DESY summer student program 2011

Imaging of viruses in the water window

Carlos González Ballestero[1]
Universidad Autónoma de Madrid

Supervisor: Andrew Martin

5 September 2011

Abstract

In this report we show the work done by the summer student in imaging of viruses. The x ray imaging experiment has been carried out in special conditions: the samples were embedded into a water jet to improve contrast. We introduce briefly the theory of imaging and the algorithms used for the reconstruction. We also present the techniques used for processing diffraction patterns so that the algorithm converges to a good image. Finally, results in two real cases are presented. The reconstructions we obtain show that the water jet configuration is very good for imaging, being a very interesting idea for next experiments.

[1] carlos.glezball@gmail.com

Contents

1 Introduction

Research in structure of matter has played an essential role in modern physics. Since more than one century ago, one of the main goal of researchers has been to understand better the behaviour of systems in a small scale. While in some fields such as subatomic physics experimental techniques are based mainly in the study of scattering proccesses between the sample and probe particles, in a larger scale (molecular physics, solid state, etc...) electromagnetic sources offer an adequate and well known way to study physycal systems.

As the technology for constructing light sources has been evolving (mainly producing more intense, coherent and monochromatic light) better tools have been developed, which allow us to study in more detail the behavior of systems in a small scale. Among these set of devices, synchrotrons and lasers are in the lead. Particularly, the Free Electron Laser (FEL), although it is not strictly a laser system, it produces radiation with similar properties, being able to achieve very short wavelengths (even in the X ray region). This property, among others, makes FEL a very important tool in many areas of physics, specially in biophysics, an emerging and really promising field.

One of the goals of medicine and biology nowadays is to reduce the damage produced to a sample (e.g. living tissue) when we modify it or measure some of its properties. Apart from the strict control of the location of the experiment inside the sample (a promising method, but still difficult to control) it is possible also to chemically control the reactions that occur in the sample. When this method is used, a good knowledge of the molecular structure (mainly of proteins) and dynamics is essential. Another key objective, specially for immunology and genetics, is the detailed study of virus and viral processes. However, if you try to get an image of these systems, or any information based on light, a problem appears because virus and proteins' size is smaller than visible wavelengths. The consequence is that it is not possible to study these systems with visible light because they are below the diffraction limit. This is the reason for what short-wavelength coherent light sources play an essential role in biophysics, as they make us able to obtain XR images of these small samples and also information about dynamic processes.

This project is mainly centered in imaging methods for biological samples. Nevertheless, these imaging techniques have important applications in many other areas, such as membrane physics or nanotechnology. Furthermore, XR light sources and specially FELs are basic tools for physics nowadays as they have lots of applications in almost every area of science.

2 X ray imaging techniques

As we know, the most used tool in imaging of systems in the sub-optic scale is the free electron laser (FEL). In this section a short amount of reasons for using it and not other sources (such as synchrotron) or other probe particles (such as electrons) are commented. In the second part, we explain how the data are collected and show the experimental setup, and finally we present a brief explanation of the algorithms used for reconstructing the image.

2.1 Advantages of using Free Electron Laser

There exist many imaging techniques at sub-optic scale which allow us to obtain very precise reconstructions of many systems. The two most used probes for imaging are photons and electrons, each of them with its own advantages and disadvantages. Electronic microscopy techniques can obtain very good resolution because it is relatively easy to have a an electron beam with very short de Broglie wavelength. Nevertheless, electrons have low penetration depth inside the sample, so this microscopy takes images only of the surface of the sample. This makes it impossible to get information about the inside of it, and this is precisely one of the most desired goals when studying biologic samples. Moreover, preparation of samples for electron microscopy requires them to be extracted from its natural environment, so that it is not possible to study these systems in their usual media. Another handicap of electrons is that the electron beam usually produces irreversible damages to all the sample, making it impossible to use it afterwards.

All this drawbacks can be solved using short wavelength photons. This particles have much larger penetration length, so that we can observe the interior of the sample and even obtain 3D images. Apart from that, they allow us to study systems in their natural environment and as a consequence the dynamics and energy interchange processes. Moreover, after the exposition, samples that have not been directly irradiated with photons stay unperturbed. In the case of proteins, virus or bacterias, this implies that sample can be analized and afterwards inoculated to a living being or reutilized for other kind of applications. For all this reasons, X ray imaging techniques are definitely the best for biophysical applications.

