Okinawa Institute of Science and Technology
Graduate University

Thesis submitted for the degree

Doctor of Philosophy

Exploring the Potential of Cryo-Electron Tomography on Protein Nanocrystals for Molecular Structure Determination

by

Märt Toots

Supervisor: **U. Skoglund**

April, 2017

Declaration of Original and Sole Authorship

I, Märt Toots, declare that this thesis entitled *Exploring the Potential of Cryo-Electron Tomography on Protein Nanocrystals for Molecular Structure Determination* and the data presented in it are original and my own work.

I confirm that:

- This work was done solely while a candidate for the research degree at the Okinawa Institute of Science and Technology Graduate University, Japan.
- No part of this work has previously been submitted for a degree at this or any other university.
- References to the work of others have been clearly attributed. Quotations from the work of others have been clearly indicated, and attributed to them.
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.
- None of this work has been previously published elsewhere, with the exception of the following:

Publications

- **M. Toots**, L.G. Öfverstedt, G. Wilken, U.Skoglund, "Visualization of 3d lysozyme nanocrystals using cryo-electron tomography", (In preparation)
- F.Mahmood, **M. Toots**, L.G. Öfverstedt, G. Wilken, U.Skoglund,"An Extended Field-based Method for Noise Removal from Electron Tomography Reconstructions", Nature Scientific Reports (Submitted).
- F. Mahmood, **M. Toots**, L.G. Öfverstedt, U.Skoglund, "Algorithm and Architecture Optimization for 2D Discrete Fourier Transforms with Simultaneous Edge Artifact Removal", IEEE Access (Submitted).
- F. Mahmood, **M. Toots**, U. Skoglund et al., "2D Fast Fourier Transforms with Edge Artifact Removal using Optimized Periodic Plus Smooth Decomposition for Real-Time Applications", IEEE Field Programmable Technology (FPT) 2015 Queenstown, New Zealand.

Patents:

- F. Mahmood, **M. Toots**, L.G. Öfverstedt & U. Skoglund, "FPGA Implementation of Multidimensional Discrete Fourier Transforms with Edge Artifact Removal", US Provisional Patent Filed (July, 2015).

Date: April, 2017
Signature:

Abstract

Exploring the Potential of Cryo-Electron Tomography on Protein Nanocrystals for Molecular Structure Determination

The three-dimensional structure of a protein molecule, challenging to determine and close to impossible to predict, plays a key role in understanding protein function and has implications in drug design. When it comes to structure determination, there exist many complementary methods, each with their specific advantages and disadvantages. Most of those methods rely on a combined signal from thousands of individuals and cannot be used for directly reconstructing an actual 3d volume as it appears inside the sample.

This thesis focuses on developing the methodology and providing proof of concept for a novel approach in structure determination by reconstructing small protein nanocrystals via cryo-electron tomography. Real-space imaging gets past the phase problem that is a challenging companion of conventional diffraction-based methods. With electron tomography we can reconstruct and visualize a 3d nanocrystal in its entirety and study the properties of small biological crystal from a new perspective.

Being a relatively unexplored territory, nanocrystal tomography sets several challenges, such as creating nanocrystals small enough for imaging with transmission electron microscope and developing algorithms for going from a tilt-series to a 3d structure. For a proof of concept we create, image and reconstruct nanocrystals of hen egg white lysozyme that, having molecular weight of only 15 kDa, is generally considered unfeasible for electron tomography. Nanocrystals make finding and determining the relative orientations of the individual molecules possible, symmetry relations help reduce the effects of missing information, and by averaging we are able to reconstruct a molecular structure at a medium resolution of around 13 Å. Using Fourier Transform (FFT) we get a direct objective measure of the resolution of details within the reconstruction in the form of a diffraction pattern and show that in specific directions the resolution reaches as high as 7 Å in a single tomogram.

Additionally, this work explores two other tightly related ideas. First, we study the concept of extended field and show with extensive simulations that extending the reconstruction space in various regularized iterative reconstruction procedures helps reduce the overall error and prevent over-smoothing. Second, calculating FFT of an image comes at a computational cost, and when the image is not periodic, the discontinuity of the opposing edges causes undesirable strong artifacts in the FFT that could obstruct important details. In this project we implemented a simultaneous 2d FFT and edge artifact removal for real-time applications on a Field Programmable Gate

Array (FPGA) reconfigurable computing system.

Acknowledgment

A number of people have played significant role in my PhD studies. Without them the completion of this thesis would not have been possible. With this I am taking the opportunity to express my warmest gratitude for their guidance and support.

First, and foremost, I would like to thank my supervisor Prof. Ulf Skoglund for introducing me to the intriguing field of structural biology and especially to cryo-electron microscopy. The long discussions about the past, the present, and the future, offered insights into the origins of the field, helped understand conventional as well as alternative viewpoints and tools of the trade, and his visionary glimpses to what the future might hold helped broaden my view and set an exciting context for the thesis research. He provided guidance in designing the experiments and interpreting the results.

I would also like to thank all other members of the Skoglund unit for creating a pleasant work atmosphere. But most notably, I'd like to thank Dr. Lars-Göran Öfverstedt for sharing his extensive knowledge about computation in general, the software package used for tomographic reconstruction in particular, and offering guidance in programming with Fortran. L.-G. also shared his know-how and gave me lots of practical tips about using the transmission electron microscope. Dr. Gunnar Wilken for mathematical insights about crystallographic lattice detection program. Dr. Faisal Mahmood for good companionship over the coffee breaks, discussions about image processing and for co-authoring research work together. Shizuka Kuda for amazing administrative powers in dealing with the paperwork and organizing travels to conferences and summer schools.

I would also like to thank my academic mentor Prof. Kenji Doya for offering his support, and Prof. Mahesh Bandi for supervising during my lab rotation work in his unit, that led to the publication of an article about the fluctuations in wind energy. I would like to express my gratitude also to the OIST Graduate School and the tireless Student Support staff for all kinds of support, and the Japanese Government for financing my PhD studies in OIST.

Finally, I would like to thank my wife Kristiina for staying strong by my side during the emotional roller-coaster of a PhD.

Abbreviations

CET	Cryo-electron tomography
COMET	Constrained maximum entropy tomography
ED	Electron diffraction
EM	Electron microscopy
ET	Electron tomography
FFT	Fast Fourier transform
HEW	Hen egg white
HEWL	Hen egg white lysozyme
microED	Electron diffraction on micro/nano-crystals
NCT	Nano-crystal tomography
PDB	Protein data bank
SFX	Serial femtosecond x-ray crystallography
SPR	Single particle reconstruction
SNR	Signal to noise ratio
tSNE	t-distributed Stochastic Neighbor Embedding
TEM	Transmission electron microscope
XFEL	X-ray free electron laser
XRC	X-ray crystallography
XRD	X-ray diffraction

To Johanna

Contents

List of Figures

List of Tables

Introduction

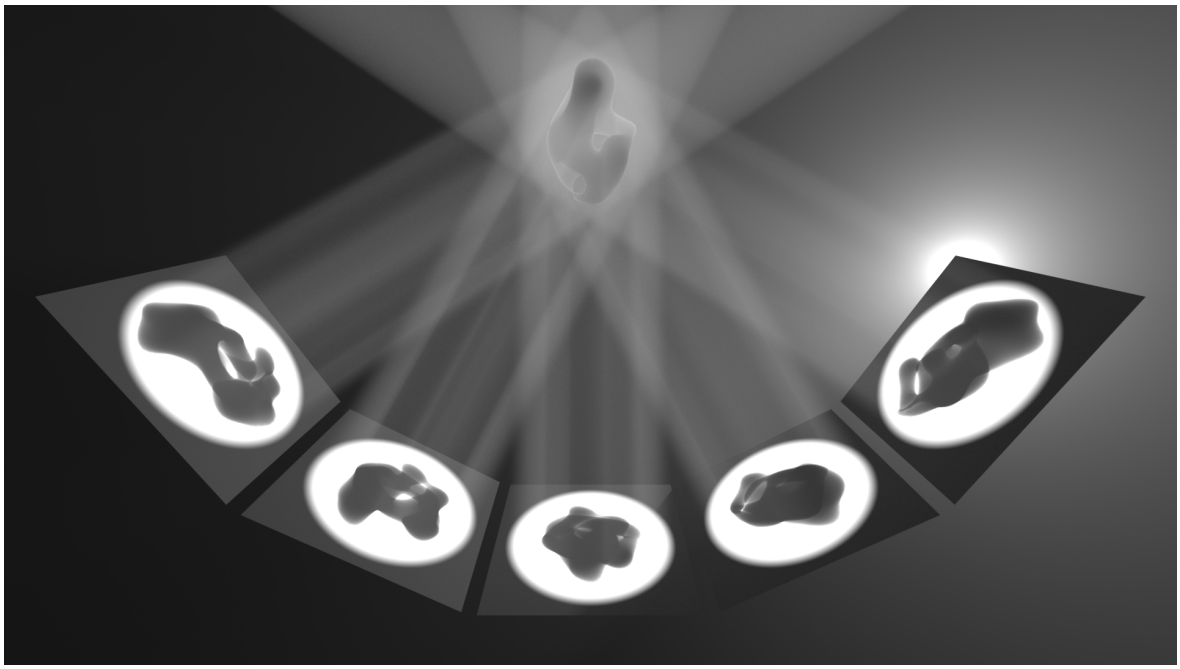

Figure 1 | An illustration of tomography. Rendered with *Blender*.

The world is full of living organisms inhabiting an astounding variety of environments from acidic hydrothermal vents deep under the sea, through warm and moist rain forests to cold and windy mountain tops. Despite the huge variety of the living systems they all share the same fundamental principles necessary to support life. A class of biological macromolecules known as proteins are responsible for most of their functions.

Proteins are biological molecules that are constructed inside a cell from amino acids. A single protein molecule is a chain with up to several thousand amino acids linked together in a linear fashion, but describing proteins by this sequence alone does not capture their biological significance.

In its natural hydrated environment a protein chain folds into a characteristic three-dimensional form due to a combination of forces acting on the amino acid chain, like the electrostatic force, the hydrogen bonds, the disulfide bridges, van der Waals force and the forces caused by the hydrophobicity and hydrophilicity. The resulting specific structure in general has different larger functional domains connected by smaller flexible linkers that are in perpetual movement around equilibrium due to the thermodynamic collisions of the surrounding molecules. This flexibility and constant change

in conformation gives proteins the ability to function in a highly specific manner [1].

These observations motivate many questions about the nature of those molecules. If a dissolved protein is not a rigid structure then how does it move? What are the statistics of its dynamics? What is the extent of its vibrations? In its natural environment, what is the time distribution of its different conformations? Can we estimate the time distribution of one molecule by studying ensemble distribution of many molecules at a single time instant? How much does the structure of a protein packed in a crystal differ from the dissolved protein? Does a protein remain active in a crystalline form and to what extent are its movements reduced? How does that freedom vary within the crystal?

Being able to answer any of these questions will take us closer to understanding the living world around us. Knowledge about the structure and dynamics of proteins will lead us to more accurate and reliable computer simulations that will aid in the development of new drugs in particular, but also understanding the nature in general.

Over the past one hundred years a variety of methods have been developed whose main goal is to determine the 3d structure of various biological molecules. These include x-ray crystallography of large well-ordered 3d crystals and serial femtosecond crystallography of small 3d crystals, electron crystallography of 2d and 3d crystals, electron tomography and sub-tomogram averaging of individual particles, electron microscopy using single particle reconstruction or helical reconstruction of tubular molecules, nuclear magnetic resonance imaging and nuclear scattering of 3d crystals. Each one of these techniques have certain strong suits and bottle-necks and the nature of the scientific question driving the experiment determines which method is optimal. All of the previously mentioned methods with the exception of electron tomography rely on averaging many molecules that can greatly enhance the signal and suppress the noise, but it inevitably brings forth the loss of individuality.

An approach to structure determination, that has for various reasons been less pursued, is using electron tomography on small protein nanocrystals, where one would have the option to average particles over a well-defined lattice while retaining the possibility of investigating the crystal also without averaging. In this thesis we set the goal to explore this avenue further, and attempt to develop a methodology that could determine the structure of a single molecule as well as provide insights about proteins in the context of small biological crystals.

Goals and Outline of the Thesis

The major goal of the current thesis is to extend the method of cryo-electron tomography to a previously little explored territory of using 3d protein nanocrystals for molecular structure investigations. Electron tomography, a comparatively low resolution method, is the only method that can reconstruct a part of a 3d volume as it appears inside the sample without any explicit need for averaging, which makes it an ideal candidate in the pursuit of knowledge about the individuality of molecules.

In the thesis we develop the methodology and show as a proof of concept using lysozyme that electron tomography is not only a feasible solution for at least medium resolution structure determination, but can also be effectively used to investigate the

nature of 3d protein nanocrystals. In tomography we have a real image of the crystal y taking advantage of a mathematical method called Fourier transform we can switch at will between real and reciprocal space representations, thereby allowing the incorporation of well developed crystallographic methodology without being stymied by the phase problem.

Motivation

There are several reasons to motivate the usage of nanocrystals and why it could be a lucrative approach.

- When embedded in amorphous ice, small organic molecules provide little contrast, which makes them hard to find. Crystals, consisting of thousands of molecules aligned on a lattice, generate plenty of contrast at specific directions. The lattice can also serve as a road map that leads towards the molecules, by looking for periodicity in the reconstruction.

- Averaging many (i.e. thousands) single particles that are floating freely in the ice layer is the conventional approach, but the relative orientations of small monomeric molecules are hard to determine, which makes the procedure challenging and error-prone. Inside a crystal the molecules are naturally aligned and related to each other by well-defined symmetry operations. only things that have to be determined are the crystallographic parameters - lattice vectors, symmetry group and the positions of the symmetry elements, given that the crystal is well ordered.

- Large well-ordered crystals, that are suitable for conventional x-ray diffraction studies, are notoriously difficult to obtain. It is believed, and backed up by recent research [2], that the production of nanocrystals is feasible for many molecules that otherwise resist growing into large crystals.

- Electron crystallography on 3d nanocrystals has recently seen daylight and is producing spectacular results [3, 4]. To get a good diffraction pattern that is suitable for crystallographic analysis a well ordered nanocrystal is necessary. Electron tomography in principle can deal with a nanocrystal of any quality and averaging can be done over only the parts that show good order, given that we can identify those. Moreover, only a single crystal in a single dataset might be all that is necessary, whereas in electron crystallography generally several crystals are needed. The compromise is that electron tomography will probably not provide similar resolution as is the norm these days for electron crystallography and other structure determination methods.

- In addition to an averaged structure, tomography offers an opportunity to directly image and reconstruct an actual part of a volume in the sample. This gives a chance to study protein nanocrystals as a separate system, giving us information about the structure of the whole crystal itself, the disorder within, molecule packing and mosaicity. It might be even possible, dependent on the quality and SNR of the data, to analyze every single molecule individually since we can pinpoint

their locations in the nanocrystal. Molecules in a nanocrystal are potentially less tightly packed than in a large crystal, therefore having greater flexibility and freedom of movement. This could serve as an ideal system for extracting a whole gallery of the molecules to provide statistics about the conformational dynamics.

While we certainly cannot address in this thesis every point that is discussed in the preceding list, we nevertheless explore the paradigm of nanocrystal tomography from many of those aspects.

Outline of the Thesis

In addition to the main overarching theme of nanocrystal tomography, the thesis also touches upon two tightly related problems to tomography and signal processing in general. The three projects, all unified under a common theme, are the following

- Speeding up fast Fourier Transform by harnessing hardware, at the same time handling the edge artifacts when transforming a non-periodic image.
- The concept of extending the size of a tomographic reconstruction beyond the region of interest and its effects on the reconstruction.
- The explorations in reconstructing 3d protein nanocrystals, extracting a single molecule structure and studying the crystalline order.

In Chapter 1 we first give an overview of some of the background of imaging in structural biology, electron microscopy, relevant details regarding image formation and the mathematics of 3d reconstruction. We also discuss the basics of the signal processing technique called Fourier transform, keeping it short on the theoretical side and putting more emphasis on the practicalities.

Chapter 2 is devoted to the discussion about the computational intricacies of 2d Fourier transform, a specific approach to edge-artifact removal in a non-periodic image, and implementing it all on hardware using field programmable gate array technology (FPGA).

In Chapter 3 we take a look at the extend field iterative reconstruction technique and compare it to its non-extended analogues. A rigorous theoretical treatment is left for the future developments, but a thorough simulation study using various iterative reconstruction procedures with regularizing capability shows its effectiveness in separating signal from the noise and in facilitating better fit with the projections.

Finally, Chapter 4 presents the core project of this thesis. We discuss the methods we used for growing 3d lysozyme nanocrystals and the subsequent imaging in transmission electron microscope. We talk about the upsides and downsides of this approach to molecular structure determination and show how medium resolution structure of lysozyme was determined from just a handful of unit cells. The last section of this paragraph explains the work-flow and the algorithmic details of the programs that were developed for crystalline image processing task.

Chapter 1

Background

1.1 Imaging in Structural Biology

X-ray Crystallography and Protein Crystallization

In structural biology the conventional method for structure determination has historically been X-ray crystallography (XRC) [5, 6]. In 1958 Kendrew et al. solved the structure of a sperm whale myoglobin by XRC [7]. This was the first ever protein structure determined by this method and served as a major milestone in structural biology. It showed to the world that the structures of proteins are in complexity far beyond what any theory had so far predicted [7].

A major obstacle in XRC is achieving protein crystals of the necessary size and quality for diffraction [8], with insoluble, flexible and otherwise unstable membrane proteins being notoriously difficult to crystallize [9]. It has been estimated that in most organisms around 25% of all genes encode for membrane proteins [10] whereas in the protein data bank (PDB) [11] out of more than 100 000 structures solved to date only ∼2% are those of membrane proteins. In spite of the difficulties associated with crystallizing membrane proteins they serve as attractive research objects since about 60% of the current drug targets are located on the cell surface [12].

X-ray Free Electron Lasers and Nanocrystallization

Until recently nanometer sized crystals were unfit for diffraction studies using conventional X-ray sources because of the radiation damage caused by the prolonged exposure [13]. This changed with the advent of x-ray free electron lasers (XFEL) that introduced an incredibly tightly focused beam that is a billion times more brilliant than what the so-called third generation synchrotron was able to deliver [14]. A paradigm shift dawned when it was realized that rather than trying to fight the inevitable radiation damage one might instead accept the destruction of the crystal as long as the diffraction pattern can be recorded [15]. "Diffraction before destruction" approach uses super intense and a mere femtosecond long XFEL pulses on protein nanocrystals to "outrun" the destruction process and record the diffraction pattern before the onset of the radiation damage [16].

In serial femtosecond x-ray crystallography (SFX) [15] millions of randomly oriented

nanocrystals are passing through the pulsating electron beam in a narrow stream, partial diffraction patterns are recorded and later combined to fill the reciprocal space [17]. Because of the restrictions set by the nanocrystal delivery system to the XFEL beam SFX can in general handle crystals in the range of 100 nm to 1 µm and each recorded diffraction pattern needs to have at least 10-16 identifiable diffraction spots for it to be useful [18]. The structure of a membrane protein Photosystem I (PSI) was solved using nano and microcrystals ranging from 200 nm to 2 µm [15].

This new application for protein nanocrystals ignited a demand for reliable and reproducible nanocrystallization methods. One approach is offered by microfluidic devices, that were also used to crystallize and sort the PSI crystals [19, 20]. Since optimal crystal growth conditions occur in the metastable zone yet the nucleation takes place in the supersaturation region [21], fine control over the system's trajectory through the phase space is necessary and offered by precisely tuned microfluidic devices.

Detecting nanocrystals in solution is a nontrivial matter as they are not resolvable to bright field light microscopy. Dynamic light scattering, fluorescence microscopy and second-order nonlinear imaging of chiral crystals have been used as a characterization method for sample crystallinity [2, 20]. Recently, it has been shown that complexes that are otherwise hard to crystallize for XRC, such as membrane proteins and multiprotein complexes, still might form nanocrystals in a crystallization drop [2]. Stevenson et al. demonstrated by using uranyl acetate stained samples and transmission electron microscope that nanocrystals with discernible crystal lattice were commonly observed . Another successful approach to producing nanocrystals is to grow large crystals the conventional way and later fractionate them using sonication, vortexing or vigorous pipetting to break larger crystals into small single crystals [3]. Crystals obtained this way are often free of the many imperfections that more frequently accompany large crystals, such as twinning, and are thus well suited for electron crystallography.

Electron Crystallography

Nanocrystals in the order of 100 nm can also be analyzed in transmission electron microscope. When an electron beam hits a crystalline sample, the electrons are scattered and a diffraction pattern can be recorded on the back focal plane of the objective electromagnetic lens.

In 2d electron crystallography 2d protein crystals are needed [22]. The structure of aquaporin (AQP0) membrane protein was solved to resolution 1.9 Å from 2d crystals by combining diffraction patterns from several hundred different crystals [23]. The downside of 2d electron crystallography is that growing and handling 2d crystals can be incredibly tricky [24]. Also with 3d protein crystallography there are several challenges that have to be overcome. The foremost is the fragility and beam sensitivity of protein crystals, which prevents tilting and collecting multiple exposures from the same crystal [25]. Thus, low SNR is always present because long exposure by the electron beam may destroy the crystals.

Recently, with the advent of direct electron detectors a technique called *microED* has seen daylight [4]. By using ultra low accumulated electron dose of $\sim 9\, e^-/\text{Å}^2$ [26], or sometimes even lower like $\sim 5\, e^-/\text{Å}^2$ [27], microED is capable of taking several exposures from different tilts by continuously rotating the same 3d nanocrystal covering a large

angular wedge. In a proof of principle study Shi et al. were able to refine the structure of lysozyme up to 2.9 Å [26], . Recently structures of many more small biological molecules have been solved to high resolution [28, 29]. When the diffraction spots reach 1.2 Å direct methods can be used for phasing [28]. When the peaks do not quite reach that resolution, then isomorphous replacement and other more challenging techniques need to be employed to solve the phase problem.

Single Particle Reconstruction

In single particle reconstruction (SPR), the particles of interest do not have to be in crystalline order and are usually scattered around the sample in random orientations. Treating the differently oriented particles as different views of the same particle allows for reconstructing a single 3d image from thousands of 2d projections. Many high resolution structures have been obtained by SPR, such as mammalian transient receptor potential TRPV1 ion channel to a resolution of 3.4 Å, breaking the side-chain resolution barrier for a membrane protein without crystallization [30], and cytoplasmic polyhedron virus to 3.88 Å resolution [31] by exploiting the icosahedral symmetry of a virus capsid [32, 33]. More recently SPR has also reached even higher resolution of 2.2 Å on β-galactosidase [34] .

A helical structure being highly symmetric contains many different views of the repeating unit and therefore makes reconstruction possible from only a single image [35]. Optical filtering to filter out non-repeating structures have been demonstrated on the helical tail of bacteriophage T4 [36].

A combination of electron crystallography and single particle analysis was demonstrated on nanocrystals in [37]. In a proof of concept study Nederlof et al. imaged less than 100 nm thick flash-frozen samples containing lysozyme nanocrystals. By using optimal Wiener filtering for lattice enhancement and single particle techniques on around 200 crystals they showed that nanocrystals can be imaged to high resolution with electron microscope.

Electron Tomography and 3d Reconstruction

Crystallography and SPR are both ingenious and incredibly useful methods for obtaining the structure of a protein molecule to high resolution, but when it comes to the study of cellular processes, they fall short on several important aspects. Firstly, the resulting structure is always obtained as a combination of a large amount of molecules, therefore losing the information about the individuality of each molecule. In SPR a classification of the images into particles in different conformations is possible and commonly done, but the particles need to be relatively large and there is a significant threat of model bias [38]. . Secondly, only rarely can specific biological functions be attributed to individual molecules. Majority of the functions in the crowded environment of the cell are taking place as a carefully coordinated balance between several molecular assemblies and their functional domains. Therefore, knowing the exact structure of a protein that is packed tightly inside a crystal will probably not provide us with a complete picture of its dynamics inside the cell [1].

Cryo-electron tomography (CET) can in principle overcome these shortcomings.

With the advance of cryo-techniques for sample preparation and the availability of powerful computers, CET has become increasingly useful to fill the niche that other structure determination methods are unable to provide. Even though CET as of yet can not provide a structure with as high a resolution as XRC or SPR, it has the advantage of reconstructing an actual part of volume inside a sample, allowing three dimensional imaging of any electron-transparent biological matter without the need for averaging. With CET, therefore, one can image individual molecules in their natural environment [39, 40]. In the context of the current work, with CET we can image protein nanocrystals of sufficiently small size and obtain information about each individual molecule.

In CET a series of images of the same sample are collected with transmission electron microscope (TEM) by tilting the sample around an axis perpendicular to the electron beam. As the focal depth of TEM is more than 100 nm the projections can be regarded as line integrals of the scattering potential of the sample. A 3d image is thereafter reconstructed by projecting the 2d images in tilt-series back to the 3d space via an operation referred to as back-projection [41, 42].

CET sets no particular constraints on the type of sample as long as it is thin enough to make it transparent for electrons. For a microscope operating at 300 keV an acceptable thickness of the sample is in the order of 100 nm, thicker samples can be sliced using cryo-microtome [43]. In cryo-electron microscopy the biological structures are preserved by either flash-freezing by plunging them into liquid ethane [44] or high pressure freezing [45] so that the water inside the sample gets vitrified, thus forming a glass-like amorphous ice instead of crystalline ice. A sample prepared in this way has to be kept below the recrystallization temperature at all times in order to avoid losing the amorphousness of the ice.

Combination of XRC and CET was performed in [39] where immunoglobulin G (IgG) molecules in many different conformations were successfully imaged using CET. After assembling a gallery of different conformations the high resolution XRC structures were individually fit into IgG domains. Thus, rigid XRC high resolution structures can be augmented with information about the dynamics of different domains by using low resolution CET [46].

Cryo-electron tomography of biological samples operates in an ultra low dose regime where a fixed total dose, that is already low, has to be distributed over all individual tilts. In such a setting images are dominated by shot noise, the signal to noise ratio (SNR) is in the order of 0.1, images are devoid of contrast, hard to align with each other and small proteins are difficult to find.

Image reconstruction in electron tomography is in itself an ill-posed mathematical problem stemming from the limitations of the data collection scheme, i.e. tilting at discrete steps and not covering the whole 180° range. The additional challenges caused by the low dose only make things more severe. Therefore, in order to cast an unsolvable ill-posed problem into a neighboring well-posed problem some *prior information* has to be included in the form of a regularizer. Several regularization schemes have been tried out on electron tomography reconstructions like Tikhonov regularization, total variation minimization [47], compressed sensing [48], shape based [49] and model based regularization, and many more.

One particular regularizing method that is used in the current work is constrained

maximum entropy tomography (COMET) [50]. COMET is an iterative tomographic reconstruction procedure with maximum entropy based regularization [51]. It finds a solution to the tomographic reconstruction problem as an interplay between two measures - data discrepancy characteristic, that pushes the solution towards better fit with the observations, and an entropy term, that smooths out the details not conclusively backed up by the data. In COMET, a low resolution version of a filtered back-projection structure is used as prior information, since we can be fairly certain in the correctness of the low resolution details. Maximizing entropy guarantees that no possible solution of the reconstruction problem is overlooked and at the same time staying maximally unbiased towards any particular one [52]. In terms of electron tomography this means that the resulting reconstruction includes only those high resolution details whose existence can be backed up by the observed data, i.e. the tilts. In [53] it was concluded, based on [54], that starting from certain axioms that a reconstruction method should satisfy, such as consistency, distinctness and continuity, it can be shown that the entropy regularization is the only method consistent with the axioms when the signal to be reconstructed is positive and real valued. .

Sub-tomogram Averaging and Symmetrization

The technique of averaging individual 3d reconstructions of supposedly identical particles is called sub-tomogram averaging. It is essentially a combination of cryo-electron tomography and single particle reconstruction [55]. Since in each reconstruction the particle is potentially in a different orientation and possibly also conformation, the reconstructions have to be classified and aligned prior to averaging together.

In [56] a least-squares refinement procedure is presented that performs alignment and averaging of individual 3d images. The method is an iterative procedure based on maximizing the correlation coefficient between the average structure and each reconstruction within a suitably chosen envelope, producing an average that is minimally biased with respect to the order of the inclusion of the structures.

One of the first cases of sub-tomogram averaging was done by combining four individual pre-messenger rhibonucleoprotein particles [57].

More modern examples of sub-tomogram averaging are based on the Bayesian view where a single object is not in one particular orientation, but instead its orientation is described by a probability distribution. Upon averaging, each particles is included in every possible orientation weighted according to its probability [3].

In the thesis project our goal is to reconstruct a 3d image of a protein nanocrystal using electron tomography. For a crystal the asymmetric units are already approximately aligned to each other according to a certain crystallographic symmetry group, so ideally no alignment search is necessary.

1.2 Technical Background

1.2.1 Fourier Transform

Probably one of the most commonly used techniques in signal processing in general, and in this work in particular, is the one called *Fourier transform.* With Fourier transform

we can represent any (reasonably well-behaved) function as an appropriately scaled infinite sum of sines and cosines.

Fourier transform $\mathcal{F}$ of a function $f : \mathbb{R}^n \to \mathbb{C}$, denoted by $\mathcal{F}f$ or with a small *hat* on top $\hat{f} : \mathbb{R}^n \to \mathbb{C}$, can be defined as

$$\hat{f}(\boldsymbol{q}) = \mathcal{F}f(\boldsymbol{q}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{i2\pi \boldsymbol{x}\cdot\boldsymbol{q}} d\boldsymbol{x} \tag{1.1}$$

and, conversely, the inverse Fourier transform as

$$f(\boldsymbol{x}) = \mathcal{F}^{-1}\hat{f}(\boldsymbol{x}) = \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{q}) e^{-i2\pi \boldsymbol{x}\cdot\boldsymbol{q}} d\boldsymbol{q}\,. \tag{1.2}$$

The question whether the integral converges or not is not a concern when the function to be transformed is a measured physical quantity. That is a valid sufficient condition for the existence of the Fourier transform, so in the context of imaging such mathematical details can be ignored [58].

When the function $f(t)$ is 1-dimensional and its argument represents the time domain, then the domain of its Fourier transform is temporal frequency. For a 2d and 3d signal $f(\boldsymbol{x})$ that measures some physical quantity over the spatial domain, such as 2d micrographs or 3d tomograms, the domain of their Fourier transform is spatial frequency. This conveniently allows the decomposition of a signal into its frequency spectrum, allowing the analysis and processing of the frequency components individually. The Fourier space is often also called the *reciprocal space*, a terminology which we will use throughout this thesis.

1.2.2 The Convolution Theorem

A convolution is a mathematical operation where one function is modified by another function, resulting in a third function that has the characteristics of both of the functions involved. The convolution of two functions $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ is defined as

$$(f \circ g)(\boldsymbol{x}) = \int_{\mathbb{R}^n} f(\boldsymbol{y}) g(\boldsymbol{x} - \boldsymbol{y}) d\boldsymbol{y}\,. \tag{1.3}$$

It is easy to show, using the definition of the Fourier transform and a change of variables, that the convolution of f and g is equivalent to a multiplication of their respective Fourier transforms followed by an inverse Fourier transform. So, the Fourier transform of $f \circ g$ is equal to the product of the Fourier transforms of f and g

$$\mathcal{F}(f \circ g)(\boldsymbol{q}) = \mathcal{F}f(\boldsymbol{q})\mathcal{F}g(\boldsymbol{q})\,. \tag{1.4}$$

The opposite is also true. The Fourier transform of a product of two functions fg is equal to the convolution of their Fourier transforms

$$\mathcal{F}(fg)(\boldsymbol{q}) = \mathcal{F}f(\boldsymbol{q}) \circ \mathcal{F}g(\boldsymbol{q})\,. \tag{1.5}$$

This convolution-multiplication duality allows us to write down elegant forms for filtering, calculating cross-correlation and expanding a motif into a repeating pattern by convolving it with a lattice.

1.2.3 Cross-correlation

The cross-correlation function $\mathcal{C}$ of two functions $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ and $g(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ is a measure of similarity between the two. Let us define the cross-correlation between $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ as

$$\mathcal{C}(\boldsymbol{x}) = \int_{\mathbb{R}^n} f(\boldsymbol{y}) g(\boldsymbol{y} + \boldsymbol{x}) d\boldsymbol{y} \,. \tag{1.6}$$

The definition of $\mathcal{C}$ is similar to a convolution, therefore it is expected that we can conveniently make use of the Fourier transform for its calculation. The difference from convolution is that one of the functions is not flipped in the cross-correlation case. It can be shown that

$$\begin{aligned} \mathcal{F}\mathcal{C}(\boldsymbol{q}) &= \mathcal{F}\left\{ \int_{\mathbb{R}^n} f(\boldsymbol{y}) g(\boldsymbol{x} + \boldsymbol{y}) d\boldsymbol{y} \right\}(\boldsymbol{q}) \\ &= \mathcal{F} f^*(\boldsymbol{q}) \mathcal{F} g(\boldsymbol{q}) \,, \end{aligned} \tag{1.7}$$

where * denotes the complex conjugate and is necessary for the imaginary parts of the transforms to also give a positive contribution to the correlation.

1.2.4 Unit Cell, Lattice and Crystal

Next let us define some of the concepts that come up when talking about crystals. In crystallography a unit cell is the smallest building block, also called the *motif*, that when periodically replicated in space generates the whole crystal. The unit cell itself can contain several identical pieces, called the *asymmetric unit*, which can all be mapped into each other by using symmetry operations, such as rotations and translations.

In 3 dimensions the shape of a unit cell is a general parallelepiped and completely specified by its 3 lattice vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$.

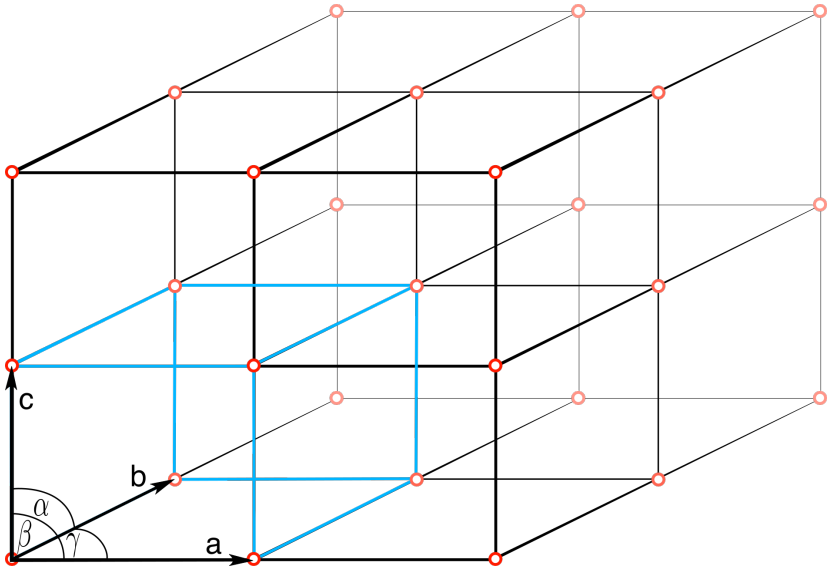

Figure 1.1 | Lattice is defined by the three vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. The unit cell is depicted in blue. α, β and γ are angles between lattice vectors.

Infinite Lattice and Infinite Crystal

To talk about crystals we have to define a lattice. A lattice can be thought of as a grid representing discrete translational symmetry in as many directions as the dimensionality of the space. In three dimensions an infinite lattice $L(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^3$ can be defined as

$$L(\boldsymbol{x}) = \sum_{n_1 \in \mathbb{Z}} \sum_{n_2 \in \mathbb{Z}} \sum_{n_3 \in \mathbb{Z}} \delta(\boldsymbol{x} - n_1\boldsymbol{a} - n_2\boldsymbol{b} - n_3\boldsymbol{c}), \tag{1.8}$$

where $\delta(\boldsymbol{x})$ is the Dirac delta function

$$\delta(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} = \boldsymbol{0}\,, \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \int_{\mathbb{R}^3} \delta(\boldsymbol{x})\, d\boldsymbol{x} = 1\,. \tag{1.9}$$

The Fourier transform of the infinite lattice can be shown to be

$$\hat{L}(\boldsymbol{q}) = \sum_{n_1 \in \mathbb{Z}} \sum_{n_2 \in \mathbb{Z}} \sum_{n_3 \in \mathbb{Z}} \delta(\boldsymbol{q} - n_1\boldsymbol{a}' - n_2\boldsymbol{b}' - n_3\boldsymbol{c}'), \tag{1.10}$$

where $\boldsymbol{a}'$, $\boldsymbol{b}'$ and $\boldsymbol{c}'$ are the reciprocal lattice vectors. As we can see, the Fourier transform of a real space lattice is in itself another lattice, a *reciprocal lattice*.

Let $\rho_0(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^3$ represent the electron density of one unit cell, and $\hat{\rho}_0(\boldsymbol{q})$, $\boldsymbol{q} \in \mathbb{R}^3$ its Fourier transform. An infinite crystal $\rho_\infty(\boldsymbol{x})$ can then be constructed by convolving the unit cell ρ_0 with the infinite lattice L

$$\rho_\infty(\boldsymbol{x}) = L(\boldsymbol{x}) \circ \rho_0(\boldsymbol{x}) = \int_{\mathbb{R}^3} L(\boldsymbol{y})\rho_0(\boldsymbol{x} - \boldsymbol{y}) d\boldsymbol{y} \tag{1.11}$$

and by the convolution theorem the Fourier transform of ρ_∞ is equivalent to the product of the transforms of ρ_0 and L

$$\hat{\rho}_\infty = \hat{\rho}_0 \hat{L}\,. \tag{1.12}$$

Creating an infinite crystal from a motif is effectively equivalent to sampling the Fourier transform of the motif at the points corresponding to the reciprocal lattice. These relations in two dimensions are visualized on Figure 1.2

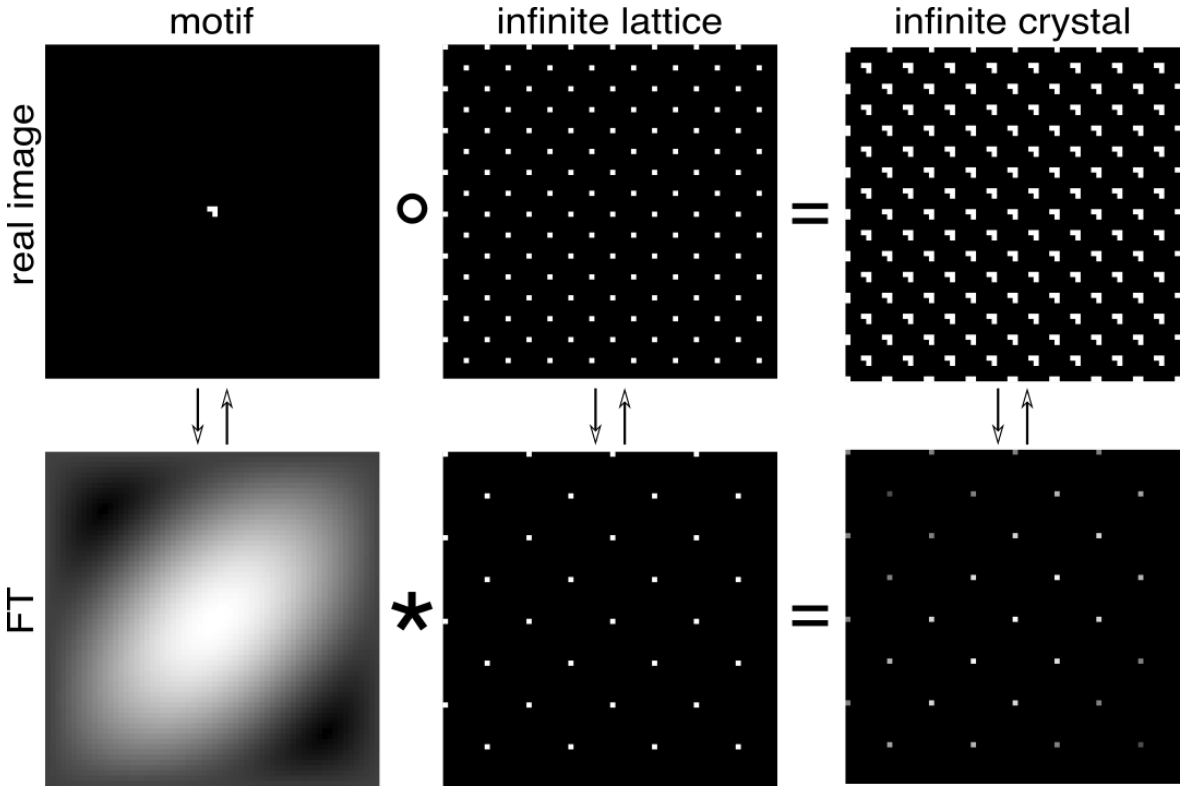

Figure 1.2 | A motif convolved with an infinite lattice produces an infinite crystal. It is equivalent to a multiplication in the reciprocal space. The infinitude of the crystal is conveyed by making the real images exactly periodic over the image boundaries.

Finite Lattice and Finite Crystal

In practice we do not have infinite crystals. Especially, when imaging a large crystal at a magnification where the periodic nature becomes visible, we are limited to capture only a small square-shaped view of the whole field. In the case of imaging nanocrystals the situation becomes even more interesting when the actual irregular shape of the crystal itself enters the equation.

A crystal must therefore be finite and bounded by some shape $S \subset \mathbb{R}^3$. Let us define the indicator function of the shape S (Figure 1.3)

$$\chi_S(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in S\,, \\ 0 & \text{if } \boldsymbol{x} \notin S\,. \end{cases} \tag{1.13}$$

The finite lattice is obtained by multiplying the infinite lattice with the shape as visualized on Figure 1.3.

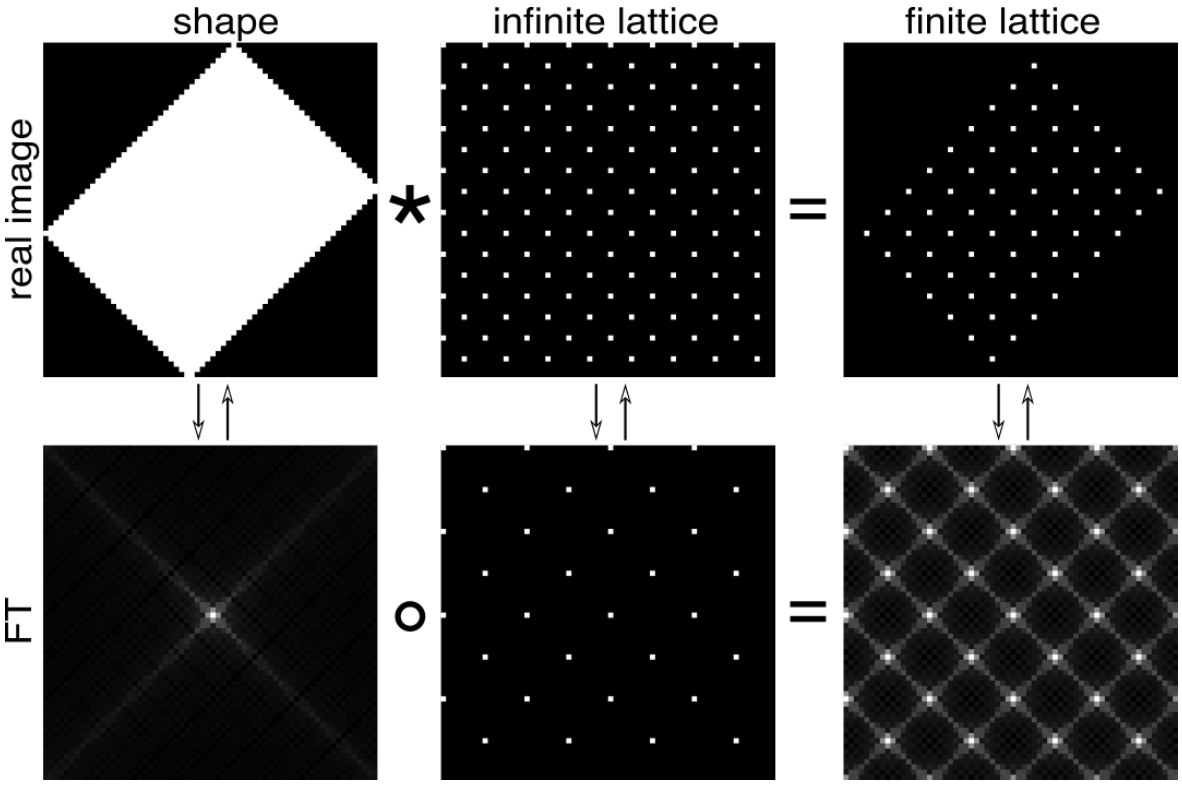

Figure 1.3 | Infinite lattice multiplied by a shape produces a finite lattice. The spots in the Fourier space are no longer crisp but a smeared out version instead. The smearing is defined by the Fourier transform of the shape.

A finite nanocrystal $\rho(\boldsymbol{x})$ of shape S is equal to convolving the motif ρ_0 with the finite lattice, and a finite lattice is equal to an infinite lattice times the shape. Thus,

$$\rho(\boldsymbol{x}) = (\chi_S(\boldsymbol{x})L(\boldsymbol{x})) \circ \rho_0(\boldsymbol{x}) \tag{1.14}$$

and its Fourier transform by the convolution theorem is

$$\hat{\rho}(\boldsymbol{q}) = \left(\hat{\chi}_S(\boldsymbol{q}) \circ \hat{L}(\boldsymbol{q})\right) \hat{\rho}_0(\boldsymbol{q})\,. \tag{1.15}$$

The Fourier transform of the finite lattice does not show crisp diffraction spots anymore but are instead a smeared out version of the infinite reciprocal lattice.

As a conclusion, we see on Figure 1.4 that the effect of a finite crystal is that the diffraction spots on the Fourier transform are getting convolved with the shape function. On top of that, for real images, the edges of the image itself are non-periodic and this adds one more square-shaped shape transform to the mix which manifests itself in the Fourier transform as additional spikes aligned with the coordinate axes. The artifacts produced by the latter shape-function can be alleviated as the shape of the image is known to be a square. By using various tricks the discontinuities at the image boundaries can be made softer which greatly reduces their impact. These techniques are explored in chapter 2.

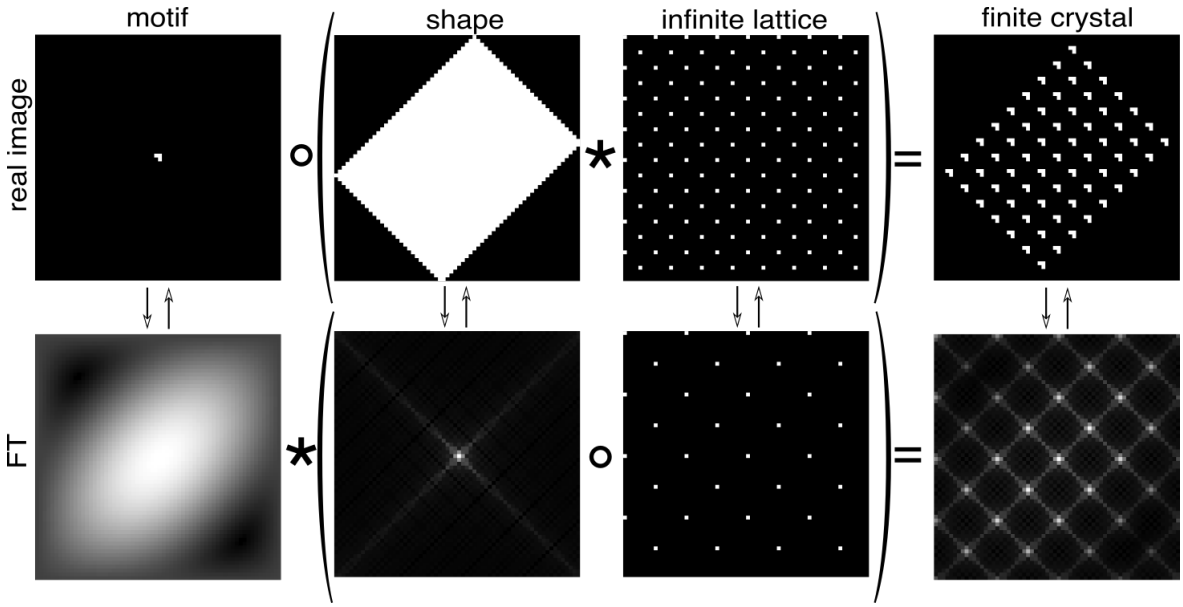

Figure 1.4 | Constructing a finite crystal from the motif, infinite lattice and the shape.

The takeaway message from this section is that the finiteness of the crystal produces artifacts that can interfere with the lattice detection and calculations in the Fourier domain.

1.3 Electron Microscopy and Image Formation

1.3.1 Electron optics

In a transmission electron microscope (TEM), a source emits electrons that are accelerated through a high voltage potential (e.g. 300 kV) and condensed onto the specimen via the condenser lens system. After transmitting through the specimen, the electron wave is focused by the optical system, consisting of a series of electromagnetic lenses, onto the image plane, where the detector records the illumination intensity. As an approximation, it is convenient to consider the electrons leaving the source as a monochromatic electron plane wave that is traveling parallel to the optical axis.

Since the electrostatic potential of he specimen does not change with time, it is sufficient to only consider standing wave solution to the Schrödinger equation governing the electron-specimen interaction [53].

The complex electron wave emerging from the source can be written as

$$\Psi(\boldsymbol{r}) = A(\boldsymbol{r})e^{-i\alpha(\boldsymbol{r})}\,, \tag{1.16}$$

where $A(\boldsymbol{r})$ is the amplitude of the wave at position $\boldsymbol{r}$ and $\alpha(\boldsymbol{r})$ is the phase.

When passing through the specimen, some of the electrons will be scattered. If the object is really thin, the scattered beam will experience a phase shift of $\pi/2$ radians and a massive reduction in amplitude, when compared to the unscattered beam. Since the

electron beam is coherent, the scattered and unscattered waves, Ψ_{us} and Ψ_{sc}, interfere. The total emerging wave is then a sum

$$\begin{aligned}\Psi(\boldsymbol{r}) &= \Psi_{us}(\boldsymbol{r}) + \Psi_{sc}(\boldsymbol{r}) \\ &= (1-\delta)A(\boldsymbol{r})e^{-i\alpha(\boldsymbol{r})} + \delta A(\boldsymbol{r})e^{-i\left(\alpha(\boldsymbol{r})+\frac{\pi}{2}\right)}\,,\end{aligned} \tag{1.17}$$

where $\delta \ll 1$ is the proportion of scattered electrons.

1.3.2 Formation of an Image

On the image plane we record the illumination intensity, which is equal to the wave function multiplied by its complex conjugate.

$$I = \Psi\Psi^*\,. \tag{1.18}$$

In bright field imaging mode of organic specimens, the two main contrast creating mechanisms are

1. **amplitude contrast** - arises from including or excluding certain electrons from the imaging system. Electrons that are scattered at a high angle will be blocked out by the objective aperture which causes amplitude contrast. The amount of contrast depends on the density of the specimen, the acceleration voltage of the microscope and the size of the cut-off aperture. This is the main contrast creating mechanism for heavy atom staining, but plays smaller role in cryo-EM.

2. **phase contrast** - arises from electron-specimen interaction where the electron wave is advanced or delayed with respect to the direct beam. The contrast depends on the interference between the scattered electrons and the direct unscattered beam. This is the main contrast creating mechanism in biological cryo-cooled vitrified specimens. To image the phase objects we take advantage of the aberrations of the microscope that add an extra phase shift to the scattered wave.

In an ideal microscope the exit wave just behind the specimen plane is identical to that on the image plane, therefore what we record on the detector is

$$I(\boldsymbol{r}) = \Psi(\boldsymbol{r})\Psi^*(\boldsymbol{r}) = A^2(\boldsymbol{r})\left((1-\delta)^2+\delta^2\right) \approx A^2(\boldsymbol{r})\,. \tag{1.19}$$

We see that imaging exactly at focus gives no contrast as the phase information is lost and we record only the variations in the amplitude. Imaging at a defocus makes the electron wave go through another frequency dependent phase shift that converts parts of the phase variation into amplitude variation.

