same spin-dependent interactions mentioned above. In a simple picture, these interactions define a spin quantization axis and cause electrons with spin-up and spin-down to scatter with different probability into a direction perpendicular to this quantization axis. A counting detector placed in this direction will thus count different rates of scattered electrons for incident spin-up or spin-down photo-electrons, for example, $I^\uparrow(E)$ and $I^\downarrow(E)$. By subsequently orienting the spin-sensitive axis of the detector along the x, y, and z-axis, we can determine all three components of the spin polarization vector $\mathbf{P}(E)$.

There is only one spin polarization analyzer so far which is based on the exchange interaction. It involves low energy scattering ($E_s = 7$ eV) at a single-domain Fe(001) surface. Detectors exploiting spin-orbit coupling either involve high-energy Mott scattering at the atomic potential (several 10 keV up to 100 keV) or low energy scattering at the periodic potential of a solid (typically 100eV). Since the strength of the spin-orbit coupling increases with the atomic number Z, all spin-orbit scattering targets comprise heavy atoms, such as Au, W, or U.

In order to see how a spin detector is interfaced with the electron spectrometer, we choose the SPLEED detector as an example (Fig. 15). In this detector one effectively performs a spin-polarized low-energy electron diffraction experiment [34]. The incoming electrons hit a W(001) surface at normal incidence with a scattering energy of about 104 eV. The diffracted beams create a four-fold symmetric LEED diffraction pattern above the surface. Of particular importance are the {20} diffraction beams, because they provide the highest spin sensitivity at these scattering conditions. Because of symmetry reasons the SPLEED detector is sensitive to two orthogonal components of the spin polarization vector. The components P_x and P_y are

determined from the intensity of the LEED reflexes by

$$P_x = \frac{1}{S} \cdot \frac{I_{[0,2]} - I_{[0,-2]}}{I_{[0,2]} + I_{[0,-2]}}$$

$$P_y = \frac{1}{S} \cdot \frac{I_{[2,0]} - I_{[-2,0]}}{I_{[2,0]} + I_{[-2,0]}} \tag{20}$$

with the spin sensitivity S. This procedure is repeated for every data point of the spectrum and yields a spin polarization spectrum $P_{x,y}(\mathbf{k}, E)$, which can be used to calculate the spin-up and spin-down contributions of one vector component to the photoemission spectrum according to

$$I^{\uparrow}(E) = \frac{I_0}{2}(1 + P(E)) \ \text{ and } \ I^{\downarrow}(E) = \frac{I_0}{2}(1 - P(E)) \tag{21}$$

with the spin-averaged total intensity I_0.

The most recent generation of spin-polarization detectors employs a two-dimensional detection scheme, which improves the detection efficiency by several orders of magnitude [35].

4.3.3 Spectromicroscopy

The instruments discussed so far are dedicated to very good energy and angular resolution, but they usually do not provide a significant lateral resolution. As a consequence, the studies are restricted to relatively large (mm) and homogeneous samples. However, the progress in nanotechnology leading to ever smaller structures and also the need to study inhomogeneous samples on small length scales is the driving force for a new class of instruments, so-called spectromicroscopes.

Basically, there are two different avenues, which are currently followed in parallel. In the scanning approach one uses a conventional electron spectrometer, but tries to focus down the light beam by means of zone plate optics or capillary optics. Depending on the wavelength of the light and the quality of the optical components, a minimum spot size of less than 50 nm seems to be within reach. This light spot is then raster scanned with respect to the sample, recording a spectrum at each location [36].

In the full-field approach, the sample is illuminated by a more extended beam, and an electron-optical immersion lens microscope is used to image the surface in real time. This image is then passed through an energy analyzer, for example, an electrostatic one similar to the one discussed above. Behind the energy analyzer the now energy-filtered image is magnified onto the image detector, usually a MCP/phosphor screen unit with a slow-scan CCD camera. With these instruments, energy-filtered imaging with a resolution below 100 nm has already been shown. In Fig. 16 we show a different design which is employed in the NanoESCA instrument [37]. The energy filtering involves a system of two hemispherical analyzers connected by a transfer lens. The advantage of this layout is that imaging errors introduced by the first hemisphere are compensated by the second one, thereby improving the overall image quality. The instrument has three different imaging modes. The first one (1) takes the image from the PEEM column
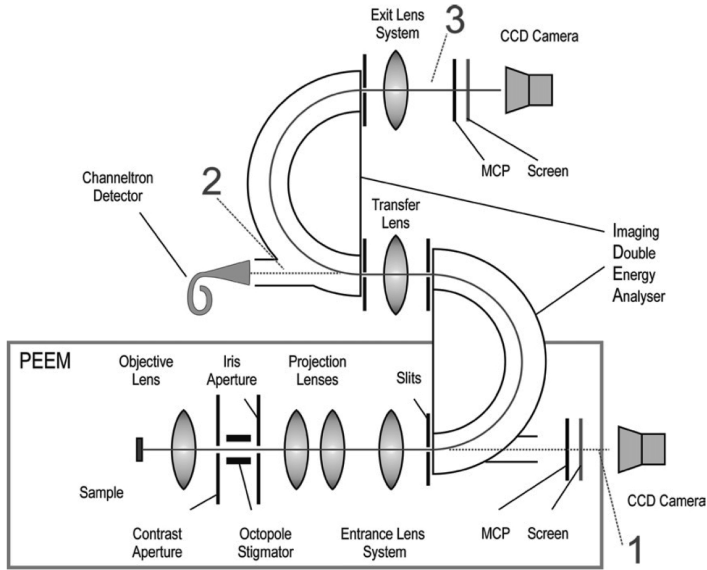
Fig. 16: Principle layout of the NanoESCA imaging spectrometer. From [37].

directly to the image detector. This corresponds to the normal unfiltered PEEM. The second operation mode (2) takes part of the electrons after the first hemisphere into a counting detector. The sample surface area where these electrons originate from is defined by means of an iris aperture and can be in the micrometer range. This mode is very interesting to measure small-spot spectra. Operation mode (3) corresponds to the energy-filtered imaging mode, i.e. the image is formed by electrons within a narrow energy window which can be adjusted, for example, to a particular core level photoemission line.

4.4   Selected Examples

In the following, we will discuss several examples illustrating different applications of photoemission spectroscopy covering band states and core levels.

4.4.1   Electronic and Chemical States

Valence state photoemission   The first example illustrates the mapping of the valence electronic states in a noble metal. A hemispherical display-type spectrometer like the one shown in Fig. 13 records an entire two-dimensional slice of the photoelectron distribution $E(\mathbf{k_f})$ in front of the sample in a single measurement. For a Ag(100) surface, which is illuminated by photons with energy $h\nu = 35$ eV, such a slice is displayed as a colour-coded map in Fig. 17 for photoelectrons emitted around an angle $\theta = 20$ with respect to the surface normal. The energy scale is renormalized to the Fermi energy E_F, and the photoelectron intensity is represented as a function of binding energy E_B and emission angle θ.

For a more detailed analysis of the spectral features one may take cuts through the $I(E_B, \theta)$ distribution, resulting in different types of spectra. A cut at fixed binding energy yields $I(\theta)$, which is sometimes called a *momentum distribution curve* (MDC). A cut at fixed angle yields $I(E_B)$, which is called an *energy distribution curve* (EDC) and corresponds to a "classical" photoemission spectrum.

Although the distribution in Fig. 17 somewhat resembles a band structure, it is important to note that the data are not a simple picture of the bands, because the energy and angular position of the intensity maxima is determined by the transition matrix elements and thus by the initial state and final state bands. Nevertheless, we can already clearly discern different types of spectral features with large and smaller dispersion. In fact, a comparison to bulk band structure calculations of silver along the [100] (Δ) direction reveals that the spectral structure that starts at the Fermi level and bends downwards to the right originates from a strongly dispersing band of symmetry Δ_1, which has a strong free-electron, *sp*-like character. The more localized *4d*-like states in silver give rise to the strong almost horizontal lines at binding energies below 4eV.

The position of the maxima in the MCD's and EDC's can be used to precisely determine the initial electronic state in the band structure by different approaches [1]. Usually this procedure is repeated for different emission directions and surface orientations in order to map the position of the initial state bands throughout the entire Brillouin zone, or at least along high-symmetry directions. Fig. 18 displays such a result for copper, with the data points being obtained from photoemission experiments and the lines representing a band structure calculation. The strongly dispersing bands starting at the Γ-point correspond to the free-electron like *sp*-type states. All other bands exhibit a weaker dispersion and have a strong *d*-type character, meaning that the electrons are more localized.

The band structure in Fig. 18 has been calculated within a relativistic scheme, i.e. it also contains the effects of spin-orbit interaction. Although copper is a material with low atomic number, spin-orbit coupling has been found to play an important role in the band symmetries, in particular, close to hybridization points. We also observe spectral signatures, which do not fit into the calculated bulk band structure $(A - F)$. A further analysis reveals that feature E
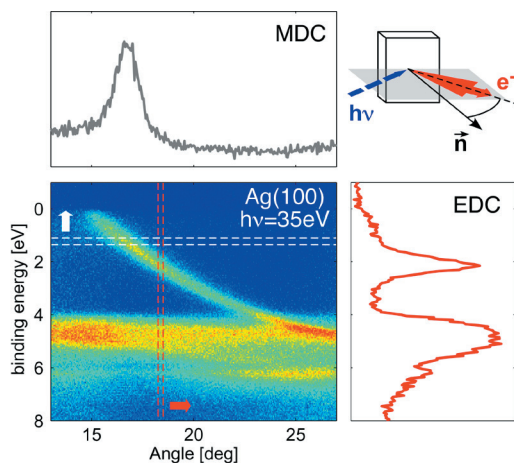
Fig. 17: Example of a two-dimensional $E(\theta)$ distribution recorded from an Ag(100) single crystal surface. The color code ranges from blue (no intensity) through yellow (medium intensity) to red (high intensity). Vertical and horizontal cuts through this distribution yield energy distribution curves (EDC) and momentum distribution curves (MDC), respectively. The broken lines bound areas of 5 lines on the detector which have been added up to the MCD (top) and EDC (right). Inset: Experimental geometry with the red triangle indicating the angular spread measured.
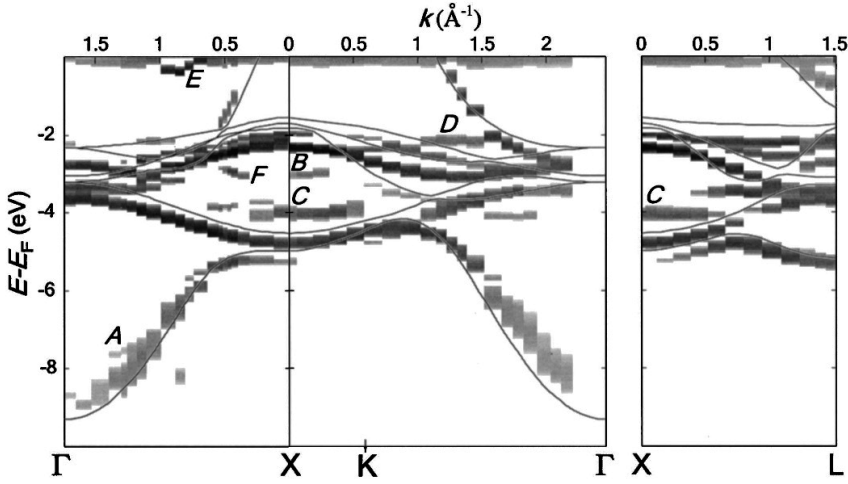
Fig. 18: Band mapping results for the bulk electronic states in a Cu single crystal along the [100] $(\Gamma - \Delta - X)$, ([110] $(\Gamma - \Sigma - K)$, and [111] $(\Gamma - \Lambda - L)$ directions. The bands are shifted from the DFT theoretical $E(\mathbf{k})$, shown by thin lines, due to excited-state self-energy effects. The constant line at E_F is due to the Fermi cutoff, and the peaks A–F are spurious structures due to multiple upper band composition, 1DOS maxima, and surface states. From [39].

is related to a surface state. The features B, C, and D are caused by transitions into regions with a strong one-dimensional density of states, where $k_\perp$ is not conserved. A hybridization of multiple upper bands leads to the appearance of feature A. Feature F is likely caused by surface state or surface resonance split off from the d bands [39]. We also note that the experimental data exhibit a systematic shift with respect to the calculated band. This is due to self-energy effects in the excited state.

Core level photoemission   The photoemission from the localized core levels gives rise to rather sharp spectral features at well-defined and characteristic binding energy values (cf. Fig. 10). These values are tabulated for the elements, for example, in the *X-Ray Data Booklet* [40], and range from several 10 eV for the shallow core levels up to 10 keV for the 1s-levels of heavy elements. Core level photoemission is often used to identify and quantify chemical species and is therefore also termed ESCA (Electron Spectroscopy for Chemical Analysis). For a sample consisting of one chemical element only, the binding energy of the spectral feature is sufficient to unambiguously identify the element[3].

An example is given in Fig. 19, which compiles different spectra recorded for the Al 2p core level. The bottom most spectrum has been obtained from a clean Al(111) surface and shows the 2p core level peak located at a binding energy of $E_B = 73$ eV. A further inspection of the spectrum reveals that the spectral line has a finite width and an asymmetric shape. The line width is mainly determined by the lifetime of the core hole created in the excitation process and by the energy resolution of the electron spectrometer. The asymmetric line shape arises mainly

---

[3] Usually one measures several core level lines at different binding energies in order to increase the accuracy of the element analysis.

due to the excitation of electron-hole pairs in the vicinity of the Fermi level. This corresponds to inelastic electron scattering of the photoelectron in the solid, effectively shifting some spectral weight to the low binding energy side of the peak. This asymmetric line shape may be modeled, for example, with the Doniach-Sunjic approach [41].

The binding energy of a given core level may change, as soon as we alter the chemical environment, for example, by a chemical reaction. Although the chemical bonds formed with an atom as a consequence of the reaction involve mainly the valence electrons, they may cause a charge transfer from or to that atom. This process modifies the electrostatic screening in the atom, ultimately resulting in a slight shift of the core level binding energy. These socalled *chemical* shifts form the basis of more elaborate ESCA approaches in determining the chemical composition of complex alloys and compounds. An illustration for chemically induced core level binding energy shifts is given by the remaining spectra in Fig. 19. These spectra are obtained by exposing the Al(111) surface to different amounts of oxygen in successive steps. After an oxygen exposure of 25 L we start to see a weak spectral feature on the high binding energy side of the 2p level. After 50 L this feature has grown into a well-defined sharp peak $E_B = 74.4$ eV, which can be attributed to photoemission from Al surface atoms onto which oxygen has chemisorbed. In addition, a third peak starts to form at still higher binding energies. After dosing 100 L onto the Al(111) surface, this third signature at $E_B = 75.7$ eV has evolved into a clear peak, which can be attributed to photoemission from Al atoms bonded in an Al$_2$O$_3$ environment. We therefore see that the oxidation from metallic aluminium to alumina
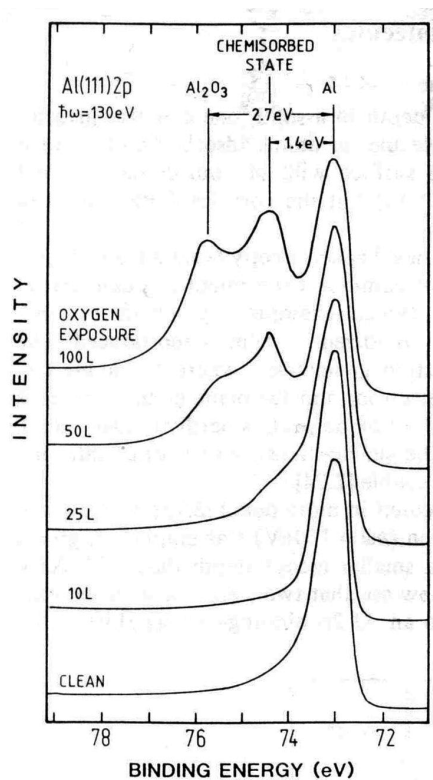
Fig. 19: Core level photoemission from Al(111). A surface reaction with oxygen leads to characteristic chemical shifts of the core level binding energies with respect to the clean surface. The amount of oxygen that the surface is exposed to is measured in units of Langmuir L ($1L = 10^{-6}mbar \cdot s$). From [1].
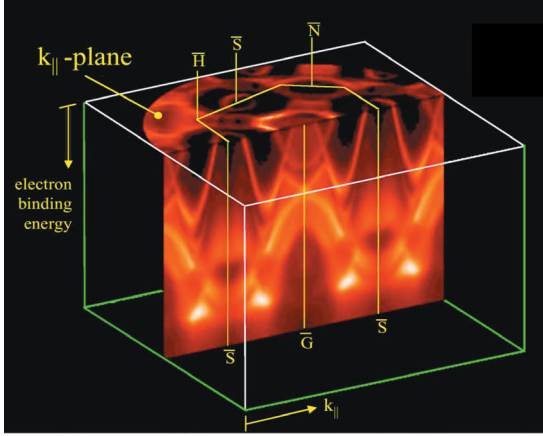
Fig. 20: (Right) Principle of Fermi surface mapping illustrated for photoemission from W(110). The plot compiles photoemission intensity distributions $I(E, \mathbf{k}_\parallel)$ for different emission angles, which can be stacked in a three-dimensional scheme. A cut through the stack at $E = E_F$ yields a two-dimensional map of the Fermi surface in the plane defined by $\mathbf{k}_\parallel$. From [42]. (Left) Fermi surface for Pb-doped Bi2212, i.e. $Bi_2Sr_2CaCu_2O_{8-\delta}$. From [43].

is accompanied by a chemical shift of the Al $2p$ core level by about $\Delta E_B = 2.7$ eV.

Fermi surface mapping   A particular aspect in modern photoemission spectroscopy is the so-called *Fermi surface mapping*. In order to see how this approach works, it is useful to recall that the data provided by the 2D display analyzers represent an intensity distribution $I(E, \mathbf{k}_\parallel)$. The electron wavevector $\mathbf{k}_\parallel$ parallel to the surface is defined by the experimental geometry. By varying the emission angles θ and ϕ (cf. Fig. 9) one obtains a set of slices through reciprocal space for different vectors $\mathbf{k}_\parallel = (k_x, k_y)$ in the surface plane. This data set can be condensed into a three-dimensional representation $E(k_x, k_y)$. An example for angle-resolved photoemission from a W(110) surface is shown in Fig. 20. The picture combines a vertical cut $I(E, k_x, k_y = 0)$ through the surface Brillouin zone (SBZ) with a horizontal cut $I(E = E_F, k_x, k_y)$. The intensity distribution $I(E, k_x, k_y = 0)$ reveals a clear dispersion of band segments along the high symmetry directions $\bar{S}-\bar{\Gamma}-\bar{S}$ in the SBZ, which can be compared to appropriate band structure calculations. The horizontal cut $I(E = E_F, k_x, k_y)$ depicts a two-dimensional map of the electronic states at the Fermi energy and can thus be *related* to the Fermi surface. In the interpretation we have to keep in mind that the map in Fig. 20 contains matrix element and photoelectron diffraction effects. These have to be taken into account when comparing the data to theoretical predictions.

The details of the Fermi surface are crucial in determining the physical properties of materials, for example, the magnetic anisotropy in magnetic systems or the origin of superconductivity. In fact, the onset of superconductivity is accompanied by the formation of a small gap around the Fermi level. It is for this reason that Fermi surface mapping has become a standard tool in the investigation of high-T_C superconductors (HTSC). The example in Fig. 20 depicts the Fermi surface of Pb-doped $Bi_2Sr_2CaCu_2O_{8-\delta}$ (short form Bi2212) [43]. The data have been recorded in the normal state (T=120 K) with a He discharge source ($h\nu = 21.2$ eV). The main
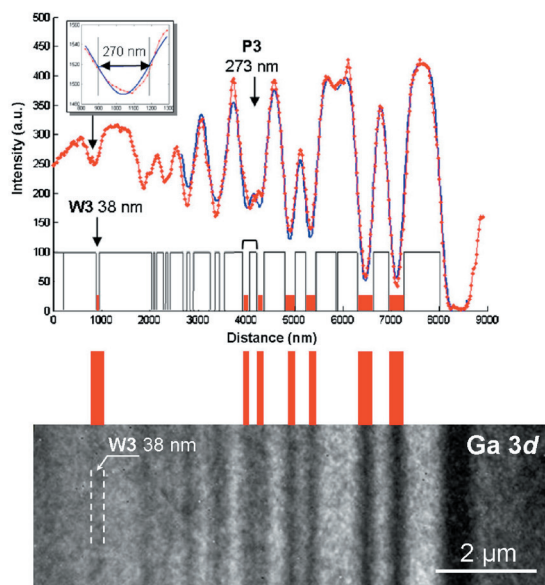
Fig. 21: Energy-filtered image from a GaAs/AlGaAs layered structure acquired at the maximum intensity of the Ga 3d core level peak (kinetic energy of 380.4 eV; incident photon beam of 400 eV). From [38].

Fermi surface is hole-like and has the form of tubes (rings) centered around the X, Y high symmetry points. In addition, weaker intensity features are observed specifically around the M point. This so-called shadow Fermi surface is attributed to a spin-related origin [43]. A detailed understanding of the Fermi surface and their change with temperature are mandatory to understand the microscopic mechanisms leading to HTSC.

Microspectroscopy   Laterally resolving photoemission techniques can yield a two-dimensional distribution of the electronic and chemical states at the surface. The current challenge in microspectroscopy is the improvement of the spectral and lateral resolution in order to increase the image contrast and reduce the minimum feature size detectable. The example in Fig. 21 has been obtained with a NanoESCA instrument (cf. Fig. 16) and represents the current status of the field. The sample is a GaAs/AlGaAs layer stack comprising layers of different thickness, which has been cut into a cross section through the layer stacking (schematic stacking sequence in the center of the image). The image has been acquired with Ga *3d* core level photoelectrons excited with photons of 400 eV energy.

The image contrast arises from the different relative Ga concentration in the GaAs and AlGaAs layers and directly reflects the layer stacking. The AlGaAs regions appear darker than the GaAs areas. The stack included a number of double layer structures, which appear in the imaged cross section as pairs of parallel lines. These can be employed to determine the lateral resolution in the experiment. The two pairs on the right hand side of the image are clearly separated. For the remaining structures on the left-hand side, a line scan analysis across the lines is necessary to extract further information. We can see that the line pair denoted as P3 in the image cannot be separated anymore into two individual features. If one assumes a 20%/80% criterion for the lateral resolution Δx, we find a value of Δx=273 nm.

On the left hand side of the image, we can discern a single line of darker contrast. This signal

(a)

Fe(001)
hν = 34.2eV

Intensity (arb.units)

A
C
B
D

spin polarization (%)

100

50

0

-8   -6   -4   -2 E_F

binding energy (eV)
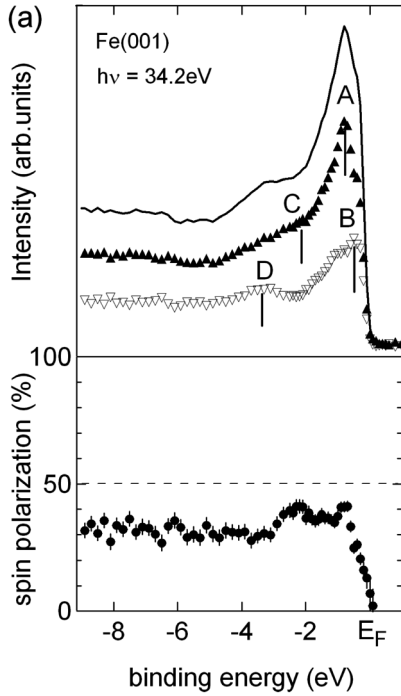
Fig. 22: Top panel: Spin-resolved photoemission spectra from Fe(001) along the surface normal at a photon energy of $h\nu = 34.2$ eV for p-polarized light impinging at 45° angle of incidence. The intensity spectrum (solid line) is decomposed into the spin-up (▲) and spin-down (▽) components. These have been extracted on the basis of the spin polarization distribution (bottom panel). The spectral features marked A-D are explained in the text.

originates from a single AlGaAs layer (W3) of 38 nm thickness. The line scan reveals that this feature causes an image line with 270 nm full-width half maximum (FWHM). This is again the lateral resolution value that we have obtained above. This example demonstrates that features well below the resolution limit can still be made visible provided that the chemical contrast is strong enough.

4.4.2 Spin Effects in Photoemission

The electron spin can give rise to very peculiar phenomena in photoemission experiments. This is due to the fact that the electronic states are subject to two spin-dependent interactions: (*i*) spin-orbit coupling, and *(ii)* exchange interaction. Whereas spin-orbit coupling is mainly an atomic property, exchange-interaction is at the heart of the many electron system and is responsible for magnetic phenomena.

Ferromagnetic systems   A ferromagnet is characterized by a finite magnetization M, i.e. a spontaneous long-range magnetic order below a critical temperature T_C. The magnetization is related to a lifting of the spin-degeneracy of the valence electronic states. As a consequence, the spin-up and spin-down bands are separated in binding energy by the exchange splitting $\Delta E_{exc}(\mathbf{k}, E)$. A spin-resolved photoemission experiment will therefore be able to directly distinguish between the spin-up and spin-down states, as the spin is preserved during the optical transition.

Fig. 22 gives an example for spin-resolved photoemission from the Fe(001) surface. The intensity spectrum (upper panel, solid line) exhibits a strong peak close to the Fermi level and a weaker spectral feature at around 3 eV binding energy. The sample is magnetized within the surface plane and the spin polarization distribution is measured along the magnetization direction (bottom panel). It reveals some structure in the region between E_F and 2 eV binding energy, and levels off to an almost constant value towards higher binding energies. From the intensity and spin polarization distributions we can calculate the spin-up and spin-down contributions to the photoemission spectrum according to eq. 21. These are represented by the closed and open symbols in the upper panel of Fig. 22. We can easily see that these partial spin spectra differ significantly from each other. In a qualitative picture the features marked by A-D can be related to transitions different initial state bands.

According to the nonrelativistic selection rules (see Appendix), the experimental set-up allows access to initial states of mainly Δ_1 and Δ_5 symmetry. By comparison to band structure calculations, the feature B can be ascribed to direct transitions from the $\Delta_5^{\downarrow}$-band lying closely below E_F. The other features appear in the majority spin spectrum: peak A is located at a binding energy of $E_B = -0.7 \pm 0.2$ eV and originates from an initial state of $\Delta_1^{\uparrow}$ symmetry, whereas feature C positioned at $E_B = -2.5 \pm 0.2$ eV results from the $\Delta_5^{\uparrow}$-band. Finally, a feature labeled D is located at $E_B = -3.4 \pm 0.2$ eV and originates from the strongly dispersive $\Delta_1^{\downarrow}$ band, which starts at the Γ-point. These findings are in good agreement with previous photoemission studies on Fe(001) as presented, for example, in Ref. [44].

Optical spin orientation   Even in the absence of magnetic interactions, it is possible to observe spin-polarized photoelectrons and relate them to the symmetry of the electronic states. This phenomenon is called "optical spin orientation" and the microscopic mechanism is provided by spin-orbit coupling, as we have already discussed in Sect. 2.2. The effects are large, if the spin-orbit coupling in the occupied states is strong.

As an example, Fig. 23 shows spin-resolved photoemission data for the W $4f$ shallow core levels obtained with linearly polarized light. The geometry was chosen such that the light impinges on the W(110) surface at a glancing angle of $17°$. Symmetry arguments require that the spin-polarization vector is oriented perpendicular to the plane spanned by the direction of incidence an the surface normal [46]. These states show a clear spin-orbit splitting of about $\Delta E_{so} \approx 2.5$ eV between the $4f_{7/2}$ and $4f_{5/2}$. We see that the partial intensity spectra of spin-up ($\blacktriangle$) and spin-down ($\triangledown$) differ significantly at the peak positions, resulting in a positive spin polarization at the $4f_{7/2}$ emission line, whereas the $4f_{5/2}$ level exhibits a negative spin polarization (bottom panel). This spin polarization reversal between the spin-orbit split levels is an instrinsic feature of the optical spin orientation process, because the total spin polarization integrated over all spin-orbit split levels is required to vanish for symmetry reasons – at least in nonmagnetic materials. Note that for a given experimental geometry the sign of the spin polarization is unambiguously connected to the symmetry of the electron states involved in the optical transition. This assertion also holds for band states in a solid and allows a detailed analysis of spin-orbit effects in the band structure on the basis of spin-polarized photoemission experiments [47].

Rashba States   The spin-polarization effect described in the previous section is due to the intraatomic spin-orbit interaction in the Hamiltonian (eq. 2.1). This term has a specific structure.
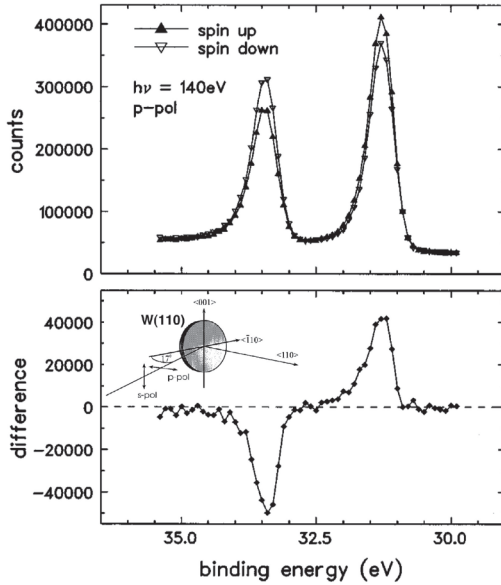
Fig. 23: Spin-resolved W 4f energy distribution curves (EDC's) measured with p-polarized light of hν=140 eV photon energy. The spin polarization vector is oriented perpendicular to the plane spanned by the direction of light incidence and the surface normal. The integral of the difference (lower panel) over the binding energy vanishes within 1% relative to the integral of the absolute value of the difference. From [45].

In the field of spin-dependent transport there is a strong desire to control the electron spin in semiconductors by electric fields. In order to describe this situation in a planar configuration, one often uses the Rasbha-Bychkov Hamiltonian [48]. Interestingly, it has a very similar mathematical form

$$H_{RB} = \alpha(-i\nabla \times \mathbf{E}) \cdot \sigma \qquad (22)$$

with the Rashba constant α and the effective electric field $\mathbf{E}$. One expects maximal effects of the Rashba Hamiltonian when the electric field, the electron momentum and the electron spin are mutually orthogonal.

In two-dimensional (2D) systems with broken inversion symmetry, this spin–orbit interaction causes spin separation of the moving electrons – which is why it is interesting for spintronics. However, the inversion symmetry of the potential is also naturally broken at any crystal surface or interface. As a consequence, electronic states localized at a surface/interface should be spin-split although this splitting can be quite small. In fact, the Rashba interaction at crystal surfaces becomes sizeable only when it couples to the large intra-atomic spin-orbit interaction. The gradient of the surface potential by itself is not sufficient to cause a directly observable splitting of the surface/interface electronic bands into spin subbands [49]. Therefore, this interaction plays an important role only if high-Z elements are involved at the surfaces or interfaces.

It is well-known that some noble metal surfaces exhibit pronounced surface states. This is also true for the unreconstructed Au(111) surface, which exhibits a Shockley-type surface state at the center of the surface Brillouin zone (SBZ), i.e. at the $\bar{\Gamma}$-point. This situation is depicted in Fig. 24. The surface state is characterized by a parabolic dispersion with $k_{\parallel}$. Due to the Rashba
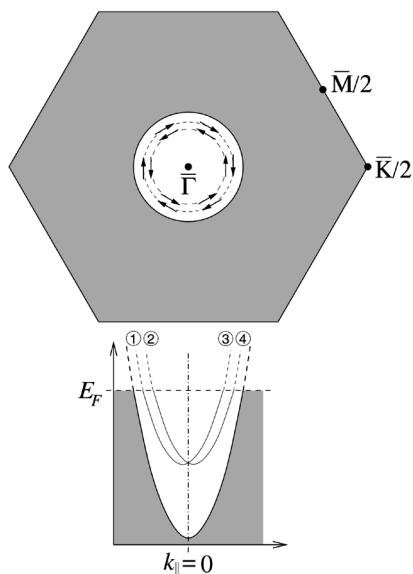
Fig. 24: Upper panel, section of the surface Brillouin zone of the unreconstructed Au(111) surface. The $\overline{\Gamma K}$ distance is $\pi\sqrt{32/3}a = 1.45 \mathring{A}^{-1}$. Lower panel, schematic view of the split surface state dispersion in a cut through $\overline{\Gamma}$. From [50].

interaction the surface state will spin-split, forming two concentric ring-shaped Fermi surfaces with opposite spin polarization in the SBZ.

Indeed this splitting can be clearly seen in a high-resolution photoemission experiment, (Fig. 25). The gray-scale intensity map represents a slice through the Brillouin zone along the direction $\overline{\Gamma K}$. Without the Rashba interaction there would be only one parabolic trace centered around $k_{\parallel} = 0$, corresponding to the dispersion of the surface state. The Rashba interaction introduces a symmetric splitting resulting in two parabolic traces with opposite spin polarization. The energy and momentum distribution curves show that the two traces are only degenerate at $k_{\parallel} = 0$, but separated otherwise.

Magnetic Dichroism in Photoemission What happens, if we have an experimental situation as described in sect. 4.4.2, but our sample is actually ferromagnetic? Let us take the example of a $2p$ core level. The ferromagnetic state is responsible for a spin-dependent energy splitting of the electronic states – not only in the valence states, but also in the core levels. These split according to their magnetic quantum number m_J, i.e. the $2p_{3/2}$ level splits into 4 sublevels ($m_{3/2}$, $m_{1/2}$, $m_{-1/2}$, $m_{-3/2}$), the $2p_{1/2}$ into two. The transition matrix elements depend on m_J *and* the orientation of the magnetization. As a consequence, the fine structure of the intensity spectrum depends on the magnetization direction. This phenomenon is called magnetic dichroism and is observed for both core levels and valence states [51].

This effect is shown in Fig. 26 for the $2p$ core level photoemission from Fe. Note that we have a very similar geometry as in experiment described in sect. 4.4.2. The magnetization vector is oriented perpendicular to the reaction plane. The upper panel compiles the photoemission spectra across the spin-orbit split $2p$ levels. We can clearly see that the spectra differ significantly for opposite magnetization directions. The difference of the two spectra is plotted in the center panel and reveals characteristic bipolar signatures at the position of the core levels. We also note
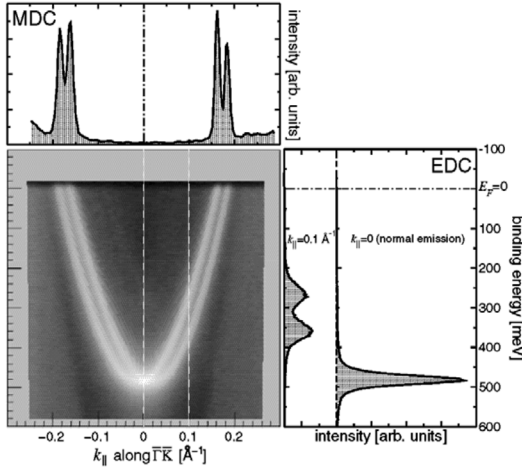
Fig. 25: Photoemission intensity of the Shockley state on Au(111) as a function of energy and momentum $I(E_B, k_{||})$ (white means high intensity). The top panel shows a cut at constant energy $E = E_F$ (MDC); the right-hand panel gives the energy distribution curves (EDCs) at $k_{||} = 0$ and $k_{||} = 0.1$ Å^{-1}. From [50].
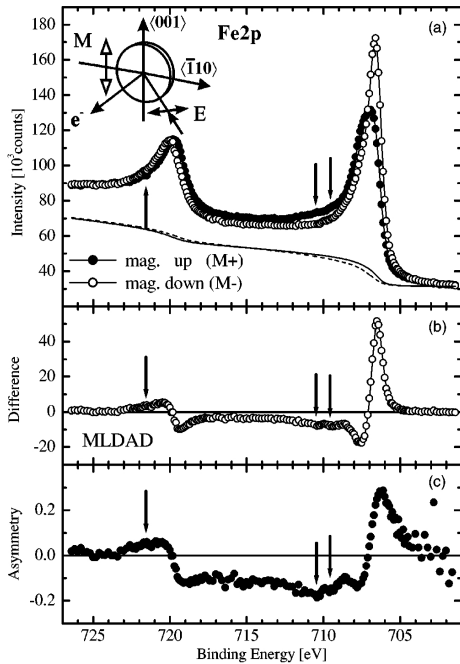
Fig. 26: (a) Fe 2p photoemission spectra and Shirley background of 15 ML Fe / W(110) excited with p-polarized radiation (hν =850 eV) for magnetization up and down (M+, M−). The inset shows the experimental geometry. (b) The intensity difference (MLDAD) of the curves from (a). (c) MLDAD asymmetry (without background). The arrows mark the position of correlation-induced satellites. From [52].

that the polarity of these features reverses between the $2p_{3/2}$ and the $2p_{1/2}$. This is consistent with the spin polarization change in the optical spin orientation experiment in sect. 4.4.2. In fact, as a general rule, optical spin orientation phenomena in nonmagnetic materials are taking the form of magnetic dichroisms in ferromagnets.

The magnetic dichroism signal is often expressed as an intensity asymmetry A

$$A = \frac{I(M+) - I(M-)}{I(M+) + I(M-)}. \tag{23}$$

which reveals a similar spectral dependence than the difference. Additional weak spectral features are related to correlation effects (see Sect. 4.4.2). As the experiment has been performed with linearly polarized light, the effect is also termed magnetic linear dichroism in the photo-electron angular distribution (MLDAD). The latter points out that the size and sign of the magnetic dichroism depends strongly on the emission angle of the photoelectrons analyzed. A closer theoretical analysis shows, that the MLDAD is actually an interference effect between the two photoemission channels into s and d final states [53].

Electronic Correlations   It is well established nowadays that photoemission spectra of narrow-band materials, such as the elements of the d transition-metal series and their compounds, cannot be entirely explained within a one-electron picture. This is due to the presence of local correlations between electrons in the partially filled d band. Experimental band mapping and its comparison with theoretical results can be a powerful tool to directly investigate correlation effects. It has to be realized, however, that the correlated electron picture is less transparent than the single particle model. The interactions due to the electronic correlations lead to a "dressing" of the single particle, i.e. when the particle moves in the solid it is always screened by these many-particle interactions. This system of particle and interaction cloud may be seen as a new quasiparticle. The respective many-electron calculations result in quasiparticle spectral functions rather than conventional band structures, which is a significant conceptual difference.

From all d transition metals, Ni has the narrowest bands and exhibits the strongest correlation effects. This can be seen in Fig. 27. Panel (a) reproduces a set of experimental angle-resolved photoemission spectra, which have been recorded for different emission angles from normal emission up to 70°. It is obvious that the spectral features disperse with the emission angle. A comparison with a single particle particle calculation in the LDA approximation (panel b), however, predicts a much stronger dispersion of the bands than observed in the the experiment. In particular, strong spectral features should also be expected at binding energies larger than 0.5 eV. This is also not observed in the experiment. Furthermore, the exchange splitting between bands of the same symmetry is calculated about twice as large, as observed in spin-resolved experiments ($\Delta E_{exc} \simeq 300\,\mathrm{meV}$ [55]).

