UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1625

Imaging Living Cells with an X-ray Laser

GIJS VAN DER SCHOT

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

ISSN 1651-6214
ISBN 978-91-513-0217-1
urn:nbn:se:uu:diva-334219

Dissertation presented at Uppsala University to be publicly examined in B/A1:111a, BMC, Husargatan 3, Uppsala, Monday, 12 March 2018 at 09:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Pierre Thibault (University of Southampton).

Abstract
van der Schot, G. 2018. Imaging Living Cells with an X-ray Laser. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1625. 79 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0217-1.

Imaging living cells at a resolution higher than the resolution of optical microscopy is a significant challenge. Fluorescence microscopy can achieve a degree of super-resolution via labeling cellular components with a fluorescent dye. Reaching nanometer or sub-nanometer resolution requires high-energy radiation with significantly shorter wavelength than that of optical light. X-rays and electrons have the requisite wavelengths and could be suitable for such studies; however, these probes also cause significant radiation damage. A dose in excess of 100,000,000 Gray (Gy, J/kg) would be required to reach nanometer resolution on a cell, and no cell can survive this amount of radiation. As a consequence, much of what we know about cells at high resolution today comes from dead material.

Theory predicts that an ultra-short and extremely bright coherent X-ray pulse from an X-ray free-electron laser can outrun key damage processes to deliver a molecular-level snapshot of a cell that is alive at the time of image formation. The principle of 'diffraction before destruction' exploits the difference between the speed of light (the X-ray pulse) and the much slower speed of damage formation. The femtosecond pulse 'freezes' motion in the cell at physiological temperatures on the time scale of atomic vibrations, offering unprecedented time resolution and a plethora of new experimental possibilities.

This thesis describes the first test experiments on imaging living cells with an X-ray laser. I present results in three essential areas of live cell imaging. (i) We have used an aerosol injector to introduce live cyanobacteria into the X-ray focus, and recorded diffraction patterns with extremely low background at very high hit rates. (ii) We demonstrated scattered signal beyond 4 nm resolution in some of these experiments. (iii) The thesis also describes image reconstruction, using a new fully automated pipeline that I developed during my studies. The reconstruction of diffraction patterns was successful for all patterns that did not have saturated pixels. The new software suite, called RedFlamingo, selects exposures with desired properties, can sort them according to sample size, shape, orientation, exposure, the number and type of objects in the beam during the exposure, their distance from each other, and so forth. The software includes validation tools to assess the quality of the reconstructions.

Keywords: Coherent diffractive X-ray imaging, flash X-ray imaging, lensless imaging, single particle imaging, cyanobacteria, phasing, image classification, substrate-free sample delivery, X-ray free-electron laser, XFEL, FEL, CDXI, CDI, CXI, FXI

Gijs van der Schot, Department of Cell and Molecular Biology, Molecular biophysics, Box 596, Uppsala University, SE-75124 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-0217-1
urn:nbn:se:uu:diva-334219 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-334219)

Dedicated to new Life: Emmanuel and Mathis

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **van der Schot, G** et al. Imaging single cells in a beam of live cyanobacteria. *Nature Communications* **6**, 5704 (2015).

II **van der Schot, G** et al. Open dataset of live cyanobacterial cells imaged using an X-ray laser. *Nature Scientific Data* **3**, 160064 (2016)

III Munke, A., **van der Schot, G** et al. Open dataset of RDV particles. *in preparation*

Reprints were made with permission from the publishers.

Supporting papers

IV G. van der Schot and A.M.J.J. Bonvin. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. Journal of Biomolecular NMR, 62(4),497-502, 2015.

V G. van der Schot *et al.* Improving 3D structure prediction from chemical shift data. Journal of biomolecular NMR, 57(1):27-35,2013.

VI T.A. Wassenaar *et al.* WeNMR: Structural Biology on the Grid. Journal of Grid Computing, 10(4):747-767, 2012.

VII A. Rosato *et al.* Blind testing of routine, fully automated determination of protein structures from NMR data. Structure, 20(2):227-236, 2012.

VIII M. Vendruscolo *et al.* Protein structure determination using chemical shifts, NMR in Mechanistic Systems Biology, Chapter 10, 107-110, 2009.

IX A. Rosato *et al.* CASD-NMR: critical assessment of automated structure determination by NMR. Nature Methods, 6(9):625-626, 2009.

X T. Gorkhover Femtosecond X-ray Fourier holography imaging of free-flying nanoparticles. Nature Photonics *Accepted*

XI R.P. Kurta *et al.* Correlations in scattered x-ray laser pulses reveal nanoscale structural features of viruses. Phys. Rev. Lett. 119, 158102, 2017.

XII H.K.N. Reddy *et al.* Coherent soft X−ray diffraction imaging of coliphage PR772 at the Linac coherent light source. Scientific Data, 4:170079, 2017.

XIII B.J. Daurer *et al.* Experimental strategies for imaging bioparticles with femtosecond hard X-ray pulses. IUCrJ, 4(3):252-262, 2017.

XIV M.F. Hantke *et al.* A data set from flash X-ray imaging of carboxysomes. Scientific data, 3:160061, aug 2016.

XV A. Munke *et al.* Coherent diffraction of single Rice Dwarf virus particles using hard X-rays at the Linac Coherent Light Source. Scientific data, 3:160064, 2016.

XVI T. Ekeberg *et al.* Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser. Physical Review Letters, 114(9):098102, 2015.

XVII M.F. Hantke *et al.* High-throughput imaging of heterogeneous cell organelles with an X-ray laser. Nature Photonics, 8(12):943-949, 2014.

XVIII A.V. Martin *et al.* Noise-robust coherent diffractive imaging with a single diffraction pattern. Optics Express, 20(15):16650, 2012.

Contents

List of abbreviations

AMO	Atomic, Molecular and Optical Sciences
FXI	femtosecond X-ray Imaging
EM	Electro Magnetic
DFT	Direct Fourier Transform
EMC	Expansion Maximization Compression
ER	Error Reduction
ESI	Electro Spray Ionisation
FEL	Free-electron laser
FIB	Focused ion beam
fs	femtosecond
GDVN	Gas Dynamic Virtual Nozzle
HIO	Hybrid Input Output
LCLS	Linac Coherent Light Source
NMR	Nuclear Magnetic Resonance
PRTF	Phase Retrieval Transfer Function
RAAR	Relaxed Averaged Alternating Reflections
SASE	Self Amplified Stimulated Emission
SFX	Serial femtosecond x-ray crystallograhpy
SPI	Single Particle Imaging
XFEL	X-ray Free-Electron Laser

1. Motivation

Cellular life and the organization of its constituents are amazingly intricate and diverse. Proteins form an interconnected and dynamic network in which specific changes to individual proteins can trigger a variety of global responses. In order to understand the factors that activate or deactivate various pathways we need to study the entire living system, including individual components and their interactions with each other. A grand challenge of the 21st century is the imaging of live cells, at or near atomic resolution, with a time resolution that allows capturing the fastest biological processes.

Super-resolution optical microscopy can image labeled parts of cells and has increased our understanding of cellular organization significantly [1]. However the technique requires the introduction of a fluorescent label, and it is ultimately limited by the size of this label. Nuclear Magnetic Resonance (NMR) can study the dynamics of proteins at atomic resolution, with picosecond time resolution [2]. Recently it has succeeded in studying labelled proteins *in vivo* [3]. In general, NMR is limited by the size of proteins. I have participated in computationally predicting protein structure from backbone chemical shift only, as shown in **Papers IV-IX**. A promising method to study cells as well as its constituents is electron cryo-microscopy. Focussed ion beams can be used to slice cryo-frozen cells into thin sections [4]. This allowed the study of internal features of the cells. Using sub-tomogram averaging, the structure of highly abundant proteins can be elucidated at resolutions beyond 4 Å[5]. For understanding rare events, however, acquisition time is a limiting factor. Although these results are incredible, cryo-electron microscopy does not study living cells. A further alternative that can be utilized to study living cells is femtosecond x-ray diffractive imaging (FXI). FXI uses ultra-short and extremely bright X-ray pulses produced by X-ray free-electron lasers (XFEL). The power of the pulse enables the measurement of interpretable signal from single bioparticles that otherwise would not scatter strong enough. The femtosecond pulse can outrun key damage processes in the sample. It is predicted that sub-nanometer resolution can be achieved on micron-sized cells with this method [6]. The femtosecond pulse gives an unprecedented time resolution that can capture the fastest biologically relevant motion at room temperature. Another big advantage of this method is the extremely high repetition rate. The recently operational European XFEL (EuXFEL) has a repetition rate of 27 000 Hz [7], potentially allowing over a billion images of a billion cells to be recorded in one day, and may open up new avenues of research in cell biology.

This thesis deals with the experimental verification of FXI on living cells, and studies if, and what, computational and experimental tools are necessary

to make high-resolution cell imaging a reality. This thesis will start by describing the general framework of the experiment: the interaction of light and matter, the generation of X-rays, the substrate-free sample delivery method, the recording of two- dimensional (2D) diffraction patterns, the reconstruction of images, how 2D images might be combined to derive three-dimensional (3D) structural information. The chapters in the results section describe a software package called RedFlamingo which is used for the classification of diffraction patterns, and the experimental results on imaging *living* cells, and how automated image classification has contributed to pattern selection and reconstruction.

2. Interaction of light and matter

Traditionally cells and cellular structures have been studied with optical light. To understand why light is a good probe to study cells, it is necessary to understand the nature of its interaction with matter. This chapter will create the mathematical framework that describes how light behaves in our diffraction experiments.

2.1 Light as electromagnetic radiation

Following the first look into the micro-world by Stelluti and Cesi [8], using a compound microscope built by Galilei, and by the spectacular discoveries of Anton Van Leeuwenhoek with his compact microscope[9], biology underwent a fundamental transformation. Light microscopy revealed the existence of cells, tissues, microorganisms and many other biological structures and phenomena. A deeper understanding of the nature of light came about much later.

Theoretical and experimental advances in the 19th century established the wave properties of light. A dramatic change took place in the 1860s, when James Clerk Maxwell unified the fields of electricity, magnetism, and somewhat surprisingly, optics. His results are known as the Maxwell equations:

$$\nabla \cdot \vec{E} = \frac{\rho}{\varepsilon_0} \tag{2.1}$$

$$\nabla \cdot \vec{B} = 0 \tag{2.2}$$

$$\nabla \times \vec{E} = -\frac{d\vec{B}}{dt} \tag{2.3}$$

$$\nabla \times \vec{B} = \mu_0 \vec{J} - \mu\varepsilon \frac{d\vec{E}}{dt} \tag{2.4}$$

Here, $\vec{E}$ is the electric field, $\vec{B}$ is the magnetic field, ε and μ are respectively the permittivity and the permeability of the material. ρ is the the charge density and $\vec{J}$ describes the local current. An important consequence of these equations is that a moving charge will induce a magnetic field, which in its turn induces a change in the electric field, and so on.

In order to understand why these laws unify magnetism, electricity and optics, we will consider a special case: vacuum. In vacuum there are no localized charges ($\rho = 0$) and no currents ($\vec{J} = 0$), which simplifies equation 2.1 and 2.4 to:

$$\nabla \cdot \vec{E} = 0 \tag{2.5}$$

$$\nabla \times \vec{B} = -\mu_0 \varepsilon_0 \frac{d\vec{E}}{dt} \tag{2.6}$$

Furthermore, experiments showed that the permittivity for vacuum, $\varepsilon_0 = 8.854 \cdot 10^{-12} Fm^{-1}$, and the permeability for vacuum $\mu_0 = 1.257 \cdot 10^{-7} NA^{-2}$. If we now ask ourselves what the field change caused by a change in the electric field is by studying the equation $\nabla \times (\nabla \times \vec{E})$, we can derive something quite amazing:

$$\nabla \times (\nabla \times \vec{E}) = \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = -\nabla^2 \vec{E}$$

The first equality is a well known mathematical relation, and for the second equality equation 2.5 is used. Now lets rewrite the equation in a different way:

$$\nabla \times (\nabla \times \vec{E}) = \nabla \times (-\frac{d\vec{B}}{dt}) = -\frac{d}{dt}(\nabla \times \vec{B}) = \frac{d}{dt}(\mu_0 \varepsilon_0 \frac{d\vec{E}}{dt}) = \mu_0 \varepsilon_0 \frac{d^2 \vec{E}}{dt^2}$$

In the first equality we used equation 2.3. In the second equality we used a general algebraic property. In the third equality we used 2.6, and in the last equality we combined the two time derivatives. Together these equations state that:

$$\frac{d^2 \vec{E}}{dt^2} = \frac{1}{\mu_0 \varepsilon_0} \nabla^2 \vec{E} = v^2 \nabla^2 \vec{E}, v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \tag{2.7}$$

Equation 2.7 is known as the wave equation, which is used to describe the propagation of a wave with the speed v. Combining this results with the experimentally determined value for μ_0 and ε_0 shows that electromagnetic (EM) waves travel at $3 \cdot 10^8 ms^{-1}$, which agreed with value of the speed of light [10]. This result strongly hints at the possibility that light is an electromagnetic wave. Further experiments succeeded in demonstrating the existence of electromagnetic waves with long wavelengths and showed that their properties are consistent with the properties of visible light, including their velocity. Visible light has now become part of the broader spectrum of electromagnetic radiation. Although Quantum Mechanics complicated the description of light further, as light demonstrates particle-like behaviour under certain conditions, many phenomena however can be described by treating light as a wave.

2.2 Photon-material interactions

Assuming that light is electromagnetic radiation, it is easy to understand that a change in electric field will affect its propagation. There are four mechanisms through which radiation interacts with matter: photoabsorption, scattering, photo-nuclear absorption and pair production. The description of these

phenomena is treated as if light will come in discrete quanta: photons. Photonuclear absorption and pair production only occur when matter is exposed to high energy gamma rays, and are therefore of little relevance to this thesis. Photoabsorption is facilitated primarily through a process called photoexcitation, in which an electron is excited to a higher level of energy, or possibly ionized in the case of photoionisation. The resulting vacancy is repopulated within femtoseconds by Auger decay and fluorescence. Auger decay results in shake-up ans shake-off excitations of other electrons [11], and therefore leads to a cascade of ionisations. Scattering can be divided into two classes: elastic scattering and inelastic scattering. Elastic scattering does not result in a change of kinetic energy of the scattering particle, nor does it change the wavelength of the radiation. Only the direction of the radiation can be changed. Inelastic scattering leads to both a change in wavelength of the radiation, and a change in kinetic energy. Photo-absorption and inelastic scattering deposit energy into the object, which will lead to structural changes within the object. Elastic scattering on the other hand can be used to gain structural information about the object, without causing damage to the object.

2.3 Scattering by a single free electron

The simplest example of elastic scattering is Thompson scattering from a free electron. A free electron will scatter or diffract the incoming radiation in all directions. An important physical concept for understanding scattering on a deeper level is the Lorentz force which describes the force exerted on a charged particle travelling in an electromagnetic field.

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B} \tag{2.8}$$

From Newton's second law of motion ($\vec{F} = m\vec{a}$) it is known that force and acceleration are proportional and parallel, thus the Lorentz force describes the acceleration of a charge in an electric and/or magnetic field. The first term of equation 2.8 is also known as the Coulombic force, and shows that a charge is accelerated in the direction of an electric field. The second term shows that a moving charge is accelerated by the presence of a magnetic field in the direction that is both perpendicular to its movement and the magnetic field, given that $\vec{v}$ and $\vec{B}$ are not parallel in which case there is no acceleration.

According to classical electromagnetic theory the electric field associated with a monochromatic plane wave of amplitude E_0, and wavelength λ, propagating in the z-direction can be described by:

$$E_{in} = E_0 e^{-2\pi i \frac{ct-z}{\lambda}} \tag{2.9}$$

When the oscillating electric field of the incident electromagnetic wave hits a stationary electron of mass m_e and charge e located at position $z = 0$, it

exerts a Coulombic force on the electron, which causes it to oscillate at the same frequency as the incident radiation. By Newton's second law of motion:

$$\vec{F} = m_e\vec{a}(t) = eE_0 e^{-2\pi\, i\frac{ct}{\lambda}} \tag{2.10}$$

$$\vec{a}(t) = \frac{eE_0}{m} e^{-2\pi\, i\frac{ct}{\lambda}} \tag{2.11}$$

Here we assumed the contribution of the magnetic part of field to be negligible, as the electron will not reach a significant velocity. An accelerated charge emits electromagnetic radiation. The oscillating electron becomes a new source of radiation that radiates spherically in all directions, at the same frequency as the incident radiation. From Maxwell's equations it follows that the electric field generated by an accelerating electron measured at point $\vec{d}$ can be described as:

$$E_s(\vec{d},t) = \frac{ea_\perp(t)}{4\pi\varepsilon_0 c^2 d} \tag{2.12}$$

$a_\perp(t)$ is the acceleration projected on a plane perpendicular to $\vec{d}$.

$$a_\perp = |a(\vec{t})|\sin(\theta) \tag{2.13}$$

Combining equations 2.10, 2.12, and 2.13, we get that the instantaneous scattered field is:

$$E_s(\vec{d},t) = \frac{e^2}{4\pi\varepsilon_0 m_e c^2}\frac{E_0\sin(\theta)}{d} e^{-2\pi\, i\frac{|\vec{d}|}{\lambda}} \tag{2.14}$$

The classical electron radius r_e can be used to simplify the equation for the scattered field E_s

$$E_s = -\frac{r_e E_0\,\sin(\theta)}{d} e^{-2\pi\, i\frac{|\vec{d}|}{\lambda}}$$

$$r_e = \frac{e^2}{4\pi\varepsilon_0 m_e c^2}$$

2.4 Two-body scattering

If more than one electron are present the scattered waves will interfere, similar to the interference of water waves. The resulting pattern of dark and light bands is called a diffraction pattern. The first person to report this phenomenon was Grimaldi in 1660. Whether the interference is constructive or destructive depends on the optical path difference (OPD) between the scattered waves. Waves with path differences close to integer numbers of wavelengths interfere constructively, if close to half integral waves will interfere destructively. The

central light band is often referred to as the zeroth diffraction order, or more colloquially as the central speckle. The light band next to it is called the first diffraction order (or first fringe), the one next to that the second order and so on. These terms will be used throughout the thesis. Figure 2.1 illustrates the phenomenon of interference.

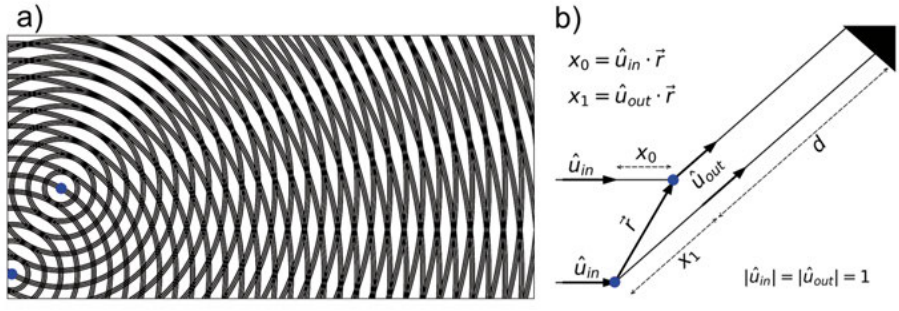


Figure 2.1. Illustration of the scattering from two free-electrons. a) The two blue dots represent two independently scattering free-electrons. The rings represent the maxima of the two scattered waves. One can see the a light-dark band structure appear, characteristic of a interference pattern. b) mathematical description of the path length difference between two waves arriving at point $\vec{d}$. $\hat{u}_{in}$ and $\hat{u}_{out}$ are unit vector pointing in the direction of the incident wave and the out going wave respectively. $\vec{r}$ is a position vector describing the relative distance between the two electrons.

If the diffraction pattern is measured at point $\vec{d}$, far away from the diffracting object itself, both scattered fields have approximately the same field strength since the distance from both electrons to the detector is about equal. In Figure 2.1 this means that $|\vec{d}+x_1| \approx |\vec{d}|$, and that both $\hat{u}_{out}$ are parallel. This approximation is called the far-field approximation. The Fresnel number (FN) is used to verify the validity of the far-field approximation.

$$FN = \frac{o^2}{d\lambda} \tag{2.15}$$

where o is the object size, d the distance from the object to the detector, and λ the wavelength of the radiation. The far-field approximation is valid when $FN \ll 1$. All diffraction patterns discussed in this thesis are taken in the far-field.