1.3.3 The Specimen Transfer Function

To see how the phase contrast is created let us start from the beginning. Let us denote the 3d electrostatic potential of our specimen as

$$\rho(x,y,z) \tag{1.20}$$

and the projected potential in z-direction as

$$\rho_t(x,y) = \int \rho(x,y,z)\,dz\,, \tag{1.21}$$

where t in $\rho_t(x,y)$ stands for thickness.

It can be shown that the specimen caused phase shift $d\alpha$ depends only on the potential that the electron "sees" when it is transmitting through the specimen, i.e. only on the thickness of the specimen $\rho_t(x,y)$ at (x,y) [59].

$$d\alpha = \frac{\pi}{\lambda E}\rho_t(x,y) = \sigma\rho_t(x,y)\,, \tag{1.22}$$

where λ is the relativistic wave length of the electron and E its energy. σ can be thought of as an interaction constant.

The specimen can be modeled as a transfer function that modifies the incoming planar electron wave. Let us take the amplitude of the incoming wave to be equal to 1. The exit wave function right after the specimen is then

$$\Psi(x,y) = e^{-i\sigma\rho_t(x,y)-\mu(x,y)}\,, \tag{1.23}$$

where μ is an absorption factor.

By assumption the specimen is thin, which implies that $\rho_t(x,y) \ll 1$, so we can use the *weak phase object approximation*. By expanding the exponential into Taylor series, neglecting the absorption μ and the higher order terms, we arrive at the following simplified expression

$$\Psi(x,y) = 1 - i\sigma\rho_t(x,y)\,. \tag{1.24}$$

1.3.4 The Contrast Transfer Function

Going through the optical system of the TEM, the wave function will be modified by the *contrast transfer function* (CTF). On the back focal plane of the objective lens, the wave function is equal to the Fourier transform of Ψ times the aberration function of the lens, usually expressed as a complex function $e^{i\chi(\boldsymbol{q})}$

$$\Psi_F(\boldsymbol{q}) = \mathcal{F}[\Psi(\boldsymbol{q})]e^{i\chi(\boldsymbol{q})}\,. \tag{1.25}$$

On the image plane the wave will be an inverse Fourier transform of Ψ_F

$$\Psi_I(\boldsymbol{x}) = \mathcal{F}^{-1}[\Psi_F](\boldsymbol{x}) = \Psi(\boldsymbol{x}) \circ \mathcal{F}^{-1}[e^{i\chi(\boldsymbol{q})}] = \Psi(\boldsymbol{x}) \circ h(\boldsymbol{x})\,, \tag{1.26}$$

that says that on the image plane the wave function is convolved with the point spread function which is an inverse Fourier transform of the CTF, denoted by $h(\boldsymbol{x})$. Since h is a complex function we can decompose it to real and imaginary parts as $h = h_{Re} + ih_{Im}$. Multiplying the wave by its complex conjugate and neglecting the terms with σ^2, the illumination intensity can be worked out to be the following,

$$\begin{aligned} I(\boldsymbol{x}) &= \Psi_I(\boldsymbol{x})\Psi_I^*(\boldsymbol{x}) \\ &= 1 + 2\sigma h_{Im}(\boldsymbol{x})\,. \end{aligned} \tag{1.27}$$

So we see that in the weak phase object approximation only the imaginary part of the CTF function $e^{i\chi(\boldsymbol{q})}$ contributes to the illumination intensity, that is $\sin\chi(q)$ by Euler's equation.

It can be shown that the CTF as a function of spatial frequency $\boldsymbol{q}$ can be written as

$$\sin\chi(\boldsymbol{q}) = \sin\left(\pi\Delta f\lambda\boldsymbol{q}^2 + \frac{\pi}{2}C_s\lambda^3\boldsymbol{q}^4\right), \tag{1.28}$$

where Δf is the defocus, λ relativistic wave length of the electrons, and C_s the spherical aberration constant [60].

1.3.5 Conclusion of the Image Formation

So, to conclude, going through the optical system of the electron microscope, the planar electron wave is first modified by the transfer function of the specimen and then by the transfer function of the objective lens. On top of the phase-modifying aberrations, various other parameters also contribute to the final shape of the CTF, such as chromatic aberration and energy spread to name a few, whose effect can be taken into account as an envelope that dampens the high resolution details.

Spatial frequencies, where the CTF is 0, are completely missing from the recorded image, and the frequencies, where the CTF is negative, are represented in the image with inverted contrast. The fact that the CTF is an oscillating function repeatedly passing through 0, makes the deconvolution process challenging. But without deconvolution we would have trouble interpreting the high resolution details as the contrast would be garbled.

1.4 Tomographic Reconstruction

1.4.1 Ill-posed Problems

Going from a given 3d object to 2d projected images is a *forward problem*, where the formation of the images is modeled by a *forward model* and is easy to compute. Tomographic reconstruction of the 3d object from the 2d projections, on the other hand, is an *inverse problem*, that is challenging to solve and usually *ill-posed*.

Problems are ill-posed (in the *Hadamard* sense) if at least one of the following is true

- The problem does not have a solution at all.
- The problem does not have a unique solution.
- The solution does not depend continuously on data - small perturbations in the data can lead to however large changes in the solution.

In electron tomography the data is commonly incomplete and very noisy, due to the limited electron dose to avoid radiation damage. This leads to an inconsistent problem where the 3d structure is not fully determined - there exists and infinite number of solutions that all fit the observed data. With the help of regularization we can solve a

nearby well-posed problem instead by using some prior information about the specimen and hope that this solution is not far from the truth. Some reasonable assumptions about the specimen that can be used as prior information are that the reconstructed density is positive and relatively smooth.

1.4.2 Data Acquisition

Let us denote our specimen as a 3d function $\rho : \mathbb{R}^3 \to \mathbb{R}$. ρ can be thought of as the electrostatic potential of the biological matter encapsulated in vitreous ice. In the following, for simplicity, we will write the equations for a 2d case. Generalization to three dimensions is straightforward.

The schematic of the data acquisition geometry in tomography is depicted in Figure 1.5 Let us denote the angle of the projection by $\varphi \in [-\pi/2, \pi/2]$. A unit vector in the direction of the projection is then $(-\sin\varphi, \cos\varphi)$ and a unit vector perpendicular to the projection direction is $(\cos\varphi, \sin\varphi)$ For a tilt angle φ a straight line, also called a ray, going through a point $(r\cos\varphi, r\sin\varphi)$, $r \in \mathbb{R}$, can be parametrized by a variable $s \in \mathbb{R}$ as

$$l_{\varphi,r} = \{(r\cos\varphi - s\sin\varphi, r\sin\varphi + s\cos\varphi) : s \in \mathbb{R}\}. \tag{1.29}$$

The parallel ray projection $\mathcal{P}_\varphi(r)$ is then integral over the line

$$\mathcal{P}_\varphi(r) = \int_{-\infty}^{\infty} \rho(r\cos\varphi - s\sin\varphi, r\sin\varphi + s\cos\varphi)ds \tag{1.30}$$

In electron tomography we are generally restricted to a limited range of angles at which the data can be collected. Several different schemes have been devised to collect data from various directions. The most used ones are perhaps

- **single-axis tilt** - the specimen is tilted only around a single axis,
- **dual-axis tilt** - the specimen is tilted first around one axis, then returned to its original position, and then tilted around a second axis,
- **conical tilt** - the specimen is tilted by changing the angles around two axes simultaneously, so that the normal vector to the specimen plane draws a circle on the unit sphere.

Single-axis scheme is possibly the most common one in cryo-ET, so let us from now on focus only on that. In general we cannot cover the whole angular wedge from $-90°$ to $90°$ due to restriction set by the electron microscope hardware. The range of possible tilt angles are generally between $-60°$ and $60°$, sometimes even higher such as $-70°$ and $70°$, but in this work (described in Methods section in chapter 4) the tilt range was from $-67°$ to $67°$.

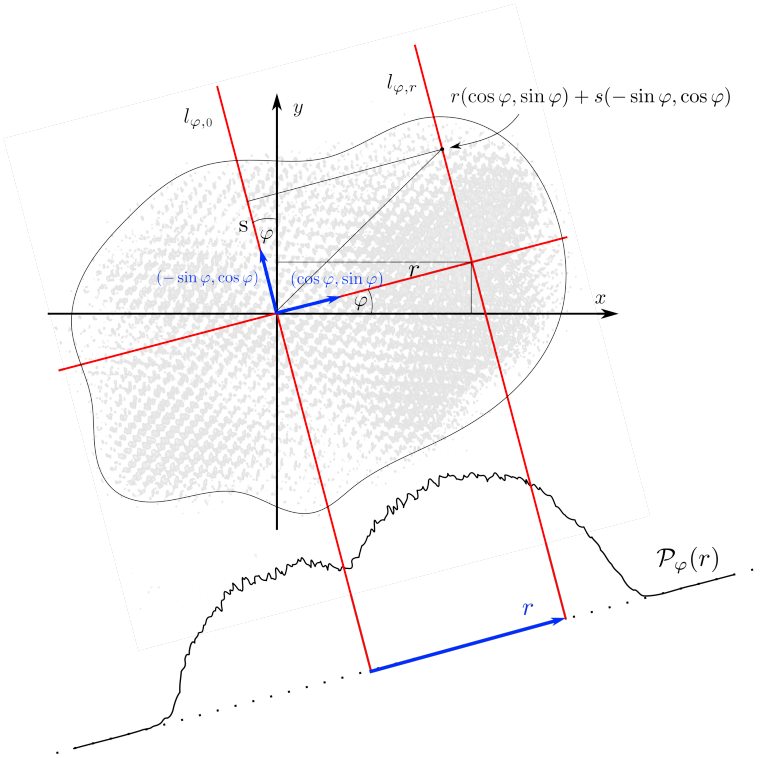

Figure 1.5 | Data acquisition geometry in 2d. An object in 2d is projected via parallel ray transform to 1d at an angle φ along the rays $l_{\varphi,r}(s)$, $r \in \mathbb{R}$.

1.4.3 Generalized Ray Transform

When illuminating a large field of view, using a straight line approximation for the electron trajectories breaks down, and the parallel ray transform must be replaced by a generalized ray transform that integrates over curved electron paths [61]. In [53] it is claimed that the generalized ray transform becomes important when the field of view extends more than 8 µm. In this thesis the TEM imaging was done at a nominal magnification of $37k$ where the whole field of view captured by the detector is about $4096 * 0.226\,\text{nm} = 926.1\,\text{nm}$. The extracts from the tilt series that were actually used for 3d reconstructions were less than $1400 * 0.226\,\text{nm} = 316.5\,\text{nm}$, which is far below 8 µm. In the current case then approximating the rays with straight lines should not pose a significant source of error, at least not before other more serious sources have been eliminated.

1.4.4 Projection-Slice Theorem

The back-bone of the analytical tomographic reconstruction methods is the *projection-slice theorem*, that says that a Fourier transform of a parallel projection of a 3d object is equal to a 2d slice in the Fourier transform of the 3d object, that is perpendicular to the direction of the projection and passing through the origin.

In 2d the projection-slice theorem then states the following

$$\hat{\mathcal{P}}_{\varphi}(r') = \hat{\rho}(r'\cos\varphi, r'\sin\varphi)\,, \tag{1.31}$$

where r' is a spatial frequency.

1.4.5 Reconstruction methods

The reconstruction methods in electron tomography can be largely divided into two classes [62]

- **Algebraic methods** - arise naturally from the discretization of the ray transform integral (1.30), which leads to a linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$, where $\boldsymbol{A}$ is the matrix describing the imaging device, $\boldsymbol{x}$ is the object and $\boldsymbol{y}$ the projection data. The reconstruction is obtained by solving the equation system for $\boldsymbol{x}$.
- **Analytical methods** - based on the projection-slice theorem. The reciprocal space is filled with the Fourier transforms of the projections and the reconstruction is obtained via 3d inverse Fourier transform.

1.4.6 Filtered Back-projection

The projection-slice theorem gives a direct approach to reconstruction - fill the Fourier space with the Fourier transformed projections and retrieve the reconstruction by inverse Fourier transform. In practice there is a problem due to uneven sampling of the Fourier space. Therefore certain adjustments need to be made.

3d reconstruction from single-axis tilt series can be carried out as a series of 2d reconstructions, slice-by-slice, where each slice is orthogonal to the tilt axis. In the following we will derive an expression that relates the projected data to the original specimen ρ.

From the definition of the Fourier transform the following is trivially true

$$\rho(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{\rho}(u,v)e^{-i2\pi(ux+vy)}dudv\,. \tag{1.32}$$

The variables u and v are on a regular Cartesian grid. We would like to perform the integration over polar coordinates and that implies a change of variables, replacing u and v with functions of r and φ. In polar coordinates u and v can be expressed as

$$u = r\cos\varphi \quad \text{and} \quad v = r\sin\varphi\,. \tag{1.33}$$

The differentials du and dv change then accordingly to

$$dudv = |r|drd\varphi, \quad r \in \mathbb{R}\,. \tag{1.34}$$

Substituting the changes into equation 1.32 yields

$$\rho(x,y) = \int_{0}^{\pi}\int_{-\infty}^{\infty} \hat{\rho}(r\cos\varphi, r\sin\varphi)e^{-i2\pi r(x\cos(\varphi)+y\sin(\varphi))}|r|drd\varphi\,. \tag{1.35}$$

According to projection-slice theorem the term $\hat{\rho}(r\cos\varphi, r\sin\varphi)$ is the Fourier transform of the projection at angle φ. So we can write

$$\rho(x,y) = \int_0^{\pi}\int_{-\infty}^{\infty} \hat{\mathcal{P}}_{\varphi}(r)|r| e^{-i2\pi r(x\cos(\varphi)+y\sin(\varphi))} dr d\varphi \,. \tag{1.36}$$

Now it is apparent that the original function $\rho(x,y)$ can be retrieved from the projections when their Fourier transforms are weighted by a factor $|r|$. This r-weighting accounts for the fact that during data acquisition the low resolution components are more finely sampled in the Fourier space than the high resolution components, and thus should be down-weighted. This method is called the filtered back-projection (FBP) or also the weighted back-projection.

1.4.7 Iterative Methods

The working principle of all iterative reconstruction methods in electron tomography is to generate an iterative sequence that converges to the least squares solution of the inverse problem [53]. Different schemes for splitting the large reconstruction problem into smaller sub-problems lead to different versions of iterative reconstruction methods, such as *algebraic reconstruction technique* (ART), which is a special case of a more general *Kaczmarz's method*, *simultaneous iterative reconstruction technique* (SIRT) and *simultaneous algebraic reconstruction technique* (SART).

In iterative methods the number of iterations becomes a regularization parameter as the initial iterates recover the low resolution details [53]. Assuming that noise in the data affects only the high resolution details, early stopping then avoids the reconstruction getting weakened by noise.

1.4.8 COMET

A particular iterative regularization method for 3d tomographic reconstruction, that has been also used in this thesis, is COnstrained Maximum Entropy Tomography (COMET) [50], where entropy is used as a regularizer that penalizes complexity of the solution.

Minimizing least squares needs little justification, whereas the question of why entropy maximization might be a good idea usually needs more motivation. Starting from certain axioms that a reconstruction method should satisfy, such as consistency, distinctness and continuity, it can been shown that the entropy regularization is the only method consistent with the axioms when the signal to be reconstructed is positive and real valued [53].

In COMET the objective is to find a solution that maximizes the entropy relative to a prior, at the same time aiming to improve the least squares fit to the observed data. The entropy functional to be maximized is

$$S(\rho) = -\int \rho(\boldsymbol{x}) \ln \frac{\rho(\boldsymbol{x})}{m(\boldsymbol{x})} d\boldsymbol{x} \,, \quad \boldsymbol{x} \in \mathbb{R}^3 \tag{1.37}$$

where $\rho(\boldsymbol{x})$ is the current density estimation and $m(\boldsymbol{x})$ a prior, generally obtained as

a low-pass filtered FBP reconstruction, that is normalized to integral 1 and describes the prior belief about the electrostatic potential.

The iterations proceed by successively updating the current solution as a linear combination of the changes that the goodness of fit measure and the entropy measure suggest

$$\rho_{new}(\boldsymbol{x}) = \rho_{old}(\boldsymbol{x}) + \alpha u(\boldsymbol{x}) + \beta v(\boldsymbol{x})\,, \tag{1.38}$$

where α and β are coefficients determining the step size in the directions given by the goodness of fit gradient u and the entropy gradient v. The result is theoretically an unbiased map that only includes features that are backed up by the data.

1.4.9 Missing Wedge

The limited angular coverage of the projections makes a gap in our knowledge about certain spatial frequencies. After filling in the 3d Fourier space with the Fourier transforms of the projections a characteristic wedge-shaped gap forms that represents the lack of frequency information mainly in the z-direction. The implication of this lack of information is that the reconstructions have anisotropic resolution where the resolution in z-direction can be much lower than in x and y.

In the general case the missing wedge cannot be filled because the frequency information is simply not present in the data. But when certain things are assumed about the specimen or when many identical but differently oriented particles are composed together, then the effects of the missing wedge can be alleviated or practically removed. For instance, when the object is assumed to consist of simple shapes, such as an atomic resolution image where the atoms are approximated by spheres, then this prior information can fill the missing wedge. Certain graph-based methods have demonstrated wedge-filling capabilities by locating similar patches in the image. Another example is sub-tomogram averaging and single particle reconstruction where the missing wedge is filled in by aligning and averaging many particles that are lying in random orientations covering all angles. In the current work, the usage of symmetries partly evens out the resolution in different directions (see section 4.4.5).

Chapter 2

2d FFT with Edge Artifact Removal

This chapter describes the work done by M. Toots in collaboration with F. Mahmood. Though the core of the project is outside the scope of this thesis, the application and the algorithmic development done by M. Toots is relevant to electron tomography. For more details the interested reader is kindly referred to [63].

2.1 Introduction

Fast Fourier transform (FFT) is one of the most widely used techniques in image processing, allowing for transforming the image from a physical space representation to a spatial frequency domain. As there is no inherent reason why images in general should be periodic, the sharp discontinuities at the opposing edges of an image cause intense cross-shaped streaking artifacts in the transform. This can become problematic and lead to errors when the transforms are used for further analysis or processing.

Several methods exist that try to alleviate the edge artifacts, such as mirroring the original image in each direction or tapering off the edges with a suitable window function. The former leads to a 4-fold (2d case) or an 8-fold (3d case) increase in the image size, which could become prohibitively large, and the latter involves altering the pixels and removing image information.

A more analytical approach, called *Periodic+Smooth (P+S) decomposition*, proposed in [64], attempts to decompose the original image into two - a *periodic component*, that captures the high-frequency details and is continuous over the edges, and a slowly varying *smooth component*, that recreates the background level and the edge discontinuities.

Though highly optimized, the computational complexity $\mathcal{O}(n^2 \log n)$ of the 2d FFT can still become a computational burden in applications that require close to real-time operations. The tendency today towards higher resolution images has only increased the amount of data transport. Adding the need for simultaneous handling of the edge artifacts leads to a situation where the computational requirements and memory bandwidth will become a bottle neck for conventional serial processing and that might not be sufficient for high throughput applications, such as machine vision and control.

A suitable target for tackling these challenges is reconfigurable Field Programmable Gate Array (FPGA) based technology. FPGAs are composed of parallel programmable

logic blocks connected by programmable routing switches (interconnects) that can implement arbitrary logic functions. FPGAs are inherently parallel, retaining the flexibility of software while gaining the speed of hardware. Compared to software, the downside of designing an algorithm for FPGA is that it is difficult, time consuming and hard to debug, but in time-critical applications the upsides far outweigh the downsides.

This chapter gives an overview of a part of the work that is fully elaborated in [63], as most of it is outside the scope of the thesis. In the paper we present a novel FPGA-based solution for 2d FFT calculation with simultaneous P+S edge-artifact removal for usage in high-performance applications. Since for the algorithm two 2d FFTs have to be calculated simultaneously, the external memory addressing can become a bottle neck.

In the following section we describe the P+S edge-artifact removal algorithm itself and a specific optimization that enables for the hardware architecture with reduced external memory access. The reduction is achieved by decreasing the number of necessary 1d FFT invocations. We also design a memory mapping scheme that can reduce row activation overhead while accessing columns of data from the external memory. This second optimization can be read about in [63]. The results presented here build on our previous work published in [65].

2.2 Credits and Contributions

Märt Toots developed the optimization of the P+S edge artifact removal algorithm that facilitates reduced external memory access for 1d FFTs. Faisal Mahmood developed and implemented an FPGA-based design for the optimized P+S algorithm and for the tile-hopping memory optimization for efficient data access from the external memory.

2.3 Periodic + Smooth decomposition

In this section we will describe the theory behind P+S decomposition edge artifact removal algorithm and demonstrate its effect on a non-periodic electron micrograph. A more theoretical mathematical treatment including proofs can be found in the original publication [64].

Let us have a discrete n by m gray-scale image $\boldsymbol{I}$ on a finite domain

$$\Omega = \{0, 1, \ldots, n-1\} \times \{0, 1, \ldots, m-1\}\,. \tag{2.1}$$

The discrete Fourier transform of $\boldsymbol{I}$ is defined as

$$\hat{\boldsymbol{I}}(s,t) = \mathcal{F}(\boldsymbol{I})(s,t) = \sum_{(i,j)\in\Omega} \boldsymbol{I}(i,j) \exp\left(-\imath 2\pi \left(\frac{si}{n} + \frac{tj}{m}\right)\right) \tag{2.2}$$

P+S decomposition attempts to separate the original image $\boldsymbol{I}$ into two components - first, a periodic component $\boldsymbol{P}$, that captures the high-resolution details and is continuous over the edges, and second, a slowly varying smooth component $\boldsymbol{S}$, that recreates

the background level and the edge discontinuities, so that

$$\boldsymbol{I} = \boldsymbol{P} + \boldsymbol{S}\,. \tag{2.3}$$

The smooth component can be found by solving a Poisson equation

$$\nabla^2 \boldsymbol{S} = \boldsymbol{B}\,. \tag{2.4}$$

This equation has a slowly varying unique solution $\boldsymbol{S}$ that can be found using Fourier transform (proof in [64]). Here $\boldsymbol{B}$ represents the boundary discontinuities that occur when crossing the edges of the image and is calculated as shown below. Let $\boldsymbol{R}$ represent the boundary differences when transitioning the edges of $\boldsymbol{I}$ row-wise and $\boldsymbol{C}$ represent the column-wise differences.

$$\begin{aligned} \boldsymbol{R}(i,j) &= \begin{cases} \boldsymbol{I}(n-1-i,j) - \boldsymbol{I}(i,j), & i = 0 \text{ or } i = n-1 \\ \boldsymbol{0}, & \text{otherwise} \end{cases} \\ \boldsymbol{C}(i,j) &= \begin{cases} \boldsymbol{I}(i,m-1-j) - \boldsymbol{I}(i,j), & j = 0 \text{ or } j = m-1 \\ \boldsymbol{0}, & \text{otherwise} \end{cases} \end{aligned} \tag{2.5}$$

The boundary image $\boldsymbol{B}$ is then defined as

$$\boldsymbol{B} = \boldsymbol{R} + \boldsymbol{C}\,. \tag{2.6}$$

It turns out that the Fourier transform of the solution to the Eq. 2.4 can be calculated by the following formula

$$\hat{\boldsymbol{S}}(s,t) = \begin{cases} \dfrac{\hat{\boldsymbol{B}}(s,t)}{2\cos\left(\dfrac{2\pi s}{n}\right) + 2\cos\left(\dfrac{2\pi t}{m}\right) - 4}, & \forall (s,t) \in \Omega \backslash \{(0,0)\} \\ 0 & (s,t) = (0,0)\,. \end{cases} \tag{2.7}$$

Setting the 0th order term of $\hat{\boldsymbol{S}}$ equal to 0 ensures that the smooth component sums to 0. The Fourier transform of the periodic component is therefore

$$\hat{\boldsymbol{P}} = \hat{\boldsymbol{I}} - \hat{\boldsymbol{S}}\,. \tag{2.8}$$

Figure 2.1 gives a visual interpretation to P+S decomposition. A crop of an image from a tilt-series of a lysozyme nanocrystal is shown on Figure 2.1a. The corresponding periodic component $\boldsymbol{P}$ can be seen on Figure 2.1b and the smooth component $\boldsymbol{S}$ on Figure 2.1c. The amplitudes of the FFTs of the triad are presented on Figure 2.1d-f. It is notable that the periodic component has no visual differences from the original image other than the background level and barely noticeable attenuation of the edge pixels to make the edges periodic. All of the high resolution information is intact and the Fourier transform shows diffraction spots without the obstructing edge artifacts.

On Figure 2.2 the original image $\boldsymbol{I}$ and the periodic component $\boldsymbol{P}$ have been replicated twice in two directions to further illustrate the effect of making the image periodic. While the original has sharp discontinuities at the edges, the periodic component transitions smoothly.

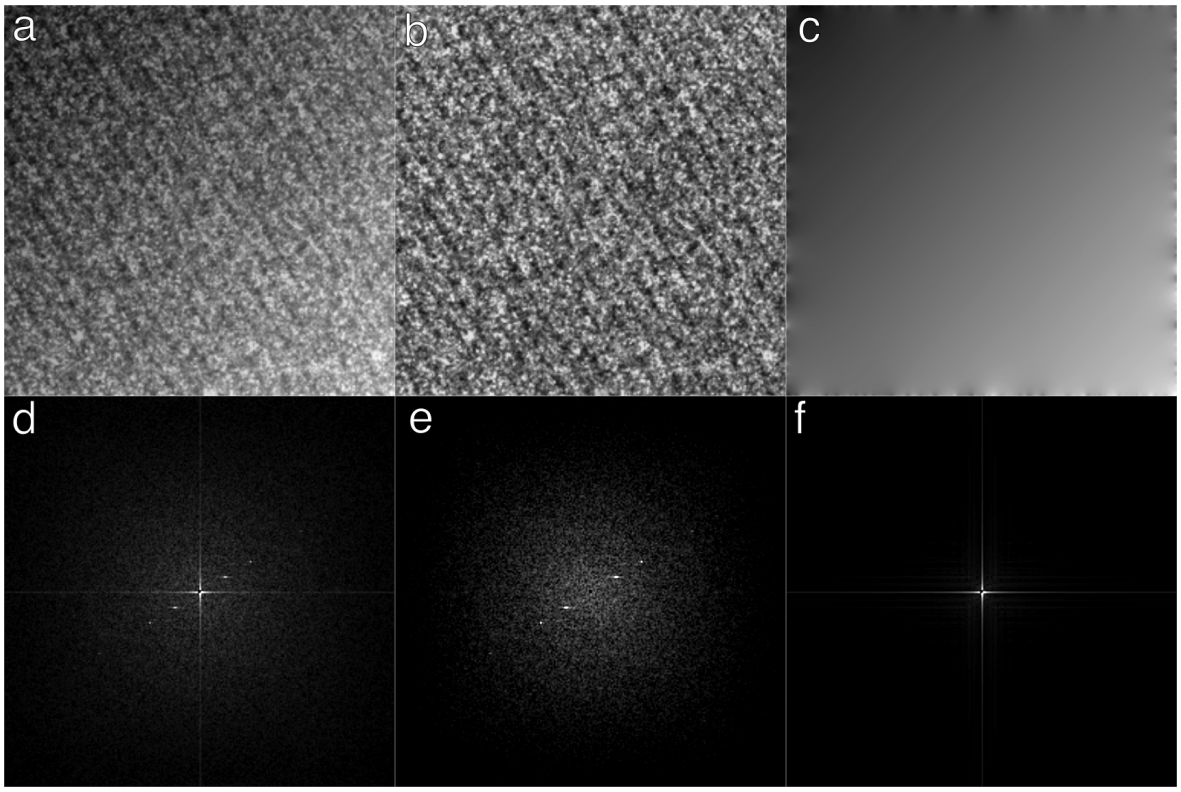

Figure 2.1 | FFT and P+S FFT of a non-periodic electron microscopy image. **a** - Electron microscopy image of a lysozyme nanocrystal, with a linear background level gradient (artificially induced). **b** - Smooth component of **a** obtained by P+S decomposition. **c** - Periodic component of **a** obtained by P+S decomposition. **d-f** - FFT amplitudes of **a-c**, respectively. The Fourier transforms have 0-order coefficient blanked and the intensity is on the square root scale to better bring out the important details in the visualization.

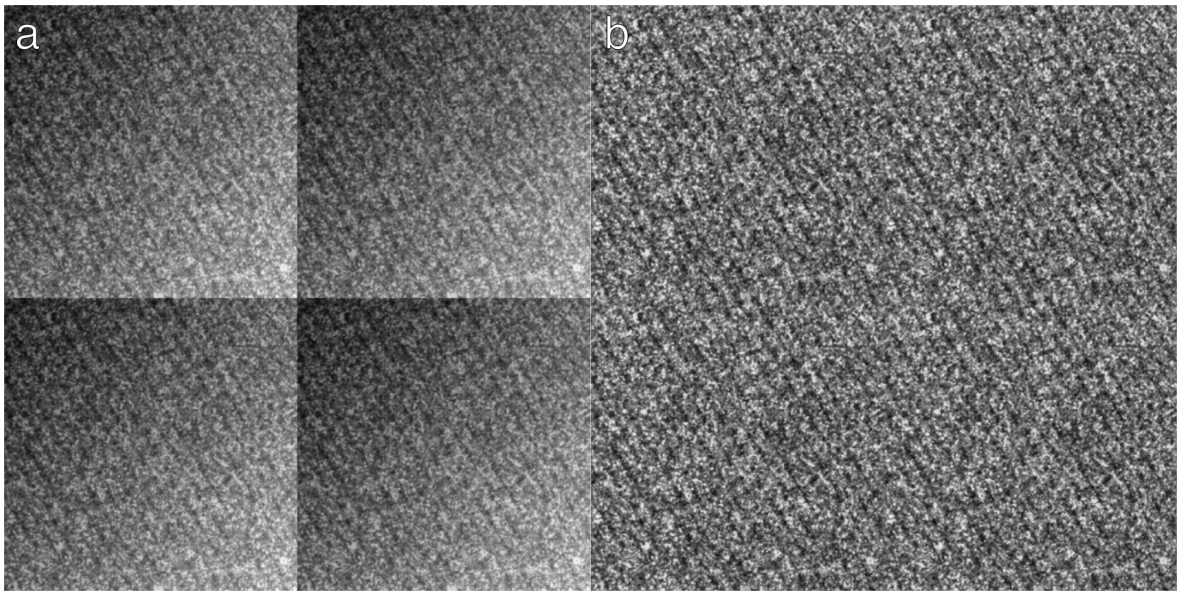

Figure 2.2 | Original and periodic version of the same image replicated to illustrate the periodicity of the periodic component. **a** - Original image from Figure 2.1a. **b** - Periodic component from Figure 2.1c.

2.4 Optimized Periodic + Smooth Decomposition

In this section we develop an optimization to computing the 2d FFT of the smooth component (Eq. 2.7). The result of the optimization is a reduced number of 1d FFT invocations necessary for calculating the 2d FFT of $\boldsymbol{B}$ by taking advantage of its simple structure.

Due to its decomposability a 2d FFT can be computed as a series of 1d FFTs, transforming first every row of an image followed by transforming every column. 2d FFT of $\boldsymbol{I}$ is equivalent to a matrix multiplication $\boldsymbol{W_n I W_m}$, where

$$\boldsymbol{W_n} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{2(n-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & w^{n-2} & w^{2(n-2)} & \dots & w^{(n-2)(n-1)} \\ 1 & w^{n-1} & w^{2(n-1)} & \dots & w^{(n-1)(n-1)} \end{pmatrix} \tag{2.9}$$

and

$$w^k = \exp\left(-\iota\frac{2\pi}{n}\right)^k = \exp\left(-\iota\frac{2\pi k}{n}\right). \tag{2.10}$$

Since w^k has period n, meaning that $w^k = w^{k+ln}$, $\forall k, l \in \mathbb{N}$, the expression for W_n can

be simplified,

$$\boldsymbol{W_n} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & w & w^2 & \dots & w^{n-2} & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{n-4} & w^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & w^{n-2} & w^{n-4} & \dots & w^4 & w^2 \\ 1 & w^{n-1} & w^{n-2} & \dots & w^2 & w \end{pmatrix} . \tag{2.11}$$

The structure of the border image $\boldsymbol{B}$ is simple with nonzero values only in the edges

$$B = C + R = \begin{pmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1,m-1} & b_{1m} \\ b_{21} & 0 & 0 & \dots & 0 & -b_{21} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ b_{n-1,1} & 0 & 0 & \dots & 0 & -b_{n-1,1} \\ b_{n1} & -b_{12} & -b_{13} & \dots & -b_{1,m-1} & -b_{nm} \end{pmatrix} . \tag{2.12}$$

In total, $\boldsymbol{B}$ has $n + m - 1$ unique elements, with the following relations between the corners

$$\begin{aligned} b_{11} &= r_{11} + c_{11} \,, \\ b_{1m} &= r_{1m} - c_{11} \,, \\ b_{n1} &= -r_{11} + c_{n1} \,, \\ b_{nm} &= -r_{1m} - c_{n1} = -b_{11} - b_{1m} - b_{n1} \,. \end{aligned} \tag{2.13}$$

In computing the 2d FFT of $\boldsymbol{B}$ one normally proceeds by first running 1d FFTs column-by-column and then 1d FFT's row-by-row (or vice versa). A 1d FFT of a column vector v with length n is $\boldsymbol{W}v$, where $\boldsymbol{W}$ is given in eq. (2.11). The column-wise FFT of the matrix $\boldsymbol{B}$ is then

$$\hat{\boldsymbol{B}} = \boldsymbol{W}\boldsymbol{B}. \tag{2.14}$$

It turns out that the structure of $\hat{\boldsymbol{B}}$ is also relatively simple.

Let us have a closer look on the first column, denoted by $\boldsymbol{B_{\cdot 1}}$. The 1d FFT of this vector is

$$\begin{aligned} \hat{\boldsymbol{B}}_{\cdot \mathbf{1}} = \boldsymbol{W}\boldsymbol{B}_{\cdot \mathbf{1}} &= \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{2(n-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & w^{n-2} & w^{2(n-2)} & \dots & w^{(n-2)(n-1)} \\ 1 & w^{n-1} & w^{2(n-1)} & \dots & w^{(n-1)(n-1)} \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \\ \dots \\ b_{n-1,1} \\ b_{n1} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n b_{i1} \\ \sum_{i=1}^n b_{i1} w^{i-1} \\ \sum_{i=1}^n b_{i1} w^{2(i-1)} \\ \dots \\ \sum_{i=1}^n b_{i1} w^{(n-2)(i-1)} \\ \sum_{i=1}^n b_{i1} w^{(n-1)(i-1)} \end{pmatrix} \end{aligned} \tag{2.15}$$

The 1d FFT of column $j \in \{2, 3, \ldots, n-1\}$ is

$$\begin{aligned}
\hat{\boldsymbol{B}}_{\cdot \boldsymbol{j}} = \boldsymbol{W}\boldsymbol{B}_{\cdot \boldsymbol{j}} &= \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & w & w^2 & \ldots & w^{n-1} \\ 1 & w^2 & w^4 & \ldots & w^{2(n-1)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & w^{n-2} & w^{2(n-2)} & \ldots & w^{(n-2)(n-1)} \\ 1 & w^{n-1} & w^{2(n-1)} & \ldots & w^{(n-1)(n-1)} \end{pmatrix} \begin{pmatrix} b_{1j} \\ 0 \\ 0 \\ \ldots \\ 0 \\ -b_{1j} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ b_{1j} - b_{1j}w^{n-1} \\ b_{1j} - b_{1j}w^{2(n-1)} \\ \ldots \\ b_{1j} - b_{1j}w^{(n-2)(n-1)} \\ b_{1j} - b_{1j}w^{(n-1)(n-1)} \end{pmatrix} = \begin{pmatrix} 0 \\ b_{1j}(1 - w^{n-1}) \\ b_{1j}(1 - w^{2(n-1)}) \\ \ldots \\ b_{1j}(1 - w^{(n-2)(n-1)}) \\ b_{1j}(1 - w^{(n-1)(n-1)}) \end{pmatrix} \\
&= b_{1j} \begin{pmatrix} 0 \\ 1 - w^{n-1} \\ 1 - w^{2(n-1)} \\ \ldots \\ 1 - w^{(n-2)(n-1)} \\ 1 - w^{(n-1)(n-1)} \end{pmatrix} = b_{1j} \begin{pmatrix} 0 \\ 1 - w^{n-1} \\ 1 - w^{n-2} \\ \ldots \\ 1 - w^2 \\ 1 - w \end{pmatrix} = b_{1j}\boldsymbol{\nu}\,,
\end{aligned} \tag{2.16}$$

where

$$\boldsymbol{\nu} = \begin{pmatrix} 0 \\ 1 - w^{n-1} \\ 1 - w^{n-2} \\ \ldots \\ 1 - w^2 \\ 1 - w \end{pmatrix} \tag{2.17}$$

The 1D FFT of the last column $\boldsymbol{B}_{\cdot\boldsymbol{m}}$ is

$$
\begin{aligned}
\hat{\boldsymbol{B}}_{\cdot\boldsymbol{m}} = \boldsymbol{W}\boldsymbol{B}_{\cdot\boldsymbol{m}} &= \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{2(n-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & w^{n-2} & w^{2(n-2)} & \dots & w^{(n-2)(n-1)} \\ 1 & w^{n-1} & w^{2(n-1)} & \dots & w^{(n-1)(n-1)} \end{pmatrix} \begin{pmatrix} b_{1m} \\ -b_{21} \\ -b_{31} \\ \dots \\ -b_{n-1,1} \\ -b_{11} - b_{n1} - b_{1m} \end{pmatrix} \\
&= \begin{pmatrix} b_{1m} - \sum_{i=1}^{n-1} b_{i1} - b_{11} - b_{n1} - b_{1m} \\ b_{1m} - \sum_{i=1}^{n-1} b_{i1} w^{i-1} - (b_{11} - b_{n1} - b_{1m}) w^{n-1} \\ b_{1m} - \sum_{i=1}^{n-1} b_{i1} w^{2(i-1)} - (b_{11} - b_{n1} - b_{1m}) w^{2(n-1)} \\ \dots \\ b_{1m} - \sum_{i=1}^{n-1} b_{i1} w^{(n-2)(i-1)} - (b_{11} - b_{n1} - b_{1m}) w^{(n-2)(n-1)} \\ b_{1m} - \sum_{i=1}^{n-1} b_{i1} w^{(n-1)(i-1)} - (b_{11} - b_{n1} - b_{1m}) w^{(n-1)(n-1)} \end{pmatrix} \\
&= \begin{pmatrix} -\sum_{i=1}^{n} b_{i1} \\ -\sum_{i=1}^{n} b_{i1} w^{i-1} + b_{11} - b_{11} w^{n-1} + b_{1m} - b_{1m} w^{n-1} \\ -\sum_{i=1}^{n} b_{i1} w^{2(i-1)} + b_{11} - b_{11} w^{2(n-1)} + b_{1m} - b_{1m} w^{2(n-1)} \\ \dots \\ -\sum_{i=1}^{n} b_{i1} w^{(n-2)(i-1)} + b_{11} - b_{11} w^{(n-2)(n-1)} + b_{1m} - b_{1m} w^{(n-2)(n-1)} \\ -\sum_{i=1}^{n} b_{i1} w^{(n-1)(i-1)} + b_{11} - b_{11} w^{(n-1)(n-1)} + b_{1m} - b_{1m} w^{(n-1)(n-1)} \end{pmatrix} \\
&= \begin{pmatrix} -\sum_{i=1}^{n} b_{i1} \\ -\sum_{i=1}^{n} b_{i1} w^{i-1} + (b_{11} + b_{1m})(1 - w^{n-1}) \\ -\sum_{i=1}^{n} b_{i1} w^{2(i-1)} + (b_{11} + b_{1m})(1 - w^{2(n-1)}) \\ \dots \\ -\sum_{i=1}^{n} b_{i1} w^{(n-2)(i-1)} + (b_{11} + b_{1m})(1 - w^{(n-2)(n-1)}) \\ -\sum_{i=1}^{n} b_{i1} w^{(n-1)(i-1)} + (b_{11} + b_{1m})(1 - w^{(n-1)(n-1)}) \end{pmatrix} \\
&= -\hat{\boldsymbol{B}}_{\cdot\boldsymbol{1}} + (b_{11} + b_{1m})\boldsymbol{\nu}\,.
\end{aligned}
\tag{2.18}
$$

So, the column-wise FFT of the matrix $\boldsymbol{B}$ is

$$
\hat{\boldsymbol{B}} = \begin{pmatrix} \hat{\boldsymbol{B}}_{\cdot\boldsymbol{1}} & b_{12}\boldsymbol{\nu} & b_{13}\boldsymbol{\nu} & \dots & b_{1,n-1}\boldsymbol{\nu} & -\hat{\boldsymbol{B}}_{\cdot\boldsymbol{1}} + (b_{11} + b_{1m})\boldsymbol{\nu} \end{pmatrix}. \tag{2.19}
$$

To conclude, to calculate column-by-column 1d FFTs of the matrix $\boldsymbol{B}$ we have to only compute the FFT of the first column and then use the appropriately scaled vector $\boldsymbol{\nu}$ to get the rest of the FFTs of the rest of the columns. To complete the whole 2d FFT of $\boldsymbol{B}$ the row-by-row 1d FFTs are then computed in a normal way. The implication of this optimization is that it can reduce significantly the number of 1d FFTs, thereby reducing the amount of external input/output.

2.5 Conclusion

Most implementations of 2d FFT rely on the row and column decomposition (RCD) of the algorithm, where the 2d FFT operation is performed by repeated invocations of row-by-row 1d FFTs followed by column-by-column invocations. We developed an

FPGA-based design for 2d FFT with simultaneous edge artifact removal using P+S decomposition, and optimized it to reduce the number of needed 1d FFTs, thereby reducing external memory access by 24%.

The optimization leads to a reduction in resources, computational time and external memory access. Before the optimization, the bottle-neck of the system was in the edge artifact removal part, whereas after the optimization, the bottle-neck has moved back to the calculation of the FFT itself. P+S could be further optimized, which could probably give a further reduction in the resources, but as the bottle-neck is still the FFT, it would not make the whole pipeline any faster.

The standard way of storing an image in DRAM is in a row-major order, so the row-wise access is fast. Reading a single column, however, is inefficient, since it needs one element from each row requiring each row to be first read into the *row buffer*. This is the major bottle neck for high-throughput 2d FFTs. To address this inefficiency we developed a second optimization that facilitates an efficient external memory access during column-wise data reads.

The essence of the proposed *tile-hopping* address mapping optimization lies in the way the results of row-wise 1d FFTs are written to DRAM. Instead of writing the results of a row-by-row 1d FFT in row-major order we remap the results in a tiled pattern, so that when later a column has to be retrieved, several elements of that column can be accessed by a single DRAM row access. Since the details of this design are outside the scope of this thesis, the interested reader is kindly referred to [63] (submitted) for further details. The memory mapping scheme is completely general, can be used for various applications and greatly reduces the run-time of 2d FFT even together with the optimized P+S edge artifact removal.

For test application we used the created 2d FFT FPGA module as an accelerator for calculating a 3d filtered back-projection (FBP) based on the projection-slice theorem. All of the 2d FFT calculations were offloaded on the FPGA, the rest was computed on a host PC. The results of reconstructing a 3d Shepp-Logan phantom of various size from 2d projections are shown on Table 2.1.

We can see that for small volumes the improvements are marginal as the price paid for losing time on data transport from the host PC to FPGA is not worth the computational speed gained from FPGA. But as the size of the problem grows, having 2d FFT implemented on FPGA becomes more and more attractive. Ultimately, having also the other components implemented on FPGA would give significant increases in speed.

3d Density	CPU (i7)	FPGA + Host PC (i7)
	seconds	seconds
$128 \times 128 \times 128$	21.3	19.5 (92%)
$256 \times 256 \times 256$	47.5	42.4 (89%)
$512 \times 512 \times 512$	94.8	81.3 (86%)
$1024 \times 1024 \times 1024$	322.3	275.3 (85%)
$2048 \times 2048 \times 2048$	1687.7	1364.4 (81%)
$4096 \times 4096 \times 4096$	16463.1	12599.4 (77%)

Table 2.1 | Comparison of filtered back-projection run time. Credits: Faisal Mahmood

Chapter 3

Extended Field Iterative Reconstruction Technique

3.1 Introduction

This chapter presents a general procedure called *extended field* (EF) reconstruction that can be used to enhance the reconstruction quality of many of the iterative reconstruction algorithms.

The basic idea behind extended field is that by reconstructing a volume that is larger than the initial region of interest (ROI), there is a chance that the measurement noise, which is not geometrically constrained, will be spread out over a larger volume, whereas the geometrically constrained signal will remain in the ROI.

The approach was originally explored in the 70s by Crowther and Klug [66] and was shown to improve the reconstruction of a 2d object from 1d projections. They used the extended field with ART and showed empirically how extending the reconstruction space can lead to faster convergence and lower noise in the ROI as some of the noise was observed to have crept out to the extended region.

We took this idea further and carried out a comprehensive simulation study to investigate the effects of extending the reconstruction field in 2d tomographic setting. We experimented with various levels on noise in the data, various types of phantoms and with many different iterative reconstruction methods. Simulations, using reconstruction methods with regularization capability, such as algebraic reconstruction technique (ART), simultaneous iterative reconstruction technique (SIRT), Tikhonov regularization and maximum entropy iterative reconstruction technique (COMET), show the effectiveness of this approach and that it significantly improves the accuracy of the reconstruction.

We also provide a heuristic model to the mechanics of EF reconstruction and derive some theoretical results for a minimalistic case when reconstructing a single pixel using Tikhonov regularization. Developing a rigorous theoretical treatment of the working principles of EF is left for the future, but simulations on 2d phantoms as well as on 3d electron tomography data clearly demonstrate its effectiveness in decreasing the noise in the reconstructions.

The core of the method has been published in [67] and a simulation study investigat-

ing the effectiveness of EF in conjunction with many different iterative regularization schemes has been submitted for publishing. In this thesis, the author will present only the main concepts and results of the simulation study, as this was work mostly done by Faisal Mahmood, and provide an analytical treatment for the simplest possible reconstruction scenario - reconstructing a single pixel from two noisy projections (Figure 3.3). For a whole story the interested reader is kindly referred to the articles.

3.2 Credits and Contributions

Märt Toots developed scripts for automated EF-reconstruction for electron tomography, wrote the section about single pixel reconstruction, and participated in discussions. Faisal Mahmood did the simulations with 2d phantoms.

3.3 Background

In a practical tomographic setting both the projections and the image area/volume to be reconstructed have to be discretized to pixels/voxels. By discretizing the parallel ray projection the imaging can be mathematically described as the following linear equation system

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\,, \tag{3.1}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the matrix representing the imaging device, $\boldsymbol{x} \in \mathbb{R}^n$ is the vectorized form of an image to be reconstructed, and $\boldsymbol{b} \in \mathbb{R}^m$ is the vectorized form of the collected data, i.e. the projections.

Every row of $\boldsymbol{A}$ corresponds to a single ray going through the density being imaged. The values in $\boldsymbol{A}$ could be interpreted as the attenuation coefficients of the density at the given pixels of the object. Since each ray only interacts with a small number of pixels, the matrix $\boldsymbol{A}$ is usually highly sparse.

The system cannot generally be solved directly via matrix inversion, as $\boldsymbol{A}$ is singular or poorly conditioned. Instead some iterative scheme is employed that starts from an initial guess and updates the solution on every iteration until some stopping criteria is met.

In an imperfect world, however, a measurement is always accompanied by noise from various sources, and some of it arises independently from the geometric constraints associated with the signal. The measurement vector $\boldsymbol{b}$ then becomes equal to $\boldsymbol{b}^* + \boldsymbol{\varepsilon}$, where $\boldsymbol{b}^*$ is the true value, which we do not know, and $\boldsymbol{\varepsilon}$ stands for the noise. Having noise in the system makes the linear equations *inconsistent*, so we set to solve it in the least square sense instead.

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} \, ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2 \tag{3.2}$$

Ill-posed inverse problems are generally unstable - small perturbations in the data lead to large perturbations in the solution [68]. For that reason regularization methods are preferred, that inject some prior knowledge into the system in form of a *regularizer*.

One widely known regularization method is *Tikhonov* regularization, where the error function to be minimized is

$$||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2 + \lambda||\boldsymbol{x}||_2^2 \,. \tag{3.3}$$

The prior knowledge in this case is that the solution must have a "reasonable" L_2 norm. The term $||\boldsymbol{x}||_2^2$ keeps the solution from diverging and suppresses large coefficients in $\boldsymbol{x}$. The parameter $\lambda \in \mathbb{R}^+$ controls the balance between the fit to the projections and the smoothness of the solution. For $\lambda = 0$ we are back at the original non-regularized problem where we only try to maximize the fit to the collected data and therefore also to the noise. As λ gets larger the more the goodness of fit term is down-weighted and the more emphasis is put on having a regular solution with small values in $\boldsymbol{x}$. Tikhonov regularization generally leads to smoother solutions and responds less to perturbations in the data.

3.4 Results

3.4.1 Extended Field

In the extended field paradigm we increase the dimensionality of the linear equation system $\boldsymbol{A}$, add new extended variables to $\boldsymbol{x}$ and pad the projections with 0-s, causing multiple 0-s to appear then at the respective positions in $\boldsymbol{b}$. The scheme of extending the field (for the simulation purposes) is illustrated in Figure 3.1.

The padding step might seem trivial at first, but the effect is that it causes the effective reconstructed area to extend, and the 0-s become constraints for the extended region to sum up to 0, so they would not diverge too far. Simulations showed that the profile of the added noise in the virtual projection creation process (Figure 3.1 Step 2) matched well with the profile of the noise that crept out to the extended region in respective directions (in the submitted article Figure 10), thereby assuring the sensibility of the EF reconstruction.

