High precision calibration in X-ray spectroscopy

Master's Thesis in Physics

Johann Jakob R. Stierhof

FAU FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

Supervisor: Prof. Dr. Jörn Wilms   27. September 2019

Contents

Chapter I
Introduction

Light, in ordinary sense, is a phenomenon that provides us with information about our surroundings. Almost all the light we perceive in our eyes is reflected from surfaces. The reflected light of objects gives information of distance, size and shape. Light, in a more general sense, as well, provides information especially about the generation process of the radiation.

This information is what is of great value to understand objects that otherwise provide no or only little insights. In astronomy the information one can receive from distant objects is the radiation astronomical objects send out due to intrinsic processes. To study and understand those objects it is important to have devices allowing a quantitative measure of the incoming radiation as well as a detailed understanding of the fundamental processes generating light.

Scientific devices were developed to gather information over all wavelengths, from long radio waves to high energetic gamma rays. These devices not only extend the accessible light range but also allow us to test theories of the world against observations. Radio receivers provide insights in the motion and distribution of gas clouds, optical devices show us the distributions and development of stars in the universe and gamma-ray detectors give access to the physics of the most violent environments in the universe.

Many of the processes which generate X-ray radiation involve accelerated charges or the transition of electrons in atomic systems. For example, the radiation which is emitted by the transition of the hydrogen atom from the first excited state to the ground state produces light waves with the energy of $E \approx 13.6\,\text{eV}$ (Demtröder 2018) located at the lower end of the X-ray spectrum.

The great insights which are provided by the observations in the X-ray regime does also face the astronomer with a problem. The high energy of X-ray photons do not allow the to pass the atmosphere of the earth without interaction. This is the reason why X-ray observatories need to be put up in space with all its complications and costs.

It is this effort, that has to be made, to obtain scientific data from X-ray sources, why one wants to get the most out of the instruments. This requires that the instruments are understood in great detail such that each value the device reports has a meaning. For the X-ray astronomy the information that gives the most information is the distribution of the photons across the energy range. From these spectra it is possible to deduce what the generating process for the photons is. The comparison with theoretical models gives detailed insights in the source and its environment. For this reason is the exact energy calibration a compelling requirement. Otherwise, derivations from miscalibrated data will end in a wrong picture of our universe.

In this work a method for calibration of the energy scale of such an X-ray detecting device will be performed. The goal is to test if the method can be used for the intended purpose and to discover any problems that might come along.

Additionally, data from a measurement campaign for measuring the oxygen edge of neutral oxygen gets calibrated. The calibration relies on a nouvelle reference method for laboratory astrophysics and should in principle increase the accuracy for line measurements drastically.

Chapter II
The Physics of Atoms

1 Electromagnetic interaction

In order to understand the atomic model and the observations that led to the development of the quantum mechanical model of the atom the knowledge of electromagnetic interactions is required. Every material that is around us in our daily life can be described and understood only with the electromagnetic interaction. Just a small number of effects require the understanding of the other fundamental forces[1] and as such a complete description of the atom down to the fundamental particles.

For the purpose of this work and for most of the effects that can be observed in e.g., astronomical sources it is sufficient to describe the atomic model by the electromagnetic interaction only. This section will give a brief overview of the development and the current understanding of the interaction between charged particles. A complete derivation may be found in e.g., Griffiths (2013); Jackson (1999).

1.1 A brief history of Electromagnetism

Electric phenomena where known to all mankind at least in the manifestation of lightning and thunder. In the ancient cultures all over the world people have realized that some materials have forces interacting between them. The obvious discoveries include natural magnets and magnetized metals. Famously, amber, found by the ancient Greeks was discovered to attract lightweight objects when rubbed (Whittaker 1910) and is the origin of the word electricity. Amber, Greek ἤλεκτρον (ēlektron), as one of the prominent curiosities that showed electrical properties led to the invention of the Latin word *electrica* by Gilbert (1600).

While most of electric phenomena, which are described in ancient text, were merely used for entertainment, there are examples which indicate a more technologically focused use case, e.g., the Baghdad battery (Keyser 1993).

The systematic study of electrical phenomena started, at least from a modern scientific point of view, in the late 16th century (Park 1898). Initially new materials were found possessing similar properties as amber. Later also properties of the electrical interaction itself got examined and described, prominently by Robert Boyle (1675).

During mid-17th century the first machines designed to show electrostatic effects appeared (e.g., the electrostatic generator by Otto von Guericke, although not recognized as such, Heathcote 1950). In the beginning of the 18th century a more systematic study of electrical effects lead to the discovery of conducting and non-conducting as well as to *sorts* of electricity by Charles François de Cisternay (Keithley 1999) laying the foundation for the concept of positive and negative charges.

Late 18th century there was good understanding of electric flows (Jean-Antoine Nollet discovered the finite speed of currents, Guarnieri 2016), the discovery of capacitors (Guarnieri 2016) and the connection between lightning and electricity by Franklins famous kite experiment. However, a concise theory of the observations was missing.

An important step towards a theoretical understanding of electric interactions was made by Charles-Augustin de Coulomb (Magie 1935, p. 408 ff.) discovering the inverse square law of charged bodies. Also, around the turn of the century Luigi Alyisio Galvani discovered the chemoelectrical effect (Magie 1935, p. 420 ff.) which led to the construction of the first battery by Alessandro Volta (Magie 1935, p. 427 ff.).

---

[1] Besides gravity, which is unavoidable on earth but also negligible when in competition with the other forces.

During the 19th century a significant fraction of the mathematical framework that is still in use today (in some form or the other) was invented. Prominent names generalizing the description of electric (and also magnetic) interactions include Farraday (Magie 1935, p. 472 ff.), Ampère (Magie 1935, p. 446 ff.) and George Green (2008). On the experimentalists side similar well known persons contributed important discoveries, most prominently Georg Simon Ohm (Magie 1935, p. 465 ff.).

The (classical) theory of electromagnetism as in use today was developed by Maxwell (Magie 1935, p. 528 ff.), Hertz (Magie 1935, p. 549 ff.), Thomson (Magie 1935, p. 583 ff.) and others. From there work the complete description of electrodynamic was developed and can be concisely expressed in Maxwell's equations in differential form (Griffiths 2013)

$$\Delta \boldsymbol{E} = \frac{\rho}{\epsilon_0} \qquad \Delta \boldsymbol{B} = 0$$
$$\Delta \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \qquad \Delta \times \boldsymbol{B} = \mu_0 \boldsymbol{I} + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} \tag{2.1}$$

where the problem describing parameters are the charge density ρ and the electric current $\boldsymbol{I}$. ϵ_0, μ_0 and c are the permittivity and permeability of free space and the speed of light in vacuum, respectively.

Einstein could show that these equations are not preserved by under Galilei transformations (Einstein 1905) which is key to Newtons equation of motion. He therefore worked out a different equation of motion obeying the same symmetry as the Maxwell equations, namely the theory of special relativity. The predictions from this theory are of great importance to understand the observations from atom transition experiments which will be described in the following sections.

1.2 Charge-Charge interactions

This section will summarize the effects that can occur between charged particles. From the first equation in Eq. (2.1) one can see that the electrical field of a point charge is radially symmetric field with field lines which (depending on the charge sign) point radially to or away from the charge (Figure 2.1).

Because Maxwell's equations are linear the superposition principle holds. This fact makes it trivial to construct the electrical fields of static point charge distributions as the solution is simply the superposition of all individual charge fields.

Any charge that is exposed to a electric field feels the force of that field and moves along the field lines either in the direction of the field if positively charged or in the opposite direction if negatively charged. This also means that a positive and a negative charge move closer to each other and eventually combine to a particle zero charge.

This is of importance as it was found that all matter consists of fundamental particles which may carry a charge that is either positive or negative. Although the same rules for electromagnetic interaction apply to those particles they have not combined to neutral particles. And still, the force that binds together the substances in our world is the attractive force between positive and negative charges.

To answer this question was one of the challenges that where faced when the theory of matter was developed during the 20th century.

Besides this rather direct electrostatic interaction between charged particles there can also be an interaction involving the magnetic field produced
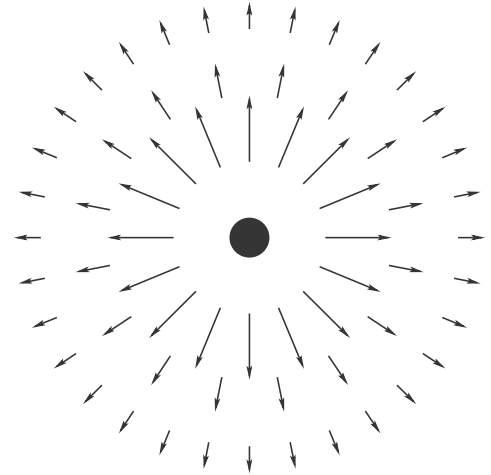

Figure 2.1: Schematic visualization of the field of a positive point charge.

by a moving charge[2]. This interaction is of importance as a small rotating charged sphere does produce a magnetic field. This effect is of importance in the atomic model as will be seen later.

1.3   Charge-Field interactions

In general all interactions between charges are really the impact the fields of all other charges have on one charged particle and as such also charge-field interactions. But there are additional solutions to Maxwell's equations that have a charge distribution of zero and still have non vanishing electric and magnetic fields.

By manipulating Maxwell's equations one gets the wave equation for electromagnetic waves, i.e., light

$$
\begin{aligned}
\left(c^2\Delta^2 - \frac{\partial^2}{\partial t^2}\right)\boldsymbol{E} &= 0 \\
\left(c^2\Delta^2 - \frac{\partial^2}{\partial t^2}\right)\boldsymbol{B} &= 0.
\end{aligned}
\tag{2.2}
$$

The solutions to this equations describe the periodic time and space dependent variations of the magnetic field that, according to those equations, propagate forever once created.

The simplest solution for the wave equation is that of a plane wave

$$
\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E_0}\cos\left(\omega t - \boldsymbol{kr}\right),
\tag{2.3}
$$

here for the electric field component, which can be simply shown by inserting in the wave equation. $\boldsymbol{E_0}$ is the amplitude of the wave, ω its angular frequency and $\boldsymbol{k}$ the wave vector giving the propagation direction. The electric field is linked to the magnetic field via

$$
c^2\frac{\partial \boldsymbol{B}}{\partial t} = \frac{\partial \boldsymbol{E}}{\partial t}
\tag{2.4}
$$

according to Farradays law (Griffiths 2013).

Electromagnetic waves can be produced by periodically moving charges. The mechanism behind the wave production lies in the finite velocity of light. When a charged particle is accelerated the change in the electromagnetic field spreads with the speed of light and propagates, as predicted by Maxwell's equations towards infinity. The knowledge of how light is generated is not only necessary to understand the origin of natural light but also to build devices which emit artificially and well defined light waves that can be used to systematically probe the interaction between light and matter.

2   Particle waves and probabilities

During the end of 19th and mainly in the 20th century it was discovered that matter consists of small particles now called atoms. And it was discovered, that the rules of the world of those particles is fundamentally different from the daily experience we have in our day to day life (Demtröder 2018).

Rutherford (1911) discovered, that those atoms any matter is made off, itself have a structure. Plucker, Hittorf, Thomson and others could show, with gas discharge tube experiments, that such atoms which have zero net charge can be separated into positive and negative charges (Demtröder 2018).

---

[2] In the framework of special relativity there is only the electric field, however, in a moving reverence frame the electric field appears as an magnetic field as defined by the Lorentz force (Griffiths 2013)

In 1897 Thomson was able to measure the ratio of e/m (charge over mass) for the negative charges and could also show the this radio is 4 orders of magnitude larger compared with the positive charges (Demtröder 2018). The former was identified as the core of the atom, the latter as electrons.

As mentioned before, was it challenging for classical theories to find a mechanism that could explain why such structures could exist. The electromagnetic interaction between the positive nucleus and the negative electrons should attract each other until they fall together. According to Rutherfords atomic model do the electrons move in circles around the nucleus, just like the planets around the sun. Bohr, however, thought of the electrical field the charges produce and the energy they must radiate away due to the circular motions (Bohr 1913; **?**). Due to this, he concluded, the atoms cannot be stabilized and additional physics need to be considered.

The solution was found by solving a different problem. It was observed that any body[3] with a temperature T emits the same spectrum of light. The explanations deduced from the classical theories could only predict parts of the spectrum and giving unphysical results (the ultraviolet catastrophe, Demtröder 2018).

The solution for this black body radiation was left to Planck (1914) who introduced the idea that energy can not be exchanged arbitrarily but only in integer values of the modes of the radiation field $\Delta E = nh\nu$ ($n \in \mathbb{N}$) where h is Planck's constant and relates the frequency ν of a radiation mode to its energy value.

This idea of discrete steps was successively applied also to the atomic model by Bohr. But instead of the energy changing in integer quantities the angular momentum of the electrons can only change in discrete steps given by $L = n\frac{h}{2\pi}$ ($n \in \mathbb{N}$). By this restriction there exists a lowest orbit that can be realized which cannot lose energy anymore due to radiation.

This discretization of the angular momentum was justified by the interpretation of matter also as waves rather than point like particles. This idea was first published by De Broglie, Louis (1925) who inserted the momentum of a particle into Planck's energy relation expressed with the photon momentum $p = E/c = h/\lambda$, giving particles a wavelength.

It was found experimentally that particles indeed show wave characteristics as predicted by de'Broglies wavelength for particles (Thomson & Reid 1927). Now, demanding that the path of an electron around the nucleus must be a standing wave one obtains the relation for the angular momentum as used in Bohrs atomic model.

Figure 2.2: The planetary model of atoms describes how the electrons move in circular paths around the nucleus

2.1 Equation of motion of microscopic particles

The discovery that also particles show wave properties, i.e., interference patterns when moving through a slit hat to be covered by the equation of motion that describes those particles. To retain the properties of point like particles the wave character of microscopic particles was interpreted in a probabilistic way. In this interpretation the time dependent position of a particle is replaced by a time dependent probability distribution of the particles position $P(\boldsymbol{r}, t) = |\psi(\boldsymbol{r}, t)|$ where ψ is the particles wave function in analogy to the wave function for electromagnetic waves.

The defining equation for the particle wave functions was first expressed by Erwin Schrödinger (1926). One can follow Schrödingers approach by starting from a plane wave of a free particle, such that $\boldsymbol{p} = \frac{h}{2\pi}\boldsymbol{k} = \hbar\boldsymbol{k}$ and
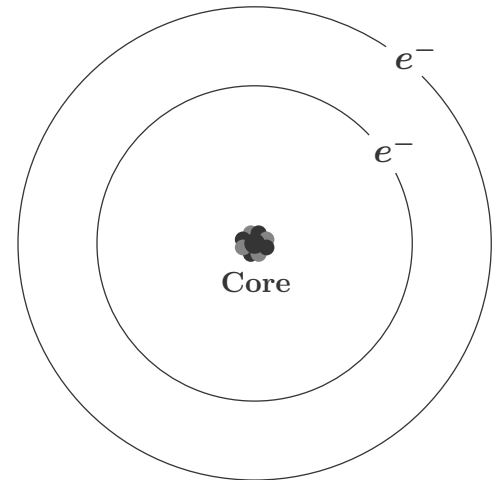
---

[3] Actually not just bodies as in solid body but all extended matter bodies as gas clouds and fluids

$E = h\nu = \hbar\omega = E_{\text{kin.}}$ ($E_{\text{pot.}} = 0$, hence, free particle). Inserting this into the wave function one obtains

$$\psi(\boldsymbol{r}, t) = A e^{\frac{i}{\hbar}(\boldsymbol{p}\boldsymbol{r} - E_{\text{kin.}}t)} \tag{2.5}$$

where $E_{\text{kin.}} = \frac{\boldsymbol{p}^2}{2m}$ given by classical theory.

Taking the second derivative of this function with respect to space and time respectively one obtains the wave equation

$$\frac{\partial^2 \psi}{\partial \boldsymbol{r}^2} = \frac{1}{u^2}\frac{\partial^2 \psi}{\partial t^2} \tag{2.6}$$

with the phase velocity $u = \omega/k$ and $\frac{\partial^2}{\partial \boldsymbol{r}^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ the Laplace operator[4]

Throwing the quantities together and adding the potential energy for a particle inside a potential independent of space and time, one obtains

$$-\frac{\hbar^2}{2m}\frac{\partial^2 \psi}{\partial t^2} + E_{\text{pot.}}\psi = E\psi. \tag{2.7}$$

The Schrödinger equation is then obtained by replacing the kinetic energy with the equation

$$\frac{\partial \psi}{\partial t} = -\frac{i}{\hbar}E_{\text{kin.}}\psi. \tag{2.8}$$

Schrödinger claimed that this equation still holds when the constant potential energy is replaced by a general potential and therefore the equation of motion is

$$i\hbar\frac{\partial \psi(\boldsymbol{r}, t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \psi(\boldsymbol{r}, t)}{\partial \boldsymbol{r}^2} + E_{\text{pot.}}\psi(\boldsymbol{r}, t). \tag{2.9}$$

A more detailed description of the development of the theory of quantum mechanics and all the experimental milestones that lead to the current picture of the microscopic world is outlined, for example, in Demtröder (2018), Bransden et al. (2003) or Griffiths & Schroeter (2018).

2.2 The physical picture of probability waves

In contrast to the classical theory of motion where a intuitive picture can be drawn of the problem, the quantum theory does not really give this intuitive access. At least not in the quantum mechanical framework as described by Schrödinger.

A more intuitive interpretation for quantum mechanical mechanical interactions was given by Richard Feynman (1948). He was thinking about the physical action[5] and that, although it plays a key role in the classical theory, it is not present in quantum mechanics.

Feynman found a different representation that is mathematical identical to Schrödingers theory but did include the action. In the classical theory solutions for a problem are given by the processes where the action is minimal. In Feynmans interpretation the probability for a process to happen is dependent on the value of the action and is largest for the classical solution.

The physical interpretation of this description is that instead of the probability waves we still have point like particles but the path a particle is may take depends on the possible paths that there are for this particle and interactions it cant take part in. This interpretation is referred to as the path integral formulation of quantum mechanics and besides very simple problems not solvable. It did, however, lay the foundation for the modern theory of fundamental particles as of today, namely the quantum field theory.

---

[4] Usually the Laplace operator is denoted with Δ, however, with this notation it is more clear that it is a derivative of the spatial part

[5] The naive interpretation of the the action is the integrated energy conversion from potential to kinetic energy and vice versa in a process.

