2D Classification and Initial Structure Generation

Ottilie von Loeffelholz

Klaholz Group, IGBMC Strasbourg

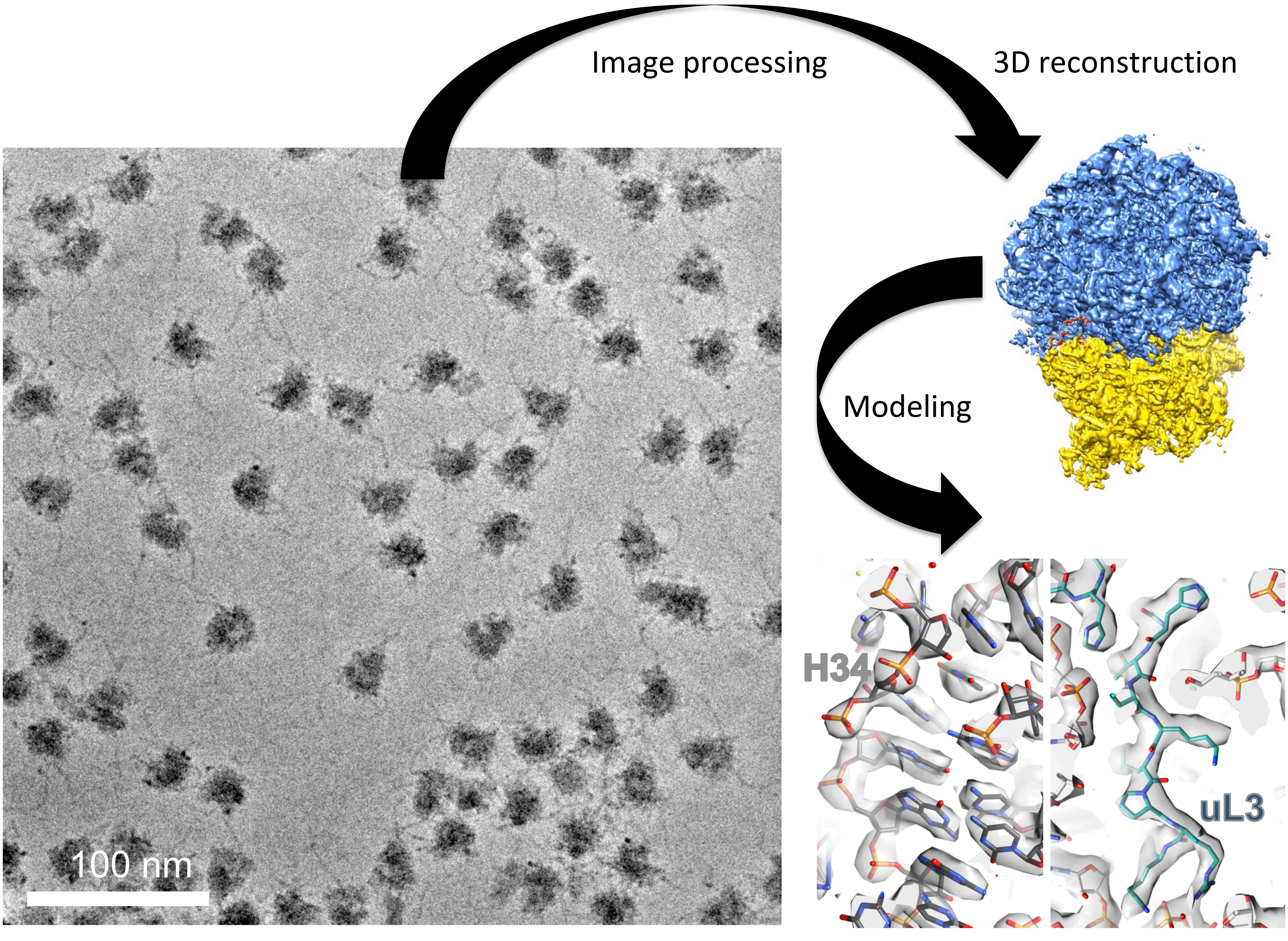


Image processing
3D reconstruction
Modeling
100 nm
H34
uL3

Image processing pipeline

Data collection -> Movies

- Movie alignment
- CTF estimation
- Particles boxing, filtering, normalisation

Structure determination

- Alignment
- Classification
- Initial structure generation (angles assignment)
- 3D reconstruction
- 3D Classifiaction
- resolution assessment
- Map interpretation/Atomic model building

Getting contrast from noisy images

- The ideal case:

  Image = Signal

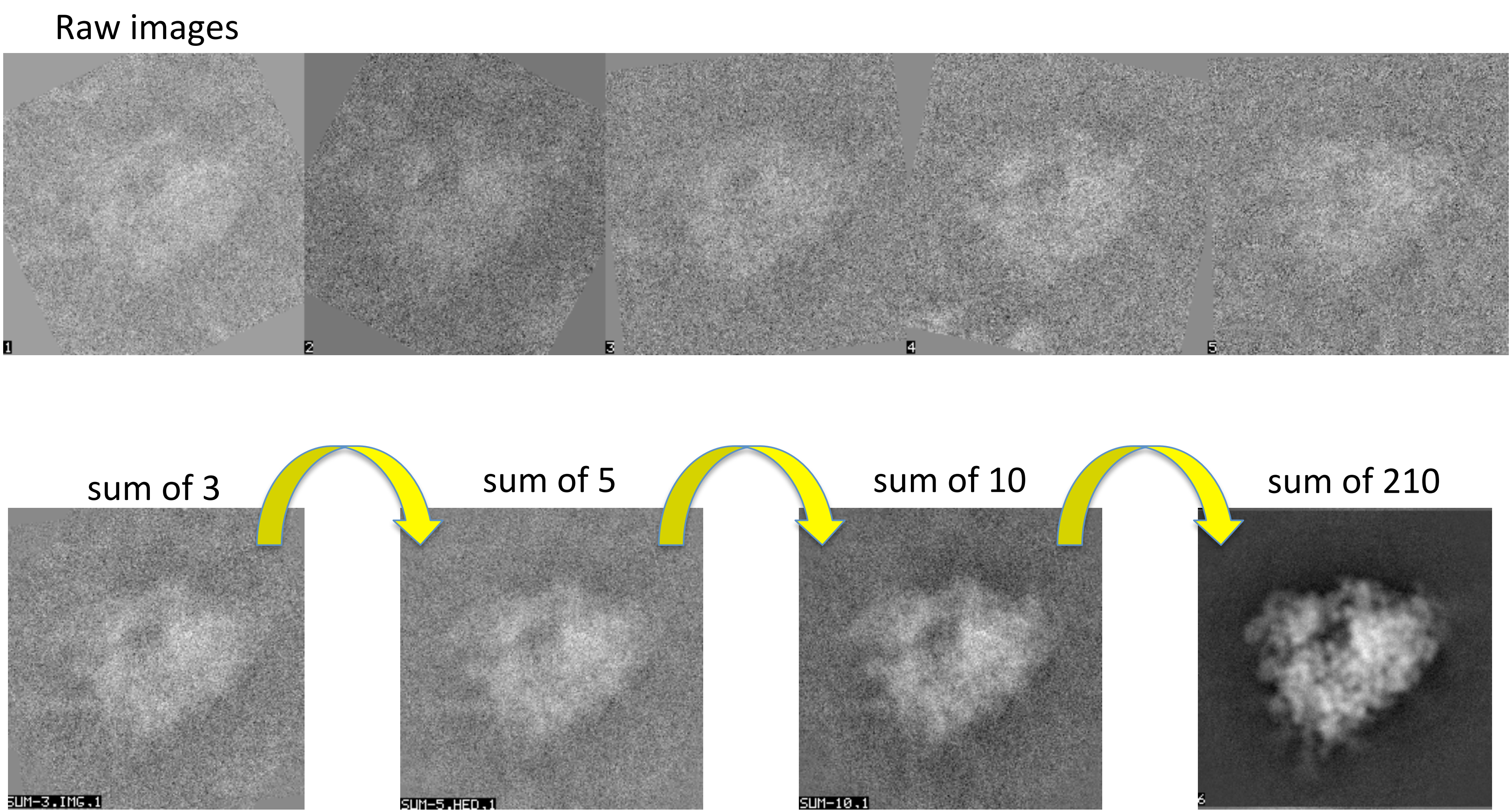
- The reality of cryo-EM:

  Image = Signal + Noise

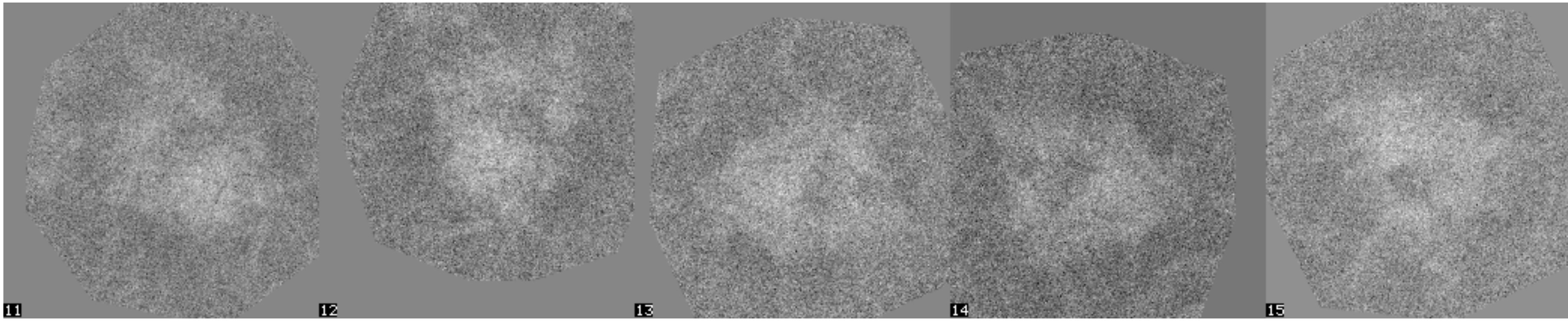
Projection of a 3D object (interaction of the beam with the sample, elastic scattering)

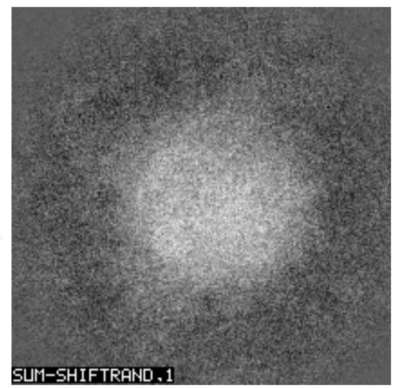
Low dose (inelastic scattering = radiation damage), ice thickness, carbon support, no interaction with the sample