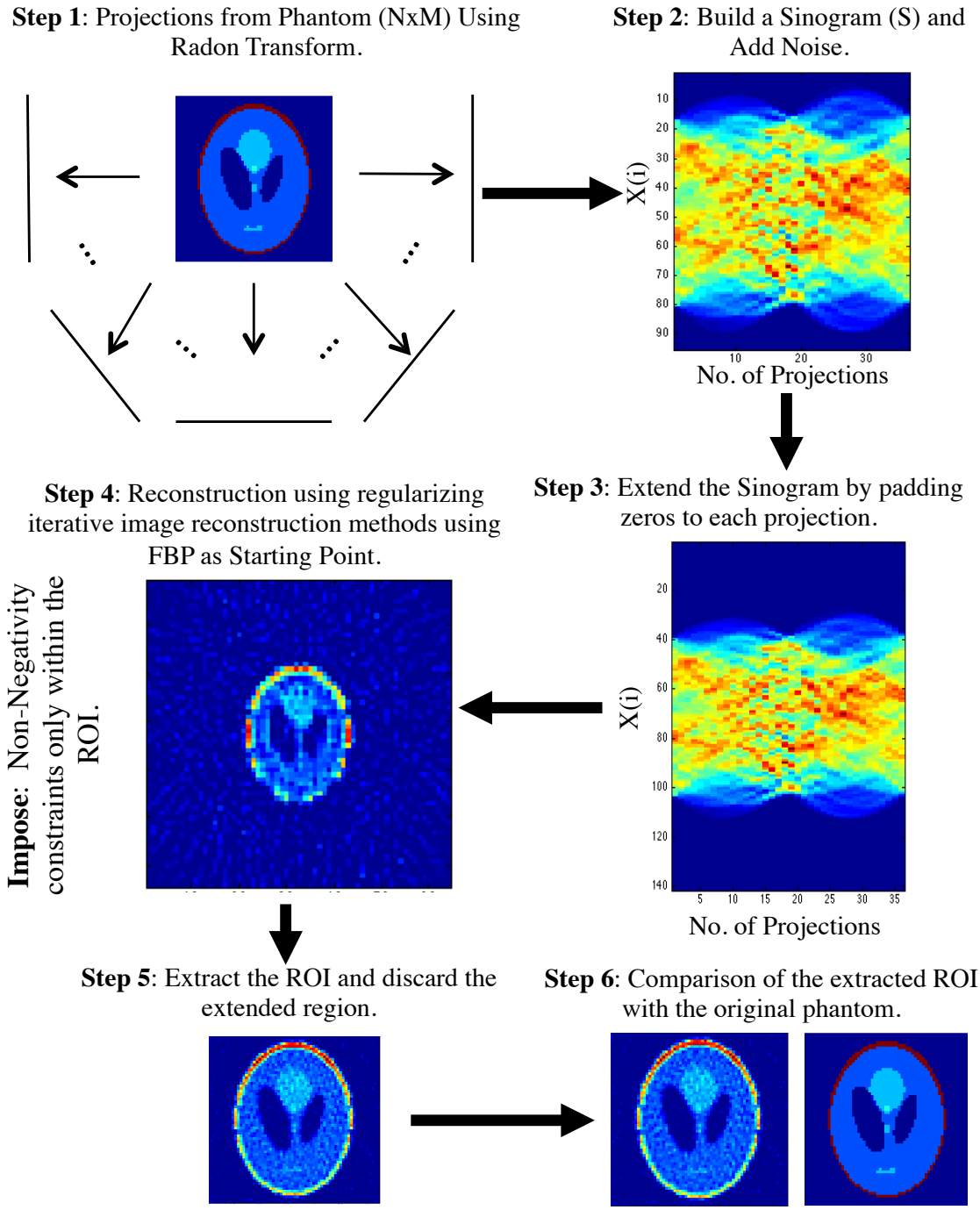

Figure 3.1 | The flow diagram of the extended field process used for the simulations. A 2d phantom is Radon transformed (Step 1) resulting in a sinogram that we impair with noise (Step 2). To create the extension of the field, the sinogram is padded with 0-s by adding trivial rows of only 0-s to the beginning of S and to the end, thereby increasing the number of rows (Step 3). After 0-padding the extended phantom is reconstructed (Step 4), the ROI extracted (Step 5), and finally the error with the original phantom computed (Step 6). **Image credit: Faisal Mahmood.**

Figure 3.2 shows some of the results of the simulations with two different phantoms, binary and Shepp-Logan, and 3 different methods, FBP, Tikhonov regularization and extended Tikhonov regularization. The EF reconstruction gives clearly the best results with the smallest error for both of the phantoms. Also, a line profile through the reconstruction shows how the extended field creates better fitting result.

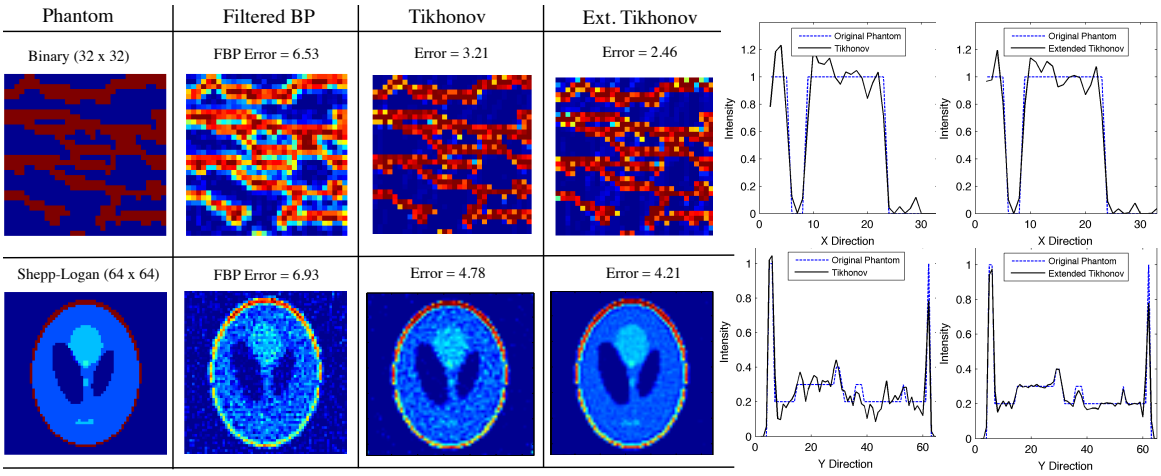

Figure 3.2 | Extend Field reconstruction results on simulations using Tikhonov regularization. The phantoms to be reconstructed are the binary phantom (first row) and the Shepp-Logan phantom. The relative magnitudes of added measurement error are 0.05 (first row) and 0.1(second row). The reconstruction error achieved with regularization is less than with filtered back-projection. Extended field Tikhonov regularization achieves even smaller error. The intensity profiles show that the extended Tikhonov performs the best and has higher regularization capability. **Image credit: Faisal Mahmood.**

3.4.2 Single Pixel Reconstruction

It is difficult to explain the effects of the extended field theoretically, as closed form solutions to the corresponding regularization problems, as far as we can derive, get quickly unwieldy beyond the most trivial cases. In the following we will attempt to solve analytically what is probably the simplest conceivable tomographic reconstruction problem, far from the complexity of a realistic tomographic problem, but mathematically tractable to provide insight into the extended field process.

Consider a tomographic reconstruction problem with the geometry given in Figure 3.3. Let us have a single pixel with density $D \in \mathbb{R}^+$, and two projections - one vertical and another horizontal. From the "experiment" we have two noisy measurements $D + e_1$ and $D + e_2$, where $e_1, e_2 \in \mathbb{R}$.

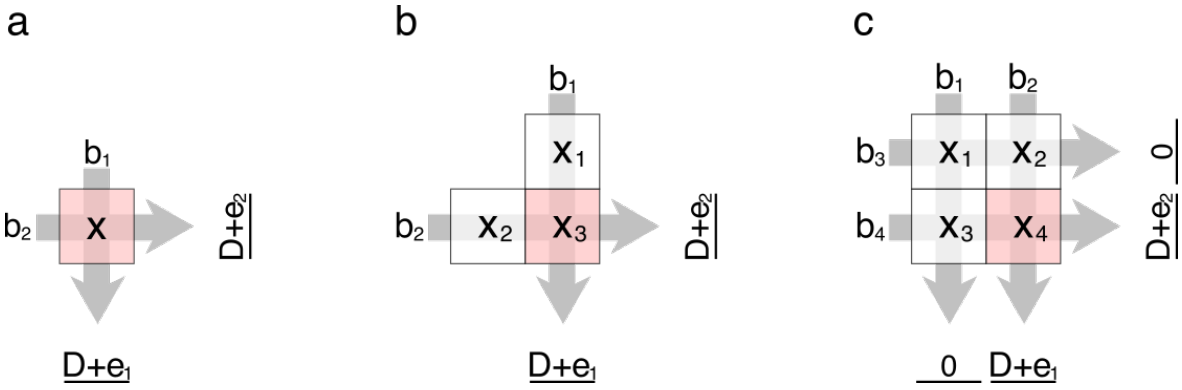

Figure 3.3 | Scheme for single pixel reconstruction with extended field. Single pixel reconstruction with **a** - two projections, no extended pixels, **b** - two projections and two extended pixels, **c** - two projections, three extended pixels and two extra conditions for the extended pixels.

We attempt to solve the reconstruction problem using Tikhonov regularization and three different strategies, two of them involving extending the field.

1. Non-extended reconstruction (Figure 3.3a). We do nothing additional and proceed in a conventional way.

2. EF reconstruction with no additional conditions (Figure 3.3b). We extend the field, but add no conditions for the extended voxels.

3. EF reconstruction with additional conditions for the extended voxels that they should sum to 0, i.e. we zero-pad the sinogram (Figure 3.3c).

Reconstruction

The least squares solution $\boldsymbol{x}^*$ to the minimization problem 3.3 can be derived analytically

$$\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I}_n)^{-1}\boldsymbol{A}^T\boldsymbol{b}\,, \tag{3.4}$$

where I_n is n by n identity matrix.

- **Non-extended reconstruction**

$$\boldsymbol{A} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\,, \quad \boldsymbol{x} = x \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} D + e_1 \\ D + e_2 \end{pmatrix}\,. \tag{3.5}$$

In this case

$$\left(\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I}_1\right)^{-1} = \frac{1}{2+\lambda} \tag{3.6}$$

and

$$x^* = (2D + e_1 + e_2)\,\frac{1}{2+\lambda} = \frac{2D + \sum e_i}{2+\lambda} \tag{3.7}$$

- **Extended Field with No Extra Conditions** Now let us extend the dimensionality of the problem by extending the field in both directions of the projections by a single pixel, as depicted in Figure 3.3b. Extending creates two extra variables x_2 and x_3, adds the corresponding coefficients to $\boldsymbol{A}$, but causes no changes in $\boldsymbol{b}$.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} D + e_1 \\ D + e_2 \end{pmatrix}. \tag{3.8}$$

Then

$$\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I}_3 = \begin{pmatrix} 1+\lambda & 0 & 1 \\ 0 & 1+\lambda & 1 \\ 1 & 1 & 2+\lambda \end{pmatrix} \tag{3.9}$$

and its inverse is

$$\begin{pmatrix} 1+\lambda & 0 & 1 \\ 0 & 1+\lambda & 1 \\ 1 & 1 & 2+\lambda \end{pmatrix}^{-1} = \frac{1}{\lambda^2+3\lambda} \begin{pmatrix} \frac{\lambda^2+3\lambda+1}{\lambda+1} & \frac{1}{\lambda+1} & -1 \\ \frac{1}{\lambda+1} & \frac{\lambda^2+3\lambda+1}{\lambda+1} & -1 \\ -1 & -1 & \lambda+1 \end{pmatrix} \tag{3.10}$$

Since we are interested in the reconstruction of only the third pixel, we need only the third row of the previous inverse. The solution to the reconstruction problem is

$$\begin{aligned} x_3^* &= \left[(\boldsymbol{A}^T\boldsymbol{A} + \lambda I_3)^{-1}\boldsymbol{A}^T\boldsymbol{b}\right]_3. \\ &= \begin{pmatrix} \frac{-1}{\lambda^2+3\lambda} & \frac{-1}{\lambda^2+3\lambda} & \frac{\lambda+1}{\lambda^2+3\lambda} \end{pmatrix} \begin{pmatrix} D+e_1 \\ D+e_2 \\ 2D+e_1+e_2 \end{pmatrix} \\ &= \frac{-(D+e_1)-(D+e_2)+(\lambda+1)(2D+e_1+e_2)}{\lambda^2+3\lambda} \\ &= (2D+e_1+e_2)\frac{1}{3+\lambda} \\ &= \frac{2D+\sum e_i}{3+\lambda} \end{aligned} \tag{3.11}$$

- **Extended Field with Extra Conditions** In the third case, as depicted in Figure 3.3c, we increase the dimensionality of the problem by one more variable and add two extra conditions that the extended pixels should sum up to 0. The

pixel to be reconstructed is x_4.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 0 \\ D + e_1 \\ 0 \\ D + e_2 \end{pmatrix}. \tag{3.12}$$

Now

$$\begin{aligned} (\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I}_4)^{-1} &= \begin{pmatrix} 2+\lambda & 1 & 1 & 0 \\ 1 & 2+\lambda & 0 & 1 \\ 1 & 0 & 2+\lambda & 1 \\ 0 & 1 & 1 & 2+\lambda \end{pmatrix}^{-1} \\ &= \frac{1}{\lambda^2 + 4\lambda} \begin{pmatrix} \frac{\lambda^2+4\lambda+2}{\lambda+2} & -1 & -1 & \frac{2}{\lambda+2} \\ -1 & \frac{\lambda^2+4\lambda+2}{\lambda+2} & \frac{2}{\lambda+2} & -1 \\ -1 & \frac{2}{\lambda+2} & \frac{\lambda^2+4\lambda+2}{\lambda+2} & -1 \\ \frac{2}{\lambda+2} & -1 & -1 & \frac{\lambda^2+4\lambda+2}{\lambda+2} \end{pmatrix} \end{aligned} \tag{3.13}$$

Here we are interested only in the reconstruction value of the 4th pixel, x_4^*. The solution to the reconstruction problem is getting increasingly more complex but it can be worked out to be

$$\begin{aligned} x_4^* &= \left[(\boldsymbol{A}^T\boldsymbol{A} + \lambda I_4)^{-1}\boldsymbol{A}^T\boldsymbol{b}\right]_4. \\ &= \begin{pmatrix} \frac{2}{\lambda(\lambda+4)(\lambda+2)} & \frac{-(\lambda+2)}{\lambda(\lambda+4)(\lambda+2)} & \frac{-(\lambda+2)}{\lambda(\lambda+4)(\lambda+2)} & \frac{\lambda^2+4\lambda+2}{\lambda(\lambda+4)(\lambda+2)} \end{pmatrix} \cdot \\ &\begin{pmatrix} 0 & D+e_1 & D+e_2 & 2D+e_1+e_2 \end{pmatrix}^T \\ &= (2D + e_1 + e_2)\frac{(\lambda+3)}{(\lambda+4)(\lambda+2)} \\ &= \frac{(2D + \sum e_i)\,(\lambda+3)}{(\lambda+4)(\lambda+2)} \end{aligned} \tag{3.14}$$

The shape of the solution is similar for each of the three cases - a data dependent term, including the sum of measurement error, times a regularization parameter dependent term

- **Method 1:** Non-extended reconstruction
  $x^* = (2D + \sum e_i)\frac{1}{\lambda+2}$

- **Method 2:** Extended field with no extra constraints
  $x^* = (2D + \sum e_i)\dfrac{1}{\lambda + 3}$

- **Method 3:** Extended field with extra hyperplanes
  $x^* = (2D + \sum e_i)\dfrac{\lambda + 3}{(\lambda + 4)(\lambda + 2)}$

A Perfect Reconstruction

Next, let us investigate for each of the three cases which regularization parameter $\lambda \in \mathbb{R}^+$ will produce a prefect reconstruction equal to D. To do that, we equate x^* with D and solve for λ. Also, to make the problem independent of the density D, let us express the sum of the measurement errors as a fraction of D and denote it by α

$$\alpha = \sum e_i / D \,. \tag{3.15}$$

- **Method 1:** Non-extended reconstruction

$$\frac{2D + \alpha D}{\lambda + 2} = D \quad \Rightarrow \quad \lambda = \alpha \tag{3.16}$$

  From $\lambda > 0$ it follows that $\alpha > 0$.

- **Method 2:** Extended Field with No Extra Conditions

$$\frac{2D + \alpha D}{\lambda + 3} = D \quad \Rightarrow \quad \lambda = \alpha - 1 \tag{3.17}$$

  From $\lambda > 0$ it follows that $\alpha > 1$.

- **Method 3:** Extended Field with Extra Conditions

$$\frac{(2D + \alpha D)\,(\lambda + 3)}{(\lambda + 4)\,(\lambda + 2)} = D \quad \Rightarrow \quad \lambda = \frac{\alpha}{2} - 2 + \sqrt{\left(\frac{\alpha}{2}\right)^2 + \alpha + 2} \tag{3.18}$$

  From $\lambda > 0$ it follows (after some straightforward algebra) that $\alpha > 2/3$.

Since the regularization parameter λ is non-negative by definition, we get that each of those three methods can produce the correct results only in the cases when the sum of error is non-negative. Moreover, for the method 2 the sum of error needs to at least match the density in magnitude, and for the method 3 the sum of errors has to be more than 66% of the density. That is to achieve a reconstruction with 0 error.

Relative Error

Finally, let us examine the reconstruction errors relative to the density D. In the following we are omitting method 2 from the discussion, as it is the worst-performing, and does not correspond to the strategy chosen in the simulations.

Let us define the relative error as

$$\varepsilon = \left| \frac{x^* - D}{D} \right| . \tag{3.19}$$

The relative errors for the methods 1 and 3 are then

$$\varepsilon_1 = \frac{\left| \frac{2D + \alpha D}{\lambda + 2} - D \right|}{D} = \left| \frac{\alpha - \lambda}{\lambda + 2} \right| \tag{3.20}$$

and

$$\varepsilon_3 = \frac{\left| \frac{(2D + \alpha D)(\lambda + 3)}{(\lambda + 2)(\lambda + 4)} - D \right|}{D} = \left| \frac{(2 + \alpha)(\lambda + 3)}{(\lambda + 2)(\lambda + 4)} - 1 \right| \tag{3.21}$$

The relations between α and λ that give relative error 0 are derived above when finding the conditions for getting a perfect reconstruction.

Another interesting question is that for which α and λ are the two methods equivalent, i.e. giving equal error, and in which regions does one outperform the other. This can be found out by solving $\varepsilon_1 = \varepsilon_3$ for λ

$$\varepsilon_1 = \varepsilon_3 \Rightarrow \lambda = \frac{1}{2} \left(\sqrt{\alpha^2 + 6\alpha + 12} + \alpha - 4 \right) . \tag{3.22}$$

The above relations are visually represented on Figure 3.4 for methods 1 and 3. The graph shows only the cases where the sum of measurement errors is a positive fraction of the true density, because in the negative part neither one of the methods is able to reach 0 error. We see that when $\alpha > 2/7$ then extended field gives a smaller relative error at the same regularization parameter value.

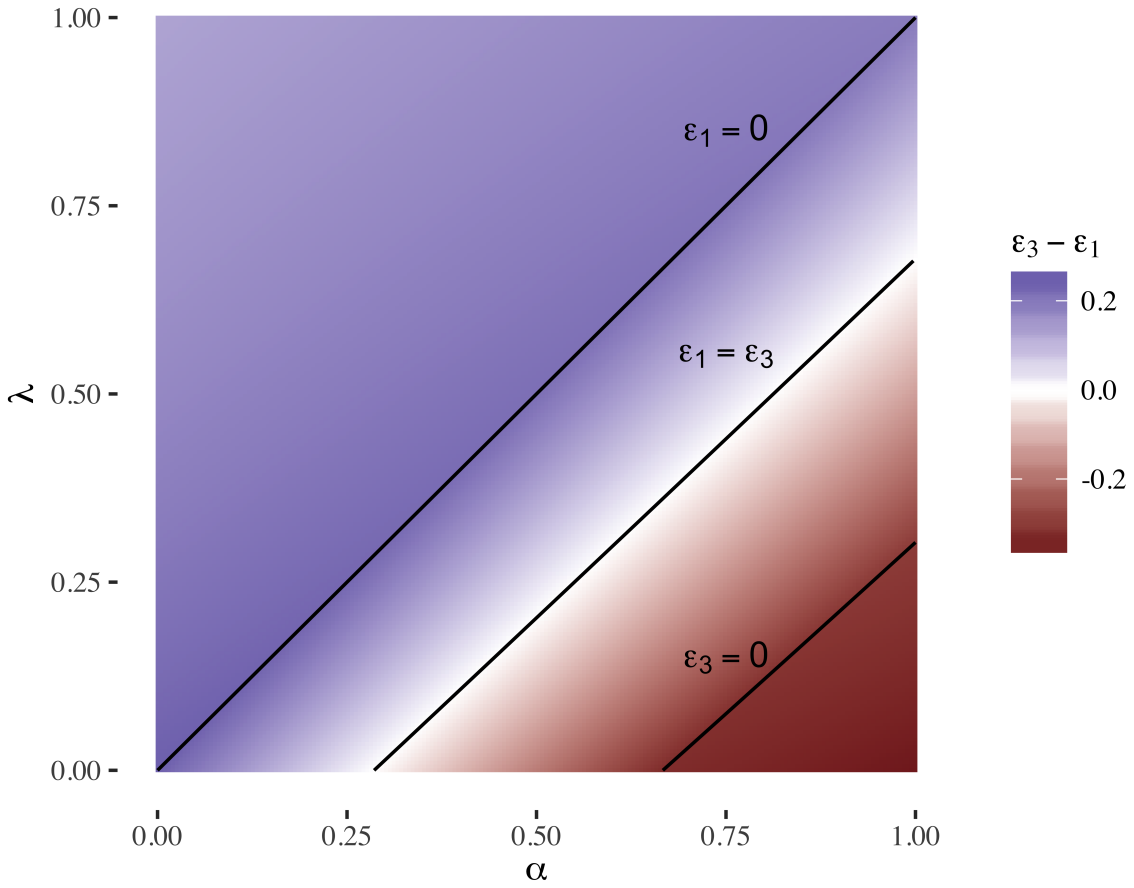

Figure 3.4 | Relative reconstruction error for non-extended field and extended field with extra constraints. This graph illustrates the difference in relative reconstruction error between non-extended reconstruction (ε_1) and the extended reconstruction (ε_3). In the purple region $\varepsilon_1 < \varepsilon_3$ and the opposite is true for the red area. Respective lines where either of the errors is 0 and where the errors are equal are depicted with black lines.

Conclusions from the Single Pixel Reconstruction

This simple exercise has shown that there are cases where it makes sense to use extended field reconstruction. We saw that for the simple single pixel reconstruction with Tikhonov regularization extended field offers an advantage in the cases where the total measurement error is positive and equal to a relatively large fraction of the true density. In such cases, even though the non-extended reconstruction scheme can obtain 0 error, it does so at a larger regularization parameter λ than the extended field reconstruction.

Another point worth mentioning is that here we have not used the prior knowledge that the density to be reconstructed must be positive. In the simulations with more complex reconstruction problems we imposed the positivity constraint to the region of interest only, whereas the extended pixels were free to take negative values.

Getting closed form solutions for any larger reconstruction problems are laborious and the equations become quickly unwieldy. At this point it is unclear how these

derivations could be generalized to more complex problems. More thorough theoretical treatment of a general reconstruction problem with extended field method remains for the future.

3.4.3 The Essence

Consider again the single pixel reconstruction from the previous section and as depicted on Figure 3.3. In each of the cases we are looking for a least squares solution to the linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. The equation system for the extended case was the following

$$\boldsymbol{A}' = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{x}' = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b}' = \begin{pmatrix} 0 \\ D + e_1 \\ 0 \\ D + e_2 \end{pmatrix}. \tag{3.23}$$

Notice that we can always rearrange the elements in $\boldsymbol{x}'$ and $\boldsymbol{b}'$ so that the first elements correspond to the non-extended field and the last elements to the extended field. To leave the system itself identical we also have to reshuffle the rows and columns of the matrix $\boldsymbol{A}'$ accordingly, which we can always do. In this case the reordered equation system looks the following

$$\begin{aligned} \boldsymbol{A}''\boldsymbol{x}'' = \boldsymbol{b}'' \Rightarrow & \left(\begin{array}{c|ccc} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right) \begin{pmatrix} x_4 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} D + e_1 \\ D + e_2 \\ 0 \\ 0 \end{pmatrix} \\ \Rightarrow & \begin{pmatrix} \boldsymbol{A} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{x}_{ex} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix}. \end{aligned} \tag{3.24}$$

After reordering it becomes clear what the effect of the extended field could be. We see that after extending the field we still have the original equation system embedded in a larger system, with the modification that for every linear condition we add a couple of new variables, that serve as a garbage bin for dumping noise. The last rows of $\boldsymbol{A}''$ concern only the extended variables as the coefficients corresponding to the ROI variables are all 0. These rows set the constraints that various linear combinations of the extended variables should add to 0.

This makes intuitively sense - for each condition (row in original $\boldsymbol{A}$) that cannot be satisfied exactly due to added noise, we give some extra variables in a hope that the noise would propagate to those extended variables, and then say that the extended variables should actually sum to 0 as they are assumed to be Gaussian with mean 0. In the iterative reconstruction schemes we can take advantage of the prior knowledge that the reconstructed density in the ROI should be positive. This condition, however, is not enforced on the extended pixels that are therefore free to take negative values.

In this light we can generalize the extended field principle away from the geometric interpretation and construct however elaborate schemes for assigning extra variables to each condition and then constraining the extra variables in turn.

Consider the following, perhaps the simplest, scheme

- For each row in original $\boldsymbol{A}$ let us give a single extra variable.
- For the extra variables add just a single constraint that the sum of all of them should be zero.

The extended version of a matrix $\boldsymbol{A}$ with n rows and m columns is then

$$\boldsymbol{A}'' = \begin{pmatrix} \boldsymbol{A} & \text{diag}(n) \\ \mathbf{0}_m & \mathbf{1}_n \end{pmatrix}, \tag{3.25}$$

where $\mathbf{0}_m$ and $\mathbf{1}_n$ are vectors consisting of only 0s and 1s with lengths m and n, respectively.

3.5 Conclusions

More thorough theoretical investigation of the extended system could shed some light what the effects on the solution are with respect to the original non-extended system, but this work is left for the future. The final conclusion is that the extended field seems to be more general than initially thought and could serve useful in certain scenarios.

Chapter 4

Nanocrystal Cryo-ET

4.1 Introduction

The main goal of this project was to explore the feasibility of and provide a proof of concept for determining the structure of a protein molecule using cryo-electron tomography of nanocrystals. In addition to molecular structure, tomographic reconstruction of a protein crystal is an interesting object in its own right as it provides a perspective into the world of small biological crystals that is not perhaps often seen. The second major goal then was to develop the tools and methodology that would enable the analysis of the nanocrystal itself, its lattice order and possible crystalline imperfections. As protein nanocrystallization, tomographic imaging of the crystals and the subsequent structure determination by harnessing symmetry is relatively unstudied in the literature, it was unclear in the beginning how many obstacles would be encountered on the road.

The Choice of a Target Protein

When developing a new method it is always a good idea to have something to validate the outcome against. That is the main reason why we decided to use hen egg white lysozyme (HEWL) as the first experimental target and did not opt for a perhaps more interesting protein with hitherto unknown structure. HEW lysozyme was the first enzyme and the second protein ever to have its structure determined via x-ray diffraction in 1965 [69], and is thereafter exceptionally well studied in the literature, making it an ideal model system. Doing a search for lysozyme in PDB results in 1785 structures from various organisms, solved using various methods, and in a large variety of different space groups. Lysozyme is easy to crystallize to large well-diffracting crystals, but also several protocols for creating nanocrystals can be found. Lastly, but importantly, lysozyme is cheap and there is easy access to large amounts.

The Challenges

The first challenges were how to create nanocrystals so small that they would be visible under 300 keV electron beam and would fit entirely into the field of view. In principle, as long as the crystals are thin enough, they do not necessarily have to fit entirely into

the field of view, especially so for electron diffraction. But in our case, getting a 3d reconstruction of a whole crystal became a separate goal in order to study the shape and the interface of the crystal. Therefore, needle-shaped crystals that have only two dimensions properly nano-sized and the third one much longer, as imaged in [37] for example, were not suitable for our goals.

Getting a high SNR image of a biological sample is notoriously challenging. When the molecule under scrutiny is large, it can be seen more easily and even individual molecules could be picked out from a processed tomogram. Lysozyme, being only 15 kDa heavy, belongs to the small end of the size spectrum and is generally considered too small for cryo-ET and single particle reconstruction (SPR).

Initially, there was a concern whether lysozyme is indeed too small for tomography. The experiments, however, showed that the size of the molecule did not become a limitation and instead gave an extra quality to the project. Namely, in a 3d reconstruction, the outline of the crystal, as opposed to the outline of an individual molecule, is almost always visible, which gives good indication as to where to look for the molecules. When investigating low SNR tomograms the question of whether one is looking at a random blob of intensity or electron density actually belonging to a molecule, is ever-present. In a crystal, the molecule makes its appearance periodically at fixed intervals, so looking for something periodic helps in locating the individuals.

The biggest challenge, perhaps, was to develop the software suite for nanocrystal analysis. The programs had to be robust to noisy data, able to identify symmetries and use that information to deliver a symmetrized and averaged version of the reconstruction. The curse of symmetrization is that after forcefully imposing symmetry on any image, the outcome always conforms perfectly to that symmetry, and the higher the symmetry, the "better" and aesthetically pleasing the end result looks. Whether the result actually makes sense and is a correct representation of the underlying signal is a separate question. So, proper sanity checking and validation of the software and the outcome becomes incredibly important.

4.2 Outline

In this Chapter we present the central project of the thesis.

Methods: In Section 4.3 we discuss the obstacle-laden road and the eventual path to successful preparation of samples containing tiny lysozyme nanocrystals in the size range of 100 nm. We also detail the evaluation of the intermediate non-optimal samples by direct imaging with TEM, and finally show images of some of the best examples of small nanocrystals and locations where the tomographic data collection was performed.

Results: Section 4.4 presents the results of the 3d reconstructions and much more. We first examine the steps taken from tilt-series to a back-projected reconstruction. After that we discuss and show the results from the lattice detection task, segmenting out the crystal and averaging of the unit cells. The section ends with showing individual as well as averaged molecular structures extracted from a larger reconstruction, discussions on the resolution, and various visualizations of the nanocrystals. We also explore the lattice order by looking at the distribution of lattice vectors, the symmetry origin and attempt to classify the unit cells into homogeneous clusters.

Programs: Section 4.5 details the algorithmic side of the methodology and the programs developed specifically for going from a general 3d reconstruction to a symmetrized version of it. Additionally, some other tools for visualization are discussed. In the final part of this section we validate the programs on simulated datasets.

4.3 Methods

Creating protein crystals in a wet-lab environment is sometimes considered more of an art than science, as conditions that work for one object of interest might prove useless or irrelevant for others. It is not uncommon for a researcher to spend years trying to grow a large well-ordered crystal from their protein of interest. Lysozyme is one of the easiest proteins to crystallize and has therefore been used extensively as a model system for research as well as for classroom experimentation. But less is known about how to produce nanocrystals suitable for electron microscopy at the same time avoiding the formation of large electron-opaque crystals. Some general principles, that favor the growth of many small crystals over few large ones, are a high concentration of dissolved protein that is made to fall out of solution relatively fast.

4.3.1 Nanocrystallization

The Trials

For our experiments we used commercially available hen egg white lysozyme purchased from Sigma Aldrich in a lyophilyzed powder form. To observe a sample in cryo-electron microscope, it has to be first deposited on an electron microscopy grid, and then flash-frozen to fix the biological material and slow all thermodynamic activity down to a halt. Unlike in a conventional approach, where the crystallization drop is prepared separately on a crystallization tray and let sit untouched for some time to give the crystals chance to form in peace, we used a different method.

Since, depending on the conditions, lysozyme crystals grow relatively quickly, and nanocrystals form effectively in an instant, a precise control over the timing of the steps was needed. The first strategy was to let the crystals grow on a crystallization tray for a certain amount of time and then try to carefully pipette out the remaining solution, all the while trying to avoid sucking in large crystals in the hope that the captured liquid will contain smaller crystals. Doing it this way it was difficult to completely avoid large crystals and the result was almost totally clogged grid which was effectively electron-opaque when observed under TEM. On top of that, we were not sure that there even were nanocrystals floating between the large ones. Also the time it took to transfer the material from crystallization tray to a grid and then freeze took too long to provide exact control over the crystallization time. Another concern was that the nanocrystals might be fragile and otherwise unstable so all excess manipulation should be minimized.

Since the crystallization process became quicker as the concentrations of protein and the precipitant increased, and for lysozyme it seemed to be a question of seconds, we decided to eliminate as many extraneous steps as possible, and grow the crystals

directly on a plasma-cleaned carbon-coated microscopy grid mounted in the flash-freezing device (FEI Vitrobot). This, at least, provided perfect control over timing. Crystallizing lysozyme directly on the grid was also easier for the human operator, but after many unsuccessful attempts, due to either a too thick of a sample, or the inability to find any crystals, the growing concerns were that the high concentration of the components makes the material too dense to allow for high resolution imaging.

In several cases it was visible to the naked eye, that upon lifting the frozen grid from the liquid ethane container into a cryo-box inside liquid nitrogen, a white layer instantly formed on the grid, which was suspected to be solidified ethane. A general belief is, that the ethane contamination will quickly sublimate in the vacuum of the electron microscope column, but this did not happen in our experience and, most importantly, did not help us get rid of the electron-opaqueness of the sample, which became a major head ache. As a remedy, a range of automatic and manual blotting conditions to wick away even the leftover liquid ethane were tested, but to no avail.

Another thought was that the nanocrystals needed more time to be able to settle on the grid and adhere to the surface, so perhaps longer waiting times were still necessary. This led to experimenting with a range of waiting times between the start of the crystallization and the flash freezing, from 5 s to 5 min.

Fortunately, it was noticed that sometimes, in spite of all the uncertainties, the grids turned out just fine, were perfectly transparent, and a large amount of protein crystals were immediately discernible. Cryo-sample preparation procedure is long, laborious and arduous, and takes time to master, so it cannot be excluded that the main source of the bad grids was in fact poor sample handling. The overall conclusion was, though, that certain combination of parameters made successful sample preparation more probable and by just persevering good grids would eventually appear.

The Final Protocol

The final crystals were produced by the following protocol.

A solution of 200 mM HEWL with 50 mM NaAc was prepared. 1 µL drop of 10 nm Au beads suspended in 50 mM NaAc was applied directly on a plasma-cleaned (Gatan Solarus) Quantifoil R2/1 grid, mounted in FEI Vitrobot at 100% humidity and temperature 4° Celsius. 3 µL of lysozyme was applied on top of the Au drop, after which the grid was gently touched with a drop of 3.0 M NaCl and 50 mM NaAc solution, which ignited a visible precipitation front and a suspected intense crystallization process. The grid was immediately blotted twice for 5 s, blotting force 4, and plunged into liquid ethane to flash-freeze the sample and encapsulate the crystals in amorphous ice. The frozen grids were directly transferred to electron microscope for imaging.

4.3.2 Cryo-Electron Tomography

The transmission electron microscope used for imaging was an FEI Titan Krios operating a field emission gun at an acceleration voltage of 300 kV and using a FEI Falcon II detector. In this study we did not use any post-column energy filters that are useful for filtering out inelastically scattered electrons.

Initially, having no idea about exactly what to look for in the samples, locating the protein crystals was challenging. The problem was that together with protein crystals the samples most probably contained also salt crystals and ice contamination and having no prior experience in the field made it difficult to distinguish between those. It also happened from time to time that the sample appeared to have no protein crystals at all.

For electron tomography it is crucial to get the sample as thin as possible to reduce background coming from inelastic scattering. Some of the first crystals that we saw were too large (See Figure 4.1). These crystals were a couple of micrometers in length covering an entire hole in the carbon support film. Crystals of that size are hopelessly large for electron tomography. Also the overall sample thickness near the crystals was too thick for good quality data.

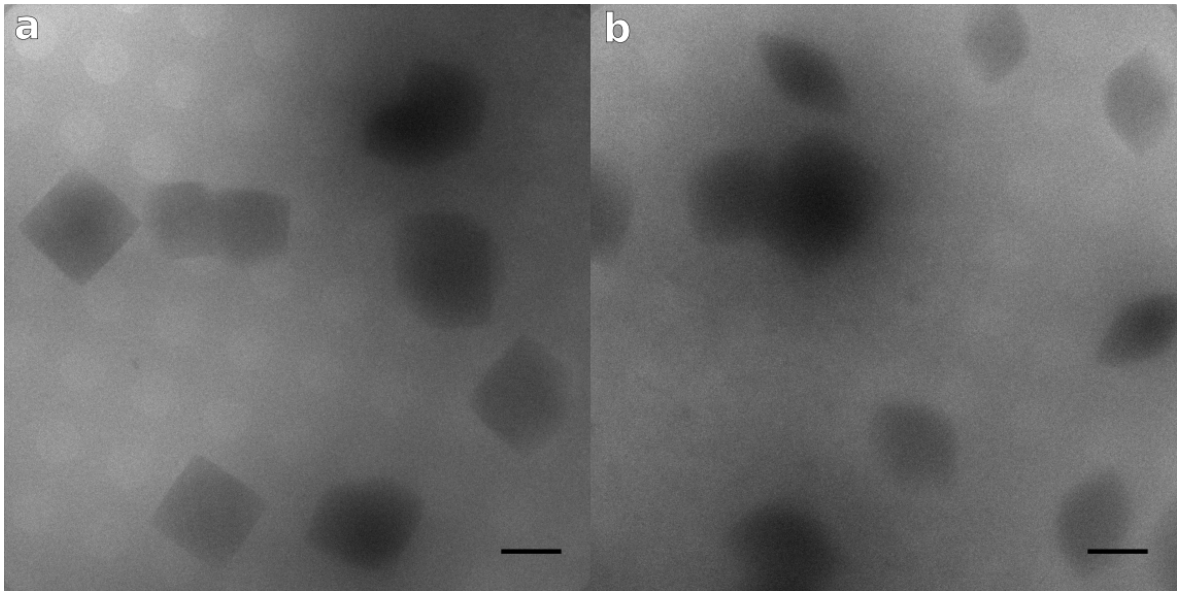

Figure 4.1 | Examples of too thick samples containing lysozyme microcrystals. The crystals are obvious but are too large for tomography and the sample is too thick for good quality imaging. These crystals can not be classified as nanocrystals anymore as they are several micrometers in size.

After some failed attempts, frustration, and reflecting on what might have gone wrong, necessary changes were introduced to the sample preparation protocol and eventually better quality samples started appearing. On Figures 4.2 and 4.3 obvious protein nanocrystals are seen, as judged by the direct observation of the lattice, and backed up by the visible long distances between the diffraction spots in the FFT, that cannot belong to a mineral crystal.

Micrographs on Figures 4.2 and 4.3 are taken at higher magnification, than what was used for electron tomography, and at a much higher electron dose, when compared to a what can be used for a single image in a tilt-series. This means, that once the images are taken, the crystals are destroyed, and no further data collection there is possible. Also, the crystals themselves with side length of around 700 nm are still on the large side, as compared to what we would ideally like for tomography.

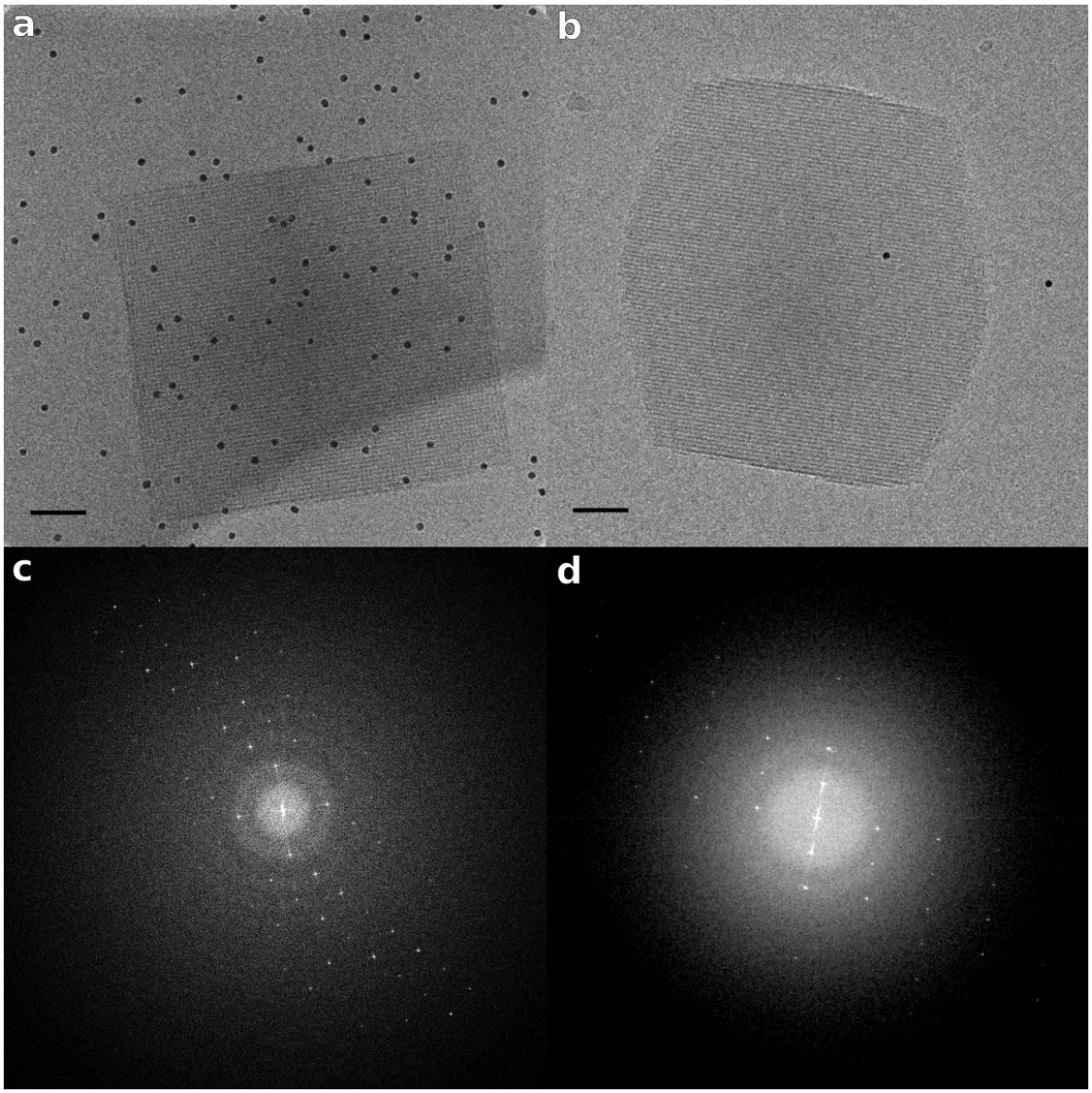

Figure 4.2 | Examples of medium sized lysozyme nanocrystals. **a** and **b** show examples of large nanocrystals. Both of the crystals are about 500 nm across. We cannot easily tell the thickness from these images, but with high likelihood they would be too large for tomography. There is no question whether these are protein crystals as we can directly see the lattice. Also, the diffraction patterns on the FFT (**c** and **d**) reveal the cell spacing to be too large for mineral crystals. The crystal on **a** has perhaps too much gold, that could derail a reconstruction algorithm and shadow the interesting details. The scale bar is 80 nm on both images.

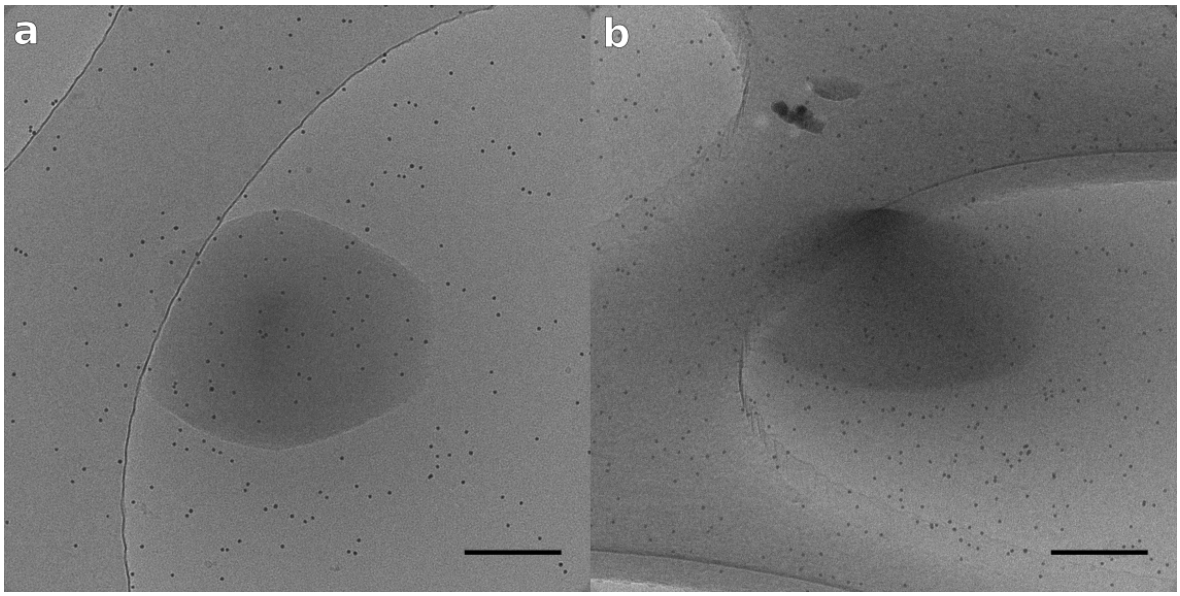

Figure 4.3 | Example of a medium sized lysozyme nanocrystal. **a** and **b** show two different views of a 700 nm lysozyme nanocrystal, that is slightly too thick for tomography. The scale bar is 250 nm.

Gradually, we managed to make the nanocrystals smaller and smaller, and the eye got trained to quickly tell lysozyme crystals apart from other uninteresting features already at low magnification. Figure 4.4 shows an overview image of a location where 320 nm and larger square-shaped lysozyme nanocrystals are scattered over a large area. This image is taken as part of the search process for potential locations for data collection, at a magnification of 2200x, binning the pixels twice, at an underfocus of 100 µm to boost the contrast of low resolution details, and all the while staying conservative with electrons not to damage the crystals.

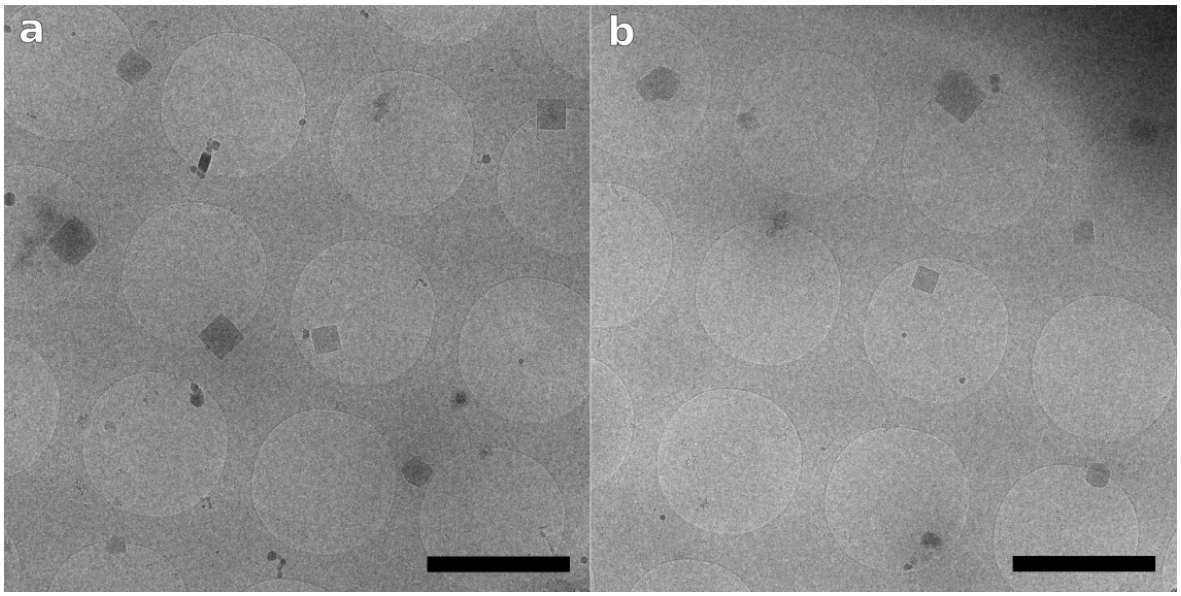

Figure 4.4 | Larger view of the grid showing plenty of nanocrystals in the size range of 350 nm **and larger**. **a** and **b** show two low magnification images featuring lots of lysozyme crystals. These positions were used for tilt-series recording and the images were taken as a part of the search process, taking care to minimize electron exposure. The scale bar is 2 µm on both images.

Finally, on Figure 4.5 two small nanocrystals are visible, named $D16$ and $D18$, with side lengths of around 130 nm and 200 nm. These locations led to acceptable datasets that were used for further processing and analysis. In general, the smaller the crystals got, the better quality tilt-series they produced.

The best dataset was collected at a location visible on Figure 4.6 that features a crystal roughly 130 nm in length and 100 nm thick. The following analysis is mostly done on this crystal, named $D17$.

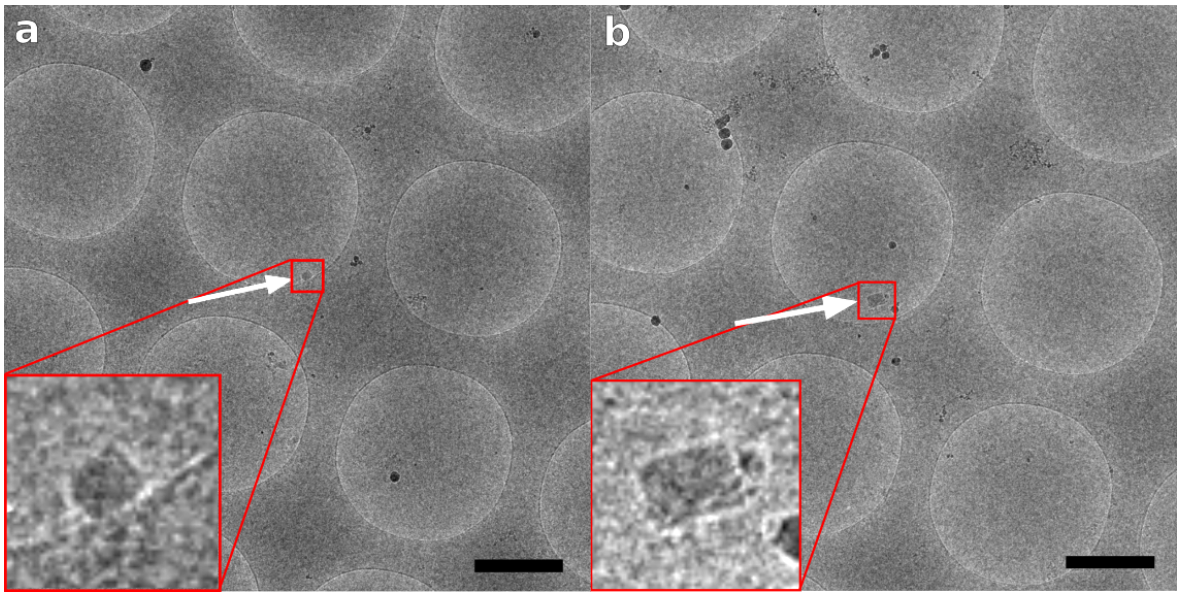

Figure 4.5 | Low magnification view of the grid featuring 130 nm **and** 200 nm **lysozyme crystals**. These positions were used for tilt-series recording and the images were taken as part of the search process, taking care to minimize electron exposure. The insets show a close-up of the crystal and have a slightly increased contrast for visualization purposes. The scale bar is 1 µm on both images.

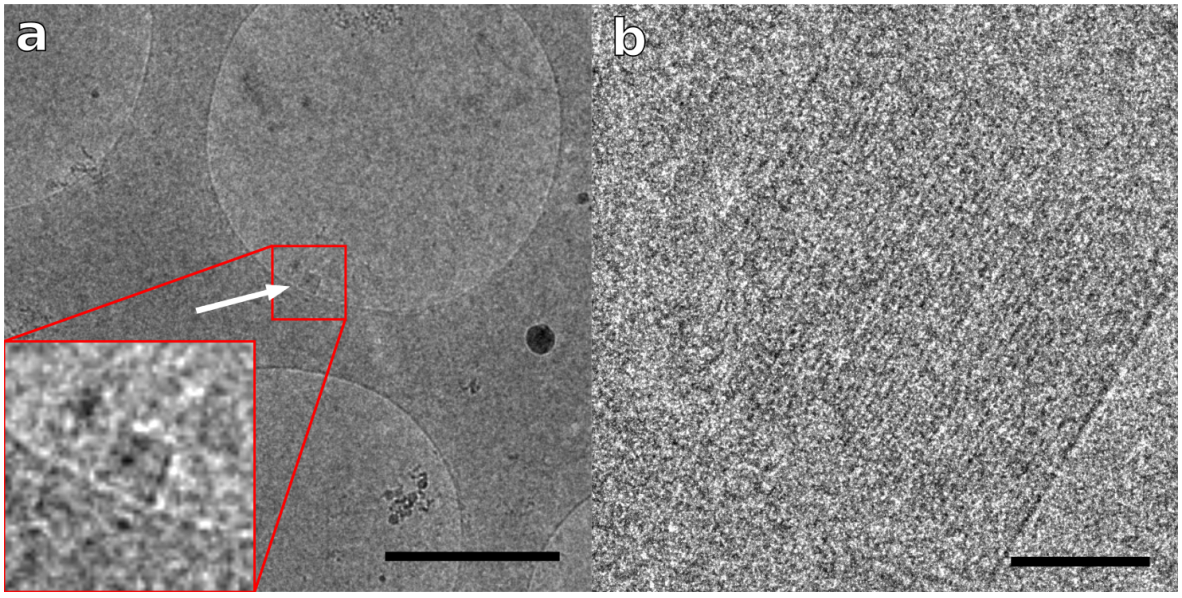

Figure 4.6 | Low magnification view of the grid featuring a 100 nm **nanocrystal and one single tilt from the collected series**. **a**, Shows one of the smallest nanocrystals we were able to find and record tilt-series of. Its side length is around 100 nm. The thickness of the crystal, as measured from the tomogram, is also about 100 nm. **b** a single projection from the collected tilt-series at a 17° angle. The lattice is clearly visible. The scale bar is 1 µm.