The electric field at point $\vec{d}$ is given by a sum of the individual scattered electric fields:

$$E(\vec{d}) = E_s(\vec{d}) + E_s(\vec{d})e^{\frac{2\pi i \Delta x}{\lambda}} \tag{2.16}$$

Δx is the optical path difference between the two scattered waves due to the relative difference in position. Δx can be determined by summing x_0 which is the path difference of the incoming radiation and x_1 which is the path difference of the outgoing radiation. Both x_0 and x_1 can be described using the

difference vector $\vec{r}$:

$$\Delta x = x_0 - x_1 = \hat{u}_{out} \cdot \vec{r} - \hat{u}_{in} \cdot \vec{r} = (\hat{u}_{out} - \hat{u}_{in}) \cdot \vec{r} \tag{2.17}$$

By defining the scattering vector $\vec{S}$

$$\vec{S} = \frac{\hat{u}_{out}}{\lambda} - \frac{\hat{u}_{in}}{\lambda} = \vec{s}_{out} - \vec{s}_{in} \tag{2.18}$$

the diffraction pattern of two electrons van be written as:

$$E(\vec{d}) = E_s(\vec{d}) + E_s(\vec{d}) e^{2\pi i \vec{S} \cdot \vec{r}} \tag{2.19}$$

Implicit in this interpretation is the assumption that the scattered field will not be scattered a second time. This approximation is called the first Born approximation. We assume this to be valid for all objects smaller than a few micrometers.

Important to note is that Δx can be at most $|\vec{r}|$, the distance between the two electrons. Abbe demonstrated that in order to resolve two electrons from each other at least two diffraction orders must be captured. This usually means the zeroth and the first order. The first order start where $\Delta x > \lambda/2$. For two electrons closer than half a wavelength apart this will never be the case. This sets a physical limit to the possible details one can resolve using a regular diffraction set-up. One would not be able to use green light ($\lambda = 500\,nm$) to image objects smaller than 250 nm. For structural studies of molecules this means that in order to distinguish single atoms, X-ray radiation is required.

2.5 Scattering from multiple electrons

The framework for describing the scattering of two electrons can easily be extended to N electrons, considering that every electron scatters independently from the others. The electric field can be described as the sum of all the individual scattered fields.

$$E(\vec{d}) = E_s(\vec{d}) \sum_n e^{2\pi i \vec{S} \cdot \vec{r}_n} \tag{2.20}$$

where n identifies the individual electrons.

In biological particles the incoming EM radiation is scattered by electrons that are part of atoms, instead of being scattered by free electrons. As we will see, this has a major effect on the scattering process. The scattering potential $\rho(\vec{r}, \lambda)$ describes the scattering from a single atom in comparison to the scattering by a free electron. At X-ray wavelengths $\rho(\vec{r}, \lambda)$ can be approximated by:

$$\rho(\lambda, \vec{r}) = -r_e(f_1 + i f_2) \tag{2.21}$$

Here, r_e is the classical electron radius. f_2 is derived from the atomic photoabsorption cross section and is a measure of absorption [12]. f_2 becomes very relevant close to an absorption edge of the material. f_1 describes the scattering power of an atom. It is related to the imaginary part by the Kramers-Kronig dispersion relation [13]. At high photon energies f_1 approaches the atomic number of an atom.

For most elements, the exact values of f_1 and f_2 have been determined experimentally for a wide range of photon-energies [14]. Figure 2.5 presents the scattering factors of two elements: carbon and oxygen. The energy range from 282-533 eV (4.40 nm - 2.33 nm) is called the water window. In this region, and especially towards the higher energies, oxygen atoms and thus water, are scattering significantly less than the carbon atoms that make up organic biomolecules. The water window is therefore a good candidate for imaging cells, as the contrast between two of their main constituents, water and biomolecules, is enhanced, but resolution is limited by the rather long wavelength.

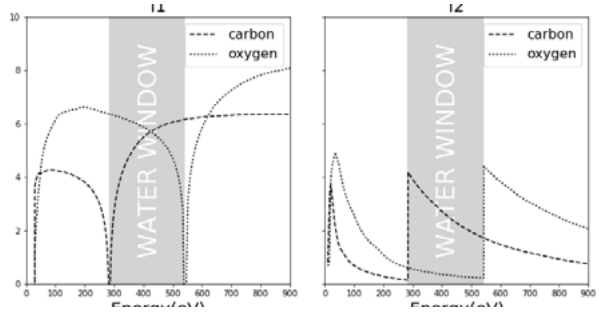


Figure 2.2. Atomic scattering factors f_1 and f_2 of carbon and oxygen, as a function of photon-energy of the X-ray radiation. The region between the absorption K-edge of carbon (282 eV, 4.40 nm) and the K-edge of oxygen (533 eV, 2.33 nm) is called the water window. The contrast between biomolecules and water is enhanced in, especially towards higher photon energies, this region. These wavelengths could be used for imaging living biological particles such as cells or organelles.

The scattered field can now be described as follows:

$$E(\vec{d}) = E_s(\vec{d}) \sum_m \rho(\lambda, \vec{r}) e^{2\pi i \vec{S} \cdot \vec{r_m}} \tag{2.22}$$

where m denotes the different atoms.

If all electrons together are assumed to form a continuous electric charge density around the many nuclei that constitute the biological particle, equation 2.20 can be written as a continuous function.

$$E(\vec{d}) = E_s(\vec{d}) \int \rho(\vec{r}, \lambda) \exp^{-2\pi i \vec{S} \cdot \vec{r}} \, d\vec{r} \tag{2.23}$$

We can now introduce the scattering factor $F(\vec{S})$.

$$F\left(\vec{S}\right) = \frac{E(\vec{d})}{E_s(\vec{d})} = F\left(\frac{\vec{d}}{|d|}\right) \tag{2.24}$$

The structure factor is independent of the distance of the detector, it only depends on the angle of measurement. Instead of structure factor, the term molecular transform is often used in the field of crystallography. Throughout this thesis we will use this term as well.

Equation 2.23 can be rewritten as:

$$F(\vec{S}) = \int \rho(\vec{r},\lambda)\exp^{-2\pi i\vec{S}\cdot\vec{r}}\, d\vec{r} \tag{2.25}$$

This equation is very similar to the well known Fourier transformation $\mathscr{F}(g(t))$ [15].

$$G(\omega) = \int g(t)e^{-2\pi i\omega t}\, dt \tag{2.26}$$

Equation 2.25 is very convenient as we now know that, in the far-field, the scattering of a plane wave is proportional to the Fourier transform of the scattering potential evaluated at $\vec{S}$.

We can now also define the inverse relation of 2.25:

$$\rho(\vec{r}) = \mathscr{F}^{-1}(F(\vec{S})) = \frac{1}{2\pi}\int F(S)\exp^{2\pi i\vec{S}\cdot\vec{r}}\, d\vec{S} \tag{2.27}$$

2.6 The ewald sphere

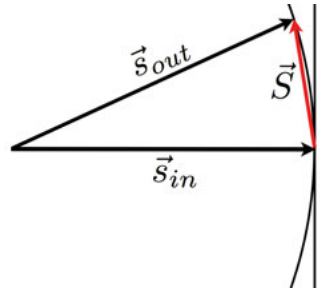


Figure 2.3. A schematic of the Ewald sphere. The scattering vector $\vec{S}$ will always reside on the surface of a sphere of length $1/\lambda$, that intersects with the origin of the three-dimensional molecular transform of the object. $\vec{S} = \vec{s}_{in} - \vec{s}_{out}$. $\vec{s}_{in}$ is constant and is determined by the direction of the incoming radiation. $\vec{s}_{out}$ has a constant length, but varies in direction with scattering angle θ.

The molecular transform $F(\vec{S})$ of a three-dimensional object is also three-dimensional. A diffraction pattern, however, is not. To understand this, we have to evaluate $\vec{S}$ more carefully.

In a diffraction experiment $\vec{S} = \vec{s}_{out} - \vec{s}_{in}$ (see equation 2.18,). $\vec{s}_{in}$ is constant, and is determined by the propagation direction of the incoming EM field. $\vec{s}_{out}$ has a fixed length ($1/\lambda$), and points in the direction of the scattered field (θ). The scattering vector $\vec{S}$ is therefore limited to reside on the surface of a sphere of radius $1/\lambda$. This sphere is called the Ewald sphere [16] (see Figure 2.3).

The Ewald sphere intersects the $F(S)$ through its center $F(0)$. At small scattering angles the curved Ewald sphere can be considered to be flat. This means that a diffraction pattern can be approximated as slice through the center of the molecular transform. This relation will prove itself to be very useful.

2.7 Properties of the Fourier transform

There is a large mathematical field describing the properties of Fourier transformations. A few of these relations are very useful to this thesis.

1. Linearity
   For any complex number a, if $h(t) = ag(t)$, then $H(\omega) = aG(\omega)$
2. Scaling
   For any non-zero real number a, if $h(t) = g(at)$, then $H(\omega) = \frac{1}{|a|} G\left(\frac{\omega}{a}\right)$
3. Translation
   For any real number t_0, if $h(t) = ag(t + \Delta t)$, then $H(\omega) = e^{2\pi i \Delta t \omega} G(\omega)$. The factor $e^{2\pi i \Delta t \omega}$ is called a 'phase-ramp'.
4. Convolution
   The Convolution theorem states that a multiplication in Fourier space is a convolution in object space (real space) and vice versa.

$$f(t) * g(t) = \int f(\tau) g(t - \tau)\, d\tau = \mathscr{F}^{-1}\left(\mathscr{F}(f(t))\, \mathscr{F}(g(t))\right) \qquad (2.28)$$

5. Projection Approximation
   The projection slice theorem states that the Fourier transform of the projection of a three-dimensional function g(t) onto a two-dimensional plane is equal to a two-dimensional slice through the origin of the three-dimensional Fourier transform of the function g(t) which is parallel to the projection plane. At small scattering angles, a diffraction pattern is a slice through the center of the three-dimensional Fourier transform. In this case the results of equation 2.27 can be interpreted as a projection of the scattering potential.
6. Discrete Fourier Transform
   In practice we are dealing with signals that are not continuous, but discrete. Most signal recording is done digitally, and to be able to do numerical computations on the signal, it has to be digitized as well. For

discrete signal the discrete Fourier transform (DFT) and the inverse discrete Fourier transform (IDFT) can be used.

$$DFT(g(T_{k,l,m})) = G(\Omega_{k,l,m}) = \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{m=1}^{M} g(T_{k,l,m})e^{-2\pi i\Omega_{k,l,m}\cdot T_{k,l,m}} \tag{2.29}$$

$$IDFT(G(\Omega_{k,l,m})) = g(T_{i,j,k}) = \frac{1}{KLM}\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{m=1}^{M} G(\Omega_{k,l,m})e^{2\pi i\Omega_{k,l,m}\cdot T_{k,l,m}} \tag{2.30}$$

where K,L and M are the number of 3D pixels (voxels) in each dimensions. $T_{k,l,m}$ and $\Omega_{k,l,m}$ are the discretized version of t and ω.

2.8 Radiation damage

The scattering power of an object is proportional to its volume, and thus varies with the third power of its size. The smaller an object is, the less scattering occurs. To compensate for this loss in scattered signal, the required dose for imaging increases. Practically this means that a dose in excess of 100 MGy is required to record an interpretable signal from a living cell. To put this in perspective, a dose of 2000 Gy is enough to kill practically any organism that live on earth [17, 18]. This means that the amount of energy deposited into the system through photoabsorption is not negligible. The main process of photoabsorption at soft-X-ray wavelengths is photoionisation. Many process follow the ionization event, of which non-radiative Auger decay can lead to significant radiation damage throughout the material. The relation between the required dose of radiation versus the maximum tolerable dose of radiation posed a physical limit to the obtainable resolution when imaging single biomolecules [19].

2.9 Diffraction before destruction

An elegant solution to the problem of radiation damage came from the realization that elastic scattering occurs on a time scale shorter than the damage processes manifest themselves structurally. The principle of diffraction before destruction, suggested in 2000 by Neutze *et al.* [20], was experimentally demonstrated on a silicon-nitride sample [21]. This experiment used the ultra short and extremely bright pulses from the first x-ray soft free-electron laser to outrun key damage processes. Although the pulse literally obliterated the sample, enough structural information was captured to retrieve the original structure.

Diffraction-before-destruction has been further validated by a wide variety of experiments [22, 23]. Serial femtosecond crystallography (SFX) showed that with this principle at least three orders of magnitude higher radiation doses are tolerated compared to regular crystallography, enabling the imaging of much smaller crystals [24]. The imaging of single particles has been shown to be feasible in 2D for a wide variety of biological particles, ranging from *living* cells (**Paper I**) to cell organelles (**Paper VXII**) to 40 nm viruses (**Paper XIII**). **Paper XVI** shows the experimental feasibility of using many 2D images to construct a 3D model of a giant virus.

3. Pulse generation by an XFEL

X-ray Free-electron lasers (XFELs) can generate multi gigawatt and femtosecond (fs) x-ray radiation. Compared to fourth-generation synchrotrons, this is a ten billion-fold increase in brightness. Typically an XFEL consists of two main structures: a linear accelerator, and an undulator. The X-ray radiation is generated in the undulator. This chapter describes the operation of both structures.

3.1 The linear accelerator

At the beginning of the linear acceleration, an electron gun releases short bunches of electrons. The electron bunches are accelerated to relativistic energies, using the electric fields generated by radio frequency (RF) amplifiers named klystrons. Each klystron can maximally add 30 MW/m of energy to the electron bunch. Under higher currents the klystron will tear itself apart. In order to reach sufficiently energetic electron bunches, the linear accelerator is typically between hundreds of meters to close to a kilometre long.

Often accelerators do not operate at the point of maximum acceleration, since one want to introduce a negative energy gradient along the particle bunch that can later be utilized for bunch compression. This acceleration method is called off-crest acceleration, and is visualized in figure 3.1. A negative energy gradient means that the electrons located at the front of the bunch have a slightly lower energy compared to the electrons located at the rear of the bunch.

Bunch Compression

Short and compact bunches are essential for an XFEL. The main bunch compression devices are chicanes. Chicanes consist of four magnetic dipoles that cause the high energy part of the bunch to deflect less than the low energy part (see Figure 3.1). Due to the negative energy gradient it thus reduces the temporal spread of the bunch. After the final compression the electron bunch is 40 fs long. Creating these short bunches is a noteworthy technical achievement.

3.2 Undulator

After the acceleration the electron bunches are sent through a periodic arrangement of magnets with alternating poles called an undulator. The magnetic field

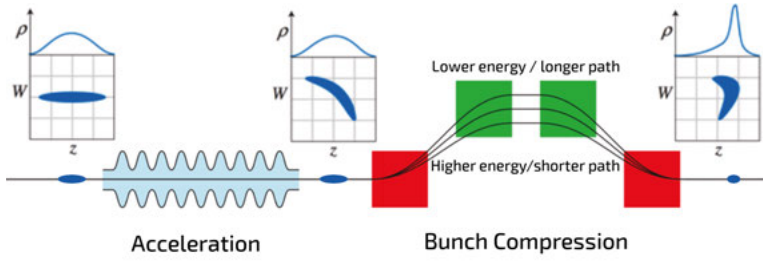


Figure 3.1. Schematic of acceleration and bunch compression at LCLS. Initially the electron bunch has little energy dispersion (W), and is extended along the direction of propagation (z), as illustrated in the left panel. The electron density distribution (ρ) is shown on top of the panel. Throughout the acceleration process the electrons at the rear of the bunch are slightly more accelerated than the electrons at the front of the bunch (shown in the middle panel). This is achieved by using the gradient of the electric field produced in the accelerating units. After the acceleration, the bunch is compressed. The green and red blocks represent magnetic dipoles that deflect the electron bunch in the opposite direction. The path of high energy electrons diverges less than lower energy electrons. As a result of the negative energy dispersion, this results in bunch compression, as shown in the right panel. Most electron density is located in a spike.

in the undulator is given by:

$$\vec{B}(z) = B_0 \cos\left(\frac{2\pi}{\lambda_u} z\right)\hat{\mathbf{y}} \tag{3.1}$$

where B_0 is the magnetic field and λ_u is the length of an undulator period. Due to the magnetic field the moving bunch of electrons experience a Lorentz force, causing it to oscillate in the transverse direction ($\hat{x}$). The movement can be described by the following equation.

$$v_x = \frac{-e B_0 \lambda_u}{2\pi m_e \gamma} \sin\left(\frac{2\pi}{\lambda_u} z\right) = \frac{Kc}{\gamma} \sin\left(k_u z\right) \tag{3.2}$$

The non-dimensional parameter K is given by:

$$K = \frac{-e B_0 \lambda_u}{2\pi m_e c} \approx 0.9337 B_0 \lambda_u$$

where the approximation is valid when the magnetic field is given in Tesla and the undulator period in centimeters.

To preserve momentum during the oscillation, the electrons emit radiation with a wavelength λ_s. As described in Chapter 2, interference effects will occur between the emitted waves from the different electrons. The essence of

this interference phenomenon lies in the longer route taken by the oscillating electrons compared to the radiation, which causes an OPD between the radiation emitted by electrons one undulator period apart. If the OPD between the electrons and the radiation is equal to a multiple of λ_s, the emitted waves will add coherently. Radiation with other wavelengths will interfere destructively. The longer an undulator, the more pronounced will this selection of specific wavelengths be. Equation 3.3 describes this so-called resonance condition.

$$c\frac{\lambda_u}{\langle v_z \rangle} - \lambda_u \cos(\theta) = n\lambda_s \tag{3.3}$$

To predict at which wavelength this happens we need to find the average longitudinal velocity $\langle v_z \rangle$. Since the total speed, v, of the electrons is not affected by the Lorentz force, the longitudinal speed is calculated using pythagoras theorem: $v_z = \sqrt{v^2 - {v_x}^2}$. Using Equation 3.2 for v_x and $\gamma^2 \equiv (1 - \frac{v^2}{c^2})^{-1}$, we can write v_z as:

$$v_z = \sqrt{c^2(1 - \frac{1}{\gamma^2}) - \frac{K^2c^2}{\gamma^2}\sin^2(k_u z)} = c\sqrt{1 - \frac{1}{\gamma^2}(1 - K^2\sin^2(k_u z))}$$

Using the mathematical identity $\frac{1}{\pi}\int_0^\pi \sin^2(x)\,dx = \frac{1}{2}$, integrating of half a undulator period gives the average velocity $\langle v_z \rangle$.

$$\langle v_z \rangle = c\sqrt{1 - \frac{1}{\gamma^2}(1 - \frac{K^2}{2})}$$

Assuming that $\gamma \gg 1$ we can expand the square root using a first order Taylor expansion: $\sqrt{1+x} \approx 1 + \frac{1}{2}x$, where x is small.

$$\langle v_z \rangle \approx c(1 - \frac{1}{2\gamma^2}(1 + \frac{K^2}{2}))$$

If we now assume that we only observe the radiation in the forward direction, we can expand $\cos(\theta) = 1 - \frac{\theta^2}{2}$. Equation 3.3 can thus be written as:

$$n\lambda_s = \lambda_u[\frac{1}{1 - \frac{1}{2\gamma^2}(1 + \frac{K^2}{2})} - (1 - \frac{\theta^2}{2})]$$

We then simplify the first fraction by recognizing that it is given by the geometrical series $\frac{1}{1-x} = 1 + x + x^2 + ...$ By approximating it to the first two terms we are left with the so-called undulator equation.

$$n\lambda_s = \frac{\lambda_u}{2\gamma^2}(1 + \frac{K^2}{2} + (\gamma\theta)^2) \tag{3.4}$$

This formula can be explained by two relativistic effects. For an observer it is the electron bunch that is moving at close to the speed of light. In frame of reference of the electron bunch it is however the undulator that is moving very fast. Fast objects are length contracted, thus the electrons observe an undulator period that is contracted to $\frac{\lambda_u}{\gamma}$, which makes it emit radiation with wavelength around $\frac{c\lambda_u}{\gamma}$. The radiation is in turn relativistically contracted by the relativistic doppler effect when observed in the laboratory's frame of reference, which adds the second gamma term in equation 3.4. The size of the relativistic doppler effect is dependent on the angle of observation relative to the direction of motion. This will cause a concentration of the energy in the forwards direction where the doppler shift is the largest.