Averaging to increase image contrast

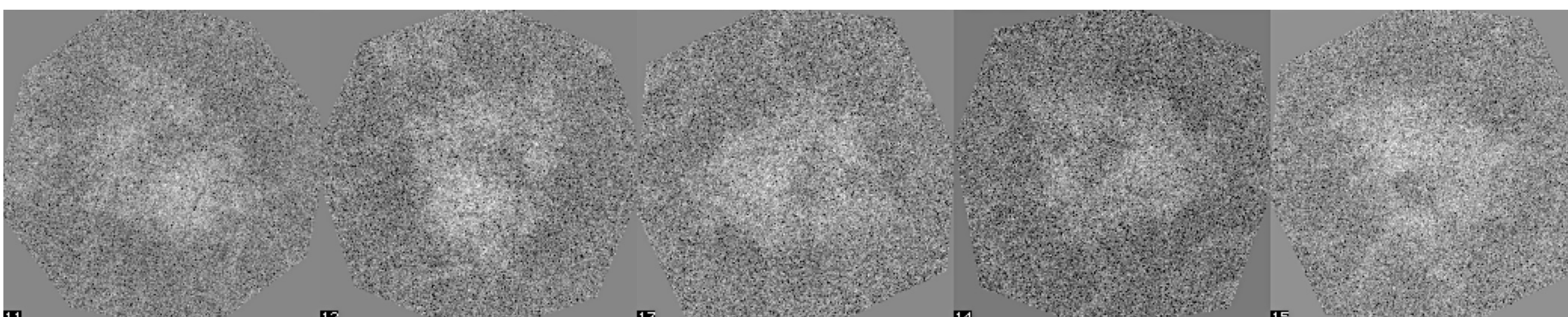


Images centering

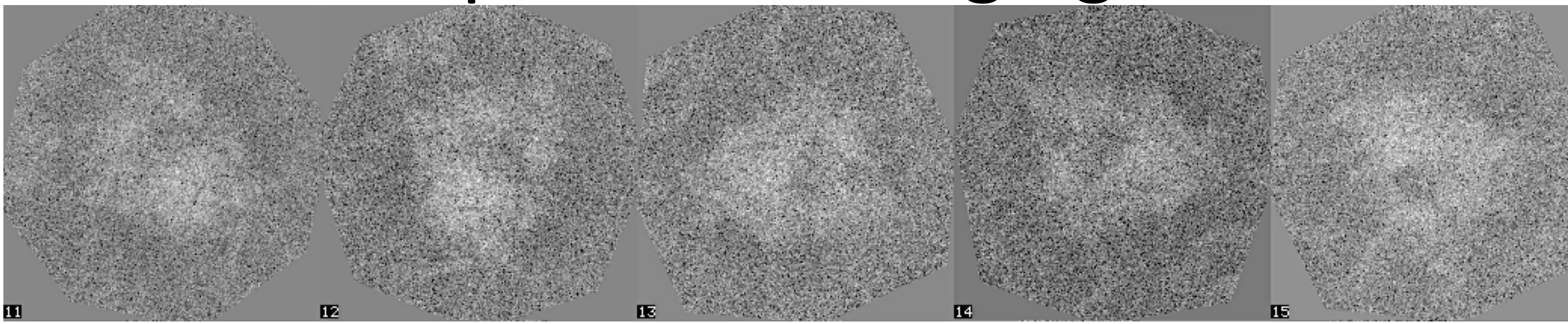


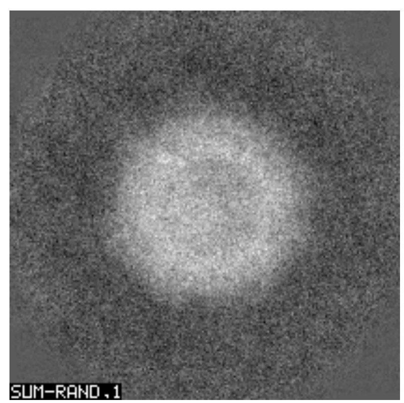


Alignment using a rotational average as reference



Images need to be rotationally aligned prior to averaging

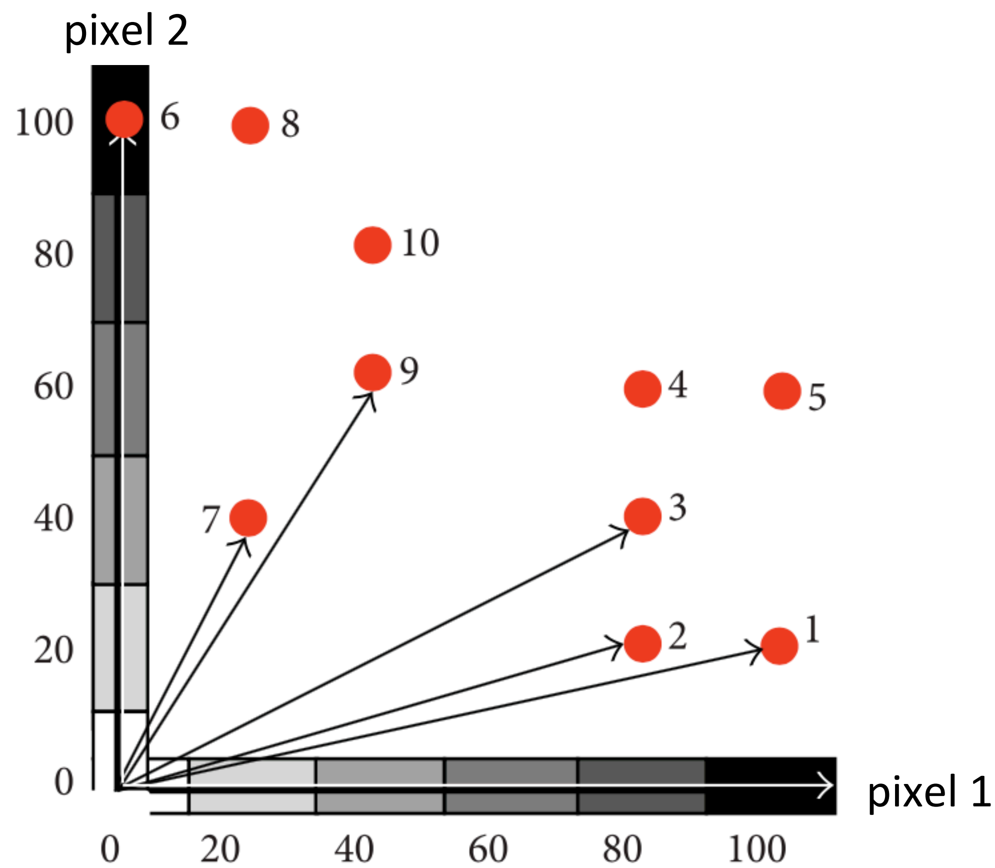




Rotational average
of centered particles

Calculation with images

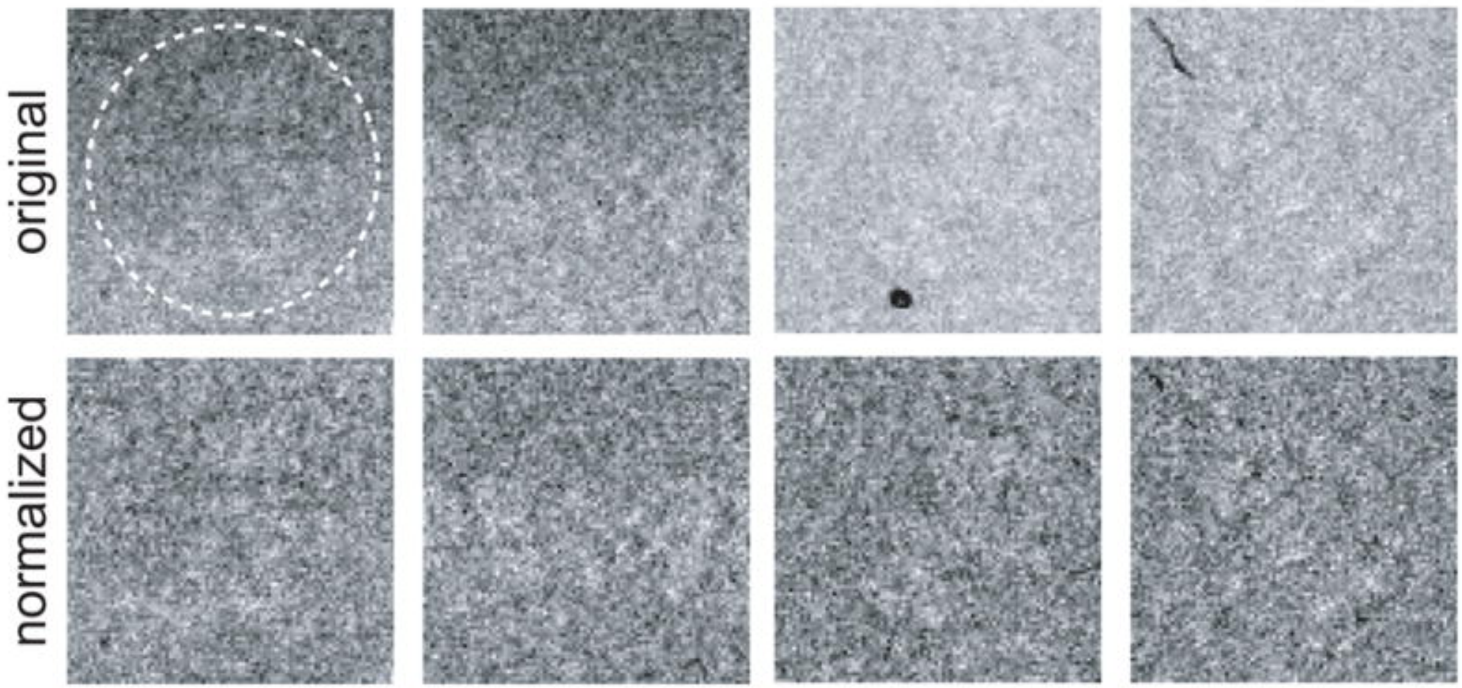
Simplified Image	pixel 1	pixel2
#1	100	20
#2	80	20
#3	80	40
#4	80	60
#5	100	60
#6	0	100
#7	20	40
#8	20	100
#9	40	60
#10	40	80



Frank and van Heel, JMB, 1982; van Heel, Ultramicroscopy, 1981;
White et al., 2017, BioMed Research International

Normalization

Normalisation: bringing all particles to similar gray value distribution. Therefore comparison between the particles is more robust



Two possibilities:

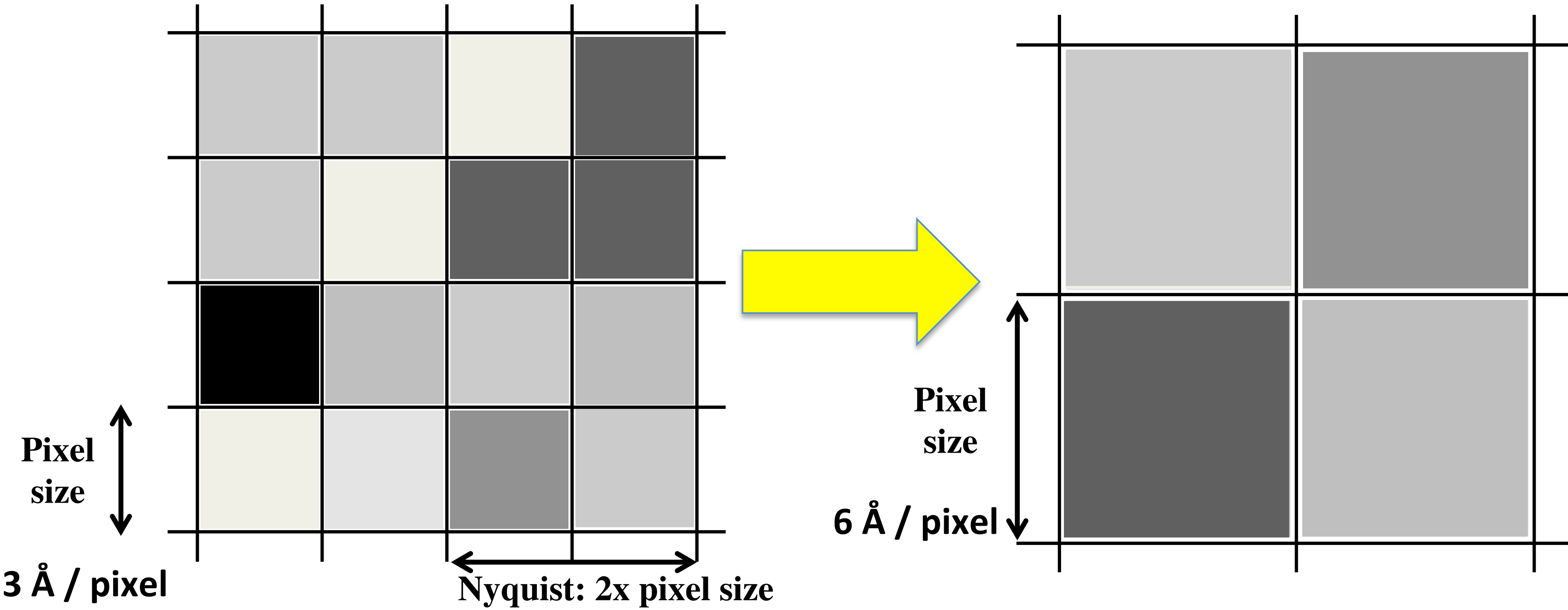
1: set mean of all images to 0 and normalize all pixel values to a given standard deviation

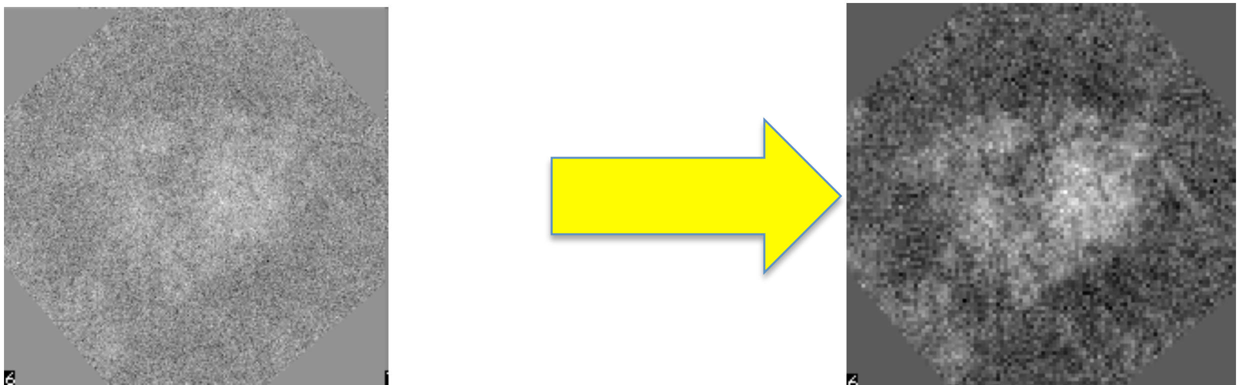
2: use pixel outside a set radius in each particles to calculate their mean and standard deviation value that is then applied on the particle

Scheres, 2010, Methods in Enzymology
Frank and van Heel, JMB, 1982; van Heel, Ultramicroscopy, 1981;

Down-sampling

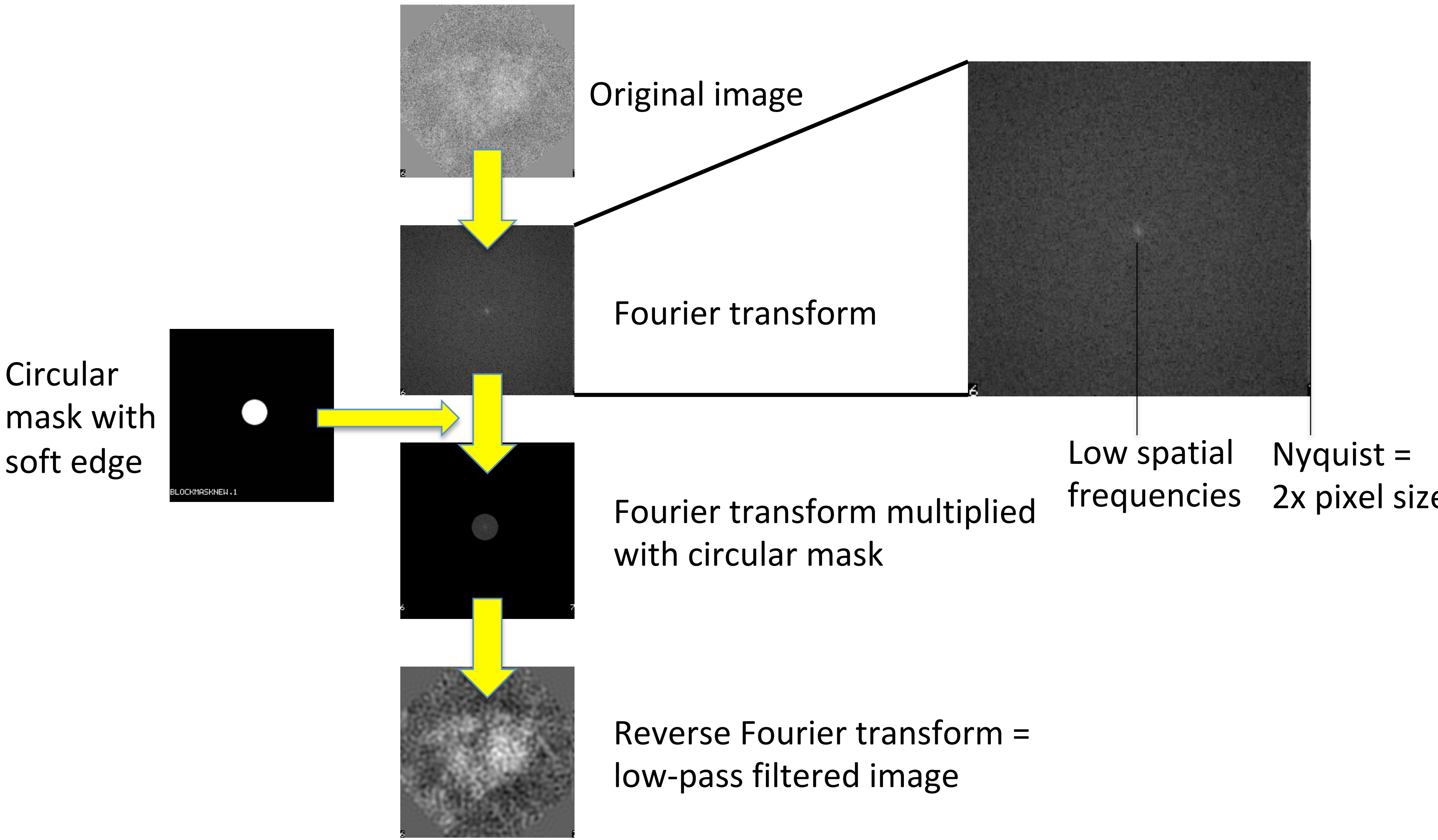
Grey values in pixels:





Increased contrast but resolution is lost

Low-pass filtering

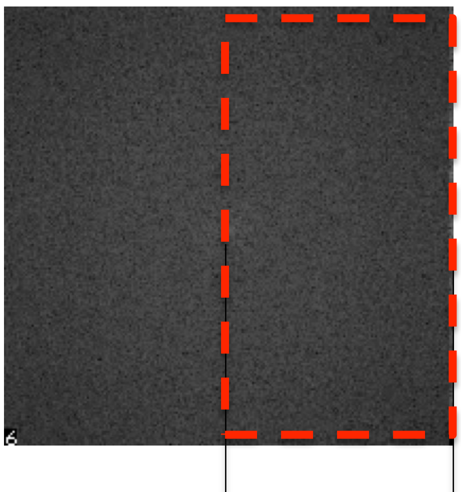


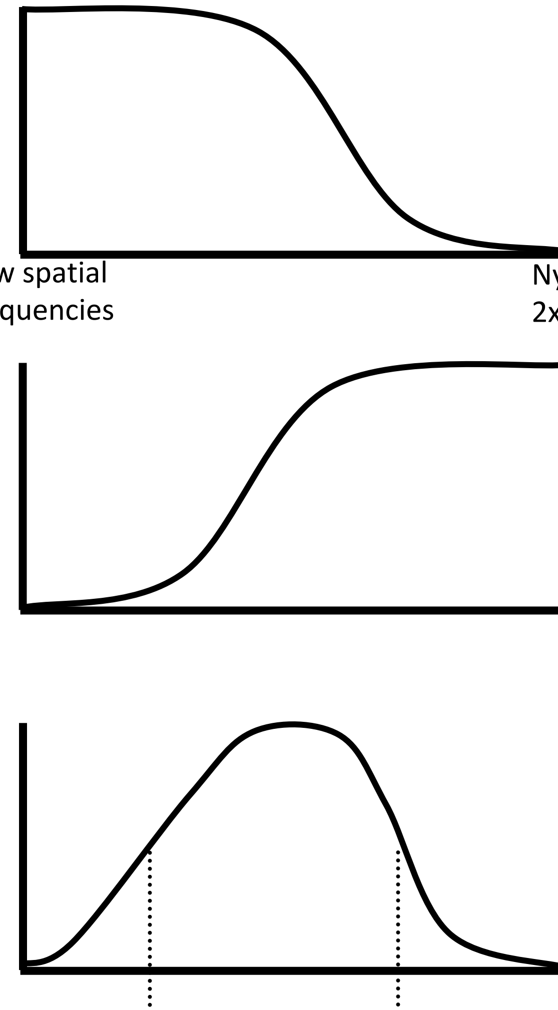
I. Pre-processing

- band-pass filtering and normalisation of particle images

Combination of high-pass and low-pass filters:

Fourier transform



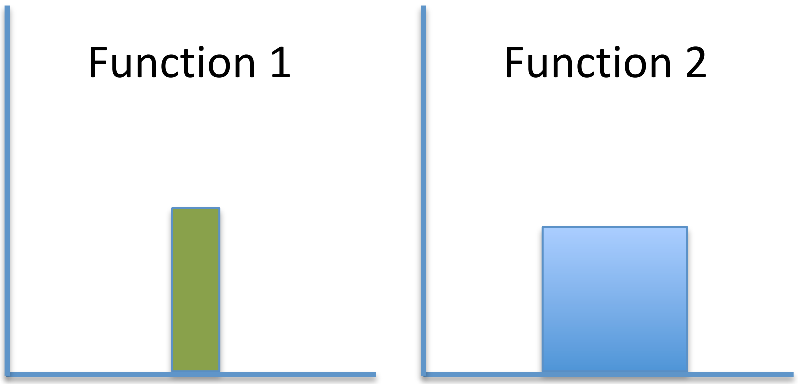


Removes:

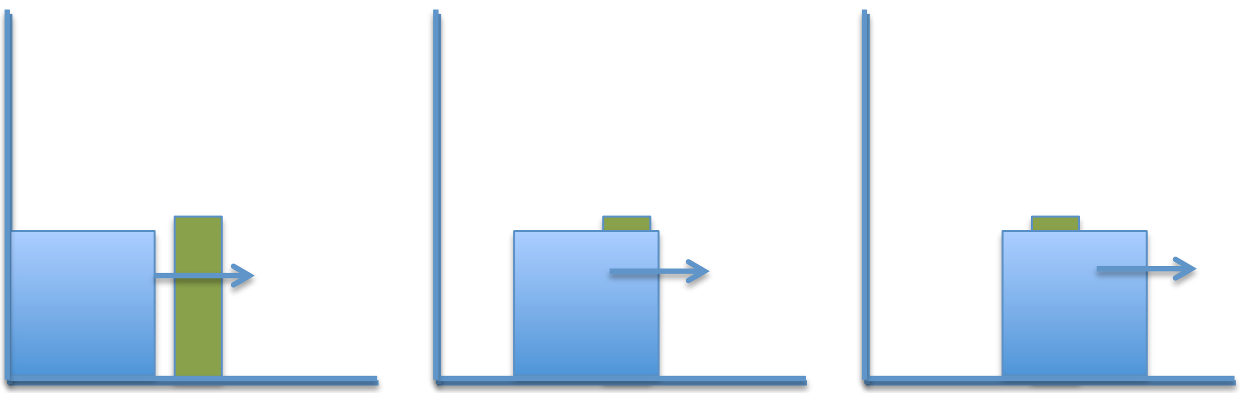
- low frequency contribution (scanner, etc.)