The tilt-series were collected at a magnification of 37000x covering an angular range from $-67°$ to $67°$ by $1°$ increments, making in total 135 images. The cryo-EM stage could have been tilted up to $70°$ on each side, but due to obstructions on the sample itself the highest tilts had to be omitted. At a magnification of 37000 the size of one pixel is 2.26 Å.

The total accumulated electron dose over the whole tilt-series was estimated to be around $40\,e^-/\text{Å}^2$. Dividing it by the number of images in the tilt series and multiplying by the area of one pixel gives $40 * 2.26^2/138 = 1.48$ electrons per pixel per image. At this low dose individual tilts are almost entirely devoid of contrast. But at certain tilt angles the molecules in a crystal all line up and the lattice, albeit noisy, becomes easily discernible (Figure 4.6b), which is the beauty of working with protein nanocrystals. While individual molecules are challenging to see in the unprocessed images, especially true for molecules as small as lysozyme, the nanocrystals can be easily found in the original sample as well as in the recorded tilt-series.

On Figure 4.7 we can see three different views from a tilt-series of a 130 nm nanocrystal $D16$. At some views (Figure 4.7b and c) the molecules line up perfectly and generate enough contrast to make the crystal lattice discernible. On a tilt just $2°$ away (Figure 4.7b) the lattice has completely disappeared and even the outline of the crystal is just barely distinguishable.

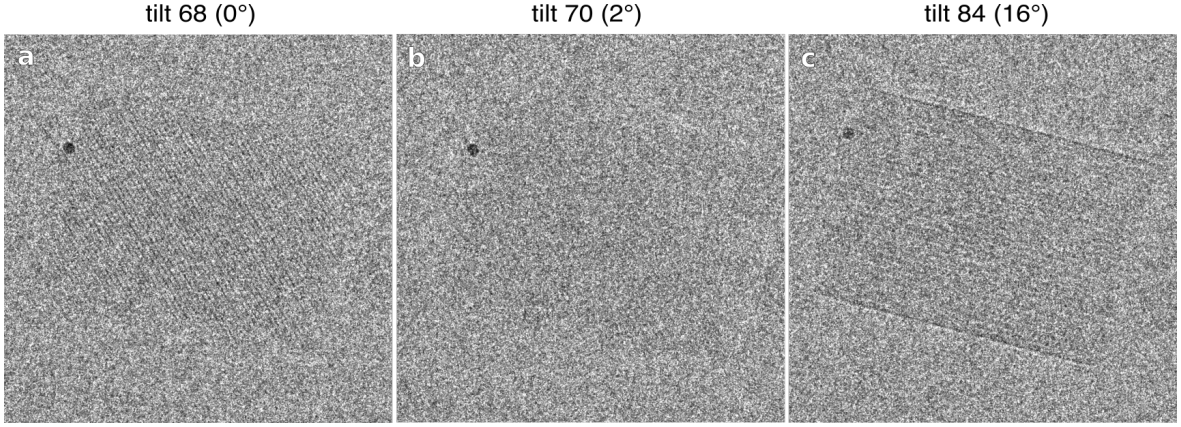

Figure 4.7 | 3 different views from the tilt-series D16 featuring a 130 nm **lysozyme nanocrystal**. **a** - view number 68 exactly in the middle of the series at a $0°$ tilt shows clear lattice. **b** - view number 70, just a $2°$ tilt from **a**, but the lattice has almost completely disappeared, and the crystal outline is just barely distinguishable. **c** - at a higher tilt angle of $16°$ the lattice becomes visible again. The black spot on all of the images is the same 10 nm gold bead.

Imaging parameters

4.3.3 Conclusion

As concluding remarks to the methods section, we saw that growing and imaging lysozyme nanocrystals is possible and feasible. The recent interest into nanocrystallization techniques, triggered by the success of serial femtosecond crystallography and the 3d electron crystallography of protein nanocrystals, is bound to make the process

even more well-defined, reliable and reproducible. The supply of the nanocrystals is not therefore going be the limiting step of nanocrystal tomography.

4.4 Results

Our recorded tilt-series consisted of 135 images from $-67°$ to $67°$, each 4096 by 4096 pixels. In this section we provide details about the process from the raw data to a molecular structure, by taking advantage of the symmetry and the programs that were developed this specific task in mind. The algorithmic details of the programs are given in the next section.

4.4.1 Tomographic Reconstruction

Going from a tilt-series to high quality 3d reconstruction is a long and difficult journey. Owing to the popularity of the field, the modern software tools, and the increase in the raw computing power, the process is becoming increasingly easier. Nevertheless, almost every step involves tuning some parameters, whose direct effect might be hard to understand, and the feedback loop can be long and sometimes not so obvious.

The general reconstruction process could be divided into the following broad steps

- Aligning the tilt-series and determining geometry.
- Back-projecting the individual views to a 3d reconstruction.
- Regularizing the reconstruction and deconvolving the contrast transfer function (CTF).

On Figure 4.8 4 different tilts from the tilt-series of the crystal $D17$ are shown. From certain angles the crystallographic lattice becomes visible. In the following we are going to reconstruct that nanocrystal.

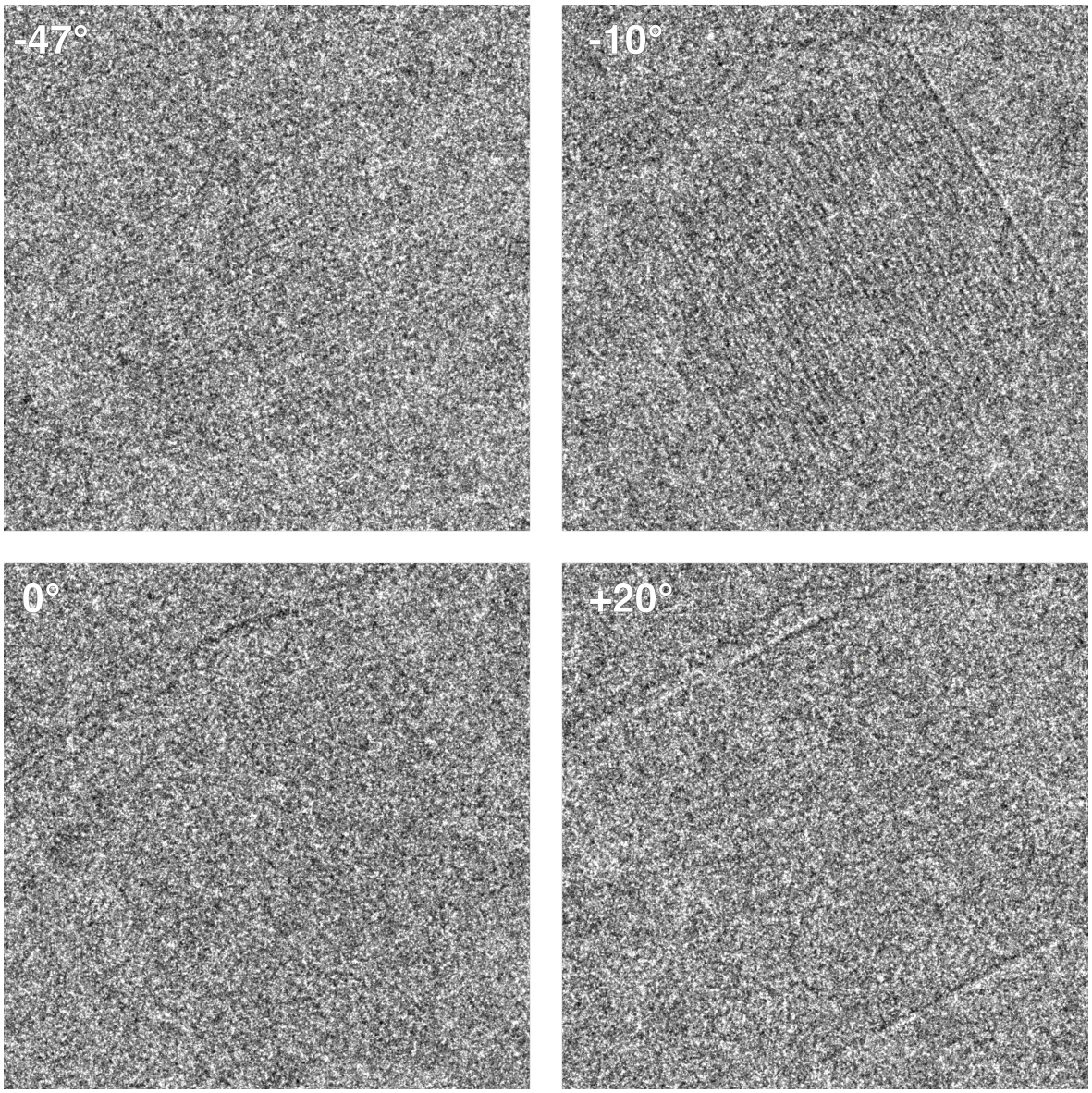

Figure 4.8 | 4 view of a D17 tilt-series 4 particular views of a tilt-series containing the crystal D17. From some angles the lattice is visible, while from others almost the whole crystal has disappeared.

Aligning the Series

Two of the conventional approaches for aligning the tilt-series are by cross-correlation, that gives a coarse alignment and involves only translational shifts, and a fine alignment by tracking gold markers. In the cross-correlation approach first the cross-correlation function between consecutive tilts is computed, a peak in the function detected, and one of the images shifted accordingly. This method seems to work fine for high-contrast images such as stained samples. On low-dose cryo-ET images the cross-correlation function can often be misleading and a peak poorly defined, so the suggested shifts become unstable and inaccurate.

The second approach, that is conventional for cryo-EM, and was also employed

here, is to take advantage of gold nano-beads. Gold, being almost entirely electron-opaque, can usually be easily seen in the individual images and therefore serves as a good marker for determining the geometry. The gold markers are picked on each image either manually or by using some automatic gold picking algorithm. Provided the gold markers have not moved individually, the positions of the markers then tell how the sample itself as a whole has moved during the imaging. The coordinates of the gold markers are then used for aligning the images to a common reference frame and for solving for parameters that determine the overall image acquisition geometry.

In general, the more gold markers we have, the higher the possibility for getting a good alignment of the images as the individual errors will be averaged out. On the other hand, too many gold markers (such as for example in Figure 4.2) will cast heavy shadows, as they are almost impenetrable to electrons, and might interfere unfavorably with reconstruction algorithms. It is also undesirable to have any gold beads in the reconstruction region of interest, unless, of course, the gold bead itself is the target of the study. In most of the datasets that we collected there were very few golds, and on some tilt-series the golds were missing altogether. The tilt-series $D17$ had two gold markers. This would probably not be enough for older electron microscopes where tilting the specimen is done manually.

Modern TEMs like FEI Titan Krios have accurate built-in machinery, that, together with fully automatized data collection procedure, constant tracking and compensation for shifts, can deliver consistent tilt-series. In that case, accurate alignments are achievable even with just a few gold beads. In addition, several of the geometry unknowns, such as change in magnification and the tilting angle, can be fixed, and thereby the number of variables to solve for reduced.

Picking the gold markers and solving for the geometry was done manually using in-house software. The average error in the alignment of the tilt-series D17 was 8 Å. Though the alignment error may seem large, it should be stressed that it is only the average error over all of the tilts and coming from mostly the poorly aligned high-tilt views that affect the resolution in z direction, i.e. parallel to the electron beam. The alignment is also improved during the reconstruction that further increases the overall achievable resolution. By unlocking the specimen rotation parameter, the alignment error went down to 2 Å, but this is likely an overfit, since with two gold beads there are too few data points to estimate that many variables.

3d Reconstruction

3d reconstruction of the tilt-series was done using 80 Å filtered back-projection reconstruction as a prior information to the COMET reconstruction algorithm [50]. The prior information is treated as a knowledge we believe to be true, and a back-projection to that low resolution is probably close to correct. In this method the CTF is deconvolved as part of the reconstruction procedure, as opposed to deconvolving the images in the tilt-series directly.

We define the CTF by making an estimation of the parameters such as the defocus and the fraction of amplitude contrast and apply it as a part of the forward operator. The defocus imaging was set to be $-2.2\,\mu$m, . In the literature one can find several attempts to estimate the value of the fraction of amplitude contrast for various types of

biological specimens. For example, in [70] the fraction was estimated to be 0.027 ± 0.01 for a bacterial flagellar filament in vitreous ice, and 0.058 ± 0.018 for amorphous carbon, at acceleration voltage of 300 keV and without electron energy filtering. It was also found there that using energy filtering markedly improves amplitude contrast, but we did not use it with our data.

We also perform local geometry correction during regularization, which means that on each iteration the algorithm tries to slightly shift and rotate the images in order to further optimize the target function.

At a binning 1, making the voxel size 2.261 Å identical to the pixel size in the individual tilts, the size of the whole reconstructed volume that captured the nanocrystal was 1024 x 1024 x 800 voxels. On a modern iMac computer with, 32Gb of RAM and 3.5GHz 8 core processor, the reconstruction took 17 hours.

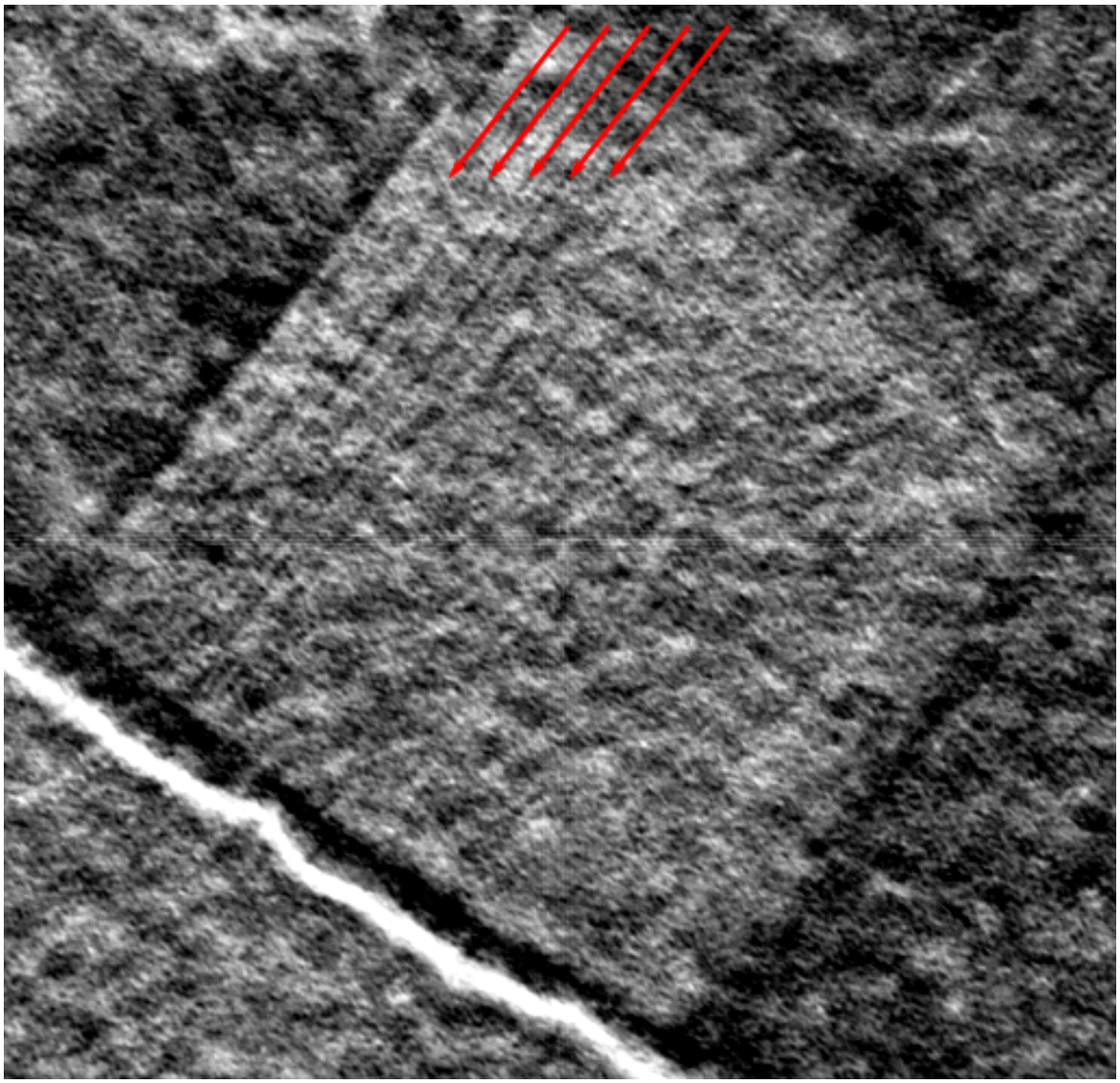

Figure 4.9 | Solid rendering of the whole 3d reconstruction of a lysozyme nanocrystal (D17). In the reconstruction the crystal together with a lattice is visible (red arrows in the top left corner). The rest of the lattice lines are not visible due to perspective projection. The lattice lines perpendicular to the indicated ones were not as clearly visible. The strong edge through the lower left corner is the edge of a hole in the carbon support film.

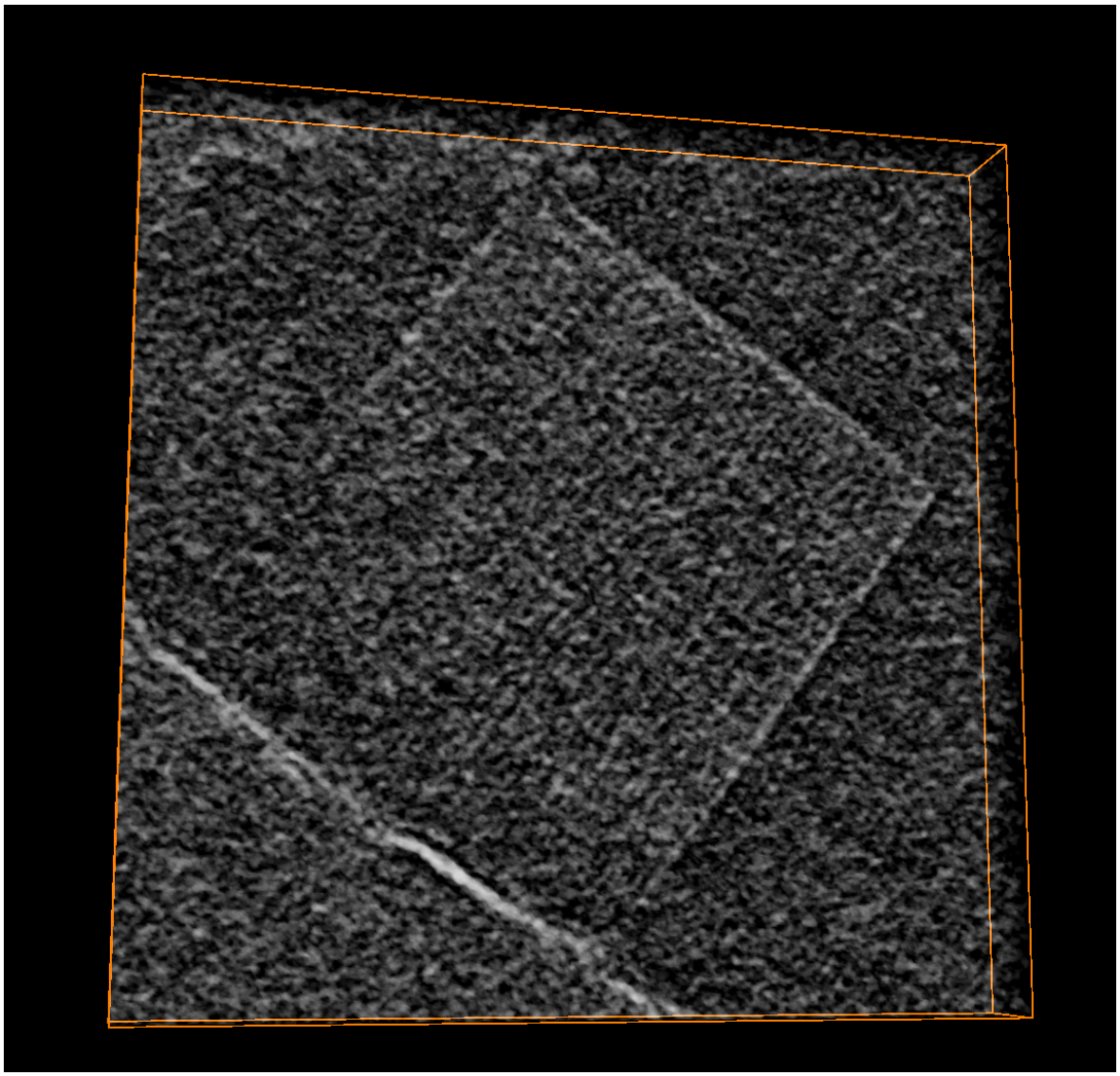

Figure 4.10 | Solid rendering of the whole 3d reconstruction of a lysozyme nanocrystal (D17) thin slice in z. The lower left corner of the crystal is missing and the edge is not as well defined as for the other straight edges.

The image on Figure 4.9 shows solid rendering of the whole tomogram, including background-containing non-crystalline sections above and below the object. The lattice is partly visible, but in general the contrast is weak. The image on Figure 4.10 is a thin z-slice through the reconstruction. Here the lattice is slightly better visible. Segmenting out the crystalline part of the image with a smooth envelope (how it was achieved is described in section 4.4.4) and calculating the FFT paints a clearer picture (Figure 4.11). The strong peaks in the frequency spectrum reaffirm that there is indeed a periodic aspect to the density.

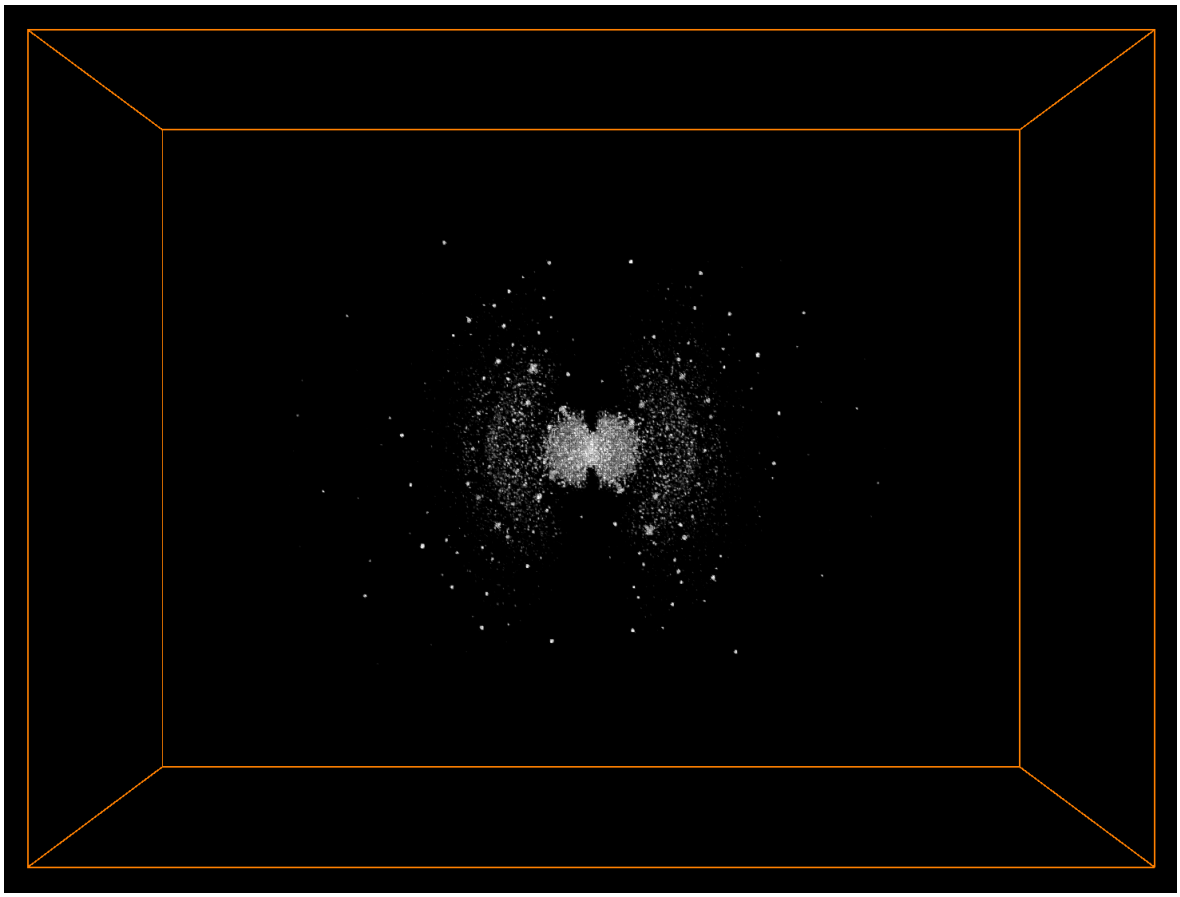

Figure 4.11 | 3d FFT of a lysozyme nanocrystal reconstruction. Fourier transform of the reconstruction on Figure 4.9 that has been first segmented out with a smooth envelope. The 3d image is viewed from the z direction (electron beam direction).

Shape of the Fourier Transform

It is a general assumption, consistent with empirical observation, that the high resolution details are the first to disappear due to accumulation of electron dose. In [71], in a subtomogram averaging setting, the authors have used a modified 3d CTF model that includes a tilt-dependent damping factor, that attempts to account for the loss of high resolution details. As the data collection for tomography proceeds, more and more electron dose is accumulated and therefore the high resolution end on the later tilts should be damped more than on the earlier tilts. Using such weighted CTF function improved the resolution in the final averaged structure from 10.2 Å to 9.2 Å [71].

In the current work we have not done such tilt-dependent CTF modification, but we can directly observe the same phenomenon at work looking at the frequency spectrum. Our data collection scheme was to first tilt the specimen through angles 0° to 67°, then return to 0°, and finally image the tilts from −1° through −67°. The Fourier transform on Figure 4.12, viewed along the tilt axis, clearly shows that there is not an equal amount of high resolution signal in the later tilts. The number of observable reflections considerably lower in the top left and bottom right quadrants, and the whole

diffraction pattern appears skewed. Looking at the diffraction pattern perpendicular to the tilt axis shows no such tendencies (Figure 4.13).

The loss of resolution caused by the radiation damage together with the decrease in resolution in the beam direction caused by the missing wedge is problematic for the data processing. Even though the high symmetry in the data set can and will alleviate the missing information, it makes the determination of unit cell vectors and the crystallographic origin much less certain, as further investigated in the sections to come.

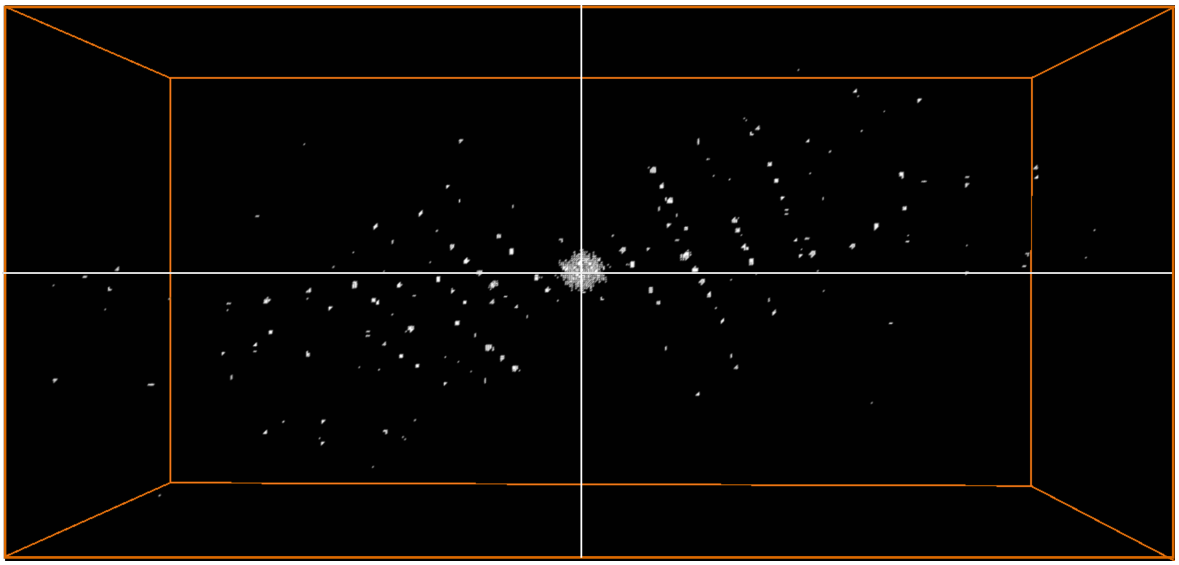

Figure 4.12 | FFT of D17 viewed along the tilt-axis (x-axis). The beam direction (z-axis) is vertical and y-axis is horizontal. This image demonstrates the skewness of the signal. The high resolution spots extend much further out on the first half of the tilt series than the second. The white cross over the image is not an artifact but drawn deliberately to aid the eye in seeing the difference in the number of spots in the respective quadrants.

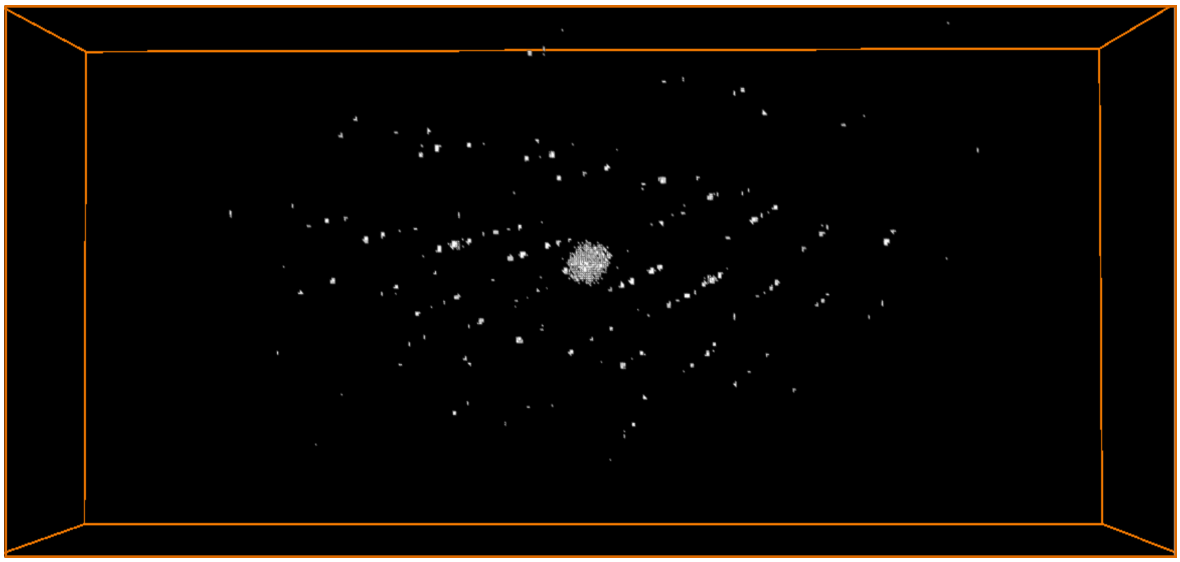

Figure 4.13 | FFT of D17 viewed along y-axis. The beam direction (z-axis) is vertical and x-axis horizontal. We see no no unevenness of the diffraction pattern (as in Figure 4.12) from this angle.

4.4.2 Peak Detection

To harness the replication of information in a crystal to our advantage we have to first figure out what the lattice of the crystal is. The lattice, defined by 3 vectors in 3 dimensions, gives the knowledge in which direction and at which step sizes the translational symmetry occurs. The Fourier transform of an image that has some feature repeating over and over again at a fixed distance will show an intense peak at a point that corresponds to the given spatial frequency. So the strategy becomes clear - to determine the lattice we have to first Fourier transform the whole image, detect the diffraction spots and find 3 such vectors that will form a basis in which the lattice points all have integer coordinates.

In our approach the lattice was determined from a filtered back-projection map reconstructed to full Nyquist spatial frequency, which in this case corresponds to 4.522 Å, i.e. twice the voxel size of 2.261 Å. The map was subsequently Fourier transformed using periodic+smooth decomposition [64] to alleviate the edge artifacts stemming from the discontinuities at the borders of the 3d image. Scattered density and the lack of low resolution information in such a map provided clear diffraction peaks in the FFT which were easier for the peak detection algorithm to pick up. The background of FFT was reduced by subtracting a low-pass filtered version of the transform (inspired by [72]), then thresholded and finally the diffraction peaks were identified by considering local intensity maximum within a spherical neighbourhood.

It should be noted that even though back-projection does not have the CTF effects removed, this does not alter the location of peaks but just their intensity as the peaks could lie close to where the CTF zero-crossings are. Moreover, there is a possibility that due to each image in the tilt-series having a slightly different defocus, many of the diffraction spots would show up, however faintly, even when the average CTF would suppress such peaks.

However refined the approach, in a low-dose cryo-ET setting the peak detection inevitably leads to a large amount of falsely detected noisy peaks along with the true lattice (Figure 4.14 top left). By setting the threshold for cutting out the background too high, we detect only the central low resolution peaks, and by thresholding lower, we detect most of the true lattice peaks, but along with it we have to somehow deal with getting rid of the noisy ones. So, to be of any use in the general case, a lattice determination method that is robust to noise is therefore crucial.

4.4.3 Lattice Vector Determination

To clean up the cloud of peaks and extract the underlying lattice we experimented with various ideas but eventually the one proving the most robust was to repeatedly apply Fourier transform on the cloud of peaks. The lattice detection algorithm iterates the following steps

1. Create a binary map with 1 where a peak was detected and 0 elsewhere

2. Calculate the amplitudes of the FFT of the binary map

3. Take the coordinates of m most intense voxels as new peaks

4. Repeat

At every iteration this approach gets rid of non-periodic peaks and usually after 2 iterations only the periodic component from the noisy cloud of peaks, i.e the lattice, is all that remains. Our approach proved to be incredibly robust and often works even when the majority of the initially detected peaks were noise. As long as there are enough points repeating in space at a fixed interval the underlying lattice will surface.

On Figure 4.14 the lattice determination algorithm from a cloud of peaks is presented graphically. The algorithm starts from a list of peak coordinates, and then iterates the previously described steps a fixed number of times. We can see that in the current case the algorithm has performed marvelously. To be fair, in this example the input list of peaks is much noisier than could have been produced by background subtraction and thresholding in a more careful way. But it was decided to take a particularly bad scenario to clearly demonstrate the power and robustness of this approach.

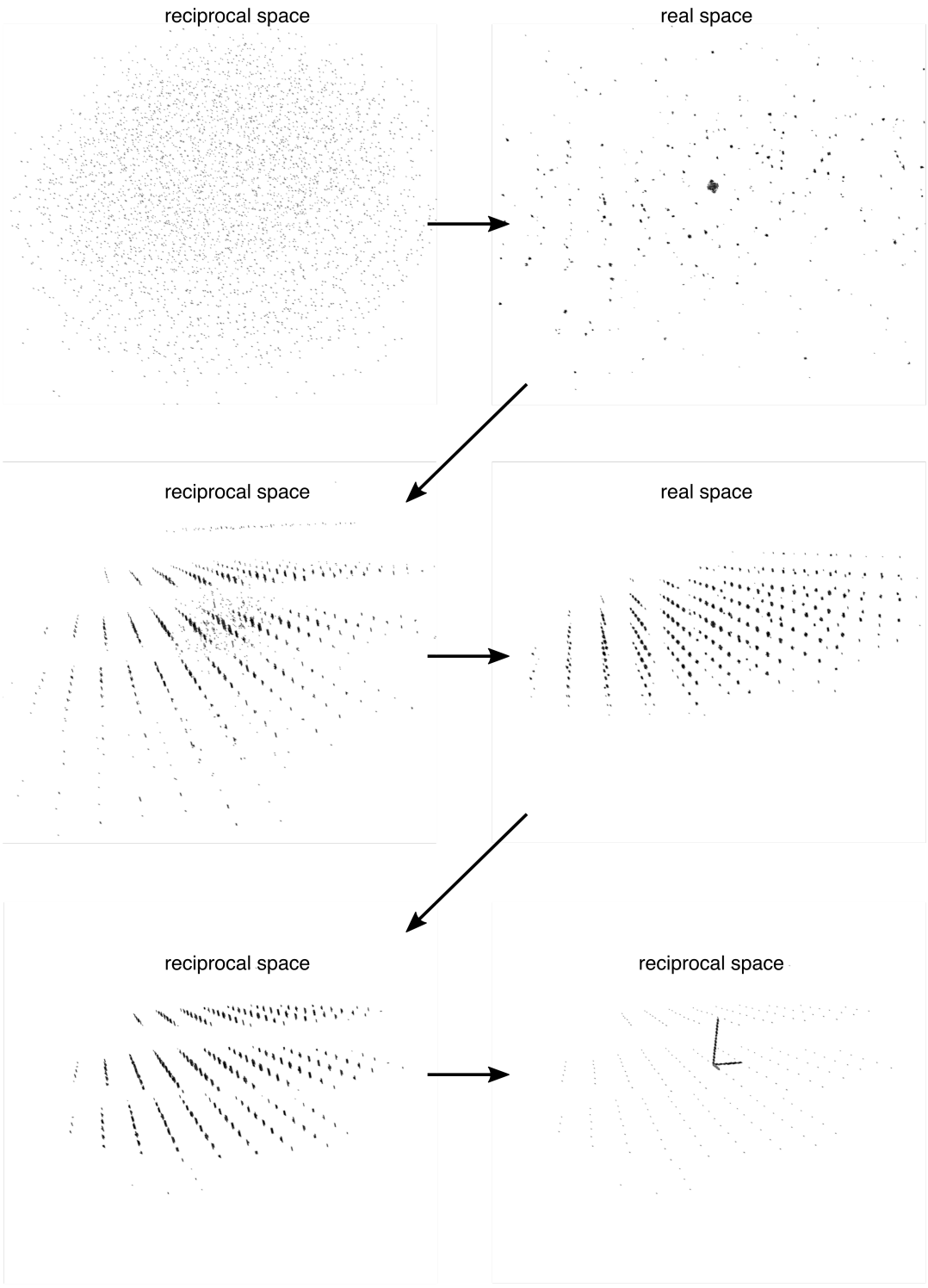

Figure 4.14 | Lattice determination from a cloud of peaks A graphical depiction of the algorithm at work.

After the lattice determination algorithm has finished and returned a good first guess about the 3 lattice vectors, a refinement of the result has to be performed. The

refinement is necessary since due to the discrete nature of the problem, the vectors are confined to have integer coordinates. In reality, a better solution probably always exists, and to find that, we perform a refinement step (further elaborated in the programs section 4.5). After the refinement, we have 3 real space vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ that describe the average periodicity in the reconstruction.

The final determined unit cell vectors were with lengths $a = 79.25\,\text{Å}$, $b = 80.28\,\text{Å}$, $c = 39.60\,\text{Å}$, and angles $\alpha = 90.77$, $\beta = 90.70$, and $\gamma = 89.36$, which are approximately in concordance with the tetragonal form and space group $P4_32_12$, as commonly found in the literature.The fact that lattice vectors a and b are not exactly equal and that the angles between the three are not exactly 90°, as required by the $P4_32_12$ space group, could be attributed to the broadening of the lattice orientation and spacing distribution during flash freezing [73], so the crystal could be under stress and the found lattice only reflects the average. The correctness of the lattice detection algorithm was validated on a dataset simulated from a PDB file (reference 1DPX) where the derived lattice vectors were correct within a fraction of an Ångstrom (described in section 4.6.3).

4.4.4 Averaging and Symmetrization

Averaging

Once the lattice vectors are determined, the averaging of the unit cells can begin. Initially, having no idea about the crystallographic space group, the obvious first step is to assume space group $P1$, i.e. just translational symmetry in the crystal with an arbitrary origin. By placing the origin of the lattice at the origin of the reconstructed volume coordinate system, we extract and averaged every single unit-cell-sized piece of the volume that fits entirely into the reconstructed volume (Figure 4.15).

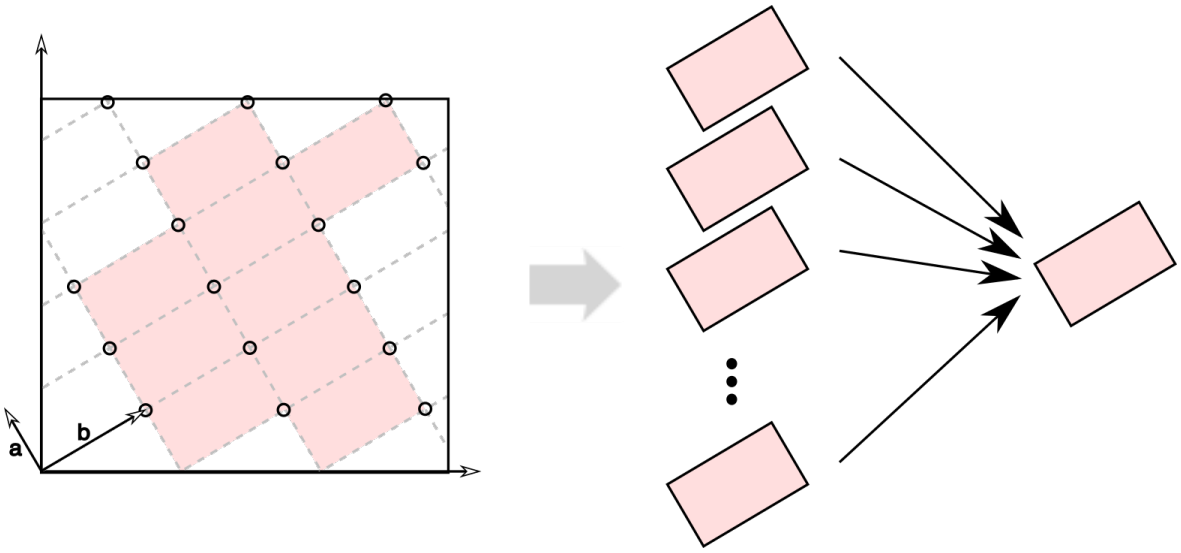

Figure 4.15 | Schematic of the averaging over lattice. All unit cells that fit entirely into the reconstruction volume are extracted and one total average is retrieved. At this point the crystal is not yet segmented out so the average includes the background as well as the crystal.

As the crystal has not yet been segmented out from the background, the total average includes everything, even the non-crystalline parts of the reconstruction (Figure 4.16a). In the next step we calculate the correlation coefficient between the total average and each individual extraction. The rationale is that the background does not conform to the lattice and therefore its total contribution to the average is just a background blur. The crystalline part, on the other hand, will get reinforced by the $P1$ averaging, and some characteristic details of the real unit cell will hopefully raise above the background level.

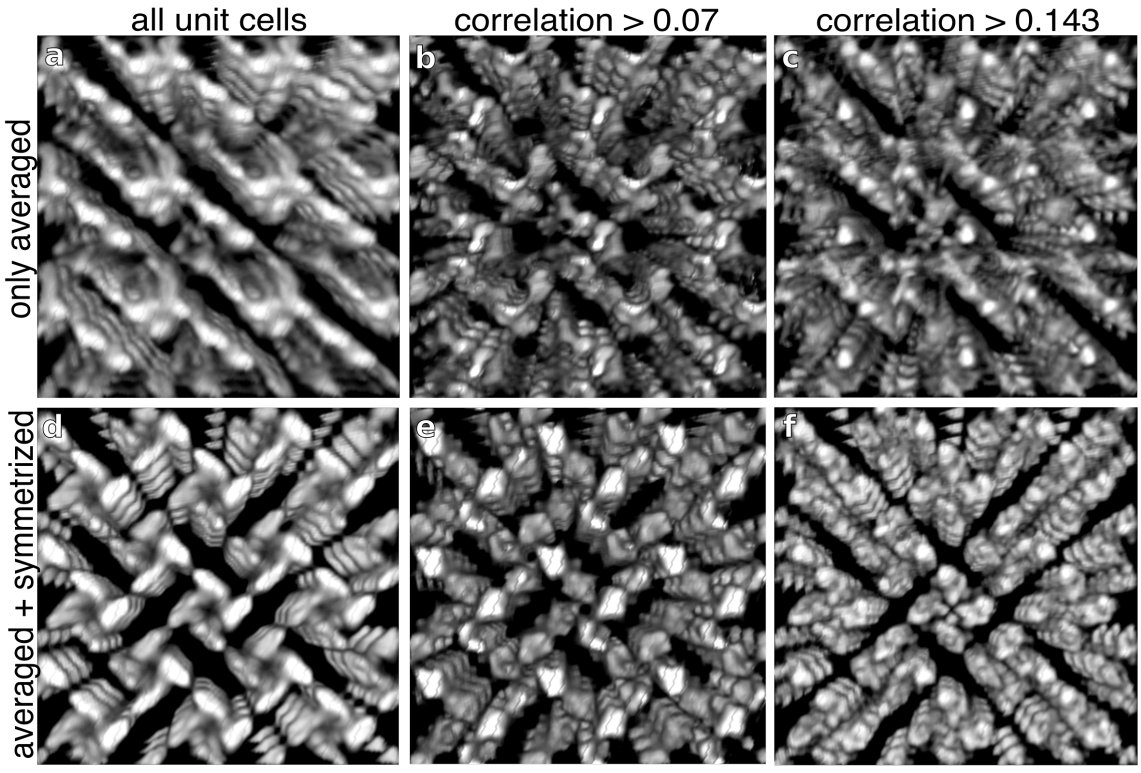

Figure 4.16 | Averaged and symmetrized unit cells differing by the subset of unit cells used. **a and d** - all possible non-overlapping unit-cell-sized subvolumes, that also includes the background, averaged and symmetrized. The result is blurry, but with the lattice-like repetition clearly visible. **b and e** - the unit cells that roughly constitute the crystal, and probably do not have any unit cells from the outside included, as the segmentation at 0.07 correlation level is on the conservative side. **c and f** - the unit cells that correlated most highly with the total average on **a**. The non-symmetrized image (**c**) gets increasingly more scattered and harder to interpret, whereas the symmetrized version (**f**) looks clean with the individual molecules nicely visible. The molecules are less well separated in the ***c*** direction (coming out of the plane of the paper).

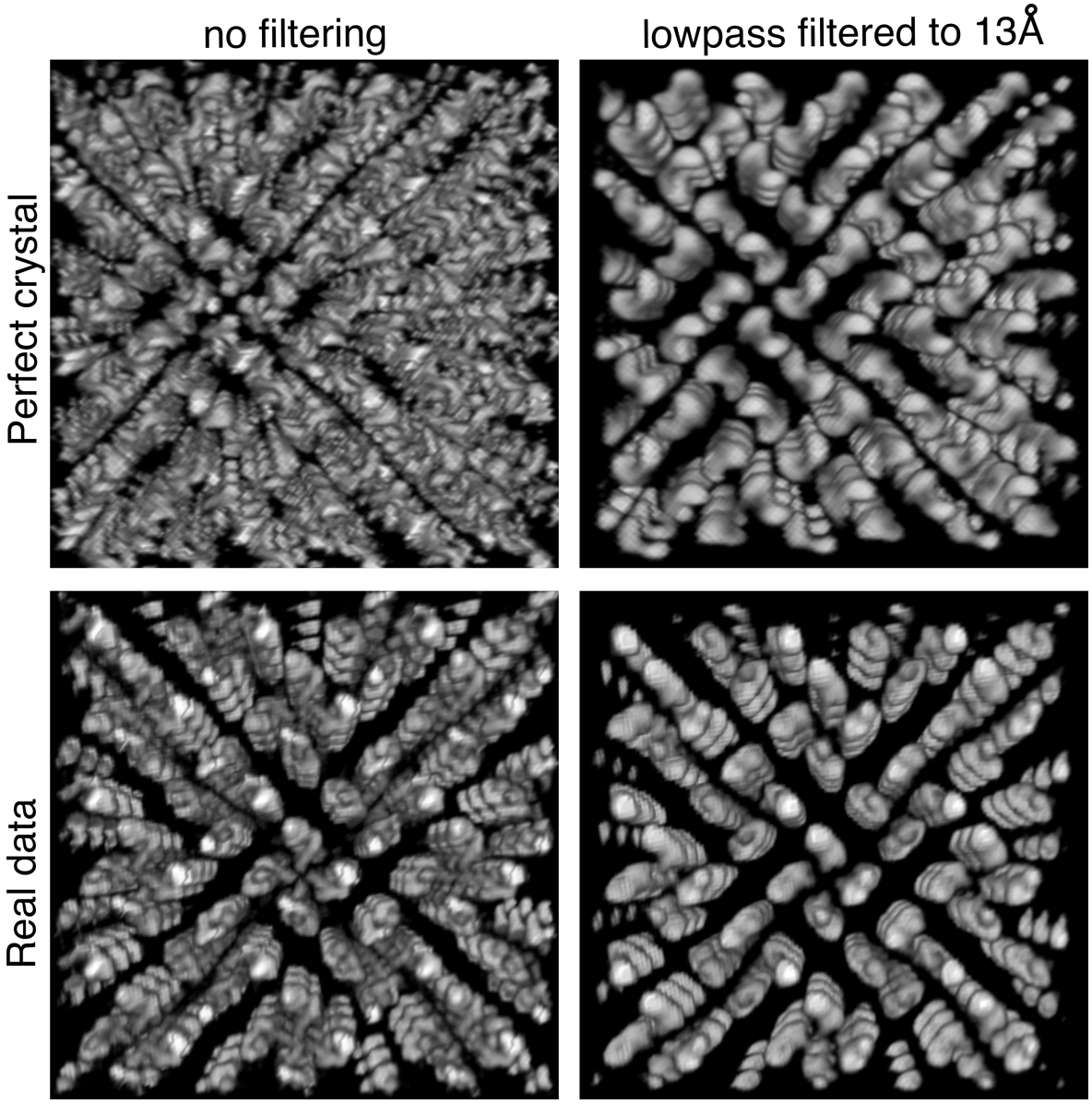

Figure 4.17 | Averaged and symmetrized unit cells. Comparison of electron density maps between real data and simulated data from the x-ray structure (pdb code 1DPX). The non-filtered simulated crystal has details up to Nyquist frequency whereas real data has much lower resolution. Low-pass filtering both of the maps to 13 Å shows that the maps are indeed quite similar.

If we plot the obtained correlation coefficients on a histogram then usually two modes emerge (Figure 4.18a) - one mode at a correlation value slightly above zero, corresponding to the cells from the background, and the other mode at some higher level, corresponding to the crystal. Some aspects to note about the correlation histogram is that, first, the correlation values are quite low, and, second, the background mode is not centered exactly around zero but instead around a slightly higher value. The former, we suppose, is due to the inhomogeneities in the crystal, and the variation of the intensity levels and the slight inhomogeneities in the crystal. The latter

aspect could be explained by the fact that since each cell contributes something to the average, then it is expected that each cell also has a small positive correlation value with the total average. For a perfect virtual crystal, that was used for validation, the correlation values are all close to 1.