The final wavelength of the emitted light is thus dependent on the energy of the electrons (v), the magnetic field in the undulator (B), and the angle of observation (θ). Control over these variables makes the selection of a specific wavelength possible.

3.3 Self-amplified stimulated emission (SASE)

So far in this explanation the electrons produce synchrotron radiation enhanced at specific wavelengths. The intensity of this radiation scales with the number of electrons in the bunch, as the radiation is not temporally coherent. The emitted photons however co-propagate with the relativistic electrons and if the undulator is long enough, the electric field of the radiation will induce an energy modulation within the electron bunch (see Figure 3.2). The resulting microbunches radiate in phase and therefore behave like giant charged particles, and emit photons proportional to the square of their total charge. At wavelengths longer than the bunch length, this radiation is coherent. This phenomenon of self-amplified spontaneous emission (SASE) is exploited to create powerful X-ray free-electron lasers, using very long undulator structures [25, 26].

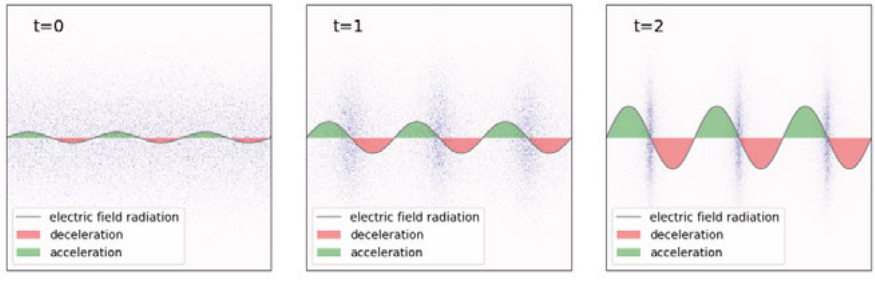


Figure 3.2. Self-amplified stimulated emission (SASE). The electric field of the radiation will induce an energy modulation within the electron bunch, that results in appearance of so-called microbunches, each spaced one wavelength apart. The three panels illustrate the evolution of the electric field of the emitted radiation, and internal modulation of the electron bunch, as the electrons progress in the undulator. The periodic nature of the electric field of the emitted radiation cause certain electrons to be accelerated and others to be decellerated, as indicated by the green and red areas, respectively. This phenomenon is called slippage. As a result, the electrons start to group in microbunches, spaced one wavelength apart. Because the microbunches are spaced an integer number of wavelengths apart, all electrons in the microbunches emit coherently. This amplifies the brilliance of the radiation, which in its turn will amplify the modulation of the electrons. This is called the high-gain mode of an undulator, and it the source of the 10-billion time increase of power of an XFEL, compared to synchrotrons. The amplification will saturate when all electrons are grouped in microbunches.

4. Substrate-free sample delivery

In FXI everything that is illuminated by the X-ray pulse is sample. In the case of weakly scattering single particles scattering from any substrate might drown the signal from the particle itself. Aerosol injection removes this clutter and assures that the sample is clearly isolated from its surroundings, and this, as we see in Chapter 6, is important for image recovery.

An aerosol injector produces small droplets of particles dissolved in a volatile buffer. The buffer evaporates under the reduced pressure inside the experimental chamber, ideally leaving behind only the particle. There are two main types of nozzles that can be used to produce small drops: the Gas Dynamic Virtual Nozzle (GDVN), and electro spray ionisation (ESI).

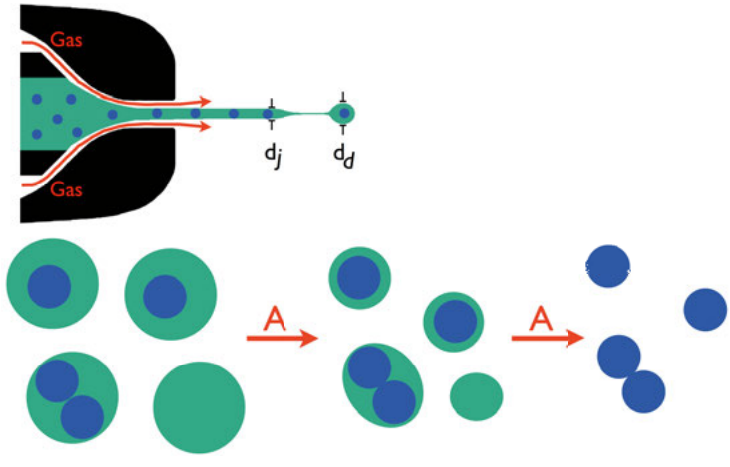


Figure 4.1. Sample delivery using a gas dynamic virtual nozzle (GDVN). The top image illustrates the GDVN. The coaxial flow of a high pressure gas accelerates a liquid jet, which reduces its diameter. If the diameter of the jet is smaller than the characteristic diameter d_j, the surface tension of the liquid will cause the jet to break into small droplets of size d_d. The bottom row of images illustrate the evaporation process. Due to the stochastic nature of the injection process different drops might have a different number of particles (colored blue) in them. In the reduced pressure of the experimental chamber, the volatile medium (colored green) will evaporate, as indicated by the letter A.

4.1 Gas dynamic virtual nozzle (GDVN)

If the coaxial flow of gas creates a jet with a diameter smaller than then a characteristic d_j the surface tension in the liquid will cause the jet to break up into a mist of small droplets. d_j can be estimated on the basis of energy conservation, and is a function of the effective pressure drop ΔP. It is assumed that all energy is transformed to kinetic energy [27].

$$d_j^{(GDVN)} = 2\sqrt{Q \cdot \left(\frac{\rho}{2\pi^2 \Delta P}\right)^{1/2}} \tag{4.1}$$

Q denotes the flow rate and ρ is the density of the liquid. The final drop diameter d_d can be related to the jet diameter using the ratio $d_d/d_j \approx 1.9$, which is the classical Rayleigh breakup. Empirically this assumption has been shown to work well for low-viscosity media such as water [28]. The GDVN can create droplets in the size range of 400 nm - 2000 nm. Sample consumption is in the range of μL/min. Figure 4.1 illustrates the GDVN and the evaporation process.

4.2 Aerodynamic lens stack

After the droplets formed, they start to evaporate in the reduced pressure environment, leaving only the non-volatile particles in the 'drop'. The aerosolised particles are guided into an aerodynamic lens stack (see Figure 4.2). This is a series of cylindrical cavities, connected by co-aligned orifices, that collimate the droplets into a narrow beam of particles [23].

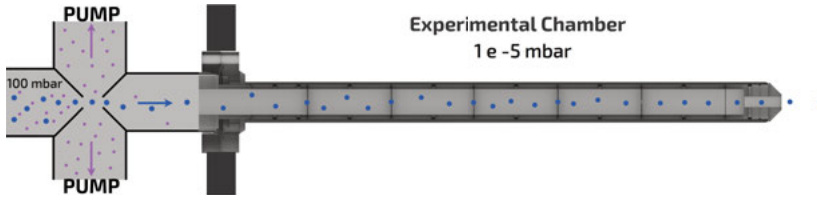


Figure 4.2. Schematic representation of the generation of a narrow particle beam from aerosol using an aerodynamic lense stack. The blue dots represent particles, and the purple dots represent gas molecules. In the first part of the focussing device excess gas molecules are pumped out. This reduces the pressure from ≈100 mbar to 1 mbar. The larger particles are guided through a series of orifices, which focuses them into a narrow stream. In the experimental chamber the pressure is below 10^{-5} mbar.

For robust particles, which are large compared to the drop size, this type of sample injection has proven to be very successful. This sample injection method made it possible to measure the diffraction patterns from isolated single particles at high signal-to-noise ratios at very high repetition rates. Up to 80% hit rates have been recorded with a 120 Hz repetition rates at the LCLS

(**Paper XVII**). The results discussed in this thesis come from datasets that were generated using the GDVN injection method.

For many samples aerosol injection is not disruptive. Molecular dynamics simulation have shown that the conformation of proteins is conserved up till the moment that the last structural water evaporates [29]. For the cyanobacterial cells described in this paper it has been shown that the shape and the autofluorescence properties of the cell membranes of the injected cells remain unchanged (**Paper I**). This is not quite unexpected. Aerosols of cyanobacteria can be carried for long distances, and metabolically active cells have been detected at altitudes of 20-70 km where atmospheric pressure drops to below a millibar [30, 31, 32]. Other sample cell lines such as *E. coli* and brewers yeast, and many types of viruses have been shown to be viable after injection. Nevertheless, not all sample types may be amenable to aerosol sample injection, and samples should be tested prior to experiments.

Electro spray ionisation (ESI)

Over the last years it became apparent that GDVN sample injection does have its limitations. If the particle is small compared to the drop size, wide size distributions of otherwise uniformly sized particles were observed (**Paper XIII**, **PaperXVII**). These observations might be explained by a combination of an incomplete evaporation process, and the build up of a significant shell of debris, originating from impurities present in the solution, around the particle. To purify the measured sample, smaller drops had to be generated. A common aerosolisation technique used in mass spectrometry called electro spray ionisation (ESI) is known to be able to produce small droplets.

ESI nozzles produce droplets through a similar droplet formation process as occurs in the GDVN. The difference lies in the process that drives jet acceleration. In ESI the jet is accelerated by an externally applied electrostatic potential. In the capillary the solvent (volatile buffer) is mixed with negatively charged ions. By applying an external field these ions are accelerated, accelerating the entire jet. The accelerated jet will ultimately break into small drops similar to what occurs in a GDVN. The characteristic d_j^{ESI} can be described as a function of the surface tension σ, electrical conductivity C, as well as the flow rate and the density of the liquid [27].

$$d_j^{(ESI)} \approx 2\sqrt{q \cdot \left(\frac{\rho \varepsilon_0}{\sigma C}\right)^{1/3}} \tag{4.2}$$

After the process of evaporation initiates, an electric potential further builds up in the drops, eventually leading to a Coulomb explosion of the drops [28]. Cycles of evarporation and explosion leads to smaller and smaller drops. Initial experiments showed that drops of mono disperse droplets of size 150-200 nm can be generated with ESI. In the final step op ESI, the charges on the

particles are neutralized. Figure 4.3 shows illustrates the ESI unit and the evaporation process. ESI can also be combined with an aerodynamic lens stack.

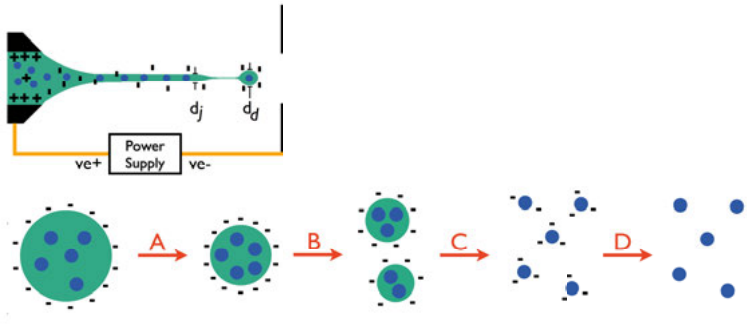


Figure 4.3. Sample delivery using electro spray ionisation (ESI). The top image illustrates ESI. The external electric field generated by a power supply accelerates the negative charges in the liquid jet, which in turn accelerate the liquid jet. This reduces its diameter. If the diameter of the jet is smaller than the characteristic diameter d_j, the surface tension of the liquid will cause the jet to break into small droplets of size d_d. The bottom row of images illustrate the evaporation process. In the reduced pressure of the experimental chamber, the volatile medium (colored green) will evaporate, as indicated by the letter A. Due to evaporation, an electric potential further builds up in the droplets, ecentually leading to a Couloumb explosion of the drops (B). Cycles of evarporation and explosion leads to smaller and smaller drops (C). In the final step the charges on the particles are neutralized (D).

The transfer of ESI to FXI is a considerable achievement that will allow the field to move forward to image smaller particles such as proteins. This would not have been possible using the GDVN aerosolisation. ESI has been shown to also function under lower flowrates, reducing sample consumption to 10 nL/min [33]. This sample consumption rate is advantageous for samples that are only available in small quantities.

Drop-on-demand

The combination of ESI and a quadrupole with an ion trap for storage allows for a pulsed delivery system that can be tuned to match the repetition rate of the XFEL [34]. This is generally referred to as drop-on-demand. This would reduce the sample consumption even further. Moreover the ion trap can also be utilized for sample selection prior to injection of sample into the x-ray pulse. This is important for single proteins as they scatter very little signal, which makes it very difficult to separate the diffraction patterns from the particle-of-interest and those diffraction patterns obtained from contaminating particles or other noise.

5. Data recording

As we have learned in Chapter 2, a diffraction pattern is the Fourier transform of the scattering potential, sampled at the Ewald sphere ($F(\vec{S})$). This Chapter will describe the essentials of the process of measuring the diffracted signal.

The value of $F(\vec{S})$ is complex. The complex amplitude corresponds to the amplitude of the measured wave, and the complex argument corresponds to the phase shift of the wave. Currently no device exists capable of measuring the phase of x-rays directly, as it changes in the attosecond time range. The amplitude of the electron magnetic wave can be determined, and this is exactly what is recorded in an intensity measurement.

$$I(\vec{S}) = |F(\vec{S})|^2 \tag{5.1}$$

where $I(\vec{S})$ is the measured intensity.

In the experiments described in this thesis we used a pnCCD detector to determine the $I(\vec{S})$. A pnCCD is a 2D array of pixels of size p. Each pixel registers the intensity of the electro magnetic wave at that location by converting the energy present in the wave into an electrical current, using a physical process called the photo-electric effect.

Figure 5.1 shows the general setup of a diffraction pattern. There are two pairs of pnCCD detectors. One pair is closed and is placed furthest away from the interaction region. Another detector pair is placed closer to the interaction region, and is opened such that it does not shadow the back detector pair. Figure 5.1 also shows an aerosol injector, as this is the injection method most used in the experiments described in this thesis.

$F(\vec{S})$ is independent from detector distance d, while a diffraction pattern is not. The relative scaling factor between the two is $\frac{1}{\lambda d}$. This means that the area covered by a pixel of size p on the detector covers an area of size $q = \frac{p}{\lambda d}$ of the Ewald sphere (given the small angle approximation holds).

Using the scaling relation between $F(\vec{S})$ sampled at the Ewald sphere, and the measured diffraction pattern at the detector, we can derive that pixel P on the detector corresponds to pixel $\vec{Q} = \frac{\vec{P}}{\lambda d}$ on the Ewald sphere.

Since $F(\vec{S})$ is measured at discrete locations, we have to use the discrete Fourier transform (DFT), and its inverse, to describe the relation between the measured diffraction pattern and the scattering potential. From now on we will use $F(\vec{Q})$ to describe $F(\vec{S})$ measured at discrete positions.

The use of the DFT implies that the scattering potential is also discretized. The pixel size in real space (scattering potential space) r is the inverse of the

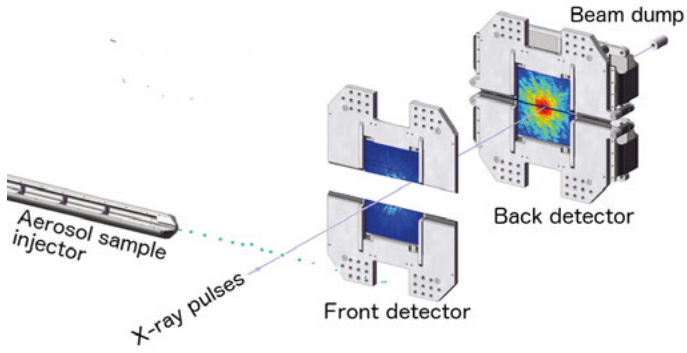


Figure 5.1. The experimental geometry. Sample is introduced into the X-ray pulse train, using an aerosol sample injector. The diffracted signal is recorded on two pairs of detectors. The front detector pair is opened such that it does not shadow the back detector. The direct beam passes trough a central hole in the back detector.

size of the detector in Fourier space (D). $D = nq$, where n in the number of pixels on the detector, q is the size of a pixel on the Ewald sphere. q is the scaled version of a detector pixel:

$$q = \frac{p}{\lambda d} \tag{5.2}$$

This makes r:

$$r = \frac{\lambda d}{np} \tag{5.3}$$

Here λ is the wavelength of the radiation, d the detector distance, and p is the size of a detector pixel.

5.1 Missing data

The geometry of the detectors used in the experiments described in this thesis consist of two moveable halves (38.4 mm by 76.8 mm), with a dead area of at least 0.8 mm in between both halves (see figure 5.1). In the center of the detector a hole is created to let the high-intensity direct beam through (each detector half misses a semicircle). For both the dead area, and the central hole no intensity information is present, and the diffraction pattern is incomplete.

In some experiments two pairs of detectors are used, in which one pair located closer to the sample (see Figure 5.1). The back detector is closed, and the front detector is opened such that it does not shadow the back detector. This opening results in a large area of missing data.

5.2 Saturation

Besides missing data due to the detector geometry, detector saturation might also lead to missing information. Each pixel on the detector can maximally hold a certain amount of electrical charge. If more charge is present in the pixel, the excess charge will overflow to neighbouring pixels, making it impossible to accurately use the original pixel as well as the affected pixels. If the amount of charge is very high, saturation might even damage the detector itself. Saturation usually occurs in the center of the detector, because these regions are typically the most intense regions of a diffraction pattern.

6. Phase retrieval

The problem with missing phase information is well known and is called the phase problem. There are many ways to overcome it: for example in crystallography, homology modelling or the anomalous scattering of heavy atoms is exploited to retrieve phases [35, 36]. In Fourier holography the interference between two wave fields is used to obtain the phases [37, 38, 39]. In ptychography the precisely known overlap and high redundancy between many exposures is used to solve the phase problem [40, 41]. In FXI the scattered field becomes oversampled, which means that, under certain conditions, phases can be retrieved from the intensity pattern itself. The next sections will describe oversampling in more detail, and explain how phases can be retrieved from an oversampled signal using iterative phase retrieval methods, and finally explain how phase retrieval can be validated.

6.1 Oversampling

In 1952 Sayre noticed, the Bragg peaks sample the molecular transform $\mathscr{F}(\rho)$ at the critical sampling rate (Π_C)[42, 43][1]. This means that if we know the phases in addition to the amplitudes at only the Bragg peaks, it is just enough to back-calculate the structure of the measured object. Single-particle imaging is free from the crystal lattice, and we obtain a continuous diffraction pattern. By choosing the detector distance appropriately such that individual pixels cover a small enough angle, we can sample the molecular transform more finely than the critical sampling rate. This sampling condition is called oversampling. Figure 6.1 illustrates both cases of sampling.

If the oversampling is large enough the redundant information given by the extra sampling point is enough to retrieve the missing phases. The linear sampling rate R_S is defined as the ratio between the actual sampling rate and the oversampling rate.

By choosing a detector setup such that the sampling rate is at least twice as high as the [44, 43] the critical sampling rate we can use the additional intensity information to recover the missing phases, and thus reconstruct the object from the measured intensities alone. It has been proven that this method often has multiple solutions when applied to 1D data [45], however, for higher

[1] It was Bernal [44] that already noticed that the signal in between Bragg peaks could be sampled by hydrating and drying crystals.