- high frequency noise

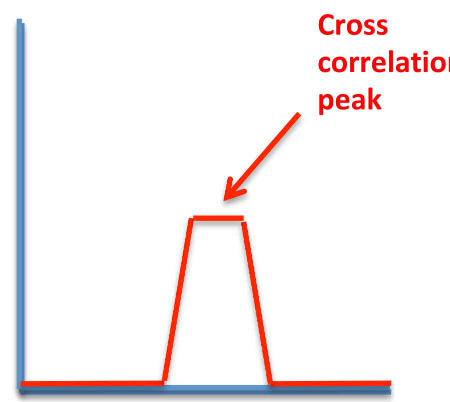
Aligning Images using Cross Correlation Function



Moving Function 2 to fit Function 1



Cross correlation



$$\text{CCF}(x') = \frac{\int Function1(x) \cdot \; Function2\,(x + x')dx}{\sqrt{\int \; Function1(x)^2 dx \cdot \int (Function2(x))^2 dx}}$$

Cross Correlation Peak

Image 1 (reference)

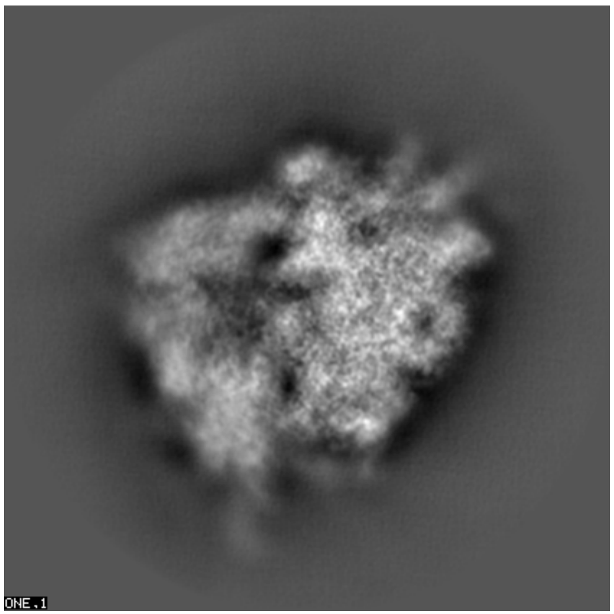
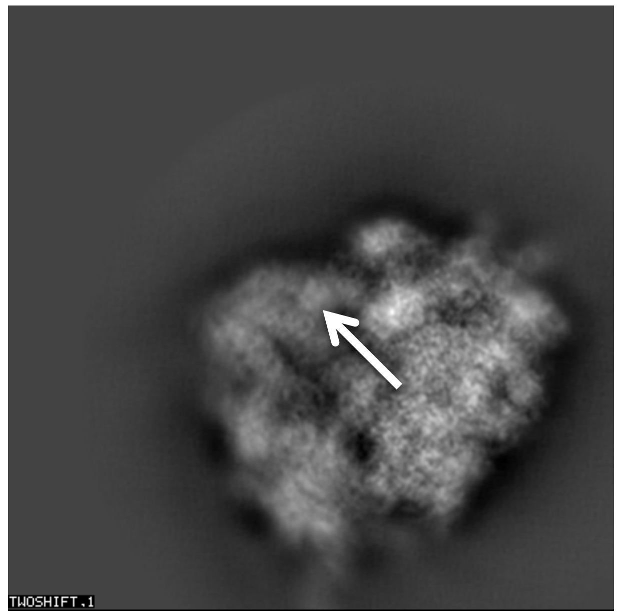
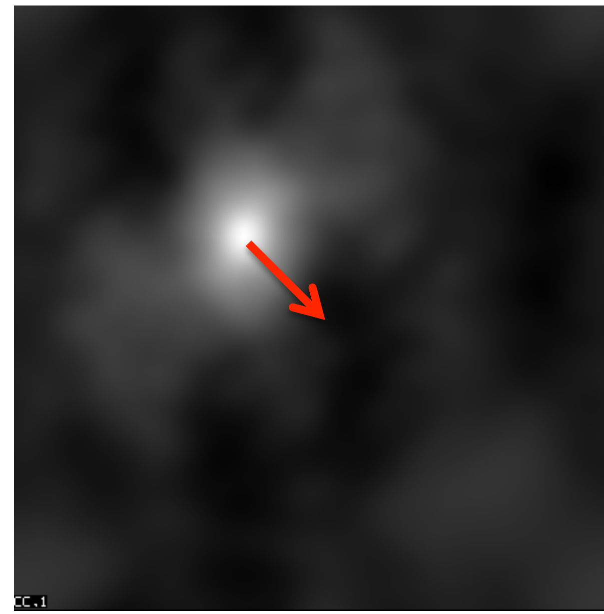


Image 2 (with shift)



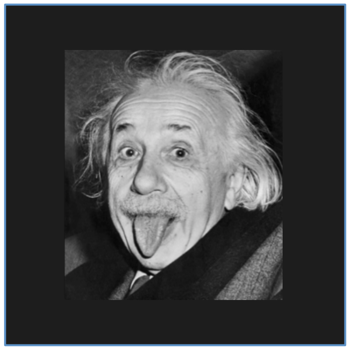
Cross correlation peak (shifted)



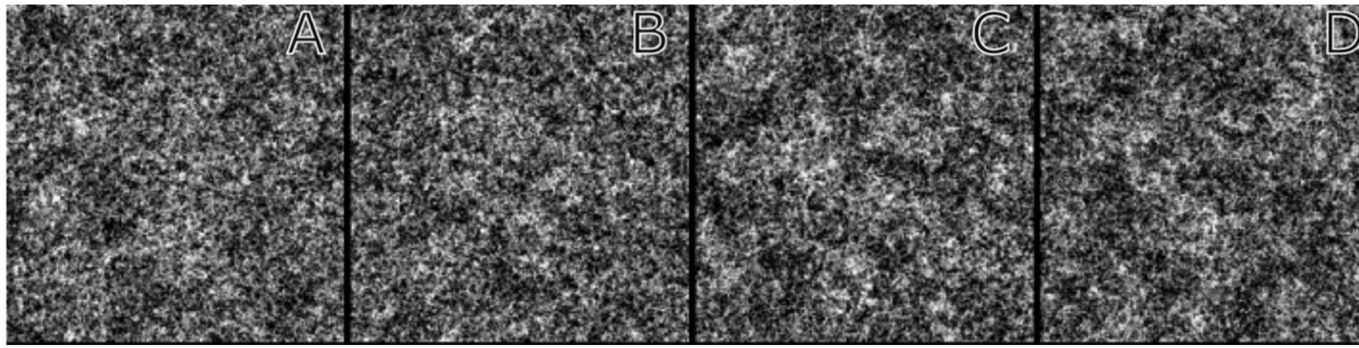
showing how to shift image 2 to match reference best

Reference-based Alignment: Model Bias: “Einstein from Noise”

Reference:



Noisy data:



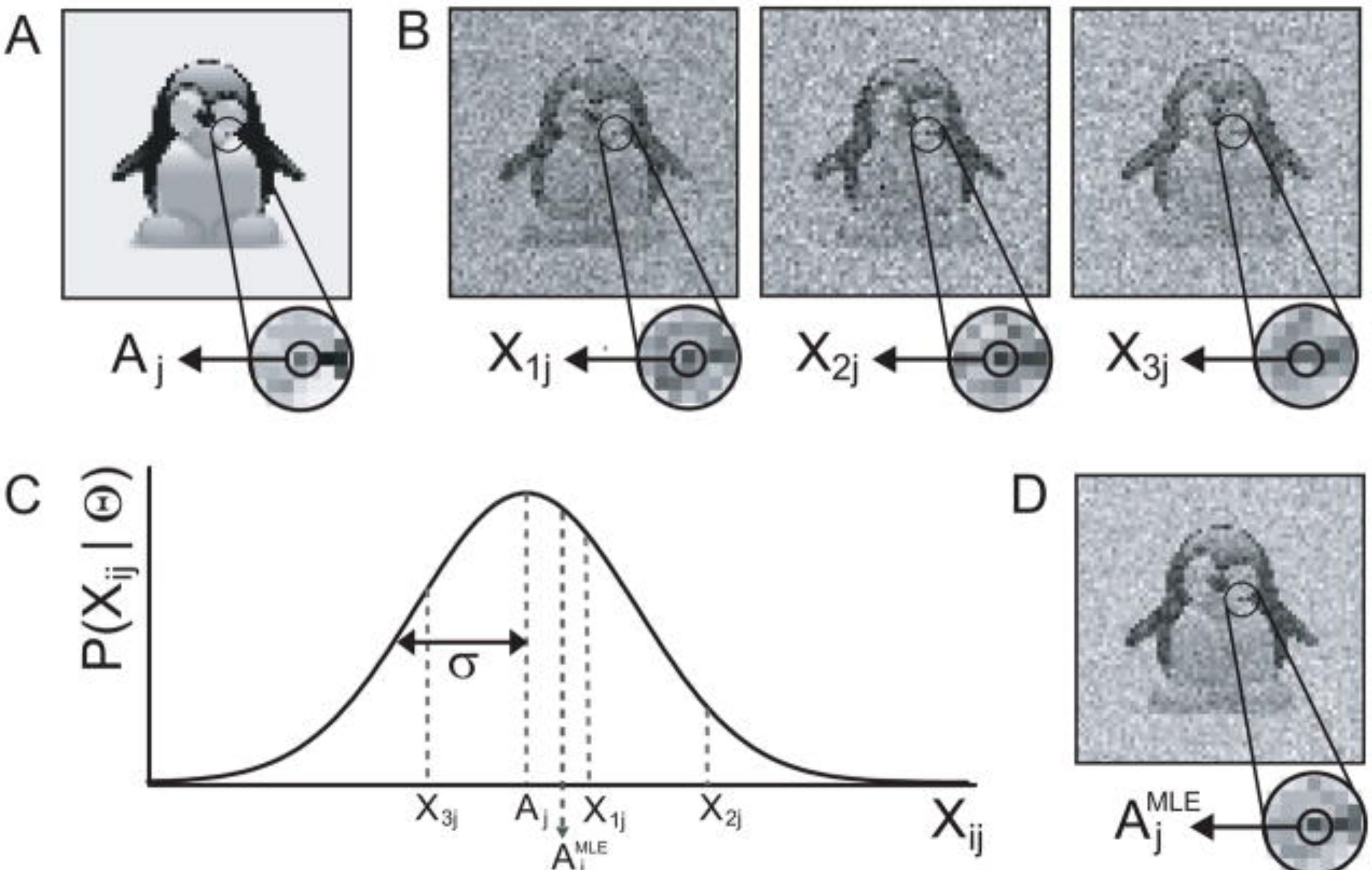
Cross correlation:

Mutual information:

“One can find anything one wishes to find in random noise!”

Shatsky et al., 2009, JSB; Henderson, PNAS, 2013; van Heel, PNAS, 2013; Subramaniam, PNAS, 2013

Maximum Likelihood



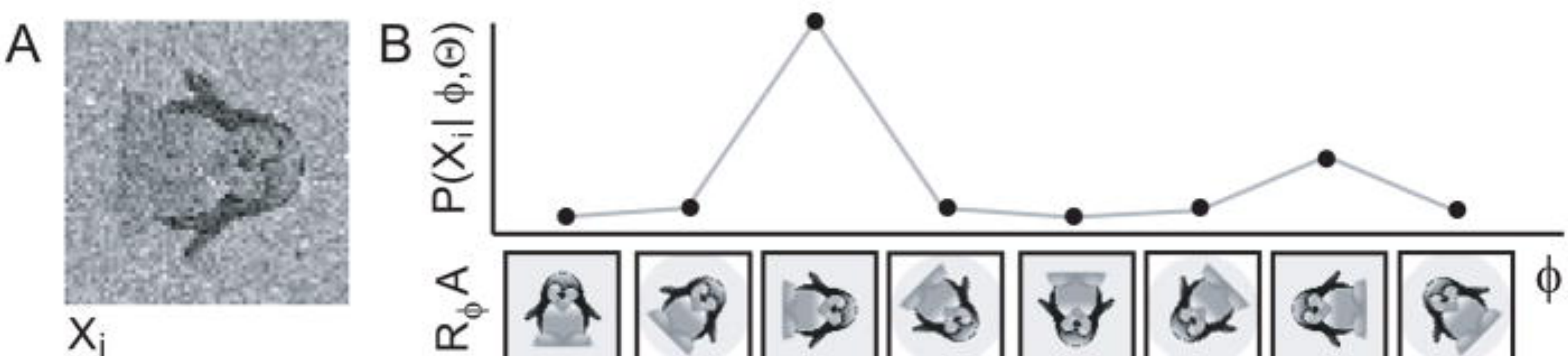
The noise has Gaussian distribution

For a pixel (J) with added noise the most likely correct gray value is A_j

Sigworth et al., 2010, Methods in Enzymology
Sigworth et al., 1998, JSB
Scheres, 2006.

The incomplete data problem in EM: Probability distribution function

How do I need to align my image to get a meaningful average?



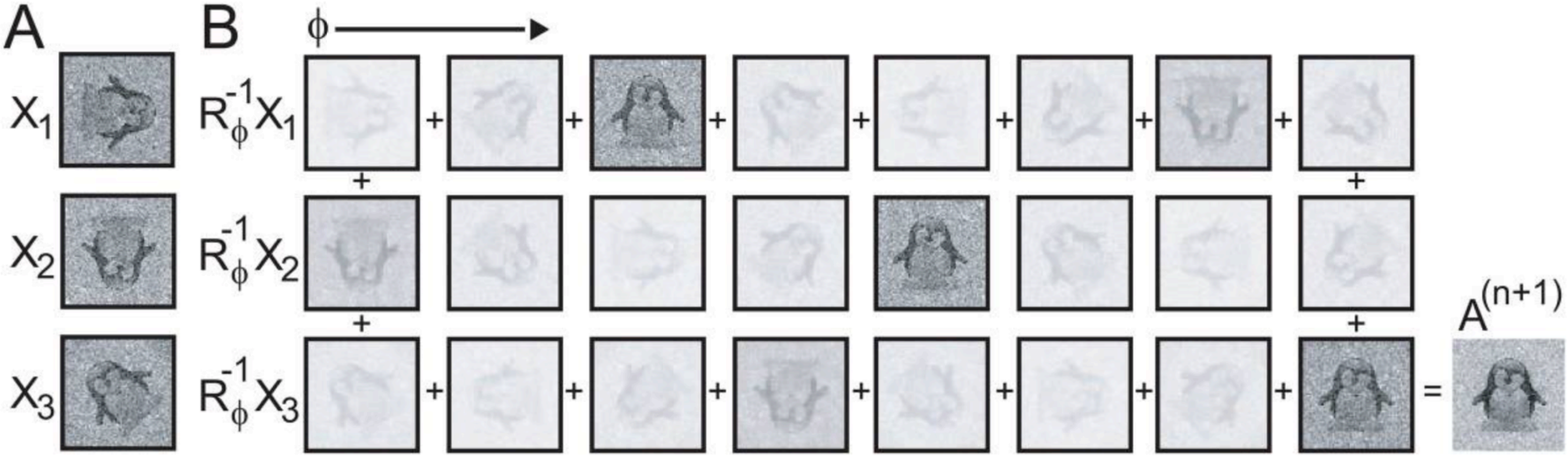
ϕ: the displacement

X: your cryo-EM image

Θ: The current model of your cryo-EM image

Sigworth et al., 2010, Methods in Enzymology

Reference image calculation by probability-weighted averaging

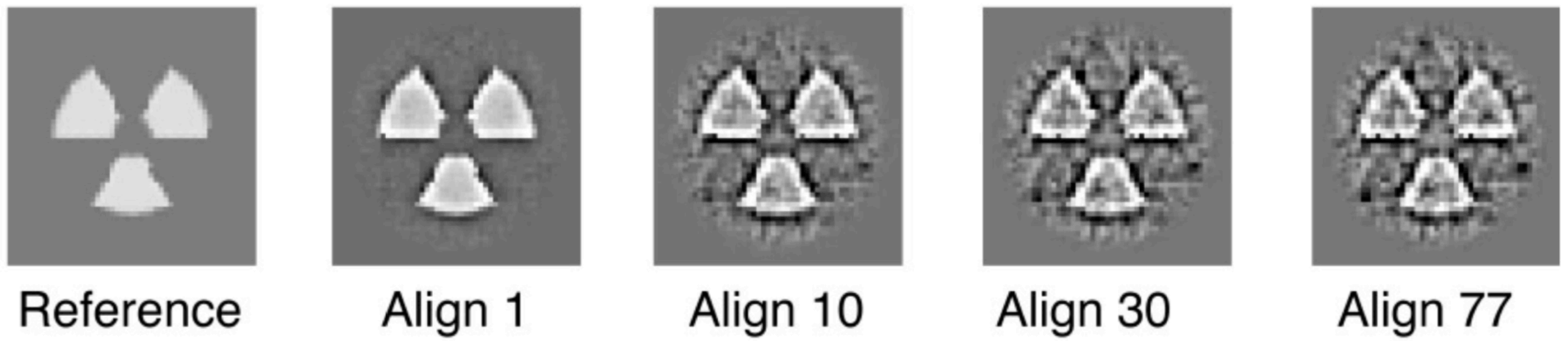


The less likely orientations are considered, but down-weighted

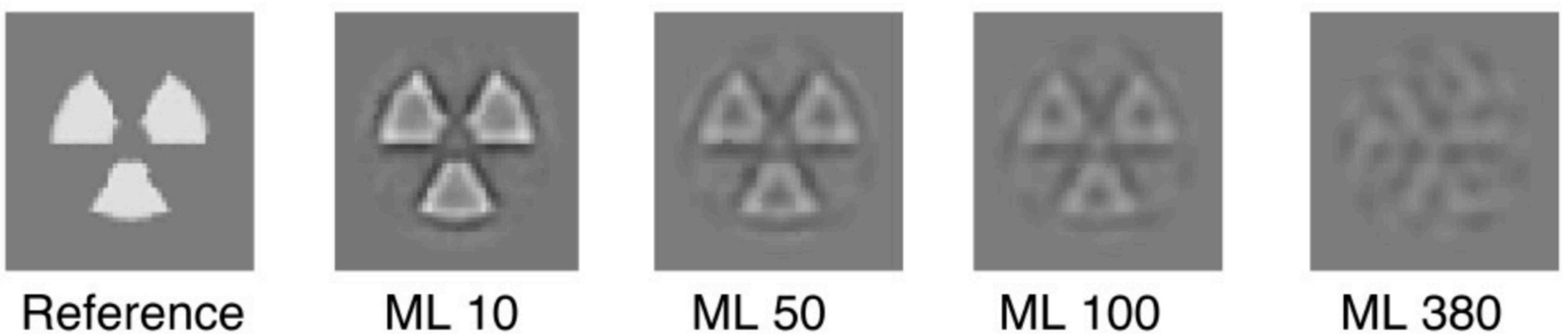
Sigworth et al., 2010, Methods in Enzymology