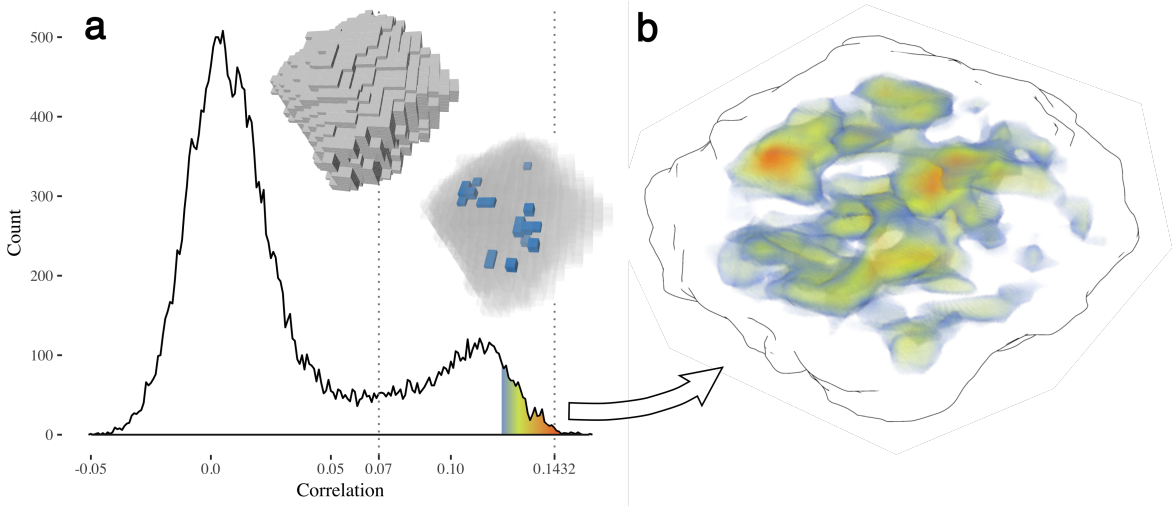

Figure 4.18 | Histogram of the correlations and the values mapped back to 3d. **a**, All unit cells in the whole reconstructed volume correlated against their average. The correlation profile is bimodal with the first mode close to 0 representing the background and the second mode at around 0.12 representing the crystal. Setting an arbitrary threshold at 0.07 selects $\sim$ 5000 unit cells from $\sim$ 24000 that roughly segment out the crystal from the background at a unit cell level accuracy. A higher threshold of 0.1432 leaves just 28 unit cells that are well aligned with each other and are later used for lysozyme molecular structure determination. The insets show the unit cells kept after selecting only the cells above respective correlation thresholds. **b**, Visualization of the correlation distribution of the crystal obtained by interpolating a 3d correlation function from the sampled points at unit cell centers. The black outline represents 0.07 level that draws out the approximate shape of the crystal. The color gradient is obtained by mapping correlation level 0.12 to blue, 0.13 to yellow and maximum level at around 0.15 to red. For visualization purposes the orientation of this image differs slightly from the crystal shape on **a**.

Segmentation of the crystal can be performed by selecting only the unit cells, that have a correlation value on the right side of the "valley", threshold chosen to be here at 0.07 (Figure 4.18a). The average over only those 5000 unit cells is visible on Figure 4.16b. An even higher threshold at 0.143, that leaves a mere 30 unit cells for averaging, obtains the average structure visible on Figure 4.16c. So a general trend was noted - the smaller the number of unit cells that go into the average, the higher the apparent resolution in the image, up to the point, where there are too few unit cells that the noise starts to dominate again.

Symmetrization

Assuming space group $P4_32_12$ we can make use of the additional symmetries present in the crystal (Figure 4.19). A unit cell in such a case has 8-fold redundancy of information. To take advantage of this redundancy an important parameter has to be determined - namely, the *symmetry origin*. When we performed only the translational averaging over the lattice, then it did not matter where the origin was placed, whereas for applying rotations it is crucial to pinpoint the locations of the rotation axes with high accuracy.

The programs section (section 4.5) describes the details of the symmetry origin finding algorithm, but the essence of the algorithm can be described as follows. In theory, the 8 symmetrically related points within the unit cell are all identical, whereas in practice there are probably variations. We look exhaustively through the whole unit cell for the best place where to put the symmetry origin, so that the discrepancy between the 8 symmetrically related points would be minimized.

The algorithm is not perfect and fails when the signal is too week. Another weak point is that since the resolution in the reconstruction is much lower in the z-directions, the best place for the symmetry origin might not be well determined. This shortcoming is further explored in the section about the unit cell distribution (section 4.4.9).

In the current case we did the symmetry origin search on the average unit cell of the whole segmented crystal (Figure 4.16b). Applying the symmetry relations on top of the translational average over the lattice obtains the maps depicted on Figure 4.16d-f. A single molecule from Figure 4.16f was then sculpted out and the surface rendering was compared to a low-passed x-ray structure on Figure 4.25.

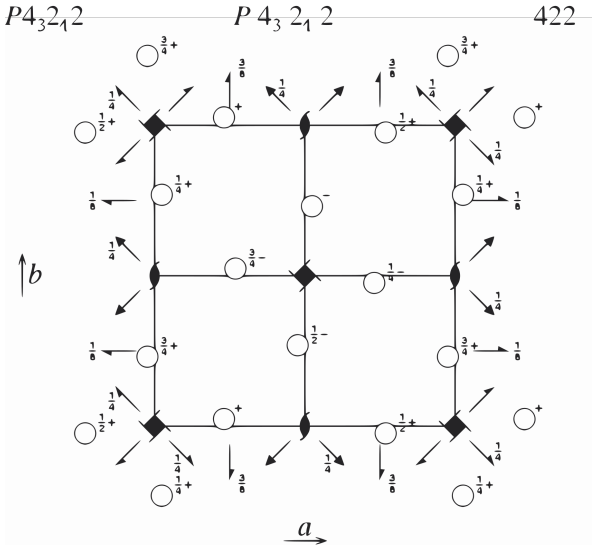

Figure 4.19 | $P4_32_12$ **space group diagram**. A unit cell following $P4_32_12$ space group, features 4-fold and 2-fold screw axes, that replicate one molecule 8 times.

Segmentation

To segment out the entire crystal, we can create a binary 3d envelope that selects the unit cells with correlation above 0.07, and multiply it with the reconstructed volume.

This will render what we consider to be the background to become equal to 0. While this is usable to simply look at the segmented volume, it will become a burden when we want to do something in the Fourier space. The shape transform of the sharp envelope will completely dominate the Fourier transform (discussed in section 1.2.1). To get rid of this problem we low-pass filter the mask by a large amount, which will obtain a smooth segmentation (Figure 4.20).

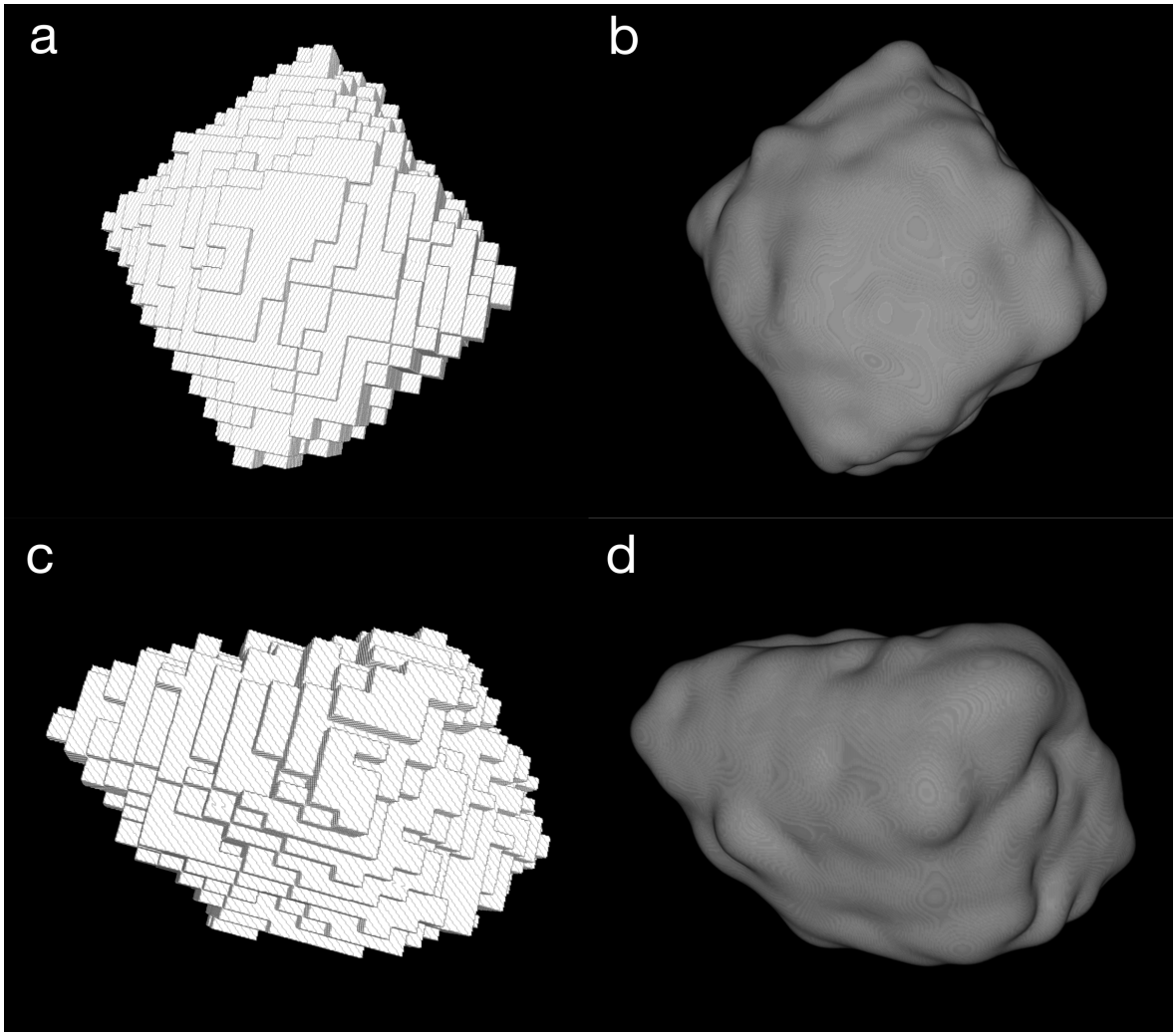

Figure 4.20 | Segmentation of the crystal D17 by correlation thresholding. **a**, Unit cells showing correlation > 0.07 with the total average. **b**, Low-passed version of **a**. **c**, Different view of **a**. **d**, Low-passed version of **c**.

Before the segmentation, the FFT of the reconstructed map showed some visible diffraction peaks, but it was difficult to see the higher order ones as they had very low intensity. Now, with the background removed, the FFT is striking (Figure 4.21. FFT of the FBP map reconstructed to full Nyquist frequency without any low-passing is full of diffraction spots. As the non-crystalline background has been suppressed, many of the faintest spots become visible. There is still some noise in the lower frequencies but it is of less nuisance now.

Usually, in tomography, it is difficult to say something certain about the resolution in the map. Proxies, like Fourier shell correlation, are often used in single particle reconstruction, but the question of where exactly should the FSC curve be cut remains. With crystals, the fact that diffraction spots appear far out in the spectrum, is a direct indication that there are details with a given spatial frequency present in the reconstruction.

4.4.5 Resolution

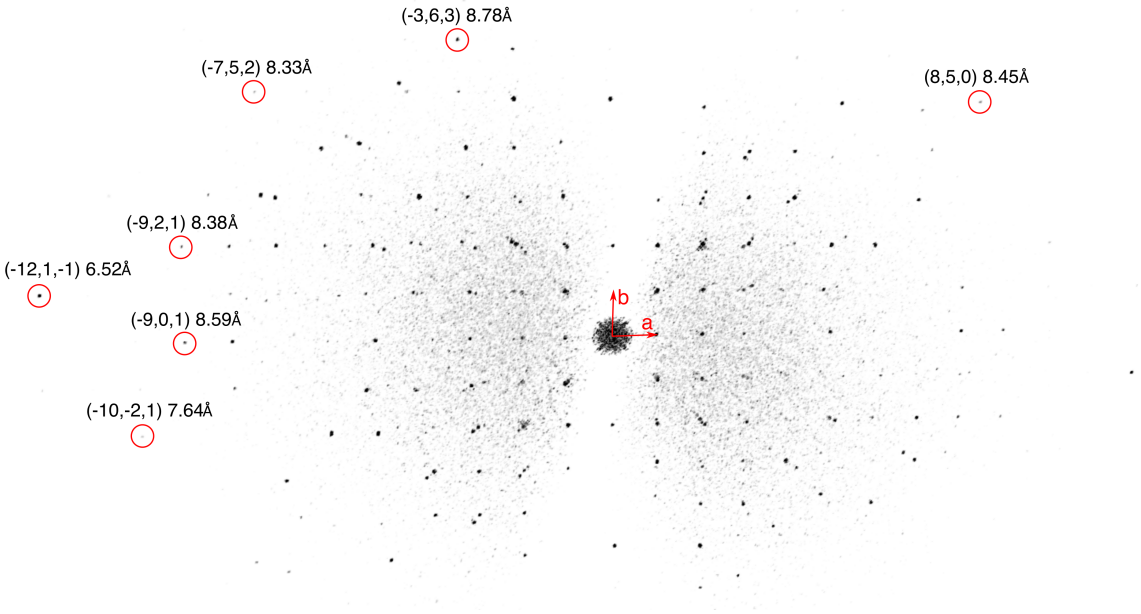

Figure 4.21 | FFT of FBP of segmented crystal D17. The transform shows reflections to at least 10th order and even two 12th order spots.

In electron tomography the question about the resolution in the data is always tricky to answer, because there are many resolution-altering factors acting isotropically or anisotropically in various directions, such as

- **The missing wedge** of information - decreases resolution in the electron beam direction
- **Sample drift** during imaging decreases resolution in the drift direction
- **Radiation damage** from long electron exposure decreases resolution in all directions, starting from the fine detail.
- **Improper CTF deconvolution** inverts contrast wrongly in the high end of the spectrum, and if the tilts each have individual defocus values, then the net effect of using one average defocus value is that the high frequencies will be damped.
- **CTF zero-crossings** removes resolution at specific frequencies.

- One or more **interpolation** steps decreases resolution in all directions.

Often there is also a lack of an objective reference structure against which to compare the results. For resolution assessment in averaging-based methods the Fourier Shell Correlation (FSC) has been traditionally used. FSC is applicable when many supposedly identical objects are averaged together, such as in single particle reconstruction and subtomogram averaging. In FSC the total pool of objects to be averaged are split randomly into two, then each of the splits is averaged independently, their respective Fourier transforms are divided into shells corresponding to resolution intervals, and finally the corresponding shells are correlated against each other. This usually results in a downward sloping curve indicating that the low resolution details of the structures are in good agreement, but the smaller the details get, the lower the agreement becomes, until after some resolution the details are effectively random. Eventually, there is a question of where to draw the line. Usually the resolution is assessed by setting a threshold, say 1/2 or 1/7, and observing at which spatial frequency the correlation drops below that value. This method has a certain dose of arbitrariness built in and is not applicable when the final result is just a single reconstruction.

In the nanocrystal case the Fourier shell correlation is definitely applicable as here, too, many unit cells and molecules are averaged together. But in addition to FSC a more direct and objective measure exists. The diffraction spots in the Fourier transform of the original non-averaged reconstruction are a clear indication of the resolution in various directions.

The original reconstruction includes large amounts of non-crystalline volume that manifests as intense background in the FFT that obstructs the view and drowns the spots. FFT of the segmented crystal, on the other hand, has greatly increased clarity and an interpretable lattice (Figure 4.21). It is remarkable that there is evidence of such high resolution details. In various dimensions the resolution reaches around 8.5 Å and some of the highest peaks go up to the 10th order, corresponding to 7.6 Å. On D17 there even exists a 12th order peak, corresponding to 6.52 Å, and some faint, but distinguishable if contoured properly, 11th order peaks.

It must be noted that this high resolution is roughly only present in the x-y directions while the resolution in z-direction is greatly reduced, mainly due to the missing wedge. Thanks to the high symmetry in the crystal and the fact that the symmetry axes do not coincide with the laboratory coordinate axes, the rotation operations can fill in a large portion of the diffraction pattern that is missing, thereby making the resolution more isotropic.

A particular weak point in our algorithms is that, if we have one object with low resolution where another object has high resolution, and high resolution on the first object where the other has low resolution, then their average will be uniformly medium resolution. This is the way our programs are implemented currently and this is reflected also in the FFT of the symmetrized map, where after symmetrization the resolution has become much more even in all directions, but we have lost the high end of the spectrum (Figure 4.22). The solution would be to, instead of naive averaging, use some kind of smart composition that would retain the high resolution from each image. This has yet to be implemented.

On Figure 4.22 we also see how the missing wedge gets partly filled in by sym-

metrization, at the expense of loosing high resolution information elsewhere. This could and should be avoided in the future implementations.

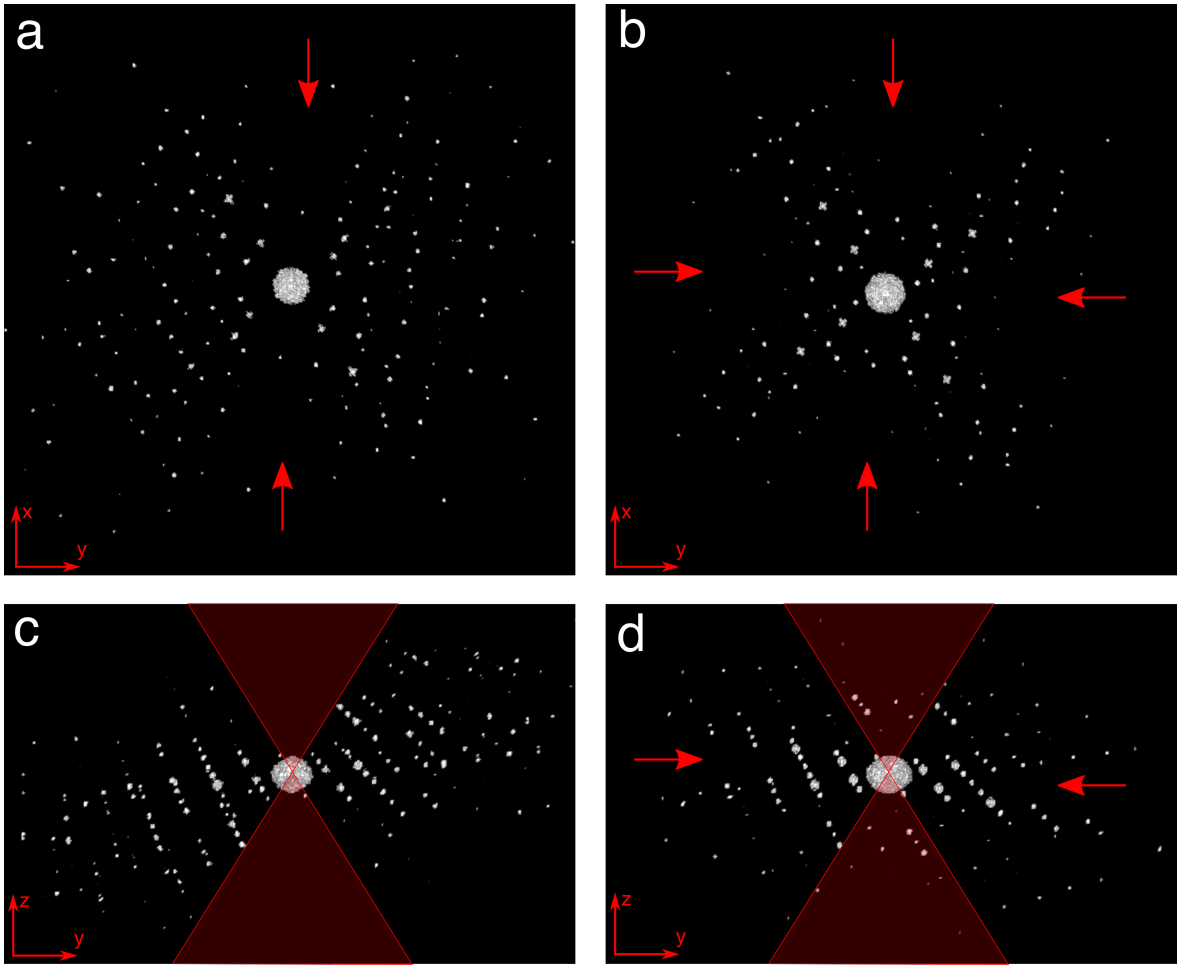

Figure 4.22 | FFT of back-projection locally averaged and symmetrized. The averaging is done locally only over 2 neighboring unit cells in each direction. The symmetrization is done using the same symmetry origin for all unit cells. The tilt-axis in data collection was x, therefore the missing wedge is visible when looking from x-direction. Red arrows indicating data that has gone missing. **a** Averaged only, view from z-direction. **b** Averaged + symmetrized, view from z-direction. **c** Averaged only, view from x-direction. **d** Averaged + symmetrized, view from x-direction.

4.4.6 Lysozyme Structure

Non-averaged reconstructions of the crystals in general have too low SNR and contrast to distinguish individual molecules. One major advantage of a nanocrystal is that it acts as a powerful guide leading directly to the molecular densities. If we take a look inside the crystal then we can identify many chunks of density that upon closer inspection strongly resemble a protein molecule. As the atomic structure of lysozyme is known, many of the found individual molecules resemble the topology of the atomic structure to a staggering degree. Possibly some reference based automatic algorithm

could extract out hundreds of individual molecules, each deformed in its own way due to various reconstruction artifacts, anisotropic resolution, radiation damage, as well as conformational variability. Distinguishing between those sources of error is difficult.

After selecting for the most highly correlating unit cells, averaging, and symmetrization, we arrive at a rather clean looking unit cell from where a single molecule can easily be seen and extracted. The general problem is that even though the molecule can be seen, when it is packed tightly in a crystal, it is hard to tell where exactly does one start and end, and which densities belong already to the neighboring molecule. Therefore, unless the resolution is really high, or the molecule is large (unlike lysozyme), or has a favorable and easily identified shape, the chance for interpretation error cannot be neglected.

The structure appears to be at a medium resolution of around 13 Å as judged by visual comparison with the simulated electron density from atomic coordinates. The reasons for not being able to obtain as high resolution as evident from the processed diffraction pattern (Figure 4.21) are probably the following

- The high resolution diffraction spots were present only at certain specific directions. Most notably, the direction parallel to the optical axis of the electron microscope has strongly decreased resolution, and therefore a simple symmetrization can only achieve resolutions somewhere in between.

- Regularization is a compromise - we lose some resolution, but gain on contrast and the interpretability of the density map.

- Each individual tilt in the tilt-series is likely imaged at a slightly varying defocus, therefore convolved with a different CTF. On top of that, there is a defocus gradient through a relatively thick sample, and even more so in an image taken at a higher tilt angle. Our regularization and CTF deconvolution procedure is not taking this variation into account. It assumes one particular CTF, determined by the input parameters, and deconvolves that. Clearly this is not optimal, but is certainly less computation-heavy. Implementing the above-mentioned additions are planned in the future developments of the procedure.

Structure from the Back-projection

It is well known that the image reconstructed by back-projection alone can be interpreted only to a relatively low resolution of around 40 Å, because of the CTF-caused contrast reversal at higher resolutions, the dominance of noise, and the inaccurate molecular volumes due to spread out density. Nevertheless, for educational purposes we went ahead and took FBP map reconstructed to the Nyquist frequency, picked out highly correlating unit cells, averaged, and subsequently symmetrized them (Figure 4.23).

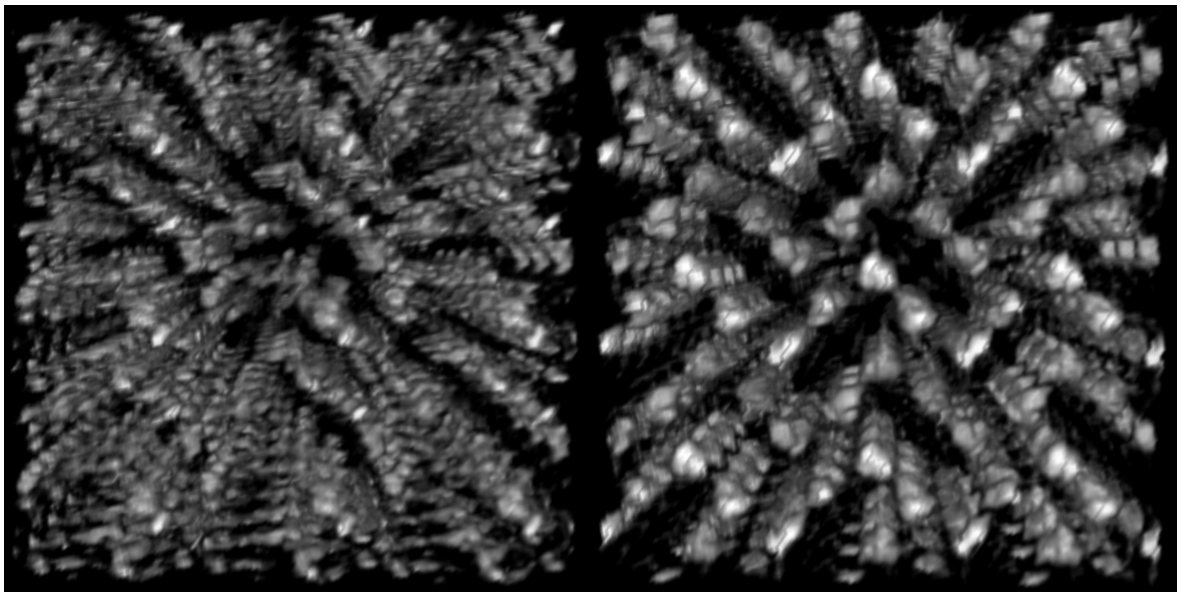

Figure 4.23 | 3x3x3 averaged and symmetrized unit cells of FBP map. **left** - average of the highest correlating unit cells. **right** - symmetrized version of **left**.

On Figure 4.24 we can see a single molecule picked out from the symmetrized unit cell (in green) side-by-side with the density representation of an x-ray structure (in red). The back-projection structure has been scaled up to match the size of the model. Even though some of the important details are missing, the overall similarity is there. The missing structural elements are presumably due to the CTF effects.

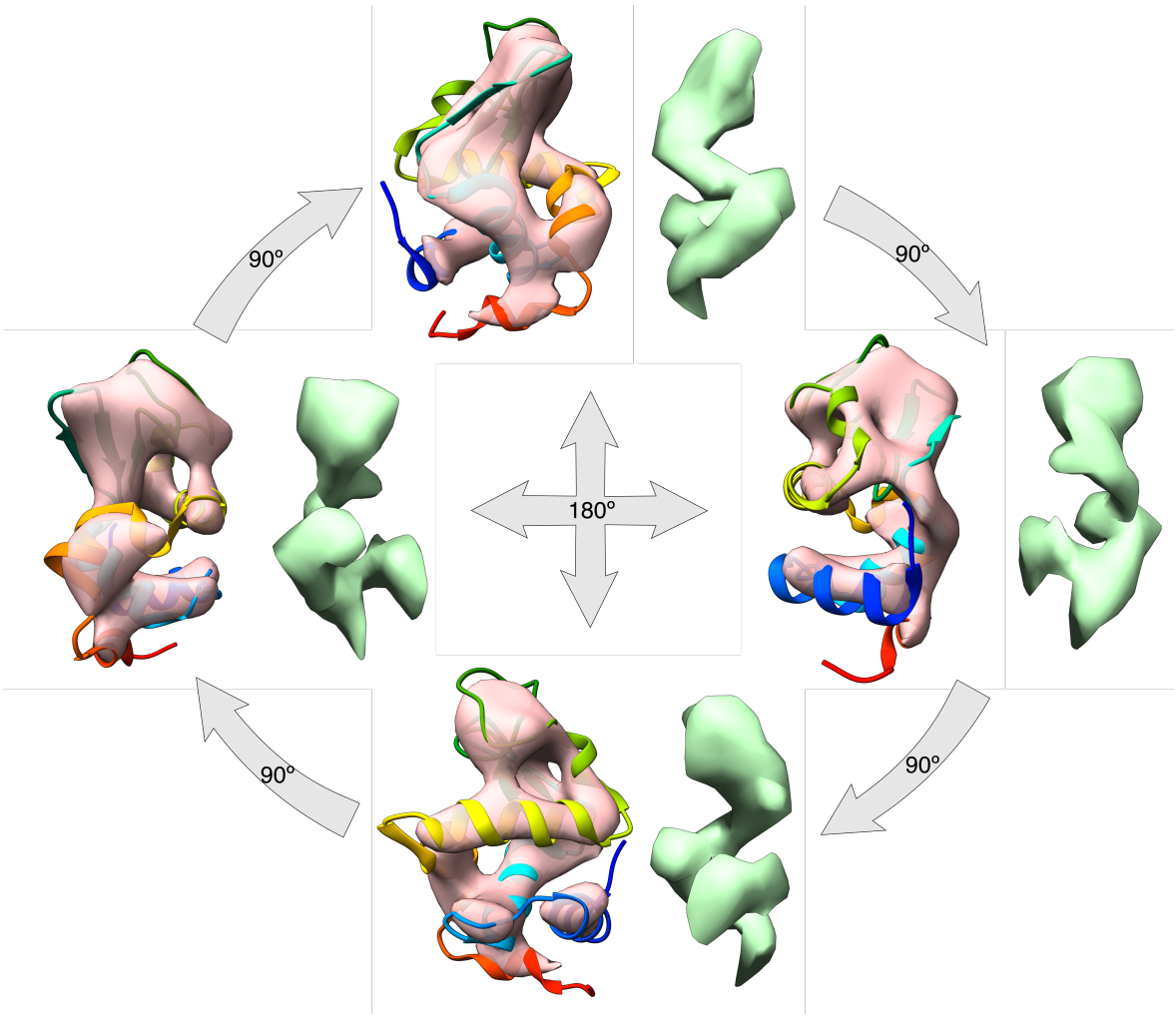

Figure 4.24 | Averaged and symmetrized lysozyme structure from back-projection. 4 different views, each 90° apart, of a suspected true structure of lysozyme (taken from PDB code 1DPX) in red color and the one reconstructed form back-projected map with no contrast function deconvolution in green. The resolution is high but some obvious details are missing, presumably due to CTF effects. Also, the reconstructed structure was much smaller but has been scaled up here to match the size of the model.

Structure with CTF deconvolution

Figure 4.25 shows a surface rendering of an individual molecule extracted from Figure 4.16f. As compared to the structure from back-projection alone (Figure 4.24), the current structure had correct volume and seems to be more complete. It however still does not agree fully with the x-ray structure even at this medium resolution level. Any number of reasons could explain the differences

- Radiation damage
- Reconstruction errors, including errors in the CTF parameters

- Errors in the symmetrization procedure
- Errors in the lattice and the symmetry origin
- The molecules are actually different

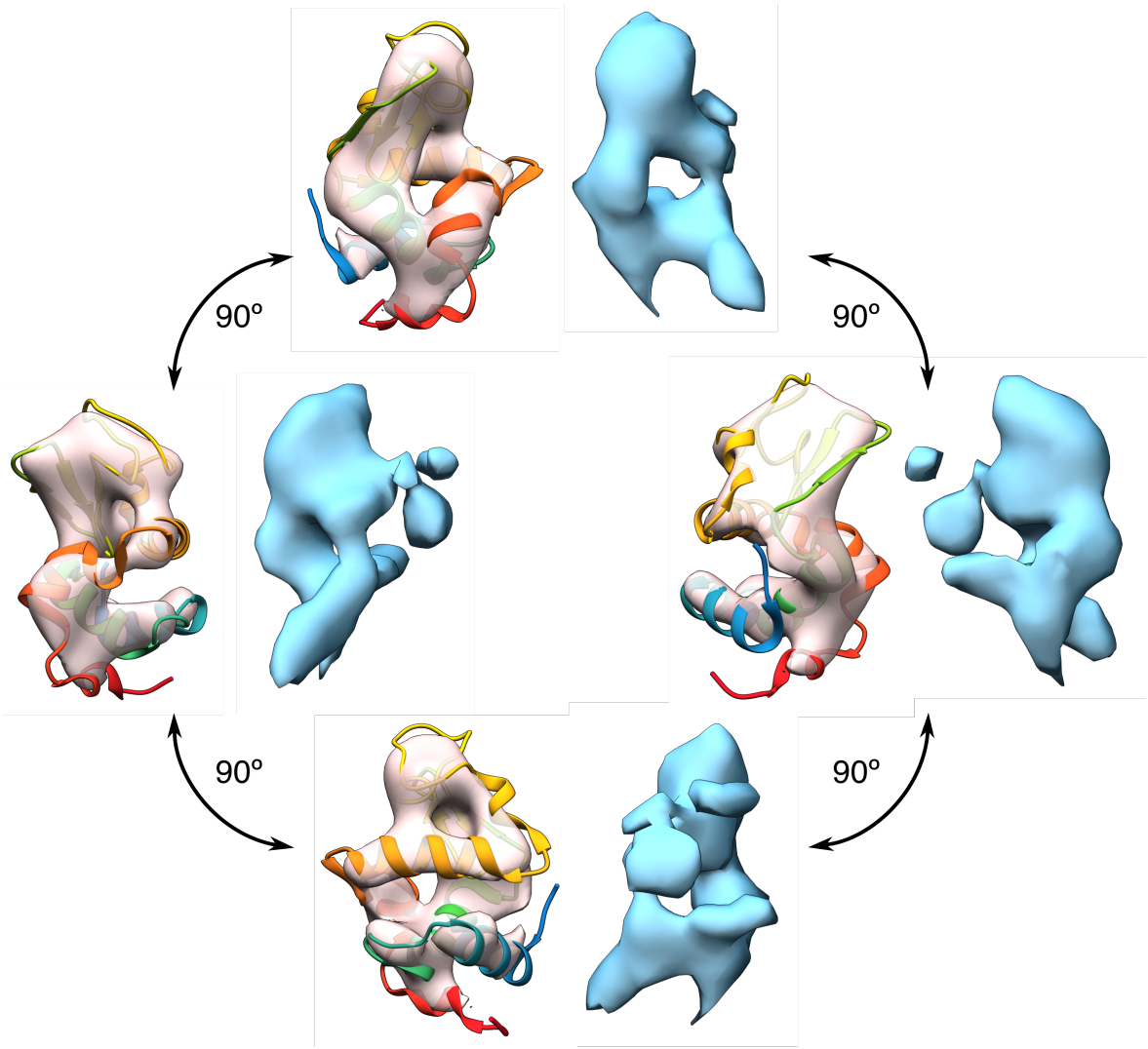

Figure 4.25 | Averaged and symmetrized lysozyme structure from 28 highest correlating unit cells. 4 different views, each 90° apart, of a suspected true structure of lysozyme (taken from PDB code 1DPX) in red color and the one reconstructed form real data in blue. This structure matched the scale well. Although the similarity between the two is noticeable, there are definitely several mismatches which could be attributed to number of reasons.

4.4.7 Local Symmetrization

We tried to get a glimpse of the variation in the lysozyme nanocrystal by not averaging and symmetrizing over the whole crystal, because that would result in a less interesting

global replication of a single average unit cell, but instead average and symmetrize locally. Replacing every unit cell in the large reconstruction with the average of its neighbours and applying symmetry operations gave a great enhancement in the signal, reduction in noise, and brought out the differences as they propagate over the crystal.

On Figure 4.26 we can see 3 different views through a nanocrystal, that has been locally averaged and symmetrized over 125 unit cells, that is 2 neighbors in each direction for each unit cell. The symmetry origin, however, is the same for each unit cell. We can definitely see some variation in intensity and variation in structure, presumably due to the fact that the average lattice parameters used here are not ideal locally, so the crystal goes in and out of sync with the average lattice.

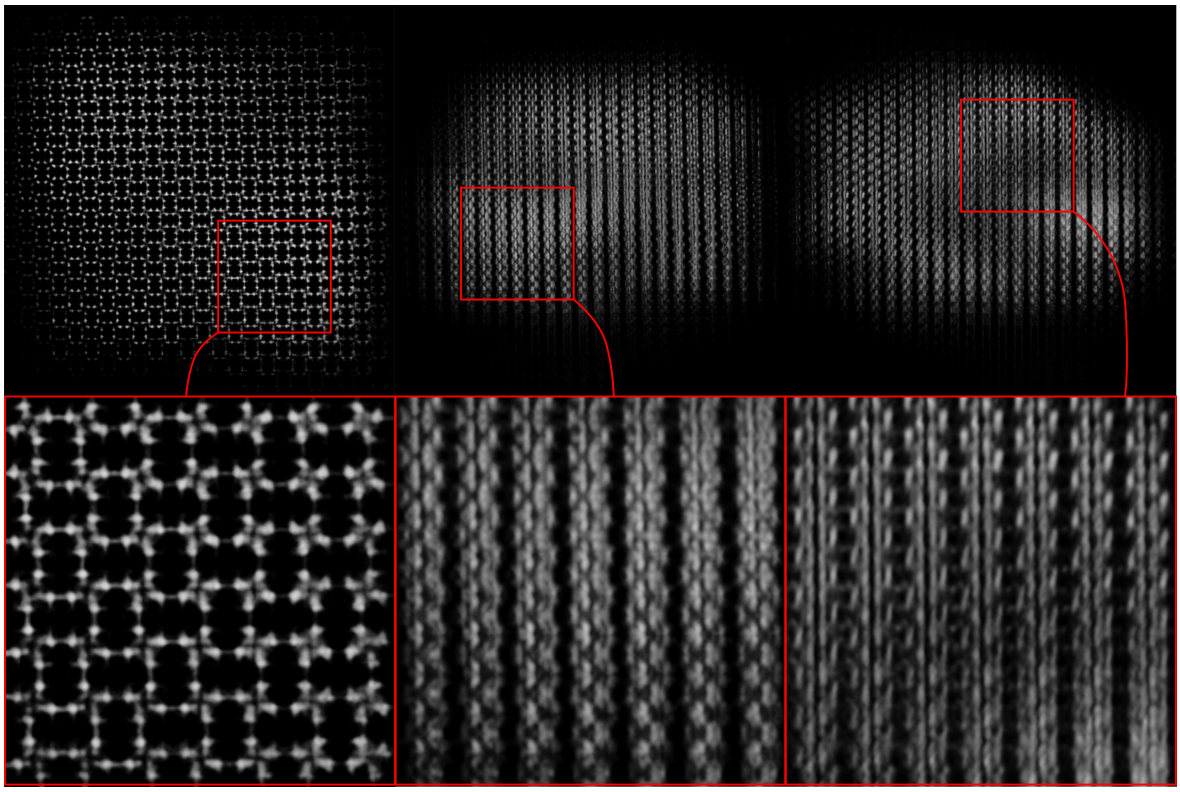

Figure 4.26 | Locally averaged and symmetrized crystal. **Left** - top view from ***c*** direction. **Center** - side view from ***b*** direction. **Right** - side view from ***a*** direction.

4.4.8 Cluster analysis

When averaging many objects, it is crucial to make sure that only similar ones get averaged together. Failing to achieve that, might result in an object that does not make sense physically, or might lead to wrong conclusions, depending on the magnitude of variation.

For the crystal D17 there was a growing suspicion that it might be composed of compartments of varying morphology. As evident from the previous sections, the less unit cells we included into the final average, the sharper the average image seemingly got, up to the point where the structure broke down, because too few unit cells in the

average did not suppress enough noise. A cluster analysis should be able to detect blocks of identical morphology, provided they exist.

The idea was to do clustering at a unit cell level, as it fits in nicely with the rest of the analysis framework we have developed. The previously determined lattice vectors were used to define the shape of the unit cell and the correlation-based segmentation gave the rough shape of the crystal.

We cut out 5104 unit cells, each of them consisting of 21880 voxels. Placing the objects into a data frame $\boldsymbol{X}$ thus gives a matrix with 5104 rows and 21880 columns. To say anything meaningful about the similarity of the objects, the dimensionality of the problem has to be reduced, because in high dimensional space odd things can happen and almost everything is equally distant from each other. A conventional method for dimensionality reduction is using *principal component analysis* (PCA).

Principal Component Analysis

The essence of PCA is to derive new features (principal components) as linear combinations of the existing ones (voxels), with the requirement that each component captures as much variation in the data as possible and the components themselves would be mutually uncorrelated. The matrix $\boldsymbol{X}$ in this case was still small enough to not require any clever treatment for computing PCA on a desktop computer, or a truncated PCA that would only try to retrieve the k first principal components.

Full PCA was computed using statistical computing environment *R*, and it took around 30 min to complete. The pairwise scatter plots of 5 of the first components showed just a random cloud of points, no hints of clustering or structure, and mapping the values of those components back to 3d reconstruction space showed no identifiable pattern. The suspicion was that the signal in the images was too weak to find meaningful principal components.

Locally Averaged Unit Cells

The next idea was to do some pre-averaging prior to PCA. Just like before, we assume that the neighboring unit cells do not differ from each other too much, and therefore it makes sense to average them. So, instead of extracting each unit cell individually, we performed local averaging at the unit cell level, by replacing each unit cell with the average of its immediate neighbors. Each unit cell thus became the average of 27 unit cells in its immediate neighborhood, including itself.

After local averaging the signal in the dataset was stronger, but the concern was that the objects were not independent anymore, which might introduce artificial similarities between the unit cells. To remedy this, we found the largest such subset of unit cells that have at least two unit cells between each member (Figure 4.27). In other words, we selected only the unit cells with *hkl* indices $(0,0,0), (0,0,3), (0,3,0), \ldots$ so that all of the objects in the dataset were independent ones and none of them shared neighbors. This left 197 averaged unit cells into the analysis.

PCA After Local Averaging

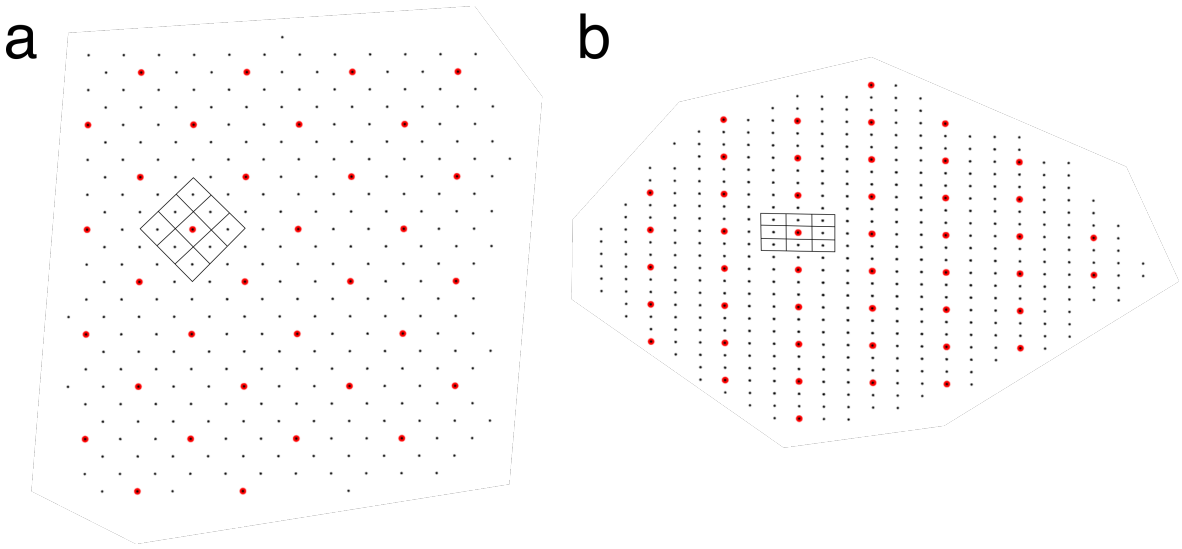

Figure 4.27 | Independent unit cells that were used for PCA. The red dots show the unit cells centers that were used for performing principal component analysis. The red grid has exactly two lattice points between each point and allows enough space to do local averaging in the immediate neighbourhood without creating artificial dependencies between the unit cells. **a** - View from $\boldsymbol{c}$ direction. **b** - View from $\boldsymbol{a}$ direction.

Calculating the principal components now took only seconds. The explanatory power of the derived principal components is illustrated on Figure 4.28. On the scree plot (Figure 4.28a) we see that there is a slight "elbow" at the 10th principal component. This observation is sometimes used as a selection rule as to how many components should be kept for further analysis. In the current case the first 10 components cumulatively explain only around 16% of the total variance in the data, which might not be enough. We decided to continue the analysis with 100 first components, that cumulatively explain 72.5%.

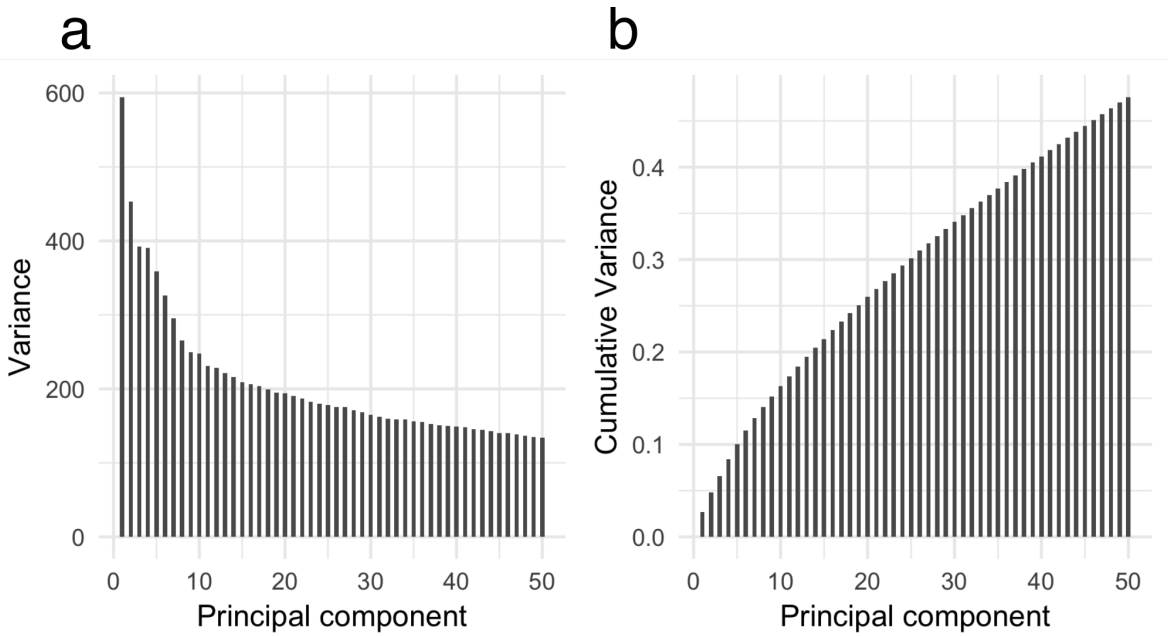

Figure 4.28 | Variance explained by the principal components. **a** - shows the so called *scree plot* of the principal components. The height of the bar shows the absolute variance explained by each of the principal components. There is a slight "elbow" at around the 10th components, whereupon the height of the bars starts decreasing slower. **b** - shows the cumulative proportion of variance explained by the principal components. We see that the first 10 components explain only around 16% of the total variance in the data.

As a sanity check, we do a quick k-means clustering of the independent 197 unit cells into 3 classes using just the 10 first principal components. We have not used any information about the physical location of the unit cells in our calculations so far. If the clusters obtained by considering just the pixel values of the independent 197 unit cells are also continuous in physical 3d space, then it reassures that we are on the right track. The picture that emerges by coloring the unit cells by the k-means assigned clusters (Figure 4.29) confirms that the clusters are indeed continuous.

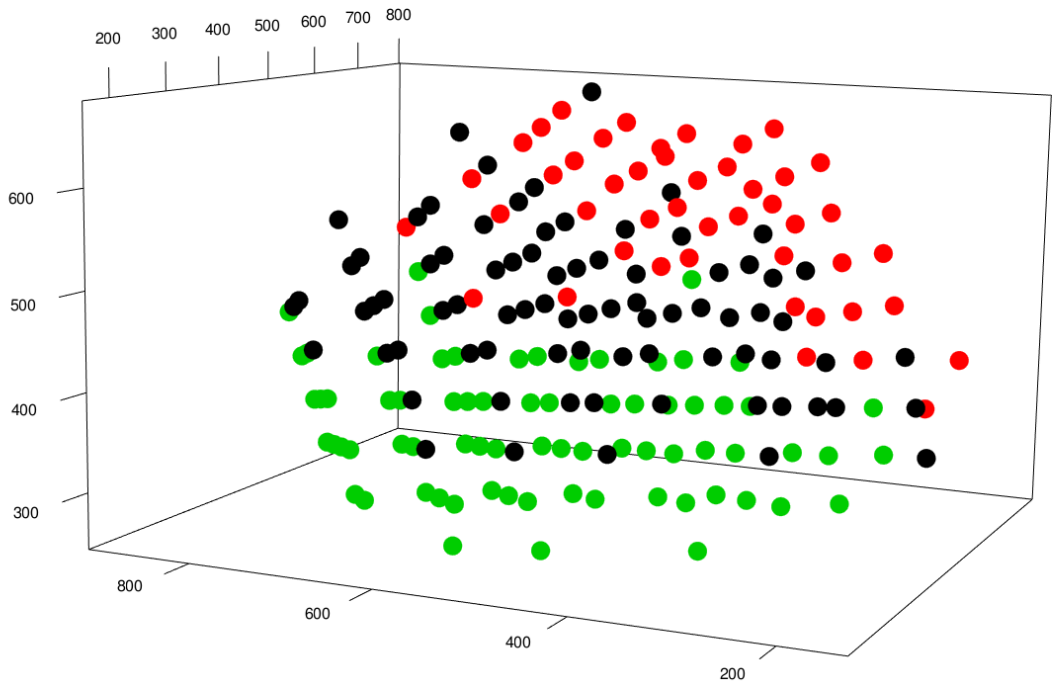

Figure 4.29 | Clustering of the independent unit cells into 3 classes using k-means on the 10 first principal components. The classes appear continuous in the physical space.

But since we are interested in classifying all of the unit cells, not only the independent ones, we project the rest of the unit cells to the same principal component space. The pairwise plots of the first 5 components from all of the unit cells are visible on Figure 4.30

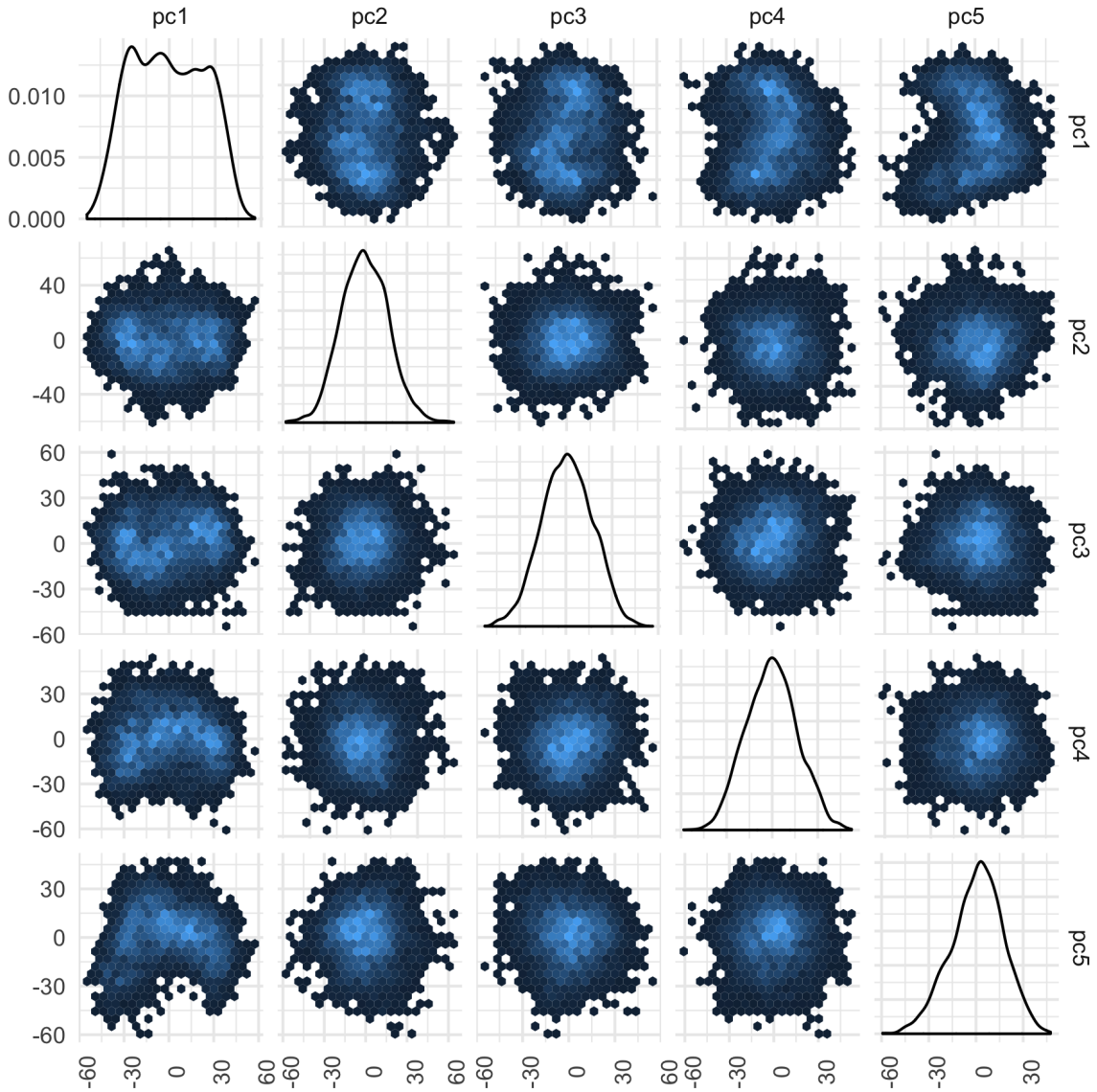

Figure 4.30 | Pairwise scatter plots of principal components. To avoid overplotting the points have been hexagonally binned. The lighter the color the higher the number of the points in a particular bin. Principal component 1 describes largest portion of the variance and plotting it against other components appears to draw out some structure.