3 The simplest atomic system: The Hydrogen Atom

In 1885 Balmer found that the emission spectrum of Hydrogen consists of several sharp lines. He also found a very simple relation between the wavelengths of the emitted lines

$$\frac{1}{\lambda} = Ry\left(\frac{1}{n_1} - \frac{1}{n_2}\right) \tag{2.10}$$

where $n_1 = 2$, $n_2 \in [2,3,4,\ldots]$ and Ry the experimentally determined Rydberg constant. Later, Lyman (1906, 1914) and Paschen (1908) found also lines that could be described with $n_1 = 1$ and $n_1 = 3$.

With Bohrs atomic model it was possible to understand this relation because the distance between core and electron is restricted to discrete values by the angular momentum constraint. The allowed distance values can be calculated by setting the Coulomb force between core and electron equal to the centripetal force and further restrict r according to $2\pi r = n\lambda_{\mathrm{dB}}$. Rearranging gives

$$r_n = \frac{n^2}{Z}a_0 \tag{2.11}$$

with the classical atomic radius or Bohr radius $a_0 = \frac{\epsilon_0 h^2}{\pi \mu Z e^2}$ of the electron (μ is the reduced mass of the electron-nucleus system).

The kinetic energy in the center of mass system is given by

$$E_{\mathrm{kin.}} = \frac{\mu}{2} = \frac{1}{2}\frac{Ze^2}{4\pi\epsilon_0 r} = -\frac{1}{2}E_{\mathrm{pot.}} \tag{2.12}$$

where the velocity is given by the equality of forces. From this the total energy is given by

$$E = E_{\mathrm{kin.}} + E_{\mathrm{pot.}} = -\frac{1}{2}\frac{Ze^2}{4\pi\epsilon_0 r}. \tag{2.13}$$

Inserting the expression for the radius one obtains the same expression as found empirically when subtracting two energy levels

$$\frac{1}{\lambda} = \frac{E_n}{hc} = -\frac{Ry}{hc}\frac{Z^2}{n^2}. \tag{2.14}$$

The Rydberg constant is therefore related to the other quantities via $Ry = \frac{\mu e^4}{8\epsilon_0^2 h^3 c}$.

This relation shows that the electron that is moving around the proton in the center of the atom can only exist in discrete states. Bohr proposed in his model that light is emitted when the electron changes from an outer radius to a inner one releasing a light wave carrying the energy difference away. In the classical picture the light is emitted by the changing dipole moment between the electron and the proton.

The rest of this chapter will summarize the predictions made by Schrödinger theory for atoms and the corrections that have to be made to contain the experimental findings.

3.1 Wave functions in the Coulomb potential

The remaining part of this chapter mainly summarizes parts of the book from Demtröder (2018) describing the theory of a particle wave in the Coulomb potential.
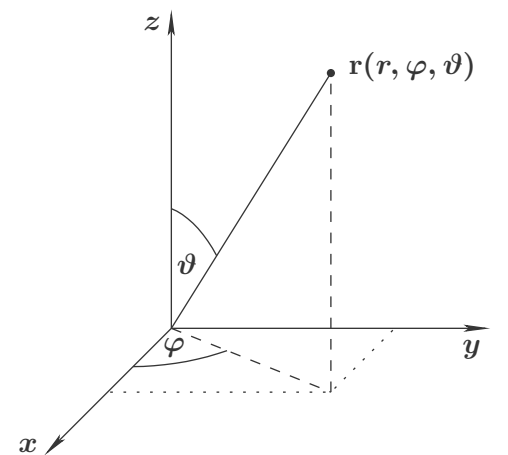

Figure 2.3: Space vector representation in spherical coordinates.

Suitable for a spherical symmetric problem is the use of spherical coordinates (Figure 2.3)

$$x = r \sin(\vartheta) \cos(\varphi)$$
$$y = r \sin(\vartheta) \sin(\varphi) \qquad (2.15)$$
$$z = r \cos(\vartheta).$$

This also changes the form of the Laplace operator to

$$\frac{\partial^2}{\partial r^2} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{r^2 \sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2}. \qquad (2.16)$$

With this form of the operator and the Schrödinger equation one tries as a ansatz for the solution a separation of all coordinates

$$\psi(r, \vartheta, \varphi) = R(r) \Theta(\vartheta) \Phi(\varphi). \qquad (2.17)$$

With this wave function the Schrödinger equation gives after multiplication with $r^2 \sin^2 \vartheta / \psi$

$$\frac{\sin^2 \vartheta}{R(r)} \frac{\mathrm{d}}{\mathrm{d}r} \left(r^2 \frac{\mathrm{d}R}{\mathrm{d}r} \right) + \frac{\sin \vartheta}{\Theta(\vartheta)} \frac{\mathrm{d}}{\mathrm{d}\vartheta} \left(\sin \vartheta \frac{\mathrm{d}\Theta}{\mathrm{d}\vartheta} \right)$$
$$+ \frac{2m}{\hbar^2} \left(E - E_{\mathrm{pot.}}(r) \right) r^2 \sin^2 \vartheta = -\frac{1}{\Phi(\varphi)} \frac{\mathrm{d}^2 \Phi}{\mathrm{d}\varphi^2}. \qquad (2.18)$$

Where the left side is solely dependent on r and ϑ and the right side solely on φ. From this one can conclude that, because the equation is valid for all combinations r, φ and ϑ, both sides in Eq. (2.18) have to equal a constant C.

For the right side the solution is then given by

$$\Phi = A e^{\pm \mathrm{i} \sqrt{C} \varphi} \qquad (2.19)$$

and because the solution has to be unique a further constraint is that $\Phi(\varphi) = \Phi(\pi + 2\pi n)$ such that follows $\sqrt{C} = m$ ($m \in \mathbb{Z}$).

Additionally, requiring that Φ is normalized when integrated over the full space, gives $A = 1/\sqrt{2\pi}$. In combination one obtains the normalized angular function

$$\Phi_m(\varphi) = \frac{1}{\sqrt{2\pi}} e^{\mathrm{i}m\varphi}. \qquad (2.20)$$

Now, dividing Eq. (2.18) by $\sin^2 \vartheta$ and rearranging such that one side has all terms dependent on ϑ and conversely all terms depending on r on the other side one obtains

$$\frac{1}{R} \frac{\mathrm{d}}{\mathrm{d}r} \left(r^2 \frac{\mathrm{d}R}{\mathrm{d}r} \right) + \frac{2m}{\hbar^2} r^2 \left(E - E_{\mathrm{pot.}}(r) \right) =$$
$$-\frac{1}{\Theta \sin \vartheta} \frac{\mathrm{d}}{\mathrm{d}\vartheta} \left(\sin \vartheta \frac{\mathrm{d}}{\mathrm{d}\vartheta} \right) + \frac{m^2}{\sin^2 \vartheta} \qquad (2.21)$$

and with the same argument as before both sides are independent and therefore have to equal a constant $\tilde{C}$.

The right side of Eq. (2.21) can be transformed into the Legendre differential equation when $m = 0$. This equation has solutions that can be written as polynomials, the Legendre's polynomials. For the case $m \neq 0$ the solutions are also known and given by the associated Legendre functions $P_l^m(cos\vartheta)$ which can be obtained from the Legendre's polynomials

$$P = a_0 + a_1 \cos \vartheta + a_2 \cos^2 \vartheta + \dots. \qquad (2.22)$$

The condition that P must be finite also for $\zeta = \pm 1$, which corresponds to $\vartheta = 0$ and $\vartheta = \pi$, requires only finitely many $a_n \neq 0$. Otherwise P would grow to infinity.

Inserting P in the Legendre equation gives a recurrent relation for the coefficients a_n depending on $\tilde{C}$. And because only finitely many $a_n \neq 0$ one can show that $\tilde{C} = l(l+1)$ $(l \in \mathbb{N})$. The associated Legendre functions are then related to the Legendre's polynomials

$$P_l^m(\cos \vartheta) = \text{const.} \cdot (1 - \cos^2)^{|m|/2} \frac{\mathrm{d}^{|m|}}{\mathrm{d}\cos^{|m|} \vartheta} P_l(\cos \vartheta). \qquad (2.23)$$

From the relation between the Legendre's polynomials and the associated Legendre functions one can see that $|m| \leq l$ has to be fulfilled, else the $|m|$th-derivative would not be defined.

Combining the solutions for Θ and Φ one obtains the spherical surface harmonic functions

$$Y_l^m(\varphi, \vartheta) = P_l^m(\cos \vartheta)\Phi_m(\varphi). \qquad (2.24)$$

Those functions have to be normalized when integrated over one sphere, otherwise the interpretation as probability is not possible

$$\int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\vartheta \; |Y_l^m(\varphi, \vartheta)|^2 \sin \vartheta = 1. \qquad (2.25)$$

It is to note that the spherical harmonic functions are independent of r and are therefore solutions for any radially symmetric potential.

Expressing the classical angular momentum $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ with the quantum mechanical momentum operator $\boldsymbol{p} = -\mathrm{i}\hbar(\boldsymbol{r} \times \frac{\partial^2}{\partial \boldsymbol{r}^2})$ one obtains the Cartesian angular momentum operators expressed in spherical coordinates as

$$
\begin{aligned}
L_x &= \mathrm{i}\hbar \left(\sin \varphi \frac{\partial}{\partial \vartheta} + \cotan\vartheta \cos \varphi \frac{\partial}{\partial \varphi} \right) \\
L_y &= \mathrm{i}\hbar \left(-\cos \frac{\partial}{\partial \vartheta} + \cotan\vartheta \sin \varphi \frac{\partial}{\partial \varphi} \right) \qquad (2.26) \\
L_z &= -\mathrm{i}\hbar \frac{\partial}{\partial \varphi}.
\end{aligned}
$$

The square of the total angular momentum is given by $L^2 = L_x^2 + L_y^2 + L_z^2$ which, when applied to the wave function of the spherically symmetric potential, gives

$$L^2\psi = l(l+1)\hbar^2\psi \qquad (2.27)$$

following directly from the Schrödinger equation. One can also see that the z-component of the angular momentum is given by

$$L_z\psi = m\hbar\psi \qquad (2.28)$$

because $\psi \sim e^{\mathrm{i}m\varphi}$. This also shows that the quantities L^2 and L_z have the same eigenfunctions and, according to quantum mechanics, can be measured simultaneously (unlike L_x and L_y).

The radial part for the hydrogen wave function solves the differential equation introduced by the Coulomb potential. The solution is known and can be expressed in terms of Laguerre potentials

$$R(r) = \sqrt{\left(\frac{2}{na_0^*}\right)^3 \frac{(n-l-1)!}{2n(n+l)!}} e^{-\rho/2} \rho^l L_{n-l-1}^{2l+1}(\rho). \qquad (2.29)$$

Here, $\rho = \frac{2r}{na_0^*}$ with the reduced Bohr radius $a_0^* = \frac{4\pi\epsilon_0\hbar^2}{\mu^2}$ (μ the reduced mass of the electron proton system, p. 147 ff., Demtröder 2018). L_k^α are the generalized Laguerre polynomials

The radial solution also gives constraints on the values for l in the same way as the spherical solution restrict m. It is such that $l < n$ ($l \in \mathbb{N}_0, n \in \mathbb{N}$). The constraint also comes from the requirement that the wave function can be normalized.

The probability of finding the electron of the hydrogen atom is given by $|\psi|^2$. The results for some probability distributions are shown in Figure 2.4. The combinations for the quantum numbers n,l and m are given in Table 2.1. Usually one labels the numbers for l with the characters $s = 0$, $p = 1$, $d = 2$, $f = 3$ and all higher from h onwards. This labeling dates back before quantum mechanics could explain the distribution and the different energy levels that observed in the hydrogen atom where called *sharp* (s), *principal* (p), *diffuse* (d) and *fundamental* (f)

3.2 Physical properties of (nlm) states

The previous section summarized the possible states a electron around a proton can have. The wave function that alone does not have direct physical meaning, only the absolute square give the probability of finding the electron at same point in space.

From the radial dependent part (2.29) one can see that the quantum number n determines the mean distance of the electron from the nucleus. One can calculate the probability of finding the electron in a spherical shell from radius r to $r + \mathrm{d}r$ via

$$P(r)\,\mathrm{d}r = \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\vartheta |\psi(r,\varphi,\vartheta)|^2 r^2 \sin\vartheta\,\mathrm{d}r. \qquad (2.30)$$

Which, for the $l = 0$ states is maximum for na_0^* identical to the results from Bohrs atomic model.

However, there are key differences between Bohrs atomic model and the quantum mechanical description. First, the expectation value of the radius given by

$$\langle r \rangle = \int_0^{\infty} \mathrm{d}r \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\vartheta r |\psi(r,\varphi,\vartheta)|^2 r^2 \sin\vartheta \qquad (2.31)$$

is $\frac{3}{2}a_0^*$ for the $(n,l) = (1,0)$ state and therefore differs from Bohrs results. Further, the value of l in Bohrs model is always positive while the wave function gives as expectation value for the angular moment 0.

Besides n determining the mean distance and therefore also the mean potential energy of the electron, the quantum number l influences the shape of the wave function. As depicted in Figure 2.4 higher values of l give more and more complex shapes of the particle distribution (p. 152 ff., Demtröder 2018).

Finally, the quantum number m determines the orientation of the wave function usually defined to the z-axis[6].

4 Atomic energy levels

The last part of this chapter will summarize the effects that are responsible for the exact location of the energy levels in the hydrogen atom. The theory that was described by Schrödinger, although very precise, was missing an important part that is responsible for the exact level energies. This missing part was the inability of Schrödingers theory to explain why electrons have a discrete magnetic moment Stern (1921); Gerlach & Stern (1922).
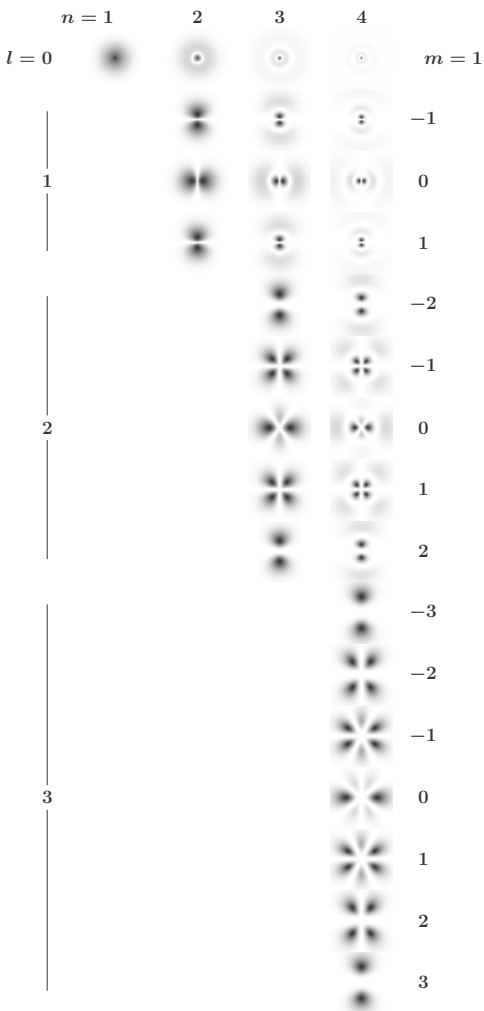
---

[6] This axis is called the quantization axis

Table 2.1: Possible combinations of the quantum numbers of the electron wave equation

n	l	m	total states
1	0 (s)	0	1
2	0 (s)	0	4
	1 (p)	-1,0,1	
3	0 (s)	0	9
	1 (p)	-1,0,1	
	2 (d)	-2,-1,0,1,2	
4	0 (s)	0	16
	1 (p)	-1,0,1	
	2 (d)	-2,-1,0,1,2	
	3 (f)	-3,-2,-1,0,1,2,3	

Figure 2.4: This figure shows the probability distribution for the first solutions of the Schrödinger equation for the hydrogen atom. The plots are cuts through the y-z-plane and scaled in size by $1/n$.
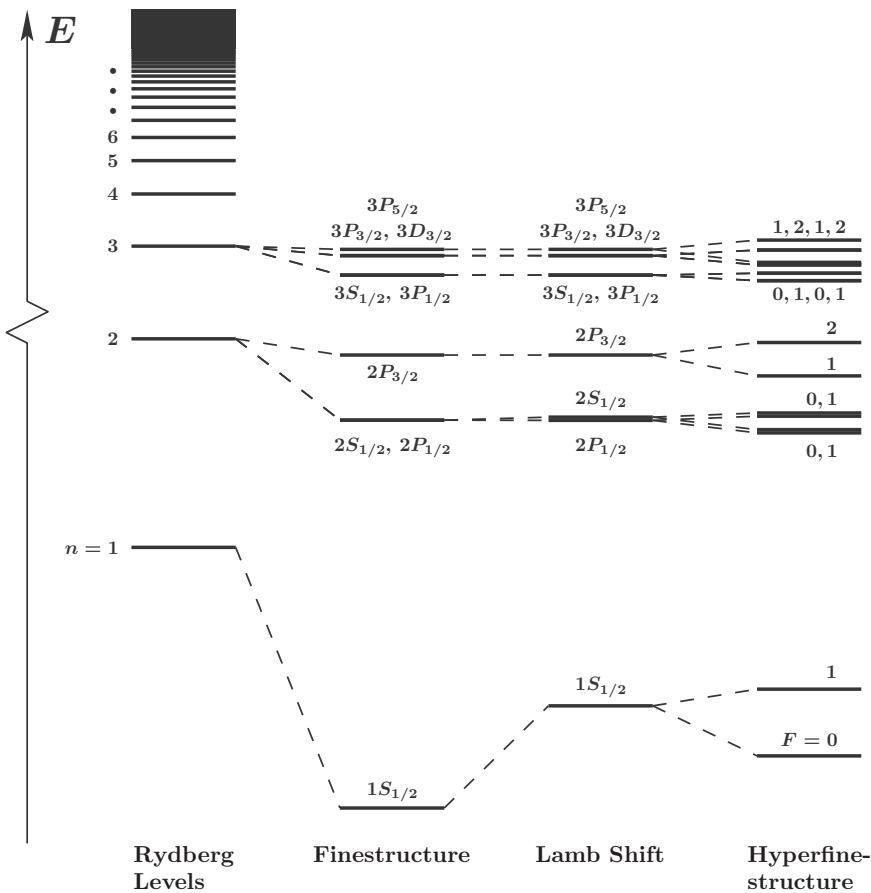
Figure 2.5: Level splitting in the hydrogen atom due to different effects on the electron. Rydberg levels: Due to mean distance between the nucleus and the electron in the Coulomb potential. Fine structure splitting (factor 10 increased w.r.t. Rydberg scale): Relativistic mass increase of the electron mass, line splitting because of spin-orbit coupling due to magnetic moment. Lamb shift (factor 20 increased w.r.t. Fine structure): Quantum-electrodynamical effect, due to the photon field interaction with the electron the energy is decreased. Hyperfine structure (factor 40 increased w.r.t. Lamb shift): Energy splitting due to spin of the nucleus and the magnetic moment of the electron (F total angular momentum).

