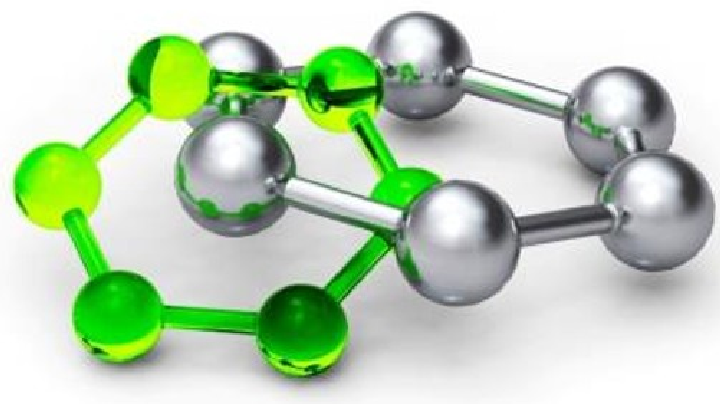
Central European Institute of Technology
BRNO | CZECH REPUBLIC

Image analysis III & 3D Reconstruction

C9940 3-Dimensional Transmission Electron Microscopy
S1007 Doing structural biology with the electron microscope



April 11, 2016

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

2007–13
OP Research and
Development for Innovation

Outline

Image analysis III

- More on last week's material
    - Dependence of SNR on $\sqrt{N}$
    - Oversampling
- Classification

3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

Outline

Image analysis III

- More on last week's material
  - Dependence of SNR on $\sqrt{N}$
  - Oversampling
- Classification

3D Reconstruction

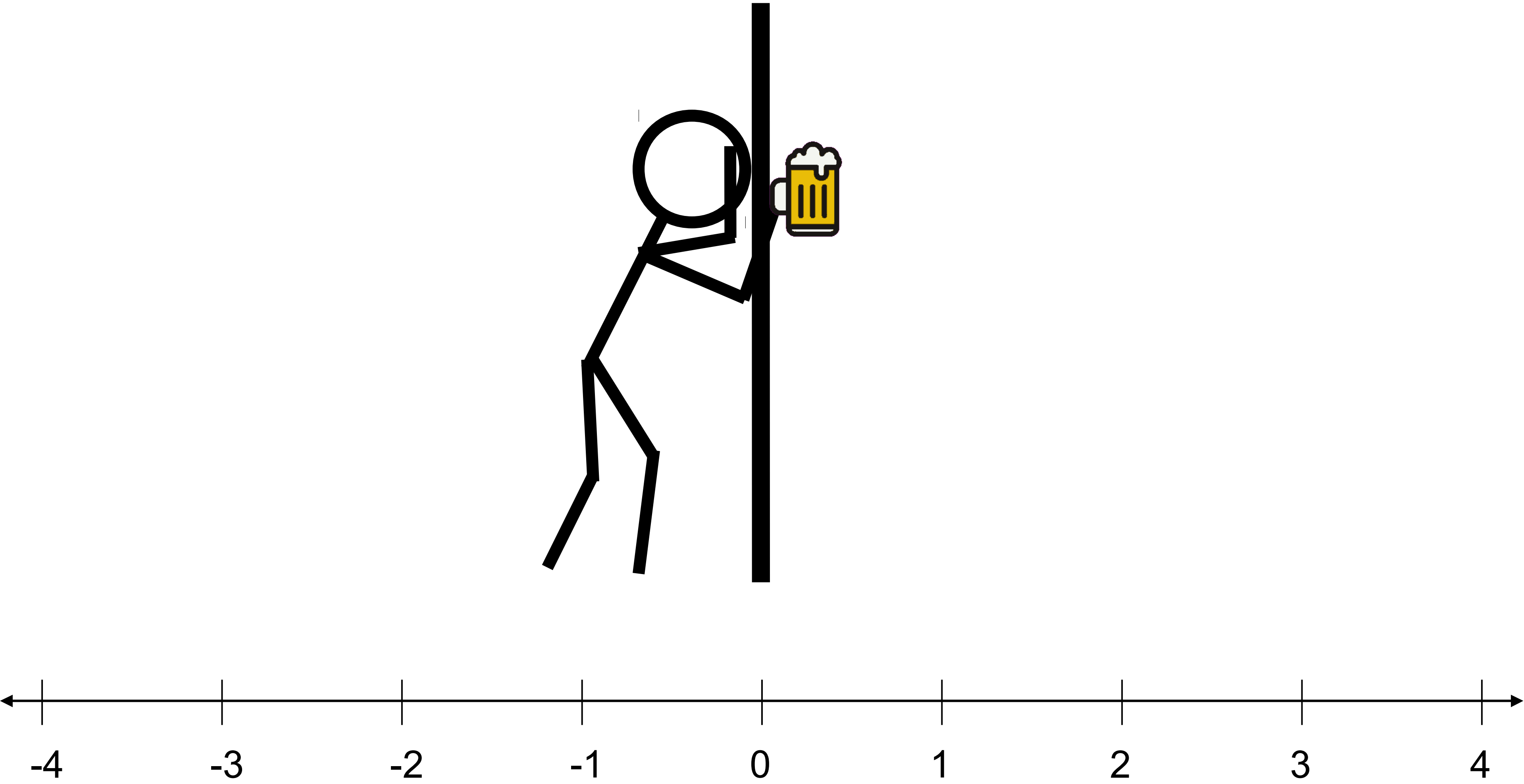
- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

Random walks:
Why signal-to-noise improves with $\sqrt{N}$

The “Drunkard's walk”

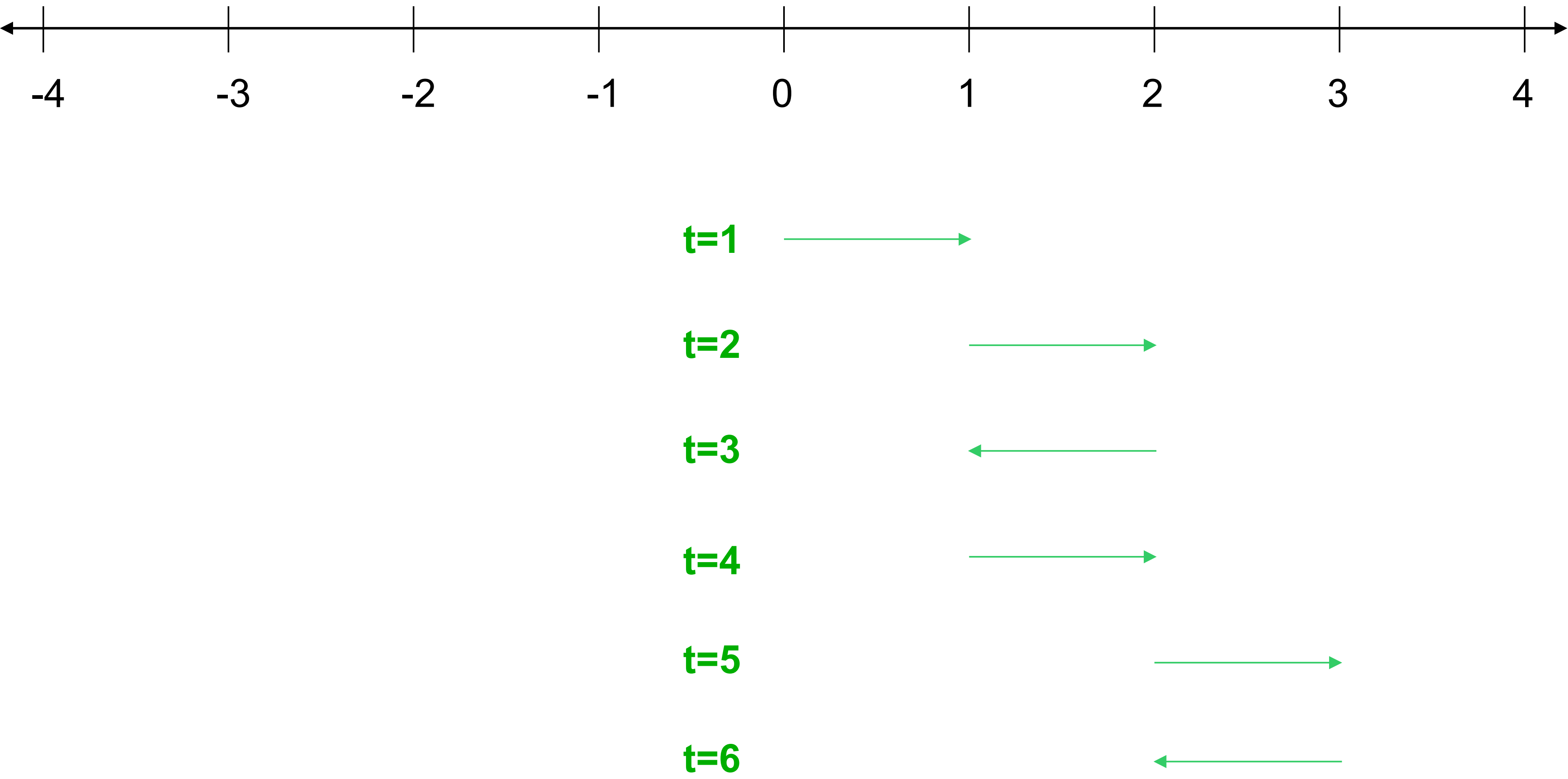
Let's conduct an experiment.

The “Drunkard's walk”

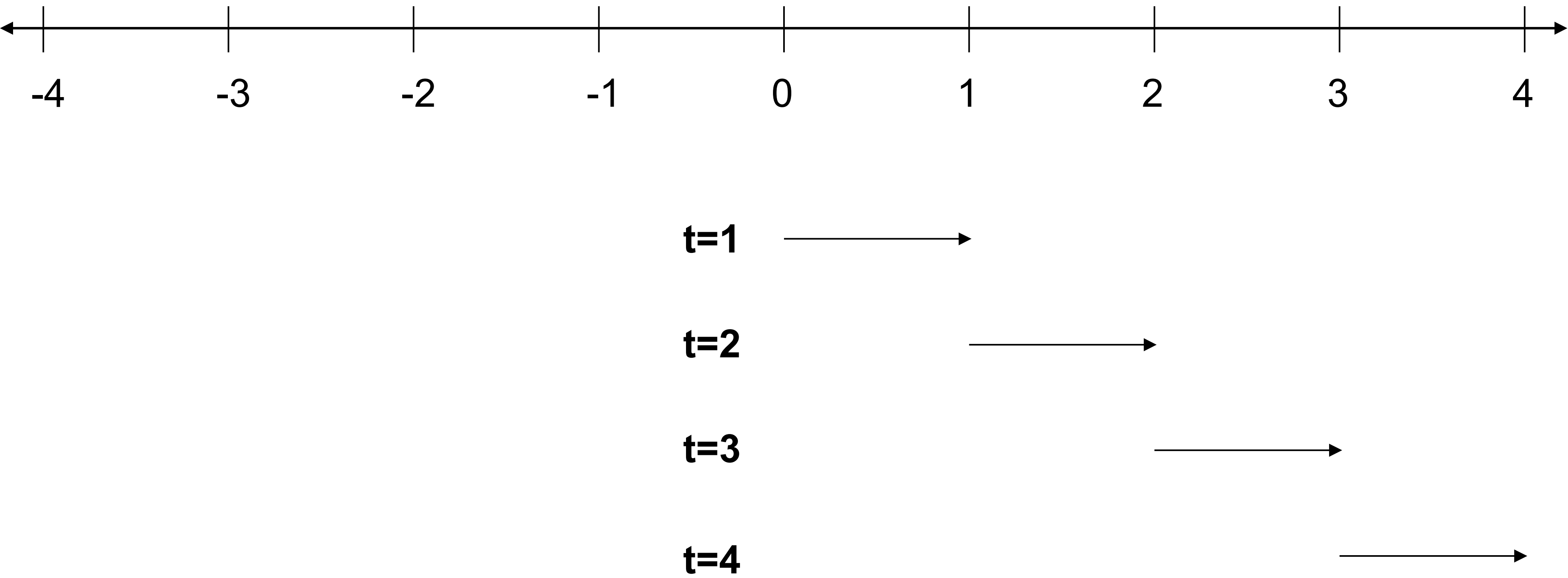


We're going to assume that each step is random and independent of previous steps.

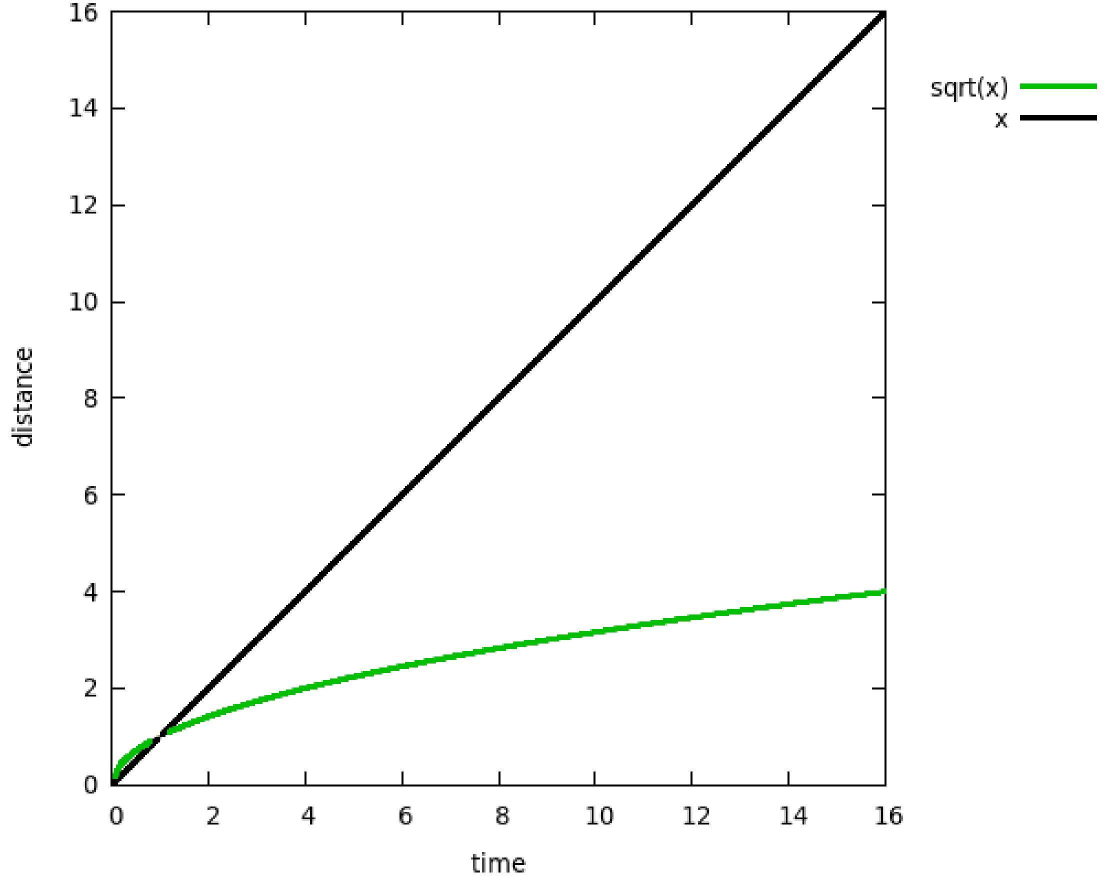
The “Drunkard's walk”



The teetotaler's walk

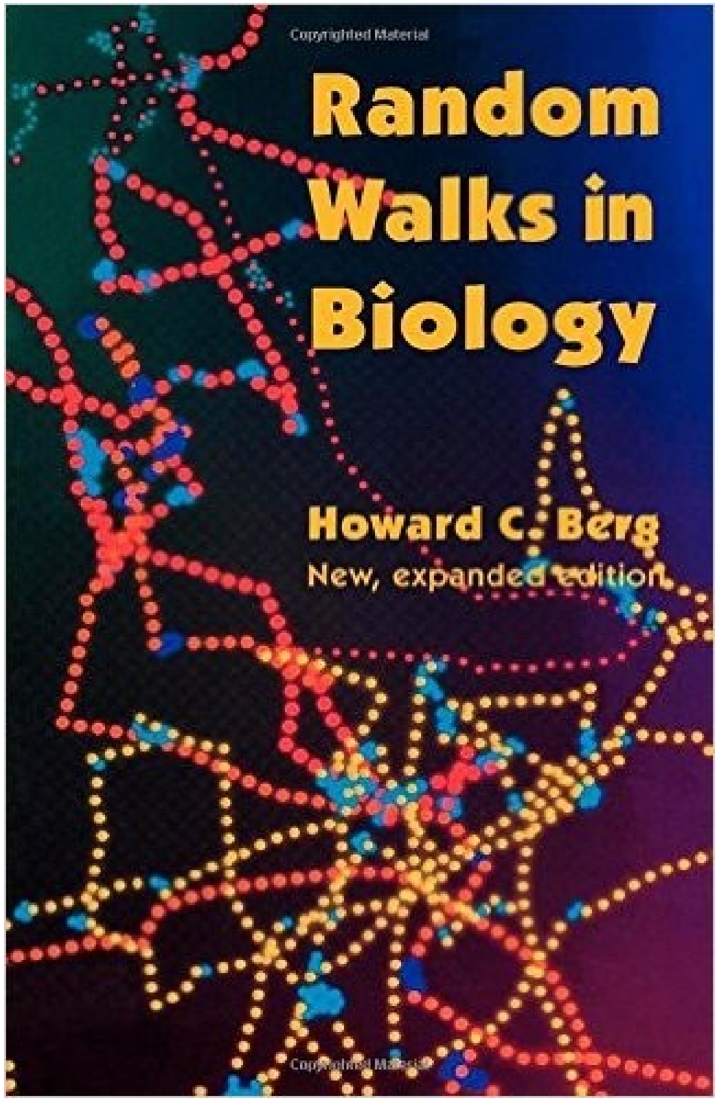


Expectation value



The expected distance that “noise” travels increases with √N.
However, it is not as fast as the distance that “signal” travels.
Thus, as we collect more data, the SNR increase by N/√N = √N

Random walks: more information

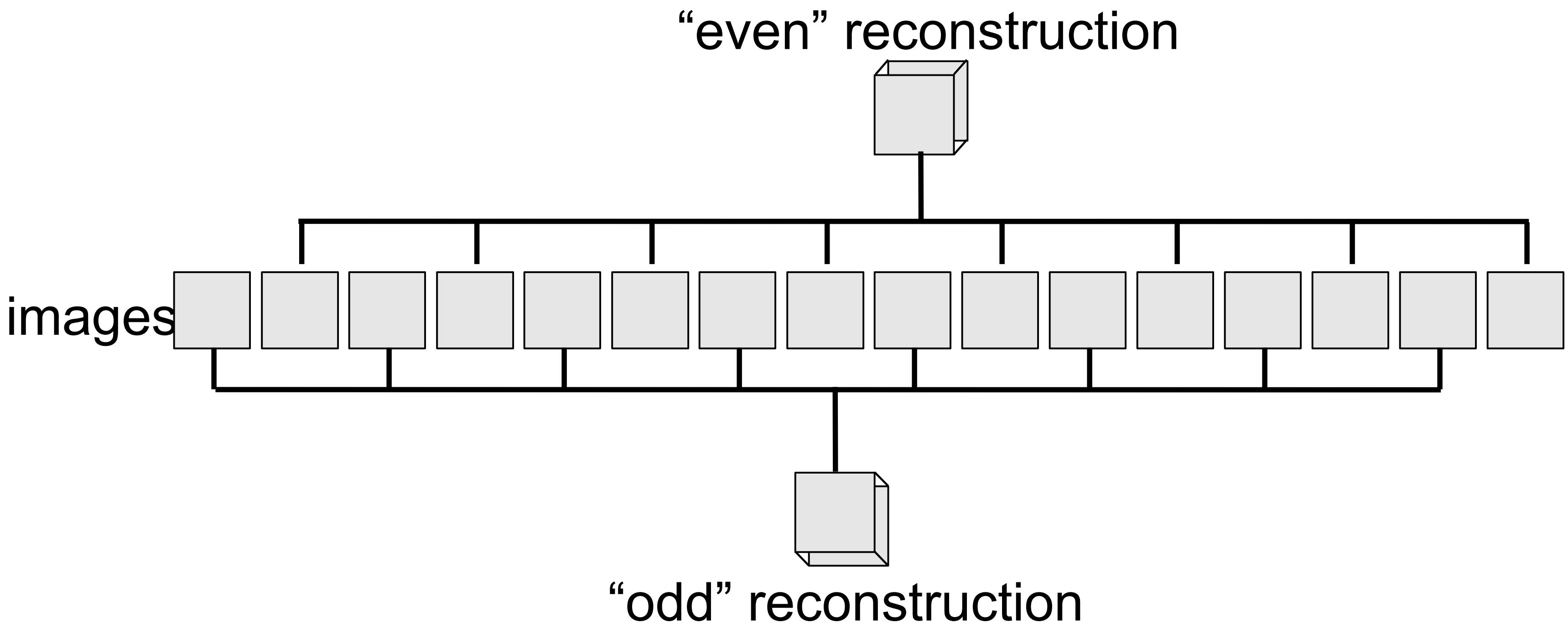


CEITEC

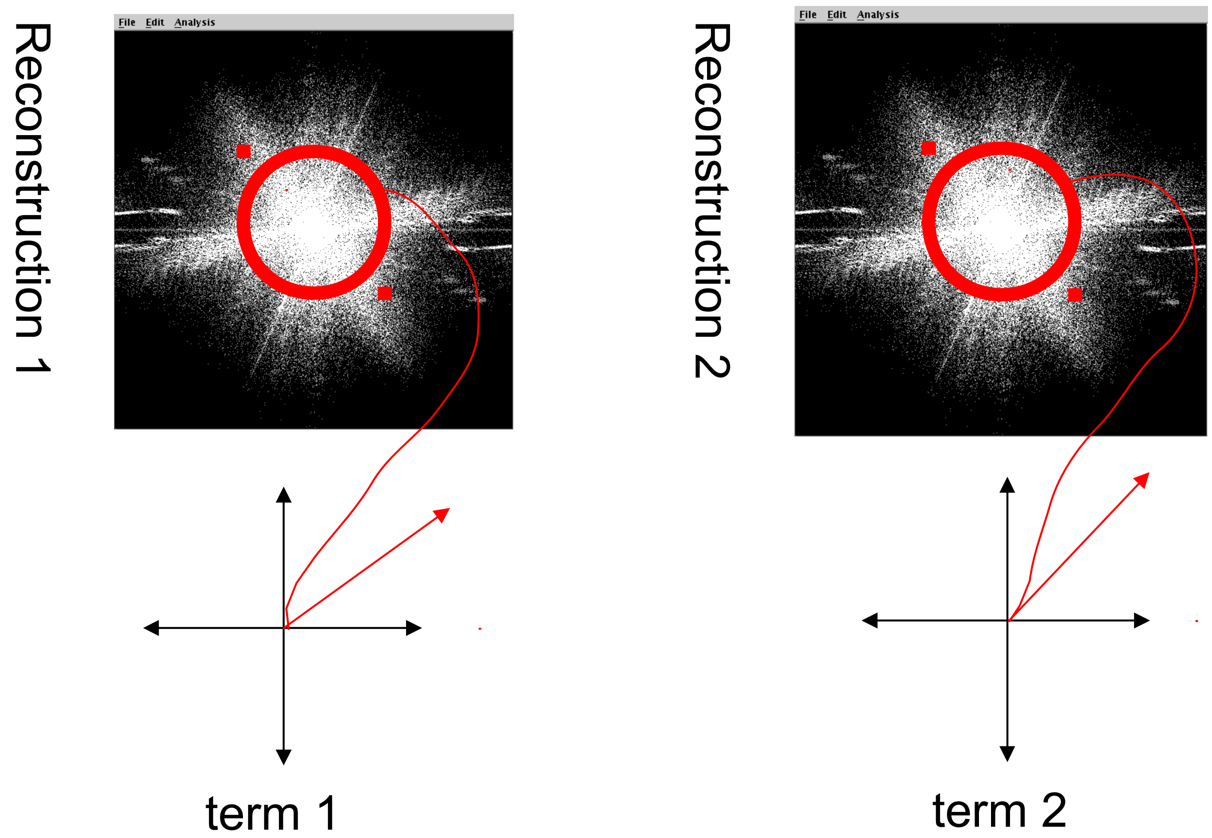
Expectation values
and how they related to resolution criteria

Review:
How do we evaluate the quality of a reconstruction?