Further Dimensionality Reduction with tSNE

Since no really clear clusters were forming in the first 2 or 3 principal components, we opted to experiment with some non-linear method that would map the first 100 principal components of each object into just 2 dimensions and hoped to obtain a sensible clustering and gain an idea about the number of clusters.

The method of choice was *t-distributed Stochastic Neighbor Embedding* (tSNE) [74]. It is a modern method that has been successfully used for visualizing high-dimensional data in numerous cases and it appears to be superior to many other non-linear mapping

techniques [74]. tSNE tries to find a low-dimensional embedding to high-dimensional data that would retain the nearest neighbor graph as faithfully as possible. The number of neighbors to consider is a tuning parameter.

The downside of tSNE is that the low-dimensional embedding is non-parametric, meaning that it cannot be directly applied to new data points, which rules out the strategy of using just the 197 independent unit cells for figuring out the mapping, and then using that mapping to map the rest of the unit cells down to 2 dimensions. So we had to run tSNE directly on all of the 5104 unit cells.

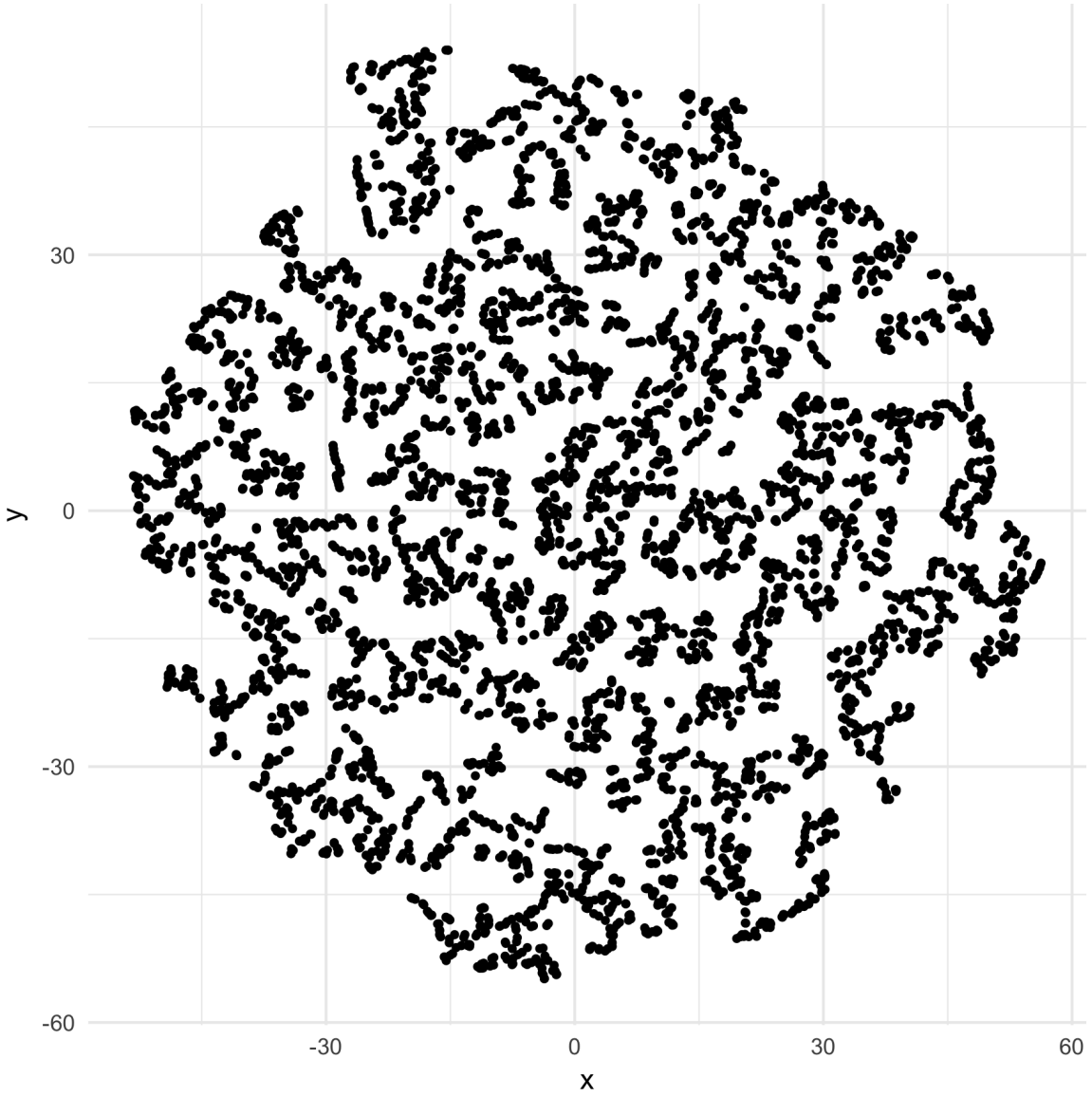

Figure 4.31 | tSNE embedding of the first 100 principal components in 2 dimensions. The first 100 components in 2 dimensions seem slightly clustered, but it is hard to tell where exactly the cluster borders run.

The results of tSNE are plotted on Figure 4.31. There seems to be some definite structure to the data, but it is hard to tell whether it is true structure and where

exactly the cluster borders run. The clusters borders are very complex, so k-means clustering would do no good here, as it can only find spherical clusters. To somehow assign the points into clusters we used single linkage hierarchical clustering.

Single Linkage Hierarchical Clustering

Single linkage clustering works by joining the closest two points on each step, until all points are merged into one cluster. By setting a distance threshold we can stop the procedure early when no connections shorter than the threshold remain. As a result the method reports the clusters that have formed by then. Single linkage clustering sounds simple, but it can be surprisingly powerful in cases where the clusters have strange shapes, just like in the current case.

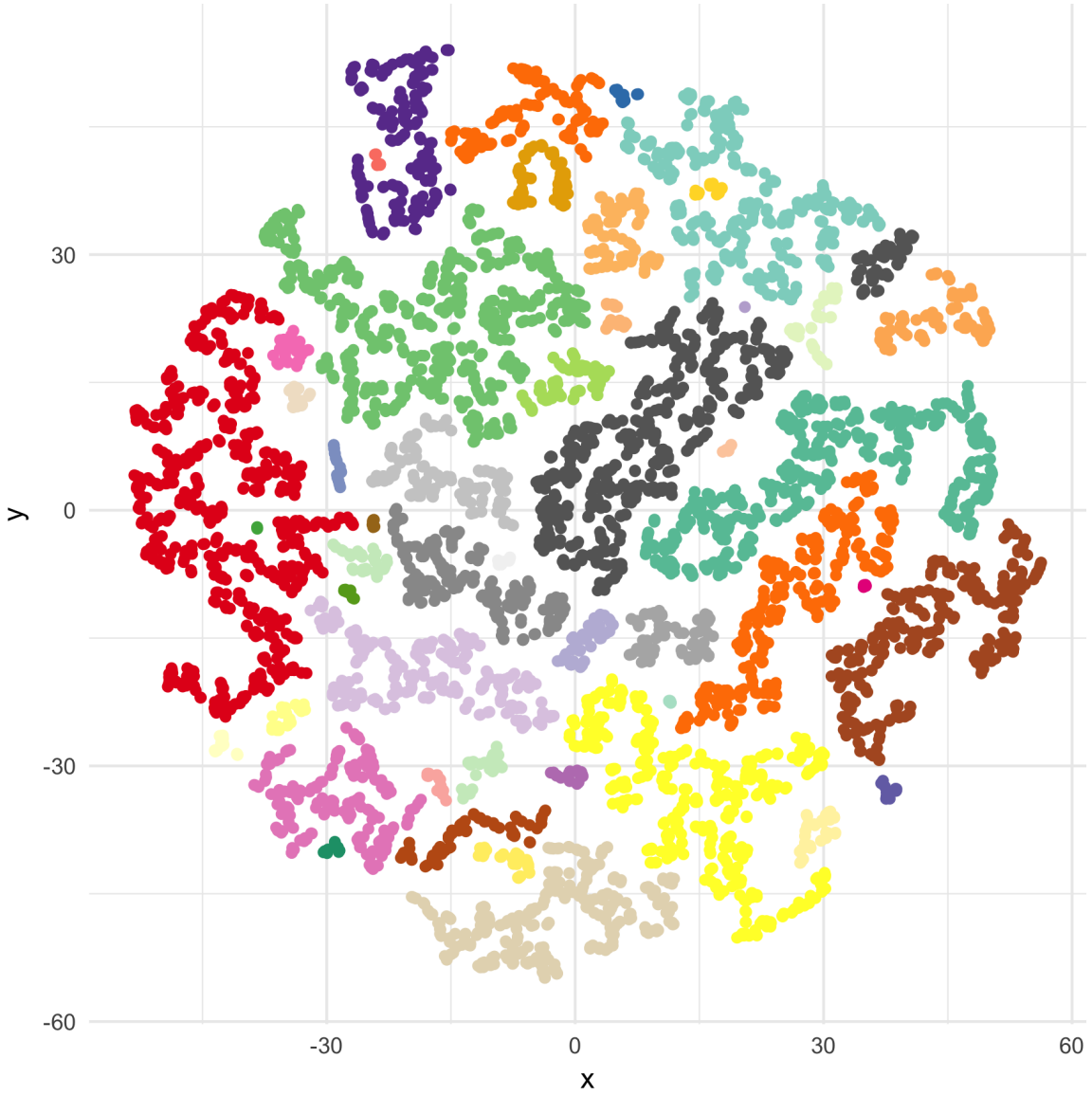

Figure 4.32 | TSNE embedding of 100 first principal components in 2 dimensions, clustered by single linkage. The 49 clusters formed by single linkage hierarchical clustering are depicted with varying colors. Some clusters are large, while others have only 1 or 2 points. The axes here represent the coordinates in the low-dimensional embedding and have no physical meaning. The clustering may seem arbitrary on the first look but in 3 physical dimensions the same clusters are mostly continuous, giving assurance that the grouping makes sense.

The 49 clusters formed by single linkage hierarchical clustering are visible on Figure 4.32. The clusters may look slightly arbitrary, but looking at the same clusters in the physical three dimensions on Figure 4.33, we see that they remain mostly continuous. The same truth is perhaps better seen on Figures 4.34-4.36 where the clusters are depicted as unit cells in 3d.

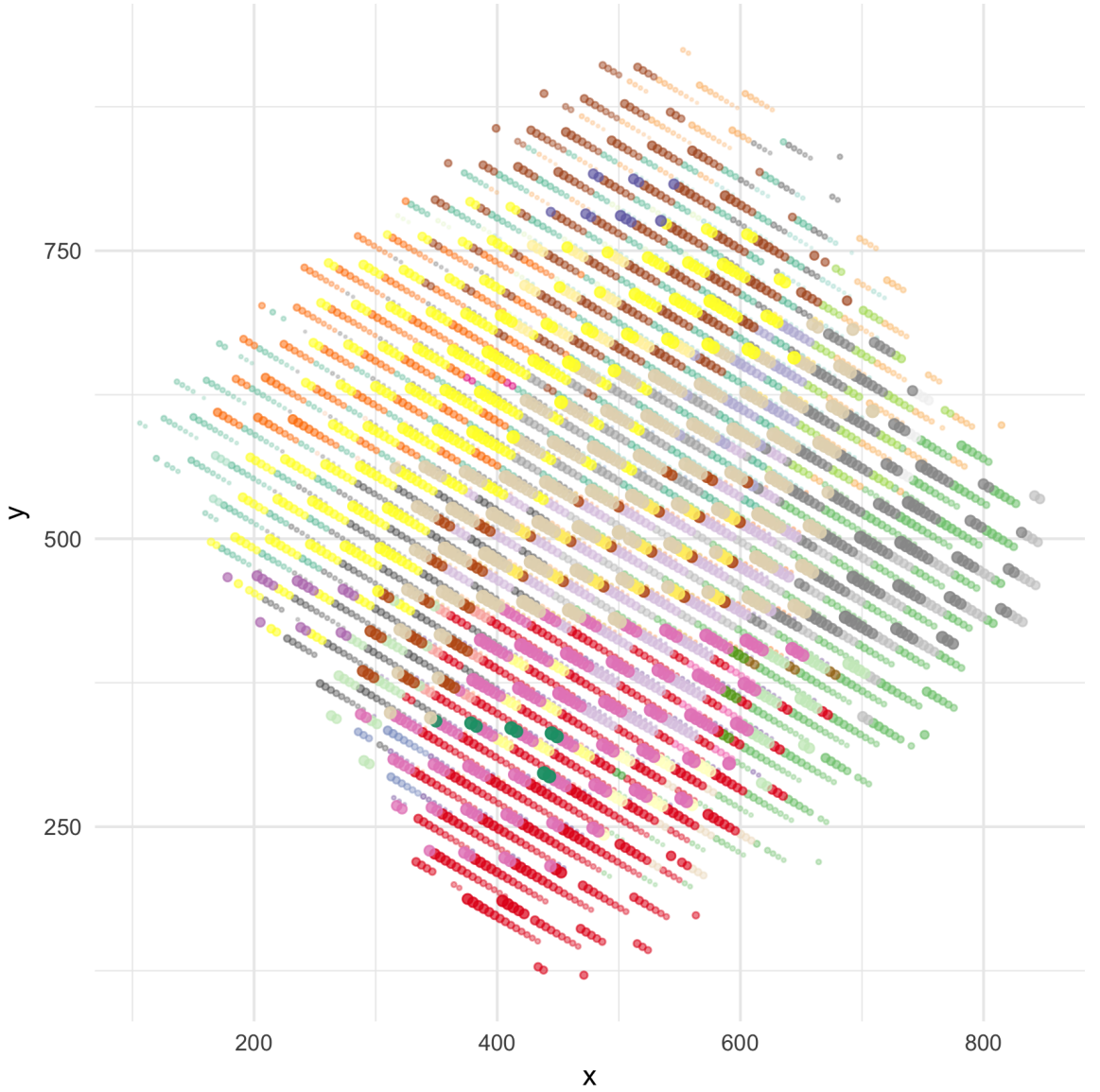

Figure 4.33 | TSNE embedding of 100 first principal components in 2 dimensions, clustered by single linkage, in 3 physical dimensions. The same clustering as in Figure 4.32 but points mapped into 3 physical dimensions. Z-coordinate is mapped to the size of the dot and alpha value to increase the depth perception, so smaller and fainter dots have smaller z value and appear "further away".

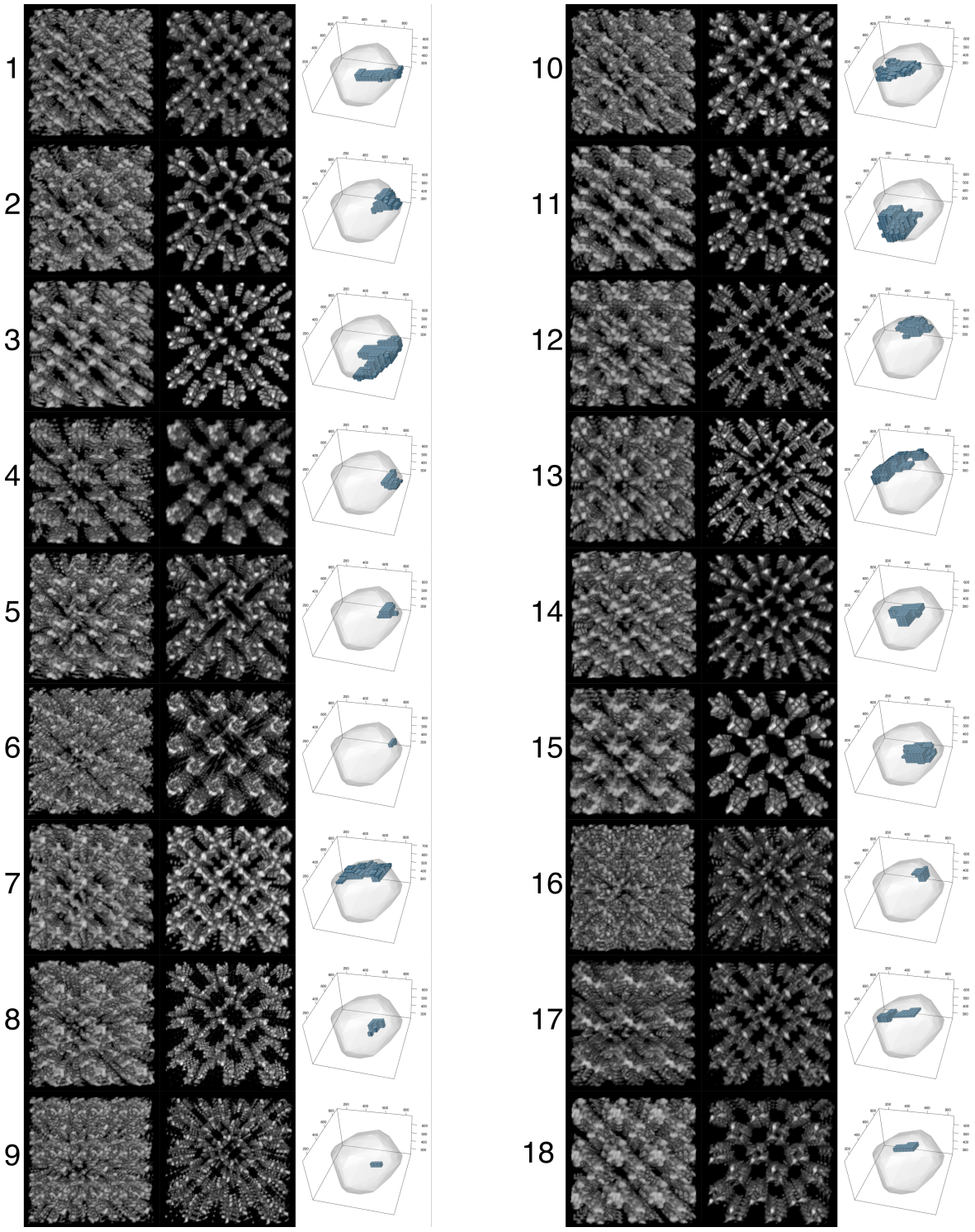

Figure 4.34

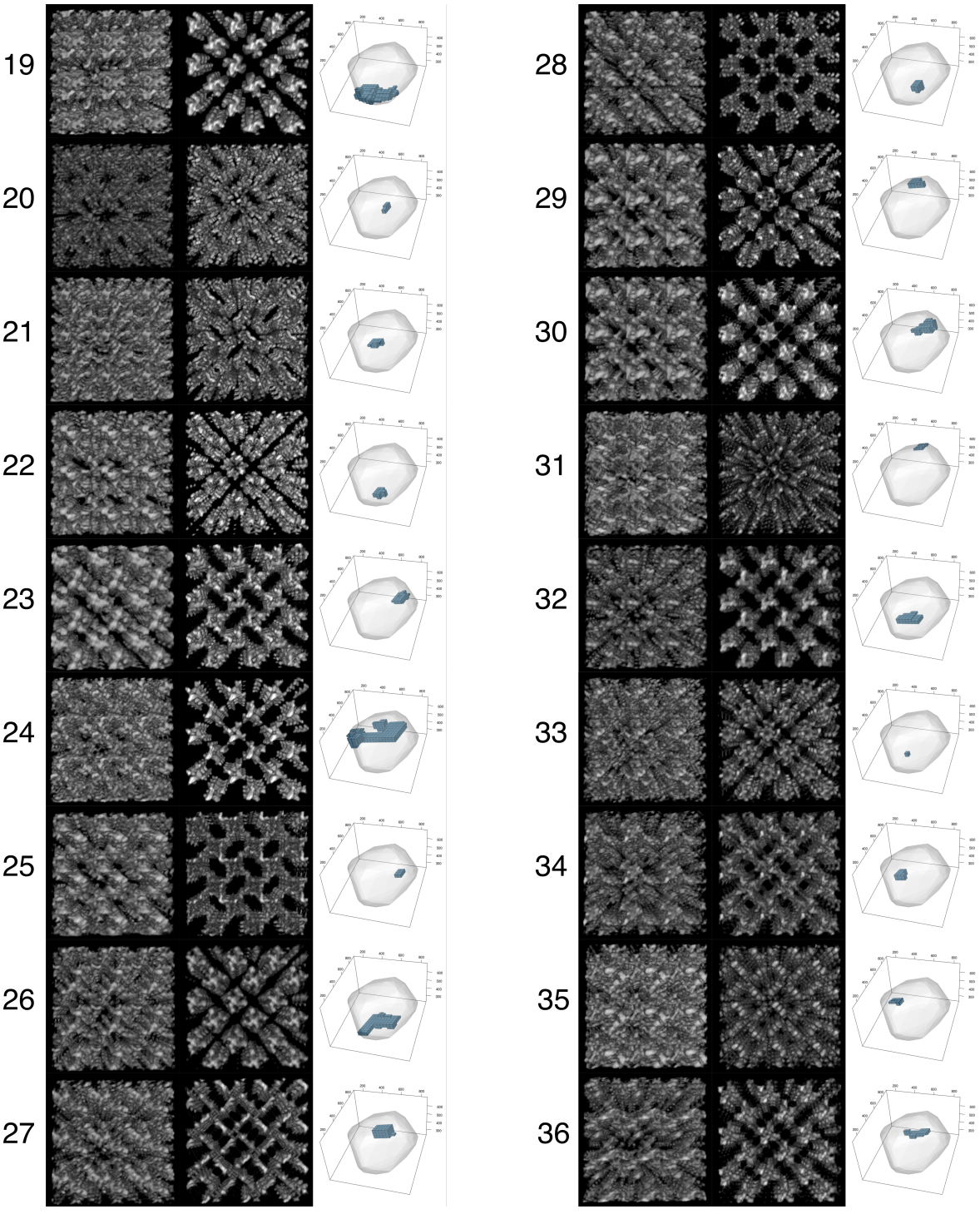

Figure 4.35

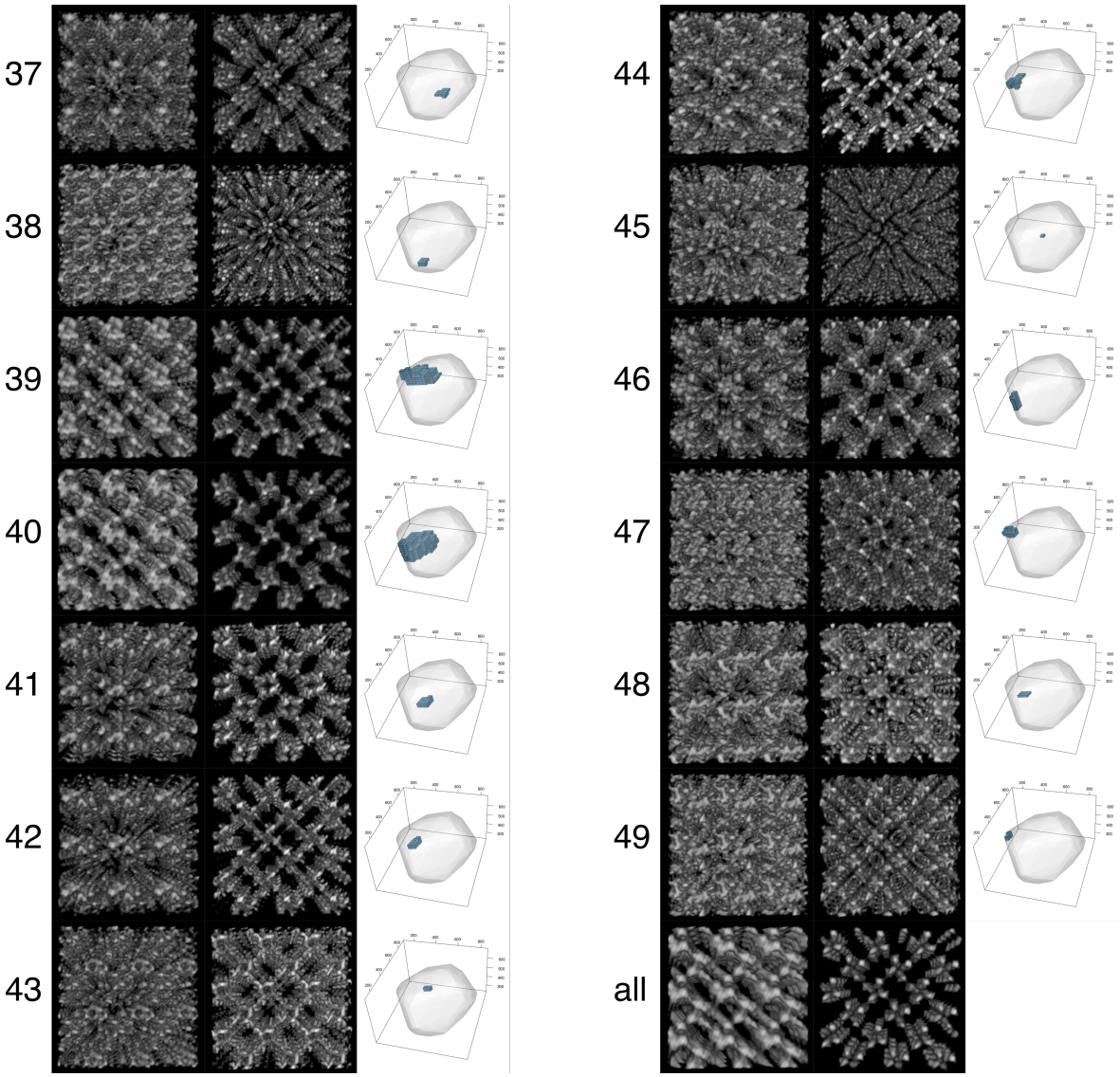

Figure 4.36 | Averaging an symmetrization within clusters. Averaged map (left) and symmetrized map (right) side-by-side. The number is a cluster indicator. The last map "all" includes unit cells from all clusters. The symmetry origin used for symmetrization is the same on all maps and was determined from the "all" map. The contouring level for the non-symmetrized maps (left map) is the same for each cluster, but it varies for the symmetrized maps (right map), because depending on how well the symmetry origin fits, the maps are more or less blurred out and therefore the gray-scale values are not directly comparable.

Conclusion

The conclusion from the cluster analysis is that the crystal is definitely not homogeneous throughout. We see a variety of different forms and shapes when dividing the crystal into clusters and averaging and symmetrizing within the clusters independently. In the current approach the number of unit cells per cluster varied considerably, from just 1 up to 538. The symmetry origin was determined on the average map, that in-

cluded all 5104 unit cells. The choice of the symmetry origin is probably suboptimal for some of the unit cells, because some clusters appear to have a shift in the origin, but using the same symmetry origin everywhere illustrates the differences between the clusters.

The question whether the variability in morphology is true variability, or it is an artifact of the reconstruction due to the defocus gradient through the reconstruction, defocus variation over the tilts or misaligned tilt-series, remains to be answered. It is remarkable, however, that the clusters formed by considering only the voxel values of the reconstruction, are continuous also in the physical 3d space. This demonstrates that there are indeed spatially varying differences between parts of the crystal.

As we will explore later in the section regarding the symmetry origin, the major problem with the current data set is that the origin can be accurately located in the $\boldsymbol{a}$-$\boldsymbol{b}$ plane, but is hard to pinpoint its coordinate on the $\boldsymbol{c}$ axis. The result is that the 4-fold screw axis stacks are well separated from each other, but the molecules the stacks are blurred together (Figure 4.34-4.36). This is currently the major obstacle, and better, more clever algorithms are needed to reliably solve this issue.

One final note is that using a different clustering method, or varying the number of tuning parameters, obtains different classes. However, the clusters themselves are almost always continuous in the physical space. A potential critique against this approach could be that the seen clusters are only there because we used local averaging and artificially created similarities between neighboring cells. However, this can be countered by pointing out that even when we clustered the 197 independent unit cells that shared no common neighbors we still saw continuity in the formed clusters. The question whether the classification reflects something about the mosaicity of the crystal cannot be conclusively answered at this point and has to be studied further.

4.4.9 Exploring Unit Cell Distribution

Despite the many challenges of cryo-electron tomography the most amazing upside is that the reconstructed 3d density is a nanoscopic chunk from the actual natural world. Most other methods used in structural biology rely one way or another on averaging which makes the study of objects that are truly unique or have a continuum of conformations difficult. While molecules inside the crystal are supposedly all identical, the whole crystal itself as a higher order structure is unique.

It has been observed that flash freezing protein crystals to cryogenic temperatures increases the *mosaicity*, defined here as the broadening of the lattice orientation and spacing distribution [73]. In this section we try to examine the distribution of unit cell vector orientations and lengths within a protein nanocrystal by harnessing cryo-electron tomography.

Coherently illuminating small crystals with x-rays makes it possible to reconstruct crystal morphology and strain field. In [75] they used Bragg Coherent Diffraction Imaging (BCDI) on 1.25 µm sized calcite crystals to visualize a network of dislocations in 3d. In [76, 77] electron tomography was used on nanoparticles to obtain an atomic resolution description of the dislocations in various materials.

The Approach

Our approach to investigate the lattice disorder is the following

1. Take a large 3d reconstruction of a protein crystal

2. Divide the reconstruction into smaller volumes, referred from here on as *chunks*

3. For each chunk calculate FFT, perform peak search and determine lattice vectors

On Figure 4.37 a 3d depiction of all of the positions where the chunks were extracted are shown as black dots. Each chunk in this case is $256 \times 256 \times 256$ voxels large. Ideally, we would have liked to use smaller chunks because chunks that large encapsulate a fairly large portion of the entire crystal, thereby losing the local information. The signal in chunks smaller than 256 cubed was too weak to for the algorithms to reliably determine the lattice vectors. The gray surface on Figure 4.37 is a convex hull around all of the lattice points that correlated with the averaged unit cell higher than a certain threshold, which approximately outlines the crystal. Based on visual inspection the crystal actually extends slightly further out than the convex hull, but the shape is approximately the same.

As there are hundreds of small chunks that each need processing the procedure would be labor intensive if performed manually. Automatic treatment needs robust or adaptive programs because each individual chunk is different and might need separate parameter tuning. Fortunately, our peak-detection and lattice determination programs are robust enough, that for a large number of chunks the automatic determination worked well and parameter tuning had to be performed only once in the beginning of the procedure. After the first round of lattice determination, we selected all chunks that gave obviously wrong results, and submitted those to the second round of the same procedure, but this time with a different set of parameters. Two rounds was enough to cover most of the crystal.

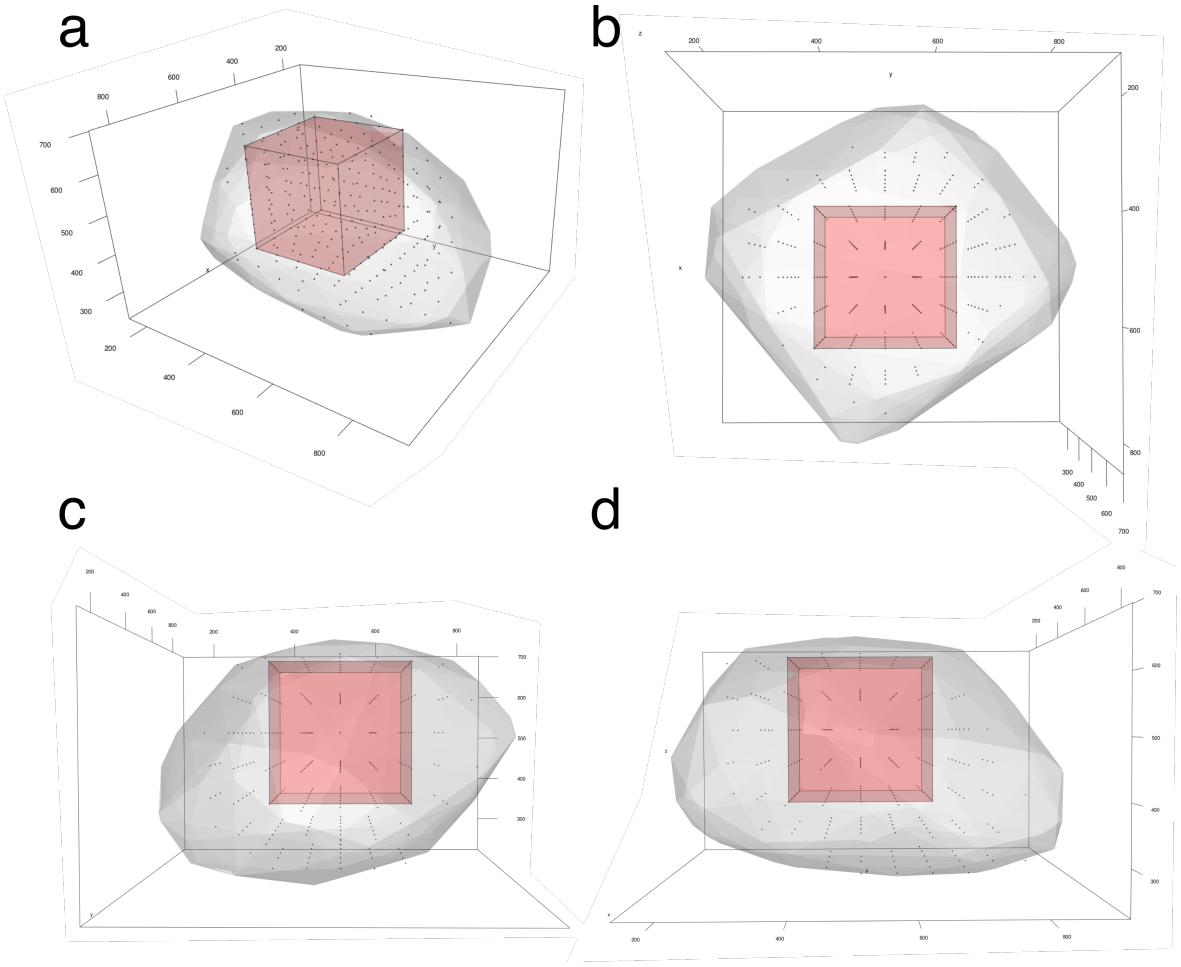

Figure 4.37 | Positions where the unit cell was determined. Black dots represent positions where a 256 × 256 × 256 voxels chunk (red cube) was extracted. The gray surface is a rough segmentation of the crystal. **a** is a tilted view, **b** is a view along z-axis, **c** is a view along y-axis and **d** is a view along x-axis.

The distribution

For 286 of the partially overlapping chunks the automatic lattice detection algorithm was able to retrieve unit cell vectors that made sense. As we expect the crystal to be in a space group $P4_32_12$ with unit cell vectors lengths around 80 Å, 80 Å and 40 Å we decided to keep only those vectors that were not further than 5 Å from those values. The locations of the chunks that gave results out of those boundaries were mostly at the periphery of the crystal thereby including a large portion of the non-crystalline background or completely outside of crystal. The fact that the lattice angles are not all exactly 90° suggests that the crystal is actually in a triclinic form rather than tetragonal, but the molecules seem to nevertheless approximately follow the $P4_32_12$ space group.

Figure 4.38 shows a 3d visualization of the unit cell vector distribution. The heavy colored lines represent the mean vector for each crystallographic direction. We can see that the $\boldsymbol{c}$ vector is the most stable while $\boldsymbol{a}$ and $\boldsymbol{b}$ vary noticeably. The variation is

mostly, but not exactly, in the z-direction. This suggests that the variation could just be from the lower spacial resolution in the z direction, caused by the missing wedge. However, that can not be the only source of variation. In some of the later visualization we see that there is also variation in the unit cell parameters dependent on x and y directions.

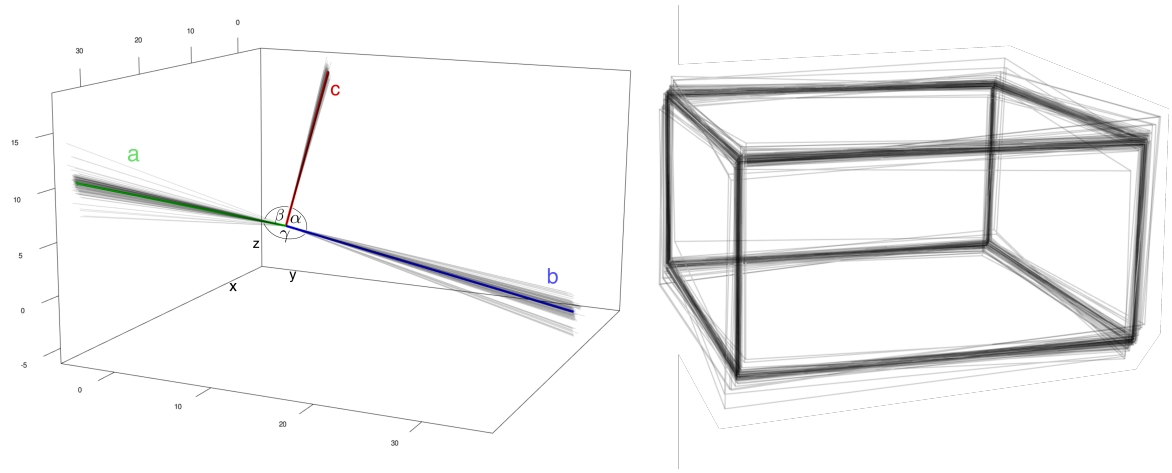

Figure 4.38 | All determined unit cell vectors plotted together. **left** - All vectors drawn together. The average vector of each bundle is drawn in color. **right** - All vectors depicted as unit cells with common center point. The unit cells are drawn partly transparent.

Distributions of coordinates

It is instructive to study which coordinates vary together, independently, or none at all. On Figure 4.39 all pairwise scatter plots as well as non-parametric density curves for each coordinate can be seen.

An interesting aspect is that several of the distributions, but especially the y-coordinate of the $\boldsymbol{c}$ vector, appears to be bimodal. The positions belonging to one mode are relatively more constant whereas the position close to the second mode seem to be more variable. It would be interesting to see how this characteristic is distributed within the crystal itself. The distribution of the y coordinate of $\boldsymbol{c}$ axis is mapped into 3d on Figure 4.40.

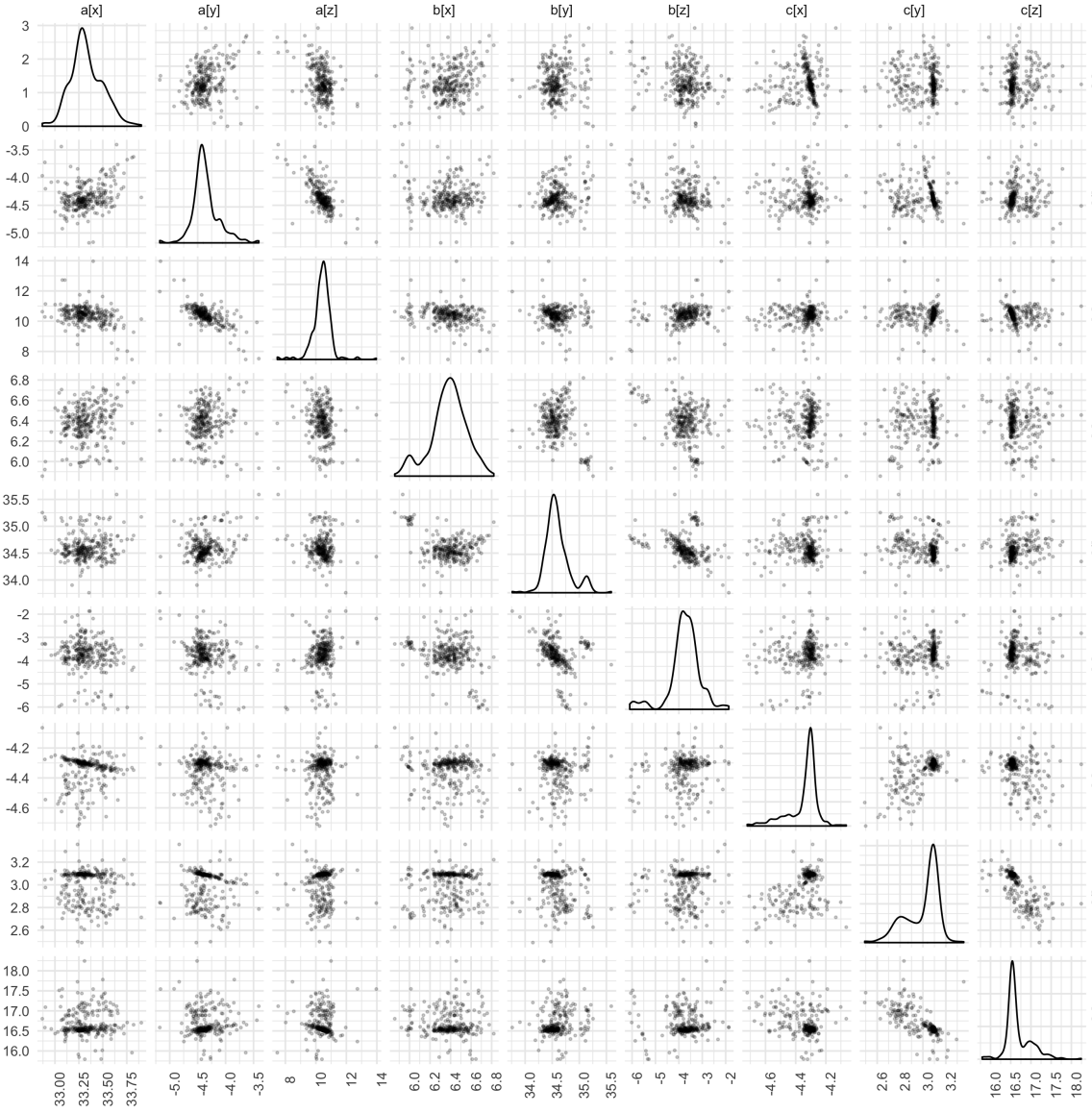

Figure 4.39 | All pairwise scatter plots of the coordinates of the unit cell vectors.

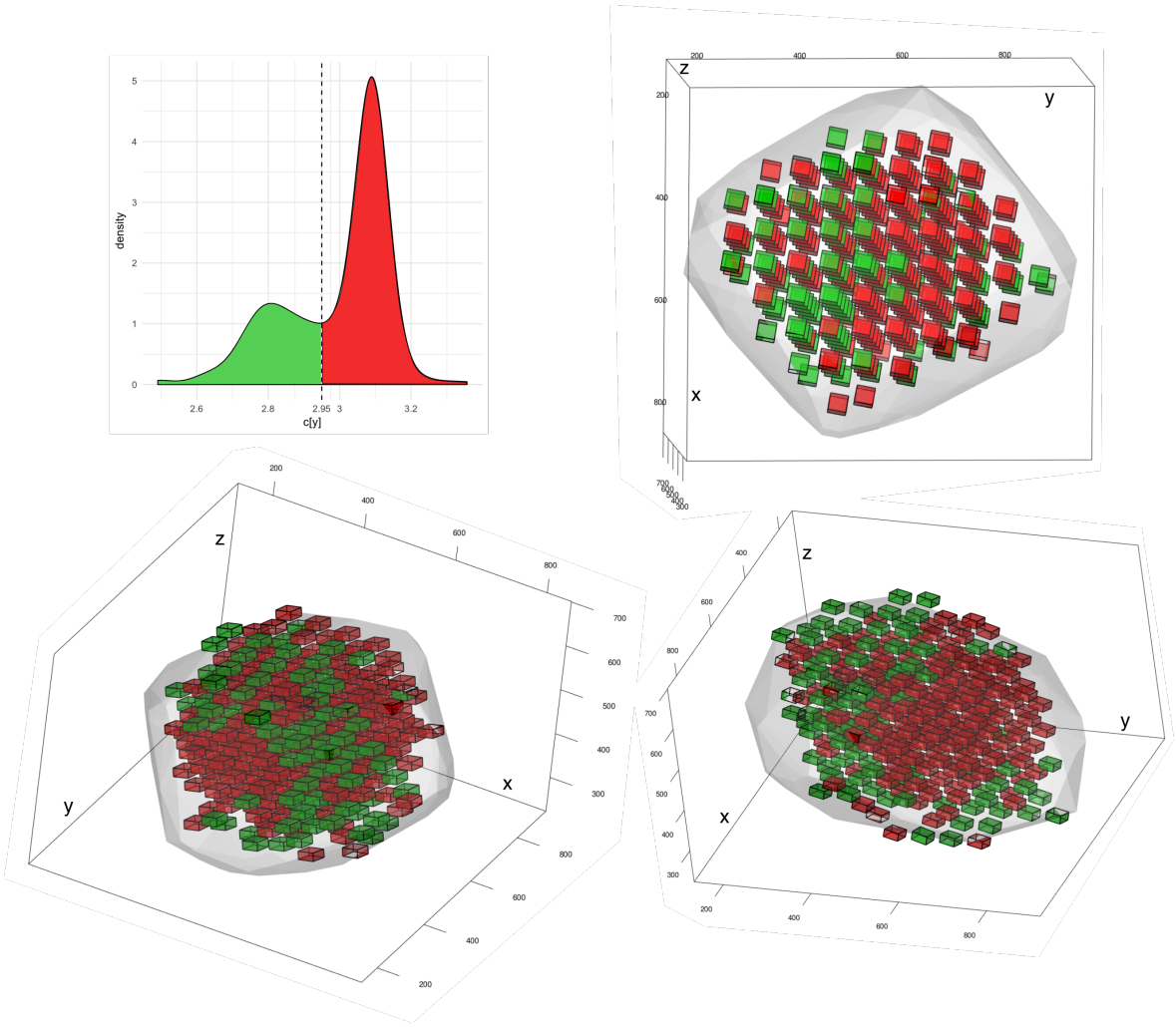

Figure 4.40 | Distribution of y-coordinate of *c* vector. The distribution of y-coordinate of ***c*** vector is split into two modes and the result is mapped into 3 dimensions. It appears that the y-coordinate of the ***c*** vector is more stable at one side of the crystal, and varies more at smaller y values and at the periphery of the crystal.

Distribution of cell parameters

The pairwise scatter plots of the unit cell parameters a, b, c, α, β and γ can be seen on 4.41. Again we can see how the length of the ***c*** vector seems to be more stable and shows a sharp peak around 39.5 Å, while ***a*** and ***b*** vary relatively more.

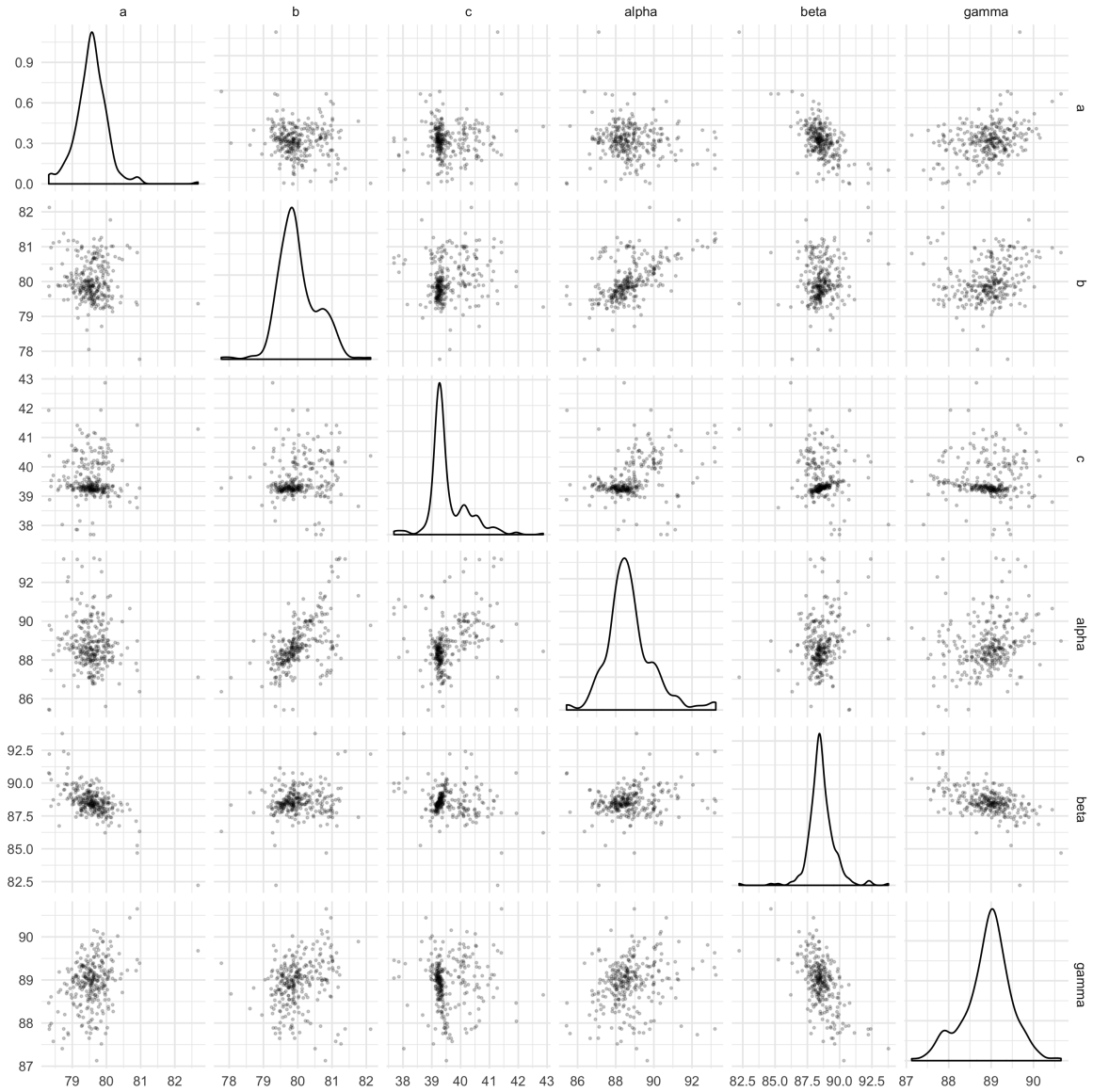

Figure 4.41 | Nonparametric density curves of the unit cell parameters.

Figure 4.42 shows the same information as on the main diagonal of Figure 4.41 but the relevant density curves have been placed on the same graph for better comparison. In a $P4_32_12$ space group all angles between the unit vectors should be 90°. On Figure 4.42a we see that only in very few cases the angles are exactly 90° and in most cases they are below, fluctuating around 88°. This might suggest that the crystal has been subjected to some external forces during the sample preparation and flash-freezing phase and is probably under some stress. In our method of crystallization, in order to avoid the formation of large crystals, we flash-froze the specimen almost immediately after the start of intense salting out. Perhaps, for future reference, the crystallization method should be altered if a more perfect crystal is desired.

