



universität  
wien

# DIPLOMARBEIT

Titel der Diplomarbeit

Electron momentum reconstruction in ACPAR

Verfasser

Michael Leitner

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Garching, 2013

Studienkennzahl lt. Studienblatt: A 405

Studienrichtung lt. Studienblatt: Mathematik

Betreuer: Univ.-Prof. Dipl.-Ing. Dr. Otmar Scherzer



# Contents

<b>1. Introduction</b>	<b>1</b>
<b>2. Physical background</b>	<b>3</b>
2.1. Angular Correlation of Positron Annihilation Radiation (ACPAR) . . . . .	3
2.1.1. History of the positron, antimatter and annihilation . . . . .	3
2.1.2. The principle of ACPAR and its implementations . . . . .	4
2.1.3. The analogy to Positron Emission Tomography (PET) . . . . .	6
2.2. The two-photon momentum density . . . . .	7
2.2.1. Connection to the experimental data . . . . .	7
2.2.2. Theoretical interpretation . . . . .	8
2.3. The crystalline state . . . . .	9
2.3.1. Crystalline matter . . . . .	9
2.3.2. Point symmetries . . . . .	10
2.3.3. Translation symmetries — the lattice . . . . .	10
2.3.4. The reciprocal lattice . . . . .	11
2.4. Microscopic principles of the electronic structure . . . . .	12
2.4.1. The Born-Oppenheimer approximation . . . . .	12
2.4.2. The wave function . . . . .	13
2.4.3. The Schrödinger equation . . . . .	14
2.4.4. Single-particle approximations . . . . .	15
2.4.5. Consequences of the crystalline state . . . . .	16
2.4.6. Occupations . . . . .	17
2.4.7. The Fermi surface . . . . .	18
2.4.8. The many-body problem . . . . .	18
2.4.9. Momentum densities . . . . .	19
2.5. Formulation of the problem . . . . .	20
<b>3. Mathematical concepts</b>	<b>21</b>
3.1. Probabilities and statistics . . . . .	21
3.1.1. Notions of probability . . . . .	21
3.1.2. Approximate solutions and their probabilistic interpretation . . . . .	22
3.1.3. Heuristic regularization . . . . .	23
3.1.4. Bayesian interpretation of regularization . . . . .	24
3.1.5. Sampling posterior distributions . . . . .	26
3.2. Denoising and deblurring . . . . .	27

3.3.	Reconstructing 3D-densities from projections . . . . .	27
3.3.1.	Direct transform methods — The Radon transform and its inversion . . . . .	28
3.3.2.	Algebraic reconstruction techniques . . . . .	30
3.3.3.	Cormack’s method . . . . .	31
3.4.	Segmentation . . . . .	31
3.4.1.	Description of shapes . . . . .	32
3.4.2.	Driving the agreement — cost functionals . . . . .	33
3.4.3.	Shape determination . . . . .	34
<b>4.</b>	<b>ACPAR data interpretation</b>	<b>35</b>
4.1.	Discussion of previous approaches . . . . .	35
4.1.1.	Data prettification . . . . .	35
4.1.2.	Interpretation of spectra . . . . .	36
4.1.3.	Reconstruction . . . . .	37
4.1.4.	Fermi surface determination . . . . .	38
4.2.	The proposed model . . . . .	39
4.3.	Implementation . . . . .	41
4.3.1.	Scales of the reconstruction problem . . . . .	41
4.3.2.	Matrix product form of forward operator . . . . .	42
4.3.3.	Describing the Fermi surface and determining band occupations . . . . .	43
4.3.4.	The projection matrices . . . . .	44
4.3.5.	Solution of the linearized subproblem . . . . .	45
4.3.6.	Sampling the Fermi surface . . . . .	46
4.3.7.	Further sophistications . . . . .	46
<b>5.</b>	<b>Results</b>	<b>49</b>
5.1.	Simulated data . . . . .	49
5.2.	General remarks on the reconstructions . . . . .	51
5.3.	Influence of model aspects . . . . .	53
5.4.	Effect of experimental uncertainties . . . . .	53
5.5.	Implications for the experiment . . . . .	54
5.6.	Conclusions . . . . .	55
<b>A.</b>	<b>Appendix</b>	<b>57</b>
A.1.	Diverse small OCTAVE functions . . . . .	57
A.2.	The Gilat-Raubenheimer method . . . . .	58
A.3.	Sparse-to-full matrix multiplication . . . . .	60
	<b>Bibliography</b>	<b>63</b>
	<b>Curriculum vitae</b>	<b>69</b>

## Abstract

This work deals with the mathematical formulation of data interpretation in connection with Angular Correlation of Positron Annihilation Radiation (ACPAR), specifically the determination of the Fermi surface, which shows up as a step in the underlying electron momentum density.

The specific aspects of the problem, i.e. noise reduction and deblurring of the resolution, reconstruction of the three-dimensional density and determination of the location of the step are instances of the general class of inverse problems. The goal of this work is to formulate and implement a model that, starting from controlled assumptions in the Bayesian sense, obtains results from given data that are both statistically optimally defined and systematically as accurate as possible.

Starting with a review of the physical fundamentals of the problem, the mathematical concepts relevant for the specific parts of the problem are presented. After a discussion of the weak points of the previous approaches a model is proposed, where a Fourier parametrization of the conduction band energy takes the role of a level set function corresponding to the Fermi surface, while the momentum density is modelled as a free function defined point-wise, but subject to an explicit regularization via its second derivative. An efficient implementation of this model, which generates samples of realizations of the Fermi surface in agreement with the experimental data, is discussed thoroughly. Finally its performance is demonstrated by way of realistic simulated data.

## Zusammenfassung

Diese Arbeit befasst sich mit der mathematischen Formulierung der Interpretation von Daten gewonnen mittels Winkelkorrelation der Positronenannihilationsstrahlung (ACPAR), speziell der Bestimmung der Fermifläche, welche als Stufe in der zugrundeliegenden Elektronenimpulsdichte aufscheint.

Die spezifischen Aspekte des Problems, nämlich Unterdrücken von Rauschen und Kompensieren endlicher Auflösung der Daten, das Rekonstruieren der dreidimensionalen Dichte, und die Bestimmung des Ortes der Stufe gehören zu der allgemeinen Klasse der inversen Probleme. Das Ziel dieser Arbeit ist die Formulierung und Implementierung eines Modells, welches ausgehend von kontrollierten Annahmen (im Bayesischen Sinne) aus gegebenen Daten sowohl statistisch möglichst gut definierte als auch systematisch möglichst akkurate Resultate gewinnt.

Beginnend mit einem Überblick über die physikalischen Grundlagen des Problems werden jeweils die für spezifische Aspekte des Problems relevanten mathematischen Konzepte präsentiert. Aufbauend auf einer Diskussion der Schwachstellen der bisherigen Zugänge wird ein Modell vorgeschlagen, in dem eine Fourier-Parametrisierung der Leitungsbandenergie die Rolle einer Level Set Funktion für die Fermifläche übernimmt, während die Impulsdichte als punktweise definierte freie Funktion modelliert wird, die einer expliziten Regularisierung ihrer zweiten Ableitung unterworfen ist. Eine effiziente Implementierung dieses Modells, das Stichproben von Realisierungen der Fermifläche in Einklang mit den experimentellen Daten generiert, wird eingehend diskutiert. Schließlich wird seine Leistung anhand von realistischen simulierten Daten belegt.

## Acknowledgements

The work described in this thesis has been performed using exclusively free software, and this also holds for its writing. I therefore want to thank here all the people whose idealism, creativity and hard work have made this possible. I want to mention especially Richard Stallman, due to whose initiative we now have complete free operating systems and who dedicated most of his life to the cause. For being able to typeset this thesis I want to thank Donald Knuth and all the contributors, where I want to explicitly mention Till Tantau, responsible for the beamer package and PGF/TikZ, which make it a pleasure to generate nice-looking presentations and graphics. My work presented here owes also a lot to the people behind GNU OCTAVE, and to a lesser extent Inkscape and POV-Ray.

On the social side I want to thank Hubert Ceeh and Josef-Andreas Weber, my group colleagues here in Garching that are the main responsables for the setting-up and operation of the ACPAR spectrometer. It is really a pleasure to discuss physics and experimental methods with you. Also I want to thank Thomas Widlak, after having lost sight of each other after our joint studies, I learned that you started your diploma and later doctoral studies, upon which we came into contact again, and from this I got the idea where do my diploma thesis. I also want to thank Otmar Scherzer for evaluating my thesis, given the problematic situation of the geographic separation.

Personally, I owe a lot to my wife Alice. Without her setting the example and insisting I would perhaps never have decided to finally finish my mathematics studies, and apart from that, the last months demanded much from both of us.

# 1. Introduction

Essentially all properties of matter relevant to everyday life depend on the material's electronic structure.<sup>1</sup> This is because the outer region of atoms, which defines how and which kinds of atoms can come together to constitute matter, is made up of the electronic "cloud". As a consequence, chemistry as the science of matter is to a large part occupied with understanding the behaviour of the electronic structure of atoms and explaining it in terms of fundamental concepts. Also many of the aspects of matter belonging to the realm of physics, such as magnetism or visual appearance, are due to the electronic structure, because common external perturbations (in the aforementioned examples a magnetic field or incident light) couple primarily to the electrons. Even the probably most primitive property of matter, its stability and how it reacts to an external force, is defined by the electronic structure acting as a potential between the atomic nuclei.

The appropriate description of the behaviour of electrons in matter under ordinary conditions is given by quantum mechanics. This set of rules is not too difficult to understand for a natural scientist, whereas its exact solution is extremely difficult (read: impossible) for any system involving more than a few electrons. However, approximate theories have proven to be of much use, as they both allow a solution (sometimes analytical, but mostly numerical) in terms of understandable concepts and are very successful in explaining and predicting most of the relevant material properties. Among these concepts are the single-particle states (indexed by a position in three-dimensional reciprocal space) and the two-dimensional Fermi surface, which separates the occupied states from the unoccupied ones.

The degree of usefulness of the Fermi surface is surprising: its existence distinguishes metals (with Fermi surface) from semiconductors or insulators (without Fermi surface), its area determines the value of several properties characteristic for metals (such as high heat or electric conductance), and a high density of states at the Fermi level in a would-be paramagnetic metal can lead to ferromagnetism. For alloys, the volume enclosed within the Fermi surface is determined by the composition, which suggests that small changes in composition can lead to large effects for the material properties if they correspond to changes in the topology of the Fermi surface.

Even though the importance of the Fermi surface is generally accepted and it is widely used for the interpretation of theoretical calculations of the electronic structure of a given system, direct experimental measurements are comparatively scarce. This is because of the difficulties associated with the various experimental techniques: the classical quantum oscillatory methods that utilize for instance the de Haas-van Alphen or the Shubnikov-de Haas effects, need high external magnetic fields, low temperatures and well-ordered samples, whereas the newer method of angle-resolved photo-emission spectroscopy (ARPES) is extremely surface-sensitive, which

---

<sup>1</sup>The exception here are nuclear properties such as radioactivity.

## 1. Introduction

casts doubt on the claim of its results to be representative of the bulk behaviour. The third possibility, Angular Correlation of Positron Annihilation Radiation (ACPAR), does not have these limitations, but suffers from the fact that a number of effects contribute to the signal, which have to be considered for deducing the Fermi surface. The method could therefore benefit greatly from a demonstration of a reliable approach for the extraction of the desired quantities from the signal, which would make it a very versatile technique for studies of the electronic structure, in particular of the Fermi surface, for a wide range of systems. This is the goal of this thesis.

From the point of view of applied mathematics, the steps of the solution of this problem belong to various subfields of *imaging*: Conceptually, in a first step the raw experimental data could be computationally processed to obtain two-dimensional intensity images. These images, which are projections of an underlying three-dimensional density distribution, would then be used to reconstruct the density by tomographic techniques, from which the Fermi surface can be determined via segmentation methods. At all stages of this inverse problem, the experimental resolution and data noise have to be considered. Actually, however, it is beneficial to treat all these requirements in an integral way by a formulation as a general Bayesian inverse problem, as will be shown during the course of this thesis.

This is the structure: First, the physical problem will be presented, beginning with the experiment and leading to the explanation of the physical laws giving rise to the observed phenomena. In the next chapter the mathematical side of the problem will be treated. For the general tasks to be done (tomographic reconstruction, segmentation, deblurring, noise reduction) a wide range of solutions have been employed, which will be discussed in relation to the specific aspects of the physical problem. Then previous approaches to the solution of the problem will be reviewed, and a new model will be proposed. Lastly, its performance will be evaluated by applying it to simulated data.

## 2. Physical background

In this chapter I will present the physical fundamentals necessary for an understanding of the problem. In contrast to most other such expositions, I chose here to proceed from the special to the general. To be specific, I will first outline the experimental principles, then give a short motivation of the effects to be observed, and finally present the relevant principles of solid state physics and derive the resulting features from first principles. The reason for this is that quantum mechanics, which is the theory appropriate for the description of the electronic structure in materials, is, although second nature to solid state physicists, possibly a rather uncomfortable terrain for non-physicists. With the present progression I hope to be able to at least communicate the experimental principle, which is easy enough to understand. Seen from another side, the chosen presentation is in line with the experimental physicist's way of work: devising an experimental method, developing a feeling for the observed results, and finally (hopefully!) being able to rationalize them by fundamental laws. I will end this chapter by a statement of the problem to be solved in this thesis in physical terms.

### 2.1. Angular Correlation of Positron Annihilation Radiation (ACPAR)

#### 2.1.1. History of the positron, antimatter and annihilation

The modern concept of antimatter initially started out as an incidental consequence of the unification of quantum mechanics and special relativity for explaining the electron spin by Dirac (1928): the equation describing the electron admitted an additional solution corresponding to a particle behaving as if it had a positive charge, but a negative energy (which runs counter to the fundamentals of physics). In a subsequent publication, Dirac (1930) elaborated on the issue and put forward the hypothesis that nearly all of these negative-energy states should be considered as occupied (and therefore unreachable), while the non-occupied states (holes) of negative energy can be considered as particles of positive energy (and positive charge), rectifying the problem of negative energies. He also tentatively identified these positive-charge solutions of the fundamental equation with the proton, the only particle of positive charge known then, and concluded that if a (positive energy) electron falls into such a hole, both the electron and the proton disappear, emitting the energy equivalent of the annihilated masses as radiation. Dirac was awarded the 1933 Nobel Prize in Physics for the breakthrough enabled by his equation.

However, it soon became apparent that under this assumption the possibility of the annihilation of electrons together with protons would preclude the existence of stable matter on relevant timescales (Oppenheimer, 1930a,b; Tamm, 1930). Also the differences in the electron and proton mass remained unaccounted for. Therefore Dirac (1931) was lead to the postulation of a new

## 2. Physical background

kind of particle with the same mass as the electron, but opposite charge, calling it anti-electron. This approach has proven successful so that in today's standard model of particle physics every particle has an anti-particle (or can annihilate with itself and is therefore its own anti-particle).

Independent of the theoretical prediction, signatures of a particle of positive charge and a mass in the range of the electron mass were found in cloud chamber photographs by Anderson (1932, 1933). He called this new particle positron (a contraction of "positive electron"<sup>1</sup>) in the expectation that the properties of this new particle would be exactly those of the electron only differing in the sign of the charge, which is nowadays in principle accepted. Anderson was awarded the 1936 Nobel Prize in Physics for this discovery. Klemperer (1934) demonstrated the experimental proof that the majority of positrons actually annihilate with the emission of two photons, each having about the energy equivalent of the electron mass.

### 2.1.2. The principle of ACPAR and its implementations

The principle of ACPAR is to probe the electronic structure of a system by implanting positrons into the sample and detecting the emitted annihilation photons in coincidence mode. Specifically, due to energy and momentum conservation, the sum of the momenta of the photons resulting from a given annihilation event is equal to the sum of the momenta of the electron-positron pair before annihilation. As follows from quantum electrodynamics (see, e.g., Berestetskii, Lifshitz, and Pitaevskii, 1982), the most probable annihilation channel is via the emission of two photons. Therefore experimentally two detectors facing the sample from opposite sides are operated in coincidence mode, i.e. detected photons are only counted if both detectors respond within a pre-defined small time window (on the order of 100 ns). Each of these pairs of photon detection events is then assumed to be due to a single annihilation event. As the single detector count rates are typically only on the order of  $10^4 \text{ s}^{-1}$ , an additional benefit is that this coincidence criterion also reduces the background to a negligible level. The primary experimental data are therefore the so-called two-photon momentum density (or rather projections of it), which is the probability density of the sum of the photon momenta and incorporates information about the electronic structure of the system, as will be detailed in Sect. 2.4.9. In principle the experiment could yield even more information by also recording the summed energy for each photon pair, but today's detection technology cannot reach the necessary energy resolution of tens of eV in the relevant regions of 500 keV for resolving the different orbitals, let alone with position-sensitive detectors which are necessary for defining the momenta.

The first measurement according to this principle was performed by Beringer and Montgomery (1942). In this study it could only be concluded that in general the angular deviations are smaller than one degree because of the poor resolution and efficiency due to the use of detectors without position resolution with an aspect ratio near one. Later DeBenedetti et al. (1950) recognized that by employing slits the resolution in one dimension can be enhanced at the expense of the second dimension and that therefore the measured quantity corresponds to the mathematically well-defined projection of the spherically averaged (due to the polycrystalline nature of the sample) three-dimensional two-photon momentum density down to one dimension (the longitudinal component is the second undefined coordinate), at the same time giving higher

---

<sup>1</sup>His calling the ordinary electron "negatron" did not take hold, however.

count rates due to the larger active area. This already enabled valuable conclusions on the electronic structure to be drawn, even more so when using single crystals (Berko and Plaskett, 1958).

The method reached its present level of sophistication with the introduction of two-dimensional arrays of detectors by Berko, Haghgooie, and Mader (1977) followed by position-sensitive detectors (West, Mayers, and Walters, 1981; Bisson et al., 1982), so that nowadays one directly obtains the two-dimensional projection of the two-photon momentum density. For the future there are plans to use position-sensitive detectors that have also some energy resolution to be able to resolve also the third (longitudinal) component via the Doppler shift, although the achievable resolution will be below what is possible with contemporary two-dimensional detectors.

The principles of a two-dimensional ACPAR experiment are illustrated in Fig. 2.1: Due to the small active sample volume the position of detection defines the respective photon propagation directions (and therefore the directions of the momentum vectors), and a deviation from collinear propagation directions is due to a transversal component of the sum momentum, as will be derived in Sect. 2.2. Essentially, summing the vectors of the detection positions  $x_1$  and  $x_2$  on the respective detectors (counted from the centre) gives a signal that is proportional to the transversal part of the sum momentum. These signals are histogrammed over two dimensions

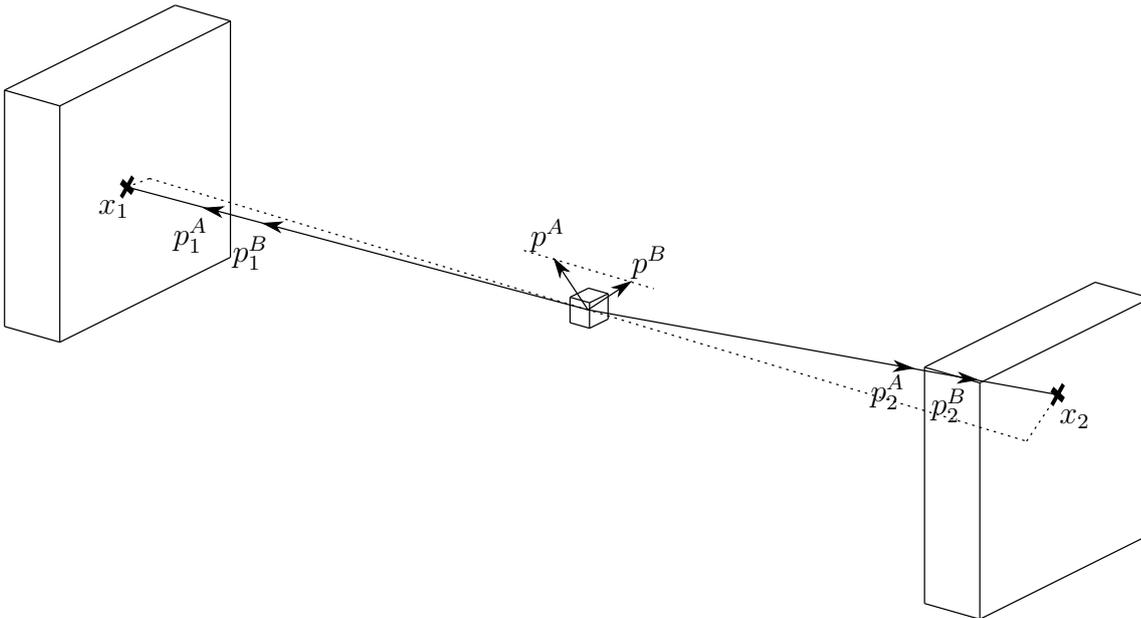


Figure 2.1.: Principles of two-dimensional ACPAR: Coincident photons have been detected at  $x_1$  and  $x_2$ , respectively, which defines the photon propagation directions (neglecting the finite sample volume). The respective longitudinal momentum components (corresponding to the energies) cannot be resolved, therefore the longitudinal component of the sum momentum  $p^X = p_1^X + p_2^X$  is unspecified (illustrated here for the experimentally indistinguishable cases  $A$  and  $B$ ).

## 2. Physical background

and give the primary experimental data, the so-called “projections”.

The following numbers will give a feeling for the relevant scales (widths and resolutions are to be understood as standard deviations of a Gaussian distribution): For a positron beam directed onto the sample from above (as will in the following always be assumed) the active sample volume and therefore the resolution loss in the vertical dimension is negligible due to the small positron penetration lengths in dense matter. In the horizontal dimension it is given by the diameter of the beam, which is on the order of 1 mm. Typical values for the deviation from collinearity  $\alpha$  are 5 mrad. With a detector resolution on the order of 2 mm, a detector-detector-distance of at least 20 m is desirable for resolving the effects. With a typical detector size of 50 cm it follows that the experimental “projections” are indeed to a good approximation the orthogonal projections of the three-dimensional two-photon momentum density into the detector plane.

### 2.1.3. The analogy to Positron Emission Tomography (PET)

Positron Emission Tomography is a classical method in radiomedicine. Due to its practical importance it has been a subject of great importance for applied mathematics (Wernick and Aarsvold, 2004). It is very much related to 2D-ACPAR both in its physical principles and the

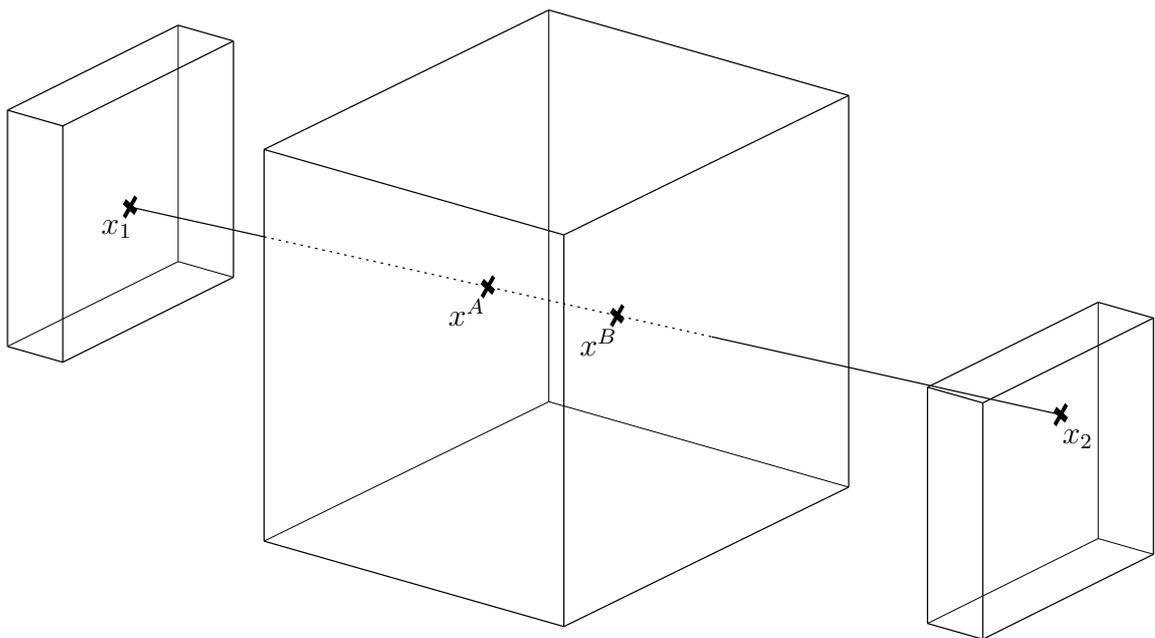


Figure 2.2.: Principles of PET: Coincident photons have been detected at  $x_1$  and  $x_2$ , therefore the annihilation event happened on the connecting line (neglecting the small deviation from collinearity). Two possible annihilation sites  $x^A$  and  $x^B$  are illustrated. In modern implementations, the difference in the photon arrival times is used as additional information on the annihilation position along the connecting line, although with worse resolution compared to the transversal components.

resulting mathematical problem, so I will work out this analogy here.