Less model bias with Maximum Likelihood

Cross correlation:



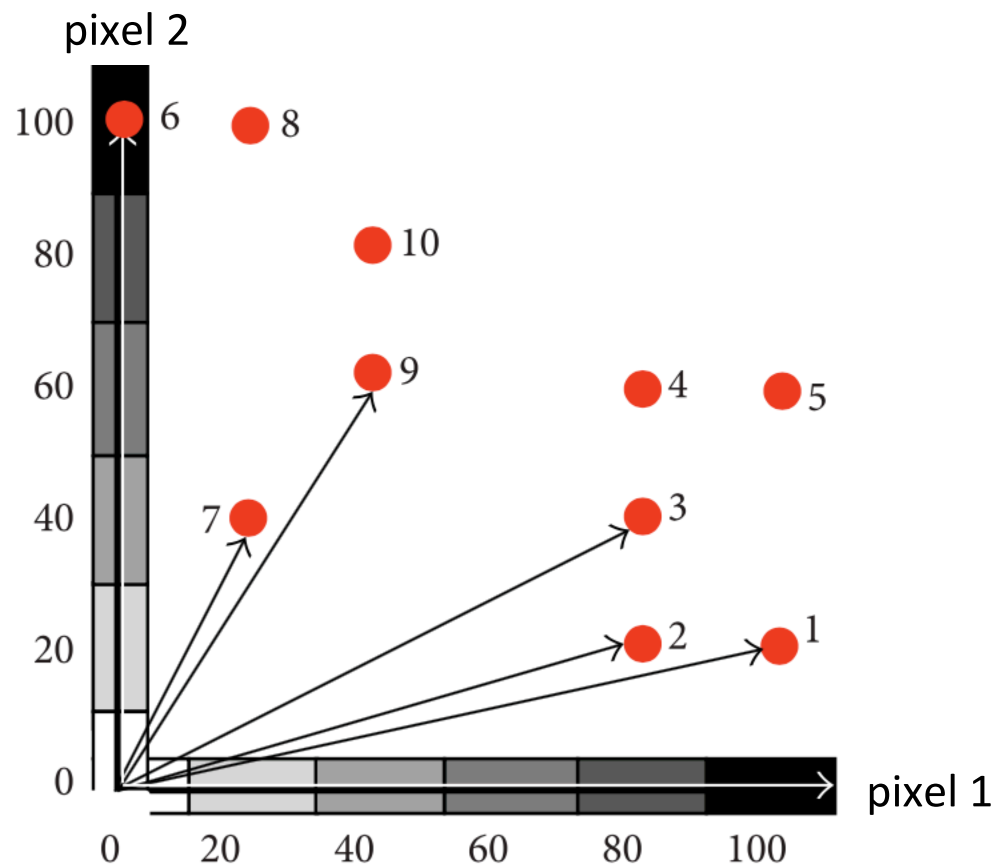
Maximum Likelihood:



Sigworth, 1998, JSB

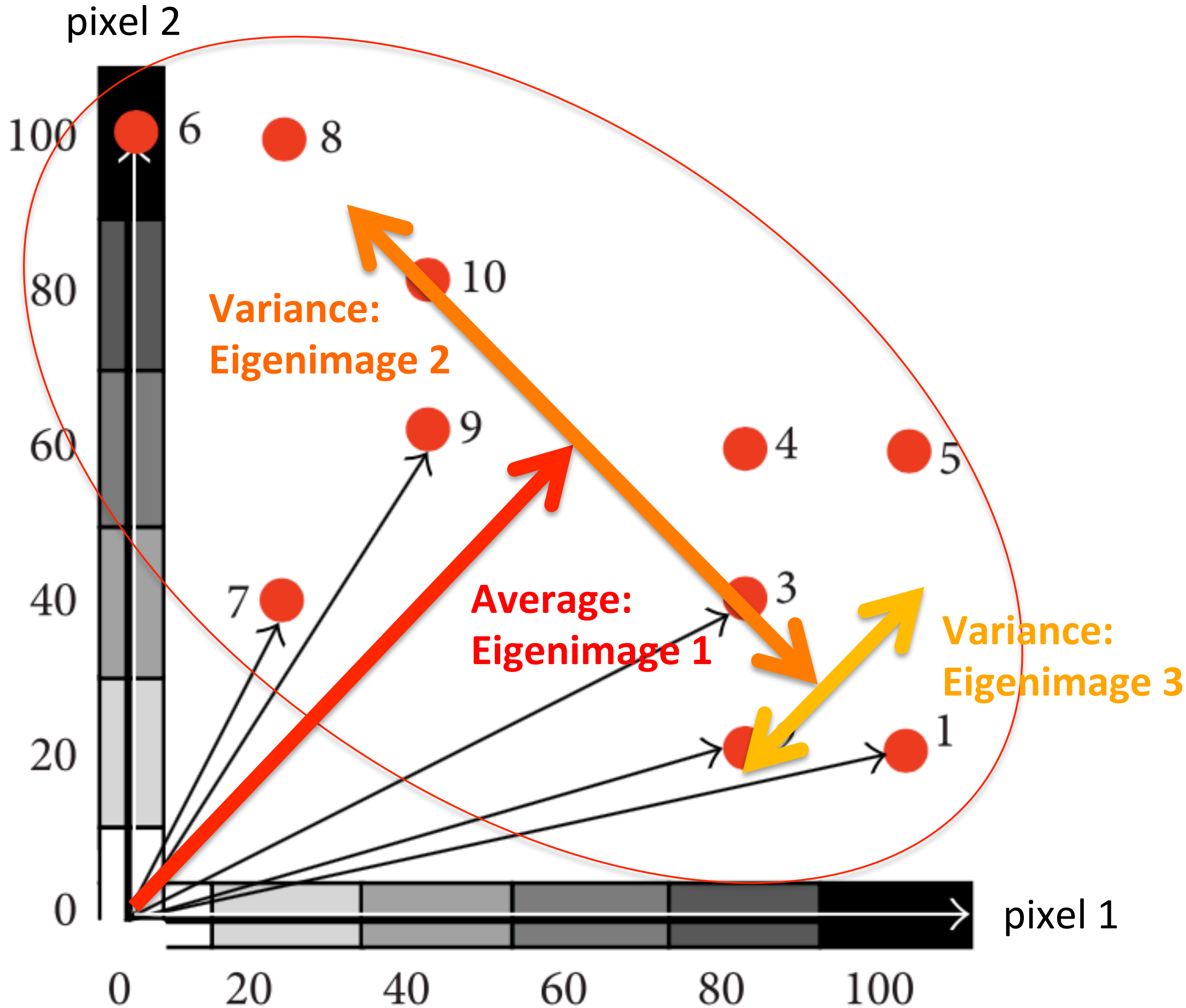
Classification of Images

Image	pixel 1	pixel2
#1	100	20
#2	80	20
#3	80	40
#4	80	60
#5	100	60
#6	0	100
#7	20	40
#8	20	100
#9	40	60
#10	40	80



Frank and van Heel, JMB, 1982; van Heel, Ultramicroscopy, 1981; White et al., 2017, BioMed Research International

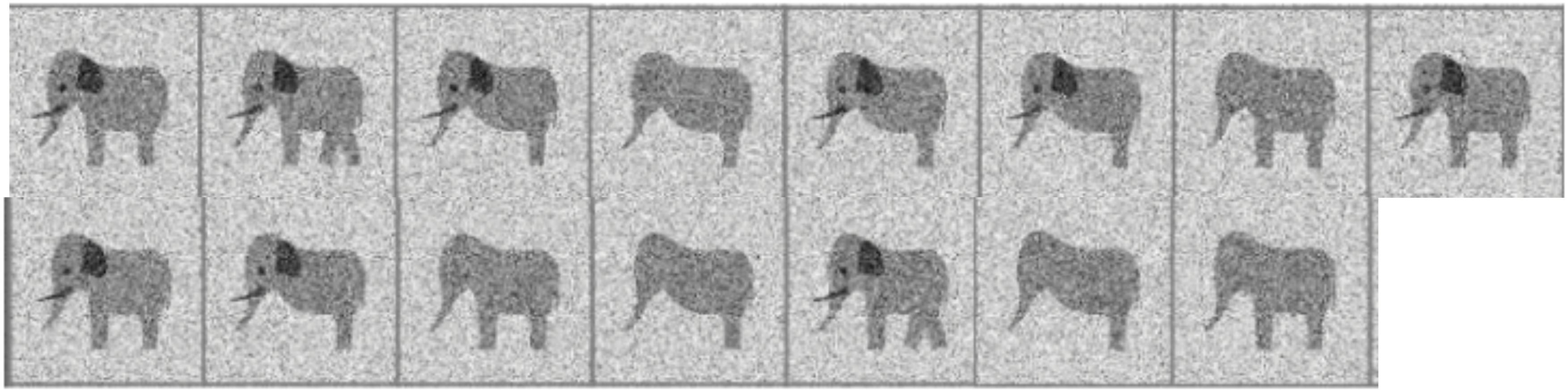
2D classification using Multivariate statistical analysis (MSA)



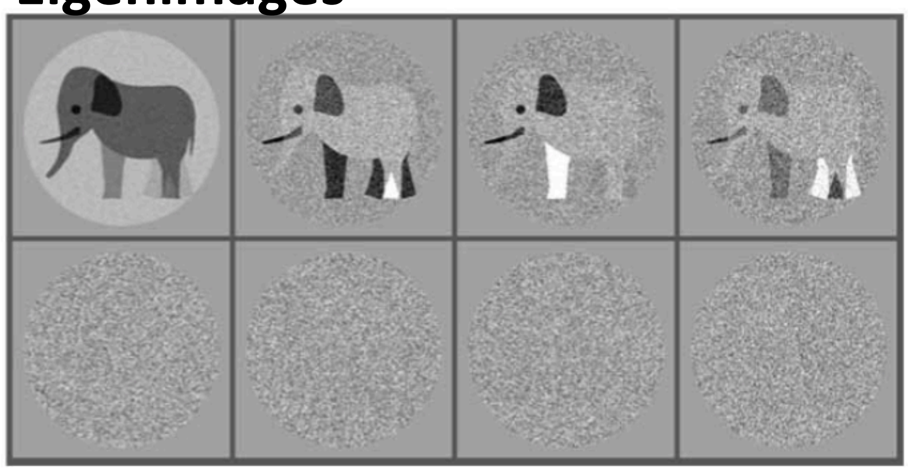
Frank and van Heel, JMB, 1982; van Heel, Ultramicroscopy, 1981; White et al., 2017, BioMed Research International

Example for Eigenimages

Dataset

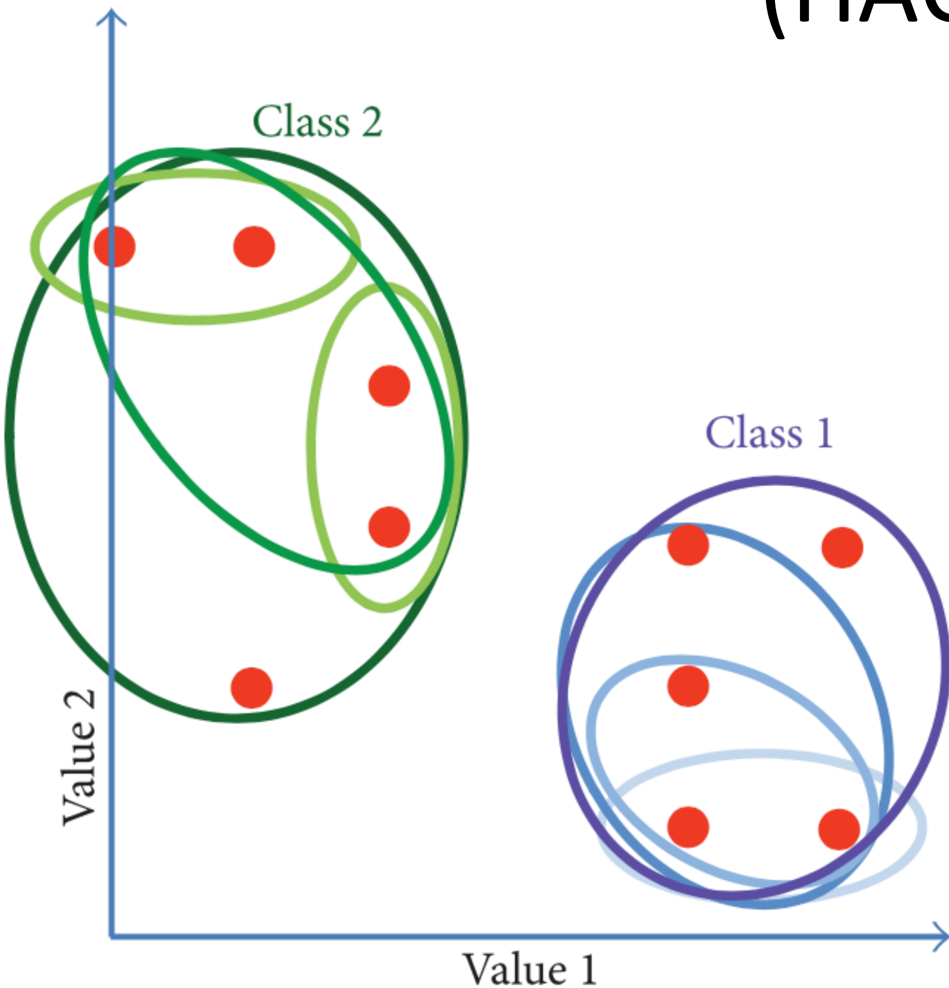


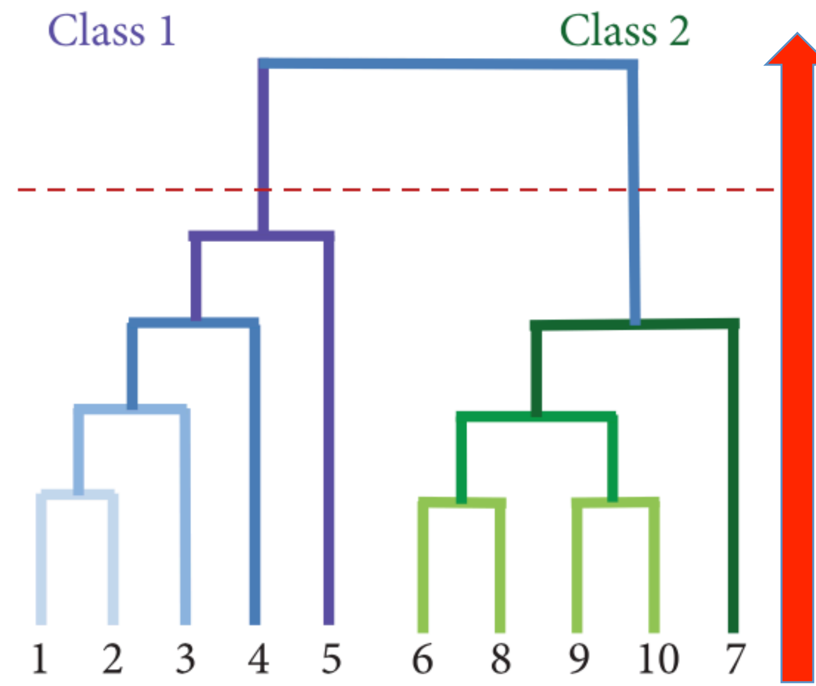
Eigenimages



White et al., 2017, BioMed Research International

Hierarchical ascendant classification (HAC)



