GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

**Master's Thesis**

# Rekonstruktion von Elektronendichten aus Rötgenstreuexperimenten an einzelnen Molekülen bei geringer Photonenzahl

# Reconstruction of electron densities from few photon singular molecule x-ray scattering experiments

prepared by

**Benjamin von Ardenne**

from Dresden

at the Max Planck Institute for Biophysical Chemistry - Computational Biophysics

**Thesis period:** 1st October 2011 until 8th January 2013

**First Referee:** Hon.-Prof. Dr. Helmut Grubmüller

**Second Referee:** Prof. Dr. Tim Salditt

**Thesis Number:** II. Physik-UniGö-Msc-2013/01

# Abstract

X-ray free electron lasers in single molecule experiments hold the promise of solving macromolecular structures. To that aim, reliable structure reconstruction from noisy scattering images with rigorous bounds for the statistical uncertainty is required, which in fact represents one of the main challenges in the field. I present studies of an approach that uses the photon correlation of single x-ray shots to reconstruct the original electron density. Using the spherical harmonics expansion of the Fourier transformation of the electron density, an analytical expression of the two and three photon correlation is derived. With that I demonstrate that the structure can be retrieved from the correlations and discuss possible solution schemes.

# Contents

# 1 Introduction

From the beginning of their discovery in 1895 by Conrad Röntgen, x-rays have been used to determine the structure of materials that surround us. One very important application of x-rays already began in 1912 when Max von Laue discovered that the radiation is diffracted by crystals. This later led to the x-ray crystallography method by William Henry and William Lawrence Bragg in which the positions of atoms in a crystal structure are determined from crystallographic diffraction patterns [1]. The method became particularly useful for large molecules such as proteins. The first of such crystallographic experiments with biological molecules were carried out in 1923 on hexamethylenetetramine [2]. Later x-ray crystallography evolved with the structure determinations of cholesterol, vitamin B12, penicillin and DNA by Dorothy Crowfoot Hodgkin between 1937 and 1954 at Cambridge University [1].

Today, structure determination is important for biological, chemical and biophysical research because the protein's function is strongly correlated to its structure. Learning about the functional behaviors, malfunctions can be understood and specially engineered drugs can be developed.

There is one problem however. More than 18 million proteins are listed in the NCBI Reference Sequence database [3] but only the structures of about 0.05% have been determined to high resolution according to the RCSB protein databank statistics on [4]. The reason is that the purification and crystallization of biological molecules is experimentally challenging. To circumvent the crystallization problem, nuclear magnetic resonance (NMR) determines the structure of the proteins in solution. For that, the resonance of the nuclei spins is measured in an external magnetic field and the strength of that field determines the quality of the signal. The structural information is encoded in the resonance frequencies of the spins because they are shifted by the chemical environment. Interpreting NMR data is challenging because the spectra of large proteins are complex with many overlapping signals and the decomposition of the individual contributions is difficult. Due to the limitations of conventional protein structure determination, new methods are needed.

To this end, scattering experiments with single biomolecules have been proposed as a novel approach to study the structure of proteins. They will become possible because newly developed x-ray sources, such as the free electron laser XFEL in Hamburg, will provide high intense illumination sources. In the experiment, a stream of hydrated proteins enters the x-ray beam at a rate of one molecule per pulse by e.g. applying electrospraying techniques. Due to the high photon flux of the incident beam, a number of photons will be elastically scattered by the electrons of the single protein. In that process, ionization destabilizes the atoms and leads to Coulomb explosion. However, Neutze *et al.* suggest in [5] that if the pulse is short enough, radiation damage can be outrun because then most of the incident photons are scattered by the unperturbed structure before the molecule degenerates.

Two major problems arise, though, which complicate structure reconstruction. First, the orientation of the molecule during a single shot is unknown. Second, the number of photons scattered by one molecule is small and this leads to low signal to noise ratios. Several solutions have been proposed to overcome the two problems.

Ourmazd *et al.* [6] considered a method in 2007 that is known from electron microscope imaging. It utilizes the fact that diffraction patterns are dissections of the molecule's Fourier space and that therefore two patterns from two different shots share a common line that allows to orient the patterns with respect to each other. Unfortunately, the method requires more than 10 photons per detector pixel and picture which is about three orders of magnitude larger than the expected scattering count in the experiment. With classification techniques, one could group the images for fixed angles but it would still require at least $10^9$ scattering images to give a successful reconstruction which is still not feasible in the experiment.

In 2009 Saldin *et al.* [7] discussed an approach in which photon correlations of diffraction patterns are used to reconstruct the shape function of the protein. This binary function only describes the nonzero regions of the electron density and cannot be used to retrieve the atomic positions of the molecule. In a follow up paper, Saldin *et al.* [8, 9] proposed a method to overcome the orientation problem. They suggested to conduct scattering experiments on molecules that only rotate about a fixed axis. This geometrical approach is a feasible method for membrane bound proteins *in situ* with the membrane surface perpendicular to the incident beam. The individual proteins are then assumed to be identical in structure but randomly oriented within the layer. Suitable proteins for this method are e.g. potassium channel proteins which are embedded in the membrane. However, the group of target proteins is

limited and the method does not apply to proteins that are not membrane bound.

A completely different approach based on a probabilistic Bayesian measure has first been introduced by Fung *et al.* in 2008 [10]. With a maximum likelihood measure, scattering patterns are sorted according to the corresponding molecule orientation, averaged and put together to reconstruct the reciprocal space of the protein. Statistical uncertainty is introduced when classifying the scattering images by orientation, and the method depends on the angular discretization of orientation space. Also the number of photons per picture needs to be at least in the order of $10^2$. As an example, they claim to have reconstructed the 500 kD chignolin with a 1.8 Å resolution at $4 \cdot 10^{-2}$ photons per pixel.

In 2010 Walczak and Grubmüller [11] developed a solution for the classification uncertainty that utilizes a seed model of the structure. For a given set of diffraction patterns the posterior probability distribution is calculated that determines the likelihood of a certain molecule orientation for each pattern. This yields the most probable orientation for the scattering images and with that information, the three dimensional Fourier transformation can be reconstructed. The seed model of the structure can be a low resolution version of the protein, obtained from NMR or x-ray crystallography, or a homology model. A variant of the method involves a Monte Carlo search in which the tertiary structure of polypeptides is determined by using the primary and secondary structure as a starting point and modifying bond angles. Very few photons per picture in the range of $N_{photons} \approx 10 - 100$ are necessary to successfully reconstruct viable structure information, but the necessity of a seed model makes it impractical for the study of large protein structures because the search space grows combinatorially.

## 1.1 Project Outline

The previously presented methods, except the model-based Bayesian method by Walczak, require more than 100 photons per scattering image. This is still beyond the number that is expected in the experiment, especially for smaller proteins which scatter less than 70 photons on average (see Section 2.2 for approximations). An experimentally determined photon scattering number cannot be given yet since the next generation free electron lasers are still under construction and very few data could be collected in the high intensity regime to this date.

Thus, the reconstruction problem should be considered in the worst case scenario

in which only very few photons per picture ($N_{photons} \ll 100$ ) are scattered. Additionally the question should be answered, what the minimum photon number is for which the atomic positions can still be determined unambiguously.

Here we aim at answering the question of the minimum photon count and developing such a few photon reconstruction method. The method should further work without *a priori* information of the proteins and it should be robust to noisy data. Walczak [11] has shown that low numbers of photons per scattering image ($N < 10$) still allow for structure retrieval which indicates that the minimum number lies in that region.

Chapter 2 reviews the single molecule scattering setup and how the scattering images are produced. An approximation of the number of expected photons per picture is derived and a review of coherent diffraction theory is given, that explains the analytical connection between the electron density and the diffraction patterns.

Chapter 3 discusses the question of how many photons per shot are needed to successfully retrieve the structure from the scattering images. For that, scattering events with only two photons per picture are analyzed with respect to their information content. The spatial information of the photons on the detector are collected for each picture in terms of a histogram of the sampled two photon correlation.

The chapter also presents a mathematical derivation of the correlations in term of the protein's Fourier transformation. The mathematical description is facilitated by using a spherical harmonics representation of three dimensional structures in which objects are decomposed into shells that are expanded in the basis of spherical harmonic functions. This framework incorporates the spherical nature of the scattering process and facilitates the representation of the symmetry of Fourier space. Since the spherical harmonics expansion is less common in describing objects, an overview will be given to present the major advantages over previous voxel-based descriptions.

The inversion of the two photon correlation will reveal, however, that the reconstruction of the original structure is not unambiguous. Thus, two approaches will be presented that aim at methodically retrieving the missing pieces of information with Monte Carlo search schemes.

In the end, the expression for the three photon correlation is presented and evidence will be given that it contains all essential information for structure reconstruction. It turns out that already a degenerate fraction of the three photon correlation can be used to eliminate the ambiguity of the two photon correlation inversion.

# 2 Single Molecule Scattering

## 2.1 Free Electron Laser

The development of x-ray sources will reach a new level with the completion of the European XFEL in Hamburg. The free electron laser will produce 27 000 flashes per second and reach a brilliance that is $10^9$ times higher than that of the best conventional x-ray radiation sources [12].

There are three important steps involved in creating such highly intense beams. First, a fine and dense electron beam has to be emitted from a high voltage electron gun. For that, an UV laser knocks out electrons of a $Cs_2Te$ cathode surface which is exposed to a high accelerating electrical field of up to 60 MV/m. In the next step, the electrons are brought to high speed in a linear accelerator comprised of resonators. They typically consist of superconducting materials such as niobium and are cooled by superfluid helium at 2 K. In this superconductive state, the resonator carries a strong oscillating microwave that transfers the energy to the electrons in the process. The resonators also focus the electrons and produce a very fine and homogeneous beam.

The emission of the actual x-rays occurs in the undulators which are periodic arrangements of magnets that force the electrons onto a tight slalom course due to the transverse Lorentz force. The sinusoidal path leads to synchrotron radiation in a narrow cone in the forward direction which overlaps with the traveling electrons. In this process, self-amplified stimulated emission (SASE) occurs in which electrons that are out of phase with the radiation get accelerated or decelerated, respectively. As a consequence, the electrons form a longitudinal fine structure which is referred to as micro-bunching. Therefore the grouped electrons emit radiation synchronously and send out high intense coherent x-ray beams whose wavelength depend on the spatial periodicity of the undulators magnets. The expected photon fluence at the XFEL will be as high as $W = 1.8 \cdot 10^{12}$ photon/pulse for a beam diameter $D = 0.1\mu$m with a pulse duration as short as 100 fs [12].

## 2.2 Single Molecule Scattering

In the experiment, a synchronized and hydrated stream of single proteins will be brought into the beam line and probed with a high intense x-ray pulse [13]. The idea to experiment with single molecules was first proposed in 2000 by Neutze *et al.* [5]. New techniques had to be developed to produce the finely bunched proteins streams by using electrospraying or Raleigh-droplet formation [14, 15]. Due to diffusive particle motion in the liquid, the orientation of the individual proteins cannot be controlled and is therefore unknown. Since the beam is short with respect to the motion of the protein, its orientation throughout a single x-ray shot is considered fixed.

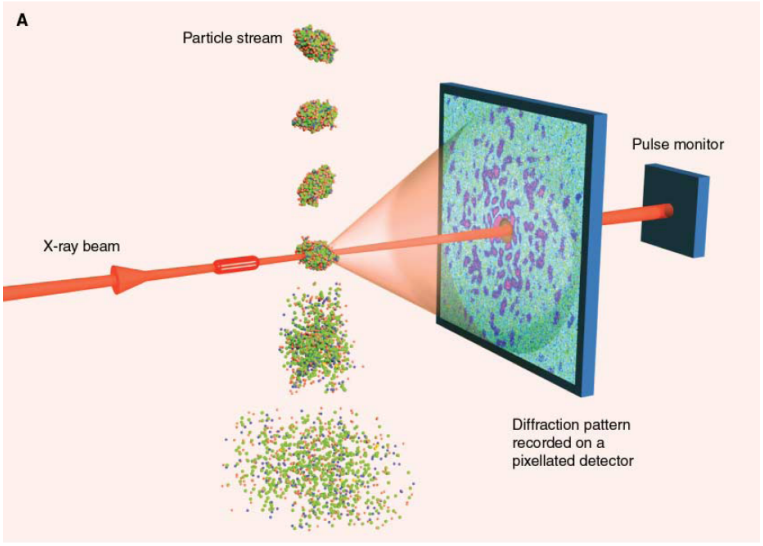



Figure 2.1: Sketch of single-particle coherent diffractive imaging with an XFEL pulse. The droplets with the single proteins are brought into the beam line and exposed to high intense x-rays. The scattered photons are recorded on a pixel detector with $10 \times 10\mu\text{m}^2$ pixel size. (Figure extracted with permission from Ref. [16])

Three types of interactions of the incident x-ray beam ($\lambda = 1 - 5\,\text{Å}$) with the bound electrons of the molecules atoms need to be considered, the photoelectric effect, inelastic scattering, and elastic scattering. Only the latter contributes to the diffraction image while the former constitute noise [17].