We split the data set into halves and compare them.



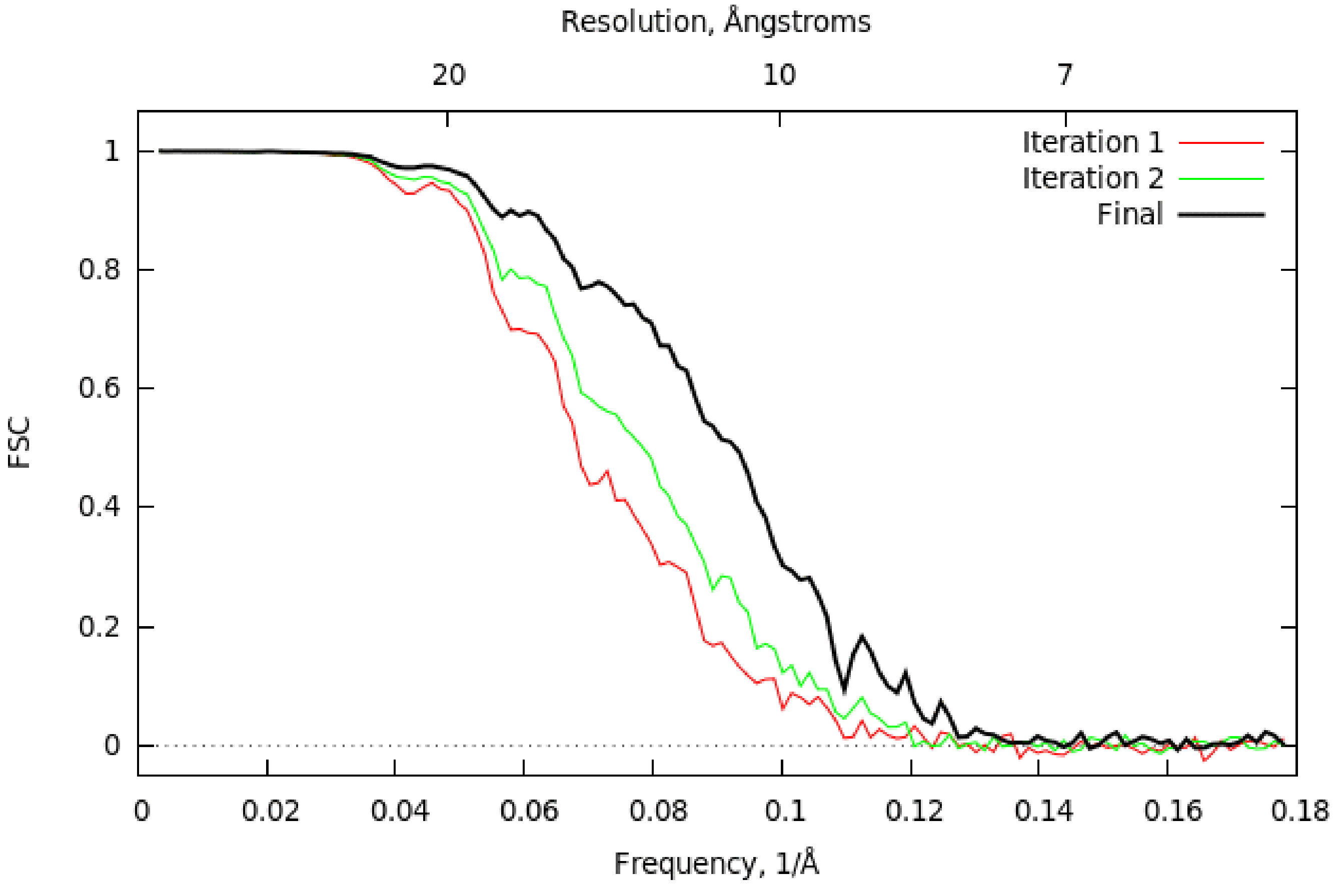
Review: Fourier Shell Correlation (FSC)



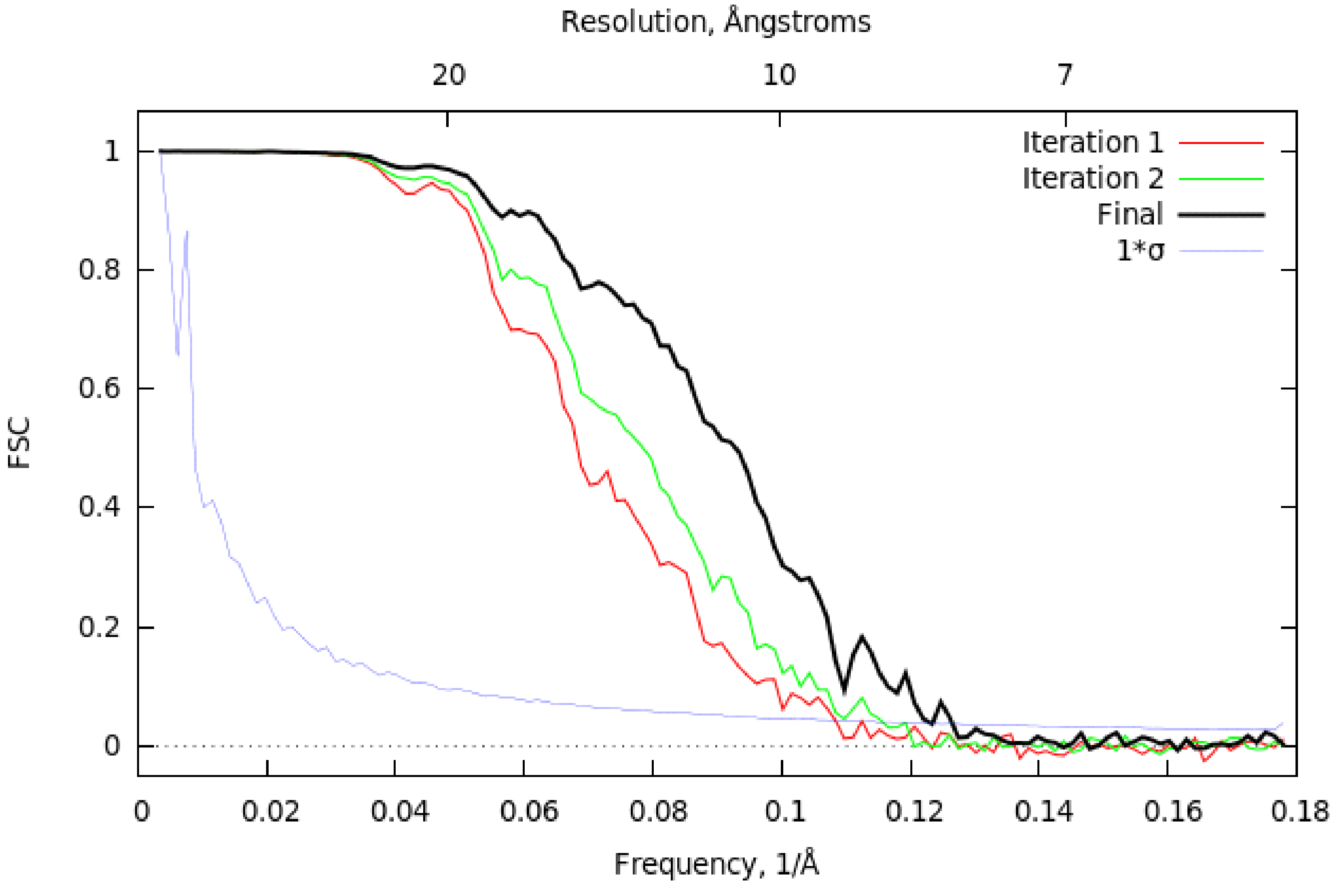
Properties:
- Fourier terms have amplitude + phase.
- Correlation values range from -1 to +1.
- Noise <u>should</u> give an average of 0.
- The comparison is done as a function of spatial frequency (or “resolution”)

CEITEC

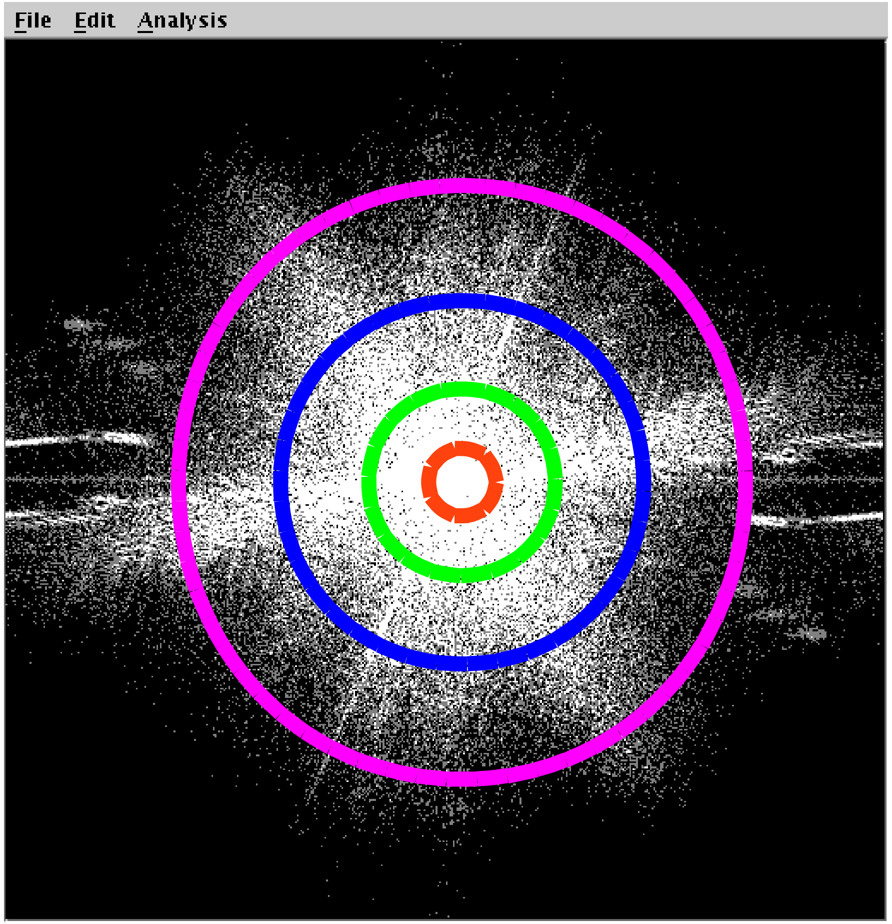
Review: Fourier Shell Correlation curve

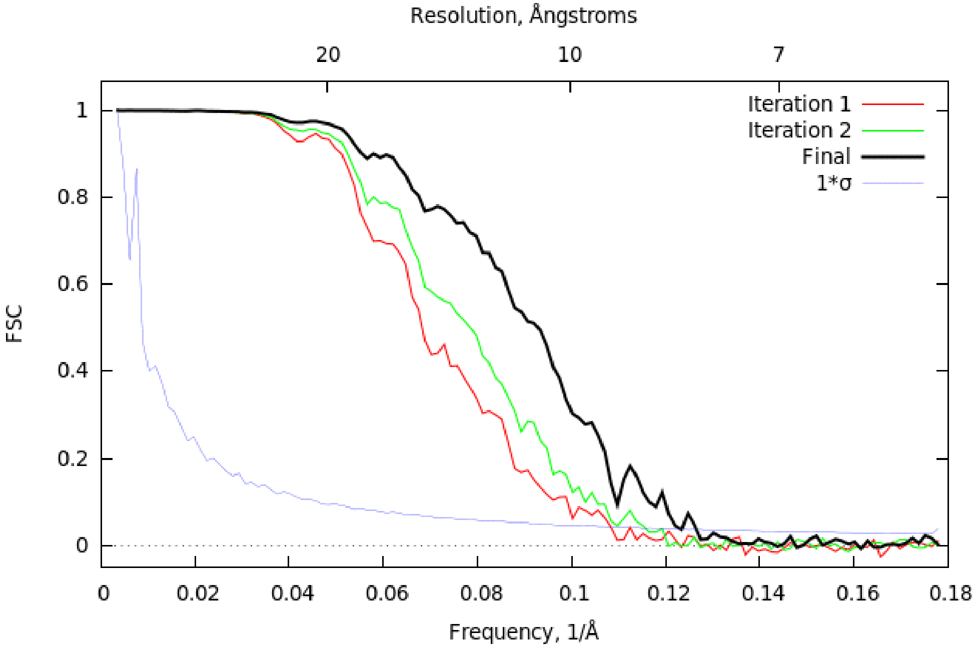


FSC curve with expectation value of noise



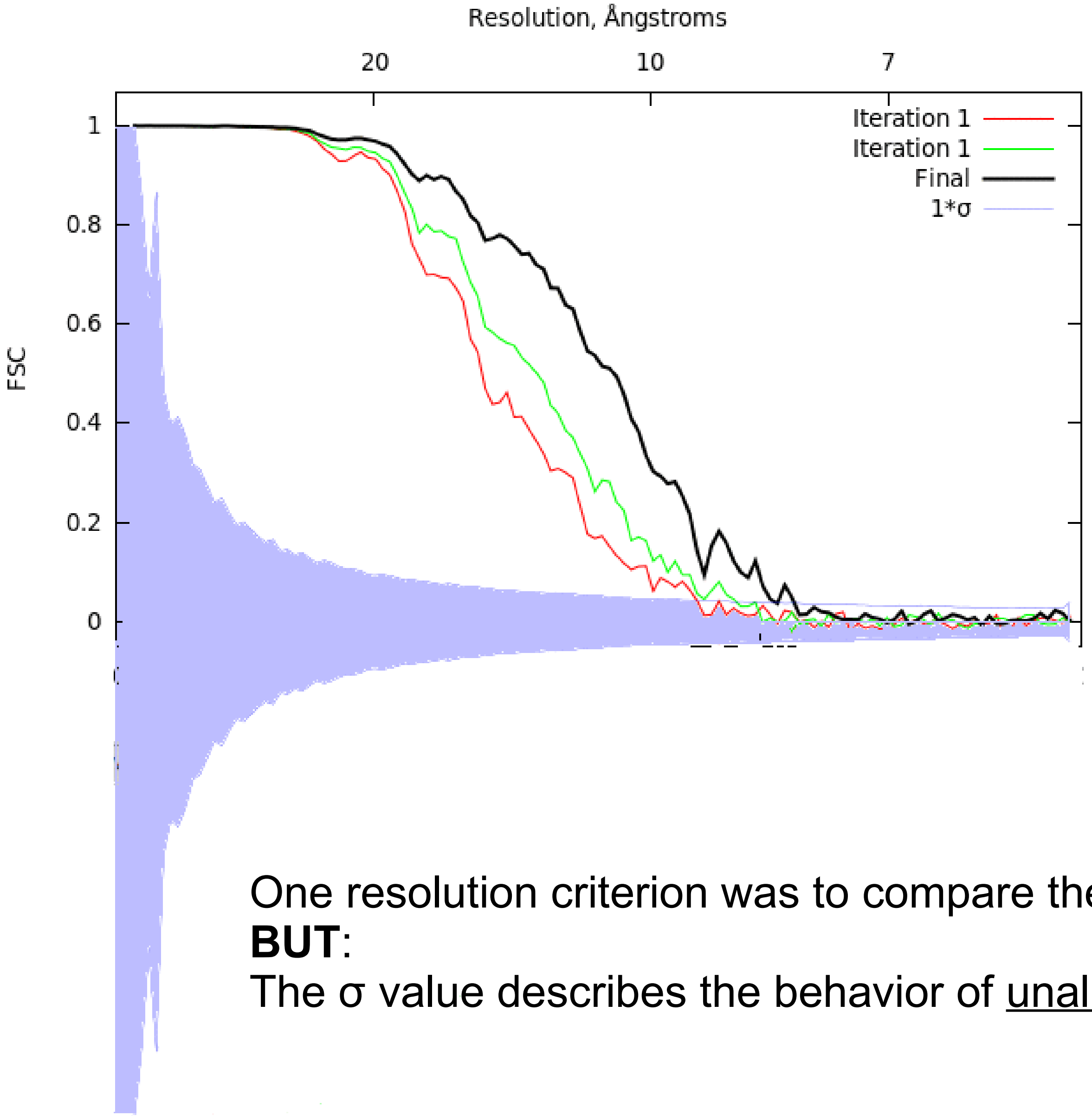
Why does σ vary with spatial frequency?





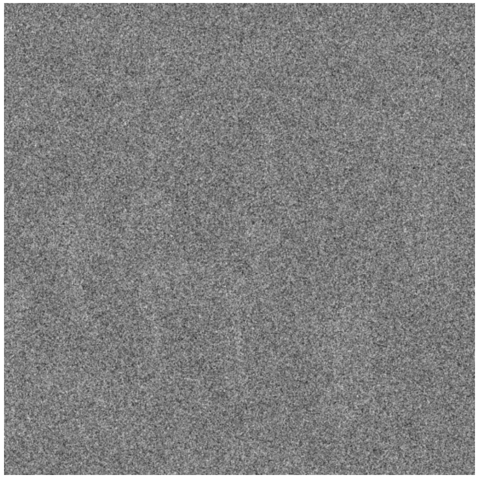
CEITEC

With small N, behavior is more unpredictable

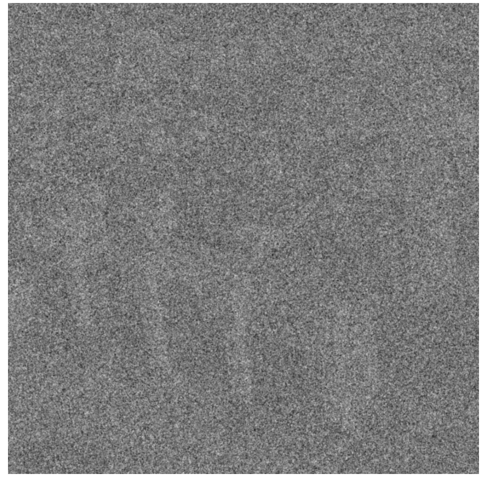


One resolution criterion was to compare the FSC to, say, 3*σ.
BUT:
The σ value describes the behavior of <u>unaligned</u> noise.