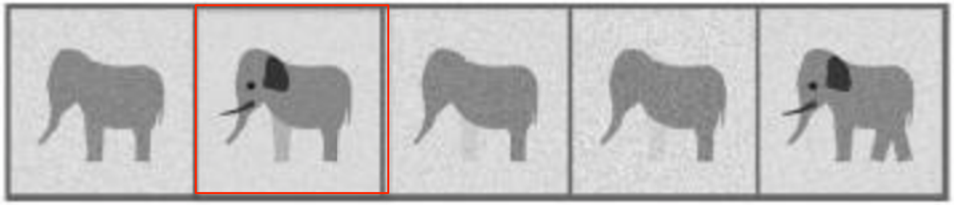


Each image is considered to be a class. The "classes" are merged in respect to their variance/eigenimage until a chosen cutoff

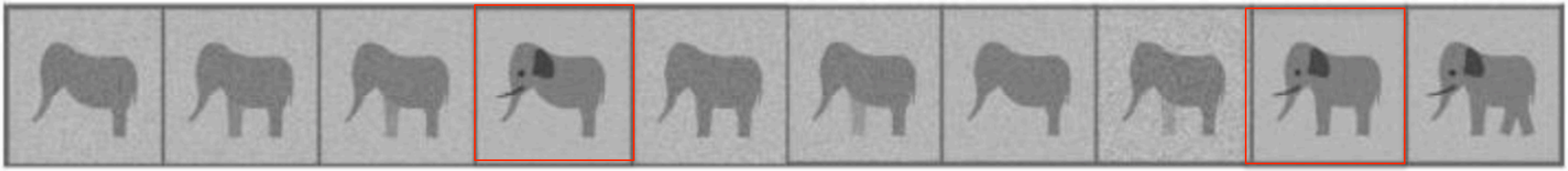
van Heel, Open Journal of Statistics, 2016
Ward, J Amer. Statist. Assoc., 1982 ; White et al., 2017, BioMed Research International

Classification using Eigenimages

Classification into 5 classes



Classification into 10 classes

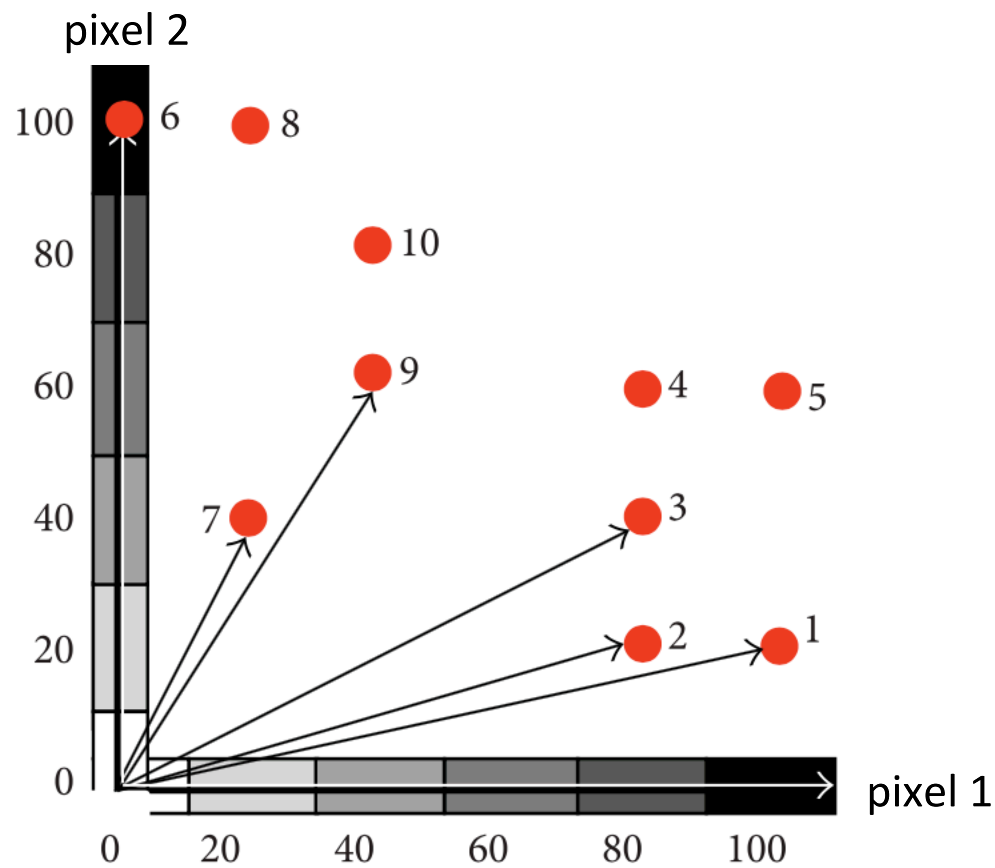


The class number must be large enough to be able to represent all heterogeneity in a dataset

White et al., 2017, BioMed Research International

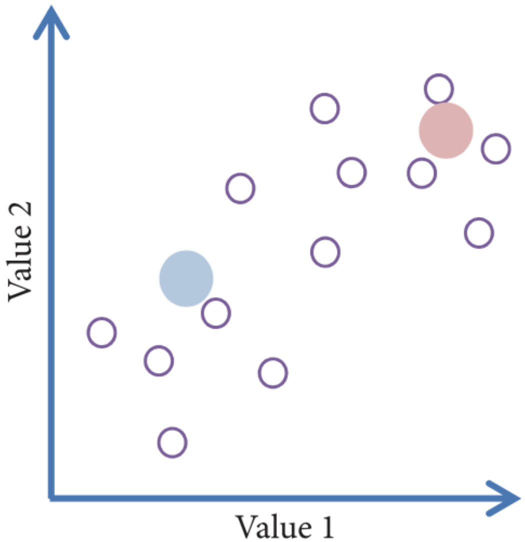
K-means clustering

Image	pixel 1	pixel2
#1	100	20
#2	80	20
#3	80	40
#4	80	60
#5	100	60
#6	0	100
#7	20	40
#8	20	100
#9	40	60
#10	40	80

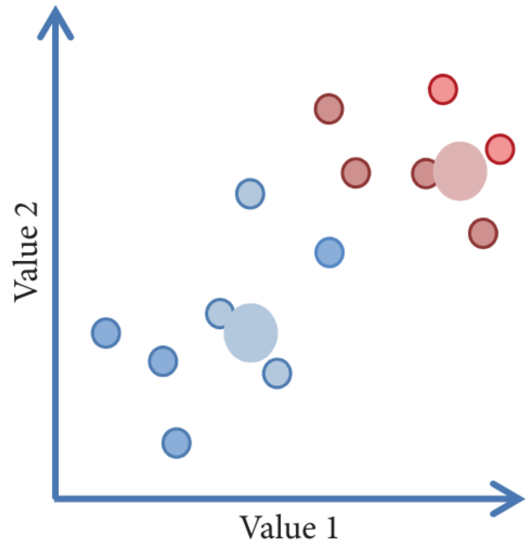


Frank and van Heel, JMB, 1982; van Heel, Ultramicroscopy, 1981; White et al., 2017, BioMed Research International

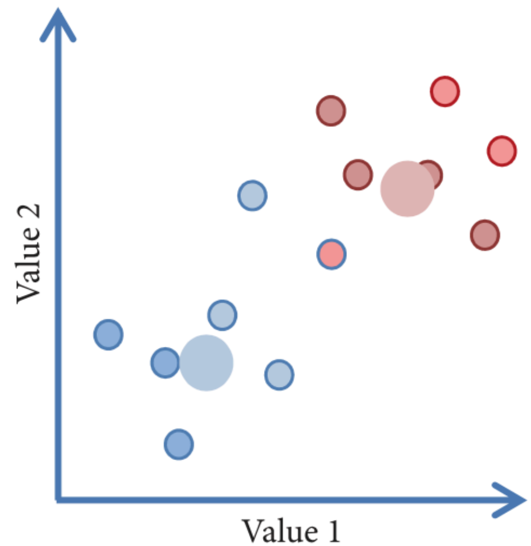
K-means clustering



K cluster centers placed randomly into a dataset



Each image is assigned to belong to the nearest cluster center and class averages are calculated

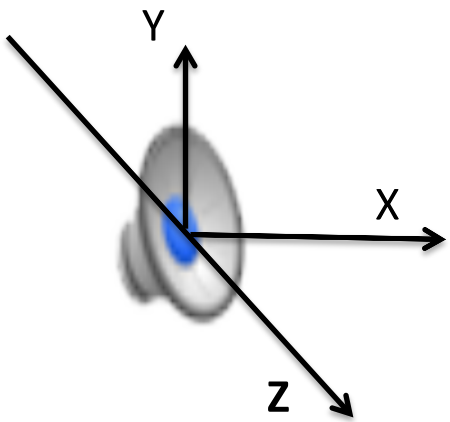


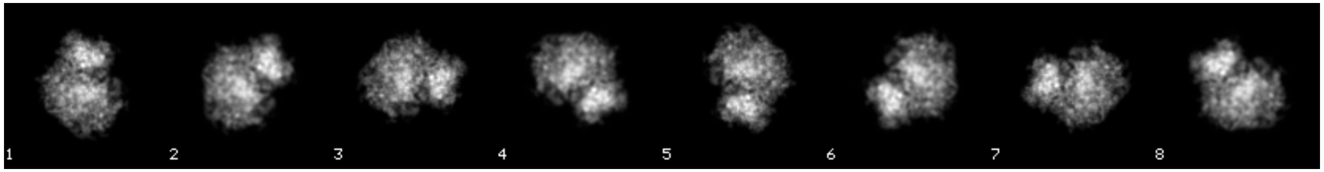
The class averages are the new cluster centers and image assignment is repeated

White et al., 2017, BioMed Research International; Penzceck, Ultramicroscopy, 1996; Diday, Rev. Stat. Appl., 1971

Euler angles

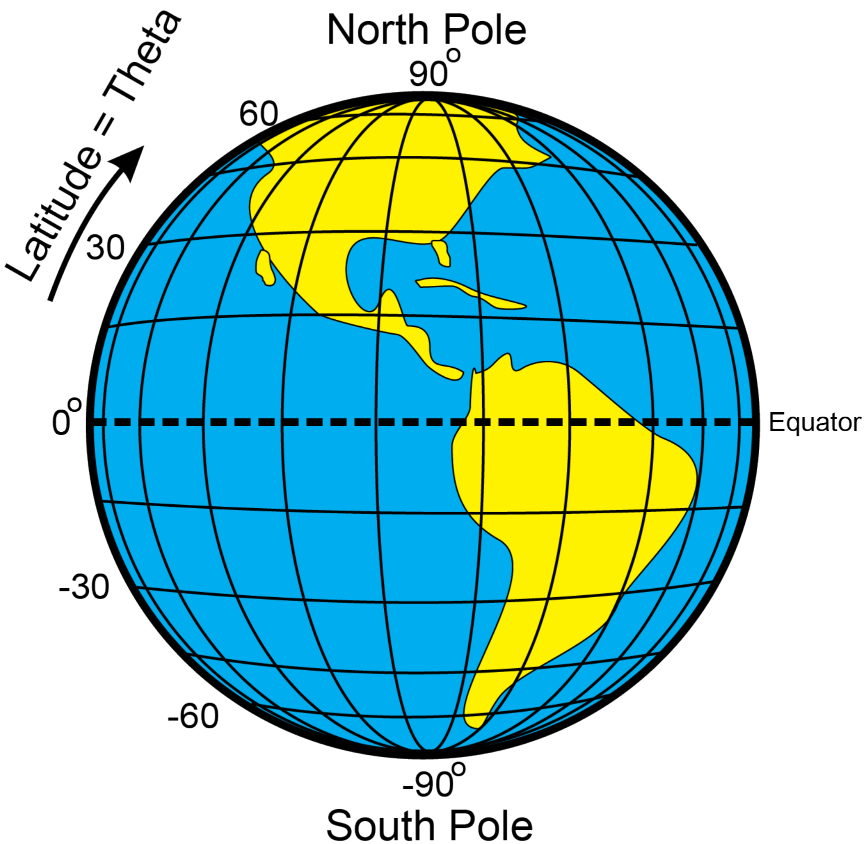
Psi: in plane rotation





Important for alignment of particles in 2D

Theta and Phi: Description of a volume



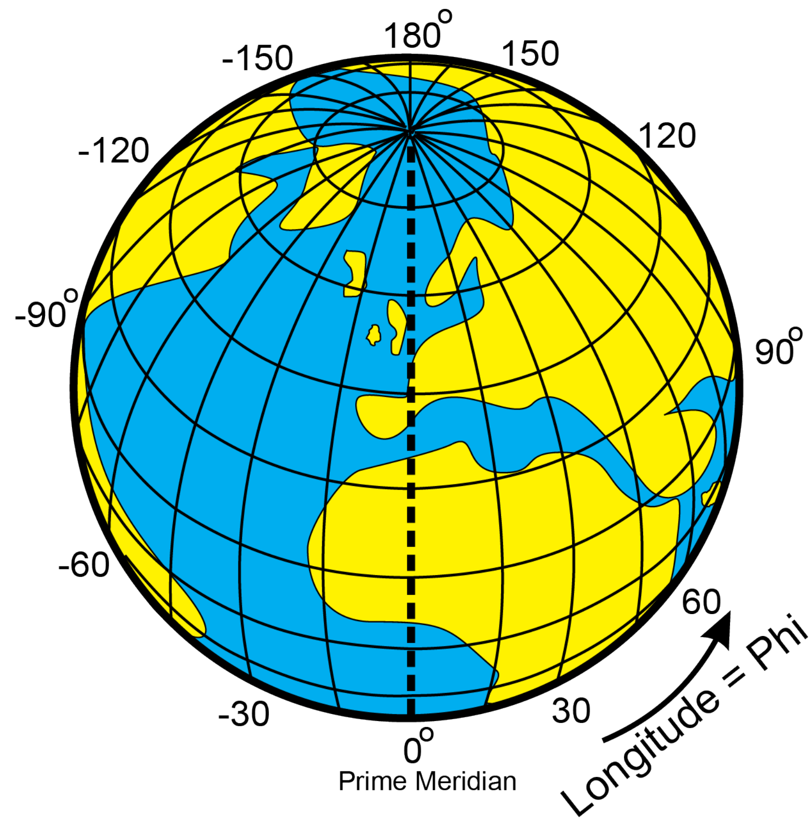
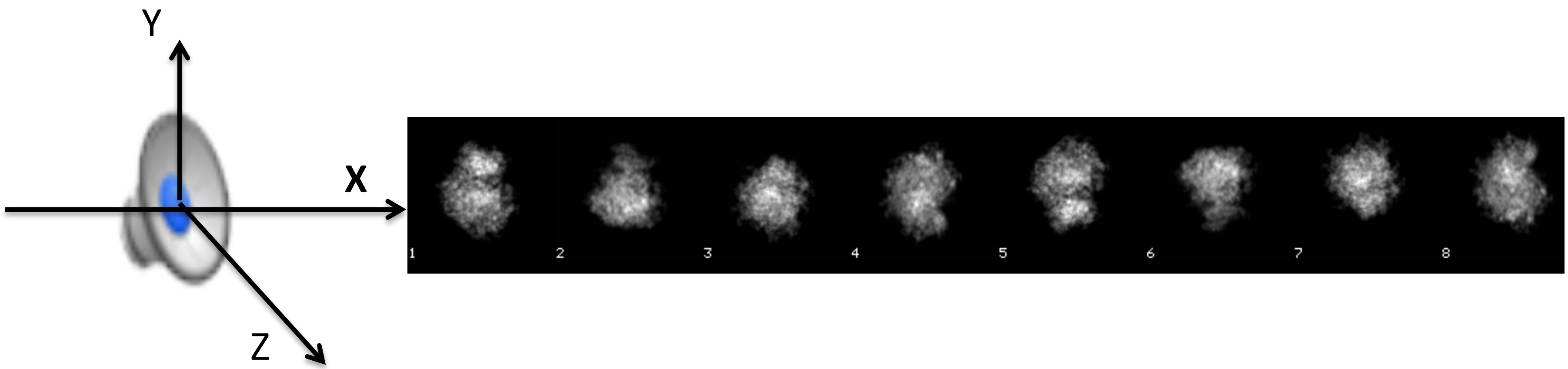


