

Title Page Dummy

This Page will be Replaced before Printing

Title page logo

Abstract Dummy Page.

Dedicated to our 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 Imaging single cells in a beam of live cyanobacteria.
- II Open dataset of live cyanobacterial cells imaged using an X-ray laser.
- III RedFlamingo: Suite for automated classification of diffraction patterns.
- IV Template-based classification of diffraction patterns.
- V Open dataset of RDV particles.

Reprints were made with permission from the publishers.

Contents

Part I: Motivation	9
1 Motivation	11
Part II: Concept	13
2 Interaction of Light and Matter	15
2.1 Light as electromagnetic radiation	15
2.2 Photon-material interactions	16
2.2.1 Scattering by a single free electron	17
2.2.2 Two-body scattering	18
2.2.3 Scattering from multiple electrons	19
2.2.4 Scattering Potential	20
2.3 Radiation Damage	22
2.4 Diffract-before-destroy	22
3 Pulse Generation by XFEL	24
3.1 Klystron	24
3.2 Bunch Compression	25
3.3 Undulator	26
3.3.1 Monochromatic X-ray radiation	26
3.3.2 SASE	27
3.4 Improvements	28
4 Sample Introduction	29
4.1 Substrate-bound Sample Delivery	29
4.1.1 Solid substrate	29
4.1.2 Micro Compartments	30
4.1.3 Liquid Jets	30
4.2 Substrate-free sample delivery	31
4.2.1 Gas Dynamic Virtual Nozzle	31
4.2.2 Aerodynamic Lens Stack	31
4.2.3 Electro Spray Ionisation	32
4.2.4 Drop-on-demand	33
5 Data recording	34
5.1 the pnCCD detector	34
5.2 Quantum Efficiency	34

5.3	Missing Data	35
5.4	Saturation	35
6	Phase Retrieval	36
6.1	Oversampling	36
6.2	Iterative Phase retrieval	37
6.2.1	Error Reduction	38
6.2.2	HIO	38
6.2.3	RAAR	38
6.3	Validation	38
6.3.1	Errors	38
6.3.2	PRTF	39
6.3.3	Missing Mode Analysis	39
6.3.4	Hierarchical clustering	40
6.4	Simulated Phase Contrast Methods	40
6.5	Shrinkwrap	42
7	3D restrictions	43
7.1	the EMC algorithm	43
7.2	Non-reproducible objects	44
7.2.1	Multi-model EMC	44
7.3	the Manifold Embedding algorithm	44
7.3.1	Tomography and sub-tomogram averaging	44
7.3.2	Holography	45
8	Image Classification	46
8.1	Template-based classification	46
8.2	Feature extraction	46
8.2.1	Size	46
8.2.2	Multiple scatterers in the focus	47
8.2.3	Edge detection	47
9	Software	48
9.1	RedFlamingo	48
	Part III: Results	49
10	Results	51
10.1	Imaging live cells	51
10.2	Classification	53
10.3	Software	53
11	Sammanfattning pa Svenska	55
	References	57

Part I:

Concept

1. Interaction of Light and Matter

Traditionally cells 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. For most work described in this thesis the classical model of describing light as a wave is sufficient. This chapter works well. In this chapter we will create the mathematical framework that describes how light behaves in our diffraction experiments.

1.1 Light as electromagnetic radiation

In the 17th and 18th century it became possible for humans to generate continuous electric charges and continuous currents [Guericke, Leiden, Volta], enabling the investigation of the phenomena of electricity and magnetism. This ultimately led to the ground breaking work of Maxwell . His four laws are the synthesis of the a ground-breaking work by Maxwell where he ties together the fields of magnetism, electricity, and somewhat surprisingly, optics. His results are described by four equations known as the Maxwell equations:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0} \quad (1.1)$$

$$\nabla \cdot \vec{B} = 0 \quad (1.2)$$

$$\nabla \times \vec{E} = - \frac{d\vec{B}}{dt} \quad (1.3)$$

$$\nabla \times \vec{B} = \mu_0 \vec{J} - \mu \epsilon \frac{d\vec{E}}{dt} \quad (1.4)$$

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

In order to understand why these laws unify magnetism, electricity and optics, we have to 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

$$\nabla \cdot \vec{E} = 0 \quad (1.5)$$

$$\nabla \times \vec{B} = -\mu_0 \epsilon_0 \frac{d\vec{E}}{dt} \quad (1.6)$$

Furthermore, experiments showed $\epsilon = \epsilon_0 = 8.854 \cdot 10^{-12} Fm^{-1}$, and $\mu = \mu_0 = 1.257 \cdot 10^{-7} NA^{-2}$ that the permitivity for vacuum, $\epsilon_0 = 8.854 \cdot 10^{-12} Fm^{-1}$, 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 mathematicala well known mathematical relation, and for the second equality equation 2.5 is used. Another way of writing this equation is: Now let's rewrite the equation in a different way:

$$\nabla \times (\nabla \times \vec{E}) = \nabla \times \left(-\frac{d\vec{B}}{dt} \right) = -\frac{d}{dt}(\nabla \times \vec{B}) = \frac{d}{dt}(\mu_0 \epsilon_0 \frac{d\vec{E}}{dt}) = \mu_0 \epsilon_0 \frac{d^2 \vec{E}}{dt^2}$$

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:

$$\frac{d^2 \vec{E}}{dt^2} = \frac{1}{\mu_0 \epsilon_0} \nabla^2 \vec{E} = v^2 \nabla^2 \vec{E}, v = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \quad (1.7)$$

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 $3 \cdot 10^8 \frac{m}{s}$, which agreed with the then known value of speed of light [37] which was already known at the time of Maxwell. This result strongly hints at the possibility that light can be is an electromagnetic wave. Further experiments succeeded in demonstrating the existence of long-wavelength electromagnetic waves 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 turned complicates the description of light as wave upside downfurther, many phenomena can be described by treating light as a pure wave.

1.2 Photon-material interactions

Assuming that light is electromagnetic radiation, it is easy to understand that a change in electric field the electric field itself will affect its propagation. There

are four mechanisms through which radiation interacts with matter: photo-absorption, scattering, photo-nuclear absorption and pair production. The latter two phenomena only occur when matter is exposed to high energy gamma rays, and are therefore of little relevance to this thesis. Photo-absorption Photoabsorption is facilitated primarily through the process of photo-excitation/photo-ionisation of electronsa process called photoexcitation, in which an electron is excited to a higher level of energy, or possibly ionized in case of photoionisation¹. Scattering can be divided into two classes: elastic scattering and inelastic scattering. In a scattering process the direction of the radiation is changed. 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 energyof the scatterer. Photo-absorption and inelastic scattering deposit energy into the object which will lead to structural changes . of the object. These phenomena are therefore not very useful to study objects structurally. Elastic scattering on the other hand can be used to gain structural information about the object, without altering the structurecausing damage to the object.

1.2.1 Scattering by a single free electron

1.3 Diffraction

The simplest example of elastic scattering is Thompson scattering from a free electron. A free electron will scatter (or diffract) or diffract the incoming radiation in all directions. If more than one electron are present interestingly the scattered waves from the different electrons 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. Figure ?? is a schematic of such scattering. Waves with path differences close to integer number of wavelengths interfere constructively, if close to half integer number of wavelengths the 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, the one next to that the second order and so on.

1.3.1 Lorentz Force

An important physical concept for understanding scattering on a deeper level is the Lorentz force . which describes the force exerted on a charged particle

¹Photoionisation is the process that underlies the photo-electric effect.

traveling in an electromagnetic field:

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B} \quad (1.8)$$

Here \vec{v} describes the speed and direction of movement of the particle and q its charge. 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 that in which case there is no acceleration).

1.3.2 Scattering by a single free electron

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

$$\vec{E}_{in} = E_0 e^{-2\pi i \frac{ct}{\lambda}} \quad (1.9)$$

When the oscillating electric field of the incident EM 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:

$$\vec{F} = m_e \vec{a} = e E_0 e^{-2\pi i \frac{ct}{\lambda}} \quad (1.10)$$

An accelerated charge emits electromagnetic radiation, since this is the only mechanism through which the electron can preserve momentum. The oscillating electron becomes a new source of radiation that radiates spherically in all directions, at the same frequency as the incident radiation. This phenomenon is called Thompson scattering. From Maxwell's equations it follows that the electric field generated by an accelerating electron measured at point \vec{d} can be described as:

$$\vec{E}_s(\vec{d}, t) = \frac{ea_{\perp}(t - |\vec{d}|/c)}{4\pi\epsilon_0 c^2 d} \frac{ea_{\perp}(t - |\vec{d}|/c)}{4\pi\epsilon_0 c^2 |\vec{d}|} \quad (1.11)$$

a_{\perp} is the acceleration projected on a plane perpendicular to \vec{d} . Combining equations 2.9, 2.10, and 2.11 we get that the instantaneous scattered field is:

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

$$E_s(\vec{d}, t) = \frac{e^2}{4\pi\epsilon_0 m_e c^2} \frac{E_0 \sin(\theta)}{d} e^{-2\pi i \frac{|\vec{d}|}{\lambda}}$$

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

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

$$r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2}$$

1.3.3 Two-body scattering Scattering from two electrons

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.

If two electrons are present, instead of one single electron, the scattered EM waves from both electrons will interfere. Figure 2.2.2 illustrates the phenomenon of interference this phenomenon.

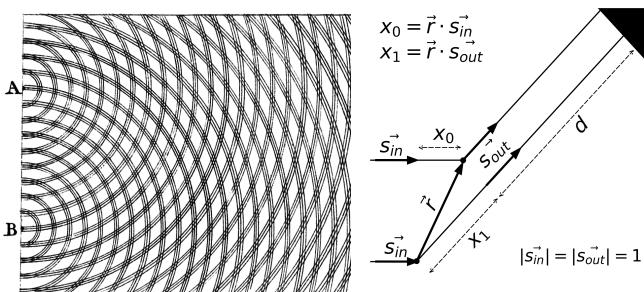


Figure 1.1. Illustration of scattering from two free-electrons A and B. a) graphic taken from Young, showing an interference pattern between two waves. b) mathematical description of the path length difference between two waves arriving at point \vec{d} . \vec{s}_{in} and \vec{s}_{out} are unit vectors pointing in the direction of propagation of the incident wave and the out going wave respectively. \vec{r} is the a position vector of describing the distance from point A , compared to point B. Note that in a) $x_0 = 0$.

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 as since the distance from both electrons to the detector is about equal: . In

figure 2.2.2 this means that $|\vec{d} + \vec{x}_1| \approx |\vec{d}|$. This approximation is called the far-field approximation. The Fresnel number (FN) is used to verify the validity of the far-field approximation.

$$FN = \frac{o^2}{d\lambda} \quad (1.12)$$

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} can be described by a is given by the sum of the individual scattered electric fields:

$$E(\vec{d}) = E_s(\vec{d}) + E_s(\vec{d})e^{\frac{2\pi i \Delta x}{\lambda}} \quad (1.13)$$

Δx is the optical path difference between both scattered the two scattered waves due to the relative difference in position. Δx can also 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. be described using the difference vector r :

$$\Delta x = x_0 - x_1 = \vec{s}_{out} \cdot \vec{r} - \vec{s}_{in} \cdot \vec{r} = (\vec{s}_{out} - \vec{s}_{in}) \cdot \vec{r} \quad (1.14)$$

By defining the scattering vector \vec{S}

$$\vec{S} = \frac{\vec{s}_{out} - \vec{s}_{in}}{\lambda} \quad (1.15)$$

the diffraction pattern of two electrons will be :

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

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

It is important Important to note is that Δx is at most can at most be $|\vec{r}|$, the distance between the two electrons $|\vec{r}|$. 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 ($\lambda = 500\text{nm}$) 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 will be required.

1.3.4 Scattering from multiple electrons

In biological particles the incoming EM radiation is scattered by many electrons. The electric field can be described as the sum of all the individual scattered fields.

$$E(\vec{d}) = E_s(\vec{d}) \sum_n e^{2\pi i \vec{S} \cdot \vec{r}_n} \quad (1.16)$$

where n identifies the individual electrons.

If all electrons together are assumed to form a continuous electric charge density around the many nuclei that constitute the biological particle, equation 2.17 can be written as a continuous function.

$$E(\vec{d}) = E_s(\vec{d}) \int \rho(\vec{r}, \lambda) \exp^{-2\pi i \vec{S} \cdot \vec{r}} d\vec{r} \quad (1.17)$$

$\rho(\vec{r}, \lambda)$ is named the scattering potential and takes the scattering strength of each nucleus into consideration.

So far we have implicitly assumed 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.

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

$$F(\vec{S}) = \frac{E(\vec{d})}{E_s(\vec{d})} = F\left(\frac{\vec{d}}{|d|}\right) \quad (1.18)$$

The structure factor is independent of the distance of the detector, it only depends on the angle of measurement. Equation 2.18 can be rewritten as:

$$F(\vec{S}) = \int \rho(\vec{r}, \lambda) \exp^{-2\pi i \vec{S} \cdot \vec{r}} d\vec{r} \quad (1.19)$$

This equation shows that a diffraction experiment is simply the 3D Fourier transform of the electron density evaluated at \vec{S} . This relation is very convenient as there is a large mathematical field describing the relations of Fourier transformations. A few of these relations are very useful to this thesis.

The Fourier shift theorem describes the change in the diffraction pattern upon object translation(i. e. shift). .

$$\mathcal{F}(f(x + \delta x)) = e^{2\pi i \delta x q} \mathcal{F}(f) \quad (1.20)$$

We see that the translation in real space turns into a multiplication with the factor $e^{2\pi i \delta x q}$ which is called a 'phase-ramp'.

The Convolution theorem states that a multiplication in Fourier space is a convolution in object space (real space) and vice versa. real space and visa versa.

$$f(x) * g(x) = \int f(\tau)g(x - \tau) d\tau = \mathcal{F}^{-1}(\mathcal{F}(f(x)) \cdot \mathcal{F}(g(x))) \quad (1.21)$$

The Fourier slice theorem states that a relates the Fourier transform of a 3D density with the Fourier transform of a 2D diffraction pattern is a projection image of the same density. More precisely it says that the transform of the projection is identical to a 2D slice through the center transform of the 3D fourier transform of the object.

It has thus become possible to move between diffraction pattern and electron density. density such that the slice goes through the origin.

1.3.5 Scattering Potential

An important term in equation the previous equations that needs further development is the scattering potential of a material $\rho(\vec{r}, \lambda)$. Mathematically $\rho(\vec{r}, \lambda)$ can be formulated as a function of the refractive index n of a material.

$$\rho(\vec{r}, \lambda) = \pi s_{in}^2 ((n(\vec{r}, \lambda))^2 - 1)) \quad (1.22)$$

$$s = \frac{1}{\lambda}$$

The refractive index is related to the scattering factors of the individual atoms in the material.

$$n = 1 - \frac{r_e}{2\pi} \lambda^2 \sum_i N_i f_i(0) \quad (1.23)$$

where r_e is the classical electron radius, λ is the wavelength, N_i is the number of atoms per unit volume. The complex scattering factor in the forward direction $f(0)$ can be described by:

$$f(0) = f_1 + i f_2 \quad (1.24)$$

The imaginary part is derived from the atomic photoabsorption cross section and is a measure of absorption [booklet]. The real part is related to the imaginary part using the Kramers-Kronig dispersion relation [Kramers]. At high photon energies f_1 approaches the atomic number of the atoms. The exact value of f_1 and f_2 at different photon-energies have been determined experimentally [Henke].