Need for a X ray source is evident since we want to observe objects which size is smaller than visible wavelengths. Between the two high frequency light sources that are more useful for biophysical applications, synchrotron and FEL, the second one appears to be the most useful in imaging due to its exceptional properties. First of all, we need a very intense beam because in non-periodic systems the difraction peaks' intensity is very low, so we need high beam intensity for them to be observed. Secondly, we need ultrashort pulses so that the sample is not destroyed before one whole pulse has passed. If not we cannot trust on our diffraction data because some of them come from an exploded (or half-

dissociated) sample. Besides, we would like to have a coherent source for being able to take enough data to obtain a reliable reconstruction. Monochromaticity is also desirable, but as we prefer shorter pulses it is a secondary objective.

According to all these characteristics we need for our source, FEL appears as the most suitable because it produces very intense, short and coherent pulses instead of the continuous, white radiation produced by a synchrotron. This is due to the internal structure of the FEL, in which electrons gather in microbunches whose emitted radiation interferes constructively, producing very intense emission peaks.

All the features of the FEL we have mentioned below are just part of the extraordinary properties of these devices. FEL seems to be nowadays an extraordinary source, and is definitely the best for many applications, including biophysics.

2.2 Experimental setup

The basic structure of an imaging experiment is rather simple: the sample is usually a beam composed of the particles which image we want to reconstruct. This beam flows perpendicularly to the laser beam from the FEL, as shown in figure 1. When the two beams meet, the light hits the sample and a certain number of photons are scattered by its electronic density and are diffracted, so that their origial direction of propagation changes. The detector records the photon impacts in a plane at a certain distance z from the interaction point, and perpendicular to the main beam. In this way we obtain the data that allow us to reconstruct the image, as a 2D diffraction pattern.

There are two important things to remark: first of all, we loose the information in the center of the diffraction pattern, because we stop the main beam using an opaque slab. The reason for this beam stop is that the main beam is too intense so even if we managed to measure it without burning the detector we would not be able to observe the secondary peaks, as they are much less intense. Secondly, detectors have to be extremely quick catching data because samples are destroyed by the beam in a very short time. Because of this fast measuring, sometimes the data you collect are spatially disordered. This is an important thing to take into accout for the analysis.

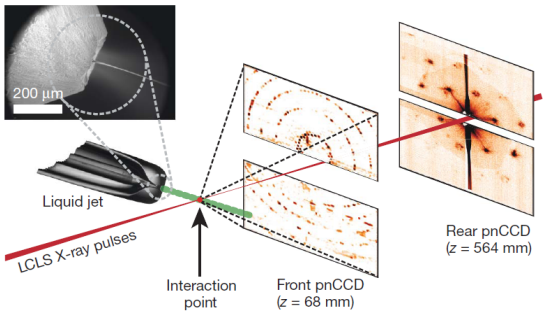

Figure 1: Experimental setup for an XR imaging experiment (image taken from [6]).

In this research we have analized diffraction patterns from a water jet ex-

periment in LCLS (Linac Coherent Light Source, California). The data was taken at the Atomic and Molecular Optics (AMO) end station, using pnCCD detectors. In this case, light from the FEL goes through a water jet in which the sample particles are contained. The objective of the water jet is to reduce losses and signal to noise ratio. This is possible because the wavelength we work with($\lambda = 2.406nm$; $E = 516eV$) is near the absorption edge of carbon but far from that of the oxygen (see figure 2) so we are going to have almost no absorption (i.e., no interaction) in oxygen. Because of this we can consider that all the diffracted photons had a previous interaction with the sample and not with the water jet, so the contrast of the image can be improved.

