Electron Microscopy and Image processing

Electron Microscope

Structure

Vacuum

The column of the electron microscope is under vacuum. The vacuum is of different quality, with the highest vacuum at the gun and the lowest vacuum at the detection system. The different qualities of the vacuum are achieved by small holes which separate the different segments but led pass most electrons unhindered.
Due to the vacuum the electrons can pass the column without colliding with residual gas molecules. Ideally, on the way from the gun to the detector the electrons are only scattered by the sample. Furthermore, a good vacuum reduces the contamination of the sample with residual gas during electron microscopy. It also increases the life-time of the filament.

Electron Emitters

Electron emission requires that energy is injected into the emitter in order to allow the electrons to overcome the work function and exit from the surface. In electron microscopy various principles are used for stimulating electron emission:
In thermionic emission the emitter is heated. This induces energy into the system and allows the electrons to leave the surface. The emission current J depends on the temperature T and the work function W and is described by the Richardson equation (k: Boltzmann constant, A material specific Richardson constant)

$$J = AT^2 e^{\frac{-W}{kT}}$$

At a certain temperature the electrons exit the surface of the emitter. Typical materials for thermionic emitters are tungsten or LaB_6. Tungsten has a work function of 4.6 eV and thus requires a relatively high temperature (ca. 2500-2800 k) to generate a suitable current. LaB_6 has a lower work function (ca. 2.5 eV) and thus requires a lower operating temperature (ca. 1800 K).
The advantage of thermionic emitters is that the technical requirements (simple design, less stringent vacuum) are low and emitters are relatively cheap. However, the emitting surface is relatively large and the brightness is relatively low thus the spatial coherence is low. This limits the resolution for images taken out of focus as required for unstained, vitrified biological samples that have a weak phase contrast. Furthermore, heating leads to a spread of the energy distribution of the emitted electrons of 1-3 eV (depending on the material and the temperature. The energy spread can be estimated by Maxwell´s speed distribution). The energy spread leads to low chromatic coherence of the electron wave which causes a defocus independent limitation in resolution.
Another emission principle is the field emission. Here a pointed tungsten filament is used. Ideally, the tip of the filament is only a few atoms thick. Emission is caused by an extraction field, which lowers the work function and allows the electrons to tunnel from the tip of the emitter. Emission only happens where the field is strong enough, which is at the very tip of the filament. Thus the emitting area is small and therefore the spatial coherence is high.

Furthermore, because the emitter is not heated the energy spread of the electrons is a magnitude lower than in thermionic emitters, giving much higher chromatic coherence.
In field emitters, the emission current J is independent of the temperature of the filament and is given by the Fowler Nordheim equation (a=1.5 x 10^{-6} A eV/V, b=6.8 eV V/nm; v: correction factor; F: field ;W: work function):

$$J = \frac{aF^2}{W} e^{-\frac{b\nu W^{\frac{3}{2}}}{F}}$$

Field emitters require a much higher vacuum than thermionic emitters. Furthermore, the sharp tip is quickly contaminated with a monolayer of gas. This reduces the extraction field and leads to large fluctuations in the emission current, which makes cold field emitter much more difficult to operate than thermionic emitters. Although, electron microscopes with cold field emitters are built, they are not or no longer used in Biology.
A compromise between the cold field emitter and the thermionic emitter is the Schottky field emitter. The Schottky field emitter uses a pointed tungsten tip, which is doped with ZrO_2. The ZrO_2 lowers the work function. Similar as in cold field emitters an external field is applied. But in case of Schottky (field) emission, the tip is less sharp and the field is not sufficient to extract the electrons on its own. However, the field lowers the work function. Therefore, a lower temperature (1400-1800 K) is sufficient to stimulate emission. Accordingly, Maxwell's speed distribution and thus the energy spread of the electrons is lower than in conventional thermionic emitters. The temperature without the field is also not sufficient to stimulate emission on its own. Therefore the emitted current depends on the extraction field and on the temperature. In principle, like for thermionic emitters the Richardson equation applies. However, the work function In the Richardson equation is lowered by the field F. The reduction in the work function ΔW is given by (ε_0 : electric constant ca. 8.85 x 10^{-12} F/m; F: field):

$$\Delta W = \sqrt{\frac{e^3 F}{4\pi\varepsilon_0}}$$

Thus Schottky field emission is an aided thermionic emission rather than a true field emission.
Because the field is highest at the tip of the emitter, emission happens only in a small area, which gives high brightness and good spatial coherence. The improved spatial coherence and the better chromatic coherence due to a lower energy spread of the emitted electrons give much better resolution than in thermionic emitters.
Furthermore, contamination of the tip of the emitter is less problematic than in cold field emitters, because the tip is less pointed. Therefore, the requirements for the vacuum are less stringent and the emission is very stable over time, making the Schottky emitter easy to operate. Schottky emitters are currently state of the art in high-resolution electron microscopy in Biology.