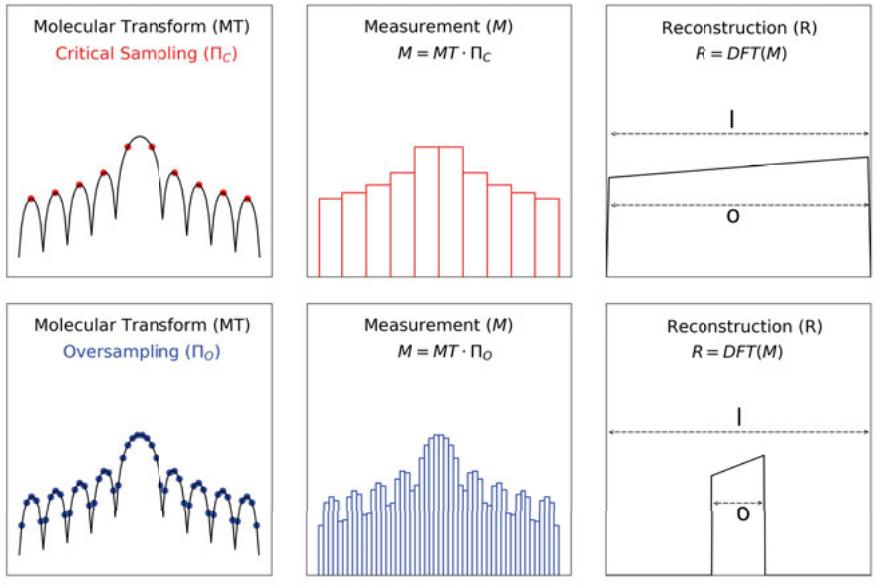


Figure 6.1. Illustration of critical sampling (top row) versus oversampling (bottom row). The right column shows two plots with an 1D signal in black. This signal corresponds to the molecular transform (MT) of our object. The MT is sampled at the critical sampling rate (top left), and an oversampling rate (bottom left). This is indicated by the red and the blue dots, respectively. The middle column shows what the measured signal, associated with the sampling rate, would look like. This bares resemblance to what is measured by a pnCCD. The right most column shows the DFT of both measurements. Important to note is that both DFTs show an similar reconstruction of the original object, but the size of the object compared to the field of view is different. The recovered object is exactly the size of the field of view when the molecular transform is sampled at the critical sampling rate. The size of the object is smaller than the field of view when the molecular transform is oversampled. Oversampling results in zero density around the recovered object.

dimensions it does, in most cases, have a unique solution[46]. Oversampling is the basis of many phase retrieval techniques in single particle imaging, as well as some clever phasing techniques applied to Serial Femtosecond X-ray Crystallography (SFX) [47, 24].

6.2 Iterative phase retrieval

In practice there are many ways to possibly retrieve the phase information from the diffraction pattern, but most common phase retrieval techniques are variations of convex optimization algorithms. This section introduces the general idea behind convex optimization, and explains three different algorithms in more detail.

As with many difficult problems, we start from the things that are known. Figure 6.1 shows that oversampling in Fourier space implies that there is an area around the object with the size $(l-o)$ for which we know the electron density $\rho(\vec{r})$ is zero. This knowledge can be used as a constraint on the possible phases. We know that a correct choice of phases would make the corresponding $\rho(r)$ zero in this area. The area that can contain the sample density and therefore contains nonzero electron density is called the support $M(\vec{r})$. This constraint is called the real-space constraint. Furthermore, we know that the recovered Fourier amplitudes should agree with the measured intensities. This is called the Fourier-space constraint.

In 1978 Fienup [48] introduced an algorithm called Error Reduction (ER) to solve the phase problem. It was inspired by an earlier algorithm by Gerchberg and Saxton [49], but applied to the phase problem. ER is an iterative approach that tries to find a solution that minimizes the disagreement with both the real-space constraint and the Fourier constraint. In words it can be described as follows:

1. Assign random phases to each pixel in Fourier space.
2. Inverse Fourier transform to get the corresponding real space.
3. Set all electron density outside the support to zero, keep the electron density inside the support unchanged.
4. Fourier transform the new density to get the corresponding Fourier space.
5. Make the recovered amplitudes match the measured intensities. Keep the phases unchanged.
6. Go back to step 2.

ER is sometimes able to find the correct solution, but because the problem is not convex it often gets stuck in local minima and does not find the global solution.

In 1984 Levi and Stark realized that the two constraints described above can be interpreted as projections in a multidimensional Hilbert space [50]. Step 3 will from now on be called the real-space projection P_r. The sequence of step 4,5 and 2 will be called the Fourier Projection P_f. In ER P_r and P_f can be defined as follows:

$$P_r\rho(\vec{r}) = \begin{cases} \rho(\vec{r}) & \text{if } \vec{r} \in M \\ 0 & \text{if } \vec{r} \notin M \end{cases} \tag{6.1}$$

$$P_f\rho(\vec{r}) = \mathscr{F}^{-1}\left(\frac{\sqrt{I}}{|\mathscr{F}(\rho(\vec{r}))|}\mathscr{F}(\rho(\vec{r}))\right) \tag{6.2}$$

The largest difference between the two projections is that the support constraint is convex, while the Fourier constraint is not. An iteration in ER can now been seen as a real-space projection followed by a Fourier projection. A common and flexible way to express one iteration of the algorithm is by the

following equation:

$$\rho_{n+1}(\vec{r}) = \begin{cases} P_f \rho_n(\vec{r}) & \text{if } \vec{r} \in M \\ 0 & \text{if } \vec{r} \notin M \end{cases} \tag{6.3}$$

Two error metrics can be constructed by measuring the change imposed by the respective projection: the Fourier error, E_f, and the real space error,E_r. Intuitively, E_r measures how much of the electron density outside the support. E_f gives the difference between the measured and the recovered amplitudes and the square root of the intensities. Mathematically E_f and E_r are defined as:

$$E_r = |P_r \rho(\vec{r}) - \rho(\vec{r})| = \left(\sum_i \rho_i^2 \right)^{\frac{1}{2}} \tag{6.4}$$

$$E_f = \left| P_f \rho(\vec{r}) - \rho(\vec{r}) \right| = \left(\sum_i \left(\frac{\sqrt{I_i}}{|F(S_i)|} - F(S_i) \right) \right)^{\frac{1}{2}} \tag{6.5}$$

The hybrid input output algorithm (HIO)

In 1982 Fienup introduced an algorithm that can escape from local minima, the so-called the Hybrid Input Output algorithm (HIO) [51]. To achieve this HIO makes use of a relaxation parameter called β. An iteration of HIO can be described as follows, using the same syntax as in equation 6.3:

$$\rho_{n+1}(\vec{r}) = \begin{cases} P_f \rho_n(\vec{r}) & \text{if } \vec{r} \in M \\ \rho_n(\vec{r}) - \beta P_f \rho_n(\vec{r}) & \text{if } \vec{r} \notin M \end{cases} \tag{6.6}$$

I consider the β parameter in a similar way as the temperature in simulated annealing. If β is large the algorithm can escape deeper local minima. Unfortunately this also means that the global minima might be escape as well. If β is small HIO will miss fewer minima, but will have greater difficulty escaping from them. As long as a minimum is not perfect (E_r and E_f are both nonzero), HIO will, given enough time, eventually be able to escape.

The relaxed averaged alternating reflections algorithm (RAAR)

Another algorithm that is often used in the work described in this thesis is the Relaxed Averaged Alternating Reflection algorithm (RAAR). RAAR does not escape all minima but it can escape shallower ones. For high-quality data RAAR seems to find the solution quicker and more reliably than HIO. An

iteration of RAAR can be described as follows:

$$\rho_{n+1}(\vec{r}) = \begin{cases} P_f \rho_n(\vec{r}) & \text{if } \vec{r} \in M \text{ and } \rho_n(\vec{r}) \geq -(1+\beta) P_f \rho_n(\vec{r}) \\ \beta\, \rho_n(\vec{r}) - (1-2\beta) P_f \rho_n(\vec{r}) & \text{if otherwise} \end{cases} \tag{6.7}$$

As both RAAR and HIO do not guarantee to end up in the bottom of a minimum, concluding phase recovery with a number of iteration of ER will ensure that the final solution is close to a minimum. This can improve the overall quality of the reconstruction, as shown in **Paper I**.

Other algorithms

There exist many other phase recovery algorithms (See [52]). A software package called Hawk [53] allows users to select and test different algorithms. Hawk is especially powerful as it is fast and gives direct graphical feedback about the reconstruction process. The latter can for example be very useful in determining the correct support size.

6.3 Shrinkwrap

While HIO and RAAR perform well when the support is reasonably well known and follows the shape of the actual object tightly. In practise this information is often not available and phase retrieval can become practically impossible if the support is too large. In 2003 Marchesini developed an algorithm called Shrinkwrap [54] that does not require an *a priori* known support as input, but instead tries to deduce the shape of the support during the reconstruction. After each n iterations the support is updated by applying a Gaussian blur to the real space image and selecting the pixels that have a value above a certain threshold. The general idea is that even with a somewhat inaccurate support, some features will be recovered well, and by using these features the support will become a bit better, which in turn allows for more feautres to be recovered. The algorithm has been very successful for experimental data where the support is usually not known in advance [23].

6.4 Validation

Errors

The most basic method to assess the difference in quality between two reconstructions is by comparing the respective errors. The reconstruction with lower errors is generally believed to be a more correct reconstruction. This method

does however not teach us anything about the biological validity of a reconstruction and is most often only used to exclude clearly failed reconstructions that will have errors that clearly deviate from the successful ones.

The phase retrieval transfer function (PRTF)

A standard tool to assess the quality of a reconstruction is the phase retrieval transfer function (PRTF). This function measures the variation within a set of independent reconstructions from the same data but different random starting phases. The PRTF can then be used to quantify the resolution of a reconstruction. The underlying assumption is that any feature that is reproducibly recovered is likely to be a true feature, but non-reproducible features are artifacts caused by the particular starting phases. There is, however, no proof of this. The variation between reconstructions is calculated for each pixel i in Fourier space by adding together all the recovered amplitudes for that pixel and dividing the total vector by the square root of the measured intensity for that pixel $v_i = |\frac{\sum A_i}{\sqrt{I_i}}|$. If the value of v_i is close to unity, all reconstructions recovered a similar amplitude for that pixel. The closer v_i is to zero the more difference is there between the individual reconstructions. The most common way to present the PRTF is to plot the radial average of v_i. This plot can be interpreted as a measure of reconstruction quality as a function of resolution. A common practice in the field is to quantify the resolution of a reconstruction by the first time the 1D PRTF drops below the, somewhat arbitrary, threshold $1 / e$ (See [23]). The corresponding resolution is given by the inverse of the distance to the origin. The real-space solution is presented as the average of all reconstructions that are included in the PRTF. In this image, non-reproducible features will be averaged out and only the reproducible features will be left.

Missing mode analysis

As described earlier, diffraction patterns often lack data in certain regions. Even in the best case the central region will be missing because the direct beam of an XFEL would otherwise damage the detector. The other main sources of missing data gaps between detector tiles and saturation.

Reconstruction algorithms deal with this missing data by recovering the amplitude as well as the phase for the pixels in these regions. In many cases the missing data does not affect the stability of phase recovery, this is however not true general.

To understand when this happens we have to note that the Fourier constraint does not constrain the amplitudes in the missing data area, and the real-space constraint does not limit the electron density inside the support. If there exists an object that can fit inside the support and has a Fourier transform that is zero outside the missing data region, this object would be completely un-

constrained. Such an object could be arbitrarily added to the solution without changing how well the solution fullfill the constraints and this would mean that there is not unique solution.

In practice completely unconstrained objects do not exist. Objects that fit inside the support and only slightly contribute outside of the missing data region can however exist. We call such objects weakly constrained. In the design of an experiment, or when deciding whether or not to phase an object it is important to predict whether weakly constrained objects exist for that particular data. This analysis is called missing mode analysis and the weakly constrained objects are usually called modes.

The DFT can be represented in matrix form, where each column represents a pixel in real space and each row represents a pixel in Fourier space. Real space and Fourier space will then each be a vector and the transform itself will be a multiplication with the DFT matrix. The elements in the input and output vectors can be rearranged arbitrarily as long as the corresponding rows and columns of the DFT matrix is rearranged in the same way. Here we have rearranged the pixels so that the unconstrained pixels, i.e. the pixels that are in the support or masked out, come first.

$$\mathscr{F} = \left(\begin{array}{c|c} \mathscr{F}_{SM} & \mathscr{F}_{\bar{S}M} \\ \hline \mathscr{F}_{S\bar{M}} & \mathscr{F}_{\bar{S}\bar{M}} \end{array}\right) \tag{6.8}$$

In the matrix above the bar indicates the inverse of the support , S, and mask, M, respectively. We are only looking for objects that only have values in S and minimize the contribution outside the mask. This means that we need to minimize $|\mathscr{F}_{S\bar{M}}\rho|$. A common method to find such objects is singular value decomposition [55]. This method decomposes $|\mathscr{F}_{S\bar{M}}\rho|$ into a diagonal matrix Σ, and two unitary matrices U and V. The values in Σ are called singular values. The smallest singular values correspond to the most weakly constraint objects, or weakly constrained modes.

Hierarchical clustering

As the reconstruction process might give different solutions, we typically use the Fourier and real-space error to select what particles to keep for the PRTF. The assumption here is that failed reconstructions have higher error scores that the successful ones.

I conducted a study to test wether this practice actually selects better reconstructions for a set of reconstructions of living cells. This was done by doing a pixel-to-pixel classification of the reconstructions and checking wether this correlates with the two error metrics. This study was included in **Paper I**.

For the classification we used a version of hierarchial clustering called Unweighted Pair Group Method with Arithmetic Mean (UPGMA)[56]. The workflow of this method is illustrated in figure 6.2.

In the first step of this method, each reconstruction is compared pairwise to every other reconstruction. The comparison score associated with each comparison is the normalized scalar product between the pair of reconstructions, after translating them to their optimal fit.

In the second step of the method similar reconstructions are grouped in clusters. Initially each reconstruction belongs to its own cluster. In every consecutive step the two most similar clusters are merged into one cluster, until only one big cluster remains. The similarity between two clusters is calculated as the average similarity between the reconstructions contained in the clusters. For each merge we log the similarity score and then plot it as a function of the number of clusters. The number of clusters where the plot makes a "kink" signifies that more clusters didn't really help in explaining the data well and therefore this number is assumed to be the number of clusters present in the set of reconstructions. The capability of estimating the number of clusters is an advantage of this particular clustering algorithm.

In a 2D plot of E_f vs. E_r, color-coded by cluster (see figure 6.2), it is possible to see whether or not there is a correlation between reconstruction similarity and error score. In the case where several large clusters remain after applying a real-space error and Fourier-error threshold, the clusters have to be examined carefully. If the remaining clusters correlate with the errors we suggest to keep only the cluster with the lowest error. If no correlation is present between score and cluster it is advised to keep all clusters for further evaluation. Otherwise there is a possibility of selecting on similarity, which would negate the validation power of the PRTF.

6.5 Simulated phase contrast methods

Once the phases of the diffraction pattern has been retrieved, we have complete knowledge of the information encoded in the wave-field. This total knowledge is powerful, because it allows us to emulate any imaging system for which an associated mathematical transform can be written, regardless of its experimentally feasibility. For example, differential interference contrast imaging, sometimes known as Nomarski imaging, is a well established technique in optical microscopy used to visualize changes in phase induced by an object which often have biological meaning, as for example the density inside certain organelles can be higher than the average density in cells. Normarski imaging will give a more pronounced image of the borders of the organelle. For an example see ref [mitochondrium).

Nomarkski imaging in its simplest form takes a scattered wave field of $\psi(x,y,z=0)$, and then interferes this wave-field with a copy of itself that

has been given both a slight transverse displacement (Δx, Δy) and a phase shift ϕ_0. Thus the intensity of the resulting wave-field F_N is [57]: $|\psi(x,y,z=0)+e^{(i\phi_0)}\psi(x-\Delta x,y-\Delta y,z=0)|^2$, from which one can show (using the Fourier shift theorem) that the transfer function becomes:

$$\begin{aligned} T_{DIC}(q_x,q_y,\tau) &= 1+e^{(i(\phi_0-q_x\Delta x-q_y\Delta y))}, \\ \tau &= (\phi_0,\Delta x,\Delta y) \end{aligned} \tag{6.9}$$

In the case of a reconstructed image $\rho(\vec{r})$, the Normarski variant N will look like:

$$N = \mathscr{F}^{-1}T_{DIC}\mathscr{F}(\rho(\vec{r})) \tag{6.10}$$

We have created simulated Nomarski images of the reconstructed images shown in the results part.

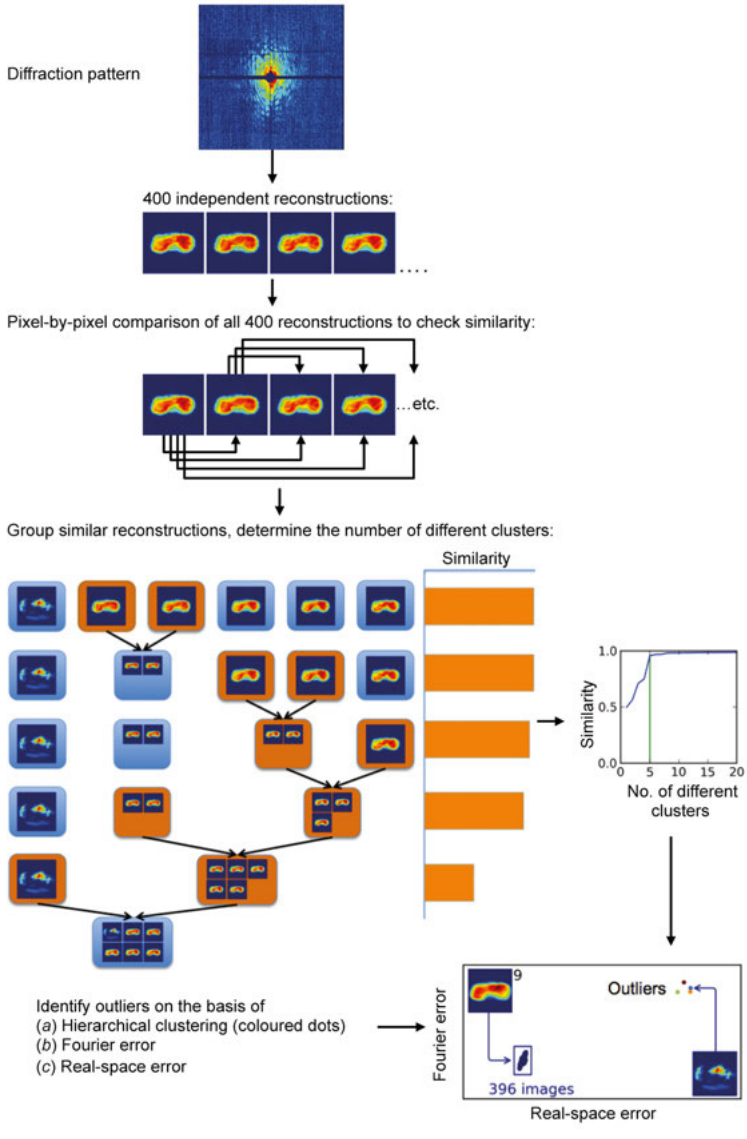


Figure 6.2. Flow chart of the UPGMA clustering algorithm, adopted from **Paper I**. In the first step of the algorithm, n independent reconstructions (400 in this example) are pairwise compared, pixel-by-pixel to check similarity. Each comparison results in a comparison score. Then each reconstruction is given its own cluster. Then a number of merging steps are performed where the two most similar clusters are merged into one. This continues until only one cluster remains. In each merge the similarity score between the clusters is given by averaging the scores calculated previously. We can now plot the similarity score of the last merge against the number of clusters. We use this plot to determine the number of clusters present in the data by choosing the point where this plot makes a kink. This is indicated by the green line in the top graph on the right. The bottom plot shows the Fourier-error and the Real-space error associated to each reconstruction, color-coded on the basis of which cluster they belong to. Here we note that there are four outliers and one main cluster. The inset image showing one of the outliers clearly show that this is a failed reconstruction.

7. Three-dimensional reconstructions

So far we have dealt with two-dimensional diffraction patterns. Although much can be learned from two-dimensional images, a three-dimensional model is essential to understand the function of biomolecules. From Chapter 2 we know that a diffraction pattern samples a curved slice through the center of the molecular transform of the object. If the same object would be illuminated from different angles, the resulting diffraction patterns together could sample the complete three-dimensional Fourier transform. The approach in which multiple two-dimensional images from different angles of illumination are combined into one three-dimensional image is called tomography.