Figure 2.2.4 presents the scattering factors of two elements: carbon and oxygen. The gray energy range from 282-533 eV (4.40 nm - 2.33 nm) is called the water window. In this region, and especially towards the endhigher energies, oxygen atoms(thus water) , and thus water, are scattering significantly less than carbon atoms(i. e. 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.

1.4 Radiation Damage

The previous section shows that scattering power of an a body is related to the wavelength of the radiation. The shorter the wavelength the less scattering

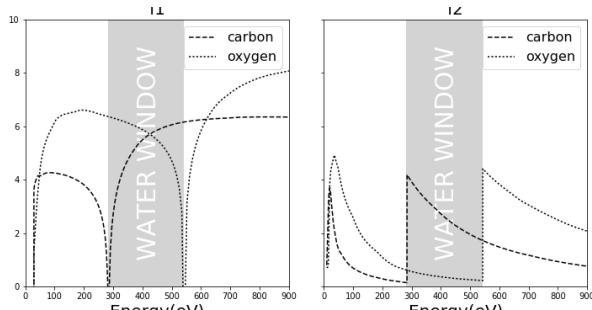


Figure 1.2. Atomic scattering factors f_1 and f_2 of carbon and oxygen, as a function of photon-energy of the X-ray radiation. The region between the absorption K-edge of carbon (282 eV, 4.40 nm) and the K-edge of oxygen (533 eV, 2.33 nm) is called the water window. The contrast between biomolecules and water is enhanced in this region, especially towards the high end of it. These wavelengths could be used for imaging living biological particles such as cells or organelles.

occurs. To compensate for this loss in 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(. This number can be put in perspective by noting that a dose of 2000 Gy is already enough to kill everything that lives on earth). Unfortunately other processes of matter-radiation interaction also become more abundant as wellat lower photon energies. At soft X-ray wavelengths the predominant damage process is photoionisation. Many process processes follow the ionization, of which non-radiative Auger decay will lead to significant radiation damage. The relation between the required dose of radiation versus the maximum tolerable dose of radiation posed used to pose a physical limit to the obtained resolution for imaging single biomolecules [Howell].

1.5 Diffract-before-destroyDiffraction before destruction

An elegant solution to the problem of radiation damage came from the realization that is possible since elastic scattering occurs on a time scale that is shorter than the damage processes manifest themselves time it takes for the damage to manifest itself structurally. This principle is of diffract-before-destroyis the diffraction before destruction principle, suggested in 2000 by Neutze et al [], was experimentally demonstrated . The principle was later demonstrated experimentally on a silicon-nitrate sample [Chapman]. This experiment used the ultra short and extremely bright pulses from the first x-ray X-ray free-electron laser to outrun key (XFEL) to outrun the damage processes. Although the pulse literally obliterated the sample, enough structural information was cap-

tured to retrieve the original structure. A single XFEL pulse has a power densities density of $10^{16} W/cm^2$, and can be between 1-100 fs long. This facilitates the imaging of single particles at room temperature.

Diffract-before-destroy Diffraction before destruction has been further validated by a wide variety of experiments. 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 []. The imaging of single particle particles has been shown feasible in 2D for a wide variety of biological particles, ranging from *living* cells (Paper I) to cell organelles (Paper 7) to 40 nm viruses (Paper 8). Paper VI shows the experimental feasibility of using many 2D images to construct a 3D model of a giant virus.

The field is I believe that we are very close to measuring the first 2D diffraction patterns of single proteins. For simulated data it has been shown that such patterns can be used to recover the 3D shape of proteins, even though the number of scattered photons might be as low as 1000 photons per frame. The ultra short ultra-short pulses also allow for the studying of the dynamics of such systems, either by separating the images from different naturally occurring conformational states, or by identifying structural changes subsequent to an external trigger []. In the more distant future these structural changes might be tracked inside living cells.

Part II:

Motivation

2. Motivation

When a famous scientist was asked why he had a horseshoe hanging over his door he replied: "They say it works, whether you believe in it or not". The same can be said about science. Science attempts to build a framework of knowledge that transcends individual beliefs. Paradoxically the knowledge acquired through the scientific process relies on models and theories that are themselves simplifications of the phenomena they try to describe. Nonetheless, the power gained through the models and theories is enormous as for example antibiotics and acquired resistance through vaccination have extended and improved our lifetimes. Furthermore, by logical inference from existing models or theories new phenomena, phenomena that had never been observed before, might even be predicted. One such examples is the bending of the path of light by a heavy object like the sun. In this thesis I will try to show a few examples of how the field of science is evolving, how ideas themselves have evolved, and that many experiments are exploiting progress from many areas of science.

One field where I particularly feel attracted to is structural biology. Cellular life and the organization of its constituents are amazingly intricate and diverse. Proteins form an interconnected network of interactions in which temporal changes to individual proteins can trigger global responses such as the reshaping of macrophages in the presence of a bacterium or a cancerous cell[1]. In order to understand the factors that activate or deactivate such pathways it is not enough to study individual components alone. A grand challenge of the 21st century is the imaging of *live* cells, at atomic resolution, at a time resolution that allows capturing even the fastest biological processes.

Super resolution optical microscopy can image cells as a whole and has increased our understanding of cellular organization significantly. However the technique is ultimately limited by the size of the fluorescent label. X-ray crystallography can image large bio-particles at atomic resolution, but requires the particles to crystallize. For cells this is not possible. Nuclear Magnetic Resonance can study the dynamics of proteins at atomic resolution, at with nanosecond time resolution. Recently it has succeeded in studying labelled proteins *in vivo* [2]. NMR is still limited by the size of the proteins it can study, as well as the collection time. It has been attempted to predict protein structure from backbone chemical shift only, which would avoid the most time consuming steps [Shen]. I have participated in the improvement of this method as shown in Paper V-X. This method works well for predicting the structure of ordered region.

The most promising method to study cells as well as its constituents at near-atomic resolution is cryo-electron microscopy. Focussed ion beams can be used to slice cryo-frozen cells into thin sections without considerably damaging them. This has solved the issue of penetration debt allowing the studying of internal features. Using sub-tomogram averaging, the structure of highly abundant proteins can be elucidated at resolutions beyond 4A [1]. For understanding rare events, however, acquisition time is a limiting factor. The time resolution of this field is currently limited to the microsecond range [2].

A promising alternative that can be utilized to study living cells is femtosecond x-ray diffraction imaging (FXI). FXI uses the ultrashort and extremely bright pulses produced by free-electron x-ray lasers. The power of the pulse allows for single particle imaging (SPI). They femtosecond pulse outrun the key damage processes that occur due to the exposure to the beam. It is predicted that sub-nanometer resolution can be achieved on micron-sized cells [12]. The femtosecond pulse gives an unprecedented time resolution that captures any biologically relevant motion, all at room temperature. Another big advantage of this method is the repetition rate. The recently operational EuXFEL can have repetition rates of 27000 Hz, potentially allowing the recording of over a billion images a day. This might enable the imaging of rare cellular events.

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 cell imaging a reality. It will start by explaining the general framework necessary for each step of the experiment: The generation of X-rays, sample introduction, the interaction of light and matter, the recording of two-dimensional (2D) diffraction pattern, the reconstruction of cell images, how 2D images might be combined to derive three-dimensional (3D) structural information, and finally how image classification might be useful for the latter two. The final chapters describe the results on life cell imaging, and how image classification has been used for imaging reconstruction and pattern selection.

Part III:

Concept

3. Interaction of Light and Matter

Traditionally cells 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. For most work described in this thesis the classical model describing light as a wave is sufficient. This chapter will create the mathematical framework that describes our diffraction experiments.

3.1 Light as electromagnetic radiation

In the 17th and 18th century it became possible for humans to generate continuous electric charges and currents [Guericke, Leiden, Volta], enabling the investigation of the phenomena of electricity and magnetism. This ultimately led to the ground breaking work of Maxwell. His four laws are the synthesis of the fields of magnetism, electricity, and surprisingly, optics.

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0} \quad (3.1)$$

$$\nabla \cdot \vec{B} = 0 \quad (3.2)$$

$$\nabla \times \vec{E} = -\frac{d\vec{B}}{dt} \quad (3.3)$$

$$\nabla \times \vec{B} = \mu_0 \vec{J} - \mu \epsilon \frac{d\vec{E}}{dt} \quad (3.4)$$

\vec{E} is the electric field, \vec{B} is the magnetic field, ϵ is the permittivity of a material, and μ is its permeability. The main concept lying within these laws is that a moving electric 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 have to 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.

$$\nabla \cdot \vec{E} = 0 \quad (3.5)$$

$$\nabla \times \vec{B} = -\mu_0 \epsilon_0 \frac{d\vec{E}}{dt} \quad (3.6)$$

Furthermore, experiments showed $\epsilon = \epsilon_0 = 8.854 \cdot 10^{-12} Fm^{-1}$, and $\mu = \mu_0 = 1.257 \cdot 10^{-7} NA^{-2}$. If we now ask ourselves what field change caused by a change in the electric field ($\nabla \times (\nabla \times \vec{E})$) we can derive something quite amazing:

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

The first equality is mathematical, and for the second equation 2.5 is used. Another way of writing this equation is:

$$\nabla \times (\nabla \times \vec{E}) = \nabla \times \left(-\frac{d\vec{B}}{dt} \right) \stackrel{\text{def}}{=} -\frac{d}{dt}(\nabla \times \vec{B}) = \frac{d}{dt}(\mu_0 \epsilon_0 \frac{d\vec{E}}{dt}) = \mu_0 \epsilon_0 \frac{d^2 \vec{E}}{dt^2}$$

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:

$$\frac{d^2 \vec{E}}{dt^2} = \frac{1}{\mu_0 \epsilon_0} \nabla^2 \vec{E} = v^2 \nabla^2 \vec{E}, v = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \quad (3.7)$$

Equation 2.7 is known as the wave equation, which is used to describe the propagation of a wave. Combining this results with the experimentally determined value for μ_0 and ϵ_0 shows that electromagnetic waves travel at $3 \cdot 10^8 \frac{m}{s}$, which agreed with the then known speed of light [37]. This result strongly hints at the possibility that light can be an electromagnetic wave. Further experiments succeeded in demonstrating the existence of long-wavelength electromagnetic waves 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 turned the description of light as wave upside down, many phenomena can be described by treating light as a wave.

3.2 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: photo-absorption, scattering, photo-nuclear absorption and pair production. The latter two phenomena only occur when matter is exposed to high energy gamma rays, and are therefore of little relevance to this thesis. Photo-absorption is facilitated primarily through the process of photo-excitation/photo-ionisation of electrons. Scattering can be divided into two classes: elastic scattering and inelastic scattering. In a scattering process the direction of the radiation is changed. Elastic scattering does not result in a change of kinetic energy of the scattering particle, nor does it change the wavelength of the radiation. Inelastic scattering leads to both a

change in wavelength of the radiation, and a change in kinetic energy of the scatterer. Photo-absorption and inelastic scattering deposit energy into the object which will lead to structural changes. Elastic scattering on the other hand can be used to gain structural information about the object, without altering the structure.

3.2.1 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.

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B} \quad (3.8)$$

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 \vec{v} and \vec{B} are not parallel (in that case there is no acceleration).

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

$$\vec{E}_{in} = E_0 e^{-2\pi i \frac{ct}{\lambda}} \quad (3.9)$$

When the oscillating electric field of the incident EM 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:

$$\vec{F} = m_e \vec{a} = e E_0 e^{-2\pi i \frac{ct}{\lambda}} \quad (3.10)$$

An accelerated charge emits electromagnetic radiation, since this is the only mechanism through which the electron can preserve momentum. The oscillating electron becomes a new source of radiation that radiates spherically in all directions, at the same frequency as the incident radiation. This phenomenon is called Thompson scattering. From Maxwell's equations it follows that the electric field generated by an accelerating electron measured at point \vec{d} can be described as:

$$\vec{E}_s(\vec{d}, t) = \frac{e a_{\perp} (t - |\vec{d}|/c)}{4\pi\epsilon_0 c^2 d} \quad (3.11)$$

a_{\perp} is the acceleration projected on a plane perpendicular to \vec{d} . Combining equations 2.9, 2.10, 2.11 we get that the instantaneous scattered field is:

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

$$E_s(\vec{d}, t) = \frac{e^2}{4\pi\epsilon_0 m_e c^2} \frac{E_0 \sin(\theta)}{d} e^{-2\pi i \frac{|\vec{d}|}{\lambda}}$$

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

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

$$r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2}$$

3.2.2 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 2.2.2 illustrates the phenomenon of interference.

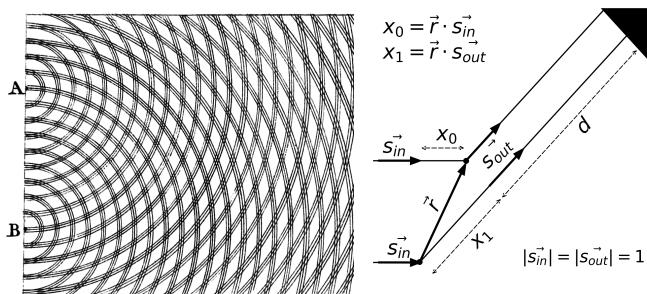


Figure 3.1. Illustration of scattering from two free-electrons A and B. a) graphic taken from Young, showing an interference pattern between two waves. b) mathematical description of the path length difference two waves arriving at point \vec{d} . \vec{s}_{in} and \vec{s}_{out} are unit vectors pointing in the direction of the incident wave and the outgoing wave. \vec{r} is the position vector of point A, compared to point B. Note that in a) $x_0 = 0$.

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 as the distance from both electrons to the detector is about equal: $|\vec{d} + \vec{x}_1| \approx |\vec{d}|$. This approximation is called the far-field approximation. The Fresnel number (FN) is used to verify the validity of the far-field approximation.

$$FN = \frac{o^2}{d\lambda} \quad (3.12)$$

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} can be described by a sum of the scattered electric fields:

$$E(\vec{d}) = E_s(\vec{d}) + E_s(\vec{d})e^{\frac{2\pi i \Delta x}{\lambda}} \quad (3.13)$$

Δx is the optical path difference between both scattered due to the relative difference in position. Δx can also be described using the difference vector \vec{r} :

$$\Delta x = x_0 - x_1 = \vec{s}_{out} \cdot \vec{r} - \vec{s}_{in} \cdot \vec{r} = (\vec{s}_{out} - \vec{s}_{in}) \cdot \vec{r} \quad (3.14)$$

By defining the scattering vector \vec{S}

$$\vec{S} = \frac{\vec{s}_{out} - \vec{s}_{in}}{\lambda} \quad (3.15)$$

the diffraction pattern of two electrons will be :

$$E(\vec{d}) = E_s(\vec{d}) + E_s(\vec{d})e^{2\pi i \vec{S} \cdot \vec{r}} \quad (3.16)$$

It is important to note is that Δx is at most the distance between the two electrons $|\vec{r}|$. 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 ($\lambda = 500\text{nm}$) 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.