The illustration is given in Fig. 2.2. The experimental set-up is exactly the same as for ACPAR, only the scales differ: the sample is much larger, and the detectors are as near as possible to the sample. The reason is that while in ACPAR one wants to be able to neglect the influence of the annihilation position to detect the deviation from collinearity, here one wants to neglect the deviation from collinearity in order to determine the annihilation position. Mathematically, these two concepts correspond to a formulation in terms of position (called real space) or directions and wavenumbers, that is in terms of Fourier components (reciprocal space in the physical jargon). The two problems are surprisingly equivalent: in both cases one measures projections of some sort, either because of insufficient energy or time resolution. For PET the mathematical problem is a bit more laborious than for ACPAR, because due to the different scales the projection lines cannot be considered as parallel any more and therefore the experimental data have to be described as higher-dimensional densities (actually, in real implementations one does not use two opposing detectors but a closed ring of detectors). Also, due to the large samples, the in-sample attenuation has to be considered.

## 2.2. The two-photon momentum density

### 2.2.1. Connection to the experimental data

Here the derivation of the connection between the photons' deviation from collinearity and the transversal components of the momentum sum is given.

Let  $\vec{p}_1$  and  $\vec{p}_2$  be the momenta of the two photons coming from a single annihilation event. Under the assumption of a point source, the photon directions (which are the directions of the respective momentum vectors) can be inferred from the impinging positions on the detectors, while contemporary detectors cannot resolve the photon energies (which correspond to the length of the momentum vectors). For simplicity, the coordinate system is chosen such that the propagation directions of the photons lie in the  $x$ - $y$ -plane, symmetric to the  $y$ -axis. The photon momenta are therefore given by

$$\vec{p}_1 = p_1 \begin{pmatrix} \cos \alpha/2 \\ \sin \alpha/2 \end{pmatrix} \quad \text{and} \quad \vec{p}_2 = p_2 \begin{pmatrix} -\cos \alpha/2 \\ \sin \alpha/2 \end{pmatrix}, \quad (2.2.1)$$

with  $\alpha$  the angular deviation from anti-parallel directions. The components of the momentum sum follow as

$$p_x^{2\gamma} = (p_1 - p_2) \cos \alpha/2 \quad (2.2.2)$$

and

$$p_y^{2\gamma} = (p_1 + p_2) \sin \alpha/2. \quad (2.2.3)$$

The relativistic expression for a particle's total energy (Jackson, 1999) is

$$E^2 = (mc^2)^2 + (pc)^2, \quad (2.2.4)$$

where  $m$  is the particle mass (which can be zero such as for photons),  $p$  is the particle momentum, and  $c$  is the speed of light. The annihilation of an electron with a positron to two photons gives

## 2. Physical background

therefore an energy balance of

$$(p_1 + p_2)c = 2m_e c^2 + \Delta E, \quad (2.2.5)$$

where the left-hand side is equal to the energy carried away by the two annihilation photons, and the right-hand side corresponds to the energy liberated by the annihilation ( $2m_e c^2$ , as the positron mass is equal to the electron mass  $m_e$ ) plus the difference in energy  $\Delta E$  of the system before and after the annihilation. Plugging this expression into Eq. (2.2.3) gives immediately

$$p_y^{2\gamma} = 2 \sin(\alpha/2) m_e c \left(1 + \frac{\Delta E}{2m_e c^2}\right). \quad (2.2.6)$$

Apart from core electrons (for which the annihilation probability is very small anyway), the relevant energies  $\Delta E$  are on the order of tens of eV compared to the electron mass-equivalent of  $m_e c^2 = 511$  keV. Due to Eq. (2.2.4), as a consequence also the momenta are much smaller than  $m_e c$  and therefore  $|\alpha| \ll 1$ . Concludingly it follows that the angular deviation is connected to the transversal part of the two-photon momentum for practical purposes by the simple expression

$$p_y^{2\gamma} = \alpha m_e c. \quad (2.2.7)$$

The general three-dimensional case obviously follows from the case treated here by splitting the so-obtained transversal part of the momentum into its horizontal and vertical components. Because of the small solid angle subtended by the detectors one can use the small-angle approximation  $\tan(\theta) = \theta$  also for the propagation directions, so that experimentally the projections of the two-photon momentum density are obtained just by histogramming the summed position vectors of pairs of coincident photons.

### 2.2.2. Theoretical interpretation

The two-photon momentum density is the primary quantity derived in an ACPAR experiment. Here I will roughly sketch what we can learn from it about the electronic structure of the solid. For a more thorough discussion see Sect. 2.4.

The momentum is a conserved quantity. As a consequence, the sum momentum of the annihilation photons is equal to the difference of the sum momentum of the system before and after the annihilation event. The two-photon momentum density therefore describes the possible transitions from the state of  $N$  electrons and one positron to  $N - 1$  electrons and no positron via the difference in total momentum, weighted by the respective probabilities. Any such transition corresponds to an excitation of a so-called “quasi-hole”.

In a metal, the distribution of these transitions, which is the distribution of occupied “quasi-particle” states, shows a discontinuity, the Fermi surface. It is due to the energy of the states being smoothly varying, with the lowest  $N$  being occupied in the ground state. The Fermi surface is then the two-dimensional manifold that separates the occupied from the unoccupied states. In a fermionic system, each state can be occupied at most once, therefore the Fermi surface determines the low-energy excitations of the system, which, for example, are responsible for the conductivity of a metal. As a consequence, its shape is the most important microscopic quantity for characterizing a given system.

## 2.3. The crystalline state

In general, the ground state of solid matter is crystalline. I will present the relevant concepts in this section. These are treated in undergraduate studies of physics, for more comprehensive presentations see standard textbooks such as Ashcroft and Mermin (1976).

### 2.3.1. Crystalline matter

Crystallinity means well-orderedness: the knowledge of the arrangement of the atoms in a small cluster already implies exactly how the atoms are arranged at any distance away. A few materials can also exist in an amorphous form, which means that a small cluster of atoms implies the arrangement of its neighbours only in a statistical sense, and over longer distances the correlation is completely lost. However, these materials, such as fused silica, can only be formed when a melt is cooled rapidly enough so that the atoms or molecules do not have time to develop the preferred arrangement, which can be promoted by the addition of further ingredients such as in window glass or ice cream. Given enough time at elevated temperatures, however, also these materials would crystallize.

The fact that the ground state of matter is crystalline is only an experimental law and in general cannot be deduced from fundamental principles. Note that already the problem of finding the ground state for the simplest conceivable model for a metal, with a potential that treats the repulsion of the closed-shell core as hard spheres and the valence electrons as totally delocalized and homogeneous, leading to an energy gain that is monotonous in the packing density, is equivalent to the Kepler conjecture.

Unless one takes special precautions to grow a single crystal, during solidification of a metallic melt normally there will be a number of nuclei, and the resulting solid will show a pattern of crystalline orientations. The grain sizes range from typically a few micrometers to a few millimetres, which can be seen by eye, for example in hot-dip galvanization. On the other hand, if the solid grows by precipitation from solutions, each new particle is deposited at the face of the crystal, which therefore grows naturally as a single crystal, leading to the popular image of mineral crystals. The regular form of these crystals is due to anisotropies of the surface energies, which can be especially high for ionic crystals, leading to a predefined shape that minimizes the integral surface energy.

Obviously, under the rigorous definition a crystal would need to be infinite, which is impossible in reality. However, for practical purposes this is no restriction, as most relevant physical properties are local in some sense, allowing the description of experiments to be separated into a bulk contribution representative of the infinite crystal and a contribution from the surface, which can be seen as a lower-dimensional crystal. By convention one admits also other deviations from perfectness: If the system is composed of atoms of different kinds, it is conceivable that the sites where atoms sit follow the symmetries, but the actual occupations of the sites by the different kinds of atoms does not. Also some random sites can be left unoccupied (termed vacancies). Both of these effects will also lead to displacements of the neighbouring atoms from the ideal positions. Finally, at finite temperatures the atoms will oscillate about their ideal positions.

## 2. Physical background

### 2.3.2. Point symmetries

A very useful description of a given crystal is via its symmetries. The point group of a crystal is the group (in the mathematical sense) of isometries on three-dimensional space that transform the crystal into itself (the positions occupied by atoms of a given kind before are also occupied by these atoms after the operation) while leaving a given point fixed. This is therefore a symmetry of directions. It can be visualized by objects that have this symmetry: Picture for example a generic rectangular parallelepiped (a matchbox without structure). The point symmetry operations are inversion (transforming a generic point  $(x, y, z) \rightarrow (-x, -y, -z)$ ), three mirror operations (for instance reflection by the  $x$ - $y$ -plane  $(x, y, z) \rightarrow (x, y, -z)$ ), and three two-fold axes of rotation (for instance rotation around the  $x$ -direction  $(x, y, z) \rightarrow (x, -y, -z)$ ). This is orthorhombic symmetry. A generic point on the surface (corresponding to a direction) belongs therefore to a class of eight equivalent points. If two of the three edges have the same length, tetragonal symmetry follows, which has additional operations in the symmetry group such as a four-fold axis of rotation and additional mirror planes. Finally, if all three edges have the same length, the crystal has cubic symmetry, so that there are 48 directions within one generic symmetry equivalence class (six permutations of the coordinates times eight possibilities for the respective signs).

There is, however, no total order on the set of point groups in the sub- and super-group sense. Picture for example a right prism with a regular pentagonal base. Here the main axis is a five-fold rotation axis, which obviously does not fit with the symmetry groups outline above.

An immediate, though important consequence of the point group of a crystal is formulated by Neumann's principle, which says that the symmetry group of any physical phenomenon contains the symmetry group of the crystal as subgroup. For example, the thermal conductivities of a crystal along two given directions will be equal if there is an element in the point group of the crystal that maps one direction to the other.

### 2.3.3. Translation symmetries — the lattice

The other symmetry concept that is relevant for crystallinity is discrete translation symmetry. This is defined by three linearly independent vectors so that a translation of the crystal by any integer linear combination of these transforms the crystal into itself. As a consequence of this it is possible to tile the volume by a periodic repetition of a single cell. An obvious choice for this unit cell is the parallelepiped that is spanned by the three basis vectors, such as in the case of the matchbox considered above. The lattice is the set of points in  $\mathbb{R}^3$  that correspond to all possible translation vectors. These translation operations also constitute a group. The choice of the basis vectors is obviously not unique. Basis vectors that yield the highest possible symmetry for a given arrangement of atoms (with the largest translation group and the smallest unit cell) are called primitive basis vectors, and the resulting unit cell is called primitive cell.

For some translation symmetries, the parallelepiped spanned by the primitive basis vectors is not optimal in the sense that it does not display the point symmetry of the lattice. An example is the face-centred cubic lattice, which is illustrated in Fig. 2.3. Sometimes it is desirable to choose the (unique) primitive cell that displays the point symmetry of the lattice, which is called the Wigner-Seitz cell. Mathematically this is just the Voronoi cell of any point in the lattice, i.e. the

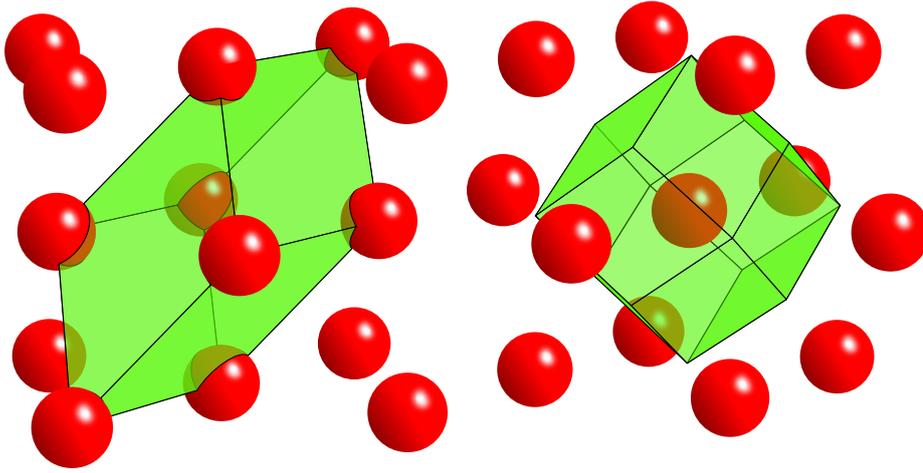


Figure 2.3.: The face-centred cubic lattice is generated by filling space with a regular arrangement of cubes and putting atoms on the cube corners (one per cube) and on the centres of the cube faces (three per cube). Here the atoms of two and a half cubic cells are plotted, with two possible primitive cells: on the left the parallelepiped generated by a given choice of primitive translation vectors, which does not display the cubic symmetry, and on the right the Wigner-Seitz cell, which displays the full point symmetry.

set of points in  $\mathbb{R}^3$  that are closer to a given lattice point than to any other lattice point.

Incidentally, the definition of crystalline matter includes systems that display only point symmetries, but no translation symmetry. In these systems called quasicrystals, the atomic interactions favour local motifs that are incompatible with translation symmetry, such as the above-mentioned pentagons (which cannot tile two-dimensional space). Still, a finite set of rules allows to extend a given nucleus uniquely to infinity. The first simple example has been derived by Penrose (1974). Subsequently, Shechtman et al. (1984) found and correctly interpreted the first physical evidence of such an order in the Al-Mn system, for which discovery Dan Shechtman was awarded the 2011 Nobel Prize of Chemistry.

### 2.3.4. The reciprocal lattice

Reciprocal space is a concept that is very valuable for the description of various phenomena in solid state physics. Essentially, it is just the Fourier dual of direct or real space. First, scattering experiments, which allow to comparatively easily access atomic length-scales by using probes with appropriate wavelengths, directly give results in reciprocal space. Additionally, the energy or frequency of delocalized excitations that are described by a modulation vector, for instance lattice oscillations, is naturally given as a function over reciprocal space. The independent variable for positions in reciprocal space is conventionally denoted  $\vec{k}$  or  $\vec{q}$ .

For a crystal, there exists a set of special positions in reciprocal space, namely those that correspond to modulations that are compatible with the real space periodicity. This is the

## 2. Physical background

reciprocal lattice

$$\Lambda' = \{\lambda' \in \mathbb{R}^3 \mid \forall \lambda \in \Lambda \exists n \in \mathbb{Z} : \lambda' \cdot \lambda = 2\pi n\}, \quad (2.3.1)$$

where  $\Lambda$  is the real space lattice. It is easy to see that this is again a lattice (a translation group), which is spanned by  $\vec{b}_1$ ,  $\vec{b}_2$  and  $\vec{b}_3$  with

$$\vec{b}_1 = 2\pi \frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)} \quad (2.3.2)$$

and cyclic permutations, when  $\vec{a}_1$ ,  $\vec{a}_2$  and  $\vec{a}_3$  are the real space basis vectors. From the definition it follows immediately that  $\Lambda'' = \Lambda$ . It is also not hard to see that  $\Lambda'$  has the same point symmetry group as  $\Lambda$ : If  $\lambda' \in \Lambda'$  and  $R$  is a given point symmetry operation it follows that  $(R\lambda') \cdot \Lambda = \lambda' \cdot (R^{-1}\Lambda) = \lambda' \cdot \Lambda = 0$  modulo  $2\pi$ .

The full-symmetry primitive cell of the reciprocal lattice is conventionally called (first) Brillouin zone. As points on the Brillouin zone boundary can display additional symmetry, they are of particular interest, as will be discussed in more detail in Sect. 2.4.5. Due to the point symmetry, reciprocal space consists of repeated sections. One such primitive cone that can cover the whole space under the point symmetry operations is called irreducible wedge.

## 2.4. Microscopic principles of the electronic structure

Here I will briefly present the most fundamental concepts of solid state physics on the atomic scale necessary for understanding the results of ACPAR measurements. Again, these are treated in undergraduate studies of physics. Illustrations of selected aspects will be given in Chap. 5 when applying the proposed algorithms to simulated data.

### 2.4.1. The Born-Oppenheimer approximation

From the point of view of condensed matter physics, matter consists of atomic nuclei and electrons. Apart from some nuclear methods in condensed matter physics, atomic nuclei can be thought of as point-like. The nuclei have positive charges given exactly by small integer multiples of the elementary charge  $e$ , where the integer nuclear charge denotes the chemical element the specific atom is a representative of. To a given nuclear charge corresponds a small number of possible nuclear masses (differing by a few percent), which incidentally can also be written (to a good approximation) as multiples of an elementary mass  $u$ . The mean elemental mass is roughly proportional to the nuclear charge. The nuclei carry an additional quantity called spin, which cautiously can be thought of as some kind of rotation, although it has no correspondence in classical macroscopic physics.

Electrons are the direct counterparts of the nuclei: they have a given fixed amount of negative charge  $-e$ , (positive) mass and spin. They have no internal structure, therefore two electrons are absolutely indistinguishable.

Despite this analogy between electrons and nuclei from the viewpoint of solid state physics, quantitative differences in their properties lead to a qualitative distinction: normalized to the same absolute charge, an atomic nucleus is about 3700 times heavier than an electron.

Additionally, the charge of a nucleus is typically a few tens of the elementary charge<sup>2</sup>. This large difference in the charge/mass ratios (and therefore in the force/mass ratios) has the consequence that due to Newton's second and third laws of motion the characteristic timescales of the electrons are much smaller than those of the atomic nuclei. Therefore the problem of calculating the time evolution of a condensed matter system (atomic nuclei and electrons) can be separated into a problem concerning only the evolution of the electronic system, treating the atomic nuclei as immobile, and the problem of the evolution of the nuclei system, which feel only the time average of the electronic system. This is called the Born-Oppenheimer approximation (Born and Oppenheimer, 1927) and corresponds to a two-step approach: compute first the electronic structure (and from this the energy and the forces) for a given arrangement of the nuclei, and then solve the ionic problem, where the electrons act as black-box potential. Here again the differing mass scales lead to a distinction: because of the large nucleus masses, the second step can for all but the lowest temperatures in general be treated within classical mechanics, i.e. describing the state of the system via the positions and velocities of all particles, subject to Newton's laws of motion. For some effects it is necessary to consider the quantization of oscillations (phonons), which also does not pose large problems. On the other hand, because of the small electron masses, the first step has to be treated fully within quantum mechanics. This is much harder, but here is where the interesting effects and open challenges of solid state physics originate from.

Actually, due to the different energy scales involved it can be beneficial for practical purposes to further distinguish between the low-lying, tightly bound core electrons and the mobile, chemically active valence electrons, and to subsume the former together with the nuclei under the concept of inert spherical ions with surplus positive charge, the forces among which being mediated by the valence electrons. From now on the focus will be on the latter, as their behaviour is responsible for most of the relevant physical effects.

### 2.4.2. The wave function

The theory appropriate for the description of electrons in condensed matter is quantum mechanics<sup>3</sup>. In an abstract setting, a state is described as a unit vector in an appropriate Hilbert space. Specifically, for a system of  $N$  electrons a convenient choice is the space of the sufficiently smooth square-integrable<sup>4</sup> complex-valued functions

$$\psi : \Omega^N \rightarrow \mathbb{C} \quad \text{with} \quad \Omega \subseteq \mathbb{R}^3, \quad (2.4.1)$$

subject to some symmetries as discussed below. The domain  $\Omega$  corresponds to the region of space to which the electrons are confined. For bounded domains additionally boundary conditions have to be specified, which, however, are normally not relevant for the resulting physics.

<sup>2</sup>As a consequence, the mass ratios in an atom are the same order of magnitude as in the solar system.

<sup>3</sup>Actually, for low-lying states relativistic effects have to be considered also, an account on the history of which has been given in Sect. 2.1.1. Most every-day physical effects, however, including those sampled by positron annihilation, are directly related only to shallow states, which are described adequately by quantum mechanics.

<sup>4</sup>For unbounded domains, however, there are physically meaningful solutions corresponding to unbound states that are only locally square-integrable.

## 2. Physical background

The physical interpretation of the wave function  $\psi$  is that  $|\psi|^2$  corresponds to the spatial probability density of the electrons, that is,  $|\psi(\vec{x}_1, \vec{x}_2, \dots)|^2 d\vec{x}^N$  is the infinitesimal probability for finding an electron at  $\vec{x}_1$ , another at  $\vec{x}_2$  and so on. As the particles are fundamentally indistinguishable, this probability cannot vary under a permutation of the three-dimensional coordinate vectors, or, equivalently, the wave function  $\psi$  can vary only by a phase factor under such a permutation. Additionally, exchanging the the same two particles two times leads to the initial situation, therefore the phase factor is  $\pm 1$ . Particles that follow the positive choice are called bosons, otherwise they are called fermions. Electrons belong to the latter kind, thus

$$\psi(\dots, \vec{x}_i, \dots, \vec{x}_j, \dots) = -\psi(\dots, \vec{x}_j, \dots, \vec{x}_i, \dots) \quad (2.4.2)$$

with all other coordinates unchanged. This leads to the Pauli exclusion principle and will be discussed in more detail in Sect. 2.4.6.

### 2.4.3. The Schrödinger equation

The time evolution of a quantum mechanical system is given by the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \tilde{\psi} = \hat{H} \tilde{\psi}, \quad (2.4.3)$$

where  $\hbar$  is the reduced Planck's constant and  $\hat{H}$  is the Hamiltonian operator corresponding to the energy of the system and  $\tilde{\psi}$  is the time-dependent wave function. For the non-relativistic case, which is relevant here, the Hamiltonian can be written as

$$\hat{H} = \frac{-\hbar^2}{2m} \Delta + V, \quad (2.4.4)$$

where the first term quantifies the kinetic energy and the second, acting as pointwise multiplication, the potential energy, i.e.  $V(\mathbf{r})$  gives the potential energy when the electrons are arranged in space as given by the  $3N$ -dimensional vector  $\mathbf{r}$ . With a real-valued potential  $V$ , the Hamilton operator (2.4.4) is symmetric, and the energy eigenvalues are therefore real, and for physically meaningful potentials (such as those that have no stronger divergences than the Coulomb potential's  $1/r$ ) the spectrum is bounded from below. With a suitable choice of boundary conditions, the operator becomes self-adjoint.

Under the assumption of a potential that is constant in time, one can make the ansatz of a separation of spatial and temporal functional dependencies

$$\tilde{\psi}(\cdot, t) = \exp(-iEt/\hbar) \psi, \quad (2.4.5)$$

which leads to the time-independent Schrödinger equation

$$E\psi = \hat{H}\psi, \quad (2.4.6)$$

describing stationary states of the system. As a consequence, the stationary states of a system are given by the eigenstates of the Hamilton operator, and the respective eigenvalues are the corresponding energies. Mathematically, the problem of finding the ground state of an electronic system corresponds therefore to solving (2.4.6) for the lowest eigenvalue  $E_0$ .

### 2.4.4. Single-particle approximations

Apart from the hydrogen atom and the singly ionized di-hydrogen molecule, both consisting of only one electron, no analytic solutions of (2.4.6) for realistic systems are known. Due to the dimensionality, numerical solutions of the full  $3N$ -dimensional partial differential equation (2.4.6) for non-trivial  $V(\mathbf{r})$  are obviously only feasible for very small  $N$ . The fundamental description is also not well suited for a qualitative understanding. The conventional approach is therefore to think about the system as a collection of distinct electrons, each governed by its own wave function, trading rigorousness for accessibility, both for understanding and, assuming some effective interaction potential, calculation. Actually, this description in terms of single-particle wave functions is also quite successful in explaining most experimental results on the majority of systems, and the remaining subsections will be restricted to this setting. Fortunately, as will be discussed at the end of Sect. 2.4.8, in most cases the single-particle view gives qualitatively correct results even for the many-particle problem. In contrast, “strongly correlated” systems, i.e. those where such a description gives qualitatively wrong results, have gained special attention in recent times due to the possibility to observe new physics there.