Lenses

Electrons are deflected by magnetic fields. Magnetic fields can be produced by passing a strong current through a wire. The field is further enhanced by coiling the wire around a soft iron core.

In electron microscopes, the electromagnetic effect is used for building lenses. The largest lens is the objective lens. It consists of a large coil surrounding a soft iron core that enhances the magnetic field. The soft iron core has a gap, which is capped by pole pieces on both sides of the gap. The pole pieces focus the magnetic field in the gap. The gap is not only the place where the electrons are deflected most, it also contains the sample. Usually the pole piece gap is just a few centimetres wide, which restricts the sample size. In contrast to light microscopy, the sample is positioned inside the objective lens. This allows a design, where the objective lens also acts as part of the condenser system that controls the illumination. An advantage of such a design is that columns can be shortened compared to a design, where the objective lens does not contribute to the condenser system.

The magnetic field of the objective lens reaches approximately 1-2 T and magnifies the image up to 50-times. For achieving an appropriate magnetic field a current of 10-20 A is passed through the wires. This leads to strong heating of the lens that is counterbalanced by water cooling. The stability of the water cooling is crucial for the resolution of the electron microscope. At this central place of the microscope close to the specimen a laminar flow of the water without air bubbles is required to keep the lens stable and free of vibrations.

Other deflection systems are pairs of coils, which point with their ends towards the electron beam. This exposes either the north or the south pole of the magnetic field to the electron beam. Deflection coils allow shifting of the beam along X and Y. For shifting in one direction, two coils are arranged co-axial, one with its north pole pointing to the electron beam, the other pointing with the south-pole to the electron beam. For deflection in two directions a second pair of coils is used, which is rotated by 90° in respect to the first pair.

For fine tuning the magnetic field and thus correcting for astigmatism quadrupoles are used. They consist of 4 coils, which are regularly arranged around the electron beam. Opposite coils point either with their North-pole or with their South-pole to the beam. Thus quadrupoles focus the beam in one direction and diverge it in the other direction. This leads to an elliptic distortion of the magnetic field. Two pairs of quadrupoles rotated by 90 ° in respect to each other are required for a full correction of astigmatic aberrations.

Detection System

Electrons are not directly visible. Thus for observing the image directly in the electron microscope, phosphorescent screens are used in the viewing chamber. The phosphorescent material is excited by the incoming electrons. The excitation is relaxed by the emission of light by a process which is called phosphorescence. This light can be observed either directly or with a binocular or can be photographed with a digital camera. The electron does not only excite the immediate area of its impact but a much larger surrounding area. This leads to some image blurring and limitations in the resolution. The thickness of the phosphorescent layer influences the sensitivity (thicker is more sensitive) and the spread of the information (thinner gives better resolution). A typical phosphorescent material is Zincsulfide (ZnS) which emits green light when hit by electrons.

Another detector is photographic film. If the electrons hit the grains in the photographic emulsion, they become exposed and can be reduced by the photographic developer. One electron is sufficient to several grains developable. The grains are typically between 0.5 and 1 μm in diameter. The electron penetrates several micrometers into the emulsion and can expose several grains. To minimize the spread of information the emulsions are thin (<10 μm). Taken together the incoming electrons on film expose a much smaller area than on phosphorescent screens. Thus, the spatial resolution of film is much better than for any recording system, which relies on a conversion of electrons into light by a phosphorescent

layer. Photographic material is relatively cheap and large areas (typically 8x10 cm) can be recorded in a single image. However, the linearity of the dynamics of film is only given over a small range and the signal to noise level is poor. Furthermore, for further digital image processing, the film has to be scanned, which leads to further compromises in the spatial resolution by the quality of the scanner. Although the spatial resolution of film is still superior to most detection systems, the cumbersome and slow scanning, the disappearance of precise scanners and the reduced availability of film, make the material slowly obsolete.

Instead of film electronic detector are used, which rely on CCD or CMOS technology. In both cases a chip with an array of light sensitive pixels converts a light signal into free electrons. In CCDs the charges are read out line by line and then converted into a voltage signal. For each line the charge of the individual pixels is transported and then individually converted into a voltage signal. Because of this process and the actual transport of charges, the read-out times of the CCD-chips are relatively slow. In the CMOS technology the conversion into voltage happens on the level of individual pixels. In principle each pixel can be addressed separately and be read out separately. Because the voltage is sensed and virtually no charge is transported, readout times are much faster. Despite this difference in read-out both detectors behave similar. They do not sense the electrons directly but require a scintilator, which converts the electrons into light (similar as the phosphorescent screen). The scintillator is applied to a glass fibre optics, which is clued onto the chip. The glass fibre optics transports the light from the scintillator to the chip. This also protects the chip from damage by the incident electrons. Other than chips found in digital cameras, the chips for electron microscopes are much larger. Today the standard size for chips are 4kx4k pixel with 15μm pixel size (relates to 16M-Pixel and approximately 6 cm x6 cm chip-size).

