Lecture 4
H Saibil

Image acquisition, noise, particle selection, alignment

Signal and noise
Detectors
Particle selection
Alignment

Image processing for cryo microscopy

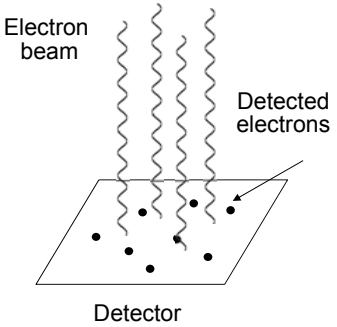
1 - 11 September 2015

FEI COMPANY
EMBO Leica JEOL
Direct Electron

Practical Course
Birkbeck College London

What is noise ?

In structural biology, radiation damage limits the useful electron dose from ice-embedded specimens to ~ **10-20 electrons per Å²**. Stochastic or counting noise arises from variations in the number of electrons which arrive at a particular point, and can be described by a Poisson distribution. The signal to noise ratio improves by a factor of √N as the electron dose increases. However, the dose must be kept low to minimise **radiation damage** to the specimen. Electrons can transfer energy to the specimen, breaking bonds and causing mass loss in biological molecules. For analysis, we assume that the image is the **sum** of the structure information plus statistical noise.



electron dose = average (N) ± √N

N = 100 electrons per unit area		
108	90	103
102	95	114
94	105	89

√N / N = 10%

N = 1000 electrons per unit area		
1025	1007	980
967	894	1016
1010	1046	964

√N / N = 3.3%

Signal, noise and detection

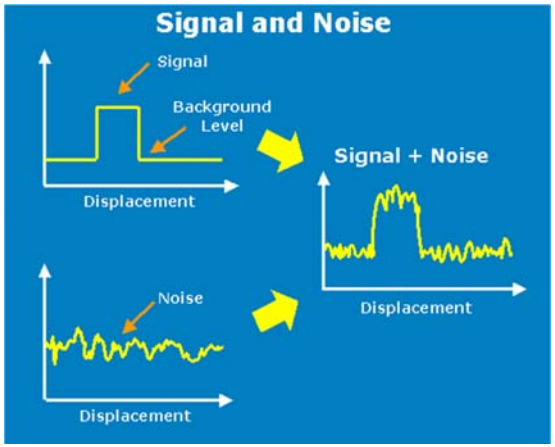
SNR: Signal-to-noise ratio

DQE: detective quantum efficiency

DQE gives a measure of how faithfully the signal is transmitted by the imaging system and is
= **SNR_output²** / **SNR _input²**

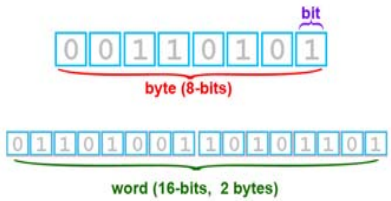
where output = digital image, input = electrons.

If DQE = 0.5, effectively half the electrons have been transmitted.



Digital images

Computer storage



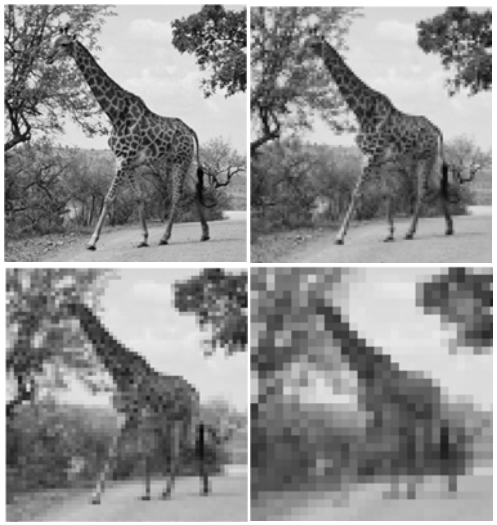
Each picture element stored in the computer, with its own grey level, is called a **pixel**. A pixel can be 8 bits (2^8 = 256 grey levels), 10 bits, 12 bits or more, depending on detector and storage formats.

Pixel grey levels

Black to white is 0-255 for an 8-bit image.

Sampling

Pixel resolution is calculated as detector pixel size (or scanning step size for film) divided by magnification. Example: for a detector pixel size of 6 □m and 42kx magnification, pixel resolution in Å is 60000/42000 = 1.43
Note : 1 □m = 10^4 Å.



The Nyquist resolution limit

Images can be represented as the sum of a series of sinusoidal waves. The highest frequency wave term defines the resolution limit. When an image is recorded or digitised, the wave components are sampled at an interval defined by the pixel size.

Harry Nyquist, Swedish physicist

Nyquist frequency is **half** the sampling **frequency** : if sampling frequency = 1/15 □m, Nyquist frequency = 1/30 □m.

Nyquist distance is **twice** the sampling **distance**.

In other words, the Nyquist frequency is the sampling needed to reconstruct a particular sine wave. For example, the reconstruction of a 10 Å map requires a sampling of 5 Å/pixel although in practice it is advisable to sample at 1/3 the required resolution, 3 Å.