Figure 4.43 shows also the co-distribution of the angles from a third perspective. We see yet again that most of the angles are below 90°. While it is possible to find positions in the crystal where the unit cell angles α and β, or α and γ, were simultaneously 90°,

there is no such position where it would hold for γ and β, and obviously then also for all of the three angles together, reflected on the 3d scatter plot.

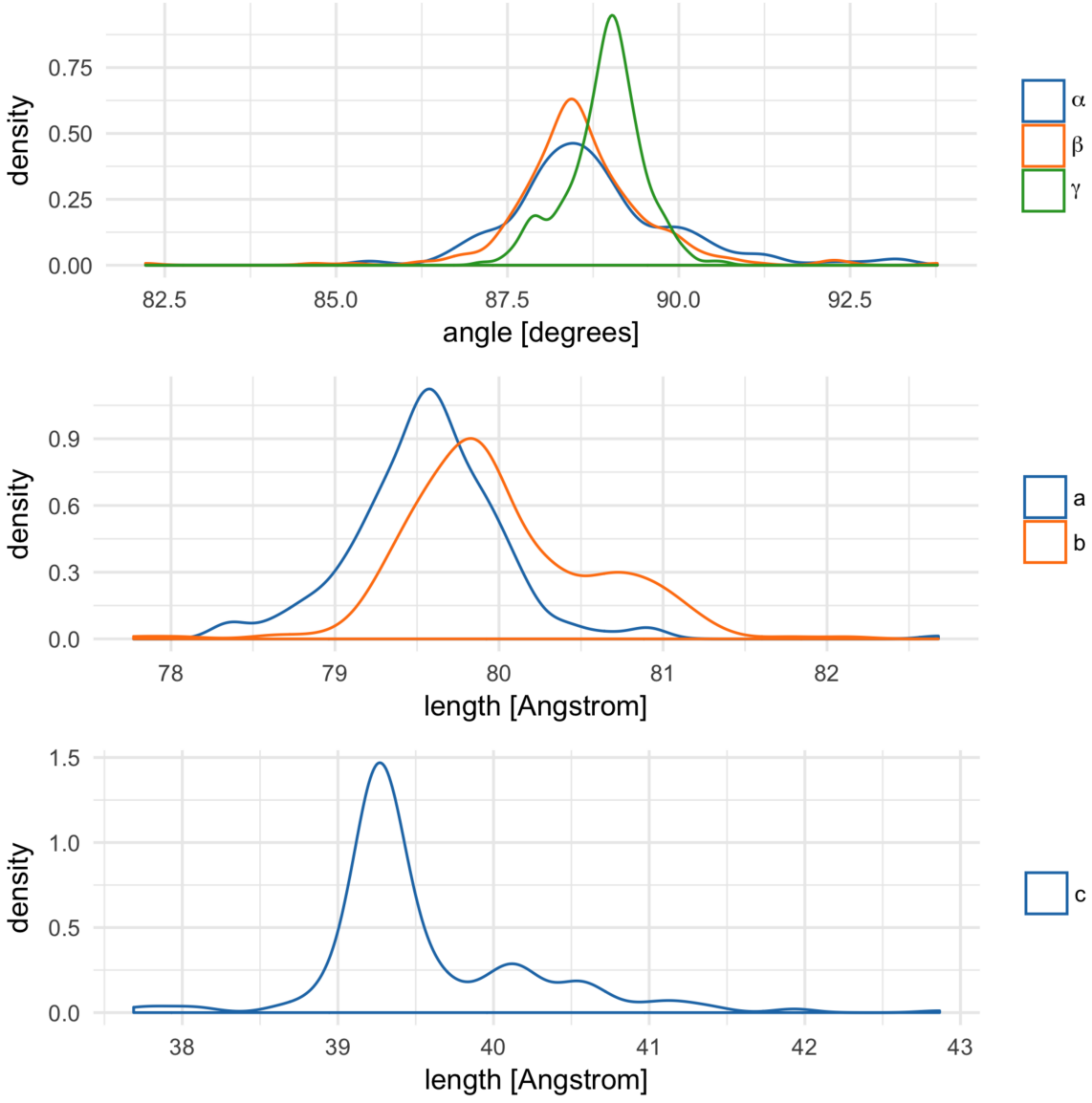

Figure 4.42 | Nonparametric density estimation of unit cell parameters.

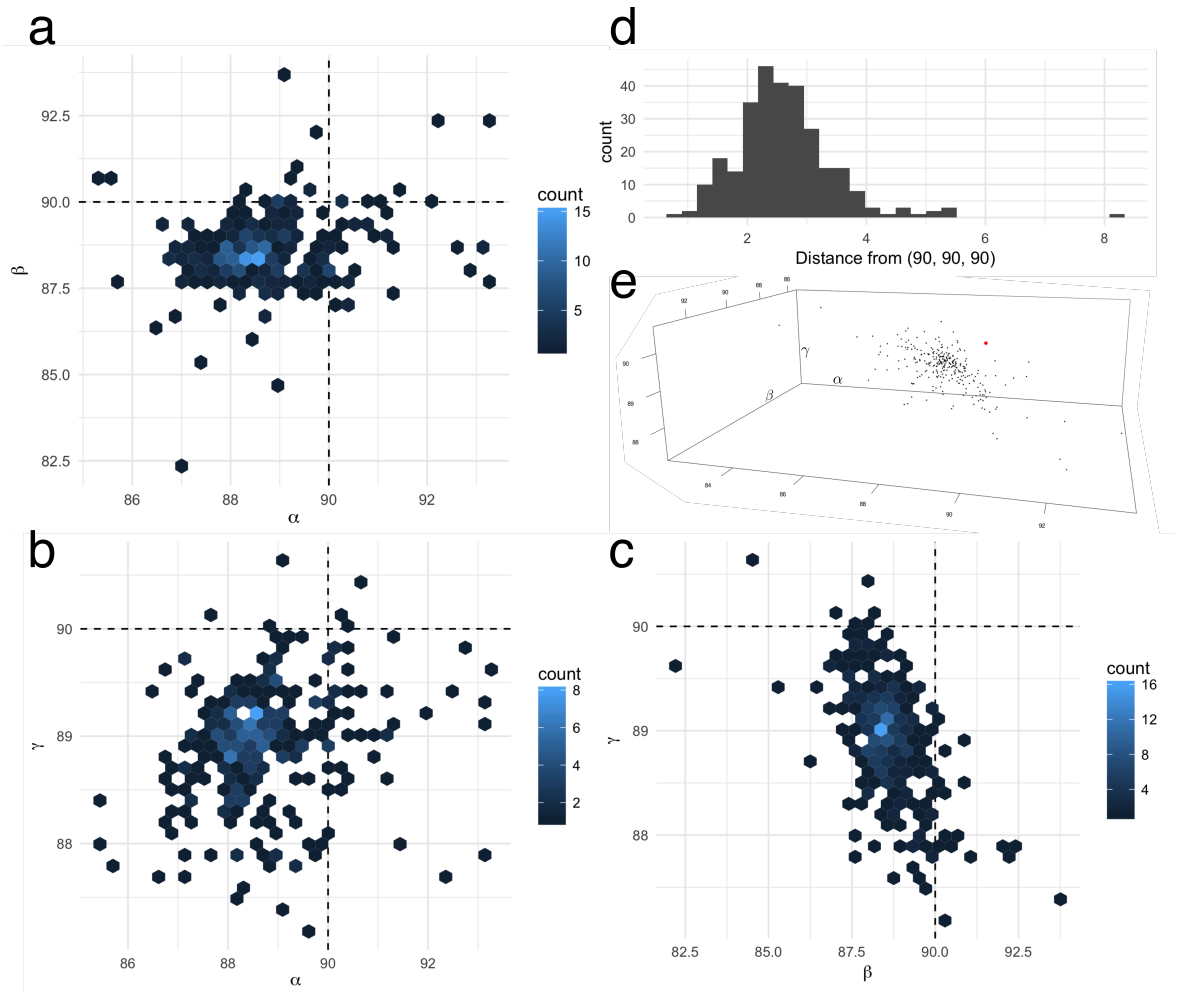

Figure 4.43 | Distribution of angles between unit cell vectors. The hexagonally binned scatter plots **a - c** show the pairwise joint distribution of the angles between the unit cell vectors. We see that most of the angles are less than 90° which suggests that the crystal, that was expected to be of a tetragonal system, is slightly sheared. **e** - On the 3d scatter plot of the angles the red spot depicts the $(90°, 90°, 90°)$ point. We see that none of the vector triplets are perfectly orthogonal. **d** - Histogram of Euclidean distance between the angles and a point $(90°, 90°, 90°)$.

On Figure 4.44 the relationship between the reference frame coordinates and two unit cell properties have been plotted. Figure 4.44a shows the variation in the length of the vector $\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c}$, i.e. the vector pointing from the origin of the unit cell to its furthest corner. We see that there is little to no dependence in x, the length is longer for smaller y and has a quadratic dependence on z. This suggests that the unit cells are slightly more stretched out in the center of the reconstruction than at the bottom or at the top. Interestingly, when the unit cell gets stretched in the center of z, there is no such dependence on the volume, which seems to remain almost constant (Figure 4.44b).

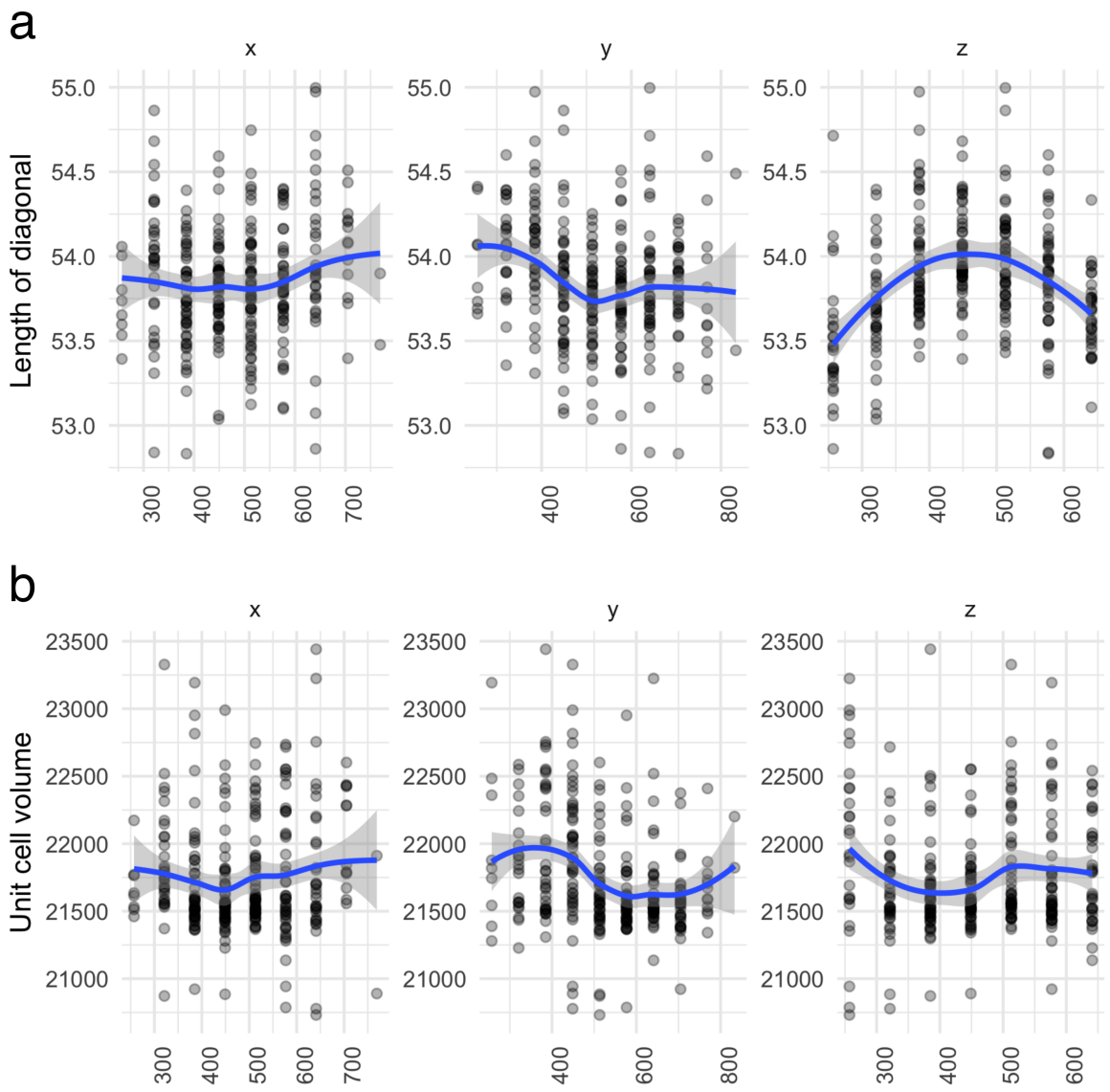

Figure 4.44 | Relationship between laboratory coordinates and the unit cell size.

4.4.10 Symmetry Origin

Knowing the lattice vectors makes possible averaging the reconstruction in P_1 space group, i.e. taking advantage only of the translational symmetry. For P_1 the origin of the lattice can be arbitrary. For any other symmetry operation, we need to know the exact locations of the symmetry elements. For a space group $P4_32_12$, we have 8 equivalent molecules within the unit cell, all mapped to each other via symmetry operations. The positions of the symmetry elements are determined once an origin is fixed.

In addition to being able to actually apply the symmetry it was an interesting question to us, whether there is a change in the relative position of the symmetry

elements over the whole crystal or does it remain constant. This information could serve as a proxy for studying dislocations or other imperfections. We already know from looking at the diffraction pattern on the Fourier transform of the large crystal that it looks like that of a single crystal.

To perform that analysis we took the same grid of positions shown on Figure 4.37. Then, taking each grid point as the center, we calculated the average of the surrounding $5 \times 5 \times 5$ unit cells. So altogether 125 unit cell sized chunks were averaged at each grid point.

Then for each averaged unit cell we performed the symmetry origin search and investigated whether there is any relative change in the origin. The results can be seen on Figure 4.45. The points are plotted in fractional coordinates, i.e. fractions of the unit cell vectors.

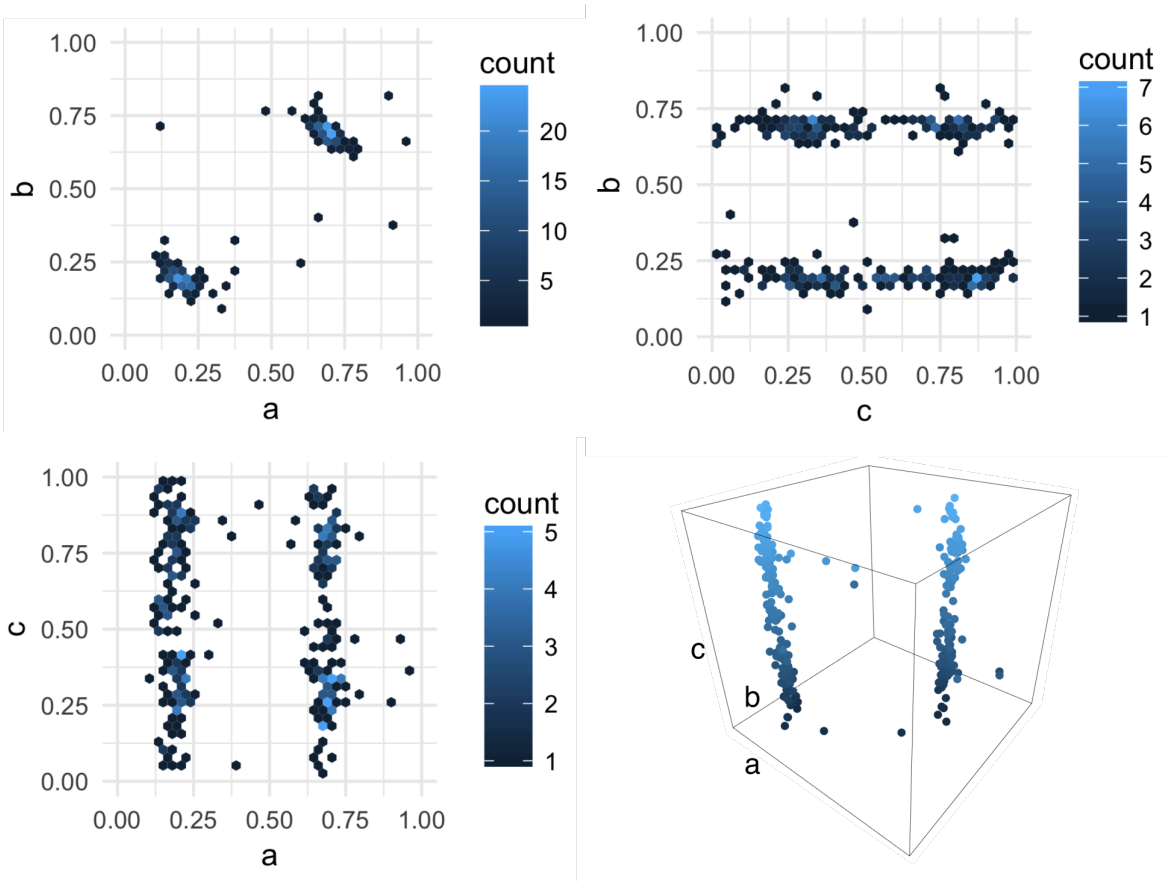

Figure 4.45 | Distribution of symmetry origins over all crystal. The coordinates are fractions of crystallographic vectors. the origin is relatively well determined on $\boldsymbol{a}$-$\boldsymbol{b}$-plane, but extremely poorly along the $\boldsymbol{c}$-axis. The color coding for the 3d scatter plot in the bottom right corner is according to $\boldsymbol{c}$ value.

Looking at the plot a couple of curious and eye-opening circumstances are revealed. First, it becomes immediately apparent that our resolution and certainty in the c-direction is not good. The crystallographic origins are without exaggeration spread through-out the unit cell along the $\boldsymbol{c}$-axis. The $\boldsymbol{c}$-axis for the current dataset is just at

a 17.3° angle with the z-coordinate, so the spread could be due to our lack of knowledge about the details in that direction. Another explanation is that there is true variation in the c-direction. Since there is little evidence to back up this claim, the first explanation is more likely.

The position of the origin is well-determined on the ***a***-***b*** plane. The reason for having two clusters is that per our current definition, the origin of symmetry lies on a 4-fold axis. $P4_32_12$ space group has two of those, so it's probably a random pick between those two by the algorithm (Figure 4.46). The same is true for the origin along ***c***-axis. So altogether we have 4 equivalent points for the symmetry origin and looking closely at Figure 4.45 we see that there indeed are 4 clusters of points.

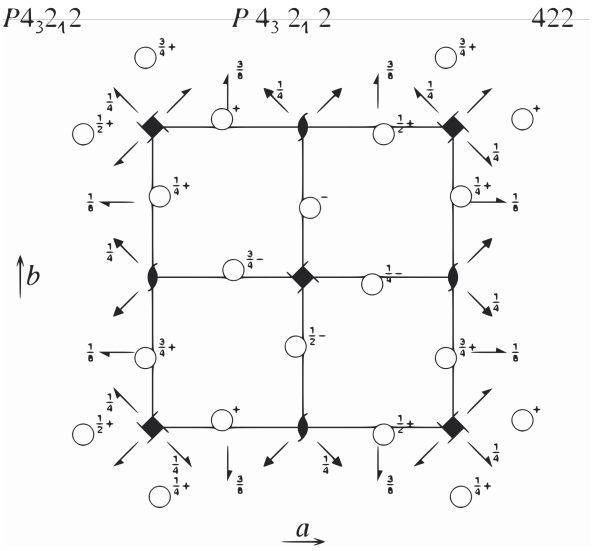

Figure 4.46 | The diagram of $P4_32_12$ space group.

Let us split the points into 4 classes. The easiest is to perhaps first split along ***b*** and say that points less than 0.5 belong to class 1 and points greater than 0.5 to class 2. Let us observe the distribution along ***c*** for the two classes separately.

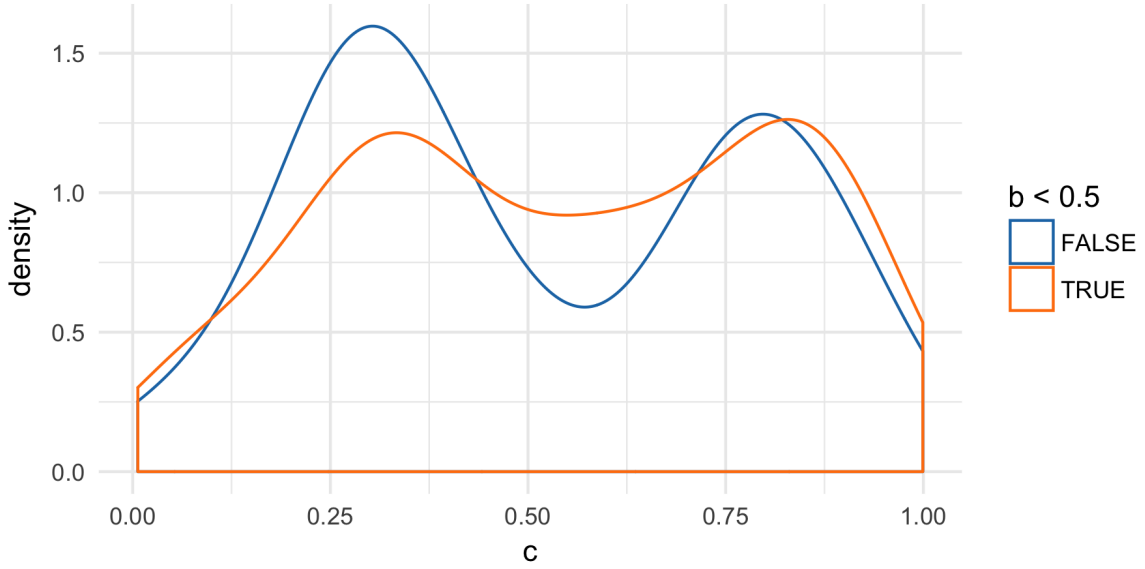

Figure 4.47 | The distribution of origin along vector c grouped by $b < 0.5$.

On Figure 4.47 we can see that the symmetry origin along vector $\boldsymbol{c}$ has two quite prominent modes for each class. To translate all points to the same interpretation we can translate all origins in fractional coordinates into the first quadrant of the unit cell by subtracting 0.5 from all coordinates that are greater than 0.5. Doing that allows us to take the median of the origins and get a reasonable estimate for the symmetry origin that will also become the new definition for the lattice origin. We take the median instead of mean to have less influence of the outliers. The median origin in fractional coordinates is then $(0.190, 0.195, 0.295)$. This number on its own carries little information as it depends on where we placed the origin previously for P_1, but it describes the shift of the lattice so that the unit cell lattice vectors would coincide with the symmetry axes.

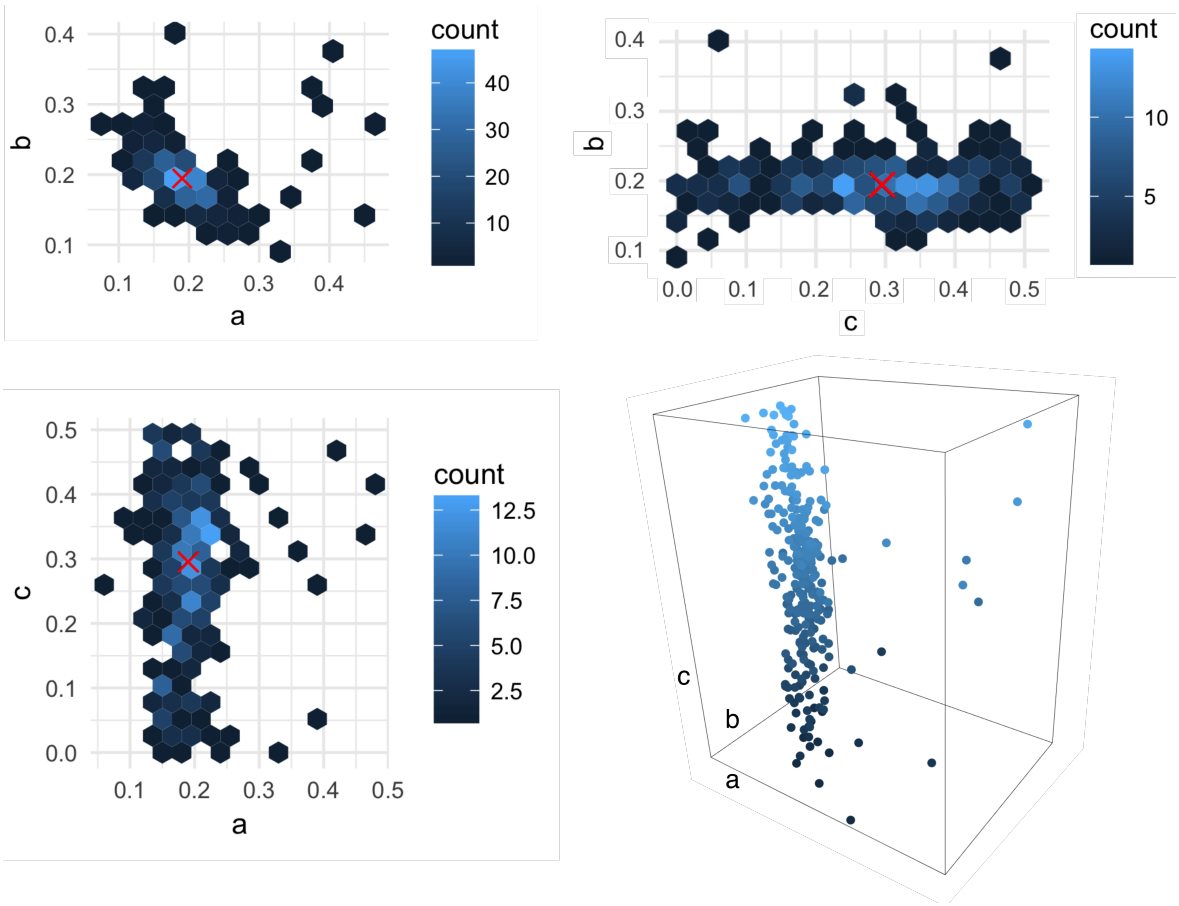

Figure 4.48 | Distribution of the symmetry origins over whole crystal. The coordinates are fractions of crystallographic vectors. The color coding for the 3d scatter plot in the bottom right corner is according to $\boldsymbol{c}$ value.

4.4.11 Shortcomings and Discussion Points

The method of nanocrystal cryo-ET being still far from maturity has its share of shortcomings that should be addressed in the future. Not all of them can be known at the time of writing, but some more important ones are being discussed in the following. The first shortcomings are more general and related to any tomographic reconstruction attempt. The later points are more specific and have to do with technicalities of nanocrystals ET.

1. **Aligning the tilt-series.** The quality of the alignment of images in the tilt-series sets an upper bound to the achievable resolution. When the gold nano-beads are scarce the chance for getting a good alignment drop. In the data sets used in the analysis in the preceding paragraphs, we had only two gold beads in the tilt-series. Though in this case we managed to get a decent alignment, in general it might not be enough. The more gold beads available for tracking, the higher the certainty of the alignment and potentially more parameters such as the distortion in the images could be solved for.

2. **Proper CTF deconvolution is crucial**. For a correct CTF model we have to know several parameters such as the defocus, fraction of amplitude contrast, chromatic and spherical aberration constants and various damping factors. It is not always easy to estimate those and, moreover, the defocus can vary substantially between the images, as well as within a single image of a thick specimen, or one that is at a high tilt. With nanocrystals we generally reconstruct a large volume that ideally includes the whole crystal. Failing to deconvolve the CTF has the potential to improperly flip the high resolution details and when the variation in the defocus is large, but only a single defocus is assumed, then it has the effect of canceling the high end of the spectrum. We saw in the filtered back-projection how the molecules largely resembled the true structure yet had important details missing and lower volume than expected.

3. **Lattice determination**. When determining the lattice, at the first step, we can only obtain an average lattice, while in actuality the crystal could have slight variations and deformations throughout. In the current implementation we basically use correlation to throw out all unit cells that do not conform well with the average unit cell that is derived from the average lattice assumption. The results are not always convincing and need some amount of luck to work. Some other ideas for better results are

    - Use the average unit cell as reference and try to align every other extraction into that, a sort of subtomogram averaging approach.
    - Extract all unit cells based on the average lattice and use clustering techniques to find groups with similar morphology.
    - Generalize the crystal unbending technique, that is used for 2d crystal images, to 3d images.

4. **Determining the symmetry group**. In this thesis we made no attempt to implement the symmetry group detection algorithm. In principle, after getting a decent 3d diffraction pattern then a lattice simulation approach could work, where all possible space groups are generated and the reflection conditions of every potential space group checked.

5. **Finding symmetry origin**. Finding the symmetry origin, as implemented now, is slow and reliable when the signal is strong. The speed and accuracy could be improved by doing a coarse search first, followed by a search on a finer grid. Another approach could be to do the search in the Fourier space.

6. **Validation**. Finally, and perhaps most importantly, the validation of the structure is something that can not be answered easily. In our proof of principle we knew the structure of lysozyme to atomic details and could use this as a reference to let us know when we have derailed. For a new structure it is obviously not as simple.

4.5 Programs

Historically, 3d crystals have been analyzed using diffraction based methods, which means that the amount of software tools available is heavily biased towards working with a diffraction pattern, where only the diffraction amplitudes are known instead of a real image of a crystal. The current work is dealing with a general tomographic reconstruction problem. Therefore, the plethora of tools developed for (cryo) electron tomography have been useful also in this context. The tools which are missing though, because of a presumed lack of need, are the ones that for a crystalline image would combine the methods developed for 3d diffraction pattern analysis in reciprocal space with methods developed for a real image processing in real space.

Starting the analysis from a real image gives the advantage of avoiding the crystallographic phase problem. Harnessing the power of Fourier transform we can hop between the two representations and though low-resolution tomography does not provide high resolution diffraction peaks, at least the phases are known for the ones we do see. In order to carry this project to its natural conclusion we had to go ahead and develop the programs that marry crystallographic approach with a real space image processing. This section describes the theoretical basis and the implementation details of the programs necessary for going from a tilt-series of a crystalline sample to a structure of a single molecule.

In the following we are going into the details of the programs that were specifically developed for handling tomograms that contain 3d crystals. The main flow of the involved procedure is the following

- Reconstruct a 3d volume
- Fourier transform the volume
- Find peaks
- Determine lattice vectors that define the unit cell
- Compute the average unit cell
- Use the average as a reference to segment out the crystal
- Compute average unit cell of the crystal
- Use the second average as a reference to select for specific parts of the crystal
- Compute average unit cell of the selected parts
- Find crystallographic symmetry origin
- Compute the third average again, this time applying also the symmetry operations
- Visualize and investigate the results

4.5.1 Peak Finder

To find the crystallographic lattice vectors, we first have to detect the high intensity peaks from within the Fourier transform. This is perhaps not as straightforward as it may sound, because in general we are dealing with maps that have very low SNR and relatively high varying background level.

The peak detection can be done directly on the Fourier transform itself, or it can be done on a processed version of it. In [25] a lattice detection algorithm was proposed for usage in electron crystallography of 3d protein nanocrystals. There the starting point is a 2d image, showing a diffraction pattern, and they propose that by calculating an auto-correlation function (ACF) of this pattern, the positions of the diffraction spots will remain unchanged, the missing peaks due to the systematic absences of a specific space group will become visible, the shadow cast by the backstop that blocks away the intense undiffracted direct beam will disappear, and the pattern will become centered by definition. The ACF is defined as follows

$$ACF(f) = \mathcal{F}^{-1}(\mathcal{F}(f)\mathcal{F}(f)^*)\,, \tag{4.1}$$

where $*$ represents the complex conjugate operation.

The ACF still has a varying background level which is much stronger in the low resolution area, and therefore directly setting an intensity threshold removes also the weaker high resolution diffraction peaks. In [72] they propose to do *adaptive background removal*, which amounts to subtracting a low-pass filtered version of the ACF from the ACF itself to remove the uneven background.

In [78] a more elaborate peak finding scheme is described, that is implemented in the software package *2dx*. The paper goes into details how to detect peaks in a Fourier transform of a 2d crystal image. They taper off the edges of the original image to lessen the edge artifacts in the transform followed by further masking of the streaking artifacts, masking out the very high resolution with the average gray value, and then doing a peak search by identifying local maxima. After that they shift they center the diffraction pattern image to each identified peak, sum the sifted images weighted by the intensity of the respective peak, and do one more peak search in the averaged pattern.

In the current work we also generalized these methods to 3d case, but eventually found that much of it was not needed for our data sets. Instead, we accept a noisy list of peaks, as long as it included the underlying lattice, and subjected this list to the lattice determination program that uses a robust algorithm that can handle a large amount of noise (described in section 4.5.2).

To perform a peak detection a custom software was written called *peak_finder*. It takes as input a 3d map, be it a modulus of a calculated FFT, or an ACF of that, or either one of those having the background adaptively removed, blanks out the part of the map that is below a certain threshold, and returns a list of voxel indices that have the highest intensity in their given neighborhood. To calculate the FFT we implemented the periodic + smooth edge artifact removal algorithm proposed in [64], that dampens a significant portion of the edge artifacts.

Implementation

Let us have a 3d map $\rho(\boldsymbol{x})$ with the dimensions of n_x, n_y and n_z. Let us define the set of all voxel coordinates as I

$$I = \{1, \ldots, n_x\} \times \{1, \ldots, n_y\} \times \{1, \ldots, n_z\} \tag{4.2}$$

The flow of the program is the following

1. Define the radius of the neighborhood $r \in \mathbb{N}$, and the type of the neighborhood (cube or a sphere). For a voxel with index $\boldsymbol{h}$ the neighborhood is defined as

   - **Cube**.
   $$N_{\boldsymbol{h}} = \{\boldsymbol{h} + \boldsymbol{k} : \boldsymbol{k} \in \{-r, \ldots, r\}^3\}, \tag{4.3}$$
   - **Sphere**.
   $$N_{\boldsymbol{h}} = \{\boldsymbol{h} + \boldsymbol{k} : \boldsymbol{k} \in \{-r, \ldots, r\}^3, \|\boldsymbol{k}\|_2 \leq r\}. \tag{4.4}$$

2. Loop over all of the voxels in ρ, and consider those voxels peaks which have the highest intensity value in their neighborhood. In this implementation we are not considering peaks that are too close to the volume boundary. Therefore, let us define the reduced set of voxels I_r that excludes all of the voxels closer to the boundary than neighborhood radius r

   $$I_r = \{1 + r, \ldots, n_x - r\} \times \{1 + r, \ldots, n_y - r\} \times \{1 + r, \ldots, n_z - r\}. \tag{4.5}$$

   The set of peaks is thus defined as

   $$P = \left\{ \boldsymbol{h} \in I_r : \max_{\boldsymbol{k} \in N_{\boldsymbol{h}}} \{\rho(\boldsymbol{k})\} = rho(\boldsymbol{h}) \right\}. \tag{4.6}$$

3. The peaks in P are all necessarily integers by the nature of discrete sampling. In the general case, where the lattice frequency is not a multiple of the fundamental frequency, the vectors do not end up exactly on the integer coordinates. The center points of the true peaks would be real numbers. *Peak_finder* does a non-linear curve fitting in 3d as implemented in *intel mkl* library to refine the center points of the peaks. In the current implementation we fit a 3d Gaussian curve f to the peak that has its intensity leaked out into the neighbouring voxels

   $$f(\boldsymbol{x}|k, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = k \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right), \tag{4.7}$$

   where k is a scaling factor,

   $$\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \tag{4.8}$$

   are the mean and variance of the Gaussian. Here we have taken the covariance terms in $\boldsymbol{\Sigma}$ to be 0, so we are fitting a spherical Gaussian function.

For each peak $\boldsymbol{p}_i$ in P the procedure consists of solving a nonlinear system of equations in the least squares sense. In the optimization we have 7 unknowns $\boldsymbol{\theta} = (k, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2)$ and $|N_{\boldsymbol{p}_i}|$ equations. So, we have a system

$$f(\boldsymbol{x}_j|\boldsymbol{\theta}) = \rho(\boldsymbol{x}_j)\,, \quad \boldsymbol{x}_j \in N_{\boldsymbol{p}_i} \tag{4.9}$$

and we are looking for $\boldsymbol{\theta}$ that minimizes the difference

$$\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{p}_i}} \left(f(\boldsymbol{x}_j|\boldsymbol{\theta}) - \rho(\boldsymbol{x}_j)\right)^2 . \tag{4.10}$$

Once the fit is done, $\boldsymbol{\mu}$ is chosen as the refined location of the peak. In simulations this sort of curve fitting improved the results, the peaks were lying much closer to a lattice line, and were less scattered around. For the real data the gain of doing the curve fitting was negligible, so it was seldom used. Since the Intel's mkl library is highly optimized there is not much computational cost on a modern machine.

4. Instead of Gaussian curve fitting the user can also choose to calculate the weighted average over the neighbourhood with weights being the intensity values. The refined position of a peak $\boldsymbol{p}_i$ is then

$$\frac{\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{p}_i}} \boldsymbol{x}_j \rho(\boldsymbol{x}_j)}{\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{p}_i}} \rho(\boldsymbol{x}_j)}\,. \tag{4.11}$$

For experimental data, considering all of the other sources of error, locating the center points of the peaks in exactly did not appear to affect the end result significantly.

5. Finally, the program outputs a list of peaks and a list of refined peaks

Experimental

We first Fourier transformed the back-projection map. The FFT did not show strong background at low resolutions and the background removal procedures did not give any benefit, therefore only the modulus of the Fourier transform of the image was input directly to the peak detection algorithm. The algorithm picked out around 600 peaks that clearly included the lattice but also a high amount of false positives (Fig 4.49)

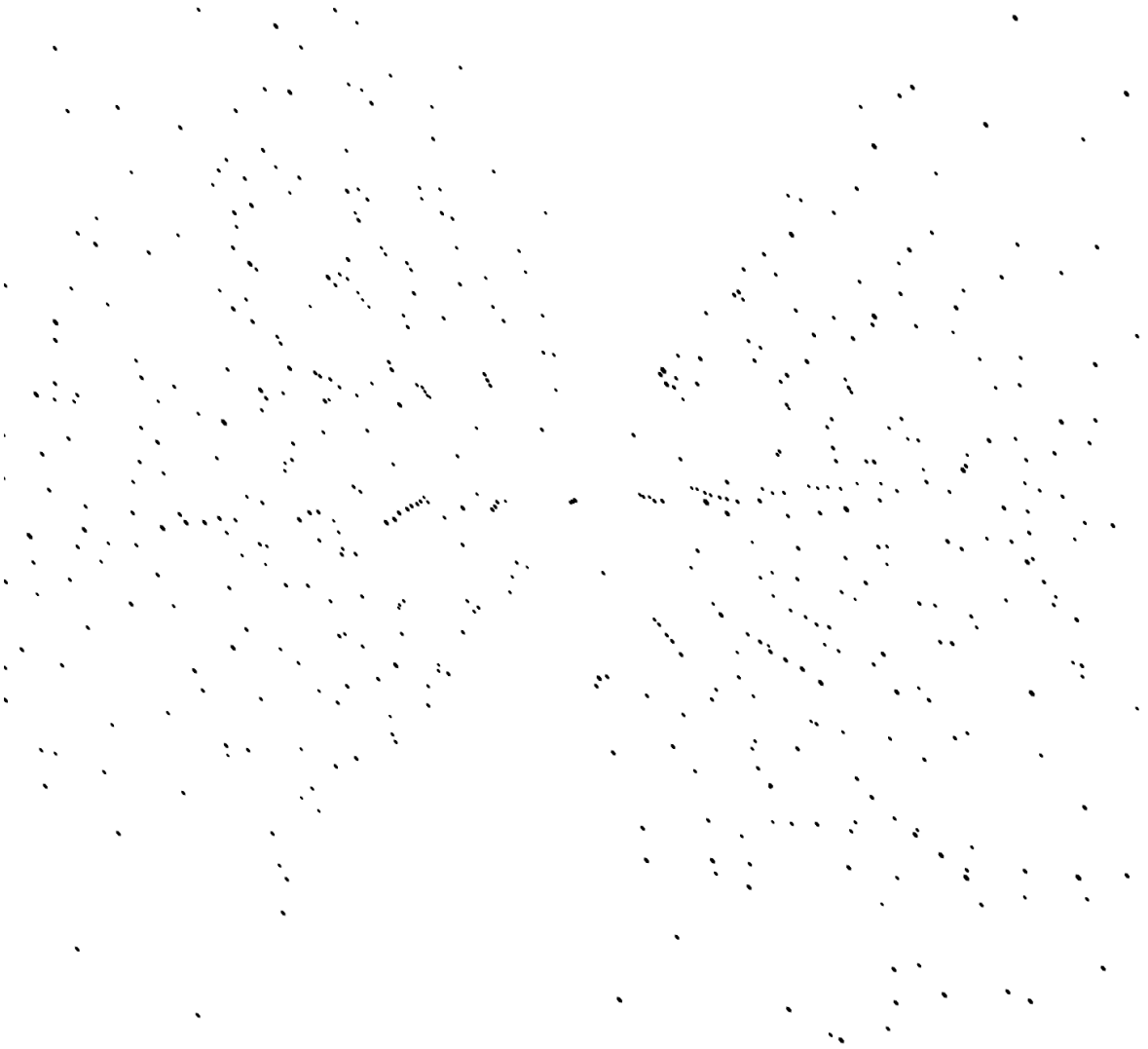

Figure 4.49 | The detected cloud of peaks. A true lattice is hidden within the noise, visible for a human eye but slightly more challenging to pick out for a computer algorithm.

4.5.2 From Peaks to Lattice

The program *peaks2lattice* is used to find the reciprocal lattice vectors based on a list of peaks. In [78] an algorithm for lattice vector determination from the FFT of electron micrographs of 2d protein crystals is described. In [25] they propose an algorithm for determining 2d lattice from a 2d measured electron diffraction pattern of 3d crystals. In both cases the search is only for two 2d vectors, whereas in this project we have a 3d calculated diffraction pattern of a 3d crystal and hence need three 3d vectors.

Here we present an algorithm where iteratively applying FFT, and at each step retaining only the brightest peaks, we manage to clear up a noisy cloud of peaks and bring out the underlying lattice.

Let us have a list of peaks P as returned by the previous program $peak\_finder$.

To go from this list of peaks to three 3d vectors we progress as follows

1. The origin $\boldsymbol{p}_0$ is determined as the peak in P which has the highest intensity value in the calculated FFT. This is always the 0-th order peak and is in the center of the FFT by definition

$$\boldsymbol{p}_0 = \underset{\boldsymbol{p}\in P}{\operatorname{argmax}}\{\rho(\boldsymbol{p})\}\,. \tag{4.12}$$

2. The following is repeated a set number of iterations. Usually 2 iterations will suffice.

   (a) Create a new binary map, with dimensions identical to the input map, place 1 where there is an identified peak and 0 elsewhere.

   (b) Calculate the inverse FFT of the binary map. This will take us to the real space.

   (c) Sort the resulting voxels based on their intensity.

   (d) Pick out a fixed number (around 10^3 or more, if the list of peaks is very noisy) of highest intensity voxels from the sorted list, and create a new binary map similarly to the first step.

   (e) FFT this new binary map and therefore go back to reciprocal space.

   (f) Sort the resulting voxels based on their intensity.

   (g) Pick out a fixed number of highest voxels from the sorted list.

   (h) repeat from (a)

3. On a final FFT of a binary map that goes from real to reciprocal space, a clean lattice has hopefully formed. Usually, the lattice does not have single distinct points but rather small blobs at the approximate location of the lattice points. The program takes the weighted average of the blob as a lattice point.

   With this algorithm the result is either a perfectly clean lattice or a clearly identifiable nonsense. There is no midway such as lattice points together with some amount of noisy peaks, because if there were, then one or more iterations would quickly take us to the mentioned two extremes. If the result is nonsense, then either the number of the kept highest intensity voxels has to be increased or the peak detection parameters fine-tuned in order to start with a cleaner set of peaks. Based on experience, if a human eye can see a lattice in the FFT then the algorithm will eventually find it, it's just a matter of optimizing program parameters

4. The goal of this step is to pick a set of 3 vectors, $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, such that they would form a right-handed basis that gives integer coordinates to the lattice points. We take the center of the lattice as origin and sort the rest of the vectors according to length (distance from the origin). It is important to note that the following algorithm works only when the result of the previous step is a perfect lattice.

   (a) Take the shortest vector as $\boldsymbol{a}$.

(b) Take the next shortest vector as $\boldsymbol{b}$ that is not approximately collinear with $\boldsymbol{a}$. The collinearity is decided by looking at the angle γ between $\boldsymbol{a}$ and $\boldsymbol{b}$.

$$\gamma = \cos^{-1} \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} \,. \tag{4.13}$$

Accept $\boldsymbol{b}$ if $\gamma \in [30°, 150°]$. The choice of the limits 30° and 150° is somewhat arbitrary and can be manipulated as a parameter.

(c) Take the next shortest vector as $\boldsymbol{c}$ that is not approximately coplanar with $\boldsymbol{a}$ and $\boldsymbol{b}$. The coplanarity is decided by looking at the ratio between the third and the first singular value from the singular value decomposition of a matrix $\boldsymbol{A}$, where

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_x & \boldsymbol{b}_x & \boldsymbol{c}_x \\ \boldsymbol{a}_y & \boldsymbol{b}_y & \boldsymbol{c}_y \\ \boldsymbol{a}_z & \boldsymbol{b}_z & \boldsymbol{c}_z \end{pmatrix} \,. \tag{4.14}$$

The singular value decomposition decomposes matrix $\boldsymbol{A}$ into a product of three matrices

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \,. \tag{4.15}$$

Matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are rotation matrices and $\boldsymbol{\Sigma}$ is a rectangular diagonal matrix with nonnegative real numbers σ_1, σ_2 and σ_3 on the diagonal, which are called the singular values. Singular values represent scaling along the coordinate axes and are ordered from largest to smallest.

The idea is that if $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are approximately coplanar, the third singular value σ_3 will be close to 0 or very small when compared to σ_1, because the first two vectors already span about the same subspace as all the three would. Therefore, if the ratio σ_3/σ_1 is smaller than a certain threshold ε, the vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are considered approximately coplanar and the search continues. Otherwise, output the matrix $\boldsymbol{A}$

Discussion and shortcomings

- The vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ have integer coordinates by definition which is not necessarily optimal. This will be dealt with by the next program that attempts to refine the lattice.

- The directions of the reciprocal vectors in the calculated FFT depend on the dimensions of the Fourier transformed volume. If the dimensions are not equal then even vectors that are all perfectly orthogonal in the real space will become skewed in the Fourier transform. In such cases it would be optimal perhaps during the vector search to transform the trial vectors to the real space and check the collinearity/coplanarity there.

Experimental example

A list of peaks from the *peak_finder* (Figure 4.49) is input to the lattice vector finder program. Figure 4.50 shows the proceeding of the lattice cleanup algorithm. Already

after the first iteration the noisy list of peaks has been filtered out so that only the periodic part remains. Figure 4.51 shows the final detected lattice overlaid on the initial noisy set of peaks.

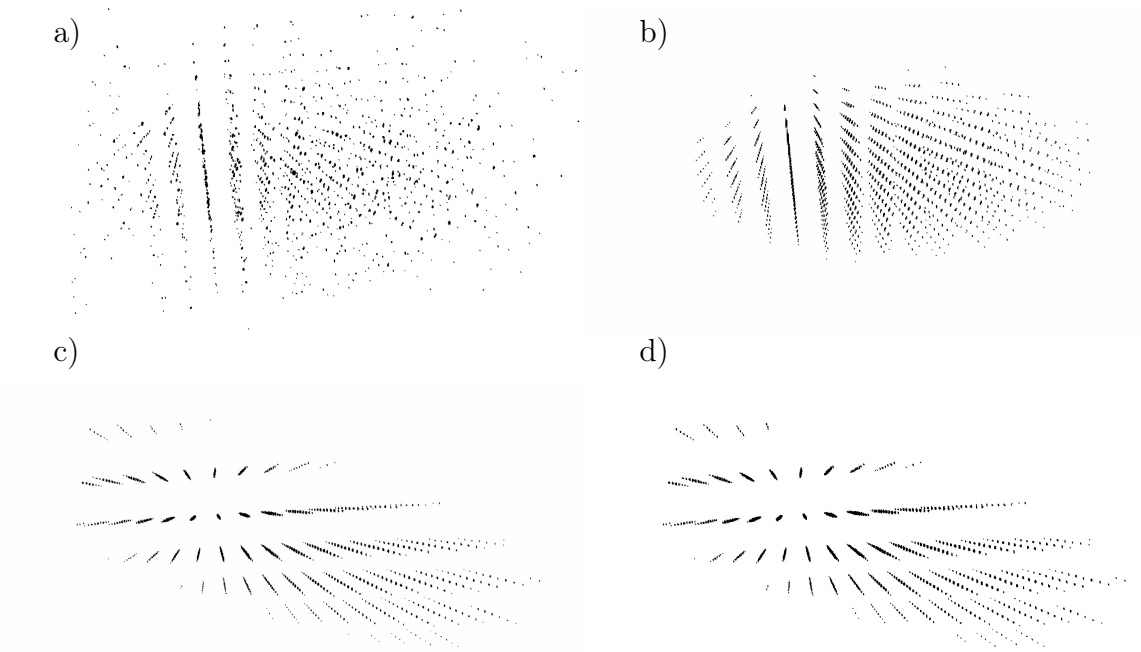

Figure 4.50 | Left column - 1st iteration. **Right column** - 2nd iteration. **Top row** - peaks in real space. **Bottom row** - peaks in reciprocal space. During the first iteration the peaks in real space are still slightly messy even though clear lattice structure is easily observable. Going back to reciprocal space after the 1st iteration leaves us with a perfect lattice. Second iteration in this case is unnecessary but is included here for illustrative purpose.

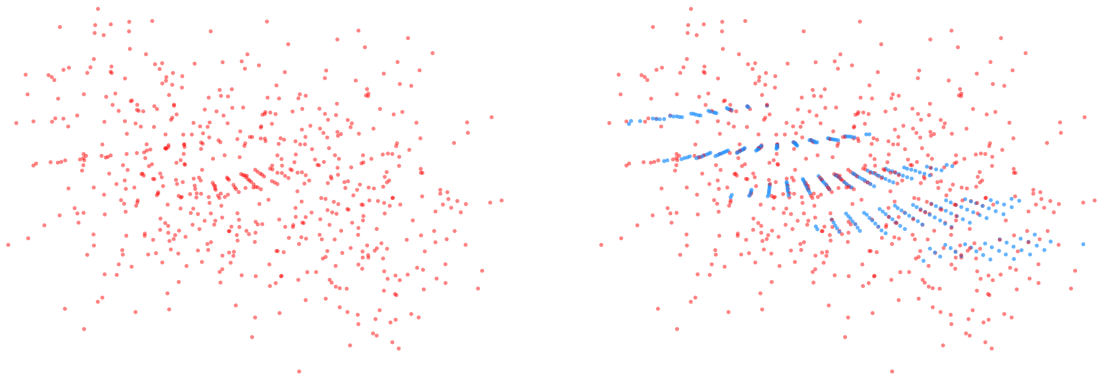

Figure 4.51 | (Left) detected peaks with red and (right) the reconstructed lattice with blue. The lattice appears to agree accurately with the detected peaks

4.5.3 Lattice Refining

As mentioned in the shortcomings of the lattice determination program, the returned lattice vectors have integer coordinates by definition. As such they give perfect integer coordinates to the first order lattice points, but will be more and more off when going further in the spectrum. The program *refine_lattice* is used to refine such lattice vectors to give an optimal fit in the least square sense to all of the peaks belonging to a lattice.