The photoelectric effect is a process in which a bound electron of the core shell completely absorbs the energy of an incoming photon and leaves the molecule's atomic shell as an Auger electron due to excitation. In a cascading process electrons

of higher shell fill the vacant position and emit a characteristic recombination radiation. As a consequence, highly charged ions are created which results in a rapid Coulomb explosion of the protein. The time scale of these Auger decay processes is approximately 100 fs and thus in the time span of the beam duration [17, 18]. In the experiment, 90% of the photons are absorbed due to the photoelectric effect and therefore add substantial noise to the recorded data.

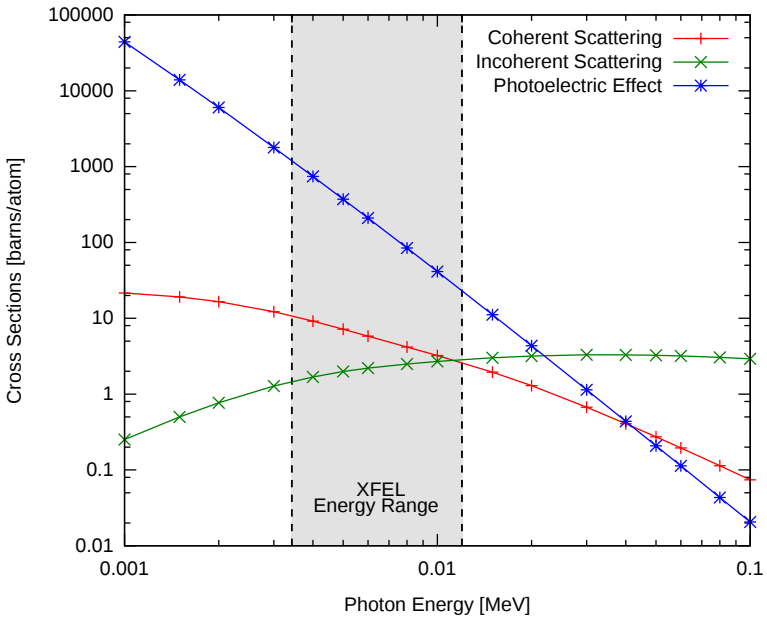



Figure 2.2: Cross sections of coherent/incoherent scattering as well as the photoelectric absorption of carbon atoms as a function of the photon energy. The energy of the XFEL radiation ranges from 2 keV to 12 keV. The data is taken from the XCOM Photon Cross Sections Database maintained by Berger et al. which archives scattering cross sections for all types of atoms and energy ranges [19].

If high energy photons interact with a weakly bound electrons, inelastic scattering may only transfer a part of the photons energy onto the electron, altering the momentum of the photon. In the energy range of the XFEL, about 3% of all the photons are scattered inelastically, loose their phase and cannot be used for structure reconstruction either.

For structure determination, one is interested in elastically scattered photons because they undergo a well understood momenta change (see Section 2.3 for coherent diffraction theory) and therefore sample the non-disturbed protein structure. Because the signal-to-noise ratio determines the quality of the reconstruction process,

we want to estimate the number of elastically scattered photons by a simple model:

$$n = N_{atoms} W \sigma_C \tag{2.1}$$

where $N_{atoms}$ is the number of atoms in the protein, $W$ is the photon fluence of the beam and $\sigma_C$ the elastic photon scattering cross section.

Here the protein model consists of 1000 Carbon atoms (average atom number according to PDB statistics [4]) with the scattering cross section of $\sigma_C = 2.6\,\text{barns/atom} = 2.6 \cdot 10^{-22}\text{mm}^2/atom$. As a result approximately 68 photons are elastically scattered during a single scattering shot.

Because, at wavelengths of $\lambda \approx 1 - 5\,\text{Å}$, the cross section of photoelectric absorption $\sigma_{C,photoelectric} = 23\,\text{barns/atom}$ is about 10 times larger than the elastic scattering cross section (see Figure 2.2) a very low signal to noise ratio of 0.1 is expected. Furthermore, at photon energies over 2 keV, the scattering cross section depends on the scattering angle and for larger angles, corresponding to higher resolution structure information, the number of scattered photons will be even lower. It is therefore crucial to understand the low photon count regime in order to be able to reconstruct the structure from the sparse scattering images. In this work we will first focus on solving the reconstruction problem without the presence of the discussed noise.

## 2.3 Coherent Diffraction

Here we review the connection between the diffraction images and the Fourier transformation of the molecule. Often the terms diffraction and scattering are used interchangeably but while diffraction describes the directional change of the light wave, scattering refers to the momentum change of the photonic particles. Still diffraction patterns are scattering images in the limes of $N_{photons} \to \infty$ and here we will use both terms to refer to the scattering data recorded by the detector.

In the experiment, a coherent planar x-ray wave is diffracted by the electron density of the protein structure. The distance between the object plane and detector is considered to be large in comparison to the size of the molecule and the detector. When regarding elastic scattering, the diffracted beam has the same wave number $k = 2\pi/\lambda$ as the incident beam and in the far field on the detector, the scattered

wave is proportional to the Fourier transformation of the electron density [20]:

$$\begin{aligned} E\left(k_x, k_y\right) &= \int \rho\left(\mathbf{x}'\right) e^{-i\mathbf{k}\cdot\mathbf{x}'} \mathrm{d}\mathbf{x}' \\ &= \mathcal{F}\left(\rho\right). \end{aligned} \tag{2.2}$$

Mapping the detector to Fourier space yields a spherical dissection, called Ewald sphere, which radius is proportional to the inverse wavelength of the photons. Here we consider the limit of a very small wavelengths of $\lambda = 1\,\text{Å}$ such that the cut can be assumed planar.

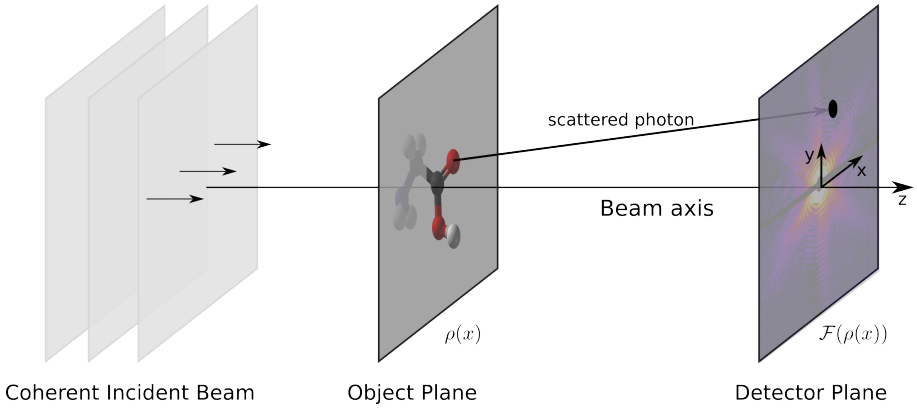



Figure 2.3: Scattering of a single photon by the electron density of the protein in the object plane. The superposition of many scattered photons give the Fourier transformation of the electron density $\rho\left(x\right)$.

On the detector only the absolute square of the Fourier transformation is recorded

$$I \propto \left|\mathcal{F}\left(\rho\right)\right|^2. \tag{2.3}$$

This data will be referred to as intensity in this work. Speaking in terms of photons, the probability of the incidence of a single photon on a particular detector pixel is proportional to the Fourier intensity at a corresponding point on the planar cut in Fourier space.

Since the experiment only yields intensities, the phase of the complex Fourier space is lost which is a common problem that also arises in classical x-ray crystallography. Methods based on Fineup's algorithm [21] have been developed to overcome this lack of the phase information but they are not in the focus of this work and will not be discussed further.

In the experiment, the molecular orientation determines the orientation of the pla-

nar dissection in Fourier space that is observed on the detector. This is because the rotation and the Fourier transform commute, i.e. $R(\mathcal{F}(\rho)) = \mathcal{F}(R(\rho))$. However, the fact that the orientation for a given pattern is unknown implies that we cannot directly relate the photon counts at the pixels with the points in Fourier space. Figure 2.4 illustrates the geometrical connection between the scattering images and the intensity in Fourier space.

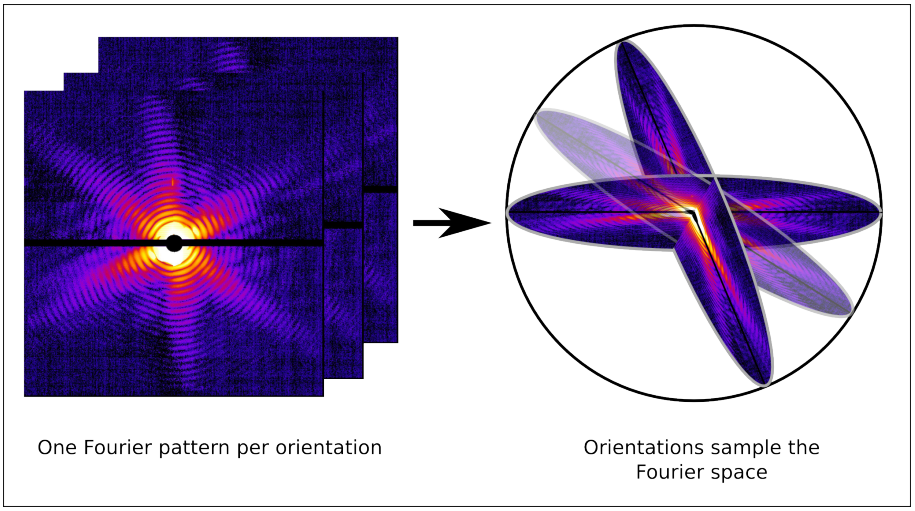



Figure 2.4: If the orientation of molecule would be known for each diffraction pattern, the images could be put together to reconstruct the 3D Fourier transformation of the molecule as sketched in the figure.[1]

Due to the stochastic nature of the experiment, the scattering images are in fact a noisy version of the planar cuts with noise according to a Poisson distribution. This effect is called shot noise and introduces additional uncertainty in the reconstruction. While it is important to keep the effect in mind, this work will only consider scattering images without this noise.

[1]Fourier pattern taken from 'SLAC National Accelerator Laboratory provides library of ultrafast images to speed lasting insights.', `http://science.energy.gov/news/in-focus/2011/12-19-11/`

# 3 Correlation Method

The diffraction patterns are the raw data of the experiment and record the photon counts per pixel for a given shot. The fact that the detector is able to count the photonic hits allows us to decompose each pattern into its single photons. The following chapter focuses on extracting photon correlations from this data and using it for structure reconstruction. It is motivated by the fact that in one dimension function correlations carry all essential information about the originating function.

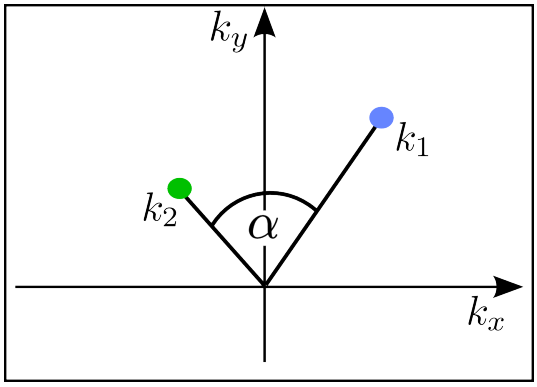



Figure 3.1: A two photon scattering event on the detector screen. One can extract the distances from the center $k_i = \sqrt{k_{x,i}^2 + k_{y,i}^2}$ and the angle between the two photons. The axis of the detector are denoted $k_x$and $k_y$ in accordance with the reciprocal frequencies of Fourier space. At each shot, the detector pattern is proportional to the intensities of a distinct planar cut in Fourier space.

In the most basic scenario, we look at two photon scattering events. As sketched in Figure 3.1, one extracts three parameters, namely the two distances from the beam center and the angle $\alpha$ between the two $\mathbf{k}$-vectors. Scattering images with more photons can always be decomposed into its two photon pairs. By counting the occurrence of two photon events with a particular relationship, one creates a histogram $c\,(k_1, k_2, \alpha)$ from the data that represents the two photon correlation of the structures intensity.