The limitations of CCD and CMOS detectors are in the scintillator and the area of the chip. Due to the excitation the information of a single incoming electron is spread over several 10^{th} of micrometers. Thus it does not make sense to generate very small pixels. In practice the typical pixel size is in the order of 15μm, which is still much smaller than the pixel size. For obtaining high resolution relatively large magnifications are required and thus with a given pixel size the field of view is limited. The very largest chips use a whole silicon waiver (12.5 cm) and have 8kx8k or 10kx10k pixels. These large chips are very expensive (ca. 500000 GBP). Furthermore they record electrons relatively far from the optical axis. In these areas distortions such as cushion distortions have a major influence on the image quality.

A new development is the direct detector, which is based on CMOS technology and uses a monolythic active pixel sensor. These detectors are optimized to convert the incoming electrons directly to a voltage signal without additional conversion step to light. Because the electrons enter the sensitive layer, they can lead to excitation several times at several places. This gives rise to a spread of the information similar as for all other detection systems. However, the information is similarly well localized as for photographic material. Furthermore, back thinning of the responsive layer of the chip reduces the spread of information further and leads to even better information transfer at high resolution. This is the first type of detector, which outperforms film. One problem of the direct detection is that the incident electrons reach the chip and lead to its ultimate destruction after recording some millions of electrons on a pixel.

The direct detectors can be read out very fast, while they are still recording electrons. In the simplest case this allows recording movies instead of a single micrograph. Alignment of the frames in respect to each other allows correcting for drift and beam induced movement, which improves resolution markedly. Some direct detectors can be read out very fast allowing the detection of single electrons. Since the signal of a single electron can vary, counting the electrons rather than measuring the change in the potential leads to an even better signal to noise ratio of the signal.

Image Formation

Interaction of Electrons with Matter

Most electrons pass the sample without any interaction. Part of the electrons is scattered close to the nuclei of the atoms of the samples. Because the electrons are much lighter than the nuclei of the sample the electrons retain their energy and change only their direction. The process is called elastic scattering, because there is no change in the energy of the electrons. Scattering angles in elastic scattering are relatively large. Some of the electrons are even back-scattered, others are scattered under such large angles, that they are removed by the apertures of the microscope. These electrons give rise to amplitude contrast. Most elastically scattered electrons pass through the electron microscope and interfere with the unscattered electrons and generate phase contrast.
Electrons, which pass further away from the nucleus, are deflected by the electrons in the shells of the atoms of the sample. Here the electrons deposit a bit of their energy onto the electrons in the sample. After the impact the direction of the electrons is only changed slightly. Because the electrons have less energy after scattering, the scattering event is called inelastic.

Sample Preparation

General Considerations

Sample preparation aims at preserving the structure of the sample in the adverse conditions inside the electron microscope. The main challenges which are faced by the sample is damage by the bombardment with electrons during the observation and the vacuum inside the electron microscope. The dose, which is experienced by the sample during a 3 s observation at high magnification, is in the order of 4000 Mrad, which is similar to the dose in the centre of an atomic bomb explosion. The deposition of energy during bombardment leads to bond breakage, which is considered as a primary event of beam damage. The resulting fragments are often no longer stable in the places where they are generated and start moving around. In addition the fragments contain highly reactive radicals, which react easily with reaction partners. The movement of fragments and further chemical reactions are considered as the secondary events of beam damage. While the primary events of beam damage are largely independent of the temperature, the secondary events are temperature dependent and thus can be reduced by simply cooling the sample. For example at 4 K most compounds are solid and diffusion is much slower. So the break-down products remain much longer where they are generated and cause less secondary damage. Therefore, at lower temperatures high-resolution information is preserved at a higher electron dose.
Another possibility to harden samples against beam damage is to preserve the structure in a state, which only responds very little to the dose challenge. One possibility is using imprints in inorganic salts, which change only slowly in response to the dose even at room temperature. This strategy is used in negative staining.
The other adverse effect in the electron microscope is the vacuum. At room temperature the vacuum inside the microscope leads to immediate evaporation of all structural water in the sample, which leads to collapse and flattening of the sample. The flattening happens mainly towards the support-film. Therefore, for untilted specimens the direction of flattening is in the same direction as the direction of projection of the image and thus a projection of a flattened sample does not look too different to the projection of an undisturbed sample. At low resolution the distortions have only a limited effect on the interpretation of the data. However, as soon as the direction of projection and the direction of flattening are no longer co-axial (as for example in tilted specimens), the distortion becomes evident. There are

various ways to reduce flattening effects. One possibility is to replace water by a substance, which can satisfy the hydrogen bonds of the protein similarly well as water but which has a lower vapour pressure and thus, does not evaporate. This property is satisfied by sugars such as glucose or trehalose, which have been successfully used for preparing two-dimensional protein crystals. However, the density of the sugar matches approximately the same density as of the protein. Therefore, there is almost no contrast from the envelope of the particles and contrast arises mainly by internal density differences inside the protein (only visible at high resolution, better than 1 nm). Alternatively the vapour pressure of water can be reduced, by cooling the sample. In the high vacuum, at some -130C the water does not evaporate and collapse of the sample is avoided. This is even beneficial for stained samples, which retain their structural water if they are cooled and thus flatten less.