7.1 Reproducible particles

In single particle FXI the illuminated particle is completely destroyed by the XFEL pulse. To date, it has therefore not been possible to image one particle multiple times using FXI. Some bioparticles however, are structurally reproducible, i.e. different particles have identical structures. This means that diffraction patterns originating from different copies sample the same molecular transform, and can thus be combined to form the 3D fourier transform of the object.

In order to successfully assemble the 3D molecular transform from multiple diffraction patterns, their relative orientations need to be known. Most of the time this information is unavailable as the sample delivery methods do not allow for orientation selection. A variety of reconstruction algorithms have been proposed for recovering the relative orientation of diffraction data, including the EMC algorithm [58], the manifold embedding (ME) set of algorithms [59], common-arc algorithm [60], and multi-particle cross-correlation analysis [61, 62, 63]. Theoretical studies suggest that the determination of diffraction pattern orientation should be possible even with photon counts as little as 100 scattered photons per image [58].

7.2 The common arc algorithm

Two diffraction patterns of the same object, will sample the same Fourier space. Because both diffraction patterns will slice the Fourier transform through the origin, both patterns must have at least an arc in common. The common

arc algorithm uses this knowledge to find the relative orientation between two diffraction patterns. In short this algorithm can be described as follows. For each pair of patterns all possible arcs between the two pattern are compared and the best match is chosen as the true one. Based on the retrieved relative orientations between many pairs of patterns a full 3D model can be assembled. This method is successful as long as the diffraction patterns are not very noisy [64]. Too much noise makes the comparison between diffraction patterns unreliable.

7.3 The expansion maximization compression (EMC) algorithm

The expansion maximization compression (EMC) algorithm is an iterative algorithm. In each iteration the measured 2D diffraction patterns (K_j), where j is the index of each pattern, are used to update a model of the molecular transform of the object (M^{MT}). The starting model can be chosen to be random, or if more is known about the model, this information could be incorporated.

Expand

Each iteration starts with the expand step in which the 3D model is expanded in all possible 2D slices throught the center, up to a specified angular accuracy. These slices are denoted W_k, where k is the index of each slice. These slices represent the possible diffraction patterns given model M^{MT}.

Maximization

In the next step, each of the slices are compared to each diffraction pattern. This results in a matrix $R_{j,k}$ that describes how well each diffraction pattern fits in each orientation. The distance metric used for comparing the measured pattern to the predicted pattern often makes use of the noise type that affected the measurement. For instance, if you know that your measurement is only affected by shot noise one could use a Poissonian distance metric:

$$d_{Poisson}(W,K) = \frac{e^{-W_i} W_i^{K_i}}{K_i!} \tag{7.1}$$

Here i indicates the i-th pixel.

If you know your measurement is affected by a source that follows a Gaussian distribution $d_Gaussian$ can be used.

$$d_{Gaussian}(W,K) = e^{-\frac{(W_i-K_i)^2}{2\sigma^2}} \tag{7.2}$$

Here σ is the width of the noise distribution. Each slice in the expanded model is then updated by summing up the measured diffraction patterns weighted by the coefficients from $R_{i,j}$. This will localize the patterns to the orientations where they fit best.

Compression

In the final step a new model M^{MT} is generated by putting back to updated slices W_k in their respective orientations. This will enforce that the slices in the next expanded model will be internally consitent with each other.

It is amazing how much noise EMC can tolerate, and still be able to retrieve orientations. This is true even if the noise model is not accurate. For the success of model assembly, it seems more important for EMC to, initially when the model is far from true, have the option to place diffraction patterns in a wide distributions of orientations, than it is to know the behaviour of the noise exactly.

8. Computational results

8.1 Pattern classification using RedFlamingo

Given the high data rates of XFELs it is impossible for humans to go through all data frames individually, and assistance of computers is needed. I wrote an open-source software framework called RedFlamingo that is designed to rapidly assess individual diffraction patterns by reducing their complexity to a small set of interpretable numbers (https://bitbucket.org/gschot/redflamingo). The power of RedFlamingo lies in its speed and modularity: users have the option to select a combination of algorithms they want to use for data evaluation, and can port their own algorithms. This is important, as each sample and experiment will come with its own set of demands on data analysis. For example the number of scattered photons vary considerably from sample to sample. This will affect the way diffraction patterns can be evaluated. The detector type often changes from experimental station to experimental station, and each experiment has its own specific artefacts due to the experimental condition at the time of operation. Within this variation it is very important to be able to know whether or not you are imaging the particle of interest, at what rate you are collecting usable data, and if the sample is intact. Furthermore, in order for algorithms such as EMC to succeed, the amount of heterogeneity within the selected set of diffraction patterns has to be limited. This chapter will describe the framework of RedFlamingo, as well as a few algorithms that I have developed to assess common sources of heterogeneity.

8.2 The framework of RedFlamingo

Figure 8.1 illustrates the workflow of RedFlamingo. As input RedFlamingo uses a set of diffraction patterns. When the selected algorithms require parameters, these can be provided in a configuration file. This particular workflow made use of eight different algorithms, but algorithms can easily be added or removed from this list.

The algorithms imported into RedFlamingo can use the results from eachother. For example, the results from the signal-distribution algorithm are used for calculating the filtered autocorrelation (see [23]) particle shape evaluation. The filtered autocorrelation is affected by the number of scattered photons.

As output RedFlamingo produces a list of scores for each individual diffraction pattern. These scores can be used for pattern selection. For example, the

scores can be used to select the most intense, but non-saturated diffraction patterns coming from particles larger than 500 nm. The selected patterns can be used by other algorithms such as EMC, or parameters such as size or particle shape can be used in automated phasing.

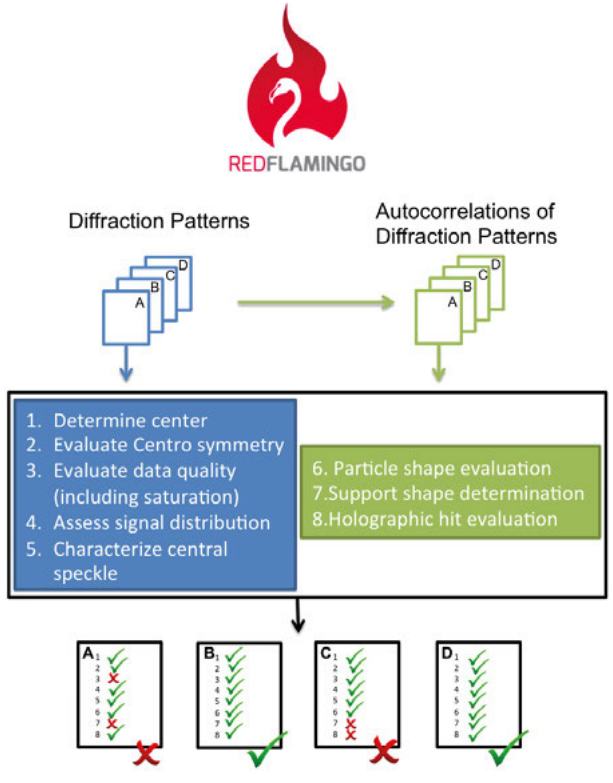


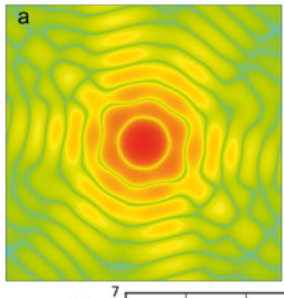
Figure 8.1. The workflow of RedFlamingo. RedFlamingo takes a set of diffraction patterns as input. In this particular workflow eigth different algorithms are selected to assess the diffraction data. Five algorithms evaluate the diffraction patterns themselves, and three are assessing the autocorrelation of the diffraction patterns. As an output RedFlamingo produces a list of scores for each diffraction pattern. Based on the scores, certain patterns can be selected (pattern B and D) or disregarded (pattern A and C).

8.3 Algorithms implemented in RedFlamingo

This section describes several feature extraction algorithms I have implemented, and made available through the RedFlamingo package.

Size

A common method to determine the size of an object is fitting the central speckle to the central speckle of a simulated diffraction pattern from a sphere. Such fitting is reliable for particles that themself are close to spherical in shape, such as icosahedrally-shaped viruses (**Paper XVII**,**Paper XIII**). The second minimum is, however, not reliable, as the location of this minimum depends on the orientation under which the icosahedron is imaged. If the third to the fifth minima are also present in the diffraction pattern, the average of these minima can also be used reliably to determine the size of the object. This method can be especially useful when the central speckle or first minimum is affected by saturation effects. Figure 8.2 shows the relative reliability of using different minima to assess the size of an icosahedral object. These results come from simulated diffraction patterns [65].



c

Minimum #	Average size (nm)	Standard deviation (nm)
first	162.6	0.8
second	162.1	3.0
third	163.2	0.7
fourth	161.7	1.1
fifth	161.5	0.6
Mean of 3rd to 5th.	162.1	0.7

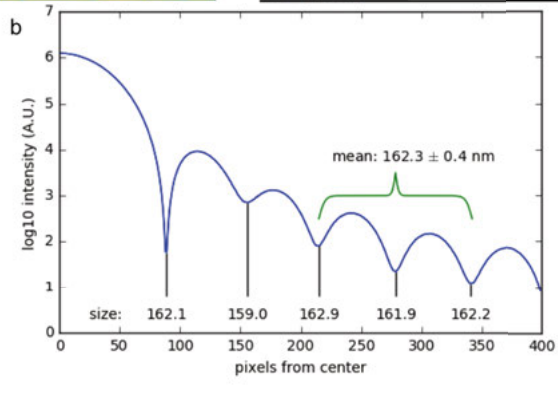


Figure 8.2. Evaluation of the size assessment of a icosahedrally shaped particle of 162 nm, using the location of different minima. a) An example of the simulated diffraction patterns used in this evaluation. Each diffraction pattern has a random orientation with respect to the beam. b) The radial average of the diffraction pattern in a with associated size estimates corresponding to the location of each minimum. c) the average size estimates based on the location of the each minima, and their respective standard deviations. Although the first minimum is the best in determining the size of an object by itself, the mean of the 3rd - 5th minima are also very good in determining the size of the object.

Edge detection

Some objects can be characterized by having sharp edges. A sharp edge in real space corresponds to a more intense speckles in the direction perpendicular to the edge in Fourier space. Objects and/or orientations of objects might be classified by determining if, and how many, edges are present. Figure 8.3 explains the edge finding algorithm, include in RedFlamingo.

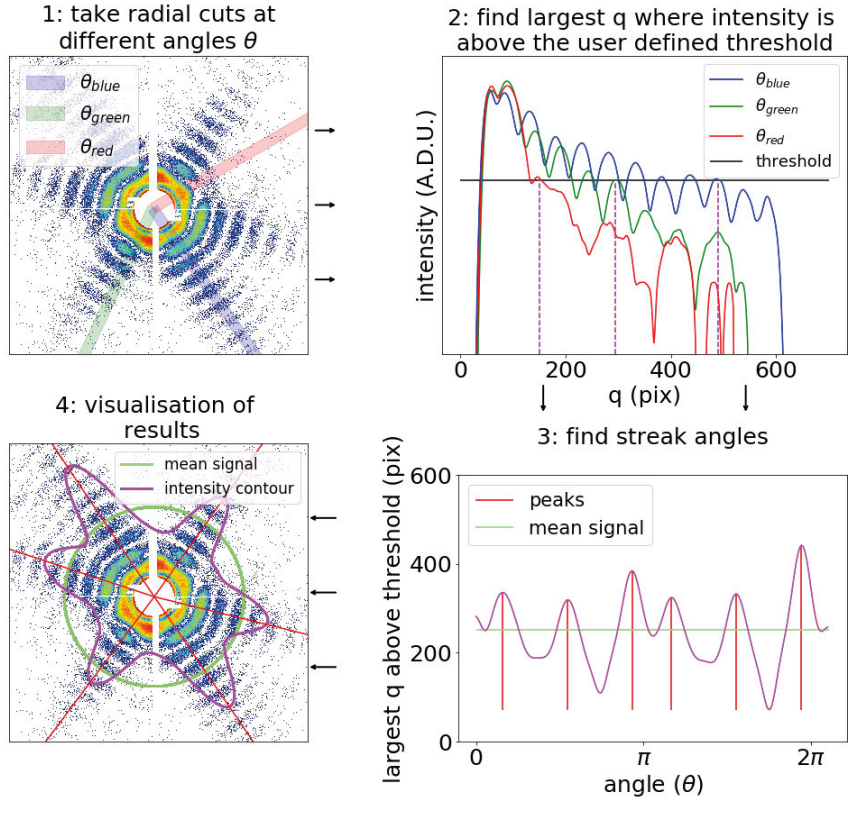


Figure 8.3. An evaluation of where in the diffraction pattern the signal is located. In the first step radial cuts are taken each n degrees. In 1) only three cuts are visualized: the blue cut is on a streak, the green one is on the edge of a streak, and the red one is in between two streaks. In step 2 the largest distance from the center where the intensity is above a user-defined thereshold is determined. These points are indicated by the dashed magenta lines in 2). The three cuts from 1) show a difference in how far the signal extends (E). In step three the maxima in the signal extend are determined (see 3). If a maximum has a 180 degree pair, we consider the pair to originate from an edge. In 3) and 4) the average E is indicated by the green line, the magenta lines contours how far the signal extends, and the red lines indicates the streak location.

Elongation

It can also be possible to determine the elongation of the particle, either by evaluating the elongation of the central speckle (**Paper XVII**,**Paper XIII**) or by evaluating the elongation of the central term in a filtered autocorrelation. This is useful in discerning between diffraction from spherical and non-spherical objects. Figure 8.4 shows two diffraction patterns: A and B. Diffraction pattern A originates from an icosahedral object and thus has a round central speckle (CS_A). The fraction of the shortest distance from the center of the central speckle to the edge of the central speckle (minor) divided by the longest distance is called the elongation ε_{DP}. Round central speckles have an elongation of 1. The central speckle of pattern B (CS_B) is much more elongated.

The evaluation of the autocorrelations of A and B show a similar result. AC_A shows a roundish particle, whereas AC_B shows a density that is more elongated. In the radial average of AC_A and AC_B this becomes visible as a mass beyond the maximum. The green area is considered to be consistent with a round particle, and the red area is considered to be consistent with an elongated particle. The fraction of the green area divided by the green plus the red area constitutes the elongation factor ε_{AC}. These methods can only be used for the evaluation of convex objects, as non-convex object might still appear roundish.

The shape of the particle

An accurate guess for the size of the particle is important for automated phasing. So far automated routines have mainly dealt with icosahedral or round particles, which means that the size of the particle, as determined from the central speckle, is enough to determine an accurate support constraint (**Paper XVII**,**Paper XIII**). The support size and shape of elongated particles such as cells and many virus species cannot be accurately guessed in this way. Figure 8.5 illustrates an algorithm that determines the support size and shape of elongated particles by tracing the contour of the central term in a filtered autocorrelation. The algorithm uses Laplace-based edge detection [66].

Multiple scatterers in the focus

Due to the stochastic nature of the injection method, two particles can end up in the interaction region at the same time. If the particles are attached to each other, a similar pattern to pattern B shown in Figure 8.4 will be measured. If the two particles are, however, separated in space, the scattered signal of the two particles will interfere. As a result so called Newton rings will be observed in the diffraction pattern (see Figure 8.6). As a result of the interference rings, the autocorrelation will show non-central densities, which are called

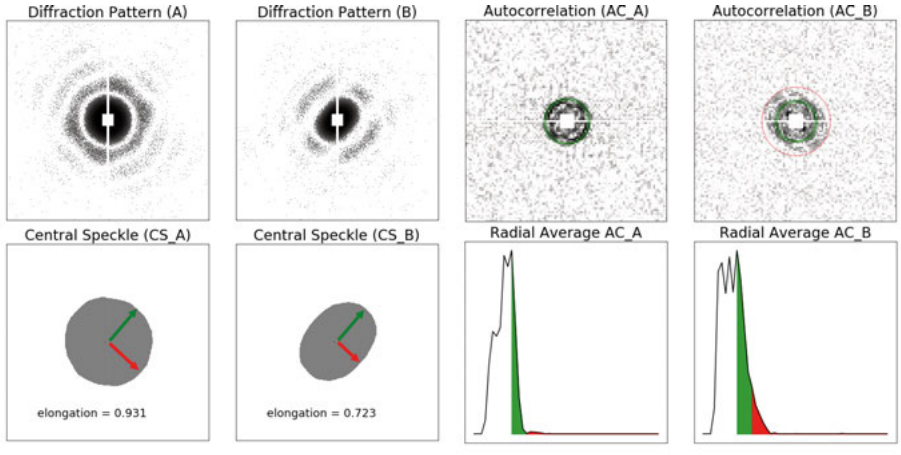


Figure 8.4. Elongation assessment of the particles that gave rise to diffraction patterns A and B, using two different methods. The central speckles of A and B, CS_A and CS_B, have a different shape. The elongation factor ε_{DP} is the fraction of the minor over the major axis of the central speckle. If ε_{DP} is close to 1, the objects that gave rise to the diffraction pattern is considered round in projection. The smaller ε_{DP} is, the more elongated the particle is considered to be. The evaluation of the autocorrelation plots show similar results. The central blob in AC_A is roundish, whereas the central blob is more elongated in AC_B. In the radial averages of both AC_A and AC_B the radius of the object is taken as the calue taken at the maximum. The green density is considered part of a spherical object. The red density is considered part of an elongated object. ε_{AC} is the fraction between the green area over the green area + the red area. The closer this fraction is to 1, the more round the particle is considered.

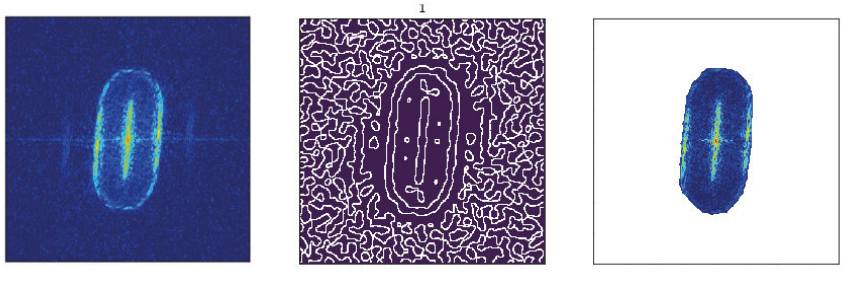


Figure 8.5. Finding the shape of the particle. a) shows the filtered autocorrelation of the particle. b) shows the autocorrelation after applying a laplacean edge detection. c) shows the automatically identified central area within the autocorrelation. It matches the central term of the autocorrelation. This shape can be used to determine the major and minor axis of the particle, and possibly as a guess for the support size and/or shape.

holograms (see **Paper X**). By determining the presence of non-central terms in the autocorrelation, it is possible to automatically separate the patterns that

have interference rings, from patterns that have not. Figure 8.6 illustrates two different methods for finding the non-central density, the first finds the density in the autocorrelation calculated over the entire diffraction pattern. The second method calculates the autocorrelation of a part of the diffraction pattern (similar to a method described in **Paper XVIII**). The latter method requires much stronger signal but avoids artefacts from the missing data.

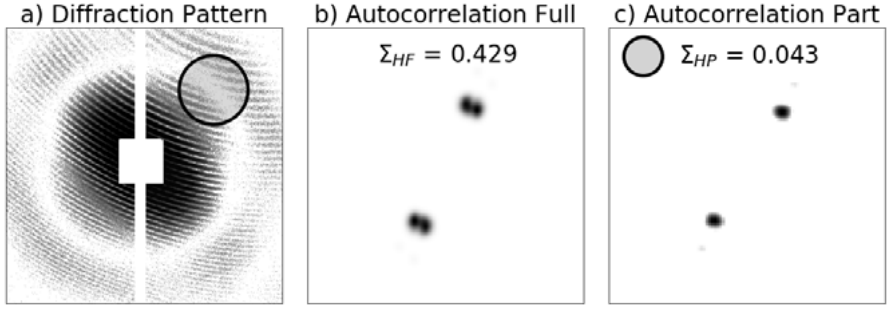


Figure 8.6. Assessing the prescence of multiple particles in the focus, located at a distance from each other. a) shows a diffraction pattern which has clear circular interference rings present. b) shows the autocorrelation of a). The subtraction helps to reduce artefacts coming from the missing data. c) shows the autocorrelation of a small circular area in a. Both methods give two peaks.