The solution to explain this phenomenon was to introduce another quantum number for the electron which can be interpreted as a rotational moment of the electron[7]. This moment together with the charge of the electron generates a magnetic moment interacting with a magnetic field. The quantum number that was found describing the observed behavior is now referred to as the spin of the electron and can only take the values $s = 1/2$ or $s = -1/2$ (Demtröder 2018, p. 161 ff.,).

Schrödingers theory can be expanded to describe electrons correctly by adding a spin phase factor to the wave function. This spin function represents the spin of the described electron and allows to apply Pauli's principle (Pauli 1926, 1940).

For a complete description of the spin property it was necessary to develop a different equation of motion properly treating special relativity. Dirac & Fowler (1928) in 1928 discovered a new description which correctly predicts the effects due to the electron spin and special relativity.

---

[7] Although this quantity mathematically behaves like the angular moment of the electron "body" one quickly runs into problems when thinking of a rotating charged sphere as model for the electron.

4.1 The Rydberg levels

As mentioned before was the driving force behind the development of quantum mechanics due to the observed discrete emission from atomic systems. It was quickly realized that the atomic system must undergo a transition where it turns from a energetic higher state to a lower state (photon emission) or vice versa (absorption). The Rydberg formula was found to empirically describe the observed line position.

After the development of quantum mechanics and the discrete energy levels which can be realized in a atomic system, the Rydberg levels follow naturally from the Coulomb potential. Here, the potential energy depends on the distance between the charges. From the wave function of the electron one can see that the distance of the electron from the nucleus is determined by the quantum number n. Because additional effects only slightly affect the energy predicted by n (Figure 2.5) it is called the *principal quantum number*.

4.2 Fine structure splitting

The spin property of the electron generates a magnetic moment according to the theory of electromagnetism. In the same way is the angular motion of the electron generating a magnetic moment depending on the angular momentum of the electron. These two moments interact and cause the energy of the electron to be increased or decreased depending on the relative orientation of the moments.

Figure 2.5 shows that the Rydberg levels split for electron states with $l > 0$. In the case of $l = 0$ the electron orbit does not generate a magnetic moment and therefore no splitting for those states occurs.

Besides the splitting of the Rydberg levels due to the spin-orbit coupling the resulting levels are also shifted in energy due to relativistic effects. One the one hand is the momentum of the particle only correctly described by relativistic theory. This gives an effective mass increase of the electron the faster it is moving reducing the total potential energy. Additionally, the interaction of the nuclear and electron charge is affected by Heißenbergs uncertainty relation. This gives an additional correction to the potential levels (Darwin 1931).

The fine structure splitting was discovered already in 1887 by Michelson & Morley but only understood much later.

4.3 Lamb-shift

In the modern interpretation of quantum mechanics is every interaction described by the exchange of particles between the interacting fields. In the case of the electromagnetic interaction the exchange particle is the photon. In this picture do the fields generated by the nucleus and the electron constantly exchange photons. This exchange is possible due to the uncertainty relation and does not violate energy conservation.

The emission and absorption of photons leads to a statistical path of the electron shifting the potential energy, called Lamb-shift (p. 176 ff., Demtröder 2018). As the electrical field is stronger in the direct vicinity of the nucleus, electrons coming closer to the core are stronger affected by the Lamb-shift compared to electrons farther away.

From the wave function on can see that the only the states with $l = 0$ have non-vanishing probability distributions at the nucleus (see Figure 2.4). Therefore those states are much stronger affected by the Lamb-shift compared to states with $l \neq 0$.

The Lamb-shift is a purely quantum electrodynamical effect and as such not contained in Diracs theory. The effect of the field interaction was experimentally discovered in 1947 by Lamb & Reutherfurd.

4.4   Hyperfine structure

The final contribution that affects the energy levels of the hydrogen atom is due to the magnetic moment of the nucleus. Just in the same way as the electron has a magnetic moment due to its spin does also the proton have a magnetic moment as consequence of the proton spin.

The couplings of the moments depends on the total angular moment of the atomic system. The splitting due to this interaction is so small that it can only be observed in Doppler-free spectroscopy (p. 170, Demtröder 2018), otherwise the levels are smeared to one peak due to the motion of the atoms.

The aforementioned effects are responsible for the exact location of the energy levels in the hydrogen atom. However, the splitting as described is only true in field free environments. As soon as the atom is located inside a magnetic field the interaction between the magnetic moment of the electron and the external field cause different splittings of the levels. The splitting that is observed depends on the relative strength of the coupling between the internal moments and the external field (Paschen & Back 1912).

As most experimental setups involve magnetic fields the observed lines undergo those effects showing a different splitting than the level scheme depicted in Figure 2.5.

The energy levels of the hydrogen atom can be populated due to collisions of the electron with other charged particles or by the absorption of light. As nature tends to minimize energy are energy levels above the ground level not stable and will transition to lower energetic states by emitting photons with the energy of the level difference. Although energetically the transition between any level is possible nature also adheres to momentum conservation such that only some processes are allowed. Many other processes can be realized by emitting more photons but happens consequently less often.

5   Multi-electron systems

So far only the hydrogen atom was discussed. For systems with more than one electron, the Schrödinger equation does not have analytical solutions. Therefore, to describe heavier elements one has to rely on approximations or numerical calculations. Although approximations usually give a better insight in the behavior of the system itself, the obtained results are valid only in a arrow range defined by the approximation used. Numerical methods on the other hand, can be, theoretically, correct to arbitrary precision. But in reality it turns out that for very complex systems the observed results do not agree with calculated ones. This is (almost always) not due to missing physics but because also those calculations rely on certain assumptions to be practical.

For the calculation of transition energies, line strengths and other quantities that are of interest, different programs were developed to simulate either collisional ionized or photo-ionized plasma. This work made use reference values calculated with Flexible Atomic Code (FAC) (Gu 2008). FAC calculates ionization processes in a fully relativistic framework of the Dirac theory.

A more theoretical treatment of multi-electron systems is given in textbooks about the atomic theory and spectroscopy: Demtröder (2018); Griffiths & Schroeter (2018); Bransden, Joachain, & Plivier (2003).

Chapter III

Spectroscopy of X-ray sources

The phenomenon of vision, light and color was always a fascinating and widely discussed topic. Already the ancient Greeks discussed several ideas about the nature and properties of light. The discovery of the interaction of light and matter under modern scientific aspects happened much later and it was René Descartes around 1600 (**?**) who first described how the colors of the rainbow are produced by light passing through glass or water. A more comprehensive description was later given by Newton & Innys (1730) in his book 'Opticks'.

In later work Wollaston (1802), who made improvements based on the work of Newton, recognized the existence of lines in observed spectra. Although Wollaston did not identify the source of the lines, his work laid out the foundation for Joseph Fraunenhofer and his foundation of astronomic spectroscopy (Frauenhofer 1817). Frauenhofer was also the first who used a grating to systematically study transitions in gases.

The improvements of the grating experiments by Herschel (1788) and Talbot (1826) later allowed Balmer (1885) to discover the $1/n^2$ dependence of the lines in the hydrogen atom and ultimately lead to the formulation of the famous Rydberg formula.

Many of the discoveries that finally forced the development of a new theory of matter and light are due to better understanding of the sources of light and systematic measurements of the emission of gas clouds. One of the most important insights was, that the generation processes of light have a characteristic spectrum and it is therefore possible to deduce many properties of a light source just from the observed spectrum. It is this relation between objects and their emitted light that make modern astronomy possible. And this is also the reason why, although many can be learned from an image, a spectrum gives much more insight in astronomical sources.

In this chapter experimental setups shall be discussed that allow to observe spectra of X-ray sources as well as methods to measure emission lines of atoms.

1   Detecting X-rays

The X-ray regime belongs to short wavelength photons and therefore corresponding high energy: $100\,\text{eV} \lesssim E_{\text{X-ray}} \lesssim 100\,\text{keV}$[1]. Due to those high energies, X-ray photons can easily penetrate materials and is therefore challenging to build X-ray detecting devices.

In general, X-ray photons are detected by measuring side effects of the photons in materials. This can be the photoelectric effect (Hertz 1887) as used in digital cameras or the measure of produced ions in a gas cloud due to photon atom collisions.

An excerpt of different detector types and there working concepts is given in the following.

1.1   Semiconductor detectors

Einstein explained in 1905 the photoelectric effect by considering photons as particles rather than waves. Photons with enough energy are able to remove electrons from a metal plate by momentum transfer. Using this interaction process one can design devices which, when exposed to light, respond with the formation of free electrons or an electron cloud within the the solid. The total charge of such a cloud can be measured and is directly proportional to the energy of the incident photon.

---

[1] This corresponds to wavelengths of 0.01 to 10 nm. But historically X-ray photons are generally given in energy instead of wavelength or frequency

Charged coupled devices

In 1970 Boyle & Smith discovered that one can use p-doped Metal Oxid Semiconductor (MOS) to store and collect charges, today called Charged Coupled Devices (CCDs). They already mentioned that the charges can be produced by incident light. Later improvements on the design lead to high performance and sensitivity for light. CCDs quickly became a standard tool for quantitative photon counting.

The working principle is schematically drawn in Figure 3.1. When a photon interacts with the semiconductor material of the MOS electrons get lifted from the valence to the conducting band producing a electron-hole pair where the electron gets attracted by a positively charged electrode. The process described by Boyle & Smith (1970) allows to move the captured charges such that they can be read out.

The number of electron-hole pairs that get created by this process depends on the energy of the incident photon and the energy gap. This can be understood quantitatively such that when a photon is interacting with the material all its energy is transferred to the semiconductor electrons. The number of electron-hole pairs that get produced is then simply given by the energy each pair requires for creation

$$N_e \approx E_\gamma / E_{pair}. \tag{3.1}$$

For silicon this energy is $3.65\,\mathrm{eV}$ (Strüder 2000, at room temperature).

From Eq. (3.1) it follows that the energy resolution of a CCD is restricted by this creation mechanism in the semiconductor. The energy that is required to create electron-hole pairs in semiconductors is of the order of some eV. With this one can estimate the energy resolution of a CCD from counting statistics by $\Delta E \approx \sqrt{N_e} E_{gap}$. This estimate gives a resolution of the order of $100\,\mathrm{eV}$ for photon energies at the Iron Kα line ($E = 6.4\,\mathrm{keV}$).

CCDs are commonly used in X-ray astronomy and many of the modern space observatories feature optics with a CCD chip, e.g., X-ray Multi-Mirror mission (XMM) (Strüder et al. 2001) or Chandra (Garmire et al. 2003) The use of CCDs is not only due to the resolution they provide, but also because they have relatively modest operation requirements which is an advantage in the harsh environments of space.

Silicon drift detectors

Compared to CCDs where the generated charges are trapped in small potentials in each pixel and iteratively moved from pixel to pixel for readout, Silicon Drift Detectors (SDDs), as their name indicates, have a steady drift of electrons to the readout electronics. This drift is due to a field that is applied to the silicon block such that electrons created in the same way as in the CCD move through the silicon to the anode (Figure 3.2).

From this it should be obvious that the intrinsic mechanics between SDDs and CCDs is the same. This also means that the energy resolution which is achieved by SDDs compares equally to CCDs. A closer inspection of the processes involved reveals, according to Strüder (2000),

$$\Delta E_{FWHM} = 2.355 w \sqrt{R_{enc.}^2 + \frac{FE}{w}} \tag{3.2}$$

where F is the Fano factor (Strüder 2000) of the semiconductor, E the photon energy, w the required energy for pair creation and the prefactor 2.355 relates the standard deviation of a Gaussian to the FWHM. The term $R_{enc.}$ contains the noise of the readout process (Strüder 2000).

Although, SDDs and CCDs have similar properties there is one crucial difference which renders SDDs in some situation the better choice. Compared
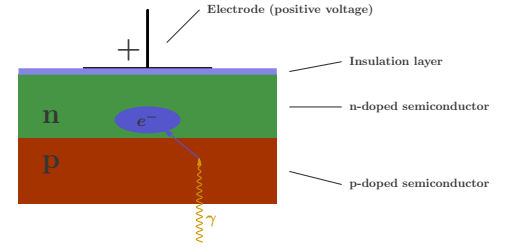

Figure 3.1: Schematic drawing of a CCD cell. Incoming photons interact with electrons in the semiconductor material and excite them to the conduction band, creating electron-hole pairs. The positive voltage applied by the electrode traps the electrons in the junction potential of the differently doped semiconductors. The number of trapped electrons depends on the energy of the impacting photon and the energy gap of the semiconductor.
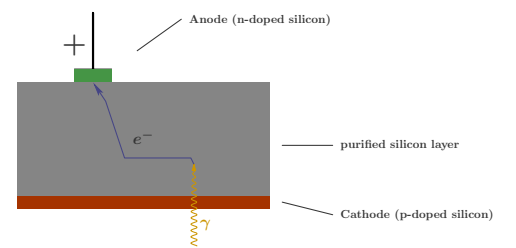

Figure 3.2: Schematic drawing of an SDD detector. Electron-hole pairs get produced on photon impact and the electrons are guided to the anode where they get detected. Compared to a CCD do these detectors not naturally provide spatial resolution.

to CCDs which are build up by layers of different material, SDDs are made of a block purified silicon (Strüder 2000) and already have a low noise signal. Therefore it is sufficient to cool the detector with Peltier elements (Strüder 2000) instead of cryogenic setups to reduce the intrinsic noise to a minimum.

1.2 Ionization detectors

In contrast to solid state detectors where the impacting photons create electron-hole pairs it is also possible to utilize the ionization power of X-rays to charge atoms or molecules in gas, liquids or solids and detect them. Two concepts for the detection of ions shall be outlined in the following.

A simple detector for X-rays can be provided with a ionization chamber (Knoll 2000). Here, the incident photons ionize gas inside the detector producing electron-ion pairs. A wire with applied voltage attracts the ions produced and a current can be measured across the wire as ions interact with it. The number of electron-ion pairs is proportional to the photon energy such that from the measured current the photon energy can be reconstructed.

Compared to simple charge detection in gas chambers one can increase the voltage that attract the ions. In this way the ions are accelerated and while moving through the gas their energy is high enough to produce additional electron-ion pairs (Knoll 2000). Therefore, the sensitivity of the detector is greatly increased. These proportional counters where also used in astronomy instrumentation, e.g., XRT-PCA **TODO:** *Insert citation.*

Instead of wire it is also possible to extract electrons which get released when ions impinge on a metal plate. The electrons are as well accelerated in an electrical field causing more electrons to be released on impact on the metal surface. The working principle of such a electron multiplier is depicted in Figure 3.3.

The resolution of ionization is in general an order of magnitude worse in comparison with solid state detectors. The reason is higher energy which is required to ionize the atoms in the gas compared with the energy required to generate electron-hole pairs in a semiconductor.
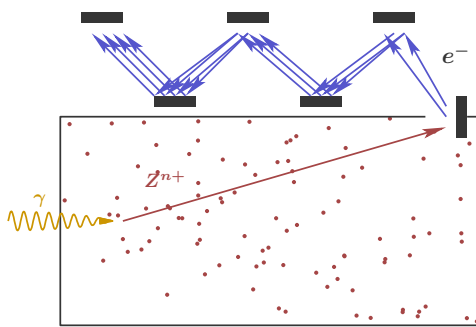

Figure 3.3: Exemplaric realization of a ion detection chamber. Impacting photons ionize the gas and a applied voltage at and anode attracts and accelerates these ions. On impact at the anode surface electrons will get liberated and successive cathodes repeat this process with the electrons, i.e., multiplying the number of electrons.

1.3 Calorimeters

Instead of measuring the energy via an excitation channel it is also possible (with allot of technological effort) to measure the energy of incident photons via the temperature change of a solid absorbing a photon.

The first time a thermometer was used to detect light was in 1800a; 1800b by Herschel for the red light diffracted by a prism. It was found that the photons transfer energy to the atoms and molecules of a solid through the excitation of one electron. This excited electron elastically collides with other electrons in the body distributing the energy over a volume and therefore heating up the solid (Figure 3.4, Hell 2017).

For the detection of this temperature change two conditions are crucial which render the operation of calorimeters as photon spectrometers very challenging. On the one hand is the temperature change very small such that thermometers are required allowing to detect this tiny change. The currently used devices either utilize thermistors (McGee et al. 1988), resistors with a strong dependence on the temperature, or Transition Edge Sensors (TES) (Ullom & Bennett 2015) which are superconductors operated just at the critical temperature where the state changes between super conducting and normal resistor.

Secondly, the temperature change that is introduced even by high energetic X-rays is still very small. Therefore, the signal is dominated by the intrinsic temperature of the detector if not cooled to a temperature just above absolute zero.
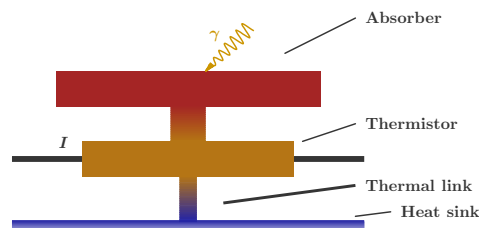

Figure 3.4: Schematic drawing of the measurement of changing temperature due to photon absorption. The signal is obtained from the thermistor where a different voltage is measured for constant current and changing temperature. For successive measurements it is necessary that the temperature dissipates as quickly as possible without affecting readout.

Despite the difficulties that are involved with cryogenic systems are calorimeters extremely valuable tools for X-ray spectroscopy. The energy resolution that can be reached with calorimeters is of eV order (Hell 2017).

The high resolution paired with high sensitivity make calorimeters the future instruments for X-ray astronomy. Despite multiple try's to run calorimeter based space instruments (Hitomi, Suzaku) it was so far not possible to keep the instruments in operation. Future missions are planed, for example XRISM (NASA) and Athena (ESA) Bering in different stages of development.

2   The EBIT Calorimeter Spectrometer

Part of this work was to test and apply a different calibration method for the EBIT Calorimeter Spectrometer (ECS)[2]. The ECS is a microcalorimeter located at LLNL and is used as a high resolution device for calibration measurements of crystals (Hell 2017, and reference therein) and transmission filters**TODO:** *Insert citation* as well as for testing the readout electronics and the cryogenic system for space missions and also the capabilities of microcalorimeters in general. This section shall provide a basic overview of the involved components and how the data is obtained.