The quasiparticle spectral functions calculated within a multiorbital Hubbard model for the experimental geometries are compiled in panel (c) of Fig. 27. The inclusion of correlation effects strongly modifies the spectra: all the structures are pushed up towards E_F by self-energy corrections reproducing much more closely the experimental results both in terms of energy position and dispersion. The spin dependence of the self-energy, arising from the different efficiencies of the scattering channels involving majority- and minority-spin electrons, strongly affects the spin polarization of the quasiparticle states. For this particular region in k space,
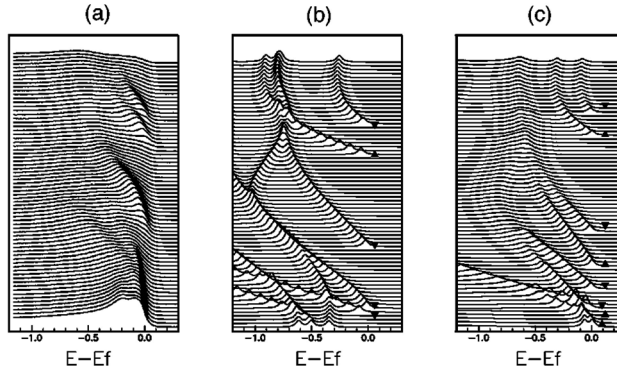
Fig. 27: Comparison between angle-resolved photoemission spectra from a Ni(110) surface at $h\nu = 21.2\,\text{eV}$ (a), single particle local-density approximation (LDA) (b), and quasiparticle calculations results (c). The polar angle ranges from 0° (bottom) to 70° (top). The spin character is indicated by ▲ and ▼. From [54].

four spin-up and four spin-down bands are theoretically predicted in the energy region of interest. While in the single-particle picture one spin-up band and four spin-down bands cross the Fermi energy, all four spin-up bands come close to E_F after the inclusion of correlation effects. Moreover, the energy separation between the spin-up and spin-down bands between $\theta = 50°$ and 60° is reduced by self-energy corrections. All this is in excellent agreement with the experimental data.

In addition to a spin- and energy dependent renormalization of the quasiparticle states due to the self energy, the correlations also lead to the appearance of new spectral features, which are completely absent in the single particle band structures. The most prominent feature in Ni is the famous "6 eV satellite". This is depicted in Fig. 28, which shows the calculated dispersion of the majority spin quasiparticle states. These are compared to spin-integrated, angle-resolved photoemission results. In the region close to E_F we observe the spin-dependent energy renormalization already discussed above. In addition, we find a strong dispersing feature corresponding to the sp-type band. At about 6 eV below the Fermi level, however, there appears a new non-dispersing feature. This is the correlation-induced satellite, which indeed turns out to be of majority-spin character in spin-resolved photoemission experiments [58].

4.4.3 Kinkology

The on-site Coulomb interactions leading to the correlation phenomena discussed above are relatively strong and thus lead to large effects in the band structure. High-resolution photo-emission nowadays provides the opportunity to study also the influence of much weaker inter-actions affecting the electronic system, for example, electron-phonon or electron-magnon inter-actions. As the analysis procedure is connected close to finding and identifying kinks and precisely measuring the spectral width in the dispersion of the quasiparticle states, this field is sometimes coined "kinkology".
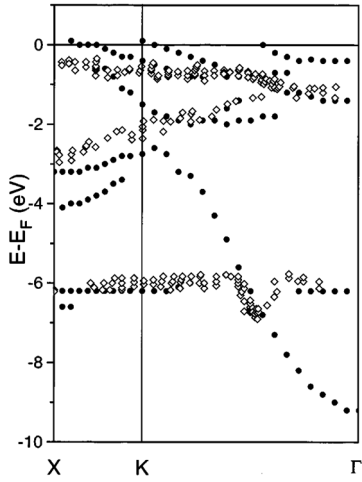
Fig. 28: Comparison between the calculated dispersion of quasiparticle states (●) for majority-spin bands and angle- resolved spin-integrated photoemission results (◊) of Ref. [56]. From [57].

Electron-phonon interaction   The interaction of different quasiparticles, such as electrons and phonons, results in a crossing and hybridization of their respective dispersion relations. At the position in k-space where such crossings occur, the states involved are shifted in energy with respect to the noninteracting case. As phonons have very low energies of the order of 100 meV the respective modifications of the dispersion behavior of the electronic quasiparticle states due to the electron-phonon interaction will be confined to a narrow region below the Fermi level. Formally, the electron-phonon interaction can be considered as an additional contribution to the self energy Σ.

All characteristics of the electron-phonon coupling (EPC) are described by the *Eliashberg function* $E(\omega, \varepsilon, \mathbf{k}) = \alpha^2(\omega, \mathbf{k})F(\omega, \varepsilon, \mathbf{k})$, the total transition probability of a quasi-particle from/to the state $(\varepsilon, \mathbf{k})$ by coupling to phonon modes of frequency ω [59]. Information about the Eliashberg function can be obtained from the angle-resolved photoemission spectra, both through the EPC distortion of the quasi-particle bands near the Fermi energy and the temperature-dependent linewidth. If $\varepsilon_0(\mathbf{k})$ is the bare quasi-particle dispersion of a surface state without EPC, then the measured dispersion $\varepsilon(\mathbf{k})$ with electron-phonen coupling is given by

$$\varepsilon(\mathbf{k}) = \varepsilon_0(\mathbf{k}) + \mathrm{Re}\Sigma(\mathbf{k}, \varepsilon) \tag{24}$$

The screening of the electrons by the lattice is represented by the self-energy function $\Sigma(\mathbf{k}, \varepsilon)$. The imaginary part of the self-energy is related to the EPC contribution to the lifetime τ of the excited electronic states,

$$1/\tau = 2\mathrm{Im}\Sigma(\mathbf{k}, \varepsilon, T) \tag{25}$$

Based on these considerations the influence of the electron-phonon coupling has been investigated in Be($10\bar{1}0$) [60]. The photoemission data in Fig. 29 (left) show the dispersion of the surface states S1 and S2 as bright features. The experimental dispersion of the quasiparticle band $\varepsilon(\mathbf{k})$ (Fig. 29, right) is compared to the expected dispersion of the surface state without
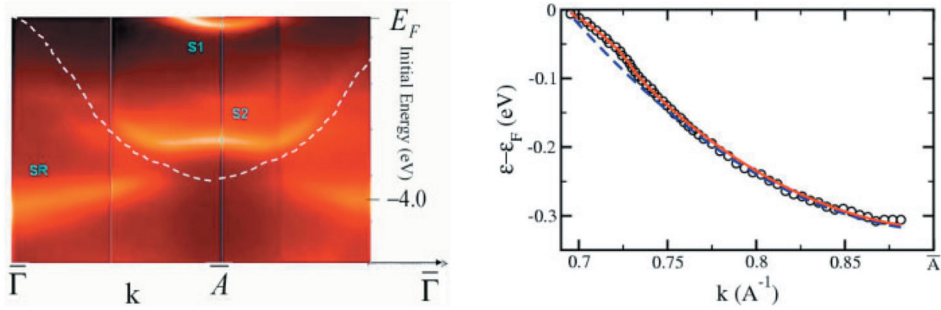
Fig. 29: (Left panel) Energy vs. momentum photoemission display of the two surface state bands S1 and S2 on Be($10\bar{1}0$). The dashed line is the bulk band edge. Data taken at 30 K at 40 eV photon energy. (Right panel) Quasi-particle dispersion determined from momentum distribution curves (circles) obtained at 24 eV photon energy. Dashed blue line is the bare particle dispersion $\epsilon_0(k)$ and the red line is the fit to the data from the extracted Eliashberg function. From [60].

additional interactions $\varepsilon_0(\mathbf{k})$. This comparison reveals a weak, but distinct deviation of the experimental data from the parabolic dispersion close to the Fermi energy. This kink is the spectral signature of the electron-phonon coupling.

Electron-magnon interaction   In a magnet we have collective excitations of the spin system – magnons. These quasiparticles have energies also in the 100 meV range. We should therefore expect that electron-magnon interaction leads to the appearance of kinks in the band structure of ferromagnetic materials. This is demonstrated for the photoemission from the Fe(110) surface. The ARPES data (Fig. 30) show the spectral distribution of the surface state photoemission close to E_F at the center of the surface Brillouin zone. A careful analysis of surface state dispersion reveals a characteristic deviation from the parabolic behavior in the regime down to 200 meV below the Fermi level. This broader kink structure can be indeed related to the electron-magnon interaction [37]. From these data it is possible to extract the strength and extension of the electron magnon interaction.

4.4.4   High-Energy Photoemission (HAXPES)

So far we have discussed effects in valence band and core level photoelectron spectroscopy at excitation energies below 1000 eV. As we know from the inelastic mean free path curves under these conditions we will have $\lambda_{in} \simeq 1\,\text{nm}$ at best, i.e. all of these experiments are more or less surface sensitive. In recent years there is a strong effort to extend photoelectron spectroscopy also to higher excitation energies up to 10 keV in order to overcome this limitation. The approach is coined HArd X-ray PhotoElectron Spectroscopy (HAXPES) and poses several experimental challenges [61]. First, the electron spectrometers must be modified to be able to measure photoelectrons with high kinetic energy and good energy resolution ($\Delta E < 100\,\text{meV}$). Second, the photoexcitation cross section for most core levels drops by 2-3 orders of magnitude, when going from 1 keV to 10 keV photon energy. As a consequence, the resulting photoelectron
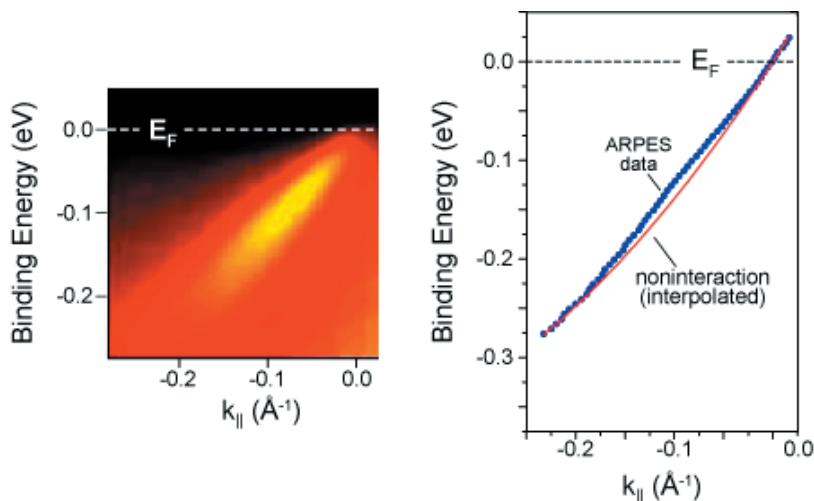
Fig. 30: ARPES data from the iron (110) surface state. Left: Raw data, showing the intense quasiparticle region. Right: The electron band dispersion (E vs. $k_\parallel$) extracted from the data reveals a weak "kink" in the region between 0.1 and 0.2 eV below E_F. From [37].

intensity will be small and difficult to measure. This can be only partially compensated on the primary side, i.e. by increasing the photon flux. At present, HAXPES experiments are still demanding and very difficult to carry out with laboratory sources.

The major advantage of HAXPES is its larger information depth which permits the access to buried layers and interfaces. The example shown in Fig. 31 is taken from the field of resistive oxides. Usually, oxides are wide band gap insulators. However, some of these materials may change their conductivity by several orders of magnitude, if a short current pulse above a certain threshold is applied to the material [62]. This current leads to the formation of conductive filaments or a local valency change in the oxide generating carriers for electrical transport. This is called the low resistive state (LRS). Interestingly, this process is reversible and the system may also be switched back into the high resistive state (HRS). This behaviour considered as a future memory principle and explains the strong interest in resistive oxides.

Ta_2O_5 is one of the promising materials that has been and still is thoroughly investigated with respect to resistive memory applications. Fig. 31 shows the comparison of photospectra taken from the Ta $4d$ core states with about 8 keV photon energy. In order to switch the conductivity of the Ta_2O_5 film a bottom and top electrode usually made from Pt is needed, through which the switching current is passed through the insulator. This means, however, that the photoemission experiment has to probe the region below the Pt electrode, which requires a sufficiently high information depth. As can be seen, the experiment is indeed able to find a difference in the relative core level intensities underneath the 10 nm thick Pt-electrode, which can be related to a change of the oxidation state from Ta^{5+} to Ta^{4+} between the HRS and LRS state [63]. This demonstrates that HAXPES is able – at least in principle – to follow and map the valency changes taking place during the resistive switching process.

Recent experiments have demonstrated that high-energy photoemission can also be used to carry out bulk-sensitive band mapping experiments [64–67].
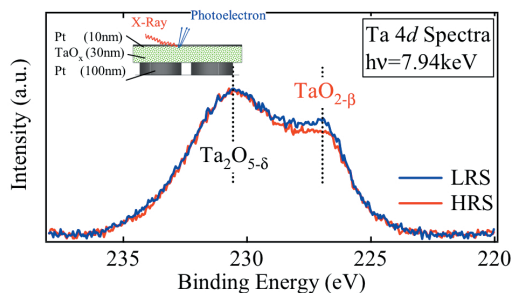
Fig. 31: Hard x-ray photoemission spectra from Ta_2O_5 in the high (HRS) and low resistive state (LRS). From [63].

In this contribution, we could only touch upon selected aspects of photon-based electronic spectroscopies and the photoexcitation processes. It should have become clear that these spectroscopies with their many facets provide powerful tools for an *en detail* electronic and chemical characterization of materials. Very important information can already be extracted by means of qualitative interpretation schemes. The full potential, however, can be unleashed only by quantitative descriptions within sophisticated photoexcitation calculations.
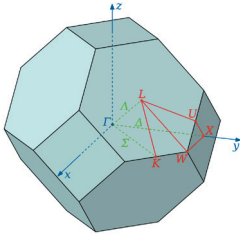
Fig. 32: Brillouin zone of an face-centered cubic crystal with high-symmetry directions (green) and high-symmetry points (red).

Appendices

Optical Selection Rules for Band States

The strength of the optical interband transitions in the electronic structure of a solid is determined by dipole selection rules. These dipole selection rules depend on the symmetry of the crystal and the experiment and have been derived from group-theoretical arguments. In the following, we will give an example for the dipole selection rules valid in a face-centered cubic crystal.

The symmetry of the electronic wavefunctions in the initial and final state of the photoemission process is described by so-called irreducible representations A_i of the symmetry group. The letter A usually refers to high-symmetry directions and points in the Brillouin zone (Fig. 32). Depending on the structure of the Hamiltonian (see, for example, eq. 2.1) one may distinguish *nonrelativistic* or *single-group representations* (spatial symmetry only), *relativistic* or *double-group representations* (including spin-orbit coupling) and *magnetic co-representations* (spin-orbit coupling and exchange interaction). The dipole operator of the photon field can also be expressed by an appropriate representation, taking into account the experimental geometry (direction of light incidence). For an fcc crystal, we have the following representations without and with spin-orbit coupling along the three high-symmetry directions

crystalline axis	symmetry label	nonrelativistic	relativistic
{001}	Δ	$\Delta_1, \Delta_2, \Delta_{2'}, \Delta_5$	$\Delta_6^1, \Delta_6^5, \Delta_7^2, \Delta_7^{2'}, \Delta_7^5$
{011}	Σ	$\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$	$\Sigma_5^1, \Sigma_5^2, \Sigma_5^3, \Sigma_5^4$
{111}	Λ	Λ_1, Λ_3	$\Lambda_{4,5}^3, \Lambda_6^1, \Lambda_6^3$

Table 1: Irreducible representations for the fcc lattice.

Electronic states with the highest symmetry have the lowest number, i.e. $\Delta_1, \Sigma_1, \Lambda_1$. In a photoemission process along the surface normal of a low-index surface, the electron is excited into a final state with the highest symmetry inside the crystal, which connects to a spherical wave (corresponding to a s wavefunction) outside the crystal propagating towards the detector. For this normal emission geometry, we have the following dipole selection rules between single-group representations [68]

In the presence of spin-orbit coupling, the interband transitions take place between double-group representations and lead to the excitation of spin-polarized photoemission due to the

initial state	final state			
	Δ_1	Δ_2	$\Delta_{2'}$	Δ_5
Δ_1	$\parallel$			$\perp$
Δ_2		$\parallel$		$\perp$
$\Delta_{2'}$			$\parallel$	$\perp$
Δ_5	$\perp$	$\perp$	$\perp$	$\perp$

initial state	final state			
	Σ_1	Σ_2	Σ_3	Σ_4
Σ_1	$\parallel$		X	Y
Σ_2		$\parallel$	Y	X
Σ_3	X	Y	$\parallel$	
Σ_4	Y	X		$\parallel$

initial state	final state	
	Λ_1	Λ_3
Λ_1	$\parallel$	$\perp$
Λ_3	$\perp$	$\parallel,\perp$

Table 2: Nonrelativistic dipole selection rules in a cubic lattice for normal emission along the three low-indexed directions for the vector potential **A** parallel ($\parallel$) and perpendicular ($\perp$) to the surface normal, and parallel to the x-axis (X) or y-axis (Y), respectively.

mechanism of optical spin-orientation. An example for the resulting relativistic selection rules with circularly polarized light are given in Table 3. For a full set of relativistic selection rules, the reader is referred to Ref. [69].

initial state	final state				
	Δ_6^1	Δ_7^2	$\Delta_7^{2'}$	Δ_6^5	Δ_7^5
Δ_6^1				$\vert\Uparrow\rangle$	$\vert\Downarrow\rangle$
Δ_7^2				$\vert\Downarrow\rangle$	$\vert\Uparrow\rangle$
$\Delta_7^{2'}$				$\vert\Downarrow\rangle$	$\vert\Uparrow\rangle$
Δ_6^5	$\vert\Downarrow\rangle$	$\vert\Uparrow\rangle$	$\vert\Uparrow\rangle$		
Δ_7^5	$\vert\Uparrow\rangle$	$\vert\Downarrow\rangle$	$\vert\Downarrow\rangle$		

Table 3: Relativistic dipole selection rules for normal emission along the Δ line and excitation with right-hand circularly polarized light. The spin-polarization is oriented along the surface normal (z-axis) and can be positive ($\vert\Uparrow\rangle$) or negative ($\vert\Downarrow\rangle$). For excitation with left-hand circularly polarized light, $\vert\Uparrow\rangle$ and $\vert\Downarrow\rangle$ have to be interchanged.

References

[1] S. Hüfner, *Photoelectron Spectroscopy* 3rd ed. (Springer, Berlin, 2003).

[2] *Solid-State Photoemission and Related Methods*, eds. W. Schattke, and M. A. van Hove (Wiley-VCH, Weinheim, 2003).

[3] F. Reinert and S. Hüfner, New J. Phys. **7**, 97 (2005).

[4] *Very High Resolution Photoelectron Spectroscopy*, ed. S. Hüfner (Springer, Berlin, 2007).

[5] F. de Groot and A. Kotani, *Core Level Spectroscopy of Solids* (CRC Press, Boca Raton, 2008).

[6] S. Suga and A. Sekiyama, *Photoelectron Spectroscopy* (Springer-Verlag, Berlin, 2014).

[7] R. M. Martin, *Electronic Structure – Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2004).

[8] V. Moruzzi, J. F. Janak, and A. R. Williams, *Calculated Electronic Properties of Metals*. (Pergamon Press, New York, 1978).

[9] K. Horn, in *Handbook of Surface Science*; Vol. 2, edited by K. Horn and M. Scheffler (Elsevier, Amsterdam, 2000), p. 383.

[10] *Optical Orientation*, edited by F. Meier and B. P. Zakharchenya (North-Holland, Amsterdam, 1984).

[11] R. N. Zare, *Angular Momentum* (Wiley, New York, 1988).

[12] D. T. Pierce and F. Meier, Phys. Rev. B **13**, 5484 (1976).

[13] J. Stöhr and H. C. Siegmann, *Magnetism: From Fundamentals to Nanoscale Dynamics*. (Springer-Verlag, Berlin, 2006).

[14] Y. Cauchois and C. Bonelle, C. R. Acad. Sci. Paris **245** (1957) 1230.

[15] R. Nakajima, J. Stöhr, and Y. U. Idzerda, Phys. Rev. B **59**, 6421 (1999).

[16] https://www-ssrl.slac.stanford.edu/stohr/xmcd.htm

[17] https://www-ssrl.slac.stanford.edu/stohr/nexafs.htm

[18] J. Stöhr, *NEXAFS Spectroscopy*. (Springer, Berlin, 1992).

[19] H. Wende, Rep. Prog. Phys. **67**, 2105 (2004).

[20] C. Schmitz-Antoniak, Rep. Prog. Phys. **78**, 062501 (2015).

[21] G. Schütz, W. Wagner, W. Wilhelm, P. Kienle, R. Zeller, R. Frahm, and G. Materlik, Phys. Rev. Lett. **58**, 737 (1987).

[22] C. T. Chen, F. Sette, Y. Ma, and S. Modesti, Phys. Rev. B **42**, 7262 (1990).

[23] I.P. Krug, *Magnetic Proximity Effects in Highly-Ordered Transition Metal Oxide Heterosystems*, PhD Thesis University Duisburg-Essen (2008), ISBN 978-3-89336-521-0.

[24] G. v. d. Laan, J. Phys.: Conf. Ser. **430**, 012127 (2013).

[25] S. Tanuma, C. J. Powell, and D. R. Penn, Surf. Interface Anal. **37**, 1 (2005); S. Tanuma, C. J. Powell, and D. R. Penn, Surf. Interf. Anal. **43**, 689 (2011).

[26] R. Feder, in: *Polarized Electrons in Surface Physics*, ed. by R. Feder (World Scientific, Singapore, 1985).

[27] G. Borstel, Appl. Phys. A **38**, 193 (1985).

[28] A. Damascelli, Z. Hussain, and Z.-X. Shen, Rev. Mod. Phys. **75**, 473 (2003).

[29] A. Oelsner, O. Schmidt, M. Schicketanz, M.J. Klais, G. Schönhense, V. Mergel, O. Jagutzki, H. Schmidt-Böcking, Rev. Sci. Instrum. **72** (2001) 3968.

[30] N. Takahashi, F. Matsui, H. Matsuda, Y. Hamada, K. Nakanishi, H. Namba, and H. Daimon, J. Electron Spectr. Rel. Phen. **163**, 45 (2008).

[31] B. Krömker, M. Escher, D. Funnemann, D. Hartung, H. Engelhard, and J. Kirschner, Rev. Sci. Instrum. **79**, 053702 (2008).

[32] C. Tusche, M. Ellguth, A. A. Ünal, C. Chiang, A. Winkelmann, A. Krasyuk, M. Hahn, G. Schönhense, and J. Kirschner, Appl. Phys. Lett. **99**, 032505 (2011).

[33] J. Kessler, *Polarized Electrons*, 2nd ed. (Springer-Verlag, Berlin, 1985).

[34] J. Kirschner, *Polarized Electrons at Surfaces*, Springer Tracts in Modern Physics Vol. 106 (Springer-Verlag, Berlin, 1985).

[35] M. Kolbe, P. Lushchyk, B. Petereit, H. J. Elmers, G. Schönhense, A. Oelsner, C. Tusche, and J. Kirschner, Phys. Rev. Lett. **107** (2011).

[36] L. Gregoratti, M. Marsi and M. Kiskinova, Synchr. Rad. News **12**, 40 (1999).

[37] M. Escher, N. Weber, M. Merkel, C. Ziethen, P. Bernhard, G. Schönhense, S. Schmidt, F. Förster, F. Reinert, B. Krömker, and D. Funnemann, J. Phys.: Condens. Matt. **17**, S1329 (2005).

[38] A. Bailly, O. Renault, N. Barrett, T. Desrues, D. Mariolle, L. F. Zagonel, and M. Escher, J. Phys.: Condens. Matt. **21**, 314002 (2009).

[39] V. N. Strocov, R. Claessen, G. Nicolay, S. Hüfner, A. Kimura, A. Harasawa, S. Shin, A. Kakizaki, H. I. Starnberg, P. O. Nilsson, and P. Blaha, Phys. Rev. B **63**, 205108 (2001).

[40] available as pdf-file at `http://xdb.lbl.gov/`

[41] S. Doniach and M. Sunjic, J. Phys. C **3**, 285 (1970).

[42] E. Rotenberg, Advanced Light Source Berkeley, (priv. communication).

[43] S. V. Borisenko, M. S. Golden, S. Legner, T. Pichler, C. Dürr, M. Knupfer, J. Fink, G. Yang, S. Abell, and H. Berger, Phys. Rev. Lett. **84**, 4453 (2000).

[44] E. Kisker, K. Schröder, W. Gudat, and M. Campagna, Phys. Rev. B **31**, 329 (1985).

[45] H. B. Rose, A. Fanelsa, T. Kinoshita, Ch. Roth, F. U. Hillebrecht, and E. Kisker, Phys. Rev. B **53**, 1630 (1996).

[46] E. Tamura and R. Feder, Europhys. Lett. **16**, 695 (1991).

[47] C. M. Schneider and J. Kirschner, Crit. Rev. Solid State Mater. Sci. **20**, 179 (1995).

[48] E. I. Rashba, Sov. Phys. Solid State **2**, 1109 (1960).

[49] S. LaShell, B. A. McDougall and E. Jensen, Phys. Rev. Lett. **77**, 3419 (1996).

[50] F. Reinert, J. Phys.: Condens. Matt. **15**, S693 (2003).

[51] W. Kuch and C. M. Schneider, Rep. Prog. Phys. **64**, 205 (2001).

[52] C. Bethke, E. Kisker, N. B. Weber, and F. U. Hillebrecht, Phys. Rev. B **71**, 024413 (2005).

[53] D. Venus, Phys. Rev. B **43**, 6144 (1993).

[54] F. Manghi, V. Bellini, J. Osterwalder, T. J. Kreutz, and P. Aebi, C. Arcangeli, Phys. Rev. B **59**, R10409 (1999).

[55] K. Ono, K. Shimada, Y. Saitoh, T. Sendohda, A. Kakizaki, T. Ishii, and K. Tanaka, J. Electron Spectr. Rel. Phen. **78**, 325 (1996).

[56] Y. Sakisaka, T. Komeda, M. Onchi, H. Kato, S. Masuda, and K. Yagi, Phys. Rev. B **36**, 6383 (1987).

[57] F. Manghi, V. Bellini, and C. Arcangeli, Phys. Rev. B **56**, 7149 (1997).

[58] R. Clauberg, W. Gudat, E. Kisker, E. Kuhlmann, and G. M. Rothberg, Phys. Rev. Lett. **47**, 1314 (1981).

[59] G. Grimvall, *The Electron-Phonon Interaction in Metals*, Selected Topics in Solid State Physics, edited by E. Wohlfarth (North-Holland, New York, 1981).

[60] S.-J. Tang, J. Shi, B. Wu, P. T. Sprunger, W. L. Yang, V. Brouet, X. J. Zhou, Z. Hussain, Z.-X. Shen, Z. Zhang, and E. W. Plummer, phys. stat. sol. (b) **241**, 2345 (2004).

[61] K. Kobayashi, Nucl. Instrum. Methods A **601**, 32 (2009).

[62] R. Waser and M. Aono, Nat. Mater. **6**, 833 (2007).

[63] H. Kumigashira (priv. communication).

[64] L. Plucinski, J. Minar, B. C. Sell, J. Braun, H. Ebert, C. M. Schneider, and C. S. Fadley, Phys. Rev. B **78**, 035108 (2008).

[65] F. Venturini, J. Minar, J. Braun, H. Ebert, and N. B. Brookes, Phys. Rev. B **77**, 045126 (2008).

[66] C. S. Fadley, J. Electron Spectr. Rel. Phen. **178-179**, 2 (2010).

[67] J. C. Woicik, in Springer Series in Surface Sciences (Springer International Publishing Switzerland, Cham, 2016).

[68] W. Eberhardt and F. J. Himpsel, Phys. Rev. B **21**, 5572 (1980).

[69] M. Wöhlecke and G. Borstel, Phys. Rev. B **24**, 2857 (1981); M. Wöhlecke and G. Borstel, Phys. Rev. B **23**, 980 (1981); G. Borstel and M. Wöhlecke, Phys. Rev. B **24**, 2321 (1981).

F 3   Scanning tunneling microscopy and spectroscopy

Ph. Ebert

Peter Grünberg Institut

Forschungszentrum Jülich GmbH

Contents

1   Introduction

Since the invention of the first scanning probe microscope the scanning tunneling microscope by Binnig and Rohrer in 1982 [1], scanning probe microscopy techniques have rapidly developed into increasingly important tools for the characterization of surface structures and material properties. The surge in applications of scanning probe microscopes is primarily due to their unique ability to provide real space images with atomic resolution of surface structures. Depending on the interaction probed, scanning probe microscopes provide the possibility to investigate periodic and nonperiodic electrical, topographic, optical, magnetic, and many other types of properties. In contrast to scattering experiments, scanning probe microscopy techniques have the advantage to provide images of a variety of surface structures in *real space*. Therefore, they are also frequently used as a complement to scattering experiments. This lecture will focus on the parent of all scanning probe microscopes, the scanning tunneling microscope, and illustrate its operation principle, the physical background, as well as measurement and detection capacities.

A scanning probe microscope works according to a simple principle (Fig. 1): A probe is scanned over the surface of interest at a small distance, where an interaction between the probe and the surface is present. This interaction can be of various nature (electrical, magnetical, mechanical, etc.) and provides the measured signal (tunnel current, force, etc). Depending on the quality and type of probe, the measured signal can be observed to reproducibly vary at atomic distances during scanning of the surface. If these scanning processes are put together line by line, an "image" of the surface is obtained. Such an image shows the spatial variation of the measured parameter, e.g., of the tunnel current.
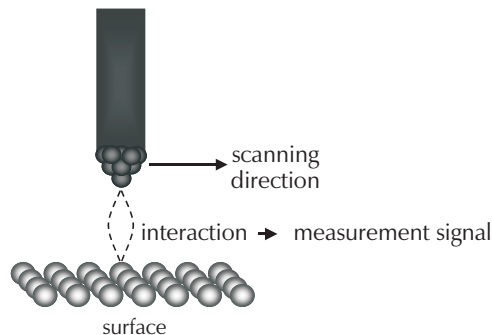
Fig. 1: Principle of a scanning probe microscope. A probe tip is scanned over a surface. The interaction with the surface yields a signal which is used to derive a real space image of the surface.

Of all the scanning probe microscopes the first invented scanning tunneling microscope (STM) still provides the highest routinely achieved resolution. It is used for semiconductors, metals, and superconductors, because it constitutes a probe for electrical properties and thus requires electrically conducting surfaces. It measures the tunnel current between a metallic, extremely fine tip and the surface. Thereby the STM probes the local density of surface states, whose atomic-scale variations allows to image surfaces with atomic resolution. Fig. 2 shows as example a STM image of the InP(110) surface. Each bright local peak represents the increased

local density of states near an atom on the surface. Black holes indicate missing atoms, i.e., vacancies [2]. Thus, one can recognize atomic rows and individual point defects.
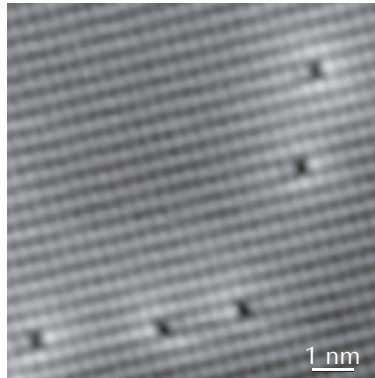
Fig. 2: Scanning tunneling microscope image of several phosphorus vacancies on a n-doped InP(110) surface. The image was obtained at negative voltages applied to the sample. Thus, the electrons tunneled from the filled InP(110) surface states into empty tip states. Adapted with permission from [2], ©1994 American Physical Society.

2 Theoretical fundamentals

A scanning tunneling microscope uses a fine metallic tip (called tunneling tip) as probe. A bias voltage is applied between the tip and the electrically conducting sample surface (see Fig. 3). Then the tip is approached toward the surface until a electrical current flows. This happens at tip-surface separations in the order of 0.5 to 1 nm. The current is a tunnel current based on the quantum-mechanical *tunnel* effect [3–5]. After a tunneling contact is established, the tip is moved laterally over the surface by a piezoelectric *scanning* unit, whose mechanical extension can be controlled by applying appropriate voltages. The scanning unit is typically capable of scanning an area of a few nm^2 up to several μm^2. Thereby a *microscopic* image of the spatial variation of the tunnel current is acquired. Hence the name *scanning tunneling microscope*.

At this stage we discuss what kind of atomic-scale structures can be made visible by utilizing the tunnel effect in the scanning tunneling microscope. These structures must by nature correspond to electrical states from or into which the electrons can tunnel. In the tunneling process, the electrons must tunnel through the vacuum gap between the tip and the sample surface. This vacuum gap represents a potential barrier, i.e. an energetically forbidden region. The tunnel effect allows a particle (here an electron) to tunnel through this potential barrier even though the electron's energy is lower than the barrier height. The probability of such a process decreases exponentially with the geometrical distance between the tip and the sample (determining the width of the potential barrier) and with increasing barrier height. An experimental apparatus making use of the tunnel effect must therefore minimize the width of the potential barrier to the degree that electrons can tunneled through it. This is realized in the scanning tunneling microscope configuration by moving the tip very close (about 1 nm or less) to the surface. The electrons can then pass between the surface and the tip. The direction of the *macroscopic*
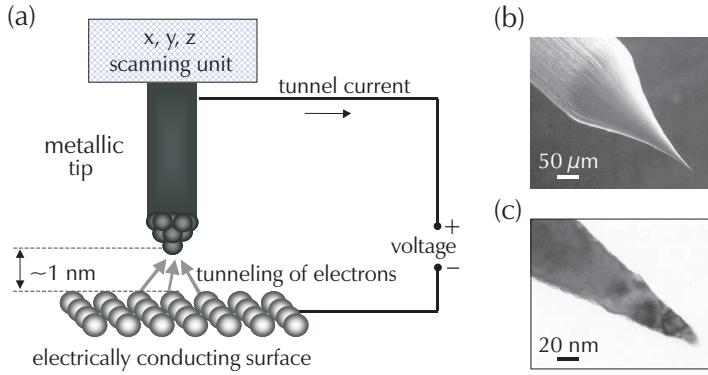
Fig. 3: (a) Schematic drawing of a classical scanning tunneling microscope. The tunnel current is used as measuring signal. (b) and (c) show scanning and transmission electron microscope images, respectively, of a typical tungsten tips used for a classical scanning tunneling microscope with no spin sensitivity. Note the sharpness of the tips, which have a radius of curvature below 10 nm.

tunnel current is fixed by applying a voltage between sample and tip, even if electrons tunnel in both directions, but with different probabilities due to the applied voltage. Note, the tunneling process of an electron through an energetically forbidden region is instantaneously and thus, the tunneling electron does not stay a measurable time span in the forbidden potential barrier region [6].

In order to explain and interpret the images of the surface states obtained in this way, efforts to develop a theory were made soon after the invention of the scanning tunneling microscope. One of the possible theoretical approaches is based on Bardeen's idea of applying a transfer Hamiltonian operator to the tunneling process [7]. This had the advantage of adequately describing the many-particle nature of the tunnel junction. In the model, a weak overlap of the wave functions of the surface states of the two electrodes (tunneling tip and sample surface) is assumed to allow a perturbation calculation. The resulting current between two planar electrodes is then given by

$$I \sim \int_{-\infty}^{\infty} |M(E)|^2 \cdot \rho_{\text{tip}}(E - eV) \cdot \rho_{\text{sample}}(E) \cdot [f(E - eV) - f(E)] dE \qquad (1)$$

with $f(E)$ being the Fermi function, M the tunneling matrix element, ρ_{sample} and ρ_{tip} the density of states of the sample and tip, respectively, and E the energy of the density of states.

On this basis, Tersoff and Hamann developed a simple theory of scanning tunneling microscopy [8, 9]. By assuming that the tunneling tip can be approximated by a metallic s-orbital with its center at the position $\boldsymbol{R}_{\text{tip}}$, as shown schematically in Fig. 4, they obtained for the tunnel current in a STM-like configuration:

$$I \sim V \cdot \rho_{\text{tip}}(E_{\text{F}}) \cdot \rho_{\text{sample}}(\boldsymbol{R}_{\text{tip}}, E_{\text{F}}) \qquad (2)$$

In addition, is was assumed that low voltages V (i.e., much smaller than the work function) are

applied in order to linearly approximate the voltage dependence (see below for the high voltage extension of Eq. 2). $\rho_{\text{tip}}(E_{\text{F}})$ is the density of states of the tip and $\rho_{\text{sample}}(\boldsymbol{R}_{\text{tip}}, E_{\text{F}})$ is that of the sample surface at the center $\boldsymbol{R}_{\text{tip}}$ of the tip orbital and at the Fermi energy E_{F}. Equation 2 shows that at low voltage *the scanning tunneling microscope thus images the electronic density of states at the sample surface near the Fermi energy*. However, this result also means that the scanning tunneling microscope images do not directly show the atoms, but rather the electronic states bound to the atoms. If we recall Fig. 2 the maxima are thus the filled states localized above the phosphorus atoms on the InP(110) surface and the dark holes are missing states arising from vacancies at the surface.
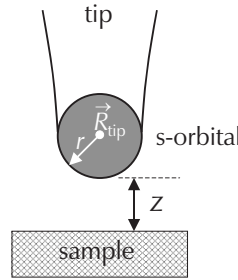
Fig. 4: Schematic representation of the tunneling geometry used in the Tersoff-Hamann model. The tip approximated by a s orbital with a radius r at the position $\boldsymbol{R}_{\text{tip}}$.

As can be seen in Eq. 2, the density of states of the probe tip enters in the measurement in the same way as the density of states of the sample. Thus depending on the exact density of states of the tip, the tunnel current will vary from tip to tip. It is therefore desirable to know the exact electronic state of the tip, but unfortunately, in actual experiments, every tip is different and the details remain almost always unknown, despite intense efforts to characterize the tips' apexes. Tip effect were nevertheless successfully distinguished from the real surface structure by careful measurements with a large number of tip configurations. Different tip configurations can be obtained during scanning over the surface by attracting individual atoms from the surface to the tip, as well as by special tip treatments, heating to high temperatures for cleaning in vacuum, ion sputtering, or field emission.

