

MASARYK UNIVERSITY
FACULTY OF INFORMATICS



Implementation of advanced motion correction methods in electron microscopy

MASTER'S THESIS

Matouš Kolařík

Brno, Spring 2018

MASARYK UNIVERSITY
FACULTY OF INFORMATICS



Implementation of advanced motion correction methods in electron microscopy

MASTER'S THESIS

Matouš Kolařík

Brno, Spring 2018

This is where a copy of the official signed thesis assignment and a copy of the Statement of an Author is located in the printed version of the document.

Declaration

Hereby I declare that this paper is my original authorial work, which I have worked out on my own. All sources, references, and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Matouš Kolařík

Advisor: Mgr. Aleš Křenek, Ph.D.

Acknowledgements

I would like to thank here to my advisor Mgr. Aleš Křenek, Ph.D. for his help and patience with this theses and me.

Abstract

Goal of this work is to implement motion correction method, based on a continuous deformation model of the sample, for movie pre-processing in cryo-electron microscopy. As part of this, synthetic data generator is developed to verify correct functionality and assist in further development.

To put the work into proper context, introduction into cryo-electron microscopy and single particle analysis is also provided.

Keywords

cryo-electron microscopy, cryo-EM, electron microscopy, Motioncor2, motion correction, movie alignment, Scipion, single particle analysis, SPA, XMIPP

Contents

Introduction	1
1 Structural biology	3
1.1 <i>X-ray crystallography</i>	3
1.2 <i>SAS</i>	4
1.3 <i>NMR</i>	4
1.4 <i>Mass spectrometry</i>	5
2 Electron Microscopy	6
2.1 <i>Structure</i>	7
2.1.1 <i>Electron gun</i>	7
2.1.2 <i>EM lens systems</i>	8
2.1.3 <i>Detectors</i>	9
2.2 <i>Specimen preparation</i>	9
2.3 <i>Image formation</i>	11
2.4 <i>Radiation damage</i>	12
3 Single particle analysis	14
3.1 <i>Particle detection</i>	14
3.2 <i>Class averages</i>	15
3.3 <i>Model construction</i>	15
3.4 <i>Iterative improvements</i>	15
3.5 <i>Available software</i>	18
4 Movie preprocessing	19
4.1 <i>Motion correction</i>	21
4.1.1 <i>Global correction</i>	21
4.1.2 <i>Noise in shift calculations</i>	22
4.1.3 <i>Local correction</i>	24
4.1.4 <i>Alternative approaches</i>	24
4.2 <i>Dose weighting</i>	25
4.3 <i>Sample deformation model</i>	27
4.3.1 <i>MotionCor2</i>	28
5 Implementation	30
5.1 <i>Model properties</i>	30

5.2	<i>Synthetic data generator</i>	34
5.2.1	<i>Deform_file</i>	35
5.3	<i>Structure of XMIPP implementation</i>	38
5.4	<i>Motion correction results</i>	41
6	Conclusion	43
6.1	<i>Further Work</i>	45
	Bibliography	47

Introduction

From the beginning of their existence in the 1930s electron microscopes have been providing insights beyond the possibilities of classical light microscopy. Since then a great number of improvements both in technique and instrumentation allowed for lower prices, higher resolution, and significantly expanded the number of suitable specimens and so also a number of possible applications. This led to the point where electron microscopes are no longer purely research tools, but have their places also in medicine or technological industry.

As can also be seen from last year's (2017) Nobel's price in chemistry, one of the rightful places of modern electron microscopes is in structural biology. The price was awarded for "developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution" and among others would not be possible without advanced computer processing of measured data.

One of the more significant and recent technological improvements is development of direct detector devices, sensors able to directly detect electron impact with high frame rate, allowing for significant improvements in usable resolution of resulting images. This work is focused on providing open-sourced implementation and acceleration of one of newest approaches for processing these sets of images, and its incorporation into processing software suit Scipion.

In the first chapters, structural biology and it's methods are introduced and general overview of electron microscopy principles and properties is presented. Third chapter describes the whole data processing workflow of single particle analysis (cryo-electron microscopy method, suitable for structural biology).

The motion correction and the processing of the sets, called movies, of high frame rate images from direct detector devices is described in the fourth chapter. Several common approaches are described, with a more detailed description of the deformation model implemented in this work and of MotionCor2, which is the only available, closed source, program implementing this approach.

The fifth chapter takes closer look on the proposed deformation model. And describes both the synthetic deformation generator and the motion correction incorporated into XMIPP that came out of this

analysis and are provided as part of this work. In the end of the chapter the correctness of the implemented motion correction is shown.

1 Structural biology

Structural biology is concerned with structural determination and visualization as well as functional and dynamical properties of organic molecules and macromolecular complexes. This is crucial for furthering our understanding of biological processes, both in the nature in general and in our own bodies allowing us to better treat or prevent various ailments.

In 1952, when Rosalind Franklin successfully crystallized and used x-ray crystallography to photograph DNA structure, it took other scientist another year to actually determine the spacial arrangement of the regular and symmetrical double helix. Although the methods used then are still relevant, the progress in resolution, computer utilization, and development of other methods allows us to routinely construct 3D models of much smaller and less regular particles in span of days or hours.

To provide better context and points of comparison with single particle analysis and cryo-electron microscopy, several other structure determination methods are briefly described in the rest of this chapter, many of which also can be and are used in tandem with SPA to speed up, verify or enhance the obtained data.

More detailed introduction to structural biology can be found for example in [1] or [2].

1.1 X-ray crystallography

When focused x-ray beam is directed through crystallized sample, diffraction patterns can be observed. These correspond to the reciprocal space representation of electron density map of average crystal unit (this is usually not equivalent to the molecule we try to observe, which is called biological unit. More on this distinction in [3].). To map these measurements to corresponding 3D model, amplitudes and phases for diffraction spots must be calculated. As X-rays cannot be easily refocused especially the later is quite problematic and is referred to as the "phase problem". Often the initial phase model is obtained from known similar structure and information obtained this way are again used to further improve the model until convergence is reached.

But the main problem is the production of suitable crystals, which have to be large enough, highly ordered and of high enough purity (both in the chemical and the conformational sense). For many interesting proteins we are unable to produce such crystals and for those we can, only a few possible conformations can often be observed. Nevertheless, when successful, very detailed models (resolution around 1.5 Å can be routinely obtained) can be produced with relatively little further effort. More details can be found in [4] or [5].

1.2 SAS

When crystallization proves to be a too difficult problem, it is possible to use the purified molecular solution in a fashion similar to the one crystals are used in crystallography. Small-angle scattering (called SAXS when used with X-rays and SANS when used with neutron beam) measures 1D scattering curve of radiation passing through such a purified solution. The limited information provided from 1D scattering curve does not lead to a single 3D model, so to limit the interpretation space, SAS techniques are not usually used only by themselves. Thanks to the easy sample preparation and relatively low demands on equipment, it can also be readily used to distinguish between several competing models generated by different methods. More about SAS in general can be found in [6], specifically about SAXS in [7] or about SANS in [8].

1.3 NMR

When an atom is placed into a magnetic field of high enough strength, its nucleus reorients itself to have a spin in parallel or antiparallel direction relative to the magnetic field. If such an atom is then irradiated by radio waves of an appropriate frequency, the atom absorbs part of the energy and changes its orientation. Position of this radio waves frequency, relative to some standard, is called chemical shift and uniquely identifies each element. In the context of a molecule the chemical shift is also further influenced by the electron cloud in the vicinity of the nucleus, so the same element bounded in different chemical context produces different absorption spectrum.

All of this leads to a complex analysis of resulting data and due to the significant spectral overlap restricts its utility to relatively smaller molecules. On the other hand the produced data are in general very reliable and with atomic resolution (the resolution decreases with increase in size of the observed molecule) and unlike other techniques on this list can be used to observe even very fast molecular processes.

Further details about NMR can be found in [9].

1.4 Mass spectrometry

In mass spectrometry the sample is vaporized or liquefied, ionized and then accelerated into bent tube surrounded by an electromagnetic field. From the speed and bent of its trajectory, it is then possible to establish its mass to charge ratio which in turn can be matched to known molecular fragments. Some qualitative and quantitative properties of the specimen can be established this way. As the molecule usually breaks into smaller molecular fragments it is possible to get useful information even for substances with very little previous knowledge.

Exhaustive overview of mass spectroscopy in general can be found in [10] and more about its place in structural biology in [11].

2 Electron Microscopy

The classical light microscopy is limited in its magnifying ability, by the wavelength of visible light, to a resolution of approximately 200 nm. One way to achieve significantly higher useful magnification, is to use different types of radiation with higher frequencies. This chapter and this work in general are focused on the use of electron radiation.

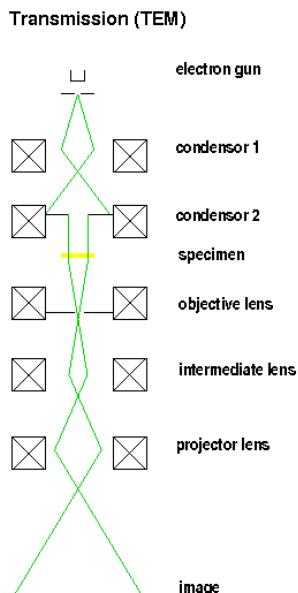


Figure 2.1: Basic diagram of a transmission electron microscope. The basic similarity to a classical light microscope can be easily seen here. Taken from [12]

Although there are several different types of electron microscopes, the one most often used with biological samples is the transmission electron microscope (TEM). For the purposes of this document, electron microscopy, unless specified otherwise, means transmission electron microscopy.

2.1 Structure

Although electrons and photons usually behave quite differently, the general structure of both TEM and light microscopes are surprisingly similar. Both have one radiation source, condenser lens to focus the beam through the sample and further lens system to correctly project the resulting image (Figure 2.1) into appropriate sensor. Rest of this section is used to provide an overview of basic structural elements of typical electron microscope.

2.1.1 Electron gun

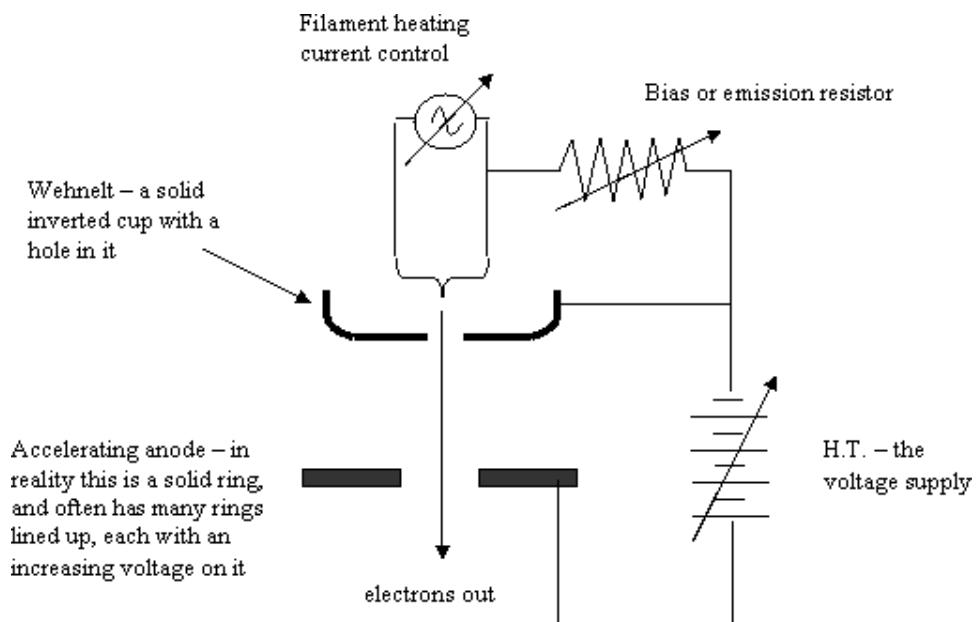


Figure 2.2: Diagram of an electron gun. Taken from [13].

An electron gun provides and accelerates electron beam through the microscope. The source of electrons (also called filament) is a heated cathode with sharp end under high current (several types of filaments are in use, for more details see [14]) covered by a so called Wehnelt cylinder which is negatively charged relative to the filament and has a small opening, to provide the electron beam with initial focus and basic directional consistency around the microscope's optical axis.

2. ELECTRON MICROSCOPY

Follows an accelerator stack, which is a system of multiple electromagnets along the optical axis, designed to accelerate and steer the electron beam in the correct direction. The resulting speed of electrons influences their interaction with the sample and contrast of the resulting images (more on image formation in section 2.3) as well as type and amount of damage caused to the sample (more on radiation damage in section 2.4).

Main characteristics of the electron gun are expressed as its spatial and time coherence. Spatial coherence describes the directional stability of provided electron beam, with the high value meaning that most of the electrons move in the same direction and the low value meaning high directional disorder. Temporal coherence similarly deals with the consistency of the electron energies i.e. speeds.

2.1.2 EM lens systems

Because of high electron energies (around 300 keV) in traditional TEM, permanent magnets cannot be used and electromagnetic lenses are utilized instead. This also allows for a better fine tuning of microscope's properties.

When the electron beam arrives at the lens system it is usually not perfectly aligned with the optical axis. It can be caused for example by the slight tilt of a preceding lens and can be corrected by adjustable magnetic field of a deflectors, which directly precedes the EM lens. For the correction of small irregularities in magnetic fields of the EM lenses, additional EM generators, called stigmators, are placed around them. By producing their own EM field, stigmators can either amplify or reduce problematic sectors of the field produced by its respective EM lens. Such irregularities can be caused, for example, by imperfections in the metal from which the lens is constructed.