The program takes a list of peaks as an input, denoted by $\boldsymbol{P}$, that could be either the list of peaks from *peak_finder*, or the list of lattice points from *peaks2lattice*. The triplet of vectors returned by the lattice determination program *peaks2lattice*, denoted here by $\boldsymbol{A}_0$, are used as an initial solution. The model for the i-th peak is

$$\boldsymbol{P}_i = \boldsymbol{A}\boldsymbol{X}_i + e_i\,, \tag{4.16}$$

where $\boldsymbol{X}_i$ are integer coordinates and e_i is an error term. The goal is to minimize the total weighted error in the least squares sense and take the minimizer $\boldsymbol{A}^*$ as our refined solution

$$\boldsymbol{A}^* = \underset{\boldsymbol{A}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i e_i^2 = \underset{\boldsymbol{A}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \left(\boldsymbol{P}_i - \boldsymbol{A}\boldsymbol{X}_i\right)^2 . \tag{4.17}$$

Let $t = \{1, 2, \ldots n_{it}\}$. For the t-th iteration do:

1. Calculate the coordinates $\boldsymbol{X}$ of each peak in the basis of lattice vectors $\boldsymbol{A}_{t-1}$ by multiplying the list of peaks $\boldsymbol{P}$ by the inverse of $\boldsymbol{A}_{t-1}$ and rounding to the nearest integer
$$\boldsymbol{X} = round\left(\boldsymbol{P}\left(\boldsymbol{A}_{t-1}^{-1}\right)^T\right) \in \mathbb{Z}^{n\times 3}\,. \tag{4.18}$$

2. Calculate the residual for the i-th peak as the maximum absolute distance from the nearest integer over the three coordinates
$$\boldsymbol{R}_i = \max_{j=1,2,3} \left| \left(\boldsymbol{A}_{t-1}^{-1}\boldsymbol{P}_i\right)_j - \boldsymbol{X}_{ij} \right| . \tag{4.19}$$

3. Assign weights to each peak by the following rules
$$\boldsymbol{W}_i = \begin{cases} 0\,, & \text{if } \boldsymbol{R}_i > 0.25 \\ \dfrac{1}{\max\left\{\boldsymbol{R}_i,\, \delta\right\}}\,, & \text{otherwise}\,, \end{cases} \tag{4.20}$$
where δ is a small number and used to avoid dividing by 0. The rationale for taking weight equal to 0 when $\boldsymbol{R}_i > 0.25$ is that the initial lattice vectors $\boldsymbol{A}_0$ are expected to be close to the optimal. If a peak is further than 0.25 lattice vector lengths from its expected position then it is probably noise.

4. Calculate the next solution as a weighted least squares estimator
$$\boldsymbol{A}_t = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}^{-1}\boldsymbol{P}\,. \tag{4.21}$$

5. If
$$\|\boldsymbol{A}_t - \boldsymbol{A}_{t-1}\| < \varepsilon \tag{4.22}$$
or $t = n_{it}$ end the program and take
$$\boldsymbol{A}^* = \boldsymbol{A}_t\,, \tag{4.23}$$
otherwise go to the next iteration.

4.5.4 From Reciprocal Lattice to Real Lattice

The program *rec2real* is short, simple, and meant to do just one task. Namely, to convert the reciprocal lattice vectors to the real space lattice vectors. Let $\boldsymbol{A}' = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{c}')$ be the reciprocal lattice vectors and $\boldsymbol{A} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ be the real space vectors. Let the dimensions of the reconstructed volume $\boldsymbol{\rho}$ be n_x, n_y and n_z. Let s be the sampling rate, e.g. 2.261 Å per voxel. Then the real space vectors can be computed as crossproducts between the properly scaled reciprocal vectors.

Let $\boldsymbol{M}$ be the scaled reciprocal vectors
$$\boldsymbol{M} = \begin{pmatrix} \frac{1}{n_x} & 0 & 0 \\ 0 & \frac{1}{n_y} & 0 \\ 0 & 0 & \frac{1}{n_z} \end{pmatrix} \boldsymbol{A}'\,. \tag{4.24}$$
The real space vectors in voxels are then
$$\boldsymbol{a} = \frac{\boldsymbol{M_{.2}} \times \boldsymbol{M_{.3}}}{|\boldsymbol{M}|}\,, \quad \boldsymbol{b} = \frac{\boldsymbol{M_{.1}} \times \boldsymbol{M_{.3}}}{|\boldsymbol{M}|}\,, \quad \boldsymbol{c} = \frac{\boldsymbol{M_{.1}} \times \boldsymbol{M_{.2}}}{|\boldsymbol{M}|}\,. \tag{4.25}$$
The lengths in Ångstrom are $|\boldsymbol{a}s|$, $|\boldsymbol{b}s|$ and $|\boldsymbol{c}s|$ and the angles α, β, γ are
$$\alpha = \cos^{-1}\frac{\boldsymbol{b}\cdot\boldsymbol{c}}{|\boldsymbol{b}||\boldsymbol{c}|}\,, \quad \beta = \cos^{-1}\frac{\boldsymbol{a}\cdot\boldsymbol{c}}{|\boldsymbol{a}||\boldsymbol{c}|}\,, \quad \gamma = \cos^{-1}\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}\,. \tag{4.26}$$

4.5.5 Average Unit Cell

Once the lattice vectors are known, the next step is to calculate the average unit cell. The averaging of the unit cells can be done in three ways

1. Averaging all of the unit cells
2. Averaging only those cells that correspond to certain criteria, e.g. correlating highly with a reference
3. Steps 1 or 2 together with the application of symmetrization operations

The first step in the work-flow is to average over the whole reconstructed volume, because initially the segmentation of the crystal and information about its symmetry group and symmetry origin are not known. Averaging over the whole crystal then necessarily includes parts that conform to the determined periodicity, and parts that belong to the background outside of crystal and contribute noise.

Finding the unit cells

Inputs to the averaging program are the reconstructed map ρ, its dimensions, and the 3 real space lattice vectors. The first question to be answered is: what combinations of the lattice vectors are actually within the reconstructed volume? Since the vectors are pointing in an arbitrary direction. when compared to the coordinate system of the volume, it is not immediately obvious which combinations we need to take in order to find all unit cells that are actually withing the volume, and not miss any. A naive solution would be to start generating the lattice indices from the smallest order and successively going higher until we have found an order that does not include any lattice points that are within the volume. This can be expensive in 3d, as we would also be looking through lattice points that are entirely in the wrong quadrant, whereas the volume is completely within the first quadrant of the coordinate system. A less expensive solution is visualized on Figure 4.52 and goes as follows

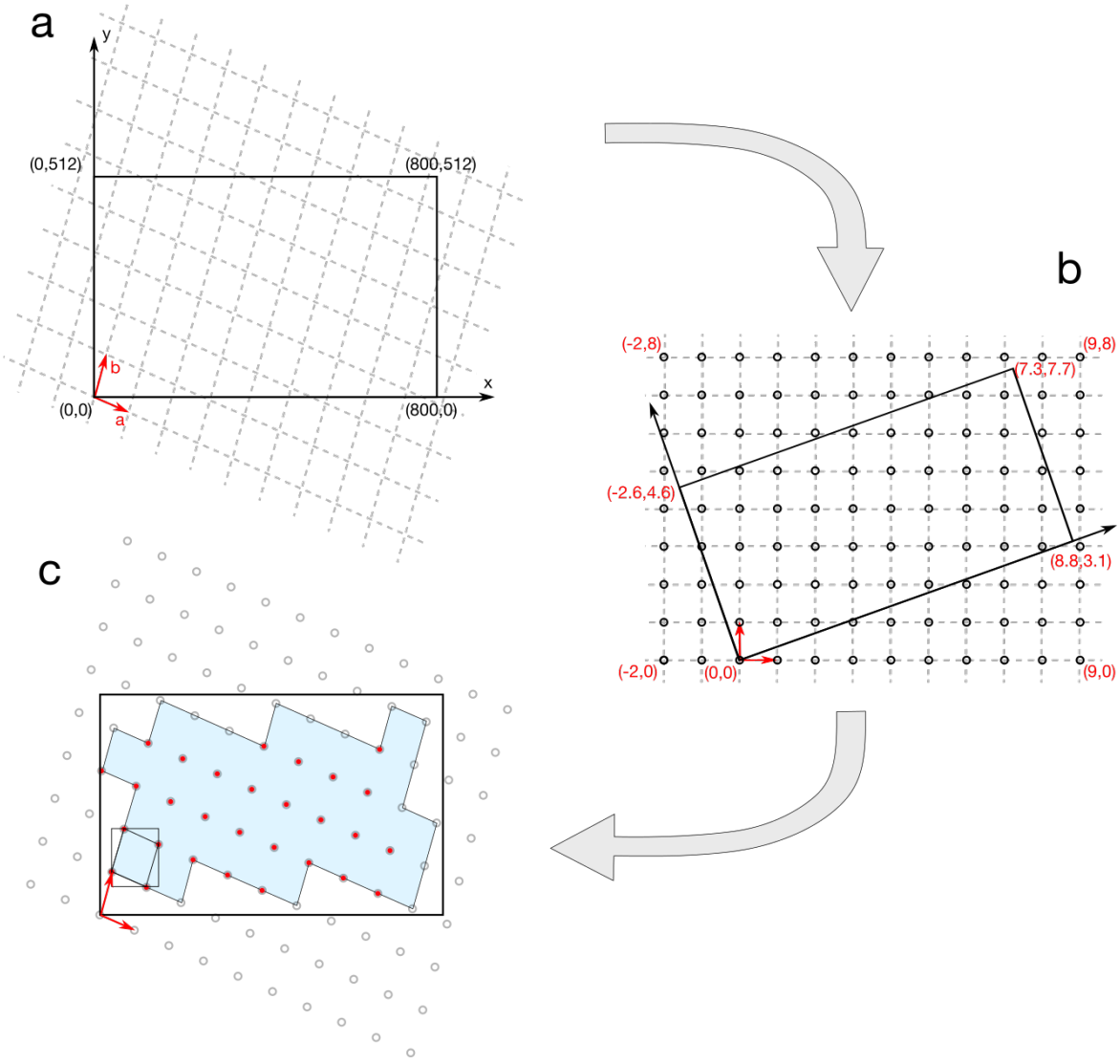

Figure 4.52 | Scheme of the process for finding every and not missing any unit cells within the reconstructed volume.

Let $\boldsymbol{A}$ be a 3x3 matrix with the lattice vectors in columns and let us have a reconstructed volume ρ with dimensions (n_x, n_y, n_z). The vectors $\boldsymbol{A}$ define a right-handed basis that represents an oblique coordinate system (Figure 4.52a). Let $\boldsymbol{C}$ be the coordinates of the corners of the volume

$$\boldsymbol{C}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & n_x & n_x & n_x & n_x \\ 1 & 1 & n_y & n_y & 1 & 1 & n_y & n_y \\ 1 & n_z & 1 & n_z & 1 & n_z & 1 & n_z \end{pmatrix} \tag{4.27}$$

Next let us transform the corner coordinates into the $\boldsymbol{A}$ coordinate system (multiplication with the inverse of $\boldsymbol{A}$), find the floor of the minimum and the ceiling of the maximum of every coordinate, and use those as limits for the lattice vector combinations (Figure 4.52b).

$$\boldsymbol{C}' = \boldsymbol{C}(\boldsymbol{A}^{-1})^T \tag{4.28}$$

and

$$\begin{aligned} a_{min} &= \lfloor \min\{\boldsymbol{C}'_{\cdot 1}\} \rfloor , \quad a_{max} = \lceil \max\{\boldsymbol{C}'_{\cdot 1}\} \rceil \\ b_{min} &= \lfloor \min\{\boldsymbol{C}'_{\cdot 2}\} \rfloor , \quad b_{max} = \lceil \max\{\boldsymbol{C}'_{\cdot 2}\} \rceil \\ c_{min} &= \lfloor \min\{\boldsymbol{C}'_{\cdot 3}\} \rfloor , \quad c_{max} = \lceil \max\{\boldsymbol{C}'_{\cdot 3}\} \rceil . \end{aligned} \tag{4.29}$$

Now we only have to loop through all of the combinations of integers that are between the found limits. Calculating the linear combinations of $\boldsymbol{A}$ columns with given coefficients gives a potential lattice point, that is either withing the reconstructed volume or not very far off. Then we check whether a rectangular box surrounding a unit cell at that give lattice point is entirely within the volume, or not (Figure 4.52c). If not, then discard the lattice point. This procedure will filter out a list of positions where a unique unit cell can be extracted and that is entirely within the reconstructed volume.

Averaging over the volume

Having found the lattice points coordinates from the previous step, the rest of the averaging is straightforward. We create a small volume with the dimensions that can exactly accommodate a desired number of unit cells. We then loop through every voxel in the small volume and replace it with the average of all of the voxels in the large volume ρ that correspond to this point in the unit cell. The lattice vectors are not necessarily whole numbers, therefore the lattice points are also real values which means we have to employ some interpolation. Every interpolation reduces the resolution so care has to be taken to avoid unnecessary interpolations and, as much as possible, refer back to the original non-interpolated data whenever processing further.

Averaging within a subregion

Similar to averaging over the whole volume, the list of possible lattice points can be filtered according to any criteria to include only a select group of unit cells in the average. Generally, the filtering occurs based on a correlation value. We can correlate

the overall average with each of the unit cells individually, obtain a distribution of correlation coefficients, and then select only the most highly correlating points.

Symmetrization

One of the bottle necks of the current approach is the symmetry origin determination. To find the symmetry origin we do an exhaustive search on as fine or coarse a grid as desired, to pinpoint a point in the unit cell that, when symmetry origin is placed there, gives the smallest average discrepancy measure between all of the other points that should in theory be equivalent due to symmetry relations. When the data is noisy the best point is not well defined and several local optima exist, which makes this approach slightly error prone. Due to high symmetry, even wrongly symmetrized unit cell looks mesmerizingly beautiful, even though the density itself might not make any sense. In our experience though, we could usually tell when the symmetry origin was far from correct, by noticing how well the individual molecules were separated and how continuous the density was.

4.5.6 Finding the Symmetry Origin

Finding the symmetry origin is one of the backbones of the whole software package. Without it only symmetrization in the P_1 space group involving only translational symmetry would be possible, as the origin here is arbitrary. For every other space group the precise locations of the symmetry elements, such as the axes of rotation, have to be determined. These are uniquely defined knowing the origin and the lattice vectors.

In our software package the search for symmetry origin is done in a naive way and is currently one of the weak points. Improving the approach should be addressed in the future.

Currently the flow of the program is the following:

1. **Input:** lattice vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, and a P_1 averaged unit cell ρ.
2. **Divide the unit cell into a grid**. The parameter k controls how fine the grid is

$$G(k) = \left\{ \boldsymbol{p} : \boldsymbol{p} = \phi_a \boldsymbol{a} + \phi_b \boldsymbol{b} + \phi_c \boldsymbol{c},\ \phi_v \in \left\{ \frac{0}{k|\boldsymbol{v}|}, \frac{1}{k|\boldsymbol{v}|}, \ldots, 1 \right\} \right\} , \tag{4.30}$$

   where $|\cdot|$ denotes the length of a vector in pixels.
3. **Loop over the grid**. For each grid point $\boldsymbol{p}$, place the symmetry origin at $\boldsymbol{p}$ and calculate the associated error metric $\varepsilon(\boldsymbol{p})$. The error $\varepsilon(\boldsymbol{p})$ is calculated in the following way
   (a) Loop over the unit cell ρ
   (b) For each voxel calculate the equivalent symmetrically related points, take their average and calculate the average squared difference.
   (c) Finally, sum the averages together.

$$\varepsilon(\boldsymbol{p}) = \sum_{\boldsymbol{x}} \sqrt{\frac{1}{n_{sym}} \sum_{i=1}^{n_{sym}} \left(S_i \rho(\boldsymbol{x}) - \mu_{\boldsymbol{x}}\right)^2}, \quad \text{where } \mu_{\boldsymbol{x}} = \frac{1}{n_{sym}} \sum_{i=1}^{n_{sym}} S_i \rho(\boldsymbol{x}), \quad (4.31)$$

where S_i stands for the i-th symmetry operation, that depends also on the symmetry origin $\boldsymbol{p}$ and takes care of the interpolation.

4. **Output**: The symmetry origin is chosen to be the point $\boldsymbol{p}^*$ that minimizes $\varepsilon(\boldsymbol{p})$

$$\boldsymbol{p}^* = \underset{\boldsymbol{p}}{\operatorname{argmin}}\, \varepsilon(\boldsymbol{p}). \quad (4.32)$$

For a high signal to noise ratio map this approach works perfectly, as is evident from the validation (section 4.6.3). For a low SNR map the approach often fails, as evident in the section describing the distribution of the symmetry origin and specifically on Figure 4.45. We see that the resolution is good in the $\boldsymbol{a}$-$\boldsymbol{b}$ plane and the origin distribution is narrowly peaked. Along the $\boldsymbol{c}$-axis, however, the distribution is spread out and the uncertainty about the true location of the origin is high.

4.6 Simulations

Important part of method development is having test cases and validation data, where the algorithm can be tested against a result that is known beforehand. In electron microscopy, validation data is not always available, especially so when studying something with a novel structure. Protein data bank features over 116000 high resolution models of various molecules from various organisms in many different compounds determined by various methods, which serves as a good reference for testing lower resolution tomography reconstructions. Provided that the molecules in a non-crystalline state or in a nanocrystal look approximately identical to the structures derived by analyzing large crystals we can use x-ray models as a reference.

The availability of TEM simulators makes it possible to construct a virtual sample from known atomic coordinates, feed it into the simulator and generate tilt-series under conditions similar to real world data collection. More can be read at [79] about the TEM simulator that was used in this work.

In this work we are not really concerned with a general tomographic reconstruction problem but rather with image processing methods that are further down the pipeline after the reconstruction. Therefore, we are using the simulator to only construct a perfect 3d electron density representation of the sample, serving as the ground truth reconstruction.

4.6.1 Generating Virtual Crystals

To generate a virtual crystal, the 3d coordinates of each atom in the molecule are needed. These are easily obtained from the protein databank (PDB) in the form of a pdb file. Since we were working with lysozyme molecule, we picked a random hen egg white lysozyme structure solved in $P4_32_12$ space group (code 1DPX), from the more than 1000 structure deposited. In addition to atomic coordinates, the pdb file

includes much more information, including the symmetry operations, crystallographic parameters and details about the experimental conditions.

For the purpose or rebuilding a nanocrystal, we wrote a short program in Fortran that takes a pdb file as input, extracts the atomic coordinates and the symmetry operations, and replicates the coordinates in 3d space according to the desired number of unit cells. Once the configuration file is generated, it can be fed into the simulator and subsequently used to generate either a realistic tomographic tilt-series or just a ground truth 3d electron density map of a perfect crystal. A 10x10x10 unit cells large virtual crystal is shown on Figure 4.53.

Alternatively, it was discovered later that the pdb file can also be read into the 3d visualization software Chimera that handles pdb files elegantly and can recreate the crystal by replicating the pdb structure according to the symmetry specified within it. It can also easily generate the electron density representation of a model to a desired resolution, rendering our reimplementation useless.

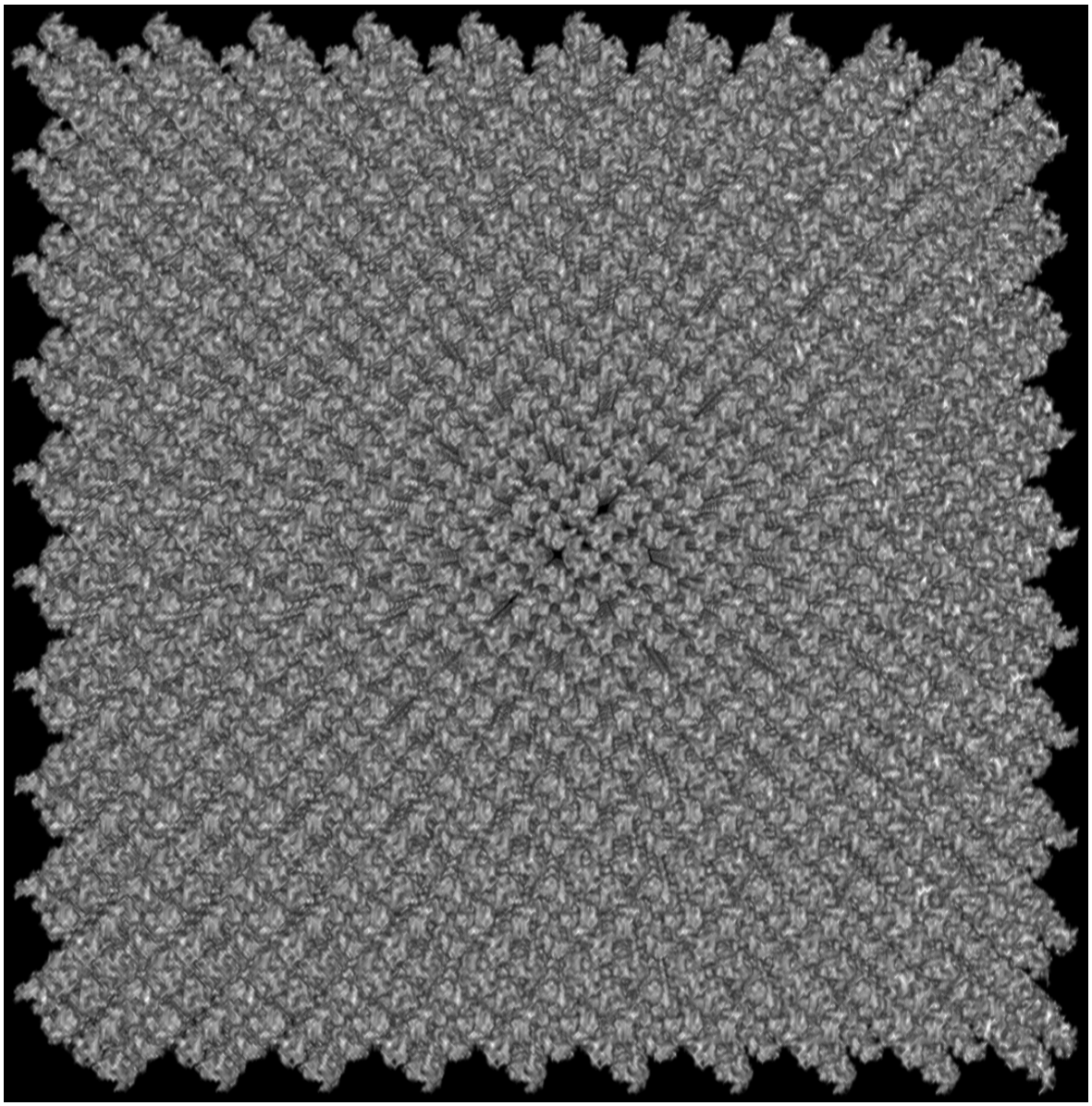

Figure 4.53 | Simulated perfect lysozyme crystals extending 10 unit cells in each crystallographic direction. The image is taken looking down the real space c axis, which coincides with the direction of the 4-fold screw axis. The 4-fold rotational symmetry is visible in the image.

4.6.2 Finding Lattice Vectors on Simulated Data

The simulated image on Fig 4.53 is tilted with respect to the coordinate axes to make the scenario more general. But because it is perfect and there is no background noise, the Fourier transform on Fig 4.54 is unrealistically clean. On a real crystal, on the other hand, the boundaries are never as sharp and well defined as in the simulated case and therefore show much weaker edge artifacts that manifest on Fig 4.54 as strong cross-shaped features, corresponding to the FFT of the shape function of the crystal. Nevertheless, it provides a useful different test case for the peak-detection algorithm.

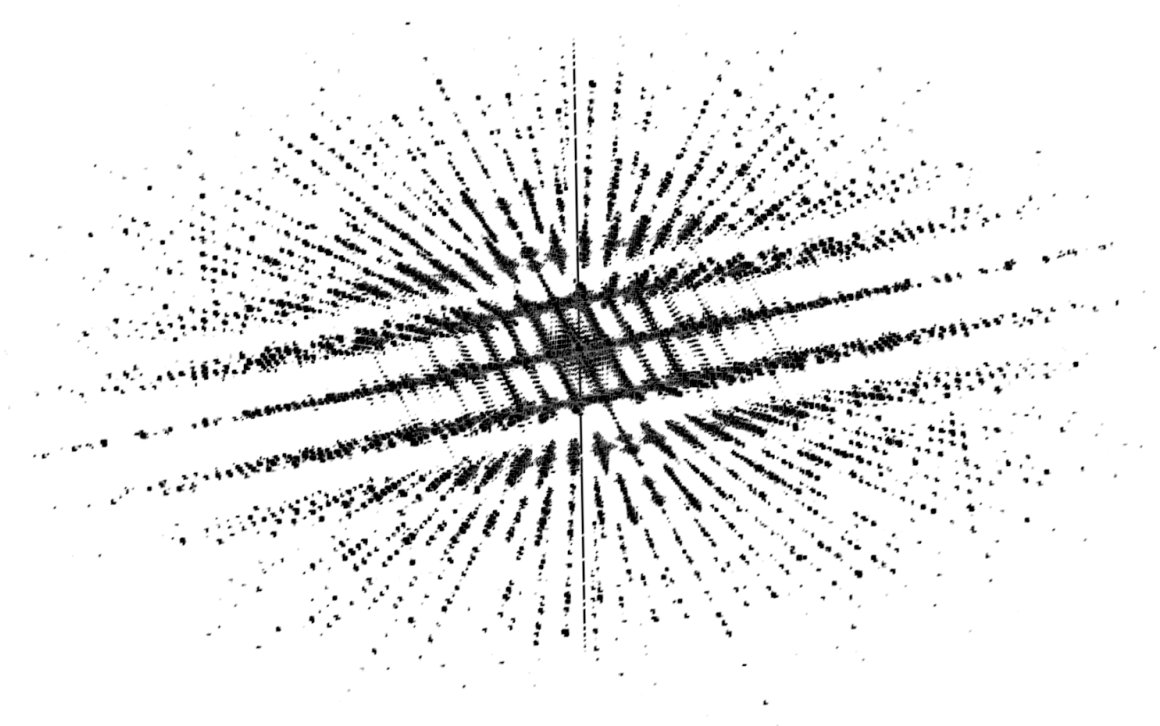

Figure 4.54 | 3d Fourier transform of the simulated perfect crystal. Strong edge artifacts are visible but no background noise.

The results after running the peak and lattice detection programs on the FFT of the simulated crystals are presented on Fig 4.55. The found crystallographic parameters are accurate to a tenth of an Ångstrom and are shown on Table 4.1.

Case	a	b	c	α	β	γ
Determined	77.056	77.023	37.231	90.037	90.014	89.991
True	77.050	77.050	37.210	90.000	90.000	90.000

Table 4.1 | Comparison of determined unit cell parameters and the true unit cell parameters. The determined parameters are accurate to at least a tenth of an Ångstrom.

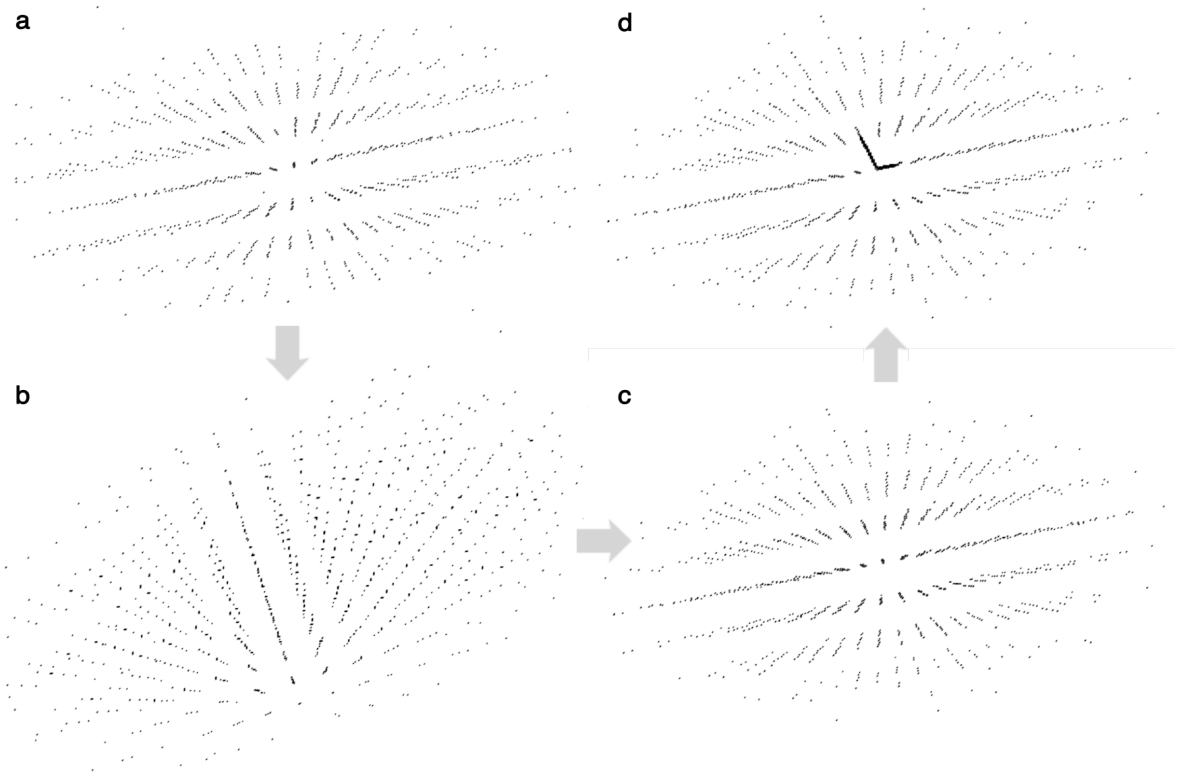

Figure 4.55 | Peak detection and lattice vector determination. **a** - The detected peaks from Fig 4.54. Some spurious peaks can be seen, but the overall result is near perfect. **b** - The first iteration of the lattice detection algorithm. The image shows the lattice in real space. **c** - The second step of the first iteration, now back in reciprocal space. The spurious peaks visible on **a** are now gone. **d** - The same as **c** with the determined lattice vectors added.

4.6.3 Averaging and Symmetrization on Simulated Data

After unit cell parameters are determined, a total averaged over all of the non-overlapping unit-cell-sized blocks from the density map Fig 4.53 is computed, including the background. The average unit cell looks identical to the original map Fig 4.56a. Next step is to cross-correlate this average to all of the extracted blocks and calculate the average again by throwing out blocks with low correlation. The result on Fig 4.56b is again identical to the original map, because in our perfect crystal case the background is identically 0.

Finally, a symmetry origin search is performed on Fig 4.56b assuming crystallographic space group $P4_32_12$. Symmetry origin defines the positions of the symmetry elements and upon applying the symmetry operations we observe no changes and arrive yet again back at the original starting map as expected Fig 4.56c. This concludes the validation of the programs. Note that some loss of resolution is inevitable because the simulated map was sampled at a Nyquist frequency and the averaging and symmetrization steps include linear interpolation.

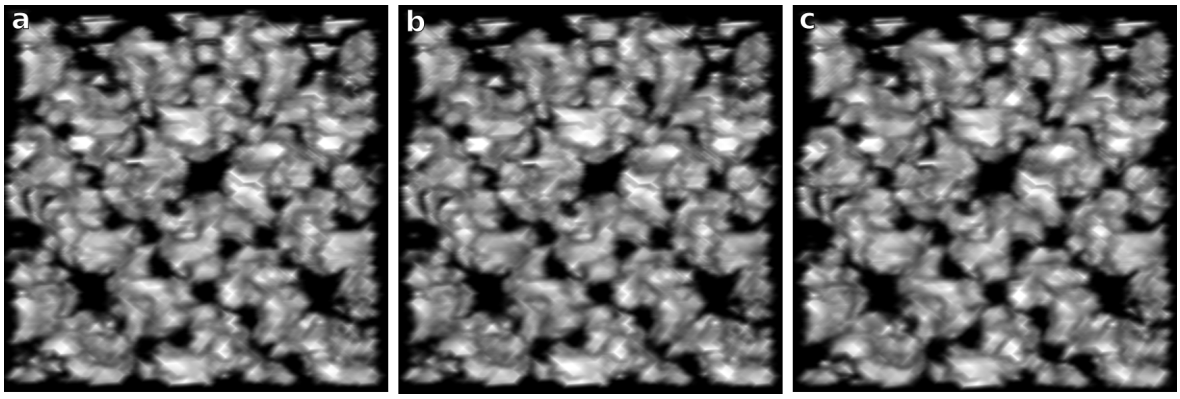

Figure 4.56 | Averaged and symmetrized unit cells. **a** - The total average of every unit-cell-sized block from the density map, including the background (which is identically 0 in this case). **b** - The average unit cell over only the crystal, excluding the background. The result is identical. **c** - Same as **b** but with applied symmetry operations. As the crystal was perfect to begin with there is no changes after the averaging and symmetrization, which serves as a validation test for the algorithms.

Local averaging

The local averaging and symmetrization program also passed the validation test without problems. The results are shown on Figure 4.57.

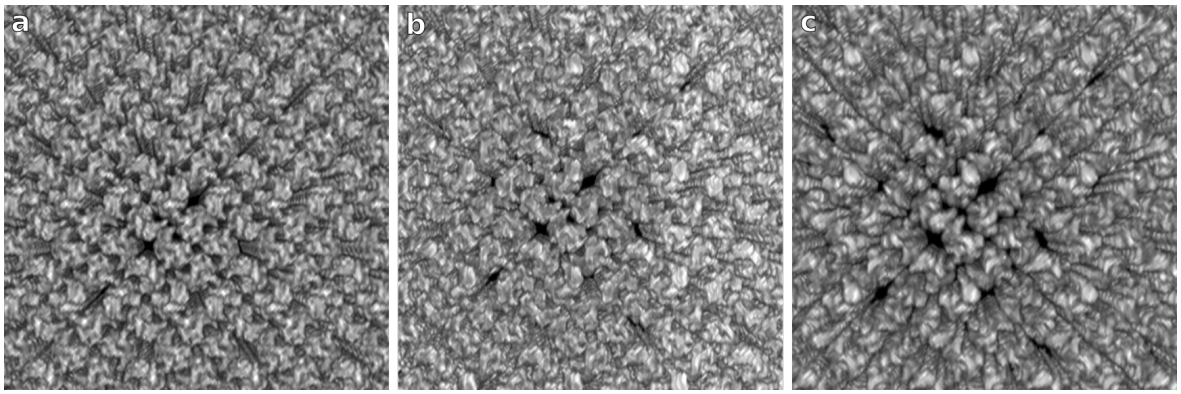

Figure 4.57 | Solid rendering of the simulated perfect lysozyme crystal. **a** - Cut-out of the center piece from a large simulated crystal (Fig 4.53). The symmetry is perfect. **b** - Locally averaged version of **a** according to the determined unit cell vectors, where each unit cell is replaced with the average of its immediate neighbours. The symmetry remains perfect. As the map in **a** is sampled at a Nyquist frequency, a small loss of resolution in the subsequent image processing steps due to interpolation is inevitable. Part of the density appears to have stronger intensity and the reason is that the moving average of the unit cells was done without segmenting the crystal. The unit cells closer to the edge of the crystal have cells that consist entirely of the blank background included in the average. **c** - Same as **b** but now the symmetry operations of the space group $P4_32_12$ have also been applied. The symmetry origin is determined by the *find_symmetry_origin* program. The symmetry still remains perfect, but the resolution has decreased due to linear interpolation.

Conclusions from the Validations

Testing the algorithms on simulated perfect crystal with known molecular structure demonstrated that our algorithms work well at the medium level of resolution that we strove for. It shows that the results from peak detection and lattice vector determination are reliable and there is a hope that the symmetry origin searching algorithm will pick up the correct point. The necessity to interpolate brings forth a loss of resolution, but we have given our best to keep the interpolation steps at minimum and always refer back to the original data when possible.

Conclusions and Future Perspectives

Protein 3d structure determination is a challenge no matter the method used. Electron tomography of protein nanocrystals is certainly no magic bullet, but as a complementary method to the preexisting ones it has a great potential for providing information about much more than just the structure of the repeating unit. It has applications ranging from medium resolution structural information, that might prove useful for other, possibly diffraction based, methods in the attempt to achieve high resolution, to studying the crystals in general and getting information about the morphology and disorder. Electron tomography has the power to retain such information.

In the thesis we approached the structural analysis of lysozyme nanocrystals from various angles. We harnessed Fourier transform to get information about the periodicity, the size and shape of the unit cell, and the average structure. We used correlation based methods as well as dimensionality reduction and clustering methods to find similar unit cells, that would give information about the order and organization within the crystal. We investigated the distribution of the unit cell vectors by dividing the large reconstruction into many subvolumes and processing each one independently. We studied the distribution of symmetry origin and saw it was well determined in two dimensions, but poorly in the third dimension.

The undertaken project showed that without perfect data there are many hurdles one has to jump over before the structure can be obtained. For a perfect crystal the process was straightforward and no tuning of the programs and parameters was needed, as was evidenced from the validation of the methods with artificial data.

The structure of the molecule was not the only goal we were chasing, because after all, the structure of lysozyme has been known for ages. If the high resolution structure is what is really desired, then investing lots of effort into sample preparation and data collection will surely pay off and yield crystals with superior quality that are not hampered by the same data-dependent issues. Much is already known about the structure of lysozyme down to atomic details, but here we developed tools that gave information about its life within a not so well ordered nanocrystal. The imperfections of real data might sometimes prove interesting in their own right. While each result individually might have errors in them, all the results together are probably a reliable indication that there was indeed variation in the data and the crystal under scrutiny was not uniform throughout.

When in a general case it is difficult to say anything about the resolution in a tomogram, then for a nanocrystal the periodicity-caused spots in the Fourier space provided a clear indication of the maximum resolution. We saw that in specific directions the resolution was incredibly high, in fact higher than we anticipated, but unfortunately

not all directions were created equal. Plausibly, due to radiation damage, and definitely, due to the missing wedge in the beam direction, the resolution was significantly decreased in certain other directions. Taking advantage of the high symmetry of the particular space group, however, helped to fill in the missing information gaps and even out the resolution, at the expense of losing the high end of the spectrum. This is one of the obvious goals for the future - to enhance the symmetrization algorithms in order to retain the high resolution in all directions.

Right now the protein nanocrystal tomography as a method is still far from maturity and there are dozens of obstacles that need to be overcome to make the method more usable and reliable. Some of those include

- Improvements in the sample preparation techniques.
- More accurate tomographic reconstruction and CTF modeling, taking into account the decrease of resolution caused by the radiation damage in the later tilts, the gradient of defocus through the volume, as well as the variation of defocus over the tilts. Being able to nail down the CTF deconvolution related aspects perfectly will definitely increase resolution and interpretability of the map.
- In the current thesis we knew the structure of lysozyme beforehand which helped to guide the decisions and let us know when we have reached a good reconstruction. For a protein with unknown structure it will not be so obvious. We also could guess with a high probability of success that the space group of the crystal was $P4_32_12$, as it is one of the most common one in the literature, and the observed data seemed to confirm it. Again, for a new protein, the question of the space group will have to be investigated further.
- Better classification methods would make the grouping of mosaic blocks more reliable, and the averaging could be done by taking these deviations into account. It would be akin to 3d unbending then, the 2d version of which is commonly used in processing 2d protein crystals.
- Currently the search for the exact location of the symmetry origin was unreliable for the real experimental data, mainly because of the lower resolution in z-direction. The method worked perfectly, however, for the simulated ideal noise free data. To make the method more reliable, the symmetry origin determination should be improved.

Time and more data will tell what the future holds for the method of protein nanocrystal tomography, and to what extent the issues we faced in the current project were data dependent or intrinsic to the paradigm. As it stands now, we saw that high resolution exists in the tomogram, but we were unable to use it to its full extent. Perfecting the methods will unquestionably bring forth an increase in the overall resolution and help shed light to the aspects of the world around us that have so far evaded us.

Bibliography

1. Alberts, B. *Molecular Biology of the Cell* (Garland Science, 2008).
2. Stevenson, H. P., Makhov, A. M., Calero, M, Edwards, A. L., Zeldin, O. B., *et al.* Use of transmission electron microscopy to identify nanocrystals of challenging protein targets. *Proceedings of the National Academy of Sciences* **111,** 8470–8475 (2014).
3. De la Cruz, M. J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., *et al.* Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nature Methods.*
4. Nannenga, B. L. & Gonen, T. Protein structure determination by MicroED. *Current Opinion in Structural Biology* **27,** 24–31 (2014).
5. Glusker, J. & Trueblood, K. *Crystal Structure Analysis: A Primer* ISBN: 9780199576340 (OUP Oxford, 2010).
6. Giacovazzo, C. Fundamentals of crystallography (2002).
7. Kendrew, J. C., Bodo, G, Dintzis, H. M., Parrish, R. G., Wyckoff, H, *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181,** 662–666 (1958).
8. Rupp, B. Predictive models for protein crystallization. *Methods* **34,** 390–407 (2004).
9. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology* **18,** 581–586 (2008).
10. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of molecular biology* **305,** 567–580 (2001).
11. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., *et al.* The protein data bank. *Nucleic acids research* **28,** 235–242 (2000).
12. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature reviews Drug discovery* **5,** 993–996 (2006).
13. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics* **28,** 171–171 (1995).

14. Emma, P, Akre, R, Arthur, J, Bionta, R, Bostedt, C, *et al.* First lasing and operation of an ångstrom-wavelength free-electron laser. *Nature Photonics* **4,** 641–647 (2010).

15. Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470,** 73–77 (2011).

16. Hunter, M. S. & Fromme, P. Toward structure determination using membrane–protein nanocrystals and microcrystals. *Methods* **55,** 387–404 (2011).

17. Kirian, R. A., White, T. A., Holton, J. M., Chapman, H. N., Fromme, P., *et al.* Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals. *Acta Crystallographica Section A: Foundations of Crystallography* **67,** 131–140 (2011).

18. Bogan, M. J. X-ray Free Electron Lasers Motivate Bioanalytical Characterization of Protein Nanocrystals: Serial Femtosecond Crystallography. *Analytical Chemistry* **85,** 3464–3471 (2013).

19. Abdallah, B. G., Kupitz, C., Fromme, P. & Ros, A. Crystallization of the Large Membrane Protein Complex Photosystem I in a Microfluidic Channel. *ACS Nano* **7,** 10534–10543 (2013).

20. Abdallah, B. G., Chao, T.-C., Kupitz, C., Fromme, P. & Ros, A. Dielectrophoretic Sorting of Membrane Protein Nanocrystals. *ACS Nano* **7,** 9129–9137 (2013).

21. Leng, J. & Salmon, J.-B. Microfluidic crystallization. *Lab on a Chip* **9,** 24 (2009).

22. Unwin, P. N. T. & Henderson, R. Molecular structure determination by electron microscopy of unstained crystalline specimens. *Journal of molecular biology* **94,** 425–440 (1975).

23. Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., *et al.* Lipid–protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* **438,** 633–638 (2005).

24. Vink, M., Derr, K. D., Love, J., Stokes, D. L. & Ubarretxena-Belandia, I. A high-throughput strategy to screen 2D crystallization trials of membrane proteins. *Journal of structural biology* **160,** 295–304 (2007).

25. Jiang, L., Georgieva, D., Nederlof, I., Liu, Z. & Abrahams, J. P. Image Processing and Lattice Determination for Three-Dimensional Nanocrystals. *Microscopy and Microanalysis* **17,** 879–885 (2011).

26. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *ELife* **2** (2013).

27. Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nature Methods* **11,** 927–930 (2014).

28. Sawaya, M. R., Rodriguez, J., Cascio, D., Collazo, M. J., Shi, D., *et al.* Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proceedings of the National Academy of Sciences* **113,** 11232–11236 (2016).

29. Nannenga, B. L. & Gonen, T. MicroED opens a new era for biological structure determination. *Current Opinion in Structural Biology* **40,** 128–135 (2016).

30. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504,** 107–112 (2013).

31. Yu, X., Jin, L. & Zhou, Z. H. 3.88A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* **453,** 415–419 (2008).

32. Crowther, R. A., AMOS, L. A., Finch, J. T., De Rosier, D. J. & Klug, A. Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Nature* **226,** 421–425 (1970).

33. Caspar, D. L. & Klug, A. *Physical principles in the construction of regular viruses Cold Spring Harbor symposia on quantitative biology* **27** (1962), 1–24.

34. Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., *et al.* 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science* **348,** 1147–1151 (2015).

35. DeRosier, D. J. & Moore, P. B. Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *Journal of molecular biology* **52,** 355–369 (1970).

36. Crowther, R. A. & Klug, A. Structural analysis of macromolecular assemblies by image reconstruction from electron micrographs. *Annual review of biochemistry* **44,** 161–182 (1975).

37. Nederlof, I., Li, Y. W., van Heel, M. & Abrahams, J. P. Imaging protein three-dimensional nanocrystals with cryo-EM. *Acta Crystallographica Section D: Biological Crystallography* **69,** 852–859 (2013).

38. Henderson, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences* **110,** 18037–18041 (2013).

39. Sandin, S., Öfverstedt, L.-G., Wikström, A.-C., Wrange, Ö. & Skoglund, U. Structure and Flexibility of Individual Immunoglobulin G Molecules in Solution. *Structure* **12,** 409–415 (2004).

40. Skoglund, U. & Daneholt, B. Electron microscope tomography. *Trends in Biochemical Sciences* **11,** 499–503 (1986).

41. Crowther, R. A., DeRosier, D. J. & Klug, A. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **317,** 319–340 (1970).

42. Gilbert, P. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology* **36,** 105–117 (1972).

43. Norlén, L, Öktem, O & Skoglund, U. Molecular cryo-electron tomography of vitreous tissue sections: current challenges. *Journal of Microscopy* **235,** 293–307 (2009).

44. Dubochet, J., Adrian, M., Chang, J.-j., Homo, J.-C., Lepault, J., *et al.* Cryo-electron microscopy of vitrified specimens. *Quarterly reviews of biophysics,* 1–100 (1988).

45. Studer, D., Humbel, B. M. & Chiquet, M. Electron microscopy of high pressure frozen samples: bridging the gap between cellular ultrastructure and atomic resolution. *Histochemistry and Cell Biology* **130,** 877–889 (2008).

46. Bongini, L, Fanelli, D, Piazza, F, De Los Rios, P, Sandin, S, *et al.* Freezing immunoglobulins to see them move. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 6466–6471 (2004).

47. Goris, B, Van den Broek, W, Batenburg, K. J., Mezerji, H. H. & Bals, S. Electron tomography based on a total variation minimization reconstruction technique. *Ultramicroscopy* **113,** 120–130 (2012).

48. Leary, R., Saghi, Z., Midgley, P. A. & Holland, D. J. Compressed sensing electron tomography. *Ultramicroscopy* **131,** 70–91 (2013).

49. Gopinath, A., Xu, G., Ress, D., Öktem, O., Subramaniam, S., *et al.* Shape-based regularization of electron tomographic reconstruction. *IEEE transactions on medical imaging* **31,** 2241–2252 (2012).

50. Skoglund, U., Öfverstedt, L.-G., Burnett, R. M. & Bricogne, G. Maximum-entropy three-dimensional reconstruction with deconvolution of the contrast transfer function: a test application with adenovirus. *Journal of structural biology* **117,** 173–188 (1996).

51. Rullgård, H, Öktem, O & Skoglund, U. A componentwise iterated relative entropy regularization method with updated prior and regularization parameter. *Inverse Problems* **23,** 2121–2139 (2007).

52. Jaynes, E. T. Information theory and statistical mechanics. *Physical review* **106,** 620 (1957).

53. Öktem, O. In *Handbook of Mathematical Methods in Imaging* 937–1031 (Springer New York, 2015). ISBN: 978-1-4939-0790-8.

54. Csiszar, I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The annals of statistics,* 2032–2066 (1991).

55. Briggs, J. A. Structural biology in situ—the potential of subtomogram averaging. *Current Opinion in Structural Biology* **23,** 261–267 (2013).

56. Öfverstedt, L.-G., Zhang, K., Isaksson, L. A., Bricogne, G. & Skoglund, U. Automated correlation and averaging of three-dimensional reconstructions obtained by electron tomography. *Journal of structural biology* **120,** 329–342 (1997).

57. Skoglund, U, Andersson, K, Strandberg, B & Daneholt, B. Three-dimensional structure of a specific pre-messenger RNP particle established by electron microscope tomography. *Nature* **319,** 560–564 (1986).

58. Bracewell, R. N. *The Fourier Transform and Its Applications* (McGraw Hill, 2000).

59. Williams, D. B. & Carter, C. B. *Transmission Electron Microscopy: A Textbook for Materials Science* (Springer, 2009).

60. Reimer, L. *Transmission electron microscopy: physics of image formation and microanalysis* (Springer, 2013).

61. Lawrence, A., Bouwer, J. C., Perkins, G. & Ellisman, M. H. Transform-based backprojection for volume reconstruction of large format electron microscope tilt series. *Journal of structural biology* **154,** 144–167 (2006).

62. Penczek, P. A. Chapter one-fundamentals of three-dimensional reconstruction from projections. *Methods in enzymology* **482,** 1–33 (2010).

63. Mahmood, F., Toots, M., Öfverstedt, L.-G. & Skoglund, U. Algorithm and Architecture Optimization for 2D Discrete Fourier Transforms with Simultaneous Edge Artifact Removal (Submitted).

64. Moisan, L. Periodic plus smooth image decomposition. *Journal of Mathematical Imaging and Vision* **39,** 161–179 (2011).

65. Mahmood, F., Toots, M., Öfverstedt, L.-G. & Skoglund, U. *2D Discrete Fourier Transform with simultaneous edge artifact removal for real-time applications* (IEEE International Conference on Field Programmable Technology (FPT), 2015), 236–239. ISBN: 978-1-4673-9091-0.

66. Crowther, R. A. & Klug, A. Three dimensional image reconstruction on an extended field—a fast, stable algorithm (1974).

67. Mahmood, F., Öfverstedt, L.-G. W. & Skoglund, B. U. *Extended field iterative reconstruction technique (efirt) for correlated noise removal* US Patent App. 14/770,245 (USPTO, WIPO, JPO, 2014).

68. Hansen, P. *Discrete Inverse Problems: Insight and Algorithms* (Society for Industrial and Applied Mathematics, 2010).

69. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., *et al.* Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **206,** 757–761 (1965).

70. Yonekura, K., Braunfeld, M. B., Maki-Yonekura, S. & Agard, D. A. Electron energy filtering significantly improves amplitude contrast of frozen-hydrated protein at 300kV. *Journal of structural biology* **156,** 524–536 (2006).

71. Bharat, T. A. M., Russo, C. J., Löwe, J., Passmore, L. A. & Scheres, S. H. W. Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging. *Structure/Folding and Design* **23,** 1743–1753 (2015).

72. Jiang, L., Georgieva, D., IJspeert, K. & Abrahams, J. P. *An intelligent peak search program for digital electron diffraction images of 3D nano-crystals 2nd International Congress on Image and Signal Processing (CISP'09)* (2009), 1–5.

73. Kriminski, S, Caylor, C., Nonato, M., Finkelstein, K. & Thorne, R. Flash-cooling and annealing of protein crystals. *Acta Crystallographica Section D: Biological Crystallography* **58,** 459–471 (2002).

74. Maaten, L. v. d. & Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* **9,** 2579–2605 (2008).

75. Clark, J. N., Ihli, J., Schenk, A. S., Kim, Y.-Y., Kulak, A. N., *et al.* Three-dimensional imaging of dislocation propagation during crystal growth and dissolution. *Nature materials* **14,** 780–784 (2015).

76. Chen, C.-C., Zhu, C., White, E. R., Chiu, C.-Y., Scott, M. C., *et al.* Three-dimensional imaging of dislocations in a nanoparticle at atomic resolution. *Nature* **496,** 74–77 (2013).

77. Barnard, J., Sharp, J, Tong, J. & Midgley, P. High-resolution three-dimensional imaging of dislocations. *Science* **313,** 319–319 (2006).

78. Zeng, X., Gipson, B., Zheng, Z. Y., Renault, L. & Stahlberg, H. Automatic lattice determination for two-dimensional crystal images. *Journal of structural biology* **160,** 353–361 (2007).

79. Rullgård, H, Öfverstedt, L. G., Masich, S, Daneholt, B & Öktem, O. Simulation of transmission electron microscope images of biological specimens. *Journal of Microscopy* **243,** 234–256 (2011).