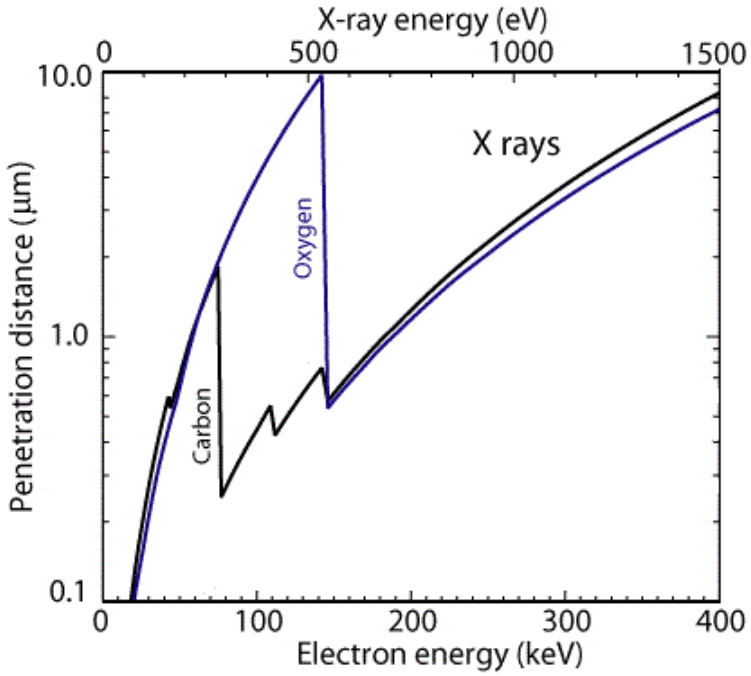

Figure 2: Absorption edges of carbon and oxigen.

The proccess of reconstructing an image using the data of the diffraction pattern can be long and difficult because each pattern needs different refinements before reconstruction (it depends on the sample, the random noise, etc...). During the summer, the work of the student has been mainly centered in processing diffraction patterns and applying the reconstruction algorithm.

2.3 Theory of imaging

Imaging techniques of non-periodic objects require some complex reconstruction algorithms, so for the sake of a better comprehension of all the text below we will describe briefly the theory and the main structure of the retrieval phase algorithm.

Our system is an object whose electronic density interact with an X ray beam. This beam is diffracted in the sample, so at a certain distance we obtain a 2D diffraction pattern, corresponding to the diffracted intensities for each spatial frequency $I(\boldsymbol{u})$. Note that this diffraction pattern is related with the wave function of the 2D image of the object, $\Psi(\boldsymbol{r})$. This function, as every continuous function, can be decomposed via Fourier transformation in components with a determined spatial frequency. In this way we obtain a diffraction pattern $I(\boldsymbol{u})$ that we now is related with the wavefunction of the object via fouriertransform:

$$I(\boldsymbol{u}) = |\int d\boldsymbol{r}\Psi(\boldsymbol{r})e^{i\boldsymbol{u}\boldsymbol{r}}|^2$$

Now, for recovering the wave function of the object (i.e. the image) we need to know also the phase of the Fourier transform, that is lost in the measuring process. So, as we only have the modulus (and not in all the plane, as we stop the main beam) we need to guess the missing information. For this reason we use some algorithms (HIO,RAAR,ER) for working out the phase numerically. The details of these iterative methods can be found in the excellent review by Marchesini[2,3] .Note that the phase in the reciprocal space is necessary for calculating the modulus in the real space. This modulus is what we consider to be the image, usually neglecting the phase in the real space. In a real experiment, this phase can be important because it is related to the inellastic scattering between light and electrons. Nevertheless, in general the ellastic scattering will be much more important so $\Psi(\boldsymbol{r})$ is real and the phase (in the real space) can be neglected. It is even possible to assume that the image is real from the beginning. This simplifies the algorithm but it is not a necessary constraint, so it is not always used.

The first step in the reconstruction algorithm is to define a support, i.e., a real function $S(x, y)$ that defines the points in which the image is different from zero. This support is a key feature of the process because it is used in every iteration of the algorithm, so if its defined size is very different from the real size of the image the program can converge slowly or even to a wrong solution.

Once we define the support, we make an arbitrary guess for the phase in the reciprocal space, $\theta(\boldsymbol{u})$ (usually we make this guess using random numbers, so that we optimize the convergence speed). As we know the modulus of the Fourier transform, we can make the first guess for the wave function in the reciprocal space:

$$\Psi_{reciprocal}(\boldsymbol{u}) = \sqrt{I(\boldsymbol{u})}e^{i\theta(\boldsymbol{u})}$$

Now is straightforward to inverse Fourier transform to obtain the guess for the wave function in the real space,i.e., the first guess for the image:

$$\Psi(\boldsymbol{r}) = \int d\boldsymbol{u}\Psi_{reciprocal}(\boldsymbol{u})e^{-i\boldsymbol{u}\boldsymbol{r}}$$

Now, we have to define a new wave function for the object using this guess and the support. Usually this new function is defined equal to the guess in the above equation for the points (x, y) inside the support. For the points outside we have a linear combination of the guess and the wave functions obtained in previous steps. The exact form of this linear combination depends of the method used, and on a variable parameter called β $(0 < \beta < 1)$.