After the lens, a metal sheet with a small aperture on the microscope's optical axis is positioned. This way, slightly different pressure can be maintained in different parts of the microscope, and electrons with too much diverging paths are filtered out from the beam.

2.1.3 Detectors

Recording of resulting image was historically done on photographic paper. Although quite high-quality results could be achieved, the time spent processing the pictures in darkroom and the inability to automatically process such pictures without the following scanning significantly limits the practicality of this approach.

Another possibility is using a sheet covered with a fluorescent substance, which in turn is being observed by a traditional camera. This is quite a simple solution, allowing for digital images and automatic processing, but it is limited by resolution of the fluorescent material, and it does not allow for rapid multiple image taking.

A more current solution is using classical CCD chips with a scintillator, translating electron signal into photons. Scintillator absorbs incoming electrons and later re-emits the energy in the form of visible light, which can be detected by a CCD chip. In this approach problems arise from the stochastic nature of the photon emissions in scintillators. One electron can produce several photons and thus blur the acquired image. Similarly, two electrons with equal energies can emit a vastly different amount of photons and produce incorrectly weighted image.

The most modern and advanced detectors are direct electron detector devices (DDD). These sensors measure arriving electrons directly and therefore provide perfectly normalized images, while also having a much higher frame rate (up to hundreds of sub-frames per second) allowing to compose images from multitudes of individual sub-frames, and accounting this way for sample for microscope movement during the exposition time. It can provide much sharper and more detailed images.

Quality of different types of detectors is usually described by detective quantum efficiency (DQE) value. DQE is a function of contrast that is dependent on spatial frequency. High values of DQE mean that lower dose of radiation is needed to acquire an image of identical quality.

2.2 Specimen preparation

Because electrons do not pass easily through most of the matter, observed samples must be very thin and specially prepared. Various

2. ELECTRON MICROSCOPY

ways, each suitable for different type of sample, have been devised since the beginning of electron microscopy. Therefore, only general ideas and the solutions most important for this work will be mentioned here (more detailed information can be found in [15] or [16]). In general, a sample must be thin enough to allow imaging, provide the best possible contrast between the actual object of interest and the carrying medium, and be prepared for vacuum and electron radiation exposure inside the microscope. Furthermore, the preparation process should keep the sample as close to the natural state as possible.

It is problematic especially for biological samples to satisfy all of these requirements. The natural state is usually in liquid water medium, which is due to the evaporation absolutely unsuitable for vacuum environment. Biological samples also tend to quickly deteriorate in electron radiation, while having very low contrast relative to the water background. Currently the most often used solution is embedding the sample in vitreous ice (usually referred to as Cryo-electron microscopy or Cryo-EM). This allows for vacuum exposure while also providing some protection against radiation damage, by restraining free molecule movement. After this, sample is sliced by a microtome and placed on a support grid (more details about Cryo-EM sample preparation can be found in [17] or [16]). The disadvantage of this solution is the necessity to keep the sample at a very low temperature at all time, to prevent formation of crystals. To achieve this, appropriately modified microscopes need to be used.

To improve contrast, negative staining technique may be used together with Cryo-EM. Negative staining uses high contrast substance added to the carrying medium. Ideally the additive should not chemically react with the sample and only helps to differentiate it from its background.

Other possible techniques use dehydration, chemical fixation, plastic embedding, or metal shadowing for sample stabilization. Although all of these methods provide sufficient protection against vacuum and sometimes improve contrast or resistance to radiation damage. They also significantly alter the structure of the sample and/or require sample specific preparation and are therefore quite laborious.

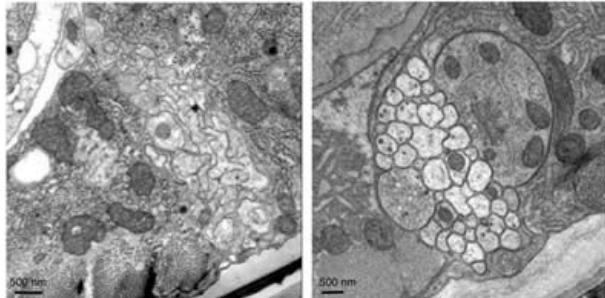


Figure 2.3: Both images show a nerve cord. Sample on the right was stabilized by chemical fixation, sample on the left was frozen under high pressure. The damage caused by chemical fixation can be easily seen. Take from [18].

2.3 Image formation

Image produced by an electron microscope is not really precise representation of reality. This is due to many factors including imperfect coherence of electron gun, astigmatism and spherical aberration of lenses, and defocus used while imaging. The relation between an ideal image (2D projection of a sample) and an image produced by microscope is usually described by the contrast transfer function (CTF), describing relation between a contrast of the acquired image and its spatial frequency. Ideal CTF would be constant across all frequencies. In reality, it is a periodic decreasing function with the highest amplitudes in the lower frequencies. This dampening in higher frequencies is caused mainly by the imperfect spatial and time coherence of electron beam. The function describing it is referred to as an envelope function. Details can be found in [19], [20] or [16].

From a different perspective, CTF is just a reciprocal space version of point spread function, which is describing distortion and blurring of one pixel of acquired image. To correct for these measuring imperfections, CTF correction is used. $R = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(I)}{CTF}\right)$. Where R is the ideal image and I is the acquired image. Such a correction nicely improves recorded values everywhere except for the frequencies where CTF is equal to 0. This information is irretrievably lost. To correct for such missing values a defocus pair can be taken. It consists of two images of the same area, taken with different focus values. The difference in

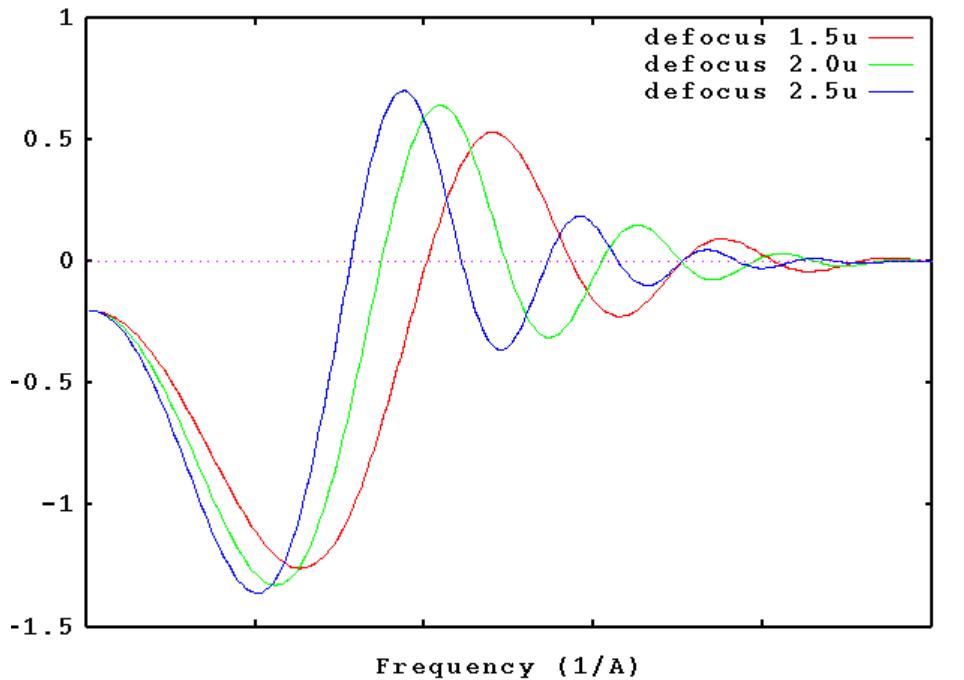


Figure 2.4: Example of CTF for different defocus values. The dampening of higher frequencies by an envelope function can be seen here. Also, the severity of the envelope function is higher the further from focus the measurement is made. Take from [21].

defocus changes the shape of the CTF and shifts the intersections with x axis to a different positions. Details can be found [16] or [22].

2.4 Radiation damage

Even in highly stable inorganic crystallic samples electron illumination causes significant structural damage over time. This damage is even more significant in biological samples and is directly related to the total amount of radiation the sample is exposed to. The exposure of only $6 e/\text{\AA}^2$ can already completely destroy diffraction patterns and made further sample imaging useless [23].

There are two primary sources of radiation damage. The knock-out damage which is caused by direct displacement of whole atoms by high energy electrons and radiolysis which is bond breakage caused

2. ELECTRON MICROSCOPY

by molecule ionization. While knock-out damage can be prevented by using electrons with energies lower than those of displacement energies of bonds present in the sample, and is virtually non-existent for lower electron energies usable in modern EM, the probability of radiosys raises for slower moving electrons. At the end this means, that some amount of damage is unavoidable.

Radiation further produces secondary electrons and phonons that cause further ionization and sample heating.

Bond breakage in organic samples also releases gas molecules (generally oxygen and hydrogen) which can form bubbles in solid medium further disrupting the sample.

More on radiation damage can be found in [24] or [25].

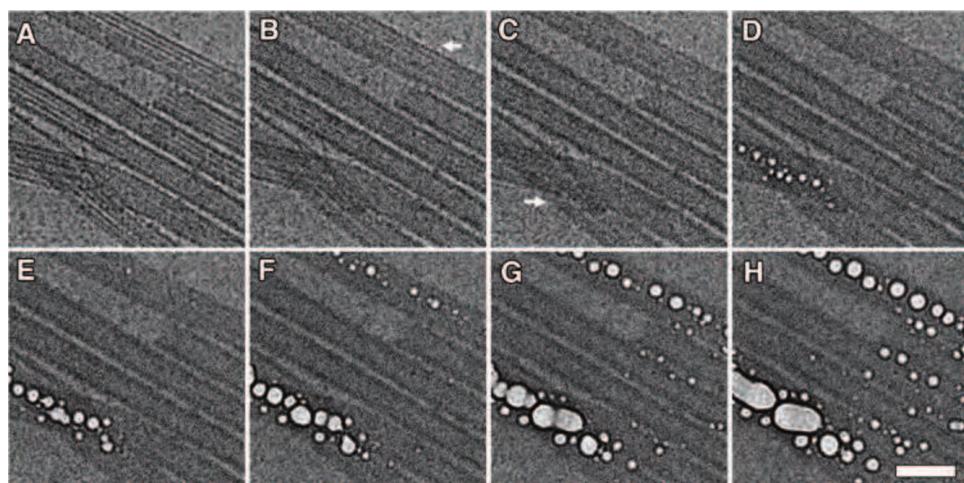


Figure 2.5: Images show microtubules with damage caused by electron radiation over the course time, with each step from A to H adding approximately $4e/\text{\AA}^2$ of exposure. Taken from [24].

3 Single particle analysis

Images obtained by TEM often suffer from very low contrast and signal-to-noise ratio (SNR). These problems get worse as attempts are made to acquire still higher resolution images. In crystallography, this problem is solved by producing the diffraction pattern as an average of large number of units in the same configuration arranged into regular crystal lattice. The single particle analysis (SPA) is trying to solve exactly this problem for electron microscopy.

As with periodical structures, SPA improves resolution and contrast of acquired images by averaging observations of multiple samples. Unlike the crystallography, these instances (referred to as *particles*) do not have to be ordered into crystals or fulfill other strict configurational criteria. The main idea is taking many images (*micrographs*) of the sample, find all visible particles, divide them into *classes* based on their configuration and orientation, and from these classes obtain *class averages*. In this way several high-resolution 2D *projections* of the particle are acquired and can be further used to construct a 3D model. Good introductory materials can be found in [14], [26], [27], or EMAN documentation [28]. Detailed can be found in [16].

3.1 Particle detection

For the SPA to work properly, first individual particles must be correctly selected. This essentially means marking the smallest rectangular area on micrograph containing the particle.

The simplest approach is manual detection, where the picking is left entirely to the user. Although this approach is very time-consuming, it still has its place, especially for particles with unknown structure. Lately, fully automatic approaches are also being used, whether based on cross-correlation with known model projections or executed by neural nets (e.g. [29]). The results are still unfortunately not good enough to allow for no human intervention. This leads to semi-automated solutions, where particles are picked automatically, with high false-positive rate, and the user is left to decide which ones to use for further processing.

3.2 Class averages

Detected particles need then to be divided into homogeneous classes, where only particles with equal orientation towards the sample plane, equal molecular state (e.g. different states of fuzzy complexes), and only one particle kind (accounting for sample contamination) per class are present. After the classification, particles inside each class needs to be rotationally and translationally aligned with each other and their class average is calculated. By averaging particles in different configurations, valuable information (especially in higher frequencies) can be lost, so correct classification is crucial to acquire correct and sharp 2D projections. Procedures based on K-means algorithm are usually used for such a class construction [26]. Alternatively maximum likelihood method [30] can be used. In this case, instead of dividing particles into disjoint classes, to each particle is assigned a vector of probabilities of belonging into every class. The class average is then calculated by summing all particles, weighted by their probability of belonging into the corresponding class.

3.3 Model construction

All class averages represent one 2D projection of the examined molecular complex. To reconstruct a 3D model the projection slice theorem [31] is used. It states that, every two such projections have, in reciprocal space, one common line called *sinogram*. Exact position and orientation of each projection in 3D reciprocal space can be determined by finding this common line, with at least two other projections. After filling the reciprocal space with all available projections, combining them along these lines, (the higher number of different 2D projections the higher resulting resolution) inverse 3D Fourier transform is performed, to acquire real space 3D representation.