Staining

For staining materials are used, which have a larger cross section for electrons than the low atomic number of biological material. Typical stains are heavy metal salts such as uranyl acetate, uranyl formate, phosphotungstic acid, silicotungstic acide or ammonium molybdate. Stains can either interact directly with the biological material or are just excluded from areas where the sample is located. The former gives rise to positive staining, because the direct interaction with the sample leads to an enrichment of the stain at the sample. Due to the accumulation of the stain at the sample, the sample appears dark. Positive staining is difficult to control and very variable. Thus it is not used for structure determination but is an occasional artefact. Positive staining often happens in highly charged areas of the biological object such as RNA in ribosomes and other complexes.
In negative staining there is no direct interaction of the sample with the stain. So the stain forms a more or less amorphous layer and is excluded from areas where the protein is located. Thus the sample leaves an imprint in the stain and depletes stain from areas which it occupies. Because the heavy metal ions have high atomic numbers, they scatter electrons much stronger than the low atomic number biological material. Thus contrast arises mainly from the pattern in the stain and to a much lesser extent from the sample. The areas, which were occupied by protein, appear bright (less scattering).
For staining, an aqueous solution of the protein is incubated on a carbon film. Some of the proteins adsorb to the carbon film. The buffer is then exchanged by washing the surface of the carbon film with a few droplets of aqueous staining solution (1-2% stain). Then, most of the sample is removed with filter paper. A thin layer of the staining solution remains on the surface and dries. After drying, the stain forms a crust at the stain accessible areas leaving a cast of the protein. This cast is quite robust, keeps for months and can be imaged in the electron microscope at room temperature.

The advantages of staining are the high contrast and the resistance to beam damage. Therefore, this method is well suited for looking at the sample and making assessments of the image quality. Furthermore, the superior contrast of the stain make small proteins and small complexes accessible (<200 kDa) to electron microscopy, which often do not generate sufficient contrast for image processing of an unstained sample.
The disadvantages of staining are:1) It is an indirect method, where not the protein is observed but only its cast in the stain. This ultimately limits the resolution. 2) In most cases the protein is exposed to high salt concentrations during staining and often also to low pH values (e.g. UoAc). This often leads to structural damage or alterations of the structure. 3) If stained samples are imaged at room temperature, structural water is lost and the samples are flattened. 4) If the sample is not completely embedded in the stain, the resulting images represent only parts of the structure. Flattening and partial embedding are especially cumbersome if images from such samples are used for image reconstruction, because

differently oriented particles are differently distorted and thus are no longer compatible with a unique three-dimensional shape.

Vitrification

To retain the water in the vacuum and thus to avoid structural collapse, the vapour pressure of the water must be below the ambient pressure. The vapour pressure can be lowered below the ambient pressure inside the column by decreasing the temperature below approximately -100C. At this temperature, the most stable modification of water is crystalline, hexagonal ice. Hexagonal ice requires more space than liquid water. This is why the ice floats on the lakes in winter. While ice crystals grow, they expand and put tremendous pressure onto the sample. This damages larger structures (e.g. filamentous structures become bent). In addition, crystalline ice in certain orientations reflects the electron beam giving rise to Bragg-reflections. These Bragg-reflections give an uneven background with dark and white areas and distract from the features in the sample.

To avoid Bragg-reflections and to reduce damage to the sample by growing ice crystals, it is desirable to maintain the liquid properties of the water (amorphous solid or glass-like). This can be achieved by rapid cooling, which prevents the water molecules to rearrange into a crystal before they become rigid in a glass-like state (vitrification). The required cooling rates for vitrification are in the order of 15000 k/s. Although the density of vitrified water is lower than of fluid water (0.93 g/cm^3 compared to 1 g/cm^3), the water does not form ice crystals and probably flows around the biological sample while it expands. Thus, vitrification does not cause any obvious damage to the structure of biological specimens.

The vitrified state is only stable at temperatures below -150C and starts to convert slowly into a cubic crystalline form at temperatures above (no vitrified state above approximately -130C). This means that after vitrification the sample needs to be handled, stored and imaged below -150C by cooling it with liquid nitrogen.