Once we have the new wave function of the object, we Fourier transform so that we go back to the reciprocal space. Here, we have a new wave function with modulus and phase (both depend on $\boldsymbol{u}$). The last step is to retain the phase as the guess for next iteration and force the modulus to be equal to the measured. That is, we change the modulus into the square root of the diffraction pattern but keep the phase. When we do this we have the second guess for the wave function in the reciprocal space,

$$\Psi_{reciprocal,2}(\boldsymbol{u}) = \sqrt{I(\boldsymbol{u})}e^{i\theta_2(\boldsymbol{u})}$$

We are again in the starting point so we can repeat the steps again and again until we achieve convergence. After some iterations (roughly 1500 in our case) the method converges into a stable image.

We can improve this algorithm using the shrinkwrap method, that consists simply on modifying the support as the algorithm runs. For this we use a lower threshold, so each n iterations (roughly 100 in our case) we define $S(x, y) = 0$ for every pixel which value is less than that threshold, and $S(x, y) = 1$ for the rest of pixels.

3 Pattern processing and results

The work of the student this summer has been mainly centered in processing diffraction patterns to prepare them for the reconstruction algorithm. This reconstruction has been carried out in one of the two cases, and partially in the second one. In this section we present the methods used and the results.

3.1 Pattern processing

Previous processing of the diffraction pattern (DP) is essential for a correct reconstruction, as can be seen in figure 3 (next page), in which we compare a raw DP with its corrected version (images a and c respectively). The reconstructions

we get for these two DPs for the same parameters are also shown (Note that the parameters are not optimized). If the DP is not correctly prepared before applying the reconstruction algorithm, it may converge into wrong solutions (note that the solution in this case is blurry and discontinuous). We can see that the modified DP produces a much sharper and accurate image.

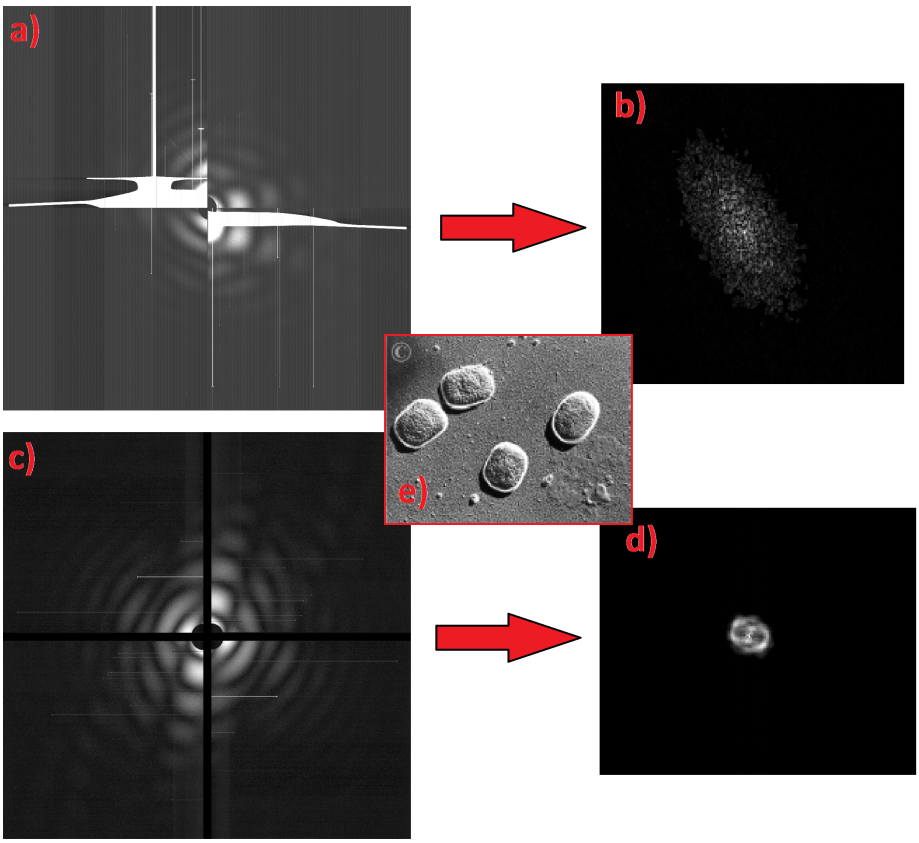

Figure 3: a) Raw DP obtained directly in the detector for a Vaccinia virus sample; b) Reconstructed image obtained applying directly the algorithm to the raw DP; c) The same DP as in case (a) after processing it; d) Reconstructed image obtained applying the same algorithm with the same parameters to the processed DP; e) Electron microscopy image of the Vaccinia virus.