3.4 Iterative improvements

The whole process is run repeatedly, while improving the resulting model by fine-tuning particle picking and classification. To better classify the particles, new 2D projections (*re-projections*) are taken from

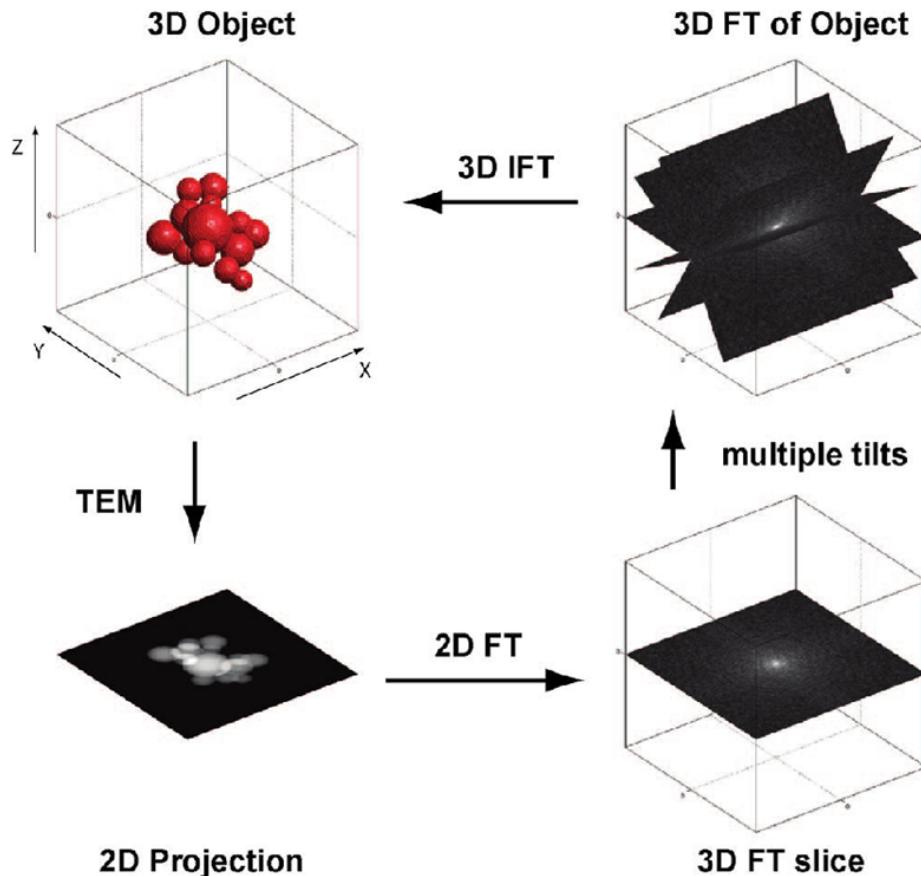


Figure 3.1: Reconstruction of 3D model from 2D projections using projection theorem. Taken from [31].

the generated model. These projections are then used as bases for a new classification, and the whole process is performed again with newly acquired averages. The cycle is usually repeated until no significant changes are produced between consecutive iterations (Figure 3.2).

The initial classification step is generally quite problematic, and erroneous classification can significantly increase the required amount of iterations and worsen the resulting model. Because of this, some kind of initial model is often supplied. This can, for example, be a result of some different structural biology techniques or just educated guess expressed as "balls and sticks" model.

3. SINGLE PARTICLE ANALYSIS

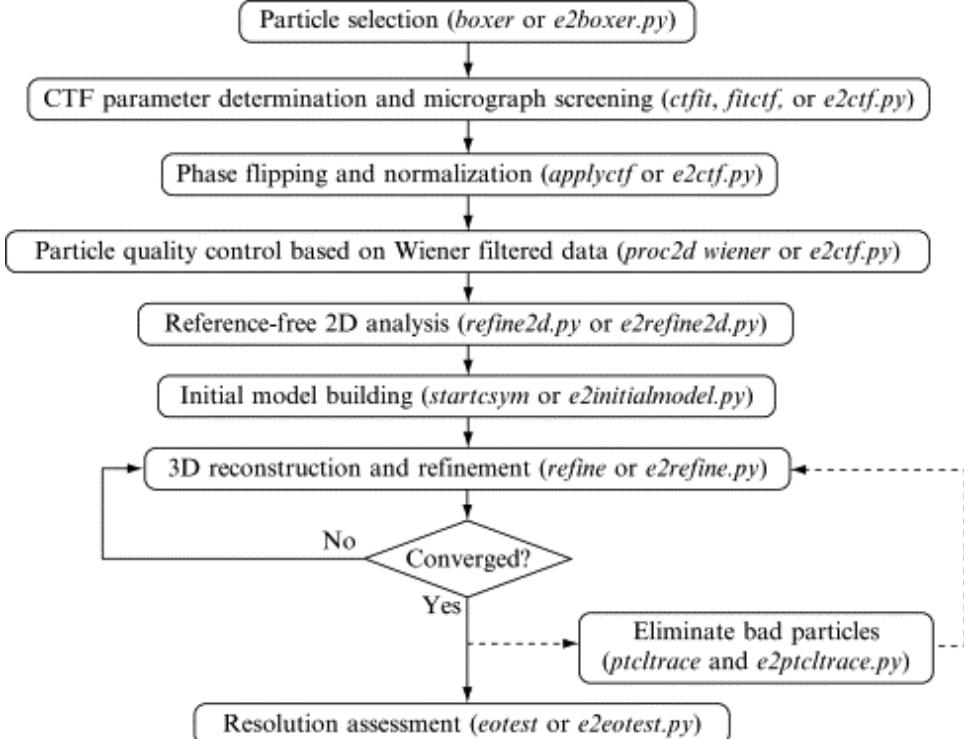


Figure 3.2: Diagram of single particle analysis workflow. The script names, next to each step, correspond the programs used by EMAN. Taken from [32].

Because of the self reinforcing nature of SPA, the model bias can be a significant problem. This happens, for example, when picking particles using cross-correlation with known or expected model of some observed molecule. When the expected model significantly differs from reality, micrographs were obtained on improperly configured microscope so just noise was recorded, or improperly prepared sample does not contain expected substances, the highest cross-correlation peaks can easily be just matches with random noise. This is known as the "Einstein from noise"[33] problem. The significance of this problem is its subtlety. "Particles" picked in such manner, can produce plausible 3D model similar to the expected one. And as it mostly occurs with low energy, low SNR data, the legitimacy of the picked particles, is hard to verify by sight, which means that it is almost impossible to

reliably and completely avoid. Therefore, at least for de novo SPA, it is preferable to either verify obtained results by means of completely different method from the arsenal of structural biology or at least splitting the dataset into two and performing two independent analysis. More on this can be found in [34].

3.5 Available software

From the complexity and amount of processed data it is clear that, even with high level of automation, model construction with SPA is a very time-consuming and demanding process requiring long training to be performed properly. So even though there are good tools for each individual step, it is preferable to integrate them into some kind of software suit, which would simplify the whole process, guide the user through it, and remember data and parameters used, and thus enable reproducibility.

The most important ones today, providing support for SPA, are probably EMAN2[35], Spider[36], Relion[37] and Scipion[38]. These are all quite different products and each performs the above mentioned particular to a varying degree and in a different style. More complete list of software libraries and frameworks suitable for electron microscopy can be found in [39].

Scipion¹ is relatively new software suit, compared for example with Spider, which has been available since 1978. Although Scipion partially evolved from and closely integrates XMIPP, software image processing library for electron microscopy, it does not primarily try to provide set of tools to perform SPA. Instead, it provides open and extendable environment for integration and cooperation between already existing tools. This means providing automatic conversion between competing styles, integrating all tools in common user interface, and allowing for easy integration of not yet incorporated tools.

1. <https://github.com/I2PC/scipion/wiki>

4 Movie preprocessing

As it was already mentioned in section 2.1.3, recent advancements in detector technology allowed for recording of high frame-rate movies instead of just one image integrated over the whole exposition time. And as it was already widely demonstrated [41, 42, 43, 40] and can also be seen in Fig. 4.1, the image quality can be significantly improved by the movie stack preprocessing.

Unfortunately, working with single frames means even lower contrast and SNR. Therefore, requiring development of some specialized approaches for successful processing

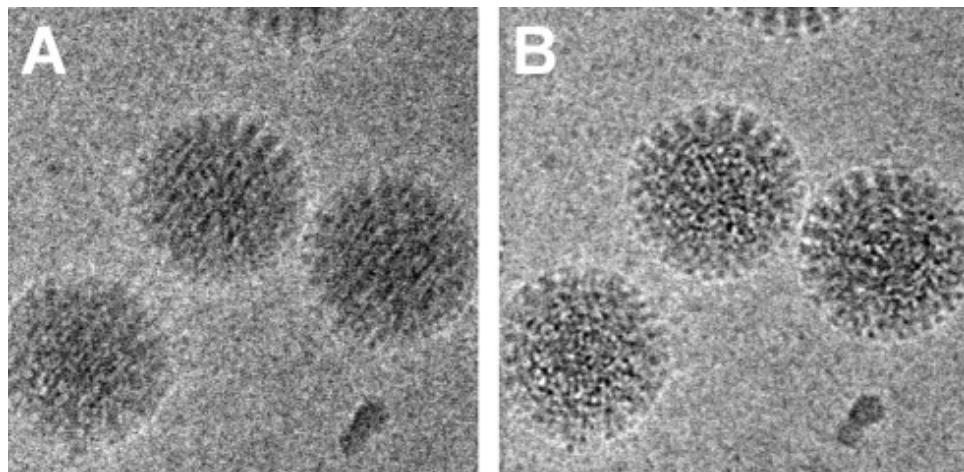


Figure 4.1: (A) Rotavirus particle before movie preprocessing. The image is blurred and high frequency details are mostly lost. (B) The same image after motion correction and frame averaging. Taken from [41]

Beside the noise arising from the background ice, there are several other possible sources of image blurring and quality degeneration:

Flexible sample movement Although, the sample is frozen in vitreous ice and kept at very low temperature, the imaged molecules do not necessarily behave like a compact solid objects. Parts of the molecule can still move relative to each other under the

electron beam. This highly depends on the strength of bonds present in each particular molecule and does not usually present significant issue. Rarely, this can become a problem when some small part has highly mobile bond with the rest of molecule significantly decreasing the resolution of such a part. With available tools, there is not much that can be done with distortions.

Fixed pattern noise Noise which is not purely random and remains static between successive images is referred to as fixed pattern noise. It is most often caused by a malfunctioning pixel in a detector (hot, cold or even dead pixels), although more complicated artifacts can also appear as it was, for example, reported in [44]. Such defects can be very problematic, when calculating shifts between frames. The presents of these patterns, especially on otherwise similar frames, can produce a false impression that the frames are not shifted to each other.

Because these defects tend to be localized to a single pixels, low-pass, filter prior to the cross-correlation, usually sufficiently reduces their impact. For example, MotionCorr uses filter of the form $e^{\frac{-B}{4H^2}}$ where H is the resolution and B is the temperature factor [41]. Further, [41] also suggests not using shift estimates from any two successive frames.

Radiation damage In section 2.4, the principles and progression of damage caused to a sample by electron radiation were described. Since this damage increases with the total radiation exposure, the earliest frames in each movie stack have the best preserved image features, and with each consecutive frame decrease in quality is experienced. When the frames are averaged to produce the final micrograph, the details present in the earlier frames are lost. This problem is further described in section 4.2.

Sample movement The movement can be caused by anything from microscopic movements of the whole microscope stage to a charge build up and sample deformation caused by electron beam. It is the the most limiting factor of this list, and treating its effects on the recorded movie is referred to as *motion correction*, and is more closely discussed in the next section.

4.1 Motion correction

Motion of particles between consecutive frames can be observed both for very small molecules and much larger molecular complexes or viral particles alike. What differs is how much it affects our ability to reconstruct accurate 3D model, and how difficult it is to correct for. The larger particles can lose only little of their higher frequency details while the smaller ones can became entirely unobservable in the averaged image.

As was already mentioned in this chapter, there are numerous possible and simultaneously contributing sources of movement. This is further complicated by our incomplete understanding of some of the processes that are causing them. But in general, the distinction is drawn between the two types of movement. There is the global movement which is homogeneous across all particles in the sample, or at least the processed frame. And there is the local movement which is slightly different for different parts of the frame or even for neighboring particles. Although similar methods can be often used and some approaches interlace both of them, the distinction is usually drawn between the *global movement correction* and the *local movement correction*.

4.1.1 Global correction

To correctly position the individual frames relative to each other before averaging, shifts between each two consecutive frames needs to be calculated (or shifts to one common reference frame which is in the end an identical problem). It is most commonly done with a help of cross-correlation.

Cross-correlation calculates similarity of two frames for different shift values, and it is widely used in signal processing and analysis. In order to perform the calculation efficiently, the cross-correlation theorem is commonly used, i.e. cross-correlation is calculated in reciprocal space as: $R = \mathcal{F}^{-1}(\mathcal{F}(A) * \mathcal{F}(B))$ where A and B are the frames of interest [45]. The shift that is producing the most similar images is then in the resulting matrix indicated by a pixel with the highest value *cross-correlation peak*.

Cross-correlation is used in MotionCorr[41], Motioncor2[44], Unblur[46], and alignframes_lmbfgs[47] and is the dominant way of dealing with shifts estimations. Alternative approaches are described in section 4.1.4.