The most drastic approximation to the problem consists in neglecting any explicit interaction between the electrons altogether and solve (2.4.6) in only three dimensions with some effective single-particle potential  $V$ . The ground state of the system is then given by populating the  $N$  states corresponding to the lowest eigenvalues (see Sect. 2.4.6), and the ground state energy is the sum over the single particle energies. By parametrizing a basic phenomenological model for  $V$  (this is emphasized by calling it “pseudo-potential”) a very satisfactory description of experimental properties has been obtained for particularly simple metals such as Al and Pb (Ashcroft, 1963; Anderson and Gold, 1965).

I will now briefly present the two classical first-principles single-particle approximations, that is algorithms that allow to calculate properties without adjustable parameters (also called *ab initio* methods, i.e. starting from the Hamiltonian). The Hartree-Fock method consists in solving the original problem (2.4.6) over a subspace of the full  $3N$ -dimensional wave function space: Take  $N$  linearly independent single-particle functions  $\psi_i$  (in this context also called orbitals) and define the many-body wave function via the Slater determinant (Slater, 1929)

$$\psi(\vec{x}_1, \dots, \vec{x}_N) = \sum_{\sigma \in S_N} \text{sgn}(\sigma) \prod_i \psi_i(\vec{x}_{\sigma(i)}), \quad (2.4.7)$$

where  $S_N$  is the  $N$ -element permutation group. This choice obviously fulfils (2.4.2). Taking the exact potential  $V$ , which describes the attractive Coulomb interaction between the nuclei and the electrons and the repulsive interaction between the electrons themselves, and minimizing the resulting energy over the form of the orbitals results in a valid state of the system, whose energy is therefore a rigorous upper bound on the ground state energy. By neglecting the electron correlation in this way the dimensionality of the problem has been greatly reduced, allowing to solve small systems such as single atoms or simple molecules.

Especially in solid state physics the predominant *ab initio* method nowadays is Kohn-Sham density functional theory (Kohn and Sham, 1965). This goes one step further and drops even an explicit treatment of the requirement (2.4.2). Again, the state of the system is described by  $N$  single-particle functions  $\psi_i$ , and the potential energy is given only via a functional of

## 2. Physical background

the overall electron density, which is the pointwise sum of the densities corresponding to the distinct electrons  $|\psi_i|^2$ . The potential energy functional is modelled as the Coulomb energy of the electron density plus the so-called exchange-correlation term, which is parametrized so as to approximately take into account the effects due to exchange (the symmetry requirement which would be treated exactly in the Slater determinant) and correlation (such as that the exact many-body wave function should be reduced when two coordinates are close to each other because of Coulomb repulsion). The most simple functional, the local density approximation, depends only on the pointwise value of the density, parametrized by comparison with the energy of a homogeneous electron gas which can be calculated exactly, and already gives quite good results for most metals. Further enhancements such as the generalized gradient approximation can improve the agreement with experimental results.

### 2.4.5. Consequences of the crystalline state

Translation symmetry of the crystal suggests a particular description for the states of the system, which I will work out here for the single-particle problem: Take some member  $R$  of the translation group, the Hamilton operator  $\hat{H}$  of Eq. (2.4.4) and a state  $\psi$ . Obviously  $\hat{H}$  commutes with  $R$ , as the potential  $V$  is symmetric with respect to  $R$  (and the Laplace operator is symmetric to any translation). Therefore the eigenstates of  $\hat{H}$  can be chosen to be also eigenstates of  $R$  (and in fact simultaneously for the whole translation group, as the translations commute). Specifically it follows that the eigenstates of  $\hat{H}$  can be chosen to fulfil

$$(R\psi_k)(\vec{x}) = \psi_k(\vec{x} + \vec{R}) = \nu_{k,\vec{R}}\psi_k(\vec{x}), \quad (2.4.8)$$

where the translation operator  $R$  is identified with its translation vector  $\vec{R}$ . Physically only states with  $|\nu_{k,\vec{R}}| = 1 \forall R$  are meaningful, as otherwise the wave function would grow in some direction without bounds.

For a more thorough analysis we observe that Fourier-transforming (2.4.6) gives

$$E\hat{\psi}(\vec{k}) = \left( \frac{\hbar^2|\vec{k}|^2}{2m}\hat{\psi} + \hat{V} * \hat{\psi} \right)(\vec{k}). \quad (2.4.9)$$

Due to its periodicity, the potential  $V$  can be written as discrete Fourier series, so that above equation links only those Fourier coefficients of  $\psi$  that differ by a reciprocal lattice vector  $\mathbf{G}$

$$\left( \frac{\hbar^2|\vec{k}|^2}{2m} - E \right) \hat{\psi}(\vec{k}) + \sum_{\mathbf{G}} \hat{V}_{\mathbf{G}} \hat{\psi}(\vec{k} - \mathbf{G}) = 0. \quad (2.4.10)$$

It therefore follows that for each  $\vec{k}$  in the first Brillouin zone (to avoid multiple counting) there exists a set of eigenvalues  $E_{i,\vec{k}}$  and eigenstates  $\psi_{i,\vec{k}}(\vec{x})$  to (2.4.6). Specifically

$$\psi_{i,\vec{k}}(\vec{x}) = e^{-i\vec{k}\vec{x}} u_{i,\vec{k}}(\vec{x}), \quad (2.4.11)$$

where  $u_{i,\vec{k}}$  displays the periodicity of the crystal. Therefore it follows that  $\nu_{\vec{k},\vec{R}} = \exp(-i\vec{k}\vec{R})$ .

The functions  $\psi_{i,\vec{k}}$  are called Bloch states. Realizing a finite crystal by assuming periodic boundary conditions, only those values of  $\vec{k}$  that are compatible with the boundary conditions are allowed. Specifically, it follows that the allowed values are equidistantly spaced and that

their number is equal to the number of unit cells in the crystal. For macroscopic systems  $\vec{k}$  can be assumed to be a continuous variable, however, and the energies  $E_{i,\vec{k}}$  vary smoothly (due to (2.4.10) and the fact that  $\hat{V}_{\mathbf{G}}$  decay fast) over the Brillouin zone for a given  $i$  and are called the energy bands. A different view, that is however equally simple and leads to the same conclusions, is provided by tight-binding theory (Slater and Koster, 1954).

Obviously the energy bands inherit the periodicity of the reciprocal lattice and can therefore be described via a Fourier series of the appropriate point symmetry. If the symmetry of the lattice is high enough (such as for the cubic cases), the boundary of the Brillouin zone will be a mirror plane. Normally this will mean that the gradient of the energy has no perpendicular component at the zone boundary, although if the relevant Fourier coefficient is zero due to additional symmetry (such as for the diamond lattice) it will lead to a symmetrical crossing of bands.

### 2.4.6. Occupations

As discussed in Sect. 2.4.4 under the framework of the Hartree-Fock-approximation, a wave function for non-interacting particles that obeys the requirement of total antisymmetry (2.4.2) can be constructed from a given choice of single particle wave functions  $\psi_i$  via the Slater determinant (2.4.7). In fact, from the uniqueness of the determinant it follows that any wave function of finitely many non-interacting fermions can be written in such a way. If now two electrons are in the same state, i.e. there exist  $i \neq j$  with  $\psi_i = \psi_j$ , the determinant is identically zero, and therefore does not correspond to a physically meaningful state. This is the fundamental reason for the Pauli exclusion principle, which states that it is impossible for two fermions to occupy the same state. Concludingly, the state of a system of non-interacting fermions is completely determined by specifying which single-particle states are occupied and which are not (as the particles are indistinguishable).

It is a principle of statistical physics that the probabilities for the system to be in some state  $\sigma$  depend only on  $E(\sigma)$ , the energy of the respective state. Specifically, the ratio of the probabilities for the system to be in the states  $\sigma_0$  or  $\sigma_1$  fulfils

$$\frac{p(\sigma_1)}{p(\sigma_0)} = \exp\left(-\frac{E(\sigma_1) - E(\sigma_0)}{k_{\text{B}}T}\right), \quad (2.4.12)$$

where  $k_{\text{B}}$  is Boltzmann's constant (in a sense, this is just the definition of temperature in statistical physics). Consider now a fermionic system with just two states, either occupied (corresponding to energy  $E$ ) or unoccupied (corresponding to zero energy), in contact with a reservoir of particles at energy  $\mu$  per particle. If the system can exchange particles with the reservoir, it will have a probability to be occupied of  $1/(1 + \exp((E - \mu)/k_{\text{B}}T))$  and to be unoccupied of  $1/(1 + \exp(-(E - \mu)/k_{\text{B}}T))$  (just check that the probabilities fulfil (2.4.12) and sum to one). As the particles do not interact, this holds also for the occupations of the single-particle states in a multi-particle system and results in the Fermi-Dirac occupation probability

$$p(\sigma_i = 1) = \frac{1}{1 + \exp\left(\frac{E(\sigma_i) - \mu}{k_{\text{B}}T}\right)}, \quad (2.4.13)$$

where  $\sigma$  is the occupation vector of the single-particle states, i.e.  $\sigma_i = 1$  means that state  $i$  is occupied. Strictly speaking, these occupation probabilities have been derived in the framework

## 2. Physical background

of the grand-canonical ensemble, which means that the number of particles in the system is not conserved. By a suitable choice of the chemical potential  $\mu(T)$  for a given temperature  $T$  it can be guaranteed, however, that the expected number of particles in the system  $\sum_i p(\sigma_i = 1)$  is equal to  $N$ , and due to the law of large numbers the relative fluctuations vanish for systems with a macroscopic number of electrons, so that (2.4.13) actually also describes the physically relevant case where the number of electrons is conserved.

For the limit of  $T \rightarrow 0$  (2.4.13) describes a step function, where exactly those states with an energy below  $\epsilon = \mu(0)$  are occupied.  $\epsilon$  is termed the Fermi energy.

### 2.4.7. The Fermi surface

In Sect. 2.4.5 it was demonstrated that the single-particle states can be indexed by a reciprocal lattice vector  $\vec{k}$  and a band index  $i$ . The band energies  $E(\vec{k}, i)$  are smooth in  $\vec{k}$ , therefore  $E(\vec{k}, i) = \epsilon$  defines a number of smooth manifolds (depending on how many bands pass through the Fermi energy). Together, they are called the Fermi surface and divide reciprocal space in regions inside the Fermi surface, where a given band is occupied at zero temperature, from the outside regions, where it is not. Their paramount importance for solid state physics follows from the fact that the typical widths of the highest occupied bands are in the range of a few eV, whereas  $k_B T \approx 0.025$  eV for room temperature. Therefore (2.4.13) is also for finite temperatures very sharp. Only electrons around the Fermi energy (with  $|E - \epsilon|$  not much larger than  $k_B T$ ) can contribute to physical effects that are due to dynamics of the electrons such as electrical or heat conductivity, while lower-lying electrons remain “frozen”, as they have no nearby empty states to move into.

From the discretization of the possible  $\vec{k}$  it follows that for a crystal consisting of  $N$  unit cells, the reciprocal volume per possible  $\vec{k}$  is equal to the  $N$ -th part of the Brillouin zone. Considering the additional binary degree of freedom of spin, it follows that for each electron per unit cell the Fermi surface encloses half of the volume of the Brillouin zone. For a system with an even number of electrons per unit cell it is therefore possible that the Fermi energy falls into a gap between bands. As a consequence, the Fermi surface is empty and the material is an insulator or a semiconductor (when the width of the band gap is small enough for some higher-lying states to be thermally occupied). If the uppermost atomic orbitals overlap to some degree in real space, as is the case for metals, the bands will become wider and overlap in energy, so that generally the Fermi surface has one (for simple metals such as Na or Cu) or more sheets (for the more complicated transition metals). Ferromagnetism leads to a splitting of the energies of the two electron spin possibilities, leading to a complex Fermi surface with a number of sheets for each spin.

### 2.4.8. The many-body problem

The treatment until now corresponded to the framework of effectively single-particle band structure models, where electron correlation is neglected. Actually, however, as was already discussed in Sect. 2.4.4, the Coulomb interaction between the electrons is of comparable magnitude as the Coulomb interaction with the ionic potential and should thus give rise to appreciable correlations between the electrons. A priori, the actual many-particle-problem can therefore not

be considered as a small perturbation to the idealized single-particle treatment. However, there is a consensus that in most cases there is a qualitative correspondence. The relevant concept here is excitation: In a non-interacting system the ground state corresponds to all single-particle states below the Fermi level being occupied and all above unoccupied. An excitation consists in some particles being moved to higher (unoccupied) states, so that now there are holes below the Fermi level. In an interacting system, single-particle states are not well-defined, instead “quasi-particles” and “quasi-holes” take their place. These can be considered as dynamical collections of excited particles having a finite life-time, which goes to infinity for small excitation energies (the dynamical collections become stable for practical purposes). These quasi-particles are in a sense a cloud of interacting particles such that the interactions with the outside cancel, therefore having made the substitution particles  $\rightarrow$  quasi-particles, they can again be thought of as non-interacting (but again each quasi-particle excitation can be occupied only once).

The remarkable point is now that in a thought experiment where one slowly dials up the interaction strength starting from zero, the state evolves smoothly, resulting in the corresponding state of the interacting system (this concept is called adiabaticity). Starting from an excited non-interacting state (some excited particles), the momentum (amongst other properties) of the resulting excited quasi-particles stays constant, while other quantities such as the energy change. As a consequence, the density of possible excitations in reciprocal space stays constant, the concept of the Fermi surface stays valid, and the occupations of the quasi-particle excitations still follow Fermi-Dirac statistics. This electronic state is called Fermi-Landau liquid.

It is possible, however, that at some point during dialling up the interaction strength a discontinuity happens and that therefore the resulting interacting state differs qualitatively from the initial non-interacting state. Such systems are called strongly correlated and are the focus of much interest. For nearly all of everyday materials the actual interaction is weaker than the critical one, however, so they can indeed be described in the framework of Fermi-Landau liquids.

### 2.4.9. Momentum densities

The last point to cover is how the two-photon momentum distributions observed in ACPAR are connected to the sample’s electronic structure. Again I will cover first the simple case of non-interacting particles and afterwards motivate the qualitative effects of correlations.

Physically, the momentum is a conserved quantity. Therefore the sum of the two annihilation photons’ momenta has to be equal to the sum of the electron and positron momentum before annihilation, under the assumption that all other electronic states are not affected by the annihilation. Quantum-mechanically, this leads to the fact that the probability for the annihilation of an electron in state  $\psi_e$  with a positron in state  $\psi_p$  into two photons with states  $\psi_i(\vec{x}) \propto \exp(-i\vec{x}\vec{k}_i)$ , respectively, is proportional to

$$\left| \int d\vec{x} (\psi_e \psi_p \overline{\psi_1 \psi_2})(\vec{x}) \right|^2, \quad (2.4.14)$$

integrated over the sample. Due to (2.4.11), it follows that this probability is non-zero only if  $\vec{k}_e + \vec{k}_p = \vec{k}_1 + \vec{k}_2 + \mathbf{G}$  for some reciprocal lattice vector  $\mathbf{G}$ . Energy conservation (see Sect. 2.2) gives another constraint on the resulting  $\vec{k}_i$ . Note that above expression essentially is just the

## 2. Physical background

absolute square of the Fourier coefficients of the product  $u_e u_p$  as in (2.4.11). The two-photon density at a given  $\vec{k} = \vec{k}_1 + \vec{k}_2$  is then just the double sum over all occupied states (electron and positron) so that the condition on the wave vectors is fulfilled, weighted by the absolute square of the respective Fourier coefficient. Crossing the Fermi surface therefore results in a jump in the density, as an additional band becomes occupied or unoccupied.

The actual many-body problem is more complicated: First, as discussed above, a description of the electronic system in terms of single-particle states is a simplification, as correlations will arise due to the interactions between the electrons. Additionally, also the positron interacts with the electrons and therefore correlations are built up. However, due to the extremely small number of positrons in the sample (not more than one at any given time), the overall spectrum of the electronic excitations will not be affected. Via an adiabatic dialling-up of the interactions as in Sect. 2.4.8, the  $\vec{k}$ -vectors of the excitations again do not change. Consequently the two-photon momentum density still shows the unperturbed discontinuities at the Fermi surface in the density of the occupied electronic states, reweighted by the positron effects. On the other hand, the question of the quantitative influence of electron-positron correlations on the values of the two-photon momentum density, which goes under the term “enhancement” (increased annihilation probability due to the attraction between electrons and positrons), is still an open problem of theoretical solid-state physics.

## 2.5. Formulation of the problem

After this exposition of the physical basics I will now formulate the problem to be solved:

The crystallographic structure of the system to be investigated is known (i.e. the geometry of the Brillouin zone and the point symmetry group). The resolution of the detectors (and additional effects limiting the resolution, such as the size of the positron beam spot on the sample and the thermal momentum of the annihilating positrons) is approximately known. Given are a number of projections of the two-photon momentum density, where the orientation of the sample crystal with respect to the detectors is known to some accuracy. The projections are *given* as *pixel data* with integer values, corresponding to uncorrelated realizations of Poisson random variables. The problem is to *find the shape of the Fermi surfaces and the three-dimensional two-photon momentum density*. From general physical knowledge (see the previous sections) it follows that the three-dimensional momentum density will be smooth away from the Fermi surface and display jumps with a sign depending on whether the Fermi surface is crossed from the inside or the outside. The symmetry of two-photon momentum density and Fermi surface are known from the crystallographic structure, and normally there is only a small number of plausible choices for the Fermi surface topology.

## 3. Mathematical concepts

The problem to be solved here (determining the Fermi surface from measured ACPAR spectra) falls into the general class of inverse problems, in the sense that for a given Fermi surface and two-photon momentum density it is straightforward to derive the resulting spectra (or stochastic realizations thereof in the presence of noise), whereas the reverse problem is not well-posed and therefore without modification does not admit a reasonable solution. Inverse problems are of great practical importance and are one of the examples where recent developments in mathematics have led to tangible beneficial effects for society. In this chapter I will review and discuss some of the concepts and tools of applied mathematics relevant for solving aspects of the problem at hand. For a more detailed treatment of specific aspects see, for instance, the relevant chapters of Scherzer (2010).

### 3.1. Probabilities and statistics

Due to the presence of uncertainties in the data and/or missing information the mathematical formulation of a data interpretation problem (deciding on fundamental parameters from measurements) will in general be cast in terms of some notion of probability. Here I will discuss the consequential concepts of approximate solutions and regularization.

#### 3.1.1. Notions of probability

Nowadays, there are two dominating interpretations of probability, going under the subjects of frequentist inference and Bayesian inference, respectively. The principal distinction is that in the frequentist interpretation the outcome of an experiment is considered to be the realization of some random process characterized by a given constant underlying parameter (vector)  $\xi$ . The probability distribution of the outcome is just the limit of the histograms of the outcomes if the experiment was repeated infinitely often. The aim of frequentist inference is then to deduce  $\xi$  from the outcome of the experiment. It has no meaning to speak of a probability distribution for  $\xi$ , as it is considered to be a fixed quantity. Rather, uncertainty is quantified via confidence intervals, which are derived in some fixed way so that for a sequence of repetitions of the experiment the respective computed confidence intervals include the true value of  $\xi$  with a certain probability. An obvious way of determining an estimate for  $\xi$  from an experiment is Maximum Likelihood Estimation (MLE), which is just the choice of  $\xi$  that maximizes the probability of the experiment to give the observed outcome.

Bayesian inference, on the other hand, treats also  $\xi$  as a random variable which can be described via probability distributions that quantify the available information on the problem. In some sense, Bayesian inference is a consistent way of updating this information by the

### 3. Mathematical concepts

outcome of experiments. It gets its name from Bayes' formula

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \quad (3.1.1)$$

which relates the marginal and conditional probability densities of two events  $X$  and  $Y$ .

A small example can serve to elucidate these concepts: Consider some shop and let  $L$  be the random variable of people that enter the shop between 9:00 and 9:15 on an average Monday. If the movements of people are uncorrelated (and the number of potential customers is large and each person's probability to enter the shop is small), then the probability density for  $L$  is given by a Poisson distribution

$$p_{\text{Poisson}}(l|\lambda) = \frac{\lambda^l}{l!} e^{-\lambda}, \quad (3.1.2)$$

characterized by the parameter  $\lambda$ , which is equal to the mean number of customers. Assume now that an experiment to determine  $\lambda$  is performed, i.e. on a given day the number of customers is actually counted and found to be  $l = 0$ . The maximum likelihood estimate is given by  $\lambda^* = \arg \max_{\lambda} p(0|\lambda) = 0$ . It is obvious that this answer is not meaningful, because while it is indeed perfectly possible that a small shop has no customer in a given quarter of an hour, the mean number of customers has to be positive, as otherwise the shop would not exist any more.

The Bayesian approach makes it possible to include such knowledge: Assume that the prior knowledge (before performing the experiment) is encoded in a probability density for  $p_{\text{prior}}(\lambda)$ . Applying Bayes' formula gives a posterior probability density for the parameter  $\lambda$  taking into account the result of the experiment

$$p_{\text{posterior}}(\lambda|l=0) = \frac{p_{\text{Poisson}}(l=0|\lambda)p_{\text{prior}}(\lambda)}{p(l=0)}. \quad (3.1.3)$$

Note that the denominator only serves to normalize the probability density and therefore normally does not need to be explicitly computed. For the case at hand, even the most primitive choice of considering each positive  $\lambda$  as equally likely (which corresponds to an improper prior distribution) results in  $p_{\text{posterior}}(\lambda|l=0) = \exp(-\lambda)$ . This quantifies the gain in information after performing the experiment. Including more prior information would give an even better defined posterior distribution. In fact, the so-obtained posterior distribution can be used as prior if an additional experiment is to be performed.

The problem to be treated in this thesis, determining Fermi surfaces from ACPAR spectra, is a quantitative problem. It is natural to treat the Fermi surface (or the parameters that describe it) as random variables and report the results of experiments in terms of probability densities. Also, the physical knowledge about the problem implies a lot of prior information. Therefore, the further discussion will be in the framework of Bayesian inference.

#### 3.1.2. Approximate solutions and their probabilistic interpretation

Generally speaking, an inverse problem is the problem of solving

$$F(x) = y \quad (3.1.4)$$

over some domain  $X$ . For a specific application,  $x$  is some unknown internal parameter (vector) and  $F$  is a known deterministic operator that specifies how some measurable data  $y$  depends on

$x$ . The task is to determine  $x$ . Even if  $F$  has an explicit inverse, the problem can be ill-posed if (i)  $F$  is not injective, if (ii) the linearization of  $F^{-1}$  around  $y$  is unbounded, or if (iii)  $y$  is corrupted by noise and therefore not in the range of  $F$ . For the case at hand all three possibilities apply: the dimension of the data vector  $y$  is smaller than the dimension of the parameter vector of the densities  $x$ , therefore the linear operator  $F$  has singular values equal to zero implying (i) and (ii), and the data are indeed noisy, which will make them incompatible with the symmetry imposed by  $F$ , implying (iii).

The solution to (iii) is to relax (3.1.4) and search for  $x$  so that  $F(x)$  is near  $y$  in some distance metric. In many cases, a sensible choice is to minimize

$$f(x) = (F(x) - y)^\top \mathbf{W} (F(x) - y) =: \|F(x) - y\|_{\mathbf{W}}^2, \quad (3.1.5)$$

where  $\mathbf{W}$  is the inverse of the covariance matrix of the noise.

The justification for this approach follows from probability theory. Suppose that the measured signal  $y$  is equal to the ideal signal  $y^* = F(x^*)$  corrupted by additive noise  $\xi$ , where the random vector  $\xi$  follows a multivariate Gaussian distribution with covariance matrix  $\mathbf{C}$

$$p(\xi) \propto \exp(-\xi^\top \mathbf{C}^{-1} \xi / 2). \quad (3.1.6)$$

Obviously  $\xi$  is equal to the difference of data and ideal signal  $\xi = y - F(x^*)$ , therefore expression (3.1.5) is nothing else than minus the double logarithm of the probabilities of the noise realization that corresponds to a given choice for  $x$  (the unspecified normalization constant of the probability gives a constant term to the logarithm, which is immaterial and will be neglected in the following). The least squares problem (3.1.5) is therefore equivalent to maximizing the probability for the corresponding residuals or, in other words, maximum likelihood estimation, an insight which is due to Gauss (1809). Thanks to the central limit theorem, assuming a Gaussian distribution for the noise is often indicated, which explains the pervasiveness of least-squares estimation.