Rapid cooling rates below the vitrification temperatures are the key for obtaining vitrification. To achieve these cooling rates, cryogens need a relatively high boiling point to prevent film boiling, when the “hot” sample enters the cryogen. Film boiling generates an insulating gas-layer, which slows down cooling despite the low temperature of the cryogen. This effect is called Leidenfrost effect. Therefore, for vitrification liquid ethane (Tb=-89C) or less commonly also liquid propane (Tb=-42C) are used as cryogens instead of liquid nitrogen (Tb=-196C), which is used for cooling the cryogens. In addition vitrification only works for thin films (some 100-500 nm thick), because for thicker films the thermal conductivity of the sample is not sufficient to remove the heat from the sample quickly enough to obtain vitrification.

The most commonly used vitrification devices are guillotine-like machines, where a grid is mounted in a pair of tweezers, which is hold by the guillotine. Approximately 2-5 µl of sample is applied to the grid. Then, most of the sample is removed by blotting with filter paper. Afterwards, the grid is plunged into liquid ethane. Variations of this setup use a computer to control blotting times and blotting forces for more reproducible results. In addition many set-ups employ environmental chambers, which keep the humidity close to the grid at 100%. This prevents evaporation of the sample, which would lead to uncontrolled, slow cooling and would increases the concentrations of the solutes and particles.

Alternatively, the sample is sprayed in a thin layer onto the grid, while the grid is plunged into the cryogen. This setup is less commonly used, because it requires more sample and the thickness of the sample is difficult to control. It depends on the size of the droplet and the surface properties of the solvent, which control how droplets spread. However, spraying is very useful for time-resolved experiments, where reaction events << 1 s are trapped.

Image Processing

In 2-dimensions, image processing aims at: Determining the parameters of the contrast transfer function and correcting for it, aligning particles in respect to each other, identifying particles which are similar to each other, measuring similarities between particles and averaging similar projections for improving the signal to noise ratio. Most aspects of 2-dimensional image processing assume that the object exists in multiple, identical copies. A prerequisite for averaging similar particle images is to know their in-plane orientations, which are given by the origin of the particles (x,y) and their in-plane rotation angles (α). If these parameters are known, the particle images can be aligned in respect to other and then coherently averaged.

In three-dimensions, image reconstruction aims at restoring the third dimension from the projection images that were acquired in the electron microscope. This requires different projections of the same object, which can either be obtained by recording a tilt series of a unique object or by combining different views from different copies of the same object. A prerequisite for reconstructing the three-dimensional volume from projections is to know the orientation of the object that was projected. The orientation in 3D is given by the three Euler angles ($\Phi\ \theta\ \psi$) and the origin of the particle (x,y).

Fourier Transforms

Many aspects of image processing use Fourier Transforms, because they have desirable properties and speed up the computation. The basic idea of Fourier Transforms is that every function (including an image) can be generated by adding up a series of cosine functions. The cosine functions are determined by the wave-length (distance between repeating features) of the spatial wave, the phase (where the cosine wave starts, if the phase is 0 the wave starts with a maximum, if the phase is 90° it starts with a minimum, etc) and the amplitude (height of the maximum). Coarse features are mainly represented by waves with large wave lengths or small frequencies ($1/\lambda$) and fine details are mainly represented by waves with short wave lengths or large frequencies. The largest wave length in a square-image is exactly the box size. Permitted wave lengths are integer fractions of the box size (the cosine wave should finish in the box with the same phase as it starts). The smallest wave length is two pixels (you need one pixel for representing the maximum and 1 for representing the minimum). The latter is also the finest feature, which can be resolved in an image. So, if an image is sampled with 5 Å/px the best resolution which can be achieved is 10Å (2px). The spatial frequency (1/2 Px), which refers to this wave length is called Nyquist frequency.

Fourier Transforms decompose the image into the spectrum of spatial frequencies, which make up the image.

Fourier space is reciprocally related to real space. So features which are very small in real space appear large in Fourier space and vice versa. Therefore, Fourier space is often referred to as reciprocal space. Each point in Fourier space has two values associated with it. One describes the phase and the other the amplitude of the wave. A diffraction image (e.g. observed in the back focal plane of the electron microscope, or determined in a diffraction experiment in X-ray crystallography) measures the intensities of the Fourier transform (amplitudes^2) but has no phase information. A Fourier transform of an image has the amplitudes and the phases. So, no information is lost in a Fourier Transform. The phase information is important for knowing, where things in the image are. If random amplitudes are assigned to correct phases, the resulting image is still recognizable whereas with the assignment of random phases to correct amplitudes, the resulting image is unrecognizable.