4.1.2 Noise in shift calculations

The low SNR means that the real cross-correlation peak can be distorted or even overshadowed by a match to a random noise, and can thus easily produce imprecise or incorrect results. To provide more reliable calculations, certain degree of redundancy and result control is, therefore, desirable. The way how it is done differs little in each software solution, but, to provide some general idea, approaches employed by MotionCorr and Unblur are presented here.

In MotionCorr, the apparent shifts are calculated for each pair of frames. The redundancy comes from an overdetermined system of linear equations composed of combinations of these calculated shifts. If $\vec{s}_{a,b}$ is apparent shift between some frames a and b then for any shift it must also obviously hold that $\vec{s}_{a,a+n} = \vec{s}_{a,a+l} + \vec{s}_{a+l,a+n}$ for any $l \leq n$ e.g. $\vec{s}_{a,a+3} = \vec{s}_{a,a+2} + \vec{s}_{a+2,a+3} = \vec{s}_{a,a+1} + \vec{s}_{a+1,a+2} + \vec{s}_{a+2,a+3} = \dots$. Such system is then solved and iteratively improved by discarding equations that differs too much from the obtained solution [41, 48].

An alternative approach is used in Unblur where individual frames are not fitted with each other, but the shifts are rather calculated between each frame and an average of all other frames (the frame itself is excluded from the average to avoid high correlation peak with itself). After it is done for each frame, the calculated shifts are used to construct improved average which, in turn, is again used to recalculate the individual shifts. The whole loop is repeated while some specified improvement is achieved in each cycle or maximum number of cycles is performed [46, 48].

Moreover, most of the tools also enforce reasonable shape of the calculated movement trail. By which we usually mean that the direction and speed of movement over the entire trail doesn't change too suddenly. It is again implemented differently in different tools. For example, Unblur approximates the calculated shifts with polynomial function or spline, while alignframes_lmbfgs, which defines

searching for shifts as optimization problem, adds penalization for sharp changes in direction to its objective function.

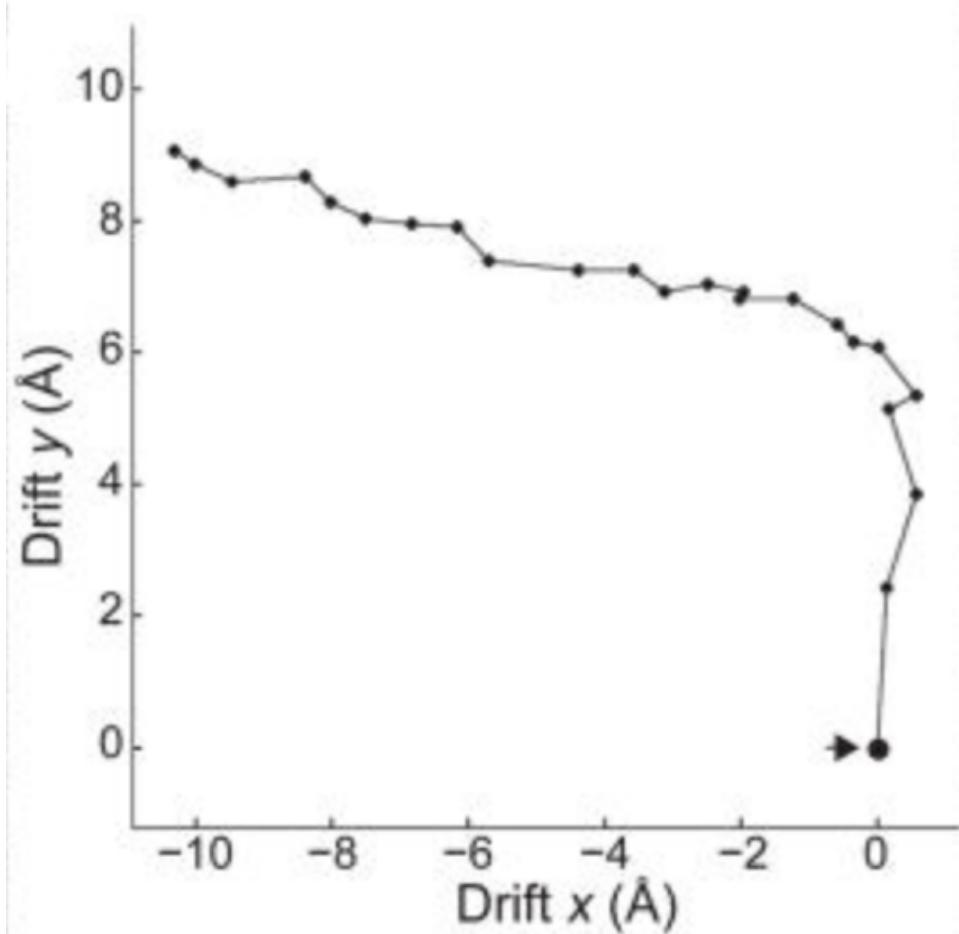


Figure 4.2: Relative movements between individual frames. The arrow indicates the first frame. Although, here it seems that the movement is roughly divided into movement along x axis and y-axis after that it is not generally the case. Taken from [41].

In addition, it was observed that the sample motion, especially its local component, is strongly time dependent. With the rate of movement highest in the few earliest frames with the intensity quickly dropping in later ones (as can be seen in Fig. 4.2). Because of this, first

few (usually around three or five) frames are often discarded, even though, these are the frames with the lowest radiation damage [41].

4.1.3 Local correction

The methods used for a global correction can be, of course, similarly used for smaller parts of frames to calculate local motion. First problem with this approach is again the noise levels. It is not possible to pick subregion of an arbitrarily small size and simply apply the global motion algorithm. The smaller the region is the larger is the chance the shift calculation will produce incorrect results base on pure noise match. For example, authors of MotionCorr explicitly mention that the method is not suitable for data smaller the 2000x2000 pixels, under typical conditions[41, 44].

Furthermore, it is not entirely clear what to do with shifts calculated for separate regions without further information about particles positions. Simple shift application would only break the resulting image into multiple fragments and destroy a lot of data along the splits. The approach used by alignparts_lmbfgs or Relion (see section 4.1.4) is to perform the local correction only later in the SPA workflow, after particles were picked. In Relion the motion correction is done on each particle individually. Alternatively, [41] suggest dividing the whole frame into overlapping regions, which are then motion corrected separately. Data for individual particles are then recovered from the frame with center closest to position of the particle.

Significantly different approach is used in MotionCor2. This solution does not require knowledge of the exact locations of the picked particles. Instead, it constructs motion field for the whole frame and applies the correction on subpixel levels. This method is more thoroughly described in section 4.3.

4.1.4 Alternative approaches

The view presented so far in this chapter is, only a simplified overview of the current state. More in-depth understanding can be gained from the cited papers or, alternatively, there is also an overview of most of the currently used motion correction tools in [48].

There are at least three noteworthy approaches to the presented problems which significantly differ from the presented view: MotionCor2 (which will be more closely described in section 4.3.1), Optical flow, and Relion.

Optical flow

Another common way to calculate relative motion between images in classical image processing is optical flow. Unlike cross-correlation, optical flow calculates movement from relative changes in pixel brightness between frames. Its use in cryo-EM motion correction was showed in [49] and is reported by the authors to work the best for a local movement calculation. And it is, therefore, a good fit for work in tandem with another tool, doing the global alignment, as is for example MotionCorr (this combination is, for example, implemented in XMIPP).

Relion

Relion is more closely integrated into the whole SPA workflow. The motion correction is not performed solely on the movie data, but it is part of the SPA loop.

In the initial step frames are assumed to be in alignment, and an 3D model is calculated from their averages. Projections of the obtained 3D model are then used to estimate shifts for individual particles, improving its estimations in each cycle [50, 48].

4.2 Dose weighting

Although, the process of combining the shifted frames was referred to as averaging in this work, it does not entirely reflect the real situation. Increase in radiation dose improves SNR in low frequency structures while also blurring the high frequency ones (this relation can be seen in Fig. 4.3). For this reason, to provide the best possible information in all frequencies, all frames cannot have the same weight across the whole frequency spectrum. The usage of different weights to frequency information of individual frames is generally known as *dose weighting* and is integrated in most of the software tools in the use.

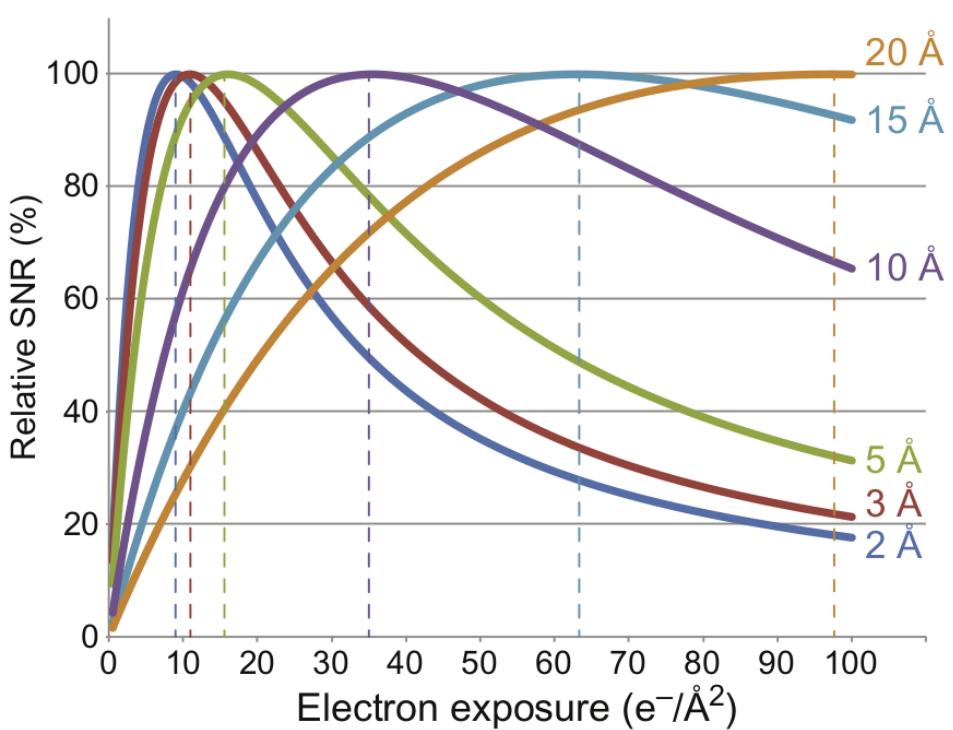


Figure 4.3: The curves represent the weighting ratios used in Unblur and alignparts_lmbfgs. The steep decline in higher resolution signal is clearly visible after $10e^-/\text{\AA}^2$, while low frequency SNR keeps improving. Taken from [48].

The weights are assigned based on the cumulative exposure of each frame, and are either based on explicit equations obtained experimentally [51, 52] (as is the case, for example, in Unblur[46]), or the frequency-dependent information decrease can be directly estimated during the SPA loop (as is done in Relion[50]). The issue of dose weighting and useful resolution calculation is in itself quite complex and lies beyond the scope of this work.

One thing to note, is that it is often desirable to produce both dose weighted and unweighted micrographs, as the CTF estimation cannot be properly performed with dose weighted average.

4.3 Sample deformation model

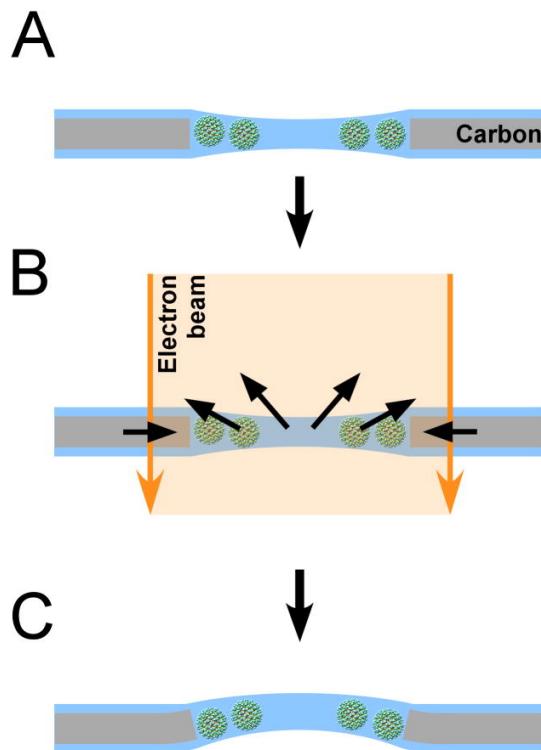


Figure 4.4: In B and C, the doming deformation of the ice caused by radiation is clearly visible. Moreover, in C, the deformation of the supporting grid can be seen. Taken from [53].

It was observed in [54] that the vitreous ice over the circular holes in carbon support grid experiences doming-like deformation under the electron beam. It was further confirmed and elaborated by [53]. Furthermore, it was suggested that significant portion of local movement can be explained by this movement of the ice. It also seems that the most of this doming motion can be attributed to changes in internal pressure of the ice caused by radiolysis, and to edge deformation of carbon support grid (see Fig. 4.4).

Unlike the approaches discussed in the beginning of this chapter, modeling the movement as continuous deformation allows us to calculate fairly exact motion correction for even sub-pixel positions. And as

no information about particles locations is required, the local correction can be performed separately to the actual SPA loop, even during the image acquisition, saving time and computational resources while providing very precise correction across the whole frames with high resistance to low SNR.

This approach was so far utilized only by MotionCor2 (see next section) providing, according to the authors, very good results even for lower SNR.