2.1   Data acquisition

As already discussed previously are the most commonly used X-ray calorimeters build with a resistor with strong temperature dependence. The signal that is is read out from the calorimeter is a voltage level which raises due to photon impact. The heat which causes the resistance change must be quickly absorbed, otherwise the detector would suffer from unnecessary long dead times. The voltage function on photon impact is therefore a sharp peak with a damping out after peak level is reached where the exact damping form depends on the characteristics of the involved solids.

In Figure 3.5 idealized signal pulses are shown for the ECS. The ECS features two different pixel types where the main difference is the thickness of the absorber and therefore different sensitivity in different photon energy ranges. The two curves in Figure 3.5 correspond to the differ end pixels where the shorter peak necessarily corresponds to the thin pixels.

The critical parameters which determine the sensitivity of the calorimeter are the heat capacity, C, of the absorber and the heat conductance, G, of thermal link to the heat sink. The temperature change of the absorber is linked to the photon energy via $\Delta T \sim E_\gamma / C$ while the damping of the signal, that is, the time required to return to the reference temperature given by the heat sink is

$$T(t) = E_\gamma / C \exp(t/\tau_0) \qquad (3.3)$$

where $\tau_0 = C/G$ (McCammon 2005).

The best energy resolution that can be achieved with calorimeters requires to reconstruct the event of an impacting photon by determining the pulse height of the voltage signal as can be seen from Eq. (3.3). It is straight forward to see that one wants to have a hight time resolution of the voltage signal for the best resolution. But this leads to a high data throughput such that storing pulse shape of every event is at best impractical but mostly just not possible. To still obtain reliable results it is necessary to analyze the pulse signal as they occur.
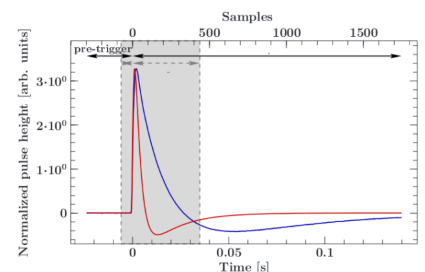

Figure 3.5: Idealized voltage function on photon impact. The two curves correspond to thin (red) and thick (blue) pixels. The gray area in the image covers the length of the mid res template for event reconstruction. The high res template covers the full range. Adopted from Hell (2017)

---

[2] This microcalorimeter is stationary placed at one of the open ports of the Lawrence Livermore National Laboratory (LLNL) Electron Beam Ion Trap (EBIT). EBITs will be discussed in the next chapter

The simplest method to obtain a measure for the peak height is by taking the highest value in the pulse shape. Although this already provides a first value the uncertainty introduced from this is very high and should be avoided if possible.

A much better estimate of the peak value is obtained by reproducing the measured event shape with a function. This is done via averaged templates of events. The detailed process is described in Hell (2017) and references therein.

The basic idea is that these templates are constructed from actual measured events and are fit with just one parameter, the peak height, to the signal. The overall process can be summarized in this way

1. Monitor voltage signal from thermistor

2. If voltage exceeds defined threshold, trigger event reconstruction

3. Fit template event to data

4. Subtract fitted event, if still above threshold, repeat

The total analyzing process is done in software after digitizing the raw signal (Thorn 2008). The best signal quality is achieved if the fit function, i.e., the template covers as many bins as possible. The reason for this will be discussed in the next section.

It is quite obvious that the quality of the measured data depends on the quality of the templates. As there is only a single parameter for fitting the shape of the templates must be as close as possible to the true event shape. To provide reasonable good templates for the data acquisition is the duty of the experimentalist. When we will discuss the calibration of the ECS data in Chapter V, we will treat the whole experimental setup as black box with no assumptions on the data quality.

2.2 Optimal filtering

A closer inspection of the signal from the thermistor reveals that the power spectrum of the noise and of an event compare equally just with a different scaling factor (McCammon 2005). In oder words, the signal is not correlated between different frequency bands such that the signal to noise ratio (SNR) for the event reconstruction gets better the larger the frequency domain is in which the signal is reconstructed. By averaging the noise signal N and compare it with the event data D in frequency space one can therefore obtain the highest possible resolution.

In practice is the transformation of data from time domain to frequency domain computationally too expensive. Therefore a technique called optimal filtering is used (Szymkowiak et al. 1993). For optimal filtering one assumes that the pulse shape solely depends on the signal strength such that it can be represented by a function f with a scaling factor h. Due to the uncorrelated noise between frequencies, the fitting is done in frequency domain. After determining the best fit $\mathrm{d}\chi^2/\mathrm{d}h = 0$ with $\chi^2 = \sum(D(\nu) - f(\nu))^2/N(\nu)^2$ and Fourier transforming the result back to time domain. One obtains for the scaling parameter

$$h = k' \sum D(t)F(t) \tag{3.4}$$

where k' a normalization constant and $F(t)$ the template function which only depends on the Fourier transform of the pulse shape f and the noise N. The process for optimal filter for the ECS is more thoroughly described in Hell (2017).

The ECS consists of an array of 36 microcalorimeters (pixels)[3] where

---

[3] Due to the hardware setup only 32 of the pixels can be in use at the same time

each pixel has its very own noise figure. Therefore, one needs to provide event functions for each pixel individually. The functions are generated by taking sample data with the ECS and average over multiple events.

From the pulse height analysis process described above one can see that the template has to fit over one event and one event only. To still obtain the pulse height for successive events a shorter template is provided for reconstruction (gray area in Figure 3.5), reducing the resolution for those events. If events with still shorter spacing in time occur one can only use the peak height itself, further degrading the resolution. This resolution degradation limits the brightness of sources for which the ECS gives the best possible results.

3 Dispersion spectroscopy

Calorimeters already provide a very high energy resolution and once a working calorimeter made it safely to space new insights into astronomical sources are almost granted. Still, the highest possible resolution is only reached with diffraction instruments. The theoretical limit for gratings is given by the Rayleigh criterion involving the size of the diffraction grating.

The diffraction pattern of gratings is given by the constructive and destructive interference of reflected waves from a grating or slit grid. The wave fronts have constructive interference when the path length of normally incident light between adjacent slits is an integer multiple of the wavelength λ. The situation is shown in Figure 3.6 from which it is easy to see that the angles under which maxima occur are given by

$$\cos \beta = mN\lambda \qquad (3.5)$$

with N the line density, i.e., the number of slits per length, m the diffraction order and θ_m the angle of the diffracted light w.r.t. to the grating plane.

Further, it follows for plane waves that for arbitrary incident and reflection angles the maxima occur at

$$\cos \alpha - \cos \beta = mN\lambda \qquad (3.6)$$

where α is the angle between the grating plane and the reflected direction.

The use of gratings does not directly give access to the energy spectrum since the photons still need to get detected. But instead of the resolution dependence of the detection process it is shifted to a spatial detection of photons. Suitable detectors are CCDs as they are used on board of Chandra and XMM to collect the photons from a dispersion grating (den Herder et al. 2001; Brinkman et al. 2000).

The tradeoff one has to take when using diffraction elements for X-ray measurements is that weak sources are extremely difficult to detect. This follows directly from the fact that the incident photons get distributed over a larger area leaving diffraction measurements only usable for bright sources or long observation times.

In the laboratory, however, where electron accelerators (e.g., BESSY, PETRAIII) provide luminous X-ray sources one can reverse the process and instead illuminate probes with nearly monochromatic X-rays. This allows very accurate measurements of transition levels in elements.

Part of this work was the analysis and calibration of data obtained from an experiment to precisely measure the oxygen Rydberg edge. X-rays from the BESSY beamline U-49/2 excited molecular oxygen in a gas cell where a electron multiplier detected the ion yield depending on the X-ray energy. For proper calibration He-like ions were produced in an EBIT providing reference lines with known energy.

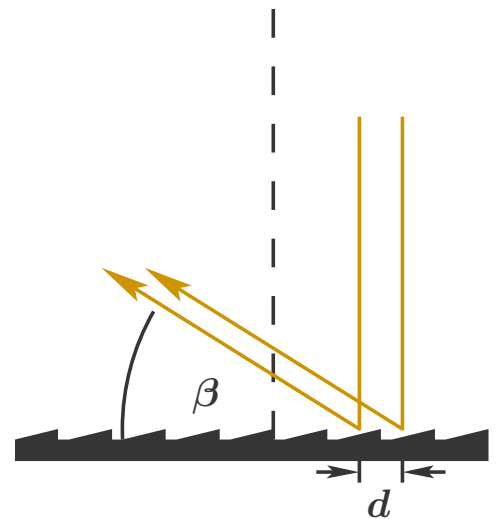The experimental setup and analysis is presented in Chapter V.

Figure 3.6: Constructive and destructive interference of light incident on a grating. The distance between adjacent lines is given in d.

Electron Beam Ion Traps

Every element has its very own characteristic energy spectrum. But not only do the energy levels differ between different elements but also between distinct ionization states of the same element.

Especially gas under conditions where X-ray radiation is produced due to hight temperatures is ionized to some fraction (depending on the temperature). For the proper identification of lines in astronomical sources and the derived quantities of them, it is important that precise and accurate reference values are available.

To measure these lines one needs a reliable way to produce plasma with desired ion distributions. For this task EBITs turned out to be an excellent and versatile tool.

1 The Livermore EBIT

The development of an easy accessible cold plasma source was driven by the need to investigate especially highly charged plasma clouds. Available devices either allowed only low charged states or generated hot plasma for hot plasma studies (Hell 2017).

The development of EBIT was driven at LLNL (Marrs 2008) providing an advancement of Electron Beam Ion Source (EBIS) which was developed as ion source for accelerator experiments (Briand et al. 1984).

As the name indicates features the EBIT an electron beam produced by a heated cathode which ionizes injected atoms. The beam is compressed by a strong magnetic field providing high current densities which are necessary for highly ionizing heavy elements. The magnetic field of the Livermore EBIT is therefore generated by superconducting Helmholtz coils.

The ions produced by collisions of the beam electrons with atoms and ions get trapped inside an electrostatic potential along the beam axis. Figure 4.1 depicts the general setup. Radial trapping is achieved due to the potential generated by the electron beam.

Due to the recombination of ions and atoms in the trap center, EBITs are a natural source of UV and X-ray radiation. For this work, the Livermore EBIT was the calibration source for the ECS providing narrow line emission at known energy.

The emitted X-rays from the EBIT were used in many publications, e.g., for transmission measurements of filters, calibrating crystal spectrometer and various other related tasks (Brown et al. 2010). But obviously is the EBIT not just a source for X-rays but also provides a tool to systematically study fundamental atomic physics of the elements. Almost all naturally abundant elements were trapped in the Livermore EBIT providing standard reference values for transition lines in highly charged plasma.
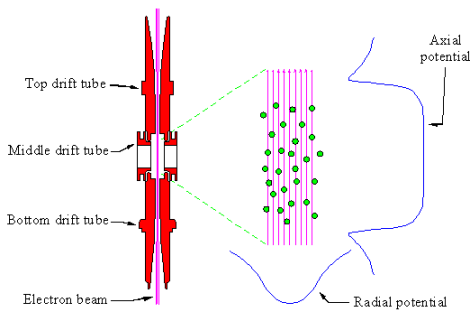
2 The PolarX-EBIT

Compared to the huge size of the Livermore EBIT does the PolarX-EBIT rightly belong to a series of so called *mini* EBITs (Micke et al. 2018) manufactured in Heidelberg. Where the Livermore EBIT fills a room completely is the PolarX-EBIT a desk sized device. This small design of all the mini EBITs is achieved by replacing of the superconducting coils with strong permanent magnets. This removes the cryogenic tank which takes up a large portion of the required space.

The permanent magnets of the mini EBITs are not able to produce a magnetic field with the same field densities as achieved by superconductors. This limits the electron beam energy but also allow to use only modest size power supplies. The graphic in Figure 4.2 shows the design of the PolarX-EBIT.
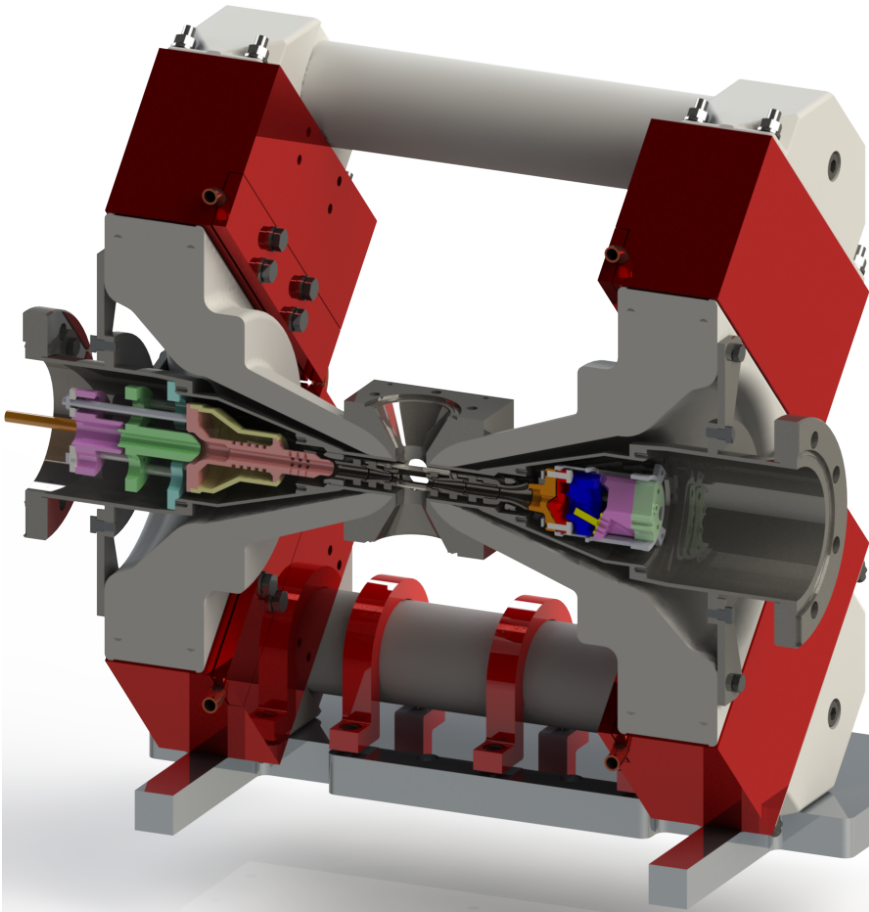

Figure 4.1: Schematic drawing of the trapping potential inside the Livermore EBIT. The careful selection of the electron beam energy gives access to the desired ion species. Adopted from https://www-ebit.lbl.gov/.

Figure 4.2: Structure of the Polar-X EBIT. The off-axis electron gun can be found in the right part below the trap axis. Courtesy of Steffen Kühn.

Compared to the other mini EBITs from the Heidelberg group is the PolarX-EBIT of special design. In the original EBIT design is the electron gun located on the trap axis and the electrons are accelerated parallel to the magnetic field lines. The PolarX-EBIT, however, has its electron gun located in a position off the trap axis. This has two implications: First, the electron beam must be accelerated in a direction that is different from the magnetic field lines applying a Lorentz force on the beam. Second, and more importantly, is the beam accessible from both sides. This design is key for the experiment and described in Chapter IV.

The Lorentz force that acts on the electron beam has the consequence that the setup of the beam is more complicated and requires additional magnetic fields for fine tuning (Micke et al. 2018).

3   EBITs in this work

The work presented in this thesis involved the Livermore EBIT as well as the PolarX-EBIT. The latter was used as a source for He-like ions interacting with an X-ray beam. The absorption and subsequent emission of the X-ray photons provide a known signal for reference. Due to the off-axis design of the PolarX-EBIT is it possible that only a fraction of the beam photons interact with the ions. The remaining photons leave the EBIT on the opposite site and can be used for the actual experiment.

Part of this work was to reconstruct the energy grid from the data obtained with the PolarX-EBIT. The detailed experimental setup and data

analysis is described in Chapter VI.

The main part of this thesis was to test an alternative calibration approach for the ECS. The calibration source for the ECS was provided by ionizing the elements Helium, Silicon, Sulfur and Argon in the Livermore EBIT. The recombination of the ions with electrons and the subsequent photon emission provide an X-ray signal of known shape. The exact calibration process and data extraction is described in Chapter V.

Chapter V

High precision detector calibration

It is very obvious that each measurement device needs a proper calibration to give reliable results. This calibration is most easily achieved by measuring a known source which can be sufficiently modeled. By comparing the data taken with the detector and the model one can find a mapping such that both agree. The aim of calibration measurements is to find this mapping. In the following we will first discuss a straight forward approach to find the mapping and later improve this strategy to calibrate the ECS.

1 Calibrating the ECS

In this section we will discuss how the ECS can be calibrated in a very easy way. As described in Hell (2017) does the ECS consist of 36 pixels which are completely independent of one another. That has the consequence that each of those pixels needs an individual calibration function. This is easily seen in Figure 5.1. If one looks closely one finds that almost no two pixels have matching data.

The calibration mapping is expected to be sufficiently described by a polynomial of 5th order. The observed and known sources are emission lines where the transition energy is known from theory or experiment to high precision and accuracy.

The energy resolution of the pixels is not high enough to resolve the line shape of the transitions. Therefore one can model the lines with Gaussian profiles and obtain the line positions by fitting profiles to the data. The count rate is in general not high enough in the individual pixels to obtain a reasonable fit. To increase the count rate for the fitting process one can find a mapping between the pixels by calculating the cross correlation of the data arrays and calculate an effective number of counts per bin for a reference pixel. The data grids of the other pixels is stretched and shifted according to the correlation by a first order polynomial.

This new effective data set can be used to obtain reasonable fits to the emission lines. From the fitted line centroid and the known line energy one finds the mapping by fitting the polynomial accordingly. To obtain the calibration mapping for the individual pixels one finally has to correct the calibration polynomial with the linear polynomial obtained by the cross correlation.

This method works reliably well and is, in particular, a very fast way to obtain a qualitative calibration. Nethertheless, the procedure as described here has some weak spots which might screw any subsequent measurements based on this calibration. In the following we will first discuss what problems this method has and later how we can increase accuracy of the calibration.

1.1 New calibration method

One of the more obvious problems with the calibration method as described previously is that it does not provide an uncertainty estimation. Although, one can calculate confidence intervals for the calibration polynomial applying Wilks theorem (Wilks 1938), does this not include the uncertainty from the cross correlation.