Grey scale resolution

2

number of grey levels

4

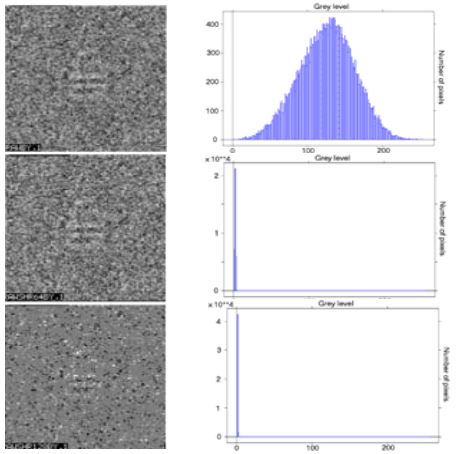
16

144

Grey scale resolution

Images are stored as 8 bits (256 grey levels) or more. But if data collection or conversion are done incorrectly this can produce a digital image represented by reduced number of grey levels, causing degradation of the image and loss of information.

Histogram of densities should be examined to ensure all the grey levels are represented.

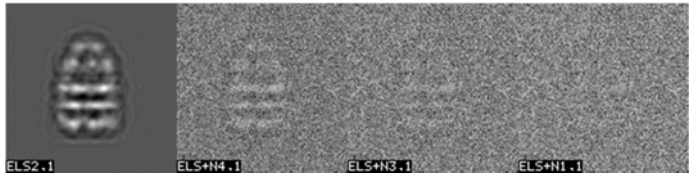


Scaled into 256 grey levels

5 grey levels

3 grey levels

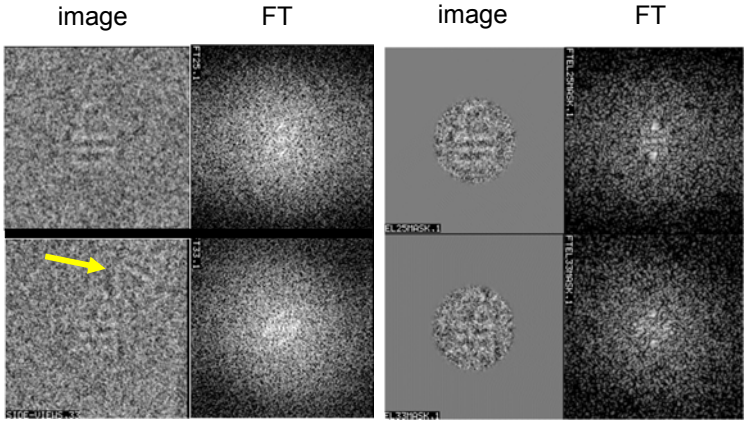
How much noise can we tolerate in single particle analysis?



Model image	Signal:Noise	Signal:Noise	Signal:Noise
2 □ defocus	4:1	2:1	1:1

At SNR of 1:1, we can barely detect the particle. Alignment programs will not work well under such conditions as noise will correlate as well as signal.

Noise reduction by real space masking

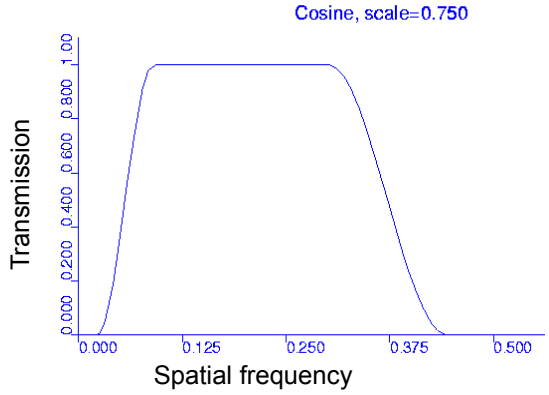


Information from neighbouring molecules contribute to the Fourier transform but can be removed by masking, leading to better alignment. The mask must not be too tight and its edges must be softened by a Gaussian or similar to prevent ripple effects which may introduce new alignment problems.

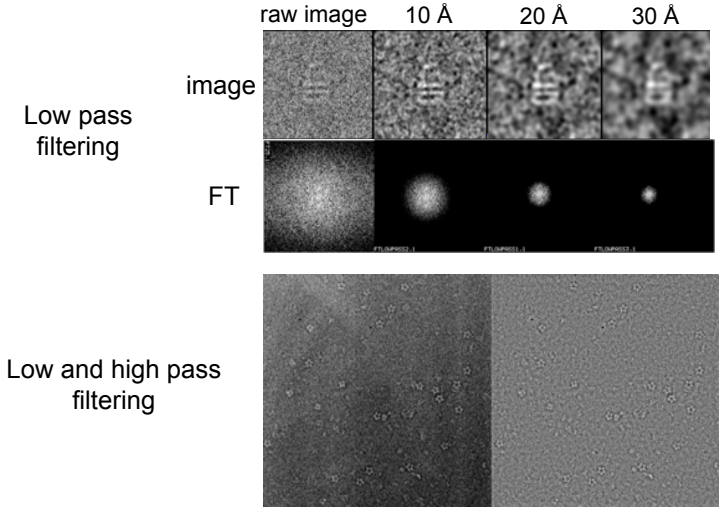
Noise reduction by Fourier bandpass filtration

This is used to remove low and high frequency noise such as uneven illumination (low) and film fog or detector noise (high). Selected frequencies of the Fourier transform of an image can be removed before back-transformation.