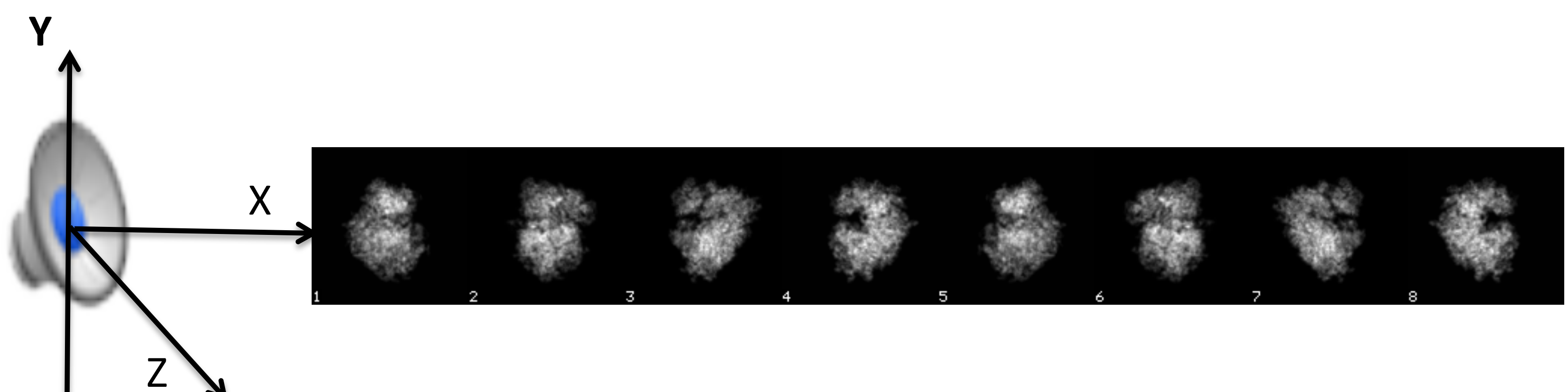
Image adapted from Wikipedia

Euler angles

Theta: out of plane rotation



Phi: rotation around Y-axis
+ 90° Theta

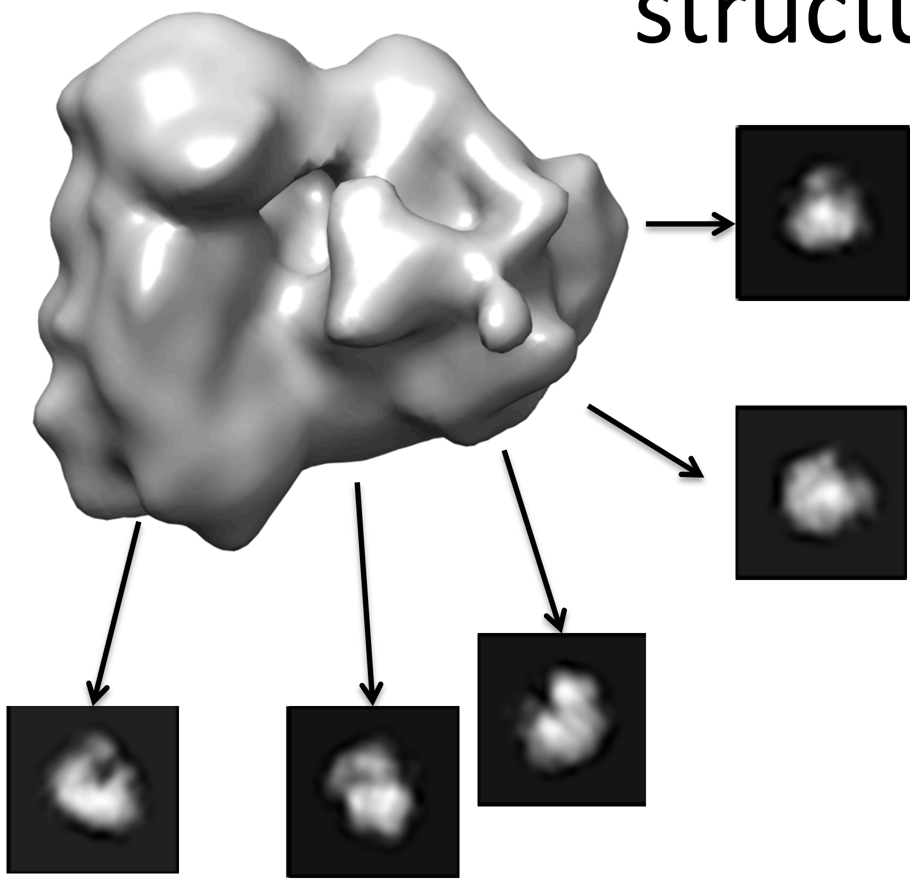


Theta and phi are used to describe the position of 2D images inside a 3D

Determination of Euler angles

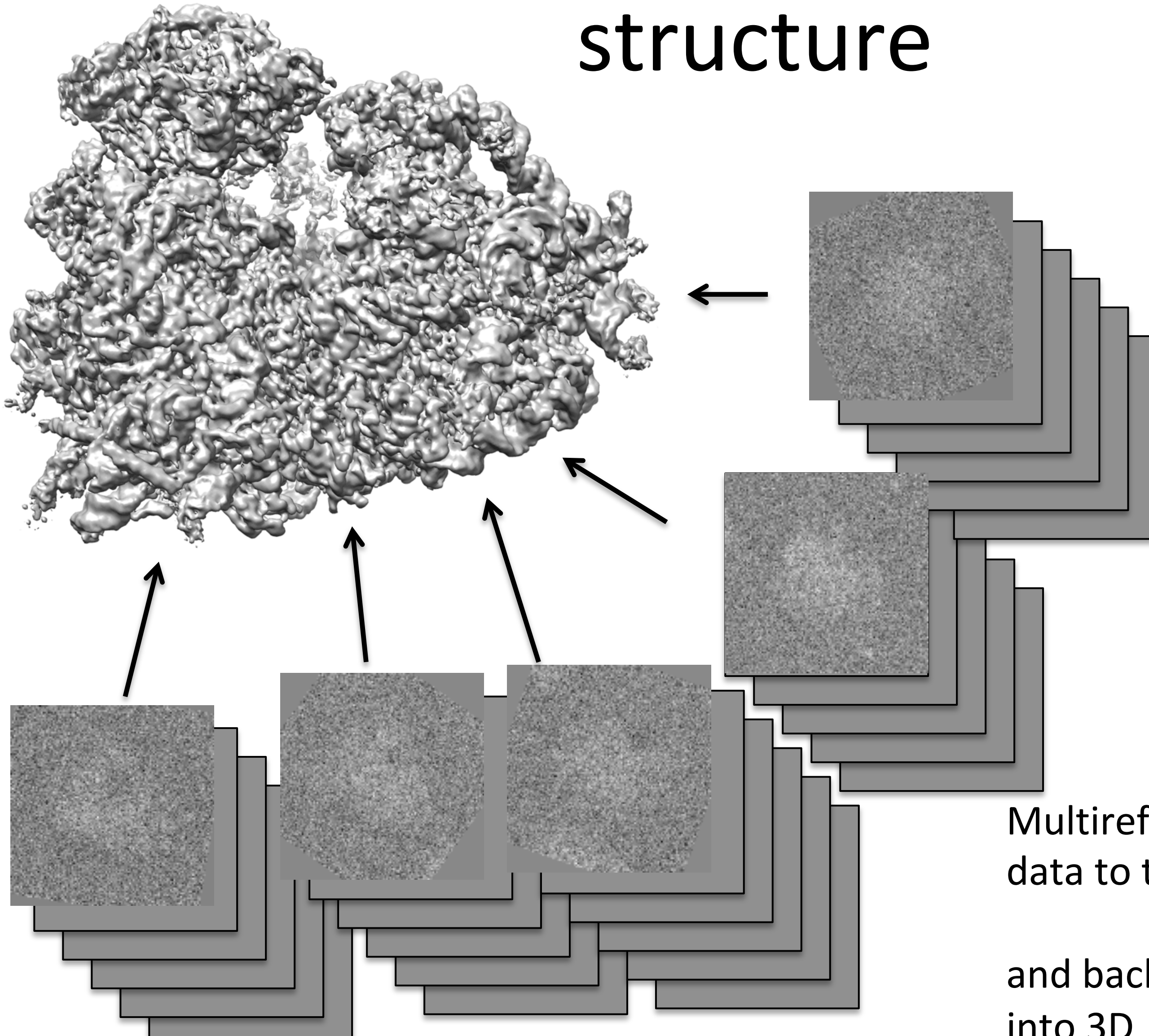
- Projection matching
- Random conical tilt/tomography
- Common lines
- Statistical Methods

1. Projection matching: You already have an idea of the 3D-structure



Generate 2D projections from a similar 3D structure, which will be used as references

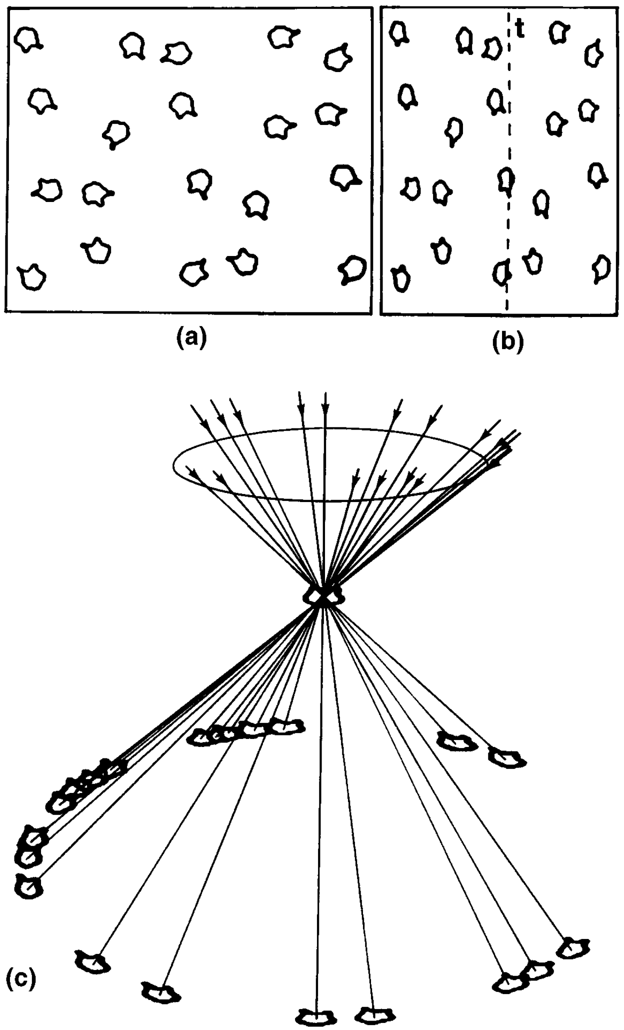
1. Projection matching: You already have an idea of the 3D-structure

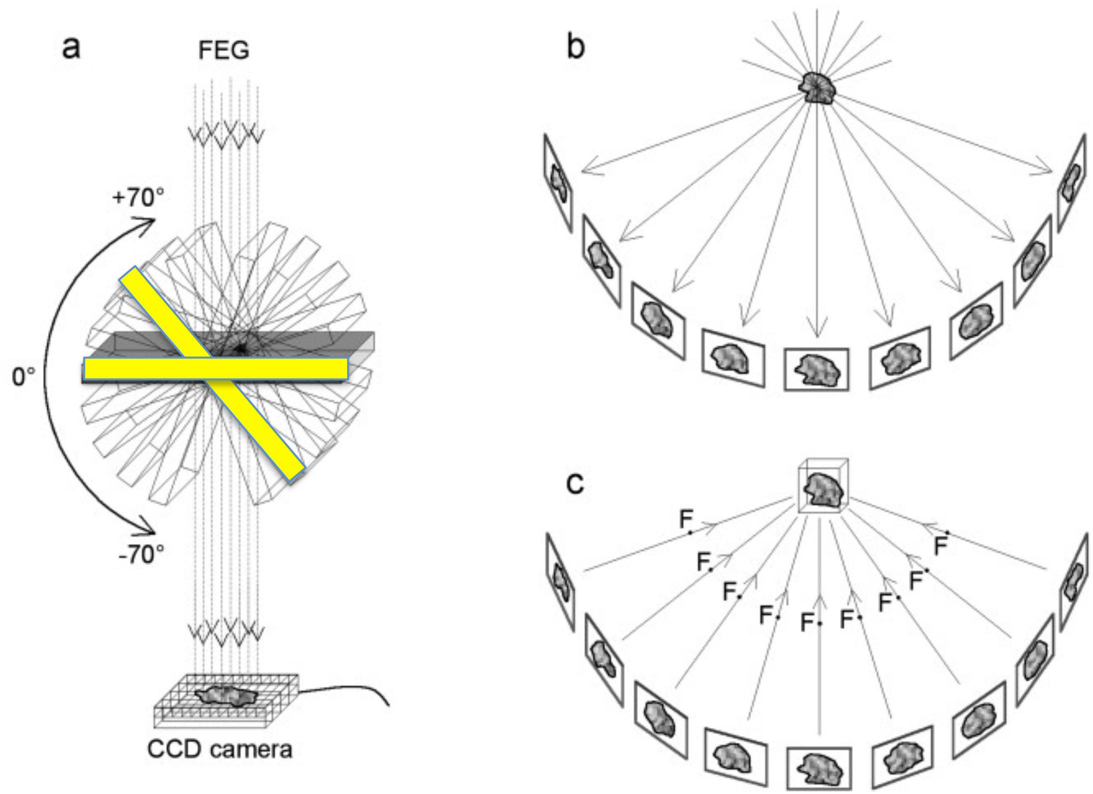


Multireference alignment of your data to the reference-projections

and back-projection of all particles into 3D

2. Random conical tilt/Tomography

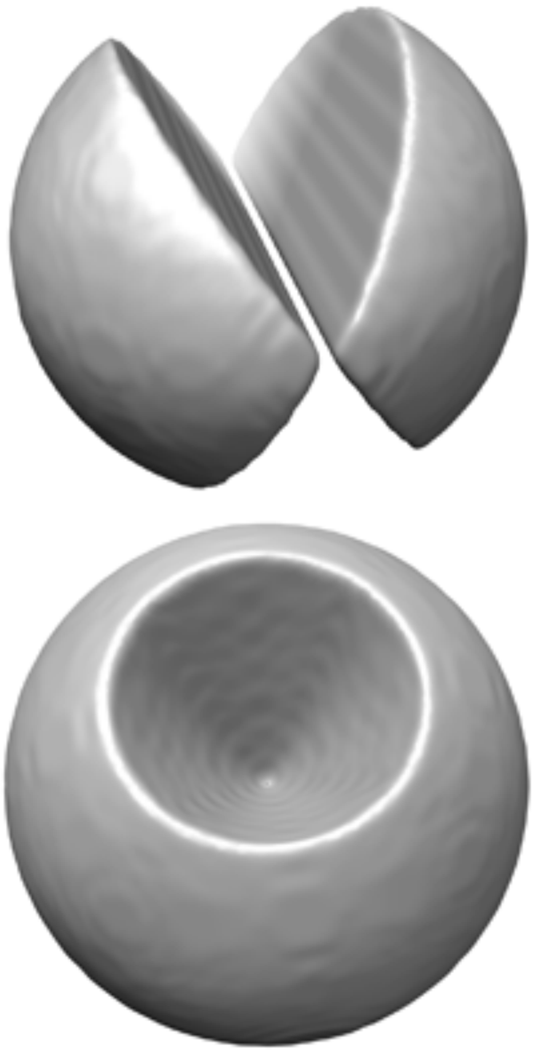


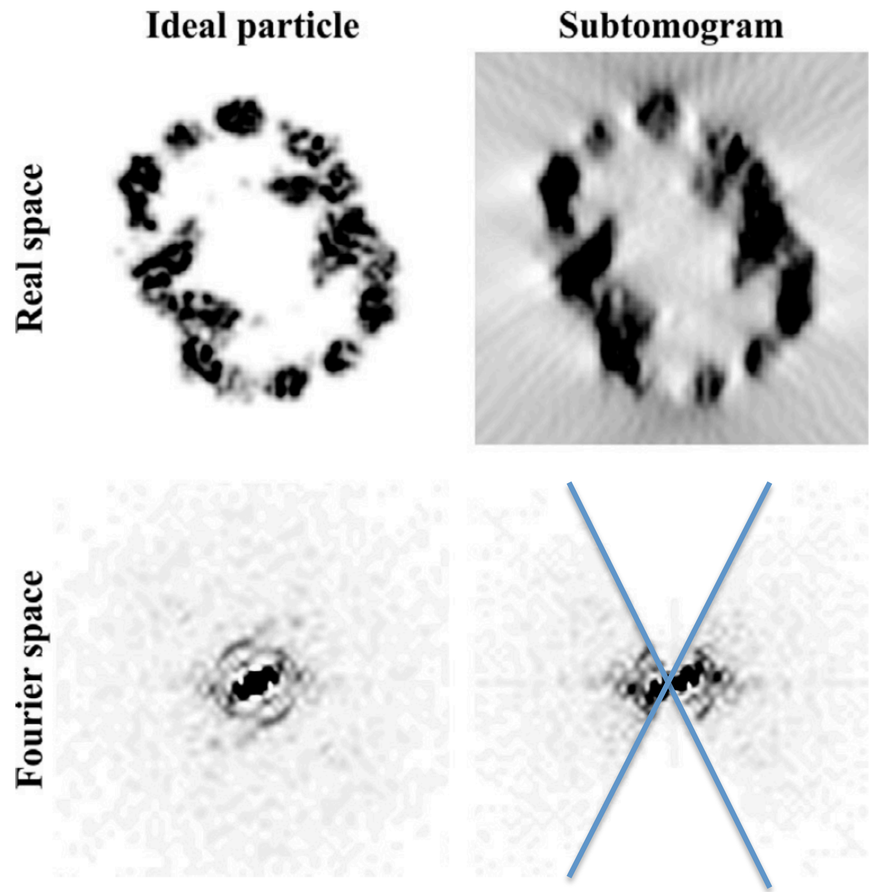


Liu, Frank, 1995, J. Opt. Soc. Am. A
Radermacher et al., J. Microscopy, 1987

Gruenewald et al., 2002, Biophys. Chem.