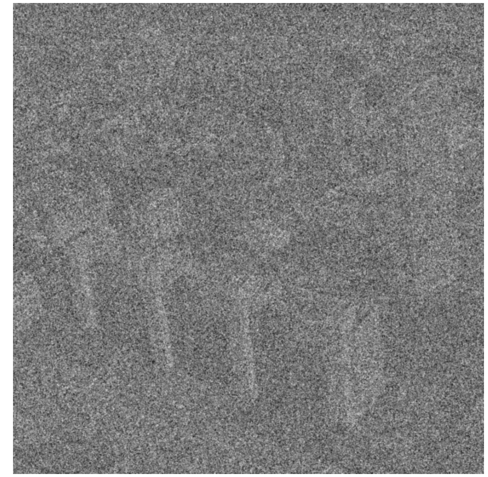
Review: model bias



N = 128

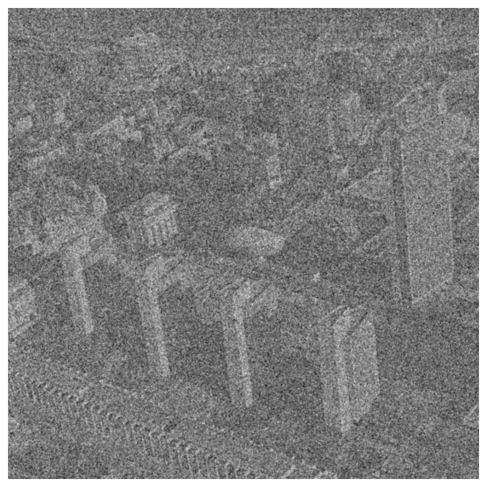


N = 256



N = 512

N = 1024

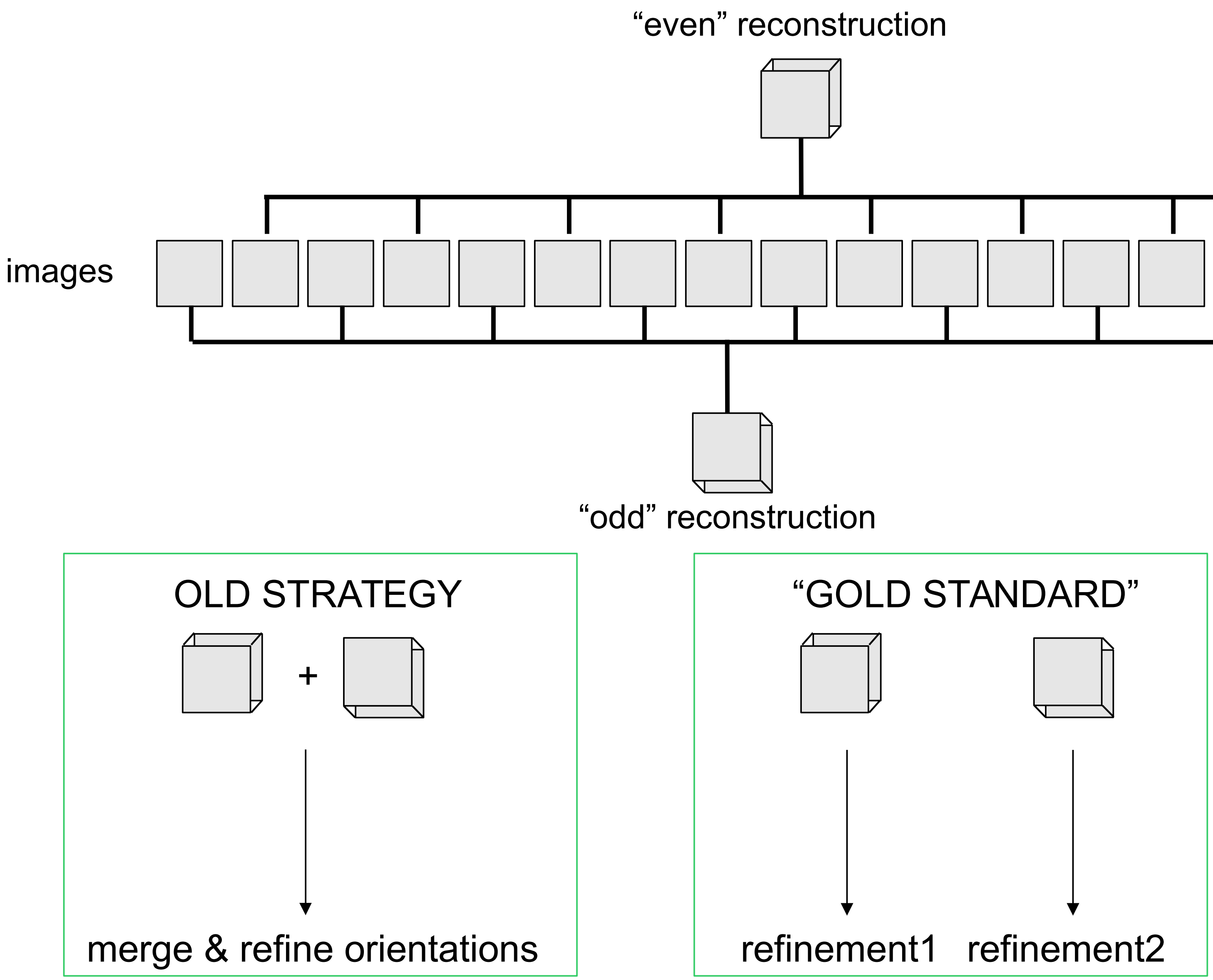


N = 2048

original

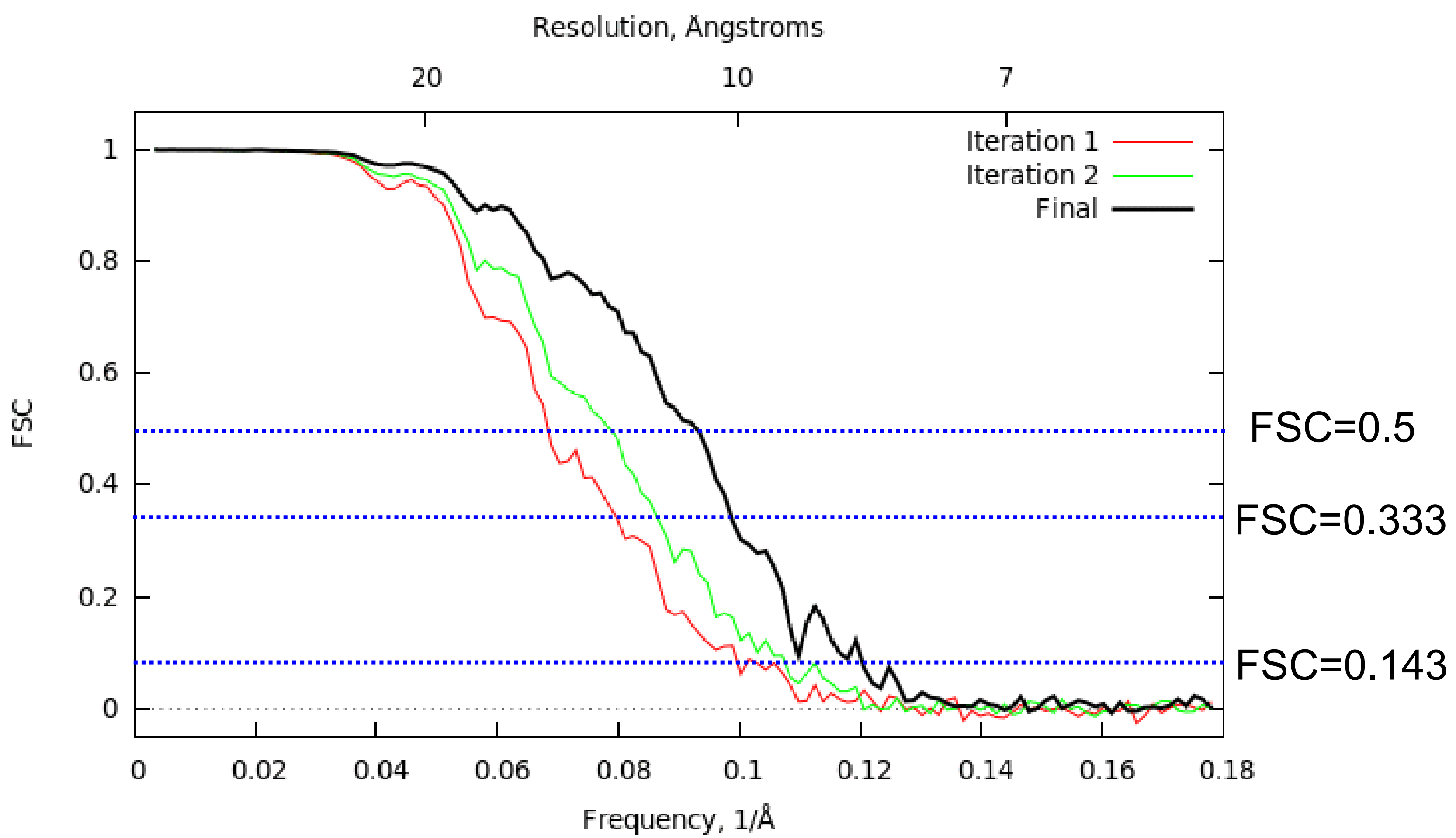
The model bias can yields false correlations in real space
is equivalent to false correlations in Fourier space.

Refinement: classical and "gold standard"



CEITEC

Different resolution criteria



Outline

Image analysis III

- More on last week's material
  - Dependence of SNR on √N
  - Oversampling
- Classification

3D Reconstruction

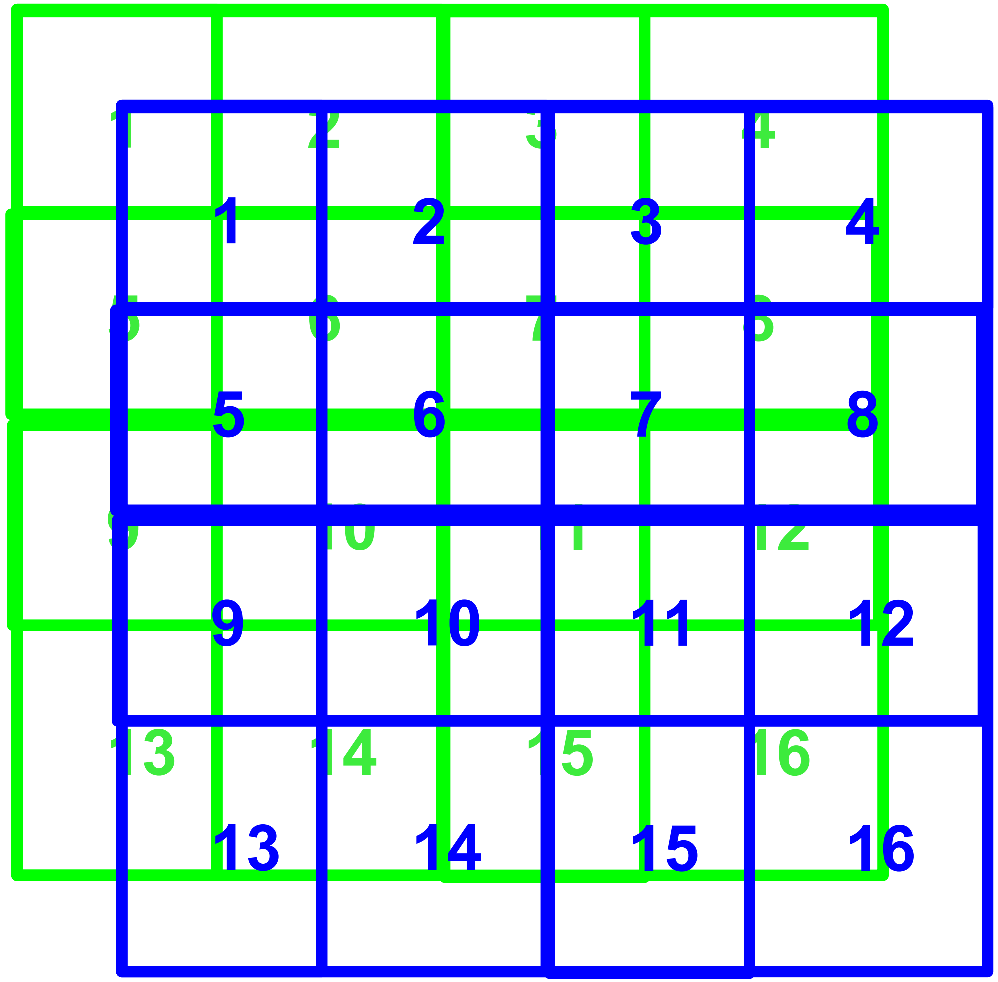
- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

CEITEC

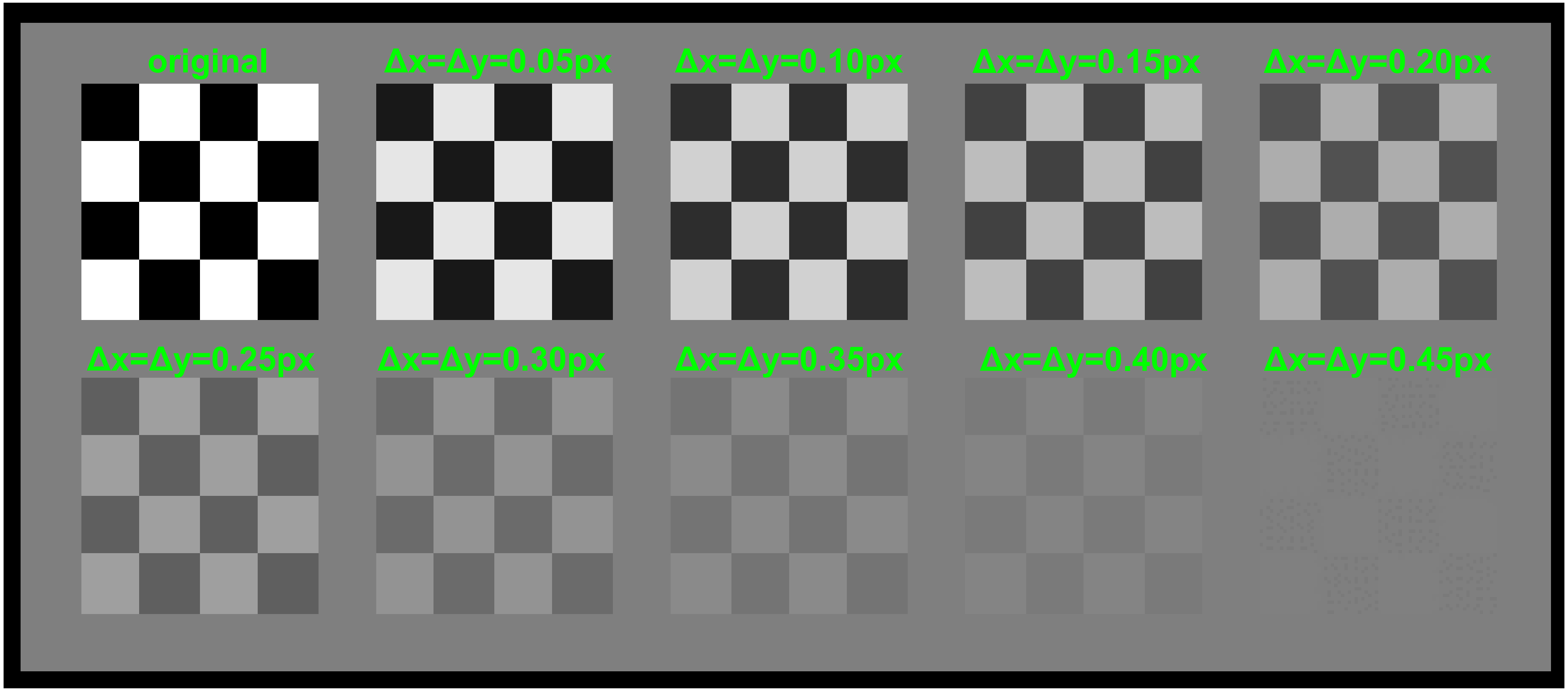
Sampling:

Oversampling an already-sampled image

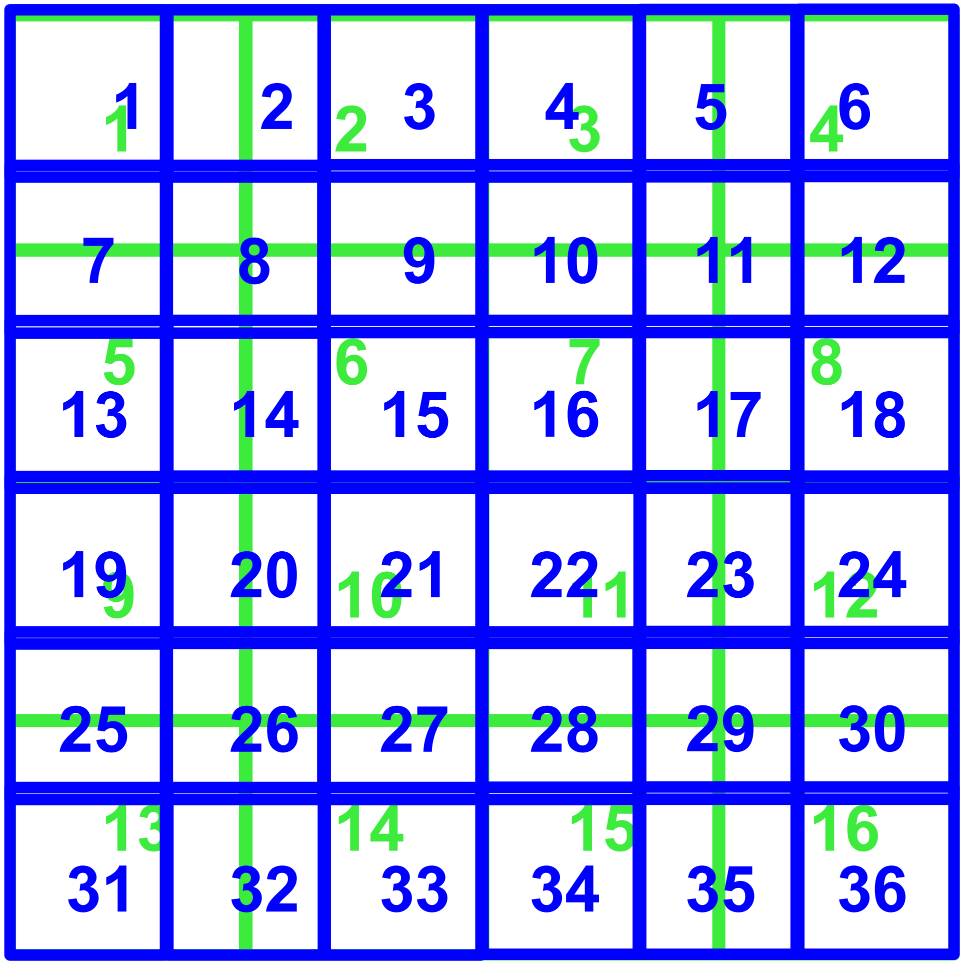
Shifts: worst-case scenario



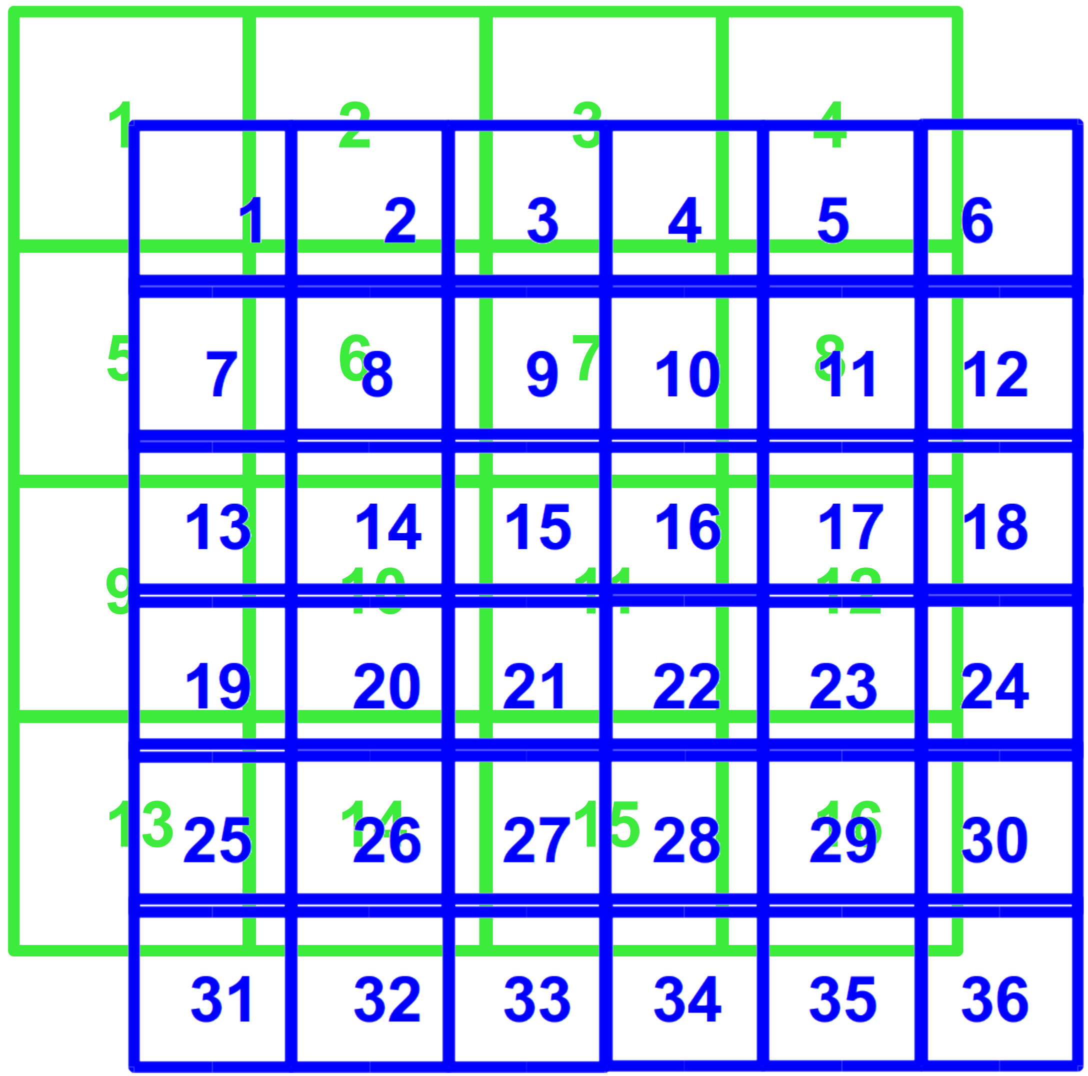
Effect of shifts

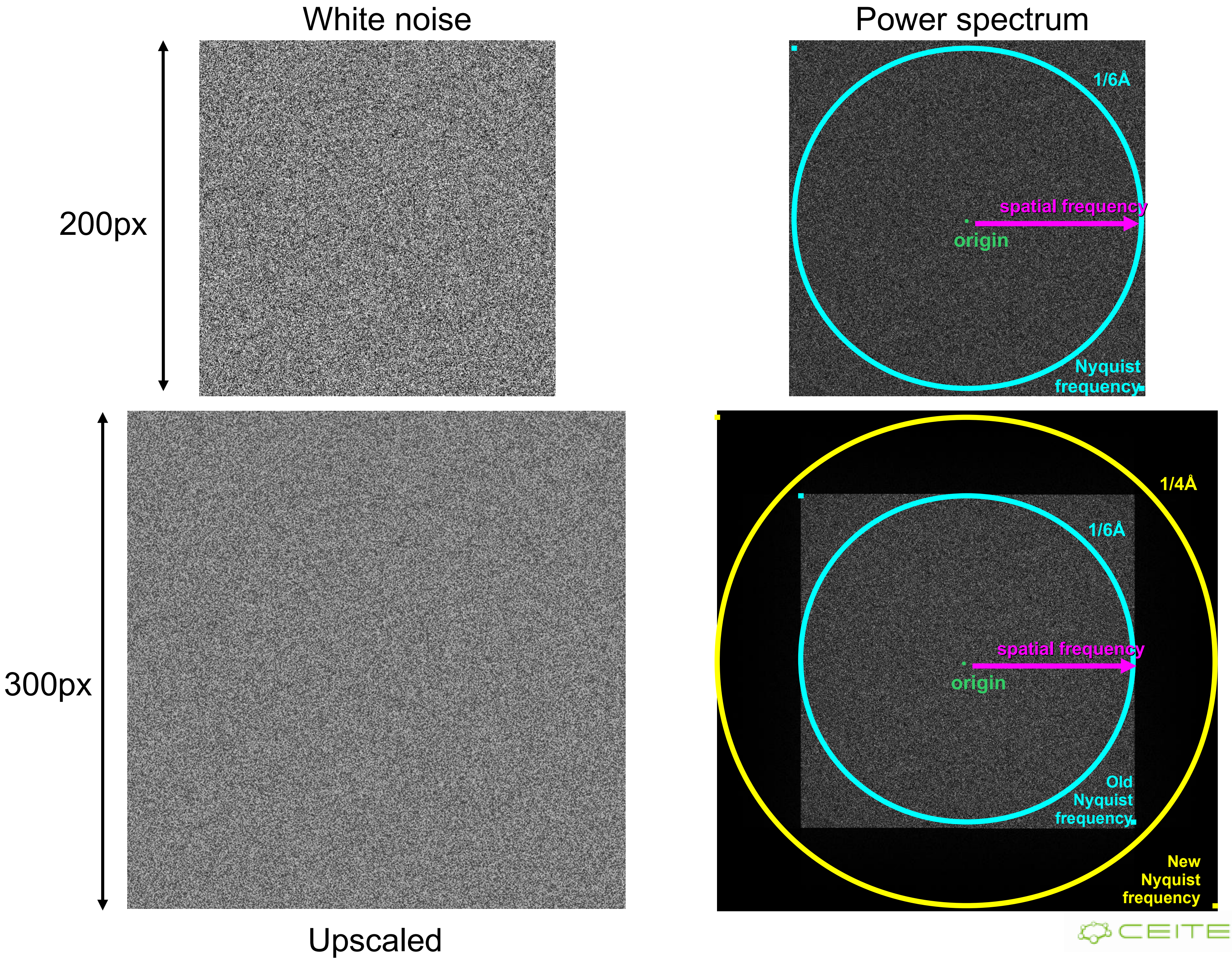


Oversampling



Worst-case scenario after oversampling




White noise
Power spectrum
200px
1/6Å
spatial frequency
origin
Nyquist
frequency
300px
1/4Å
1/6Å
spatial frequency
origin
Old
Nyquist
frequency
New
Nyquist
frequency
Upscaled

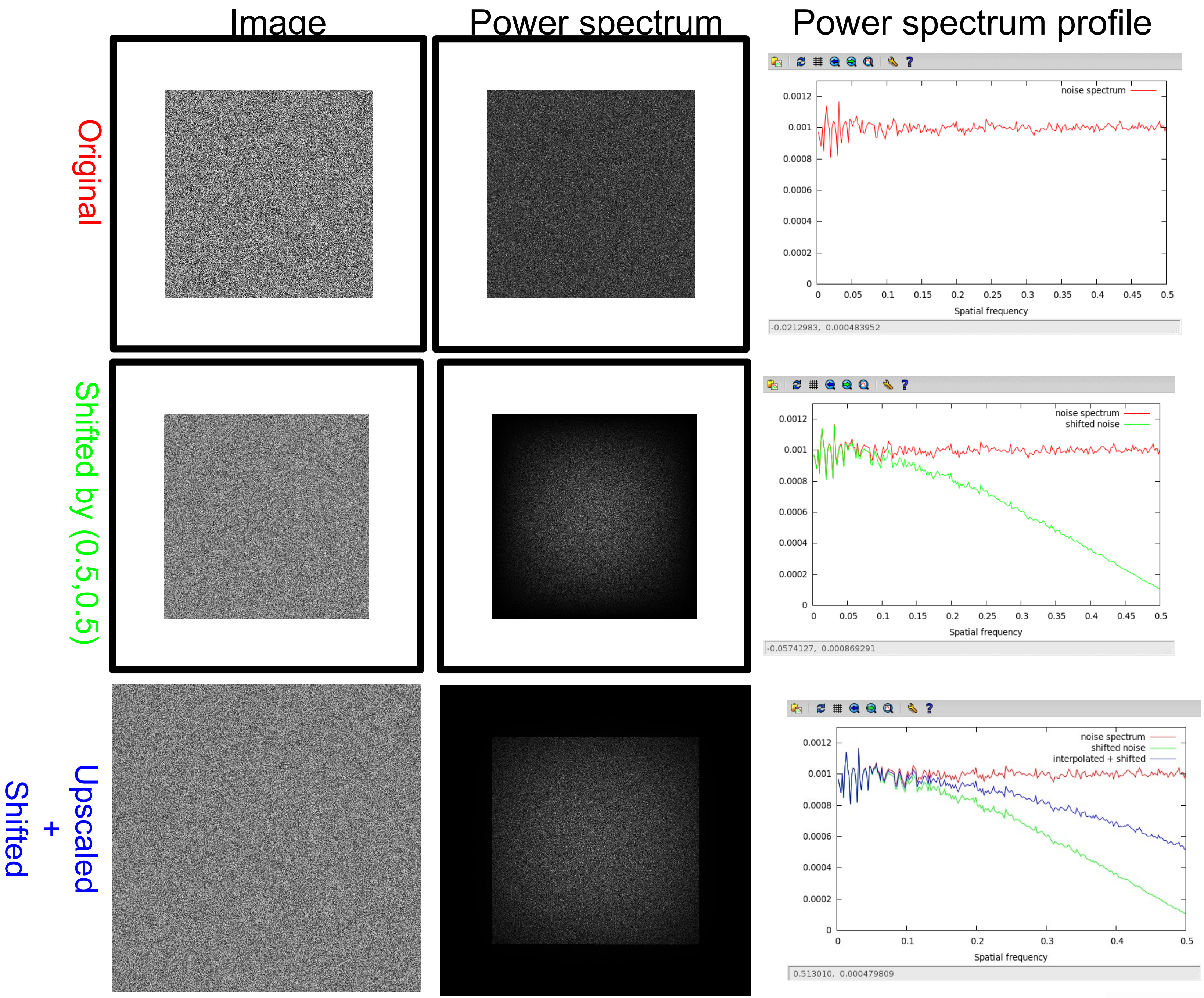
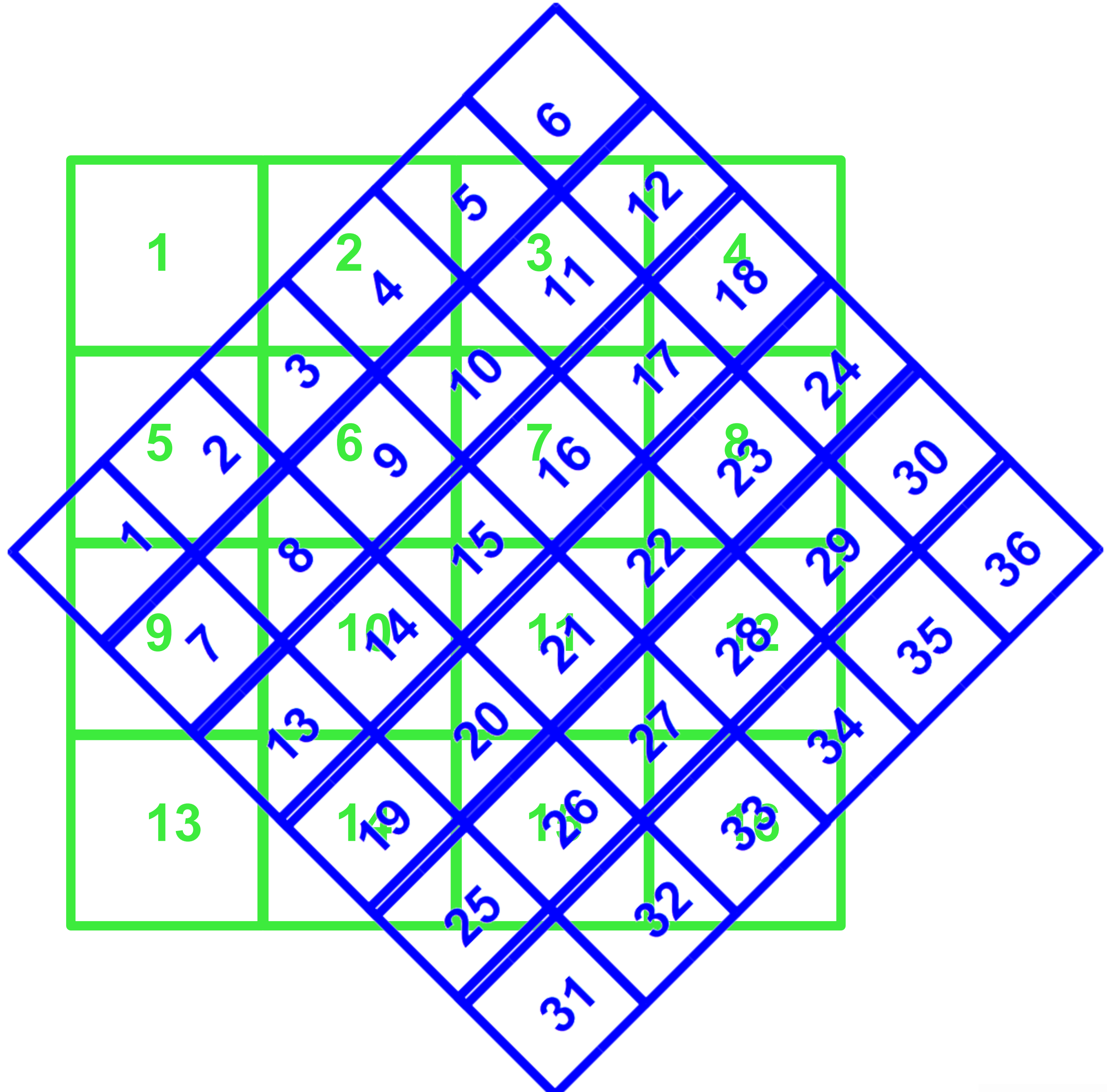


Image
Power spectrum
Power spectrum profile
Original
Shifted by (0.5,0.5)
Upscaled + Shifted
noise spectrum
shifted noise
interpolated + shifted
Spatial frequency
-0.0212983, 0.000483952
-0.0574127, 0.000869291
0.513010, 0.000479809


1
2
3
4
5
6
7
8
9
10
11
12
13
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

CEITEC

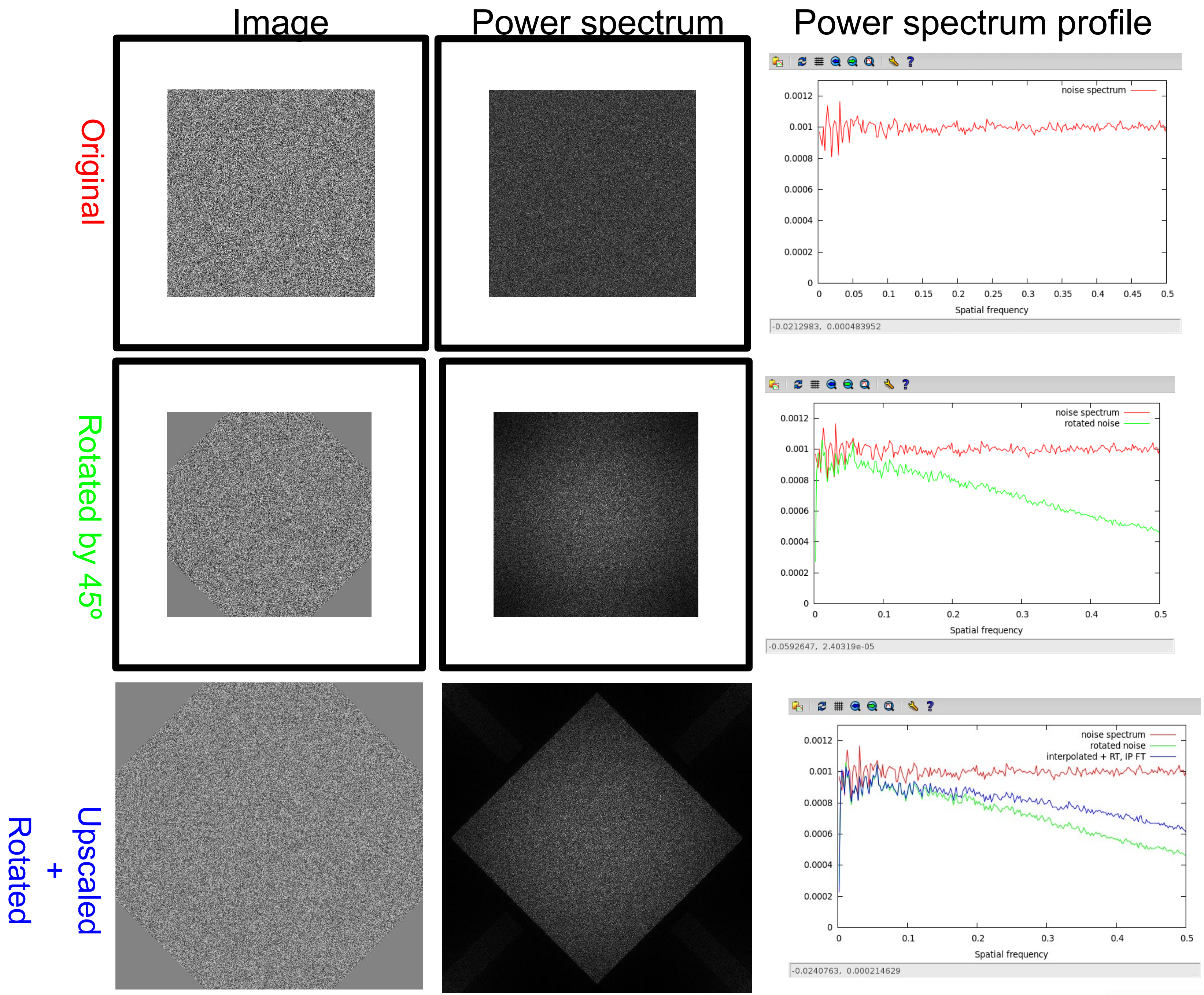


Image
Power spectrum
Power spectrum profile
Original
Rotated by 45°
Upscaled
+
Rotated
noise spectrum
rotated noise
interpolated + RT, IP FT
Spatial frequency
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
-0.0212983, 0.000483952
-0.0592647, 2.40319e-05
-0.0240763, 0.000214629

Oversampling: Conclusion

You can do a little better by oversampling.

Bammes... Chiu (2012) J. Struct. Biol.

Outline

Image analysis III

- More on last week's material
    - Dependence of SNR on √N
    - Oversampling
- Classification

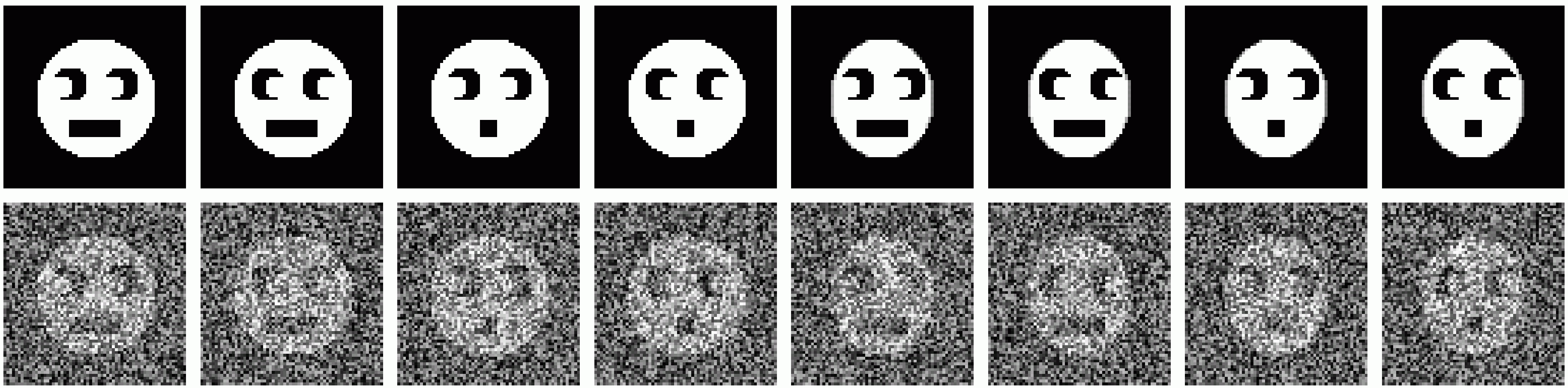
3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

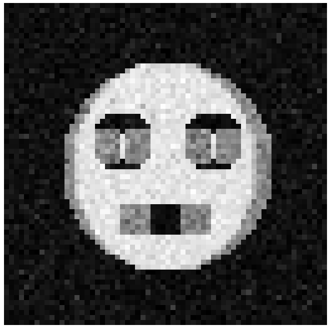
Classification

Reiteration of the problem

8 classes of faces, 64x64 pixels



With noise added

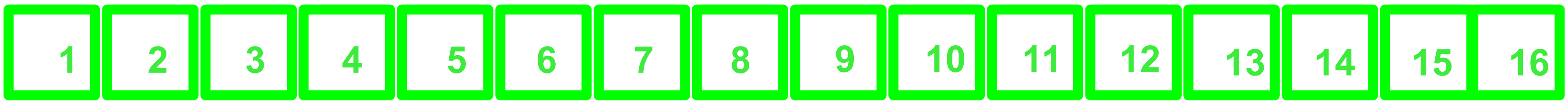
Average: 

Before we can average the data, we first should find homogeneous subsets.