Fourier bandpass filters should have smooth edges (e.g. Gaussian, Cosine bell, Cauchy) to avoid introducing ripple artifacts.



Effects of low and high pass Fourier bandpass filtration



Calculation of filter radii

Processing programs need **filter radii** in transform pixels or fractional units :

R = S x N / F

where **S** is sampling in Å/pixel

N is box size

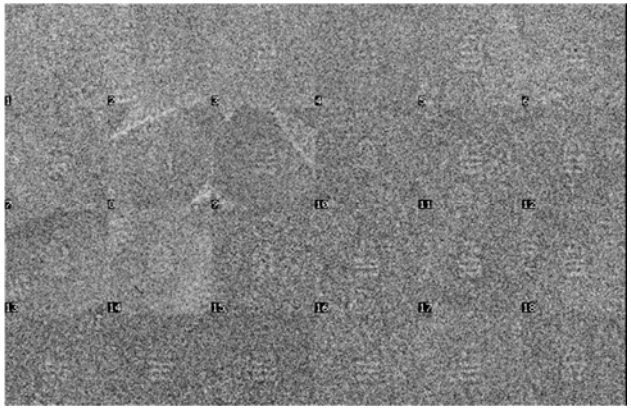
F is required resolution in Å

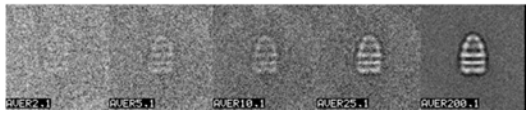
Since the maximum radius is half the box size, the maximum theoretical resolution is double the sampling (related to Nyquist frequency). The fractional value used in the programs (**R/N** or **S/F**) uses the range 0 – 0.5 (e.g. in Spider, EMAN) and 0-1 in IMAGIC.

In practice, the **high pass filter** reduces background gradients: its radius must exceed the particle diameter. The resolution of the calculated model will be restricted by the **low pass filter**. For example, with a pixel size of 5 Å and maximum particle diameter 150 Å, you might choose a high pass filter radius corresponding to a spacing of 180 Å and low pass to 20 Å (fractional units 5.0/180 = .028 and 5.0/20.0 = 0.25 respectively).

Noise reduction by averaging

Raw images

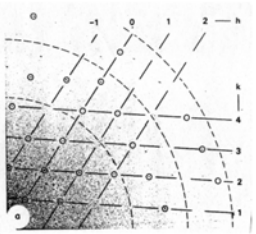




Averages of 2 5 10 25 200 images

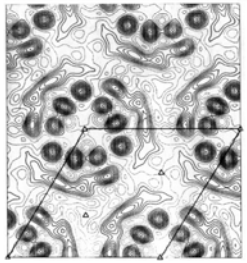
Noise reduction by filtering out everything except diffraction peaks - Bacteriorhodopsin

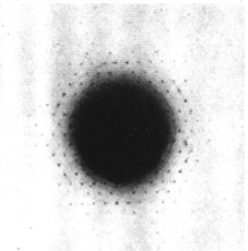
image



FT of image

Map after filtering, unbending, averaging

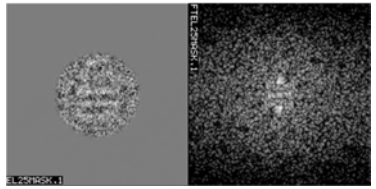


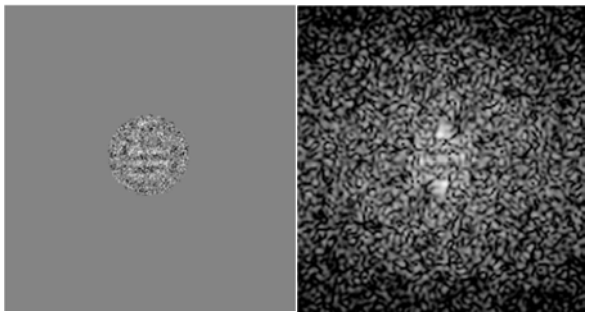


Electron diffraction

Henderson & Unwin, 1974

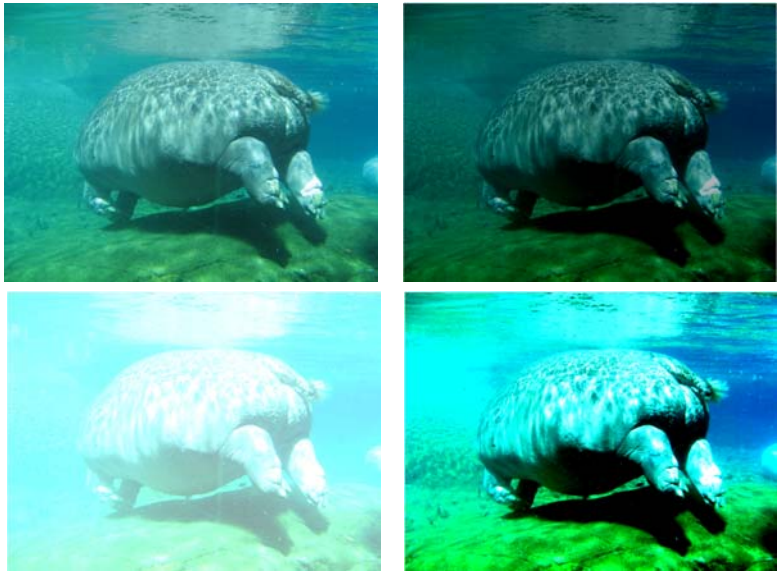
Padding images increases the sampling of the transform