Before the correlation is derived analytically in Section 3.2, we will discuss the most suitable representation framework for the structural data, both in real and Fourier space. The way we describe the data mathematically is important for the analytical description as well as the numerical storage of the data.

A voxel-based representation is usually the first choice when describing and storing complicated three dimensional functions. The space is divided into cubic cells that sit on a regular three dimensional grid and the function values are stored for each cell. The size of the voxels then limit the resolution of the description.

There are two major problems with the cubic representation that makes it impractical for this work. First, the photon information is encoded by the distance of the pixel to the beam center pixel. This distance represents a sphere in three dimensional Fourier space which is difficult to represent in a cubical framework.

Second, the photon counts near the center of the detector will be much higher due to the physics of low and high angle scattering. This effect means that the space closer to the center has a higher information density that requires a different sampling. The cubic framework, however, samples the whole space with even resolution and makes it difficult to oversample the central region without using more sophisticated dynamic cube size systems.

Also, due to the properties of the transformation, the intensity is highly symmetric with $I(\mathbf{k}) = I(-\mathbf{k})$. Consequently the representation should incorporate these symmetries, be spherical by nature and allow for adaptive resolutions both for the whole structure and for particular shells. In light of these requirements, we choose the spherical harmonics framework for structure representation and review it in the following section.

## 3.1 Spherical Harmonics Representation

The spherical harmonic functions, denoted $Y_l^m(\theta, \varphi)$, are a basis of the Hilbert space of square-integrable functions. They allow to expand any function on the unit sphere and will be used to describe the three dimensional electron densities and their Fourier transformations in this work.

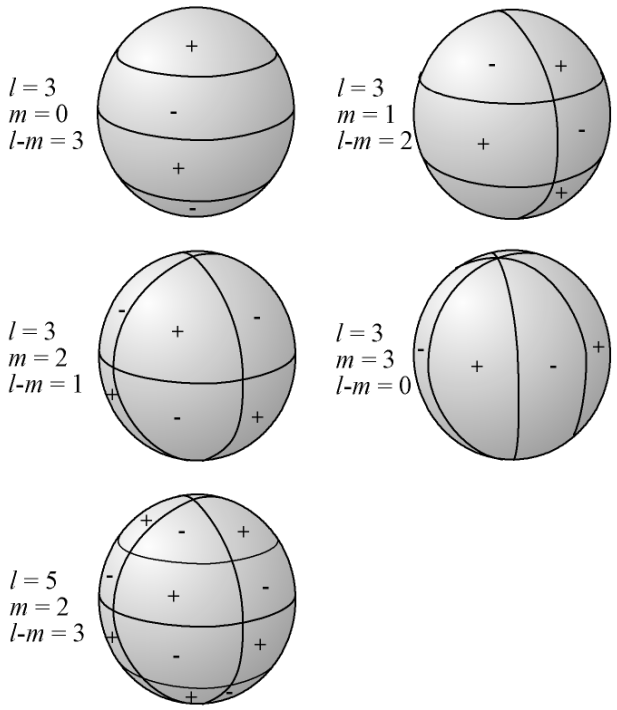



Figure 3.2: Nodal line plot of $Y_l^m(\theta, \varphi)$ on the unit sphere. $\mathrm{Re}\,[Y_l^m]$ is equal to 0 along the black circles. There are $m$ great circles passing through the poles and $l - m$ circles of equal latitude. The function changes sign each time it crosses one of these lines and the number of "patches" increases with $l$ and $m$ [22].

The spherical harmonic functions are comprised of a $\theta$-dependent Legendre polynomial and a $\varphi$-dependent complex exponential part:

$$Y_{lm}(\theta, \varphi) = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}}\, P_\ell^m(\cos\theta)\, e^{im\varphi}. \tag{3.1}$$

These functions form an orthonormal basis of the euclidean space with the orthogonality theorem:

$$\int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} Y_{lm}\, Y_{l'm'}^*\, d\Omega = \delta_{\ell\ell'}\, \delta_{mm'}. \tag{3.2}$$

This allows us to expand any square-integrable function on the unit sphere in terms of spherical harmonic functions:

$$I(\varphi, \theta) \;=\; \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\varphi, \theta) \tag{3.3}$$

with the spherical harmonics coefficients $f_{lm}$ specifying the contributions of the correpsonding harmonic basis function $Y_{lm}$. In Figure 3.2 the first modes of the spherical harmonics are depicted. The so called magnetic number $m$ modulates the function in azimuthal direction whereas the orbital momentum number $l$ varies it in longitudinal direction. With increasing indices $l$ and $m$ the spherical harmonics exhibit higher modes and are therefore used to express finer structures. This fact allows to control the resolution of the description by limiting the orbital momentum number $l$ with a bandlimit $L$.

To describe a three dimensional object with spherical harmonics, one needs to decompose the structure into shells with different radii. Each shell with radius $r$ is then described by a separate set of spherical harmonics (SH) coefficients $f_{lm}(r)$ and the number of shells determine the resolution in radial direction.

The expansion a three dimensional structure with a finite number of spherical harmonics coefficients is expressed as follows:

$$I(\mathbf{r}(r,\varphi,\theta)) = \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm}(r) Y_{lm}(\varphi,\theta) \tag{3.4}$$

where $L$ is the bandlimit, $K$ the number of shells and $f_{lm}(r)$ the coefficients of the equidistant shells at radii $r = k \cdot \Delta r$ with $1 < k < K$. An illustration of the dependency of the resolution on the number of coefficients will be given in Section 3.4.2.

Rotations are going to be important in our usage of spherical harmonics. To describe all possible rotations of a point on a sphere, two angular coordinates are necessary. The angle $\alpha$ describes a rotation about the y-axis which is in plane with the equator, while the angle $\beta$ denotes the rotation about the polar z-axis. With the help of the rotation factors $R^l_{m'm}(\omega)$ [23]

$$\begin{aligned} R^l_{m'm}(\alpha,\beta) &= e^{im\beta} r^l_{m'm}(\alpha) && (3.5)\\ &= e^{im\beta} \int_0^{2\pi} \int_0^{\pi} Y_{lm}(R_y(\alpha)(\theta,\varphi)) \overline{Y_{lm'}(\theta,\varphi)} \sin(\theta)\, \mathrm{d}\theta \mathrm{d}\varphi && (3.6) \end{aligned}$$

the spherical harmonic coefficients can be expressed after an arbitrary rotation as

$$f^*_{lm} = \sum_{m'=-l}^{l} f_{lm} R^l_{m'm}. \tag{3.7}$$

The new coefficients are then a linear combination of the unrotated coefficients for a fixed index $l$.

### 3.1.1 Spherical Bessel Transformation

In the next step, we will review the expression of the Fourier transformation and the intensity in spherical harmonics representation so all calculations can be done in this framework. We start with the spherical harmonics coefficients of the electron density $\rho(\mathbf{r})$ that are given by the integrals with the corresponding basis functions:

$$\rho_{lm}(r) = \int \rho(\mathbf{r}) Y_{lm}^{*}(\omega) \, \mathrm{d}\omega. \tag{3.8}$$

The Fourier transformation of the density $\rho$, $[\mathcal{F}(\rho)](k)$, is important for the experiment but conventional Fast Fourier Transform algorithms work with voxel-based structures only. To avoid back and forth transformation, the spherical Bessel transformation is used to calculate the coefficients $F_{lm}$ describing the Fourier transformation from the coefficients $\rho_{lm}$ describing the electron density [7, 23–26]:

$$F_{lm}(k) = i^{l} \sqrt{\frac{2}{\pi}} \int \rho_{lm}(r) \, j_l(kr) \, r^2 \mathrm{d}r. \tag{3.9}$$

$$\begin{array}{ccc} \rho(x) & \longrightarrow \mathcal{F}(\rho(x)) \longrightarrow & SH(\mathcal{F}(\rho(x))) \\ \| & & \| \\ SH(\rho(x)) & \longrightarrow & SB(SH(\rho(x))) \end{array}$$

Figure 3.3: Connection between real space, the Fourier transformation and the spherical Bessel transformation.

All coefficients for a fixed $lm$-combination, lying on different shells $r$, will be convolved with the spherical Bessel function $j_l$ of order $l$. While the spherical harmonics transformation can be seen as a 2D Fourier transformation on the sphere, the spherical Bessel transformation is the Fourier transformation of the radial part. Figure 3.3 summarizes the connection between Fourier (F), spherical harmonics (SH) and spherical Bessel (SB) transformation.

### 3.1.2 Intensity in Spherical Harmonics

In the next step, we want to express the spherical harmonics coefficients $A_{lm}(k)$ of the intensities $I(k) = |\mathcal{F}(\rho)|^2$ in terms of the coefficients of the complex Fourier transformation $F_{lm}(k)$. This operation involves a more complex convolution of coefficients $F_{lm}$ as shown by Stuhrmann *et. al.* in [27]:

$$\begin{aligned} A_{lm}(k) &= (-1)^m \sum_{l_1 l_2 m_1 m_2} (-1)^{m_1} \sqrt{\frac{(2l_1+1)(2l_2+1)(2l+1)}{4\pi}} \\ &\quad \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l \\ m_1 & \bar{m_2} & \bar{m} \end{pmatrix} F_{l_1 m_1}(k) F^*_{l_2 m_2}(k). \end{aligned} \tag{3.10}$$

The quantities in the parenthesis are scalar numbers called Wiegner-3j symbols that are products of integrals over 3 spherical harmonics basis functions according to the following equation[1]:

$$\begin{aligned} \int Y_{l_1 m_1}(\theta,\varphi) Y_{l_2 m_2}(\theta,\varphi) Y_{l_3 m_3}(\theta,\varphi) \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\varphi &= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \\ &\quad \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \end{aligned}$$

The calculation of a single coefficient $A_{lm}(k)$ involves the summation over many coefficients which makes it computationally expensive. In this case, the simple operation of taking the absolute square of a function becomes more complicated in spherical harmonics representation but the following section will show, that other operations, such as the convolution of functions, are represented much simpler.

Incorporating the symmetry of the molecule's intensity with Friedel's rule $I(\mathbf{k}) = I(-\mathbf{k})$ [7] leads to a reduction of descriptors in spherical harmonics representation. This is easily shown by expressing the rule in the spherical harmonics expansion of the intensity:

$$\begin{aligned} \sum_{lm} A_{lm}(k) Y_{lm}\left(\hat{\mathbf{k}}\right) &= \sum_{lm} A_{lm}(k) Y_{lm}\left(-\hat{\mathbf{k}}\right) \\ &= \sum_{lm} A_{lm}(k) (-1)^l Y_{lm}\left(\hat{\mathbf{k}}\right). \end{aligned} \tag{3.11}$$

[1] `http://en.wikipedia.org/wiki/Wigner_3-j_symbols`

This expression only holds for even $l$ which is why spherical harmonics coefficients for $l = 1, 3, 5, ...$ will be zero when describing symmetric functions. With the presented framework, it is now possible to to describe and calculate the electron density, its Fourier transformation and the intensity exclusively with spherical harmonics coefficients.

## 3.2 Two Photon Correlation

With the spherical harmonics representation of the intensity at hand, we now want to derive the two photon correlation in terms of the intensity coefficients in this section. Then one can study the inversion of this equation and the question will be answered, if the structure can be retrieved from the two photon correlation exclusively.

We start with spherical harmonics expansion of the intensity (see Section 3.1):

$$I(\kappa) = \sum_{lm} A_{lm}(|\kappa|) Y_{lm}(\theta, \varphi) \tag{3.12}$$

in which the coefficients $A_{lm}(|\kappa|)$ describe the structure. As discussed in Section 2.2, the orientation of the molecule during a single scattering shot is unknown. Therefore we introduce an arbitrary 3-dimensional rotation $\omega$ of the expansion:

$$I_\omega(\kappa) = \sum_{lmm'} A_{lm}(|\kappa|) Y_{lm'}(\theta, \varphi) R^l_{m'm}(\omega). \tag{3.13}$$

In this case, the rotation $R^l_{m'm}$ changes the basis functions $Y_{lm}$ and expresses the rotated spherical harmonics as linear superposition of harmonics with the same orbital number $l$. In the experiment, we observe data that is proportional to a planar dissection of the intensity. Thus, without loss of generality, we look at points that lie in the xy-plane ($\theta = \pi/2$ ):

$$\begin{aligned} I_\omega(\mathbf{K}) &= \sum_{lmm'} A_{lm}(|\mathbf{K}|) Y_{lm'}\left(\frac{\pi}{2}, \varphi\right) R^l_{m'm}(\omega) \\ &= I_\omega(k_1, \varphi). \end{aligned} \tag{3.14}$$

Here we choose the xy-plane as the planar cut because it is simple to describe. Later, the averaging over all possible orientations of that plane makes this choice irrelevant. The 2-dimensional vector in this plane is denoted $\mathbf{K}$ to distinguish it from the 3-dimensional vector $\kappa$. Now, the projection is exactly the planar cut that

is sampled during a single molecule scattering shot.