Note that there are problems where the measured data cannot be described as the ideal signal plus additive noise with known distribution (let alone Gaussian), for instance with counting noise such as in the example of the number of people entering a shop, which follows a Poisson distribution. Here the expected deviations of the data from the ideal signal depend on the mean, that is, on the parameter to be determined. In such situations it is necessary to replace the weighted squared deviations by a general expression for the probability  $p(y|F(x))$  (or its logarithm, respectively).

### 3.1.3. Heuristic regularization

In the sense of approximate solutions as discussed above, (3.1.4) will admit a solution, but due to the unboundedness of the inverse operator and the presence of noise this solution will not be stable. The concept of modifying the question so that the regularized solution becomes a stable approximation is called *regularization*.

There is a number of distinct ways this can be done:

1. describe the search space by a manifold of lower dimension

### 3. Mathematical concepts

2. stop an iterative method before convergence has reached
3. use an explicit regularization functional

The first approach is evident: if  $x$  for instance describes two-dimensional data that are known to be smooth, then a formulation in terms of a truncated Fourier series or a two-dimensional polynomial will be able to describe the data (nearly) as well as a pixel-by-pixel description, but uses much less degrees of freedom and is therefore much more stable. In other cases a more localized formulation, such as in terms of blobs (Lewitt, 1990), can be better suited. As a rule, for a given problem it is advisable to choose a formulation of the unknowns that is able to describe the features to a given accuracy with as few degrees of freedom as possible.

The second approach requires more elaboration: Consider minimizing (3.1.5) numerically by an iterative algorithm. Under certain conditions (Chung, Knepper, and Nagy, 2010) the features that are reconstructed in early iterations will be those that correspond to large singular values of  $\mathbf{A}$ , where  $\mathbf{A}$  is the linearization of  $F$ , and which are therefore not much affected by noise. Eventually, however, it will converge towards the true solution of (3.1.5), which is unstable with respect to noise. The trick is now to stop the iterative algorithm at a stage where the “essential” features of the solution have been reconstructed, but before it becomes corrupted by noise. Deciding on when to stop is nontrivial. This also applies to the case of an expectation-maximization algorithm (Dempster, Laird, and Rubin, 1977), which is a particular recipe for iteratively solving maximum-likelihood problems. This idea of stopping iterative solution algorithms before convergence will not be followed here further, because firstly for the case at hand the solution can be arrived at directly as will be shown in Sect. 4.3.5, secondly it does not allow for generating probability densities for the solution in the Bayesian sense, and thirdly it is less clear than using an explicit regularization functional.

The third approach in its simplest form is called Tikhonov regularization (Tikhonov, 1963) and corresponds to replacing (3.1.5) by

$$f(x) = \|F(x) - y\|_{\mathbf{W}}^2 + \alpha \|\mathbf{R}x\|^2 \quad (3.1.7)$$

with a positive regularization parameter  $\alpha$ . Taking for instance  $\mathbf{R}$  as the identity operator, it is obvious that the spectrum of the Hessian of  $f$  at any solution of the unregularized problem is bounded from below by  $\alpha$ , and that therefore the solution of (3.1.7) for small enough  $\alpha$  and reasonable  $F$  is unique and stable with respect to noise on  $y$ . The magnitude of  $\alpha$  decides on the relative weighting of the conflicting requirements of stability and faith to the data  $y$ . More general regularization functionals quantify for instance the difference of the histogram of  $x$  or its derivatives from a specified distribution. A priori, adding such regularization functionals to the objective function might seem arbitrary. An interpretation in terms of Bayesian probabilities will be given in the next section.

#### 3.1.4. Bayesian interpretation of regularization

Having identified the objective function in unregularized least-squares estimation with the negative logarithm of the probabilities of the residuals corresponding to a given choice for  $x$ , it is obvious to interpret the addition of a regularization functional also in terms of probabilities.

In fact, assuming the noise to be independent of the actual model parameters  $x$ , the joint probability is equal to the product of the respective probabilities, and therefore its logarithm is equal to the sum of the respective logarithms. In a Bayesian interpretation the regularization functional is thus the negative logarithm of the probability density of  $x$  as given by prior knowledge. A frequentist probabilist could consider it as subsuming the outcomes of different prior experiments and conceptually do joint maximum likelihood estimation on the data of all experiments at once, which would be called maximum a posteriori estimation.

I will now explicitly discuss this concept, taking as example a popular choice for a regularization functional, the principle of maximum entropy (MaxEnt). Consider an abstract situation where there are  $I$  categories and  $N$  entities that initially get assigned to the individual categories randomly, independent of each other and with equal probability for each category. Denote the corresponding random vector of entities per category with  $x_i$ . The probability distribution for  $x_i$  is the multinomial distribution with probability density

$$p(x) = \frac{N!}{\prod_i x_i!} \frac{1}{I^N} \quad (3.1.8)$$

with  $\sum_i x_i = N$ . If  $N$  is much larger than  $I$ , that is the expected number of entities per category is large, Stirling's formula gives

$$\log(p(x)) = - \sum_i x_i \log(x_i) + \text{const.} \quad (3.1.9)$$

Now for simplicity suppose that in an experiment you have measured the number of entities in each category with additive Gaussian errors with standard deviations  $\sigma_i$  and denote this data vector with  $y$ . The resulting expression for the negative logarithm of the posterior probability density is then

$$f(x) = \sum_i \frac{(x_i - y_i)^2}{2\sigma_i^2} + \sum_i x_i \log(x_i) \quad (3.1.10)$$

up to an additive constant. The maximum a posteriori estimate would be the minimizer of this functional. The maximum-entropy principle gets its name from the similarity of the regularization functional to the expression for thermodynamic entropy (which in statistical physics is defined by way of just such a multinomial probability consideration).

A specific example for the abstract model above could be the case that it is known that  $N$  individuals of a given animal species are living in a given area, and that some estimate for the number of individuals for subdivisions is known, for instance via a survey on the animals' droppings. If it can be assumed that the animals do not interact with each other, then (3.1.9) is indeed the correct expression. However, this regularization functional has come to be applied with arbitrary regularization parameter  $\alpha$  also for situations where there is no a-priori reason to suppose that the unknowns result from a random assignment of quanta to categories as above, see for instance the examples in Sect. 4.1.

In contrast to the example above, especially for the case of regularization via subjective prior information it is often comparatively easy to formulate this knowledge in a qualitative way (such as "the square norm of the second derivative should be small"), but much harder to constrain quantitatively. In other words, the "correct" value for the regularization parameter  $\alpha$  is unknown. If the probability distribution of the noise is known sufficiently well (as is usually the case), then this problem can be solved by choosing  $\alpha$  in such a way that the resulting residuals

### 3. Mathematical concepts

are “typical” representatives. To be specific, for a multivariate Gaussian distribution it is known that  $\xi^\top \mathbf{C}^{-1} \xi$  is a random variable that follows the  $\chi^2$ -distribution for the given number of degrees of freedom  $n$ . If  $n$  is large, then this distribution is clustered sharply around its expected value of  $n$ , and  $\alpha$  can be chosen so that the resulting  $F(x)$  neither over- nor underfits the data  $y$ . This is known as Morozov’s principle (Morozov, 1966).

#### 3.1.5. Sampling posterior distributions

The goal of interpreting experimental data in a Bayesian view consists in deriving the posterior distribution  $p$  of the unknowns  $x$ , or at least its mean and covariance matrix. In the simplest case of a linear forward map, a Gaussian prior and additive Gaussian errors, the posterior distribution is again Gaussian, which is uniquely determined by maximizing the expression of the posterior distribution and computing the Hessian matrix of its logarithm. In the general case, this is not possible, however, and lacking an analytic expression of the distribution, such information can only be realized via a sample of the distribution.

Sampling a multivariate distribution  $p(x)$  directly (by testing the probability of random values) will be very inefficient if there is no tight estimate of its support known a priori. The solution to the problem lies in using Markov chain Monte Carlo algorithms, which generate correlated sequences of samples. Here I will explicitly consider the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), which exploits the regularity of the distribution to be sampled by choosing a new sample  $x_{N+1}$  iteratively in the vicinity of the previous one  $x_N$ . As the old one likely corresponded to a high value of probability, so will the new one, which leads to its efficiency. In its most basic formulation the algorithm consists in generating random trial displacements  $\Delta x_{N+1}$  according to some symmetric probability density  $p_\Delta(\Delta x) = p_\Delta(-\Delta x)$ , setting the trial value  $\tilde{x}_{N+1} = x_N + \Delta x_{N+1}$ , and accepting  $\tilde{x}_{N+1}$  as  $x_{N+1}$  with the probability

$$P = \min(p(\tilde{x}_{N+1})/p(x_N), 1), \quad (3.1.11)$$

otherwise setting  $x_{N+1} = x_N$ . It follows that if the probability density for  $x_N$  is given by  $p(x)$ , so will it be for  $x_{N+1}$ . As a consequence, the elements of the Markov chain are a sample of  $p$  after the bias introduced by the choice of  $x_0$  has dissipated, which will happen within finite time under mild assumptions on  $p$  and  $p_\Delta$ .

A particular advantage of this algorithm for sampling posterior distributions lies in the fact that it evaluates only quotients of probabilities, and that it is therefore not necessary for the expression of the posterior distribution to be normalized, which would be hard to achieve in the general case.

The efficiency of the sampling, i.e. how fast the random walk moves through parameter space, follows from the choice of the displacement probability density  $p_\Delta(x)$ . If this distribution is too wide, most of the trials will be rejected, as they lie in regions of low probability, and the Markov chain will consist of long runs of identical elements. If on the other hand it is too narrow, most of the trials will be accepted, but successive values will be highly correlated. This is most critical in the case when  $p(x)$  is highly anisotropic. The optimal choice for  $p_\Delta(x)$  should reflect this anisotropy, although this information is obtained only during sampling. A useful generalization of the Metropolis-Hastings algorithm is therefore to use an evolving  $p_\Delta^N(x)$ ,

either by parametrizing the yet obtained sample  $x_1 \dots x_N$  as a multivariate Gaussian, or in a model-free way to take for the new displacement the (perhaps contracted) vector between two randomly chosen past sample points additional to some small isotropic component.

## 3.2. Denoising and deblurring

One of the most basic ill-posed inverse problems is denoising and/or deblurring. Consider the problem of a measurement of a spatially varying quantity  $y$  that is known to be smooth in some sense, while the actually measured data are affected by additive noise. If the noise is spatially independent, then it is intuitive that the smooth signal can be recovered. A primitive approach would be to convolve the data by a smoothing kernel. This is a well-posed problem, but first it will only decrease the noise level, and second it can compromise significant features of the data.

A better approach is to view it as an Bayesian approximation problem: Take the operator  $F$  in (3.1.7) as the identity operator and  $\mathbf{R}$  as a derivative operator of some order. Minimizing (3.1.7) with respect to  $x$  will then give a smooth approximation to the data, where the conflicting requirements of smoothness and fit to the data are weighted by the regularization parameter  $\alpha$ . If the noise level is known, Morozov's principle will provide a well-defined solution.

The scope of this example can be extended to the case of finite resolution of the measurement process (for optical measurements, this can follow from the object being out of focus). Mathematically, this corresponds to the measured data being a convolution of the actual spatially varying quantity with a resolution kernel. In principle, the inverse operator is known, it is just the convolution with the kernel that is given by inverting the resolution kernel pointwise in the Fourier domain. This problem is ill-posed, however, because this deconvolution kernel will grow unbounded at high spatial frequencies, and noise in the data at these frequencies will be accordingly amplified. The solution consists again in regularization. As data come always with some noise, it is actually most natural to view the problem as a combined denoising/deblurring problem, i.e. as an instance of (3.1.7), with  $F$  the blurring operator.

Sometimes it can be the case that only the form of the blurring kernel is known or assumed (such as Gaussian), subject to some unknown parameters (such as the width). This problem is known as blind deconvolution. It can be modelled as a Bayesian hierarchical model, where one considers not only the to-be-determined quantity  $y$  as unknown, but also the operator's parameters, the prior information about which being given by an additional prior distribution.

## 3.3. Reconstructing 3D-densities from projections

The problem of reconstructing three-dimensional densities from lower-dimensional projections is probably most familiar from medical applications such as computerized X-ray tomography (CT) or positron emission tomography (PET, see Sect. 2.1.3). In fact, "tomography" implies that the problem is reduced to determining two-dimensional densities from projections along lines, which correspond to slices through the three-dimensional data. I will review the classical results here. For a more pedagogical treatment see, e.g., Feeman (2010).

### 3.3.1. Direct transform methods — The Radon transform and its inversion

Johann Radon (1917) was the first who defined and solved the mathematical problem of determining a two-dimensional function from its integrals over all lines (he also treated the higher-dimensional generalization of integrals over hyperplanes). Consider the Radon transform of a two-dimensional function  $f$

$$p(\theta, s) = (\mathcal{R}f)(\theta, s) = \int dl f(\cos(\theta)s - \sin(\theta)l, \sin(\theta)s + \cos(\theta)l). \quad (3.3.1)$$

This corresponds to the integrals over the set of all lines in the plane, described by  $\theta$ , the enclosed angle to the second coordinate axis, and  $s$ , the distance from the origin.

There exist a number of analytical expressions for the inverse operator  $\mathcal{R}^{-1}$ , of which the probably easiest follows from the projection-slice theorem. For its derivation observe that the Fourier transform of  $f$  along the first coordinate axis is

$$\mathcal{F}(f)(q_1, 0) = \int dx_1 dx_2 f(x_1, x_2) e^{-iq_1 x_1} = \int dx_1 p(0, x_1) e^{-iq_1 x_1} = \mathcal{F}(p(0, \cdot))(q_1), \quad (3.3.2)$$

that is just the Fourier transform of the projections orthogonal to the first coordinate axis. For any higher-dimensional Fourier transform an orthogonal transformation of coordinates results in the Fourier coordinates being subjected to the same transform (as this leaves the inner product in the Fourier exponential invariant), therefore the above expression can be generalized to the projection-slice theorem: The two-dimensional Fourier transform of  $f$  evaluated along any slice through the origin is equal to the one-dimensional Fourier transform of the Radon transform of  $f$  evaluated for the corresponding angle

$$\mathcal{F}(f)(q \cos(\theta), q \sin(\theta)) = \mathcal{F}(p(\theta, \cdot))(q). \quad (3.3.3)$$

A possibility for inverting the Radon transform is therefore to Fourier transform the projections with respect to the radial coordinate  $s$ , interpolate the values star-like on a Cartesian grid, and transform back. This method is called direct Fourier inversion. In higher dimensions it allows generalizations in multiple ways, for instance in three dimensions it either connects the Fourier transform over the integrals over parallel planes with the Fourier transform evaluated on the orthogonal line through the origin, or it connects the Fourier transform over the integral over lines with the three-dimensional transform evaluated on the orthogonal plane.

A different algorithm can be obtained by applying the operator

$$\frac{1}{(2\pi)^2} \int_0^\pi d\theta \int_{-\infty}^\infty dq |q| e^{iq(x_1 \cos \theta + x_2 \sin \theta)} \quad (3.3.4)$$

to both sides of (3.3.3). This operator is obviously just the inverse two-dimensional Fourier transform in polar coordinates (albeit with a non-canonical coordinate domain), therefore the left-hand side gives  $f(x_1, x_2)$ . For the right-hand side it can be interpreted as multiplication by the modulus of the radial coordinate followed by a one-dimensional inverse Fourier transform evaluated at  $q_1 \cos \theta + q_2 \sin \theta$ , which leads to

$$f(x_1, x_2) = \frac{1}{2\pi} \int_0^\pi d\theta \left( \mathcal{F}^{-1} \left( |\cdot| \mathcal{F}(p(\theta, \cdot)) \right) \right) (x_1 \cos \theta + x_2 \sin \theta). \quad (3.3.5)$$

Defining the back-projection operator

$$(\mathcal{B}g)(x_1, x_2) = \frac{1}{\pi} \int_0^\pi d\theta g(\theta, x_1 \cos \theta + x_2 \sin \theta) \quad (3.3.6)$$

this can be written more succinctly as

$$f(x_1, x_2) = \frac{1}{2} \mathcal{B}_{(\theta, r) \rightarrow (x_1, x_2)} \left( \mathcal{F}_{t \rightarrow r}^{-1} \left( |t| \mathcal{F}_{s \rightarrow t} (p(\theta, s))(\theta, t) \right) (\theta, r) \right), \quad (3.3.7)$$

where for clarity all transformed coordinates have been indicated. This method is called filtered back-projection, as it filters the data in the radial coordinate with a Fourier filter that has the absolute value function as frequency response, and then back-projects them to the Cartesian plane, where the action of back-projection can be understood as distributing all data points in the projections additively back on the projection lines they correspond to. Incidentally, the operator  $\mathcal{BR}$  without filtering is equal to a convolution with  $1/\sqrt{x^2 + y^2}$ , which would give yet another method to invert the Radon transform, albeit at the price of having to de-convolve a two-dimensional kernel.

The last method to be mentioned here uses the fact that multiplying the Fourier transform of a function by  $it$  is equal to transforming the first derivative of the function. Therefore, the filtering employed in (3.3.7) is equivalent to a filter with a frequency response of  $i$  times the sign function applied to the negated derivative with respect to the radial coordinate. Such a filtering is equivalent to the Hilbert transform  $\mathcal{H}$ , which leads to the expression

$$f = -\frac{1}{2} \mathcal{B} \mathcal{H} \nabla_s p. \quad (3.3.8)$$

By the way, the Hilbert transform can be equivalently written as the distributional convolution with  $1/x$  in the Cauchy principal value sense.

All these methods have in common that they are exact inverse operators to the Radon transform in a continuous setting. They are called direct transform methods because they give an explicitly computable solution in terms of integral transforms. In order to use them for actual applications, additional aspects have to be considered, such as that measured data are discrete so that a coordinate transform between Cartesian and polar coordinates necessitates interpolation, and that noise has to be treated robustly, which for the filtered back-projection can be fulfilled by using a band-limited filter instead of  $|t|$  (and equivalently for the other filtering methods). Still, for medical applications they are preferably employed, primarily because they give the result directly without much computational effort, in contrast to the general case for algebraic reconstruction techniques, as discussed in the next section.

However, for the case of three-dimensional reconstruction in ACPAR, the case looks different: in contrast to the cylindrical shape of the human body, which lends itself to sectioning approaches<sup>1</sup>, for the two-photon momentum distribution it is in general not reasonable to restrict oneself to projection directions in a given plane. In fact, in the cases of high symmetry a given direction has a number of equivalent directions, so that automatically there is information also about out-of-plane projection directions. As a consequence, there is a conflict of dimensions: while the functions that the Radon transform and its inverse act upon are three-dimensional

<sup>1</sup>Actually, it would even be impossible to measure the absorption of a scan line from the head to the toes with justifiable radiation doses.

### 3. Mathematical concepts

(either by considering the Radon transform for integrals over planes or on independent sections), the data that are in principle accessible are four-dimensional (two pixel dimensions per projection direction, and two dimensional freedom in choosing the projection direction), but can be sampled only rather sparsely (a small number of projection directions). Using direct transform methods would therefore entail neglecting a large amount of available information and, at the same time, having to interpolate large ranges of the data.

#### 3.3.2. Algebraic reconstruction techniques

An alternative way to the direct transform methods are algebraic reconstruction techniques<sup>2</sup>. Where the former use analytical properties of the Radon transform to give an explicit expression of its inverse, for the latter the Radon transform is just an instance of a general operator  $F$  connecting the unknowns  $x$  with the data  $y$  in the sense of Sect. 3.1.2. In general, arriving at the solution is computationally harder than the discrete Fourier transforms that are to be computed for the direct transform methods, but the general formulation allows all the concepts presented in Sect. 3.1 to be applied, such as a rigorous treatment of noise and prior information. Also note that in contrast to the direct transform methods these techniques can cope with situations where some projections are missing, such as for tomosynthesis (Dobbins and Godfrey, 2003) or when metal implants in the patient render certain projections useless (Boas and Fleischmann, 2011), or, on the other hand, where some aspects of the problem are over-determined. For all these reasons, algebraic reconstruction techniques are to be preferred for ACPAR reconstructions.

The Radon transform is a linear operator. Therefore, after choosing some basis for the solution space and a Tikhonov regularization functional, the maximum a posteriori estimate corresponds to minimizing (3.1.7) with a linear operator  $F$

$$x^* = \arg \min_x (\mathbf{A}x - y)^\top \mathbf{W}(\mathbf{A}x - y) + \alpha x^\top \mathbf{R}^\top \mathbf{R}x \quad (3.3.9)$$

The solution is attained where the gradient of above expression with respect to  $x$  is the zero vector. Expanding the expression, using the fact that  $\nabla_x (w^\top x) = \nabla_x (x^\top w) = w$ , that as a covariance matrix  $\mathbf{W}$  is symmetrical and employing the Leibniz product formula leads to

$$0 \stackrel{!}{=} \nabla_x f(x) \Big|_{x^*} = 2(\mathbf{A}^\top \mathbf{W} \mathbf{A} x^* - \mathbf{A}^\top \mathbf{W} y + \alpha \mathbf{R}^\top \mathbf{R} x^*) \quad (3.3.10)$$

or

$$(\mathbf{A}^\top \mathbf{W} \mathbf{A} + \alpha \mathbf{R}^\top \mathbf{R}) x^* = \mathbf{A}^\top \mathbf{W} y. \quad (3.3.11)$$

This expression is called the normal equation (or normal equations when understood as one scalar equation per datum), reducing the problem to the field of linear algebra, which justifies the name of the technique. The solution to the normal equation is conveniently written as

$$x^* = (\mathbf{A}^\top \mathbf{W} \mathbf{A} + \alpha \mathbf{R}^\top \mathbf{R}) \setminus (\mathbf{A}^\top \mathbf{W} y), \quad (3.3.12)$$

---

<sup>2</sup>Note that the nomenclature here is perhaps a bit unfortunate. What I call here algebraic reconstruction techniques is canonically called iterative reconstruction techniques or series expansion methods (Herman and Lent, 1976), of which more narrowly defined algebraic reconstruction techniques are a strict subset (Gordon, Bender, and Herman, 1970). Oddly, though, in principle the solution does not need to be arrived at iteratively, nor is it obvious in which sense a series would be expanded here.

where the backslash operator means multiplication by the inverse from the left. For the actual implementation any algorithm for numerically solving a system of linear equations can be chosen. For sufficient  $\alpha$ , the problem will be well-conditioned.

### 3.3.3. Cormack's method

Unaware of Radon's work, Allan McLeod Cormack (1963) gave a different solution to the tomographic projection problem, with the explicit intention to apply it for medical imaging. In the initial formulation he expanded both the unknown density  $f$  and the projections  $p$  into Fourier series for each choice of the radial coordinate and gave an explicit expression for a given  $r$ -dependent Fourier component of  $f$  in terms of an integral equation involving the corresponding Fourier component of  $p$ , still in the vein of the direct transform methods. In a later modification, however, he expanded  $f$  also in the radial coordinate into a series of orthogonal polynomials and showed that the Radon transform of these basis polynomials gives again orthogonal polynomials (Cormack, 1964). In other words, he gave an explicit expression for the singular value decomposition of the Radon transform, and therefore a simple recipe for its inversion. For this work, he received the 1979 Nobel Prize for Medicine and Physiology, together with Godfrey Newbold Hounsfield, who built the first working computerized X-ray tomograph.

This approach is somehow intermediate between the direct transform methods and algebraic reconstruction techniques: It can be seen as an instance of the latter, where the bases in projection and reconstruction space have been chosen so that the operator matrix  $A$  is diagonal. Restricting the basis set implicitly regularizes the problem due to the orthogonality of the polynomials, and it allows for non-constant weights to be specified when expanding the data in the basis set. The disadvantages of the method consist in the fact that, analogous to the direct transform methods, the orthogonality of the basis is conditional on the projection angles being evenly spaced and that it again is a strictly tomographical method, i.e. that it can include information only from projection directions in a single plane. In this sense, the algorithm dictates how the measurements have to be made. In addition, it is only stable for quite small bases (Cormack, 1964) and is therefore unable to describe sharp edges.