4.3.1 MotionCor2

MotionCor2[44] is a recent (2016) addition to the motion correction software developed by a team consisting, for great part, of people who developed the original motion correction algorithm MotionCorr. The software is built on the before described doming-like theory of ice deformation, confirming the observed doming as significant source of local movement. [44] also reported the ability to accurately motion correct even movies with shorter exposition times and lower defocus values.

The frames are first corrected for global motion by an algorithm similar to the solution used in Unblur (see section 4.1.2). The individual frames are aligned to a higher SNR average of the whole movie without the current frame. This is performed repeatedly while improving the obtained average until specified value of residual error is or maximum number of iterations is reached.

This global corrected frame stack is then divided into 25 sub-frame stacks (each frame is divided into 5x5 partitions) and the local shifts are calculated for each stack separately using the same approach as for the global motion. Calculated shifts are then assumed to accurately represent the residual movement of the respective sub-frame centers. Then the information is used to estimate coefficients of the doming deformation model using least square fitting, individually for shifts along x and y axis resulting in two functions $S_x(x, y, t)$ and $S_y(x, y, t)$, where the function $S(x, y, t)$ used to generally approximate the doming effect is:

$$S(x, y, t) = (c_0 + c_1x + c_2x^2 + c_3y + c_4y^2 + c_5xy) \times (c_6t + c_7t^2 + c_8t^3) \quad (4.1)$$

Resulting functions S_x and S_y are then used to calculate shifts for each pixel on a super-resolution frames. The exact values on integer pixel coordinates are approximated by bilinear interpolations.

MotionCor2 is also capable of producing both dose weighted and unweighted results, and contains bad pixel correction. This should not only correct for single hot and cold pixels but also for more complicated artifacts caused by detector firmware. Even though, closer details on how exactly both of these functions are performed are not provided.

Although the MotionCor2 is available for free academic and non-profit use, the authors decided not to publish their source codes. Making it impossible for other researchers to further experiment and built on this solution. This is only more unfortunate because of the extensive work that was done on the GPU acceleration of MotionCor2. For this reason, the basic functionality of MotionCor2 is implemented into XMIPP framework as part of this work, to allow for further research and improvements in this line of motion correction algorithms.

5 Implementation

As a part of this work, two software pieces were developed. A tool for a reproduction of the effects of the doming described in the section 4.3 and a XMIPP program performing cryo-em movie motion correction using the same approach. Both of them are included as attachments to this thesis. The doming simulator is called doming_generator and is implemented as a Python module. The XMIPP implementation is build as a part of the Scipion project and is attached with the whole Scipion framework source code.

This chapter serves to describe the interface, implementation, and performance of both of these tools. As well as to more closely examine the effects of the doming and possibilities of motion correction based on this approach.

5.1 Model properties

The deformation model, as is expressed by equation 4.1, can be separated into two parts. The first part (5.2) describes the spatial-dependent shifts localizing the frame onto a quadratic surface. The second part (5.2) describes the time-dependent dynamics of shifts changes by an cubic equation. Both of these parts combined are enough to sufficiently well describe the doming of ice, even with global shifts. However, it is also able to describe deformation shapes and dynamics not conforming to the pattern described previously. Although, this doesn't matter much in the purely algorithmic sense, as the coefficients are determined by a least square fitting to actual measurements, and should, therefore, if correctly estimated, produce physically feasible model. Establishing proper ranges and restrictions on values of the coefficients of Eq. (4.1) is beneficial for better understanding of the model, and more important for this work, to generate feasible synthetic data.

$$f_t(t) = c_6t + c_7t^2 + c_8t^3 \quad (5.1)$$

$$f_s(x, y) = c_0 + c_1x + c_2x^2 + c_3y + c_4y^2 + c_5xy \quad (5.2)$$

There are several restrictions that can be extracted from section 4.3, on the time-dependent dynamics that must hold over the measuring:

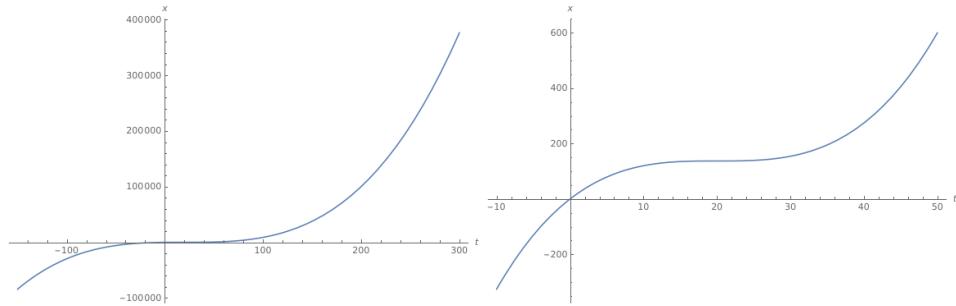


Figure 5.1: Both images are the same function $f_t(t)$, randomly generated with the restrictions described in this chapter. All functions generated this way have this kind of shape. Generated using WolframAlpha.

1. There is no deformation before the beginning of irradiation.
2. The absolute value is non-decreasing across the whole interval (the shift can be negative or positive). Although it can be easily assumed that the value is always positive and the up or down shape is left to be described by the spatial part of equation.
3. The value changes fast for first few frames with the change slowing in the further ones. After high enough amount of irradiation the doming effects mostly stops.

This in general means that $f_t(0) = 0$, which is trivially fulfilled by the absence of a absolute element in Eq. (5.1), and on the measurement interval $< t_1, t_n >$ the function is monotonous non-decreasing and concave (t_1 is the time point of frame, t_n of the last) and it, further, holds that $0 < t_1 < t_n$. As the intervals between t_k a t_{k+1} for any $k \in < 0, n - 1 >$ are generally equivalent, the interval of quick value increase is relatively short when compared to the interval of slow or no increase, it leads to function of general shape showed by Figure 5.1 (no local maxima, one inflection point). Such shape means that there is only one solution to $f'_t(t) = 3c_8t^2 + 2c_7t + c_6$. For model generation purposes it is also useful to specify the maximal value on the interval to some value x_n , and determine the position of the inflection point to

t_n . From all this we can establish equation system:

$$\begin{aligned} f_t(t_n) &= c_6 t_n + c_7 t_n^2 + c_8 t_n^3 = x_n \\ (2c_7)^2 - 4(3c_8 c_6) &= 0 \\ 3c_8 t + 2c_7 &= 0 \\ x > 0 \\ t > 0 \end{aligned}$$

By solving this system and eliminating all, for other reason, unreasonable solutions, for example $c_8 = 0 \wedge c_7 = 0$, for specific x_n and t_n we can define the coefficient to:

$$\begin{aligned} c_6 &= \frac{3c_8 t_n^2}{4} = \frac{3x_n}{t_n} \\ c_7 &= \frac{-3c_8 t}{2} = \frac{-6x_n}{t_n^2} \\ c_8 &= \frac{4x_n}{t_n^3} \\ t > 0 \end{aligned}$$

It is less straightforward to establish similar requirements for the spatial-dependent part of the model Eq. (5.2). In general, the function is defined across any region in \mathbb{R}^2 , and it is complex enough to describe both the local and global movement. From what we now the model should have these properties:

1. The doming is not ideal, but the overall shape should be close to doming.
2. The greatest shifts are on the edges of the dome (i.e. around the edges of the circular holes in support grid).
3. There is a global minimum in the middle of the dome

From 1), it can be reasoned that the coefficients next to the quadratic elements (c_2, c_4) should have equal sign. 2) and 3) then, further determines this sign to be positive giving it the general shape of an convex paraboloid (Fig. 5.2). The values of c_2 and c_4 should also be kept relatively close. As with quadratic functions in general, the

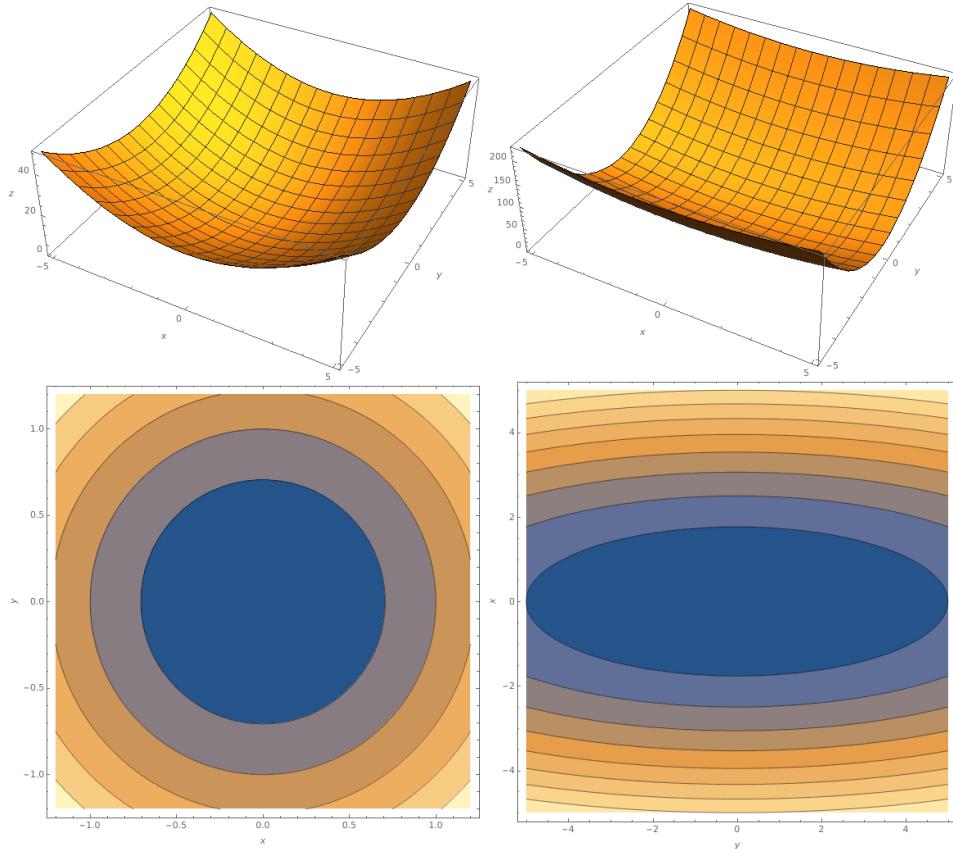


Figure 5.2: The two upper graphs are plots of the possible shapes of the (5.2) model. With the lower two being contour plot of the same functions with blue being the lower value. The left column features regular dome obtained from $z = x^2 + y^2$, while the right illustrates the effect of greater difference between c_2 and c_4 coefficient ($z = x^2 + 8y^2$). Generated using WolframAlpha.

5. IMPLEMENTATION

mixed element c_5xy is result of rotation around the z axis. Rotation of $z = c_1x + c_2x^2 + c_3y + c_4y^2$ by an angle α around z axis produces coefficients:

$$\begin{aligned}c_1 &= c_1\cos(\alpha) + c_3\sin(\alpha) \\c_2 &= c_2(\cos(\alpha))^2 + c_4(\sin(\alpha))^2 \\c_3 &= c_3\cos(\alpha) - c_1\sin(\alpha) \\c_4 &= c_2(\sin(\alpha))^2 + c_4(\cos(\alpha))^2 \\c_5 &= c_4\sin(2\alpha) - c_2\sin(2\alpha)\end{aligned}$$

The paraboloid can be further arbitrarily shifted. Which, beside the possibility of moving the deformation, means that, as long as the choice is consistent during the whole process, it does not matter where in the image is located the point of origin. To perform translation of the point of origin by $-s_x$ along x axis and $-s_y$ along y axis the coefficients are changed as follow:

$$\begin{aligned}c_0 &= c_0 + c_1s_x + c_2s_x^2 + c_3s_y + c_4s_y^2 + c_5s_xs_y \\c_1 &= c_1 + 2c_2s_x + c_5s_y \\c_2 &= c_2 \\c_3 &= c_3 + 2c_4s_y + c_5s_x \\c_4 &= c_4 \\c_5 &= c_5\end{aligned}$$

It is also worth noting, that the described overall properties remain also valid in the combined model (4.1). The only difference are specific values at each given point which are uniformly scaled by the time-dependent part at each time point.

5.2 Synthetic data generator

The doming_generator Python module was developed as both a quick check of the feasibility of the approach described in [44], as a convenient tool to generate synthetic data for XMIPP implementation

testing, and for closer examination of the effects the doming has on image.

The module is implemented to run under Python3, although, it should be compatible with earlier versions as well. The calculations are performed with help NumPy¹ and SciPy² libraries, and mrcfile³ module is used for data saving. Two general functions are made available as part of the module: `deform_file` and `motion_correct_files`.

5.2.1 Deform_file

```
deform_file(path=None, shape=None,
            time_points=None, coefficients=None, save=None,
            add_grid=True, movie=False, verbose=True)
```

`deform_file` is designed for synthetic data generation. Based on the provided parameters, it either applies specified deformation on selected gray-scale image, or tries to generate reasonable random deformation on a dummy image from SciPy library. The result is a list of images with deformation corresponding to selected exposure times, with the option to generate cryo-em movie file.

path is a string path leading to a gray-scale image to be deformed. In the case, this parameter is set to None, the image is selected from `scipy.misc.face(True)`.

shape defines the size of the output image as tuple of height and width values in pixels. For None value, the default smaller size is used.

time_points is list of exposure times for which deformed images should be generated.

coefficients describe the actual deformation. It should be two dimensional array of shape (2, 9). The first 9 coefficients defining $c_{0..9}$ from equation 4.1 for calculating shifts along y axis and the second 9 coefficients for shifts along x axis.