Equation 2 can be better interpreted by considering the exponential decay of the density of surface states into the vacuum with the effective inverse decay length κ_{eff}:

$$\kappa_{\text{eff}} = \sqrt{\frac{2m_{\text{e}}B}{\hbar^2} + \left|\boldsymbol{k}_{\|}\right|^2} \tag{3}$$

m_{e} is the effective mass of the electron, $\boldsymbol{k}_{\|}$ is the parallel wave vector of the tunneling electrons, which enters into the Eq. 3 due to the momentum conservation in the tunneling process [10]. B is the barrier height of the vacuum gap between the tip and the sample surface. The barrier height is a function of the applied voltage V and the work functions Φ_{sample} and Φ_{tip} of the sample and tip [11], respectively. It can be approximated to:

$$B = \frac{\Phi_{\text{tip}} + \Phi_{\text{sample}}}{2} - \frac{|eV|}{2} \tag{4}$$

The tunnel current thus decreases exponentially with the tip-sample distance z:

$$I \sim \exp\left[-2\kappa_{\text{eff}}z\right] \tag{5}$$

The exponential current–tip-sample distance dependence is essential for the high accuracy of a scanning tunneling microscope: First, very small changes in the tip-sample separation cause large changes in the tunnel current. This yields a high vertical resolution. Second, the tip just needs one nanotip, only about 0.1 nm closer to the surface than all other neighboring nanotips. Then essentially all the tunnel current flows only over this closest nanotip. Thus, even apparently wide and blunt tips can yield atomic resolution along the surface due to the exponential current-distance dependence and the presence of nanotips on the macroscopic tunneling tip.

The description of the tunnel current by Eq. 2, however, has an important restriction: it only applies to low voltages V, which multiplied by e must be much smaller that the work function ($eV \ll \Phi_{\text{sample}}$). This is reasonably correct for the tunneling conditions used to image metal surfaces. However, for the investigation of semiconductor surfaces, voltages of the order of 2 to 3 V, sometimes even larger as in case of GaN cleavage surfaces [12, 13] are required due to the existence of a wide band gap. Therefore the applied voltages times the electron charge e are in the same magnitude as the work function, and the above used approximations are insufficient. Thus the theory must be extended. The simplest extension yields:

$$I \sim \int_{E_{\text{F,tip}}}^{E_{\text{F,tip}}+eV} \rho_{\text{tip}}(W)\rho_{\text{sample}}(W + eV, \boldsymbol{R}_{\text{tip}})T(W, V)dW \tag{6}$$

$T(W,V)$ is a transmission coefficient, which depends on the energy of the electrons and the applied voltage. The transmission coefficient arises from the increased tunneling probability for surface states with smaller ionization energy (leading to a smaller effective tunneling barrier) and for the voltage dependence of the tunneling barrier. The transmission coefficient can be well approximated, if one considers the exponential decay of the density of states into the vacuum and the fact that the tunnel current is based of the density of states of the sample at the position of the tip $\boldsymbol{R}_{\text{tip}}$. This position corresponds to a tip-sample separation z and thus the transmission coefficient describes the z dependence of the density of states for a given energy and voltage. In case of positive voltages and zero parallel wave vector (tunneling from the Γ point) $T(W, V)$ can be thus approximated by [14]:

$$T(W, V) = \exp\left[-2z \cdot \sqrt{\frac{2m_{\text{e}}\left[\frac{\Phi_{\text{tip}}+\Phi_{\text{sample}}}{2} + \frac{eV}{2} - W\right]}{\hbar^2}}\right] \tag{7}$$

The tunnel current is thus composed of the product of the density of states of the tip and sample at all the different electron energies that are allowed to participate in the tunneling process (Fig. 5). For example, an image measured at -2 V applied to the sample, consequently shows all occupied sample states with an energy between the Fermi energy and 2 eV below the Fermi energy. In analogy tunneling at a positive voltages applied to the sample provides a measurement of the empty surface states in an energy interval determined again by the voltage.
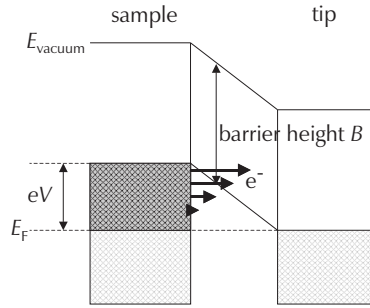
Fig. 5: At high voltages not only the states near the Fermi energy E_F contribute to the current but all states whose energy ranges between E_F and $E_F + eV$.

This effect can be illustrated further using the InP(110) surface, which has two surface states: an occupied state below the valence band edge and an empty state above the conduction band edge (Fig. 6). All other states are located geometrically deeper in the crystal or energetically deeper in the bands. They thus contribute only at high voltages [15], which will not be considered here.

In the special case of the InP(110) surface, the occupied surface state is spatially located above the P atoms, whereas the empty state is bound to the In atoms (Fig. 6c1,c2). The P and In atoms are alternately arranged in zigzag rows. At negative sample voltages, the scanning tunneling microscope probes the occupied states located at the P sublattice, whose electrons tunnel into the empty states of the tunneling tip (Fig. 6a). Conversely, only the empty surface states at the In sublattice are probed at positive voltages applied to the sample (Fig. 6b) [15]- [17]. If the voltage polarity is changed every scan line, i.e. the occupied and the empty states are probed each alternating scan line, the two resulting images can be superimposed and the zigzag rows of alternating indium and phosphorus atoms become visible (Fig. 6c3).

Apart from the spatial distribution of the density of states, its energy dependence can be determined from current-voltage characteristics using Eq. 6. In order to do so, however, information is required about the transmission coefficient, which turns out to be a great obstacle even if approximations [18] are used. Therefore, in most cases, an experimentally viable approach is used, in which the density of states is approximated to [19] [20]:

$$\rho_{\text{sample}}(eV) \approx (dI/dV)/(\overline{I/V}) \tag{8}$$

In this case the transmission coefficient in Eq. 6 is approximated by $\overline{I/V}$ to remove the distance dependence (division by I) and the voltage dependence (multiplication by V). The overline means that the voltage dependence of I/V is strongly smoothened to avoid singularities in the density of states ρ in the band gap region where no current is measured. Despite these limitations it is well possible to experimentally measure the density of states as a function of the energy relative to the Fermi level using Eq. 8.
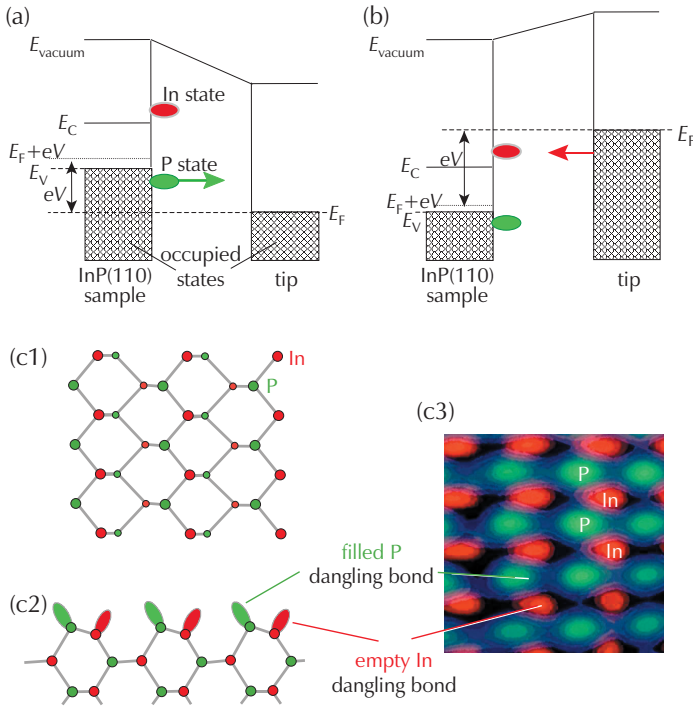
Fig. 6: Tunneling process at (a) negative and (b) positive voltages applied to the InP(110) surface. At negative voltages occupied states at the P atoms contribute to the tunnel current, while at positive voltages empty In-derived states dominate the current flow. (c1) Schematic top view and (c2) side view of the (110) surfaces of III-V compound semiconductors. (c3) Superposition of two scanning tunneling microscope images measured at positive (red) and negative (green) voltage. The density of state maxima correspond to the surface states at the In and P atoms, respectively.

3   Operating modes

At this stage the experimental operation of a scanning tunneli ng microscope is addressed. The simplest manner to obtain a scanning tunneling microscope image is to directly measure the variation of the tunnel current as a function of the scanning position while keeping the distance z between tip and sample surface constant. A so-called current image is then obtained. Instead of directly recording the atomic variation of the current, however, the usual procedure is to keep the tunnel current constant while scanning over the surface. This is done by changing the distance z between tip and surface using a feedback loop (Fig. 7a). In order to get an image, one records the voltage applied to the piezoelectric crystal (z-piezo), which adjusts the tip-sample distance such that the tunnel current is kept constant as schematically illustrated in Fig. 7b and 7c. This yields a constant-current STM image.

The constant-current mode is the preferred operation mode for most measurements, because it compensates with help of the feedback drifts in the tip-sample separation, which in current
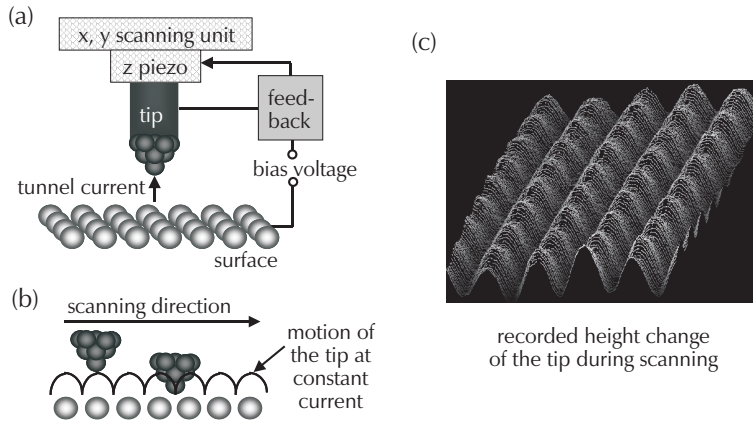
Fig. 7: (a) Schematic drawing of a scanning tunneling microscope with feedback loop used to keep the tunnel current constant while scanning by adjusting the tip-sample separation using a z piezoelectric element. (b) Motion of the tip in the constant-current mode due to the adjusting of the tip-sample separation. (c) The STM image is obtained by recording the voltage necessary to adjust the tip-sample separation in the constant-current mode for a large number of individual scan lines. The voltage is proportional to changes in the tip-sample separation.

STM images would lead to a huge change in current due to the high tip-sample separation sensitivity (see Eq. 5). Such drift-induced current changes would otherwise obscure the atomic scale information.

A further operation mode is the spectroscopy acquisition by STM. It is usually done by interrupting the feedback in order to keep the tip-sample separation constant during acquisition of the $I - V$ spectroscopy data. This can be done at any desired surface spot or for every pixel in a STM image. Although the shortly interrupted feedback loop keeps the tip-sample separation in principle constant, there are two effects, which may lead to changes in the tip-sample separation. First a possible drift changing the tip-sample separation can be controlled by a sufficiently long equilibration time of the system and a fast $I - V$ data acquisition. Second, fluctuations in the electronic structure, e.g., dopant atoms or fluctuations of the concentration of dopant atoms, lead to different tip-sample separations for identical set conditions (set current and set voltage). For a proper comparison of spectra from different locations, the individual spectras' tip-sample separations need to be recalibrated by the acquisition of additional current–tip-sample separation curves from which one can determine κ_{eff} in Eq. 5 [21] [22]. Once this is done an exact and quantitative comparison of spectroscopy data from different surface spots is possible.

4   Experimental realization

A large variety of different scanning tunneling microscope designs were developed, in order to adjust it best the needs of the individual research projects. Of course, it is not possible to discuss all designs here and it is preferable to refer to selected references [23]- [25]. In the following, a design developed at the Research Center Jülich [26] will be discussed in more detail to outline

the general operating principles, which are – with modifications – the same as for other scanning tunneling microscopes. In particular, the surface is always scanned with the aid of piezoelectric adjusting elements.
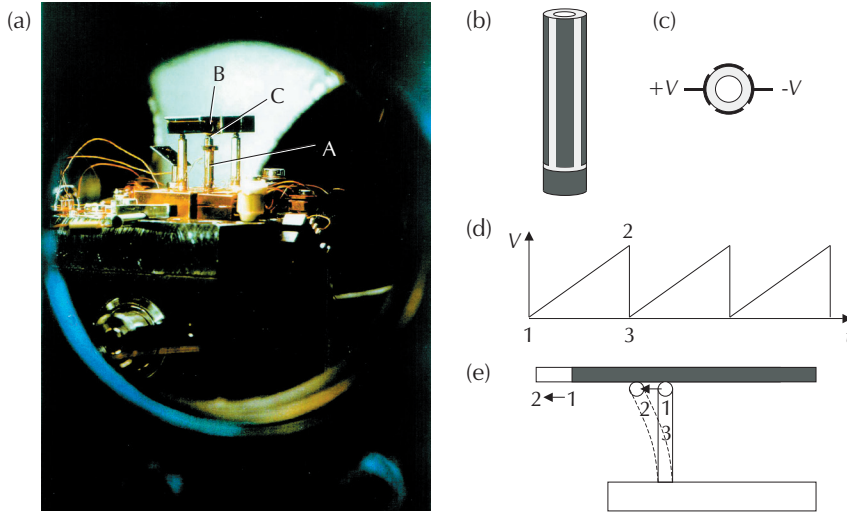
Fig. 8: Example of a scanning tunneling microscope. (a) View through a window flange at the vacuum chamber showing the STM with the scanner tube (A), the tip (B), and the sample holder (B). (b) Detailed view of a piezoelectric tube with four metallization fields. (c) View from top on a piezoelectric tube showing the metallization fields in dark and the voltage connections for bending the tube. (d) Voltage applied on the piezoelectric tube to shift the sample holder as shown in (e) using its inertia during the fast retreat of the bended tube back to its initial position.

In order to obtain atomically resolved images, the scanning tunneling microscope must have a high mechanical stability, such that no uncontrolled movements take place between the tip and the sample surface during the measurement. How critical the mechanical design of a microscope is, may be recognized by the fact that the tip must be positioned relative to the sample surface with a precision one order of magnitude better than the measuring accuracy required, i.e. horizontally within approx. 10 pm and vertically within 1 pm. The desired mechanical properties are achieved, for example, with a microscope that consists of radially polarized piezoelectric tubes arranged to form an equilateral triangle (Fig. 8). These three (outer) tubes carry the sample holder with the sample (B in Fig. 8). A fourth tube, the scanning tube, is glued in the center of the triangle. A small z-piezoelectric element, which holds the tip (C in Fig. 8), is mounted on the scanning tube for decoupling the z-motion from the x- and y-scanning motions. The inner metallization of the piezoelectric tubes are electrically connected to ground. Each piezoelectric tube has four additional metallizations on the outside to which, e.g., the scanning voltages or voltages required to move the sample laterally, are applied (Fig. 8b and c). Due to the radial polarization of the piezoelectric material, the tubes can be bent, elongated or shortened.

One of the three outer tubes supporting the sample holder is mounted on a mobile base plate, whereas the other two and the scanning tube are mounted on a common fixed base plate. The coarse approach is achieved by raising and lowering the piezoelectric tube mounted on the

mobile base plate. This allows us to adjust mechanically the tip-sample separation until the separation is small enough for the z-piezo's extension to control the adjustment of the tip-sample separation, e.g., keep the tunnel current constant.

The sample holder can be moved by bending the three outer piezoelectric tubes. If one prefers to move the sample over larger distances, then voltage pulses as shown in Fig. 8d are applied. They bend the outer tubes slowly in the desired direction and retract them very fast. Due to the inertia of the sample holder, it can only follow the slow bending, but not the fast retraction (Fig. 8e). Then the piezo tubes will glide back, while the sample holder rests in its bended position. A repetition of this process makes it possible to reach any location on the sample holder.

The whole microscope rests on several damping rings in a ultrahigh vacuum chamber, which is positioned on a compressed-air-damped table of approximately 1 t weight. The measurements are performed at a pressure of less than 1×10^{-8} Pa in the vacuum chamber to ensure that the surfaces remain clean. Fig. 8a shows such a scanning tunneling microscope viewed through one of the window flanges of the vacuum chamber.

The preparation of the tunneling tips is one most crucial part is operating a scanning tunneling microscope, because the tunneling tip as probe is directly affecting the quality of the measurement results. One possibility of preparing tunneling tips is the electrochemical etching of polycrystalline tungsten wire with NaOH. The tips produced in this manner have a radius of curvature of only 5 nm as shown in Fig. 3b and c.

5   Applications of the scanning tunneling microscope

The scanning tunneling microscope covers a wide field of applications wherever information about the surface structure is required in real space. The applications are so widely distributed over many research fields, ranging from biology to crystallography, that it is essentially impossible to provide a full overview of the possibilities to apply the scanning tunneling microscope. For a more complete overview consult Refs. [25] [27] [28]. Here only selected examples will be presented, which illustrates the potential of a scanning tunneling microscope.

5.1   Atomic-scale investigation of surface defects

Due to its high spatial resolution, the scanning tunneling microscope is an ideal instrument for the examination of lattice defects. In particular, the electronic and structural properties of point defects such as individual vacancies and dopant atoms can be measured, which has not yet been possible with other methods on the atomic scale. As an example, we consider P vacancies on InP(110) surfaces. In order to produce a vacancy, an atom must be removed from the surface. Therefore, three bonds are broken and so-called dangling bonds, i.e. unsaturated bonds, are left. These unsaturated bonds do not represent the energetically most favorable configuration and they reconstruct forming three defect energy levels. Defect energy levels are electron states located at or in a vacancy. In the case of high defect concentrations, the defect energy levels change the electrical properties of whole crystals, which is utilized e.g. in doping semiconductor crystals with impurities. As shown in Fig. 6, a scanning tunneling microscope only images either the occupied or the empty states. Since the occupied states correspond to the positions

of the phosphorus atoms in the surface, a missing occupied surface state is the signature of a phosphorus vacancy. This missing occupied surface state can be seen in Fig. 9 in the case of phosphorus vacancies on p-doped InP(110) surfaces. In addition, Zn dopant atoms are visible, which are surrounded by a local elevation (Fig. 9) due to their negative charge. In contrast on p-doped InP(110) surfaces the phosphorus vacancies are surrounded by a depression (Fig. 9), because of their positive charge [2, 29, 30] .
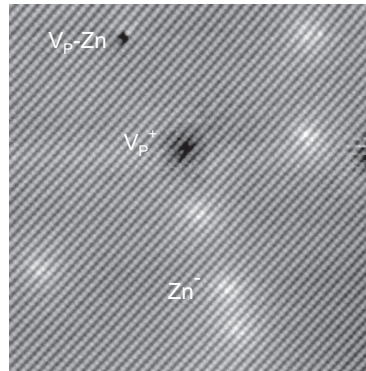
Fig. 9: Overview of defects occurring on p-doped InP(110) surfaces. In addition to phosphorus vacancies appearing as black depressions (V_P^+, white elevations surounding Zn dopant atoms (Zn^-) can be observed (sample voltage -2.2 V). Adapted with permission from [29], ©1996 American Physical Society.

5.2   Probing the local potential

The above example shows that local screened Coulomb potentials surrounding charged defects and dopant atoms [31] are visible in STM images. The question is now, how can the local potential influence the tunnel current and thus be visible in STM images? Figure 10a shows a STM image of a two-dimensional semiconducting $\sqrt{3} \times \sqrt{3}$ Ga overlayer on Si(111). Each maximum in the empty state STM image corresponds to one empty dangling bond above a Ga adatom. The weaker maxima (marked D) arise from Si atoms located on $\sqrt{3} \times \sqrt{3}$ Ga sites. These Si atoms act as donors and provide the free electrons. The resulting positive charges of the Si dopants induce a redistribution of the free charge carriers and thereby a potential change, which gives rise to the surrounding bright contrast on which the atomic corrugation is superimposed. The local potential change also shows up in the tunneling spectra: the valence E_V and conduction band E_C edges shift 0.15 eV to higher energies with increasing spatial separation from the dopant site (dotted lines in Fig. 10c) [14].

In addition, the STM images exhibit long-range height changes with lateral extensions of 5 to 15 nm (see dashed and dotted elliptical lines in Fig. 10a). Furthermore, regions with dark and bright contrast appear in surface areas with low and high dopant concentration, respectively. The above observed band edge shifts indicate again that potential changes along the surface give rise to the long-range height changes in STM images.

The sensitivity of the tunnel current to the potential can be illustrated using Fig. 11a, which
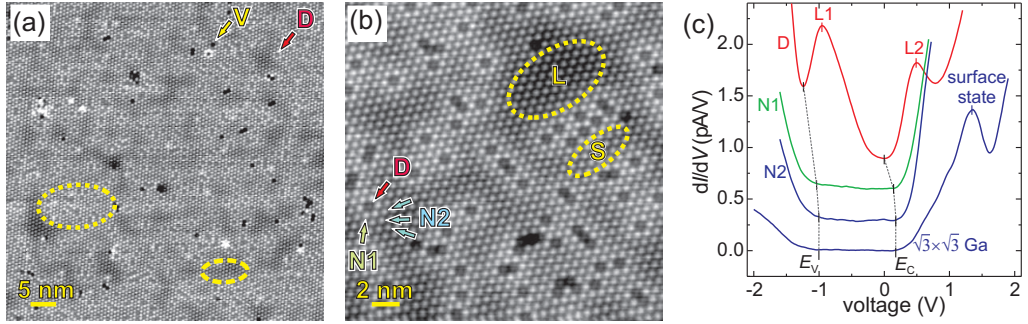
Fig. 10: (a) STM image of a 2D semiconducting $\sqrt{3} \times \sqrt{3}$ Ga overlayer on Si(111) measured at 2 V. In addition to atomic-sized features arising from vacancies (V) and Si dopant atoms (D), long-range changes in the contrast occur. Examples are indicated by dashed (depression) and dotted (elevation) elliptical lines. (b) High-resolution STM image. Each charged Si dopant is surrounded by a bright contrast. (c) dI /dV tunneling spectra above dopants (D), directly neighboring Ga atoms (N1), Ga atoms further away (N2) [see (b)], and for the $\sqrt{3} \times \sqrt{3}$ Ga overlayer (all Ga sites not neighboring to dopants). For enhanced sensitivity, the spectra D, N1, and N2 were taken at a tip-sample separation 0.06 nm smaller and each curve is offset by 0.3 pA/V for clarity. The conduction (E_C) and valence (E_V) band edges shifts are indicated by a dotted line. L1 and L2 are localized states related to the Si donors. Reprinted with permission from [14]. ©2009 American Institute of Physics

shows schematically a tunnel contact between a metallic tip and a semiconductor. With no potential change (black lines) the tunnel current is the sum of all electrons tunneling from the electron states between E_F and $E_F + eV$ into the empty sample states. This is schematically shown by the black triangle. In the presence of a band bending (or any potential fluctuation) the band edges is shifted (red lines) and as a result the barrier is modified (red double ended arrow) and the tunnel current increases as shown by the red triangle (in case of a negative potential change). Thereby the tunnel current sensitively changes with any potential fluctuation. The feedback loop keeps the tunnel current nevertheless constant by changing the tip-sample separation, which yields the contrast changes in Fig. 10 a and b.

The underlying potential fluctuations in Fig. 10 a and b can be quantitatively derived from the contrast fluctuations by relating quantitatively the *local* height change Δz from the spatial average z_{avg} of tip-sample separation ($z = z_{avg} + \Delta z$) to the *local* potential [14]. This allows to relate the local potential with the local dopant concentration n_{local} as shown in Fig. 11b. This data can be well described [14] by

$$n_{\text{local}} = \frac{6m_{eff}kT}{\pi\hbar^2} ln \left[1 + e^{-(E_{C,\text{avg}} - E_F + \Delta E_C)/kT}\right] \qquad (9)$$

as shown by the solid fit curve and the effective masses obtained in Fig. 11c. They correspond well to the effective mass of a Si(111) plane of 0.37 m_0.

This example thus illustrates that it is possible to investigate quantitatively the local potential induced by local fluctuations in the distribution of dopant atoms and/or defects with the aid
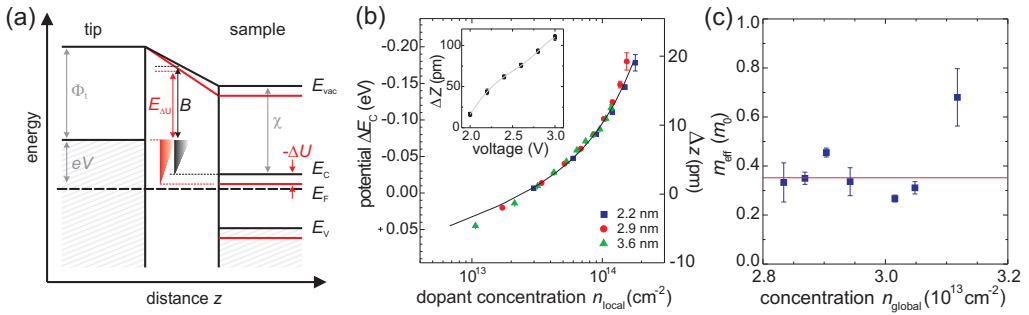
Fig. 11: (a) Energetic diagram showing the potential sensitivity of the scanning tunneling microscope. The tunnel contact between a metallic tip and a semiconducting surface is shown in black lines. The barrier B is indicated by a black double ended arrow. The total tunnel current is indicated by the black triangle. In case of a potential change in the semiconductor $(-\Delta U)$ the barrier changes (red double ended arrow) and the total tunnel current increases in the case of a negative potential change (red triangle). (b) Local deviation from the average potential ΔE_C (left axis) derived from the local height change $-\Delta z$ as a function of the local dopant concentration n_{local} for different resolutions. The solid line is a fit to the data. Inset: corresponding heightvoltage curve at a set current of 0.1 nA, probing the energy (voltage) sensitivity of the tip. (c) Effective mass m_{eff} vs the global dopant concentration n_{global} obtained from fitting different data sets.

of a scanning tunneling microscope. The additional deeper analysis performed in Ref. [14] shows that one can even determine the exact origin of the different potential fluctuations in a two-dimensional semiconductor with disordered dopants.

5.3   Scanning tunneling microscopy of nano-scale device structures

Using the so-called cross-sectional technique, where a grown sample is cleaved perpendicular to its growth direction to expose a cross-section of the growth structure, it is also possible to investigate hetero- and homostructures with atomic resolution. Figure 12a shows a large scale cross-sectional scanning tunneling microscopy (STM) overview of several 30 nm wide p- and n-doped GaAs layers cleaved along a (110) plane (C and Si dopant atom concentrations of $(5\pm1)\times10^{18}$ and $(4\pm1)\times10^{18}$ cm^3, respectively). An atomically resolved image is shown in Fig. 12b. The p- and n-doped layers are separated by lines with a darker contrast, whereas the doped layers themselves appear both bright. The n- and p-doped layers were identified on basis of the growth sequence, secondary ion mass spectra, and tunneling spectra showing a typical p and n type behavior.

The dark lines between the p- and n-doped layers were found to be the image of the depletion zones localized at $p-n$ interfaces, where the Fermi-energy is close to midgap [33, 34]. Thus the dark lines mark the *electronic* interface between the n-doped and the p-doped layers. It is important to note that this interface is *not* the atomically sharp metallurgical or chemical interface (marked by dashes), where the doping changes from C$_{As}$ to Si$_{Ga}$ or vice versa. Fig. 12 shows that the electronic interface (dark lines) exhibits rather a roughness much larger than
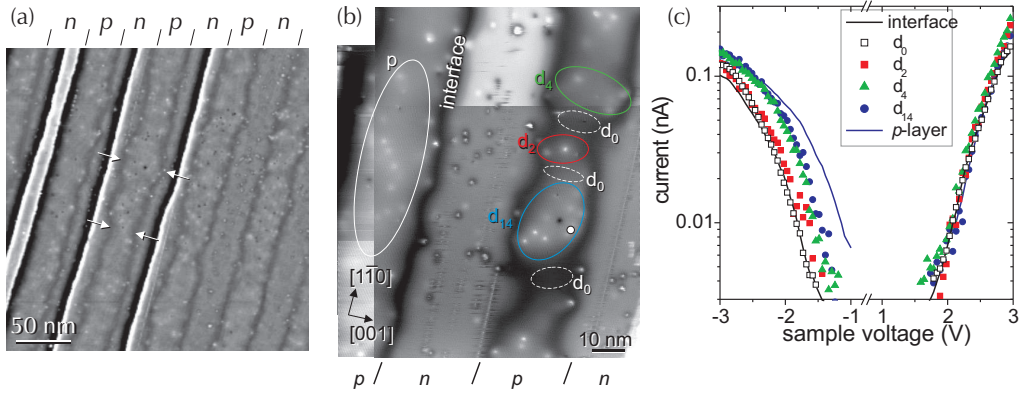
Fig. 12: Large scale (a) and atomically resolved (b) cross-sectional scanning tunneling microscopy images of multiple p- and n-doped GaAs layers. The bright hillocks marked by Si_{Ga} and C_{As} arise from individual dopant atoms. The dark lines between the p- and n-type layers are the signatures of the depletion zone at each $p-n$ interface and show the position of the electronic interface. Note its pronounced roughness and its correlation with the dopant atoms. The bright hillocks are signatures of dopant atoms. The encircled p-doped areas d_2, d_4, and d_{14} are dopant-induced dots confined by potential barriers due to the doping of the surrounding areas (n-type and lack of dopants (d_0)). (c) Current-voltage curves acquired in these encircled areas in (b). The spectra were normalized to a common tip-sample separation. The growth direction is [001]. Adapted with permission from [21], ©2002 American Physical Society and [32] ©2003 American Institute of Physics.

one atomic layer. The white arrows in Fig. 12a point out examples of a local electronically very narrow and wide p-type 'layer'. Note that the individual bright hillocks with about 3 to 5 nm diameter, visible in the p- as well as n-doped layers, are the signatures of negatively charged C_{As} and positively charged Si_{Ga} dopant atoms, respectively [28]. Their contrast is essentially given by the image of the screened Coulomb potential as outlined above [31]. Figure 12b shows that the depletion zone imaged as dark line circumvents each individual dopant atom. Undoubtedly each dopant atom near the metallurgical interface causes a short range meandering of the electronic interface on the scale of 2-5 nm.

Careful inspection of large scale STM images and a quantitative analysis of the interface roughness reveals a further contribution to the interface roughness on a longer length scale, leading to the electronically wide and narrow layers (see examples marked by white arrows in Fig. 12). The physical origin of this second roughness contribution is nicely illustrated in Fig. 12b, which shows that dopant atoms within nominally homogeneously doped layers exhibit large variations in the local concentration leading to clusters of dopant atoms as those encircled in Fig. 12b. Above and below are areas locally free of dopant atoms (marked d_0). This clustering not only induces the long range roughness of the electronic interfaces with correlation lengths of about 25 nm and amplitudes of approximately 2.5 nm, but it also drastically modifies the electronic properties.

In order to illustrate this effect the encircled areas, labeled d_2, d_4, and d_{14} according to the num-

ber of dopant atoms visible within the area, are of particular interest. These areas are bordered along the growth direction by *n*-doped layers and perpendicular to the growth direction by zones with no dopant atoms (dark contrast areas labeled d_0). These zones with no dopant atoms exhibit tunneling spectra (Fig. 12c) with typical characteristics of a depleted region (as discussed below). Thus the areas (d_2, d_4, and d_{14}) are semiconductor dots, whose confining potential for free holes is defined by the doping of the surrounding and thus by build-in potentials in the order of a few tenths of eV (border toward depleted zones) to 1.4 eV (toward n-doped layer).

Figure 12c shows local tunneling spectra measured above the different cluster areas. In order to allow a proper comparison of the spectra, they were normalized to a common tip-sample separation [34]. All spectra are essentially identical at positive sample voltages. In contrast, at negative sample voltages the current-voltage curves are shifted relative to each other. The fewer dopants are inside the cluster, the greater is the shift towards more negative voltages relative to the spectrum of the spatially extended *p*-doped layer (labeled p in Fig. 1).

For a quantitative discussion of the spectra, we recall that the spectra consist of the current from valence and into conduction band states at negative and positive voltages, respectively, with the band gap region in between [11, 16]. At a fixed voltage the tunneling current is determined by the energetic positions of the band edges underneath the tip:

At negative voltages, electrons tunnel from all filled valence band states lying between the valence band edge at the surface and the Fermi level of the tip. The size of this energy window is determined by the degree the tip bends the bands at the surface. This tip-induced band bending arises from the fact that the electric field between the tip and the surface penetrates into the semiconductor surface, due to the limited free charge carrier in a semiconductor. Therefore the shifts of the spectra at negative voltages (Figs. 2, 3) indicate that the tip pulls downward the position of the valence band edge at the surface of our dopant-induced dots the more, the less dopants are enclosed within the dot. This suggests a reduced screening ability of the dots with smaller numbers of dopant atoms. This situation is schematically shown in Fig. 13.

The electric field of the tip is screened by negatively charged acceptors, whose free holes have to be pushed away. On this basis, a reduction in the ability to screen the field of the tip is due to a combination of (i) the impossibility to deplete the dopant-induced dots from the free holes and (ii) the number of acceptors within the dot.

(i) The first effect depends on the size of the dot. If the dimension of the dot is much larger than the depletion width induced by the screening of the tip's field, i.e., the field can be screened entirely by the acceptors within the dot, then the band bending is determined by the concentration of dopant atoms in the dot (band edges drawn in black lines in Fig. 13). This effect is applicable to the spatially extended p-doped layer on the left hand side of Fig. 12a, which is electronically homogeneous over dimensions of more than 100 nm, which is much larger than the estimated depletion width of 10 to 25 nm for acceptor concentrations of $4 \times 10^{18} cm^{-3}$ at -2.5 V sample voltage.

If the dots' dimensions are similar to the width of the depletion zone, one also needs to consider whether the holes can actually be pushed away sufficiently to accommodate the screening of the tip's electric field. For the small dot sizes, this means that holes need to be pushed out of the enclosed dot areas to deplete them . We recall that our dots are confined by potential wells. These wells act as a barrier for the free holes and impedes the holes to be pushed out (see band edges drawn in red lines in Fig. 13), such that the remaining negatively charged
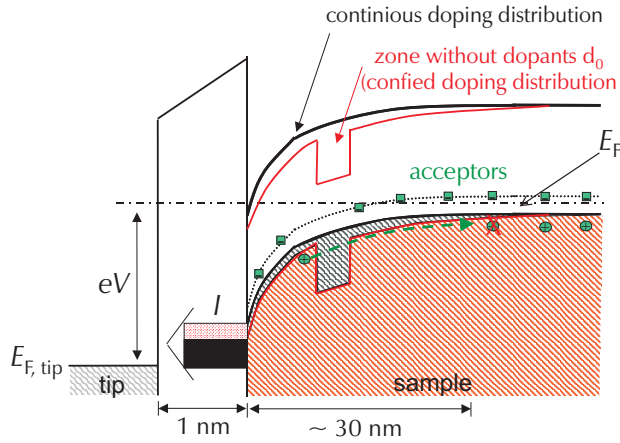
Fig. 13: Lateral band diagram showing the conduction and valence band edges in the presence of a tip with negative voltages applied to the sample. Two cases are shown. In black: model of continuous doping throughout the sample. In red: inhomogeneous doping with confining barriers arising from areas free of dopants. In that case the depletion zone is extended and therefore the band edge positions at the surface are lowered compared to the unconfined case. As a result the tunnel current is reduced. Adapted from [22].

acceptors could screen the electric field of the tip. If the dot cannot be sufficiently depleted, the band bending increases, causing lower tunneling currents than expected for infinitely sized bulk GaAs crystals with the same dopant concentration. The effect of confinement of the free holes on the tunneling current becomes smaller the higher the applied voltage, because the relative fraction of states inhibited to tunnel by the locally increased tip-induced band bending diminishes as more valence band states are involved in the tunneling process.

At positive sample voltage no dependence of the current on the number of enclosed dopant atoms is found, implying that the energetic position of the conduction band edge underneath the tip is the same for all areas investigated here. At positive sample voltage the electric field of the tip is screened by free holes accumulating at the p-doped surface. These holes feel no barrier to accumulate at the surface and thus can effectively screen the tip's field. Furthermore, the number of accumulated holes is determined by the density of valence band states, which is the same for all investigated areas, because the material is the same. This explains the almost invariant current-voltage spectra obtained on the various dots at positive sample voltages.

(ii) The above discussion also shows, that the ability to screen the electric field of the tip is directly proportional to the number of acceptors available within the electronically isolated dot. Thus the fewer acceptors within the dot, the larger the tip-induced band bending and the larger the voltage shift observed in the tunneling spectra.

These results show that local variations in the dopant atom distribution lead in nanoscale semiconductor structures to uncontrolled Fermi-energy positions in space and energy, effectively limiting the miniaturization of semiconductor devices.

5.4   Spin-polarized scanning tunneling microscope

5.4.1   Conservation of energy, momentum, and spin

The STM can be under suitable experimental conditions sensitive to the spin structure of a surface. A close look at Eq. 1 shows that during the tunneling process through the vacuum barrier, the tunneling electrons must conserve their energy. In fact inelastic tunneling processes, which are necessary to allow for energy changes, have only a very small contribution to the total tunnel current and can be neglected in a first approximation. Similarly the momentum [10] and the spin of the electrons must be conserved, too. [35–40]

The spin conservation is irrelevant for non-magnetic tungsten, platinum, or platinum-iridium tips usually used, because these materials exhibit an equal number of states with both spin polarizations. However, using magnetic tips significant spin-dependent tunneling currents can arise. Figure 14 shows the tunneling of electrons under consideration of spin conservation in a simplified schematic. Due to the different energy dependence of the density of states for the spin-up and spin-down electrons (e.g. intrinsic exchange splitting of (anti)ferro-magnetic materials), the tunnel current differs for the two spin polarizations, when tunneling into a magnetic tip, which is characterized by different density of states for the spin-up and spin-down polarizations. [35] Figure 14 shows that the degree of current change depends on the relative orientation of the spin structure between the tip and the sample (and on the voltage). As a result the tunneling of spin-up and spin-down electrons must be considered as independent processes. Hence, Eqs. 1 and 2 need to be modified and replaced with a sum of four different equations, one for each sample-tip spin combination:
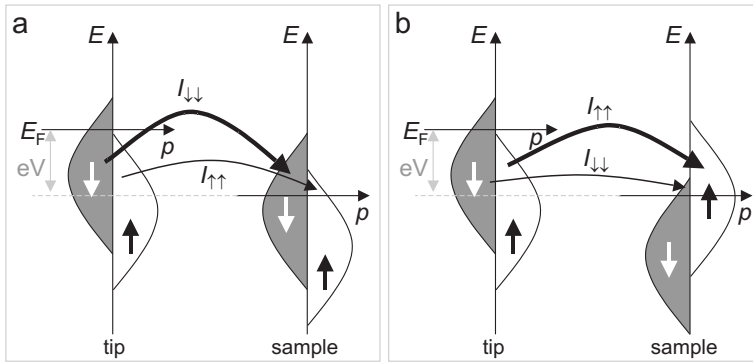
Fig. 14: Principle of spin-polarized tunneling between magnetized electrodes. Shown are the examples of two (a) parallel and (b) antiparallel magnetized electrodes referred to as sample and tip. In case of elastic tunneling the spin is conserved and thus the tunnel current for the spin-up and spin-down electrons differ in magnitude (as visualized by the width of the arrows).

$$\left. \frac{\mathrm{d}I_{\uparrow\uparrow}}{\mathrm{d}V} \right|_{V\approx 0} \propto \rho^{\uparrow}_{\text{sample}}(E_{\text{F}}) \cdot \rho^{\uparrow}_{\text{tip}}(E_{\text{F}}) \cdot |M^{\uparrow\uparrow}(E_{\text{F}})|^2 \qquad (10)$$

$$\frac{\mathrm{d}I_{\downarrow\downarrow}}{\mathrm{d}V}\bigg|_{V\approx0} \propto \rho^{\downarrow}_{\text{sample}}(E_{\text{F}}) \cdot \rho^{\downarrow}_{\text{tip}}(E_{\text{F}}) \cdot |M^{\downarrow\downarrow}(E_{\text{F}})|^2 \tag{11}$$

$$\frac{\mathrm{d}I_{\uparrow\downarrow}}{\mathrm{d}V}\bigg|_{V\approx0} \propto \rho^{\uparrow}_{\text{sample}}(E_{\text{F}}) \cdot \rho^{\downarrow}_{\text{tip}}(E_{\text{F}}) \cdot |M^{\uparrow\downarrow}(E_{\text{F}})|^2 \tag{12}$$

$$\frac{\mathrm{d}I_{\downarrow\uparrow}}{\mathrm{d}V}\bigg|_{V\approx0} \propto \rho^{\downarrow}_{\text{sample}}(E_{\text{F}}) \cdot \rho^{\uparrow}_{\text{tip}}(E_{\text{F}}) \cdot |M^{\downarrow\uparrow}(E_{\text{F}})|^2 \tag{13}$$

The notation is in analogy to Eq. 2 only that the derivative form is used here. Furthermore these equations are valid for small voltages applied to the sample. Small means here again that the voltage applied must be significantly smaller than the tunneling barrier. When higher voltages are needed one has to use the analogous form of Eq. 6. A detailed analysis of Eqs. 10, 11, 12, and 13 shows two features depending on the magnetic properties of the tip.