As stated before, the correlation entries are proportional to products of intensities. Thus, with the xy-plane in Fourier space, we can write a two photon correlation entry as the average over all possible orientations of the product of two points in that plane with the respective angle $\alpha$ and distance $k_1$ and $k_2$ to the center:

$$\begin{aligned} c_{k_1,k_2,\alpha} &= \langle I_\omega(\mathbf{K_1}) \cdot I^*_\omega(\mathbf{K_2}) \rangle_\omega & (3.15)\\ &= \Bigg\langle \sum_{l_1 m_1 m_1'} A_{l_1 m_1}(k_1)\, Y_{l_1 m_1'}\left(\frac{\pi}{2}, 0\right) R^{l_1}_{m_1' m_1}(\omega) & (3.16)\\ &\qquad \sum_{l_2 m_2 m_2'} A^*_{l_2 m_2}(k_2)\, Y^*_{l_2 m_2'}\left(\frac{\pi}{2}, \alpha\right) R^{l_2 *}_{m_2' m_2}(\omega) \Bigg\rangle_\omega . \end{aligned}$$

In other words, for a given $(k_1, k_2, \alpha)$-characterization, the above equation sums up the probabilities of all possible two photon pairs with this particular property by going over all molecule orientations.

Figure 3.4 illustrates the averaging process of these pairs and shows how the 2-dimensional projection of the Fourier transformation is connected with the 3-dimensional structure. After rewriting Equation 3.16 to:

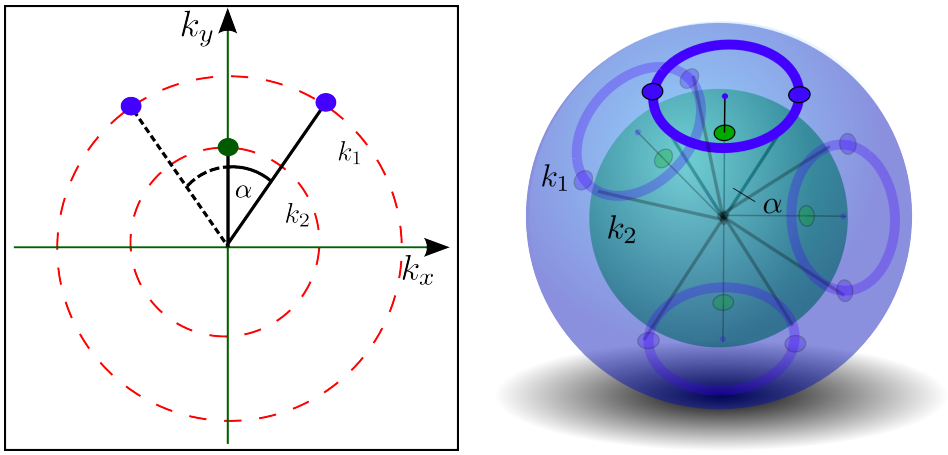
$$\begin{aligned} c_{k_1,k_2,\alpha} &= \sum_{l_1 m_1 m_1'} \sum_{l_2 m_2 m_2'} A_{l_1 m_1}(k_1)\, Y_{l_1 m_1'}\left(\frac{\pi}{2}, 0\right)(\omega)\, A^*_{l_2 m_2}(k_2)\, Y^*_{l_2 m_2'}\left(\frac{\pi}{2}, \alpha\right) \\ &\qquad \left\langle R^{l_1}_{m_1' m_1}(\omega)\, R^{l_2 *}_{m_2' m_2}(\omega) \right\rangle_\omega & (3.17) \end{aligned}$$

one sees that the averaging only concerns the rotation factors. They are defined via integrals over spherical harmonic functions and a distinct orthogonality relation for rotation averages exists, that facilitates the equation above:

$$\left\langle R^{l_1}_{m_1' m_1}(\omega)\, R^{l_2 *}_{m_2' m_2}(\omega) \right\rangle_\omega = \frac{1}{2l+1} \delta_{l_1 l_2} \delta_{m_1 m_2} \delta_{m_1' m_2'} . \quad (3.18)$$

This relation eliminates the averaging term in the correlation and is one of the reasons why the spherical harmonics representation is so powerful in expressing the two photon correlation. After inserting Equation 3.18 into 3.17, a much simplified version of the two photon correlation is obtained:

$$c_{k_1,k_2,\alpha} = \sum_{l m m'} A_{lm}(k_1)\, Y_{lm'}\left(\frac{\pi}{2}, 0\right)(\omega)\, A^*_{lm}(k_2)\, Y^*_{lm'}\left(\frac{\pi}{2}, \alpha\right) . \quad (3.19)$$




(a) Detector screen with a two photon event, its mirrored counterpart and the two reciprocal shells $k_1$ and $k_2$ depicted in dashed red lines. Rotating the molecule around an axis that rotates the Fourier projection around the center of the detector produces photon pairs originating from different Fourier points but that have the same $(k_1, k_2, \alpha)$ characterization as the others.

(b) The solid points denote the same two photon event $(k_1, k_2, \alpha)$ as in Figure (a) but depicted in the 3-dimensional Fourier space with the two exemplary $k_1$ (blue) and $k_2$ (green) shells. The circle illustraltes that upon rotation about the axis parallel to the $\mathbf{k}_2$ intensity vector other two photon pairs are sampled that still have the same $\alpha$-angle. The opaque copies of these two photon scenarios show that an additional moving of the points on shell $k_2$ is necessary to gather all possible combinations for a single $(k_1, k_2, \alpha)$ slot.

Figure 3.4: Averaging over all possible orientations of two photon pairs on shell $k_1$ and $k_2$ that have an $\alpha$ distance.

It is further simplified by using the fact that a summation of products of spherical harmonics over the magnetic number $m'$ facilitates to [28]

$$\sum_{m'=-l}^{l} Y_{lm'}\left(\frac{\pi}{2}, 0\right)(\omega) \cdot Y^*_{lm'}\left(\frac{\pi}{2}, \alpha\right)_{m'} = \frac{2l+1}{4\pi} P_l\left(\cos\left(\alpha\right)\right), \tag{3.20}$$

which gives to a compact expression for the two photon correlation that is comprised only of a sum of products of spherical harmonics coefficients of the intensity:

$$c_{k_1,k_2,\alpha} = \sum_l P_l\left(\cos\left(\alpha\right)\right) \sum_m A_{lm}\left(k_1\right)(\omega)\, A^*_{lm}\left(k_2\right). \tag{3.21}$$

This equation now allows us to analyze the information content of the correlation by looking at the inversion.

## 3.3 Inversion Problem

In this section, the inversion of the two photon correlation from Equation 3.21 is discussed. A successful inversion would allow to retrieve the coefficients of the intensity from the two photon correlation alone.

As a start, we rewrite the two photon correlation in vector and matrix form. For that, we define $\mathbf{a}_l(k)$:

$$\mathbf{a}_l(k) = \begin{pmatrix} A_{lm}(k) \\ A_{l(m-1)} \\ \dots \quad (k) \\ \dots \\ A_{l-m}(k) \end{pmatrix} \tag{3.22}$$

as the $(2l+1)$-dimensional coefficient vector with all the spherical harmonics coefficients $A_{lm}$ of fixed $l$ and $\{m \in \mathbb{Z} | -l < m < l\}$. Then we can write the correlation partly in vector form:

$$c_{k_1,k_2,\alpha} = \sum_l P_l(\cos(\alpha)) \underbrace{\mathbf{a_l}(k_1)\,\mathbf{a_l^*}(k_2)}_{A_l(k_1,k_2)} \tag{3.23}$$

by expressing the sum over $m$ as a scalar product of coefficient vectors. If we further use the vectors with correlation entries $\mathbf{c}_{k_1,k_2} = \begin{pmatrix} c_{k_1,k_2,\alpha_1} & \dots & c_{k_1,k_2,\alpha_N} \end{pmatrix}$ and scalar product entries $\mathbf{a}(k_1,k_2) = \begin{pmatrix} A_0(k_1,k_2) & \dots & A_l(k_1,k_2) \end{pmatrix}$as well as the matrix

$$\mathbf{P} = \begin{pmatrix} P_0(\cos(\alpha_1)) & \dots & P_0(\cos(\alpha_N)) \\ \dots & \dots & \dots \\ P_{l_{max}}(\cos(\alpha_1)) & \dots & P_{l_{max}}(\cos(\alpha_N)) \end{pmatrix}, \tag{3.24}$$

we can use the correlation to retrieve the scalar products $A_l(k_1,k_2)$ by solving the following linear problem:

$$\mathbf{c}_{k_1,k_2} = \mathbf{P}\mathbf{a}(k_1,k_2). \tag{3.25}$$

In Equation 3.25 the vector $\mathbf{c}$ is known from the experiment and the matrix $\mathbf{P}$ is fixed. For structure reconstruction we are interested in the individual $A_{lm}(k)$ because those coefficients carry the structure information of the molecule but the vector $\mathbf{a}$, that is the solution of the linear equation above, only contains scalar products. Still, for a given orbital momentum number $l$ and the case that we have

sampled enough shells, namely $k_{max} \geqq 2l + 1$, there are enough scalar products $A_l(k_i, k_j)$ are present to express the *Gram* matrix [29]

$$\mathbf{G}_l = \mathbf{F}_l^T \mathbf{F}_l \tag{3.26}$$

$$= \begin{pmatrix} A_l(k_1, k_1) & \dots & A_l(k_1, k_{k_{max}}) \\ \dots & \dots & \dots \\ A_l(k_{k_{max}}, k_1) & \dots & A_l(k_{k_{max}}, k_{k_{max}}) \end{pmatrix}. \tag{3.27}$$

It is comprised of the known scalar products between the coefficients vectors belonging to a shell $k$ and an index $l$ gathered from the solutions from Equation 3.25. The matrix $\mathbf{F}_l$ with

$$\mathbf{F}_l = \begin{pmatrix} \mathbf{a}_l(k_1) & \mathbf{a}_l(k_2) & \dots & \mathbf{a}_l(k_{k_{max}}) \end{pmatrix} \tag{3.28}$$

contains the original spherical harmonics coefficients for a fixed $l$. So we want to solve this system for the matrix $\mathbf{F}_l$ which can be done by diagonalizing $\mathbf{G}_l$

$$\mathbf{L}_l = \mathbf{Y_l} \mathbf{G_l} \mathbf{Y_l^T}. \tag{3.29}$$

Solving this eigenvalue problem gives

$$\mathbf{F}_l = \sqrt{\mathbf{L}_l} \mathbf{Y}_l \tag{3.30}$$

with the diagonal matrix $\mathbf{L}_l$ and the transformation matrix $\mathbf{Y}_l$. The solution matrix $\mathbf{F}_l$ contains a valid set of spherical harmonics coefficients that represent an intensity structure that produces the originating two photon correlation. However, this is an eigenvalue problem and the solution is only obtained up to an arbitrary rotation in the $2l+1$-dimensional eigenspace. There exists an infinite set of additional solutions that also produces equivalent correlations and is therefore valid. This degeneracy originates from the fact that we only have scalar products between all the coefficient vector that only give us the relative position of the vector in terms of the respective angles. Unfortunately, the scalar products carry no information about the overall orientation of the connected set of coefficient vectors but the appearance of the underlying structure depends on the correct rotation in that space. Thus, we now have only found the correct spherical harmonics coefficients up to an unknown rotation $\mathbf{U}_l$ for each $l$ separately:

$$\mathbf{A}_l(k) = \mathbf{U}_l \mathbf{A}_l^0(k). \tag{3.31}$$

The fact that the arbitrary rotation can not be retrieved from the two photon correlation itself means that it is not possible to retrieve all the structural information from such a two photon correlation.

The degeneracy of the reconstruction is not directly connected to the arbitrary rotation of the molecule during a scattering shot, as one might think. Here we have one unknown $2l+1$-dimensional rotation for every $l$ which leads to many more unknown than the 3 missing Euler rotations of the molecule.

The main problem with the two photon correlation in spherical harmonics representation is that the scalar products only concern coefficients with equal momentum numbers $l$. Thus, the problem is decomposed into sub-problems for each $l$ and coefficients for different $l$ cannot be related. In other words, the rotation of the different $2l+1$-dimensional spaces cannot be correctly aligned relative to each other because no information is available in the correlation that interconnects these spaces.