1. <http://www.numpy.org/>
2. <http://www.scipy.org/>
3. <http://pypi.org/project/mrcfile/>

save In case of None value, the results are only returned as an array.

Other values are expected to be string paths to some folder where the resulting images should be saved.

add_grid Allows to draw a black grid over the image to better visualize the occurring deformations.

movie If save argument also set, it saves the results also as a cryo-em movie.

verbose If this option is set to True, then informations about the calculation progress are printed out.

The deformation is done simply by the same deformation model application as is motion correction. The difference is only in the used model coefficients and required time point. In case the coefficients are generated by the application, the restriction described in previous section are kept in mind.

The time coefficients c_6 , c_7 and c_8 are chosen so that scaling of at most 2 is generated on the used time interval ($f_t(t_n) = 2$).

The space-dependent coefficient are generated in steps. First the c_2 (coefficient of x^2) coefficient is randomly generated to produce maximal shift of 3% of the image width over the length of the image and then c_4 (coefficient of y^2) is also randomly generated so that the combined shift is at most 5% of the images longer dimension. Both the 3% and 5% limits were defined mostly arbitrarily after experimenting with different values. The result is randomly rotated along the z axis and the minimum (center of the dome) is randomly moved. The center is kept within an area defined by the image with added 10% additional space to all the edges.

For saving the resulting movie (in case movie option is selected), mrc file format is used together with XMIPP defined STAR format. The mrc format [55] is a standard way to save image and volume data in both cryo-em and tomography. Unfortunately, the standard does not define a way to store time stamp data together with the images. It would be possible to store them in some non-standard way (in extra header space, for example), but it was decided to go in a more complicated but also more transparent direction and use STAR file format defined in XMIPP [56]. The data are saved in a data block movie_stack with tags images and time.

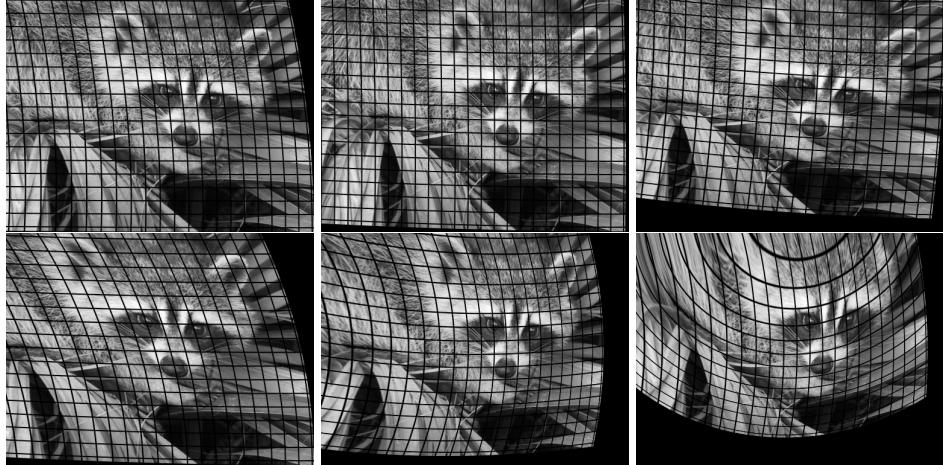


Figure 5.3: Several example of deformations caused to the `scipy.misc.face(True)` produced by the synthetic data generator. The upper row are images which could be normally produced (although, to provide better idea, these are also intentionally on the more intense side). The lower row was generated without checks on the coefficients value (although, to produce some recognizable image, these were kept in the lower spectra). The black grid was added over the image for better visualization.

```
motion_correct_files(paths=[], time_points=[],
                     coefficients=None, save_path=None,
                     save_partial=False, verbose=True):
```

This is a proof of concept function executing the entire motion correction process on the provided deformed data. And it is not intended to be used to actually perform any real world motion correction, but to provide easily changeable prototyping implementation. Some steps, important for low SNR images, are not implemented (for example, the low pass filter for fixed patter noise described in the beginning of chapter 4 and, for performance reasons, the maximum number of iterations for shift calculations is set very low). Furthermore, the performance is starting to significantly slow down already on even small images.

paths Paths to all files containing the deformed input images.

time_points Exposition times at which the images were taken (the order should correspond to the paths argument).

coefficients If the coefficients are supplied (see description of `deform_file` function) they are applied to the images to perform the reconstruction. If the argument is set to None the whole motion correction process, as it was described in section 4.3.1, is executed, and the coefficients are estimated by the program.

save_path This argument behaves the same way as `save` in `deform_file` function.

save_partials In case the argument is set to True not only the final reconstructed image is saved, but also every image after its final correction is saved to the folder defined by the `save_path` argument.

The example usage of both of these functions is provided inside `main.py` file.

5.3 Structure of XMIPP implementation

XMIPP is an image processing framework designed mainly for use in the context of SPA. It is organized into a number of utilities, called here programs. These tend to have a very specific purpose (correct CTF, find common lines between two projections, etc.⁴) and more complex tasks are intended to be performed by chaining them onto shared data structure files. To maintain a reasonable common interface, `XmippProgram` virtual class is provided. When implemented, it takes care of proper argument passing, implements `-help` option with arguments descriptions, or is able to handle exceptions. This interface can also be used to auto-generate simple GUI for the programs.

To specify a new program, it must be specified in the list of programs in the `scipion_scons` file. The actual code is then placed in the `${XMIPP_HOME}/applications/programs/program_name` folder. There usually lies only one very simple `cpp` file containing either minimal main function or only call to the `RUN_XMIPP_PROGRAM` macro

4. <http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/ListOfProgramsv3>

with an argument containing a name of a class implementing the `XmippProgram`.

The calculation performing code is situated in the libraries which in turn are further divided into thematic modules. For this work, the most important ones are the `data`, holding shared data structures and some shared operations on them, and the `reconstruction`, which, among others, contains the provided motion correction program.

In the reconstruction module, there already is a motion correction program present. Unlike our implementation, it is based on the old MotionCorr approach and is mostly able to perform only global correction. Nevertheless, to provide consistency, the interface of our implementation, was based on the provided one.

The implemented program is named `xmipp_movie_alignment_mc2` and have the following options:

- i** Mandatory argument providing path to an meta data file with, at least, labels `image` and `time` (the format was described in previous section).
- o** Mandatory argument providing path the the output motion corrected micrograph file.
- maxIterations** This argument specifies the maximal number of iterations in the iterative frame alignment. The default value is 10, which is the same value Unblur uses.
- ounaligned** Optional argument containing path to an output micrograph file constructed from uncorrected frames.
- j** Defines number of threads available for program run. By default only 2 threads are used.
- fnDarkImage** Path to an image containing data for a dark-frame subtraction, to eliminate hot pixels in fixed pattern noise. If no path is provided, the correction is not performed.
- fnGainImage** Path to an image containing data for a gain image correction, to normalize the pixels sensitivity and so remove cold pixels in fixed pattern noise.

5. IMPLEMENTATION

The actual calculation is in the class *ProgMovieAlignmentDeformationModel*, inheriting from *XmippProgram*, and implemented in files *movie_alignment_deformation_model.**.

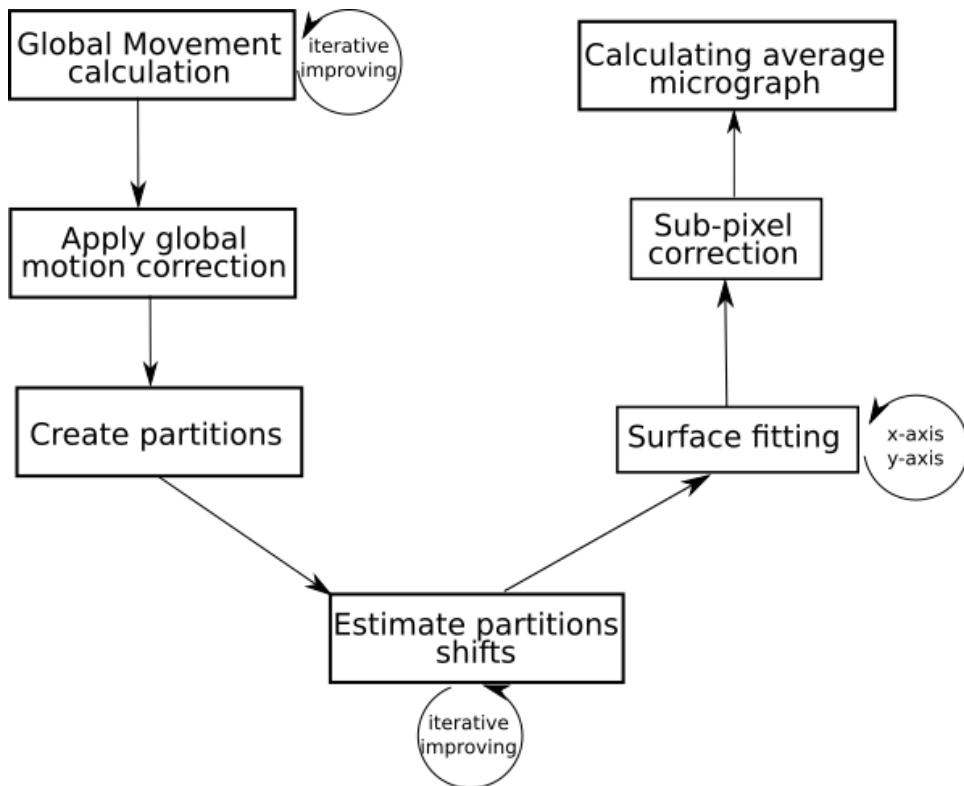


Figure 5.4: The overall schema of the motion correction calculation.

The individual loaded frames are stored into *MultidimArray<double>* and are the dark and gain corrections are immediately applied to them. From this, the unaligned micrograph is calculated, in case such option was selected.

The relative global motion is iteratively calculated, as was described in section 4.3.1. The individual frames are not actually shifted until the final shifts are determined, to minimize the accumulated data degradation.

The necessary cross-correlation is calculated by XMIPP using the cross-correlation theorem and internally utilizing the FFTW⁵ library.

5. <http://www.fftw.org/>

The motion corrected frames then need to be divided, each into 25 equally large partitions (each frame is divided into five rows and five columns). They are, afterward, treated as individual stacks each consisting of partitions lying in the same position in a different frame. For each stack, the shifts are estimated using the same method as in the case of global motion. From the obtained data, four-dimensional points are obtained. Each point of the form (x, y, t, s) where x and y are coordinates of the center of the partition (relative to the original frame), t is time point of the original frame, and s is the obtained shift.

The four-dimensional points are used to estimate the coefficients of the (4.1). For this purpose, *ALGLIB*⁶ numerical library is utilized to calculate least square fitting with s as the functional value at point (x, y, t) . The whole calculation is done twice, separately for shifts along x-axis and y-axis, producing two sets of model coefficients.

Each pixel of every frame is then separately shifted along x-axis and y-axis to the position it was supposed to be in time 0, based on the obtained models. This is executed in a 2times upsampled image, to provide smoother results. The resulting micrograph is then the average of these calculated frames. The overall diagram can be seen in Fig. 5.4.

5.4 Motion correction results

Figure 5.5 and 5.6 shows the difference between simple averaging and the results provided by the implemented motion correction. It can be easily seen, that the correction improves the image details especially around the edges.

The deformation is clearly correctly reversed. But it is also easily visible that the approach has its limitation. There are instances, as is the case in Fig. 5.8 where the distortion is corrected for but not sufficiently enough. There are also instances where the model overcompensates and distorts the area into the other direction, as is the case in Fig. 5.7. This could be further improved by increasing the number of partitions per frame. Although, the increased number of data points would produce more accurate model, the effect would be most likely more

6. <http://www.alglib.net/>

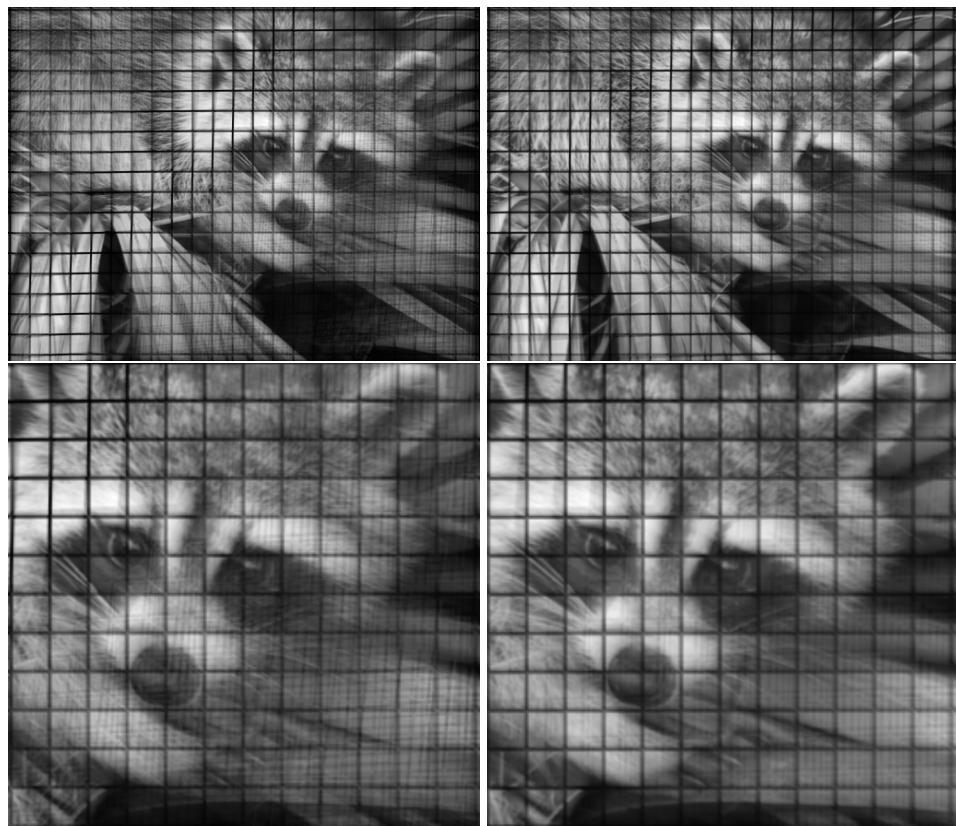


Figure 5.5: The same data set was used to produce simple global corrected average (on the left side) and deformation corrected image (on the right side). Overall and detailed view.

then offset by the decrease in the accuracy of motion estimation in low SNR images.