A similar problem arises from the fact that the cross correlation is only applied asymmetrical between the pixels. This has the consequence that when the reference pixel is a statistical outlier, the calibration is biased towards the reference pixel.

Other problems arise when one tries to use the calibration far from the energy where the reference lines were measured. In this case one can not trust the polynomial any more and the result will be erroneous. This, however, is a problem of data itself and not so much dependent on how the detector is calibrated.
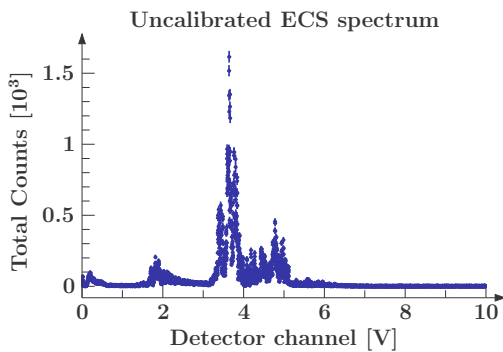

Figure 5.1: Uncalibrated data summed over all pixels. The emission lines are from Ne, Si and S. One can clearly see that the pixels do not agree and in sum create a badly resolved spectrum

1.2 Fitting the pixels simultaneously

To improve the calibration we describe the data in individual by emission line profiles whose line centroids are tied together by the calibration polynomial. This will not improve the situation on its one but when fitting the model to all pixel data at once we obtain the same statistical significance as if we would add up all the data. This task is performed within the Interactive Spectral Interpreting System (ISIS) environment.

ISIS allows one to define arbitrary complex models by modifying how the data is treated. We use this to define only one model for all the data from all pixels. For now we say that each data set is tagged by the element that was measured and that the model returns the corresponding line at the theoretical line position (e.g., in energy). It is import to note that some parameters of the model are global and other are local to the data or to the pixel (which will be groups of multiple data sets).

The parameters that are free for the fitting procedure are of course the coefficients of the calibration polynomial and the normalizations for the emission lines. As the detector resolution is not high enough to resolve the line shape we use delta peaks. The detector resolution is modeled by convolving the model with a Gaussian kernel. The overall structure of the model is displayed in Table 5.1.

By fitting this model to the data one obtains the calibration polynomial and by examining the fit statistic one can calculate the confidence intervals for each parameter (see 5). This might not necessarily improve the calibration but gives direct access to the uncertainty.

1.3 Adding measurement lines

In the next step we also use the lines we are really interested in to match the calibration internally between the pixels. For this we add line profiles to the model such that the line energy is fixed and gets mapped by the calibration polynomial onto the detector grid. Again, the normalization is data set specific as not only the pixels are different in response but also the measurement times differ.

Due to those additional lines we can ensure that the calibration is aligned between the pixels. This does not improve the calibration itself, it just makes sure that the pixels match upon unknown lines. Provided that one has enough calibration lines, e.g., lines where the centroid energy is known around the lines of interest the calibration should also improve for the measured lines.

The fit process is than to first find a good agreement between the model of the reference lines and the measurement and add unknown line model components gradually to improve the calibration. When the fit succeeds the result should not only give a decent calibration but also reliable results for the unknown line positions.

2 Data extraction and model assignment

Now that the overall strategy is clear we will in detail examine how the data gets extracted from the raw observation. Further we will see how the model is working in detail such that the fitting process can be understood.

As the working principle of the ECS was already described in Chapter IV we focus only on the measurement data as they were recorded in event files. Each measurement campaign measures four (relevant) parameters for one event: Timestamp, quality, pixel and peak height. The timestamp is used to group the events such that events from the same experimental settings are recognized as one data set. An exemplary measurement campaign is

Table 5.1: General structure of the calibration model

| Model parameter	Description
X1_lines.norm(1)	Normalization for transition line 1 to ground level for element X
X1_lines.norm(2)	Normalization for line 2 to ground level
⋮	⋮
X2_lines.norm(1)	Normalization for transition line 1 to first excited level
⋮	⋮
gauss_conv(1).sigma	Resolution for pixel 1
⋮	⋮
poly(1).c0	Constant parameter of calibration polynomial for pixel 1
poly(2).c1	Linear parameter of polynomial
⋮	⋮
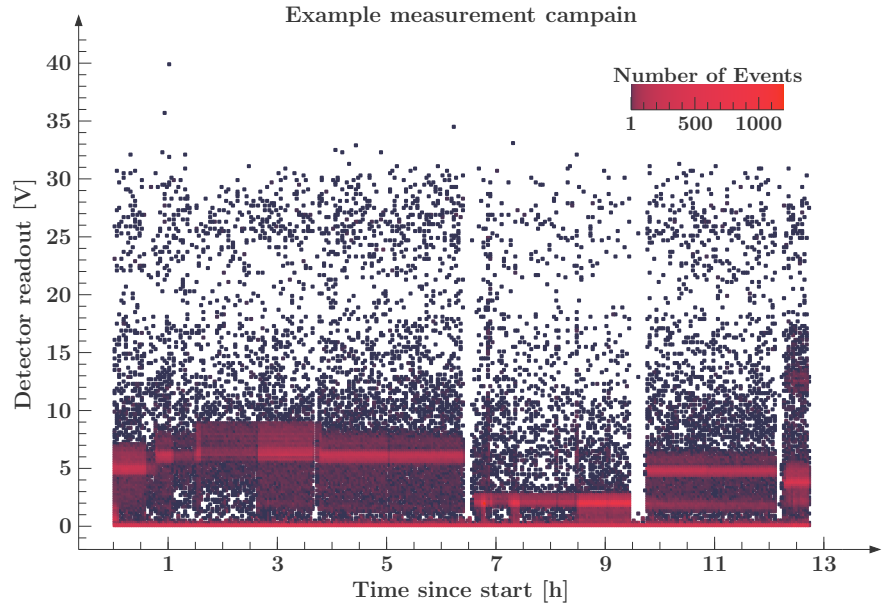
Figure 5.3: This plot shows how the events occurred during the measurement. Events closed to each other where grouped for displaying purposes. Gaps in time direction indicate markers where the experimental conditions have changed.

displayed in Figure 5.3. Here a change in settings is usually marked by a gap in the events[1,2].

2.1 Constructing data histograms

From the event list regions which contain the events that such be part of the calibration are selected. The original intention of the experiment was to measure the absorption of different filters[3]. Therefore one can see intervals with different count rate in Figure 5.3 (especially in the lower energy region) showing the same emission lines. For the calibration we consider all events that were taken for one element injected no matter if a filter was inserted or not. This does not change the calibration outcome it only reduces the number of events per line.

From the chosen (time) regions in the event list one can construct the data by sorting the events into a histogram. For each pixel a bin width of 0.02 V was chosen. This choice is somewhat arbitrary. Arguments for this particular binning are that the resolution is still sufficiently good and the count rates in the bins not to low.

The exposure time for the different emission lines was in general long enough such that we can pick only events with the highest quality flags. After sorting the events into the histogram one obtains a spectrum as already shown in Figure 5.1. The spectrum for the individual pixels is displayed in Figure 5.4.

The different data sets consist of the events from one pixel in one specific selected time interval. This is necessary as each data set is tagged with the element that was measured together with a flag that tells if the data is from an element used for calibration or not. This information is later used

---

[1] Due to how the data is displayed not all gaps are visible in this plot

[2] The plot not only shows differences when the injected gas was changed, but also when filters where used. This reduces the number of events in certain regions

[3] Essentially one measures the thickness by comparing the transmission with and without filter

in the model to decide what parameters of the line are independent. The model itself only contributes counts to a data set if the line is present in the data. In other words, the model is not just one model but different for each pixel and element just that some parameters like the coefficients for the polynomials are forced to be the same for each pixel.

3   Calibration line model

This section will describe how the data is modeled and also how the calibration is constructed from the resulting best fit. In the end we will obtain a functional description of (parts) of the ECS from which one can construct a response matrix. This matrix can then be assigned to the data such that the data can be loaded in a way that is very common in X-ray astronomy: Instead of manipulating the data by calculating calibrated data from the raw detector signal, one provides a description of the detector. When fitting the data with models, the model function is convolved with the response such that the detector idiosyncrasies are taken into account.

3.1   The model for one pixel

For a better understanding on how the model works we first look at the data for one pixel. For further simplification let us only view at data for one calibration element. in this case our model predicts the counts as

$$C(h; p) = R(h; \sigma) * F(P(h; c_0, c_1, c_2, c_3, c_4, c_5); X, N_1, N_2, \ldots) \qquad (5.1)$$

for detector bin h depending on parameter set p. Here R denotes the resolution model by a Gaussian convolution with width σ, F the line model for element X with normalization N_i for transition line i and P the polynomial with its 6 coefficients. The model for the emission lines does not depend on any additional parameter as the line position in energy as well as its shape is predicted by theory and fixed. Note that F gives the predicted counts for the energy grid and only the polynomial P translates the detector grid to the energy grid.

As mentioned previously is the ECS resolution not sensitive enough to resolve the line shape. Therefore F only consists of a series of delta peaks whose final width is determined by R.

The events also contain photon impacts from background emission of the EBIT. This can be described by the exponential function

$$B(h; \lambda, c) = \exp(-\lambda h) + c. \qquad (5.2)$$

Further, as the exposure of the different measurements might not be the same, an additional normalization term is introduced. The total model is then given by

$$C'(h; p') = N_n \cdot (C(h; p) + B(h; \lambda, c)) \qquad (5.3)$$

where n is a number unique to each data set.

If more calibration elements are involved one gets the general model via the replacement $F(E, X, N_1, \ldots) \to F(E, \bar{X}, \bar{N})$:

$$\begin{aligned} F(E, \bar{X}, \bar{N}) = {} & F(E, X_1, N_{X_1,1}, N_{X_1,2}, \ldots) \\ & + F(E, X_2, N_{X_2,1}, N_{X_2,2}, \ldots) + \ldots \quad . \end{aligned} \qquad (5.4)$$

The model is now almost complete, the only remaining thing to add is a model component for the unknown lines measured. The only difference between the calibration lines and the lines yet to measure are the additional parameters for the central energy and the line shape. In general those lines
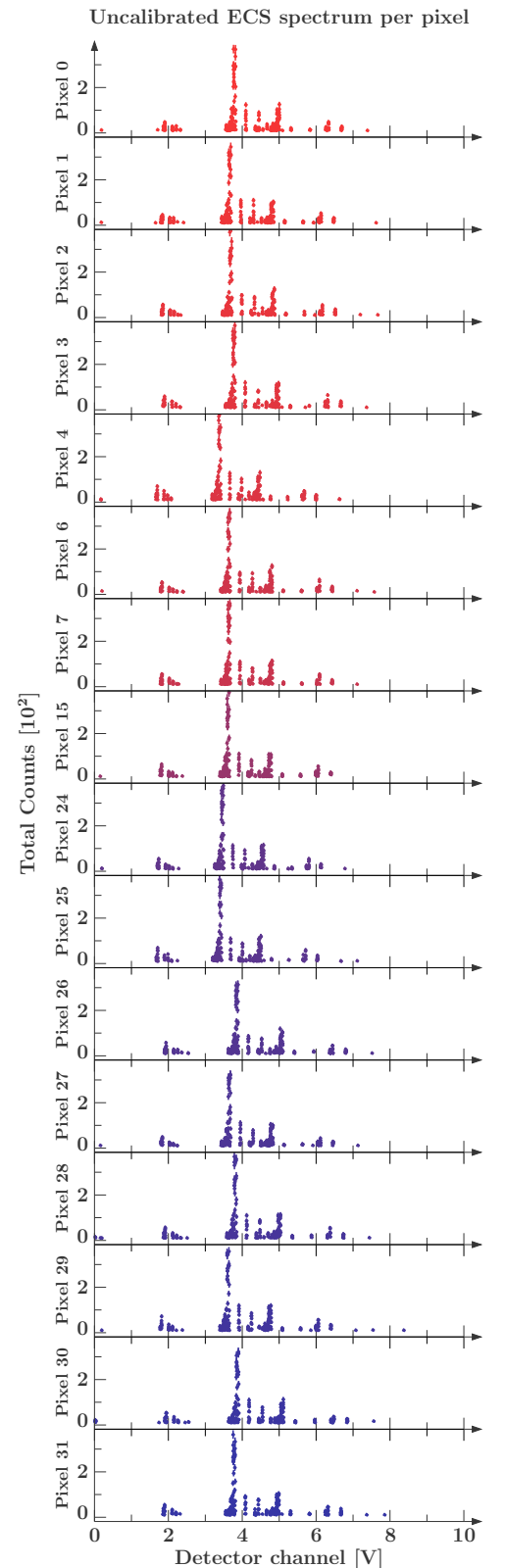

Figure 5.4: This figure shows the binned histograms for each pixel. Not every of the 32 pixel was working and gave a constant signal of 0 V which is why they are ignored here.

are described by a Voigt profile, but again, the shape cannot be resolved such that we only use delta peaks. The line function is in the end defined as

$$
\begin{aligned}
F(E, \bar{X}, \bar{N}, \bar{E}_c, \bar{N}_u) = & F(E, \bar{X}, \bar{N}) \\
& + N_{u,1} \cdot \delta(E_{c,1} - E) \\
& + N_{u,2} \cdot \delta(E_{c,2} - E) + \ldots \quad .
\end{aligned}
\tag{5.5}
$$

Things get much more complicated when we consider this model for multiple pixel. In that case, not only do we need to keep track of the parameters that are the same for different pixel, but also ensure that each model only contributes to the data where components are set. For this task we use ISIS as it has a very powerful mechanism build in that almost does what we need.

3.2   Implementing the model

In ISIS one can very simply load data into a histogram as long as one can provide values, errors and a binned grid that is defined by lower and upper boundaries. Grids that are not in ascending order or have non matching boundaries as, e.g., $\text{low}_{i+1} \neq \text{high}_i$ are invalid and will raise an exception. This is mentioned explicitly here because when calculating a grid from a polynomial this might be violated[4].

Defining a function for fitting is very easy using ISIS. Through the call `fit_fun("function(1)+function(2)")` a fit model consisting of to (different) instances of the function with name *function* will be used. Both instances have individual parameters which can be accessed by the integer given in the function argument.

In our case, we have one data set for each pixel for each measurement that was taken. For the data we consider here this gives a total of ? data sets[5] where ? are for calibration (Si,S,Ne) and the remaining data is considered as unknown[6].

ISIS allows to define fit functions in a way such that there is an instance of the function for each data set. With some modifications this can be used such that there are function instances for each pixel (the calibration polynomial) and for each measured set (the observed spectrum). This is useful as it reduces number of available parameters to the required minimum compared to the straight forward approach where parameters that describe the same quantity for different data sets are tied together.

The model that is used to describe the data is therefore constructed in a way such that for each pixel there is one calibration polynomial that influences the lines. Each line, however, has only one instance for every measurement set.

The final fit function is expressed as   where the `poly_fold` function returns 0 but sets the polynomial internal for the `ecs_lines`. Both, the known and unknown lines are hidden inside the `ecs_lines` function and undergo the broadening of the Gaussian convolution expressed in `gauss_conv`. The `response` allows to adjust the signal gain between the pixels and `exp_bkg`

---

[4] There are methods to construct polynomials that are monotonic increasing in an arbitrary interval (the required setup here) but are difficult to implement in a general context. One can, however, find a parameterization that is monotonic increasing for the argument in $[0, \infty)$. Because it is not clear if the polynomial can be assumed to have this form, this was not considered.

[5] From the 32 pixels not all are working or give reasonable signals. Those were ignored thought the whole process.

[6] The unknown transition lines originate from Argon. Only the low charge states of Argon are not known but treating the high charge states also as unknown gives a measure for the quality of the calibration.

adds an additional exponential component to compensate for background photons.

Although the function looks not very complicated as expressed term it expands to numerous parameters that are used to match the data. For the used data set a total number of ? parameters are generated. The full form of the fit function is in that sense complicated as it is not very easy to perform the fit and obtain a reasonable description of the data. A problem that occurs and which is quite easy to understand comes from the line model. One can imagine that when a narrow peak[7] is modeled with a Gaussian (or equivalent) profile the used optimization routine will only converge to a decent solution if the peak model has sufficient overlap with the data peak.

A simple (and not very sophisticated) test was made to find the overlap that is at least required. For the test a set of random data points were draw from a Poisson distribution where there mean was given by a Gaussian profile with width 0.5 and centered at 5.5. The fit was then performed by setting the normalization and width of the fit function to the true values and the center in the range from 1 to 10. For each center position 1000 trials were performed to fit the data. Figure 5.5 displays the result for different optimization methods based on the χ^2 statistic.

Now, in the model used for calibrating the micro calorimeters the position of the lines are set by the polynomial. This means that when the initial polynomial coefficients are not close to the true values, the fit will horribly fail. The next section outlines the procedure in order to find reasonable start values and how the different components can be fit.

4 Fitting the data

As with every model that is of more complexity than a simple linear regression the selection of initial parameter sets is crucial. Not only is the time for the optimization procedure reduced with good starting values it often allows to find optimal solutions in the first place. For our calibration model the initial parameters that influence the convergence the most are the polynomial coefficients.

The reason is on the one hand the overlap of model and data line profiles as discussed in the previous section but also the inherit behavior of the polynomial. On imagines that some of the lines are modeled quite well while others are far off[8]. In this situation a optimization routine tries to modify one of the coefficients with the result that now *all* the lines are off. So the algorithm comes to the conclusion that varying the selected coefficient does not improve the model. But this is of course true for all coefficients and it is more and more difficult for the higher order ones.

In order to circumvent this problem two general solutions are feasible. The straight forward way is to use the results obtained from earlier calibration methods as discussed in Section 1 as starting points. The other solution is to place a subset of the modeled lines at the observed positions by hand. This method has two advantages: First it is not affected by any bias that might be introduced during previous calibrations. And second, although the polynomial might qualitatively hit the same points in the first method, the shape can be very different. When setting the initial positions of some lines the shape of the polynomials between the pixels should be very similar depending on the chosen lines.