The process of preparing the DP for its posterior analysis depends on each DP, because each of them need some characteristics to be modified. As a consequence, during the summer we have developed some routines that allow us to modify the DP in different ways. The most important implemented routines are the following:

- Routine for reordering of the quarters of the image: the raw DP can have disordered quarters, so the first we have to do is to reorder them to have the correct distribution, as we show in figure 4.

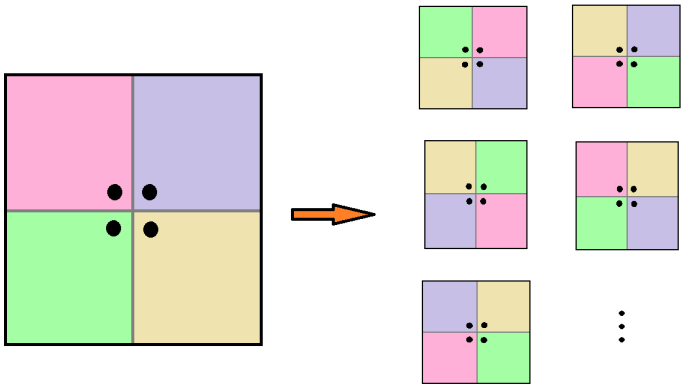

Figure 4: Effect of the routine which allows us to interchange quarters of the image.

- Routine that selects one half of the DP and generates a new DP only with that half. That is, it takes the half we want to use and performs inversion symmetry with respect to the center of the image. With this we are assuming that the two halves are equal and the DP is symmetric under inversion, i.e. we assume that the wave function of the image is real. This is equivalent to saying that there is no inellastic component in the scattering process. In figure 5 you can see how this routine acts.

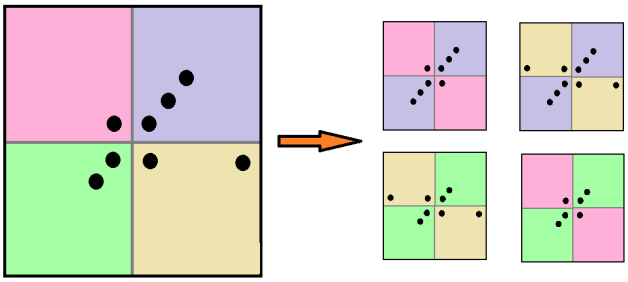

Figure 5: Effect of the routine which selects one half and performs inversion symmetry with respect to the centre of the image.

- Routine for opening gaps in the image (horizontal and vertical) covered by a mask, so that we can correct some possible deviations of the centre of the diffraction pattern with respect to the centre of the image. The effect

of this routine is shown in figure 6, and it is strictly necessary when we use just one half of the DP.

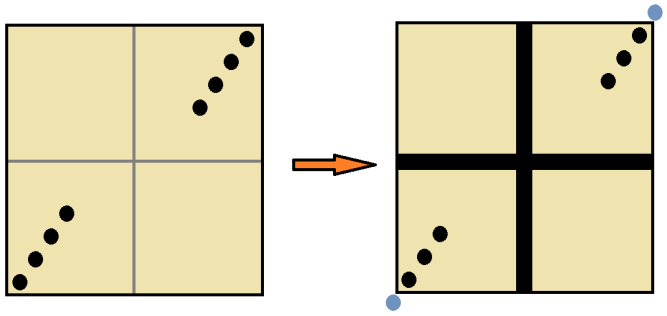

Figure 6: Effect of the routine which opens a horizontal and/or a vertical gap in the diffraction pattern. The black points are covered by a mask and guessed by the algorithm.

- Routine for performing inversion symmetry with respect to a point different from the central pixel of the image (see figure 7 in next page). This program is used in case that some deviation occurs, so the centre of the DP and the centre of the image don't coincide. In this case, we take a n x n square (n is roughly 50 pixels in our simulations) centered in the image and we perform inversion symmetry over the DP with respect to every point in that square. In this way we obtain n^2 different DPs. Amongst them we select for working the one which optimizes this inversion symmetry, finding with this the real centre of the diffraction pattern and making it coincide with the center of the image. The DP that optimizes this symmetry is that which minimizes the next quantity:

$$\frac{1}{2N} \sum_{usefulpixels} \frac{(p_{ij} - \overline{p}_{ij})^2}{\mid p_{ij} \mid\mid \overline{p}_{ij} \mid}$$

  where N is the number of useful pixels, defined as that pixels whose value and the value of its central symmetric is different from zero. p_{ij} is the pixel with coordinates i, j inside the image, and $\overline{p}_{ij}$ is the central symmetric of it (the exact coordinates of this pixel depend on the centre chosen for the inversion).