3.2.3 Scattering from multiple electrons

In biological particles the incoming EM radiation is scattered by many electrons. The electric field can be described as the sum of all the individual scattered fields.

$$E(\vec{d}) = E_s(\vec{d}) \sum_n e^{2\pi i \vec{S} \cdot \vec{r}_n} \quad (3.17)$$

If all electrons together are assumed to form a continuous electric charge density around the many nuclei that constitute the biological particle, equation 2.17 can be written as a continuous function.

$$E(\vec{d}) = E_s(\vec{d}) \int \rho(\vec{r}, \lambda) \exp^{-2\pi i \vec{S} \cdot \vec{r}} d\vec{r} \quad (3.18)$$

$\rho(\vec{r}, \lambda)$ is named the scattering potential and takes the scattering strength of each nucleus into consideration.

So far we have implicitly assumed 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.

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

$$F(\vec{S}) = \frac{E(\vec{d})}{E_s(\vec{d})} = F\left(\frac{\vec{d}}{|d|}\right) \quad (3.19)$$

The structure factor is independent of the distance of the detector, it only depends on the angle of measurement. Equation 2.18 can be rewritten as:

$$F(\vec{S}) = \int \rho(\vec{r}, \lambda) \exp^{-2\pi i \vec{S} \cdot \vec{r}} d\vec{r} \quad (3.20)$$

This equation shows that a diffraction experiment is simply the 3D Fourier transform of the electron density evaluated at \vec{S} . This relation is very convenient as there is a large mathematical field describing the relations of Fourier transformations. A few of these relations are very useful to this thesis.

The Fourier shift theorem describes the change in the diffraction pattern upon object translation (i.e. shift).

$$\mathcal{F}(f(x + \delta x)) = e^{2\pi i \delta x q} \mathcal{F}(f) \quad (3.21)$$

The factor $e^{2\pi i \delta x q}$ is called a 'phase-ramp'.

The Convolution theorem states that a multiplication in Fourier space is a convolution in object space (real space) and vice versa.

$$f(x) * g(x) = \int f(\tau)g(x - \tau) d\tau = \mathcal{F}^{-1} \mathcal{F}f(x) \times \mathcal{F}g(x) \quad (3.22)$$

The Fourier slice theorem states that a 2D diffraction pattern is a slice through the center of the 3D fourier transform of the object. It has thus become possible to move between diffraction pattern and electron density.

3.2.4 Scattering Potential

An important term in equation that needs further development is the scattering potential of a material $\rho(\vec{r}, \lambda)$. Mathematically $\rho(\vec{r}, \lambda)$ can be formulated as a function of the refractive index n of a material.

$$\rho(\vec{r}, \lambda) = \pi s_{in}^2 ((n(\vec{r}, \lambda))^2 - 1)) \quad (3.23)$$

$$s = \frac{1}{\lambda}$$

The refractive index is related to the scattering factors of the individual atoms in the material.

$$n = 1 - \frac{r_e}{2\pi} \lambda^2 \sum_i N_i f_i(0) \quad (3.24)$$

where r_e is the classical electron radius, λ is the wavelength, N_i is the number of atoms per unit volume. The complex scattering factor in the forward direction $f(0)$ can be described by:

$$f(0) = f_1 + i f_2 \quad (3.25)$$

The imaginary part is derived from the atomic photoabsorption cross section and is a measure of absorption [booklet]. The real part is related to the imaginary part using the Kramers-Kronig dispersion relation [Kramers]. At high photon energies f_1 approaches the atomic number of the atoms. The exact value of f_1 and f_2 at different photon-energies have been determined experimentally [Henke].

Figure 2.2.4 presents the scattering factors of two elements: carbon and oxygen. The gray energy range from 282-533 eV (4.40 nm - 2.33 nm) is called the water window. In this region, and especially towards the end, oxygen atoms (thus water) are scattering significantly less than carbon atoms (i.e. organic biomolecules). The water window therefore a good candidate for imaging cells, as the contrast between two of their main constituents, water and biomolecules, is enhanced.

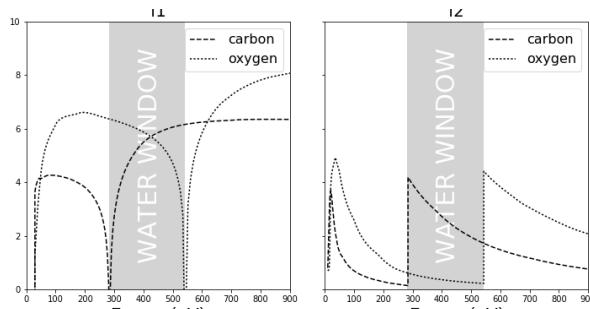


Figure 3.2. Atomic scattering factors f_1 and f_2 of carbon and oxygen, as a function of photon-energy of the X-ray radiation. The region between the absorption K-edge of carbon (282 eV, 4.40 nm) and the K-edge of oxygen (533 eV, 2.33 nm) is called the water window. The contrast between biomolecules and water is enhanced in this region, especially towards the high end of it. These wavelengths could be used for imaging living biological particles such as cells or organelles.

3.3 Radiation Damage

The previous section shows that scattering power of an body is related to the wavelength of the radiation. The shorter the wavelength the less scattering occurs. To compensate for this loss in 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 (a dose of 2000 Gy is already enough to kill everything that lives on earth). Unfortunately other processes of matter-radiation interaction become more abundant as well. At soft X-ray wavelengths the predominant damage process is photoionisation. Many process follow the ionization, of which non-radiative Auger decay will lead to significant radiation damage. The relation between the required dose of radiation versus the maximum tolerable dose of radiation posed a physical limit to the obtained resolution for imaging single biomolecules [Howell].

3.4 Diffract-before-destroy

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. This principle is of diffract-before-destroy, suggested in 2000 by Neutze et al [], was experimentally demonstrated on a silicon-nitrate sample [Chapman]. This experiment used the ultra short and extremely bright pulses from the first x-ray 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. A single XFEL pulse has a power densities of $10^{16} W/cm^2$, and can be between 1-100 fs long. This facilitates the imaging of single particles at room temperature.

Diffract-before-destroy has been further validated by a wide variety of experiments. 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 []. The imaging of single particle has been shown feasible in 2D for a wide variety of biological particles, ranging from *living* cells (Paper I) to cell organelles (Paper 7) to 40 nm viruses (Paper 8). Paper VI shows the experimental feasibility of using many 2D images to construct a 3D model of a giant virus.

The field is very close to measuring the first 2D diffraction patterns of single proteins. For simulated data it has been shown that such patterns can be used to recover the 3D shape of proteins, even though the number of scattered photons might be as low a 1000 photons per frame. The ultra short pulses also allow for the studying of the dynamics of such systems, either by separating the images from different naturally occurring conformational states, or by

identifying structural changes subsequent to an external trigger []. In the more distant future these structural changes might be tracked inside living cells.

4. Pulse Generation by XFEL

Free-electron lasers (XFELs), invented by J. Madey and demonstrated experimentally in the 1970s, can generate multi gigawatt and femtosecond (fs) x-ray radiation. Compared to fourth generation synchrotrons, this is a ten billion-time increase in brightness. Typically an XFEL consists of three sections: i) the electron gun, ii) a linear accelerator, and iii) an undulator. The x-ray radiation is generated in the undulator. Figure 1 shows a schematic of the XFEL at the Linear Accelerator Coherent Light Source (LCLS) at Stanford. This chapter will describe the operating of the LCLS in specific. Other XFELs might use different hardware, but the fundamental principles that drive an XFEL will be very similar. Please note that the words bunch and pulse are used to discriminate between a collection of electrons and radiation respectively.

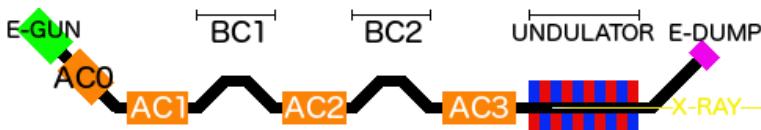


Figure 4.1. Schematic of LCLS. Electron bunches are created by the electron gun (E-gun). The bunches are consecutively accelerated by linear accelerators (AC0-3) and compressed by bunchcompressors (BC0-1). After the final acceleration the bunch is guided into the undulator, where an x-ray pulse is generated. After the undulator the x-ray pulse continues to the beamline, and the electron bunch is dumped (E-dump).

In the electron gun, a copper plate is irradiated by a powerful ultraviolet laser pulse, which releases a bunch of electrons. At this moment the electron bunch is a few picoseconds long and has a peak current less than 100 A. As shown in figure 3.1a, the energy spread within the bunch is minimal at this point. The electron bunch is directly accelerated to 99.7% of the speed of light, using a standing wave in a RF cavity (AC0). Due to cooling reasons in AC0 the LCLS produces electron pulses at maximally 120 Hz.

4.1 Klystron

Consecutively the bunches are guided in a long series of accelerators named klystrons. See AC1-3 in figure 3, and for more detail figure 3.1d. In essence a klystron is a cylinder on which a strong MHz alternating current is applied. The AC current induces an EM field in the MHz range (RF), similar to the way radiowaves are produced in an antenna. Due to the spacing

between irises the relative phase velocity of the wave is matched to the speed of the electrons, thus through the Lorentz force longitudinal acceleration can be achieved throughout the whole klystron. In the center of the cavity the magnetic field is zero, the acceleration is therefore proportional to the strength of the electric field. 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 15 GeV electron bunches at least 500 m of klystrons are needed. In reality the accelerating unit of LCLS are 889 m long. Often accelerators do not operate at the point of maximum acceleration, since one wants 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 b and d. A negative energy gradient means that the electrons located at the front of the bunch have a slightly lower energy compared to electrons located at the rear of the bunch.

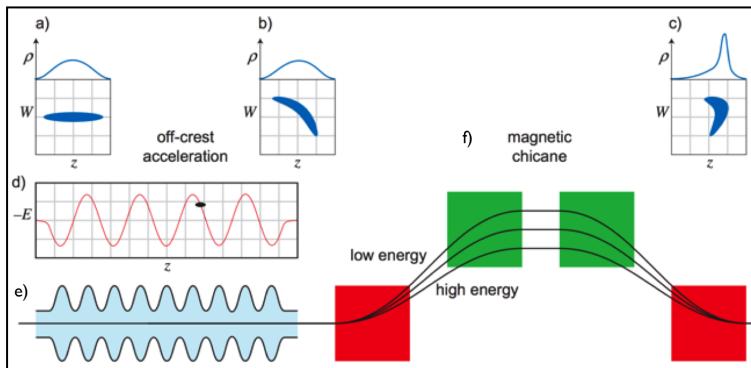


Figure 4.2. Schematic of Acceleration and bunch compression at LCLS. a-c) change of bunch shape throughout the accelerator. d) illustration of the off-crest acceleration principle. a) shows that this creates an energy chirp. e) schematic of a single klystron. f) schematic of a chicane. The green and red blocks are magnetic dipoles that deflect the electron bunch in opposite direction. The path of high energy electrons diverges less than low energy electrons. If the bunch has a negative energy chirp, this results in bunch compression.

4.2 Bunch Compression

Short and compact bunches are essential for the functioning of an XFEL. One of the main bunch-compression devices are chicanes. Chicanes consist of four magnetic dipoles that diverge the high energy part of the bunch 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, has a peak current of 4.3 kA, and travels at 99.999% of

the speed of light. The relativistic $\gamma \sim 2000$. Creating these type of bunches is a noteworthy technical achievement,

4.3 Undulator

4.3.1 Monochromatic X-ray radiation

The undulator is the heart of an XFEL. It is a periodic arrangement of magnets with alternating poles. The magnetic field of a periodic magnetic structure can be written as:

$$\vec{B}(z) = B_0 \cos\left(\frac{2\pi}{\lambda_u} z\right) \hat{\mathbf{y}} \quad (4.1)$$

B_0 is the magnetic field, λ_u is the wavelength 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 (x). The movement can be described by the following equation.

$$v_x = \frac{-e B_0 \lambda_u}{2\pi m_e \gamma} \sin\left(\frac{2\pi}{\lambda_u} z\right) = \frac{Kc}{\gamma} \sin(k_u z) \quad (4.2)$$

Usually the non-dimensional parameter K is written as:

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

where the magnetic field is measured in Tesla and the undulator period in centimeters.

Preserving momentum the oscillating electrons emit radiation (λ_s). As described in chapter 1, interference effects will occur between the emitted waves from the different charges (remember the electrons come in a bunch which has a size). 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 an integer of λ_s , the emitted waves will add coherently. Radiation with other wavelengths will interfere destructively. The longer an undulator the more pronounced the selection of specific wavelengths. Equation 3.3 describes this so-called resonance condition.

$$c \frac{\lambda_u}{< v_z >} - \lambda_u \cos(\theta) = n\lambda_s \quad (4.3)$$

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

¹Using Pythagoras' theorem

we can write v_z as:

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

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 $\langle v_z \rangle$.

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

$\gamma \gg 1$, thus we can expand the root $\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \dots$ (first order Taylor expansion).

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

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:

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

The geometrical series $\frac{1}{1-x} = 1+x+x^2+\dots$ allows us to obtain the well known undulator equation.

$$n\lambda_s = \frac{\lambda_u}{2\gamma^2} \left(1 + \frac{K^2}{2} + (\gamma\theta)^2\right) \quad (4.4)$$

This formula can be explained by two relativistic effects. For an observer it is the electron bunch that is moving close to the speed of light. In the electron bunch's frame of reference however it is the undulator that is moving very fast. Fast objects are length contracted, thus the electrons observes a undulator period of $\frac{\lambda_u}{\gamma}$, which makes it emit radiation with λ around $\frac{c\lambda_u}{\gamma}$. The radiation is relativistically contracted when observed in the laboratory's frame of reference (relativistic doppler effect), adding the second gamma term. The size of relativistic doppler effect is dependent on the angle of observation relative to the direction of motion.

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. LCLS can produce X-ray radiation in the range of 400 eV -10 keV.

4.3.2 SASE

So far the emitted radiation is monochromatic, but not temporally coherent. This means the total emitted power is proportional to the number of the elec-

trons in the bunch (N_e). If all emitted radiation would emit in phase, the total irradiated power would be proportional to N_e^2 . Pellegrini and others realized that for short pulses, and long undulators, temporal coherence could be achieved by a process called Self Amplified Stimulated Emission (SASE). SASE exploit the slight non-uniformities in the electron bunch, that cause radiation with a certain phase to be slightly more prevalent than others. The Coulombic force exerted by the electric field of this radiation will start grouping electrons at its nodes (See figure 2). Electrons are accelerated when the electric field is positive, and decelerated when negative, causing the so-called microbunches to appear. This process forms a positive feedback loop, causing a exponential increase in radiation power. The exponential growth cannot continue indefinitely, and the power must saturate at a certain level. The SASE process ultimately saturates when either the electrons loose so much energy the resonance condition is not met anymore, or when the energy of the radiation surpasses that of the electron bunch. In the latter case the EM wave will start to transfer energy to the electron bunch instead².