General Principle of Image reconstruction

There are different ways how to think about 3D-image reconstruction. One was formulated by De Rosier and Klug in 1968. The basic idea is that due to the large depth of focus, an electron micrograph is a projection of the object. In Fourier space, the 2D-Fourier transform of the 2D-projection is a central section through the 3D-Fourier space. Different projections of an object give differently oriented sections of the Fourier space, which all go through the origin of the Fourier Space. By combining the information from different projections the Fourier space can be filled and the three-dimensional image information can be recovered by Fourier inversion. All what is required is an even sampling of the Fourier space.

Back projection

Image reconstruction can also be formulated in real space. The basic idea is that the 2D-projections are expanded into the third-dimension back into the direction from where they were originally generated by the electron beam. Projections from different directions are projected back onto the same volume and are averaged. This generates a 3D-map, where density, which represents the real volume, adds up much faster than in the background. Isotropic resolution is achieved, if the projections are evenly distributed.

Back projection and Fourier reconstruction follow the same basic idea. In real space expanding the 2D-projection into the third dimension, means that in Fourier space it is compressed into a section. The direction of back-projection determines the orientation of the Fourier-section in 3D-Fourier space

Principle of Tomography

A simple method to obtain evenly distributed views of an object is to tilt it inside the microscope and to record a series of projections from the differently tilted objects in regular angular intervals. A tilt range of +/- 90° around a single axis would be required for a complete sampling of the three-dimensional space. However, the mounting of the sample generates a shadow, which limits the practical tilt range to approximately +/-70°. Furthermore, with increasing tilt angles, the length of the sample that is penetrated by the electron beam increases (ca. 3 times at 70°, 2 times at 60° and 1.4 times at 45°). The increased thickness makes the quality of the images much worse than of the untilted object (e.g. due to increased multiple scattering of electrons) and in many cases limits the tilt-range even further (e.g. +/- 60°).

Irradiation of the sample with electrons causes beam damage. At the surfaces of organic material in vitrified water bubbling becomes evident at approximately 100 e/Å^2 . This means that the dose for recording the whole tilt series should remain below this critical dose and the permissible dose needs to be fractionated over all images of the series. As a consequence the individual projections are very noisy and also the details in the resulting three-dimensional reconstructions are limited by the signal-to noise ratio.

For an accurate reconstruction, all the images of the tilt series need to be centered in respect to each other and the position of the tilt axis and of the tilt angle need to be known. For the accurate determination of these parameters, electron dense fiducial markers are used. These are usually small colloidal gold particles (5-10 nm), which are randomly distributed across the sample. Due to the high scattering cross section of gold and the size of the colloidal gold particles they generate a strong contrast (black dots) and can be easily recognized in the individual micrographs even at low dose. The distribution of the fiducial markers changes in dependence of the relative position of the projections (x,y), the orientation of the tiltaxis and the tilt angle. By tracking the positions of several markers throughout the tilt series, the parameters of the origin, the tilt angle and tilt axis can be accurately determined.

For resolving features of certain spatial frequencies in 3-dimensions the angular increment of the tilt needs to be adapted to the thickness of the object. For resolving a detail of size r in an object of diameter D, the angular increment β is:

$$\frac{r}{D} = \arctan\beta \approx \beta$$

This relation can be best understood in Fourier space. The thickness of the object determines the thickness of the section of a single projection in Fourier space (1/D). In Fourier space the sections are tilted around a single axis. For a given angular increment there is a certain radius (1/r) up to which the Fourier space is completely sampled/covered. This radius gives the largest spatial frequency, which can be resolved. These considerations are often referred to as the Crowther criterion. The Crowther criterion is a geometrical consideration, which does not account for signal-to-noise considerations. Furthermore, the Crowther criterion assumes that the tilt angle is +/-90°. In case of tomography with a limited tilt range, there is no structural information in the direction of the missing wedge. Therefore, tomograms have an anisotropic resolution, which is much worth in the z-direction. A practical consequence is for example that tomograms of vesicles and liposomes miss the upper and lower part and resolve only the sides of the spherical structure.

Random Conical Tilt reconstruction

Many biological objects come in multiple identical copies, which can have random orientations. Random conical tilt reconstruction is based on particle images, which show a projection in the same direction and only a variation of the in-plane rotational angle. For the random conical tilt reconstruction pairs of micrographs are required instead of a complete tilt series in tomography. The first micrograph of the pair is of the highly tilted sample (used for the reconstruction), the second one is of the untilted sample. Because only two micrographs are taken, each micrograph can be taken with a higher dose (fractionating of the total dose over 2 micrographs) than in a tomographic series, which gives a better signal to noise ratio in the micrographs. Now, particle images are selected in pairs from both micrographs (pair: same particle in both micrographs). Comparison of the distribution of the particles on both micrographs informs about the position of the tilt axis and the tilt angle (similar to the fiducial markers in tomography). From the micrographs of the untilted particles the relative in plane rotations are determined. These in-plane rotations, together with the orientation of the tilt axis and the tilt angle, give the complete information of the direction of projection for the tilted mates. For the random conical tilt reconstruction this information is used to calculate a reconstruction of the centered tilted particle images. The images of the untilted particles are not included in the image reconstruction.
The in plane rotation of the particles is random (“random” in Random Conical Tilt reconstruction). A good coverage of the in-plane angles is required for a good three-dimensional resolution. If a good distribution of in plane rotations is obtained, the resulting image reconstruction has only a missing cone (“cone” for conical in Random Conical Tilt) in Fourier space. The opening of the cone is given by the tilt angle (half opening angle of the cone is 90°- tilt angle). Because the missing cone is much smaller than the missing wedge in tomography and because averaging over many individual particle images gives a much better signal-to-noise ratio, the resulting reconstruction has a better resolution than the resolution obtained from individual tomograms. However, the resolution is also unisotropic, because of the missing cone.