Multivariate data analysis (MDA)

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

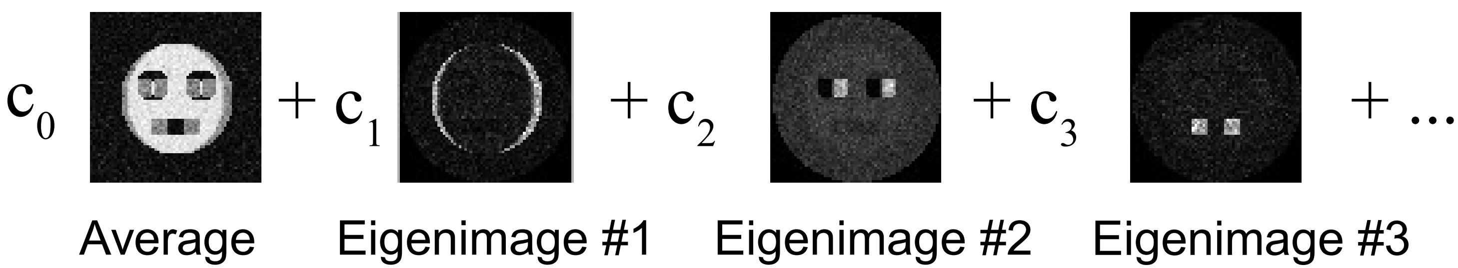
Multivariate data analysis (MDA), or Multivariate statistical analysis (MSA)

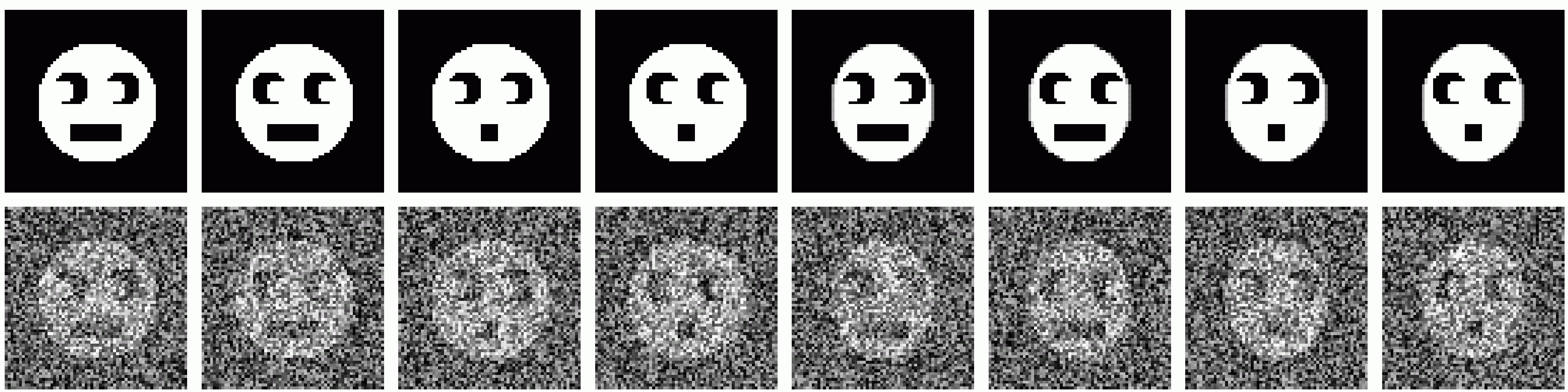


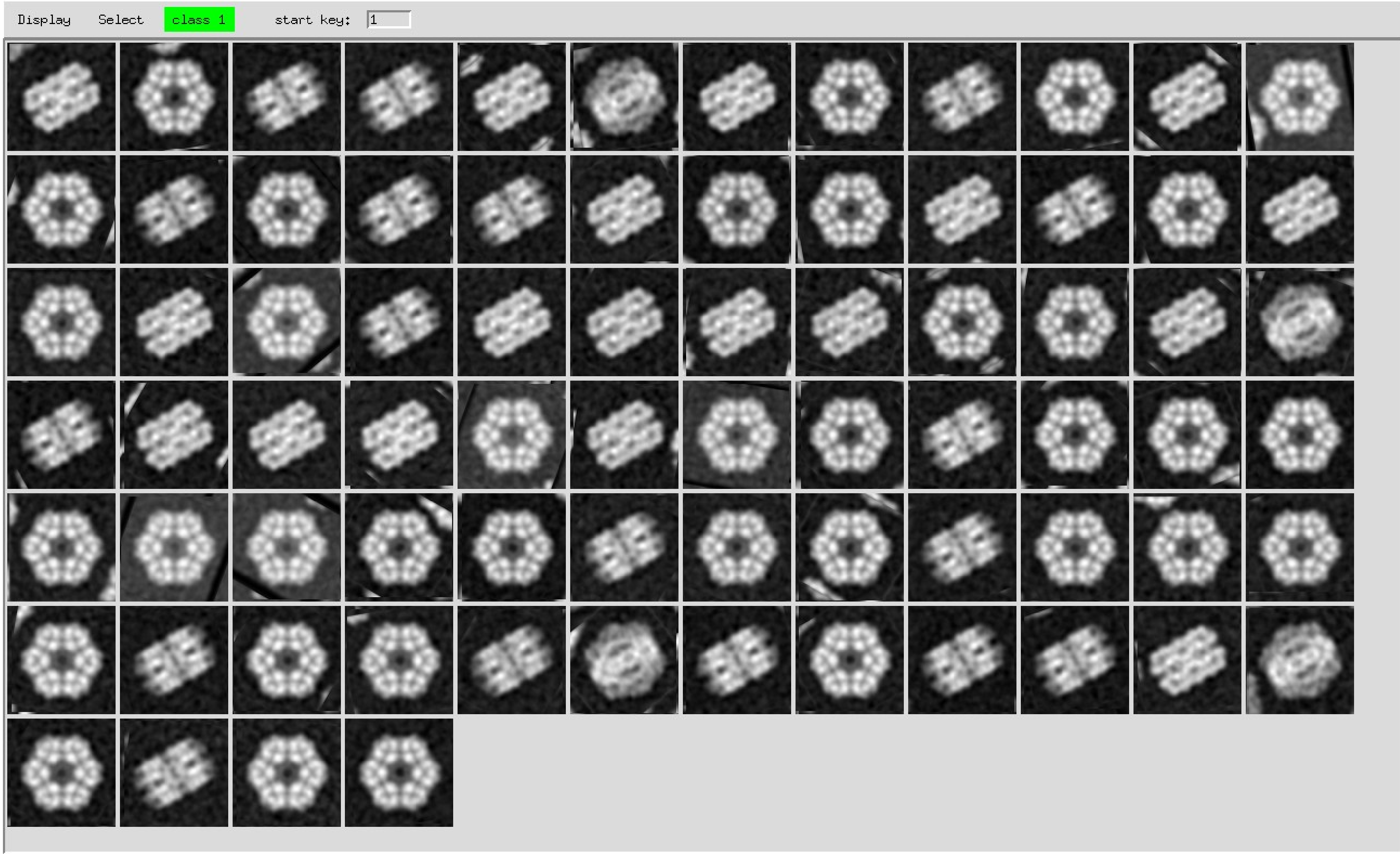
Our 16-pixel image can be reorganized into a 16-coordinate vector.

MDA: Reconstituted images

Linear combinations of these images will give us approximations of the images that make up the data.



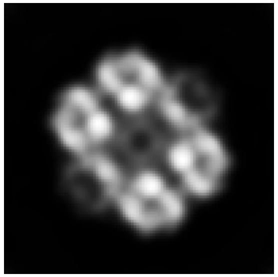


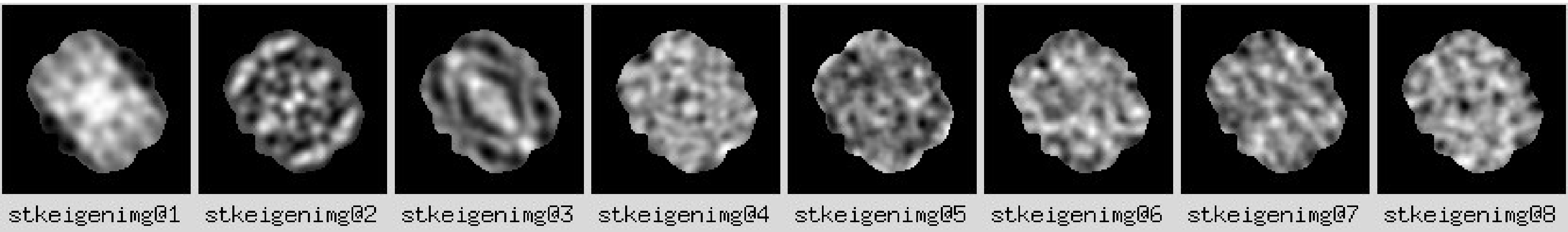


Phantom images of worm hemoglobin

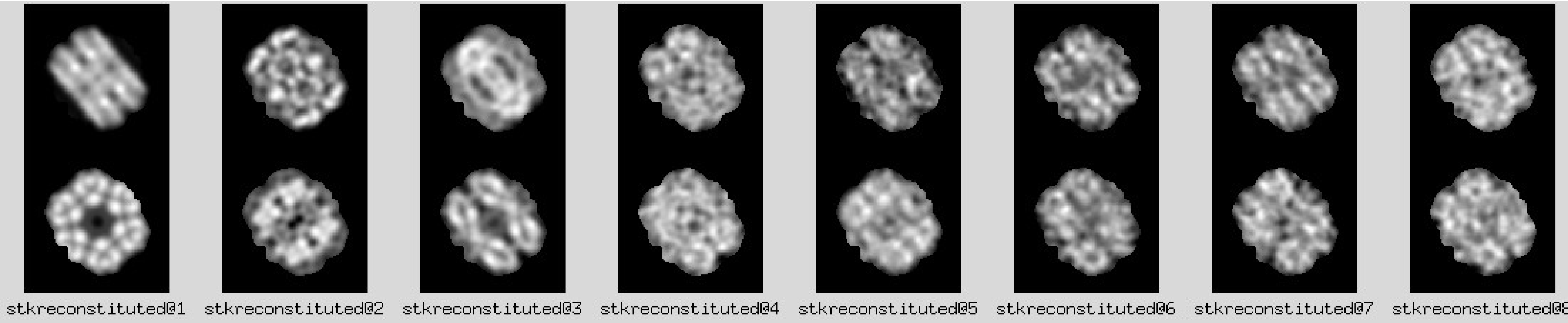
MDA of worm hemoglobin

Average:



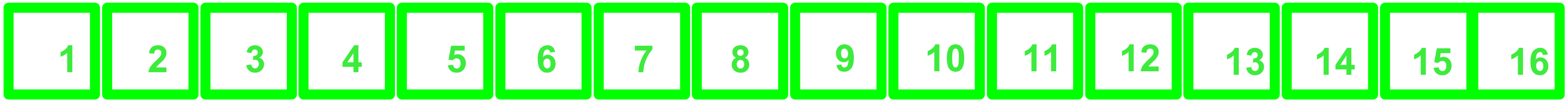


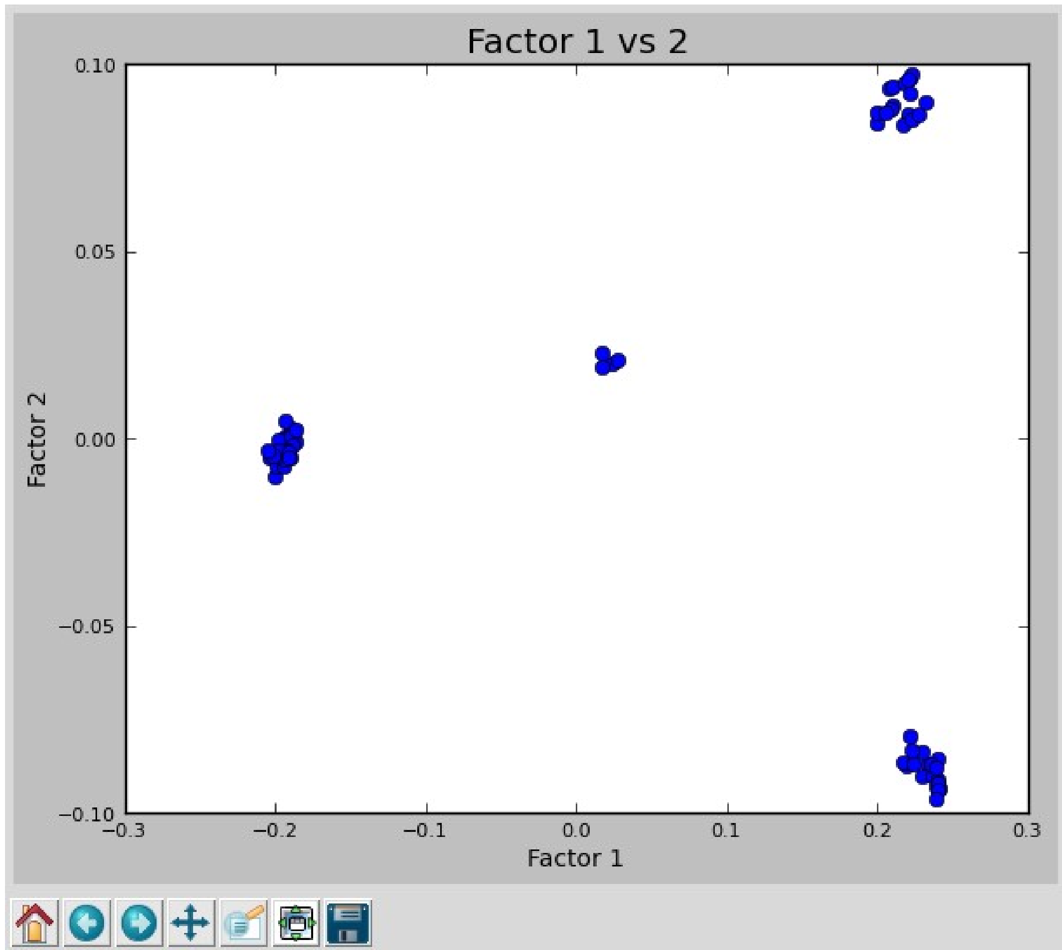
$+c_0$ $+c_1$ $+c_2$ $+c_3$ $+c_4$ $+c_5$



$-c_0$ $-c_1$ $-c_2$ $-c_3$ $-c_4$ $-c_5$

Classification

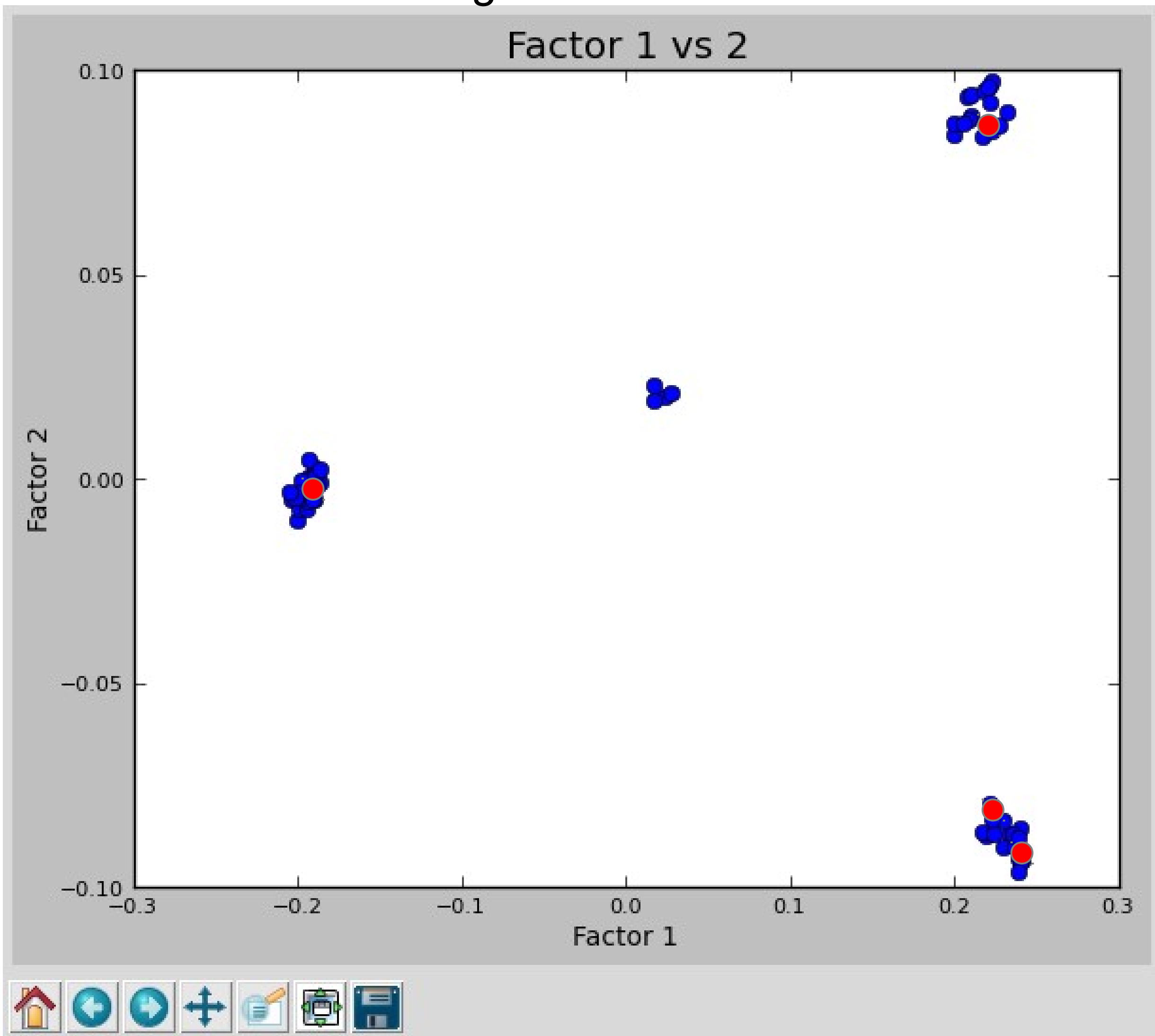




How do we categorize/classify the images?

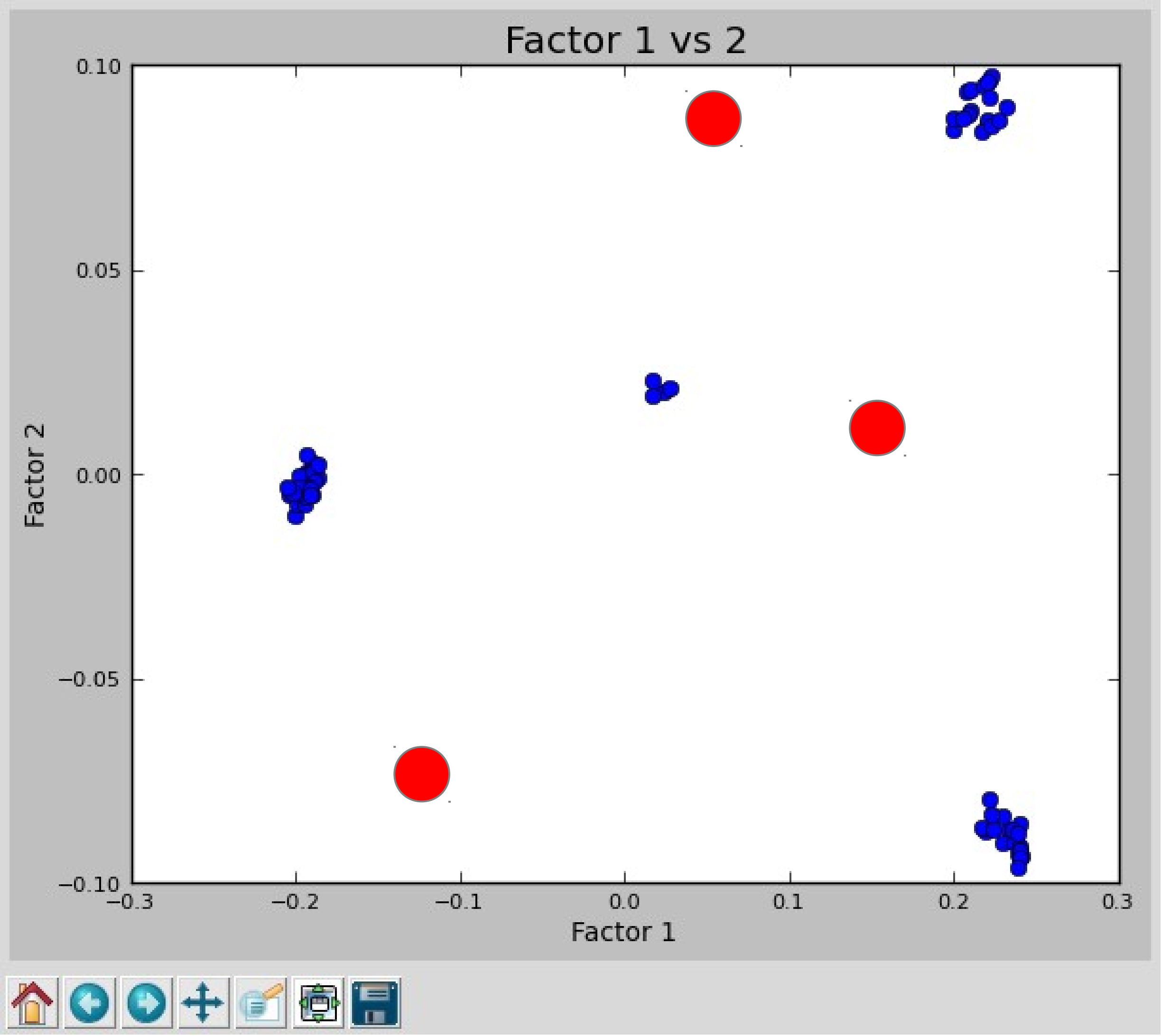
K-means classification

A number *K* of images are chosen as seeds.

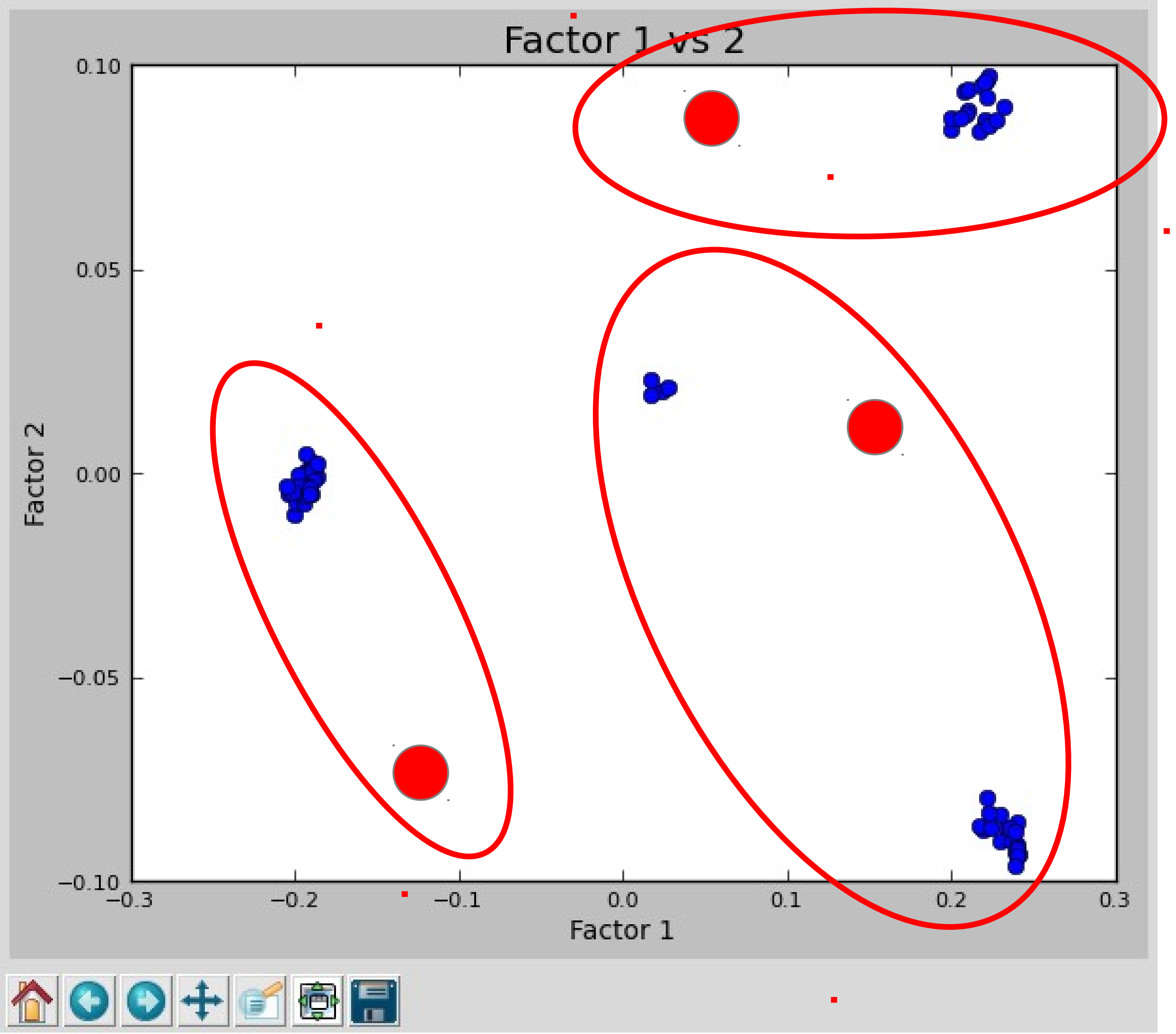


BAD: Some clusters may be overrepresented/underrepresented.