For those reasons and also to avoid to end up with a self fulfilling calibration we set the coefficients by hand for each pixel in the following
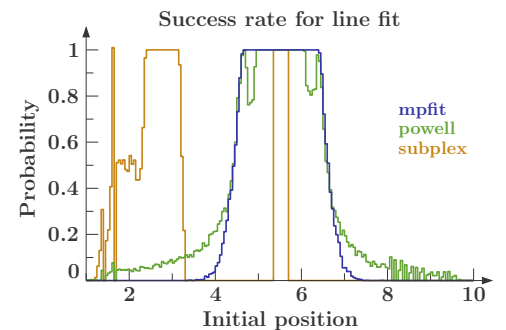

Figure 5.5: Fit test for different start positions of a Gaussian line to randomly generated gauss data centered at 5.5. The figure shows the success rate for each fit try where a success is defined where $\chi^2_{\text{red.}} <= 2$. Generally it depends on the optimization method how large the initial difference between model and data might be. The tested optimization methods were MPFIT (Levenberg 1944), POWELL (Powell 1964) and SUBPLEX (Rowan 1990) for minimizing the χ^2 statistic.

---

[7] Narrow in the sense that the line width only covers a small fraction of all data points

[8] The modeled lines are of course present but not at the observed detector units

way:

1. Set the zeroth order coefficient to zero $c_0 = 0$.

2. Set the first order coefficient to a value such that the range is roughly correct. In our case $c_1 = 500$ (I.e., the energy is roughly 500 times the detector unit).

3. Select a predicted line position at low energies/detector units but not lower as already selected lines and select the true position of the line.

4. Solve the linear equation(s) for the new coefficient(s) such that the polynomial hits the selected points.

5. Repeat from Step 3 until the number of selected lines matches the order of the polynomial.

This process worked quite reliable only at certain points the solution for the linear equation was a polynomial where higher detector units mapped to lower energy values. In this case the process was repeated with different lines selected.

5   Fitting, fit method and fit statistic

As one has starting values for the polynomial coefficients the best strategy to fit the calibration function is to first find good solutions for the pixels individually, i.e., apart from the calibration lines no other line components are added. The best strategy here is to iteratively fit the normalization for the calibration lines without any other free parameter. And then the polynomial coefficients without any other free parameter. If one tries to fit both parameter classes right from the beginning, the optimization algorithm might find a better solution by minimizing the counts in the lines and shift them away from the true values.

For the fitting itself several established algorithms are available. The most widely used algorithm is probably the Levenberg-Marquardt (LM) least squares method (Levenberg 1944). The algorithm minimizes the least squares, generally also named χ^2, so the sum of the square difference between model M and data D

$$\chi^2 = \sum_i^N w_i (D_i - M_i)^2.$$

The weights w differ for different realizations and different assumptions. For the ordinary least squares $w_i = 1$. The method introduced by Marquardt solves the minimization problem by iteratively taking steps towards steepest decent (Marquardt 1963). Levenberg contributed a modification to allow for faster convergence which is similar to the Gauss-Newton method of root finding.

Further contributions from various authors made the algorithm a fast converging, robust and scale invariant method to minimize a function. In the particular case of curve fitting it minimizes the χ^2.

The χ^2-statistic in ISIS can be used with different weights to control where the minimum is located. The usual case for χ^2 is to use as weights the squared inverse of the data uncertainty. This ensures that measured data points with large error do contribute little to the statistical value when they are not very well described by the model.

The first introduction of the χ^2-statistic goes back to Pearson (1900) who found that the χ^2 is a good measure to test for acceptance or rejection of

a predictive model. He also concluded that this can be used to find better solutions for the same model when varying the parameters. However, the χ^2 statistic assumes uncorrelated and normal distributed data!

When an experiment is performed where the outcome is the counted number of event occurrences the data is always Poisson distributed. The experiment is in our case the counting of photons per time interval (and energy). For a large number of counts N the Poisson distribution is very accurately described by a normal distribution with mean $\sqrt{N}$. This can be justified by treating a Poisson distribution of N variables as N independent Poisson distributions of 1 variable[9]. By the central limit theorem those are distributed normally.

So it is clear that for counting experiments with a large number of counts (usually ≥ 20 is considered as the limit) the data is sufficiently described by a normal distribution and, hence, can be analyzed using the χ^2-statistic. However, for our case the number of counts is rarely larger than 15. And those are only reached in the peak centers of the emission lines. Most of the data bins have a value of 0, 1 or 2. So, equally clear as before, the χ^2-statistic is not applicable for our particular data.

5.1   The Cash statistic

The underlying idea behind the minimization of is to maximize the probability that a set of data points, which can be truly described by the implied model, result in the observed numbers. This combined probability is called the maximum likelihood (for given data and model, Wilks 1938). And is, in the case of uncorrelated data points, simply given by the joint probability of the individual measurements.

The χ^2-statistic is a straight forward realization of the maximum likelihood function. Assuming that the observed data is truly following the model of interest and that the data is uncorrelated[10]. Then, when further assumed, that the distribution of the data is normal with mean given by the model, the maximum likelihood is simply

$$\mathcal{L} = \prod_i^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(D_i - M_i)^2}{2\sigma^2}}.$$

For practical reasons one takes the logarithm of the maximum likelihood so the product turns into a summation. Additionally the sign is inverted such that minimizations algorithms can be applied. Therefor the log likelihood function is

$$-2\ln\mathcal{L} = \sum_{i=1}^N (D_i - M_i)^2/\sigma^2 \qquad (5.6)$$

ignoring constant terms and multiplied by two.

But we have already seen that the normal distribution is not suitable for our case. This is very easy to see when plotting the normal distribution as approximation for Poisson distribution and the Poisson distribution together. In Figure 5.6 the difference between both distributions at low expectation values is fairly large, while for higher values the difference gets small.

The normal approximation for low count numbers has a bias towards lower count numbers. That means, when used for minimization, the χ^2 minimum predicts lower count numbers for the model as observed. This is

---

[9] This can be done as a sum of Poisson distributed variables is again Poisson distributed.
[10] To clarify, correlation means here that each measurement made is completely independent of any other measurement. No matter if measured for a different quantity (here, e.g., energy) or at a different time.
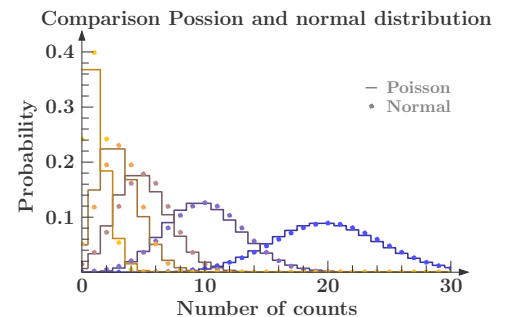

Figure 5.6: Comparison between normal and Poisson distribution for various expectation values. The difference at low expectation values proves that the normal distribution underestimates the true value for low count numbers.

fairly easy to understand as the normal distribution extends beyond zero which accumulate while the Poisson distribution has a finite probability only for non-negative integers.

To circumvent this problem Cash (1979) introduced a better statistical measure for experiments with low number of counts. He derived the statistical contribution by essentially replacing the normal distribution in the log likelihood by a Poisson distribution. The general assumption that the measurements are independent is again assumed which implies that the new statistic be expressed as

$$C = -2 \sum_{i=1}^{N} D_i \ln M_i - M_i. \qquad (5.7)$$

Here, all terms containing only the data are dropped as for comparison between two parameter sets of the same model they cancel anyway. It is convenient to write the Cash statistic as $C = 2 \sum_{i=1}^{N} (M_i - D_i) + D_i \ln(D_i/M_i)$[11] because when dividing by the number of degrees of freedom $C_{\text{red.}} = C/\text{ndf}$ provides a qualitative measure for the goodness of the fit. If $C_{\text{red.}}$ is much smaller than unity the model primarily describes the noise while if it is much larger than unity the model does not describe the data at all.

Since in X-ray experiments the measured quantity is in many cases the counting of events it is in general a good idea to use the Cash statistic instead of the χ^2-statistic. Neither is it computationally more expensive nor does it break down for large count numbers.

Others have proposed modifications to the χ^2 statistic to circumvent problems that are associated with low count rates. For example is the modification introduced by Mighell (1999) intended to give a good statistic for the observation of zero counts. As the usual χ^2-statistic is not defined for this case (division by 0) usually an error of 1 is assumed for measurements with outcome 0. But this introduces a bias when fitting the data and also impacts the calculation of confidence intervals.

When the proper statistic for counting experiments is used in the first place, non of the problems will occur. A minor problem arises, however, with the Cash statistic (not exclusively) that is related to the minimization method in use.

5.2 Minimization methods

Besides the LM algorithm that we already discussed shortly in the previous section there are many different approaches to solve the minimization problem for a given function with parameters. All practical solutions describe an iterative process that makes parameter changes based on decision criterions. In the case of the Marquardt algorithm the decision is made according to the (numerical derived) steepest decent vector in the parameter space.

Most algorithms are, like the LM, based on gradient calculation methods. The advantage here is, that the number of function evaluations is rather small and they can utilize the will tested and optimized libraries for solving linear problems (e.g., MINPACK or LAPACK). On the other hand, those algorithms are very often prone to local minima traps, e.g., the algorithm will not be able to find a possible global minimum when the gradient vector pulls the parametric solution always back to a local minimum. Other methods try to avoid this problem either by randomly testing other parameter combinations or not using a gradient method at all.

---

[11]   This modification essentially adds an offset term such that the minimum of the statistic for each data bin is zero.

Table 5.2: Optimization methods (general referred to as fit-methods) implemented in ISIS (except for `emcee` which is provided by the Isisscripts developed at the Dr. Karl Remeis Sternwarte Bamberg.

Method	Description
`mpfit, lmdif`	LM algorithm described in Marquardt (1963)
`marquardt, plm`	Alternative LM algorithms provided by ISIS
`diffevol`	Differential evolution algorithm based on stochastic population.
`simplex`	Nelder-Mead algorithm for general unconstraint minimization.
`powell`	Non-gradient method describe by Powell (1964).
`simann`	Probabilistic method based on simulated annealing.
`subplex`	Derivate of the `simplex` algorithm requiring less function evaluations compared to simplex.
`emcee`	Markov-Chain Monte-Carlo method for probabilistic exploration of parameter landscape following a Markov-Chain model.

In Table 5.2 a list of the available optimization methods which are implemented in ISIS is given. A subset of those will be discussed in the following.

mpfit

The `mpfit` routine is the already discussed LM algorithm for finding solutions near the true minimum of a function. The algorithm was first described by Marquardt (1963) as a solution for the minimization problem using a strategy that was described by Levenberg (1944) already 1944.

The general idea behind the algorithm is that it follows the gradient of the χ^2 landscape when far from a minimum and approaches analytical solution by second order approximation when near a minimum. The controlling parameter is usually called λ and acts as a damping factor for the curvature matrix such that when λ is large the matrix equation degenerates into independent equations. At each step λ is adjusted based on the difference between the current and the newly calculated function value.

The user can control the behavior of the minimization strategy by controlling when and how the damping parameter is updated. Therefore several strategies are can be used for problems of different complexity.

Despite its very general applicability LM can cause a large number of function evaluations when the path through the χ^2 landscape follows a narrow canyon (Lampton 1997). A particular problem in our case is that estimated minimum from the Cash statistic is notably away from the true solution at least when the standard settings are used.

This can be understood as when the Cash statistic is inserted in the gradient equation of the LM algorithm it is identical to zero in first order approximation. That said, when the function is close to the true minimum, the gradient is evaluated to (or at least nearly) zero. But this means, for the algorithm, it has converged. To avoid this one has to tweak the control parameters for LM.

powell

The fit method first described by Powell (1964) has the advantage to not calculating any derivatives of the functions. Therefore it is not as easily confused by complicated parameter landscapes compared to usual gradient descent methods.

The algorithm is, concept wise, explained very briefly. One starts from the initial set of parameters and search for the minimum in each parameter direction. In each step one calculates new search directions such that the final directions are conjugate.

Unfortunately, Powell (1964) made a mistake in the argumentation allowing for directions to emerge which span only a subspace of the parameter space. In consequence, the algorithm could reach a state at which a dimension of the parameter could never again change. Therefore, the solution could be incorrect.

? corrected the procedure accordingly and, with further improvements, showed that the method converges to the minimum for any strictly convex function.

subplex

In his dissertation about the analysis of algorithmic stability **?** described a variation of the famous Nelder-Mead Simplex (NMS) algorithm that tries to avoid the general pitfalls that NMS has. A brief description of NMS follows (a better explanation is given by **?**).

A simplex is the convex hull enclosing $n+1$ points in an n-dimensional space. In the 2d case the simplex is a triangle, in 3d a tetrahedron, etc. At each step, NMS evaluates the function at the points of the simplex and, depending on the outcome, will either stretch or contract the simplex.

Now, NMS is not very efficient when the number of dimension is large neither is the convergence fast. To compensate for this, subplex runs the NMS method only in a subspace of the full parameter space. The subspaces are selected and modified following essentially the same idea as the `powell` algorithm does for its search directions. The obvious difference: Instead of directions subplex searches within orthogonal subspaces.

emcee

Goodman & Weare (2010) described a Monte-Carlo Markov-Chain (MCMC) method for exploring and optimizing difficult functions. The interesting concept makes use of the observation that algorithms like NMS are invariant under afine transformation and therefore have superior performance when applied to highly skewed distributions. Consequently, the MCMC method described is also invariant under afine transformations.

In the usual way of Monte-Carlo methods is a non-probabilistic value described by a random sample. Each sample variable is drawn from a (mostly uniform) distribution and accepted if it falls inside the parameter space outlined by the problem else rejected. The standard example for a Monte-Carlo method is to approximate π by drawing N times two random numbers (x, y) from a uniform distribution $(0, 1)$. Each time the pair (x, y) fulfills $sqrt(x^2 + y^2) \leq 1$ the sample is accepted. The (approximated) value for π is then given by $4A/N$ where A is the number of accepted samples.

On the other hand, Markov-Chains are a simple way of modeling a stochastic process. A sequence of stochastic steps is said to be Markov when one can deduce the future and past of the current state from this said state.

The algorithm to use MCMC for fitting data is described in . The idea is to start with an ensemble of walkers placed on the parameter surface. At each iteration each walker is updated, i.e., moves in a certain direction. The move is calculated by selecting one of the other walkers and go an amount Z of the vector between the position of the walkers[12].

The acceptance rule is here imposed by a normal distribution where the probability of acceptance is calculated from $\Delta\chi^2 = \chi^2_{\text{current}} - \chi^2_{\text{new}}$. If the new χ^2 is better then the old one (that means, smaller) the step is always accepted.

It should be clear that this implementation is only working when the underlying fit statistic is χ^2. Otherwise the acceptance rule has to be modified to be used for, e.g., Cash.

From those optimization methods discussed above the `subplex` method was chosen as it does not underestimate the Cash statistic like `mpfit` does. `emcee` was not chosen because the implementation in the Isisscripts was only handling proper χ^2-statistic. And finally, `powell` was discarded because it performed notably worse than `subplex` in convergence as well as number of function evaluations.

---

[12] The stretch factor Z is drawn randomly from a symmetric distribution $g(z)$. Symmetric here means that $g(1/z) = zg(z)$ such that the reversed step is equally likely.

6 Confidence intervals

Besides finding the best description of a data set giving a certain model it is also important to decide if the model is describing the data in the first place and how confident the analyst is about the found parameters.

The first criterion was already briefly mentioned, that is, deciding if a model does indeed describe the data. It should be fairly obvious that one can derive a measure directly from the log likelihood function. Since the maximum likelihood is the probability that the observed data was drawn from the tested model and the data is independent one should expect that on average the probability should be more than $\frac{1}{2}$. It turns out that for this measure the arithmetic mean is not a good indicator rather the geometric mean should be used. So by dividing the χ^2 by the number of data points (that is the number of degrees of freedom) one has a qualitative measure for the fit. It is relatively easy to show that this is indeed the geometric mean of the maximum likelihood probability and that when $\chi^2_{\text{red.}} \sim 1$ the geometric mean is $e^{-\frac{1}{2}} > 0.5$.

In the extreme cases, if $\chi^2_{\text{red.}}$ is much smaller than unity the mean probability is close to one. Which is highly unlikely for error prone data. Conversely, if $\chi^2_{\text{red.}}$ is much larger the mean probability is zero, indicating that the data is not drawn from the model distribution[13].

6.1 Computing confidence

As soon as a reasonable fit is found, that means a fit where the reduced statistic is of order unity, one can calculate the confidence intervals.

There are several approaches to obtain an estimate of the model uncertainty of which we will discuss three here. However, it is necessary to understand what a confidence interval is and more importantly how it is to be understood.

Whenever there is measured data there is also an uncertainty involved. This uncertainty is tightly bound to the data and how the data was generated. Under general circumstances those uncertainties can only be estimated and are given either as absolute or relative errors to the measured value.

In many situations those uncertainties are given as 1σ standard normal errors. That means, that in 68% (that is the 1σ correspondence) of repetitions of the measurement the value will be within those uncertainty bounds and they will be normal distributed around the (true) value.

The confidence interval calculated for model parameters is naively also interpreted as the *trust region* of the best fit parameter. However, the parameter confidence intervals have no information about the model but about the data. In other words, a given confidence interval for a model parameter does not tell that with such and such confidence the true parameter is within the uncertainty bounds[14].

The interpretation of parameter confidence intervals has to be the following: If the true value (that means, a value that is obtained after many repetitions of the experiment) is *not* within the confidence interval, the performed experiment is exceptional, i.e., there is only a $(100-68)\%$ chance (assuming 1σ confidence) that this has happened.

Now, how can confidence intervals be computed? The first and (almost always) wrong way is to use the covariance matrix that is calculated during

---

[13] The reduced χ^2 is only a qualitative measurement for the quality of a fit given a certain model. It does not provide a resilient decision criterion between two competing models. For this one has to do a proper hypothesis testing.

[14] The parameter has the definite value as found by the optimization method. Also, either the parameter is in the confidence interval (good fit) or outside (bad fit).

the optimization procedure with LM. The estimate here is in most cases not good as the covariance matrix is calculated from the Hessian matrix which, in turn, is estimated from the square of the Jacobian matrix. However, if the residuals are non-linear the result underestimates the true uncertainty.

A much better way to calculate uncertainties is by examining the statistical landscape. The standard approach in astronomy to obtain the confidence interval of the i-th parameter Θ_i is to fit all other parameters while Θ_i is fixed to some value. The resulting best fit is inside the confidence interval if the statistical difference to the global best fit is smaller than the χ^2-distribution for one degree of freedom and the requested confidence level (Wilks 1938).

By carefully stepping the parameter value one may find the confidence boundary. It should be obvious that this only gives the confidence interval up to a certain accuracy.