Averaging

For single particle applications, which aim at higher resolution, micrographs are recorded with doses between 5 and 20 e/Å^2 for minimizing radiation damage and for keeping structural information below 10Å resolution intact. At this dose the average spectral signal to noise ratio for a vitrified biological complex is between 0.01 and 0.1 (with peak values between 0.1 and 1 depending on the size of the object). To recognize features clearly, a signal to noise ratio of 1-3 is required. This means, that the information of images from different identical particles in identical conformations need to be coherently averaged to improve the signal to noise ratio.

There are different definitions of the signal to noise ratio (SNR). In image processing the signal to noise ratio is often defined as the signal amplitude (A) divided by the standard deviation of the noise (σ):

$$SNR = \frac{A}{\sigma}$$

In this case if the "same" image is measured n-times and added, the signal amplitude grows by a factor of n whereas the uncorrelated noise grows only by the sqrt(n). Following this definition the signal to noise ratio grows with the sqrt(n) with the number of averaged observations (n).

Averaging is a very powerful method for improving the signal-to-noise ratio in image processing. In praxis in the few cases, where resolutions of approximately 3.5Å were reached approximately 10^5-10^6 asymmetric units were averaged. For 20Å resolution only approximately 10^4 particle images are required.

It is also possible to average different volumes. This is done in subtomogram averaging or in averaging of random conical tilt volumes. For effective averaging, the volumes need to be aligned in respect to each other. Special care must be taken for areas, where information in the missing wedge or missing cone of one volume is averaged with properly sampled areas in another volume. Adequate averaging of differently oriented volumes fills in the missing information and can lead to isotropic resolution in the averaged volumes.

Alignment and Correlation

Effective averaging requires that the averaged images are in the same orientation. For measuring relative in plane shifts the cross correlation function is used. Cross correlation can be understood as a physical experiment, where two translucent images show the same features but are shifted in respect to each other. The two images are superimposed and light is shone through both images. At the other side the cumulated light intensity is measured. Now the images are shifted in respect to each other. If the intensity of the light becomes maximal, the two images are properly aligned in respect to each other. In mathematical terms this is formulated in the cross correlation function CCF:

$$CCF(r_{pq}) = \frac{1}{N'} \sum_x \sum_y imageA(r_{x,y}) \cdot imageB(r_{x,y} + r_{pq})$$

In this formula the grey values in image A are multiplied pixel by pixel ($r_{x,y}$) with the grey values in image B, which is displaced by a displacement vector (r_{pq}) and are summed and normalized by the number of overlapping pixels (N'). The best correlation is given for the displacement vector for which the function becomes maximal.

For finding the in plane rotation, the two images are transformed into cylindrical coordinates. In cylindrical coordinates one axis depicts the radius r and the other the angle α. The relation between cylindrical coordinates and Cartesian coordinates is given by

$$r = \sqrt{x^2 + y^2}; x = r * \cos\alpha; y = r * \sin\alpha$$

A rotation in the Cartesian coordinate system is a shift in the cylindrical coordinate system and thus the rotation can also be determined by the Cross correlation function using cylindrical coordinates.

The cross correlation function is also a measurement for the similarity between two images. If the images become more and more dissimilar, the peak of the cross correlation function becomes smaller and smaller. However, the absolute height of the peak depends on the grey value distribution in the two images. So for global comparisons, the peak of the cross correlation function needs to be normalized by the grey value distribution. This normalized value is referred to as the cross correlation coefficient CCC.

$$CCC = \frac{\sum_x \sum_y (\text{imageA}(r_{xy}) - \overline{\text{greyA}}) * (\text{imageB}(r_{xy}) - \overline{\text{greyB}})}{\sqrt{\sum_x \sum_y (\text{imageA}(r_{xy}) - \overline{\text{greyA}})^2 * \sum_x \sum_y (\text{imageB}(r_{xy}) - \text{greyB})^2}}$$

In the formula greyA and greyB refer to the average grey value of the respective images. The normalized cross correlation coefficients can have values between +1 and -1. 1 refers to identical images and thus to a perfect match. Zero indicates that there is no similarity and negative values indicate an anti-correlation, e.g. by comparing a positive image with a negative image. The cross correlation coefficient is also a measure for the signal to noise ratio (SNR=CCC/1-CCC). So a cross correlation value of 0.5 refers to a SNR of 1.
Cross-Correlation-coefficients are used in multi-reference alignment to measure to which reference image a noisy image matches best (projection matching). Often this is used for supervised classification (reference based classification). It is also assumed that particle images, which group into the same class (align best to the same reference) have similar properties, e.g. similar spatial orientation or similar conformation or similar composition as the reference image.