The Missing Wedge problem

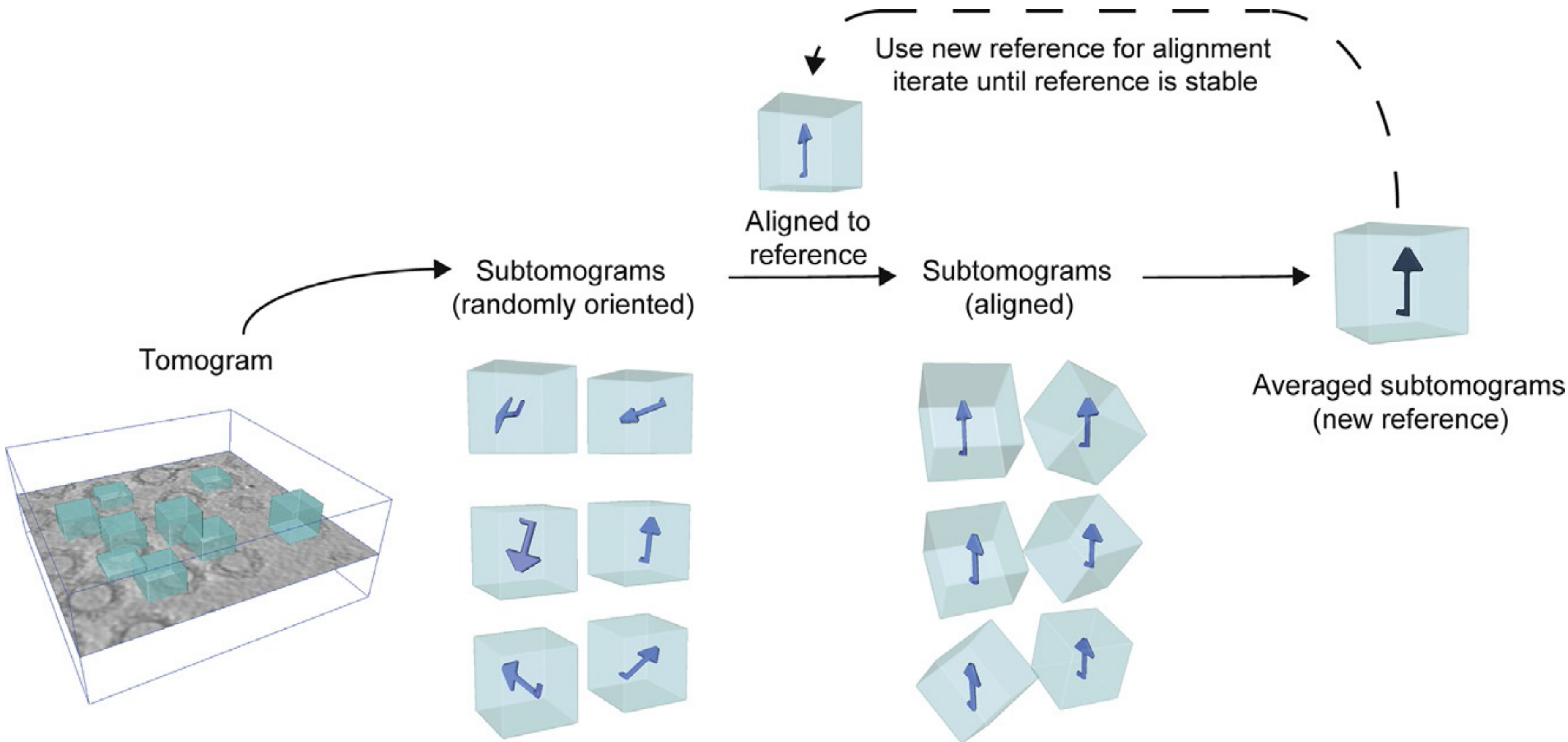




Wan,Briggs, 2016, Methods in Enzymology

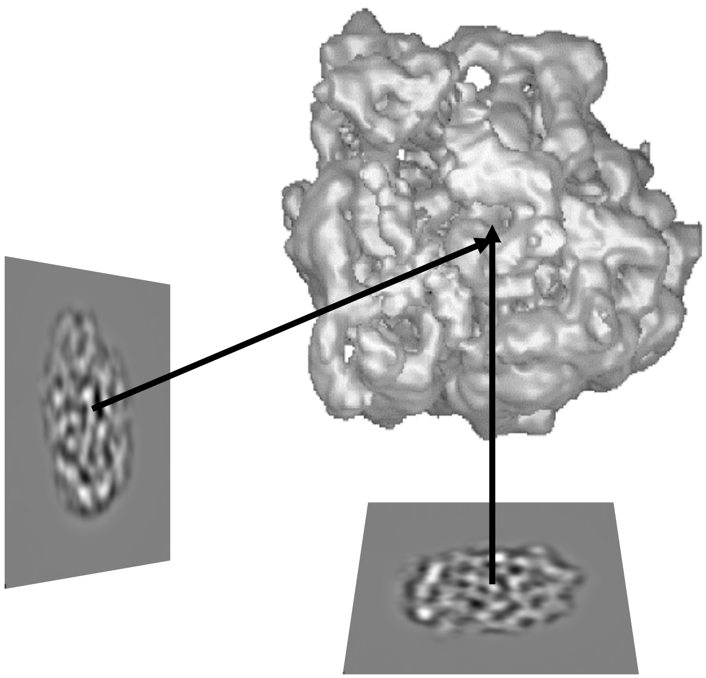
Taken from Bsoft website: http://www.msg.ucsf.edu/local/programs/bsoft/howto/bsoft_missing.html

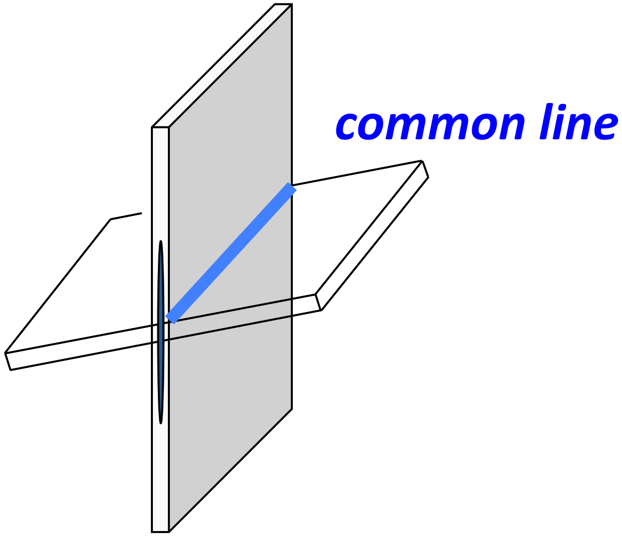
Sub-Tomogram Averaging



Wan,Briggs, 2016, Methods in Enzymology

3. Angular reconstitution with Common lines





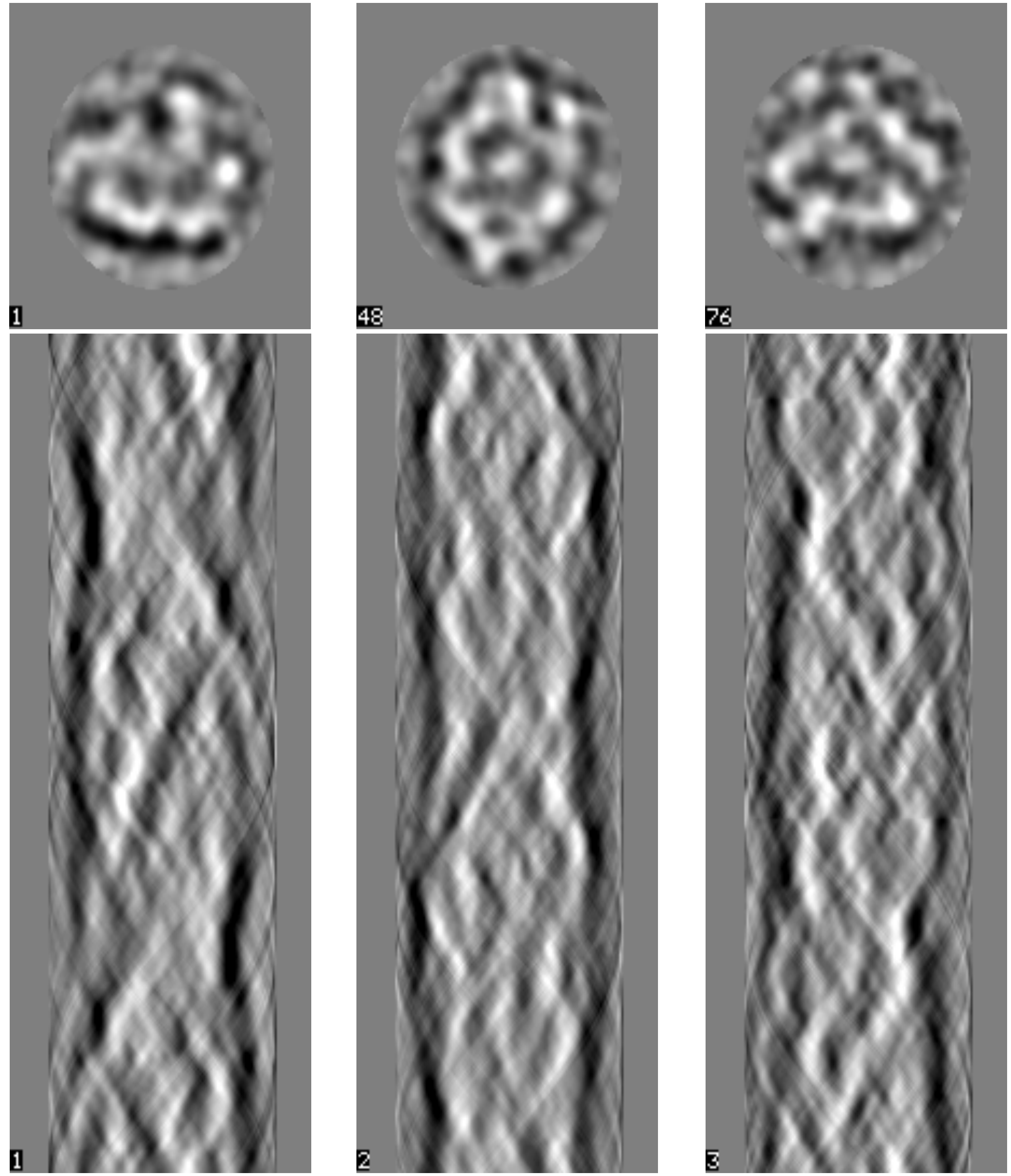
common line projections theorem
Theorem of the central section.

II. Structure determination
- angle assignment
- angular reconstitution

sinogram = **line-projection** of the 2D image
(also called Radon transform)

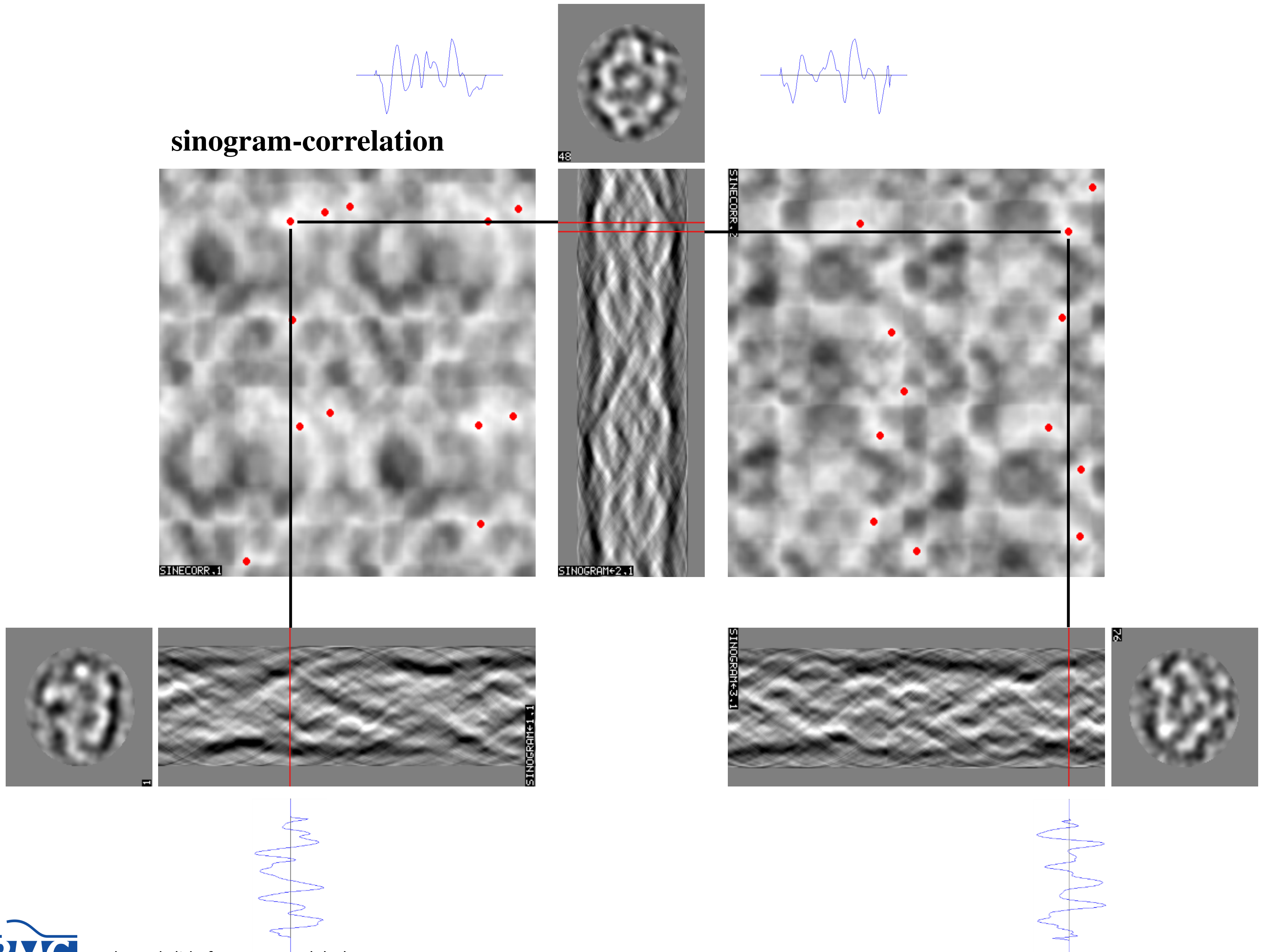
amplitude-square-root filtered

Select 3 clearly different views (here: class average numbers 1,48,76):



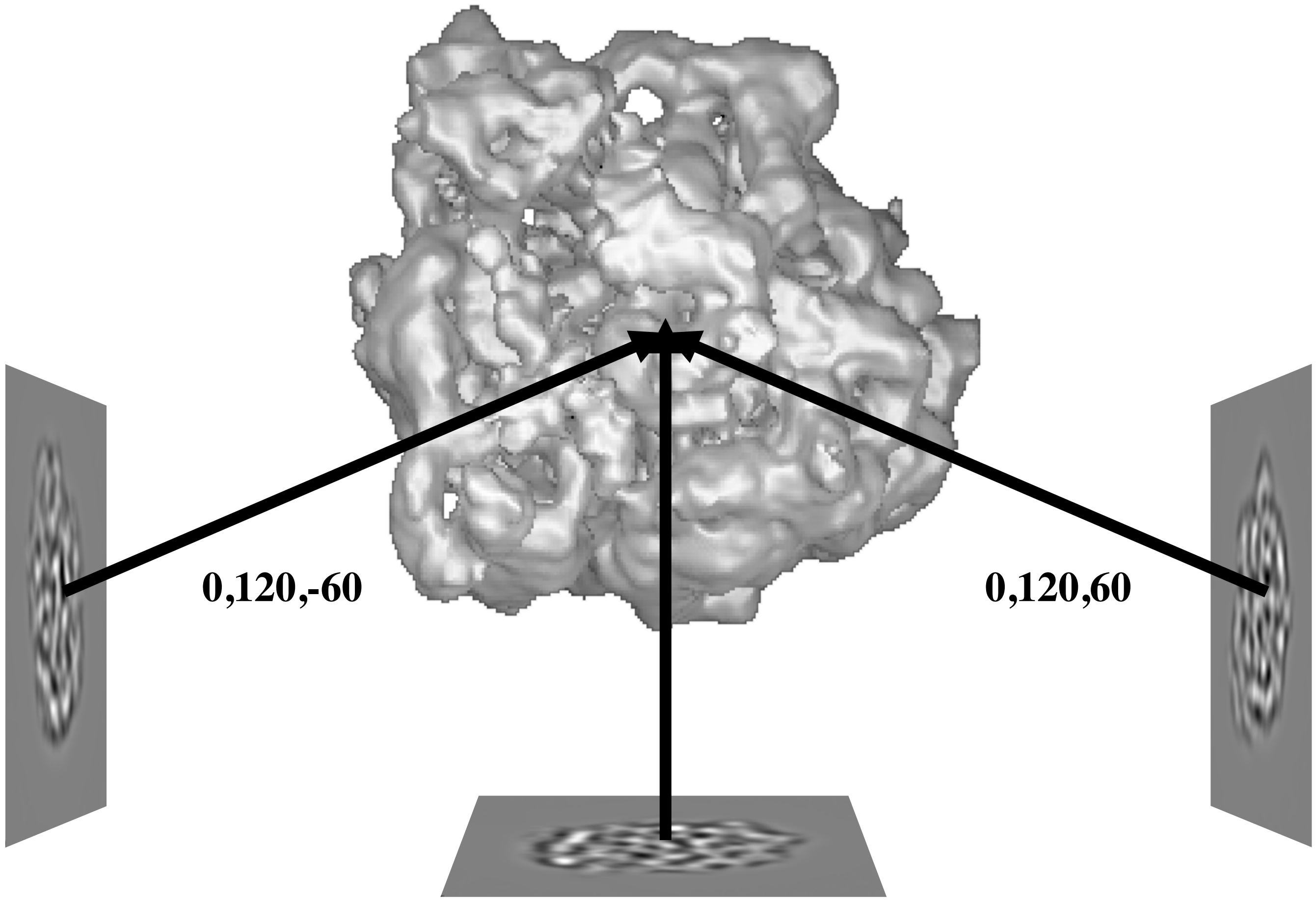
adapted slide from Bruno Klaholz

sinogram-correlation



adapted slide from Bruno Klaholz

II. Structure determination
- 3D reconstruction

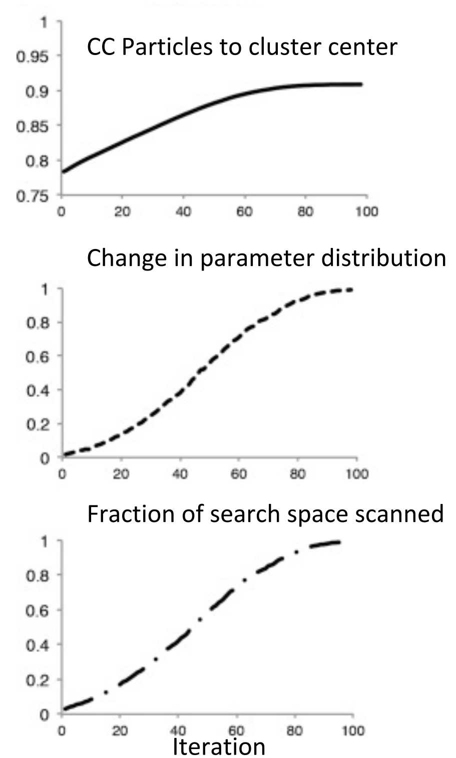


back-projection → 3D reconstruction

4. Stochastic approach

hill climbing

1. Assign images random angles
2. sort images into x number of random classes.
3. For each individual particle image, identify the in-plane rotation and cluster assignment that improves the correlation in comparison to last round.
4. Iterate the process until convergence.