Each of these routines have been written when necessary for modifying the diffraction pattern, i.e., the reconstruction is a continuous refinement process: if the image is not clear (and doesn't turn clearer when modifying the parameters) we change the input DP. After that we reconstruct it again and observe the results. The reconstruction program (that we had since the beginning), as the routines enumerated above, has been written in IDL.

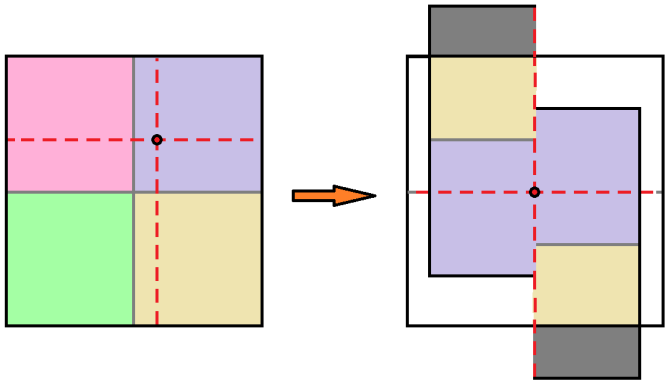

Figure 7: Effect of the routine which performs the inversion symmetry with respect to a point that doesn't coincide with the centre of the image. The right image shows the final DP we insert into the algorithm.

3.2 Results

Along the summer we have studied two different systems. The image of the first one, the mimi virus, has been completely reconstructed, while the image of the second one (Vaccinia virus) has been reocnstructed only in part. Here we present the results we obtained.

3.2.1 Mimi virus

The first sample analized was a mimi virus (see figure 8). The DP was measured in the back detector, and can be seen in its raw version in figure 9 (next page). For the reconstruction we took the lower half and performed inversion symmetry with respect to the centre of the image, leaving a gap between both halves which width was a new parameter of the program (in this case the optimum value was 50 pixels).Figure 10 (next page) shows the amplitude of the reconstructed wavefunction. The shape of the mimi virus is clearly visible, and even it is possible to see part of the spacing between the two parts of the core. The size (around 400 nm) is also in perfect agreement with what we know. Figure 11 (page 14) shows the phase, that is shown to be zero in the region in which the amplitude is not, so the scattering is mainly ellastic (note that this is not a consequence of the (forced) inversion

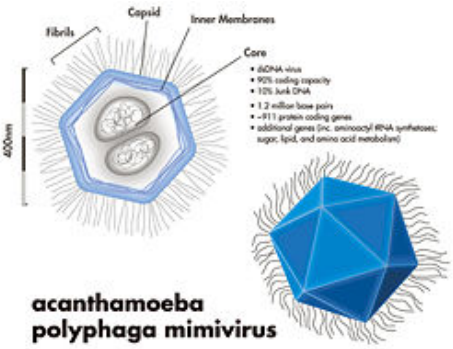

Figure 8: Artistic image of a mimi virus.

symmetry of the DP because we have a gap and the algorithm guesses the points in there, so it can converge into a non-real solution).

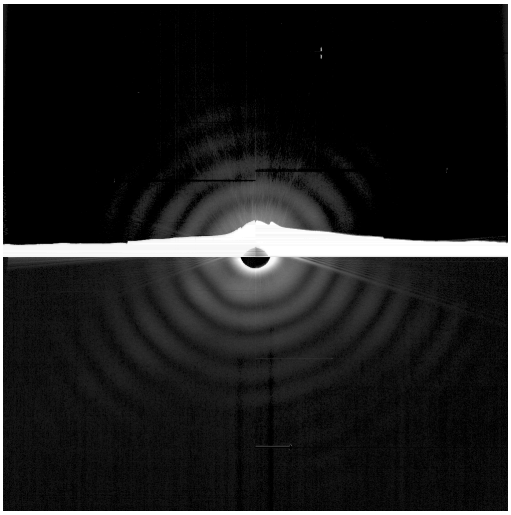

Figure 9: Raw diffraction pattern for the mimi virus, measured on the back detector.