Adding a frame of background around the image samples the transform more finely and reduces rounding errors during interpolation. This is important for cross correlation calculations used in alignment.

What is normalization and why is it important?



What is normalization and why is it important?

For an individual image, the range of pixel values may not matter because display programs scale the image values to the screen, but what if you are doing a series of cross correlations to align images to a set of references? If some references have very big numerical values, they will give higher cross correlation peaks and will falsely appear to provide the best match.

Moreover, the value of the FT at the origin is equal to the mean value of the image, so a high mean value will give a huge central peak on the FT which is unrelated to the structural information.

You MUST **normalize** all images and references used for cross-correlation.

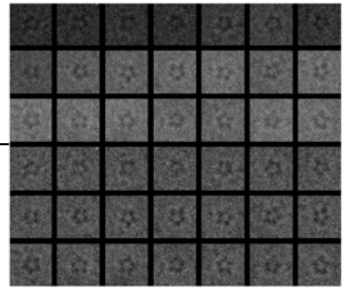
This means rescaling their pixel values to a **mean value of zero** and to the same **standard deviation**.

Normalisation

- $X_{norm} = (X - m) / s$, X pixel density, m mean, s standard deviation.
- Normalisation scales the data so the variation in each particle is equivalent.
- Care should be taken when normalising model projections.

Raw images

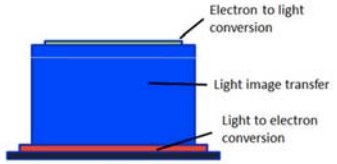
normalised images



Electron detectors

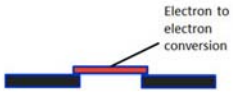
CCD

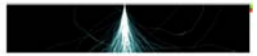
- Incident electrons excite phosphor screen to emit photons, which are converted to charge by a semiconductor (charge-coupled device)
- Charge read via coupling mechanism



Direct

- Incident electrons excite semiconductor electrons, generating a charge
- Charge read and converted to voltage
- Technical issues solved: radiation hardness, fast readout, back thinning





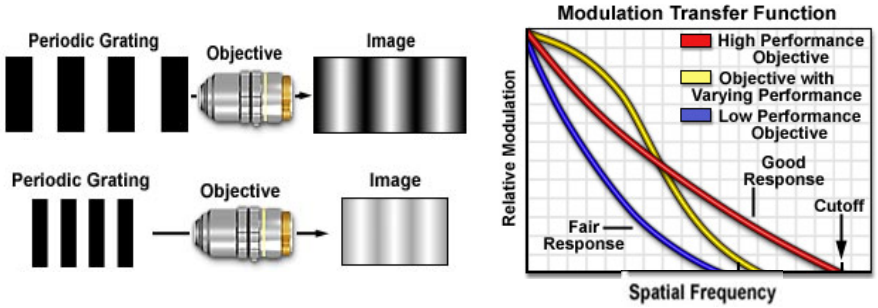
Greatly improved speed, resolution and sensitivity

Gatan web site

Optical performance as a function of resolution

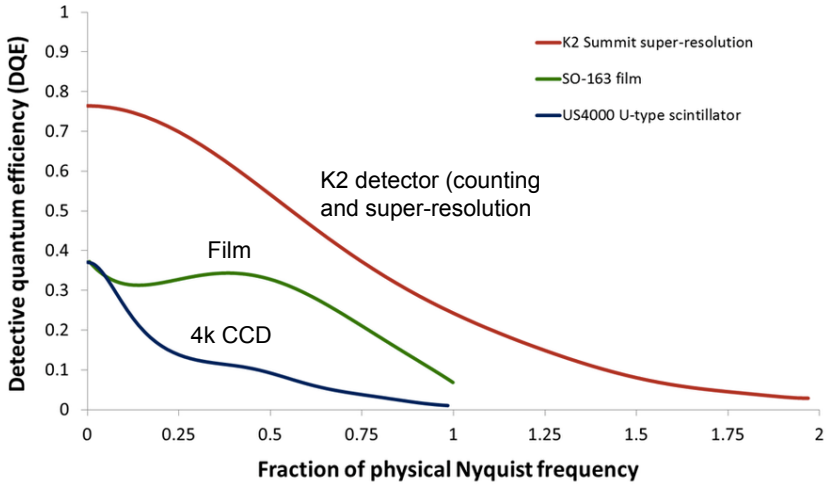
MTF: modulation transfer function gauges resolution and performance. It gives a measure of the contrast transmitted from the object to the image at a particular resolution and = **Image modulation/Object modulation**