9. Experimental results

9.1 Imaging live cells

For the study described in **Paper I** we used cells of two species of cyanobacteria: *Cyanobium gracile* and *Synechococcus elongatus*. Cyanobacteria are photosynthetic bacteria that can be found in almost any habitat on earth, ranging from hot volcanic areas to cold polar ice caps, and play an important role in the global carbon and nitrogen cycle. In a 25 day cycle an algal bloom of 1000 km^2 can sequester around 22,000 tonnes of atmospheric carbon into organic carbon [67].

Cyanobium gracile cells were selected for this experiment because of their small size and their robustness with respect to the injection procedure. Single *C. gracile* and *S. elongatus* cells have an oval-to-cylindrical shape, and vary in size between 0.25-0.4 μm in diameter and 0.4-4.0 μm in length [68]. Cells divide symmetrically by binary fission. The two daughter cells separate from each other after reaching the size and shape of the mother cell [69]. We used non-synchronised cell cultures undergoing active growth, which means our sample contained cells in various stages of their cell cycle.

The experiments described in **Paper I** were carried out at the atomic, molecular, and optical science (AMO) endstation at LCLS [70], at a photon energy of 512 eV (corresponding to a wavelength of 2.40 nm) and 1100 eV (1.13 nm). Figure 5.1 shows the arrangement of the experiment. The length of the photon bunch was about 70 fs. Far-field diffraction patterns were recorded on a pair of pnCCD detectors [71] in the CFEL-ASG Multi Purpose (CAMP) instrument [71]. The detectors were placed at 741 mm downstream from the intersection between the X-ray beam and the stream of sample. The detector read-out rate matched the 120 Hz repetition rate of the LCLS.

We collected diffraction patterns of *C. gracile* cells for an hour at a hit ratio of 43%. The strongest 7,500 hits were selected for further analysis, using the Cheetah software package [72]. The linear sampling ratio of the diffraction patterns were around 20-fold, which allowed phase recovery from the measured intensity patterns. Phase retrieval was not a trivial problem because strong hits saturated the detectors at low diffraction angles. As a compromise, we manually selected medium-strong hits, which contained either no, or only few saturated pixels, while still providing scattered signal to reasonably high resolution. Missing mode analysis revealed no unconstrained modes for all reconstructed cells presented in **Paper I**.

Phases were retrieved using the Hawk software package [53]. For each pattern 400 reconstructions were made, each starting from different random initial phases. These reconstructions consisted of 5000 iterations with the RAAR

algorithm [73], using a Shrinkwrap algorithm [54] for support determination, and concluded with 100 iterations with the ER algorithm [48, 51]. The initial and final support sizes were manually determined. No real-value constraints were used since we anticipated the absorption in the thick cells to give effects similar to a phase object.

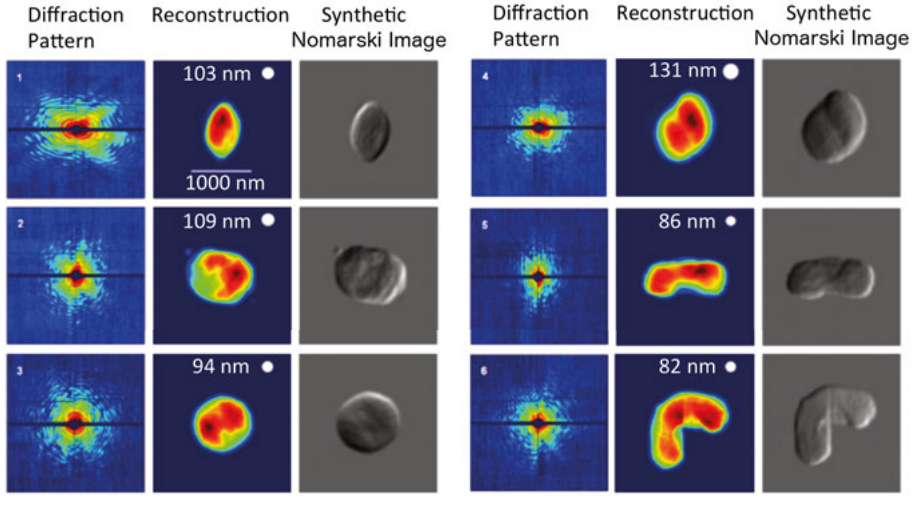


Figure 9.1. Diffraction patterns and reconstructed electron densities for six live C. gracile cells. The cells were alive at the moment the femtosecond pulse traversed them, but exploded within picoseconds after. The photon energy used was 517 eV, which is in the so-called water window. The distance between sample and detector was 740 mm. The total number of scattered photons recorded in the different diffraction patterns varied between 0.5 and 5 million. Each reconstructed image is the average of up to 400 independent reconstructions. The phase retrieval transfer function was used to estimate the resolution of the reconstructed images. White circles in the reconstructions indicate the resolution relative to the size of the sample. Features smaller than the circle should not be interpreted as sample features. Reconstructions are normalized, where dark blue is 0 density and dark red is the most dense part of the cell.The cells have been sorted according to cell size. Synthetic X-ray Nomarski images were calculated from the complex-valued reconstructions to show the reconstructed phase shift properties of the object together with its density.

Figure 9.1 shows the reconstructed exit wave-fronts for six *C. gracile* cells together with the corresponding diffraction patterns, and a synthetic Nomarski image. The reconstructions represent 2D projections of the electron density of the cells. The images show the expected morphologies of cells during division [68, 69]. The resolution of each reconstruction is indicated by the size of the round white dot. This means that features smaller than the dot were not recovery reproducibly and should not be interpreted as features of the sample.

The resolution of the reconstructions were estimated from the PRTF (See Figure 9.2), using the 1/e threshold. Before calculating the PRTF and averag-

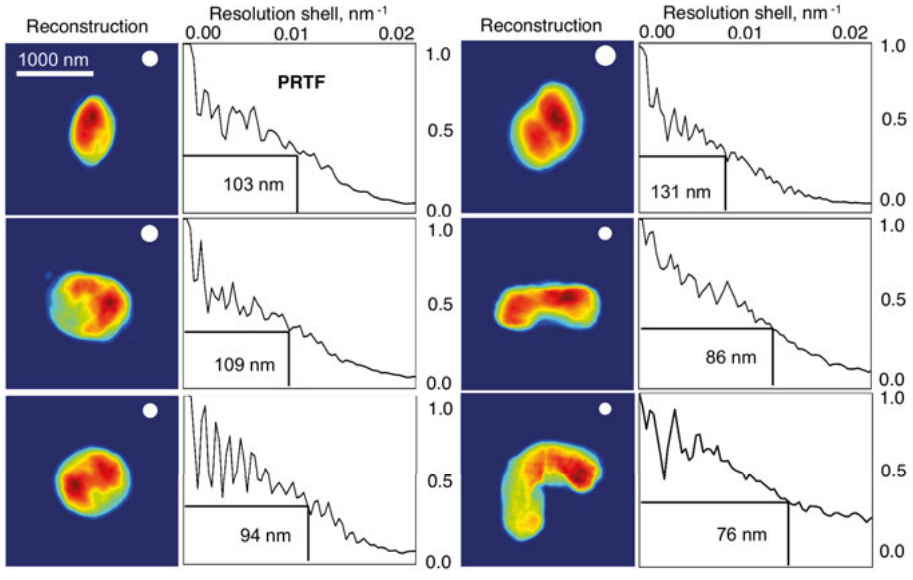


Figure 9.2. Image resolution of the reconstruction. For each reconstruction shown in Fig. 9.1, the corresponding PRTF is shown. Resolution is determined as the first time the PRTF function drops below 1/e [23]. The white dot in the reconstruction has the size of the resolution determined from the PRTF.

ing the repeated reconstructions we removed outliers between the reconstructions by applying a threshold to the Fourier error. Clustering validates the results from using a threshold on the Fourier error and the real-space error. On average the main cluster contained about 370 out of 400 reconstructions (93%), except for one case where only 96 reconstructions formed the biggest cluster (Figure 9.3). That the main cluster often incorporates most reconstructions, and has the lowest error scores made us believe that the average image of the main cluster is likely to represent the best reconstruction minimum.

Detector saturation limited the achievable resolution. A number of much stronger exposures than shown in Figure 9.1 were also recorded, and in some of these exposures the diffraction signal extended to nanometer resolution. Figure 9.4 shows one such pattern for a live *S. elongatus* cell at 1,100 eV photon energy. Two pnCCD detector pairs were used to record this pattern. The configuration of the central back detector in Figure 9.4 is identical to the detector to record the diffraction patterns shown in Figure 9.1. The front detector is the same type as the back detector but is placed at 220 mm from the interaction region and has a much wider gap. In strong exposures, a large part of the back detector was saturated, which prevented reliable phasing. The signal however extended beyond 4 nm resolution on the front detectors. This is the size of a small protein molecule. More than 58 million scattered photons were recorded on the back detectors, and 1.3 million on the front detectors.

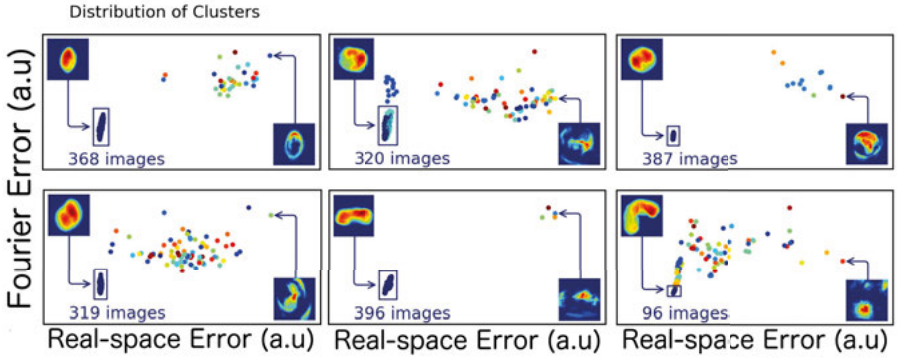


Figure 9.3. Scatter plots of the Fourier vs real-space error for each individual reconstruction for each average reconstruction shown in Fig. 9.1. Each cluster has its own color. The image in the top left corner is the average reconstruction of all the reconstructions selected within the blue box. The threshold for cluster selection has always been the Fourier error, which avoids selecting for similarity.

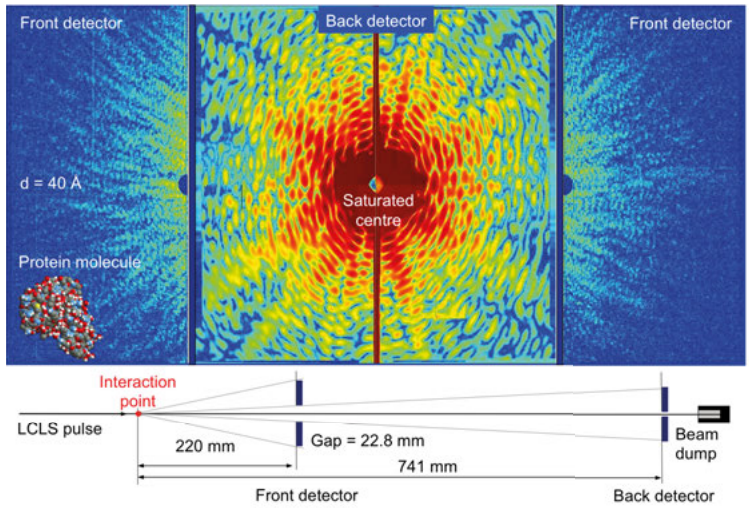


Figure 9.4. Combined diffraction signal of front and back detector. In strong hits the scattered signal from a micron-sized *S. elongatus* cell extends to 4 nm full-period resolution. This is comparable to the size of a small protein molecule. The cell was alive during the exposure. The central area of the diffraction pattern is saturated, which prevented image reconstruction. The bottom image shows the experimental geometry of the front and back detector pairs. The front detector pair is opens such that it does not shadow the back detector. Both detectors have a hole in the center, which allows the direct beam to pass.

9.2 Large data and data deposition

The ability to record millions of diffraction patterns in a day at X-ray free-electron lasers (XFELs) opens up new opportunities for experiments on cells.

The massive amount of data will represent more than just individual projection images of cells. There is however a need to develop algorithms to create abstract models of cells from the large amounts of data. With so many images per day, even statistically rare events could be pinpointed and studied. XFEL beam time is scarce nonetheless, and many researchers have limited access to experimental XFEL data. To aid the development of algorithms that can interpret the wealth of diffraction data we have released the data sets used in the cell study (**Paper II**). We deposited both raw and pre-processed data at the Coherent X-ray Imaging Data Bank (CXIDB) [74] (http://www.cxidb.org/id-37.html), and published a descriptor of the data that includes experimental details, as well as the structure of the deposited data, including the parameters used for data selection.

9.3 RedFlamingo

RedFlamingo has been tested on several datasets. This chapter shows four real-world cases that illustrate different demands on the data analysis. In the first example, RedFlamingo is used for pattern classification. The main discriminatory features were size, particle elongation, and number of particles in the beam. The second example shows that RedFlamingo can be used to size particles within a wide range of sizes and shapes. The third example shows that diffraction patterns from single particles can be separated from diffraction patterns originating from multiple particles located at a distance from each other. The final example shows how the shape of elongated particles can be determined more accurately.

Pattern Classification of an heterogeneous RDV data set

Rice Dwarf Virus (RDV) is the causal agent of rice dwarf disease. It can result in severe crop losses in rice and other gramineae plants in East Asian countries due to stunted growth and chlorotic specks. The structure of RDV has previously been solved to 3.5 Å resolution by X-ray crystallography [75] (PDB 1UF2). The RDV capsid has an icosahedral symmetry, and it is approximately 72 nm in diameter across the 5-fold axis.

As part of a large international collaboration called the Single Particle Imaging initiative[76], an extensive dataset of RDV hits was collected. Most diffraction patterns in this data set are not affected by detector saturation, which makes it possible to estimate the size of the icosahedral-shaped particle by determining the location of the first minimum in the diffraction patterns (see Chapter 8). RDV was classified on the size and shape of the central speckle, and the size and shape of the central term in the autocorrelation. The results of the classification are shown in Figure 9.5. If the central speckle and the central term of the autocorrelation are circular or only slightly elongated the

particle was classified as a single particle. If the central speckle or the central term of the autocorrelation is very elongated, the pattern is classified as cluster. If the autocorrelation has non-central densities, the pattern was classified as multiple.

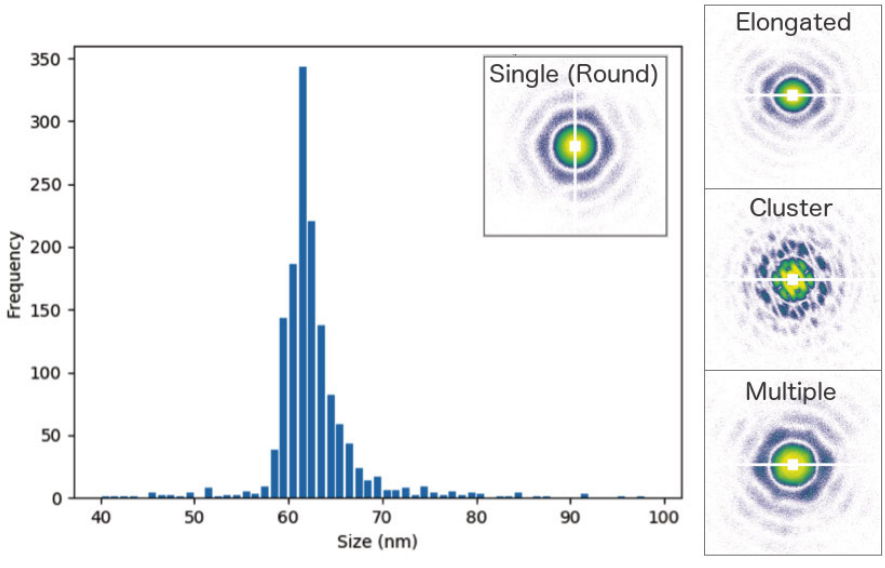


Figure 9.5. Classification of particles. The size distribution of the 3608 diffraction patterns that were classified as single hits shows a peak at 63 nm. Right panels shows an example from a diffraction pattern of each class. The gap between the two detector halves and the beam stop are masked out.

The diffraction patterns that have a round central speckle show features commonly associated with icosahedral particles.

Pattern selection based on edges

The second example deals with the selection of diffraction patterns based on the presence of sharp edges. The sample used was the 331-kbp chlorovirus Paramecium bursaria chlorella virus 1 (PBCV-1). PBCV-1 is the type member of the genus Chlorovirus that infects certain chlorella-like green algae from freshwater sources [77]. It is a quasi-icosahedral particle with a diameter of 190 nm across the 5-fold axis.

In the experiment PBCV particles were injected into the XFEL focus using a GDVN. As expected for particles significantly smaller than the drop size, this resulted in a wide size distribution, of which elongation assessment showed that most particles are very round. Using the edge finding algorithm described in Chapter 8, we could find patterns that show clear features related to edges, as shown in Figure 9.6, separating these patterns from objects with spherical appearance.

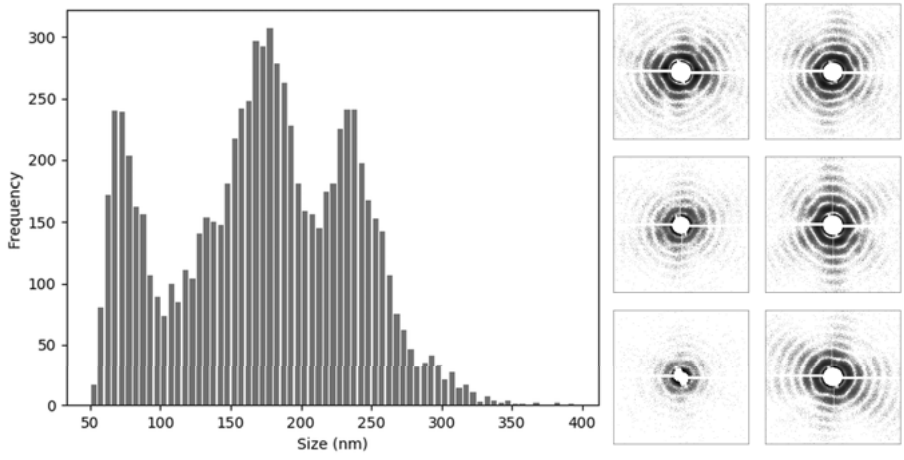


Figure 9.6. Measured particle size-distribution when shooting Paramecium bursaria chlorella virus 1 (PBCV-1) and xenon clusters during an X-ray holography experiment (Paper X). Sizing of hits by RedFlamingo shows a wide size distribution. Particles with a diameter of around 70 nm are likely to be Xenon-clusters. These clusters were co-injected with the PBCV particles. The peak around 170 matches the size of the PBCV particles, and the peak of particle-size 240 might be a contamination by an earlier imaged virus of approximately that size. Most particles were very spherical, but using the edge detection algorithm, patterns orginating from particles with clear edges were selected. Six examples of such patterns are shown on the right.

Particle Shape Assessment

Pithovirus sibericum is a 30,000 year-old giant virus (about 500 nm x 2000 nm) from the Siberian permafrost and its shape resembles a “pithos”, i.e. a large amphora used by the ancient Greeks. We used RedFlamingo for the automated shape assessment of the elongated virus particle. Using the shape assessment algorithm of RedFlamingo illustrated in Figure 8.4, we determined the shape of many particles. Figure 9.7 shows a scatter plot of the maximum and minimum size for each particle. The results show that a large fraction of the particles is roundish in shape. The size variations follow expectations. In the future this methods might find itself useful for the automation of the reconstruction of heterogeneous particles of unknown size, such as cells.

X-ray holography on the fly - Multiple objects in the X-ray focus

I participated in an experiment at the LCLS to capture holographic images of viruses, using two injectors simultaneously to shoot virus particles and reference objects into the X-ray beam. One of these injectors was our aerosol injector, used to inject virus particles into the X-ray focus. The other injector one was a xenon cluster source, used to create a beam of reference objects. The cluster beam intersected the sample beam in the X-ray focus (**Paper X**).