The high power density within the pulse means that there are more than one photon in the volume of a λ^3 . Which means that so far diffraction patterns come from single photons interfering with themselves []. Interference between more than one photon might be different. For example there might be special interactions between photons.

4.4 Improvements

Seeding Multi-angle High rep rate

²This phenomenon is similar to inverse Compton scattering, in which electrons absorb energy from the radiation field.

5. Sample Introduction

This chapter will discuss different methods of delivering sample into the x-ray pulse. Sample delivery can be divided into roughly two classes: substrate-bound and substrate free sample delivery. Figure 4 shows a schematic of a general set-up of an experiment, including several sample delivery systems. Usually one sample delivery system is chosen, which is aligned such that the sample and the beam intersect. If the x-ray pulse intersect with the sample (and/or substrate) a diffraction pattern is recorded on one or two sets of detectors, placed at different distances from the interaction region.

It is important to note that the whole system is under reduced pressure, 10^{-6} mbar. This is done to reduce the background scattering from the gas molecules themselves. Pumping a chamber down to 10^{-6} mbar takes hours, and is a factor which has to be taken into account in the planning of the experiment.

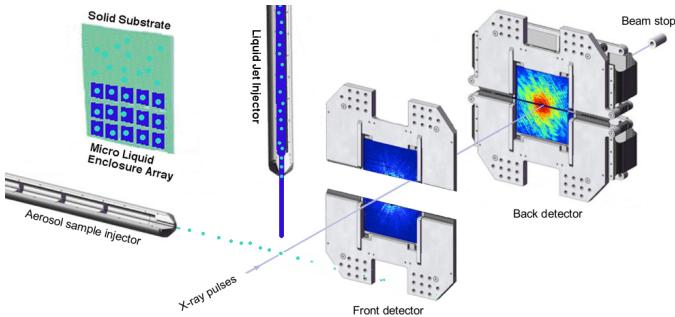


Figure 5.1. Experimental geometry, including different sample delivery system. From the bottom left the x-rays pulses enter the chamber. The experimentalist can now select a sample delivery system. Shown are an aerosol sample delivery system, liquid jet injection system, and two types of substrate bound sample delivery systems. If the beam intersect the sample (and/or substrate) a diffraction pattern is recorded on two sets of detectors, placed at different distances from the interaction region.

5.1 Substrate-bound Sample Delivery

5.1.1 Solid substrate

One of the easiest ways of introducing samples is by attaching them to a solid surface. Shortly after the operation of the first XFEL (FLASH) the principle

of diffract-before-destroy was demonstrated on a inorganic sample attached to a fixed membrane [Chapman]. Soon after the first cells were imaged on a silicon-nitrate window. The main disadvantages of this sample delivery system are i)the sample has to be able to tolerate prolonged exposure to reduced pressure, ii) the substrate causes significant background scattering, and iii) the repetition rate is limited by the time it takes to mechanically move the solid substrate from one position to the next. The latter becomes a significant problem at the high repetition rate XFEL sources.

5.1.2 Micro Compartments

In order to circumvent the problem of the prolonged exposure to the vacuum, a system called Micro Liquid Enclosure Array (MLEA) can be used. A MLEA system consists of an array of micro-compartments filled with liquid attached to a regular grid (see figure 4. The microcompartments are closed off with by a membrane. This method has been successfully used in the imaging of cells. Getting your sample into the microcompartments is tedious and time consuming. The liquid in the microcompartments produces significant background scattering.

5.1.3 Liquid Jets

Liquid jet systems can produce stable streams of particles in solution. Mainly two types of liquid jet nozzles are used for sample injection into the XFEL pulse train: i) liquid nozzles (Rayleigh jets), ii) liquid nozzles assisted by coaxial gas flow also referred to as gas dynamic virtual nozzle (GDVN).

Vacuum compatible systems of the Rayleigh jets with diameters ranging approximately from 5 to 100 μm exist. Typical liquid flow velocities of such jets is 120 m/s. The sample consumption within such a jet is in the order of ml/s. For many biological samples this rate of sample consumption is too high. Sample consumption can be reduced by slowing down the stream or by reducing the diameter of the nozzle. However, for orifice diameters smaller than 10 μm , clogging of the nozzles becomes a serious issue. To reduce the jet diameter without the danger of clogging the sample delivery, GDVN liquid jets were introduced [DePonte]. In a GDVN nozzle a coaxial stream of high velocity He gas is used to accelerate the liquid stream. Similar to the flow of water from a tap, the accelerating stream is reduced in diameter. Jet diameters below 1 μm have been recently achieved [W].

Liquid jets have the advantage to study biological samples in their natural environment, to refresh the sample throughout the experiment with sample consumption in the order of 10 $\mu\text{l}/\text{min}$ [Weierstall 2012], and allows for imaging at high repetition rates. This GDVN setup is commonly used in serial femtosecond crystallography (SFX) experiments. The Bragg peaks of nano

crystals are sufficiently strong to surmount the background scattering from the surrounding jet.

5.2 Substrate-free sample delivery

In FXI everything that is illuminated by the X-ray pulse is sample, including the structure of the sample holder, the liquid column of a liquid jet, or materials that make up microfluidic devices. Such sample holders contribute to scattering and increase unwanted noise. Especially for weakly scattering single particles the noise might drown the signal. Aerosol injection removes this clutter and assures that the sample is clearly isolated from its surroundings, and this is, as we see in a later chapter, important for image recovery.

An aerosol injector produces small droplets of particles dissolved in a volatile buffer, typically ammonium acetate. The buffer evaporates under the reduced pressure, 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, and electro spray ionisation (ESI).

5.2.1 Gas Dynamic Virtual Nozzle

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 [Calvo].

$$d_j^{(GDVN)} = 2\sqrt{Q \cdot \left(\frac{\rho}{2\pi^2 \Delta P}\right)^{1/2}} \quad (5.1)$$

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 shown to work well for low viscosity media such as water. The GDVN can create droplets in the size range of 400 nm - 2000 nm. Sample consumption is in the range of $\mu\text{l}/\text{min}$.

5.2.2 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). This is a series of cylindrical cavities, connected by co-aligned orifices, that collimate the droplets into a narrow beam of particles [].

For robust particles, which are large compared to the drop size, this type of sample injection has shown to be very successful. This sample injection method really opened up the possibility to measure the diffraction pattern of isolated single particles at high signal-to-noise ratios at very high repetition rates. Up to 80% hit ratios of the 120 Hz repetition rates of the LCLS have been recorded [Hantke]. The results discussed in this thesis come from datasets that were generated using the GVDN 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 []. 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 []. Other sample cell lines such as E coli and brewers yeast, or many types of viruses have shown to be viable after injection. Nevertheless, not all sample types may be amenable to aerosol sample injection, and any sample type should be tested prior to experiments.

5.2.3 Electro Spray Ionisation

Over the last years it became apparent that GDVN sample injection does have its limitations. If the particle becomes small compared to the drop size, wide size distributions of otherwise uniformly sized particles were measured [Daurer, Hantke]. These observations might be explained by a combination of an incomplete evaporation process, and the build up 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 as happens in a GDVN. The characteristic $d_j^{(ESI)}$ can be described as a function of the surface tension σ , vacuum permittivity ϵ_0 , electrical conductivity C , as well as the flow rate and the density of the liquid.

$$d_j^{(ESI)} \approx 2\sqrt{Q \cdot \left(\frac{\rho \epsilon_0}{\sigma C}\right)^{1/3}} \quad (5.2)$$

After the process of evaporation initiates an electric potential starts to further build up in the drops, eventually leading to a coulomb explosions of the drops [Cole]. Cycles of evaporation 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. ESI can be combined with an aerodynamic lens stack as well.

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, allowing to reduce sample consumption to 10 nl/min [1]. This sample consumption rate is compatible with samples that are only available in small quantities.

5.2.4 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. 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 diffraction patterns from the real particle from diffraction patterns from potential contamination.

6. Data recording

After the interaction between light and our material we have to measure the scattered waves. Equation shows that the scattered wave is determined by the complex amplitude $F(S)$. The real part of $F(S)$ is called the amplitude, and the imaginary part of $F(S)$ is called the phase (ϕ). Currently there no device is capable of measuring the phase of x-rays directly, as it changes in the attosecond range. The former can be measured, but based on the low probability of interaction between light and matter, as well as the vacuum condition in which the detector is placed, this is not trivial. In the satellite mission XMM-Newton of the European space agency, two types of x-ray detectors were used: the CMOS and the pn-junction charge coupled device (pnCCD). The experiments described in this thesis used a modified version of the original pnCCD type detector.

6.1 the pnCCD detector

The exact working of a pnCCD detector requires an understanding of semiconductor physics, which goes outside of the scope of this thesis. A pnCCD consists of three layers of material. First a layer of ... protecting the other layers from the high vacuum. Secondly there is a layer of high-purity n-type silicon. In this layer the incoming x-ray photons are absorbed and electrons are released from the bulk material through the photoexcitation. A third layer consisting p-type silicon is placed behind the second layer. By applying a positive charge to the layers electrons move towards the back of the detector where they are stored in potential wells. At the frame rate these wells are depleted, and the current is amplified and then measured.

6.2 Quantum Efficiency

An important parameter describing the pnCCD is the quantum efficiency (QE). This parameter describes the fraction of incident photons converted to charge carriers. Often this is close to 100% but for each material the QE varies as a function of wavelength. A high QE improve the signal to noise ratio, which becomes important in low the low signal areas of the detector. record the diffracted signal in that low scattering parts of the detector. High energy X-ray radiation (larger than 5keV) would require another type of detector than the pnCCD.

6.3 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. In the center of detector a hole is created to let the high-intensity direct beam through (each detector halve misses a semicircle). In this area there is no intensity information present.

6.4 Saturation

Besides missing data due to the detector geometry, detector saturation might also lead to missing information. The potential wells described in 5.1 can maximally hold a certain amount of charge. If more charge is present (many photons scatter in them, the excess charge overflows to neighboring wells, making it impossible to accurately use the original and the affected wells. If the amount of charge is very high, it might even damage the detector itself. If no bragg peaks are present on the detector, all saturation will occur in the center of the detector.

7. Phase Retrieval

The problem of 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. Holography uses the interference between two wave fields to obtain the phases. Ptychography uses the precisely known overlap and high redundancy between many exposures to solve the phase problem. CXI makes use of the fact that pnCCD detectors so-called oversample the diffraction pattern, which means that phases can under certain conditions be retrieved from the intensity pattern itself. The next sections will describe oversampling in more detail, explain how phases can be retrieved from an oversampled signal using an iterative phase retrieval method, and finally how phase retrieval can be validated.

7.1 Oversampling

In 1952 Sayre noticed that the Bragg peaks sample the molecular transform $\mathcal{F}(H)$ at the critical sampling rate (S_C)[71] (See Shannon [75]). This means that knowing the phases belonging to the Bragg peaks is just enough to back-calculate the structure of the measured object. Single particle imaging is free from the crystal lattice, thus isolated Bragg peaks give way to 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. The sampling rate S is determined by the wavelength λ , the detector distance d the pixel size p , and the size of the object.

$$S = \frac{\lambda d}{s p} \tag{7.1}$$

In single particle imaging we do not measure the phases, and loose half of the information necessary to determine the molecular transform at the sampling point. By choosing a detector setup such that the sampling rate is at least twice the critical sampling rate we can use the additional intensity information to recover the phases. It would thus be possible reconstruct the object from the measured intensities alone. Oversampling is the basis of many phase retrieval techniques in SPI, as well as some clever phasing techniques in SFX [7, 21].

7.2 Iterative Phase retrieval

In practice there are many ways of actually retrieving the phase information from the diffraction pattern, but most common phase retrieval techniques are variations of convex optimization algorithms. In this section the general idea behind convex optimization is explained.

As can be seen in Figure oversampling in Fourier space implies that there is an area the size of $(l - o)$ around the object for which we know the electron density is zero. This knowledge can be used as a constraint: we know that a correct choice of phases would make the corresponding $\rho(r)$ be zero in this area. The area that can contain positive electron density is called the support mask M . Furthermore, we know the recovered Fourier amplitudes should agree with the measured intensities.

An iterative approach can be designed to find the solution that minimizes the disagreement to both constraints:

1. Assign random phases to each pixel.
2. Inverse Fourier transform $F(s)$.
3. Set all electron density outside of M to zero, keep other electron density.
4. Fourier transform $\rho(r)$
5. Make sure that the recovered amplitudes match the measured intensities.
Keep the phases.
6. Go back to step 2.

A way to formally write down these steps is found in constraint of step 3 and 5 can be found in equation 6.2.

$$P_r \rho(\vec{r}) = \begin{cases} \rho(\vec{r}) & \text{if } \vec{r} \in S \\ 0 & \text{if } \vec{r} \notin S \end{cases} \quad (7.2)$$

$$P_f \rho(\vec{r}) = \mathcal{F}^{-1} \left(\frac{\sqrt{I}}{|F(\rho(\vec{r}))|} \mathcal{F}(\rho(\vec{r})) \right) \quad (7.3)$$

Two error metrics can be associated with both constraints; the fourier error (E_f) and the real space error (E_r). E_f and E_r are defined as:

$$E_r = |P_r \rho(\vec{r}) - \rho(\vec{r})| = \left(\sum_i \rho_i^2 \right)^{\frac{1}{2}} \quad (7.4)$$

$$E_f = |P_f \rho(\vec{r}) - \rho(\vec{r})| = \left(\sum_i \left(\frac{\sqrt{I_i}}{|F(S_i)|} - F(S_i) \right)^2 \right)^{\frac{1}{2}} \quad (7.5)$$

7.2.1 Error Reduction

Convex optimization assumes that hyperplane of all possible solutions that forfill to one of the constraints is convex. A most straightforward implementation of the above-described algorithm is Error Reduction, invented Fienup based on algorithm by Gerchberg, Saxton. would in this case be enough to retrieve the true solution.

7.2.2 HIO

It has been shown that the solution set that for fills the support constraint is convex. Due to noise and missing data our solution set however are often not convex, which means that there are multiple (local) minima. In this case ER will likely not find the global minimum or true solution. Other algorithms had to be invented. Historically an algorithm named HIO was one of the first that can escape from local minima. This is achieved by slightly adjusting step 3.

$$P_r \rho(\vec{r}) = \begin{cases} \rho(\vec{r}) & \text{if } \vec{r} \in S \\ \rho(\vec{r}) - \beta P_f \rho(\vec{r}) & \text{if } \vec{r} \notin S \end{cases} \quad (7.6)$$

7.2.3 RAAR

Another algorithm that is often used is Relaxed A Raar

Other algorithms that are often used for Phase retrieval are diffusion map (DM), GHIO, .

7.3 Validation

In iterative phase retrieval every reconstruction results in an image, irrespective of it having biological relevance or not. It is therefore very important to have tools to validate a reconstruction. This section will describe different validation tools used in the field.