The equivalent in cryo EM is the **Contrast Transfer Function**



The objective modifies the image by a **point spread function**

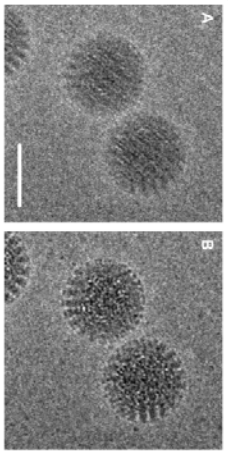
DQE = SNR_output² / SNR _input²



Gatan web site

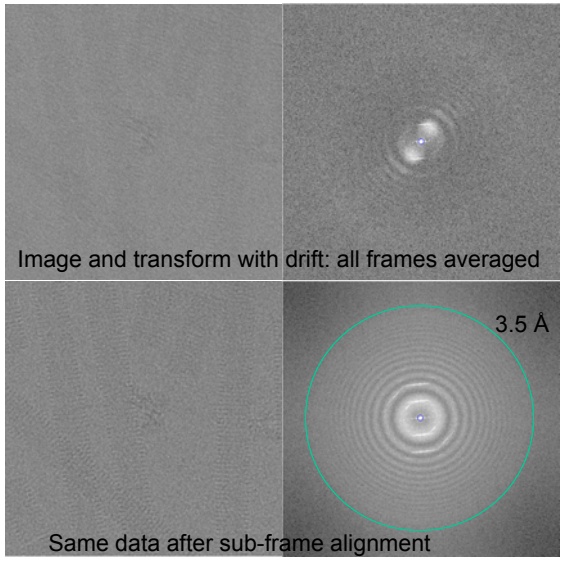
Fast frame rate of direct detectors enables motion correction

Motion correction on rotavirus



Brilot et al JSB (2012)

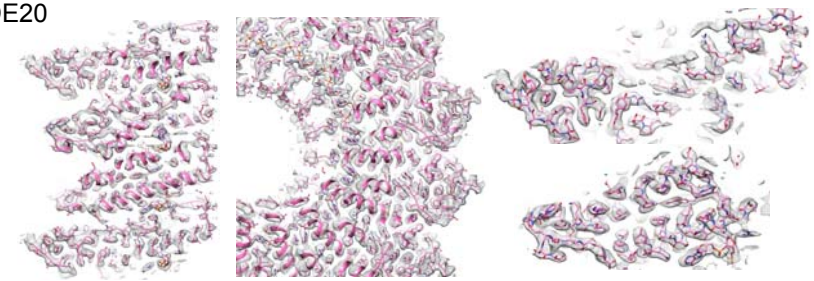
Direct electron detector: TMV on carbon film



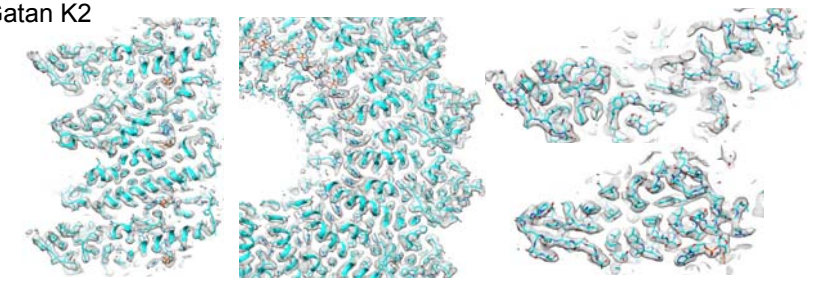
Dan Clare

TMV reconstructions with refined coordinates

DE20



Gatan K2



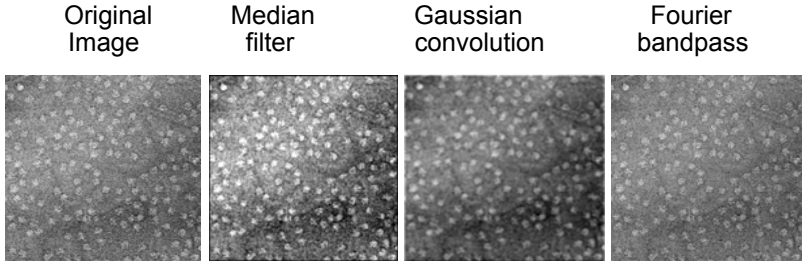
90% of TMV side chains defined; 3.3 Å resolution

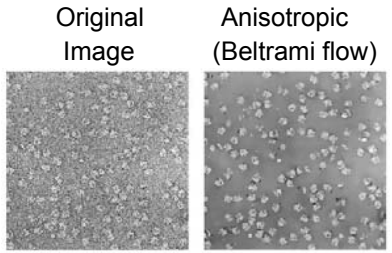
Dan Clare

Automatic particle detection methods

- Correlation-based template matching
- Edge detection
- Texture-based
- Intensity comparison
- Selection by classification

Noise reduction techniques

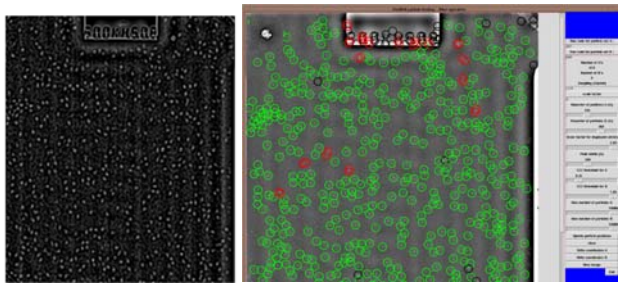




Correlation based template matching

Reduce noise by Gaussian convolution, anisotropic diffusion or Fourier filtering. Calculate correlation map between micrograph image and templates in various orientations. Find peaks, select particles from the unfiltered micrograph according to adjustable thresholds.