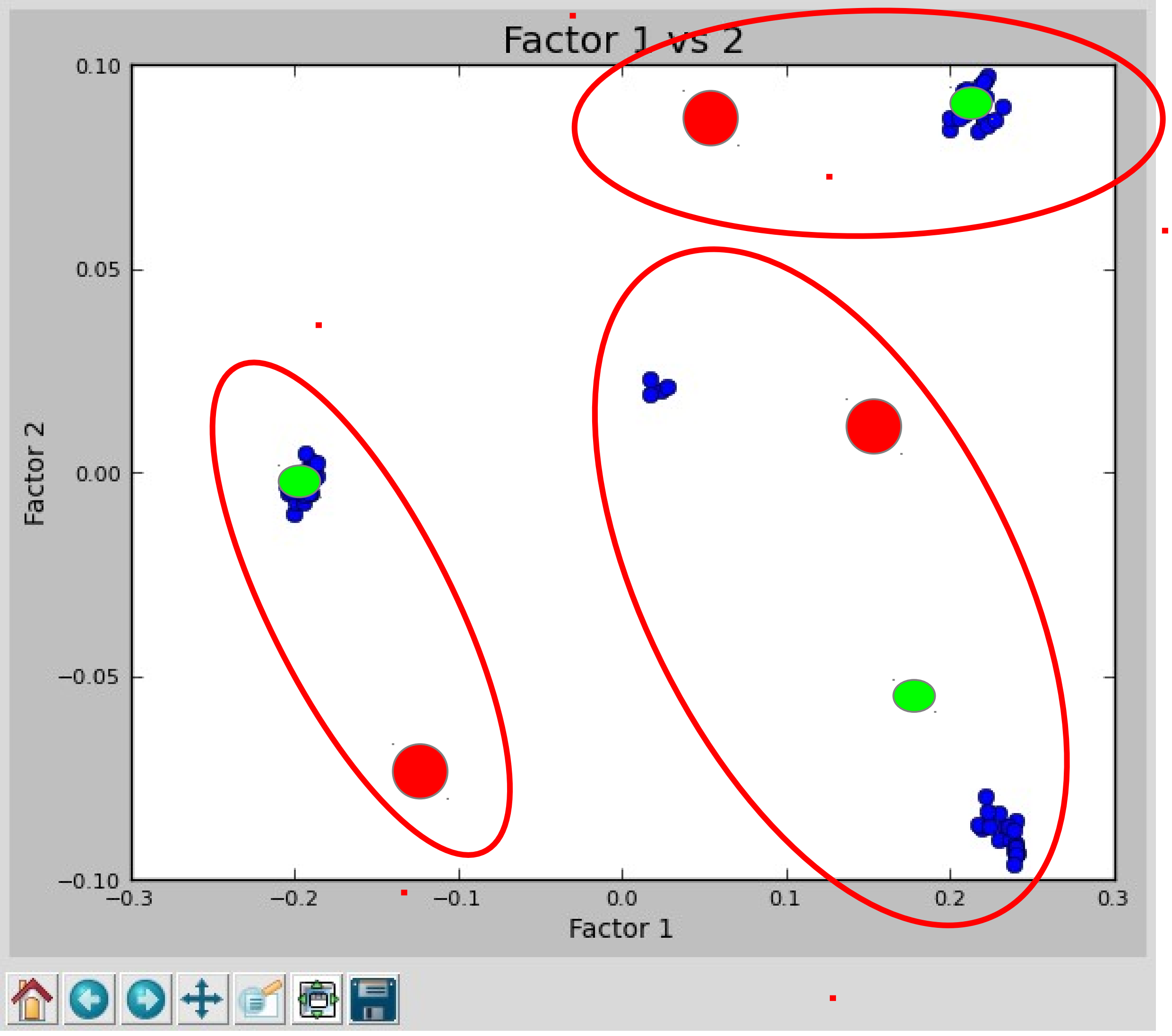
Diday's method of moving centers



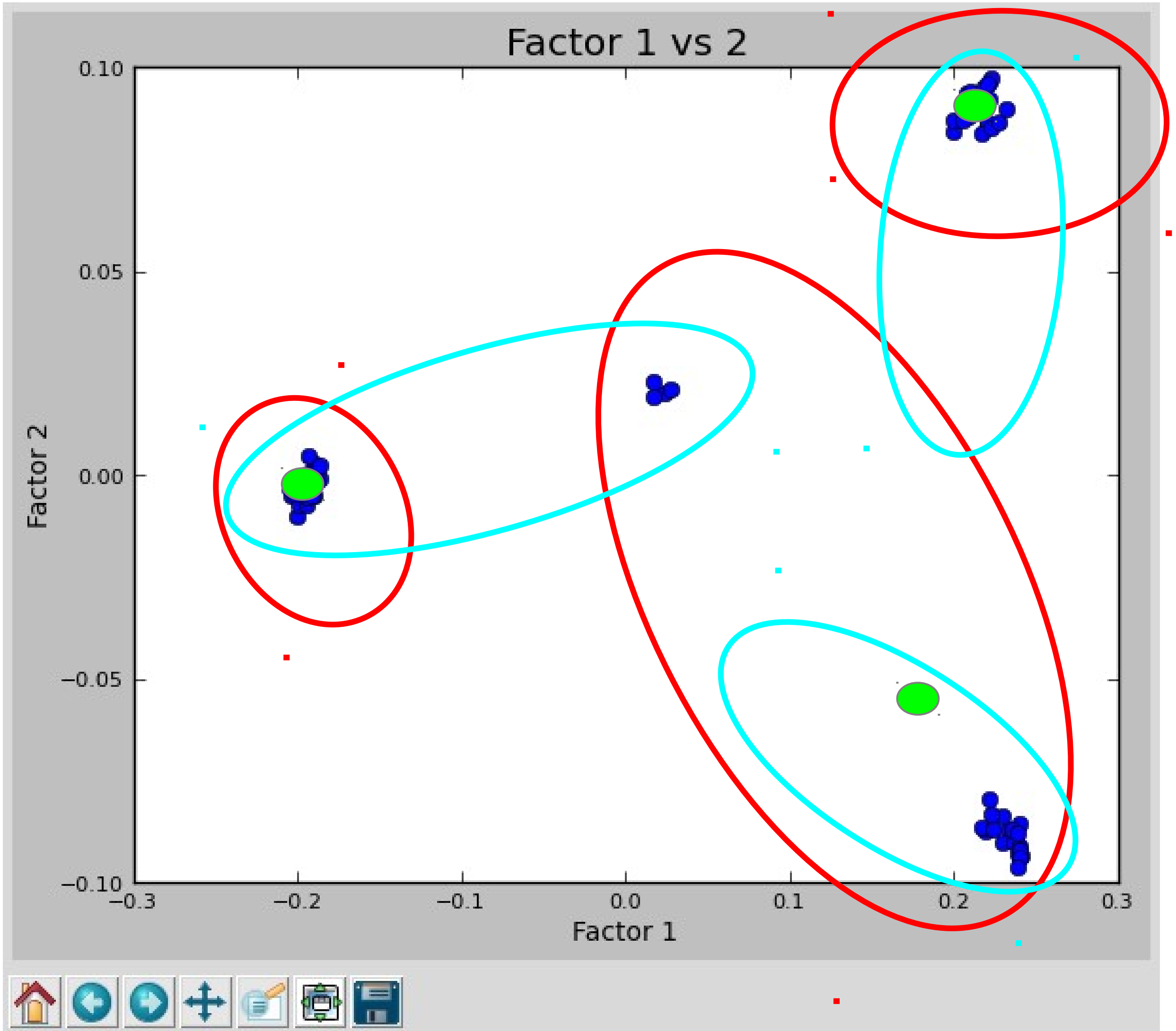
Diday's method of moving centers



Diday's method of moving centers

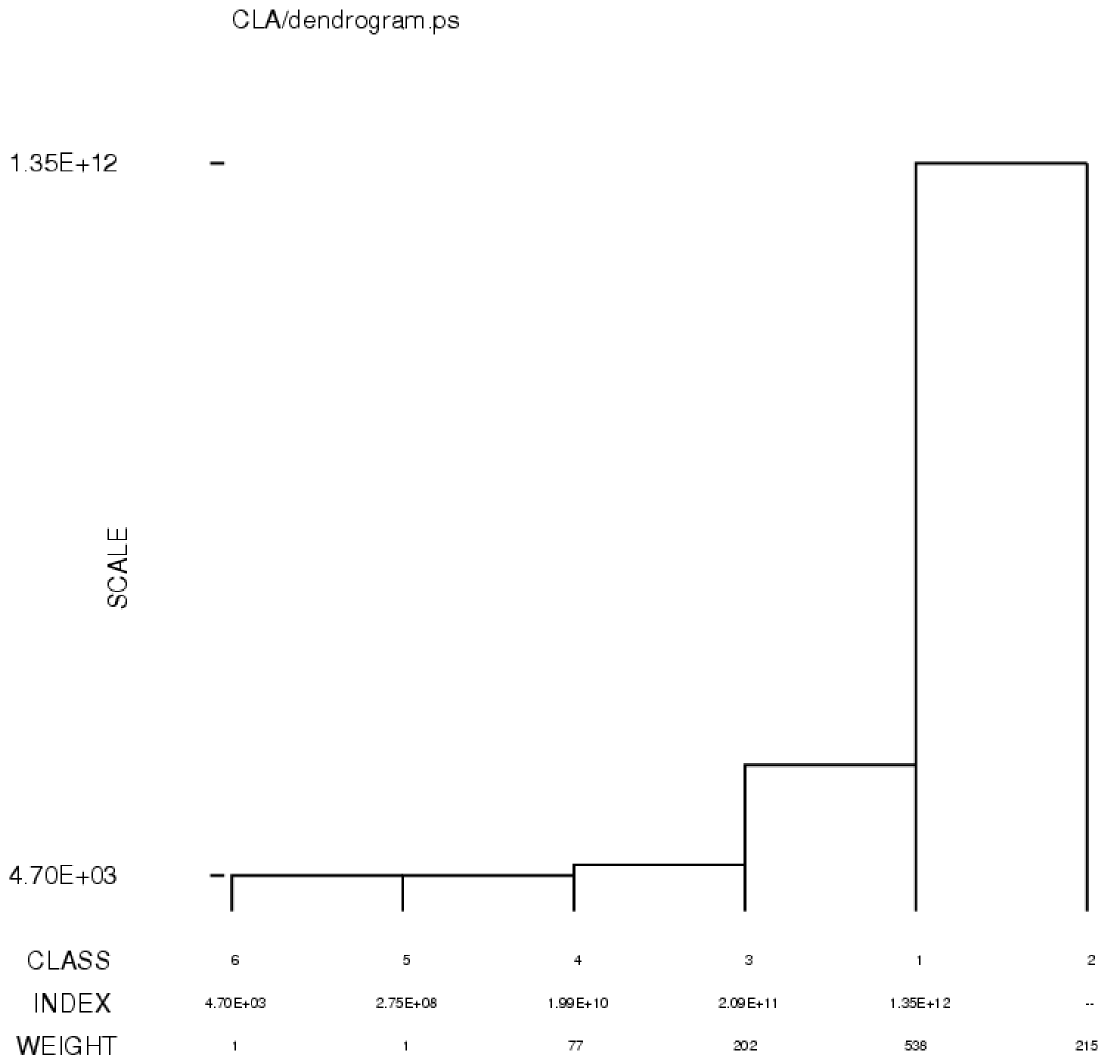


Diday's method of moving centers

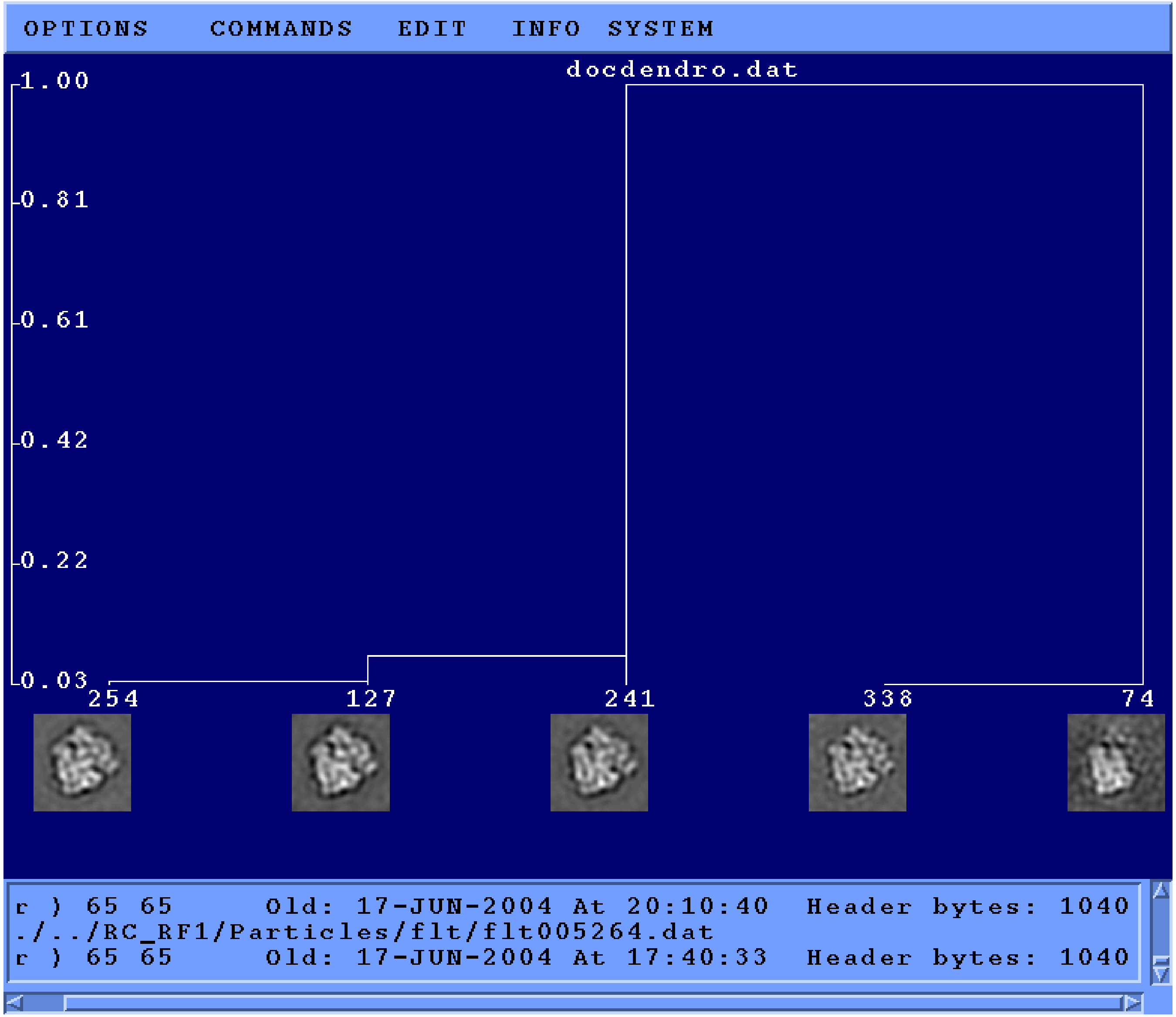


We will note the images that always “travel” together, and will call them a class.

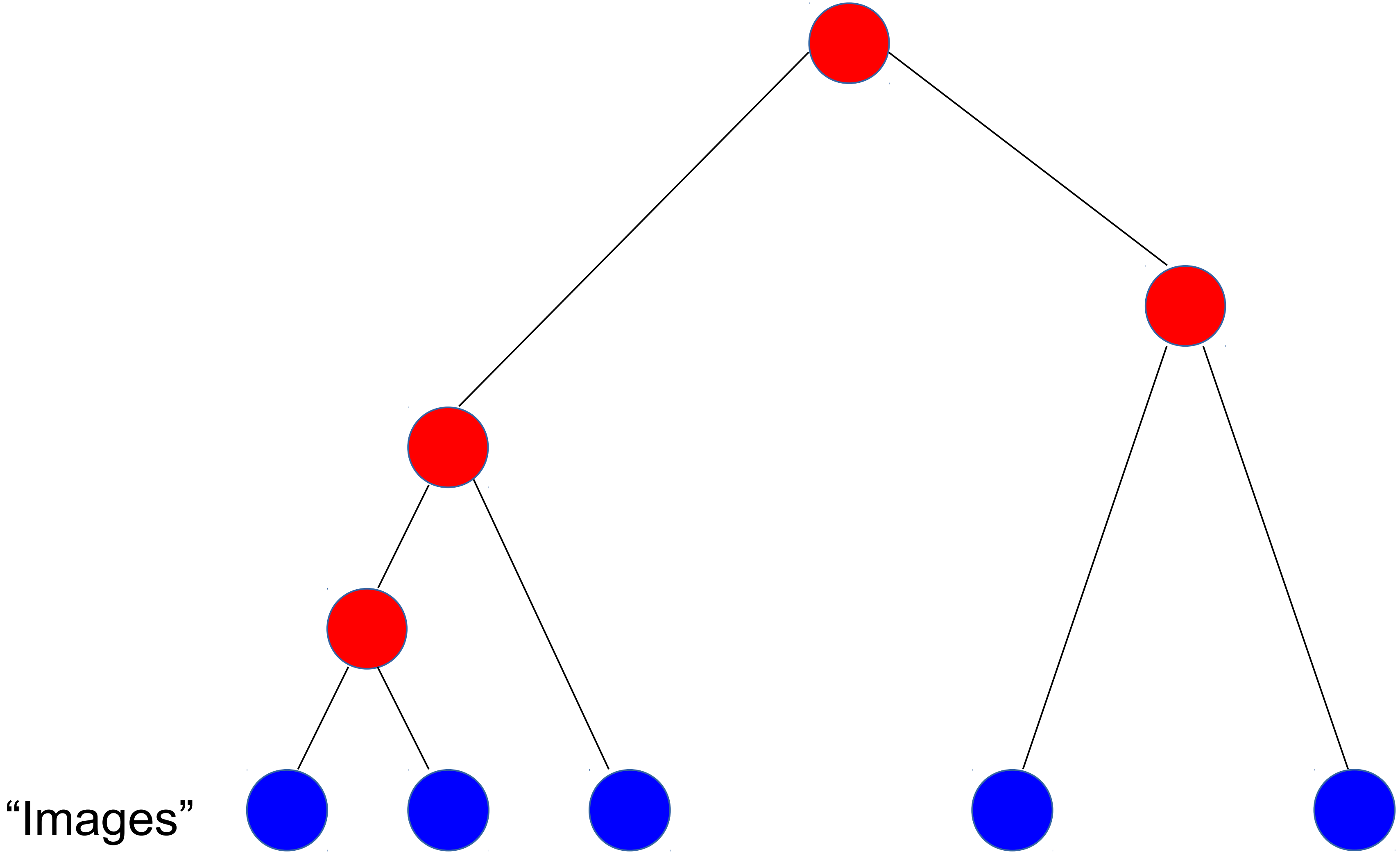
Dendrogram


CLA/dendrogram.ps
1.35E+12
SCALE
4.70E+03
CLASS
6
5
4
3
1
2
INDEX
4.70E+03
2.75E+08
1.99E+10
2.09E+11
1.35E+12
--
WEIGHT
1
1
77
202
538
215