(i) For a non-magnetic metallic tip the density of states $\rho^{\uparrow}_{\text{tip}}(E_{\text{F}})$ and $\rho^{\downarrow}_{\text{tip}}(E_{\text{F}})$ and the corresponding matric elements will be equal for both spin polarizations. As a result the total current is comprised of Eqs. 10, 11, 12, and 13. This results again in a current given by Eq. 2 for the classical STM.

(ii) However, with magnetic tips, where the density of states for the up- and downward spin polarizations, $\rho^{\uparrow}_{\text{tip}}(E_{\text{F}})$ and $\rho^{\downarrow}_{\text{tip}}(E_{\text{F}})$, are not equal, significant differences in the tunnel current can exist for electrons in the sample with up- and downward spin orientation. [36] This is the basic principle on which the spin-sensitivity of the STM is founded.

5.4.2 Layered spin structure of Cr(001)

Figure 15 shows a measurement of step heights on the Cr(001) surface in the constant-current mode. [36] With a non-magnetic tip all monoatomic steps were found to have the same height of 0.149 ± 0.8 nm in good agreement with the bulk layer separation of 0.144 nm. However, when using a ferromagnetic CrO_2 tip two different step height occur in the constant current images, 0.12 and 0.16 nm (see lower right set of height profiles in Fig. 15a). This can be explained when recalling the spin structure in the bulk of Cr. Cr crystalizes in the cubic body centered crystal structure. The atoms on the outer corners of the unit cell all have one spin polarization, which is antiparallel with that of the atom in the center of the unit cell (see Fig. 15b). When cutting the crystal on a {100} surface, one exposes on each terrace only surface atoms with one spin polarization. However, on neighboring terraces the exposed atoms have opposite spin polarizations. [41] As a result depending of the spin polarization of the whole terraces the tunnel current is either bigger or smaller than in case of a non-magnetic tip. This is compensated by increasing (by Δs_1) or decreasing (by Δs_2) the tip-sample separation until a constant current is achieved, respectively (see Fig. 15c). As a result two different step heights labelled h_1 and h_2 are observed. This result convincingly proves that spin-sensitive tunneling is possible in the constant-current mode and thus SP-STM in general, [36] however thus far without atomic resolution.
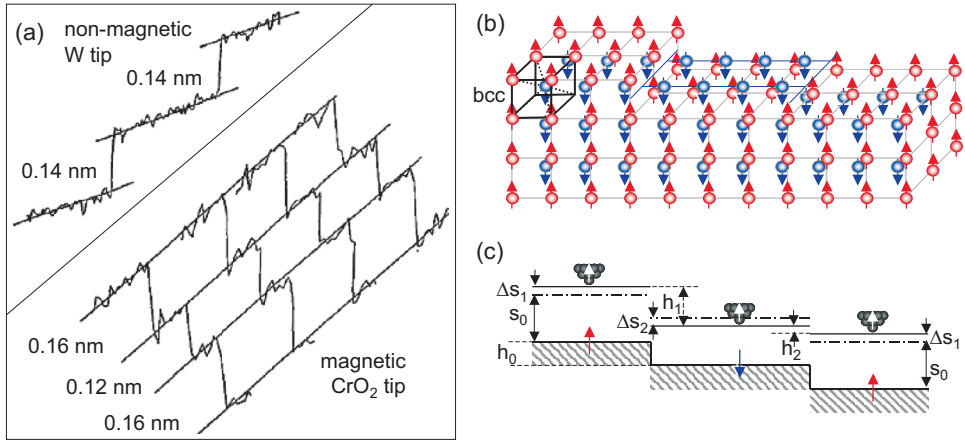
Fig. 15: (a) lower right corner: arbitrarily chosen (not successive) line scans over the same three monoatomic layer high steps obtained with a magnetic CrO_2 tip. Alternately 2 different step heights are observed. When the surface is imaged with a non-magnetic W tip only one step height is found (upper left corner). Adapted with permission from [36]. ©1990 American Physical Society. (b) Topological antiferromagnetic order of the Cr(001) surface with terraces separated by monoatomic layer high steps. Different terraces are magnetized in opposite direction [41]. (c) Schematic drawing of a ferromagnetic tip scanning over alternatively magnetized terraces separated by monoatomic layer high steps (step height h_0).

5.4.3 Atomically-resolved spin-polarized scanning tunneling microscopy

Thus far the spin polarization structures were much larger than the atomic lattice, i.e. whole terraces or large domain walls. The spin-polarized scanning tunneling microscope has, however, the capacity of atomic resolution and thus is able to image a surface magnetic structure with atomic resolution too. An almost perfect system to test the ultimate resolution are Mn monolayers on W(110) surfaces. [42] The monolayers exhibit an antiferromagnetic on the atomic scale, i.e. the Mn atoms have alternatingly a spin up and down polarization, such that the magnetic moment is canceled out on an atomic length scale.

Figure 16a shows such a Mn monolayer thick film partially covering a stepped W(110) surface. The film structure is schematically shown in the height profile below the constant-current STM image in Fig. 16a. When zooming in one one of these Mn overlayers with a non-magnetic W tip, one can image every Mn atom in the monolayer Mn lattice (Fig. 16b). For comparison an atomic ball model and a density functional calculation of the local density of states in included and labelled 'theory'. The good agreement shows that every Mn atom is imaged equally.

The atomically resolved image changes, however, fundamentally when using a magnetic Fe coated tip (Fig. 16c). Now, the apparent unit cell is changed. Only every second Mn atom is visible and the lattice constant along the [112] direction (line direction) is doubled (see line profiles in Fig. 16d). This doubling implies that now not every Mn atom is equal. In fact the result is consistent with the antiferromagnetic spin ordering on the atomic scale in the Mn film. Every second Mn atom along the [112] direction has the same spin polarization, while the atoms inbetween have the antiparallel spin orientation (see atomic ball model with spins in Fig. 16c).
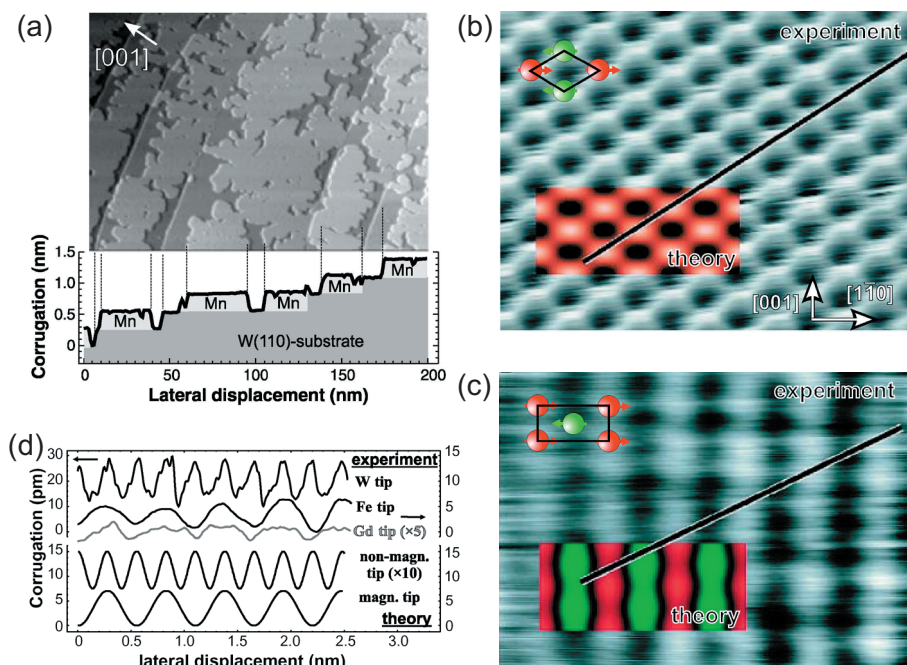
Fig. 16: (a) Topography of 0.75 ML Mn film grown on a stepped W(110) substrate. A line-section is shown at the bottom edge of the image. The structure of the sample is schematically represented by different gray levels. The image size is 200×200 nm^2. (b), (c) Comparison of experimental and theoretical high-resolution STM images of a Mn ML thick film on W(110) with (b) a nonmagnetic W tip and (c) a magnetic Fe tip. (d) Experimental and theoretical line sections for the images in (b) and (c). The unit cell of the calculated magnetic ground-state configuration is shown in (b) and (c) for comparison. Tunneling parameters for both images are $I = 40$ nA and $V = 3$ mV. The image size is 2.7×2.2 nm^2. Adapted with permission from [42]. ©2000 AAAS

This leads to a larger magnetic unit cell as shown in the theoretical calculation (inset in Fig. 16c) and in the experimental image. The doubling of the magnetic lattice constant as compared to the atomic one is also reflected in the theoretical line scans shown in Fig. 16d.

This example shows that spin-polarized scanning tunneling microscopy can also be used to probe the spin polarization of individual atoms. This provided in the case presented here a direct proof of the predicted two dimensional antiferromagnetic state of Mn monolayer thick films on W(110). The atomically resolved spin-polarized scanning tunneling microscopy will provide new insights into many other magnetic systems, whose atomic scale magnetic structure remains debated.

6   Summary

The application examples presented here, illustrate only a very small fraction of the vast application range of scanning tunneling microscopy and spectroscopy. Nevertheless, the examples show that the scanning tunneling microscopes developed within a rather short period of time to indispensable tools in surface sciences, physics, chemistry, biology, materials development, and even technological developments in industry. The wide applicability is primarily due to the simple principle of scanning a probe tip over a surface and obtaining an atomically resolved image based on a probe tip – sample interaction. surfaces can even routinely These advantages were to date not achievable using other techniques, which make the scanning probe techniques so unique.

References

[1] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Appl. Phys. Lett. **40**, 178 (1982), Phys. Rev. Lett. **49**, 57 (1982).

[2] Ph. Ebert, K. Urban, and M.G. Lagally, Phys. Rev. Lett. **72**, 840 (1994).

[3] G. Gamov, Z. Phys. **51**, 204 (1928).

[4] E. U. Condon and R. W. Gurney, Nature **122**, 439 (1928)

[5] L. Nordheim, Z. Phys. **46**, 833 (1927).

[6] P. Eckle, M. Smolarski, P. Schlup, J. Biegert, A. Staudte, M. Schöffler, H. G. Muller, R. Dörner, and U. Keller, Nature Physics **4**, 565 (2008).

[7] J. Bardeen, Phys Rev. Lett. **6,** 57 (1961).

[8] J. Tersoff and D.R. Hamann, Phys. Rev. **B 31**,805 (1985).

[9] J. Tersoff and D.R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983).

[10] N. D. Jäger, E. R. Weber, K. Urban, and Ph. Ebert, Phys. Rev. B **67**, 165327 (2003).

[11] J.A. Stroscio, R.M. Feenstra, D.M. Newns, and A.P. Fein, J.Vac. Sci. Technol. A **6**, 499 (1988).

[12] L. Ivanova, S. Borisova, H. Eisele, M. Dhne, A. Laubsch, and Ph. Ebert, Appl. Phys. Lett. **93**, 192110 (2008).

[13] Ph. Ebert, L. Ivanova, and H. Eisele, Phys. Rev. B **80**, 085316 (2009).

[14] S. Landrock, Y. Jiang, K.H. Wu, E.G. Wang, K. Urban, and Ph. Ebert, Appl. Phys. Lett. **95**, 072107 (2009).

[15] Ph. Ebert, B. Engels, P. Richard, K. Schroeder, S. Blügel, C. Domke, M. Heinrich, and K. Urban, Phys. Rev. Lett. **77**, 2997 (1996).

[16] R.M. Feenstra, J.A. Stroscio, J. Tersoff, and A.P. Fein, Phys. Rev. Lett. **58**, 1192 (1987).

[17] Ph. Ebert, G. Cox, U. Poppe, and K. Urban, Surfac. Sci. **271**, 587 (1992).

[18] A. Selloni, P. Carnevali, E. Tosatti, and D.C. Chen, Phys. Rev. B **31**, 2602 (1985), idem **34**, 7406 (1986).

[19] J.A. Stroscio, R.M. Feenstra, and A.P. Fein, Phys. Rev. Lett. **57**, 2579 (1986).

[20] C.J. Chen, J.Vac.Sci. Technol. A **6**, 319 (1988).

[21] N. D. Jäger, K. Urban, E. R. Weber, and Ph. Ebert, Phys. Rev. B **65**, 235302 (2002).

[22] N. D. Jäger, *Effect of individual dopant atoms on the electronic properties of GaAs investigated by scanning tunneling microscopy and spectroscopy*, Thesis, RWTH-Aachen (2003).

[23] K. Besocke, Surf. Sci. **181**, 145 (1987).

[24] Y. Kuk and P.J. Silverman, Rev. Sci. Instrum. **60**, 165 (1989).

[25] H.-J. Güntherodt and R. Wiesendanger, *Scanning Tunneling Microscopy*, Vol. 1 and R. Wiesendanger and H.-J. Güntherodt, *Scanning Tunneling Microscopy*, Vols. **2** and **3**, Springer Series in Surface Science Vols. **20**, **28**, and **29**, Ed. R. Gomer, Springer, Berlin, 1992, 1993.

[26] G. Cox, *Untersuchung von Grenzflächen und Gitterbaufehlern in GaAs mit Hilfe der Rastertunnelmikroskopie*, Thesis, RWTH Aachen, published as Forschungszentrum Jülich GmbH Bericht 2382 (1990).

[27] S. Chiang (Ed.), Special Issue of Chemical Reviews **97** (4), June 1997.

[28] Ph. Ebert, Surf. Sci. Rep. **33**, 121 (1999); Current Opinion in Solid State and Materials Science **5**, 211 (2001).

[29] Ph. Ebert, M. Heinrich, M. Simon, C. Domke, K. Urban, C.K. Shih, M.B. Webb, and M.G. Lagally, Phys. Rev. B **53**, 4580 (1996).

[30] Ph. Ebert, M. Heinrich, M. Simon, K. Urban, and M.G. Lagally, Phys. Rev. B **51**, 9696 (1995).

[31] R. B. Dingle, Phil. Mag. **46**, 831 (1955).

[32] N. D. Jäger, K. Urban, E. R. Weber, and Ph. Ebert, Appl. Phys. Lett. 82, 2700 (2003).

[33] R. M. Feenstra *et al.,* in: *Semiconductor Interfaces at the Sub-Nanometer Scale*, Eds. H. W. M. Salemink and M. D. Pashley (Kluwer Academic, Dordrecht, 1993), p. 127; R. M. Feenstra *et al.,* Appl. Phys. Lett. **61**, 795 (1992).

[34] N. D. Jäger, M. Marso, M. Salmeron, E. R. Weber, K. Urban, and Ph. Ebert, Phys. Rev. B **67**, 165307 (2003).

[35] R. Wiesendanger, Rev. Mod. Phys. **81**, 1495 (2009).

[36] R. Wiesendanger, H.-J. Güntherodt, G. Güntherodt, R. J. Gambino, and R. Ruf, Phys. Rev. Lett. **65**, 247 (1990).

[37] M. Bode, Rep. Prog. Phys. **66**, 523 (2003).

[38] S. F. Alverado and P. Renaud, Phys. Rev. Lett. **68**, 1387 (1992).

[39] D. T. Pierce, Phys. Scripta **38**, 291 (1988).

[40] D. Wortmann, S. Heinze, Ph. Kurz, G. Bihlmeyer, and S. Blügel, Phys. Rev. Lett. **86**, 4132 (2001).

[41] S. Blügel, D: Pescia, and P. H. Dederichs, Phys. Rev. B **39**, 1392 (1989).

[42] S. Heinze, M. Bode, A. Kubetzka, O. Pietzsch, X. Nie, S. Blügel, and R. Wiesendanger, Science **288**, 1805 (2000).

F 4   Scanning Electron Microscopy

B. Förster

Complementary Techniques

Ernst-Ruska Centre

Forschungszentrum Jülich GmbH

Contents

Introduction

The average human eye is able to see objects of about 0.1 mm in size. For smaller objects, an optical or electron microscope can be used. Scanning Electron Microscopes (SEM) are particularly used for observing the fine structure of the sample surface in a range of low to high magnifications. They show a greater depth of focus than optical microscopes and can be combined with an X-ray detector to conduct compositional analysis of the examined area.

1   SEM: Instrumentation and Principles

The major components of a Scanning Electron Microscope (SEM) are shown in figure1. Electrons are generated in the electron gun and accelerated up to 30 keV. Electromagnetic lenses are used to focus the beam. The sample is scanned line by line and leads to elastic and inelastic scattering or absorption of electrons, as well as variations of electromagnetic radiation. Subsequently the various secondary signals (including secondary electrons (SE), backscattered electrons (BSE), Auger electrons, cathodoluminescence signals or X-rays) will be detected with appropriate detectors (see left side of figure 5). The intensity of the secondary signal depends on the surface morphology, chemistry, physical states, etc. of the specimen. The amplified detected signal is displayed as an image on the monitor. To investigate samples in the SEM, they usually should be conductive, dry and stable in vacuum.
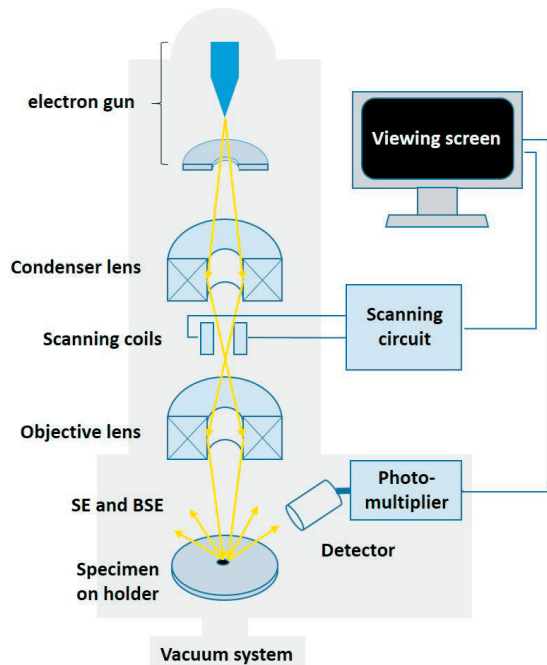
Figure 1: Schematic of a Scanning Electron Microscope

1.1   Electron Guns

The purpose of the electron gun is to provide a stable beam of electrons of adjustable energy. It generates electrons and accelerates them to an energy in the range 0.1–30 keV. There are two main types of electron guns: thermal and field emission guns. In a thermal emission (thermionic) filament a tungsten filament or a LaB_6 crystal is heated by a filament current.

This results in the emission of thermal electrons. A field emission gun consists of a sharp metal tip (usually Tungsten) with a radius of less than 100 nm. When a negative potential is applied to the cathode, the electric field is concentrated at the tip, which facilitates electron emission (emission current). [3]

A constant beam current is required to create a good quality image because all image information is recorded as a function of time. Since the micrograph is acquired over a period of time of minutes (for high quality images), any changes in the filament emission will affect the image intensity at that point in the scan. This will produce a poor quality image because the brightness will vary across the image.

1.2   Electromagnetic Lenses

A series of electromagnetic lenses and apertures are used to reduce the diameter of the electron beam and to place a small, focused spot of electrons onto the specimen (figure 2). A magnetic lens consists of an iron case enclosing a coil of wire, which generates a magnetic field across the lens gap between the pole pieces [3]. These produce a focal length, which can be changed by varying the current through the coil. The magnetic field bends electron paths in a similar way that solid glass lenses bend light rays.

In a SEM, there are usually two lens sets. The condenser lens is at the top and the objective lens at the bottom of the column. The condenser lens converges the cone of the electron beam to a spot below it, before the cone diverges out again and is converged back again by the objective lens and down onto the sample (compare figure1). The objective lens also has some influence on the diameter of the spot size of the electron beam on the specimen surface; however, its main role is in focusing the beam onto the sample.
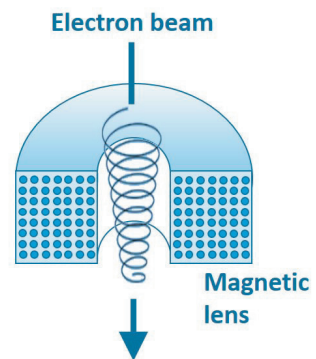
Figure 2: Electromagnetic lens

Contrary to Transmission Electron Microscopy (TEM) or optical microscopy, the magnification of the image in a SEM is done by the scanning system: an area of the sample is scanned with a given number of pixels. To magnify a detail of this area, the smaller cut out is scanned with the same number of pixels as shown in figure 3.
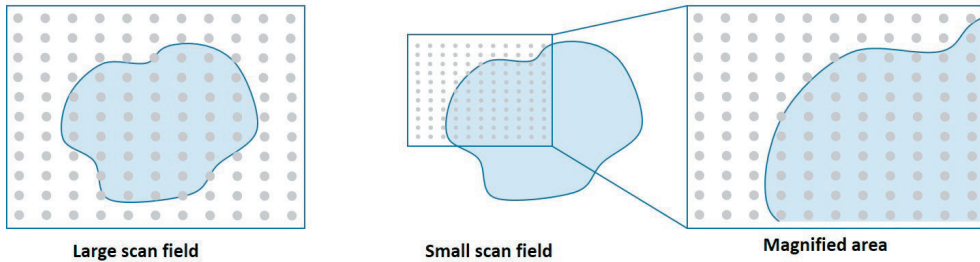
Figure 3: Magnification in a SEM

1.3 Lens Aberrations and Resolution

Magnetic lenses suffer from a number of defects or aberrations in their performance. However, in electron optics, by contrast to the situation in light optics, the effects of aberrations cannot be cancelled by using combinations of lenses. The only recourse therefore is to try to minimize these effects.

Spherical Aberration
Spherical aberration arises because electrons further away from the optical axis are refracted more strongly towards the optical axis by the magnetic field than those electrons, which are travelling close to it [3]. This can be minimised by using a smaller objective lens aperture. Unfortunately using a smaller aperture will reduce the probe current and introduces aperture diffraction. Nevertheless, by using a smaller working distance (WD) the effect of spherical aberration can be improved.

Aperture Diffraction
For very small apertures, the wave nature of electrons gives rise to a circular diffraction pattern instead of a point at the image plane. Electrons diffract at the edge of the small aperture and appear in the image plane as a broad intensity distribution ("Airy disk") surrounded by smaller subsidiary maxima [3].

Chromatic Aberration
As the electron beam has a certain energy spread, electrons within the beam will have slightly different energies. Thus, they will experience different forces at the same location within the field of the magnetic lens. Electrons with higher energies will be bent more strongly than electrons with lower energies. This results in an area of minimum confusion rather than in a sharp focal point. Chromatic aberration cannot be corrected by the operator and is more visible when using low accelerating voltages [5].

Astigmatism
Because of lens defects (machining errors, asymmetry in lens winding) and contamination on

aperture or column, the cross section of the electron beam can vary in shape. Usually an elliptical cross section is formed instead of a circular one, which results in an elliptical distorted image [4]. The astigmatism can be corrected by applying stigmator coils around the objective lens.

Limit of resolution

The resolution of an optical system is defined as the smallest distance between two points that still can be determined as two separate entities. As a rough guideline, to resolve two objects clearly, their distance should be in the same range as the used wavelength. The theoretical resolution in the SEM is much higher than the effectively achievable resolution. In the SEM the resolution mainly is limited by the following three factors:

- Lens aberrations
- Size of the electron spot (spot size)
- Size of interaction volume (spatial resolution)

Furthermore, contrast and signal-to-noise ratio are also important factors for the resolution.

The spot size mainly is defined by the electron optical system and the electron gun. Thereby, field emission guns can achieve higher resolutions, compared to tungsten or LaB_6 cathodes.

1.4   Depth of Field

The part of the image, appearing acceptably in focus, is called "depth of field". Compared to light microscopes, the depth of field in SEMs is much bigger. An electron beam with a small convergence angle provides a large depth of field, a beam with a large convergence angle a small depth of field. It can be influenced by changing either the size of the aperture or the working distance:

small aperture – large depth of field
large aperture - low depth of field
long WD – large depth of field
short WD – low depth of field

A large depth of field is important, if a specimen with a large topological variation has to be observed [4] as shown in the example in figure 4.
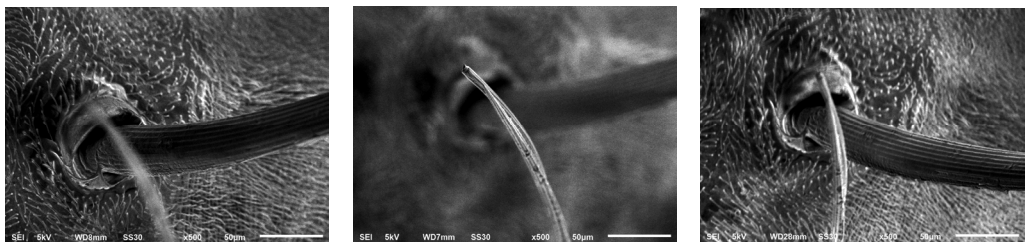
Figure 4: Depth of field, left: background is in focus, middle: foreground is in focus, right:for- and background are in focus because of a larger working distance providing a larger depth of field

1.5   Electron Matter Interaction

The interaction of primary electrons with the specimen can be divided into two general classes: elastic and inelastic scattering. The different signals that occur when the electron beam hits the sample are summarized in figure 5. The size of the interaction volume is determined by the penetration depth of the electrons according to the energy of the primary electrons, the angle of incidence and the observed material. In general, secondary and backscattered electrons are used for image formation. To perform element analysis most SEMs are also equipped with an X-ray detector.
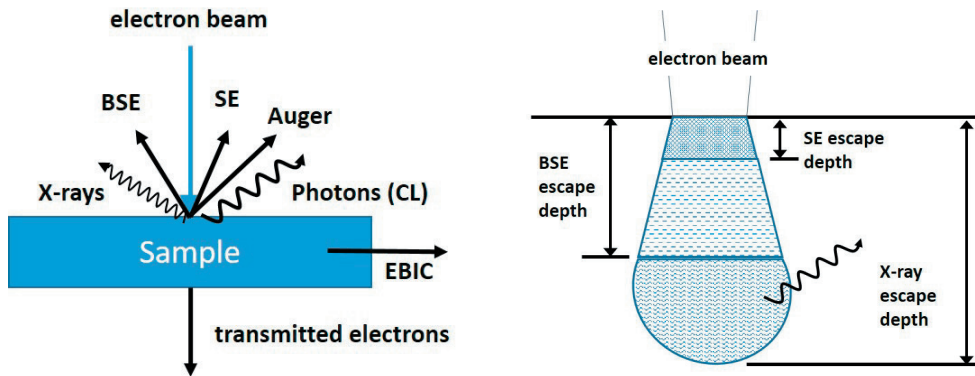
Figure 5: Electron sample interaction in the SEM
left: generated signals, right: sample interaction volume

1.6   Secondary Electrons

Secondary electrons (SE) are generated by inelastic scattering of incident electrons (electron beam) with the outer atomic shell of the specimen atoms. By definition, secondary electrons have kinetic energies up to 50 eV, but the most probable energy ranges from 2 to 5 eV. Due to their low energy, they only can escape from the specimen, if they are very close to the surface (few nm). Thus, images from SE electrons show very good surface details (topological contrast) and are used for high-resolution images [1].

1.7   Backscattered Electrons

Elastic scattering of the incident beam electrons within the specimen generates backscattered electrons (BSE). As their name tells, the BSEs are scattered backwards and are emitted out of the specimen. Since BSEs have higher energies than SEs, information from a relatively deep region of the sample is contained in the signal of the backscattered electrons. The BSE image also contains information about the specimen composition, because the number and the

directions of the BSEs are determined by the average atomic number of the specimen. In practice, the number of BSEs increases with increasing atomic number Z. Thus BSE images show contrast due to the variation of the chemical contrast of the specimen, whereas SE images mainly reflect the surface topography [1,2].

1.8 X-Ray Radiation

When the electron beam hits the specimen, characteristic X-rays are emitted.
The primary electron can be scattered inelastically by an inner shell electron causing the electron to undergo a transition to an outer shell and leaving the atom with a vacancy (hole) in its inner shell. To release the energy from this excited atom, an electron from the outer shell fills the vacancy by emitting a photon. The energy of the photon is the difference between the energy of the inner and outer shell electrons. These photons are called characteristic X-rays, because their energies are characteristic of individual elements and thus can be used for elemental analysis.
Despite the characteristic X-rays also a continuous X-ray spectrum occurs because of the deceleration of the electrons within the electric field in the region of the atomic nucleus.
Instead of emitting X-rays, the energy can also be transferred to another electron, which is ejected from the atom. This is called an Auger electron. Auger electrons are usually emitted by elements with a low atomic number, whereas characteristic X-ray radiation is dominated by elements with high atomic numbers.

2   Detectors

In a SEM different kind of detectors are used. Usually there is an Everhart-Thornley detector for SE, a semiconductor detector for BSE and an EDX detector for X-ray microanalysis. Additionally there are detectors available for high-resolution measurements or for other special applications, for example a STEM detector for transmitted electrons, an In-lens or in-column detector for high resolution images or low voltage BSE detection or a WDX detector for microanalysis.

2.1 SE Detectors

The most common electron detector used in the SEM is the Everhart-Thornley detector (figure 6). Mainly it is used to detect SEs but it also detects BSEs. The detector usually is mounted laterally on one side of the specimen chamber. The electrons are attracted towards the detector by a positive bias, which is applied to a collector at the
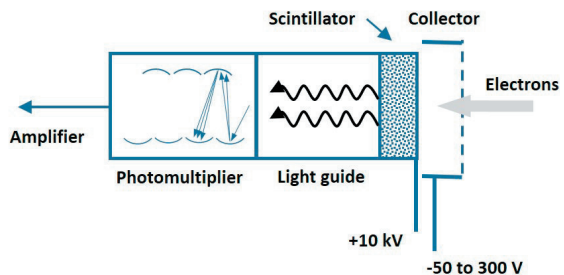
Figure 6: Everhart-Thornley detector

front end of the detector. After passing the collector, the electrons are accelerated towards the scintillator by applying a positive voltage of about 10 kV. This results in the production of photons, which travel through a light guide to a photomultiplier. There the light is converted to electrons, which are amplified as an electric signal. By applying a negative bias to the collector, it is also possible to detect BSEs only.

For high-resolution imaging even at low kV conditions often an Inlens or Incolumn SE-detector is used (figure 7). It is located inside the microscope column above the objective lens and arranged rotationally symmetric around the optical axis. The position of the detector enables the efficient detection of SEs, particularly at small working distances.

2.2 BSE Detectors

Usually the BSE detector is mounted directly below the final lens to offer a large solid angle to detect back scattered electrons (figure 7). The BSE detector is made of semiconductor material. When BSEs impinge on it, electron/hole pairs are formed in the depletion layer of the semiconductor. A reversed bias at the depletion layer leads to a charge carrier transport and thus forms the signal. A metal coating on top of the detector absorbs impinging secondary electrons. BSE detectors often consist of multiple rings or segments. By adding or subtracting the signal of the different segments, different kind of contrasts may be displayed in the image (material or topological contrast).

Similar to the in-lens or in-column SE detectors also BSE detectors can be located inside of the microscope column. These detectors are mostly used for low kV applications.
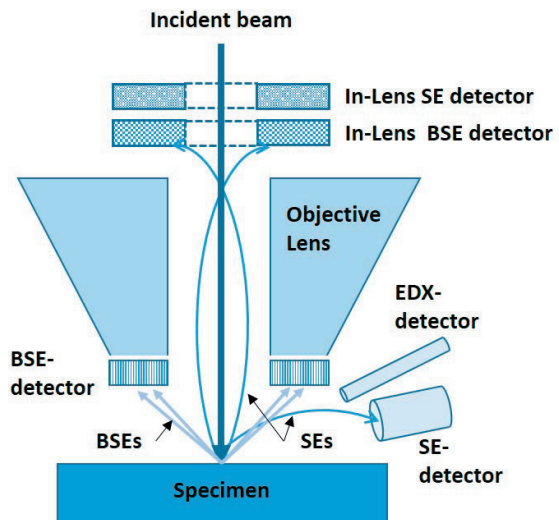
Figure 7: Arrangement of detectors in the SEM chamber

2.3 X-Ray Detector

Nowadays mostly semiconductor detectors are used to measure X-rays. The silicon drift detector (SDD) mainly consists of an undoped silicon wafer. The SDD detector is built up of two opposing p-doped layers. Only a small part on the backside is n-doped and serves as collector electrode. A field effect transistor (FET) is located in the middle of the wafer, which acts as amplifier. The main characteristic of a SDD detector is its transversal field generated

by a series of ring electrodes. By applying a voltage, a depletion zone with a potential minimum between the layers in the middle of the wafer is formed. Through the electrodes a field gradient is generated, which guides the electrons to the anode and the holes to the drift rings or the contact on the backside. Generally, X-ray detectors measure the energy of an incident photon by the amount of ionization processes it produces in the detector material. As the energy of the incident photon is proportional to the amount of generated charge carriers (height of the pulse), conclusions about the composition of the material can be drawn (elemental analysis).
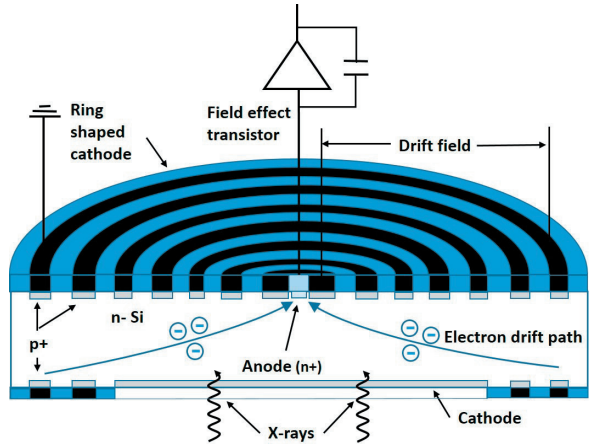
Figure 8: SDD detector

3   Contrast mechanisms

There are different contrast mechanisms in a SEM. Depending of the origin of the scattered electrons and the acceleration voltage different images will be obtained using different kind of detectors.