7.3.1 Errors

The most basic method to assess the difference in quality between two reconstructions is comparing the respective errors. The reconstruction with lower errors is generally considered to be a more successful reconstruction. This method however does not tell anything about the biological validity of a reconstruction and is most often only used to exclude outliers (reconstructions that have clearly failed).

7.3.2 PRTF

The standard tool to assess the quality of a reconstruction is the phase retrieval transfer function (PRTF). This function considers the variation within a set of independent reconstructions (each reconstruction starting from a random set of phases), and uses this variation to quantify the resolution of a reconstruction. The basic assumption behind the method is that if a similar object is retrieved repeatedly, this object is supposed to be similar to the true object. The variation between reconstructions is calculated for each pixel i by adding all recovered amplitudes for that pixel and dividing the total vector by square root of the measured intensity for that pixel $v_i = \left| \frac{\sum A_i}{\sqrt{I_i}} \right|$. If the value of v_i is close to unity all reconstructions recovered a similar amplitude for pixel i . The closer v_i is to zero the more difference there is between individual reconstructions. The standard PRTF plot shows the radial average of v_i . A common practice is the field is to define the resolution of a reconstruction as the first time the 1D PRTF drops below $\frac{1}{e}$. This threshold is arbitrary. For spherical objects a periodic drop in the 1D PRTF, corresponding to a drop in measured intensity, can be observed. This drop can be explained by the fact that the phase of a pixel measuring 0 intensity is unconstrained.

7.3.3 Missing Mode Analysis

As described in previous sections diffraction patterns lack data. Even in the best case the central region will be missing because the direct beam will otherwise damage the detector. The main other sources of missing data are detector geometry and saturation.

Reconstruction algorithms deal with this missing data by recovering the phase as well as the amplitude for the pixels in these regions. In many cases the missing data does not affect the stability of phase recovery, however this cannot be said in general. For this we have to note that the Fourier constraint does not limit 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 could be arbitrarily added to the solution and thus would form an ambiguity in the reconstruction process.

In general completely unconstrained objects do not exist. Objects that fit the support constraint and only slightly contribute outside the missing data region, i.e. weakly constrained objects, can however exist. In the design of an experiment, or when deciding whether or not to phase an object it is important to have a tool that predicts whether weakly constrained objects exist or not. Missing mode analysis is such a tool.

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. The

rows and columns can be reordered according to whether they conform to the fourier and real space constraints or not.

$$\mathcal{F} = \left(\begin{array}{c|c} \mathcal{F}_{SM} & \mathcal{F}_{\bar{S}M} \\ \hline \mathcal{F}_{S\bar{M}} & \mathcal{F}_{\bar{S}\bar{M}} \end{array} \right) \quad (7.7)$$

We are after objects that minimize $|\mathcal{F}_{SM}\rho|$. A common method to find such objects is singular value decomposition [Eckart]. This method decomposes $|\mathcal{F}_{SM}\rho|$ into a diagonal matrix Σ , and two unitary matrices U and V in the following way.

EXPLAIN EIGENVALUES AND BASIS

7.3.4 Hierarchical clustering

To test whether the Fourier Error and Real space error are good measures to assess the variance between individual reconstructions, a clustering analysis can be performed. An example of such an analysis is UPGMA (Unweighted Pair Group Method with Arithmetic Mean) hierarchical clustering [Sokal, Michener]. This method checks the similarity between reconstructions by making pairwise pixel-to-pixel comparisons between all reconstructions. The similarity score is the normalized scalar product between a pair reconstructions after translating them to their optimal fit. We plot the similarity score of each cluster as a function of the number of clusters and choose the agglomeration step where the plot makes a "kink". This is a standard way to estimate the number of clusters present in the set of reconstructions. This capability is an advantage of the UPGMA clustering algorithm. In a 2D plot of FE vs RE, color-coded by cluster, it is possible to see whether or not there is a correlation of reconstruction similarity and error score. An illustration of this method can be found in 6.3.4

In the case that several large clusters remaining 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.

7.4 Simulated Phase Contrast Methods

Once one has retrieved the phase of the diffraction pattern, one has complete knowledge of the information encoded in the said wave-field. This total

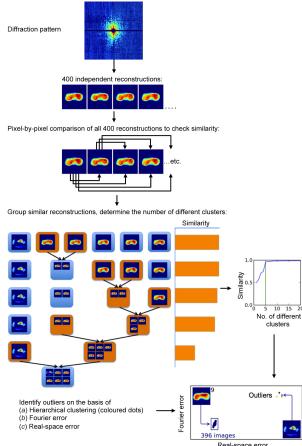


Figure 7.1. UPGMA Clustering.

knowledge is powerful, because it allows the emulation of the action of any imaging system for which an associated mathematical transform can be written, regardless of whether that system is experimentally realizable in hardware.

For example, differential interference contrast imaging or Nomarski imaging is experimentally a well established technique to visualize the changes in phase. This change can 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 [mitochondrion].

Nomarksiki imaging in its simplest form takes a coherent wave-field of interest, say $\psi(x, y, z = 0)$, and then interferes this wave-field with a copy of itself that has been given both a slight transverse displacement ($\Delta x, \Delta y$) and a phase shift ϕ_0 (Nomarski and Weill, 1955). Thus the intensity of the resulting wave-field is [Paganin]: $|\psi(x, y, z = 0) + e^{(i\phi_0)}\psi(x - \Delta x, y - \Delta y, z = 0)|^2$, from which one can show (using the Fourier shift theorem) that the transfer function becomes:

$$T_{DIC}(k_x, k_y, \tau) = 1 + e^{(i(\phi_0 - k_x \Delta x - k_y \Delta y))}, \quad \tau = (\phi_0, \Delta x, \Delta y) \quad (7.8)$$

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

$$N = \mathcal{F}^{-1} T_{DIC} \mathcal{F}(\rho(\vec{r})) \quad (7.9)$$

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

7.5 Shrinkwrap

While HIO and RAAR perform well when the support is tight and well known, phasing can practically become impossible if the support is too large. In 2003 Marchesini developed an algorithm that does not require a priori known support as input, but instead tries to deduce the shape of the support during the reconstruction. It starts by guessing the support from the autocorrelation. It does so by blurring the autocorrelation and selecting all pixels above a certain threshold to be included in the support. Consecutively 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. By varying the amount of blur and/or the selection threshold as the iterations progress, the support can slowly wrap and shrink to the true shape of the object. The general idea behind the algorithm is that even with a non-accurate support some features will be recovered well, and by using these features finally the an accurate support will be found. The algorithm has been very successful for experimental data [].

8. 3D restrictions

So far we have dealt with 2D diffraction patterns. Although much can be learned (and has been learned) from 2D images, having a structural 3D model is essential to understand the functioning of biomolecules. From the Fourier slice theorem we know that a diffraction pattern is a 2D slice through the center of the 3D diffraction pattern of the object. Different slices could therefore build up the 3D fourier transform of the object. In the current CXI setup it is impossible to image one object multiple times, making it difficult to reconstruct the complete 3D diffraction volume from one particle alone. If the particle of interest is reproducible the diffraction pattern from each copy could be used to construct the 3D fourier transform of the object. This does pose a new problem: the relative orientation of each diffraction pattern compared to the others has to be recovered. A variety of reconstruction algorithms have been proposed for recovering the relative orientation of diffraction data, including the EMC algorithm [Loh], the manifold embedding algorithms [Giannakis], simple best-fit models [Tegze], and multi-particle cross-correlation analysis [Saldin, Kirian, Kam, Kurta]. Theoretical studies suggest that the determination of diffraction pattern orientation should be possible even with very low photon counts of as little as 1000 scattered photons per image [Loh], making EMC and ME very suitable to recover the 3D structure of sparsely scattering single proteins.

8.1 the EMC algorithm

The EMC algorithm is an iterative program. On the one side you have the 2D measured intensity patterns (IM_i), and on the other side a 3D model of the Fourier space of the object (M). Each iteration starts with the expand step in which the 3D model is expanded in n central 2D slices (IP_i). N is dependent on the sampling rate of the 3D model. Each of the slices are compared to each diffraction pattern, using a model for the noise, to predict a probability score that the measured diffraction pattern is coming from the model diffraction pattern. To paraphrase: imagine that you would image each particle at the same orientation, the measured diffraction patterns will not be identical, since the noise is varying from shot to shot. Each pixel however will have an associated distribution of measured intensity values. Based on this distribution, which can for example be Poissonian in the case of pure shot noise, one can assign a probability that IM_i is coming from slice IP_i . In a good model not only will

the measured values for a given 3D pixel (voxel) be similar for different IM , the distribution of measured values will have to be similar to the known noise distribution. This is a very subtle idea and utilizes the data very efficiently.

Comparing all IM_i to each IP_i , lead to n probability distributions predicting the likelihood of the orientation of the measured diffraction patterns. This step is called the maximize step. Using the probability distributions as weights, a new model M_{n+1} is generated from the measured diffraction patterns in the so-called compress step. The initial model M_0 can be a random model. After one iteration a similar patterns will end up in similar orientations which hopefully will make M_1 a bit more similar to the real Fourier pattern of the particle. After repeating this Expand-Maximize-Compress cycle multiple times a realistic model might be generated.

8.2 Non-reproducible objects

The addition of diffraction patterns from non reproducible object (i.e. incoherent addition) is very problematic in Fourier space [Maia], as all features of the resolution of variation will be washed out. This makes it very difficult to retrieve 3D information from an heterogeneous object such as cells and the fold of many proteins. To tackle the heterogeneity problem many adaptations of existing algorithms or new algorithms have been developed.

8.2.1 Multi-model EMC

Currently a version of EMC is on its way that can generate multiple models of the same particle. It is very similar to single state EMC, but differs in the fact that it build multiple self-consistent models.

8.3 the Manifold Embedding algorithm

8.3.1 Tomography and sub-tomogram averaging

Tomography used a set-up in which the object is imaged from m different but known orientations. Sometimes this is enough to gain enough information about the object in 3D. Crowther's criterion describes the relation between resolution and number of required exposures (m). For large objects and/or high resolutions it becomes difficult to have enough exposures to recover the full object at atomic resolution. In this case one could use multiple low resolution parts of the tomogram (sub-tomogram) to build up a model of semi-reproducible object in a method similar to EMC. The limit of the number of beams (m) for which this is possible in an XFEL setup is currently under investigation.

8.3.2 Holography

At high resolutions the detector does not measure the Ewald sphere and the Friedel symmetry of the diffraction pattern breaks. This is often very visible at longer wavelengths (<400 eV). This curvature can be exploited to resolve semi-3D information about the object, but the extend at which this is possible is disputed. In some cases much is known about the possible shape or buildup of the particle and then it is possible to reconstruct a 3D structure [Miao, Fennel, ...]

9. Image Classification

In order for algorithms such as EMC to work, the amount of heterogeneity within the diffraction data set has to be limited [TOMAS is there an article describing this?]. Furthermore, each of the steps in the experiment introduces its own type of noise to the measured diffraction pattern. For example think of the debris clumping around small particles compared to the drop when sprayed with the GDVN, detector malfunctioning or saturation, or intensity fluctuations due to the random start of the SASE process. Often the results of the first two types of noise can not be tolerated by EMC, and image classification before the EMC step is necessary. Also fast feedback about the first type of noise is very useful to have during the experiment. So far a very robust sizing method has been developed, but more extended methods might come in useful. Several methods shown here can give rapid feedback on the heterogeneity of the particle.

9.1 Template-based classification

If your object has a known shape it could be possible to only select the diffraction patterns that are similar to a set of expected diffraction patterns from the object called templates. Paper III explores the possibility of template-based classification. In general this method is highly dependent on the choice of template, as well as the amount and type of variation present in your sample.

9.2 Feature extraction

Another way of selecting diffraction patterns is by extracting general features from the diffraction pattern such as size, particle shape, amount of saturation, number of particles in the beam. Based on the relative values associated with the features patterns might be selected or not, or the experimental conditions might be rapidly adjusted. This section describes several methods to extract features.

9.2.1 Size

A common method to determine the size of an object is fitting the central speckle to the central speckle of simulated diffraction pattern from a sphere.

This method has shown to be successful for particles that have an icosahedral shape to spherical shape [41, 24]. If the diffraction pattern also has the 3rd to the 5th minima present, the average of these minima can also be used to describe the size of the object. This method might be especially useful for strong scatterers, because the central speckle or first minimum might not be reliable due to saturation effects.

It might also be possible to determine the size and shape of an object from a filtered autocorrelation. This is useful when the object cannot be approximated by a sphere, but is rather elongated in one axis. Due to aerosol injection, the strongest contrast in the object is coming from the edge of the object. Therefore the edge is often one of the most clear features in an autocorrelation. This method is limited to convex objects.

9.2.2 Multiple scatterers in the focus

9.2.3 Edge detection

Some objects are characterized by having sharp edges. A sharp edge in real space corresponds to a streak in the direction perpendicular to the edge in Fourier space. Objects and/or orientations of objects might be classified by determining if streaks are present, and how many. Figure explains the idea behind a streak finding algorithm described in paper IV.

10. Software

10.1 RedFlamingo

Part IV:

Results

11. Results

11.1 Imaging live cells

For the study described in paper I we used cells of two species of cyanobacteria as a sample (*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. They play an important role in the global carbon 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 before cyanophages caused the bloom to collapse.

Cyanobium gracile cells were selected for their small size, their robustness with respect to the injection procedure, and convenient autofluorescence properties to assess cell viability. Solitary *C. gracile* and *S. elongatus* cells have an oval-to-cylindrical shape, and vary in size between $0.25\text{-}0.4 \mu\text{m}$ in diameter and $0.4\text{-}4.0 \mu\text{m}$ in length [Komarek]. Cell divide symmetrically by binary fission. The two daughter cells separate from each other after reaching the size and shape of the mother cell [Komarek2]. We used non-synchronised cell cultures undergoing active growth providing cells in various stages of their cell cycle.

The experiments described in paper I was carried out at the atomic, molecular, and optical science (AMO) endstation at LCLS at 512 eV (2.40 nm) and 1100 eV (1.13 nm) photon energy. Figure 10.1 shows the arrangement of the experiment. The length of the photon bunch was about 70 fs. The pulse was focused to a spot of $3 \mu\text{m} \times 7 \mu\text{m}$. The average photon density in the focus was about $1.1 \times 10^{11} \text{ photons/pulse}/\mu\text{m}^2$ at 517 eV, and $8.6 \times 10^{10} \text{ photons/pulse}/\mu\text{m}^2$ at 1,100 eV. Far-field diffraction patterns were recorded on a pair of pnCCD detectors [Struder] in the CFEL-ASG Multi Purpose (CAMP) instrument [Struder]. The detectors were place at 741 mm downstream from the interaction region of the beam and the sample. The detector read out rate matched the 120 Hz repetition rate of the LCLS.

We collected diffraction patterns of *C. gracile* cells for 60 minutes at a hit ratio of 43% and selected the 7,500 clear single hits for further analysis, using the Cheetah software package. The oversampling rate of the particle was around 12-fold, which allowed direct phase recovery from the measured intensity patterns. This was not a trivial problem because strong hits saturate the detectors at low diffraction angles. As a compromise, we 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 cells presented in paper I.