Related methods that expand the density in spherical harmonics (and therefore can enforce given three-dimensional point symmetries) have been derived by Mijnaerends (1967) for plane integrals and Pecora (1987) for line integrals.

## 3.4. Segmentation

The other major point to be considered for Fermi surface determination from ACPAR data, apart from the reconstruction of three-dimensional densities from the projections, is how to determine the Fermi surface(s) from the three-dimensional densities. As detailed in Sect. 2.4.9, the two-photon momentum density shows a step of varying height at each Fermi surface sheet, with a higher density in the regions inside the Fermi surface than outside, and is smooth away from the Fermi surface. In the community of applied mathematics, the equivalent problem of partitioning a domain into subdomains along discontinuities is known as (image) *segmentation*,

### 3. Mathematical concepts

here transferred to three-dimensional space.

#### 3.4.1. Description of shapes

In the most general setting, the edges are just a general set  $\Gamma \subset \mathbb{R}^3$ . However, such a general description is not amenable to analysis, therefore alternative descriptions that give edges that are inherently regular in some sense are preferred. I will present the two main approaches in the following (Delfour and Zolésio, 2011):

In the one view, the edge shapes themselves are the fundamental variables. They can be either defined directly by a specific parametrization on a two-dimensional domain, or they can be obtained by applying deformations defined on three-dimensional space to a given initial shape (such as a sphere). In either case, the induced topology of the resulting shape (e.g. whether it has holes or not) is obviously fixed, given either by the domain of the parametrization or the initial shape. Another approach is to consider functions on three-dimensional space  $f_i(\vec{k})$  as the fundamental variables and define the edges implicitly as the solution set of  $f_i(\vec{k}) = 0$ . This concept is known as level set formulation (Osher and Sethian, 1988).

Principally, all these mentioned models can be implemented both in a parametrized way (so that the actual free variables are, for instance, Fourier coefficients that describe the deformation or the level set function, respectively) or in a generic parameter-free way (describing the functions by a numerical discretization that is updated in some way). Also, the regularity of the resulting shapes has to be guaranteed, either implicitly by the construction of the parametrization (such as by restricting the number of considered Fourier components) or explicitly by a regularization functional on the respective functions in the parameter-free case.

Apart from questions of implementation, the various approaches differ qualitatively in the kinds of shapes they can describe: Level sets in the basic formulation can describe only manifolds that are topologically closed (but not necessarily compact). In contrast, treating the shapes explicitly also allows for holes in the shapes. On the other hand, continuous perturbations of the level set functions can lead to changes in the topology of the edges, such as that one manifold splits into two disconnected manifolds, which clearly cannot happen via continuous perturbations of the shapes themselves. However, explicit shape parametrization, in contrast to the other approaches, is able to smoothly lead to a manifold that intersects itself. By generalizing above approaches, a multitude of qualitatively different situations can be described: First, it is obvious that several manifolds that behave independently can be obtained trivially by a union of independent manifolds. More interesting is the possibility of prescribing a given shape topology by subjecting an initial shape composed of a number of manifolds (such as intersecting spheres) to a common deformation. In the level set approach, nested shapes follow from considering different levels of the same level set function. The problem of segmentation in the narrow sense of the word (dividing a domain  $\Omega$  into subsets  $R_i$  with  $\Omega = \bigcup_i R_i$  that overlap only on the boundary) is probably most naturally treated by considering one level set function  $E_i$  for each subdomain and defining  $R_i = \{x \in \Omega | f_i(x) \leq f_j(x) \forall j \neq i\}$ . Finally, also complicated situations can be described, e.g. a formulation  $\Gamma = \{x \in \Omega | f_1(x) = 0 \wedge f_2(x) \leq 0\}$  allows for holes in the shapes.

These different characteristics of the respective methods suggest that for a given problem the a priori knowledge on the shapes leads to an evident choice of the framework to be preferred.

### 3.4.2. Driving the agreement — cost functionals

For determining an edge set that describes the data well, the agreement has to be quantified. For the problem of finding edges  $\Gamma$  within otherwise smoothly varying data  $g$  essentially two basic approaches are mainly being used, both originally proposed for computer vision. On the one hand there is the Mumford-Shah functional (Mumford and Shah, 1989). Its idea is to find a parsimonious edge set so that the data can be well described away from the edges by a smooth function. Specifically it reads

$$E(f, \Gamma) = \mu^2 \int_{\Omega} (f - g)^2 + \int_{\Omega \setminus \Gamma} \|\nabla f\|^2 + \nu |\Gamma|, \quad (3.4.1)$$

where  $|\Gamma|$  denotes some concept of measure, e.g. the Hausdorff measure. The parameters  $\mu$  and  $\nu$  decide on the weights given to the conflicting demands, which can be understood via the extreme cases of dropping one of the three terms: if  $\mu$  is zero, the minimum will be attained where  $f$  is any constant function and  $\Gamma$  the empty set, while if  $\mu$  and  $\nu$  go to infinity, the resulting  $f$  will be equal to  $g$  and again  $\Gamma$  will be the empty set. Finally, the requirement for parsimony follows from the insight that if  $\nu$  is zero,  $\Gamma$  will tend towards the largest set within the considered space (if it exists), and  $f$  will be piecewise constant on the subdomains cut out by  $\Gamma$ . Minimizing this functional with respect to  $f$  and  $\Gamma$  gives an optimal set of edges  $\Gamma$ .

Above expression corresponds to the original one given by Mumford and Shah (1989). They prove that in their original two-dimensional formulation the minimum of expression (3.4.1) actually exists, when  $\Gamma$  is allowed to vary over all finite sets of  $C^2$ -arcs. Obviously, the approach can be generalized, such as by considering measures of agreement other than the  $L^2$ -norm, for instance the  $L^1$ -norm, leading to total variation regularization (Rudin, Osher, and Fatemi, 1992), or criteria for smoothness different from a small norm of the first derivative. For the actual implementation, often an approach along the lines of Ambrosio and Tortorelli (1990) is followed, where the edges are again described implicitly via a smooth function.

A different idea was followed by Kass, Witkin, and Terzopoulos (1988): they proposed to use a functional

$$E(\Gamma) = \mu \int ds |\Gamma'(s)|^2 + \nu \int ds |\Gamma''(s)|^2 - \int ds |\nabla g(\Gamma(s))|^2 \quad (3.4.2)$$

that considers only the edges  $\Gamma$  as independent variables and actively searches for the regions of highest gradient in the data  $g$ . In fact, above choice for the data-dependent term is only one of several possibilities considered in the original publication. Again, the measure of the edge set (first term in (3.4.2)) and its curvature (second term) have to be regularized. The authors termed shapes subject to this cost functional *snakes* due to the way they converge to the edges during shape optimization.

The different formulations correspond to different notions of edges: while for the Mumford-Shah functional the optimal edges separate regions of different average  $g$  (in order to minimize the necessary gradient in  $f$ ), snakes will converge to sharp features. Obviously, the former concept will show a much better resistance to small-scale uncorrelated noise such as present in count-rate limited experiments, while snakes will be preferable for applications such as computer vision, where noise is of no concern.

### 3.4.3. Shape determination

For determining the edges that optimally describe the data, in principle two conceptual approaches are followed: In the majority of cases one considers incremental modifications of the shapes. If the cost functional is defined in local terms (i.e. in terms of integrals) then an intuitive image for this concept is the problem from condensed matter physics where the material can react by viscous flow to forces that result from some energy functional, which is identified with the cost functional of shape determination. Variational calculus then gives descent directions, which can be interpreted as local velocities in the physical model and be used to propagate the shape (Santosa, 1996; Delfour and Zolésio, 2011). This concept is called *shape optimization*.

This way of incrementally optimizing the shapes by small deformations is applicable to both explicit descriptions of the shapes as well as level set descriptions (where the values of the level set functions flow according to the velocity field). The more natural implementation would be in a parameter-free way, leading to the standard problem of numerically solving a partial integro-differential equation on a discretization grid, although with more effort also an application with parametrized shapes/level set functions is conceivable. Such a time evolution will converge to a locally optimal solution; standard schemes such as starting from different initial conditions (in this case corresponding to different shape topologies) can be used to find the global optimum with some certainty.

While this approach naturally gives the “best” shapes for given data, more work would have to be done for obtaining further statistical information, such as the estimated variance due to noise in the data. Conceivably, a way to generalize this approach would be to add a stochastic term to the force equation, leading to a Langevin-type equation for the evolution of the shapes (Langevin, 1908). Varying the magnitude of the stochastic term so that the mean data misfit corresponds to the experimental noise would in the end lead to a sequence of representative reconstructions, although additional analysis would be needed to obtain the functional form of the prior probability distribution such an approach corresponds to. Also, it is obvious that the obtained sample reconstructions are highly correlated, indicating the inefficiency of such a method.

Especially for applications where the results are to be determined in a Bayesian sense, i.e. as posterior probability distributions, treating the problem as an instance of a general inverse problem seems to be superior. Here a parametrized description will be preferable. Standard approaches for sampling the posterior distribution of the corresponding parameter vectors as discussed in Sect. 3.1.5 can then be used.

## 4. ACPAR data interpretation

With the exposition of the physics behind the problem and selected mathematical approaches relevant to its solution in the previous chapters, I will in this chapter first review past approaches in the physical community, then state my chosen formulation, and finally discuss its implementation.

### 4.1. Discussion of previous approaches

Here I will give a survey over representative approaches to some of the points considered above that have been employed by the ACPAR community, and discuss their specific shortcomings. The experimental technique of two-dimensional ACPAR originated in the 1980s as discussed in Sect. 2.1.2, a time where digital computing was still much less powerful than today. In the meantime, there have been no notable experimental break-throughs, which perhaps explains that the used data interpretation approaches still seem to confine themselves to those past restrictions. A general remark pertains to all published approaches: the different aspects of the problem (deconvolving the experimental resolution, reconstruction, edge detection) are always approached in a sequential manner, which is undesirable as different biases are introduced at each step, compromising the later steps.

For a recent review of ACPAR articles using three-dimensional reconstruction methods (which nowadays comprise the largest part of ACPAR publications) with some claim of completeness see Kontrym-Sznajd (2009).

#### 4.1.1. Data prettification

The primary data as obtained in an ACPAR experiment are not the abstract projections as assumed by the mathematical algorithms presented in Chap. 3. The initial data handling that is conventionally applied is rather uninspiring and therefore mostly detailed in PhD theses such as Biasini (1995) and Kruseman (1999). In most cases, the experimental data take the form of two-dimensional histograms of the deviations from collinear propagation directions as detected on the two detectors, of a size of a few hundred bins squared. The initial steps consist normally in centring these histograms (so that the exactly collinear events, which can be off by a few pixels due to miscalibration of the detector positions, end up in the centre of the histogram) and rotating (so that a high-symmetry direction of the crystal ends up for instance in the exactly vertical direction, again to correct miscalibrations on the order of one degree). It is clear that these operations require interpolation and therefore lead to a resolution loss on the order of one pixel and destroy the Poissonian statistics of the data. However, with current communication and storage capacities it is possible to record all events on a PC instead of online

#### 4. ACPAR data interpretation

histogramming, so that the histogram can be computed afterwards on the properly transformed event coordinates without loss of resolution (Leitner, Ceeh, and Weber, 2012).

Another aspect is the so-called momentum sampling function. Because of the finite size of the detectors, the probability for a pair of photons with a given deviation of propagation directions to fall on the respective detectors depends on the deviation: for a collinear pair it suffices if one photon reaches its detector for the other to reach it also (under the assumption of symmetric detector positions), while for non-zero deviations the probability decreases. The resulting momentum sampling function is given by the convolution of the respective detector response functions and is normally used to normalize the histograms to give (noisy) projections in the mathematical sense. Of course, the absolute error of the outermost pixels is greatly amplified in this way.

If the projections are along high-symmetry directions of the crystal, they should reflect this symmetry. In this case the data are often folded to conform to this symmetry, which decreases the pixel-wise relative statistical error, but introduces additional correlations between the values. Also, it is conceivable that such a symmetrization amplifies systematic errors, if the actual alignment of the crystal deviates from the nominal high-symmetry one.

The last step of initial data handling that is sometimes applied concerns the resolution. With the present detectors, statistical and systematic errors in position assignment, additional to the anisotropic contribution due to the positron spot size, are quite large compared to the features to be resolved (Leitner, Ceeh, and Weber, 2012). As discussed in Sect. 3.2, correcting the data for these effects is an ill-posed problem. The used approaches rely on maximum-entropy regularization (Fretwell et al., 1995) or related concepts (Gerhardt et al., 1998). However, there is no a-priori reason why the entropy of the projections should be small, in fact it is clear that the projections will have the highest intensity in the centre and decay over some four orders of magnitude towards large deviations, in contrast to the a-priori assumption of a multi-nomial distribution with equal values for each pixel that justifies the maximum-entropy approach (see the example in Sect. 3.1.4). A still more elaborate approach that expressly smooths the data is given by Chiba et al. (2007). It is needless to say that the properties of the resulting statistical and systematic errors in the corrected projections will be virtually impossible to state. Finally, Kruseman (1999) proposes a diametrically opposed solution to the problem of anisotropic resolutions, which is to worsen the resolution in the vertical dimension by convolution with a smearing kernel in order to avoid introducing artefacts in deconvolution.

##### 4.1.2. Interpretation of spectra

In the early times, the interpretation of the data was done directly in terms of the projections, after having cleansed them of above-mentioned effects. Here two ways have been demonstrated. The first is based upon the observation that the actual signal of modulations in the conduction electron density projections due to the Fermi surface breaks is dwarfed by the more-or-less isotropic signal of filled bands, which has its maximum at the centre and decays outwards. The solution is then to subtract the radial average from the projections so that the anisotropic contributions, which should be mostly due to conduction bands, remain. An example is given by Haghighi et al. (1991), where faint anisotropic signatures in a high-temperature superconductor have been isolated that supposedly conform well to theoretical predictions.

The starting point of the second approach lies in the non-interacting particle expression for the two-photon momentum density (2.4.14) for the case of a constant positron wave function  $\psi_p$ . Each electron wave function  $\psi_e$  has an  $L^2$ -norm of one (as the absolute square corresponds to its location probability density), therefore by Parseval's theorem its Fourier transform is also normalized. Due to its Bloch form (2.4.11), its contribution to the two-photon momentum density is discrete, located at the positions of the suitably shifted reciprocal lattice. If now one folds all the higher-momentum components back into the first Brillouin zone, it will sum to one. As a consequence, the contribution of filled bands after this backfolding is constant in reciprocal space, and all the modulation is due to the jumps in occupation at the Fermi surface sheets. This statement is called the LCW theorem (Lock, Crisp, and West, 1973).

For enhancing the edges filtering techniques have been proposed, such as the method of O'Brien et al. (1995), which is effectively a band-limited second derivative, or an effective abuse of the maximum-entropy resolution deconvolution mentioned above (Dugdale et al., 1994). The rationale behind these approaches lies in the fact that while for a realistic positron wave function and many-body effects the contribution of a filled band is not exactly constant, its variation after backfolding will be smooth, so that the filtering techniques enhance the edges compared to these smooth variations.

For layered systems like the cuprate superconductors studied by Haghighi et al. (1991), where the relevant aspects of the electronic structure are effectively two-dimensional, it may be a valid approach to interpret directly the projection orthogonal to the layers. For general systems, however, the best one can hope for with such an interpretation is to be able to identify the most prominent features qualitatively. More information can be obtained by the reconstruction techniques to be discussed in the next section.

### 4.1.3. Reconstruction

Nowadays, in most cases there is the ambition to reconstruct the full three-dimensional information from the results of an ACPAR experiment, where algorithms along the lines of Sect. 3.3 are used. I will discuss the used methods here only generally, for a comprehensive list of the pertinent publications see Kontrym-Sznajd (2009).

Direct Fourier inversion in the two-dimensional formulation, the filtered back-projection and Cormack's method or variants thereof have been used with approximately equal shares in nearly all published ACPAR reconstructions. In these methods, the problem of the determination of a three-dimensional density is reduced to independently determining the density on parallel planes. Such an approach is sub-optimal even in the absence of symmetries, such as for medical imaging, as in general the three-dimensional density to be determined will show as high smoothness within the slices as between them (which is neglected by independent reconstructions). For ACPAR, where the density will display the strict symmetries required by the crystallinity, this is even more severe, as discussed at the end of Sect. 3.3.1. In principle it is possible to average the reconstructed data so that they display the symmetry, although this propagates the between-slices non-regularity into all dimensions. Kontrym-Sznajd et al. (2004) give a somewhat circuitous solution to this problem: they reconstruct the density in slices and then regularize these data by reparametrizing in terms of the spherical harmonics that are in accordance with the crystal symmetries. Another disadvantage of these methods lies in

#### 4. ACPAR data interpretation

their requirement of specific projection directions: deviations within the reconstruction plane, perhaps due to misalignment, make the reconstruction algorithm only more complicated, but these algorithms are in principle not able to treat an out-of-plane deviation.

A problem specific to the direct transform methods lies in the fact that different from computerized X-ray tomography, where the maximum absorption is on the order of 50% and the Poisson noise level has therefore a comparable magnitude for each data point, in ACPAR the density varies by orders of magnitude, and the absolute noise level varies accordingly. For direct transform methods the weighting can only be specified locally, however, and therefore they cannot account for this fact.

The original rationale behind all these historic methods (either direct transform methods or expansion in suitable basis function) was that they allowed for an efficient numerical implementation (in terms of discrete Fourier transforms or the solution of small, sparse systems of linear equations). Even in recent publications claims on the prohibitive inefficiency of treating reconstruction as an instance of a generic inverse problem along the lines of algebraic reconstruction techniques as presented in Sect. 3.3.2 can be found (e.g. Feeman, 2010). While this may be true for medical applications, in ACPAR the acquisition of a set of projections of high statistical quality takes on the order of a few weeks. Thus it seems highly warranted to employ the most potent algorithms for reconstruction, even if these take noticeable numerical effort.

Only very recently such an algebraic approach has been used for the reconstruction of ACPAR data (Pylak, Kontrym-Sznajd, and Dobrzyński, 2011). However, the unphysical tendency of their maximum-entropy formulation towards flatness of the reconstructed densities has been discussed already above.

##### 4.1.4. Fermi surface determination

In principle, the Fermi surface should give rise to sharp steps in the reconstructed three-dimensional density. Actually, however, these edges are only poorly defined: band-limited direct transform methods and reconstructions by expansion in finite series cannot describe edges and lead to a smoothing (additional to the effect of finite resolution, if this is not corrected for), while in contrast a maximum-entropy reconstruction does not favour smoothness, so that the statistical noise in the data will be reflected in voxel-to-voxel noise in the reconstructions. In most cases the reconstructed three-dimensional density is therefore interpreted only qualitatively in terms of the topology of the Fermi surface.

A few attempts have been made at quantitative Fermi surface reconstructions: Biasini (2000) proposed a phenomenological parametrization of the Fermi surface and computed directly the LCW-folded density for a few high-symmetry projections in the non-interacting particle model with a constant positron wave function. He then determined the Fermi surface parameters by minimizing the deviations from the LCW-folded experimental projections. He conceded that positron wave function and many-body effects (i.e. that the LCW theorem is not exact in this case) will lead to systematic errors.

Later he also proposed a radically different, model-free approach (Biasini et al., 2002): after reconstructing the three-dimensional density, he defined the Fermi surface as the iso-contour to a certain density level  $\rho_0$  so that the variation of the enclosed volume with respect to  $\rho_0$  relative

to the surface area is minimized (i.e. that the mean steepness on the surface is maximized). However, also this approach depends on the LCW theorem being valid, as otherwise the Fermi surface is no iso-contour at finite resolutions. This is reflected in the fact that the maximum in the steepness is not very well defined.

Probably the most involved published method for determining Fermi surfaces is due to Laverock et al. (2010): they calculated the electronic wave functions by ab-initio methods, and then fitted the resulting LCW-folded projections of the two-photon momentum densities to the experimentally measured projections by applying both shifts in energy to the distinct bands and varying the contribution of the states (which should mimic a different enhancement for different atomic states). While it is questionable in how far the fitted parameters are physically meaningful, it can be seen as just a phenomenological model that tries to reproduce the sharp edges due to the Fermi surface and the smooth variations due to positron wave function effects and many-body effects. However, such an interpretation cannot be counted any more as experimental physics, because it requires the calculated band structure to capture the physics of the system, as the only freedom in the shape of the Fermi surface sheets lies in the rigid band shift.

## 4.2. The proposed model

Here I motivate a model to solve the physical problem as outlined in Sect. 2.5.

A major shortcoming of the previous approaches lies in the fact that they comprise many sequential steps of data manipulation. An exemplary chain of data handling could consist in deconvolving the resolution out of the projections (with some regularization, perhaps including explicit smoothing), centring and rotating the projections, dividing by the momentum sampling function, symmetrizing the projections, reconstructing by some series expansion, symmetrizing the reconstructed three-dimensional density, applying some edge-enhancing filter, and finding the Fermi surface. Propagating the estimated statistical and systematic errors through these steps seems impossible, and at some steps there are even additional systematic biases introduced due to regularization (such as deconvolving, reconstructing and filtering). In contrast, the formulation as a general inverse problem allows to concatenate all these steps into the forward operator, which is then solved under explicit regularization by prior information.

A formulation as a general inverse problem will also be more efficient with respect to the information obtainable with a given statistical investment during the experiment (i.e. the radiation dose for computerized X-ray tomography, and the data collection time for count-rate limited applications such as ACPAR): Radon has shown that the set of all line integrals determines the density in a plane, therefore sectioning approaches work. Experimentally, however, the accessible data are the line integrals in three-dimensional space (apart from geometrical constraints due to the cylindrical shape of the human body in medical tomography). Direct inversion methods in principle cannot be generalized to such situations, as the inverse problem would be overdetermined. For algebraic reconstruction techniques this is possible, however, and as will be discussed in more detail in Sect. 5.5, the matrix to be inverted in a linear inverse problem will be better conditioned if the available data is more independent, resulting in less uncertainty in the reconstruction with a given level of statistics in the data. For the reconstruction of densities with certain known three-dimensional symmetries, the advantage

#### 4. ACPAR data interpretation

of a general formulation is even more obvious. Additionally, such an approach allows also for modelling unknown misorientations during the experiment (treating the corresponding Euler angles as additional variables).

Here I propose to formulate the model by explicitly taking the steps at the Fermi surface, together with the smooth variation in between, into account. Specifically, the unknowns are to consist of a geometrical variable  $\Gamma$  (the Fermi surface) and a functional variable  $\rho$  (the density). I will allow for spatially varying densities, but require smoothness by explicit regularization in terms of the  $L^2$ -norm of the discrete second derivative away from the Fermi surface (corresponding to an  $l^2$ -norm regularization of the densities' Fourier components with a weight proportional to the square of the reciprocal coordinate), as dictated by physics. In such an approach already the reconstruction drives the evolution of the Fermi surface so that the features of the projections can be described by densities that are as smooth as possible. The model can also be seen as a generalization of the Mumford-Shah problem (3.4.1), where in the first term not the discrepancy between fitted data  $f$  and measured data  $g$  is quantified, but between  $F(\rho)$  and  $g$ , where  $F$  is the projection operator and  $\rho$  the density. Such a model has already been proposed for the concurrent inversion and segmentation of X-ray tomography data by Ramlau and Ring (2007) and Klann and Ramlau (2013), but with piece-wise constant densities away from the edges. In contrast, the tendency to maximize the steps at the edges of a formulation guided by the energy functional for (two-dimensional) snakes would contradict physical a priori knowledge.

For the description of the Fermi surface a level set formulation, as opposed to an explicit parametrization, is preferable. This is because the Fermi surface physically *is* the level set of the band energies, see Sect. 2.4.7, and therefore a level set description leads to the correct topology (such as no holes and no self-intersecting). For the general case of multiple Fermi surface sheets, the correct positions and crossing behaviours with respect to each other can be guaranteed by enforcing that the band energies conform to the required symmetries, which is conceptually easier. Also, the physical a priori knowledge about the Fermi surface shapes follows from the knowledge on the behaviour of the band energies, which leads to a much more natural Bayesian interpretation of regularization.