- Depends on quality of template(s)
- Sensitive to noise



Huang & Penczek(2004), Rath & Frank(2004), Roseman(2003), Sigworth(2004), Wong et al., Ludtke(1999)

Software packages for automatic particle detection

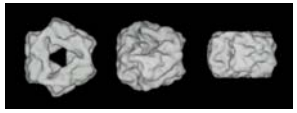
- Signature (Chen) Correlation/template matching.
- SwarmPS (Hankamer) Correlation / template matching or edge detection.
- EMAN boxer (Ludtke) Correlation/template matching.
- FindEM (Roseman) Correlation based, template matching.
- DoG picker/Tiltpicker (Carragher) Difference of Gaussians/thresholding.

http://en.wikibooks.org/wiki/Software_Tools_For_Molecular_Microscopy/Application_tools

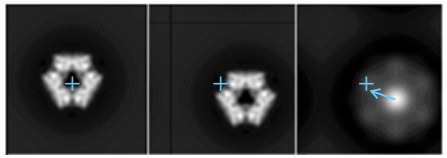
Comparative study: Zhu *et al* (2004)

Alignment and cross-correlation

1. Translational alignment to a reference



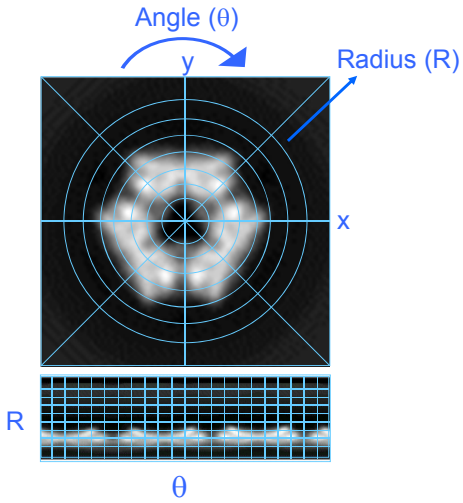
Small hsp with 32 symmetry



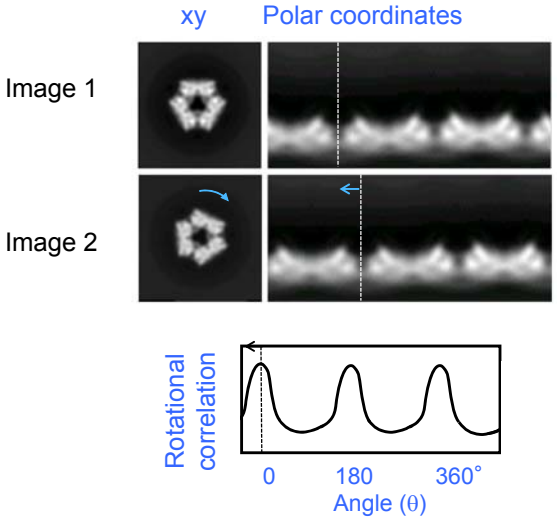
Shift of image 2 to match image 1

Image 1 Image 2 Cross-correlation

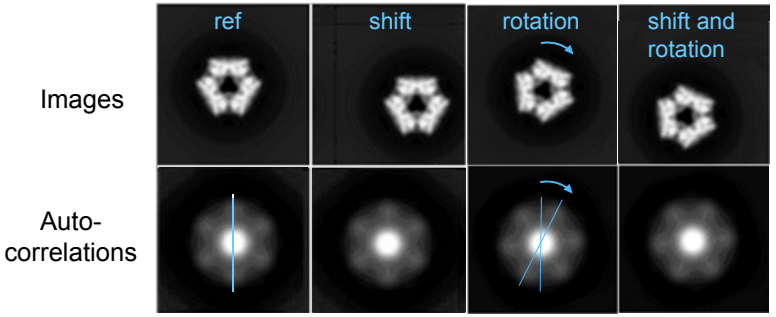
Polar coordinates



Rotational alignment to a reference

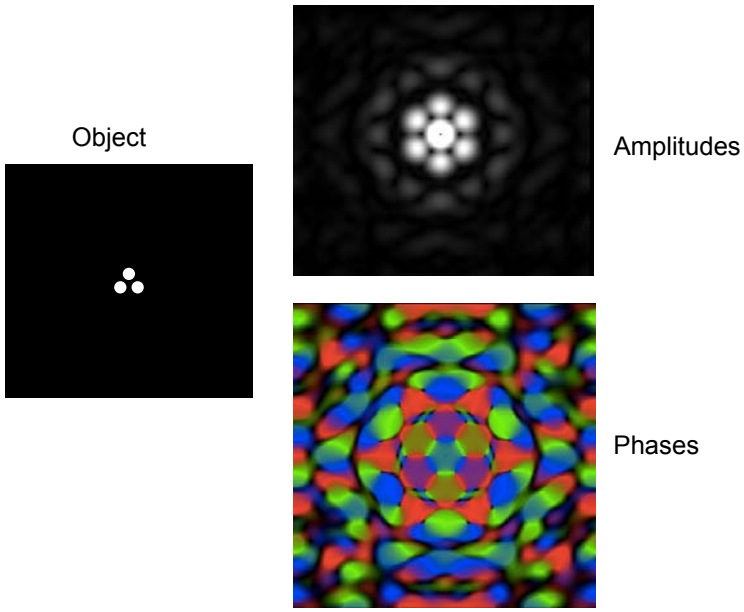


Autocorrelation