Figure 2 a-j shows the reconstructed exit wave-fronts (images) for ten live *C. gracile* cells together with the corresponding diffraction patterns, and a synthetic DIC image. The reconstructions represent 2D projections of the electron density of the cells. The images are arranged by increasing size, and show the expected morphologies of cells during division [Komarek,Komarek2]. The resolution of each reconstruction is indicated by the size of the round white dot. Features smaller than the dot need to be interpreted with care.

Phases were retrieved using the Hawk software package. For each pattern 400 reconstructions were made, each starting from different random initial phases. These reconstructions consisted of 5000 iterations with the RAAR algorithm [Luke], using a Shrinkwrap algorithm[Machesini] for support determination, and concluded with 100 iterations by the ER algorithm[Fienup, Gerchberg-Saxton]. The initial and final support size was selected manually. No additional constraints were used since we anticipated the effects of absorption in the thick cells to give effects similar to a phase object.

Resolution for the reconstructions was estimated from the PRTF (See Figure 10.1), using the $1/e$ cutoff. Before calculating a PRTF and averaging the images we removed outliers among 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(Figures 7 a-j). 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 7j). This made us believe that the average image of the main cluster is the true reconstruction minimum. Th failed reconstructions did not find the true minimum. The scatter plots of Figure 7 a-j show that the real-space error is more reliable than the Fourier-error for identifying failed reconstructions. Furthermore, clustering aids in the identification of failed reconstructions, as it does for reconstruction 7, even if the error measures are not different.

Detector saturation limited the achievable resolution. In fact, the reconstructions shown in Figure 10.1 come from exposures that did not saturate the detectors. A number of much stronger exposures were also recorded, and in some of these exposures the diffraction signal extended to nanometer resolution. Figure 10.1 shows one such pattern for a live *S. elongatus* cell at 1,100 eV photon energy, 70 fs pulse length, about 10^{11} photons μm^{-2} on the sample. Four pnCCD detectors were used to record this pattern (Figure 10.1b). The configuration of the central back detector in Figure 4 is identical to the detector used in Figure 2. The front detector is the same type as the back detector but is placed at 220 mm from the interaction region. In this strong hit, a large part of the back detector was saturated, preventing reliable phasing. The signal however extended beyond 4 nm resolution on the front detectors (at sigma 3.7), which 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. The size of the cell was derived from the autocorrelation. Figure 4c shows that in a log/log representation the drop-off of the signal

is linear with spatial frequency in the range covered by our measurements, and the exponent of the signal decay is -3.31 ± 0.01 , matching simulations [Jakobson].

DICUSSION

11.2 Classification

Sizing diffraction pattern

Sizing Autocorrelation

Identification multiple scatterers

1) Multiple Particles

2) Holographic Images

Edge detection

Particle shape detection

1) Round particles

2) Elongated Particles

Success on wide variety of sizes and scattering strengths.

Template-based Classification

11.3 Software

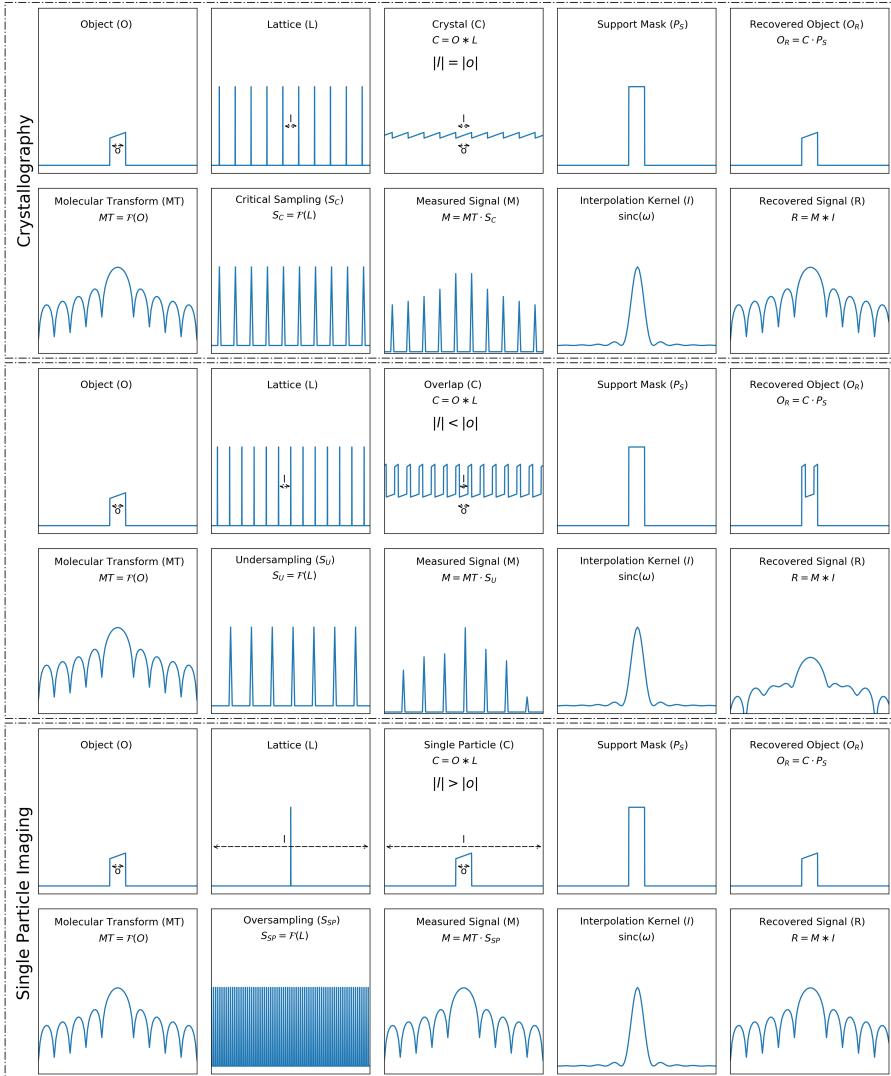


Figure 11.1. Different rates of sampling explained

12. Sammanfattning pa Svenska

In this thesis we want to develop new methods to image biological diversity at the molecular level. Understanding this diversity is very important as it can play a major role in development of deceases. The thesis focuses on two topics: the imaging *living* cells and the development of a computational tool to assess variability within a population of images.

In molecular biology structure and function are closely related. This however does not mean that each object has one single static shape, as for example a key has. On the contrary, many objects have multiple conformations and only through understanding the variation, their function becomes interpretable. A further complication arises from the fact that the variation is often highly dependent on the exact environment the objects are in. Outside of their natural environment, molecular machines might start operating differently. A tool that can images molecular machinery inside living cells (this is their natural environment) is therefore required to truly understand the structure of molecular life.

The amount of detail within objects (resolution) that one can image is limited by the wavelength of the light. Molecular machines are build up out of atoms, so light with a wavelength of the size of an atom is needed understand the functioning of molecular machines. This type of light is X-ray radiation. Another fundamental principle is that the smaller an object is, the stronger the beam of light that shines on the object needs to be in order to record enough information to record an image. Strong beams have the disadvantage that they damage the objects themselves. This led to the limitation that single objects smaller than 100 atoms in diameter could not be resolved structurally with X-rays.

The development of a new type of X-ray laser, x-ray free-electron lasers, created the solution to the damage-imposed resolution limit. XFELs produce ultra-short and extremely brilliant X-ray pulses at repetition rates up to 1 billions images a day. The power of the pulse ensures that even individual molecular machines can be imaged in atomic detail. These pulses do damage the objects they image, but because they are ultra-short the damage only happens after the pulse. The recorded image is therefore an image of the undamaged object. This principle is called diffract-before-destroy.

The word diffract indicates that we are not recording images of an object in the way a photo camera captures images. Instead we measure a so called diffraction pattern. The 2D diffraction pattern can be converted to a real image (2D) in a process called image reconstruction. The feasibility of the diffract-before-destroy principle as well as the image reconstruction process has been shown for a variety of non-biological and biological samples in 2D. And recently the first 3D model of a biological object has been generated from many 2D diffraction patterns.

In 2008, researchers showed that it is theoretically possible to image small living cells at sub-nanometer resolution in 2D. The main work of the thesis was an experimental verification of this prediction. It showed that, using a relatively weak XFEL compared to the predicted parameters, it is possible to record images with signal upto 40 atoms resolution. Unfortunately, due to detector saturation, image reconstruction was not successful for these images. To avoid saturation the pulse power had to be limited, which in turn posed the limit to resolution of the recovered object to 82 nm. We have suggested improvements to avoid detector saturation, but these suggestions have not been tested experimentally.

Due to the variability between different cells it is not trivial to fully automate the process of object recovery from the recorded image, something that is needed if one want to study the variability of cells, or the molecular machines inside them. For the automation to succeed it necessary to predict the shape of the object from the recorded image itself. Furthermore, other parameters such as saturation or having two cells in the beam at once also have to automatically assessed.

This led to the development of the software suite called RedFlamingo. Red Flamingo can be used to assess the quality of individual images, and deduces several essential features for image reconstruction. In some cases it can also determine whether the observed variety originate from a biological source, or whether it is a result from he experiment. It turns out that this feature can be useful in the process of deriving a 3D model from many 2D images.

I am very excited to see the development of this technique. Measuring images without saturation effects might open the door to observing the biological processes occurring inside living cells. As the field is heading now, the initial focus will most likely be on the elucidation of conformational variability in isolated molecular machines. After this we can hopefully observe these machines acting and interacting inside their natural environment, allowing us to visualize the organisation of life.

References

- [1] 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*, 23(6):065009, jun 2013.
- [2] Massimo 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.
- [3] Jakob Andreasson, Andrew V Martin, Meng Liang, Nicusor Timneanu, Andrew Aquila, Fenglin Wang, Bianca Iwan, Martin Svenda, Tomas Ekeberg, Max Hantke, Johan Bielecki, Daniel Rolles, Artem Rudenko, Lutz Foucar, Robert Hartmann, Benjamin Erk, Benedikt Rudek, Henry N Chapman, Janos Hajdu, and Anton Barty. Automated identification and classification of single particle serial femtosecond X-ray diffraction data. *Optics express*, 22(3):2497–2510, 2014.
- [4] 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*, 2(4), 2015.
- [5] Kartik Ayyer, Ti Yen Lan, Veit Elser, and N Duane Loh. Dragonfly: An implementation of the expand-maximize-compress algorithm for single-particle imaging. *Journal of Applied Crystallography*, 49(4):1320–1335, 2016.
- [6] Kartik Ayyer, Hugh T Philipp, Mark W Tate, Jennifer L Wierman, Veit Elser, and Sol M Gruner. Determination of crystallographic intensities from sparse data. *IUCrJ*, 2(1):29–34, 2015.
- [7] Kartik Ayyer, Oleksandr M Yefanov, Dominik Oberthür, Shatabdi Roy-Chowdhury, Lorenzo Galli, Valerio Mariani, Shibom Basu, Jesse Coe, Chelsie E Conrad, Raimund Fromme, Alexander Schaffer, Katerina Dörner, Daniel James, Christopher Kupitz, Markus Metz, Garrett Nelson, Paulraj Lourdu Xavier, Kenneth R Beyerlein, Marius Schmidt, Iosifina Sarrou, John C H Spence, Uwe Weierstall, Thomas A White, Jay-How Yang, Yun Zhao, Mengning Liang, Andrew Aquila, Mark S Hunter, Joseph S Robinson, Jason E Koglin, Sébastien Boutet, Petra Fromme, Anton Barty, and Henry N Chapman. Macromolecular diffractive imaging using imperfect crystals. *Nature*, 530(7589):202–6, feb 2016.
- [8] Anton Barty. Time-resolved imaging using x-ray free electron lasers. *J. Phys. B: At. Mol. Opt. Phys*, 43:194014–15, 2010.
- [9] Anton Barty. Single molecule imaging using X-ray free electron lasers. *Current Opinion in Structural Biology*, 40:186–194, 2016.

- [10] Anton Barty, Carl Caleman, Andrew Aquila, Nicusor Timneanu, Lukas Lomb, Thomas a White, Jakob Andreasson, David Arnlund, Saša Bajt, Thomas R M Barends, Miriam Barthelmess, Michael J Bogan, Christoph Bostedt, John D Bozek, Ryan Coffee, Nicola Coppola, Jan Davidsson, Daniel P Deponte, R Bruce Doak, Tomas Ekeberg, Veit Elser, Sascha W Epp, Benjamin Erk, Holger Fleckenstein, Lutz Foucar, Petra Fromme, Heinz Graafsma, Lars Gumprecht, Janos Hajdu, Christina Y Hampton, Robert Hartmann, Andreas Hartmann, Günter Hauser, Helmut Hirsemann, Peter Holl, Mark S Hunter, Linda Johansson, Stephan Kassemeyer, Nils Kimmel, Richard a Kirian, Mengning Liang, Filipe R N C Maia, Erik Malmerberg, Stefano Marchesini, Andrew V Martin, Karol Nass, Richard Neutze, Christian Reich, Daniel Rolles, Benedikt Rudek, Artem Rudenko, Howard Scott, Ilme Schlichting, Joachim Schulz, M Marvin Seibert, Robert L Shoeman, Raymond G Sierra, Heike Soltau, John C H Spence, Francesco Stellato, Stephan Stern, Lothar Strüder, Joachim Ullrich, X Wang, Georg Weidenspointner, Uwe Weierstall, Cornelia B Wunderer, and Henry N Chapman. Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nature photonics*, 6(December):35–40, jan 2012.
- [11] Anton Barty, Richard a Kirian, Filipe R N C Maia, Max Hantke, Chun Hong Yoon, Thomas a White, and Henry Chapman. Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of applied crystallography*, 47:1118–1131, jun 2014.
- [12] Magnus Bergh, Gösta Huldt, Nicusor Timneanu, Filipe R. N. C. Maia, and Janos Hajdu. Feasibility of imaging living cells at subnanometer resolutions by ultrafast X-ray diffraction. *Quarterly Reviews of Biophysics*, 41(3-4):181, nov 2008.
- [13] H. M. Berman. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, jan 2000.
- [14] Michael J Bogan, W Henry Benner, Sébastien Boutet, Urs Rohner, Matthias Frank, Anton Barty, M Marvin Seibert, Filipe Maia, Stefano Marchesini, Saša Bajt, Bruce Woods, Vincent Riot, Stefan P Hau-Riege, Martin Svenda, Erik Marklund, Eberhard Spiller, Janos Hajdu, and Henry N Chapman. Single Particle X-ray Diffractive Imaging. *Nano Letters*, 8(1):310–316, jan 2008.
- [15] Michael J Bogan, Dmitri Starodub, Christina Y Hampton, and Raymond G Sierra. Single-particle coherent diffractive imaging with a soft x-ray free electron laser: towards soot aerosol morphology. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 43(19):194013, oct 2010.
- [16] Gábor Bortel and Miklós Tegze. Common arc method for diffraction pattern orientation. *Acta crystallographica. Section A, Foundations of crystallography*, 67(Pt 6):533–543, nov 2011.
- [17] 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*, 46:164003, aug 2013.