3.1   Topological Contrast

 As mentioned in chapter 1.6, SEs are very sensitive to the surface topology. Using the Everhart-Thornley detector leads to typical micrographs often showing a 3-dimensional impression. Areas with electrons emitted from a surface that faces away from the detector or which is blocked by the topography of the specimen, will appear darker than surfaces that face towards the detector. This topographical contrast due to the position of the SE detector is a major factor in the "life-like" appearances of SE images (figure 9).
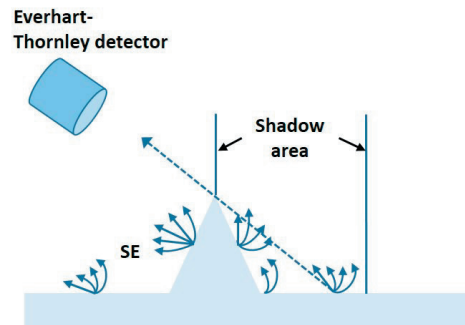
Figure 9: Contrast effect due to the position of the detector

Another contrast effect of SEs originates from the dependence of the secondary electron yield on the angle between the primary beam and the surface normal to the specimen. Tilted surfaces yield more SE than surfaces normal to the beam and therefore appear brighter (figure 10). Thus, SEs provide particularly good edge detail. Edges (and often-pointy parts) look brighter than the rest of the image because more SE can leave the specimen surface [5].
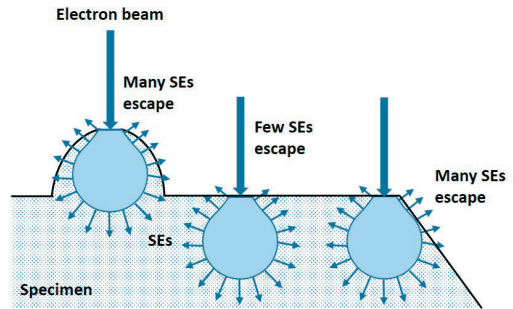
Figure 10: Occurrence of edge effect in fine structure specimen

Influence of the acceleration voltage on the contrast

Additionally to the topological contrast, also the chosen accelerating voltage has effects on the image. In general, increasing the acceleration voltage decreases the spherical aberration and therefore increases the resolution. However, an increasing acceleration voltage increases the interaction volume of the sample, with the consequence, that fine structures are no longer visible. An example is shown in figure 11. The micrographs display core shell nanoparticles with a soft polymer shell and a metal core. On the left side of the figure, a low accelerating voltage of 2 kV was chosen, showing the whole nanoparticle. If the accelerating voltage is doubled to 4 kV the interaction volume becomes larger and so the metal core becomes visible. Therefore, for very small surface details, a low acceleration voltage has to be chosen. On the other hand, by applying a higher acceleration voltage objects underneath the surface layer can be revealed.
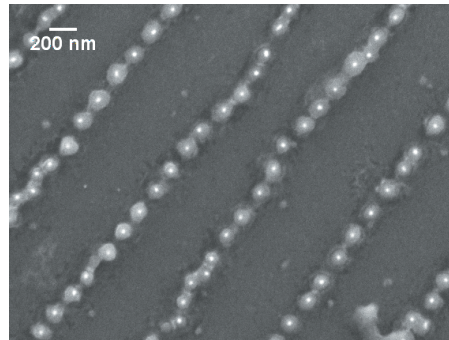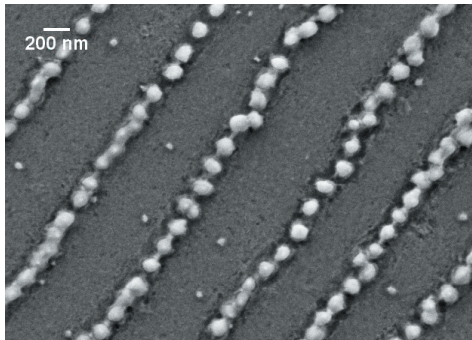
Figure 11: Influence of the accelerating voltage on the image, left: core shell particles investigated with an accelerating voltage of 2 kV. Only the core is visible. Right: same sample investigated with 4 kV, now the core particles are also visible

3.2   Material contrast / Z contrast

The material contrast arises from differences in the local material composition within the specimen. The amount of back scattered electrons increases with the atomic number Z. Thus, especially in BSE images regions of high atomic number will appear brighter relative to those with low atomic number [5].
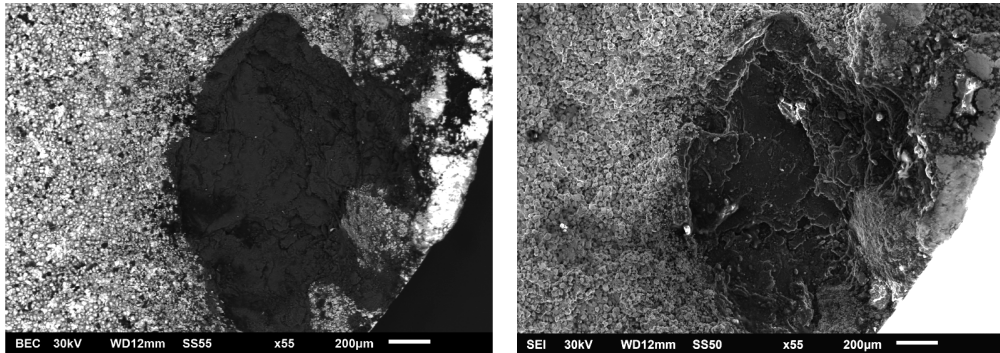
Figure 12: images of a broken screw, left: BSE image with material contrast: materials with higher atomic number appear brighter, right: SE image with topological contrast and edge effect

References

[1] R.F. Egerton, *Physical Principles of Electron Microscopy: An Introduction to TEM, SEM and AFM* (Cham, Springer, 2016)

[2] SEM A to Z, JEOL Ltd

[3] Goldstein J.I. et al. (2003) The SEM and Its Modes of Operation. In: Scanning Electron Microscopy and X-ray Microanalysis. Springer, Boston, MA

[4] W. Zhou, Z.L. Wang, *Scanning Microscopy for Nanotechnology* (Springer 2006)

[5] L. Reimer, *Scanning Electron Microscopy*, (Springer, 1998)

F 5   Cryo-Transmission Electron Microscopy

M.S. Appavou

Jülich Centre for Neutron Science

Forschungszentrum Jülich GmbH

Contents

Introduction

Electron microscopy allows structural investigation in a complementary [as well as competitive] manner to crystallography, small angle neutron scattering and nuclear magnetic resonance.

This technique is based on the wave-like property of the electron and on their interactions with the electrostatic potential distribution of the atoms in the sample.

The transmission electron microscopy principle is similar to the one of light microscopy except that instead of photons, one uses electrons and instead of glass lenses, electromagnetic lenses are used.

Although its use needs some technical requests like maximum thickness, essentially for soft matter material, it is a powerful tool to investigate structures.

1   Historical aspect

Historically, the electron microscopy techniques get its origin in the discovery of De Broglie who explained the wave-like property of the electron in 1925 [1]. The term "Electron microscopy" appears for the first time in the 1931 Ernst Ruzka paper [2], one year after he built the very first working electron microscope. This instrument contained the first version of electromagnetic lenses for electrons described by H. Busch in 1926 [3]. Ruzka received the Nobel Prize for physics in 1929. From then, the technique was subject of evolution and improvement in terms of instrumentation with Field Emission Gun in 1954 by V.E. Cosslet and M.E. Haine [4] and with Crewe in 1968 [5] as well as in terms of specimen preparation with the use of negative staining during the 1960′s [6]. In 1984, a consequent step in improvement of the technique was the use of Cryo-TEM by Dubochet and co-workers [7]. They developed methods to rapidly freeze the aqueous based specimen in a quasi-native hydrated state avoiding beam damage effect in a larger extent than at room temperature. Jacques Dubochet, Richard Henderson and Joachim Frank received the Nobel Prize for this work in 2017.

2   Instrument Description

2.1 Electron properties

Electrons are a type of subatomic particle with a negative charge and a mass of $9.10^9 \times 10^{-31}$ kg. As they behave both as wave and particle, this dual-property allows the application of the De Broglie relationship [1] : h=λp where λ is the wavelength of the electron and p the momentum with p=mv.

As electrons are negatively charged particles, they can be influenced by electromagnetic field according to :

$$q\Phi = (1/2)*(p^2/m)$$

where q and m are respectively the charge and the mass of the particle, p the equilibrium momentum.

2.2   Principle of Transmission Electron Microscopy

As the name suggests, the electron beam is transmitted through the sample. It utilises electrons as source of illumination which gives much improved resolution over a light microscope (around a thousand-fold better: c.a 0.2nm compared with 0.2mm) this is mainly because the effective wavelengths of accelerated electrons are extremely shorter than those of light. The effective wavelengths in the TEM is $\lambda = (1.5/V)^{1/2}$ nm where V = the accelerating voltage of the electron beam (Table1).

Voltage	λ	N.A.	Resolution
25,000	0.0077nm	0.01	0.47nm
50,000	0.0055nm	0.01	0.33nm
75,000	0.0045nm	0.01	0.27nm
100,000	0.0039nm	0.01	0.24nm
200,000	0.0027nm	0.01	0.17nm
1,000,000	0.0012nm	0.01	0.07nm
3,000,000	0.0007nm	0.01	0.04nm

Table 1: Effective wavelenghts and resolution in TEM as a function of acceleration voltage

At the acceleration voltages used in TEM, relativistic effects have to be taken into account (Table 2):

$$\lambda = h / [2m_0eV (1 + eV/2m_0c^2)]^{1/2}$$

V_{acc} / kV	Non relativistic wavelength / pm	Relativistic wavelength / pm	Mass x m_0	Velocity x 10^8 m/s
100	3.86	3.70	1.20	1.64
200	2.73	2.51	1.39	2.09
300	2.23	1.97	1.59	2.33
400	1.93	1.64	1.78	2.48
1000	1.22	0.87	2.96	2.82

Table 2: Relativistic wavelenghts, electron mass and velocity in TEM as a function of acceleration voltage.
Rest mass of an electron: $m_0 = 9.109$ x 10^{-31} kg
Speed of light in vacuum: $c = 2.998$ x 10^8 m/s

The resolution (nm) is equal to 0.61 X λ /N.A. where N.A. = the numerical aperture of the objective lens. The resolution improves and the effective wavelength decreases with increased accelerating voltage.

As express before, the TEM principle is similar to the one of light microscopy: Electron microscopy cover a space scale from micrometer to nanometer and is complementary with other techniques such as light microscopy.

2.2.1 Electron beam generation

Two main methods are used in electron microscope to generate electron beams: thermo-ionic and field emission gun. In the case of thermos-ionic gun, which include Tungsten or LaB6 as material for the filament, an electric current is applied to this cathode and electrons are then extracted from the crystal. A tandem form by a negatively charged Wehnelt cylinder surrounding the loop and a positively charged anode conduct and accelerate the extracted electrons through the TEM column toward the specimen.

In Field Emission gun (FEG) which is more and more used in the nowadays EM, high-density electrons are emitted when a strong electric field is applied to a cathode (electron emission element) with a sharpened tip made of material such as zirconium oxide (which emits electrons by Schottky effect)

2.2.2 Electromagnetic lenses and apertures

There are three main sets of electromagnet to conduct the electrons inside the TEM column through the specimen until the phosphoric image detector plate: the condenser lenses which have for purpose to focus the electron beam, the objective lenses after the specimen to interplay with the phase of the produced image and thus used as phase contrast enhancer and the projector lenses which are used for the magnification. Intermediate and projector lenses magnify and project the focused image onto the fluorescent screen (converts electrons to photons) at the base of the column or to a CCD (charge-coupled device) camera beneath. The condenser and objective lenses are in tandem with specific aperture to limit the amount of electrons striking the specimen (protecting it from excessive irradiation) and limit the number of x-rays generated from electrons hitting parts of the microscope column. The larger the aperture angle, the greater the maximum illumination intensity, but in general the poorer the image quality.

The electromagnetic lens defects are similar to those of optical lenses and these affect from achievement of the maximum theoretical resolution.

2.2.2.1 Spherical aberration

Actually, as electron rays passing at large angles through the lens, i.e. at the periphery are refracted more than those passing along the lens axis and therefore do not have the same focal point, there is the so-called spherical aberration, thus the resolution limit is lowered. In terms of geometrical optics, it is given d the diameter of the disc of minimum confusion $d = (0.5 \times M \times Cs \times \alpha^3)$ where M is the magnification, Cs is the spherical aberration coefficient and α is the angular aperture. Apertures are also used in the TEM to limit the peripheral electrons and minimise this aberration.

2.2.2.2 Chromatic aberration

In this case, electrons of different energies converge at different focal points and this is essentially equivalent to chromatic aberration in light microscopy. This can be minimised by: increasing the accelerating voltage, an improved vacuum and the use of the thinnest possible specimen.

2.2.2.3 Astigmatism

This phenomenon occurs when the field within the electromagnetic lens is not perfectly symmetrical. This can be due to imperfect boring of the lens pole pieces or contamination of the column, specimen or apertures. TEM instruments have astigmatism controls called stigmator to correct for this.

2.2.3 Interaction between the electron and the matters at the specimen

As the focused electrons interact with the specimen, several events occur at this level:
Secondary electrons: Electrons scattered from sample atoms by interactions with beam electrons. Those electrons are used for topographical study via SEM.
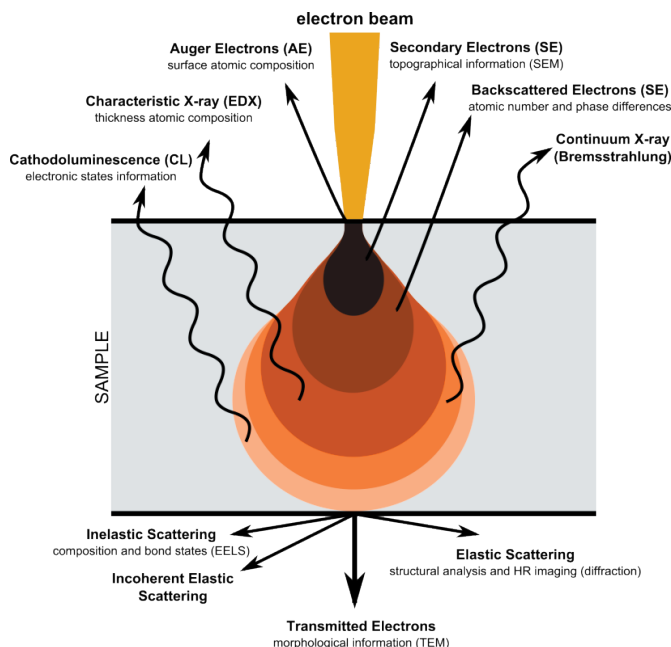
Fig.1: Electron Interaction with Matter. Claudionico~commonswiki [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)], from Wikimedia Commons.

Backscattered electrons: Primary (beam) electrons that have been deflected by the specimen through an angle generally greater than 90° so that they exit the sample with little or no loss of energy. Those BSEs are used to detect contrast between areas with different chemical compositions via SEM.

Bremsstrahlung (continuum x-rays) : Decceleration of electrons in the Coulomb field of the nucleus. This lead to the emission of x-ray carrying the surplus energy ΔE and to uncharacteristic x-rays.

Auger electrons: Surplus energy is transferred to another electron, which is emitted.

EDX or Energy dispersive X-ray analysis or spectrometry (sometimes EDS). An EDX spectrometer makes a spectrum of X-rays emitted by the specimen on the basis of their energy.This is used for elemental analysis and chemical characterization [8].

Cathodoluminescence: The emission of light photons by a material under electron bombardment. CL detectors either collect all light emitted by the specimen or can analyse the wavelengths emitted by the specimen and display an emission spectrum or an image of the distribution of cathodoluminescence emitted by the specimen in real color.

An electron of the primary beam is scattered by the electrostatic interaction with the positively charged nucleus of an atom at an angle of more than 90°, yielding backscattered electrons. This type of electrons has practically the same energy as the ones of the primary beam.

An electron of the primary beam is scattered by the electrostatic interaction with the positively charged nucleus of an atom at an angle of less than 90°, yielding elastically scatterd electrons. Also these electrons do not loose energy and therefore are referred to as elastically scattered electrons.

Elastic coherent electrons are scattered with an angle below 10° whereas elastic incoherent electrons are scattered with an angle above 10°.

Electrons can also loose energy while interacting with the "electron cloud" of the atom. These are called inelastically scattered electrons.

This interaction can lead to the following processes :

- Inner-shell ionisation : An electron is pushed out of the electron cloud. The electron „hole" is filled by an electron of an outer shell: Surplus energy is either emitted as characteristic x-ray or transferred to another electron, which is emitted (Auger electron).

- Phonons are lattice vibrations (heat) and induce beam damage

- Plasmons originates from oscillations of loosely bound electrons in metals

Therefore, elastic scattering brings the information for TEM image and diffraction pattern whereas inelastic scattering are exploited for electron energy loss spectroscopy (EELS) to determine the atomic structure and chemical properties of a specimen, i.e. elemental mapping, this includs the type and quantity of atoms present, chemical state of atoms and the collective interactions of atoms with their neighbors [9].

2.2.4 Image formation

The image formation in TEM is basically the same as for regular light microscopy, since the properties of electrons are very similar to the ones of light. As for light, electrons can be observed as particles or as waves depending on the physical phenomenon to be explained (wave particle duality).

The primary electron beam passes through the thin specimen and generates non-diffracted and diffracted electromagnetic waves. The objective lens collates the diffracted and non-diffracted waves in the image plane and forms the primary image.

Focusing

Contrast of biological specimens can be further improved by slight underfocusing.
Underfocusing leads to Fresnel rings (deletion/amplification of signal), which improve the visibility of biological structures (Figure 2). Note: Too much underfocus or overfocus introduces artifacts and the image gets unsharp. True focus (Gaussian focus) corresponds to minimum contrast.
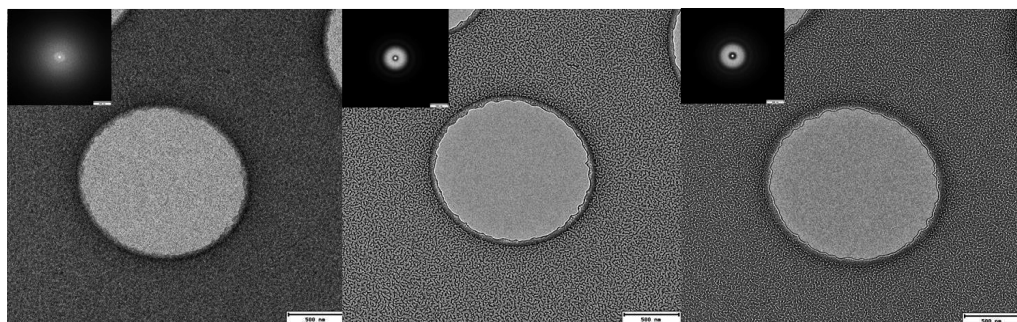
Fig. 2: True focus, Under-focus and over-focus of a hole in a commercial carbon coated grid. Minimum contrast corresponds to true Gaussian focus. Too much over- or underfocus provides an unsharp image with artefacts. In insets are the corresponding power spectra.

Different kind of contrast

There are four types of contrast in microscopy:
The amplitude or mass-thickness contrast is related to the thickness of the specimen and therefore to the density; a denser region will appear darker than a thinner one.
The phase contrast is made by using the objective aperture (in TEM) selecting only small angle scattering rays coming from the specimen. This contrast is made by the interferences between diffracted and transmitted waves after scattering through the specimen.
The Energy contrast (Energy-filtered TEM or EFTEM) will use the difference in atomic weight to make the distinction between different phases in a specimen. Only electrons of particular kinetic energies are used to form the image. In our case inelastic scattered electron are filtered and the only elastically scattered ones are used. This kind of contrast is quite thickness/density dependant.
The Diffraction contrast will enhance the differences between amorphous and crystalline phases in a specimen. It somehow depends on the grain orientation.
All these above mention contrast are inter-dependant.
In a standard TEM, mass thickness is the primary contrast mechanism for non-crystalline specimens, while phase contrast and diffraction contrast are the most important factors in image formation for crystalline specimens.

2.2.5   Image detection and recording

Until some years ago, projected images of specimen were made on photographic films which needed skills and patience to developp and produce images. The main inconvenient, essentially for Cryo-TEM measurements, was that beam induced artefact are discovered at the very end of the procedure. Nowadays, CCD, and more recently CMOS (Complementary metal–oxide–semiconductor) camera with capacities up to 4 k x 4 k are accessible and allows more cumfortable images recording and processing to improve their quality.

2.2.6   Image visualisation and processing

Integrated image visualisation program are available with most of the CCD or CMOS camera and are provided by the respective company. However, open source program such as Image J [10,11,12] are also freely available and are quite powerful to enhance brightness and contrast for example, to evaluate size distribution or to perform Fourier Transform. Image filtering to get rid of inelastic scattering for example are also possible. Cautions have, of course, to be taken and any processing more than contrast and brightness enhancement has to be explicitly mentionned in published images.

3   Cryo-Transmission Electron Microscopy

3.1   The Leidenfrost effect

The Leidenfrost effect [13] is a physical phenomenon in which a liquid, in near contact with a mass significantly hotter than the liquid's boiling point, produces an insulating vapor layer keeping that liquid from boiling rapidly. Because of this 'repulsive force', a droplet hovers over the surface rather than making physical contact with it. This is what happening when putting an object at room temperature in liquid nitrogen : the di-atom nitrogen N2, although in cold liquid state at -196°C, goes into its gas phase as far as it is in contact with the specimen which is approximatively 171°C warmer. In the best case, there is gradient of freezing sample with a high probability of water ice crystals formation leading the specimen to be not useable because of the consecutive leak of transparency to the neutrons and, in the worst case, the destruction of the specimen.

3.2   Cryogen

Ethane is the most commonly used cryogen for Cryo-TEM because of its higher cooling efficiency and easier handling regards to others. With a melting point at -183°C, a viscosity 9.10^{-3} poise at the melting point and a boiling point at -89°C, it has a relatively good coolig efficiency compare to liquid nitrogen (melting point at -210°C with viscosity of $0,2.10^{-3}$ poise and a boiling point of -196°C which is too close to the melting point and therefore doesn´t have a good cooling efficiency). An alternative is the liquefied propane which is sometime

used, even in mixture with ethane. However, although it has a melting point of -189°C and a boiling point of -42°C allowing more time to manipulate the frozen sample, it has a much larger viscosity (87.10^{-3} poise) [14].

Case of organic solvent

By investigating samples in organic solvent, one encounters a dilemma: at low temperature, i.e. -180°C, the ethane is in the liquid state and is therefore a solvent which may interact with the one of the samples and this interaction, thus, may induces artefacts. For some cases, liquid nitrogen is directly used and it was shown by the group of E. Talmon to work properly [15]. For other cases, see the sections 3.4 Freeze-Fracture-Direct-Imaging.

3.3 Cryo-Direct Imaging

The Cryo-Direct Imaging, or Cryo-DI, consists on the flash cooling of a thin layer of liquid sample within a grid, generally a holey carbon coated grid.

1) few microliter of liquid sample is deposited onto the grid which is hold by a tweezer inside the thermoregulated and humidity-controlled chamber of a cryo-plunge.
2) after a certain time to give the opportunity to the sample for relaxation, the grid is blotted with a filter paper. One has to take into account that a change in concentration might occur since the blotting has for purpose to remove excess of solvent/buffer.
3) the grid is therefore plunged rapidly into a container filled with liquefied ethane. The ethane liquefaction is performed by pouring ethane in its gas phase in the container surrounded by liquid nitrogen.
4) once plunged in liquid ethane, the specimen (i.e. grid + thin film of liquid sample) is transferred quickly in liquid nitrogen environment prior settlement on the tip of a pre-cooled cryo-specimen holder.
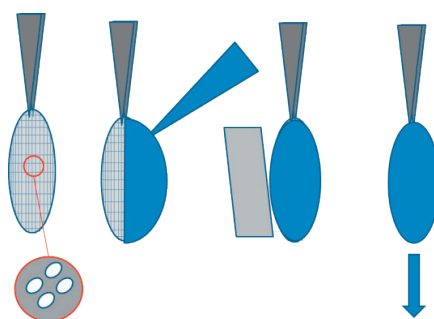
Fig. 3: Schematic for cryo-fixation by Cryo-plunge.

One has to keep in mind that by using holey carbon coated grid, relevant area of interest would be within the holes of the grid. However, the thin film of aqueous solution would have a concave shape leading particles, vesicles, etc… to be redistributed according to the available space. Another kind of grids are the lacey carbon grid where a network of carbon wire are

suspended or deposited onto the metallic grid. There is more space for the liquid to be spread between the wires and offer more probabilities to encounter the structures of interest.
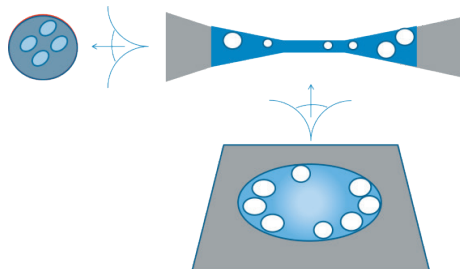
Fig. 4: close up of a commercial holeay carbon coated grid with eventual feature regard to nanoparticle dispersion in a thin water layer.

Nowadays, Controlled-Environment Vitrifiction System (CEVS) [16] are available with relative humidity and temperature controlled environment before plunging onto the cryogen.

3.4   Freeze-Fracture-Direct Imaging

To avoid interactions of liquid ethane as cryogen with sample containing organic solvent, one can use the Freeze-Fracture and Direct Imaging, or FFDI, method described by Belkoura et al in 2004 [17]. It consists of using a pair of grid fixed on cupper plates. Both combination of grid and plates are hold with a tweezer, each glued on each tip of the tweezer. Then, the grids are dipped onto the sample, which can be kept at a certain temperature. By closing the tweezer inside the solution, one traps a certain amount of sample. The whole "sandwich" system is then plunged in liquid ethane in the same way as for Cryo-DI. The plates are separated and the grids opened in liquid nitrogen environment. Each grid will present a freeze fracture feature where some areas will be thin and transparent enough for the electrons to go through.
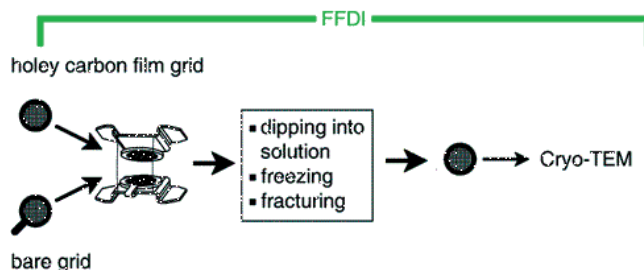
Fig. 5: Schematic drawing of the one-step preparation technique for FFDI, describing the main steps of the technique: sandwich building, dipping into the solution to be investigated, freezing in L Eth, and fracturing under liquid nitrogen. The sandwich is then separated to directly image one half. Adapted with permission from [17]. Copyright (2004) American Chemical Society.
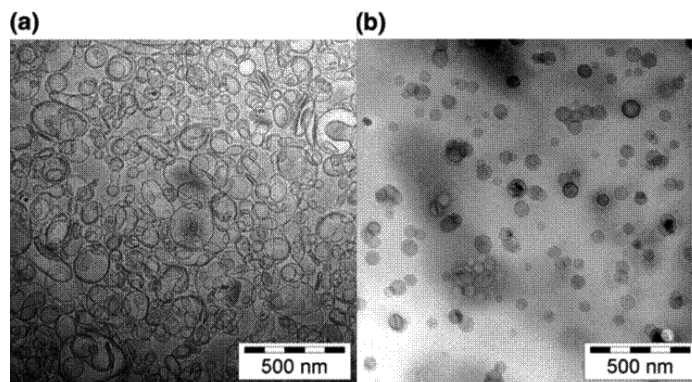
Fig. 6: Micrographs (instrument magnification 10 000×) of a physiologically buffered liposome solution containing 1.67 wt % vesicles. (a) Conventional cryo-TEM technique *with blotting* at room temperature. The total electron dose was 17 e Å^{-2}. Note the wide variety of vesicle shapes (shear effect) and sizes (size segregation). Moreover, the number density of vesicles is much too high to correspond to 1.67 wt % (concentration change). The gray values are modified with a $\gamma^{1/2}$ function. (b) Newly developed FFDI technique *without blotting* at room temperature. The total electron dose is 50 e Å^{-2}. All vesicles are spherical with a measured mean diameter of 61 ± 11 nm. The gray values are modified with a $\gamma^{1/2}$ function. Adapted with permission from [17]. Copyright (2004) American Chemical Society.

3.5   Cryo-Ultramicrotomy

3.5.1   Cryo-sectionning on bulk polymer sample

Some bulk polymer need to be cooled down at a certain temperature in order to be correctly sectioned. The ideal temperature is material dependent and a DSC measurement is usually sufficient to find it. Otherwise, time has to be spent to determine this temperature Tg. For instance, a too cold block specimen become too brittle to produce sections and a not enough cooled block would be too soft. Sectionning of bulk polymer composed of newly studied material is challenging essentially because of frequently existing heterogeneity of the sample. Bulk polymer sections sometime need to be hydrated by aqueous solvent and then frozen in order to be investigated by Cryo-TEM under vacuum.

3.5.2   Cryo-sectionning on frozen aqueous sample (CEMOVIS)

The Cryo-Electron Microscopy Of Vitreous Sample (CEMOVIS) method was developed by the groupd of J. Dubochet [18]. Frozen block specimen are produced by high pressure freezing at 2000 bar before freezing into liquid nitrogen. The frozen block specimen can be sectionned at RT after freeze substitution. The frozen sections, once collected onto a grid, can be transferred in a Cryo-holder and investigated by Cryo-TEM as described in the previous paragraphs or in the following ones.

3.6   Cryo-TEM for structural investigation

Recent progress in the TEM technology allows higher resolution structural investiogations of particles and particularly proteins. There are three groups according to analytical purpose and features of the biological sample. Cryo-electron crystallography, Cryo-single-particle reconstruction and Cryo-electron tomography (Cryo-ET).

3.6.1   Cryo-electron crystallography

Cryo-electron crystallography produces images and electron diffraction patterns. The method was developped by Henderson et al and the first electron crystallographic protein structure at atomic resolution was bacteriorhodopsin in 1990 [19]. It needs a regular 2D array of molecules and image enhancement by Fourier averaging. In a first step a regular image of 2D object with 3-Fold symmetry is taken by Cryo-TEM as described before. Then, phases are collected by performing Fourier transform of the image. An electron diffraction of noisy image is taken in order to measure the intensities of the peaks. By combining diffraction amplitudes with image phases one can obtain a clean 3-fold symmetry average image. The above mentioned step has to be repeated by tilting the specimen before building up a 3D Fourier space (amplitudes from electron diffraction, phases from Fourier inversion of image) and thus to calculate 3D electron density map.

3.6.2   Cryo-single-particle reconstruction

Although the principle of single particle reconstruction is from the early 70´s [20,21] , development of Cryo-EM [22] and, particularly progress in terms of computing development [23] allows the expansion of the technique. Cryo-single-particle reconstruction is produced from identical particles viewed in different orientations (Fig.7). However, the lower limit is 250 – 500 kDa. The advantage is that missing cone is not a problem as long as multiple views are present. Actually 1000s of particles has to be considered. The problem is the finding relative orientation of each particle which is essentially time consuming. First achieved with icosahedral viruses – 60 – fold symmetry reduces number of particles needed with 9Å resolution at best [20,21].

3.6.1   Cryo-electron tomography (Cryo-ET)

Similar to the tomography technique used in medicine, cryo-electron tomography, or Cryo-ET, is based on multiple images of the same specimen recorded at different tilt angles (Fig.8). The technique was developped for Cryo-TEM by Baumeister et al during the 90´s [24,25] and has a recent impressive development due to progress in images acquisition by CCD camera and in computing for automated tilt series acquisition, especially by J. Frank [26].

Fig. 7 : Cryo-Single Particle Reconstruction. First (top-left), a large number of particle are pictured by Cryo-TEM, then (top-right) the acquired images are sorted and oriented. Finally (bottom-left), the resulted aligned images are projected to reconstruct the 3 D particle.
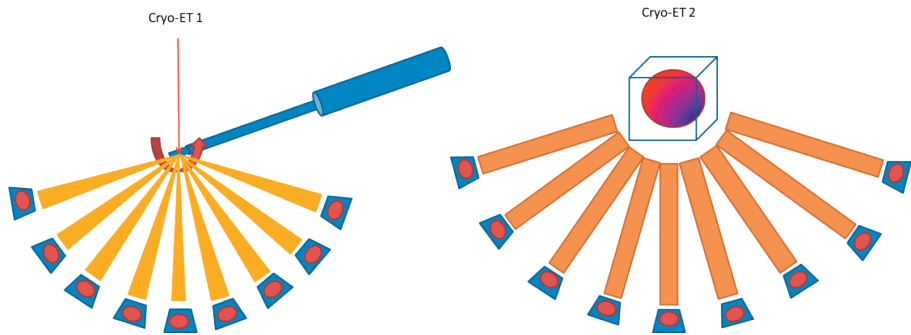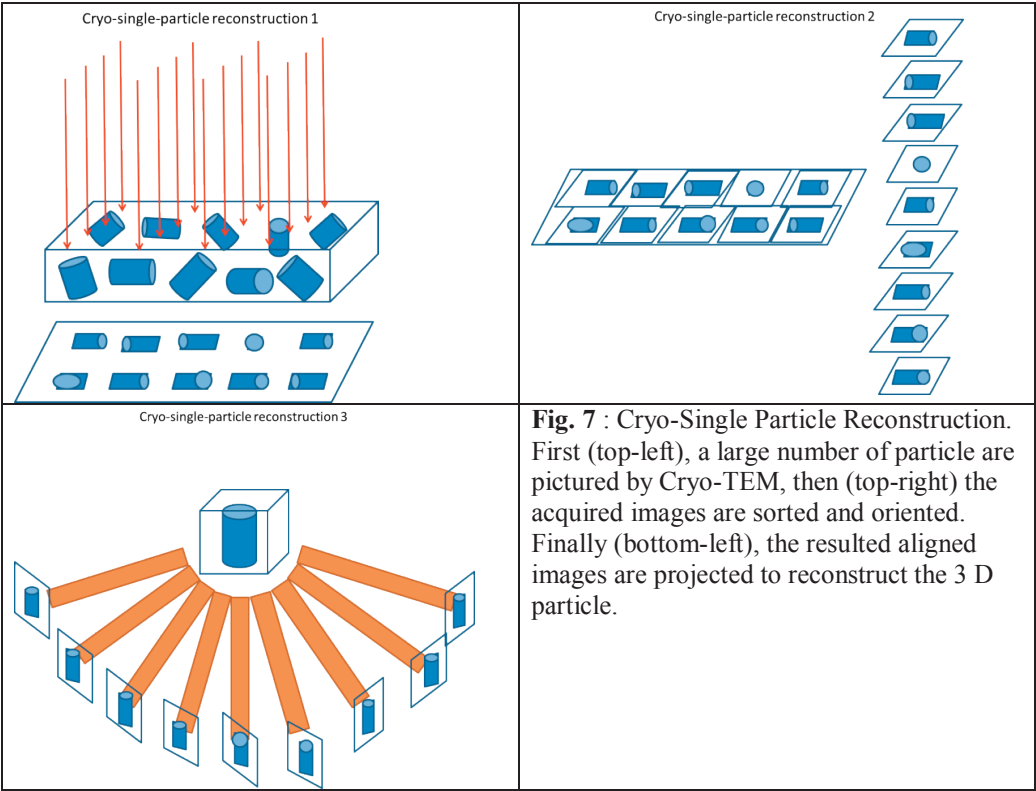
Fig. 8 : Schematic of principle for Cryo-Electron Tomograpy. Tilt series of images of a single particle are taken at different angles (left). The 3D structure of the particle is then reconstructed from the acquired images series (right).

3.7   Cryo-TEM as complementary techniques for neutron scattering

We present here couple of examples involving complementary techniques, essentially Small Angle Neutron Scattering and Cryo-TEM. As this can be seen, it is mainly a matter of comparison between values for typical distances such as radius, inter-particle or inter-vesicles spacing, etc…

3.7.1   Example 1 : Liposomes for Effective Drug Delivery

In the frame of the development of drug delivery system and particularly for cancer treatment, It is proposed by Acampora et al [27] to replace the usually used Poly-Ethylene-Glycol (PEG) but having some side effects [28-30] by the Lipo-oligosaccharide (LOS) from the Gram-negative bacterium Rhizobium rubi to improve the life time of liposomes particularly regard to the opsonisation process. To prepare the liposomes, de-O-acylated LOS (de-LOS) was co-formulated with 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid and the anticancer nucleolipid-based Ru(III) complex, ToThyRu.

What was shown by SANS and Cryo-TEM is the existence of co-existence of unilamellar and oligolamellar ToThyRu/POPC/de-LOS liposomes.

The results validate the use of lipooligosaccharides in formulating liposomes and pave the way to their use in drug delivery applications.

In fig. 9 is shown micrograph of ToThyRu/POPC/de-LOS 10:85:5 vitrified sample. Presence of liposomes is displayed. Note the oligo-lamellar structures (i. e., constituted by a limited number of concentric bilayers).
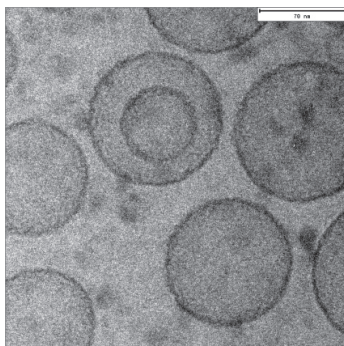
Fig. 9: Cryo-TEM image of ToThyRu/POPC/de-LOS 10:85:5 liposomes.

Figure 10 presents SANS results on the pure POPC and ToThyRu/POPC/de-LOS at 10:85:5 and 10:70:20 molar ratio samples: a power law decay for all the curves is observed, in the range of the scattering vector 0.020 <q/A< 0.035, slightly higher than- 2. As confirmed by Cryo-TEM, such value for power law indicates coexistence of uni-lamellar and oligo-lamellar liposomes.
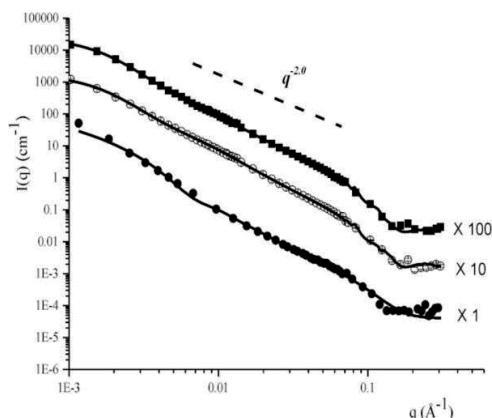
Fig. 10:. Scattering cross-sections of on pure POPC (open circles) and ToThyRu/POPC/de-LOS at 10:85:5 (closed circles) and 10:70:20 (closed squares) molar ratio samples in D2O. Solid lines correspond to the best fit of the Lamellar paraCrystal model to the experimental data. Reprinted with permission from [27] © 2016 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

The Lamellar paraCrystal Model [31] was used to fit the experimental data to extract the bilayer numbers, their thickness and the spacing (Figure 10). This model calculates the scattering from a stack of repeating lamellar structures of infinite lateral dimensions. The parameters extracted from the fitting suggest that the system is mainly constituted of unilamellar liposomes, the average number of lamellae is about 2 (considering the presence of oligolamellar liposomes) and their spacing is about 10 nm.

The use of Cryo-TEM is justified here to confirm the presence of the two populations of vesicles.

3.7.2   Example 2 : Complex Nanoassemblies

In the context of interpolyelectrolyte complexes (IPC) synthesis and particularly electrostatic coassembly, Delisavva et al [32] have used the combination of Cryo-Transmission Electron Microscopy (Cryo-TEM), Small Angle Neutron Scattering (SANS) and light scattering to investigate the electrostatic coassembly of the double-hydrophilic copolymer poly(2-vinypyridine)-blockpoly (ethylene oxide) (P2VP−PEO) with the anionic gemini surfactant 6,6′-(ethane-1,2-diylbis(oxy))bis (sodium 3-dodecylbenzenesulfonate) (G2). Comparison between single tail and Gemini surfactants have been made in this study.
Single-tail surfactants are, for example, sodium dodecyl sulfate (SDS) or dodecylpyridinium chloride (DPCl). Gemini surfactants consist of two hydrophilic head groups, two aliphatic chains (tails) attached to the heads, and a spacer connecting the heads [33].
Figure 11 shows SANS curves for P2VP− PEO/G2 mixtures at charge ratio Z ranging from 0.15 to 2; for P2VP−PEO concentration of 1 mg/mL.
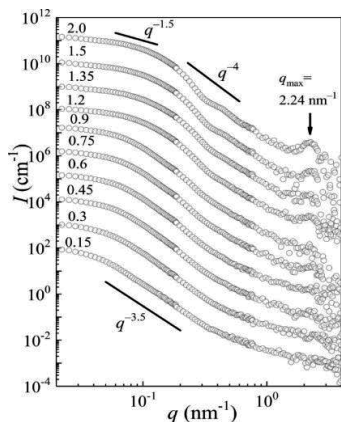
Fig. 11: SANS curves for P2VP−PEO/G2 complexes (polymer concentration, cpol = 1 mg/ mL) in 0.1 M DCl. Charge ratios, Z, are indicated above the individual curves. Reprinted with permission from [32]. Copyright (2017) American Chemical Society.
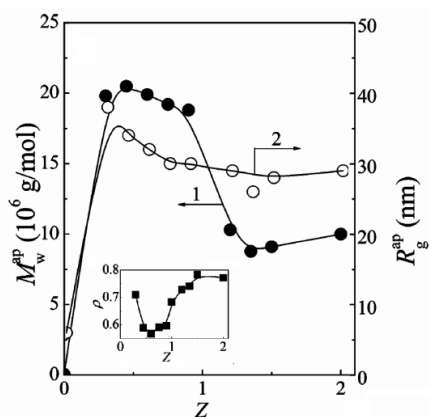
Fig. 12: Analysis of SANS data: molar masses (closed circle 1) and gyration radii (open circle 2) of P2VP−PEO/G2 complexes (cpol = 1 mg/mL) as functions of charge ratios Z. Inset: R_g/R_H ratios plotted as functions of Z. Adapted with permission from [32]. Copyright (2017) American Chemical Society.