More specifically, it can be expected that the level set functions (corresponding to the band energies) will be smooth in the sense that they can be described by a limited number of Fourier components, which motivates a corresponding parametrization. Physically, this corresponds to a tight-binding formulation (see Sect. 2.4.5). The crystalline symmetry of the band energies can readily be enforced by restricting the set of non-zero Fourier components. Also, for a quantitative determination of the Fermi surface shape in ACPAR the topology of the Fermi surface can be assumed as known. If this topology is enforced in another way, it is therefore not necessary for the Fermi surface area to be regularized, as the Fourier-limited level set functions suffice to guarantee regular edges in accordance with the physical requirements.

A main difference of the considered problem to the general Mumford-Shah model is the fact that the contributions from the distinct bands are strictly positive. The Fermi surface sheets have a well-defined orientation, and when crossing from the outside to the inside the density will jump by a positive amount. Therefore (and for implementational reasons detailed in Sect. 4.3.3) it is actually beneficial to consider for each band that crosses the Fermi surface an unknown positive density defined on the whole reciprocal space, additional to a density due to core bands.

The sum density is then the sum over all band densities multiplied by the characteristic functions of the regions inside the respective Fermi surface sheets.

In contrast to the more popular case of image processing, where the result will be interpreted qualitatively by human vision and pattern recognition, with physical experiments the uncertainty is an integral constituent of quantitative results. This implies a Bayesian framework, where a representative sample of Fermi surface realizations in agreement with the data is to be produced. As already discussed in Sect. 3.4.3, this practically excludes the possibility to employ shape optimization, so the problem will be treated as an instance of a general inverse problem. Taking in all above considerations, its formulation can be cast in

$$f(\rho, \Gamma) = -\log(P(b|\mathbf{A}_\Gamma\rho)) + \alpha\|\mathbf{D}_2\rho\|^2, \quad (4.2.1)$$

where  $f(\rho, \Gamma)$  is the negative logarithm of the posterior probability,  $b$  are the measured data,  $\rho$  the band densities,  $\mathbf{A}_\Gamma$  is the linear projection operator taking into account all the experimental parameters and the occupations via  $\Gamma$  (described over the Fourier coefficients),  $P$  the probability for observing  $b$  under Poissonian statistics, and  $\mathbf{D}_2$  the discrete Laplacian operator. The regularization parameter  $\alpha$  can be determined via Morozov's principle, as the noise level is known due to Poisson statistics. Marginalizing the resulting posterior distribution with respect to  $\rho$  leads directly to the resulting distribution of  $\Gamma$ , which is the objective of the experiment. Additionally, an explicit dependence of  $\mathbf{A}_\Gamma$  on small misalignments or unknown resolution can be included. Depending on the information content of the data, these additional degrees of freedom can either be determined, or they lead to a wider posterior distribution in  $\Gamma$ , treating possible systematic errors in a statistically rigorous way.

### 4.3. Implementation

Here I will present the numerical implementation of the model proposed in Sect. 4.2 in some detail. It was accomplished for the most part in GNU OCTAVE, which is an interpreted high-level framework for numerical mathematics. The implementation is quite efficient, which for an interpreted language means that apart from initialization much data should be processed per instruction in the innermost loops (this concept is known as vectorization). The most critical operations have been implemented in low-level code (see Sect. A.3).

As a demonstration, in Chap. 5 the algorithms will be applied to the case of copper, which is a face-centred cubic system with one Fermi surface sheet. For efficiency of labour, these parameters have been hard-coded into the implementation given below; a generalization to different symmetries or more Fermi surface sheets would be trivial but time-consuming.

#### 4.3.1. Scales of the reconstruction problem

The principal scale of the reconstruction is set by the numbers  $N_e$  (e for extended), which gives the number of reconstructed cubic voxels per dimension, and  $N_c$  (c for cell), the number of voxels per dimension of the reciprocal space cubic unit cell (note that in the case of a face-centred lattice, the cubic cell is not primitive). This formulation restricts the unit cell dimensions to be a multiple of the voxel dimension, which is beneficial for the implementation. The projections are

#### 4. ACPAR data interpretation

discretized into square pixels, with  $N_p$  pixels per dimension (p for projection). The ratio  $s$  of the projection pixel size to the voxel size can be specified. Values close to one are desirable (as for  $s \ll 1$  resolution is lost, while for  $s \gg 1$  the reconstruction is unnecessarily hard without gain in information), but for the reconstruction of experimental data a deviation from one will be necessary, because the pixel size is fixed by the detector, and the voxel size has to be some fraction of the unit cell. For a more efficient formulation, the densities are described as a linear combination of localized basis functions, with  $N_b$  degrees of freedom per dimension (b for basis).  $N_e$ ,  $N_c$  and  $N_b$  are restricted to even numbers, and the formulation is symmetric, i.e. there are  $N_e/2$  voxels per dimension on each side of the centre of reconstructed space. Due to the cubic symmetry, the irreducible wedge contains only the 48th part of the whole space, which corresponds to  $N_i = \frac{1}{48} N_b(N_b + 2)(N_b + 4)$  inequivalent coefficients (i for irreducible). The total number of unknowns describing the density for the case of one Fermi surface sheet is therefore  $2N_i$ .

For the demonstration of the algorithm in Chap. 5, the values have been chosen as  $N_e = 144$ ,  $N_c = 48$ ,  $N_p = 144$ , and  $N_b = 36$ , leading to  $N_i = 1140$  independent coefficients per band. With these values the reconstruction allocates about 1.5 GB of main memory, for much larger problems some optimization would have to be done.

#### 4.3.2. Matrix product form of forward operator

The forward operator (mapping the coefficients  $\rho$  describing the band densities to the resulting projections) for a given Fermi surface and experimental parameters (such as orientations and resolutions) is linear, so it can be described as the product of matrices corresponding to the distinct steps, acting on the density vector. Each item in the list below corresponds to one such matrix.

- Apply the operations of the point symmetry group to the coefficients of the basis functions in the irreducible wedge to restore the full description of the first octant. This corresponds to a block diagonal matrix with the same  $(N_b/2)^3$ -by- $N_i$  matrix two times on the diagonal (as the action on both band densities is the same).
- Compute the fine-grained densities on the octant via a  $(N_e/2)^3$ -by- $(N_b/2)^3$  matrix, again as two blocks of a block-diagonal matrix. For the localized basis functions a variety of choices are possible, here a tent-like form corresponding to the fine-grained densities being tri-linear interpolations between values on a coarser grid with mirror boundary conditions has been chosen.
- Multiply the densities pointwise by the occupation of the respective voxels in the respective bands, given by the Fermi surface geometry, and collapse to a sum density. This is effected by the horizontal concatenation of two  $(N_e/2)^3$ -by- $(N_e/2)^3$  diagonal matrices. For the density due to the filled core bands, this is just the identity matrix, while for the conduction band it is a diagonal matrix that has as its entries the fractional occupation of the respective voxels. For the computation of these occupations see Sect. 4.3.3.
- Mirror this sum density on one octant to the other octants by a  $N_e^3$ -by- $(N_e/2)^3$  matrix.

- Project the three-dimensional densities along the chosen projection directions, implemented as the vertical concatenation of  $N_p^2$ -by- $N_e^3$  matrices. For their definition see Sect. 4.3.4.
- Convolve the projections with the resolution kernel (modelled as Gaussian with distinct horizontal and vertical component) with periodic boundary conditions, which is also implemented as multiplication by a block diagonal matrix with identical  $N_p^2$ -by- $N_p^2$  matrix entries. In order to ensure sparsity, the Gaussian is truncated after three standard deviations.
- Multiply the projections by the respective momentum sampling functions, which corresponds to a diagonal matrix.

Finally, add a constant background to the projections, which for notational simplicity can be considered as an additional degree of freedom of the densities and propagated through above matrix products. All these matrices are very sparse, and even their product, which corresponds to the forward operator, has only about 25% filling. Code for the smaller functions mentioned above is given in Sect. A.1.

The regularization operator in (4.2.1) is implemented as discrete Laplacian acting with mirror boundary conditions on the fine-grained densities on the octant (its code is also given in Sect. A.1), i.e.  $D_2$  is to be understood in the following as the product of this Laplacian and the matrices corresponding to the first two items in the above list.

### 4.3.3. Describing the Fermi surface and determining band occupations

The occupations of the voxels for the respective bands are determined by the fraction of the voxel that is within the Fermi surface. For a face-centred cubic lattice, the atoms sit at positions  $\frac{1}{2}(m_1, m_2, m_3)$ , where the  $m_i$  are integer numbers with  $\sum_i m_i$  even (see Fig. 2.3 for an illustration). With the definition (2.3.1), the corresponding reciprocal lattice is given by  $2\pi(n_1, n_2, n_3)$  with either all  $n_i$  even or odd. This is the body-centred cubic lattice. A function that is invariant with respect to translation by any reciprocal lattice vector (i.e. periodic within the first Brillouin zone), such as the band dispersion, can have only Fourier coefficients for the reciprocal lattice of the reciprocal lattice, that means again the direct space face-centred lattice. If the function in addition has to display the point symmetry of the lattice, the Fourier coefficients for lattice vectors that can be related by a point symmetry operation have to agree.

With these considerations the band dispersion can be described by a vector of Fourier coefficients  $f_i$  corresponding to an ordered list of inequivalent direct space lattice vectors  $\vec{s}_i$ , specifically  $\vec{s}_0 = (0, 0, 0)$ ,  $\vec{s}_1 = (\frac{1}{2}, \frac{1}{2}, 0)$ ,  $\vec{s}_2 = (1, 0, 0)$  and so on, so that

$$f(\vec{q}) = - \sum_i \gamma_i \sum_{\vec{s} \equiv \vec{s}_i} e^{-i\vec{q}\vec{s}_i}. \quad (4.3.1)$$

Here two vectors  $\vec{s}$  are considered equivalent if they can be mapped onto each other by a point symmetry operation.

The Fermi surface  $\Gamma$  is the level set of this function so that the volume enclosed within the Fermi surface is correct (half the volume of the Brillouin zone per conduction electron). The

#### 4. ACPAR data interpretation

corresponding implicitly defined energy is the Fermi energy. As the value of  $\gamma_0$  would only shift the Fermi energy, it can be set to zero. Additionally, the resulting level set is not affected by a scalar multiplication of the coefficient vector, so  $\gamma_1$  can be set to one. The additional entries of  $\gamma$  correspond to the Fermi surface shape's degrees of freedom.

For calculating the occupations of the voxels within the cubic cell from a given coefficient vector  $\gamma$  the energy and its gradient at the centres of all voxels are computed via (4.3.1), implemented as fast Fourier transform, and as an initial guess the Fermi energy is set to zero (corresponding to the mean energy over the Brillouin zone, as  $\gamma_0 = 0$ ). Then the corresponding voxel occupations are computed via the Gilat-Raubenheimer method, see Sect. A.2. These occupations are summed and compared to the required value, and the Fermi energy is updated by Newton's method. The necessary derivatives of the occupations are computed as by-product of the Gilat-Raubenheimer method.

The Gilat-Raubenheimer method computes for each voxel the fraction within the Fermi surface analytically under the assumption of a linear variation of the band energy within the voxel. If the discretization is fine enough to capture the Fermi surface shape accurately, then the linear approximation is clearly valid. Note that in this formulation the sum density varies continuously with the Fermi surface coefficients  $\gamma$  in any  $L^p$ -norm with  $p < \infty$  (but not point-wise) in a continuum setting. In the numerical discretization it is obviously continuous even voxel-wise, which is in contrast to a formulation that checks only the value at the voxel centre and sets the whole voxel to zero or one. This simple implementation is made possible by considering distinct densities for each band; a formulation of one density with allowed edges would have to specify two densities only for those voxels that are touched by the Fermi surface, which would be much more complex to implement correctly.

#### 4.3.4. The projection matrices

The problem of constructing the projection matrices, that is, to which projection pixels and with which weights each voxel in a given cubic grid in general orientation contributes, is analogous to the problem solved by the Gilat-Raubenheimer method. To be specific,  $G(\omega)$  in Sect. A.2 specifies the fraction of a cube oriented along the coordinate axes that is on one side of a given plane in arbitrary orientation, or equivalently the fraction of a cube in arbitrary orientation that is below a plane normal to a coordinate axis. In other words, it is the exact analytical solution for the problem of discrete plane projections (or rather slice projections) of a three-dimensional density described as a function that is piece-wise constant on cubic voxels. It is evident that the exact analytical solution for the discrete line-projection problem is much harder, as both the number of cases to be considered is much larger and the distinct expressions are functions of two variables. Therefore this problem is solved here only approximately.

It is more convenient to consider the projection plane fixed and the cubic grid to be in some arbitrary orientation, described by Euler angles. For an approximate solution first a single cube is subdivided into small cubes. Then the  $x$  and  $y$  coordinates of the centres of the small cubes are two-dimensionally histogrammed (with comparable resolution). Actually, for high-symmetry orientations of the cube it is advisable to apply dithering on a very small scale to avoid moiré patterns, as it is possible that planes of the cube centre lattice end up directly at the histogram bin boundaries in exact arithmetic. This histogram is then summed up over both dimensions

to give an approximation to the cumulative distribution function of the projection of the cube. Then, for each voxel and for each projection pixel that has an overlap with the projection of the voxel the pixel corners relative to the projected voxel centre are mapped to coordinates into the look-up matrix, the corresponding values are computed by linear interpolation, and the weight by which the voxel contributes to the pixel is obtained by cross-wise subtraction of those values. These operations are implemented as vectorized OCTAVE code with only a single loop over the distinct projections, but as a consequence the code is rather lengthy and will not be reproduced here.

The approximation can be made arbitrarily accurate by a dense sampling without much additional cost in computing time, additionally the projection matrices are afterwards convolved with the resolution kernel anyway, so that the described algorithm can be considered as numerically exact for practical purposes.

#### 4.3.5. Solution of the linearized subproblem

Taking up the expression for the negative logarithm of the posterior probability (4.2.1), it can be seen that approximating the Poissonian distribution by a Gaussian with a covariance matrix  $\mathbf{C}$  assumed as known leads for given  $\gamma$  and experimental parameters to a quadratic expression in  $\rho$

$$f(\rho, \gamma) = \frac{1}{2}(\mathbf{A}_\gamma \rho - b)^\top \mathbf{W}(\mathbf{A}_\gamma \rho - b) + \alpha \|\mathbf{D}_2 \rho\|^2, \quad (4.3.2)$$

where  $\mathbf{W}$  is the inverse of the covariance matrix  $\mathbf{C}$  and  $\mathbf{A}_\gamma$  is the product matrix form of the forward operator detailed in Sect. 4.3.2, depending on the Fermi surface coefficients  $\gamma$  and the experimental parameters. As discussed in Sect. 3.3.2, the maximum a posteriori estimate of  $\rho$  is then

$$\rho_\gamma^* = (\mathbf{A}_\gamma^\top \mathbf{W} \mathbf{A}_\gamma + \alpha \mathbf{D}_2^\top \mathbf{D}_2) \setminus (\mathbf{A}_\gamma^\top \mathbf{W} b). \quad (4.3.3)$$

For Poissonian statistics  $\mathbf{C}$  is diagonal and can be approximated by its maximum-likelihood estimate given the data, which is just the measured number of events per pixel  $b$  (or for better statistics a locally averaged version thereof, assuming smooth variations of the projections underneath the noise), where special care has to be taken for pixels with a low number of events, in which case the maximum-likelihood estimate gives too small values that would bias the reconstruction (see the discussion in Sect. 3.1.1). A consistent alternative choice would be the posterior mean under the assumption of a flat prior, which gives an estimate for the variance of  $b + 1$  for  $b$  counts. For a rigorous solution of (4.2.1) see Sect. 4.3.7 below.

The computational complexity for solving this problem lies mainly in the computation of the product  $\mathbf{A}_\gamma^\top \mathbf{W} \mathbf{A}_\gamma$ : For varying the Fermi surface  $\Gamma$  while keeping the other experimental parameters fixed, the determination of the voxel occupations is comparatively easy (about 0.2 CPU-seconds on a contemporary PC), while all other matrices in Sect. 4.3.2 can be precomputed. The part of  $\mathbf{A}_\gamma$  that corresponds to the filled bands can also be precomputed, and the sparse matrix products leading to the other part take about 2 CPU-seconds. For the computation of  $\mathbf{A}_\gamma^\top \mathbf{W} \mathbf{A}_\gamma$  it is actually more efficient to define  $\mathbf{B} = \sqrt{\mathbf{W}} \mathbf{A}_\gamma$  and compute  $\mathbf{B}^\top \mathbf{B}$ , so that the symmetry of the result is explicit to the algorithm. Still, even when computing only the  $\gamma$ -dependent blocks and exploiting the symmetry, this is the numerically most costly step. A discussion of an efficient implementation in C is given in Sect. A.3, which computes the

#### 4. ACPAR data interpretation

product on four processor threads in about 3 seconds. The final solution of the linear system of equations takes then only about half a second, so that the maximum a priori estimate for  $\rho$  and consequently the optimal fit to the data for a given  $\gamma$  can be computed in about 7 seconds.

##### 4.3.6. Sampling the Fermi surface

The unknowns in the present formulation of the ACPAR reconstruction problem are the band density coefficients  $\rho$  and the Fermi surface coefficients  $\gamma$ . The different nature of these two vectors of unknowns suggests a two-step approach: For given  $\gamma$ , the  $\rho$  most compatible with the data are given by (4.3.3). The form of the conditional probability density of  $\rho$  given by (4.3.2) is a multivariate Gaussian, therefore it follows that the probability density for  $\gamma$  marginalized over  $\rho$  is proportional to the joint probability evaluated at  $\rho_\gamma^*$  multiplied by the determinant of the covariance matrix of  $\rho$  given  $\gamma$ , and its negative logarithm is given by

$$f(\gamma) = f(\rho_\gamma^*, \gamma) + \log(\det(\mathbf{A}_\gamma^\top \mathbf{W} \mathbf{A}_\gamma + \alpha \mathbf{D}_2^\top \mathbf{D}_2)) \quad (4.3.4)$$

up to an immaterial constant. It is not obvious whether this expression can be considered in good approximation as quadratic, actually if  $\rho$  was fixed and resolution was perfect, above expression would not be guaranteed to be anything more than continuous in  $\gamma$ , although the freedom in  $\rho$  will smooth this dependence.

For sampling the posterior distribution of  $\gamma$  defined as above in a model-free way, the Metropolis-Hastings algorithm is used (see Sect. 3.1.5). The resulting sample of the Fermi surface coefficients is then used to compute samples of some dimensions of the Fermi surface by solving (4.3.1) for the Fermi energy in distinct directions of  $\vec{q}$ . The expected values and standard deviations of these samples constitute the final result of the experiment.

##### 4.3.7. Further sophistications

There are a variety of conceivable aspects how the basic model given above can be improved. Here I will discuss some of those, their influence on the behaviour and results of the algorithm will be given in Sect. 5.3.

The first point concerns the choice of the localized basis functions used to describe the density. As described in Sect. 4.3.2, in the basic formulation these are piecewise tri-linear, corresponding to simple linear interpolation. As the actual densities to be reconstructed are smooth, such a choice seems suboptimal. Smoother basis function can easily be obtained with the present implementation by performing the linear interpolation successively: for the present case of  $N_e/N_b = 4$ , writing the  $(N_e/2)^3$ -by- $(N_b/2)^3$  linear interpolation matrix as the product of a  $(N_e/2)^3$ -by- $(N_e/4)^3$  and a  $(N_e/4)^3$ -by- $(N_e/8)^3$  matrix corresponds to piecewise tri-quadratic basis functions.

The next point to consider lies in the approximation of the Poisson distribution by a Gaussian distribution. Using the actual expression for the likelihood of observing the given data  $b$  when in the mean  $\mathbf{A}_\gamma \rho$  are expected in the expression (4.3.2) would also obviate the problem that the covariance matrix  $\mathbf{C}$  is unknown. The necessary analysis corresponds to a Taylor expansion of

the logarithm of the Poissonian likelihood  $g(\rho) = \log(P(b|\mathbf{A}_\gamma\rho))$  in (4.2.1)

$$\nabla g(\rho - \rho_0) = \mathbf{A}_\gamma^\top \left( \frac{y}{\mathbf{A}_\gamma \rho_0} - 1 \right) - \mathbf{A}_\gamma^\top \mathbf{W}(\rho_0) \mathbf{A}_\gamma (\rho - \rho_0) + O((\rho - \rho_0)^2) \quad (4.3.5)$$

with  $\mathbf{W}(\rho) = y/(\mathbf{A}_\gamma \rho)^2$ , where the divisions and squares are understood as entrywise. This expression can be applied to find the zero of the gradient iteratively. With this non-Gaussian formulation, the marginal posterior distribution for  $\gamma$  is not any more proportional to the likelihood of the maximum a posteriori estimate times the determinant, so this aspect has to be checked also.

Lastly, it could be beneficial to regularize not the Laplacian of the densities, but of the logarithm of the densities. While above formulation corresponds to the prior information that the absolute curvature of the densities should be small, regularizing the logarithm would mean that the relative curvatures are small, which is physically more plausible. It would have the additional benefit that the reconstructed densities are strictly positive, a property which is not guaranteed by above formulation. Again, this would necessitate to solve for the zero of the gradient iteratively.



## 5. Results

Here I will apply the Fermi surface reconstruction algorithm as proposed in Sect. 4.3 to simulated realistic data. To be specific, I will first present the method used to compute the model and the resulting densities and projections. Then I will discuss general observations with respect to the reconstruction and test the effects of modifying certain aspects of the reconstruction algorithm. Further I will evaluate the proposed reconstruction model in terms of the expected accuracy of the Fermi surface dimensions obtained from experimental data, due to both statistical and systematic effects. Unfortunately, such an evaluation seems to be lacking for the previous approaches in Fermi surface reconstruction. Finally, I will discuss the implications for the experiment, i.e. how for a given measuring time the optimal accuracy in the reconstructed Fermi surface can be reached.

### 5.1. Simulated data

The electronic structure of copper was computed by the free software package ABINIT version 6.12.3 (Gonze et al., 2009). This is a Kohn-Sham density functional theory code (see Sect. 2.4.4) that describes the valence single-particle wave functions in the Fourier domain, and the effect of the ionic cores in the projector-augmented wave formulation (Torrent et al., 2008). For the exchange-correlation functional a generalized gradient approximation (Perdew, Burke, and Ernzerhof, 1996) was used, and the self-consistent calculations were converged with respect to the density of points in the Brillouin zone and the considered Fourier components. The 1s, 2s and 2p states were treated as core states. The principal point here, however, lies not in a claim to quantitatively reproduce the physical properties of copper, but rather to give qualitatively plausible model data on which to evaluate the performance of the Fermi surface reconstruction algorithm.

The calculation of ACPAR spectra is no native feature of ABINIT, but it was not hard to implement: In order to compute the lifetime of positrons in matter, ABINIT can self-consistently compute also positron states for a given electronic density, and the plane-wave nature of the description of electron and positron states is very amenable to further treatments. To be specific, with the electron density obtained from the self-consistent calculation the positron wave function in the lowest band at the  $\Gamma$  point (i.e. with a Bloch wave vector equal to zero) was computed in the limit of vanishing positron concentration (as is appropriate for delocalized positrons, in contrast to the case of a defect). Additionally the electron wave functions and energies were computed non-self-consistently for a dense grid within the Brillouin zone, and the Fourier components of the electron-positron wave function were derived as given in (2.4.14), i.e. in the independent particle model. Then these densities together with the band energies were linearly interpolated from  $96^3$  to the final  $144^3$  grid, the occupations were computed from

## 5. Results

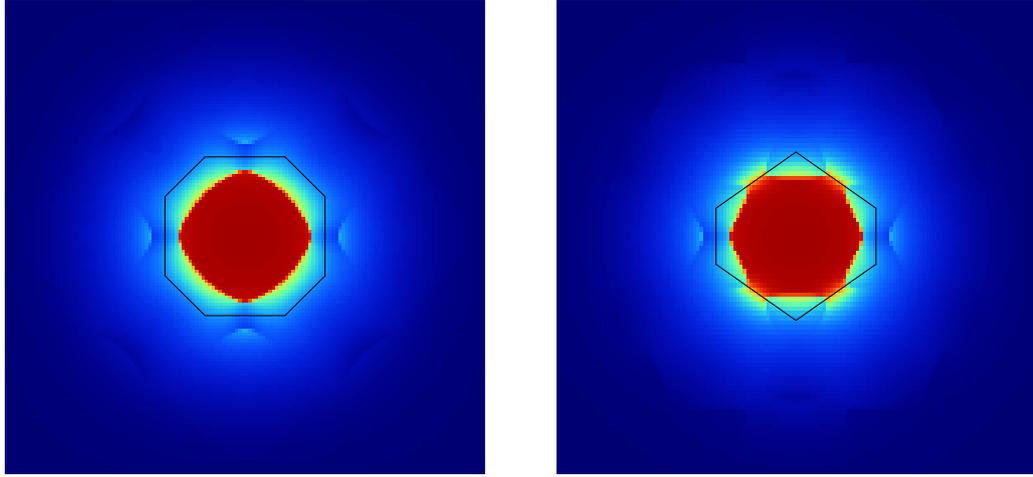


Figure 5.1.: Cuts through the simulated two-photon momentum density on a (100)-plane (left) and a (110)-plane (right) through the origin, with the boundaries of the first Brillouin zone superimposed.

the energies by the Gilat-Raubenheimer method, and the densities were summed over the bands, weighted by the occupations. Two cuts through the resulting three-dimensional two-photon momentum densities are given in Fig. 5.1. The sharp edges due to the Fermi surface are evident inside the first Brillouin zone together with some aspects of the Fermi surface in higher Brillouin zones (compare the Fermi surface rendering in Fig. 5.4). Away from the Fermi surface, the variation of the density is smooth, with the highest variation in the Fermi surface necks that cross the Brillouin zone boundary in the (110)-cut.