A different approach to explain the information deficiency of the the two photon correlation using mode counting was given by Elser 2011 [30]. He brought the correlation expression in the form

$$c\left(k_1, k_2, l\right) = N_{k_1 k_2 l} \sum_{m=-l}^{l} A_{klm} A_{klm}^{*} \tag{3.32}$$

in which the $\alpha$-dependency in $P_l\left(\cos\left(\alpha\right)\right)$ was integrated out due to orthogonality. The resulting independent constraint equations are counted as

$$E = \sum_{l}^{L} \sum_{k_1=1}^{K(l)} \sum_{k_2=k_1}^{K(l)} 1 \tag{3.33}$$

with upper limits $L$ and $K\left(l\right)$ determined by mode counting. Comparing with the number of free variables in intensity space, $V = \left(\pi/6\right)\tilde{R}^3$, with $\tilde{R}$ being the dimensions of the box the covers the particle, one sees that the reconstruction problem is underdetermined: $E/V \approx 0.698$.

After all, it seems additional angular information needs to be extracted from the experiment in order to unambiguously reconstruct the structure analytically. Hence, Section 3.6 will continue with the analysis of the three photon correlation. Beforehand, the question will be addressed if the remaining unknown orientations can be retrieved by a method that poses additional constraints on the search in the solutions and that exploits additional knowledge of the protein systems such as positivity of the electron density and the known support. For that, a numerical implementa-

tion of the discussed processes was programmed and will be briefly discussed in the following section.

## 3.4 Numerical Evaluation

In the next section, the computational tools are presented that were used to numerically implement the mathematics presented in the previous sections. Then, an evaluation of the resolution capabilities of spherical harmonics descriptions was carried out and is discussed in Section 3.4.2. With the help of the two photon correlation form found in Equation 3.21, a brief comparison between an analytically calculated correlations and Monte Carlo generated one will also be shown. In all evaluations, a Glutathione molecule as well as some smaller atom aggregations were used as simple test systems.

### 3.4.1 Tools

The underlying code was written in the *Lua* scripting language[2] which was used with a genius Just-In-Time compiler *LuaJIT*[3] that generates system-dependent bytecode before running the instructions. On top of that, the *GSL-Shell extension*[4] offered fast access to the mathematical library *GSL (GNU Science Library)*[5] from within Lua [31]. GSL contains advanced functions for random numbers (taus2 a maximally equidistributed combined Tausworthe generator by L'Ecuyer was used) and spatial functions such as Legendre Polynomials, Spherical Bessel functions and Wiegner-3j symbols which are necessary to carry out the introduced calculations. Spherical harmonics transformations of a given structure can be computationally expensive. Highly optimized algorithms have been developed and implemented for that purpose and can be reviewed in [32–34]. A well tested implementation is the *S2Kit Lite*[6] which was used in this project for fast back-and-forth transformation in spherical harmonics space. The library is implemented in Standard C and was easily imported via the *FFI library* which allows calling external C functions and using C data structures from native Lua code.

[2] `http://www.lua.org/about.html`
[3] `http://www.luajit.org`
[4] `http://www.nongnu.org/gsl-shell/`
[5] `http://www.gnu.org/software/gsl/`
[6] `http://www.cs.dartmouth.edu/~geelong/sphere/`

### 3.4.2 Spherical Harmonics Resolution

The resolution of a spherical harmonics description is influenced by two different factors. First, the number of coefficients that is used to describe the structure and second, the accuracy with which the single coefficients are determined experimentally.

In the experiment, the bandlimit $L$ is determined by the number of shells $K$ the experiment can resolve (as discussed in Section 3.3). Here, we first want to illustrate the influence of the number of coefficients on the quality of the description. For that, we look at a simple tripeptide called Glutathione (GSH) and vary the bandlimit $L$ and number of shells $K$.

For a given set of spherical harmonics coefficients with $l < L$ and $k < K$, there are

$$K \sum_{l=0}^{L} (2l+1) = K\,(L+1)^2 \tag{3.34}$$

different complex numbers describing the structure. Glutathione consists of 20 atoms which means that we have 60 unknown parameters in the three dimensional position space. Using the same number of coefficients would equal an upper limit for the coefficients of $K = L \approx 4$. However, the spherical harmonics basis functions are not a suitable descriptor for atomic positions in space and therefore we will need more than 60 coefficients.

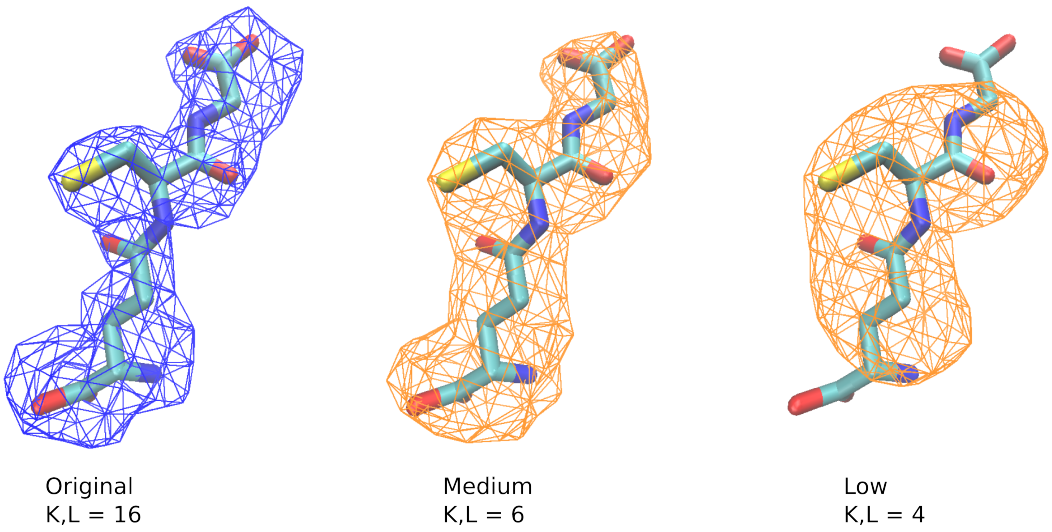



Figure 3.5: Comparison of the resolution of the structure description with $L, K = 6$ and $L, K = 4$ for Glutathione sampled with high resolution.

To evaluate the dependency on the coefficient number, the Glutathione tripeptide

was sampled at high resolution ($K, L = 16 \equiv 4624$ coefficients). In Figure 3.5 the original electron density is plotted on the left side as an isosurface plot.

Next, the spherical Bessel transformation was carried out to obtain the coefficients describing the peptide's Fourier transformation. In this coefficient set, higher order terms above $L, K$ were removed and the coefficients were back transformed. This operation is equivalent to deleting Fourier data above a certain radii and reducing the resolution of the remaining spherical data, just like a low pass filter.

For a cutoff at $K, L = 4$, which equals 100 coefficients, we clearly see that the rod like shape of the protein cannot be sufficiently described and the resulting structure is more dense in the center. The angular resolution as well as the radial resolution of the structure is compromised, as one would expect. However, quite remarkably with $K, L = 6$, i.e. 294 coefficients, the overall structure of the protein is reconstructed well with most losses in the outer regions of the structure and the thickness of the strand. This visual evaluation supports, that already for a limited amount of coefficients, one can gain valuable information about the tertiary structure such as the approximate position or bending of the backbone strands. Using the same number of descriptors in a cubic framework equals an approx. $6x6x6$ cube which would not represent the structure with the same quality of resolution as the spherical harmonics description.

The second important influence on the resolution of a structure is the accuracy of the coefficient determination. Photons scattered by small angles which correspond to low $k$ will occur more likely in the experiment, as discussed in Section 2.2. Thus, the experiment will collect more photons in that region and reconstruct coefficients for low $k$ with higher statistical accuracy due to better sampling.

As a consequence, we will have a higher certainty about low resolution description of the data in analogy to classical x-ray crystallography where higher resolved structure information can only be obtained from intensities farther away from the detector origin where naturally the signal-to-noise ratio is worse.

### 3.4.3 Monte Carlo Method

To simulate the experiment, two points on the sphere were generated with an even distribution. This was done by choosing a random polar angle $\varphi \in [0, 2\pi]$ and another random number $x \in [-1, 1]$ that converts to the azimuthal angle $\theta = \operatorname{acos}(x)$ and ensures that that the equator region is as densely sampled as the poles. Then the product of the intensity at the two points was sorted into the corresponding

$(k_1, k_2, \alpha)$ entry of the histogram and afterwards the data was normalized by the number of sampled pairs per slot.

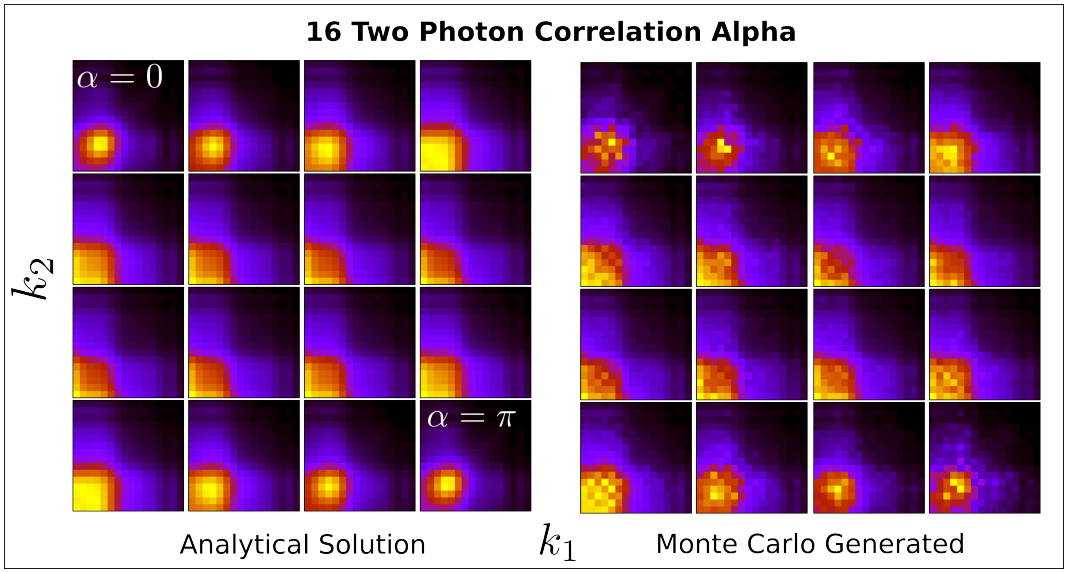



Figure 3.6: Two photon correlation of Glutathione analytically calculated and Monte Carlo generated ( $N_{pairs} = 10^4$ samples per $(k_1, k_2)$pair).

On the left side of Figure 3.6, the two photon correlation calculated from Equation 3.21 is shown, while the right part shows the same correlation generated by the Monte Carlo scheme with $N_{pairs} = 10^4$ two photon pairs for each $(k_1, k_2)$-combination. Both correlations were generated from a structure with a limit on the number of coefficients of $K, L = 16$ and an angular discretization $N_\alpha = 16$. Each frame in the graph shows a single $\alpha$-slice of the correlation.

The two correlation datasets exhibit the same fundamental features. In analogy to Fourier space, most of the intensity is close to the origin which is why each frame shows significant peaks in the lower $k$ region. The varying shapes of these peaks encode parts of the Fourier data but with a lack of information as shown before. For high $N_{pairs} > 10^5$ both correlations became visibly indistinguishable for the structure of Glutathion which is why they are not shown here. Figure 3.6 depicts the Monte Carlo generated correlation with a low $N_{pairs}$ to see a sparsity and to emulate the low photon count of the experiment. The value $10^4$ is chosen such that an effect due to sparse sampling becomes visible without making the correlation unrecognizable. In the experiment, the number of photon pairs $N_{pairs}$ for a $(k_1, k_2)$-combination is determined by the number of shots and the number of photons per shot. In comparison to other methods, this method can compensate low photon

numbers per picture with many pictures.

A $k$-dependent scattering intensity is not included in the Monte Carlo model. Taking this effect into account would lower the signal in the upper regions of $k_1$ and $k_2$. Both shot noise and background noise have not been considered in these comparisons.

## 3.5 Solution Schemes

This section addresses the question, if the missing information of the two photon correlation, namely the orientations in the $2l+1$-dimensional coefficient spaces, can be retrieved with additional constraints or adequate iteration/Monte Carlo schemes. Before the two different approaches will be discussed, a suitable metric will be presented to evaluate the difference between two correlations.