From the above presented results in reciprocal space, confirmation of the results in the real space is needed. Therefore Cryo-TEM experiments were performed on the same samples in the same conditions and corresponding micrographs are presented in Figure 13.
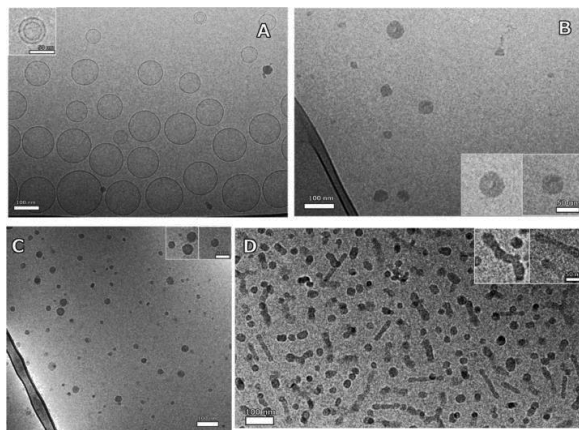
Fig. 13: Cryo-TEM images of G2 in 0.1 M DCl (A) and of P2VP−PEO/G2 at charge ratio Z = 0.3 (B), 1.0 (C), and 2.0 (D). Reprinted with permission from [30]. Copyright (2017) American Chemical Society.

Relevant results from SANS and Cryo-TEM are gathered in table 3:

Charge ratio Z	SANS (Figs. 11,12)	Cryo-TEM (Fig.13)
Z < 1	No structures. $Rg/R_H < 1$: the compactness of the aggregates is characteristic of core−shell micelles [34,35].	The PE−S nanoparticles are formed (Figure 13B). The particles are polydisperse in size, which is consistent with the scattering data. (radii from 10 to 50 nm). The hydrated coronas of the particles are not visible because of the low contrast.
Z = 1	$I(q)$ decays changes from $q^{-3.5}$ to $q^{-1.5}$. This indicates a transition from rough interfaces of compact particles [36] to more elongated scatterers. Decrease in the molar mass by a factor of 2.2 is observed, without any drastic change in the gyration radius which means the formation of core/shell particles.	(Figure 13C), a mixture of smaller spherical particles is observed.
Z > 1	Distinct oscillations are visible and are typical for cylindrical shapes of the particles. Increase in Rg/R_H indicating an elongation of the scatterers.	(Figure 13D) fusion of the small spheres into wormlike particles are present, which is in agreement with the SANS data. Although the spherical particles are touching each other's in several cases, no full merging into compact cylindrical particles is observed.

Table 3 : Sum up of the results from SANS and Cryo-TEM of P2VP−PEO/G2 complexes as functions of charge ratios Z

Additionally in the SANS data, for Z =2, a correlation peak with its maximum at 2.24 nm^{-1} appears and is due to dense packing of G2 in the PE−S core. In the Cryo-TEM image presented in Figure 13A, 20 to 100 nm polydisperse spherical single-wall vesicles are observed for G2 in 0,1 M DCl. Such structures were already described in pure water [37]. Multilayer vesicles are seen in a few images (insert).

From these complementary results between SANS, DLS and Cryo-TEM, in addition with Isothermal Titration Calorimetry, it was deduced that the structure of Gemini surfactants promote the formation of cylinders, but the structural reorganization is prevented by the high kinetic barrier due to the compactness of the structures that were formed at Z = 1. The structural transitions, essentially located at the cores, occur in a fast regime at Z close to 1 and particles are organized in elongated structure at Z > 1 in longer time scale [32].

3.7.3 Example 3 : Decorated Vesicles

Still in the frame of drug delivery system development, interactions between nanoparticles and vesicles have been investigated by Hoffman et al [38]. In particular, the influence of small

adsorbed silica nanoparticles (SiNPs) on the structure of zwitterionic DOPC vesicles have been studied with a combination of SANS, cryo- TEM and DLS.
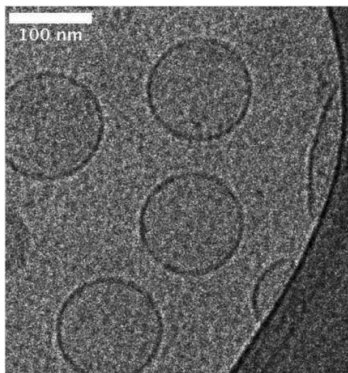
Fig.14: Cryo-TEM image of undecorated vesicles. From [38] Published by The Royal Society of Chemistry.
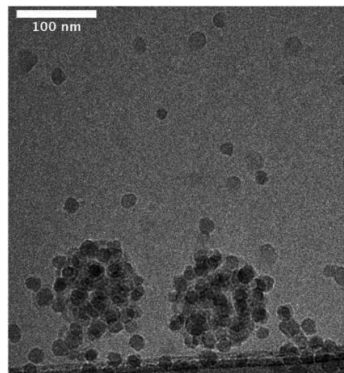
Fig. 15 : Cryo-TEM image of vesicles decorated with a [NP]/[vesicle] ratio of 12. From [38] Published by The Royal Society of Chemistry.

In figs 14 and 15 are evidence the adsorption of the nanoparticle on the vesicles. The size and shape of the vesicles remain unchanged to the situation without NPs. Even though cryo-TEM does not provide 3-dimensional information, it is obvious that the nanoparticles are located on the membrane and not inside the vesicle.

SANS measurements confirmed the unchanged structure of the vesicles with and without adsorbed nanoparticles since their rather low contrast in D2O make them hardly visible. Nearly no changes between the spectra with and without added NPs can be seen (Fig. 15).
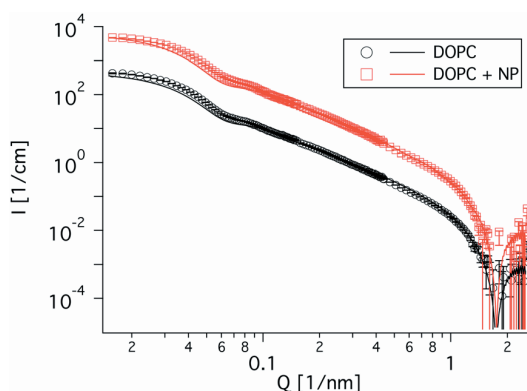
Fig. 16 : SANS curves of pure DOPC vesicles (0.1 wt%) and DOPC vesicles with 0.085 wt% NPs added. Almost no structural change can be seen. The curve with DOPC and NPs is shifted by a factor of 10. From [38] Published by The Royal Society of Chemistry.

The ratio between the intensities (with and without NPs) reveal a peak at 0,2 nm^{-1} related to the inter-particle correlation distance on the vesicles (Figure S2 in SI of [38]). In addition of DLS measurements, radii of the vesicles and inter-particle distances have been found to be in agreement by all the used methods. Cryo-TEM measurement was, here, extremely important to confirm the presence of the nanoparticles on the vesicles.

In the frame of tunable supramolecular complex formation, Houston et al [39] needed to attain space scale outside of the one accessible by SANS in their study. For instance, if the diameter of a vesicle is well outside the q-window of a SANS experiment (3.0 to 70.0 nm), the scattering would lead to deduce lamellar sheets (q^{-2}). Direct imaging by cryo-TEM has therefore been used to visualise the particle size and morphology in D2O and it was shown remarkable changes in morphology for phosphonium-functionalised homo- (P3HTPMe3) and diblock (P3HT-b-P3HTPMe3) ionic conjugated polythiophenes co-assembly upon addition of SDS. The Cryo-TEM technique has the advantage to preserve the micelles in a hydrated state and to keep the aqueous environment remains undisturbed.

In Crassous et al 2009 [40], two types of colloidal latex particles in dilute suspension are investigated by SAXS and Cryo.TEM: (i) core particles made of polystyrene with a thin layer of poly(N-isopropylacrylamide) (PNIPAM) and (ii) core-shell particles consisting of core particles onto which a network of cross-linked PNIPAM is affixed. It was shown that although good agreement was found by the two techniques for the core particle, only Cryo-TEM was able to reveal the buckling of the network affixed to the surface for the core-shell particles as SAXS is only sensitive to the average radial structure.

4   Conclusion

Transmission electron microscopy and particularly Cryo-TEM is a powerful tool more and more used in complementarity with other techniques (dynamic light scattering, NMR, FTIR,…) and even large scale instrument method such as neutron or X-ray scattering for soft matter and biology for example. Recent progress in TEM technology in parallel with improvement in computing allows to perform high resolution structural invetigations of proteins in complementarity (sometime in competition) with other techniques such as X-ray and NMR crystallography. In the frame of soft matter studies, Cryo-TEM and the requiring methods of sample preparation allows comparison of results with investigations made in liquid state or involving water in the system of interest: It allows to access space scale which are not reachable with techniques like light scattering, small angle neutron or X-ray scattering and thus complement the obtained results. Cryo-TEM gives also the possibilty to observe structurale features not visible with the above mentionned techniques.

References

[1] L. de Broglie, thesis, University of Paris 1924; L. de Broglie, Ann. Phys. 3, 1925, 22–128.
[2] E. Ruska und M. Knoll, In: Z. techn. Physik. 12, 1931, S. 389–400
[3] H. Busch, Ann. Physik 1926 ,81, 974,

[4] V.E. Cosslet and M.E. Haine, Proc. Conf. Electron Microscopy, London 1954 (Royal Microscopical Soc., London; 1956 p. 639.
[5] A.V. Crewe et al, Rev. Sci. Instr. 1968, 39, 4
[6] S. Brenner and RW Horne, Biochim Biophys Acta 1959, 34, 103–110
[7] M. Adrian, J. Dubochet, J. Lepault, and A.W. McDowall, Nature 1984, 308, 32-36
[8] J. Goldstein, Scanning Electron Microscopy and X-Ray Microanalysis, Springer US, 2003
[9] R. F. Egerton, Reports on Progress in Physics, 2009, 72, 016502
[10] W.S. Rasband, ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2012.
[11] C.A. Schneider et al, Nature Methods 2012, 9, 671-675.
[12] M.D. Abramoff et al, Biophotonics International, 2004, 11, 36-42
[13] J. G. Leidenfrost, Hermann Ovenius, Duisburg 1756
[14] P. Echlin, in Low-Temperature Microscopy and Analysis, Springer Science + Business Media, LLC 1992 p 71
[15] D. Danino et al, J. Colloid. Interf. Sci., 2002, 249, 180-186
[16] J.R. Bellare et al, J Electron Microsc Tech 1988, 10, 87–111
[17] L. Belkoura et al Langmuir 2004, 20, 4391-4399
[18] A. Al-Amudi et al EMBO J. 2004 Sep 15; 23(18): 3583–3588
[19] R. Henderson et al J Mol Biol. 1990, 213 (4): 899–929,
[20] R. A. Crowther, , et al. Nature 1970, 226, 421–425,
[21] D. J.DeRosier, and P. B. Moore, J. Mol. Biol. 1970, 52, 355–369,
[22] Y.Cong, , et al. Proc. Natl. Acad. Sci. 2010, USA 107, 4967–4972.
[23] J.Frank, , et al. J. Struct. Biol. 1996, 116, 190–199.
[24] K. Dierksen et al, Ultramicroscopy,1993,49, 109-120
[25] AJ Koster, et al J Struct Biol, 1997, 120, 276-308.
[26] J. Frank, 2006, Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell, 2nd edn. London, New York: Springer.
[27] F. Acampora et al, ChemistrySelect 2016, 1, 2129 – 2139
[28] A. VertutDoi, et al, Bba-Biomembranes 1996, 1278,19–28.
[29] A. Vergara, et al, Phys Chem Chem Phys 1999, 1, 4875–4879.
[30] K. Knop, et al, Angew Chem Int Edit 2010, 49, 6288–6308.
[31] M. Kotlarchyk, et al, J Appl Crystallogr 1991, 24, 753–758.
[32] F. Delisavva et al Macromolecules 2017, 50, 21, 8745-8754.
[33] M. J. Rosen et al, Surfactants Deterg. 1998, 1, 547−554.
[34] J . Clara-Rahola, et al, J. Chem. Phys. 2012, 136, 214903.
[35] A. Fernández-Nieves et al J. Chem. Phys. 2004, 120, 374−378.
[36] M. A. Crichton et al, Langmuir 2005, 21, 10028−10031.
[37] Z. Hordyjewicz-Baran et al Colloids Surf., A 2015, 484, 336−344.
[38] I. Hoffmann et al Nanoscale, 2014, 6, 6945
[39] J. Houston et al, Nanoscale, 2017, 9, 17481
[40] J. Crassous et al, Langmuir 2009, 25, 7862–7871

The first part of this chapter (section 2) is based on the following text book:
D.B. Williams, C.B. Carter. Transmission Electron Microscopy : A Textbook for Materials Science, Springer Science + Business Media, LLC 1996, 2009

F 6   Advanced Transmission Electron Microscopy

R.E. Dunin-Borkowski[1,2], J. Mayer[1,3],

L. Houben[1,4], M. Feuerbacher[2], M. Heggen[1], L. Jin[1],

A. Kovács[1], J. Caron[1], T. Denneulin[1], M. Luysberg[1], A. Thust[1]

[1]Ernst Ruska-Centre for Microscopy & Spectroscopy with Electrons, Forschungszentrum Jülich GmbH, Jülich, Germany

[2]Institute for Microstructure Research, Peter Grünberg Institute, Forschungszentrum Jülich GmbH, Jülich, Germany

[3]Central Facility for Electron Microscopy, RWTH Aachen, Germany

[4]Department of Materials and Interfaces, Weizmann Institute of Science, Rehovot, Israel

Contents

Introduction

Modern transmission electron microscopes (TEMs) can be used to obtain quantitative information about the structural, electronic and chemical properties of materials on length scales down to the sub-Å level [1-3]. In recent years, electron microscopy has been revolutionized by improvements in aberration correctors, monochromators and imaging filters, as well as by advances in computing power for microscope control, image analysis and image simulation. Specialized techniques have also been developed, including the use of *in situ* gas reaction and liquid cell electron microscopy to study growth processes and chemical reactions in materials and the application of bright and dark field electron holography and diffraction-based phase contrast techniques to measure variations in electrostatic potential, magnetic induction and crystallographic strain in materials. Photographs of two generations of modern aberration corrected transmission electron microscopes are shown in Fig. 1.

Fig. 1: (Left) FEI Titan 80-300 and (right) FEI Titan G3 50-300 PICO field emission transmission electron microscopes at Forschungszentrum Jülich. The instrument on the left is equipped with a spherical aberration corrector on the objective lens and has an information limit of 0.08 nm. Photograph courtesy of Ralf-Uwe Limbach (Forschungszentrum Jülich). The instrument on the right is a fourth generation transmission electron microscope equipped with a monochromator, a spherical aberration corrector on the condenser lens and a combined spherical and chromatic aberration corrector on the objective lens. It has an information limit for coherent high-resolution TEM of below 50 pm at 200 kV. Photograph courtesy of Christian Lüning (www.arbeitsblende.de).

The basic operation of a TEM is in many respects analogous to that of a light optical microscope. The electron source at the top of the column is either a thermionic emitter such as a LaB_6 single crystal or a field emitter such as a zirconia-coated W tip. The primary advantage

of a field emission source is that it is brighter, as a result of the smaller extraction area for electrons, which can be only a few nm in size. After emission, the electrons are accelerated by a voltage of typically between 50 and 300 kV and focused onto a specimen by 2 or 3 condenser lenses. The specimen is commonly a 3 mm disk or chip that is prepared to be extremely thin in the region of interest, as the typical specimen thickness for high-resolution imaging is below a few tens of nm. The specimen stage allows movement of the sample in three spatial directions and tilting about two axes. The objective lens is located directly below the specimen and has a focal length of only a few mm. Its design and the stability of its power supply are crucial for the electron optical performance of the microscope. The intermediate and projector lenses in the lower part of the column are used to magnify the image and have a relatively small influence on image quality. After passing the projector lens, the electrons can be observed on a fluorescent screen or recorded on a charge coupled device camera or a direct electron detector. Vibration damping systems and electromagnetic field compensation systems are normally required to create a sufficiently stable environment for a state of the art instrument.

1 Aberration correction in high-resolution TEM

High-resolution TEM (HRTEM) involves the acquisition of images with a spatial resolution that is sufficient to separate single atomic columns. The interpretation of such images is, however, not straightforward, as the recorded intensity is not a direct representation of the specimen but an interference image that is affected by the strength of interaction of the incident electrons with the specimen and by the contrast transfer of the microscope [4].

The interaction of an incoming electron wave with a TEM specimen can be described by solving the relativistically corrected Schrödinger equation for the electron wavefunction $\Psi(\mathbf{r})$ in a crystal potential $V(\mathbf{r})$ according to the Bethe-Bloch formalism [5]. In practice, several approximations are often used to understand the image formation process. In the phase object approximation (POA) for a thin specimen, atoms in the sample are described by a projected potential that is continuous and constant in the direction of the incident electron beam. The electron wavefunction after specimen thickness t can then be written in the form

$$\Psi(\mathbf{r},t) \approx \exp[i\phi(\mathbf{r},t)] = \exp[i\sigma V_P(\mathbf{r})t] \; , \tag{1}$$

where ϕ is the phase of the electron wave, σ is an interaction constant that depends on the microscope accelerating voltage and V_P is the projected crystal potential. The weak phase object approximation (WPOA) further assumes that the phase modulation of the electron wave is small in a very thin specimen, resulting in the expression

$$\Psi(\mathbf{r},t) \approx 1 + i\phi(\mathbf{r},t) = 1 + i\sigma V_p(\mathbf{r})t \; . \tag{2}$$

The influence of the electron microscope lenses on image formation is described by modifying the exit plane wavefunction using a phase factor $\exp[-i\chi]$ according to the expression

$$\Psi_i(\mathbf{g}) = \Psi(\mathbf{g})\exp[-i\chi(\mathbf{g})] \; , \tag{3}$$

where, for the rotationally symmetric aberrations of a round lens, the aberration function

$$\chi(g) = 2\pi \left(\frac{1}{4} C_S \lambda^3 g^4 + \frac{1}{2} Z \lambda g^2 \right) . \tag{4}$$

In Eq. 4, λ is the wavelength of the electron beam, Z is the defocus and C_S is the coefficient of spherical aberration of the objective lens, which cannot be avoided for a round electromagnetic lens [6] and describes the deviation of rays that pass the outer part of the objective lens compared with near-axis rays and the resultant blurring of the object.

For a thin specimen, the WPOA results in an expression for the linear image intensity of the form

$$I_L(\mathbf{g} \neq 0) \approx 2\sigma V(\mathbf{g}) t \sin \chi(\mathbf{g}) , \tag{5}$$

from which it is apparent that the optimum image contrast is obtained when the coherent contrast transfer function (CTF) $\sin \chi = \pm 1$, *i.e.*, when $\chi(\mathbf{g})$ is an odd multiple of $\pi/2$ for all values of g. However, $\sin \chi$ is in general a function that oscillates strongly with g. As a result, atomic columns that are arranged with a spacing of $d = 1/g$ are imaged as black dots only at selected spatial frequencies for which the CTF is close to -1. At spatial frequencies for which the CTF is close to +1, the atomic columns are imaged as white dots, while they may be invisible if the spatial frequency corresponding to their interatomic spacing coincides with a zero of the CTF. In order to approach ideal phase contrast transfer behaviour, a defocus setting can be chosen that balances the g^2 and g^4 terms in the aberration function, allowing for a relatively broad band of frequencies to be transferred with a CTF close to -1. This defocus

$$Z_S \approx -\sqrt{\frac{4}{3} C_S \lambda} \tag{6}$$

is known as Scherzer defocus [7]. The point resolution $d_S = 1/g_S$ is defined by the first zero crossing of the CTF at Scherzer defocus and is given by the expression

$$g_S \approx \left(\frac{3}{16} C_S \lambda^3 \right)^{-\frac{1}{4}} . \tag{7}$$

Typical values of point resolution for commercially available non-aberration-corrected medium voltage transmission electron microscopes at accelerating voltages of 200-400 kV are in the range 0.24 to 0.17 nm.

The limited coherence of the electron source and electronic instabilities have the additional effect of multiplying the coherent CTF by envelope functions, which result in a cut-off of the contrast transfer at higher spatial frequencies. The spatial frequency at which the partially coherent CTF falls below a threshold value defines the information limit of the microscope. For an uncorrected field emission TEM, the information limit can be higher than the point resolution, leading to strong oscillations in the partially coherent CTF and blurring in images. Figure 2 shows $\sin \chi$ for a 200 kV instrument for two different values of C_S. At higher spatial frequencies, the CTF oscillates rapidly up to the information limit. If C_S is reduced, then a broad transfer band extends up to the information limit of 0.125 nm, improving the point resolution.
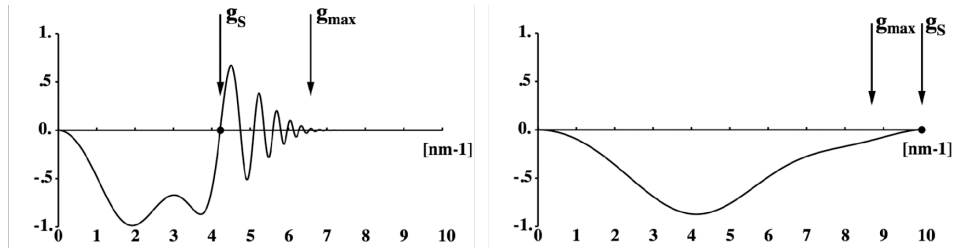
Fig. 2: Partially coherent contrast transfer functions at Scherzer defocus calculated for a CM20 field emission microscope operated at 200 kV for (left) C_S = 1.2 mm and (right) C_S = 0.04 mm. Figure created by A. Thust (Forschungszentrum Jülich) and reproduced from the notes for the 38th IFF Spring School "Probing the Nanoworld" (2007).

An important approach that can be used to overcome the influence of the strongly oscillating part of the CTF between the point resolution and the information limit for an uncorrected transmission electron microscope is to use numerical reconstruction of the exit plane wavefunction from a defocus series of 10-20 images recorded from the same object area [8-10], as shown schematically in Fig. 3. According to the phase object approximation, the heights of the phase maxima in the wavefunction are approximately proportional to the projected potential for a thin specimen, permitting the chemical distinction of atomic species. The resulting availability of the complex-valued wavefunction also enables correction of aberrations in software. In the resulting wavefunction, it becomes possible to resolve light elements such as C, N and O close to or even below 0.1 nm spatial resolution.
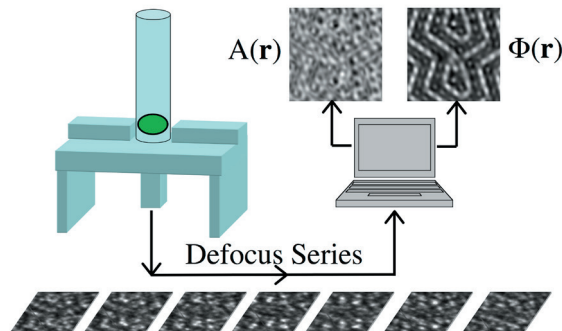
Fig. 3: Principle of focal series restoration. A series of images is recorded from the same object area using different settings of the objective lens defocus. The quantum mechanical exit plane wavefunction, which consists of amplitude and phase, can be retrieved from the series of images by means of numerical procedures. Figure created by A. Thust (Forschungszentrum Jülich) and reproduced from the notes for the 38th IFF Spring School "Probing the Nanoworld" (2007).

An example of the elimination of delocalisation by applying focal series restoration using non-linear reconstruction to 20 images is shown in Fig. 4 for a twin boundary in $BaTiO_3$. The left side of Fig. 4 shows one image from the focal series, which was recorded using an uncorrected Philips CM20ST field emission microscope. Blurring of the image due to objective lens aberrations is visible. The right side of Fig. 4 shows the phase of the restored wavefunction, in which blurring has been reduced and the atomic positions are resolved up to the information limit of the electron microscope [11].
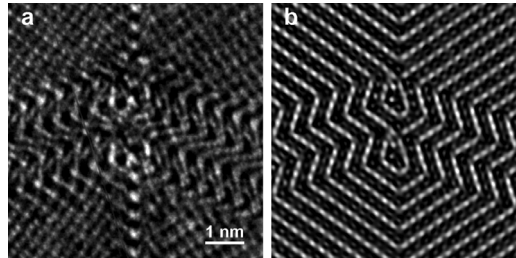
Fig. 4: Example of focal series restoration. (a) Representative image from a series of 20 images of twinned [110] oriented $BaTiO_3$. (b) Phase of the restored exit plane wavefunction, from which aberrations have been removed numerically. The positions of atomic columns are revealed by phase maxima. Figure created by C.L. Jia, R. Rosenfeld and A. Thust (Forschungszentrum Jülich) and reproduced from Ref. [11].

Significantly, the elimination of spherical aberration using *hardware* is now possible with multipole lenses. The first successful demonstration of spherical aberration correction was achieved in the late 1990s in a project funded by the Volkswagenstiftung involving Forschungszentrum Jülich [12, 13]. By tuning the spherical aberration coefficient and other higher order aberrations, optimum contrast transfer and a dramatic improvement in resolution can be achieved. Figure 5 shows a comparison between an experimental defocus series of [110] $SrTiO_3$ recorded using a C_S corrected Philips CM200ST field emission microscope and simulated images. A structure image predicted by simulations for a defocus of +10 nm appears in the experimental dataset.
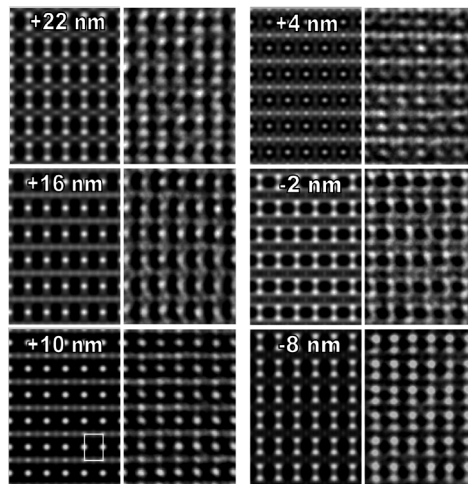
Fig. 5: Comparison of an experimental defocus series of $SrTiO_3$ oriented along [110] with simulated images. The experimental images were recorded using a C_S corrected CM200ST field emission microscope using a value for C_S of –0.04 mm. Simulated images are shown on the left. Experimental images are shown on the right. The defocus values of the simulated images are indicated. The specimen thickness used for the simulations was 3.5 nm. Half of a $SrTiO_3$ cell projected along [110] is indicated by the white frame in the lower left simulated image. Figure created by C.J. Jia (Forschungszentrum Jülich) and reproduced from the notes for the 34th IFF Spring School "Fundamentals of Nanoelectronics" (2003).

The ability to tune the spherical aberration coefficient of the microscope opens up new ways to record directly interpretable images. In particular, the choice of a small negative value for C_S results in strong contrast from light atom columns, which appear bright [14]. When using this negative C_S imaging (NCSI) technique, paraxial rays travelling through the objective lens come to focus ahead of outer rays. Light elements such as oxygen in $Pb(Zr_{0.2}Ti_{0.8})O_3$, which are adjacent to columns of heavy atoms, can then be resolved [15].

It should be noted that the complete aberration function is made up of many individual aberrations. Sub-Å imaging requires both a knowledge of all significant wave aberrations (typically up to sixth order in g) and the stability of these aberrations over the length of an experiment [16]. The decomposition of a typical aberration function into its components is illustrated in Fig. 6, while the time stability of the defocus measured for two different electron microscopes is show in Fig. 7.
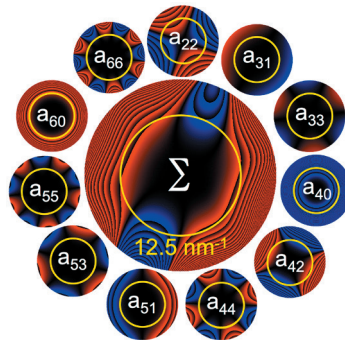
Fig. 6: Aberration function (centre) and its decomposition into basic aberrations. Positive values of the aberration function are shown in red, while negative values are shown in blue. Sawtooth jumps occur at intervals of $\pi/4$. The yellow circle marks a value of $g = 12.5$ nm^{-1}, which corresponds to a resolution of 0.08 nm. "Aberration-free" imaging is achieved when the modulus of the total aberration function does not exceed $\pi/4$ within a circle that defines the resolution limit. Figure created by J. Barthel (Forschungszentrum Jülich).
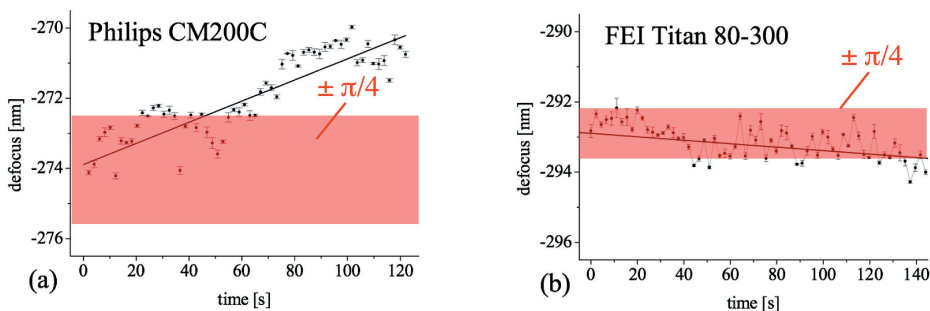
Fig. 7: Time variation of the objective lens defocus measured over a period of approximately 2 minutes for (a) a 200 kV Philips CM200 TEM with a resolution of 0.12 nm and (b) a 300 kV FEI Titan TEM with a resolution of 0.08 nm. Both thermally-induced long-term drift and short-term fluctuations caused by instabilities in the accelerating voltage and the objective lens current are observed. Figure created by J. Barthel and A. Thust (Forschungszentrum Jülich) and adapted in part from Ref. [16].

The combined application of chromatic aberration (C_C) correction and spherical aberration correction in HRTEM results in a significant improvement in spatial resolution to 50 pm at an accelerating voltage in the range of 200-300 kV. An illustration of the improvement in image resolution and interpretability with microscope generation and the use of aberration correction is shown in Fig. 8 for simulated images of AlN.

CM30-UT (1992)	CM200-Cs (1998)	Titan 80-300 (2005)	PICO (2012)
300 kV FEG	200 kV FEG	300 kV FEG	300 kV X-FEG
no corrector	Cs corrector	Cs corrector	Cc & Cs corrector
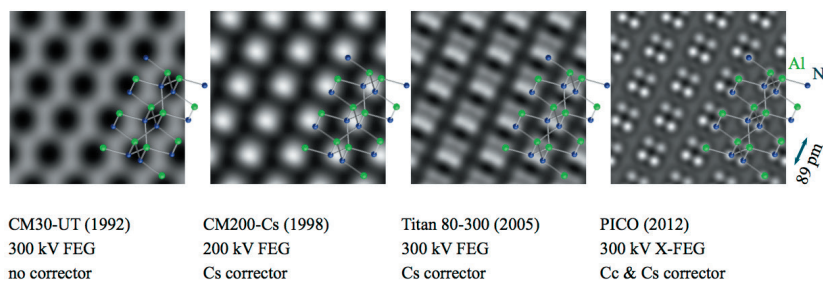
Fig. 8: Simulations of an AlN crystal viewed along the [113] zone axis for different transmission electron microscopes operated under optimised conditions (Scherzer defocus and Lentzen focus). Figure created by K. Tillmann (Forschungszentrum Jülich).

Figure 9 illustrates the resolving power of the C_S/C_C corrected PICO TEM in Forschungszentrum Jülich for **b**-axis-oriented yttrium-aluminium perovskite (YAP) [17]. At room temperature, YAP has the orthorhombic structure (space group *Pnma*) with lattice parameters $a = 0.5330$ nm, $b = 0.7375$ nm and $c = 0.5180$ nm, as shown schematically in Fig. 9a. Along the **b**-axis, Y-Y-atom pairs (marked by bold lines) have projected separations of ~57 pm along two directions. The tilts of corner-shared oxygen octahedra are indicated by grey rectangles. Figure 9b shows an experimental C_S/C_C corrected HRTEM image of YAP recorded without any post-filtering.
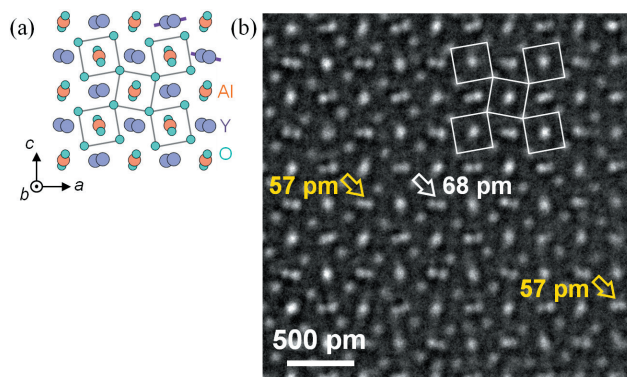
Fig. 9: (a) Schematic diagram of $2 \times 2 \times 2$ orthorhombic unit cells of YAP viewed along the crystallographic **b** axis. 57 pm Y-Y-atom pair separations are indicated along two directions by bold lines. The corners of the marked squares indicate the positions of O atoms, while their centres indicate the positions of Al atoms overlapping with O atoms. (b) Experimental C_S/C_C corrected HRTEM image of YAP viewed along the **b** axis. Arrows denote Y-Y atom pairs, with measured pair separations indicated. Figure adapted from Ref. [17].

The apparent separations of the Y-Y-atom pairs marked in Fig. 9b were measured by locally fitting two Gaussian functions to the image intensity. Although they take values of between 55 and 85 pm with a mean value of 69 pm, the nominal 57 pm separations are resolved along two directions, as indicated by yellow arrows. To the best of our knowledge, this is the highest direct resolution in coherent HRTEM atomic imaging that has been achieved at 200 kV to date. The pure O atomic columns are imaged with very high contrast, as outlined by white lines in the experimental image and marked in the structural model shown in Fig. 9a.

2 Aberration correction in high-resolution STEM

Figure 10 shows a schematic illustration of a scanning TEM (STEM), in which a focused electron beam is scanned across a specimen. For each position of the electron probe, a bright-field detector, a dark-field detector or a pixelated detector collects electrons that have interacted with the specimen. A microscope such as those shown in Fig. 1 can be used for both HRTEM and STEM if it is equipped with a scanning unit above the specimen and detectors in the diffraction plane. In order to achieve a resolution in the sub-Å range, it must also be equipped with an aberration corrector for the condenser lens system.
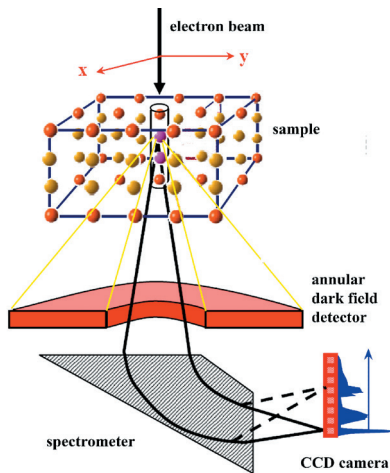
Fig. 10: In STEM, an electron probe is scanned across a specimen. For each position of the probe, electrons are collected using a detector located in a diffraction plane. A spectrometer can also be used to record an electron energy-loss spectrum.

Spatial resolution in STEM is determined primarily by the size of the electron probe. The probe function is, in turn, determined by parameters that include the diameter of the condenser aperture and the aberrations of the condenser lens system. Correction of the aberrations of the condenser lens system is crucial to obtain sub-Å resolution.

Under certain assumptions, the image intensity in an annular dark-field (ADF) image is given by a convolution of the square of the probe function and the square of the specimen transmission function. The image formation process can be regarded as incoherent if the inner angle of the ADF detector is much larger than the diameter of the condenser aperture that defines the convergence angle of the illumination. Typically, only electrons that have been scattered to

sufficiently high angles (on the order of 100 mrad, compared to about 20 mrad for the semi-angle of the condenser aperture) must be collected by the detector. For a sufficiently thin specimen, to a first approximation the image intensity in such a high-angle ADF (HAADF) image is proportional to the atomic number Z raised to the power 1.7. However, the image can also be affected by beam broadening as the electron probe travels through the specimen and by dynamical diffraction.

An example of the application of aberration corrected STEM to the characterization of a complex metallic alloy (CMA) sample is illustrated in Fig. 11. Such materials have unusual electronic transport [18], magnetic [19] and plastic [20] properties and contain novel types of dislocations, which are referred to as metadislocations [21]. In the T-Al-Mn-Pd phase, which is orthorhombic with lattice parameters $a = 1.47$ nm, $b = 1.25$ nm and $c = 1.26$ nm, a deformation mechanism based on the movement of a novel type of dislocation was found. Figure 11 shows an aberration corrected HAADF STEM image of a dislocation recorded along the b direction of the T-phase structure [22]. The unit cell (left blue rectangle), which contains 156 atoms, can be divided further into structural subunits, which are represented by elongated hexagonal tiles (white and yellow polygons) that are arranged in rows of alternating orientation and mutually tilted by 36°. The white contrast dots located at the centres of the hexagons correspond to atomic columns that contain Pd ($Z = 46$). Lower contrast dots at the edges and vertices of the hexagons correspond to atomic columns that contain Mn ($Z = 25$) and Al ($Z = 13$).
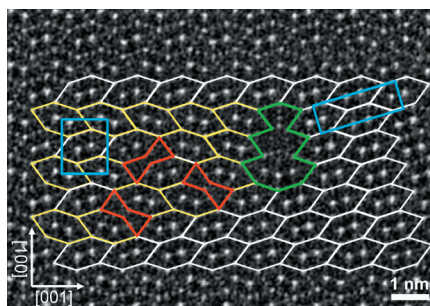
Fig. 11: High-resolution HAADF STEM image of a dislocation in the CMA T-Al-Mn-Pd. White and yellow hexagons represent structural subunits of the T-phase (left) and R-phase (right). The corresponding unit cells are shown as blue rectangles. Pd columns ($Z = 46$) at the centres of the hexagons generate stronger HAADF image contrast than Mn ($Z = 25$) and Al ($Z = 13$) columns at the edges and vertices of the hexagons. Figure adapted from Ref. [22].