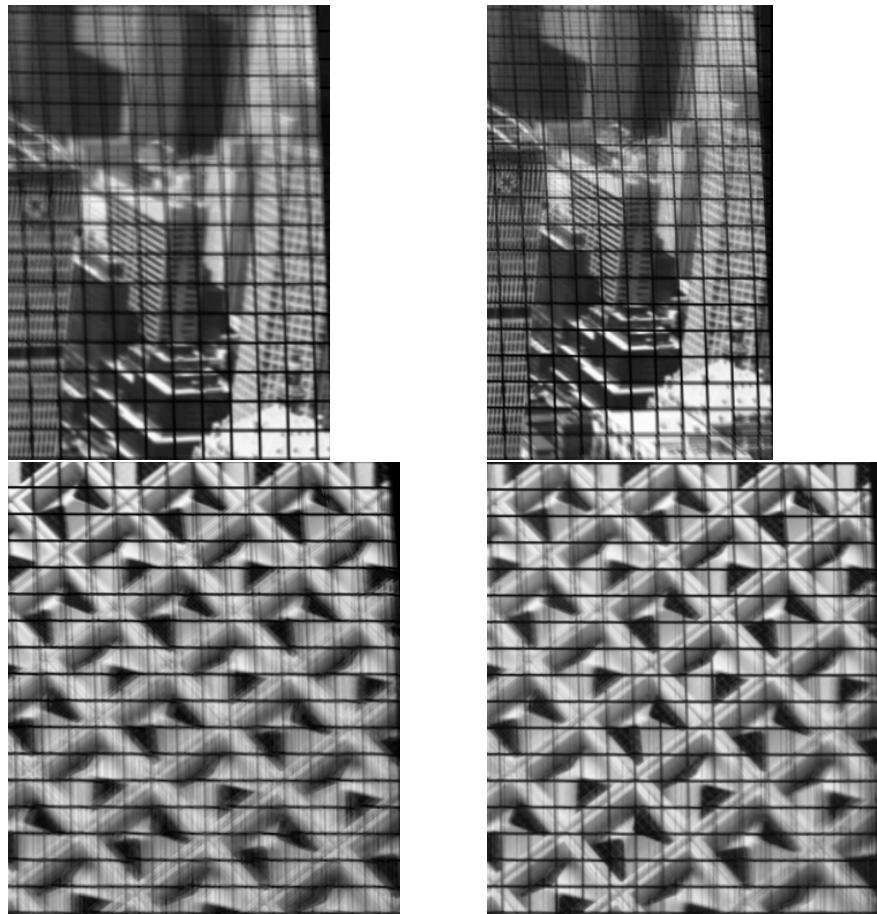


Figure 5.6: Zoomed in view of detail improvement of deformation correction. (Left column was only global corrected, right column was deformation corrected). The cityscape is take from BBC ^c and the pattern was created by Bryan Colosky^d.

-
- a. <http://www.bbc.com/future/tags/city>
 - b. <https://unsplash.com/photos/3ChXTytXDUl>
 - c. <http://www.bbc.com/future/tags/city>
 - d. <https://unsplash.com/photos/3ChXTytXDUl>

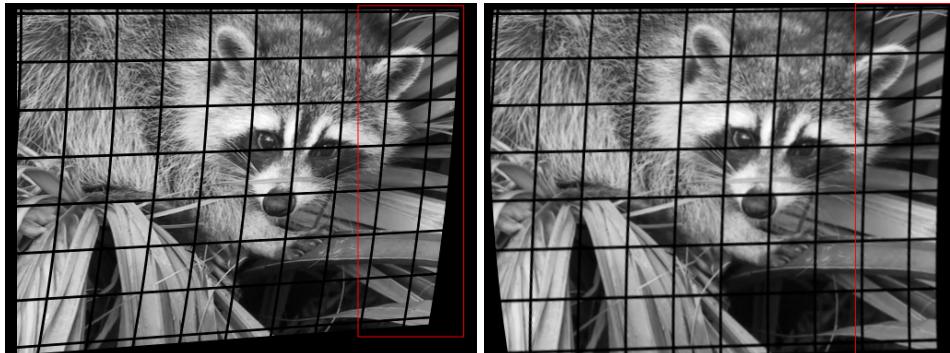


Figure 5.7: The frame on the left was only globally aligned while the frame on the right was motion corrected by the deformation model. This instance shows the overcompensation (highlighted by the red rectangle). The vertical lines are convex before the correction and concave instead of straight afterwards.



Figure 5.8: The frame on the left was only globally aligned while the frame on the right was motion corrected by the deformation model. In this case the model does not compensate enough. Although, the situation is improved, the horizontal lines (highlighted by the red rectangle) are still significantly bent on their right end upwards even after correction.

6 Conclusion

This work provided introductory material with references for further reading for both electron microscopy and single particle analysis as

well as closer examination of current motion correction approaches. Extra attention was dedicated to description of the doming model and its usage in MotionCor2 software.

The synthetic deformation generator and motion correction program for XMIPP was developed. Using both of these tools, it was demonstrated, on a qualitative level, what kind of deformations the proposed model can account for and how good are the particular results.

Included implementations provide both a basic confirmation of the applicability of the algorithm described in [44], and a possible stepping stone in a future open source development of this approach.

6.1 Further Work

For future work, there still remain several issues to be addressed before the implementation should be used in any real world setting. One already mentioned point is the missing dose weighting. This could be implemented by incorporating, for example, the dose weighting scheme described in [46]. As it does not require closer integration into the SPA loop, and thus preserves this advantage of the sub-pixel motion correction.

Further, there was not enough time and space in this work to perform more detailed quantitative estimations of the effects on acquired resolution. To address this, it could be interesting to compare the resolution improvements, relative to simple frame averaging, with MotionCor2 and ideally also other mentioned tools. Although, the MotionCor2 is accepted as very well working tool (for example, [57] stating that the results acquired were "for all practical purposes, virtually identical to that from particle polishing"), there seems to be no work systematically comparing quality of the obtainable results on wider range of movies and microscope settings, as was, for example, done on smaller scale in [40].

The reason for this thesis was to provide a first step for a further improvement and development of the deformation modeling approach. It may be, for instance, interesting to see the effects of changing the number of partitions, used to calculate local movement. Or to calculate the local movement by different approach, for example, using

6. CONCLUSION

optical flow. But this would need to be performed together with the before mentioned quantitative measurements of the motion correction impact on the overall acquired resolution.

Bibliography

1. MINOR, Daniel L. The Neurobiologist's Guide to Structural Biology: A Primer on Why Macromolecular Structure Matters and How to Evaluate Structural Data. *Neuron* [online]. 2007, vol. 54, no. 4, pp. 511–533 [visited on 2018-03-08]. ISSN 08966273. Available from DOI: 10.1016/j.neuron.2007.04.026.
2. NATIONAL SCIENCE FOUNDATION (DBI-1338415); US DEPARTMENT OF ENERGY; THE NATIONAL INSTITUTES OF HEALTH. *Introduction to PDB Data*. Available also from: <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction>.
3. SHUCHISMITA DUTTA; CATHERINE L. LAWSON; RACHEL KRAMER GREEN. *Introduction to Biological Assemblies and the PDB Archive*. Available also from: <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/biological-assemblies>.
4. BRAGG, Lawrence. X-RAY CRYSTALLOGRAPHY. *Scientific American*. 1968, vol. 219, no. 1, pp. 58–74. Available also from: <http://www.jstor.org/stable/24927456>.
5. SMYTH, M. S.; MARTIN, J. H. x ray crystallography. *Molecular pathology*: MP. 2000, vol. 53, no. 1, pp. 8–14. ISSN 1366-8714.
6. SVERGUN, Dmitri I; KOCH, Michel H J. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* [online]. 2003, vol. 66, no. 10, pp. 1735–1782 [visited on 2018-03-10]. ISSN 0034-4885, 1361-6633. ISSN 0034-4885, 1361-6633. Available from DOI: 10.1088/0034-4885/66/10/R05.
7. LIPFERT, Jan; DONIACH, Sebastian. Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes. *Annual Review of Biophysics and Biomolecular Structure* [online]. 2007, vol. 36, no. 1, pp. 307–327 [visited on 2018-03-10]. ISSN 1056-8700, 1545-4266. ISSN 1056-8700, 1545-4266. Available from DOI: 10.1146/annurev.biophys.36.040306.132655.
8. JACROT, B. The study of biological structures by neutron scattering from solution. *Reports on Progress in Physics* [online]. 1976, vol. 39, no. 10, pp. 911–953 [visited on 2018-03-10]. ISSN 0034-4885, 1361-6633.

BIBLIOGRAPHY

- ISSN 0034-4885, 1361-6633. Available from DOI: 10.1088/0034-4885/39/10/001.
9. FERENTZ, Ann E.; WAGNER, Gerhard. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Quarterly Reviews of Biophysics* [online]. 2000, vol. 33, no. 1, pp. 29–65 [visited on 2018-03-10]. ISSN 00335835. Available from DOI: 10.1017/S0033583500003589.
 10. HOFFMANN, Edmond de; STROOBANT, Vincent. *Mass spectrometry: principles and applications*. 3rd ed. Chichester, West Sussex, England ; Hoboken, NJ: J. Wiley, 2007. ISBN 978-0-470-03310-4.
 11. WYSOCKI, Vicki H.; RESING, Katheryn A.; ZHANG, Qingfen; CHENG, Guilong. Mass spectrometry of peptides and proteins. *Methods* [online]. 2005, vol. 35, no. 3, pp. 211–222 [visited on 2018-03-10]. ISSN 10462023. Available from DOI: 10.1016/jymeth.2004.08.013.
 12. LEADLEY, David. *Transmission Electron Microscopy (TEM)*. 2010. Available also from: <https://warwick.ac.uk/fac/sci/physics/current/postgraduate/regis/mpagswarwick/ex5/techniques/structural/tem/>.
 13. J. M. RODENBURG. *The gun*. 2004. Available also from: <http://www.rodenburg.org/guide/t400.html>.
 14. MILNE, Jacqueline L. S. et al. Cryo-electron microscopy - a primer for the non-microscopist. *FEBS Journal* [online]. 2013, vol. 280, no. 1, pp. 28–45 [visited on 2017-12-02]. ISSN 1742464X. Available from DOI: 10.1111/febs.12078.
 15. WATT, Ian M. *The principles and practice of electron microscopy*. 2nd ed. Cambridge ; New York: Cambridge University Press, 1997. ISBN 978-0-521-43456-0 978-0-521-43591-8.
 16. FRANK, J. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. 2nd ed. Oxford ; New York: Oxford University Press, 2006. ISBN 978-0-19-515096-4 978-0-19-518218-7. OCLC: ocm56368337.
 17. DUBOCHEZ, Jacques; ADRIAN, Marc; CHANG, Jiin-Ju; HOMO, Jean-Claude; LEPAULT, Jean; McDOWALL, Alasdair W.; SCHULTZ, Patrick. Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics* [online]. 1988, vol. 21, no. 02, pp. 129 [visited on

BIBLIOGRAPHY

- 2017-12-03]. ISSN 0033-5835, 1469-8994. ISSN 0033-5835, 1469-8994. Available from DOI: 10.1017/S0033583500004297.
18. THE UNIVERSITY OF UTAH. *Advanced Microscopy*. The University of Utah, 2011. Available also from: <http://advanced-microscopy.utah.edu>.
 19. WADE, R.H. A brief look at imaging and contrast transfer. *Ultra-microscopy* [online]. 1992, vol. 46, no. 1-4, pp. 145–156 [visited on 2017-12-04]. ISSN 03043991. Available from DOI: 10.1016/0304-3991(92)90011-8.
 20. KIRKLAND, Earl J. *Advanced Computing in Electron Microscopy* [online]. Boston, MA: Springer US, 1998 [visited on 2017-12-04]. ISBN 978-1-4757-4406-4. Available from: <http://public.eblib.com/choice/publicfullrecord.aspx?p=3086530>. OCLC: 851818896.
 21. LEITH, ArDean. *SPIDER & WEB*. Available also from: <https://spider.wadsworth.org>.
 22. ZANETTI, Giulia; RICHES, James D.; FULLER, Stephen D.; BRIGGS, John A.G. Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *Journal of Structural Biology* [online]. 2009, vol. 168, no. 2, pp. 305–312 [visited on 2017-12-04]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2009.08.002.
 23. GLAESER, Robert M. *LIMITATION TO SIGNIFICANT INFORMATION IN BIOLOGICAL ELECTRON MICROSCOPY AS A RESULT OF RADIATION DAMAGE*. Lawrence Berkeley National Laboratory, 1970. Available also from: <https://escholarship.org/uc/item/8d33n38m>.
 24. B., David; E., James. Low-Dose Imaging Techniques for Transmission Electron Microscopy. In: MAAZ, Khan (ed.). *The Transmission Electron Microscope* [online]. InTech, 2012 [visited on 2018-03-05]. ISBN 978-953-51-0450-6. Available from DOI: 10.5772/36614.
 25. EGERTON, R.F.; LI, P.; MALAC, M. Radiation damage in the TEM and SEM. *Micron* [online]. 2004, vol. 35, no. 6, pp. 399–409 [visited on 2018-03-05]. ISSN 09684328. Available from DOI: 10.1016/j.micron.2004.02.003.