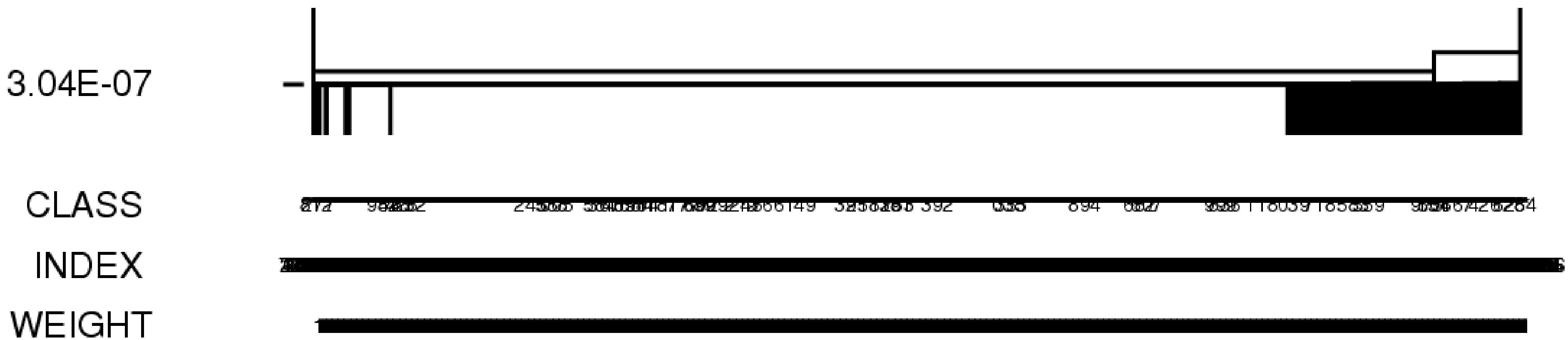
Dendrogram


OPTIONS
COMMANDS
EDIT
INFO
SYSTEM
docdendro.dat
1.00
0.81
0.61
0.42
0.22
0.03
254
127
241
338
74
r) 65 65 Old: 17-JUN-2004 At 20:10:40 Header bytes: 1040
./../RC_RF1/Particles/flt/flt005264.dat
r) 65 65 Old: 17-JUN-2004 At 17:40:33 Header bytes: 1040

Hierarchical ascendant classification

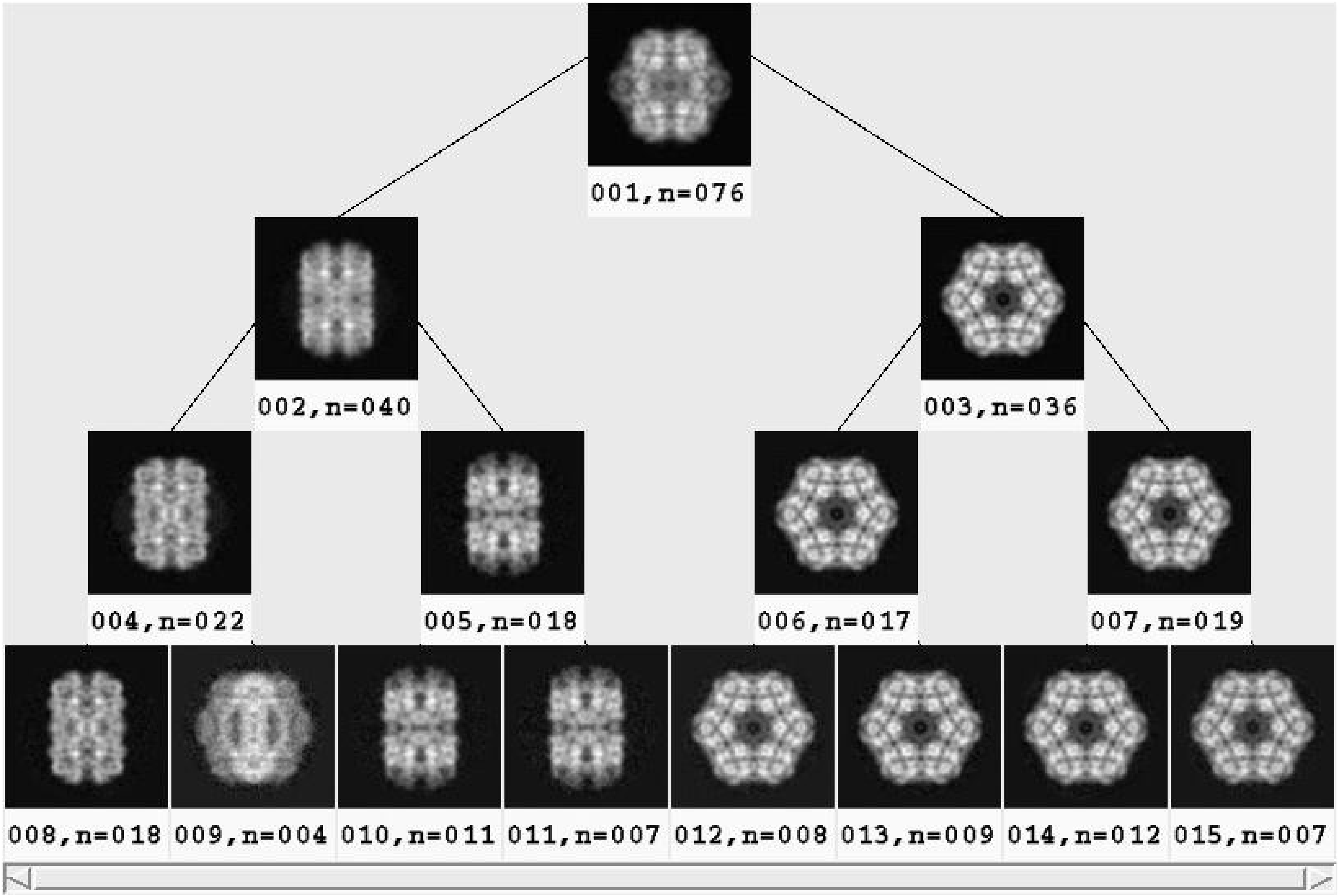


Hierarchical Ascendant Classification



All images are represented.
The dendrogram will be too heavily branched to interpret without truncation.

Binary-tree viewer



BAD: Information about the height of the branch is lost.

Outline

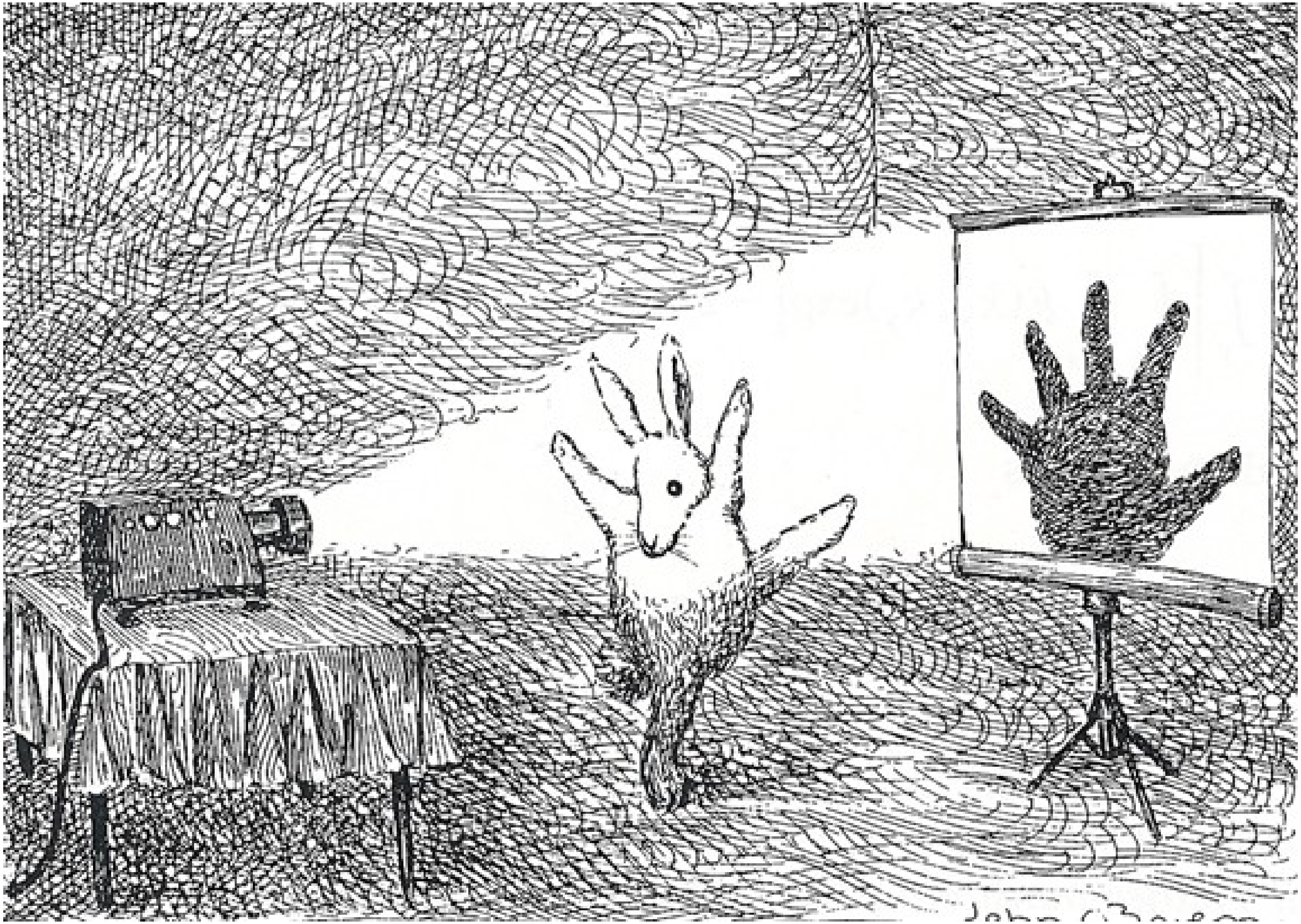
Image analysis III

- More on last week's material
    - Dependence of SNR on √N
    - Oversampling
- Classification

3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

How do you go from 2D to 3D?



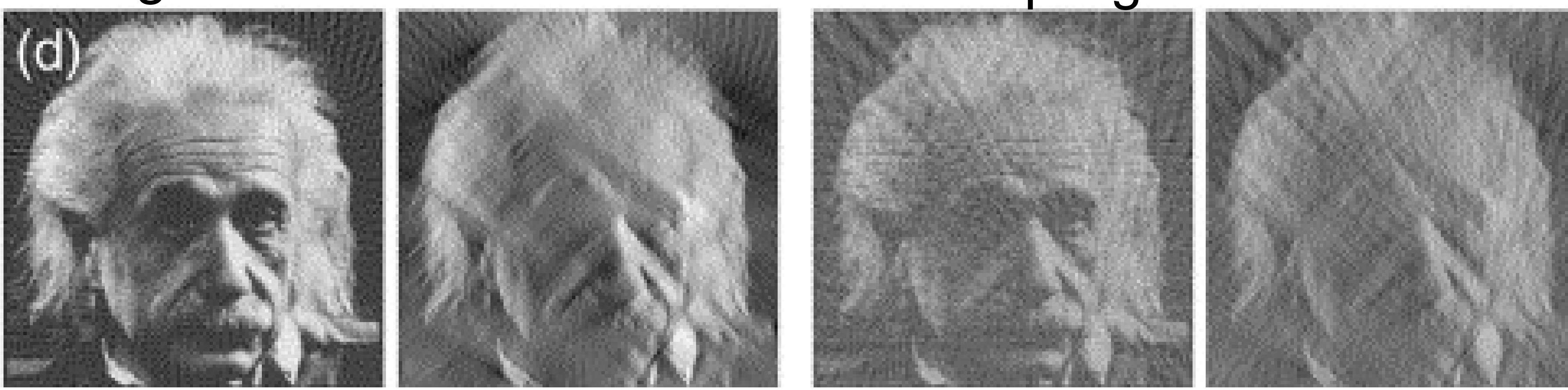
John O'Brien, 1991, *The New Yorker*

What information do we need for 3D reconstruction?

1. different orientations
2. known orientations
3. many particles

What happens when we're missing views?

good | missing views | sparse sampling | sparse sampling + missing views



Baumeister et al. (1999), *Trends in Cell Biol.*, **9**: 81-5.

Your sample isn't guaranteed to adopt different orientations, in which case you many need to explicitly tilt the microscope stage.

(more later...)

What information do we need for 3D reconstruction?

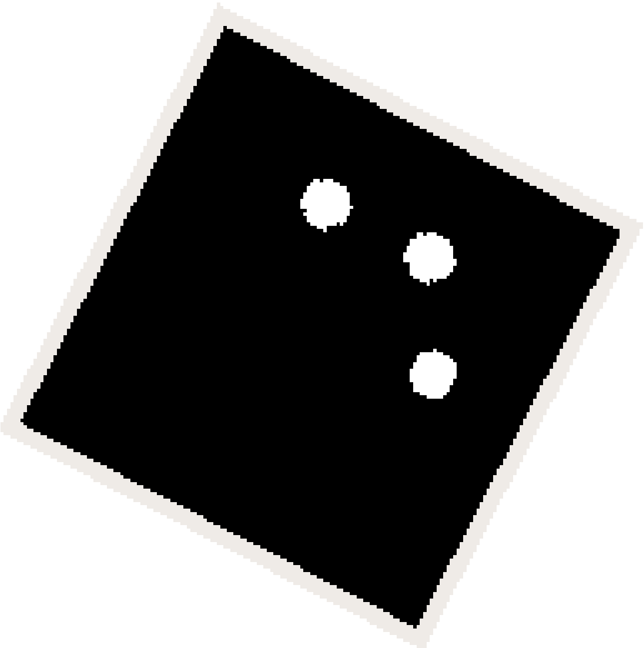
1. different orientations
2. known orientations
3. many particles

I have all of this information.
Now what?

There are two general categories of 3D reconstruction

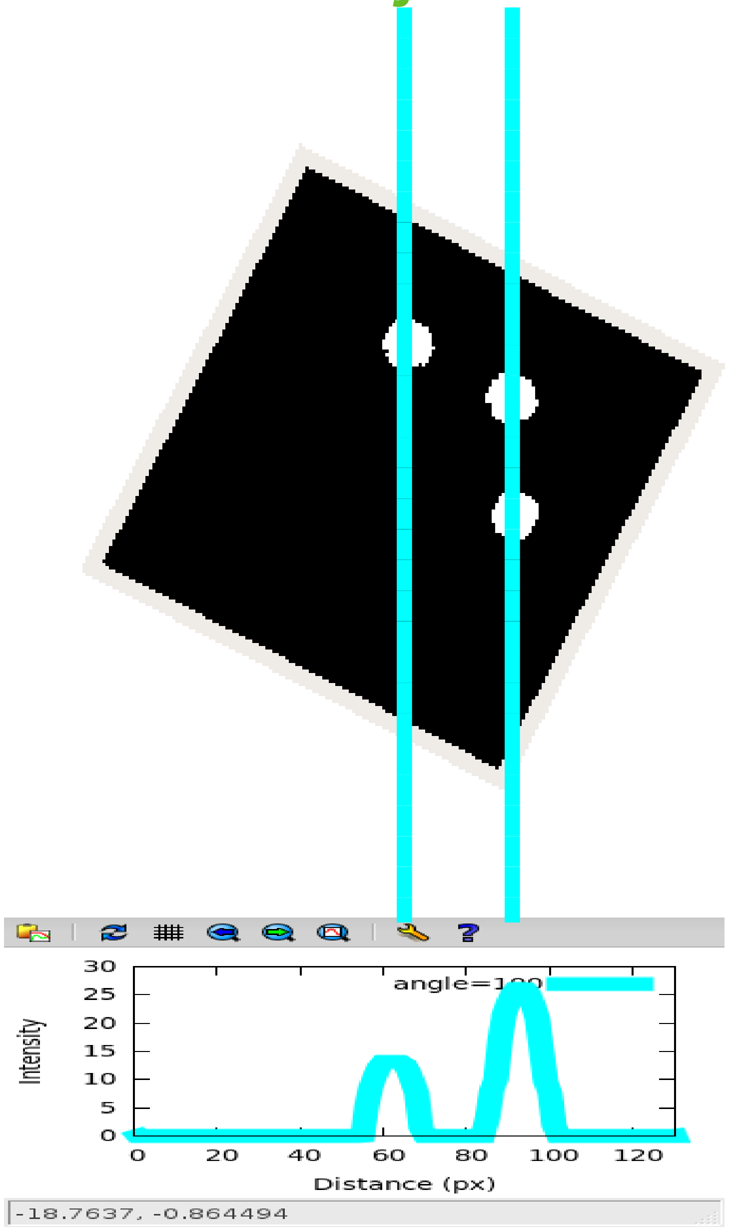
1. Real space
2. Fourier space

Reconstruction in real space

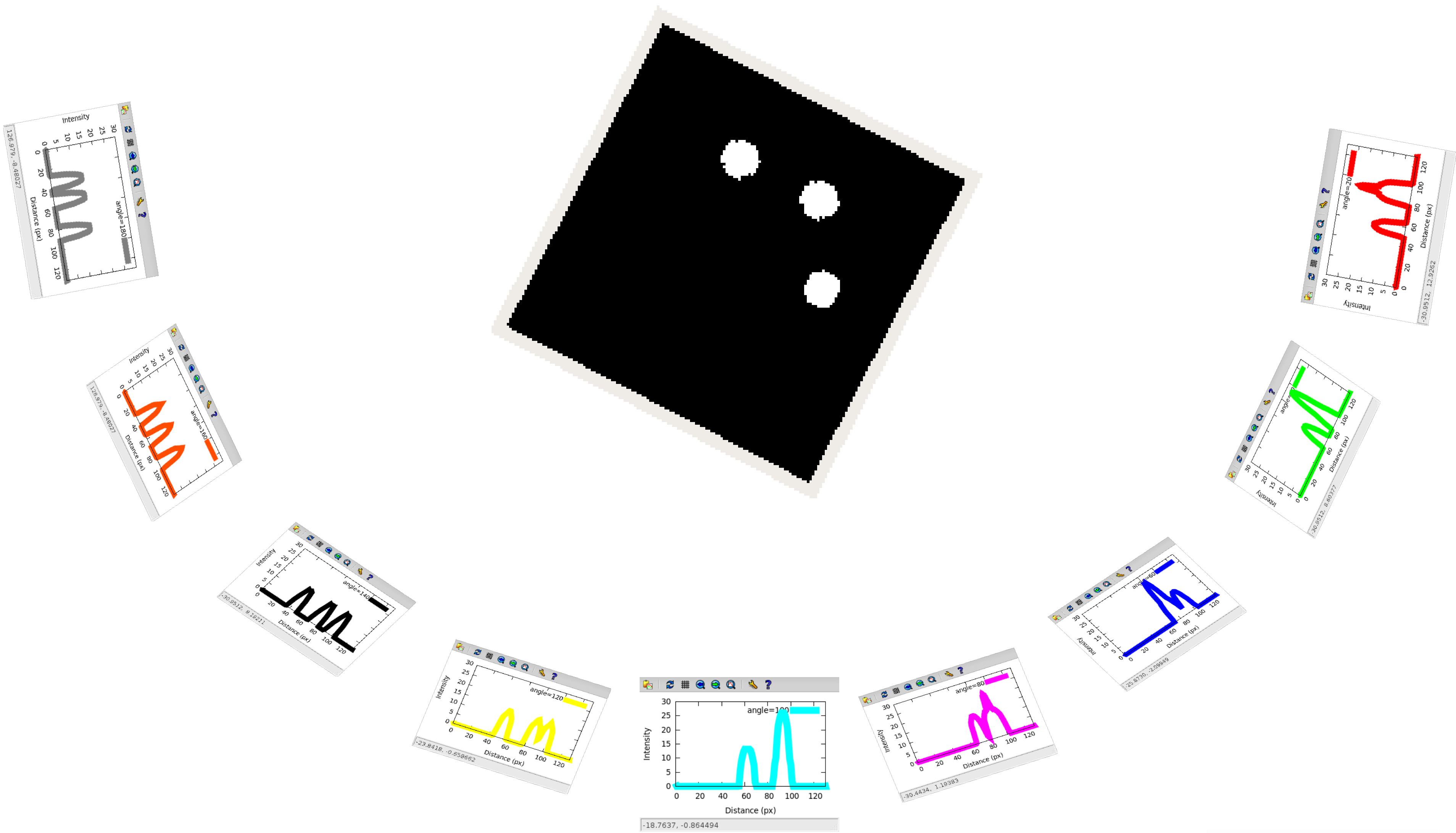


We are going to reconstruct a 2D object from 1D projections. The principle is the similar to, but simpler than, reconstructing a 3D object from 2D projections.