- [18] A T Brünger. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355:472, 1992.
- [19] H N Chapman, C Caleman, and N Timneanu. Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647):20130313, jun 2014.
- [20] Henry N. Chapman, Anton Barty, Stefano Marchesini, Aleksandr Noy, Stefan P. Hau-Riege, Congwu Cui, Malcolm R. Howells, Rachel Rosen, Haifeng He, John C. H. Spence, Uwe Weierstall, Tobias Beetz, Chris Jacobsen, and David Shapiro. High-resolution ab initio three-dimensional x-ray diffraction microscopy. *Journal of the Optical Society of America A*, 23(5):1179, may 2006.
- [21] Henry N. Chapman, Petra Fromme, Anton Barty, Thomas A. White, Richard A. Kirian, Andrew Aquila, Mark S. Hunter, Joachim Schulz, Daniel P. DePonte, Uwe Weierstall, R. Bruce Doak, Filipe R. N. C. Maia, Andrew V. Martin, Ilme Schlichting, Lukas Lomb, Nicola Coppola, Robert L. Shoeman, Sascha W. Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Lutz Foucar, Nils Kimmel, Georg Weidenspointner, Peter Holl, Mengning Liang, Miriam Barthelmess, Carl Caleman, Sébastien Boutet, Michael J. Bogan, Jacek Krzywinski, Christoph Bostedt, Saša Bajt, Lars Gumprecht, Benedikt Rudek, Benjamin Erk, Carlo Schmidt, André Hömke, Christian Reich, Daniel Pietschner, Lothar Strüder, Günter Hauser, Hubert Gorke, Joachim Ullrich, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Kühnel, Marc Messerschmidt, John D. Bozek, Stefan P. Hau-Riege, Matthias Frank, Christina Y. Hampton, Raymond G. Sierra, Dmitri Starodub, Garth J. Williams, Janos Hajdu, Nicusor Timneanu, M. Marvin Seibert, Jakob Andreasson, Andrea Rocker, Olof Jönsson, Martin Svenda, Stephan Stern, Karol Nass, Robert Andritschke, Claus-Dieter Schröter, Faton Krasniqi, Mario Bott, Kevin E. Schmidt, Xiaoyu Wang, Ingo Grotjohann, James M. Holton, Thomas R. M. Barends, Richard Neutze, Stefano Marchesini, Raimund Fromme, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhov, Inger Andersson, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Björn Nilsson, and John C. H. Spence. Femtosecond X-ray protein nanocrystallography. *Nature*, 470(7332):73–77, feb 2011.
- [22] G F Christopher and S L Anna. Microfluidic methods for generating continuous droplet streams. *Journal of Physics D: Applied Physics*, 40(19):R319–R336, oct 2007.
- [23] F. Cruz-Mazo, M. A. Herrada, A. M. Gañán-Calvo, and J. M. Montanero. Global stability of axisymmetric flow focusing. *Journal of Fluid Mechanics*, 832:329–344, dec 2017.
- [24] Benedikt J. Daurer, Kenta Okamoto, Johan Bielecki, Filipe R. N. C. Maia, Kerstin Mühlig, M. Marvin Seibert, Max F. Hantke, Carl Nettelblad, W. Henry Benner, Martin Svenda, Nicusor Timneanu, Tomas Ekeberg, N. Duane Loh, Alberto Pietrini, Alessandro Zani, Asawari D. Rath, Daniel Westphal, Richard A. Kirian, Salah Awel, Max O. Wiedorn, Gijs van der Schot, Gunilla H. Carlsson, Dirk Hasse, Jonas A. Sellberg, Anton Barty, Jakob Andreasson, Sébastien Boutet, Garth Williams, Jason Koglin, Inger Andersson, Janos Hajdu, and Daniel S. D. Larsson. Experimental strategies for imaging bioparticles with

- femtosecond hard X-ray pulses. *IUCrJ*, 4(3):251–262, may 2017.
- [25] 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*, 41(19):195505, oct 2008.
- [26] Jeffrey J Donatelli, Peter H Zwart, and James a. Sethian. Iterative phasing for fluctuation X-ray scattering. *Proceedings of the National Academy of Sciences*, page 201513738, 2015.
- [27] Stefan Eisebitt. X-ray holography: The hole story. *Nature Photonics*, 2008.
- [28] T Ekeberg, S Engblom, and J Liu. Machine learning for ultrafast X-ray diffraction patterns on large-scale GPU clusters. *International Journal of High Performance Computing Applications*, 29(2):233–243, 2015.
- [29] Tomas Ekeberg. *Flash Diffractive Imaging in Three Dimensions*. Doctoral thesis, Uppsala University, 2012.
- [30] Tomas Ekeberg, Martin Svenda, Chantal Abergel, Filipe R. N. C. Maia, Virginie Seltzer, Jean-Michel Claverie, Max Hantke, Olof Jönsson, Carl Nettelblad, Gijs van der Schot, Mengning Liang, Daniel P. DePonte, Anton Barty, M. Marvin Seibert, Bianca Iwan, Inger Andersson, N. Duane Loh, Andrew V. Martin, Henry Chapman, Christoph Bostedt, John D. Bozek, Ken R. Ferguson, Jacek Krzywinski, Sascha W. Epp, Daniel Rolles, Artem Rudenko, Robert Hartmann, Nils Kimmel, and Janos Hajdu. Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser. *Physical Review Letters*, 114(9):098102, mar 2015.
- [31] Tomas Ekeberg, Martin Svenda, M. Marvin Seibert, Chantal Abergel, Filipe R.N.C. Maia, Virginie Seltzer, Daniel P. DePonte, Andrew Aquila, Jakob Andreasson, Bianca Iwan, Olof Jönsson, Daniel Westphal, Duško Odić, Inger Andersson, Anton Barty, Meng Liang, Andrew V. Martin, Lars Gumprecht, Holger Fleckenstein, Saša Bajt, Miriam Barthelmess, Nicola Coppola, Jean-Michel Claverie, N. Duane Loh, Christoph Bostedt, John D. Bozek, Jacek Krzywinski, Marc Messerschmidt, Michael J. Bogan, Christina Y. Hampton, Raymond G. Sierra, Matthias Frank, Robert L. Shoeman, Lukas Lomb, Lutz Foucar, Sascha W. Epp, Daniel Rolles, Artem Rudenko, Robert Hartmann, Andreas Hartmann, Nils Kimmel, Peter Holl, Georg Weidenspointner, Benedikt Rudek, Benjamin Erk, Stephan Kassemeyer, Ilme Schlichting, Lothar Strüder, Joachim Ullrich, Carlo Schmidt, Faton Krasniqi, Günter Hauser, Christian Reich, Heike Soltau, Sebastian Schorb, Helmut Hirsemann, Cornelia Wunderer, Heinz Graafsma, Henry Chapman, and Janos Hajdu. Single-shot diffraction data from the Mimivirus particle using an X-ray free-electron laser. *Scientific Data*, 3:160060, aug 2016.
- [32] V Elser. Noise Limits on Reconstructing Diffraction Signals From Random Tomographs. *IEEE Transactions on Information Theory*, 55(10):4715–4722, 2009.
- [33] Ken R Ferguson, Maximilian Bucher, John D Bozek, Sebastian Carron, Jean-Charles Castagna, Ryan Coffee, G Ivan Curiel, Michael Holmes, Jacek Krzywinski, Marc Messerschmidt, Michael Minitti, Ankush Mitra, Stefan Moeller, Peter Noonan, Timur Osipov, Sebastian Schorb, Michele Swiggers, Alexander Wallace, Jing Yin, and Christoph Bostedt. The Atomic, Molecular

- and Optical Science instrument at the Linac Coherent Light Source. *Journal of Synchrotron Radiation*, 22(3):492–497, may 2015.
- [34] J R Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.
 - [35] Jim Fienup and C C Wackerman. Phase retrieval stagnation problems and solutions, 1986.
 - [36] Lutz Foucar, Anton Barty, Nicola Coppola, Robert Hartmann, Peter Holl, Uwe Hoppe, Stephan Kassemeyer, Nils Kimmel, Jochen Küpper, Mirko Scholz, Simone Techert, Thomas a. White, Lothar Strüder, and Joachim Ullrich. CASS CFEL ASG software suite. *Computer Physics Communications*, 183:2207–2213, oct 2012.
 - [37] K D Froome, L Essen, and Richard A Rhodes. The Velocity of Light And Radio Waves. *Physics Today*, 24(5):49, may 1971.
 - [38] Alfonso M Gañán-Calvo, Daniel P DePonte, Miguel A Herrada, John C H Spence, Uwe Weierstall, and R Bruce Doak. Liquid Capillary Micro/Nanojets in Free-Jet Expansion. *Small*, 6(7):822–824, apr 2010.
 - [39] R W Gerchberg and W Owen Saxton. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.
 - [40] Max F Hantke, Tomas Ekeberg, and Filipe R N C Maia. Condor: a simulation tool for flash X-ray imaging. *Journal of Applied Crystallography*, 49:1356–1362, 2016.
 - [41] Max F. Hantke, Dirk Hasse, Filipe R. N. C. Maia, Tomas Ekeberg, Katja John, Martin Svenda, N. Duane Loh, Andrew V. Martin, Nicusor Timneanu, Daniel S. D. Larsson, Gijs van der Schot, Gunilla H. Carlsson, Margareta Ingelman, Jakob Andreasson, Daniel Westphal, Mengning Liang, Francesco Stellato, Daniel P. DePonte, Robert Hartmann, Nils Kimmel, Richard A. Kirian, M. Marvin Seibert, Kerstin Mühlig, Sebastian Schorb, Ken Ferguson, Christoph Bostedt, Sebastian Carron, John D. Bozek, Daniel Rolles, Artem Rudenko, Sascha Epp, Henry N. Chapman, Anton Barty, Janos Hajdu, and Inger Andersson. High-throughput imaging of heterogeneous cell organelles with an X-ray laser. *Nature Photonics*, 8(12):943–949, nov 2014.
 - [42] Robert Hartmann, Sascha Epp, Sven Herrmann, Peter Holl, Norbert Meidinger, Christian Reich, Daniel Rolles, Heike Soltau, Lothar Struder, Joachim Ullrich, and Georg Weidenspointner. Large format pnCCDs as imaging detectors for X-ray free-electron-lasers. *2008 IEEE Nuclear Science Symposium Conference Record*, pages 2590–2595, 2008.
 - [43] Stefan P Hau-Riege, Sébastien Boutet, Anton Barty, Saša Bajt, Michael J Bogan, Matthias Frank, Jakob Andreasson, Bianca Iwan, M Marvin Seibert, Janos Hajdu, Anne Sakdinawat, Joachim Schulz, Rolf Treusch, and Henry N Chapman. Sacrificial tamper slows down sample explosion in FLASH diffraction experiments. *Physical Review Letters*, 104(6):1–4, 2010.
 - [44] a Hosseinzadeh, P Schwander, a Dashti, R Fung, R M D’Souza, and Abbas Ourmazd. High-resolution structure of viruses from random diffraction snapshots. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369:20130326, 2014.
 - [45] G Huldt, a. Szoke, and J Hajdu. Diffraction imaging of single particles and

- biomolecules. *Journal of Structural Biology*, 144(1-2):219–227, 2003.
- [46] Slavica Jonić. Cryo-electron Microscopy Analysis of Structurally Heterogeneous Macromolecular Complexes. *Computational and Structural Biotechnology Journal*, 14:385–390, 2016.
- [47] Takashi Kimura, Yasumasa Joti, Akemi Shibuya, Changyong Song, Sangsoo Kim, Kensuke Tono, Makina Yabashi, Masatada Tamakoshi, Toshiyuki Moriya, Tairo Oshima, Tetsuya Ishikawa, Yoshitaka Bessho, and Yoshinori Nishino. Imaging live cell in micro-liquid enclosure by X-ray laser diffraction. *Nature Communications*, 5, jan 2014.
- [48] Peng Liu, Paul J. Ziemann, David B. Kittelson, and Peter 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*, 22(3):293–313, jan 1995.
- [49] N D Loh. Effects of extraneous noise in Cryptotomography. *SPIE Optical Engineering + Applications*, 8500:85000K, oct 2012.
- [50] N D Loh, M J Bogan, V Elser, A Barty, S Boutet, S Bajt, J Hajdu, T Ekeberg, F R N C Maia, J Schulz, M M Seibert, B Iwan, N Timneanu, S Marchesini, I Schlichting, R L Shoeman, L Lomb, M Frank, M Liang, and H N Chapman. Cryptotomography: Reconstructing 3D Fourier Intensities from Randomly Oriented Single-Shot Diffraction Patterns. *Physical Review Letters*, 104(22):225501, jun 2010.
- [51] N Duane Loh. A minimal view of single-particle imaging with X-ray lasers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647):20130328, jun 2014.
- [52] Ne-Te Duane Loh and Veit Elser. Reconstruction algorithm for single-particle diffraction imaging experiments. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 80(2 Pt 2):26705, aug 2009.
- [53] Filipe R N C Maia. The Coherent X-ray Imaging Data Bank. *Nature Methods*, 9(9):854–855, aug 2012.
- [54] Filipe R. N. C. Maia, Tomas Ekeberg, Nicusor Timneanu, David van der Spoel, and Janos Hajdu. Structural variability and the incoherent addition of scattered intensities in single-particle diffraction. *Physical Review E*, 80(3):031905, sep 2009.
- [55] Filipe R N C Maia, Tomas Ekeberg, David Van Der Spoel, and Janos Hajdu. Hawk : the image reconstruction package for coherent X-ray diffractive imaging. *Journal of Applied Crystallography*, 43(6):1535–1539, 2010.
- [56] Filipe R N C Maia and Janos Hajdu. The trickle before the torrent-diffraction data from X-ray lasers. *Scientific data*, 3:160059, aug 2016.
- [57] S Marchesini. Invited Article: A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments*, 78(1):11301, jan 2007.
- [58] 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*, 68(14):140101, oct 2003.
- [59] A V Martin, A J Morgan, T Ekeberg, N D Loh, F.R.N.C. Maia, F Wang, J C H Spence, and H N Chapman. The extraction of single-particle diffraction patterns from a multiple-particle diffraction pattern. *Optics Express*, 21(13):15102,

2013.