The statistically correct approach to obtain the true confidence interval is by repeating the measurement several times. From the distribution of the measured values one obtains the confidence interval. Of course, this is not practical in reality such that Monte-Carlo methods are used. By simulating measurements according to the global best fit one obtains the distribution of the measured values from the simulated data. However, this implies that the best fit model is the optimal solution and one knows the statistical distribution of the data.

Since the method is well know and allows to find improved fit models we used the confidence estimation from the χ^2-distribution. This assumes that the used measure, the fit-statistic, does follow a χ^2-distribution. Cash (1976) has shown that this is indeed true for any underlying statistical distribution as long it is well behaved and convergent.

7   Calibration results

As of the end of this work the model was not fully converged which is why it is difficult to give absolut revolts. This already shows one major problem that comes with the approach as described: It is very impractical. The iterative process that was performed to slowly improve the model just takes an incredible amount of time and computation power.

The main reason why the calibration following this method is difficult is the polynomial function. Due to the dependence of the higher order terms on the lower orders is difficult for fit routines to handle. Additionally, many coefficient combinations cause the polynomial to be invalid in the data range. The invalid combinations are those such that the polynomial has an extremum in the data range causing a reversing of the bin borders. In other words, only those coefficients, where the polynomial turns out to be monotic increasing in the data range, are allowed. This splits the parameter search into different regions of the parameter space and a transition between separated but allowed parameter regions is, at least in this model, not possible.

Methods to statistically discover the parameter space additionally suffer from the large number of free parameters. Methods like `emcee` require computational power proportional to the number of parameters. Also, the invalid parameter regions must be excluded for such a search without disturbing the search process.

Despite the difficulties and the additional amount of time that is required to fully converge the model one can see that the overall method is working. The current results are presented below.
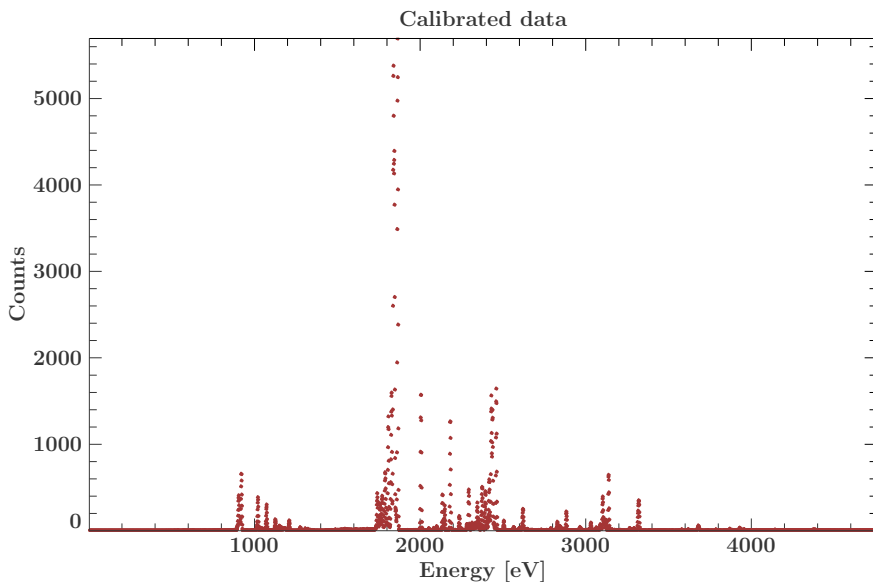
Figure 5.8: Total calibration of all pixels summed up. In general does the calibration method work however it takes a lot of time to fit the data with the presented model.

7.1 Best fit and checks

In Figure 5.8 is the current best overall calibration depicted. From the comparison with the uncalibrated data one can see that the calibration in general works. At least no strong deviation between the uncalibrated data of one pixel and calibrated data occur such that qualitatively the method gives correct results.

A close inspection of the model and the data in the calibrated space reveals that the lines are not correctly reproduced the intensity either indicating that a simple Gaussian as model for the limited resolution does not suffice or the model has other, intrinsic, problems.

In the current state it does not make lots of sense to compare the different calibration methods as there is no quantitative measure for "how close" the model is to the best fit value. However, a simple check was done generating the calibrated data and fitting some of the known line with Gaussian profiles. All lines tested in this way are at least close to their reference values. It is to mention that even if one finds a deviation between the fitted lines of the calibrated data and the reference values it may only hint at a problem of the calibration but it can as well just be an artefact that vanishes once the fit has converged.

The calibration is shown together with its model in detail in Figure 5.7. Besides the already mentioned problems with the reproduced intensity of the lines is the overall fit quite good. The reduced Cash statistic gives a value of $\Delta C_{\mathrm{red.}} = 1.28$ indicating a decent fit. The problems can only be seen by looking at the fit to the raw data. In Figure 5.9 the sum of data and model in detector space is shown. Here one can easily find lines in the model that have no corresponding signal in the data and vice versa.

One can see that the model fits very well in the low energy range and gets worse towards higher energies. The reason for this is also located in the polynomial. The low energy part can be quite simply described by the linear coefficient while the high energy part depends on the higher order coefficients. Also, although there are known lines in the highest energy range (argon line region) only some of them are known. From the data set it was not very clear where the separation between the known and unknown argon
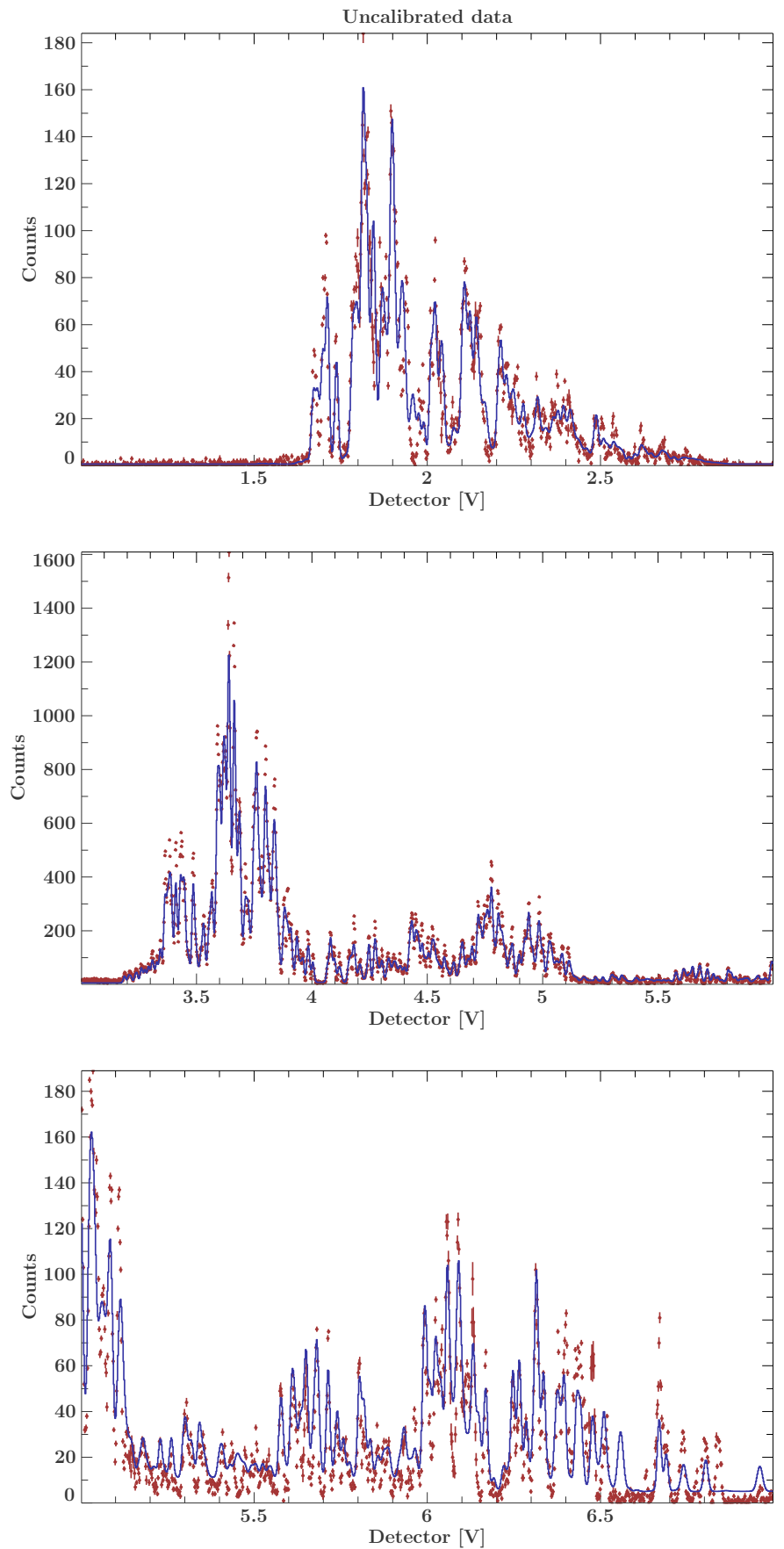
Figure 5.9: Summation of all pixels in detector space together with the current best fit. Although the model fits quite well at lower peak heights, there are still many iteration steps required until the model also gives good results for higher energies.
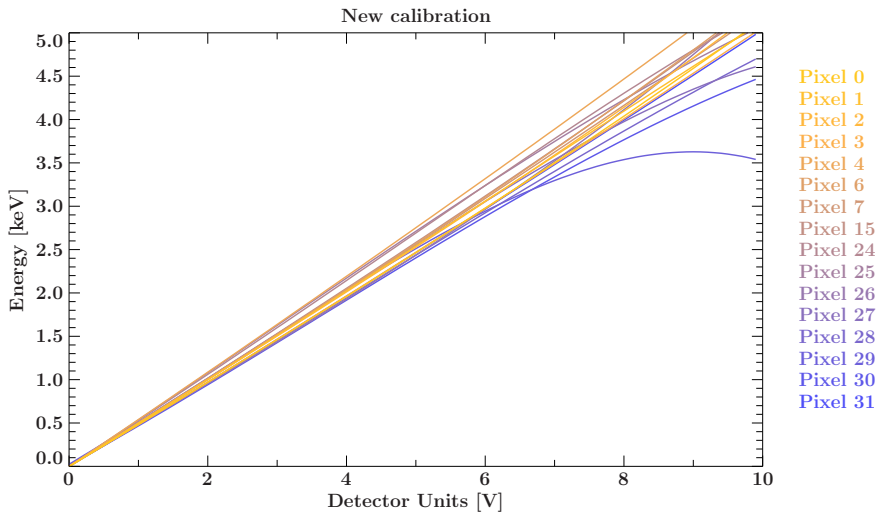
Figure 5.10: Current best fit for the calibration polynomial. Most of the pixels are well described by the model. However, especially pixels 28, 29 and 30 require some more fine tuning.

line measurement was such that all argon lines were treated as unknown[15]. Due to this it is expected, that the calibration in this region is not very good as it is lacking any reference point.

7.2 Model improvements

Does the impracticality of using this method of calibration mean it should not be used at all? The answer to this question is not easy and does depend on the use case. For example, the ECS is often shut down (warm up) and needs a calibration after each re-initialization because the noise environment and therefore also the templates changed. In this case it is just not practical to do a calibration with a time consumption as required here. On the other hand, the issues that are involved with the calibration method presented are also present in the "standard" calibration. Most of the issues are just hidden in the process when calculating the cross correlation.

The time that is required with the new model highly depends on the complexity of the calibration function. Although a polynomial is mathematically quite easy to handle is it challenging to use for finding best fitting functions. A major step in practically can be made if one replaces the calibration polynomial by a function with less free parameters. The resulting polynomials are shown in Figure 5.10. From those one might think of using a spline function where the parameters determine start and end points and the turning point of the slope.

A different approach might be to use a physical model of the detector where the parameters are independent. Such a mode might provide a better target for known fit algorithms. Even if such a model is more expensive to compute, if an algorithm is able to find the fest fitting set of parameters without lots of user interaction it should be preferred over a function where it is the other way around.

Finally, for a more widely applicable calibration procedure there are additional components which might necessary to consider. So far the resolution of the detector was assumed to be constant over the full energy range. In reality this is certainly not the case instead the resolution depends on the

---

[15] This is also the reason why, although lines are visible in the data, the model does not have any line components at all positions
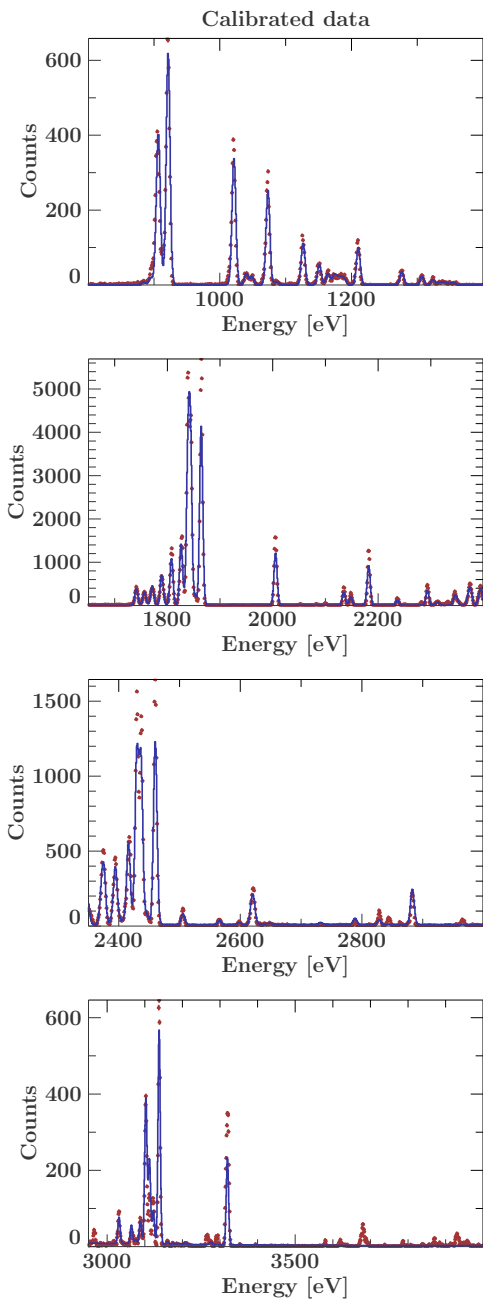
Figure 5.7: Calibrated data in different energy ranges. Blue curve shows calibration model translated to energy grid. The model fits best for the low energy range as this is controlled by the low order coefficients of the polynomial.

energy.

So to conclude, the calibration by fitting a model of the emission lines simultaneous with the detector properties does work. The effort which is required to obtain a reasonable result, however, is in many situations just not in relation to the outcome. This calibration methods should only be used if one can provide a simpler calibration function or the calibration is required to be as accurate as possible, e.g., for space missions.

The true benefit this method provides, that is, calculating the confidence of calibration function, can not be shown here. For this it is required that the fit has converged. Once the function truly is capable of representing each pixel it is possible to do the analysis of the unknown lines in the data. The function then gives direct access to the uncertainty and the uncertainty contributions from the calibration and the line fit.

Line calibration measurements

The observation of spectra reveals insights to the processes involved that are generating the light. In that way the most basic conclusion that can be drawn from a spectrum is if the process is statistical (e.g., black body radiation) or a precise change of the involved system (e.g, line emission/absorption). Once these processes are understood well enough they provide a powerful tool to study astronomical objects. And in most cases, the energy distribution of the light is the only piece of information that is accessible for us.

One of the most basic information that a spectrum with emission lines contains is the motion of the emitting object. As the emitted light undergoes the Doppler shift when the source is moving in the direction of line of sight the emission lines are shifted towards higher or lower wavelengths depending on the object moving towards us or away from us, respectively.

To measure this relative motion it is crucial to know where the emission lines would be observed when the source had a relative velocity of zero. Commonly those emission lines originate from atoms that are part of the object, e.g., gas clouds. The characteristic patterns of those atomic transitions allow a straight forward comparison with measured laboratory data.

For this comparison it is crucial that the reference data, so the data measured in the laboratory, has sufficient accuracy and precision. In this chapter the calibration of new and independent measurements for the neutral oxygen Rydberg series will be presented. The results, resolving a mismatch between astronomical and laboratory data, are presented elsewhere.

1 Gas on the run

One of the many targets that reveal fundamental insights are the emission lines from gas clouds drifting through the galactic disc. The observation of emission lines from those clouds give rise to a better understanding of the structure and evolution of the milky way.

For this reason, observations in many directions were performed with high energetic resolution (Chandra grating observations, **?**). The observed line shifts due to the Doppler effect imposed that in general the gas is gravitationally bound to the galactic disc. However, the observation of neutral oxygen revealed a shift of $\sim 500\,\mathrm{meV}$ averaged over many lines of sight (**?**).

Interpreting this shift as the relative velocity between us and the gas the conclusion would be: Oxygen is moving away from us with a velocity of $\sim 300\,\mathrm{km\,s^{-1}}$! This highly unlikely result was interpreted, at least from astrophysical side, as a hint to inaccurate reference data.

In the following a new calibration method will be described that is completely independent from prior calibrations and gains its accuracy only from first principals.

2 Energy calibration at the BESSY II beam line U49/2

The conceptual idea behind the calibration is remarkably simple: One uses a monochromatic X-ray source that interacts simultaneously with the atoms of interest (here oxygen) and highly ionized atoms of any element.

For practical reasons the ions must have transition lines close to the energetic range of interest. Also, to avoid any bias from prior assumptions, the ionized gas should be hydrogen or helium like.

Now, stepping the energy of the incident X-ray beam provides the information of interest as well as points for calibration. The system for one or

two electrons can be solved analytically which in turn sets the scale of the measurement.

This setup is made possible thanks to the off-axis design of the Polar-X EBIT (Micke et al. 2018) assembled in Heidelberg.

2.1 Coincident readout

The experimental setup is show in Figure 6.1. Here, one can see that the X-ray beam from the beam line is processed by a grating and mirror couple which is used to select a certain energy range. In principle only the grating is necessary to split the incoming beam into near-monochromatic light. However, the mirror allows one to adjust the light path such that the beam focus stays at the same position. From the geometry one can see that this is fulfilled if $\sin(\alpha)/\sin(\beta) = c_{\text{ff}}$ where c_{ff} is constant[1]. This is generally referred to as fixed focus condition.

The, now quasi-monochromatic, photons enter the Polar-X EBIT and eventually interact with the ions produced in the drift tubes. If the energy of the photons is just right they will transfer the ions from ground state to an excited state[2]. When this excited state decays a surplus of photons with the energy corresponding to the energetic difference of the two states is detected. The radiation characteristic depends on the decaying state and the polarization of the photons. Therefore, we measured the photons parallel and perpendicular to the polarization. The emission was measured with SDDs.