### 3.5.1 Comparing Correlations

Suppose an experiment produces a correlation $c_{exp}$ for a given structure. With a model structure at hand, one could calculate the two photon correlation $c_{mod}$ of that model. When comparing the experimental with the model correlation, a suitable metric is needed to measure how similar the two correlations are. The hypothesis is that a similarity in correlations also means a similarity in structure.

The most basic idea is to sum up the differences of the correlation entries:

$$\Delta = \sum |c_{i,mod} - c_{i,exp}| \,. \tag{3.35}$$

This does not consider, that some regions in the correlation (such as the high $k$ regions) generally have smaller intensities but are equally important in the comparison. One could compensate for that with a $k$-dependent normalization factor but here we introduce a completely different metric based on probabilities which is more intuitive[35]. Consider the probability $p$ that a set of diffraction patterns with many two photon pairs give rise to a certain reference correlation $c_{ref}$.

$$p = \prod_{j(pairs\,in\,exp.)} c_{mod}\left(k_{1,j}, k_{2,j}, \alpha_j\right) \tag{3.36}$$

$$\ln p = \sum_{j(pairs\,in\,exp.)} \ln c_{mod}\left(k_{1,j}, k_{2,j}, \alpha_j\right) \tag{3.37}$$

Now suppose in our set of pairs, we have multiple two photon events that belong to one $k_1, k_2, \alpha$ characterization. The number of those pairs is proportional to the $c_{exp}(k_1, k_2, \alpha)$ value. With that factor, we can let the sum run over all correlation entries:

$$\ln p = \sum_{\{\mathbf{r}_i\}} c_{exp}(\mathbf{r}_i) \cdot \ln c_{mod}(\mathbf{r}_i) \tag{3.38}$$

to get the log of the probability $p$ in terms of the model and experimental data. This value gives a measure for how close the two correlations are and it reaches maximum when the correlations $c_{exp} = c_{mod}$ are identical.

### 3.5.2 Interatomic Distance Variations

In structure determination experiments, the sequence of the proteins is always known. With that information one can use a model in which the inter-atomic distances are the unknowns instead of the coefficients describing the Fourier intensity. This saves the work for phase retrieval and the fitting of the atom positions in the back transformed electron density. Working with atom distances also lowers the space of unknowns drastically in comparison to spherical harmonics descriptions. Each set of atomic positions maps to a two photon correlation and the correct position set yields the reference correlation $c_{ref}$. The question remains how distinct the atomic position are in terms of their correlation difference in this interatomic distances space.

The Monte Carlo search in atom position space looked as follows:

1. Start by obtaining the experimental correlation $c_{ref}$ either directly in the experiment or with Equation 3.21 for simulation purposes.

2. Initialize the atomic positions of the model with an educated guess according to typical interatomic distances or purely random with overlap restrictions.

3. Repeat:

   a) Modify atomic positions by moving each atom into a random direction with Gaussian distributed displacements $\mathbf{\Delta d}_{gauss}$. Avoid drifting by eliminating the center of mass motion of the system.

   b) Calculate the resulting model correlation $c_{mod}$ and the corresponding difference to the experimental results $\ln p$ . If the proposed step increases the

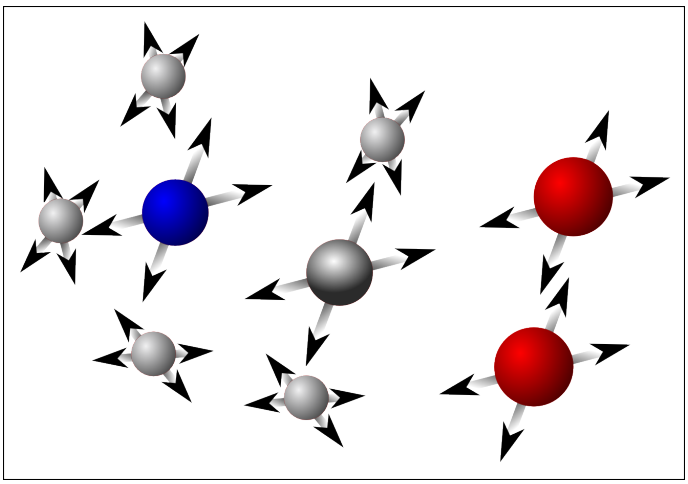


Figure 3.7: Interatomic distance variation scheme in which at each step, the positions of all atoms were varied with random direction and Gaussian distributed displacements.

similarity between the correlations, meaning $\Delta = \ln p - (\ln p)_{previous} < 0$, accept the step. If $rand\,() \in [0,1] < \exp(-\Delta/T)$ also accept the step with appropriate temperature $T$ to avoid being stuck in local minimum.

c) Decrease temperature after each accepted step for a "simulated annealing" method.

d) Adjust displacement $\mathbf{\Delta d}_{gauss}$ to keep acceptance rate in the range of 20 - 50 % .

For a two and three atom system, this approach correctly determines the interatomic distances with correlation as the correct measure. For more than 3 atoms however, multiple local minima of the measure ln$p$ become visible in the space of distances. Repeating the simulation with different initial atomic coordinates and randomly permuting atoms in the simulation does not yield convergence to the known solution.

Due to the variety of minima, the time to find the global minima with a slow temperature decrease seems to become extremely long with just as few as 5 atoms in the system ($N\,(N+1)\,/2 = 15$ unknown distances). For such a system the simulations never found the correct set of interatomic distances. However, correlations are found that belong to a different set of distances and still coincides with the original correlation up to small fluctuations. For this solution scheme, the correlation does not seem to hold enough detailed information to distinguish between these fine distances in the range of Å*ngström*.

### 3.5.3 Coefficients Monte Carlo

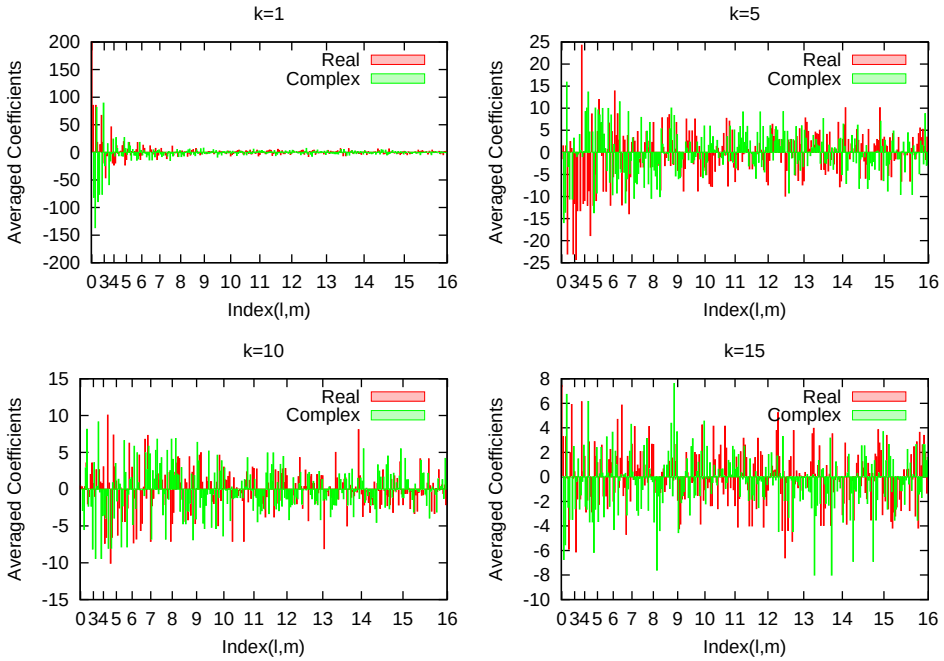



Figure 3.8: Spherical harmonics coefficients averaged over 129 small proteins from the PDB with $K, L = 16$ and $r_{max} = 45\,\text{Å}$. The coefficients are plotted as a function of their indices. The tics count the $l$ and coefficients in between tics range from $m = [-l, l]$.

The second solution scheme aims toward exploiting the resolution property of the spherical harmonics framework. Instead of modifying atomic positions, a Monte Carlo search is performed on the spherical harmonics coefficients of complex Fourier space. Each step, the coefficients are varied with a small Gaussian distributed change. The changes are $l, m$-dependent and their values were chosen to be proportional to the average value of such a coefficient when averaging over 100 small proteins from the PDB (see Figure 3.8). An advantage of this method is to have the phases in complex Fourier space without the need of phasing algorithms.

After each coefficient modification, the $\ln p$ is calculated and compared to the previous one. To lower the search space and to introduce additional constraints on the coefficients, each proposed set of coefficients is forced to give a positive electron density with the correct support. The support is defined to be the space in which the

electron density $\rho(\mathbf{x})$ is nonzero. This procedure was implemented by transforming the coefficients from spherical harmonics to real space, imposing the constraints and back transforming. A convenient feature of this solution scheme is, that the search process can be started with a small bandwidth $L$ meaning that the Fourier structure is determined at low angular resolution. The idea is to increase $L$ progressively once the lower $lm$-coefficients have converged.

As a result, this method also renders multiple solutions that produce similar correlations but represent different structures that differ from each other in more than just a three dimensional Euler rotation as can be seen in Figure 3.9. Despite additional constraints the correct solution could not be successfully isolated.

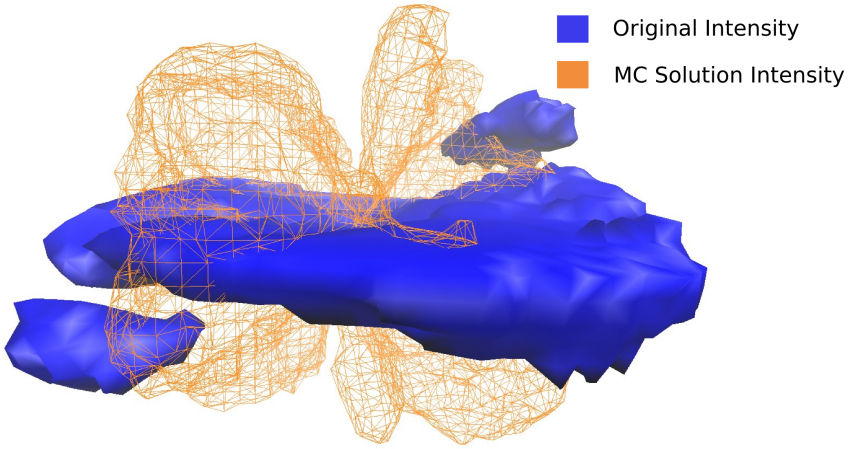



Figure 3.9: Isosurface plot of 3D intensity of the Glutathion and a solution found in the coefficient space Monte Carlo search. Both intensities differ in shape but give the same two photon correlations.

The two presented approaches have shown that this is a general problem and that it is hard to solve the problem of degeneracy with additional constraints. However, it is not a proof that such constraints, that unambiguously isolates the correct structure in terms of its two photon correlation, do not exist. Still the studies were carried out with small proteins whereas common applications would face much larger proteins in which the degeneracy would be even larger and harder to compensate for. Therefore, the next section will analyze the triple photon correlation to see if the inversion is possible and if this data structure represents a unique descriptor of the molecule structure.

## 3.6 Three Photon Correlation

It has been shown that the two photon correlation lacks information about the structure. The missing information has been identified as the unknown rotations in the cryptic $2l + 1$-dimensional space of coefficients. The question remains, if these unknowns can be obtained from higher order correlations or if reconstruction from the next higher triple correlation is possible all together. To answer that, we derive the analytical expression of the three photon correlation in analogy to Section 3.2.

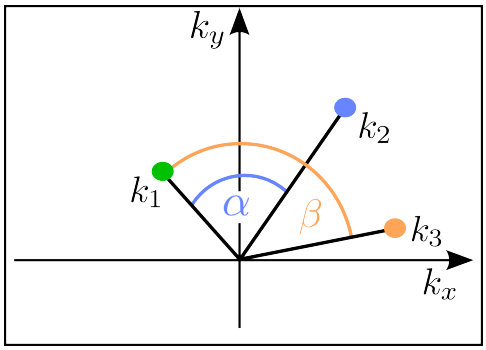



Figure 3.10: Three photon scattering event on the detector screen. One can extract the distances to the center $k_1, k_2$ and $k_3$ and the respective angular relationships $\alpha \in [0, \pi]$ and $\beta \in [0, 2\pi]$.

The definition of the correct angle assignment in the three photon correlation has to be chosen carefully. The two photon correlation uses the smallest angle between two points in the plane. However, introducing a third point relative to the two makes it important to distinguish between then two situations sketched in Figure 3.11. Using the smallest angle definition for $\beta \in [0, \pi]$ would introduce ambiguity. Therefore $\beta$ needs to be defined clockwise in the range $[0, 2\pi]$ to cover the complete circle.