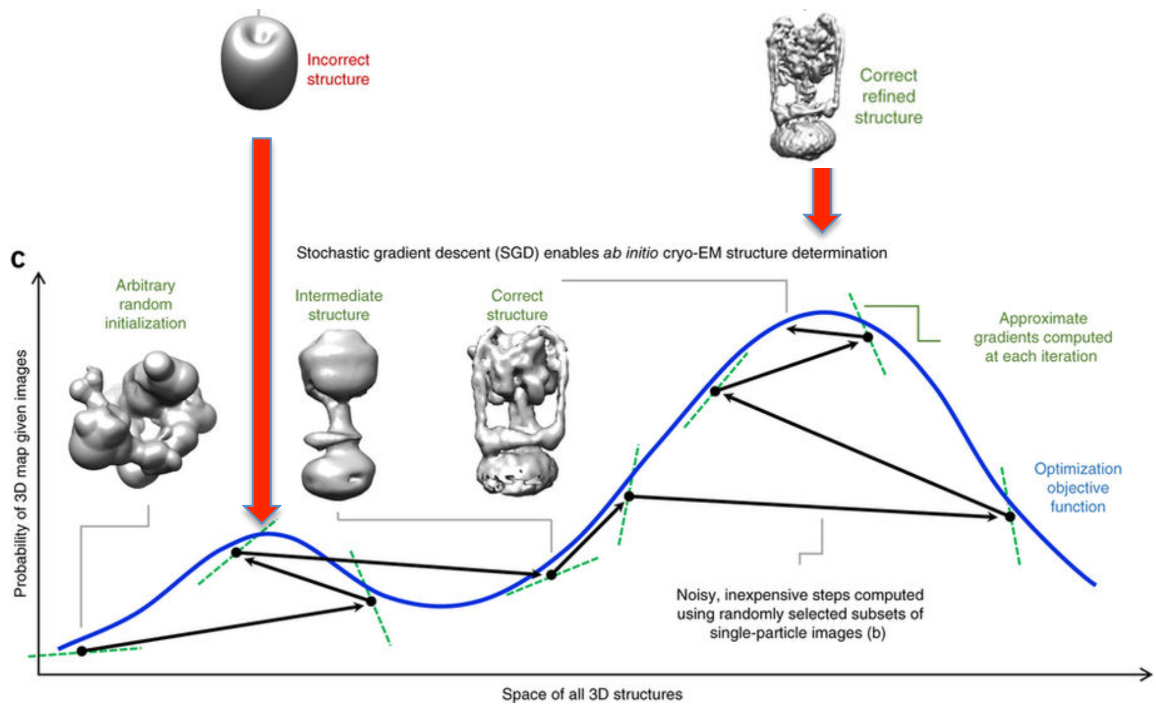


gradient descendent

1. Overall likelihood of a small number of randomly chosen images to correspond to a 3D structure is calculated
2. Each iteration a different subset is chosen and the gradient between overall likelihood compared to previous round is computed
3. 3D structure is updated each iteration based on computed gradient

local maximum

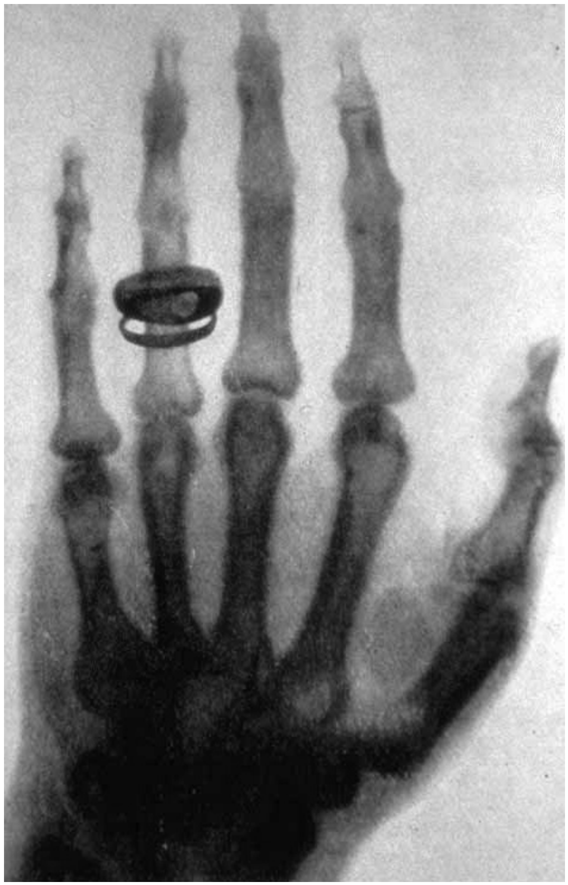
true solution



Reboul et al., 2016, Structure

Punjani et al., 2017, Nature Methods

The problem of the right hand

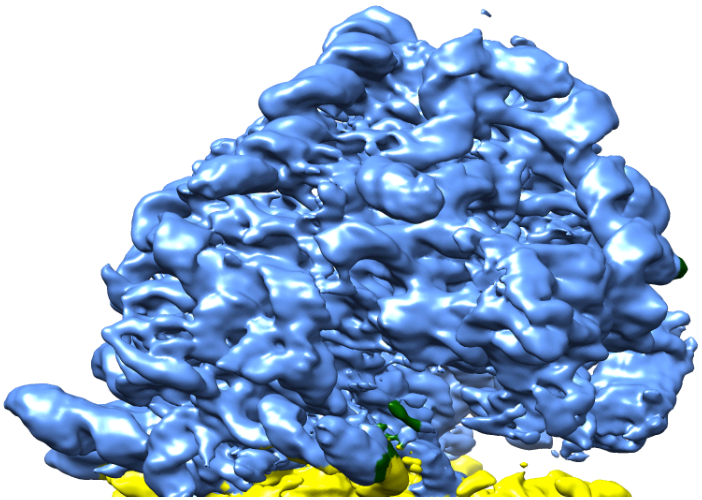


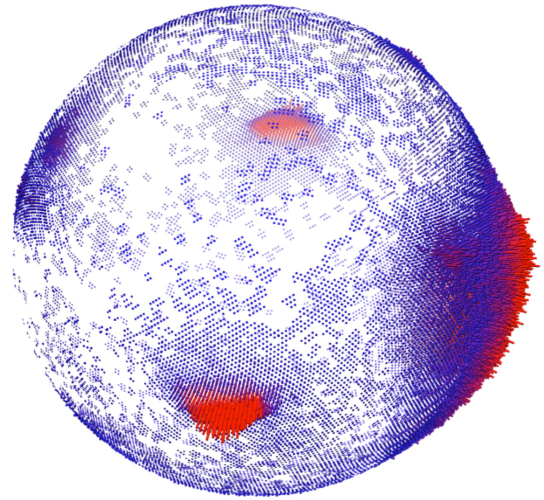
One of the first x-ray images taken by C. Roentgen

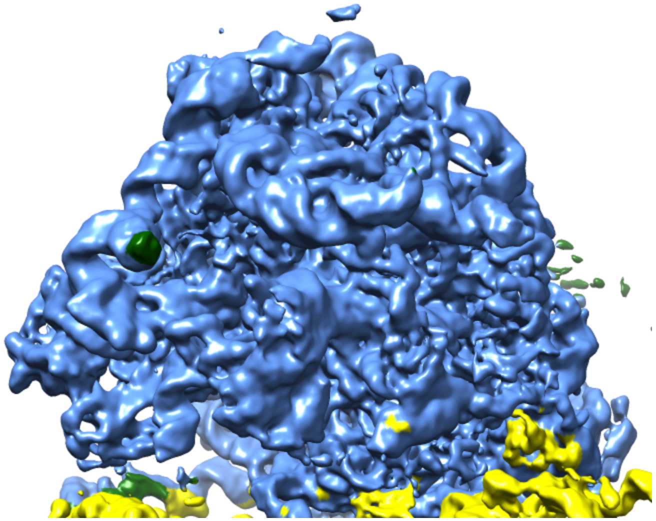
In transmission microscopy it is not possible to distinguish between the left and the right hand from one image

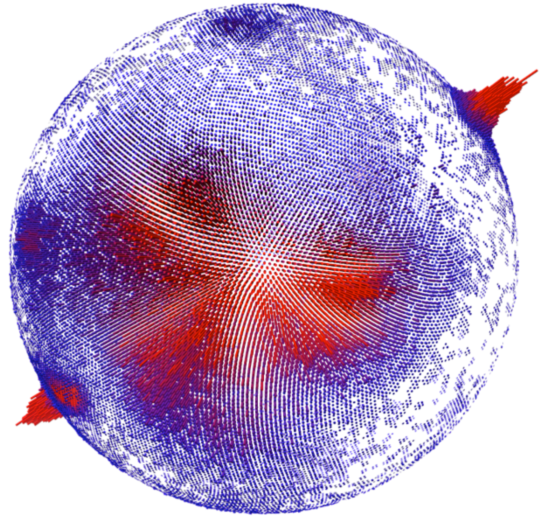
Issue with common lines and statistical approach. For RCT/Tomography the hand is fixed by tilting the specimen

Uneven angular distribution (preferential orientation)

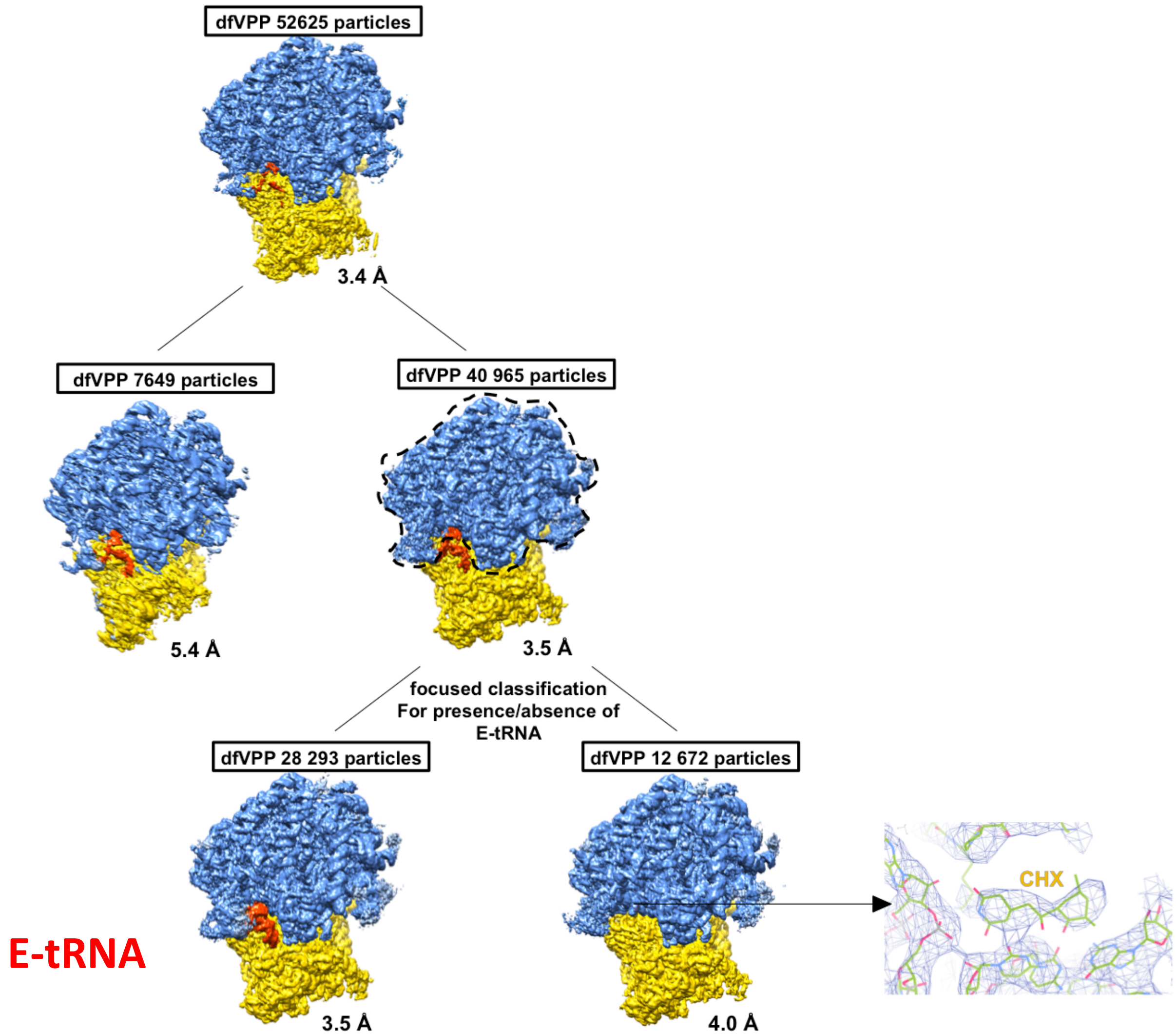








Heterogeneity



Only after sorting for heterogeneity it is possible to see full density for CHX/E-tRNA

von Loefffelholz et al, JSB, 2018

Further reading

- Classification and Alignment

  - Elad, N., Clare, D.K., Saibil, H.R., Orlova, E.V., 2008. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. J Struct Biol 162, 108–20. https://doi.org/10.1016/j.jsb.2007.11.007
  - Scheres, S.H., 2010. Classification of structural heterogeneity by maximum-likelihood methods. Methods Enzym. 482, 295–320. https://doi.org/10.1016/S0076-6879(10)82012-9
  - Sigworth, F.J., 1998. A maximum-likelihood approach to single-particle image refinement. J Struct Biol 122, 328–39. https://doi.org/10.1006/jsbi.1998.4014
  - Sigworth, F.J., Doerschuk, P.C., Carazo, J.-M., Scheres, S.H.W., 2010. An introduction to maximum-likelihood methods in cryo-EM. Methods Enzymol. 482, 263–294. https://doi.org/10.1016/S0076-6879(10)82011-7
  - White, H.E., Ignatiou, A., Clare, D.K., Orlova, E.V., 2017. Structural Study of Heterogeneous Biological Samples by Cryoelectron Microscopy and Image Processing. BioMed Res. Int. 2017, 1032432. https://doi.org/10.1155/2017/1032432

- Initial Structure Generation

  - Crowther, R.A., DeRosier, D.J., Klug, A., 1970. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc R Soc Lond A 317, 319–340. https://doi.org/10.1098/rspa.1970.0119
  - Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296. https://doi.org/10.1038/nmeth.4169
  - Radermacher, M., Wagenknecht, T., Verschoor, A., Frank, J., 1987. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J. Microsc. 146, 113–136.
  - Reboul, C.F., Bonnet, F., Elmlund, D., Elmlund, H., 2016. A stochastic hill climbing approach for simultaneous 2D alignment and clustering of cryogenic electron microscopy images. Structure 24, 988–996. https://doi.org/10.1016/j.str.2016.04.006
  - Van Heel, M., 1987. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy 21, 111–123. https://doi.org/10.1016/0304-3991(87)90078-7

- Book about cryo-EM and image processing

J. Frank, Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State, Oxford University Press, 2nd edition, 2008.