Right behind the Polar-X EBIT two gas cells were attached. Separated from the EBIT and from each other by nanometer thin windows. In each of those cells the gases of actual interest were injected. Undergoing the same process of excitation due to the incident X-rays the gas in the cells got ionized. We detected the amount of ionized atoms with electron multipliers which are a direct response to the X-ray photons.

Each event in the gas cells as well as in the EBIT detectors was recorded. However, only events that were coincident with arriving beam line photons are considered as valid. This greatly reduces the background counts in each of the detectors as the re-emitted photons highly exceed the background photons during the arrival of a photon bunch at least at the energy where the emission line is located. Further enhancement is obtained by a threshold cut described below.

2.2 Data reduction and model construction

The recorded data consists of a bunch of useful information which was later used to calibrate the measurements of the gas cells. The first idea for calibrating the data was to use the nominal grid given by the beam line. From the observation of the calibration lines it was known that the nominal energy is shifted by $\sim 3\,\text{eV}$. By reconstructing an energy grid with a linear approximation to the calibration lines gives results already with a very good precision[3]. However, the underlying physics for the experimental setup are well known such that this information gives rise to even better calibration.

It is know that a grating reflects incident light into different angles depending on the wavelength. This behavior is easily described by the

---

[1] For the whole time of the experiment this value was set to $c_{\text{ff}} = 2.25$. By comparing the form of the sine function it is clear that this condition is not true in general but only to first order approximation for small α and β.

[2] Generally any low-level to high-level excitation could happen. Yet, the life time of the excited states is extremely short and therefore this is very unlikely to happen.

[3] "Very good" here means sufficient to resolve the mismatch between astronomical and laboratory observations.
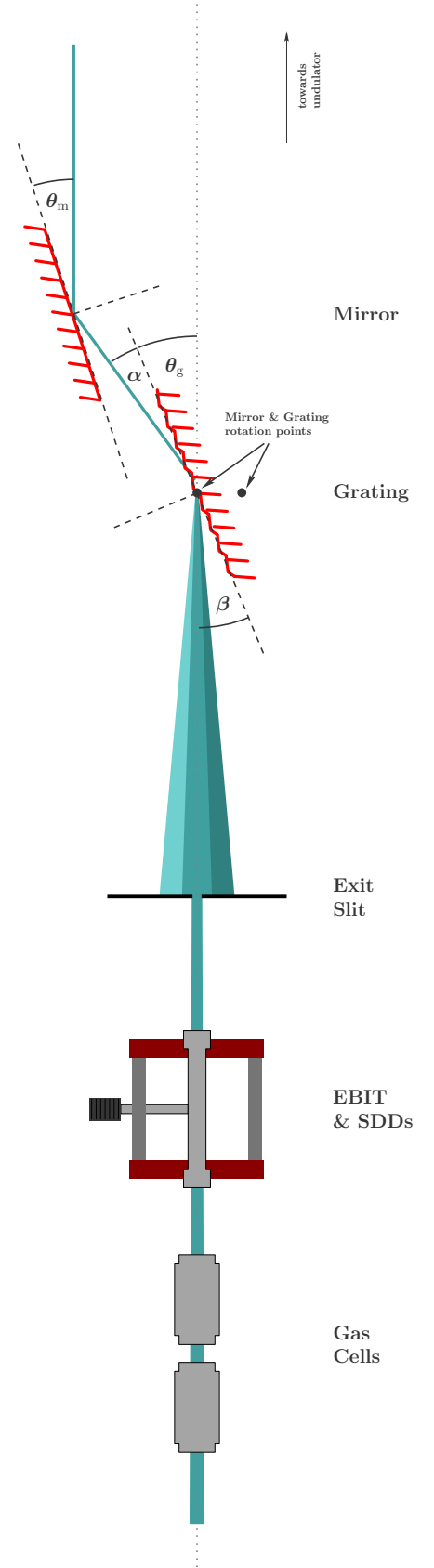

Figure 6.1: Experimental setup at the BESSY II beam line U49/2. Focusing optics are not shown.

grating equation from Section 3. By replacing the wavelength with the corresponding energy and inserting the geometric setup as displayed in 6.1 one obtains

$$E = \frac{mhc}{d} \frac{1}{\cos(\alpha) - \cos(\beta)}. \tag{6.1}$$

Where $\alpha = 2\theta_m - \beta$ and $\beta = -\theta_g$ with the mirror and grating angles θ_m and θ_g respectively, m is the reflection order and d the line separation of the grating.

Each of the variable values was recorded during the experiment. However, the dependent angles α and β as well as the energy were calculated from the measured angles $\theta_i (i = m, g)$. Here it is to note that the recorded angles θ_i did not coincide with the true mirror and grating angles but were shifted by a constant offset.

The calibration function was therefore constructed in the following way: First, we determined the offset in mirror and grating angle by reproducing the grating equation as used by the beam line. Next, we added an additional offset parameter to those angles such that the fixed focus conditions remains fulfilled in first order approximation. As one can easily verify is this shift also a shift of the angle selection along the (mirror, grating) direction in angle angle space. Therefore, the distance in this space stays the same, but may change the energy range.

With this translation of the energy grid one can now model the data by adding line profiles at the theoretical energy. The parameter pair that describes the data does therefore also give the true[4] grating equation which can be used to translate the data of the gas cells.

Unfortunately was the recorded data of the angles affected by significant noise. We convinced ourself that this noise is originating from the readout process and not inherit to the motors that drive mirror or grating. The values for both angles scatter by the same relative amount although they differ by one order of magnitude. Also, the mean of both angles was perfectly fitting the requested positions during one measurement. This made us very confident that the noise really comes from the readout electronics.

To overcome this issue we made use of the fact that for each measurement the steps where chosen equidistant in energy and therefore also in angles (to first order approximation). By this we defined the data to be counts per step which allowed us to compute a linear relation between the steps and the measured angles.

The counted events measured with the SDDs cover a large energy range with low energetic resolution. Therefore, the emission of the resonantly excited transitions with the X-ray beam are barely visible in the SDD spectrum. But the high energy resolution comes not from the detector itself but the grating that illuminates the gas in the EBIT in a narrow energy range.

As the SDDs measure all photon events from the EBIT also photons which are produced from the excitation and de-excitation because of the electron beam are measured. To reduce the number of photons not originating from the X-ray excitement not only were events rejected which did not coincide with the beam bunches but also a low energy threshold was applied. This threshold was determined such that the SNR was maximized. Here, the noise level was estimated by the median of the data.

---

[4] Under the assumption that the theoretical values are the true, i.e., natural values
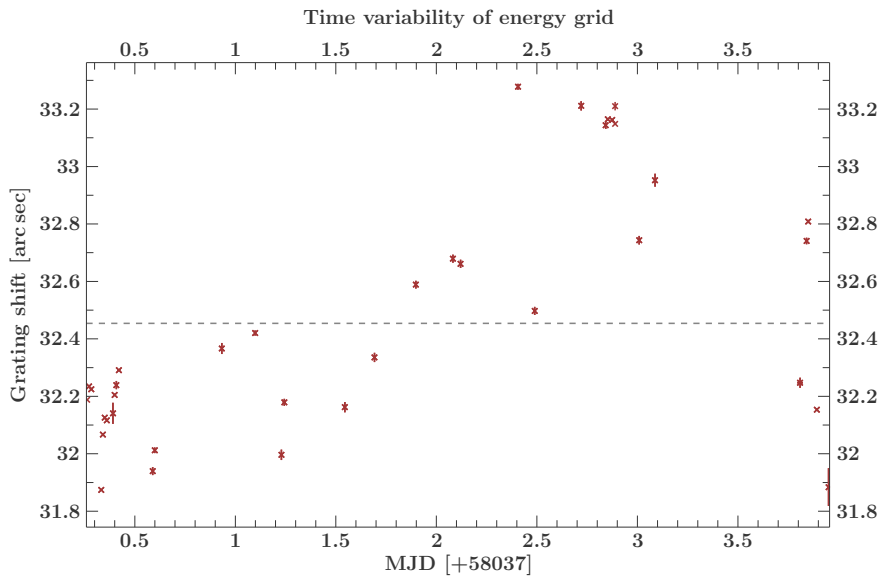
Figure 6.2: The determined shift parameter from various scans during the measurement campaign. The dashed line indicates the mean mean value of the data points. The data used in this figure were taken during experimental setup. It can be clearly seen that the value varies with time most likely due to instabilities of the beam line.

3 Calibration result

The lines were modeled with Voigt profiles with a narrow Lorentz component and multiplied by the exposure time. In that way the normalization of the profiles directly gives the count rate of photons within the line. The fit procedure is essentially the same as laid out in 5 with one important difference. That is, one measurement point obtained by a specific setup of the grating is a direct measurement of the counts per energy contrary to a histogram. This is important as when fitting histogram data the fit function must give the total counts in the respective bin, that is, the integral of the model function over the bin width.

The fit function for the beam line measurements are, however, not integrated. Also, the data has in principle errors in the dependent as well as the independent variables which influences the fit statistic significantly. But as we described the data as counts per measurement step (and also the functions are defined as counts per measurement step) which are per definition exact, we could treat the data simply by evaluating the Cash statistic for data and model.

The best fitting model of the grating equation was found by the simplex method. The nominal value of the grating period is given as $1/d = 602.4\,\mathrm{mm^{-1}}$ leaving the fit function solely dependent on the shift parameter. However, as can be seen in Figure 6.2, the determined shift parameter varies during the measurement campaign.

For a better understanding of this shift we calculate confidence intervals for the fit with the same method as described in 5. This indicates that the energy uncertainty is of the order of $7\,\mathrm{meV}$[5]. However, the variability in time is significantly larger than that. From this we conclude that the uncertainty is highly dominated by systematic errors which stem from the stability of the beam line. A comparison between data sets gathered by the detectors of the gas cells shows significant shifts within parts of the data. Therefore, the variability occurs already on short time scales. The source

---

[5] Corresponding to 68% uncertainty

if this variation could not be determined but one may think of a drifting photon beam hitting slightly different spots of the grating or encoder errors of the step motors driving mirror and grating.

From the short time variations we could conclude that the overall uncertainty does not exceed 30 meV. Hence, the energy calibration that was reached with the presented setup is enough to resolve the oxygen puzzle. The results of the analysis of the oxygen data will be presented in a separate paper.

The conclusion one can draw from this experiment is, that the presented setup is very generic and allows to measure basically every element with the calculated precision, provided one can produce H-, or He-like ions in the EBIT with emission lines in the region of interest. If the X-ray source provides a better stability one may actually reach a precision down to a few meV[6].

With this setup we could show that the next generation of X-ray observatories can be supplemented with accurate and precise data. The original intention for this experiment, resolving the oxygen puzzle, shows the need to revisit measurements of transition lines. This will not only correct the reference list for possible errors but also open the doors to use the full capability of high resolution instruments.

---

[6] The final precision that was reached at the BESSY beam line is still remarkable for an accelerator designed and build 20 years ago. The design was never meant to reach a precision of this level.

Acknowledgments

Bibliography

Balmer, J. J. 1885, Annalen der Physik, 261, 80

Bohr, N. 1913, Philosophical Magazine, 26, 1

Boyle, R. 1675, Experiments, notes, &c., about the mechanical origine or production of divers particular qualities: among which is inserted a discourse of the imperfection of the chymist's doctrine of qualities; together with some reflections upon the hypothesis of alcali and acidum

Boyle, W. S. & Smith, G. E. 1970, Bell System Technical Journal, 49, 587

Bransden, B., Joachain, C., & Plivier, T. 2003, Physics of Atoms and Molecules, Pearson Education (Prentice Hall)

Briand, J. P., Charles, P., Arianer, J., et al. 1984, Phys. Rev. Lett., 52, 617

Brinkman, B. C., Gunsing, T., Kaastra, J. S., et al. 2000, in Proc. of SPIE, Vol. 4012, X-Ray Optics, Instruments, and Missions III, ed. J. E. Truemper & B. Aschenbach, 81–90

Brown, G. V., Beiersdorfer, P., Clementson, J., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7732, Proc. of SPIE, 77324Q

Cash, W. 1976, Astronomy and Astropysics, 52, 307

Cash, W. 1979, Astrophysical Journal, 228, 939

Darwin, C. G. 1931, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 130, 632

De Broglie, Louis. 1925, Ann. Phys., 10, 22

Demtröder, W. 2018, Atoms, Molecules and Photons (Berlin, Heidelberg: Springer Berlin Heidelberg)

den Herder, J. W., Brinkman, A. C., Kahn, S. M., et al. 2001, Astronomy and Astrophysics, 365, L7

Dirac, P. A. M. & Fowler, R. H. 1928, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 117, 610

Einstein, A. 1905, Annalen der Physik, 322, 132

Einstein, A. 1905, Annalen der Physik, 322, 891

Feynman, R. P. 1948, Rev. Mod. Phys., 20, 367

Frauenhofer, J. 1817, Bestimmung des Brechungs- und Farbzerstreuungs-Vermögens verschiedener Glasarten in Bezug auf die Vervollkommnung achromatischer Fernröhre No. Bd. 5 (Bayerische Akademie der Wissenschaften), 193–231

Garmire, G. P., Bautz, M. W., Ford, P. G., Nousek, J. A., & Ricker, Jr., G. R. 2003, in Proc. of SPIE, Vol. 4851, X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy., ed. J. E. Truemper & H. D. Tananbaum, 28–44

Gerlach, W. & Stern, O. 1922, Zeitschrift fur Physik, 9, 349

Gilbert, W. o. C. 1600, On the magnet, magnetick bodies also, and on the great magnet the earth a new physiology, demonstrated by many arguments & experiments (Chiswick)

Goodman, J. & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Green, G. 2008, An Essay on the Application of mathematical Analysis to the theories of Electricity and Magnetism, arXiv:0807.0088

Griffiths, D. J. 2013, Introduction to electrodynamics; 4th ed. (Boston, MA: Pearson), re-published by Cambridge University Press in 2017

Griffiths, D. J. & Schroeter, D. F. 2018, Introduction to Quantum Mechanics, 3rd edn. (Cambridge University Press)

Gu, M. F. 2008, Canadian Journal of Physics, 86, 675

Guarnieri, M. 2016, Proceedings of the IEEE, 104, 467

Heathcote, N. d. V. 1950, Annals of Science, 6, 293

Hell, N. 2017, PhD thesis, Friedrich-Alexander Universität Erlangen-Nürnberg

Herschel, J. W. F. 1788, Transactions of the Royal Society of Edinburgh., 445, 66 v.

Herschel, W. 1800a, Philosophical Transactions of the Royal Society of London Series I, 90, 284

Herschel, W. 1800b, Philosophical Transactions of the Royal Society of London Series I, 90, 255

Hertz, H. 1887, Annalen der Physik, 267, 983

Jackson, J. D. 1999, Classical electrodynamics, 3rd edn. (Wiley)

Keithley, J. F. 1999, The story of electrical and magnetic measurements (IEEE)

Keyser, P. T. 1993, Journal of Near Eastern Studies, 52, 81

Knoll, G. F. 2000, Radiation detection and measurement

Lamb, W. E. & Retherford, R. C. 1947, Phys. Rev., 72, 241

Lampton, M. 1997, Computers in Physics, 11, 110

Levenberg, K. 1944, Quarterly of Applied Mathematics, 2, 164

Lyman, T. 1906, Astrophysical Journal, 23, 181

Lyman, T. 1914, Nature, 93, 241

Magie, W. F. 1935, A source book in physics (Harvard University Press)

Marquardt, D. W. 1963, Journal of the Society for Industrial and Applied Mathematics, 11, 431

Marrs, R. E. 2008, Canadian Journal of Physics, 86, 11

McCammon, D. 2005, Thermal Equilibrium Calorimeters - An Introduction, ed. C. Enss, Vol. 99, 1

McGee, T., Science, I. S. U. M., & Department, E. 1988, Principles and Methods of Temperature Measurement, A Wiley-Interscience publication (Wiley)

Michelson, A. A. & Morley, E. W. 1887, American Journal of Science, 34, 427

Micke, P., Kühn, S., Buchauer, L., et al. 2018, Review of Scientific Instruments, 89, 063109

Mighell, K. J. 1999, Astrophysical Journal, 518, 380

Newton, I. & Innys, W. 1730, Opticks: Or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light, Opticks: Or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light (William Innys at the West-End of St. Paul's.)

Park, B. 1898, A history of electricity (The intellectual rise in electricity) from antiquity to the days of Benjamin Franklin, by Park Benjamin (Wiley)

Paschen, F. 1908, Annalen der Physik, 332, 537

Paschen, F. & Back, E. 1912, Annalen der Physik, 344, 897

Pauli, W. 1926, Zeitschrift für Physik, 36, 336

Pauli, W. 1940, Phys. Rev., 58, 716

Pearson, K. F. 1900, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50, 157

Planck, M. 1914, The theory of heat radiation (P. Blakiston's Son & Co)

Powell, M. J. D. 1964, The Computer Journal, 7, 155

Rowan, T. 1990, PhD thesis, University of Texas at Austin

Rutherford, E. F. 1911, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 21, 669

Schrödinger, E. 1926, Annalen der Physik, 385, 437

Stern, O. 1921, Zeitschrift fur Physik, 7, 249

Strüder, L. 2000, Nuclear Instruments and Methods in Physics Research A, 454, 73

Strüder, L., Briel, U., Dennerl, K., et al. 2001, Astronomy and Astrophysics, 365, L18

Szymkowiak, A. E., Kelley, R. L., Moseley, S. H., & Stahle, C. K. 1993, Journal of Low Temperature Physics, 93, 281

Talbot. 1826, The Edinburgh Journal of Science, 77

Thomson, G. P. & Reid, A. 1927, Nature, 119, 890

Thorn, D. B. 2008, PhD thesis, University of California, Davis

Ullom, J. N. & Bennett, D. A. 2015, Superconductor Science and Technology, 28, 084003

Whittaker, E. T. 1910, A history of the theories of aether and electricity : from the age of Descartes to the close of the nineteenth century (Longmans, Green & Co)

Wilks, S. S. 1938, Ann. Math. Statist., 9, 60

Wollaston, W. H. 1802, Philosophical Transactions of the Royal Society of London, 92, 365

Hiermit bestätige ich, dass die vorliegende Arbeit von mir verfasst und nur unter zuhilfenahme der angegebenen Quellen und Hilfmittle erstellt wurde.

Unterschrift   Datum

Jakob Stierhof