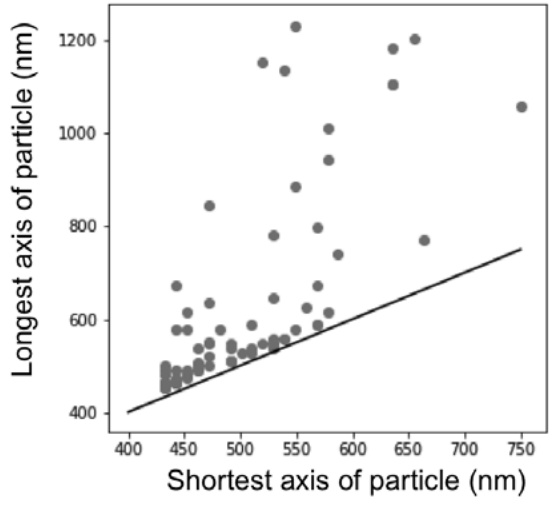


Figure 9.7. Scatter plot of the shortest axis vs. the longest axis of each particle, derived from the filtered autocorrelation.

We used clusters of approximately 70 nm in diameter. A holographic diffraction pattern was recorded when the X-ray pulse simultaneously captured a sample particle and one or more reference objects. Such a pattern is similar to Figure 8.6. During the experiment, some of the hits came from single reference objects or sample particles, others from various combinations of these, including single and multiple sample particles, single and multiple reference objects, or any mixture of sample particles and reference objects. The various types of shots had to be sorted out and this was achieved by RedFlamingo, using the algorithm described in Chapter 8. Table 9.1 shows the results of the classification for samples varying between 70 and 2000 nm in size.

Sample	Size (nm)	Single (#)	Holographic (#)
Pithovirus Sibericum	500 x 2000	257	84
Mimivirus	450	398	43
MelV virus	220	258	136
PBCV1	190	258	133
RDV	72	483	14

Table 9.1. *Identification of multiple particles in the focus. Diffraction patterns from five different samples. The first column shows the name of the sample, the second the sample size. The third and the fourth column indicate the number of images classified as: Single or Holographic. Patterns with strong saturation were automatically excluded from the analysis, due to the potential artefacts in the autocorrelation.*

10. Discussion

In synchrotron X-ray microscopy, the maximal attainable resolution on non-living biological particles is limited to about 10 nm. Unfortunately, synchrotron radiation kills live cells long before any measurable signal can be accumulated, and, as a consequence, no living cell has ever been imaged at high resolution at a synchrotron. 'Diffraction before destruction' overcomes this problem and can give high-resolution data, but it only permits one shot from the sample, corresponding to a spherical section through the Fourier amplitudes of the object. Three-dimensional structure determination is possible for identical objects exposed to the beam one-by-one in different orientations; this however cannot be done easily with non-identical objects, such as cells. Methods have been proposed for the simultaneous illumination of cells from multiple directions to provide a 3D view of the object, and instrumentation to achieve this is under development. However, the number of simultaneous views one can hope for are few.

Although 3D imaging would be highly desirable, high-resolution projections are very valuable. Clinical and research laboratories around the world utilize 2D images of cells, often enhanced by various phase-contrast techniques. This is still possible using diffraction-before-destruction methods, but at a significant higher resolution.

According to predictions, data to sub-nanometer resolution may be recorded from micron-sized living cells using 'diffraction-before-destruction' methods. Physical limits to the resolution in the pattern are related to sample size and composition, pulse duration, pulse intensity, wavelength and and radiation damage to the sample during the exposure. No fundamental limit however has been encountered so far.

When evaluating the resolution we make a distinction between resolution available in the signal and the resolution of the reconstruction. We have recorded data beyond 4 nm resolution, and reconstructed images up to a resolution of 76 nm. These are still the highest resolution recordings and reconstructions of living cells using coherent diffractive imaging.

The diffraction signal from cells decreases as a function of scattering angle with an exponent of about 3.31. Accurate measurement of the signal at nanometer resolutions therefore requires very low background. Container-free sample injection delivers truly isolated samples into the X-ray beam and allow the recording of diffraction patterns with low scattered background. Under these conditions, signal from the sample can be measured above the flat background even at high resolution. The contrast between the sample and its

surroundings (wet helium gas expanding into a vacuum chamber) is high. The clean background and the high contrast are important for the finite support constraint in phase retrieval.

The resolution in a 2D reconstruction from a single exposure depends on the success of the phase retrieval, and is also reduced by the lift-off of the Ewald sphere from the projection plane at high angles (shorter wavelengths would alleviate this problem). The projection approximation also presumes that the Born approximation is valid. This means that for samples thicker than 1-2 μm, harder X-rays than those used here will be required.

The maximal size of an object for successful reconstruction is currently limited to about 1-2 micron at the LCLS for a number of reasons. First, the bandwidth of an LCLS pulse of about 0.2% will put an upper boundon the number of resolution elements to about 250. If a target resolution of 4 nm is aimed for, the object size cannot be bigger than about 1 micron. Smaller bandwidth would allow studies for larger objects at higher resolutions. Second, the focus must be large enough to cover the sample yet contain enough photons to produce a strong scattered signal from the cell. A larger focus requires more photons per pulse and these extra photons are currently not available from the LCLS. This limits the maximal useful focus size, which in turn limits the object size to about 1-2 micron.

Missing low-resolution data pose another large problem in image reconstruction. Low-resolution terms are crucial for the determination of the object support. The X-ray detector has a hole at its centre to let the direct beam pass through. The size of this blind spot limits the maximal object size to 1-2 μm at the relevant wavelengths. In strong exposures, there is a further and significant loss of low-resolution data due to detector saturation.

The current limitations are technical. A femtosecond exposure 'freezes' all cellular processes at room temperature, including diffusion, and thus eliminates blurring through particle motion. This is an advantage over other cell-imaging methods and will become important if or when nanometer / sub-nanometer resolution will be achieved on micron-sized cells. At the moment, we still need stronger pulses that what is currently available to make this a reality. This is illustrated in Figure 10.1. The European XFEL could provide the stronger and shorter pulses that are needed to image cells at nanometer resolution.

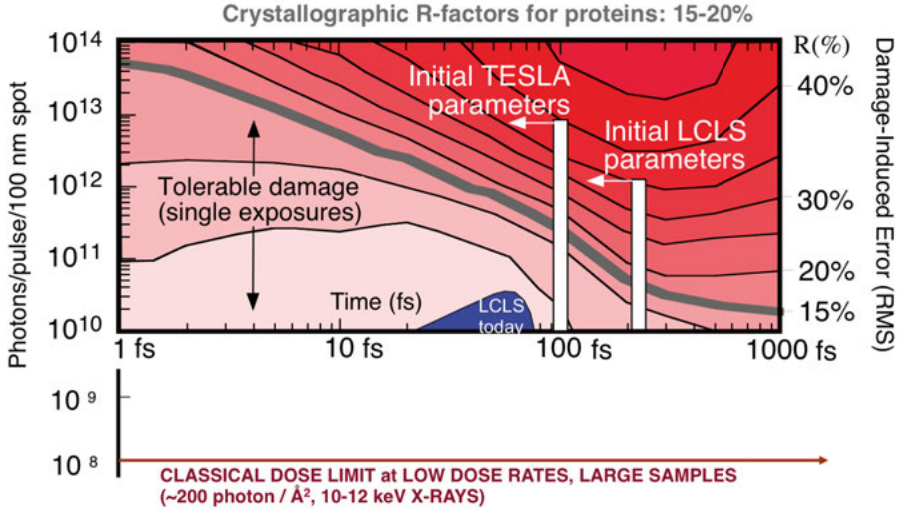


Figure 10.1. The landscape of damage tolerance. Contour plots of the weighted electronic R factors [from equation (2) in [20]] as functions of the X-ray flux and pulse duration at 12 keV photon energy. The plot illustrates the extent to which the information content of the elastically scattered X-rays is degraded by radiation damage. Damage as regarded acceptable if the R-factor is below 15%, indicated by the grey line. Photon pulses from the LCLS are weak, and have not reached the grey line anywhere in the parameter space.

11. Sammanfattning på Svenska

För att verkligen förstå hur en cell fungerar, måste man studera celler som är vid liv. Ändå bedrivs mycket forskning på celler som är nedfrusna eller har dött av strålskador. I denna avhandling presenterar jag ett steg på vägen mot att studera molekylära maskiner i deras naturliga miljö inuti cellen. Detta kan på sikt lära oss mer om vitt skilda fenomen som virusinfektioner, celldelning, fotosyntes och andra processer som är viktiga både för biologin och människors hälsa och miljö.

Den detaljnivå på vilken vi kan avbilda ett objekt är begränsad till ungefär våglängden hos det ljus som används för avbildningen. De minsta byggstenarna i celler är atomer och för att kunna se cellerna i fullständig detalj behövs alltså ljus med en matchande våglängd. Denna våglängd är ca 0.0000000001 meter, vilket även kallas för 1 Ångström och ljus med denna våglängd kallas röntgenstrålar.

Ju mindre ett objekt är desto starkare ljus behövs för att samla in tillräckligt med information för att skapa en tydlig bild av objektet. Stark strålning har dock nackdelen att den skadar det föremål som undersöks. Detta medför att objekt som är mindre än 100 Ångström inte kan avbildas med vanliga tekniker utan kommer att förstöras innan en bild har kunnat framställas.

En ny typ av röntgenkälla, kallad röntgenfrielektronlaser, kan producera mycket korta och extremt starka röntgenpulser och kan möjliggöra ett sätt att undvika provskador. Pulsens styrka gör att även enskilda molekylära maskiner kan avbildas i detalj på atomnivå. Pulserna skadar förvisso provet men eftersom de är extremt korta så uppstår skadorna först efter att pulsen har passerat provet. Bilden föreställer därför det oskadade provet. Denna princip brukar kallas för diffraktion före destruktion.

De bilder som vi pratar om här är inte framställda på samma sätt som med en vanlig kamera. Istället uppmäter vi ett så kallat diffraktionsmönster vilket beskriver hur ljuset sprids efter att det har träffat provet. Detta diffraktionsmönster kan sen analyseras för att återskapa en 2D bild av provet. Att denna typ av experiment är möjliga att genomföra har tidigare visats för en mängd olika prover.

År 2008 förutsa forskare att det i teorin är möjligt att använda metoden för att ta 2D bilder av små levande celler med en upplösning på 10 Ångström. Huvudarbetet i den här avhandlingen är en experimentell verifiering av denna förutsägelse. Vi sprayade in levande cyanobakterier i röntgenpulsen och samlade över 100 000 bilder på levande celler under endast en timme. Trots att denna laser var betydligt svagare än den som användes i forskarnas förutsägelse kunde vi se diffraktion till en upplösning på 40 Ångström.

Problemet med dessa bilder var att signalen var för stark för att hanteras av röntgendetektorn. Vi var därför tvungna att nöja oss med att endast analysera svagare diffraktionsmönster som därmed gav en sämre upplösning. Därför har vi designat ett förbättrat experiment där vi med hjälp av ett filter framför vissa delar av detektorn skulle undvika de tidigare problemen. Denna uppställning har dock inte kunnat testas experimentellt ännu då vi väntar på ett tillfälle att använda någon av de få röntgenfrielektronlasrar som finns i världen.

Eftersom vi kan samla in många tusentals diffraktionsmönster från celler i varje experiment så är det viktigt att inte behöva ägna tid åt varje enskild analys av ett diffraktionsmönster. Att automatisera processen är dock svårt eftersom cellers storlek och form har en mycket stor naturlig variation. Utan denna automatisering är det dock väldigt svårt att göra några som helst studier som sammanför information från mer än en handfull celler. För att underlätta denna utveckling så har vi upplåtit våra diffraktionsdata så att andra, som inte har haft möjlighet att göra egna experiment ändå kan bidra till utvecklingen av analysmetoder.

För att automatiseringen ska lyckas är det viktigt att kunna förutsäga cellens storlek och ungefärliga form direkt från diffraktionsmönstret. Dessutom måste man även kunna identifiera andra aspekter, som detektormättnad eller om flera celler befann sig i röntgenstrålen samtidigt. För att lösa dessa problem har jag utvecklat programmet RedFlamingo. RedFlamingo kan snabbt och automatiskt bedöma kvaliteten hos ett diffraktionsmönster och beräkna flera viktiga parametrar som sen kan användas i den vidare analysen. RedFlamingo kan även användas för att separera diffraktionsmönster som kommer från den partikel vi vill studera från diffraktion från andra partiklar som kan ha hamnat i röntgenstrålen.

Jag är väldigt entusiastisk över att vara en del av utvecklingen av denna teknik. Vi har hittills inte stött på någon fundamental gräns för hur långt tekniken kan nå. De nuvarande begränsningarna är tekniska och kan komma att undanröjas med tillgång till starkare och kortare röntgenpulser. Nästa generations röntgenfrielektronlasrar kommer troligtvis att erbjuda just detta och kanske ge oss möjligheten att studera cellers byggstenar med atomär upplösning på plats i cellerna själva.

Acknowledgements

Science is team effort and I want to thank a very large international network of researchers for many ideas and a wonderful time to work on them.

In particular I want to thank my supervisor Tomas. Thank you for your endless patience, clarity of explanations and deep and quick insight into a wide variety of material. You have taught me much on what it entails to communicate an idea, and how to go for one. Good luck with your grant and I am looking forward to your results on conformational heterogeneity. And Janos: thank you for offering me a position in your group. Your vision and your boundless passion to make anything a reality are very inspirational. You have really given me many opportunities to develop!

Ida, thank you for being a great office mate and a kindred spirit. May Stockholm treat you and Federico well. Kenta, thanks for your joy for pathogenic viruses, and for introducing me to the wonderful Japanese culture. This was fun. Martin, Daniel L, Hemanth, Dirk, Gunilla, Anna S-L, Anna M, Margaretha, Karin, kiwi Laura, and Inger: without biology imaging would be rather empty. Daniel W., Marvin, Kerstin, Johan, Carl, Alberto, Jonas, Max, Benedikt, Filipe, Bianca, and Jakob and all former lab members for the many fun breaks, ski trips and for sharing a passion for a wide variety of ideas. It was very much fun to spend five years and many world trips together. I want to thank all international collaborators including Akifumi-san, Prof Nagakawa, Duane, Andrew M., Anatoli, Tais, Andrew, Olena, and Leonie for a wide variety of hardcore experiments as well as the surrounding social activities. I want to thank Prof Bonvin to giving me my first opportunity to explore structural biology, and Prof Förster for offering a great opportunity and very smooth transition into a new job.

Privately I want to thank Bart, Brina, Fay, Roeland, Laurens, Mijke, Bram, Marsela, Sara, Floris, Dien, Frerk, Mariana, Gabriel, and Warren for being great friends. Jos, Ina en Femke: het is fijn om zo een diepe liefde als basis te hebben. Emmanuel and Mathis: thanks for just being there. Sharing time with two wonderful sons is my greatest happiness. Syrinx: you have enriched my life and gave me the confidence to explore the end of reason.

Author Contribution

Paper I

I performed the data analysis, excluding the preprocessing. That includes the pattern selection, phase recovery, missing mode analysis, and the clustering analysis.

Paper II

I performed the data analysis, including the preprocessing and the pattern selection.

Paper III

I performed the data analysis, including the preprocessing, hit selection, and pattern classification. For the latter we used a software package I wrote.

References

[1] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution.," *Science (New York, N.Y.)*, vol. 313, pp. 1642–5, sep 2006.

[2] P. J. Sapienza and A. L. Lee, "Using NMR to study fast dynamics in proteins: methods and applications.," *Current opinion in pharmacology*, vol. 10, pp. 723–30, dec 2010.

[3] Y. Ye, X. Liu, Z. Zhang, Q. Wu, B. Jiang, L. Jiang, X. Zhang, M. Liu, G. J. Pielak, and C. Li, "(19) F NMR spectroscopy as a probe of cytoplasmic viscosity and weak protein interactions in living cells.," *Chemistry (Weinheim an der Bergstrasse, Germany)*, vol. 19, pp. 12705–10, sep 2013.

[4] M. Marko, C. Hsieh, R. Schalek, J. Frank, and C. Mannella, "Focused-ion-beam thinning of frozen-hydrated biological specimens for cryo-electron microscopy," *Nature Methods*, vol. 4, pp. 215–217, mar 2007.

[5] F. K. M. Schur, M. Obr, W. J. H. Hagen, W. Wan, A. J. Jakobi, J. M. Kirkpatrick, C. Sachse, H.-G. Kräusslich, and J. A. G. Briggs, "An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation," *Science*, vol. 353, pp. 506–508, jul 2016.

[6] M. Bergh, G. Huldt, N. Tîmneanu, F. R. N. C. Maia, and J. Hajdu, "Feasibility of imaging living cells at subnanometer resolutions by ultrafast X-ray diffraction," *Quarterly Reviews of Biophysics*, vol. 41, p. 181, nov 2008.

[7] M. Altarelli and Deutsches Elektronen-Synchrotron (Center). X-Ray Free-Electron Laser Project., *XFEL : the European X-ray free-electron laser : technical design report*. DESY XFEL Project Group, 2006.

[8] F. Stelluti and F. Cesi, "Apiarium," 1625.

[9] A. van Leeuwenhoek, "Letter to Oldenburg," 1676.

[10] K. D. Froome, L. Essen, and R. A. Rhodes, "The Velocity of Light And Radio Waves," *Physics Today*, vol. 24, p. 49, may 1971.

[11] P. Persson, S. Lunell, A. Szöke, B. Ziaja, and J. Hajdu, "Shake-up and shake-off excitations with associated electron losses in X-ray studies of proteins," *Protein Science*, vol. 10, pp. 2480–2484, mar 2001.

[12] "X-Ray Data Booklet."

[13] R. de L. Kronig, "On the Theory of Dispersion of X-Rays," *Journal of the Optical Society of America*, vol. 12, p. 547, jun 1926.

[14] B. Henke, E. Gullikson, and J. Davis, "X-Ray Interactions: Photoabsorption, Scattering, Transmission, and Reflection at E = 50-30,000 eV, Z = 1-92," *Atomic Data and Nuclear Data Tables*, vol. 54, pp. 181–342, jul 1993.

[15] A. Vretblad, "Graduate texts in mathematics," *Fourier Analysis and its applications*, 2003.

[16] P. Ewald, “Introduction to dynamical theory of x-ray diffraction,” *Acta Crystallographica Section A - Crystal Physics Diffraction Theoretical And General Crystallography*, vol. A 25 (Part, p. 103, 1969.

[17] B. P. Fairand, *Radiation sterilization for health care products : x-ray, gamma, and electron beam*. CRC Press, 2002.

[18] J. C. Game, M. Williamson, and C. Baccari, “X-Ray Survival Characteristics and Genetic Analysis for Nine Saccharomyces Deletion Mutants That Show Altered Radiation Sensitivity,” *Genetics*, vol. 169, pp. 51–63, oct 2005.

[19] M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. M. Holton, C. J. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D. A. Shapiro, J. C. H. Spence, and D. Starodub, “An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy,” *Journal of Electron Spectroscopy and Related Phenomena*, vol. 170, pp. 4–12, 2009.

[20] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, “Potential for biomolecular imaging with femtosecond X-ray pulses,” *Nature*, vol. 406, pp. 752–757, aug 2000.

[21] H. N. Chapman, A. Barty, M. J. Bogan, S. Boutet, M. Frank, S. P. Hau-Riege, S. Marchesini, B. W. Woods, S. Bajt, W. H. Benner, R. A. London, E. Plönjes, M. Kuhlmann, R. Treusch, S. Düsterer, T. Tschentscher, J. R. Schneider, E. Spiller, T. Möller, C. Bostedt, M. Hoener, D. A. Shapiro, K. O. Hodgson, D. van der Spoel, F. Burmeister, M. Bergh, C. Caleman, G. Huldt, M. M. Seibert, F. R. N. C. Maia, R. W. Lee, A. Szöke, N. Timneanu, and J. Hajdu, “Femtosecond diffractive imaging with a soft-X-ray free-electron laser,” *Nature Physics*, vol. 2, pp. 839–843, dec 2006.