The autocorrelation function is the correlation of an image with itself. Since the 2 images being correlated are at the same position, the central peak is always at the origin and the autocorrelation does not vary when the object is shifted. Other features of the autocorrelation reflect features in the structure and rotate with the image. Note the additional 2-fold symmetry in the autocorrelation.

Phases reveal the true symmetry



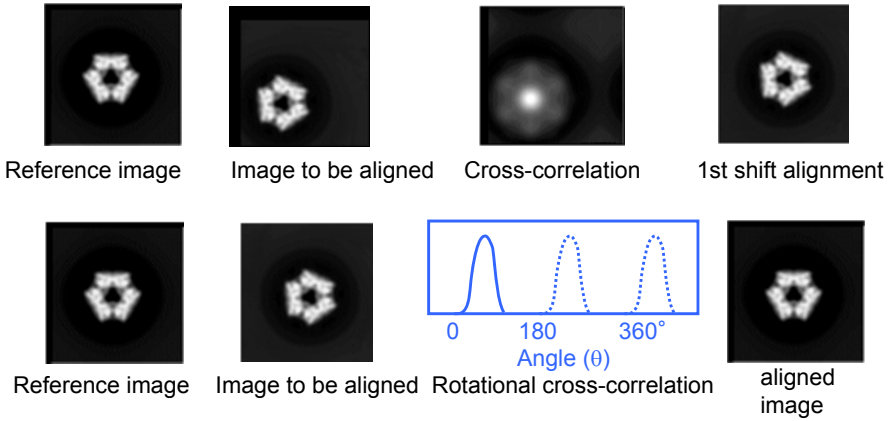
Autocorrelation of atoms

The **Patterson function** in crystallography is the autocorrelation function of an atomic structure. Each pair of atoms (similar density features) gives rise to a peak in the Patterson function, which is a collection of inter-atomic vectors. It is obtained by calculating the FT of the diffraction pattern. It does not contain any phase information, which is another way of saying that it does not change when the object is shifted.

For small molecules, the atomic positions can be deduced from the peaks in the Patterson function.

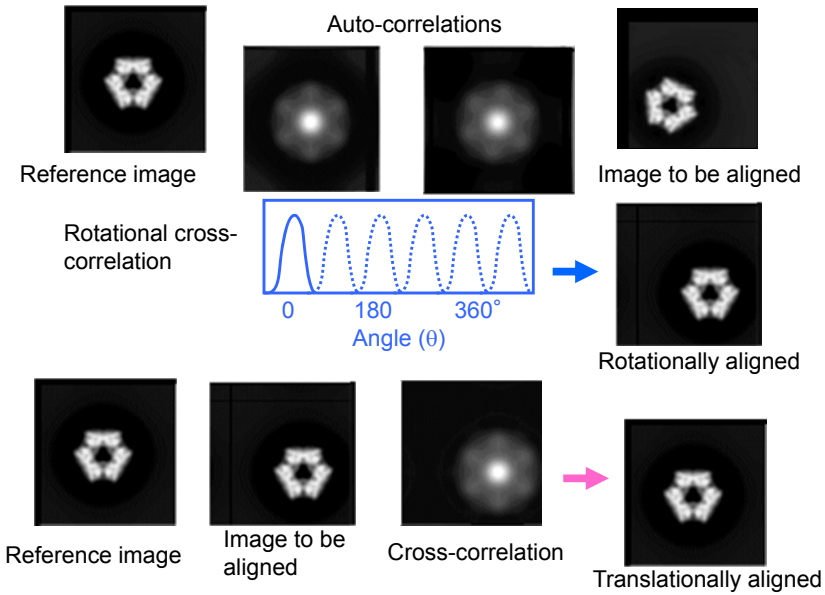
Alignment

Algorithm for alignment of individual images with respect to one reference image by an iterative sequence of translational and rotational alignments using cross-correlation functions. 3-5 iterations are sufficient to achieve good alignment.