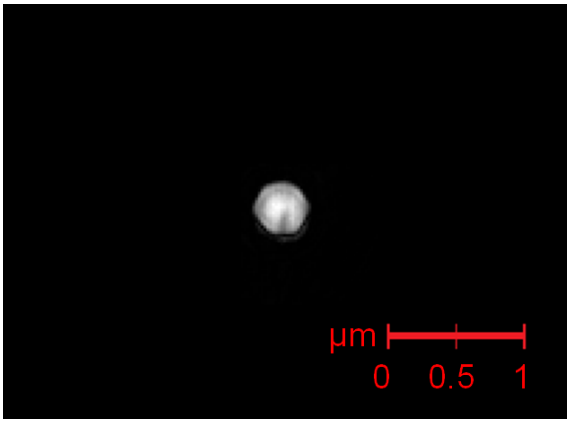

Figure 10: Amplitude of the reconstructed image of the mimi virus.

After the reconstruction, we can chek wether it is correct or not by a simple Fourier transform of the image. We can compare the modulus of this Fourier transform with the original DP and see how similar they are. In the figure 12 (next page) we can see this "'simulated DP"'. If we compare with the original in figure 9 we can see that they look very similar, and also the guessed parts (inside

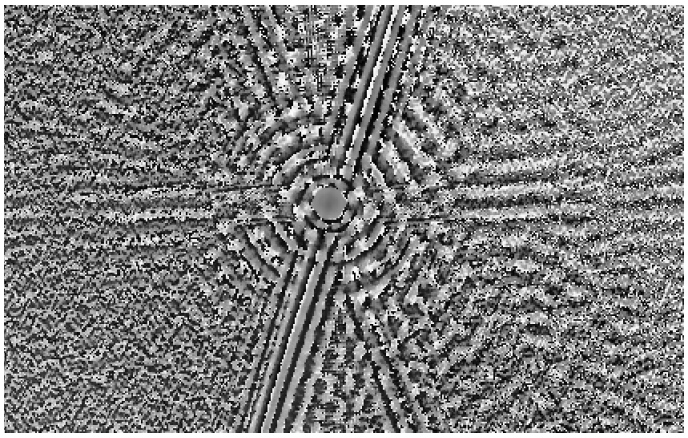

Figure 11: Phase of the reconstructed image of the mimi virus.

the gap) coincide rather good with what is expected: a brilliant central spot and continuous rings. According to these results we can say that our reconstruction is complete and our image accurate.

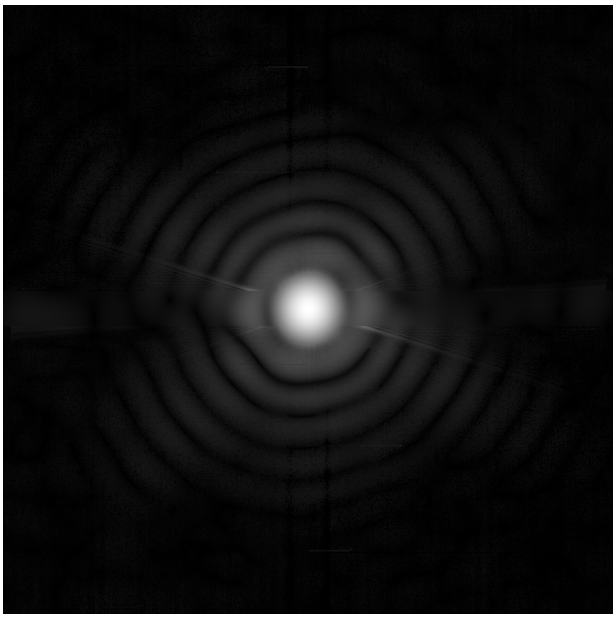

Figure 12: Fourier transform of the reconstructed image (simulated diffraction pattern).

3.2.2 Vaccinia virus

The second sample we studied was the Vaccinia virus. We can see an image in figure 13, obtained with electron microscopy. The raw diffraction pattern and the processed one can be seen in figure 3a and figure 3c respectively. In this case, the data came from the front detector, so the reconstruction was specially difficult. After the basic process of reordering quarters of the DP, we had a problem to solve: during the measurement process the detector was slightly deviated from the centre of the beam. The consequence is that the centre of the DP doesn't coincide with the central pixel of the image, so if we try to use only one half of it (as we did) we must first find the real center and perform inversion symmetry with respect to that pixel. For this reason we wrote the routine explained above (pages 11/12). As we spent a lot of time with the processing (specially with this last routine) we didn't have much for the proper reconstruction, so it was impossible for us to complete it. The best reconstruction we got is displayed in figure 14. We can see it is a little blurry but has the correct shape and size. The edges are not very cear because the parameters of the simulation are not optimized. This is consequence of another difficulty we found during the reconstruction: apart from the horizontal gap (unavoidable because of the beam stop) we had in this case a vertical gap