[22] M. Seibert, S. Boutet, M. Svenda, T. Ekeberg, F. R. N. C. Maia, M. J. Bogan, N. Tîmneanu, A. Barty, S. Hau-Riege, C. Caleman, M. Frank, H. Benner, J. Y. Lee, S. Marchesini, J. W. Shaevitz, D. A. Fletcher, S. Bajt, I. Andersson, H. N. Chapman, and J. Hajdu, “Femtosecond diffractive imaging of biological cells,” *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 43, p. 194015, oct 2010.

[23] M. M. Seibert, T. Ekeberg, F. R. N. C. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, D. Westphal, M. Hantke, D. P. DePonte, A. Barty, J. Schulz, L. Gumprecht, N. Coppola, A. Aquila, M. Liang, T. A. White, A. Martin, C. Caleman, S. Stern, C. Abergel, V. Seltzer, J.-M. Claverie, C. Bostedt, J. D. Bozek, S. Boutet, A. A. Miahnahri, M. Messerschmidt, J. Krzywinski, G. Williams, K. O. Hodgson, M. J. Bogan, C. Y. Hampton, R. G. Sierra, D. Starodub, I. Andersson, S. Bajt, M. Barthelmess, J. C. H. Spence, P. Fromme, U. Weierstall, R. Kirian, M. Hunter, R. B. Doak, S. Marchesini, S. P. Hau-Riege, M. Frank, R. L. Shoeman, L. Lomb, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, C. Schmidt, L. Foucar, N. Kimmel, P. Holl, B. Rudek, B. Erk, A. Hömke, C. Reich, D. Pietschner, G. Weidenspointner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, I. Schlichting, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kühnel, R. Andritschke, C.-D. Schröter, F. Krasniqi, M. Bott, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, H. N. Chapman, and J. Hajdu, “Single mimivirus particles intercepted and imaged with an X-ray laser,” *Nature*, vol. 470, pp. 78–81, feb

2011.
[24] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.-D. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and J. C. H. Spence, "Femtosecond X-ray protein nanocrystallography," *Nature*, vol. 470, pp. 73–77, feb 2011.
[25] A. M. Kondratenko and E. L. Saldin, "GENERATION OF COHERENT RADIATION BY A RELATIVISTIC ELECTRON BEAM IN AN ONDULATOR*," *Particle Accelerators*, vol. 10, pp. 207–216, 1980.
[26] R. Bonifacio, C. Pellegrini, and L. Narducci, "Collective instabilities and high-gain regime in a free electron laser," *Optics Communications*, vol. 50, pp. 373–378, jul 1984.
[27] A. J. Acero, N. Rebollo-Muñoz, J. M. Montanero, A. M. Gañán-Calvo, and E. J. Vega, "A new flow focusing technique to produce very thin jets," *Journal of Micromechanics and Microengineering*, vol. 23, p. 065009, jun 2013.
[28] R. Cole, *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*. New York, USA: Wiley, 1997.
[29] E. G. Marklund, D. S. D. Larsson, D. van der Spoel, A. Patriksson, and C. Caleman, "Structural stability of electrosprayed proteins: temperature and hydration effects," *Physical Chemistry Chemical Physics*, vol. 11, p. 8069, sep 2009.
[30] A. A. Imshenetsky, S. V. Lysenko, and G. A. Kazakov, "Upper boundary of the biosphere.," *Applied and environmental microbiology*, vol. 35, pp. 1–5, jan 1978.
[31] D. W. Griffin, "Terrestrial Microorganisms at an Altitude of 20,000 m in Earth's Atmosphere," *Aerobiologia*, vol. 20, pp. 135–140, jun 2004.
[32] A. M. Womack, B. J. M. Bohannan, and J. L. Green, "Biodiversity and biogeography of the atmosphere.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, pp. 3645–53, nov 2010.
[33] U. Weierstall, J. C. H. Spence, and R. B. Doak, "Injector for scattering measurements on fully solvated biospecies," *Review of Scientific Instruments*, vol. 83, p. 035108, mar 2012.
[34] J. Schulz, S. Bari, J. Buck, and C. Uetrecht, "Sample refreshment schemes for high repetition rate FEL experiments.," in *Advances in X-ray Free-Electron*

Lasers II: Instrumentation, pp. 1–11, 2013.
[35] M. Rossmann, *The molecular replacement method a collection of papers on the use of non-crystallographic symmetry.* New York: Gordon and Breach, 1972.
[36] U. Arndt, "Phase determination in X-ray crystallography by anomalous scattering techniques," *Nuclear Instruments and Methods*, vol. 152, pp. 307–311, jun 1978.
[37] I. McNulty, J. Kirz, C. Jacobsen, E. H. Anderson, M. R. Howells, and D. P. Kern, "High-Resolution Imaging by Fourier Transform X-ray Holography," *Science*, vol. 256, pp. 1009–1012, may 1992.
[38] S. Eisebitt, J. Lüning, W. F. Schlotter, M. Lörgen, O. Hellwig, W. Eberhardt, and J. Stöhr, "Lensless imaging of magnetic nanostructures by X-ray spectro-holography," *Nature*, vol. 432, pp. 885–888, dec 2004.
[39] S. Marchesini, S. Boutet, A. E. Sakdinawat, M. J. Bogan, S. Bajt, A. Barty, H. N. Chapman, M. Frank, S. P. Hau-Riege, A. Szöke, C. Cui, D. A. Shapiro, M. R. Howells, J. C. H. Spence, J. W. Shaevitz, J. Y. Lee, J. Hajdu, and M. M. Seibert, "Massively parallel X-ray holography," *Nature Photonics*, vol. 2, pp. 560–563, sep 2008.
[40] J. Rodenburg, A. Hurst, and A. Cullis, "Transmission microscopy without lenses for objects of unlimited size," *Ultramicroscopy*, vol. 107, pp. 227–231, feb 2007.
[41] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, "Probe retrieval in ptychographic coherent diffractive imaging," *Ultramicroscopy*, vol. 109, pp. 338–343, mar 2009.
[42] D. Sayre, "Some implications of a theorem due to Shannon," *Acta Crystallographica*, vol. 5, pp. 843–843, nov 1952.
[43] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, pp. 10–21, jan 1949.
[44] J. Bernal, I. Fankuchen, and M. Perutz, "An X-ray study of chymotrypsin and hemoglobin.," *Nature*, pp. 523–524, 1938.
[45] A. Walther, "The Question of Phase Retrieval in Optics," *Optica Acta: International Journal of Optics*, vol. 10, pp. 41–49, jan 1963.
[46] Y. Bruck and L. Sodin, "On the ambiguity of the image reconstruction problem," *Optics Communications*, vol. 30, pp. 304–308, sep 1979.
[47] K. Ayyer, O. M. Yefanov, D. Oberthür, S. Roy-Chowdhury, L. Galli, V. Mariani, S. Basu, J. Coe, C. E. Conrad, R. Fromme, A. Schaffer, K. Dörner, D. James, C. Kupitz, M. Metz, G. Nelson, P. L. Xavier, K. R. Beyerlein, M. Schmidt, I. Sarrou, J. C. H. Spence, U. Weierstall, T. A. White, J.-H. Yang, Y. Zhao, M. Liang, A. Aquila, M. S. Hunter, J. S. Robinson, J. E. Koglin, S. Boutet, P. Fromme, A. Barty, and H. N. Chapman, "Macromolecular diffractive imaging using imperfect crystals.," *Nature*, vol. 530, pp. 202–6, feb 2016.
[48] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Optics Letters*, vol. 3, p. 27, jul 1978.
[49] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.
[50] A. Levi and H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude," *Journal of the*

Optical Society of America A, vol. 1, p. 932, sep 1984.
[51] J. R. Fienup, "Phase retrieval algorithms: a comparison.," *Applied optics*, vol. 21, no. 15, pp. 2758–2769, 1982.
[52] S. Marchesini, "Invited Article: A unified evaluation of iterative projection algorithms for phase retrieval," *Review of Scientific Instruments*, vol. 78, p. 11301, jan 2007.
[53] F. R. N. C. Maia, T. Ekeberg, D. Van Der Spoel, and J. Hajdu, "Hawk : the image reconstruction package for coherent X-ray diffractive imaging," *Journal of Applied Crystallography*, vol. 43, no. 6, pp. 1535–1539, 2010.
[54] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, "X-ray image reconstruction from a diffraction pattern alone," *Physical Review B*, vol. 68, p. 140101, oct 2003.
[55] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, sep 1936.
[56] R. Sokal and C. Michener, "A statistical method for evaluating systematic relationships," *Univ. Kans. Sci. Bull.*, vol. 38, pp. 1409–1438, 1958.
[57] D. Paganin, T. E. Gureyev, S. C. Mayo, a. W. Stevenson, Y. I. Nesterets, and S. W. Wilkins, "X-ray omni microscopy.," *Journal of microscopy*, vol. 214, no. Pt 3, pp. 315–327, 2004.
[58] N.-T. D. Loh and V. Elser, "Reconstruction algorithm for single-particle diffraction imaging experiments.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 80, p. 26705, aug 2009.
[59] D. Giannakis, P. Schwander, and A. Ourmazd, "The symmetries of image formation by scattering. I. Theoretical framework," *Optics express*, vol. 20, no. 12, pp. 5591–5596, 2010.
[60] G. Bortel and M. Tegze, "Common arc method for diffraction pattern orientation.," *Acta crystallographica. Section A, Foundations of crystallography*, vol. 67, pp. 533–543, nov 2011.
[61] Z. Kam, "Determination of Macromolecular Structure in Solution by Spatial Correlation of Scattering Fluctuations," *Macromolecules*, vol. 10, pp. 927–934, sep 1977.
[62] D. K. Saldin, H. C. Poon, V. L. Shneerson, M. Howells, H. N. Chapman, R. A. Kirian, K. E. Schmidt, and J. C. H. Spence, "Beyond small-angle x-ray scattering: Exploiting angular correlations," *Physical Review B*, vol. 81, p. 174105, may 2010.
[63] R. A. Kirian, K. E. Schmidt, X. Wang, R. B. Doak, and J. C. H. Spence, "Signal, noise, and resolution in correlated fluctuations from snapshot small-angle x-ray scattering," *Physical Review E*, vol. 84, p. 011921, jul 2011.
[64] M. Tegze and G. Bortel, "Atomic structure of a single large biomolecule from diffraction patterns of random orientations," *Journal of Structural Biology*, 2012.
[65] M. F. Hantke, T. Ekeberg, and F. R. N. C. Maia, "Condor: a simulation tool for flash X-ray imaging," *Journal of Applied Crystallography*, vol. 49, pp. 1356–1362, 2016.
[66] Mathworks, "Laplace Edge Detection."
[67] Y. Lehahn, I. Koren, D. Schatz, M. Frada, U. Sheyn, E. Boss, S. Efrati, Y. Rudich, M. Trainic, S. Sharoni, C. Laber, G. R. DiTullio, M. J. L. Coolen,

A. M. Martins, B. A. S. Van Mooy, K. D. Bidle, and A. Vardi, "Decoupling physical from biological processes to assess the impact of viruses on a mesoscale algal bloom.," *Current biology : CB*, vol. 24, pp. 2041–6, sep 2014.

[68] J. Komarek, J. Kopecky, and V. Cepak, "Generic characters of the simplest cyanoprokaryotes Cyanobium, Cyanobacterium and Synechococcus," *Cryptogamie Algologie*, vol. 20, pp. 209–222, jul 1999.

[69] G. Stanier (Cohen-Bazire), "Fine structure of cyanobacteria," *Methods in Enzymology*, vol. 167, pp. 157–172, jan 1988.

[70] C. Bostedt, J. D. Bozek, P. H. Bucksbaum, R. N. Coffee, J. B. Hastings, Z. Huang, R. W. Lee, S. Schorb, J. N. Corlett, P. Denes, P. Emma, R. W. Falcone, R. W. Schoenlein, G. Doumy, E. P. Kanter, B. Kraessig, S. Southworth, L. Young, L. Fang, M. Hoener, N. Berrah, C. Roedig, and L. F. DiMauro, "Ultra-fast and ultra-intense x-ray sciences: first results from the Linac Coherent Light Source free-electron laser," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 46, p. 164003, aug 2013.

[71] L. Struder, S. Epp, D. Rolles, R. Hartmann, P. Holl, G. Lutz, H. Soltau, R. Eckart, C. Reich, K. Heinzinger, C. Thamm, A. Rudenko, F. Krasniqi, K.-U. Kuhnel, C. Bauer, C.-D. Schroter, R. Moshammer, S. Techert, D. Miessner, M. Porro, O. Halker, N. Meidinger, N. Kimmel, R. Andritschke, F. Schopper, G. Weidenspointner, A. Ziegler, D. Pietschner, S. Herrmann, U. Pietsch, A. Walenta, W. Leitenberger, C. Bostedt, T. Moller, D. Rupp, M. Adolph, H. Graafsma, H. Hirsemann, K. Gartner, R. Richter, L. Foucar, R. L. Shoeman, I. Schlichting, and J. Ullrich, "Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 614, pp. 483–496, mar 2010.

[72] A. Barty, R. a. Kirian, F. R. N. C. Maia, M. Hantke, C. H. Yoon, T. a. White, and H. Chapman, "Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data," *Journal of applied crystallography*, vol. 47, pp. 1118–1131, jun 2014.

[73] D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," *Inverse Problems*, vol. 21, pp. 37–50, feb 2005.

[74] F. R. N. C. Maia, "The Coherent X-ray Imaging Data Bank," *Nature Methods*, vol. 9, pp. 854–855, aug 2012.

[75] A. Nakagawa, N. Miyazaki, J. Taka, H. Naitow, A. Ogawa, Z. Fujimoto, H. Mizuno, T. Higashi, Y. Watanabe, T. Omura, R. H. Cheng, and T. Tsukihara, "The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins.," *Structure (London, England : 1993)*, vol. 11, pp. 1227–38, oct 2003.

[76] A. Aquila, A. Barty, C. Bostedt, S. Boutet, G. Carini, D. DePonte, P. Drell, S. Doniach, K. H. Downing, T. Earnest, H. Elmlund, V. Elser, M. Gühr, J. Hajdu, J. Hastings, S. P. Hau-Riege, Z. Huang, E. E. Lattman, F. R. N. C. Maia, S. Marchesini, A. Ourmazd, C. Pellegrini, R. Santra, I. Schlichting, C. Schroer, J. C. H. Spence, I. A. Vartanyants, S. Wakatsuki, W. I. Weis, and G. J. Williams, "The linac coherent light source single particle imaging road map," *Structural Dynamics*, vol. 2, no. 4, 2015.

[77] J. L. Van Etten and D. D. Dunigan, "Chloroviruses: not your everyday plant virus.," *Trends in plant science*, vol. 17, pp. 1–8, jan 2012.
[78] F. R. N. C. Maia, T. Ekeberg, N. Timneanu, D. van der Spoel, and J. Hajdu, "Structural variability and the incoherent addition of scattered intensities in single-particle diffraction," *Physical Review E*, vol. 80, p. 031905, sep 2009.
[79] T. Ekeberg, *Flash Diffractive Imaging in Three Dimensions*. Doctoral thesis, Uppsala University, 2012.
[80] D. P. DePonte, U. Weierstall, K. Schmidt, J. Warner, D. Starodub, J. C. H. Spence, and R. B. Doak, "Gas dynamic virtual nozzle for generation of microscopic droplet streams," *Journal of Physics D: Applied Physics*, vol. 41, p. 195505, oct 2008.
[81] G. Nomarski and A. Weill, "Application a la metallographie des methodes interferentielles a deux ondes polarisees," *Rev. de Metallurgie*, vol. 2, no. 52, p. 121, 1955.
[82] C. H. Yoon, P. Schwander, C. Abergel, I. Andersson, J. Andreasson, A. Aquila, S. Bajt, M. Barthelmess, A. Barty, M. J. Bogan, C. Bostedt, J. Bozek, H. N. Chapman, J.-M. Claverie, N. Coppola, D. P. DePonte, T. Ekeberg, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, H. Graafsma, L. Gumprecht, J. Hajdu, C. Y. Hampton, A. Hartmann, E. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Kiskinova, M. Liang, N.-T. D. Loh, L. Lomb, F. R. N. C. Maia, A. V. Martin, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, I. Schlichting, J. Schulz, M. Seibert, V. Seltzer, R. L. Shoeman, R. G. Sierra, H. Soltau, D. Starodub, J. Steinbrener, G. Stier, L. Strüder, M. Svenda, J. Ullrich, G. Weidenspointner, T. A. White, C. Wunderer, and A. Ourmazd, "Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering," *Optics Express*, vol. 19, no. 17, p. 16542, 2011.
[83] A. M. Gañán-Calvo, D. P. DePonte, M. A. Herrada, J. C. H. Spence, U. Weierstall, and R. B. Doak, "Liquid Capillary Micro/Nanojets in Free-Jet Expansion," *Small*, vol. 6, pp. 822–824, apr 2010.
[84] A. T. Brünger, "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures," *Nature*, vol. 355, p. 472, 1992.
[85] B. J. Daurer, M. F. Hantke, C. Nettelblad, and F. R. N. C. Maia, "Hummingbird: monitoring and analyzing flash X-ray imaging experiments in real time 1," *J. Appl. Cryst*, vol. 49, pp. 1042–1047, 2016.
[86] P. Liu, P. J. Ziemann, D. B. Kittelson, and P. H. McMurry, "Generating Particle Beams of Controlled Dimensions and Divergence: I. Theory of Particle Motion in Aerodynamic Lenses and Nozzle Expansions," *Aerosol Science and Technology*, vol. 22, pp. 293–313, jan 1995.
[87] J. C. H. Spence, R. A. Kirian, X. Wang, U. Weierstall, K. E. Schmidt, T. White, A. Barty, H. N. Chapman, S. Marchesini, and J. Holton, "Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals," *Optics Express*, 2011.
[88] F. Cruz-Mazo, M. A. Herrada, A. M. Gañán-Calvo, and J. M. Montanero, "Global stability of axisymmetric flow focusing," *Journal of Fluid Mechanics*, vol. 832, pp. 329–344, dec 2017.
[89] J. Andreasson, A. V. Martin, M. Liang, N. Timneanu, A. Aquila, F. Wang,

B. Iwan, M. Svenda, T. Ekeberg, M. Hantke, J. Bielecki, D. Rolles, A. Rudenko, L. Foucar, R. Hartmann, B. Erk, B. Rudek, H. N. Chapman, J. Hajdu, and A. Barty, "Automated identification and classification of single particle serial femtosecond X-ray diffraction data.," *Optics express*, vol. 22, no. 3, pp. 2497–2510, 2014.

[90] N. D. Loh, M. J. Bogan, V. Elser, A. Barty, S. Boutet, S. Bajt, J. Hajdu, T. Ekeberg, F. R. N. C. Maia, J. Schulz, M. M. Seibert, B. Iwan, N. Timneanu, S. Marchesini, I. Schlichting, R. L. Shoeman, L. Lomb, M. Frank, M. Liang, and H. N. Chapman, "Cryptotomography: Reconstructing 3D Fourier Intensities from Randomly Oriented Single-Shot Diffraction Patterns," *Physical Review Letters*, vol. 104, p. 225501, jun 2010.

[91] A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, "Quantitative optical phase microscopy," *Optics Letters*, vol. 23, p. 817, jun 1998.

[92] T. Tajima and J. M. Dawson, "Laser Electron Accelerator," *Physical Review Letters*, vol. 43, pp. 267–270, jul 1979.

[93] L. Foucar, A. Barty, N. Coppola, R. Hartmann, P. Holl, U. Hoppe, S. Kassemeyer, N. Kimmel, J. Küpper, M. Scholz, S. Techert, T. a. White, L. Strüder, and J. Ullrich, "CASS CFEL ASG software suite," *Computer Physics Communications*, vol. 183, pp. 2207–2213, oct 2012.

[94] K. Ayyer, T. Y. Lan, V. Elser, and N. D. Loh, "Dragonfly: An implementation of the expand-maximize-compress algorithm for single-particle imaging," *Journal of Applied Crystallography*, vol. 49, no. 4, pp. 1320–1335, 2016.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1625

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-334219

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018