- [60] A. V. Martin, F. Wang, N. D. Loh, T. Ekeberg, F. R. N. C. Maia, M. Hantke, G. van der Schot, C. Y. Hampton, R. G. Sierra, A. Aquila, S. Bajt, M. Barthelmess, C. Bostedt, J. D. Bozek, N. Coppola, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, M. Frank, H. Graafsma, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Liang, L. Lomb, S. Marchesini, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, J. Schulz, R. L. Shoeman, H. Soltau, D. Starodub, J. Steinbrener, F. Stellato, L. Strüder, J. Ullrich, G. Weidenspointner, T. A. White, C. B. Wunderer, A. Barty, I. Schlichting, M. J. Bogan, and H. N. Chapman. Noise-robust coherent diffractive imaging with a single diffraction pattern. *Optics Express*, 20(15):16650, jul 2012.
- [61] A V Andrew V Martin, Jakob Andreasson, Andrew Aquila, Saša Bajt, Thomas R M Barends, Miriam Barthelmess, Anton Barty, W Henry Benner, Christoph Bostedt, John D Bozek, Phillip Bucksbaum, Carl Caleman, Nicola Coppola, Daniel P DePonte, Tomas Ekeberg, Sascha W Epp, Benjamin Erk, George R Farquar, Holger Fleckenstein, Lutz Foucar, Matthias Frank, Lars Gumprecht, Christina Y Hampton, Max Hantke, Andreas Hartmann, Elisabeth Hartmann, Robert Hartmann, Stephan P Hau-Riege, Günther Hauser, Peter Holl, Andre Hoemke, Olof Jönsson, Stephan Kassemeyer, Nils Kimmel, Maya Kiskinova, Faton Krasniqi, Jacek Krzywinski, Mengning Liang, Ne-Te Duane Loh, Lukas Lomb, Filipe R N C Maia, Stefano Marchesini, Marc Messerschmidt, Karol Nass, Duško Odic, Emanuele Pedersoli, Christian Reich, Daniel Rolles, Benedikt Rudek, Artem Rudenko, Carlo Schmidt, Joachim Schultz, M Marvin Seibert, Robert L Shoeman, Raymond G Sierra, Heike Soltau, Dmitri Starodub, Jan Steinbrener, Francesco Stellato, Lothar Strüder, Martin Svenda, Herbert Tobias, Joachim Ullrich, Georg Weidenspointner, Daniel Westphal, Thomas A White, Garth Williams, Janos Hajdu, Ilme Schlichting, Michael J Bogan, and Henry N Chapman. Single particle imaging with soft X-rays at the Linac Coherent Light Source. *SPIE Optics+ . . .*, page 807809, 2011.
- [62] M Marvin Seibert, Sébastien Boutet, Martin Svenda, Tomas Ekeberg, Filipe R N C Maia, Michael J Bogan, Nicusor Tîmneanu, Anton Barty, Stefan Hau-Riege, Carl Caleman, Matthias Frank, Henry Benner, Joanna Y Lee, Stefano Marchesini, Joshua W Shaevitz, Daniel A Fletcher, Sasa Bajt, Inger Andersson, Henry N Chapman, and Janos Hajdu. Femtosecond diffractive imaging of biological cells. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 43(19):194015, oct 2010.
- [63] A. A. Michelson. Measurement of the Velocity of Light Between Mount Wilson and Mount San Antonio. *The Astrophysical Journal*, 65:1, jan 1927.
- [64] R. Neutze. Opportunities and challenges for time-resolved studies of protein structural dynamics at X-ray free-electron lasers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647):20130318–20130318, jun 2014.
- [65] Richard Neutze, Remco Wouts, David van der Spoel, Edgar Weckert, and Janos Hajdu. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature*, 406(6797):752–757, aug 2000.
- [66] D Paganin, T E Gureyev, S C Mayo, a W Stevenson, Ya I Nesterets, and S W

- Wilkins. X-ray omni microscopy. *Journal of microscopy*, 214(Pt 3):315–327, 2004.
- [67] Hemanth K.N. Reddy, Chun Hong Yoon, Andrew Aquila, Salah Awel, Kartik Ayyer, Anton Barty, Peter Berntsen, Johan Bielecki, Sergey Bobkov, Maximilian Bucher, Gabriella A. Carini, Sebastian Carron, Henry Chapman, Benedikt Daurer, Hasan DeMirici, Tomas Ekeberg, Petra Fromme, Janos Hajdu, Max Felix Hanke, Philip Hart, Brenda G. Hogue, Ahmad Hosseiniزاده, Yoonhee Kim, Richard A. Kirian, Ruslan P. Kurta, Daniel S.D. Larsson, N. Duane Loh, Filipe R.N.C. Maia, Adrian P. Mancuso, Kerstin Mühlig, Anna Munke, Daewoong Nam, Carl Nettelblad, Abbas Ourmazd, Max Rose, Peter Schwander, Marvin Seibert, Jonas A. Sellberg, Changyong Song, John C.H. Spence, Martin Svenda, Gijs Van der Schot, Ivan A. Vartanyants, Garth J. Williams, and P. Lourdu Xavier. Coherent soft X-ray diffraction imaging of coliphage PR772 at the Linac coherent light source. *Scientific Data*, 4:170079, jun 2017.
- [68] Antonio Rosato, Anurag Bagaria, David Baker, Benjamin Bardiaux, Andrea Cavalli, Jurgen F Doreleijers, Andrea Giachetti, Paul Guerry, Peter Güntert, Torsten Herrmann, Yuanpeng J Huang, Hendrik R A Jonker, Binchen Mao, Thérèse E Malliavin, Gaetano T Montelione, Michael Nilges, Srivatsan Raman, Gijs van der Schot, Wim F Vranken, Geerten W Vuister, and Alexandre M J J Bonvin. CASD-NMR: critical assessment of automated structure determination by NMR. *Nature Methods*, 6(9):625–626, sep 2009.
- [69] M Rossmann. *The molecular replacement method a collection of papers on the use of non-crystallographic symmetry*. Gordon and Breach, New York, 1972.
- [70] 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*, 81(17):1–6, 2010.
- [71] D. Sayre. Some implications of a theorem due to Shannon. *Acta Crystallographica*, 5(6):843–843, nov 1952.
- [72] P Schwander, R Fung, and Abbas Ourmazd. Conformations of macromolecules and their complexes from heterogeneous datasets. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369:20130567, 2014.
- [73] P Schwander, R Fung, G N Phillips, and a Ourmazd. Mapping the conformations of biological assemblies. *New Journal of Physics*, 12(3):35007, mar 2010.
- [74] M. Marvin Seibert, Tomas Ekeberg, Filipe R. N. C. Maia, Martin Svenda, Jakob Andreasson, Olof Jönsson, Duško Odić, Bianca Iwan, Andrea Rocker, Daniel Westphal, Max Hantke, Daniel P. DePonte, Anton Barty, Joachim Schulz, Lars Gumprecht, Nicola Coppola, Andrew Aquila, Mengning Liang, Thomas A. White, Andrew Martin, Carl Caleman, Stephan Stern, Chantal Abergel, Virginie Seltzer, Jean-Michel Claverie, Christoph Bostedt, John D. Bozek, Sébastien Boutet, A. Alan Miahnahri, Marc Messerschmidt, Jacek Krzywinski, Garth Williams, Keith O. Hodgson, Michael J. Bogan, Christina Y. Hampton, Raymond G. Sierra, Dmitri Starodub, Inger Andersson, Saša Bajt, Miriam Barthelmess, John C. H. Spence, Petra Fromme, Uwe Weierstall, Richard Kirian, Mark Hunter, R. Bruce Doak, Stefano Marchesini, Stefan P. Hau-Riege, Matthias Frank, Robert L. Shoeman, Lukas Lomb, Sascha W. Epp, Robert

Hartmann, Daniel Rolles, Artem Rudenko, Carlo Schmidt, Lutz Foucar, Nils Kimmel, Peter Holl, Benedikt Rudek, Benjamin Erk, André Hömke, Christian Reich, Daniel Pietschner, Georg Weidenspointner, Lothar Strüder, Günter Hauser, Hubert Gorke, Joachim Ullrich, Ilme Schlichting, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Künnel, Robert Andritschke, Claus-Dieter Schröter, Faton Krasniqi, Mario Bott, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Björn Nilsson, Henry N. Chapman, and Janos Hajdu. Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature*, 470(7332):78–81, feb 2011.

- [75] C E Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, jan 1949.
- [76] C E Shannon. Communication Theory of Secrecy Systems*. *Bell system technical journal*, (July), 1949.
- [77] David Shapiro, Pierre Thibault, Tobias Beetz, Veit Elser, Malcolm Howells, Chris Jacobsen, Janos Kirz, Enju Lima, Huijie Miao, Aaron M Neiman, and David Sayre. Biological imaging by soft x-ray diffraction microscopy. *Proceedings of the National Academy of Sciences*, 102(43):15343–15346, oct 2005.
- [78] V L Shneerson, a Ourmazd, and D K Saldin. Crystallography without crystals. I. The common-line method for assembling a three-dimensional diffraction volume from single-particle scattering. *Acta crystallographica. Section A, Foundations of crystallography*, 64(Pt 2):303–315, mar 2008.
- [79] J C H Spence, U Weierstall, and H N Chapman. X-ray lasers for structural and dynamic biology. *Reports on Progress in Physics*, 75(10):102601, oct 2012.
- [80] John C. H. Spence, Richard A. Kirian, Xiaoyu Wang, Uwe Weierstall, Kevin E. Schmidt, Thomas White, Anton Barty, Henry N. Chapman, Stefano Marchesini, and James Holton. Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals. *Optics Express*, 2011.
- [81] Miklós Tegze and Gábor Bortel. Atomic structure of a single large biomolecule from diffraction patterns of random orientations. *Journal of Structural Biology*, 2012.
- [82] Pierre Thibault, Veit Elser, Chris Jacobsen, David Shapiro, and David Sayre. Reconstruction of a yeast cell from X-ray diffraction data. *Acta Crystallographica Section A Foundations of Crystallography*, 62(4):248–261, jul 2006.
- [83] Gijs van der Schot and Alexandre M. J. J. Bonvin. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. *Journal of Biomolecular NMR*, 62(4):497–502, aug 2015.
- [84] Gijs van der Schot, Martin Svenda, Filipe R. N. C. Maia, Max Hantke, Daniel P. DePonte, M. Marvin Seibert, Andrew Aquila, Joachim Schulz, Richard Kirian, Mengning Liang, Francesco Stellato, Bianca Iwan, Jakob Andreasson, Nicusor Timneanu, Daniel Westphal, F. Nunes Almeida, Dusko Odic, Dirk Hasse, Gunilla H. Carlsson, Daniel S. D. Larsson, Anton Barty, Andrew V. Martin, Sebastian Schorb, Christoph Bostedt, John D. Bozek, Daniel Rolles, Artem Rudenko, Sascha Epp, Lutz Foucar, Benedikt Rudek, Robert Hartmann, Nils Kimmel, Peter Holl, Lars Englert, Ne-Te Duane Loh, Henry N. Chapman, Inger

- Andersson, Janos Hajdu, and Tomas Ekeberg. Imaging single cells in a beam of live cyanobacteria with an X-ray laser. *Nature Communications*, 6:5704, feb 2015.
- [85] Gijs van der Schot, Zaiyong Zhang, Robert Vernon, Yang Shen, Wim F Vranken, David Baker, Alexandre M J J Bonvin, and Oliver F Lange. Improving 3D structure prediction from chemical shift data. *Journal of biomolecular NMR*, 57(1):27–35, 2013.
- [86] Marin van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3):250–262, sep 2005.
- [87] Michał Walczak and Helmut Grubmüller. Bayesian orientation estimate and structure information from sparse single-molecule x-ray diffraction images. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2014.
- [88] Tsjerk A. Wassenaar, Marc van Dijk, Nuno Loureiro-Ferreira, Gijs van der Schot, Sjoerd J. de Vries, Christophe Schmitz, Johan van der Zwan, Rolf Boelens, Andrea Giachetti, Lucio Ferella, Antonio Rosato, Ivano Bertini, Torsten Herrmann, Hendrik R. A. Jonker, Anurag Bagaria, Victor Jaravine, Peter Güntert, Harald Schwalbe, Wim F. Vranken, Jurgen F. Doreleijers, Gert Vriend, Geerten W. Vuister, Daniel Franke, Alexey Kikhney, Dmitri I. Svergun, Rasmus H. Fogh, John Ionides, Ernest D. Laue, Chris Spronk, Simonas Jurkša, Marco Verlato, Simone Badoer, Stefano Dal Pra, Mirco Mazzucato, Eric Frizziero, and Alexandre M. J. J. Bonvin. WeNMR: Structural Biology on the Grid. *Journal of Grid Computing*, 10(4):743–767, dec 2012.
- [89] Chuan Xiao, Paul R Chipman, Anthony J Battisti, Valorie D Bowman, Patricia Renesto, Didier Raoult, and Michael G Rossmann. Cryo-electron microscopy of the giant Mimivirus. *Journal of molecular biology*, 353(3):493–496, oct 2005.
- [90] Rui Xu, Huaidong Jiang, Changyong Song, Jose a Rodriguez, Zhifeng Huang, Chien-Chun Chen, Daewoong Nam, Jaehyun Park, Marcus Gallagher-Jones, Sangsoo Kim, Sunam Kim, Akihiro Suzuki, Yuki Takayama, Tomotaka Oroguchi, Yukio Takahashi, Jiadong Fan, Yunfei Zou, Takaki Hatsui, Yuichi Inubushi, Takashi Kameshima, Koji Yonekura, Kensuke Tono, Tadashi Togashi, Takahiro Sato, Masaki Yamamoto, Masayoshi Nakasako, Makina Yabashi, Tetsuya Ishikawa, and Jianwei Miao. Single-shot three-dimensional structure determination of nanocrystals with femtosecond X-ray free-electron laser pulses. *Nature communications*, 5(May):4061, jan 2014.
- [91] Chun Hong Yoon, Peter Schwander, Chantal Abergel, Inger Andersson, Jakob Andreasson, Andrew Aquila, Saša Bajt, Miriam Barthelmess, Anton Barty, Michael J Bogan, Christoph Bostedt, John Bozek, Henry N Chapman, Jean-Michel Claverie, Nicola Coppola, Daniel P DePonte, Tomas Ekeberg, Sascha W Epp, Benjamin Erk, Holger Fleckenstein, Lutz Foucar, Heinz Graafsma, Lars Gumprecht, Janos Hajdu, Christina Y Hampton, Andreas Hartmann, Elisabeth Hartmann, Robert Hartmann, Gunter Hauser, Helmut Hirsemann, Peter Holl, Stephan Kassemeyer, Nils Kimmel, Maya Kiskinova, Mengning Liang, Ne-Te Duane Loh, Lukas Lomb, Filipe R N C Maia, Andrew V Martin, Karol Nass, Emanuele Pedersoli, Christian Reich, Daniel Rolles, Benedikt Rudek, Artem Rudenko, Ilme Schlichting, Joachim Schulz, Marvin Seibert, Virginie Seltzer, Robert L Shoeman, Raymond G Sierra, Heike Soltau, Dmitri Starodub, Jan Steinbrener, Gunter Stier, Lothar Strüder, Martin

- Svenda, Joachim Ullrich, Georg Weidenspointner, Thomas A White, Cornelia Wunderer, and Abbas Ourmazd. Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering. *Optics Express*, 19(17):16542, 2011.
- [92] L. Young, E. P. Kanter, B. Krässig, Y. Li, A. M. March, S. T. Pratt, R. Santra, S. H. Southworth, N. Rohringer, L. F. DiMauro, G. Doumy, C. A. Roedig, N. Berrah, L. Fang, M. Hoener, P. H. Bucksbaum, J. P. Cryan, S. Ghimire, J. M. Glownia, D. A. Reis, J. D. Bozek, C. Bostedt, and M. Messerschmidt. Femtosecond electronic response of atoms to ultra-intense X-rays. *Nature*, 466(7302):56–61, jul 2010.