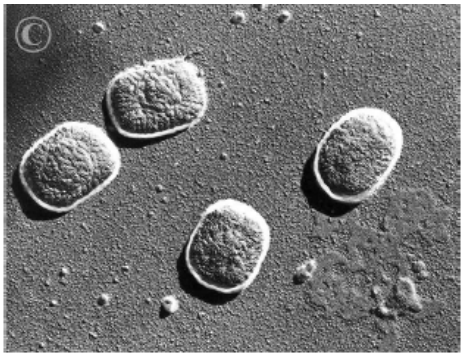

Figure 13: Electron microscopy image of a Vaccinia virus.

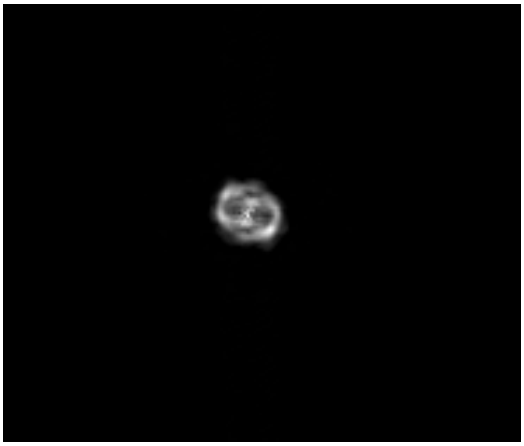

Figure 14: Amplitude of the reconstructed image of the Vaccinia virus.

as a result of the deviation of the centre of the DP. The consequence of this is that we have to handle with two parameters (widths of the two gaps) instead of one, so the reconstruction is much more complicated and requires much more time. Anyway, the results obtained are promising and it seems it would not be a problem for us to reconstruct the image in some more weeks of work.

4 Conclusion

The X ray imaging method is, as we have seen, a powerful tool for biophysics among other areas. The algorithm for phase retrieval is complex and sometimes slow, but it can obtain very accurate solutions and clear images. We have shown that a previous processing of the DP is essential for the algorithm to converge to a clear and sharp image, or even to converge to the correct solution. The process of refining the DP and the posterior reconstruction is difficult because the data have a lot of noise, so we have to "'clean"' the diffraction pattern and sometimes even use just one half (so we have to call an additional subroutine). Besides, we don't have enough information from the experimental setup to work out the correct width of the gaps, so we have to add a new parameter (or sometimes even two) and optimize them, so much time is lost. Nevertheless, we achieved at least one good reconstruction (and one partial one) that shows a sharp and defined image. This shows that the water jet configuration for the imaging experiment is at least as good as the default one, appearing as a promising technique for future experiments.

5 Acknowledgments

First of all, I want to thank DESY organization and staff, for giving me the opportunity of researching here. Secondly, to Henry Chapman and all the CFEL group and collaborators, for taking the data and providing me with all the tools I needed. Finally, I want to thank specially Andrew Martin for helping me with this project and being so patient with me.

References

[1] *Femtosecond diffractive imaging with a soft-X-ray free-electron laser*, H.Chapman et al., Nature physics Vol.2 (December 2006)

[2] *Invited Article: A unified evaluation of iterative projection algorithms for phase retrieval*, S. Marchesini, Review of Scientific Instruments 78, 011301 (2007)

[3] *X-ray image reconstruction from a diffraction pattern alone*, S. Marchesini et al., physical review B 68, 140101(R) (2003)

[4] *Extending X-Ray Crystallography to Allow the Imaging of Noncrystalline Materials, Cells, and Single Protein Complexes*, Jianwei Miao, Tetsuya Ishikawa, Qun Shen and Thomas Earnest, Annu. Rev. Phys. Chem. 2008. 59:387-410

[5] *Single mimivirus particles intercepted and imaged with an X-ray laser*, M. Marvin Seibert et al., Nature Vol.470 (3 february 2011)

[6] *Femtosecond X-ray protein nanocrystallography*, H. Chaoman et al., Nature Vol.470 (3 february 2011)

[7] *Coherent lensless X-ray imaging*, Henry N. Chapman and Keith A. Nugent, Nature Photonics Vol.4, December 2010