An example of the application of aberration corrected HRTEM and aberration corrected STEM to similar materials is shown in Fig. 12. The images were obtained from semiconductors that contain transition metal atoms introduced with the intention of combining ferromagnetic and semiconducting properties for the design of spin-electronic devices. In such materials, ferromagnetism can result from the presence of nanoscale clusters of magnetic atoms or randomly located diluted transition metal impurities or defects. Care was required to minimize artefacts when preparing the cross-sectional TEM specimens. Figure 12a shows an aberration corrected HRTEM image of Mn-doped GaAs grown by molecular beam epitaxy containing hexagonal (NiAs-type) MnAs crystals, voids and As crystals [23]. Figure 12b shows an aberration corrected HAADF STEM image of Fe-doped GaN grown by metalorganic chemical vapour deposition containing Fe-rich nitride crystals and N_2-filled bubbles [24].
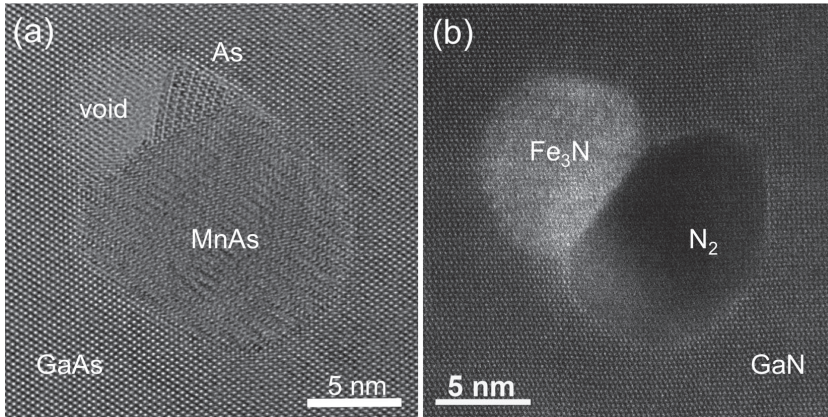
Fig. 12: (a) Aberration corrected HRTEM image of hexagonal MnAs, orthorhombic As and a void in a GaAs host. This figure is adapted from Ref. [23]. (b) Aberration corrected HAADF STEM image of Fe_3N and an adjacent N_2 bubble in a GaN host. The ADF inner detector semi-angle used was 47.4 mrad. Figure adapted from Ref. [24].

Figure 13 shows aberration corrected bright-field and dark-field aberration corrected STEM images of the topological insulators Bi_2Te_3 and Sb_2Te_3, which comprise quintuple layers such as Te-Bi-Te-Bi-Te that are bound together by van der Waals forces. The present example shows a 6-nm-thick Bi_2Te_3 layer covered by a 15-nm-thick Sb_2Te_3 layer. The layer system is grown on a Si (111) substrate and shows perfect epitaxial growth. Since Bi_2Te_3 is intrinsically *n*-type, while Sb_2Te_3 is intrinsically *p*-type, the Fermi level at the surface of the stack can be tuned by choosing appropriate layer thicknesses. In this way, chosen topological surface states can be formed, resulting in dissipationless transport [25, 26].
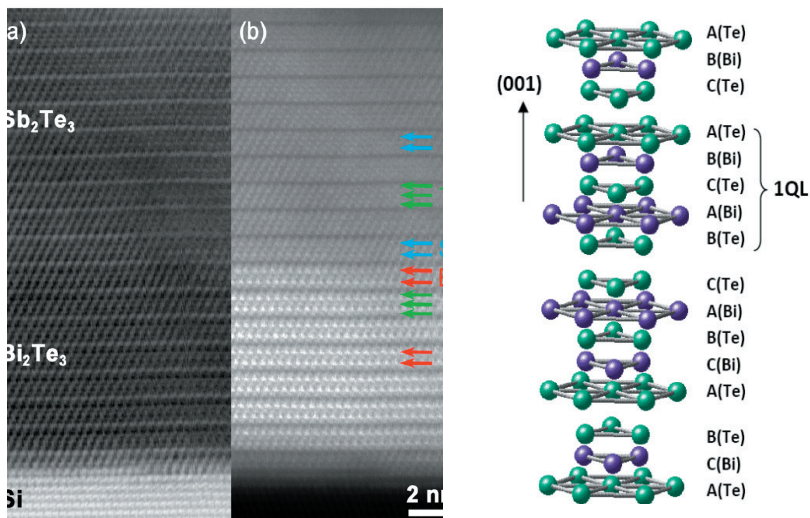
Fig. 13: Bright-field and dark-field aberration corrected STEM images of a 6-nm-thick Bi_2Te_3 layer covered by a 15-nm-thick Sb_2Te_3 layer on a Si (111) substrate. Figure adapted from Ref. [25].

3   Electron energy-loss spectroscopy

Electron energy-loss (EEL) spectra can be recorded using either post-column or in-column spectrometers to reveal the inelastic interactions of electrons with a specimen, which involve both energy and momentum transfer. Figure 14 shows the characteristic features in an EEL spectrum. The intensity is plotted as a function of energy loss, with zero energy corresponding to elastically scattered electrons. Small energy losses result from valence losses, *i.e.*, plasmon excitations, interband and intraband transitions. Losses at higher energies are associated with inner shell ionization. Since the latter signals are weak compared to the low loss region, different spectrometer settings are typically used to record the low loss and high loss regions of a spectrum. As the total signal in an ionization edge is proportional to the number of excited atoms, it can be used for quantification of the local chemical composition. Evaluation of the integrated signal requires subtraction of the background in the spectrum. In most cases, a power law function of the type

$$I(\Delta E, A, r) = A \exp(-r \Delta E) \qquad\qquad (8)$$

can be used, where A and r are fitting parameters. Details about quantitative fitting of EEL spectra using calculations of inelastic scattering cross-sections and the treatment of specimen thickness effects can be found elsewhere [27].
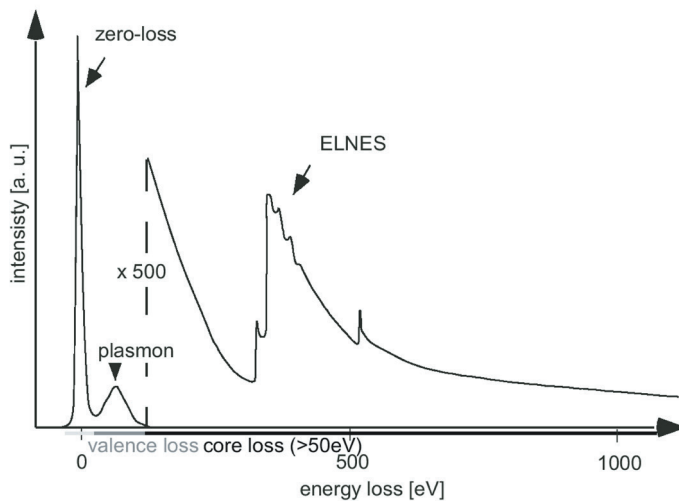
Fig. 14: Schematic EEL spectrum showing a wide range of energy losses, including elastically scattered electrons (zero-loss peak), valence losses (plasmons, interband and intraband transitions) and core ionization edges. EEL spectra have a large dynamic range. The ionization edges are superimposed on a background signal.

The energy loss near edge structure (ELNES) marked in Fig. 14 contains information about the local chemical environment of an excited atom. Two models, which are shown in Fig. 15, are typically used to understand the edge structure: (a) a band structure model, in which an electron from a shell is excited into unoccupied states above the Fermi level; (b) a multiple scattering

model, in which ELNES is described by an outgoing wave that is scattered by surrounding atoms. The intensity of the edge scales primarily with the density of final states.
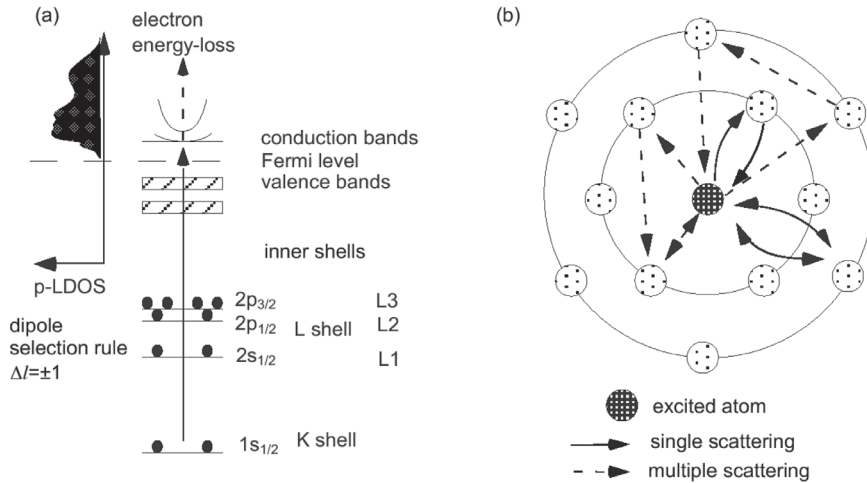
Fig. 15: Schematic representations of two models for the origin of electron energy-loss near-edge structure (ELNES) for core ionization edges. (a) Transition of strongly bound core electrons into unoccupied states. (b) Multiple scattering description of ELNES.

Figure 16 shows background-subtracted EEL spectra recorded from an Fe-doped GaN semiconductor. In this material, depending on the growth temperature used, Fe can either be distributed homogenously in the GaN host lattice or it can accumulate in the form of Fe-N nanocrystals. In the present specimen, Fe-N nanocrystal formation was observed in samples that had been deposited at temperatures higher than 850 °C. Most of the Fe-N nanocrystals were found to be associated with closely adjacent void-like features. EEL spectroscopy (EELS) was used to show that these features are bubbles filled with molecular N_2 (see Fig. 12b). In order to interpret the experimental results, N K edge spectra were calculated for GaN using self-consistent real-space multiple scattering calculations with FEFF 9.05.
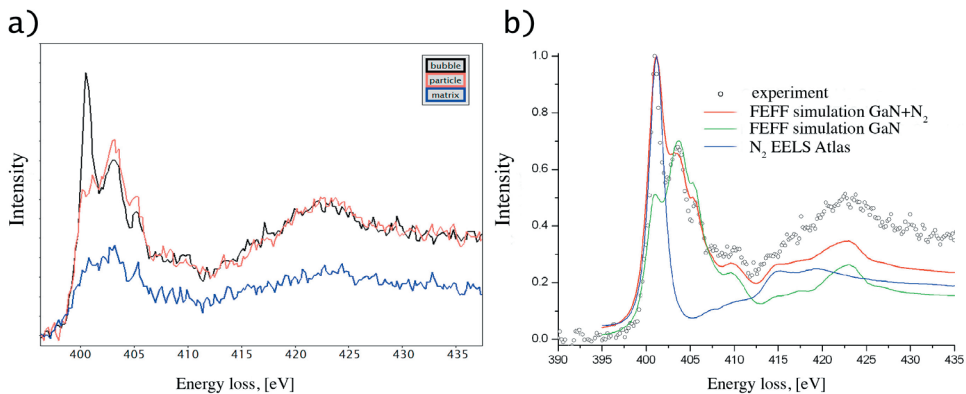
Fig. 16: (a) Background-subtracted N K edge spectra recorded from an Fe-N nanocrystal, a N_2 bubble and a GaN host in Fe-doped GaN. (b) Experimental spectra recorded from a nitrogen bubble alongside an experimental measurement from N_2 taken from the EELS Atlas and simulated spectra calculated for GaN. Figure adapted from Ref. [24].

In Figure 16, the EELS measurement was performed on a single nanocrystal embedded in the GaN host. A 100 kV accelerating voltage and a distributed-dose acquisition routine [28] were used to minimize and control electron beam induced damage during the experiment. Figure 16a shows N K edge spectra recorded from the nanocrystal, the adjacent N_2-containing region and the GaN host. The N K edge shows a three-peaked structure between 400 and 407 eV. Figure 16b shows a spectrum recorded from the bubble, alongside an experimental spectrum from N_2 taken from the *EELS Atlas* [29] and a simulation of a GaN spectrum.

The simultaneous application of high-resolution STEM and EELS can be used to access chemical information about a specimen on the atomic scale. Both the HAADF signal and the EEL signal are recorded at each (x, y) position when a focused electron beam is scanned across the sample. Figures 17-19 illustrate the application of HAADF STEM and EELS to a study of $DyScO_3/SrTiO_3$ interfaces. The atomic structure of a $DyScO_3$ layer in $SrTiO_3$ is revealed in the aberration corrected HAADF STEM image shown in Fig. 17, which was recorded using a probe size of 0.08 nm. The intensity of each atom column reflects differences in atomic number. Here, $DyScO_3$ is imaged along the [101] direction of the orthorhombic unit cell, resulting in a zigzag arrangement of Dy columns parallel to the interface [30]. Sc and Ti atoms are located at the centres of oxygen octahedra (displayed in orange in the structure models) and have weaker contrast when compared to Sr and Dy. Oxygen atoms cannot be resolved.
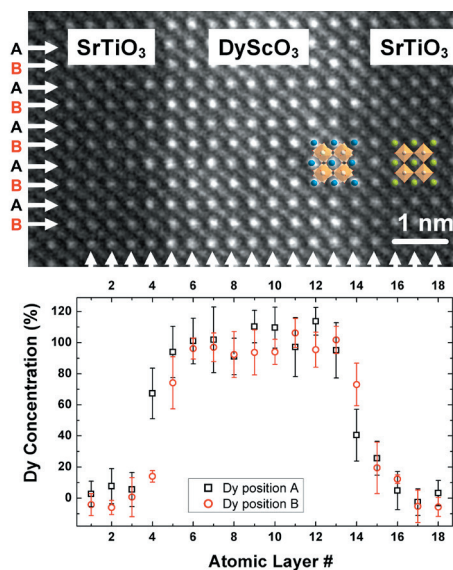
Fig. 17: (Top) HAADF STEM image of a $SrTiO_3/DyScO_3$ multilayer. (Bottom) For each Dy layer, the concentrations at A and B were determined. The interface layers (rows 4 and 14) show differences in contrast between neighbouring positions A and B, suggesting the formation of an ordered interface structure. Figure taken from Ref. [31].

At the interfaces in rows 4 and 14, the contrast at the Dy/Sr positions alternates between neighbouring columns. This behaviour is seen in both the image and the concentration profile. Concentrations were determined by quantifying the intensities using Gaussian fits to the contrast at each atomic column position. In each row, the intensities of five equivalent positions (A and B) were averaged. Error bars denote the standard deviation obtained from concentration

values in regions of constant composition. The Dy concentration was obtained by calculating the integrated intensity to the power of 0.5 and normalizing to 100% within the $DyScO_3$ layer and 0% within the $SrTiO_3$ layer. In row 4 of Fig. 17, the positions labelled "A" are brighter than positions B. The opposite behaviour is seen in row 14, where positions B have the highest intensity, suggesting the formation of an ordered interface structure.

Since Sc and Ti are close in atomic number (21 and 22, respectively), EEL spectra were recorded simultaneously with the HAADF image, in order to assign a spectrum to each atomic layer [32]. Since the edges show only a weak signal, a larger probe size of ~0.2 nm and an acquisition time of 1 s per spectrum were used. Figure 18 shows an HAADF STEM image of $SrTiO_3$/$DyScO_3$/$SrTiO_3$. Figure 19 shows EEL spectra recorded from the layers marked in colour in Fig. 18. The spectra show the Sc and Ti L_{23} edges, which have their onsets at 402 eV and 456 eV, respectively. For both elements, four characteristic lines are present as a result of splitting into the e_g and t_{2g} levels of the L_2 and L_3 edges due to the octahedral crystal field imposed by the O atoms. The Sc signal decreases on moving from the $DyScO_3$ layer into the $SrTiO_3$. The opposite behaviour is seen for the Ti edge. The concentrations of Ti and Sc were determined by extracting the integrated signals of the edges and making use of the fact that the Sc concentration in the $DyScO_3$ is 100% and falls to 0% in the $SrTiO_3$. The Ti and Sc concentrations were determined in this way for each atomic layer and are plotted in Fig. 20, superimposed on the HAADF image of Fig. 18. The Dy concentration is also displayed as an average value for positions A and B in each atomic layer. Neither the Dy nor the Sc concentration changes abruptly at the interface. Instead, intermixing is observed, extending over 2-3 atomic layers. Both Sc and Ti show typical edge shapes for octahedral coordination. The fact that the valence of Ti is close to 4+ for all of the atomic layers can be concluded from the observation that the shapes of the lines do not change. For a Ti^{3+} ion, the line shape would be expected to consist of two broad peaks, instead of the four lines that are observed experimentally. The measured compositions were used to infer the charge distributions in the layers, as well as the fact that a predicted "polar catastrophe" is overcome by adjustment to the chemical compositions of the individual atomic layers.
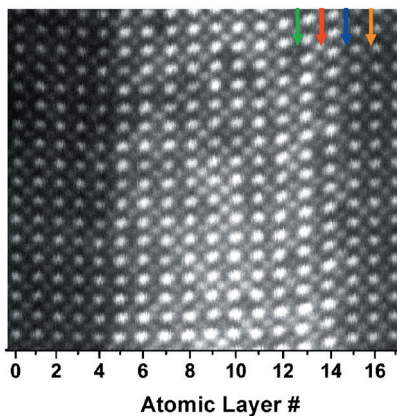
Fig. 18: Z contrast image of a $DyScO_3$/$SrTiO_3$ multilayer recorded in 100 s. Arrows mark layers that correspond to the EEL spectra shown in Fig. 19. Figure taken from Ref. [31].
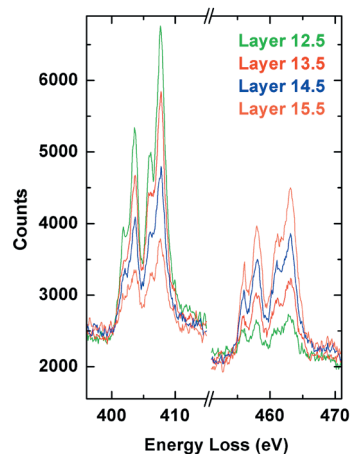
Fig. 19: EEL spectra from layers 12.5 through 15.5, revealing the Sc L_{23} edge at 402 eV and the Ti L_{23} edges at 456 eV. Figure taken from Ref. [31].
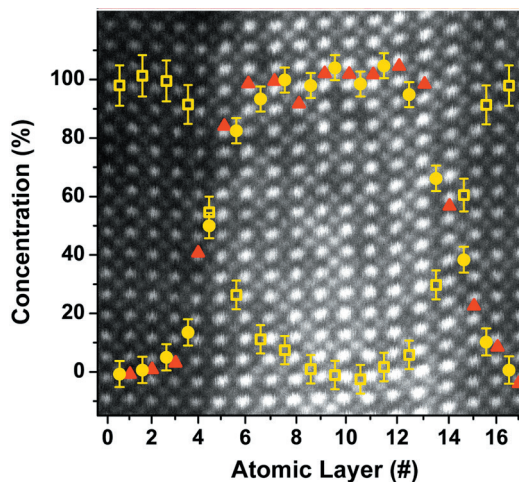
Fig. 20: HAADF STEM image of SrTiO₃/DyScO₃. The fast scan direction runs from bottom to top. Superimposed are the average Dy concentrations as red triangles (average of positions A and B in Fig. 17), as well as Sc (yellow circles) and Ti (yellow open squares) concentrations for each row. Error bars are standard deviations obtained from concentrations in regions of constant composition. Figure taken from Ref. [31].

Figure 21 shows an example of a combined HAADF STEM/ EELS study of a ZnPd/ZnO nanoparticle. ZnPd/ZnO is an outstanding catalyst for methanol steam reforming. Its high catalytic activity has been attributed to the interaction between small ZnO patches on the surface of the particles and ZnPd particles [33]. The ZnPd particle shows non-uniform contrast with dark regions across the particle (Fig. 21a). In order to reveal the nature of these contrast features, an EEL spectrum profile was recorded across the particle along the white arrow. The individual line profiles (Fig. 21b) represent the elemental distribution of Pd, Zn, and O across the particle. The intensity represents the amount of each element in the viewing direction. A comparison between the HAADF micrograph and the EELS data shows that the dark areas (small arrows in Fig. 21a) correspond to a local depletion of Pd, which result in local minima in the Pd distribution at about 20 and 33 nm (local dip in the HAADF curve in Fig. 21b). The amount of Zn shows an increase from the outside of the particles, which is consistent with the geometrical increase in the thickness of the ZnPd particle towards its centre. At about 20 and 33 nm, two additional bumps are visible. These observations show that the local Pd depletions in the dark areas are accompanied by an enrichment in Zn. The O EELS signal is almost zero along most of the scan; only at about 20 and 33 nm are two peaks visible. The results in Fig. 21 indicate that Zn-rich regions in chemically inhomogeneous ZnPd/ZnO methanol steam reforming catalysts, penetrating the particle surface, are capped by ZnO [34].
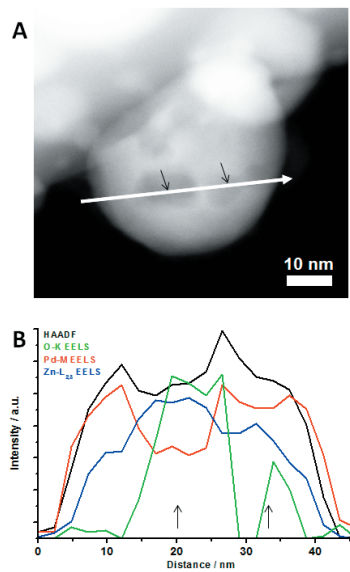
Fig. 21: (a) HAADF STEM micrograph of a ZnPd/ZnO nanoparticle showing non-uniform contrast. An EELS line scan across the particle was performed along the white arrow. (b) EEL spectrum profiles for Pd, Zn and O shown alongside a corresponding HAADF intensity profile. Dark spots (black arrows) in the HAADF image correspond to regions that are Pd-depleted and enriched in Zn. O is only present at the dark spots. Figure taken from Ref. [34].

4   Electron tomography

An important development in electron tomography for the measurement of the three-dimensional morphology, density or composition of an object is illustrated in Figs 22 and 23. The advance involves the development of an improved alignment procedure for a tilt series of recorded images. Previously, algorithms for three-dimensional reconstruction, such as backprojection or algebraic reconstruction, assumed a precise knowledge of the projection geometry and the absence of any relative image displacements prior to reconstruction. The most widely used techniques for image alignment in electron tomography were developed for the biological sciences and are based on pattern matching by correlation techniques [35] or by the tracking of fiducial markers [36, 37]. These methods are useful at low and medium magnification, with automated schemes for alignment facilitating the refinement process [38, 39]. However, cross-correlation and marker based methods have disadvantages. Successive cross-correlation of images in a tilt series can only produce an approximate alignment since three-dimensional rotation around a common axis is not taken into account and the image changes with viewing direction. Alignment errors can accumulate during the procedure [40, 41], resulting in a loss of high-resolution detail in the tomogram. Tracking of marker points in sinogram space does account for three-dimensional rotation, but relies on the presence of object details that serve as a markers or landmarks. Methods that use the reconstruction feedback of the tomogram iteratively to improve image alignment have been proposed and demonstrated in the literature. These methods include manual feedback control, iterative optimization in the form of alignment to pseudo-projections of a tomogram sequence [42], three-dimensional model-based approaches that align with respect to expected projections of automatically or pre-

identified specimen features and area matching as an extension of the cosine stretching used in cross-correlation alignment [43]. An alternative variant of feedback optimization involves the iterative refinement of tomogram contrast and resolution [44] and does not depend on user-assisted motif recognition or the presence of extrinsic markers. Flow charts of the conventional and alternative algorithm are shown in Fig. 22.
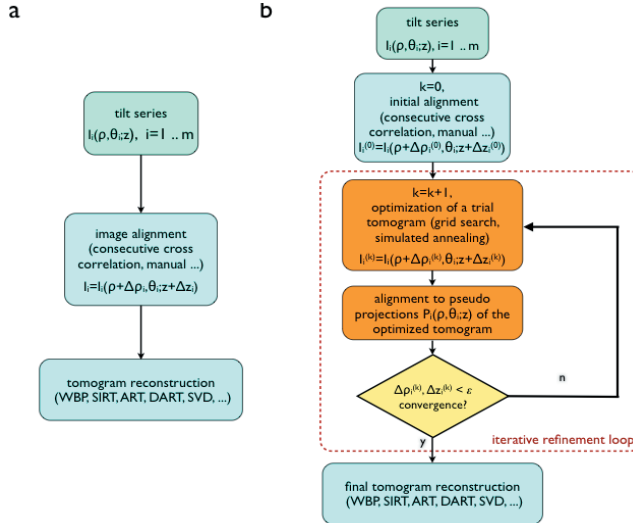
Fig. 22: Flowchart for refinement of image alignment for tomographic reconstruction from a tilt series of images. (a) Conventional procedure involving alignment of images prior to reconstruction. (b) Alternative algorithm. Figure taken from Ref. [44].
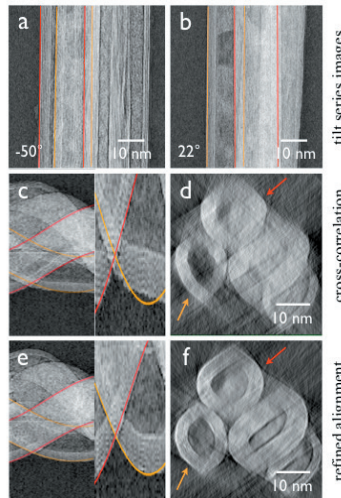
Fig. 23: Comparison of conventional and alternative alignment for reconstruction of a bundle of WS_2 nanotubes. (a) and (b) are images from the tilt series. The lines mark the outer shells of two nanotubes. (c) shows a sinogram (left) after conventional cross-correlation alignment. The magnified lower part reveals a lack of correlation. (d) Tomogram slice showing the bundle in cross-section. (e) and (f) Sinogram and a tomogram slice after refinement of the alignment. Figure taken from Ref. [44].

The optimization loop of the refinement algorithm provides a correct model of the three-dimensional motion of the object and maximizes tomogram resolution. Simulations demonstrate that translational displacement accuracy better than the width of the point spread function in an individual image can be achieved. Experimental tomographic reconstructions at medium resolution show that the marker-free alignment procedure is competitive with fiducial marker alignment and removes artefacts after consecutive cross-correlation alignment [44]. Figure 23 demonstrates the application of the procedure to the reconstruction of WS_2 nanotubes. In this example, no obvious nodal points can be identified and self-markers cannot be utilized. A further complication is that the bundle of nanotubes shows significant motif change in projection during tilt. In such an example, image alignment based on cross-correlation is likely to fail. After refinement, the sinogram traces of the tube shells follow the expected projection and no shape distortion is seen in the tomogram slices, apart from residual missing wedge artefacts.

5 Electron holography of magnetic and electric fields

Off-axis electron holography is a powerful TEM technique that can be used to characterize magnetic fields and electrostatic potentials in materials [45]. It is the only technique that provides direct access to the phase of the electron wave that has passed through a thin specimen, in contrast to TEM techniques that record only spatial distributions of image intensity (described above). The technique can be divided into two primary modes: (i) high-resolution electron holography, in which the interpretable resolution in lattice images is improved by the use of software phase plates to correct for electron microscope lens aberrations [46] and (ii) medium-resolution electron holography, in which magnetic fields and electrostatic potentials within and around materials are studied, usually with nm spatial resolution [47]. Some of the most successful early examples of the application of medium-resolution electron holography were the experimental confirmation of magnetic flux quantization in superconducting toroids and the study of magnetic flux vortices in superconductors [48]. In this section, more recent developments and applications of medium-resolution off-axis electron holography for the study of nanoscale materials are described. A schematic diagram showing the typical electron-optical configuration for off-axis electron holography is shown in Fig. 24a.
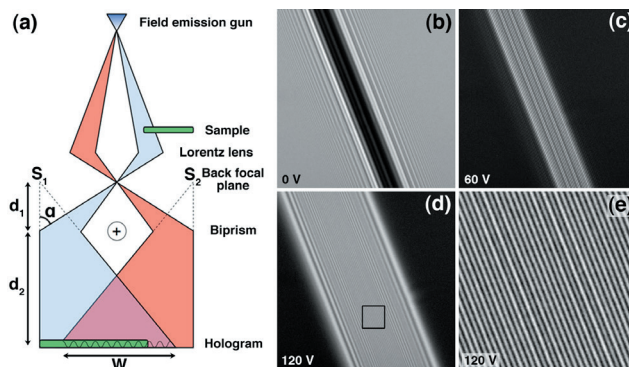
Fig. 24: (a) Schematic diagram of the setup for off-axis electron holography. (b-e) Interference patterns (holograms) recorded from vacuum using biprism voltages of (b) 0, (c) 60 and (d, e) 120 V. In (b), only Fresnel fringes from the edges of the biprism are visible. (e) corresponds to the box marked in (d). The field of view in (b)-(d) is 916 nm.

The region of interest is positioned so that it covers about half the field of view. A biprism (a thin conducting wire), which is usually located close to the first image plane, has a positive voltage applied to it to overlap a "reference" electron wave that passed through vacuum with the electron wave that passed through the specimen. The overlap region contains interference fringes and an image of the specimen, in addition to Fresnel fringes from the edge of the biprism (Figs 24b-24e). The configuration is equivalent to using two electron sources S_1 and S_2.

The intensity in an off-axis electron hologram consists of the reference image intensity, the specimen image intensity and a set of cosinusoidal fringes. A representative hologram of a specimen that contains a chain of magnetite (Fe_3O_4) crystals is shown in Fig. 25a, alongside a vacuum hologram (Fig. 25b) and a magnified region of the specimen hologram (Fig. 25c). In order to extract phase and amplitude information from an off-axis electron hologram, it is Fourier transformed. The Fourier transform contains a peak at the origin corresponding to the Fourier transform of the reference image, a peak at the origin corresponding to the Fourier transform of a bright-field image of the specimen, a peak at $q = -q_c$ corresponding to the Fourier transform of the image wavefunction and a peak at $q = +q_c$ corresponding to the Fourier transform of the complex conjugate of the wavefunction. Figure 25d shows the Fourier transform of the hologram shown in Fig. 25a. In order to recover the complex wavefunction, one of the "sidebands" in the Fourier transform is selected (Fig. 25e) and inverse Fourier transformed. The amplitude and phase are then calculated. The phase image is initially calculated modulo 2π, meaning that 2π discontinuities appear where the phase exceeds this amount (Fig. 25g). The phase image can be "unwrapped" by using suitable algorithms, as shown in Fig. 25h. In Fig. 25d, the streak from the "centreband" towards the sidebands is attributed to the presence of Fresnel fringes from the biprism wire, which can lead to artefacts in the reconstructed amplitude and phase. Its effect can be minimized by masking the streak from the Fourier transform before inverse Fourier transformation (Fig. 25e).
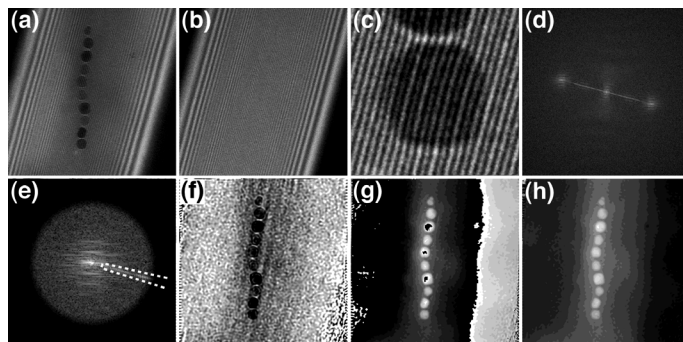
Fig. 25: Processing steps used to convert an electron hologram into an amplitude and phase image. (a) Hologram of a chain of magnetite crystals. The overlap width and holographic interference fringe spacing are 650 and 3.3 nm, respectively. (b) Vacuum hologram. (c) Magnified region of the specimen hologram. (d) Fourier transform of (a). (e) One of the sidebands extracted from the Fourier transform, shown after applying a circular mask with smooth edges. The streak from the Fresnel fringes can be removed by assigning a value of zero to pixels inside the region marked by the dashed line. Inverse Fourier transformation of the sideband provides a complex image wave, which is displayed in the form of (f) an amplitude image and (g) a modulo 2π phase image. Phase unwrapping algorithms can be used to remove the 2π phase discontinuities from (g) to yield the unwrapped phase image shown in (h).

The phase recorded using electron holography is sensitive to both the electrostatic potential and the in-plane component of the magnetic induction in the specimen integrated in the beam direction. Neglecting dynamical diffraction (*i.e.*, assuming that the specimen is thin and weakly diffracting), the phase can be written in the form

$$\phi(x,y) = \phi_e(x,y) + \phi_m(x,y) = C_E \int\limits_{-\infty}^{+\infty} V(x,y,z)\,dz - \frac{e}{\hbar}\int\limits_{-\infty}^{+\infty} A_z(x,y,z)\,dz \quad , \tag{9}$$

where V is the electrostatic potential, A_z is the component of the magnetic vector potential parallel to the electron beam direction,

$$C_E = \left(\frac{2\pi}{\lambda}\right)\left(\frac{E + E_0}{E(E + 2E_0)}\right) \tag{10}$$

is a constant that depends on the accelerating voltage $U=E/e$, λ is the (relativistic) electron wavelength and $E_0 = 511$ keV is the rest mass energy of the electron. C_E takes a value of 6.53×10^6 rad/m^{-1}V^{-1} at an accelerating voltage of 300 kV. If no external charges or applied electric fields are present, then the primary electrostatic contribution to the phase originates from the mean inner potential V_0 coupled with variations in specimen thickness $t(x,y)$. If the specimen has uniform composition, then the electrostatic contribution to the phase is proportional to $t(x,y)$. If $t(x,y)$ or its gradient is known, then φ_e can be used to measure V_0.

The magnetic contribution to the phase carries information about the magnetic flux within and around the specimen. The difference between its value at any two points (x_1,y_1) and (x_2,y_2)

$$\Delta\phi_m = \phi_m(x_1,y_1) - \phi_m(x_2,y_2) = -\frac{e}{\hbar}\int\limits_{-\infty}^{+\infty} A_z(x_1,y_1,z)\,dz + \frac{e}{\hbar}\int\limits_{-\infty}^{+\infty} A_z(x_2,y_2,z)\,dz \tag{11}$$

can be written in the form

$$\Delta\phi_m = -\frac{e}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l} \tag{12}$$

for a rectangular loop formed by two parallel electron trajectories that cross the sample at coordinates (x_1,y_1) and (x_2,y_2) and are joined, at infinity, by segments perpendicular to the trajectories. By virtue of Stokes' theorem,

$$\Delta\phi_m = \frac{e}{\hbar}\iint \mathbf{B}\cdot\hat{\mathbf{n}}\,dS = \frac{\pi}{\varphi_0}\Phi(S) \quad , \tag{13}$$

where $\varphi_0=h/2e=2.07\times10^{-15}$ Tm^2 is a flux quantum. The difference between any two points in a phase image is therefore a measure of the magnetic flux through the region of space bounded by two electron trajectories that crossed the sample at the positions of these two points. A graphical representation of the magnetic flux distribution can be obtained by adding contours to a recorded phase image.

The relationship between the magnetic contribution to the phase and the magnetic induction can be established from the gradient of φ_m. In the special case when stray fields can be neglected, the sample has a constant thickness and the magnetic induction does not vary with z in the specimen, the magnetic contribution to the phase gradient is proportional to the magnetic induction.

When characterizing magnetic materials, the objective lens is usually turned off because it creates a large (>2 T) magnetic field at the position of the specimen. Instead, a non-immersion Lorentz lens is used as the imaging lens. This lens allows the specimen to be imaged at relatively high magnification in magnetic-field-free conditions. The spherical aberration coefficient of this lens is large (*e.g.*, 8400 mm for an FEI Titan), resulting in a point resolution of ~2 nm.

It should be noted that the separation of electrostatic and magnetic contributions to the phase is almost always mandatory in order to obtain quantitative magnetic information from a phase image. The few instances when this step may be avoided include the special case of magnetic domains in a film of constant thickness (far from the specimen edge). The contribution of the mean inner potential to the phase can be much greater than the magnetic contribution for small (sub-50-nm) magnetic nanocrystals. Separation of the magnetic contribution from the recorded phase is then essential. In principle, the most accurate way of achieving this separation involves turning the specimen over after acquiring a hologram and acquiring a second hologram from the same region. The sum and difference of the holograms are then used to determine twice the mean inner potential and twice the magnetic contribution to the phase, respectively. An alternative method, which is often more practical, involves performing a magnetization reversal experiment in the microscope and then selecting pairs of holograms that differ only in the magnetization direction in the specimen. If the magnetization in the specimen does not reverse perfectly, then such reversal measurements may need to be repeated many times so that non-systematic differences between reversed images are averaged out.

Figure 26 shows the application of electron holography to study the current-induced motion of transverse and vortex-type magnetic domain walls in permalloy (NiFe) zigzag lines [49].
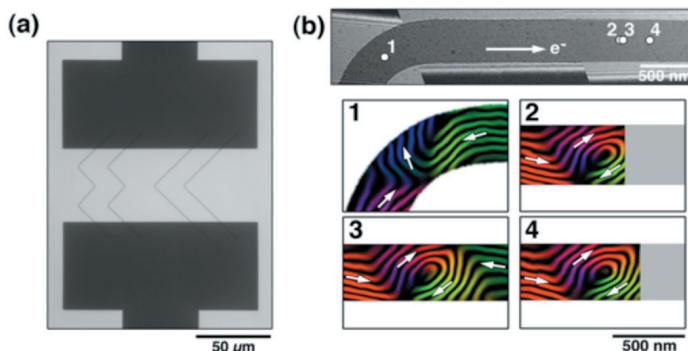
Fig. 26: (a) Bright-field image of permalloy (NiFe) zigzag wires (width 430 nm, thickness 11 nm) on Si_3N_4. (b) Higher magnification image of a wire and magnetic induction maps measured using electron holography showing a magnetic domain wall after the injection of 10 μs current pulses with a current density of 3.14×10^{11} A/m^2. The arrow in the top image indicates the direction of electron flow. The phase contour spacing in the induction maps is 0.785 rads. Figure adapted in part from Ref. [49].

Figure 26b shows the sequential positions of a magnetic domain wall that was subjected to 10 µs pulses with a current density of 3.14×10^{11} A/m^2. Off-axis electron holograms were recorded at each of these positions. A transverse domain wall initially formed at a kinked region of the wire after the application of a magnetic field. After applying a current pulse, the wall moved by 2330 nm in the direction of electron flow and transformed into a vortex-type wall. After a second pulse, the vortex-type wall moved in the same direction and became distorted, with the long axis of the vortex increasingly perpendicular to the wire length. This behaviour may be associated with roughness or defects, which may restrict the movement of the wall. After a third pulse, the domain wall moved 260 nm further and retained its vortex character.