To calculate the triple correlation, one averages again over all rotations of the product of three intensities in the 2-dimensional Fourier projection corresponding to a certain angular relation $(\alpha, \beta)$:

$$c_{k_1,k_2,k_3,\alpha,\beta} = \langle I_\omega(k_1, 0) \cdot I_\omega(k_2, \alpha) \cdot I_\omega^*(k_3, \beta) \rangle . \tag{3.39}$$

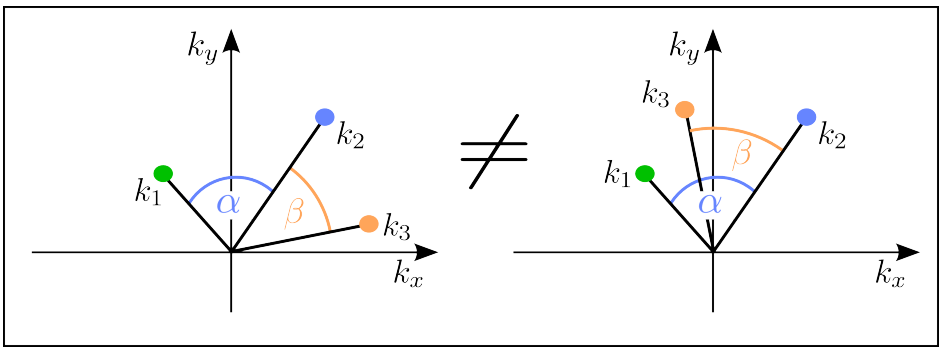



Figure 3.11: Two three photon scenarios that would not be distinguishable in a smallest angle definition framework because the angle $\beta$ would be identical in both cases although photon $k_3$ is positioned at two different positions relative to the other two photons.

By averaging, every possible triplet combination that contributes to a correlation entry is accounted for. Inserting the spherical harmonics expansion of the projection from Equation 3.14 into Equation 3.39 then yields:

$$\begin{aligned} c_{\alpha,\beta} \quad = \quad & \sum_{l1\,l2\,l3} \sum_{m1\,m2\,m3} \sum_{m1'\,m2'\,m3'} A_{l1m1}\left(k_1\right) A_{l2m2}\left(k_2\right) A^*_{l3m3}\left(k_3\right) \\ & Y_{l1m1'}\left(\frac{\pi}{2}, 0\right) \cdot Y_{l2m2'}\left(\frac{\pi}{2}, \alpha\right) \cdot Y^*_{l3m3'}\left(\frac{\pi}{2}, \beta\right) \\ & \left\langle R^{l1}_{m1m1'} \cdot R^{l2}_{m2m2'} \cdot R^{l3}_{m3m3'} \right\rangle_\omega . \end{aligned} \tag{3.40}$$

The averaging again only concerns the rotation factors and therefore the coefficients as well as the basis functions with fixed parameters can be pulled out of the averaging operator. It is possible to rewrite the product of two rotation factors $R$ into one as follows[36]:

$$\begin{aligned} R^{l1}_{m1m1'} R^{l2}_{m2m2'} \quad = \quad & \sum_{L=|l_1-l_2|}^{l_1+l_2} \sum_{MM'} \left(2L+1\right) \left(-1\right)^{M-M'} \cdot \\ & \begin{pmatrix} l_1 & l_2 & L \\ m_1 & m_2 & -M \end{pmatrix} \\ & \begin{pmatrix} l_1 & l_2 & L \\ m_1' & m_2' & -M' \end{pmatrix} R^L_{MM'}. \end{aligned} \tag{3.41}$$

Together with the simple orthogonality theorem of two such rotations:

$$\left\langle R^L_{MM'} R^{l3*}_{m3m3'} \right\rangle_\omega = \frac{1}{2L+1} \delta_{l3L} \delta_{m3M} \delta_{m3'M'}, \tag{3.42}$$

the integral free version of the three photon correlation can be expressed as follows:

$$\begin{aligned} c_{k_1,k_2,k_3,\alpha,\beta} &= \sum_{l1\,l2\,l3}\sum_{m1\,m2\,m3} A_{l1m1}(k_1)\,A_{l2m2}(k_2)\,A^*_{l3m3}(k_3) \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & -m_3 \end{pmatrix} \\ &\quad \sum_{m1'\,m2'\,m3'} (-1)^{m3-m3'} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1' & m_2' & -m_3' \end{pmatrix} \\ &\quad Y_{l1m1'}\left(\frac{\pi}{2},0\right)\cdot Y_{l2m2'}\left(\frac{\pi}{2},\alpha\right)\cdot Y^*_{l3m3'}\left(\frac{\pi}{2},\beta\right). \end{aligned} \tag{3.43}$$

To facilitate the analysis the basis function is merged into the complex function $f\,(l_1,l_2,l_3,m_1,m_2,m_3,\alpha,\beta)$:

$$\begin{aligned} c_{k_1,k_2,k_3,\alpha,\beta} &= \sum_{l1\,l2\,l3}\sum_{m1\,m2\,m3} A_{l1m1}(k_1)\,A_{l2m2}(k_2)\,A^*_{l3m3}(k_3)\cdot \\ &\quad f\,(l_1,l_2,l_3,m_1,m_2,m_3,\alpha,\beta)\,. \end{aligned} \tag{3.44}$$

In comparison to the two photon correlation, coefficients for different $l$ appear in the same product and the basis function $f$ determines the index combination for which this is the case. The Wiegner-3j functions are well studied and give nonzero functions for cases in which $m_1 + m_2 = m_3$ and $|l_1 - l_2| \leq l_3 \leq l_1 + l_2$ (triangular inequality). Further analysis on the triple correlation basis function $f$ needs to be done to determine if inversion of this equation is possible. In analogy to Section 3.5 Monte Carlo schemes could be used to search for a distinct structure with the triple correlation as a similarity measure. However there is one approach that is presented in the following section that only utilizes parts of the three photon correlation to eliminate the missing rotations of the reconstruction from the two photon cases.

## 3.7 Degenerate Three Photon Correlation

Equation 3.43 contains the sum over an irregular set of coefficient indices combinations which makes it difficult to invert. Already in 1980 Kam discussed a similar problem in the context of electron micrography imaging [37]. Experimentally, more photons are available for such a setup which is why he proposed to only regard the degenerate form of the triple correlation. In a degenerate event two out of the three photons hit the same detector pixel and overlap $\mathbf{K_1} = \mathbf{K_2}$ (see Figure ). In these cases the mathematical description of the problem can be facilitated.

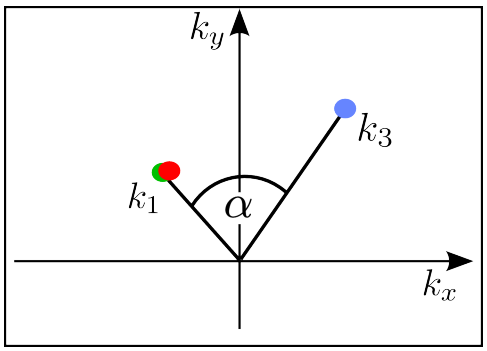



Figure 3.12: Degenerate three photon event with two photons detected at the same detector pixel.

The following identity is used:

$$Y_{l1m1'}(\Omega_1) \cdot Y_{l2m2'}(\Omega_1) = \sum_{\lambda=|l_1-l_2|}^{\lambda} \sum_{\mu=-\lambda} (-1)\,\mu \sqrt{\frac{(2l_1+1)(2l_2+1)(2\lambda+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & \lambda \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & \lambda \\ m_1' & m_2' & -\mu \end{pmatrix} Y_{\lambda\mu}(\Omega_1)\,. \tag{3.45}$$

that expresses the product of two spherical harmonics with different indices $l_1, m_1^{'}$ and $l_2, m_2^{'}$ at a particular point on the sphere $\Omega_1$ as a sum of single harmonic functions going over a range of indices. Additionally, the sum over Wiegner-3j symbols can be facilitated with the following expression:

$$\sum_{m_1'm_2'} \begin{pmatrix} l_1 & l_2 & \lambda \\ m_1' & m_2' & -\mu \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1' & m_2' & -m_3' \end{pmatrix} = \frac{1}{2l_3+1} \delta_{\lambda l_3} \delta_{\mu m_3'}. \tag{3.46}$$

Inserting the two equations above into the triple correlation from Equation 3.20, one finds the following form:

$$c_{k_1,k_1,k_3,\alpha,\alpha} = \sum_{l1\,l2\,l} \sum_{m1\,m2\,m} A_{l1m1}(k_1)\, A_{l2m2}(k_1)\, A^*_{l3m3}(k_3)\, P_l(\cos(\alpha)) \\ (-1)^m \sqrt{\frac{(2l_1+1)(2l_2+1)(2l+1)}{(4\pi)^3}} \\ \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & -m \end{pmatrix}. \tag{3.47}$$

In this degenerate version, the triple correlation $c_{k_1,k_1,k_3,\alpha,\alpha} = t_{k_1,k_3,\alpha}$ entries contains the same number of unknowns as the conventional two photon correlation $c_{k_1,k_3,\alpha}$, namely the spherical harmonics coefficients $A_{lm}(k_1)$ and $A_{lm}(k_3)$ of only two shells. The idea is to use this additional information to eliminate the missing rotations $\mathbf{U}^l_{mm'}$ of the two photon reconstruction by iteratively inverting the triple correlation and solving for the rotations. In this approach, the rotations are the only unknowns while the coefficients $\mathbf{A}^0_l(k)$ are considered to be known and chosen as one of the many degenerate solutions:

$$\mathbf{A}_l(k) = \mathbf{U}_l \mathbf{A}^0_l(k). \tag{3.48}$$

More detailed information about the equations required for this iterative rotation retrieval can be found in Kam's work Appendix B [37]. The simulations that have been done for this work have revealed that the correct rotations are no fixed points and convergence was never reached although Kam claims that reconstruction is possible. More fundamentally, in the experiment there will not be sufficient numbers of degenerate three photon pairs to determine this part of the triple correlation up to necessary accuracy. Still, the degenerate photon correlation is sufficient to obtain the missing information from experimental scattering images and this fact indicates that three photon correlations has all information to unambiguously reconstruct the electron density. However, it will require a more detailed analysis to completely understand the information content of the triple correlation. If no analytical inversion can be found, a nonlinear solver might be applied to find the correct coefficients. Unfortunately, due to the number of unknown coefficients, the search space is large which is a problem for most solvers. Finding a solution for low resolution coefficients might be a feasible solution for that problem.

# 4 Summary and Conclusion

Single molecule scattering experiments hold the promise of protein structure determination without the need of crystallization. However, the experimental setup has two major problems. The orientation of the protein at the moment of exposure is unknown and the expected number of scattered photons per image is small with approximately 68 photons for a protein with 1000 atoms. Averaging over diffraction patterns that belong to the same molecule orientation involves an error-prone orientation classification of the pictures. A reconstruction method is needed, that is rotation invariant, stable to noise and capable of structure reconstruction from scattering images with small number of photons.

Here we have presented the idea to use two and three photon correlations for the reconstruction. The correlations are histograms of photon pairs and triplets in the pictures with particular distances from the detector center and angles between the photons. The approach makes it unnecessary to determine the most probable protein orientation belonging to each picture because it operates on the level of photon pairs and triplets. Also, by using the relative photon positions in the correlation, one ensures that the resulting data is rotation invariant.

In order to assess the information content of the two and three photon correlation, an expression of the correlations in terms of the Fourier intensity was derived. To facilitate the calculations, a spherical harmonics representation of three dimensional structures was utilized because the description framework supports the spherical nature of the correlation, incorporates the symmetry of the intensity and allows to vary the resolution of the description by limiting the number of coefficients.

In a next step, the inversion of the two photon correlation was analyzed to answer the question if the spherical harmonics coefficients of the intensity can be retrieved. The derivation showed that the two photon correlation is comprised of quasi scalar products of spherical harmonics coefficients with the same orbital momentum number $l$. As a result, upon reconstruction, these coefficients can only be retrieved up to an arbitrary rotation in the high dimensional coefficient spaces. Thus, we showed

that the two photon correlation itself does not contain all necessary information to reconstruct the electron density unambiguously.

The arbitrary rotations represent a set of unknowns. In this thesis two ideas have been discussed to overcome that information deficiency by using the correlation as a similarity measure in two different Monte Carlo search schemes. In the first approach, a model of the protein was built as a linear superposition of Gaussian spheres representing the electron densities of the atoms. The atomic positions were varied until the two photon correlation of this model structure coincided with the correlation of the reference structure. For that purpose, a difference measure based on probability was proposed to compare correlations and to determine their similarity.