As described in Sect. 4.3, these three-dimensional densities were then projected along a given direction, convolved with the instrumental resolution (here assumed as a Gaussian with a realistic standard deviation of two pixel horizontally and one pixel vertically) and used to simulate realizations of the Poissonian process for the respective pixels. Except where noted otherwise, for the reconstruction projections along (100), (110) and (111) were used, each with the realistic value of  $3.3 \times 10^7$  counts. Two of these simulated noisy projections are displayed in Fig. 5.2.

For a rendering of the calculated Fermi surface see Fig. 5.4. It shows the distinct features of the Fermi surface of copper (along with those of the other noble metals), which are the bulges along the (100)-directions (towards the square parts of the Brillouin zone boundary) and the necks in (111)-direction, where the Fermi surface crosses the Brillouin zone boundary (to be periodically repeated in the neighbouring Brillouin zones).

For the further quantitative discussion it is convenient to define the length-scale via the radius of the free-electron Fermi sphere (that is the sphere that encloses the correct reciprocal space volume, in this case half the Brillouin zone)  $r_f = \sqrt[3]{3/16\pi}$  in reciprocal lattice units (i.e. the distance between opposing squares of the Brillouin zone boundary). The variation of the conduction band energy over reciprocal space is notably smooth, which is reflected in the fact that the Fermi surface can be satisfactorily described with only three non-trivial Fourier coefficients (the root-mean-square deviation of the so-parametrized Fermi surface from the

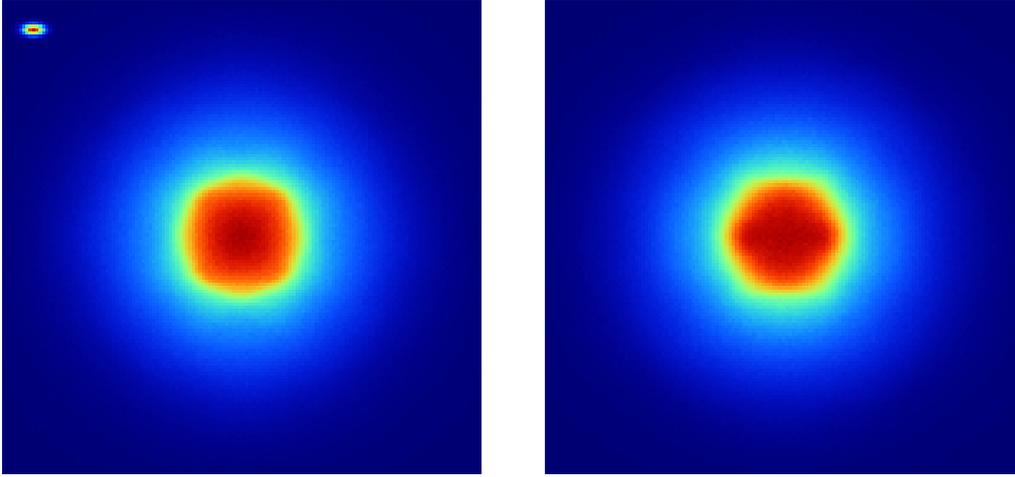


Figure 5.2.: Projections of the two-photon momentum density along (100) (left) and (110) (right). Orientations and scale are equal to Fig. 5.1. The assumed resolution kernel is given in the corner of the left panel. The pseudo-colour range is  $[0, 15000]$ , corresponding to  $3.3 \times 10^7$  counts per projection.

simulated one is less than  $0.001 r_f$ ). From this it can be concluded that at least for the conduction band around the Fermi energy a tight-binding description captures the essential physics.

The computed occupations correspond to Fermi surface radii along (100) of  $1.063 r_f$ , along (110) of  $0.946 r_f$ , and a neck radius of  $0.203 r_f$ . These values agree surprisingly well with those of Roaf (1962) obtained by fitting de Haas-van Alphen data. In contrast, the agreement of the Fourier coefficients is much worse, as already with three parameters the variations of the Fermi surface with respect to the parameters are nearly linearly dependent. Experimentally, independent information obtained from measurements of the anomalous skin effect (Pippard, 1957) has been used to further constrain the parameters.

## 5.2. General remarks on the reconstructions

Apart from the missing noise, the projections calculated from the sampled Fermi surfaces and corresponding maximum a posteriori density estimates are visually indistinguishable from the data and hence are not given here. In contrast, the reconstructed densities (Fig. 5.3) are obviously different from their source (Fig. 5.1): where in the original densities the part within the Fermi surface in the first Brillouin zone is nearly flat and starts to decay only when passing through the necks, in the reconstructions there is a depression in the zone centre and also around the Fermi surface of about 5%. Additionally, the (100)-cut shows a faint cross in the outer regions, which is reminiscent of the classical pictures of the edges along the projection lines when doing a filtered back-projection reconstruction with too large angular step (e.g. Feeman, 2010). It seems therefore plausible that this artefact is due to a too low number of projections. Perhaps also a more sophisticated regularization functional that promotes isotropic densities in the outer regions, which actually conforms to the physical expectation, could ameliorate the situation, but

## 5. Results

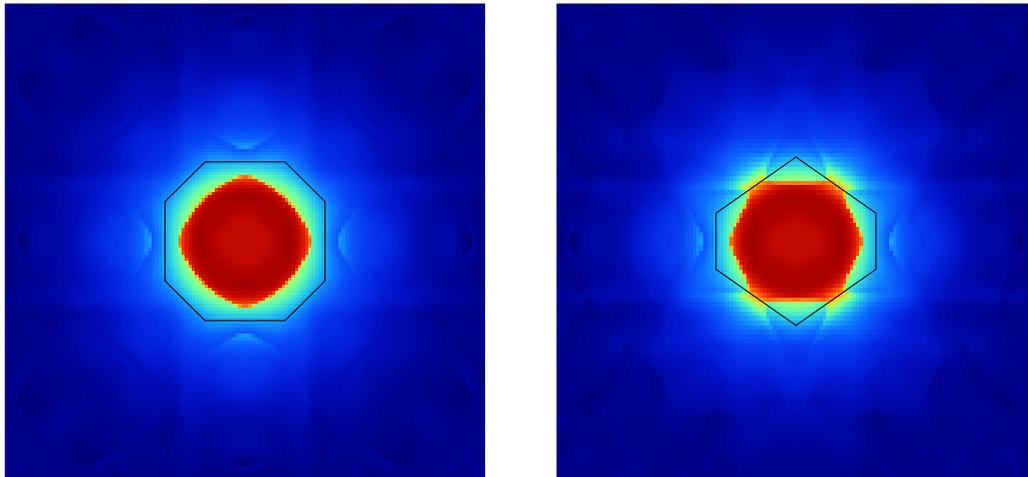


Figure 5.3.: Cuts through the reconstructed two-photon momentum density, same orientations as Fig. 5.1.

I want to stress here that the presented reconstruction at only three projections already shows the superiority of the used formulation compared to classical tomographical reconstruction methods. Apart from these differences in the densities, the primary goal of the reconstruction, that is the Fermi surface, is visually perfectly reproduced.

This assessment is also reflected in a quantitative determination of the Fermi surface dimensions, which evaluate as  $1.064(5) r_f$ ,  $0.949(1) r_f$  and  $0.184(5) r_f$ , corresponding to the radius along (100) and (110) and the neck radius, respectively. The values in parentheses are the standard deviation in the last significant digit over the reconstructions from different realizations of the Poisson noise. A comparison with the original values shows that the agreement is quantitatively very good, although significant differences remain, where the relative error is most prominent in the neck radius.

As mentioned already above, however, the first three free Fermi surface parameters are nearly linearly dependent, therefore they are not well defined and the covariance matrix of the generated samples has a single large eigenvalue<sup>1</sup>. There are two possibilities for treating such a situation: On the one hand it follows from physics that the envelope of the Fourier components will decay quite fast towards higher indices. Therefore it would be justified to add a regularization on the values of the Fourier components, which would concentrate the distribution in the region of the original ellipsoid nearest to the origin. A different way would be to consider more free Fourier components, calculate the derivatives of the Fermi surface shape with respect to these components, and choose the modes corresponding to the highest singular values. This approach could be expected to make the sampling of the distribution most efficient, but is not grounded in physics. For keeping the model as simple as possible, here neither of these possibilities was followed, instead the results are reported in terms of the principal dimensions of the Fermi surface, which are well defined as discussed above.

<sup>1</sup>By the way, at this statistical level of the data, the posterior probability density of  $\gamma$  is constrained to such a small region in parameter space that its logarithm is in a good approximation quadratic in this region, so the question raised in Sect. 4.3.6 can be answered in the affirmative.

### 5.3. Influence of model aspects

Here I will report the effects of varying details of the model as proposed in Sect. 4.3.7.

The localized basis functions used above were obtained by repeated linear interpolation, corresponding to piecewise tri-quadratic functions. In fact, when using tri-linear basis functions, this lesser degree of smoothness would be visually recognizable in the reconstructed densities. This can also be quantified: Using a regularization parameter  $\alpha$  that gives a  $\chi^2$  of about 62300 depending on the realization of the Poisson random variables (where the target value as per Morozov's principle would be  $3N_p^2 = 62208$ ) in the Gaussian approximation with the quadratic basis functions yields a  $\chi^2$  of about 63700 with the linear basis functions. This shows that the quadratic basis functions can describe the data better and are therefore to be preferred, additional to the fact that they agree better with the physical expectation. The effects on the reconstructed Fermi surface parameters are not drastic, however: With the linear basis functions, the neck radius would be even smaller, but the (100)-radius would be larger, in both cases by about  $0.002 r_f$ .

The next point concerns the difference in computing the marginalized posterior probability density for  $\gamma$  by either modelling the conditional density of  $\rho$  as a Gaussian as in (4.3.4), or taking directly the conditional probability of the maximum a posteriori estimate for  $\rho$ . With the parameters chosen as described above, the logarithm of the determinant scatters only by a value of about one for the sampled  $\gamma$ , and the effect on the computed Fermi surface dimensions is below  $0.002 r_f$ , which is hard to determine significantly. As a consequence, in the present situation it is justified to neglect the corresponding term in (4.3.4). The reason for this is that the data constrain the variation of the Fermi surface to a very small region, where the width of the conditional probability density for  $\rho$  cannot vary much. In other cases, this could be different.

The difference between using the Gaussian approximation to the Poisson distribution of the counts and using the correct expression is equally small. Again, this is due to the fact that the count numbers per pixel are not too small, so that the approximation is justified. Considering this aspect is numerically not hard, however; as the maximum a posteriori densities do not vary much, a single iteration of (4.3.5) for each new  $\gamma$  (taking the estimate at the old  $\gamma$  as starting point) suffices to arrive at the solution.

The last point concerns the possibility of regularizing the logarithm of the densities. This was also tried by solving for the zero of the gradient in the Taylor approximation iteratively, but it proved to be harder than in the case of the correct expression for the Poisson probability. This can be either due to the fact that in contrast to above, here there is no approximation to be used as a starting point, or that the problem is much more non-linear.

### 5.4. Effect of experimental uncertainties

In the actual experimental setting, the forward operator, which maps the densities and the Fermi surface to the projections, depends on additional parameters that are only approximately known. The two most grave uncertainties are expected to be the actual orientation of the sample and the instrumental resolution. The natural way to account for these uncertainties is to model them also in a Bayesian framework by enlarging the set of unknowns and adding respective priors.

## 5. Results

Here I will explore the consequences of the above-mentioned principal effects.

The posterior probability is very sensitive to the resolution kernel used in the reconstruction. Using for the reconstruction a kernel that is a few percent larger than the kernel used for simulating the projections results in a much lower posterior probability. Therefore the reconstructed results are robust with respect to such an uncertainty, as such larger kernels are identified as being very unlikely. On the other hand, the probability increases for smaller kernels up to a minimum at a kernel width of 88% of the correct one. That the data can be better reconstructed by an incorrect forward operator is at first sight surprising, but it can be understood by considering the effect of regularization: for describing a given variation in the projections with a smaller resolution kernel, the densities themselves can be smoother than they would have to be with a wider kernel, which is rewarded by the regularization. For even smaller kernels, the posterior probability drops again, because the Fermi surface effects in the projections would become too sharp.

However, this preference of the model for too small resolution kernels is not critical, because on the one hand, for the interpretation of an experiment the effect can be determined from simulations and added to the prior information, while on the other hand, the simulations with the above-mentioned 88%-wide kernel yield Fermi surface dimensions of  $1.061(4) r_f$ ,  $0.949(1) r_f$  and  $0.186(3) r_f$ , which is indistinguishable from the results with the actual kernel.

For misalignment, the situation is comparable. Experimentally, one is confident that the misalignment is bounded better than within  $1^\circ$  in each Euler angle. The simulations show that misalignments in the angle that rotates the projection around the projection line are easy to spot, half a degree would already lead to a much unlikelier reconstruction, presumably because this would rotate the mirror symmetries in the projections, which however are quite well defined. Misalignments in the other two angles do not affect the posterior probability of the reconstruction much, but again, the Fermi surface dimensions reconstructed with a generous choice of  $1^\circ$  misalignment are  $1.062(8) r_f$ ,  $0.950(2) r_f$  and  $0.183(4) r_f$ , showing no significant deviation.

### 5.5. Implications for the experiment

As the last point, I want to consider the effect of the measurement parameters that are free for the ACPAR experimenter to choose. In contrast to the case of medical imaging, where both geometrical constraints and the used reconstruction method define how the measurement has to proceed, here the number of projections, the projection directions and the respective statistical weight can be chosen freely. In ACPAR, the limiting factor is the count rate (the parameters considered here correspond to overall measurement times of a few days), so the relevant question is how to allocate a certain integral measurement time most economically.

First the question of the effect of the statistical accuracy has to be settled. Here the surprising conclusion has to be drawn that the resulting Fermi surface dimensions are largely unaffected by the integral number of counts in the projections: with a five-fold increase they evaluate as  $1.066(5) r_f$ ,  $0.950(2) r_f$  and  $0.186(6) r_f$ , and with a five-fold decrease as  $1.063(11) r_f$ ,  $0.948(2) r_f$  and  $0.186(7) r_f$ . This stability is due to regularization; it shows that the present reconstructions describe all the information content in above-mentioned three projections even

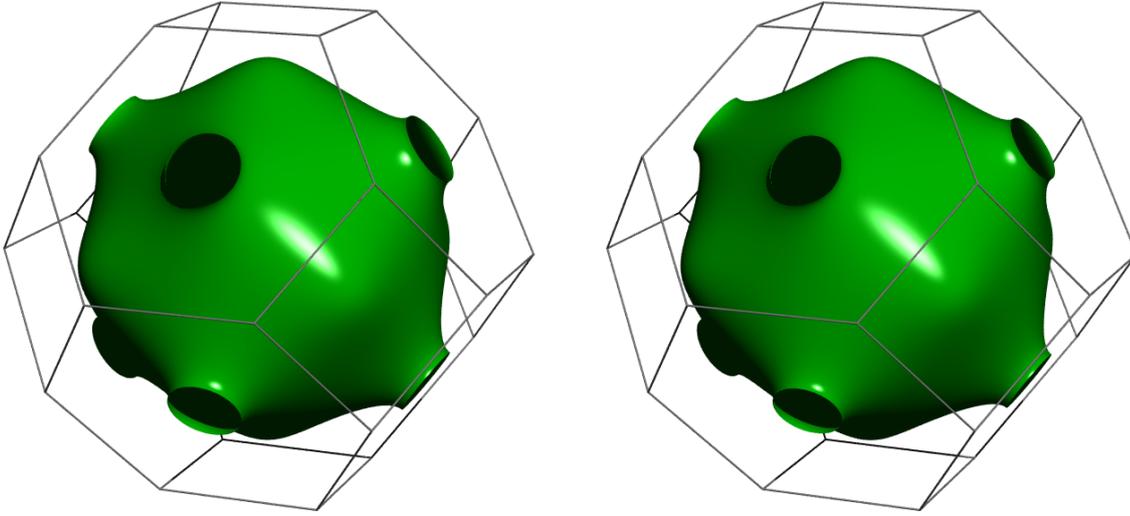


Figure 5.4.: Rendering of portion of Fermi surface within first Brillouin zone, actual ab-initio simulation (left) and representative reconstruction with four projections and  $10^8$  counts (right).

in the absence of noise.

A better reconstruction should be obtained by distributing the counts over a larger number of independent projections. This can be understood by considering (4.3.3): The system of equations is (essentially) underdetermined, therefore  $\mathbf{A}_\gamma^\top \mathbf{W} \mathbf{A}_\gamma$  has eigenvalues equal to zero. The effect of regularization is to raise these eigenvalues away from zero so that the solution becomes stable, at the price that the corresponding components of the solution are not determined by the data, but by the regularization. Actually,  $\mathbf{A}_\gamma^\top \mathbf{W} \mathbf{A}_\gamma$  is a sum over symmetric positive semi-definite matrices corresponding to the different projections. By adding an additional independent projection the relative spread of the eigenvalues will therefore decrease, and less aspects of the solution will have to be provided by the regularization.

It is conceivable that it would require a much larger number of projections to observe a significant improvement of the reconstructed Fermi surface dimensions. A trial with a fourth projection direction at the same overall number of counts resulted in  $1.068(8) r_f$ ,  $0.949(2) r_f$  and  $0.187(3) r_f$ , which is not much different from the previous values, although it comes nearer to the actual value for the (111)-neck, which is consistently reconstructed too small, than all quoted values above. Also, the corresponding densities do not show the back-projection artefact discussed in Sect. 5.2 any more.

## 5.6. Conclusions

Fermi surface determination by ACPAR has always been a rather qualitative method, due to the limitations in the methodological approaches to data interpretation. In this thesis I have proposed a method based on established concepts in applied mathematics with the explicit aim to allow for determining quantitative results on the Fermi surface dimensions. The application to simulated

## 5. Results

data shows that this goal has largely been reached: assuming realistic values for the resolution and statistical accuracy of the projections gives a reconstruction with systematic errors of about one percent of the free electron Fermi momentum, and statistical errors even lower. As has been shown, these reconstructions are stable with respect to errors in the experimental assumptions such as misalignments and inaccuracies in the instrumental resolution. For an actual application to experimental data, the systematic errors can even be reduced by determining an estimate for them from simulations such as was done here, and correcting the experimental dimensions for these errors.

Lastly, the resulting images are perhaps the best demonstration of the quality of the attained reconstruction: Figure 5.4 shows renderings of both the simulated Fermi surface and a representative reconstruction. The visual agreement is unambiguous, only by knowing where to look one can perhaps discern that the reconstructed neck radius is too small.

# A. Appendix

## A.1. Diverse small OCTAVE functions

Here I give selected efficient OCTAVE implementations of functionalities described in 4.3.

```
function Aro=mat_Aro(Ne)
#Propagation from irreducible cubic wedge to octant
[x,y,z]=ndgrid(.5:Ne/2);
[s,i]=sortrows(sort([x(:) y(:) z(:)],2));
Aro=sparse([1 cumsum(diff((s-.5)*[Ne^2/4 Ne/2 1]')!=0)'+1],i,1);

function Ainterpl=mat_lin_blowup(N1,N2)
#linearly interpolates grid of N1^3 to N2^3
[x,y,z]=ndgrid((.5:N2)/N2*N1);
xyz=[x(:) y(:) z(:)];
oxyz=round(xyz);
uxyz=max(0,oxyz-1);
weight=xyz-oxyz+.5;
oxyz=min(oxyz,N1-1);
xo=oxyz(:,1);yo=oxyz(:,2)*N1;zo=oxyz(:,3)*N1^2;
xu=uxyz(:,1);yu=uxyz(:,2)*N1;zu=uxyz(:,3)*N1^2;
weight=[(1-weight(:,1)).*(1-weight(:,2)).*(1-weight(:,3));
         weight(:,1).*(1-weight(:,2)).*(1-weight(:,3));
         (1-weight(:,1)).*weight(:,2).*(1-weight(:,3));
         weight(:,1).*weight(:,2).*(1-weight(:,3));
         (1-weight(:,1)).*(1-weight(:,2)).*weight(:,3);
         weight(:,1).*(1-weight(:,2)).*weight(:,3);
         (1-weight(:,1)).*weight(:,2).*weight(:,3);
         weight(:,1).*weight(:,2).*weight(:,3)];
Ainterpl=sparse(1+[xu+yu+zu;xo+yu+zu;xu+yo+zu;xo+yo+zu;xu+yu+zo;xo+yu+zo;xu+yo+zo;xo+yo+zo],
               (1:N2^3)'(:,ones(1,8))(:),weight);

function Aof=mat_Aof(Ne)
#Propagation from cubic octant to full space
[x,y,z]=ndgrid(-Ne/2+.5:Ne/2);
xyz=[x(:) y(:) z(:)];
Aof=sparse((abs(xyz)-.5)*[Ne^2/4 Ne/2 1]'+1,1:Ne^3,1);

function Afinres=mat_finres(Np,sigma1,sigma2)
#Convolution with resolution kernel
[x,y]=ndgrid(floor(-3*sigma1):ceil(3*sigma1),floor(-3*sigma2):ceil(3*sigma2));
kern=exp(-((x/sigma1).^2+(y/sigma2).^2)/2);kern/=sum(kern(:));
N=prod(size(kern));
[i,j]=ndgrid(0:Np-1);
```

## A. Appendix

```

ind=mod(i(:)(:,ones(N,1))+x(:)'(ones(Np^2,1),:),Np)+...
    mod(j(:)(:,ones(N,1))+y(:)'(ones(Np^2,1),:),Np)*Np+1;
Afinres=sparse(i(:)(:,ones(N,1))+j(:)(:,ones(N,1))*Np+1,ind,kern(:)'(ones(Np^2,1),:),Np^2,Np^2);

function D=mat_D(Ne)
#Laplacian regularization operator on octant
D=sparse(Ne^3/8,Ne^3/8);
index=((1:(Ne/2-1))'*ones(1,(Ne/2)^2)+ones(Ne/2-1,1)*(0:(Ne/2)^2-1))*Ne/2(:);
D+=sparse(index,index,1,Ne^3/8,Ne^3/8)-sparse(index,index+1,1,Ne^3/8,Ne^3/8);
D+=sparse(index+1,index+1,1,Ne^3/8,Ne^3/8)-sparse(index+1,index,1,Ne^3/8,Ne^3/8);
index=((1:(Ne/2)^2-Ne/2))*ones(1,Ne/2)+ones((Ne/2)^2-Ne/2,1)*(0:(Ne/2-1))*(Ne/2)^2(:);
D+=sparse(index,index,1,Ne^3/8,Ne^3/8)-sparse(index,index+Ne/2,1,Ne^3/8,Ne^3/8);
D+=sparse(index+Ne/2,index+Ne/2,1,Ne^3/8,Ne^3/8)-sparse(index+Ne/2,index,1,Ne^3/8,Ne^3/8);
index=(1:(Ne/2)^3-(Ne/2)^2);
D+=sparse(index,index,1,Ne^3/8,Ne^3/8)-sparse(index,index+(Ne/2)^2,1,Ne^3/8,Ne^3/8);
D+=sparse(index+(Ne/2)^2,index+(Ne/2)^2,1,Ne^3/8,Ne^3/8)-...
    sparse(index+(Ne/2)^2,index,1,Ne^3/8,Ne^3/8);

```

## A.2. The Gilat-Raubenheimer method

The problem of computing a density of states is relevant for many fields of solid-state physics. The density of states is defined via

$$d(\omega) = \int_B d\vec{x} \delta(f(\vec{x}) - \omega), \quad (\text{A.2.1})$$

where  $B$  is the Brillouin zone (or any other domain, depending on the problem). Mathematically, an equivalent problem is to compute the probability density of a scalar-valued random variable that is an arbitrary function of a vector-valued random variable with known probability density.