BIBLIOGRAPHY

26. CHENG, Yifan; GRIGORIEFF, Nikolaus; PENCZEK, PawelA.; WALZ, Thomas. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* [online]. 2015, vol. 161, no. 3, pp. 438–449 [visited on 2017-12-04]. ISSN 00928674. Available from DOI: 10.1016/j.cell.2015.03.050.
27. CHENG, Yifan. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* [online]. 2015, vol. 161, no. 3, pp. 450–457 [visited on 2017-12-04]. ISSN 00928674. Available from DOI: 10.1016/j.cell.2015.03.049.
28. *EMAN2.2 Reconstruction Tutorial*. Baylor College of Medicine, National Center for Macromolecular Imaging, 2017. Available also from: <http://blake.bcm.edu/emanwiki/EMAN2/Tutorials?action=AttachFile&do=get&target=EMAN2-2Tutorial.pdf>.
29. WANG, Feng; GONG, Huichao; LIU, Gaochao; LI, Meijing; YAN, Chuangye; XIA, Tian; LI, Xueming; ZENG, Jianyang. DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. *Journal of Structural Biology* [online]. 2016, vol. 195, no. 3, pp. 325–336 [visited on 2017-12-04]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2016.07.006.
30. SIGWORTH, Fred J.; DOERSCHUK, Peter C.; CARAZO, Jose-Maria; SCHERES, Sjors H.W. An Introduction to Maximum-Likelihood Methods in Cryo-EM. In: *Methods in Enzymology* [online]. Elsevier, 2010, vol. 482, pp. 263–294 [visited on 2017-12-04]. ISBN 978-0-12-384991-5. Available from DOI: 10.1016/S0076-6879(10)82011-7.
31. THE RESEARCH UNIVERSITY IN THE HELMHOLTZ ASSOCIATION. *Electron Tomography Developments*. 2017. Available also from: <https://www.int.kit.edu/1731.php>.
32. CONG, Yao; LUDTKE, Steven J. Single Particle Analysis at High Resolution. In: *Methods in Enzymology* [online]. Elsevier, 2010, vol. 482, pp. 211–235 [visited on 2018-03-08]. ISBN 978-0-12-384991-5. Available from DOI: 10.1016/S0076-6879(10)82009-9.
33. HENDERSON, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences* [online]. 2013, vol. 110, no. 45, pp. 18037–18041 [visited on 2018-05-18]. ISSN 0027-8424, 1091-6490. ISSN 0027-8424, 1091-6490. Available from DOI: 10.1073/pnas.1314449110.

BIBLIOGRAPHY

34. SHIGEMATSU, H.; SIGWORTH, F.J. Noise models and cryo-EM drift correction with a direct-electron camera. *Ultramicroscopy* [online]. 2013, vol. 131, pp. 61–69 [visited on 2018-05-06]. ISSN 03043991. Available from DOI: 10.1016/j.ultramic.2013.04.001.
35. TANG, Guang; PENG, Liwei; BALDWIN, Philip R.; MANN, Deepinder S.; JIANG, Wen; REES, Ian; LUDTKE, Steven J. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology* [online]. 2007, vol. 157, no. 1, pp. 38–46 [visited on 2017-12-03]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2006.05.009.
36. FRANK, Joachim; RADERMACHER, Michael; PENCZEK, Pawel; ZHU, Jun; LI, Yanhong; LADJADJ, Mahieddine; LEITH, Ardean. SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. *Journal of Structural Biology* [online]. 1996, vol. 116, no. 1, pp. 190–199 [visited on 2017-12-04]. ISSN 10478477. Available from DOI: 10.1006/jsbi.1996.0030.
37. MRC LABORATORY OF MOLECULAR BIOLOGY. *Relion*. Available also from: https://www2.mrc-lmb.cam.ac.uk/relion/index.php?title>Main_Page.
38. INSTRUCT IMAGE PROCESSING CENTER. *Scipion*. Available also from: <https://github.com/I2PC/scipion/wiki>.
39. *Software Tools For Molecular Microscopy*. Available also from: https://en.wikibooks.org/wiki/Software_Tools_For_Molecular_Microscopy.
40. RAWSON, S.; IADANZA, M.G.; RANSON, N.A.; MUENCH, S.P. Methods to account for movement and flexibility in cryo-EM data processing. *Methods* [online]. 2016, vol. 100, pp. 35–41 [visited on 2017-12-04]. ISSN 10462023. Available from DOI: 10.1016/j.ymeth.2016.03.011.
41. LI, Xueming; MOONEY, Paul; ZHENG, Shawn; BOOTH, Christopher R; BRAUNFELD, Michael B; GUBBENS, Sander; AGARD, David A; CHENG, Yifan. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* [online]. 2013, vol. 10, no. 6, pp. 584–590 [visited on 2017-12-04]. ISSN 1548-7091, 1548-7105. ISSN 1548-7091, 1548-7105. Available from DOI: 10.1038/nmeth.2472.

BIBLIOGRAPHY

42. CAMPBELL, MelodyG. et al. Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy. *Structure* [online]. 2012, vol. 20, no. 11, pp. 1823–1828 [visited on 2018-05-05]. ISSN 09692126. Available from DOI: 10.1016/j.str.2012.08.026.
43. BAI, Xiao-chen; FERNANDEZ, Israel S; MCMULLAN, Greg; SCHERES, Sjors HW. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* [online]. 2013, vol. 2 [visited on 2018-05-06]. ISSN 2050-084X. Available from DOI: 10.7554/eLife.00461.
44. ZHENG, Shawn; PALOVCAK, Eugene; ARMACHE, Jean-Paul; CHENG, Yifan; AGARD, David. Anisotropic Correction of Beam-induced Motion for Improved Single-particle Electron Cryo-microscopy. *Methods* [online]. 2016 [visited on 2017-12-04]. Available from DOI: 10.1101/061960.
45. WEISSTEIN, ERIC W. *Cross-Correlation Theorem*. Available also from: <http://mathworld.wolfram.com/Cross-CorrelationTheorem.html>.
46. GRANT, Timothy; GRIGORIEFF, Nikolaus. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* [online]. 2015, vol. 4 [visited on 2017-12-04]. ISSN 2050-084X. Available from DOI: 10.7554/eLife.06980.
47. RUBINSTEIN, John L.; BRUBAKER, Marcus A. Alignment of cryo-EM movies of individual particles by optimization of image translations. *Journal of Structural Biology* [online]. 2015, vol. 192, no. 2, pp. 188–195 [visited on 2017-12-04]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2015.08.007.
48. RIPSTEIN, Z.A.; RUBINSTEIN, J.L. Processing of Cryo-EM Movie Data. In: *Methods in Enzymology* [online]. Elsevier, 2016, vol. 579, pp. 103–124 [visited on 2017-12-04]. ISBN 978-0-12-805382-9. Available from DOI: 10.1016/bs.mie.2016.04.009.
49. ABRISHAMI, Vahid; VARGAS, Javier; LI, Xueming; CHENG, Yifan; MARABINI, Roberto; SORZANO, Carlos Óscar Sánchez; CARAZO, José María. Alignment of direct detection device micrographs using a robust Optical Flow approach. *Journal of Structural Biology* [on-

BIBLIOGRAPHY

- line]. 2015, vol. 189, no. 3, pp. 163–176 [visited on 2018-05-05]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2015.02.001.
50. SCHERES, Sjors HW. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* [online]. 2014, vol. 3 [visited on 2018-05-06]. ISSN 2050-084X. Available from DOI: 10.7554/eLife.03665.
51. BAKER, Lindsay A.; SMITH, Eric A.; BUELER, Stephanie A.; RUBINSTEIN, John L. The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals. *Journal of Structural Biology* [online]. 2010, vol. 169, no. 3, pp. 431–437 [visited on 2018-05-06]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2009.11.014.
52. HAYWARD, Steven B.; GLAESER, Robert M. Radiation damage of purple membrane at low temperature. *Ultramicroscopy* [online]. 1979, vol. 4, no. 2, pp. 201–210 [visited on 2018-05-06]. ISSN 03043991. Available from DOI: 10.1016/S0304-3991(79)90211-0.
53. BRILLOT, Axel F.; CHEN, James Z.; CHENG, Anchi; PAN, Junhua; HARRISON, Stephen C.; POTTER, Clinton S.; CARRAGHER, Bridget; HENDERSON, Richard; GRIGORIEFF, Nikolaus. Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* [online]. 2012, vol. 177, no. 3, pp. 630–637 [visited on 2017-12-04]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2012.02.003.
54. WRIGHT, Elizabeth R.; IANCU, Cristina V.; TIVOL, William F.; JENSEN, Grant J. Observations on the behavior of vitreous ice at 82K and 12K. *Journal of Structural Biology* [online]. 2006, vol. 153, no. 3, pp. 241–252 [visited on 2018-05-06]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2005.12.003.
55. CHENG, Anchi et al. MRC2014: Extensions to the MRC format header for electron cryo-microscopy and tomography. *Journal of Structural Biology* [online]. 2015, vol. 192, no. 2, pp. 146–150 [visited on 2018-05-18]. ISSN 10478477. Available from DOI: 10.1016/j.jsb.2015.04.002.
56. CARAZO, JOSE M. ET AL. *File Formats*. Available also from: <http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/FileFormats>.

BIBLIOGRAPHY

57. DANEV, Radostin; TEGUNOV, Dmitry; BAUMEISTER, Wolfgang. Using the Volta phase plate with defocus for cryo-EM single particle analysis. *eLife* [online]. 2017, vol. 6 [visited on 2018-05-13]. ISSN 2050-084X. Available from DOI: 10.7554/eLife.23006.
58. WILLIAMS, David B.; CARTER, C. Barry. The Transmission Electron Microscope. In: *Transmission Electron Microscopy* [online]. Boston, MA: Springer US, 1996, pp. 3–17 [visited on 2017-12-03]. ISBN 978-0-306-45324-3 978-1-4757-2519-3. Available from DOI: 10.1007/978-1-4757-2519-3_1.
59. BELL, James M.; CHEN, Muyuan; BALDWIN, Philip R.; LUDTKE, Steven J. High resolution single particle refinement in EMAN2.1. *Methods* [online]. 2016, vol. 100, pp. 25–34 [visited on 2017-12-03]. ISSN 10462023. Available from DOI: 10.1016/j.ymeth.2016.02.018.
60. AFANASYEV, Pavel et al. Single-particle cryo-EM using alignment by classification (ABC): the structure of *Lumbricus terrestris* haemoglobin. *IUCrJ* [online]. 2017, vol. 4, no. 5, pp. 678–694 [visited on 2017-12-04]. ISSN 2052-2525. Available from DOI: 10.1107/S2052252517010922.
61. JOYEUX, Laurent; PENCZEK, Paweł A. Efficiency of 2D alignment methods. *Ultramicroscopy* [online]. 2002, vol. 92, no. 2, pp. 33–46 [visited on 2017-12-04]. ISSN 03043991. Available from DOI: 10.1016/S0304-3991(01)00154-1.
62. *EM software*. EMDataBank.org. Available also from: <http://www.emdatabank.org/emsoftware.html>.
63. SCHERES, Sjors H W; NÚÑEZ-RAMÍREZ, Rafael; SORZANO, Carlos O S; CARAZO, José María; MARABINI, Roberto. Image processing for electron microscopy single-particle analysis using XMIPP. *Nature Protocols* [online]. 2008, vol. 3, no. 6, pp. 977–990 [visited on 2017-12-04]. ISSN 1754-2189, 1750-2799. ISSN 1754-2189, 1750-2799. Available from DOI: 10.1038/nprot.2008.62.
64. XUE, Guoliang. An O(n) time hierarchical tree algorithm for computing force field in n-body simulations. *Theoretical Computer Science* [online]. 1998, vol. 197, no. 1-2, pp. 157–169 [visited on 2017-12-17]. ISSN 03043975. Available from DOI: 10.1016/S0304-3975(97)00232-6.

BIBLIOGRAPHY

65. CONFERENCE, Supercomputing (ed.). *Supercomputing '94: proceedings ; Washington, D.C., November 14 - 18, 1994*. LosAlamitos, Calif.: IEEE Computer Soc. Press, 1994. ISBN 978-0-8186-6605-6 978-0-8186-6607-0. OCLC: 246646775.
66. VERNON-PARRY, K.D. Scanning electron microscopy: an introduction. *III-Vs Review* [online]. 2000, vol. 13, no. 4, pp. 40–44 [visited on 2018-03-08]. ISSN 09611290. Available from DOI: 10.1016/S0961-1290(00)80006-X.
67. REIMER, Ludwig. *Scanning Electron Microscopy* [online]. Ed. by HAWKES, Peter W.; LOTSCH, Helmut K. V. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998 [visited on 2018-03-08]. Springer Series in Optical Sciences. ISBN 978-3-642-08372-3 978-3-540-38967-5. Available from DOI: 10.1007/978-3-540-38967-5.