Although the correlation could be used to determine the distances of a 3 atom system, for more than 3 atoms, the atomic position displacement method has not converged to the correct solution. Instead structures were found which render very similar correlations but which are different from the original system.

A second approach performed a search in the spherical harmonics coefficient of the complex Fourier space. The idea was to first optimize the low resolution coefficients in a hierarchical approach and later proceed with higher order coefficients to increase the resolution of the solution. Unfortunately, this scheme also rendered many solutions. Posing additional constraints on the coefficients such that they make the corresponding electron density positive and within the correct support, did not eliminate the multitude of solutions.

The two presented solution approaches were not able to eliminate the missing unknowns but this work cannot exclude that constraints exist that allow to use the two photon correlation for structure reconstruction. Here we showed however, that the degeneracy of the solution space is difficult to eliminate with constraints and is expected to grow combinatorially with molecule size.

The derivation of the three photon correlation in Section 3.6 has given a similar sum of products like the two photon correlation. The resulting form consists of triple products which make the inversion mathematically more challenging. In comparison to the two photon case, the triple correlation links coefficients belonging to different orbital momenta numbers $l$. The loss of these mixed products was the reason why a complete reconstruction of the coefficients from two photon correlations was impossible. Hence, the occurrence of these coefficient products with different $l$ indicates that the structure can be retrieved from this correlation. Furthermore the study

of the degenerate three photon correlation revealed that it can be used to retrieve the arbitrary rotations of the solution of the two photon correlation inversion. This fact answers one of the main question of this work, namely that three photons per picture are sufficient to unambiguously reconstruct the protein structure. Unfortunately, the studied iterative solver did not converge and this problem remains to be solved.

## 4.1 Outlook

In the follow up work of this thesis, a suitable solution scheme to retrieve the coefficients from the two and three photon correlation needs to be developed. There are two possible ways to proceed.

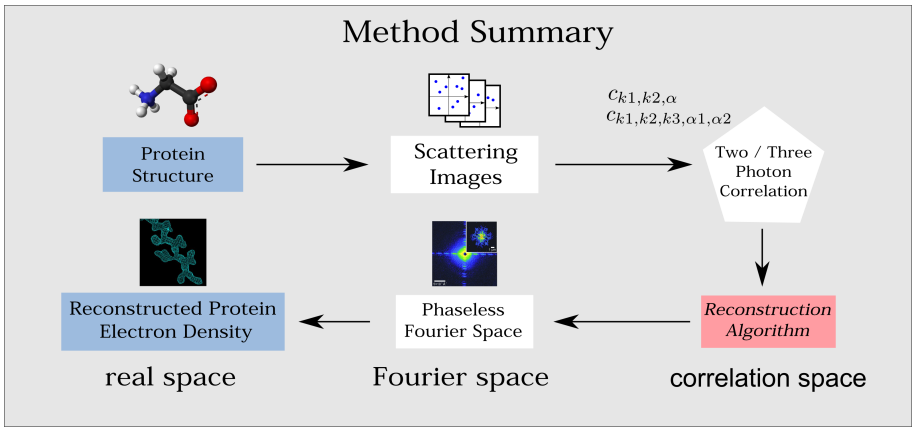



Figure 4.1: A method summary for the reconstruction of protein's electron density with the help of two and three photon correlations.

First, one might develop a solver for the unknown rotations with the help of the degenerate triple correlation. Unfortunately, this degenerate part is only a small fraction of the experimental histogram data and methods are preferred in which all the acquired data, i.e. the whole three photon correlation, is used.

Thus the second path aims toward inverting the whole triple correlation. Equation 3.44 expressed the correlation in its most basic form in which triple products belonging to a set of indices contribute to a certain entry according to a complex basis function $f\left(l_1, l_2, l_3, m_1, m_2, m_3\right)$. Studying the orthogonality of these functions might open a way to invert the equations. Figure 4.1 summarizes the steps that are involved in retrieving a protein structure from a single molecule scattering experiment along the two described paths.

As an alternative approach to the analytical solution, a Monte Carlo search could be performed on the complex Fourier space coefficients in analogy to the two photon case in Section 3.5.3. The hypothesis is that the three photon correlations can distinguish different structures better than the two photon version and thus the degeneracy of solutions is lower or nonexistent.

For future photon correlation based reconstruction schemes, the shot noise needs to be considered. As a simple denoising scheme one may simply calculate the correlations of Poissonian noise and subtracting it from the experimental correlation with the signal to noise ratio as a scaling factor.

# Acknowledgment

I would like to thank my supervisor Prof. Dr. Helmut Grubmüller for giving me the opportunity to work on this project. He provided the initial concept of the photon correlations and offered an inspiring collaboration with many interesting and helpful discussions. It was fun to develop new ideas in a team and to benefit from his experience and expertise. I would also like to thank Prof. Dr. Tim Salditt for being the Second Referee of this thesis.

A big thanks goes to Michal Walczak for all the feedback and idea discussions and for sharing his single molecule scattering know-how. I also like to thank Christian Blau, Carsten Kutzner, Lars Bock and Frank Wiederschein for their practical tips with all kinds of computer and math problems and Petra Keller for sharing her knowledge in writing scientific publications. My final thanks the whole department of Computational and Theoretical Biophysics at the Max-Planck Institute for Biophysics for the kind and welcoming atmosphere that I appreciated a lot.

# References


[1] A. Assmus. Early history of x-rays. *SLAC Stanford*, pages 10–24, 1995.

[2] D.W.J. Cruickshank. Time-resolved macromolecular crystallography: introductory remarks and a little history. *Philosophical Transactions: Physical Sciences and Engineering*, pages 169–173, 1992.

[3] K.D. Pruitt, T. Tatusova, G.R. Brown, and D.R. Maglott. Ncbi reference sequences (refseq): current status, new features and genome annotation policy. *Nucleic Acids Research*, 40(D1):D130–D135, 2012.

[4] Research Collaboratory for Structural Bioinformatics (RCSB) consortium. Protein databank statistics. URL `http://www.pdb.org/pdb/static.do?p=general_information/pdb_statistics/index.html`.

[5] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu. Potential for biomolecular imaging with femtosecond x-ray pulses. *Nature*, 406(6797): 752–757, 2000.

[6] V.L. Shneerson, A. Ourmazd, and D.K. Saldin. Crystallography without crystals. i. the common-line method for assembling a three-dimensional diffraction volume from single-particle scattering. *Acta Crystallographica Section A: Foundations of Crystallography*, 64(2):303–315, 2008.

[7] D.K. Saldin, V.L. Shneerson, R. Fung, and A. Ourmazd. Structure of isolated biomolecules obtained from ultrashort x-ray pulses: exploiting the symmetry of random orientations. *Journal of Physics: Condensed Matter*, 21:134014, 2009.

[8] D.K. Saldin, V.L. Shneerson, M. Howells, S. Marchesini, H.N. Chapman, R. Kirian, K.E. Schmidt, and J.C.H. Spence. Structure of a single particle from scattering by many particles randomly oriented about an axis: toward structure solution without crystallization. *New Journal of Physics*, 12:035014, 2010.

[9] D.K. Saldin, H.C. Poon, P. Schwander, M. Uddin, and M. Schmidt. Reconstructing an icosahedral virus from single-particle diffraction experiments. *Optics Express*, 19(18):17318–17335, 2011.

[10] R. Fung, V. Shneerson, D.K. Saldin, and A. Ourmazd. Structure from fleeting illumination of faint spinning objects in flight. *Nature Physics*, 5(1):64–67, 2008.

[11] M. Walczak. Bayesian orientation estimate from single molecule x-ray scattering data. *Diploma-Thesis*, 2010.

[12] M. Altarelli, R. Brinkmann, M. Chergui, W. Decking, B. Dobson, S. Duesterer, G. Gruebel, W. Graeff, H. Graafsma, J. Hajdu, et al. The european x-ray free-electron laser. *Technical Design Report, DESY*, 97, 2006.

[13] J. Hajdu. Imaging of single particles and biomolecules. *Presentation Uppsala University*.

[14] J.C.H. Spence, K. Schmidt, J.S. Wu, G. Hembree, U. Weierstall, B. Doak, and P. Fromme. Diffraction and imaging from a beam of laser-aligned proteins: resolution limits. *Acta Crystallographica Section A: Foundations of Crystallography*, 61(2):237–245, 2005.

[15] J. Fenn. Electrospray ionization mass spectrometry: How it all began. *Journal of Biomolecular Techniques*, 13(3):101, 2002.

[16] K.J. Gaffney and H.N. Chapman. Imaging atomic structure and dynamics with ultrafast x-ray scattering. *Science*, 316(5830):1444–1448, 2007.

[17] J. Hajdu. Single-molecule x-ray diffraction. *Current Opinion in Structural Biology*, 10(5):569–573, 2000.

[18] L. Inhester, C.F. Burmeister, G. Groenhof, and H. Grubmueller. Auger spectrum of a water molecule after single and double core ionization. *The Journal of Chemical Physics*, 136:144304, 2012.

[19] S.M. Seltzer J. Chang J.S. Coursey R. Sukumar D.S. Zucker K. Olsen M.J. Berger, J.H. Hubbell. Xcom: Photon cross sections database, 1998 - 2010. URL `http://www.nist.gov/pml/data/xcom/index.cfm`.

[20] W. Demtroeder. *Experimentalphysik 2*. Springer, 2006.

[21] J.R. Fienup. Reconstruction of an object from the modulus of its fourier transform. *Optics letters*, 3(1):27–29, 1978.

[22] C.D.N Chan. Spherical harmonics schematic - wikimedia picture. URL `http://commons.wikimedia.org/wiki/File:Harmoniques_spheriques_positif_negatif.png`.

[23] N. Baddour. Operational and convolution properties of three-dimensional fourier transforms in spherical polar coordinates. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 27(10):2144–2155, Oct 2010.

[24] Masayuki Toyoda and Taisuke Ozaki. Fast spherical bessel transform via fast fourier transform and recurrence formula. *Computer Physics Communications*, 181(2):277–282, 2010. ISSN 0010-4655. doi: 10.1016/j.cpc.2009.09.020. URL `http://www.sciencedirect.com/science/article/pii/S0010465509003105`.

[25] B. Leistedt, A. Rassat, A. Refregier, and J.L. Starck. 3dex: a code for fast spherical fourier-bessel decomposition of 3d surveys. *Astronomy & Astrophysics*, 540, 2012.

[26] W. Wieder. A generalized debye scattering formula and the hankel transform. *Zeitschrift fuer Naturforschung*, 54:124–130, 1999.

[27] H.B. Stuhrmann. Interpretation of small-angle scattering functions of dilute solutions and gases. a representation of the structures related to a one-particle scattering function. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 26(3):297–306, 1970.

[28] E.W. Weisstein. Spherical harmonic, 2012. URL `http://mathworld.wolfram.com/SphericalHarmonic.html,http://functions.wolfram.com/HypergeometricFunctions/SphericalHarmonicYGeneral/`.

[29] M. Vesper. Verfeinerung der proteinkristallographie durch flexibilitaetsvorhersagen. Master's thesis, Max-Planck-Institut fuer Biophysikalische Chemie, 2008.

[30] V. Elser. Strategies for processing diffraction data from randomly oriented particles. *Ultramicroscopy*, 111(7):788–792, 2011.

[31] F. Abbate, B. von Ardenne, and P. De Cruz. Gsl-shell lua extension math library online manual, 2010 - 2012. URL `http://www.nongnu.org/gsl-shell/doc/`.

[32] D.M. Healy, D.N. Rockmore, P.J. Kostelec, and S. Moore. Ffts for the 2-sphere-improvements and variations. *Journal of Fourier Analysis and Applications*, 9 (4):341–385, 2003.

[33] R. Suda and M. Takami. A fast spherical harmonics transform algorithm. *Mathematics of Computation*, 71(238):703–716, 2002.

[34] J. Blais. Discrete spherical harmonic transforms: Numerical preconditioning and optimization. *Computational Science–ICCS*, pages 638–645, 2008.

[35] A.J. McCoy. Liking likelihood. *Acta Crystallographica Section D: Biological Crystallography*, 60(12):2169–2183, 2004.

[36] E.W. Weisstein. Wigner 3j-symbol, 2012. URL `http://mathworld.wolfram.com/Wigner3j-Symbol.html`.

[37] Z. Kam. The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology*, 82(1):15–39, 1980.

**Erklärung** nach §18(8) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder singemäSS aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den January 8, 2013

(Benjamin von Ardenne)