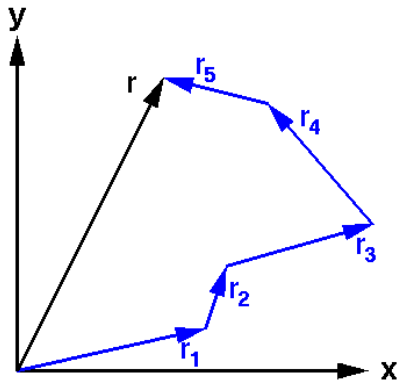


Steinkilberg, M. And Schramm, H.J. (1980) Eine verbesserte Drehkorrelationsmethode fur die Structurbestimmung biologischer Macromolecule durch Mittelung elektronen-mikroskopischer Bilder. Hopper-Seylers. Z. Physiol.Chem. 361, 1363-1369

2-step alignment using autocorrelations



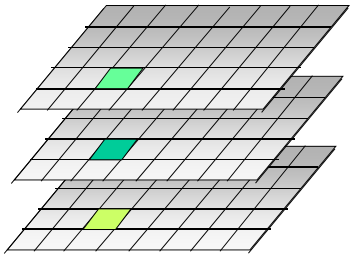
Vector addition of successive alignment parameters to avoid degradation by repeated interpolation of images



Alignment: How to assess it

- Clarity of averaged image
- Resemblance to raw data
- Improvement of correlation coefficients with successive iterations
- Low ***variance*** inside molecule projection
- Similar features after classification

Variance



For a stack of aligned images, the variance can be calculated for each pixel, to give a map of variations between the images in the data set. This can help to assess the reliability of features seen on the average image, and can reveal if images of different structures are mixed up in the same data set.

The variance is determined for each pixel as the difference between the pixel value in a given image and the average value of that pixel in all the images. This difference is squared and the sum of these squares is calculated for all the images in the stack.

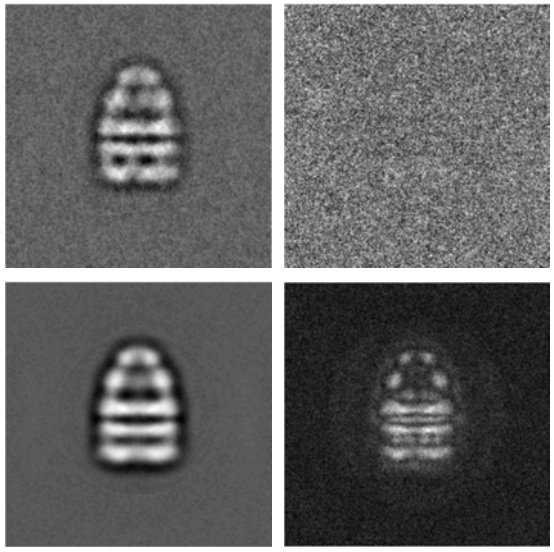
$$\text{Variance} = [1/(N-1)] \sum_{i,,j} [P_i(r_j) - P_{av}(r_j)]^2$$

where $P_i(r_j)$ is the value of pixel j in image i and $P_{av}(r_j)$ is the average value of pixel j in all the images, for a set of N images.

Average and variance of homogeneous and mixed data sets

Set of views with same orientation

Mixture of side views - different orientations around 7-fold (vertical) axis



average variance

Current issues in alignment and classification

• **Projection matching** by **cross correlation**: the conventional method is to simply assign the match according to the highest score, which is often not correct, due to noise or model limitations. **Maximum likelihood/Bayesian** methods, as in Relion and Frealign, include noise models and assign multiple matches with different weights.

• Projection matching can be done with class averages or single images, in real or Fourier space

• **Classification** to sort out heterogeneity: e.g. supervised classification (such as ribosome ± factors), separation by multivariate statistical analysis in 2D or 3D, 3D variance calculation by bootstrapping, multiparticle refinement in 3D

• **Model bias** – the starting model must be appropriately filtered to remove features that are not known to be represented in the data:

Einstein from noise – alignment of random noise to an image

References

Faruqi, A.R. & McMullan, G. (2011). Electron detectors for electron microscopy. Quart. Rev. Biophys **44**, 357-390.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Quart. Rev. Biophys. **28**, 171-193.

McMullan, G., Chen, S., Henderson, R.,Faruqi, A. R. , Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy* **109**, (2009) 1126-1143.

Li, X. et al (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods 10, 584-590.

Bammes, B. et al (2012) Direct electron detection yields cryo-EM reconstructions at resolutions beyond ¾ Nyquist frequency. J. Struct. Biol. 177, 589-601.

Roseman, A.M. (2004). FindEM - a fast, efficient program for automatic selection of particles from electron micrographs. *J. Struct. Biol.* **145**, 91-99.

Chen, J.Z. & Grigorieff, N. (2007). SIGNATURE : A single-particle selection system for molecular electron microscopy. *J. Struct. Biol.* **157**, 168-173.

Elmlund, D & Elmlund, H (2015) Cryogenic electron microscopy and single-particle analysis. Ann Rev Biochem 84, 499-517.

Bai, X, McMullan, G & Scheres, SHW (2015) How cryo-EM is revolutionizing structural biology. TIBS 40, 49-57.

Cheng, Y (2015) Single-particle cryo-EM at crystallographic resolution. Cell 161, 450-457.