Gilat and Raubenheimer (1966) gave an efficient numerical algorithm for this problem when  $f$  and its gradient can be computed for any given point. It consists in splitting the integration domain  $B$  into cubes (special care has to be devoted to the case where this is not possible due to the shape of  $B$ ), for each cube approximating  $f$  by its linearization at the cube's centre, and computing the integral explicitly. Geometrically, the problem thus corresponds to determining the area of the intersection of a plane with a cube.

Consider now a cube centred at the origin with an edge length of 2 and a linear function  $f(\vec{x}) = x_1 l_1 + x_2 l_2 + x_3 l_3$ . Due to the symmetry of the cubic integration domain, its density of states is invariant with respect to a permutation of the coordinates or a change in sign, therefore we can assume  $l_1 \geq l_2 \geq l_3 \geq 0$ . Define

$$\begin{aligned}
 \omega_4 &= l_1 + l_2 + l_3 \\
 \omega_3 &= l_1 + l_2 - l_3 \\
 \omega_2 &= l_1 - l_2 + l_3 \\
 \omega_1 &= -l_1 + l_2 + l_3
 \end{aligned} \quad (\text{A.2.2})$$

and, for notational convenience

$$k = (\omega_4^2 - \omega_3^2 - \omega_2^2 - \omega_1^2)/2. \quad (\text{A.2.3})$$

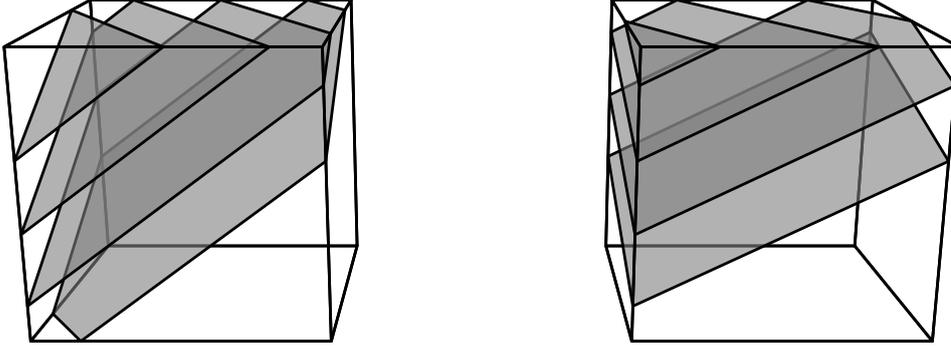


Figure A.1.: Intersections of the iso-planes of  $f$  with the cube for the three cases  $i = 0 \dots 3$  (from bottom to top plane) and the two qualitatively different geometries a (left) and b (right).

Above-defined  $\pm\omega_i$  correspond to the values of  $f$  on the cube corners. Between those points (A.2.1) evaluates to a polynomial in  $\omega$ . Note that apart from  $\omega_1$ , all  $\omega_i$  are positive. The sign of  $\omega_1$  qualifies the linear function  $f$  into two generic cases a (with  $\omega_1 > 0$ ) and b ( $\omega_1 < 0$ ). The geometry of the iso-plane corresponding to a given  $\omega$  is defined by the cases  $i$ :  $|\omega_i| \leq \omega \leq |\omega_{i+1}|$  for  $i = 0 \dots 4$ , with  $\omega_0 = 0$  and  $\omega_5 = \infty$ . For an illustration see Fig. A.1.

It then follows that the density of states over the cube for this linear function is given by

$$d(\omega) = \frac{1}{l_1 l_2 l_3} g(|\omega|) \quad (\text{A.2.4})$$

and its integral

$$\int_{-\infty}^{\omega} d\nu d(\nu) = 4 + \frac{\text{sgn } \omega}{l_1 l_2 l_3} G(|\omega|) \quad (\text{A.2.5})$$

with  $g$  and  $G$  as defined below. Expressions for the degenerate cases where one or more of the  $l_i$  are equal to zero can easily be obtained from these general expressions by cancelling.

Case	$g(\omega)$	$G(\omega)$
0a	$k - \omega^2$	$k\omega - \omega^3/3$
0b	$4l_2 l_3$	$4l_2 l_3 \omega$
1	$\frac{1}{2}k + 2l_2 l_3 - \omega_1 \omega - \frac{1}{2}\omega^2$	$-\frac{1}{6}\omega_1^3 + (\frac{1}{2}k + 2l_2 l_3)\omega - \frac{1}{2}\omega_1 \omega^2 - \frac{1}{6}\omega^3$
2	$2l_3(l_1 + l_2 - \omega)$	$4l_3(l_1 l_2 - \frac{1}{3}l_3^2) - 2l_3^2(\omega - \omega_4) - l_3(\omega - \omega_4)^2$
3	$\frac{1}{2}(\omega - \omega_4)^2$	$4l_1 l_2 l_3 + \frac{1}{6}(\omega - \omega_4)^3$
4	0	$4l_1 l_2 l_3$

The expressions for  $g(\omega)$  have been derived by Gilat and Raubenheimer (1966). In this thesis the algorithm was used for determining the proportion that is within the Fermi surface for a given voxel. The solution to this problem for a given deviation of the energy at the cube centre from the Fermi energy and a given energy gradient follows immediately from the expressions for  $G(\omega)$ . The derivative of these occupations with respect to the energy is obviously just  $g(\omega)$ .

### A.3. Sparse-to-full matrix multiplication

Different from full matrices that are stored element for element, for sparse matrices only the positions and values of non-zero elements are stored. This can be very useful when the scale of the problem (number of equations and unknowns) is large, but each equation couples only a small number of unknowns. Apart from storage, arithmetic operations on sparse matrices can also be performed much faster when only the non-zero elements have to be explicitly treated. If the density of a sparse matrix (i.e. the ratio of non-zero elements to the product of the dimensions) becomes high, however, the efficiency becomes inferior compared to a treatment in terms of full matrices, both in terms of storage (as with sparse matrices for each element also the position has to be recorded), but even more so in terms of run-time for arithmetic operations such as multiplications, as the consecutive order of full matrices allows for much more streamlined data handling (and actually fast matrix multiplication algorithms exist that improve on the time complexity bounds of the naive algorithm).

In 4.3.5, the product of moderately sparse matrices is singled out as the bottleneck of the implementation. To be specific, two matrices with dimensions of about 1000 by 60000 and a density of about 0.25 are to be multiplied along the large dimension, so that the result is a square matrix of 1000 by 1000. It is clear that the product matrix will be full, as the number of scalar products to be calculated is much larger than the product matrices' number of elements. A general-purpose sparse matrix multiplication algorithm will therefore be very inefficient, as it has to generate a well-formed sparse matrix result (which implies explicit bookkeeping on the results of the scalar multiplications, sorting them and eventually adding them together). Converting to full matrices before multiplication would be an alternative, but this needs about 16 times more scalar multiplications (if performed by naive full matrix multiplication) than sparse matrix multiplication (not counting the more involved calculation of memory addresses with sparse matrix multiplication). It therefore seemed that a coherent implementation of a function that takes sparse matrices as input and gives a full matrix as output could outperform either alternative.

The resulting code is available at <http://homepage.univie.ac.at/michael.leitner/sparse.zip> and offered under the GNU General Public License, version 2. It consists of a function for calculating the product of two sparse matrices  $\mathbf{A}$  and  $\mathbf{B}$  so that  $\mathbf{AB}^T = \mathbf{C}$ , where  $\mathbf{C}$  is a full matrix (the transpose allows for a simpler and more efficient implementation when both matrices are given in compressed column storage), and basically the same function for the special case  $\mathbf{B} = \mathbf{A}$ , which is about a factor of two faster by using the symmetry of the product matrix. Both functions are written in C99 for the efficiency of native code created by C compilers and the possibility to link to C libraries from nearly any language. For the case at hand, these functions are called by wrapper functions written in C++ that can be natively compiled to OCTAVE's so-called oct-files.

The algorithm works by successively computing blocks of the resulting full matrix on the order of the L2 cache size, so that neither reading from nor writing to main memory constitute communication bottlenecks. This can be demonstrated by a simple estimation with the numbers quoted above: Take a block size of 250 by 250 elements. With 8 byte doubles this corresponds to about 490 kB, which fits comfortably into the 512 kB L2 cache. Then the algorithm iterates over the columns of the sparse input matrices. If they have random structure at a density of 0.25,

then for both matrices in each column about 60 non-zero elements will contribute to this block, corresponding to about 1.5 kB of data (8 byte doubles for the value and 4 byte integers for the address per element), which fits conveniently into the L1 cache. Each pair of these elements is then multiplied together and added at the corresponding position in the block, corresponding to 3600 times doing one double multiplication and addition, some integer multiplications and additions for the address, and a store to the L2 cache. Fetching the 1.5 kB into L1 cache from main memory will take much less time, and only after all 60000 columns the block has to be stored into the resulting matrix in main memory. This demonstrates that this is probably the most efficient implementation possible for naive sparse matrix multiplication for such densities. Only for extremely sparse matrices, where also the resulting matrix is rather sparse, it would be more efficient to store the scalar multiplication results as a list, and afterwards sort them and merge colliding entries.

Another benefit of the blocked assembly of the resulting matrix is that the algorithm can be parallelized over the blocks without effort. This was done in the framework of POSIX threads with four worker threads, matched to the number of processor threads. The resulting functions are inferior to OCTAVE's stock sparse matrix multiplication at very low densities, both approaches have equal run-time at a density of 0.0012 in the input matrices (giving a density of 0.08 in the resulting matrix), while for an input matrix density of 0.03 the custom implementation is already ten times faster, after which the discrepancy grows only slowly until a factor of 15 at the target density of 0.25. By the way, converting the input matrices to full and using OCTAVE's full matrix multiplication operator (which uses some BLAS implementation that runs on two threads, different from the rest of OCTAVE) performs equally compared to the custom implementation at an input density of 0.7, while it is on a par with the stock sparse matrix multiplication already at an input density of 0.09.

The comparison above has been done on sparse matrices with random structure. However, sparse matrices that originate from a model will have structure and so will their product, which will affect the relative merits somewhat.



# Bibliography

- L. Ambrosio and V. M. Tortorelli, Approximation of functional depending on jumps by elliptic functional via  $\Gamma$ -convergence, *Commun. Pure Appl. Math* **43**, 999 (1990).
- C. D. Anderson, The apparent existence of easily deflectable positives, *Science* **76**, 238 (1932).
- C. D. Anderson, The Positive Electron, *Phys. Rev.* **43**, 491 (1933).
- J. R. Anderson and A. V. Gold, Fermi Surface, Pseudopotential Coefficients, and Spin-Orbit Coupling in Lead, *Phys. Rev.* **139**, A1459 (1965).
- N. W. Ashcroft, The Fermi surface of aluminium, *Philos. Mag.* **8**, 2055 (1963).
- N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976).
- V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, no. 4 in Course of theoretical physics (Butterworth-Heinemann, 1982).
- R. Beringer and C. G. Montgomery, The Angular Distribution of Positron Annihilation Radiation, *Phys. Rev.* **61**, 222 (1942).
- S. Berko, M. Haghgooie, and J. J. Mader, Momentum density measurements with a new multi-counter two-dimensional angular correlation of annihilation radiation apparatus, *Phys. Lett. A* **63**, 335 (1977).
- S. Berko and J. S. Plaskett, Correlation of Annihilation Radiation in Oriented Single Metal Crystals, *Phys. Rev.* **112**, 1877 (1958).
- M. Biasini, *Positron annihilation studies of heavy-fermion metals and related compounds*, Ph.D. thesis, University of Bristol (1995).
- M. Biasini, Study of the Fermi surface of molybdenum and chromium via positron annihilation experiments, *Physica B* **275**, 285 (2000).
- M. Biasini, G. Ferro, G. Kontrym-Sznajd, and A. Czopnik, Fermi surface nesting and magnetic structure of  $\text{ErGa}_3$ , *Phys. Rev. B* **66**, 075126 (2002).
- P. E. Bisson, P. Descouts, A. Dupanloup, A. A. Manuel, E. Perreard, M. Peter, and R. Sachot, A high angular resolution 2-D correlation apparatus for positron annihilation studies, *Helv. Phys. Acta* **55**, 100 (1982).

## Bibliography

- F. E. Boas and D. Fleischmann, Evaluation of Two Iterative Techniques for Reducing Metal Artifacts in Computed Tomography, *Radiology* **259**, 894 (2011).
- M. Born and R. Oppenheimer, Zur Quantentheorie der Molekeln, *Ann. Phys. (Leipzig)* **84**, 457 (1927).
- T. Chiba, T. Toyama, Y. Nagai, and M. Hasegawa, Resolution deconvolution method applied to 2D-ACAR measurements, *phys. stat. sol. C* **4**, 3993 (2007).
- J. Chung, S. Knepper, and J. G. Nagy, Large-Scale Inverse Problems in Imaging, in O. Scherzer (ed.), *Handbook of Mathematical Methods in Imaging*, pp. 43–86 (Springer, New York, 2010).
- A. M. Cormack, Representation of a function by its line integrals, with some radiological applications, *J. Appl. Phys.* **34**, 2722 (1963).
- A. M. Cormack, Representation of a function by its line integrals, with some radiological applications. II, *J. Appl. Phys.* **35**, 2908 (1964).
- S. DeBenedetti, C. E. Cowan, W. R. Konneker, and H. Primakoff, On the Angular Distribution of Two-Photon Annihilation Radiation, *Phys. Rev.* **77**, 205 (1950).
- M. C. Delfour and J.-P. Zolésio, *Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization* (Society for Industrial and Applied Mathematics, Philadelphia, 2011), second edn.
- A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Statist. Soc. B* **39**, 1 (1977).
- P. A. M. Dirac, The Quantum Theory of the Electron, *Proc. R. Soc. Lond. A* **117**, 610 (1928).
- P. A. M. Dirac, A Theory of Electrons and Protons, *Proc. R. Soc. Lond. A* **126**, 360 (1930).
- P. A. M. Dirac, Quantised Singularities in the Electromagnetic Field, *Proc. R. Soc. Lond. A* **133**, 60 (1931).
- J. T. Dobbins, III and D. J. Godfrey, Digital x-ray tomosynthesis: current state of the art and clinical potential, *Phys. Med. Biol.* **48**, R65 (2003).
- S. B. Dugdale, M. A. Alam, H. M. Fretwell, M. Biasini, and D. Wilson, Application of maximum entropy to extract Fermi surface topology from positron annihilation measurement, *J. Phys.: Condens. Matter* **6**, L435 (1994).
- T. Feeman, *The Mathematics of Medical Imaging. A Beginner's Guide* (Springer, New York, 2010).
- H. M. Fretwell, S. B. Dugdale, M. A. Alam, M. Biasini, L. Hoffmann, and A. A. Manuel, Reconstruction of 3D Electron-Positron Momentum Densities from 2D Projections: Role of Maximum-Entropy Deconvolution Prior to Reconstruction, *Europhys. Lett.* **32**, 771 (1995).

- C. F. Gauss, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium* (F. Perthes et I. H. Besser, Hamburg, 1809).
- U. Gerhardt, S. Marquardt, N. Schroeder, and S. Weiss, Bayesian deconvolution and analysis of photoelectron or any other spectra: Fermi-liquid versus marginal Fermi-liquid behavior of the 3d electrons in Ni, *Phys. Rev. B* **58**, 6877 (1998).
- G. Gilat and L. J. Raubenheimer, Accurate Numerical Method for Calculating Frequency-Distribution Functions in Solids, *Phys. Rev.* **144**, 390 (1966), see also the erratum.
- X. Gonze et al., ABINIT: First-principles approach to material and nanosystem properties, *Comput. Phys. Comm.* **180**, 2582 (2009).
- R. Gordon, R. Bender, and G. T. Herman, Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography, *J. theor. Biol.* **29**, 471 (1970).
- H. Haghghi, J. H. Kaiser, S. Rayner, R. N. West, J. Z. Liu, R. Shelton, R. H. Howell, F. Solal, and M. J. Fluss, Direct observation of Fermi surface in  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ , *Phys. Rev. Lett.* **67**, 382 (1991).
- W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, 97 (1970).
- G. T. Herman and A. Lent, Iterative reconstruction algorithms, *Comput. Math. Biol.* **6**, 273 (1976).
- J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1999).
- M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models, *Int. J. Comput. Vision* **1**, 321 (1988).
- E. Klann and R. Ramlau, Regularization Properties of Mumford–Shah-Type Functionals with Perimeter and Norm Constraints for Linear Ill-Posed Problems, *SIAM J. Imaging Sci.* **6**, 413 (2013).
- O. Klemperer, On the annihilation radiation of the positron, *Proc. Camb. Phil. Soc.* **30**, 347 (1934).
- W. Kohn and L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev.* **140**, A1133 (1965).
- G. Kontrym-Sznajd, Fermiology via the electron momentum distribution (Review Article), *Low Temp. Phys.* **35**, 599 (2009).
- G. Kontrym-Sznajd, M. Samsel-Czekala, M. Biasini, and Y. Kubo, Band structure of  $\text{LaB}_6$  by an algorithm for filtering reconstructed electron-positron momentum densities, *Phys. Rev. B* **70**, 125103 (2004).
- A. C. Kruseman, *Two-Dimensional ACAR and Low-Background DBAR Studies on Materials with Defects*, Ph.D. thesis, Delft University of Technology (1999).

## Bibliography

- P. Langevin, Sur la théorie du mouvement brownien, *C. R. Acad. Sci. Paris* **146**, 530 (1908).
- J. Laverock, T. D. Haynes, M. A. Alam, and S. B. Dugdale, Experimental determination of the state-dependent enhancement of the electron-positron momentum density in solids, *Phys. Rev. B* **82**, 125127 (2010).
- M. Leitner, H. Ceeh, and J.-A. Weber, Eliminating spatial distortions in Anger-type gamma cameras, *New J. Phys.* **14**, 123014 (2012).
- R. M. Lewitt, Multidimensional digital image representations using generalized Kaiser-Bessel window functions, *J. Opt. Soc. Am. A* **7**, 1834 (1990).
- D. G. Lock, V. H. C. Crisp, and R. N. West, Positron annihilation and Fermi surface studies: a new approach, *J. Phys. F* **3**, 561 (1973).
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
- P. E. Mijnarends, Determination of Anisotropic Momentum Distributions in Positron Annihilation, *Phys. Rev.* **160**, 512 (1967).
- V. A. Morozov, On the solution of functional equations by the method of regularization, *Sov. Math. Dokl.* **7**, 414 (1966).
- D. Mumford and J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, *Commun. Pure Appl. Math* **42**, 577 (1989).
- K. M. O'Brien, M. Z. Brand, S. Rayner, and R. N. West, The enhancement of Fermi-surface images in positron ACAR spectra, *J. Phys.: Condens. Matter* **7**, 925 (1995).
- J. R. Oppenheimer, On the Theory of Electrons and Protons, *Phys. Rev.* **35**, 562 (1930a).
- J. R. Oppenheimer, Two Notes On the Probability of Radiative Transitions, *Phys. Rev.* **35**, 939 (1930b).
- S. Osher and J. A. Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations, *J. Comput. Phys.* **79**, 12 (1988).
- L. M. Pecora, 3D tomographic reconstruction from 2D data using spherical harmonics, *IEEE Trans. Nucl. Sci.* **NS-34**, 642 (1987).
- R. Penrose, The role of aesthetics in pure and applied mathematical research, *Bull. Inst. Math. Appl.* **10**, 266 (1974).
- J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- A. B. Pippard, An experimental determination of the Fermi surface in copper, *Phil. Trans. A* **250**, 325 (1957).

- M. Pylak, G. Kontrym-Sznajd, and L. Dobrzyński, Electron-positron momentum density distribution of Gd from 2D ACAR data via Maximum Entropy and Cormack's methods, *Appl. Phys. A* **104**, 587 (2011).
- J. Radon, Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten, *Ber. Verh. Sächs. Ges. Wiss. Leipzig, Math. Phys. Kl.* **69**, 262 (1917).
- R. Ramlau and W. Ring, A Mumford-Shah level-set approach for the inversion and segmentation of X-ray tomography data, *J. Comput. Phys.* **221**, 539 (2007).
- D. J. Roaf, The Fermi surfaces of copper, silver and gold. II. Calculation of the Fermi surfaces, *Phil. Trans. A* **255**, 135 (1962).
- L. I. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D* **60**, 259 (1992).
- F. Santosa, A level-set approach for inverse problems involving obstacles, *ESIAM Contr. Optim. Calc. Var.* **1**, 17 (1996).
- O. Scherzer (ed.), *Handbook of Mathematical Methods in Imaging*, Springer Reference (Springer, New York, 2010).
- D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn, Metallic Phase with Long-Range Orientational Order and No Translational Symmetry, *Phys. Rev. Lett.* **53**, 1951 (1984).
- J. C. Slater, The Theory of Complex Spectra, *Phys. Rev.* **34**, 1293 (1929).
- J. C. Slater and G. F. Koster, Simplified LCAO Method for the Periodic Potential Problem, *Phys. Rev.* **94**, 1498 (1954).
- I. Tamm, Über die Wechselwirkung der freien Elektronen mit der Strahlung nach der Diracschen Theorie des Elektrons und nach der Quantenelektrodynamik, *Z. Phys.* **62**, 545 (1930).
- A. Tikhonov, Solution of incorrectly formulated problems and the regularization method, *Sov. Math. Dokl.* **5**, 1035 (1963).
- M. Torrent, F. Jollet, F. Bottin, G. Zérah, and X. Gonze, Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under pressure, *Comput. Mater. Sci.* **42**, 337 (2008).
- M. N. Wernick and J. N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT* (Elsevier Academic Press, San Diego, London, 2004).
- R. N. West, J. Mayers, and P. A. Walters, A high-efficiency two-dimensional angular correlation spectrometer for positron studies, *J. Phys. E* **14**, 478 (1981).



# Curriculum vitae

## Personal data

Name: Michael Leitner  
Date of birth: October 13<sup>th</sup>, 1982  
Place of birth: Mistelbach, Austria  
Nationality: Austria  
Family status: Married, one child

## Education

09/1989 – 06/1993: Volksschule Neudorf  
09/1993 – 06/2001: Bundesgymnasium u. Bundesrealgymnasium Laa a.d. Thaya  
09/2001 – 04/2002: Military service  
09/2002 – 04/2013: undergraduate studies (mathematics) at the Faculty of Mathematics, University of Vienna  
03/2003 – 03/2007: undergraduate studies (physics) at the Faculty of Physics, University of Vienna  
03/2007: Mag. rer. nat. (MSc) in physics, University of Vienna  
Diploma thesis: “Diffusion studies with X-ray photon correlation spectroscopy”, Advisor: ao. Univ.-Prof. Dr. Bogdan Sepiol  
03/2007 – 02/2010: doctoral studies (physics) at the Faculty of Physics, University of Vienna  
02/2010: Dr. rer. nat. (PhD) in physics, University of Vienna  
Doctoral thesis: “Studying Atomic Dynamics with Coherent X-rays”, Advisor: ao. Univ.-Prof. Dr. Bogdan Sepiol

## Positions

03/2007 – 12/2008: Research Assistant supported by the Initiativkolleg I022-N, research group “Dynamics of Condensed Systems”, Faculty of Physics, University of Vienna  
01/2009 – 04/2010: Research Assistant at the research group “Dynamics of Condensed Systems”, Faculty of Physics, University of Vienna  
05/2010 – 09/2011: Postdoctoral Research Assistant at the research group “Dynamics of Condensed Systems”, Faculty of Physics, University of Vienna, interrupted by:  
10/2010 – 03/2011: Parental leave  
09/2011 – present: Postdoctoral Research Assistant at the Forschungsreaktor Heinz Maier-Leibnitz, Technische Universität München