Some low symmetry magnetic materials and thin film magnetic heterostructures exhibit a weak asymmetric exchange coupling, which is termed the Dzyaloshinskii-Moriya interaction (DMI) and is responsible for the formation of *magnetic skyrmions*. Their existence was predicted theoretically in 1989. However experimental proof was not available until 2009, when neutron diffraction and electron microscopy confirmed skyrmion lattice formation in B20-type MnSi. The nature and symmetry of the DMI can affect the spin structure of a skyrmion. In particular, there are both vortex-like skyrmions, in which intermediate spins are aligned tangentially to circular structures (Bloch-type) and hedgehog-like skyrmions, in which the intermediate spins point along a radial direction (Néel-type). However, both structures are topologically equivalent. The recent observation of anti-skyrmions in acentric tetragonal Heusler alloys and skyrmion bubbles in centrosymmetric frustrated magnets extends the family of skyrmions to four members. Figures 27a and 27b show the structure of B20-type FeGe and the magnetic spin arrangement in a Bloch-type skyrmion. Figure 27c shows a temperature *vs* external magnetic field diagram for the FeGe system. Figures 27d and 27e show magnetic phase shift images and corresponding magnetic induction maps of a helical magnetic texture and a skyrmion lattice in FeGe in the presence of a small magnetic field applied using the conventional microscope objective lens. The magnetic phase contribution was obtained by recording phase images below and above the critical temperature of FeGe, where the system is paramagnetic. The pairs of phase images could then be aligned and subtracted from each other on the assumption that the mean inner potential contribution to the phase is the same at both temperatures.

Advances in specimen preparation using focused Ga ion beam milling have allowed tailor-made specimens to be fabricated to study magnetic skyrmions in confined geometries. Figure 28a shows a bright-field TEM image of an FeGe nanostripe embedded in PtC$_x$. The nanostripe was prepared using a lift-out method in a dual beam focused ion beam system with a thickness of 110 nm, a length of 2.6 µm and a width that varied from ~10 nm to ~180 nm. At low temperatures and in the presence of an out-of-plane magnetic field, the helical structure transformed to a single chain of skyrmions. The skyrmions adopted a sequence of compressed, regular and stretched morphologies with increasing nanostripe width, as shown in Fig. 28b. Based on a detailed analysis of the experimental results and on atomistic simulations of the predicted magnetic spin textures, a width-field magnetic phase diagram for chiral magnetic stripes was constructed [50]. Figure 28c shows an FeGe nanorod that was used to prepare a nanodisk (inset) with a diameter of 160 nm to study target skyrmion formation [51]. Theoretical calculations predicted that a target skyrmion is stable in zero magnetic field and that it can exist in two states with different chirality and polarity. Figures 28d and 28e show magnetic induction maps of two such target skyrmion states in an FeGe nanodisk measured using off-axis electron holography at 95 K in zero field.
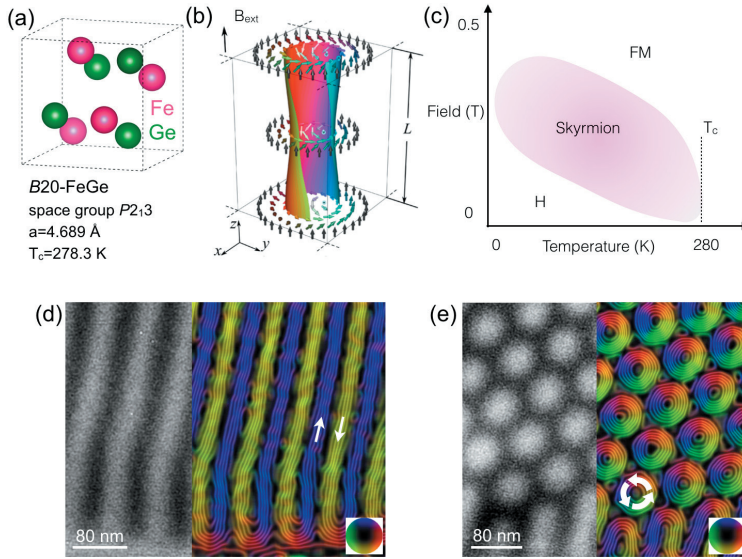
Fig. 27: (a) Unit cell of B20-type FeGe, which has a cubic structure and a magnetic transition temperature of 278.3 K. (b) Spin structure of a Bloch-type skyrmion in a thin film of thickness L in a field B_{ext}. (c) Schematic B-T phase diagram for FeGe. Regions "H", "Skyrmion" and "FM" denote a helical magnetic structure, a skyrmion and a saturated ferromagnetic state, respectively. (d) Magnetic phase image and induction map of a helical magnetic structure in FeGe recorded at 95 K in zero applied field. (e) Magnetic phase shift and magnetic induction map of a skyrmion lattice recorded at 95 K in the presence of a 100 mT out-of-plane field. In each induction map, the phase contour spacing is 0.098 rads. Figure reproduced from the notes for the 48th IFF Spring School "Topological Matter - Topological Insulators, Skyrmions and Majoranas" (2017).
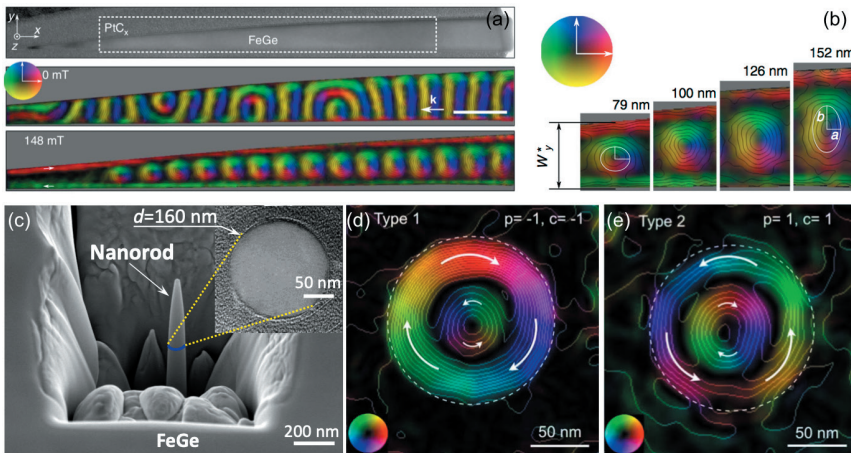
Fig. 28: (a) Helical and skyrmion magnetic states in a wedge-shaped FeGe nanostripe at 220 K in zero field and in a 148 mT out-of-plane field. (b) Enlarged induction maps of skyrmions at different nanostripe widths W_y. (c) Scanning electron microscopy image of the fabrication of an FeGe nanodisk using focused ion beam milling. (d, e) Magnetic induction maps of target skyrmion states in the nanodisk with opposite polarities. Figures (a) and (b) are taken from Ref. [50]. Figures (c)-(e) are taken from Ref. [51].

Recently a model-based method was developed to determine the magnetisation from a recorded magnetic phase image quantitatively [52]. Figures 29a and 29b show a magnetic phase image and a corresponding magnetic induction map of Bloch-type skyrmions in FeGe. The projected in-plane magnetisation was determined using the model-based iteration reconstruction algorithm. The enlarged map of the magnitude of the projected in-plane magnetisation shown in Fig. 29c reveals the hexagonal shape of an individual skyrmion in the lattice arrangement. Figure 29d shows a corresponding vector map of the projected in-plane magnetisation. The measured saturation magnetization is approximately 110 kAm^{-1} at 200 K, which is consistent with the expected value for FeGe for the given temperature [53].
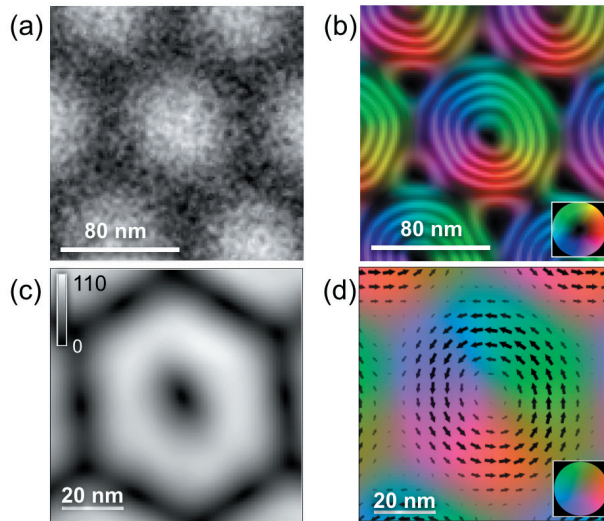
Fig. 29: (a) Magnetic phase image and (b) corresponding magnetic induction map of a skyrmion lattice recorded using off-axis electron holography at 200 K in the presence of a 100 mT out-of-plane magnetic field. The phase contour spacing is 0.098 rads. (c) Projected in-plane magnetisation determined from the magnetic phase image using a model-based iterative reconstruction algorithm. The magnetisation is displayed in units of kAm^{-1}. (d) Colour-coded vector map of the projected in-plane magnetisation. Figure reproduced from Ref. [53].

One of the most important applications of medium-resolution off-axis electron holography is the measurement of electrostatic potentials in semiconductor devices, in order to fulfil the requirements of the semiconductor industry for a quantitative dopant profiling technique that has nm-scale spatial resolution. Such measurements are often made from site-specific regions of semiconductor devices by using focused ion beam milling with Ga ions to prepare the TEM specimen.

Unfortunately, focused ion beam milling can affect the electrical properties of the specimen as a result of the creation of defects that can extend to a depth of more than 100 nm. Approaches such as low energy ion milling can then be used to reduce the thicknesses of the damaged regions, although they do not remove them entirely. The effects of specimen preparation and illumination by the high energy electron beam on the measured electrical properties are presently being addressed by several research groups [*e.g.*, 54].

Figure 30 illustrates the application of off-axis electron holography to the measurement of the electrostatic potential within and around a carbon nanotube that has a voltage applied to it *in situ* in the TEM [55]. Figures 30a and 30b show defocused bright-field images of the specimen, while Fig. 30c shows a phase image of a nanotube recorded using off-axis electron holography. The charge distribution along the nanotube was determined by analysing the gradient of the recorded phase image. In this way, the nanotube in Fig. 30c was inferred to have two different charge densities along its length (2.3 and 4.5 e/nm).
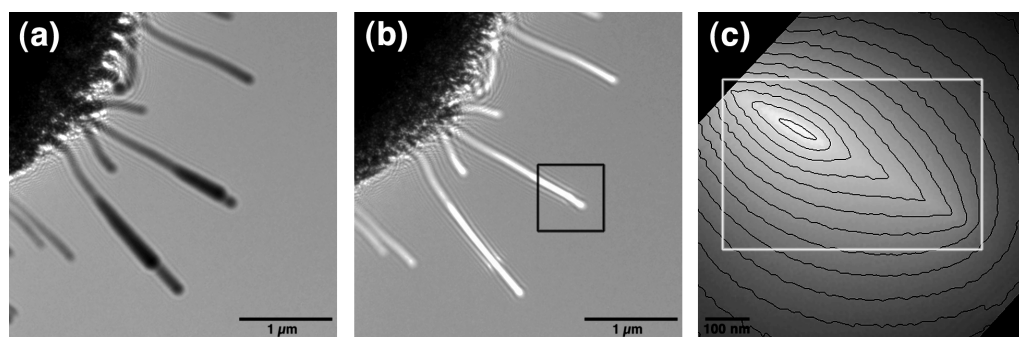
Fig. 30: (a) Underfocus and (b) overfocus bright-field TEM images (defocus ~3 mm) of carbon nanotubes protruding from a nanotube bundle in the presence of an applied bias of 50 V created by a Au electrode (not shown) placed 5 μm away. (c) Phase image recorded using off-axis electron holography from the region marked in (b), with 1.5 rad phase contours superimposed. See text for details. Figure adapted from Ref. [55].

Other recent advances in medium resolution bright-field off-axis electron holography include the development of new approaches for the acquisition, analysis and simulation of electron holograms, the design and use of specimen holders that allow electrical contacts to be applied to devices *in situ* in the electron microscope, and the comparison of recorded magnetic induction maps both with micromagnetic simulations and with three-dimensional information about the local compositions and morphologies of the same specimens acquired using electron tomography. In the future, the development of bright-field off-axis electron holography will require new approaches for the careful separation of weak magnetic signals from unwanted contributions to recorded phase images. Further work is also required to increase the sensitivity of the technique for measuring weak fields and to improve its time resolution. The characterization of magnetic vector fields *inside* nanocrystals in three dimensions by combining electron tomography with electron holography is also of great interest.

The technique of dark-field off-axis electron holography is used to map microstructural deformations in epitaxial samples [56, 57]. It involves the use of an electron biprism to overlap an electron wave that has been diffracted by a crystalline region of interest (the object wave) with an electron wave that has been diffracted by the underlying crystalline substrate (the reference wave), as shown in Fig. 31a. The sample is usually oriented in a two-beam condition to maximise the intensity of the diffracted beam. The incident illumination is tilted so that the diffracted beam travels on the optical axis to minimize the influence of aberrations. The corresponding diffracted spot is selected using a small objective aperture inserted in the back focal plane of the microscope. The biprism is placed so that both the region of interest and part

of the substrate are visible in the hologram. Just as in conventional off-axis electron holography, the hologram is recorded and a phase image is reconstructed using a Fourier-transform-based approach. In addition to the previously described electrostatic and magnetic contributions, the phase contains a term that is reeferred to as the geometric phase ϕ_G, which can be written in the form

$$\phi_G = -2\pi \mathbf{g}_{\mathrm{ref}} . \mathbf{u} \quad , \tag{14}$$

where $\mathbf{g}_{\mathrm{ref}}$ is the reciprocal lattice vector in the reference region (substrate) and $\mathbf{u}$ is the displacement (or translation) vector of the lattice planes [58]. If two orthogonal directions (x, y) are defined, where $\mathbf{g}_{\mathrm{ref}}$ is parallel to the x axis, the pure deformation ε_{xx} and the shear (or rotation) component ε_{xy} can be obtained from the gradients of the displacement in the x and y directions, respectively, according to the expreesions

$$\varepsilon_{xx} = \frac{\partial u_x}{\partial x} ; \varepsilon_{xy} = \frac{\partial u_x}{\partial y} \quad . \tag{15}$$

For quantitative measurements of the deformation, $\mathbf{g}_{\mathrm{ref}}$ must be known, which means that the lattice parameter of the substrate, the Miller indices of the chosen set of lattice planes and the image calibration must be well determined. In order to reconstruct the four components of the 2D deformation matrix, it is necessary to record at least two dark-field off-axis electron holograms using non-colinear $\mathbf{g}$-vectors.
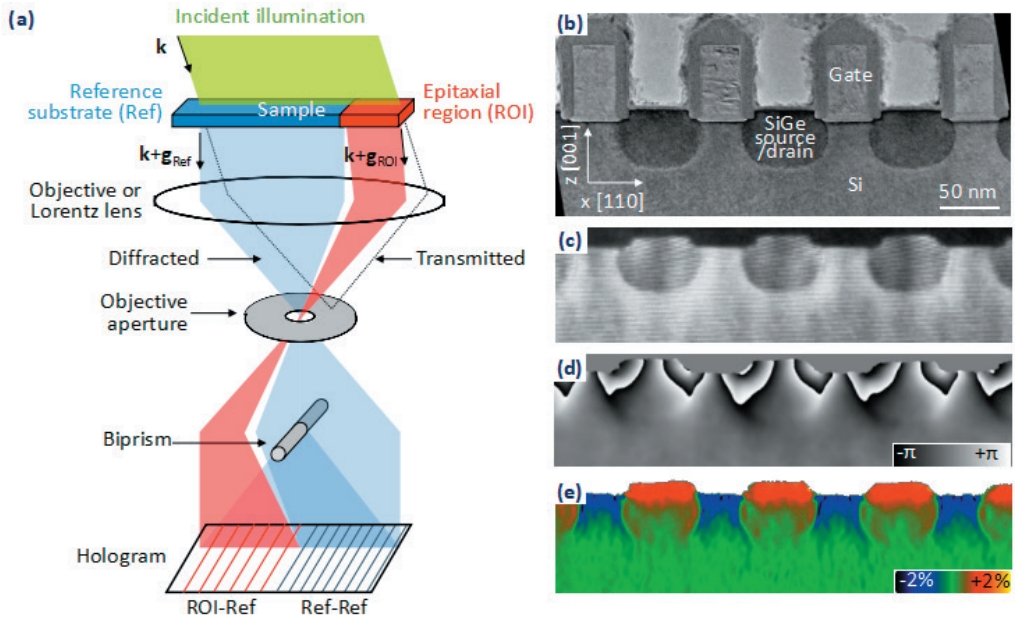
Fig. 31: (a) Schematic diagram of the setup for off-axis dark-field electron holography. (b) TEM image of transistors with recessed SiGe source/drains. (c) (220) dark-field electron hologram. (d) Reconstructed phase image. (e) ε_{xx} horizontal deformation map. Figure adapted from Ref. [65].

Dark-field electron holography is particularly useful in microelectronics, where epitaxial strain is used to control the mobility of charge carriers in semiconducting materials. It has been used to investigate thin SiGe layers [59] and transistor structures [60]. Figures 31b-31e show results obtained from transistor structures with recessed SiGe source/drains [61]. SiGe has a larger lattice parameter than Si and it is grown in the source/drain to compress the Si channel. Figure 31b shows a TEM image of the sample, Fig. 31c is a (220) dark-field electron hologram, Fig. 31d is the reconstructed phase image and Fig. 31^e is the resulting horizontal ε_{xx} deformation map. A positive deformation indicates an elongation of the crystal lattice, while a negative deformation indicates a compression. In the present example, the compression measured in the channel region is about -1%.

Figure 32 illustrates the application of dark-field off-axis electron holography to measure the strain distribution in an InAs quantum dot grown in InP using low pressure metal organic vapour-phase epitaxy at 520 °C. The maximum strain measured in the centre of the dot, relative to the unstrained substrate, is 5.4±0.1%. Comparisons of the strain measurements with computer simulations suggest that the As concentration in the wetting layer is 29% InAs and the composition of the centre of the dot is 100% InAs [62].
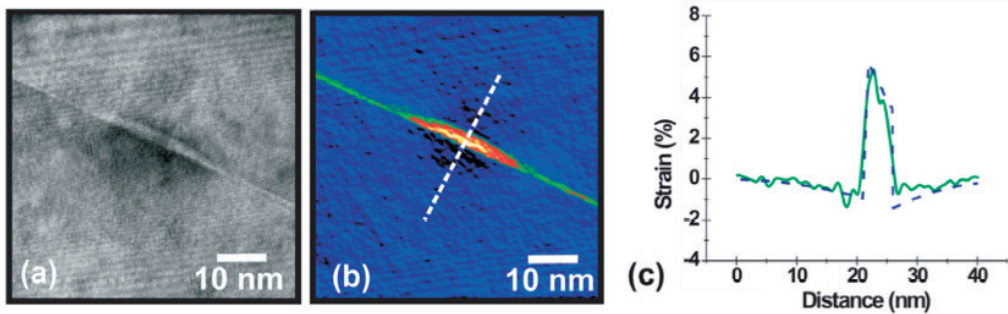
Fig. 32: (a) Dark-field off-axis electron hologram of an InAs quantum dot in InP, with the 004 diffraction spot selected. (b) Strain map for the 004 growth direction derived from (a). (c) Strain profile extracted from the region indicated by the dashed line in (b) (solid line) compared to simulation (dashed line). Figure adapted from Ref. [62].

In order to achieve a precise understanding of such deformation measurements, it may be necessary to study strain relaxation effects induced by thinning of the specimen and to take into account dynamical diffraction [63]. Strain relaxation in thin foils can be simulated using finite element modeling based on anisotropic elastic theory. If the sample exhibits gradients in composition or variations in thickness, then it may be necessary to remove the influence of these non-geometric phase variations before calculating the deformation, for example by subtracting a phase image obtained from a bright-field hologram or using two phase images obtained in dark-field conditions using opposite diffracted beams.

Dark-field electron holography can be operated either in Lorentz mode for a large field-of-view or in norrmal mode for high spatial resolution [64]. Alternative configurations to the off-axis mode have also been developed, such as in-line dark-field electron holography [65], which is based on the analysis of focal series of dark-field images and differential phase contrast dark-field electron holography, which requires the use of a pre-specimen electron biprism [61].

Acknowledgments

References

[1] J.C.H. Spence, *High-Resolution Electron Microscopy* (Oxford University Press, New York, 2003).

[2] L. Reimer, H. Kohl, *Transmission Electron Microscopy: Physics of Image Formation* (Springer, New York, 2008).

[3] D.B. Williams and C.B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science* (Plenum Press, New York, 2009).

[4] A. Thust, in: Vorlesungsmanuskripte des 29. Ferienkurses des Instituts für Festkörperforschung, Schriften des Forschungszentrums Jülich, Reihe Materie und Material, Band 1, 1998.

[5] H. Bethe, Ann. Phys. **87**, 55 (1928).

[6] O. Scherzer, Z. Phys. **101**, 593 (1936).

[7] O. Scherzer, J. Appl. Phys. **20**, 20 (1949).

[8] P. Schiske, Proc. 4th Eur. Conf. on Electron Microscopy, Rome, 145 (1968).

[9] W.M.J. Coene, A. Thust, M. Op de Beeck, D. van Dyck, Ultramicrosc. **64**, 109 (1996).

[10] A. Thust, W.M.J. Coene, M. Op de Beeck, D. van Dyck, Ultramicrosc. **64**, 211 (1996).

[11] C.L. Jia, Phil. Mag. Lett. **79**, 99 (1999).

[12] M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius, K. Urban, Ultramicrosc. **75**, 53 (1998).

[13] M. Lentzen, B. Jahnen, C.L. Jia, A. Thust, K. Tillmann, K. Urban, Ultramicrosc. **92**, 233 (2002).

[14] C.L. Jia, M. Lentzen, K. Urban, Science **299**, 870 (2003).

[15] C.L. Jia, S.B. Mi, K. Urban, I. Vrejoiu, M. Alexe, D. Hesse, Nat. Mater. **7**, 57 (2008).

[16] J. Barthel, A. Thust, Ultramicrosc. **111**, 27 (2010).

[17] L. Jin, J. Barthel, C.L. Jia, K. Urban, Ultramicrosc. **176**, 99 (2017).

[18] M. Andersson, M. Feuerbacher, Ö. Rapp, Phys. Rev. B **78**, 024201 (2008).

[19] J. Dolinšek, J. Slanovec, Z. Jagličić, M. Heggen, S. Balanetskyy, M. Feuerbacher, K. Urban, Phys. Rev. B **77**, 064430 (2008).

[20] S. Roitsch, M. Heggen, M. Lipinska-Chwalek, M. Feuerbacher, Intermetallics **15**, 833 (2007).

[21] M. Feuerbacher, M. Heggen, Phil. Mag. **86**, 935 (2006).

[22] M. Heggen, L. Houben, M. Feuerbacher, Nat. Mater. **9**, 332 (2010).

[23] A. Kovács, J. Sadowski, T. Kasama, M. Duchamp, R.E. Dunin-Borkowski, J. Phys. D: Appl. Phys. **46**, 145309 (2013).

[24] A. Kovács, B. Schaffer, M.S. Moreno, J.R. Jinschek, A.J. Craven, T. Dietl, A. Bonanni, R.E. Dunin-Borkowski, J. Appl. Phys. **114**, 33530 (2013).

[25] M. Lanius, J. Kampmeier, C. Weyrich, S. Kölling, M. Schall, P. Schüffelgen, E. Neumann, M. Luysberg, G. Mussler, P.M. Koenraad, T. Schäpers, D. Grützmacher, Cryst. Growth Des. **16**, 2057 (2016).

[26] M. Eschbach, E. Młyńczak, J. Kellner, J. Kampmeier, M. Lanius, E. Neumann, C. Weyrich, M. Gehlmann, P. Gospodarič, S. Döring, G. Mussler, N. Demarina, M. Luysberg, G. Bihlmayer, T. Schäpers, L. Plucinski, S. Blügel, M. Morgenstern, C.M. Schneider, D. Grützmacher, Nat. Commun. **6**, 8816 (2015).

[27] R.F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Springer, New York, 2011).

[28] K. Sader, B. Schaffer, G. Vaughan, R. Brydson, A. Brown, A. Bleloch, Ultramicrosc. **110**, 998 (2010).

[29] C.C. Ahn, O.L. Krivanek, *EELS Atlas* (Gatan, 1983).

[30] M. Boese, T. Heeg, J. Schubert, M. Luysberg, J. Mater. Sci. **41**, 4434 (2006).

[31] M. Luysberg, M. Heidelmann, L. Houben, M. Boese, T. Heeg, J. Schubert, M. Roeckerath, Acta Mater. **57**, 3192 (2009).

[32] M. Heidelmann, J. Barthel, L. Houben, Ultramicrosc. **109**, 1447 (2009).

[33] M. Friedrich, S. Penner, M. Heggen, M. Armbrüster, Angewandte Chemie Int. Ed. **52**, 4389 (2013).

[34] M. Heggen, S. Penner, M. Friedrich, R.E. Dunin-Borkowski, M. Armbrüster, J.Phys. Chem. C **120**, 10460 (2016).

[35] J. Frank, B.F. McEwan, in: J. Frank (Ed.), *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope* (Plenum Press, New York, 1992).

[36] M.C. Lawrence, in: J. Frank (Ed.), *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope* (Plenum Press, New York, 1992).

[37] D.N. Mastronarde, in: J. Frank (Ed.), *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell* (Springer: New York, 2006).

[38] J.C. Fung, W. Liu, W.J. De Ruijter, H. Chen, C.K. Abbey, J.W. Sedat, D.A. Agard, J. Struct. Biol. **116**, 181 (1996).

[39] S. Brandt, U. Ziese, J. Microsc. **222**, 1 (2006).

[40] W.O. Saxton, W. Baumeister, M. Hahn, Ultramicrosc. **13**, 57 (1984).

[41] J. Frank, B.F. McEwen, M. Radermacher, J. Electron Microsc. Techn. **6**, 193 (1987).

[42] C.H. Owen, W.J. Landis, Ultramicrosc. **63**, 27 (1996).

[43] H. Winkler, K.A. Taylor, Ultramicrosc. **106**, 240 (2006).

[44] L. Houben, M. Bar Sadan, Ultramicrosc. **111**, 1512 (2011).

[45] D. Gabor, Proc. Roy. Soc. A **197**, 454 (1949).

[46] M. Lehmann, H. Lichte, D. Geiger, G. Lang, E. Schweda, Mater. Character. **42**, 249 (1999).

[47] R.E. Dunin-Borkowski, M.R. McCartney, D.J. Smith, in H.S. Nalwa (Ed.), *Encyclopedia of Nanoscience and Nanotechnology*, Vol. 3, pp.41-100 (American Scientific Publishers, Stevenson Ranch, 2004).

[48] T. Matsuda, S. Hasegawa, M. Igarashi, T. Kobayashi, M. Naito, H. Kajiyama, J. Endo, N. Osakabe, A. Tonomura, R. Aoki, Phys. Rev. Lett. **62**, 2519 (1989).

[49] F. Junginger, M. Kläui, D. Backes, U. Rüdiger, T. Kasama, R.E. Dunin-Borkowski, L.J. Heyderman, C.A.F. Vaz, J.A.C. Bland, Appl. Phys. Lett. **88**, 212510 (2007).

[50] C. Jin, Z.-A. Li, A. Kovács, J. Caron, F. Zheng, F.N. Rybakov, N.S. Kiselev, H. Du, S. Blügel, M. Tian, Y. Zhang, M. Farle, R.E. Dunin-Borkowski, Nat. Commun. **8**, 15569 (2017).

[51] F. Zheng, H. Li, S. Wang, D. Song, C. Jin, W. Wei, A. Kovács, J. Zang, M. Tian, Y. Zhang, H. Du, R.E. Dunin-Borkowski, Phys. Rev. Lett. **119**, 197205 (2017).

[52] J. Caron, Ph.D. thesis, Forschungszentrum Jülich, 2017.

[53] A. Kovács, J. Caron, A. Savchenko, N.S. Kiselev, K. Shibata, Z.-A. Li, N. Kanazawa, Y. Tokura, S. Blügel, R.E. Dunin-Borkowski., Appl. Phys. Lett. **111**, 192410 (2017).

[54] D. Cooper, R. Truche, P. Rivallin, J.M. Hartmann, F. Laugier, F. Bertin, A. Chabli, J.L. Rouviere, Appl. Phys. Lett. **91**, 143501 (2007).

[55] M. Beleggia, T. Kasama, R.E. Dunin-Borkowski, S. Hofmann, G. Pozzi, Appl. Phys. Lett. **98**, 243101 (2011).

[56] M. Hÿtch, F. Houdellier, F. Hüe, E. Snoeck, Nature **453** (7198), 1086 (2008).

[57] K.J. Hanszen, J. Phys. D: Appl. Phys. **19**, 373 (1986).

[58] M.J. Hÿtch, F. Houdellier, F. Hüe, E. Snoeck, Ultramicroscopy **111**, 1328 (2011).

[59] T. Denneulin, J.L. Rouvière, A. Béché, M. Py, J.P. Barnes, N. Rochat, J.M. Hartmann, D. Cooper, Semicond. Sci. Technol. **26**, 125010 (2011).

[60] D. Cooper, A. Béché, J.M. Hartmann, V. Carron, J.L. Rouvière, Appl. Phys. Lett. **96**, 113508 (2010).

[61] T. Denneulin, F. Houdellier, M.J. Hÿtch, Ultramicrosc. **160**, 98 (2016).

[62] N. Cherkashin, S. Reboh, M.J. Hÿtch, A. Claverie, V.V. Preobrazhenskii, M.A. Putyato, B.R. Semyagin, V.V. Chaldyshev, Appl. Phys. Lett. **102**, 173115 (2013).

[63] E. Javon, A. Lubk, R. Cours, S. Reboh, N. Cherkashin, F. Houdellier, C. Gatel, M.J. Hÿtch, Ultramicrosc. **147**, 70 (2014).

[64] D. Cooper, J.L. Rouvière, A. Béché, S. Kadkhodazadeh, E.S. Semenova, K. Yvind, R.E. Dunin-Borkowski, Appl. Phys. Lett. **99**, 261911 (2011).

[65] C.T. Koch, V.B. Ozdol and P.A. van Aken, Appl. Phys. Lett. **96**, 091901 (2010).

Index

The IFF Spring School in Jülich, its 50th anniversary and the Jülich Centre for Neutron Science

web: *www.iff-springschool.de*
email: *springschool@fz-juelich.de*
phone: *++49 2461 61-4750*

The annual IFF Spring School has a long-standing tradition. It was established by 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 behaviour of soft matter. Since the restructuring in 2011, research in the area of electronic systems, their phenomena, as well as their applications in information technology, became part of the Peter Grünberg Institute (PGI) named after the IFF scientist who received the Nobel Prize in Physics in 2007. Biophysics and soft matter research are now located at the Institute of Complex Systems (ICS). These institutes are linked together and supported by more method-oriented institutes: The Institute for Advanced Simulation (IAS), which focuses on developing and applying high-performance computing, the Jülich Centre for Neutron Science (JCNS), which operates advanced neutron scattering instruments, and the Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons (ER-C), which hosts some of the most advanced electron microscopes. The IFF Spring School is now organized in turn by one of these institutes. Every year it focuses on a topical subject of condensed matter research, selected from one of the areas of expertise of the above-mentioned institutes.

2019 marks a very special year for the IFF Spring School, as we celebrate its 50th anniversary! This year's school is organized by the Jülich Centre for Neutron Science in cooperation with all the other institutes originating from the former IFF.

The Jülich Centre for Neutron Science uses neutrons as microscopic probe to conduct research on condensed matter and life science systems. To this end, JCNS develops and builds neutron scattering instruments at leading national and international sources. Major outstations exist at the Heinz Maier-Leibnitz Zentrum (MLZ) in Garching, Germany, at the Institute Laue-Langevin (ILL) in Grenoble, France, and at the Spallation Neutron Source (SNS) in Oak Ridge, USA. JCNS operates these instruments for its own research and offers them to a broad national and international user community from science and industry. Research at JCNS has four major thrusts, reflected in the institutes' departments: (i) neutron scattering and soft matter (JCNS-1), (ii) quantum materials and collective phenomena (JCNS-2), (iii) neutron analytics for energy materials (JCNS-3) and (iv) neutron methods (JCNS-4). Through double memberships with ICS (JCNS-1 / ICS-1) and PGI (JCNS-2 / PGI-4), neutrons become available as a powerful probe in the respective institutes. New opportunities for research with neutrons open up at the future European Spallation Source (ESS), where JCNS is heavily engaged. Finally, for the longer-term future of research with neutrons in Germany and Europe, JCNS is developing a novel type of scalable, compact accelerator-driven high brilliance neutron source, the HBS project.

Lecture Notes of the IFF Spring School

Scattering! Soft, Functional and Quantum Materials
50th IFF Spring School **2019**
11.-22.03.2019
M. Angst, T. Brückel, S. Förster, K. Friese, R. Zorn
ISBN: 978-3-95806-380-8

Physics of Life
49th IFF Spring School **2018**
26.02.-09.03.2018
G. Gompper, J. Dhont, J. Elgeti, C. Fahlke, D. Fedosov, S. Förster, P. Lettinga, A. Offenhäusser
ISBN: 978-3-95806-286-3

Topological Matter - Topological Insulators, Skyrmions and Majoranas
48th IFF Spring School **2017**
27.03.-07.04.2017
S. Blügel, Y. Mokrousov, T. Schäpers, Y. Ando
ISBN: 978-3-95806-202-3

Memristive Phenomena - From Fundamental Physics to Neuromorphic Computing
47th IFF Spring School **2016**
22.02.-04.03.2016
R. Waser and M. Wuttig
ISBN: 978-3-95806-091-3

Functional Soft Matter
46th IFF Spring School **2015**
23.02.-06.03.2015
J. Dhont, G. Gompper, G. Meier, D. Richter, G. Vliegenthart and R. Zorn
ISBN: 978-3-89336-999-7

Computing Solids - Models, ab-initio methods and supercomputing
45th IFF Spring School **2014**
10.-21.03.2014
S. Blügel, N. Helbig, V. Meden and D. Wortmann
ISBN: 978-3-89336-912-6

Quantum Information Processing
44th IFF Spring School **2013**
25.02.-08.03.2013
D. DiVincenzo
ISBN: 978-3-89336-833-4

Scattering Methods for Condensed Matter Research: Towards Novel Applications at Future Sources
43rd IFF Spring School **2012**
05.-16.03.2012
M. Angst, T. Brückel, D. Richter and R. Zorn
ISBN: 978-3-89336-759-7

Macromolecular Systems in Soft- and Living-Matter
42nd IFF Spring School **2011**
14.-25.02.2011
J. K.G. Dhont, G. Gompper, P. R. Lang, D. Richter, M. Ripoll, D. Willbold and R. Zorn
ISBN: 978-3-89336-688-0

Electronic Oxides - Correlation Phenomena, Exotic Phases and Novel Functionalities
41st IFF Spring School **2010**
08.-19.03.2010
S. Blügel, T. Brückel, R. Waser and C.M. Schneider
ISBN: 978-3-89336-609-5

Spintronics – From GMR to Quantum Information
40th IFF Spring School **2009**
09.-20.03.2009
S. Blügel, D. Bürgler, M. Morgenstern, C. M. Schneider and R. Waser
ISBN: 978-3-89336-559-3

Soft Matter - From Synthetic to Biological Materials
39th IFF Spring School **2008**
03.-14.03.20108
J.K.G. Dhont, G. Gompper, G. Nägele, D. Richter and R.G. Winkler
ISBN: 978-3-89336-517-3

Lecture Notes of the IFF Spring School

**Probing the Nanoworld -
Microscopies, Scattering and
Spectroscopies of the Solid State**
38th IFF Spring School **2007**
12.-23.03.2007
K. Urban, C. M. Schneider, T. Brückel,
S. Blügel, K. Tillmann, W. Schweika,
M. Lentzen and L. Baumgarten
ISBN: 978-3-89336-462-6

**Computational Condensed Matter
Physics**
37th IFF Spring School **2006**
06.-17.03.2006
S. Blügel, G. Gompper, E. Koch, H.
Müller-Krumbhaar, R. Spatschek and R.
G. Winkler
ISBN: 978-3-89336-430-5

**Magnetism goes Nano - Electron
Correlations, Spin Transport,
Molecular Magnetism**
36th IFF Spring School **2005**
14.-25.02.2005
S. Blügel, T. Brückel and C. M.
Schneider
ISBN: 3-89336-381-5

**Physics meets Biology - From Soft
Matter of Cell Biology**
35th IFF Spring School **2004**
22.03.-02.04.2004
G. Gompper, U. B. Kaupp, J. K. G.
Dhont, D. Richter and R. G. Winkler
ISBN: 3-89336-348-3

Fundamentals of Nanoelectronics
34th IFF Spring School **2003**
10.-21.03.2003
S. Blügel, M. Luysberg, K. Urban and R.
Waser
ISBN: 3-89336-319-X

**Soft Matter - Complex Materials on
Mesoscopic Scales**
33rd IFF Spring School **2002**
04.-15.03.2002
J.K.G. Dhont, G. Gompper and D.
Richter
ISBN: 3-89336-297-5

**Neue Materialien für die
Informationstechnik**
32. IFF-Ferienkurse **2001**
05.-16.03.2001
R. Waser (Editor)
ISBN: 3-89336-279-7

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies

Band / Volume 184
Translation Initiation with 70S Ribosomes: A Single Molecule Study
C. Remes (2018), iv, 113 pp
ISBN: 978-3-95806-358-7

Band / Volume 185
Scanning tunneling potentiometry at nanoscale defects in thin films
F. Lüpke (2018), iv, 144 pp (untersch. Pag.)
ISBN: 978-3-95806-361-7

Band / Volume 186
Inelastic neutron scattering on magnetocaloric compounds
N. Biniskos (2018), iii, 92 pp
ISBN: 978-3-95806-362-4

Band / Volume 187
**Magnetic Order and Excitation in Frustrated
Pyrochlore 5d - Transition Metal Oxides**
E. Feng (2018), iv, 182 pp
ISBN: 978-3-95806-365-5

Band / Volume 188
**Finite-Difference Time-Domain Simulations Assisting to Reconstruct
the Brain's Nerve Fiber Architecture by 3D Polarized Light Imaging**
M. Menzel (2018), ix, 296 pp
ISBN: 978-3-95806-368-6

Band / Volume 189
**Characterization of the cell-substrate interface
using surface plasmon resonance microscopy**
E. M. Kreysing (2018), xiii, 260 pp
ISBN: 978-3-95806-369-3

Band / Volume 190
Scattering! Soft, Functional and Quantum Materials
Lecture Notes of the 50th IFF Spring School 2019
11 – 22 March 2019, Jülich, Germany
ed. by M. Angst, T. Brückel, S. Förster, K. Friese, R. Zorn (2019),
ca 1000 pp
ISBN: 978-3-95806-380-8

Weitere *Schriften des Verlags im Forschungszentrum Jülich* unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp

JÜLICH
Forschungszentrum