Iterative Refinement in Single particle Image Processing

Single particle image processing aims at coherently averaging as many particles as possible in a three-dimensional map. In order to do this effectively, for each particle image the in-plane origin (x,y) and the spatial orientation (Euler angles) need to be known. These parameters are required for combining the particle images into a 3D-map either by Fourier inversion or by back projection.
Iterative refinement techniques use a starting three-dimensional model (e.g. determined by random conical tilt reconstruction or by sub-tomogram averaging). This model is projected with equally spaced angular increments into different directions. These projections are used as reference projections. In multi reference alignment all noisy particle images are aligned in-plane in respect to the reference projections and the correlation coefficient is determined. It is now assumed that a noisy particle image is most similar to the reference projection, with which it has the highest cross-correlation coefficient. It is then assumed that after in plane

alignment it also has the same spatial orientation/Euler angles as the best matching reference projection. This gives all parameter (x,y, $\Phi\ \theta\ \psi$) that are required for adding the particle image correctly into the 3D-image reconstruction in back projection or by Fourier inversion. The new image reconstruction is used to generate new reference projections. The whole process is iterated until no further improvement occurs. Ideally, during refinement the 3D-map becomes more accurate thus the in-plane alignment becomes more precise and the correlation coefficients increase because reference projections and real images match better. In addition if the reconstruction is pushed towards higher resolution a finer angular sampling is required, increasing the number of unique reference projections. This makes projection matching a very computing intensive method, which requires compute clusters with many processors.
Iterative refinement is not limited to single volumes. It can also be used to recognize and separate different conformations. This requires that reference projections from several 3D-maps (e.g. presenting different conformations or different compositions) are included in the reference projections for comparison.

Further Reading

Classical papers and reviews:

DeRosier, D.J. and Klug, A. (1968) Reconstruction of three dimensinal structures from electron micrographes. *Nature*, **217**, 130-134.

The paper describes the general principle of 3D-reconstruction by Fourier inversion. Here it is used for a helical object. Absolutely classical!

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A.W. and Schultz, P. (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys*, **21**, 129-228.

Dubochet is the inventor of electron cryo microscopy of vitrified samples. This review is an excellent source of information about the relevant phases of water, freezing techniques, beam damage etc. It is still used as general reference for “cryo microscopy”. There is a lot of information which is relevant for the lecture. Definitely, worth to have a look

Radermacher, M. (1988) Three-Dimensional Reconstruction of single particles from random and nonrandom tilt series. *J.of Elec.Micr.Tech.*, **9**, 359-394.

Radermacher has introduced random conical tilt reconstruction. Although the math might be difficult, the figures are helpful to understand the geometry and the general considerations.

Text-Books and related recourses

Frank (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. New York: Oxford University Press. ISBN 0-19-518218-9.

This is the only proper text book on single particle image processing in Biology. It incorporates the math and the general principles. In addition, in the introduction it explains some general aspects of electron microscopy (sample preparation, electron microscope, contrast transfer function)

Reimer and Kohl (2008) Transmission electron microscopy. Heidelberg: Springer. ISBN 0387400931, 9780387400938

Relevant information on the hardware, the optics, image formation and contrast transfer. This might be hard to read and requires a lot of physics. But if you are keen on the details, this is where I would look them up

http://www.els.net/WileyCDA/ElsTopics/L1-STB,L2-EMS.html

Wiley has a collection of reviews on electron microscopy. Most of them are written for undergraduates. There is a lot of useful and easily understandable information here.

Highlights

Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C. and Grigorieff, N. (2008) Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc Natl Acad Sci U S A*, **105**, 1867-1872.

This example was briefly shown in the lecture. It is a very good paper, which dissects, what is important to get this resolution. It is the milestone, which proves that near atomic resolution can be achieved by single particle image analysis.

Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V. and Stark, H. (2010) Ribosome dynamics and tRNA movement by time-resolved electron cryo microscopy. *Nature*, **466**, 329-333.

Awesome single particle analysis of some 2 million ribosomes. It resolves approximately 50 different conformations of the ribosome, during translation. Starting such an analysis requires courage and absolute commitment.