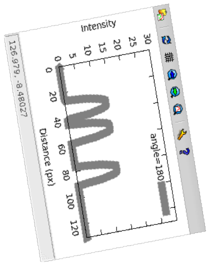
Projection of our 2D object

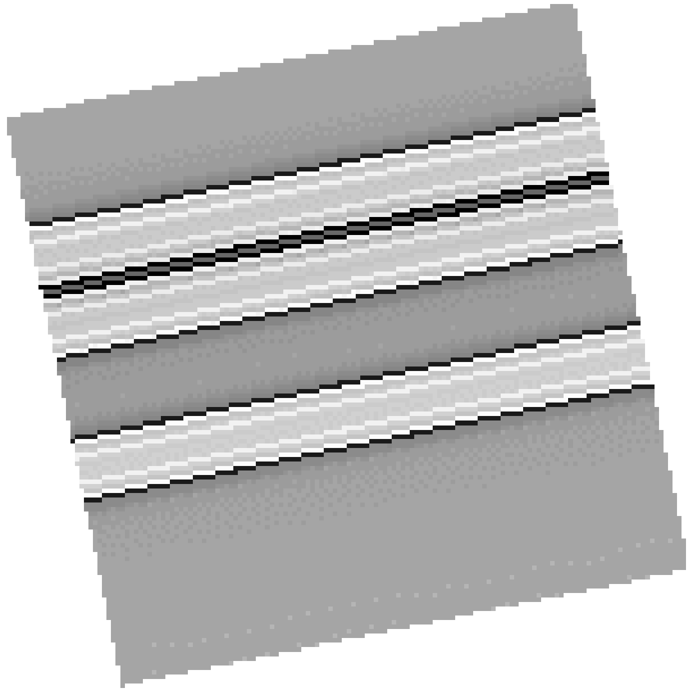


Now, project in several directions

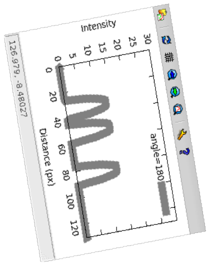


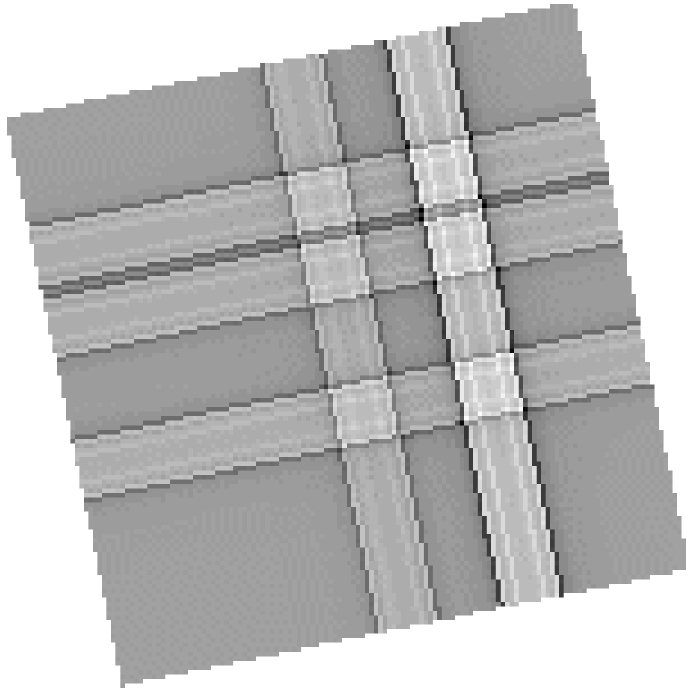
Reconstruction is the inversion of projection

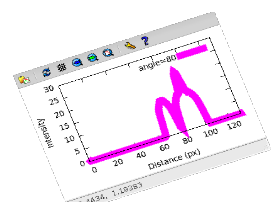




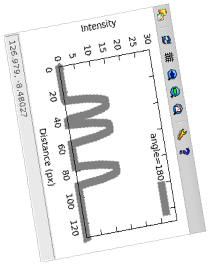
Reconstruction is the inversion of projection

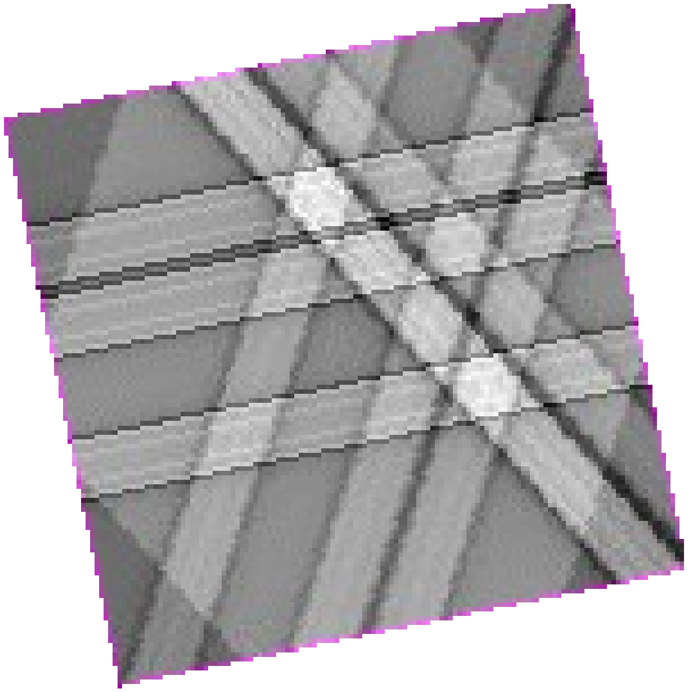


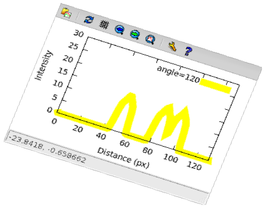


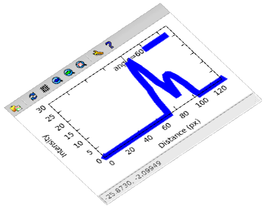


Reconstruction is the inversion of projection

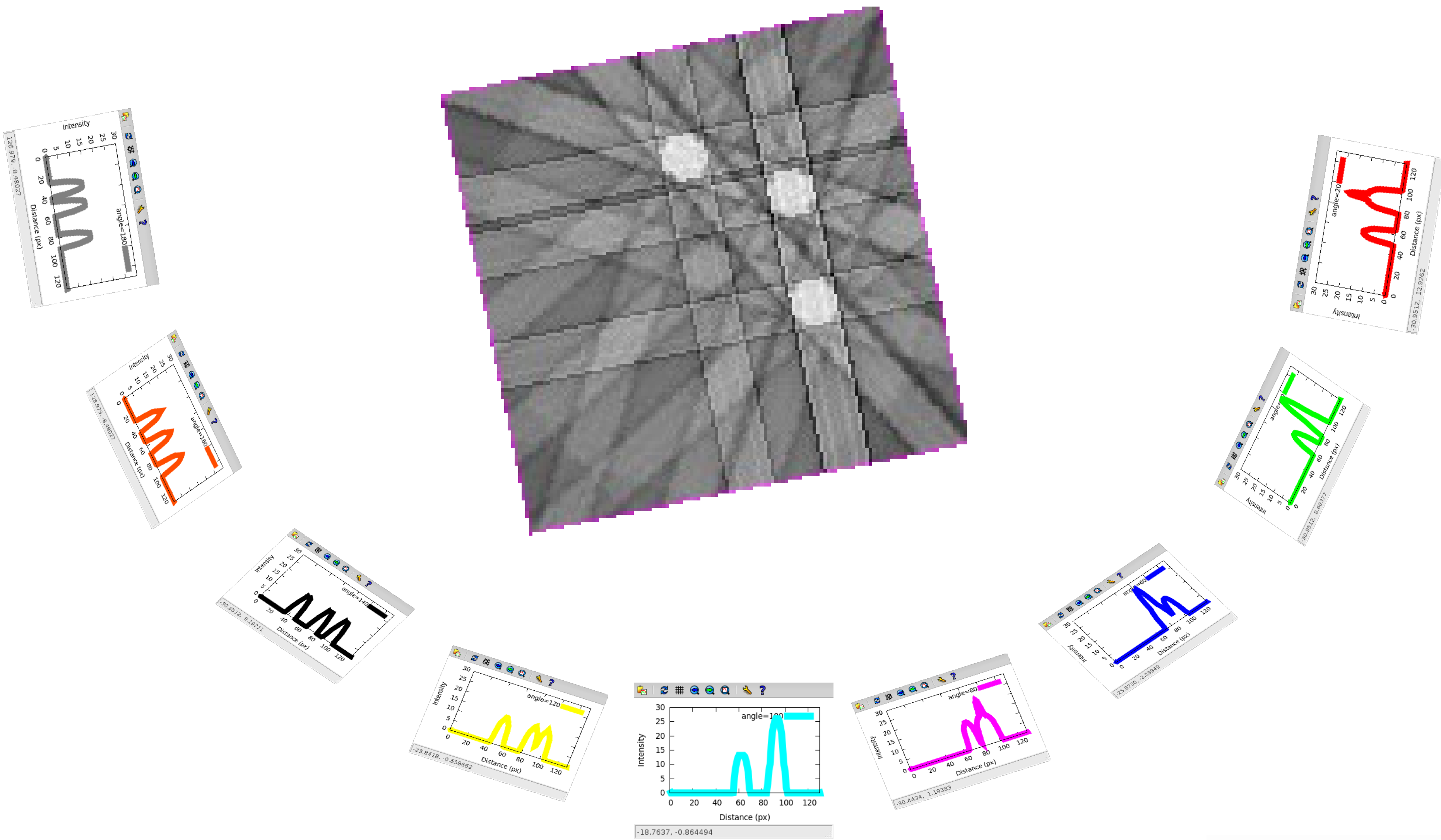




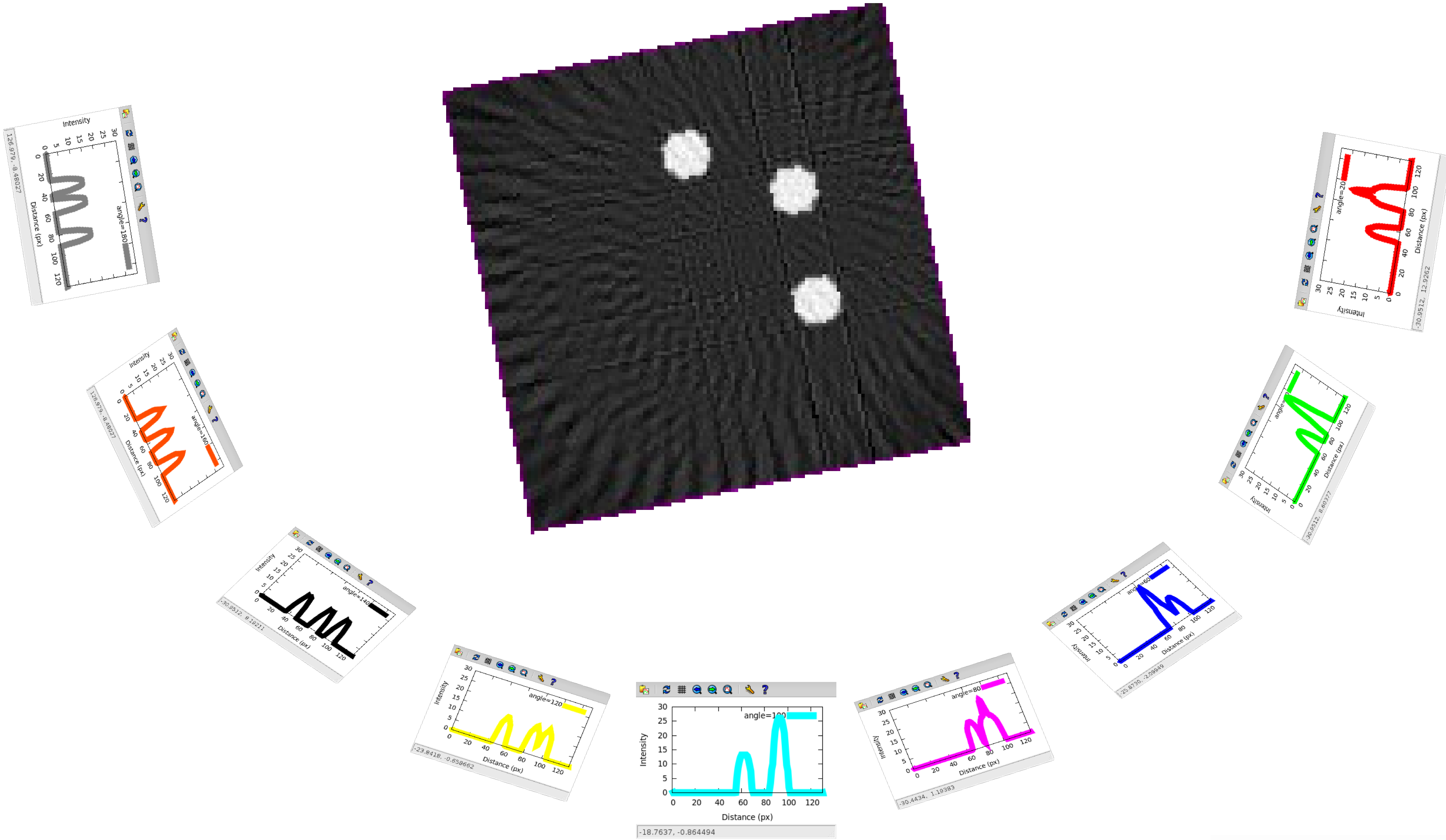




Reconstruction is the inversion of projection



Reconstruction is the inversion of projection



The reconstruction doesn't agree well with the projections. What can we do?

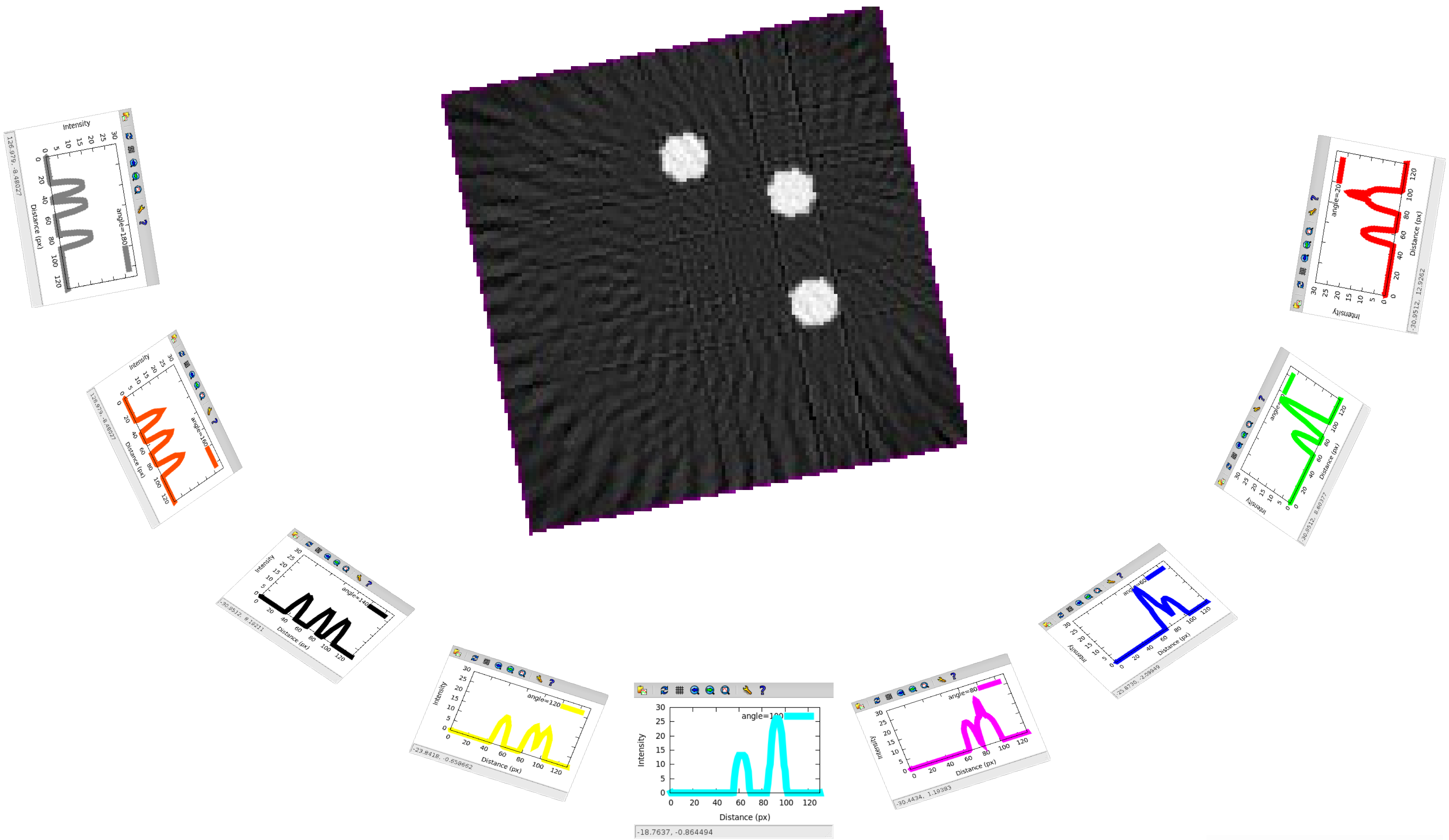
(one) ANSWER:
Simultaneous Iterative Reconstruction Technique

Simultaneous Iterative Reconstruction Technique

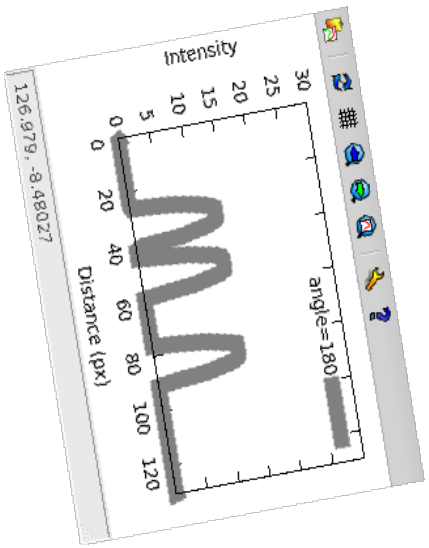
The idea:

- You compute re-projections of your model.
- Compare the re-projections to your experimental data.
    - There will be differences.
- You weight the differences by a fudge factor, λ.
- You adjust the model by the difference weighted by λ.
- Repeat.

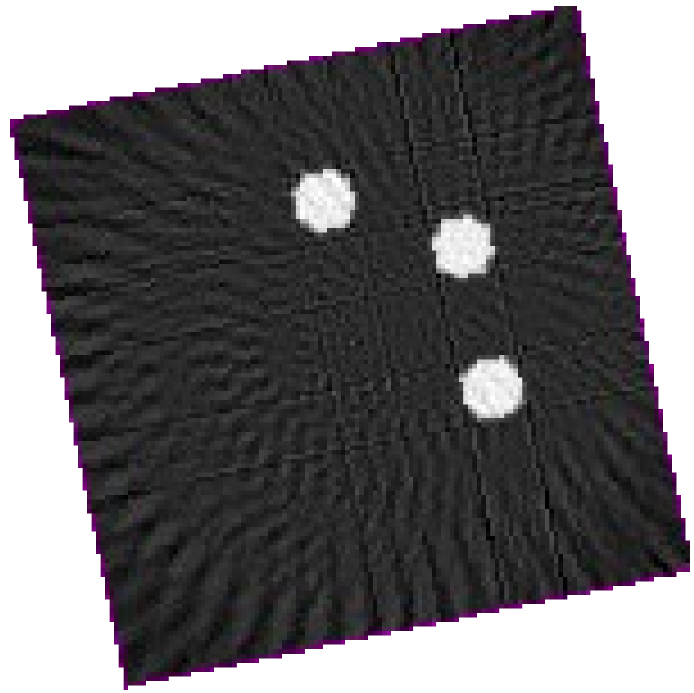
Simultaneous Iterative Reconstruction Technique



Simultaneous Iterative Reconstruction Technique



Experimental projection

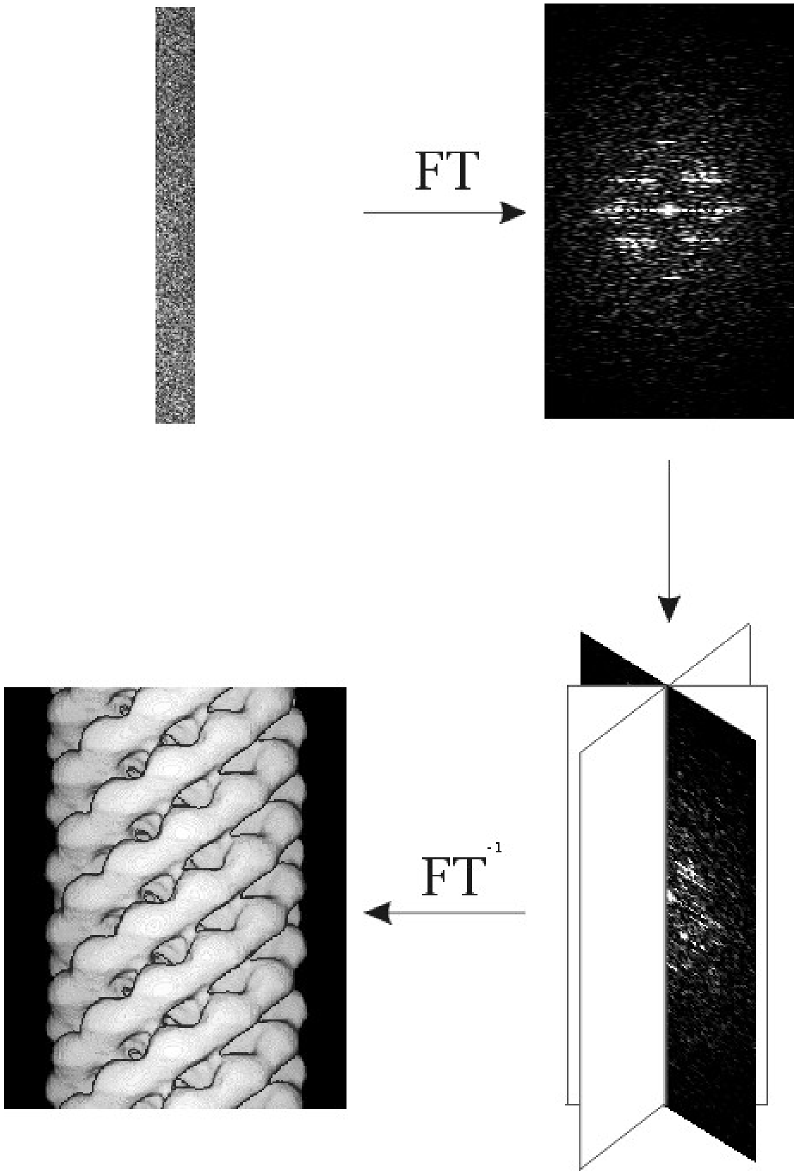


Model

Here, the differences (which will be down-weighted by λ) are the ripples in the background.

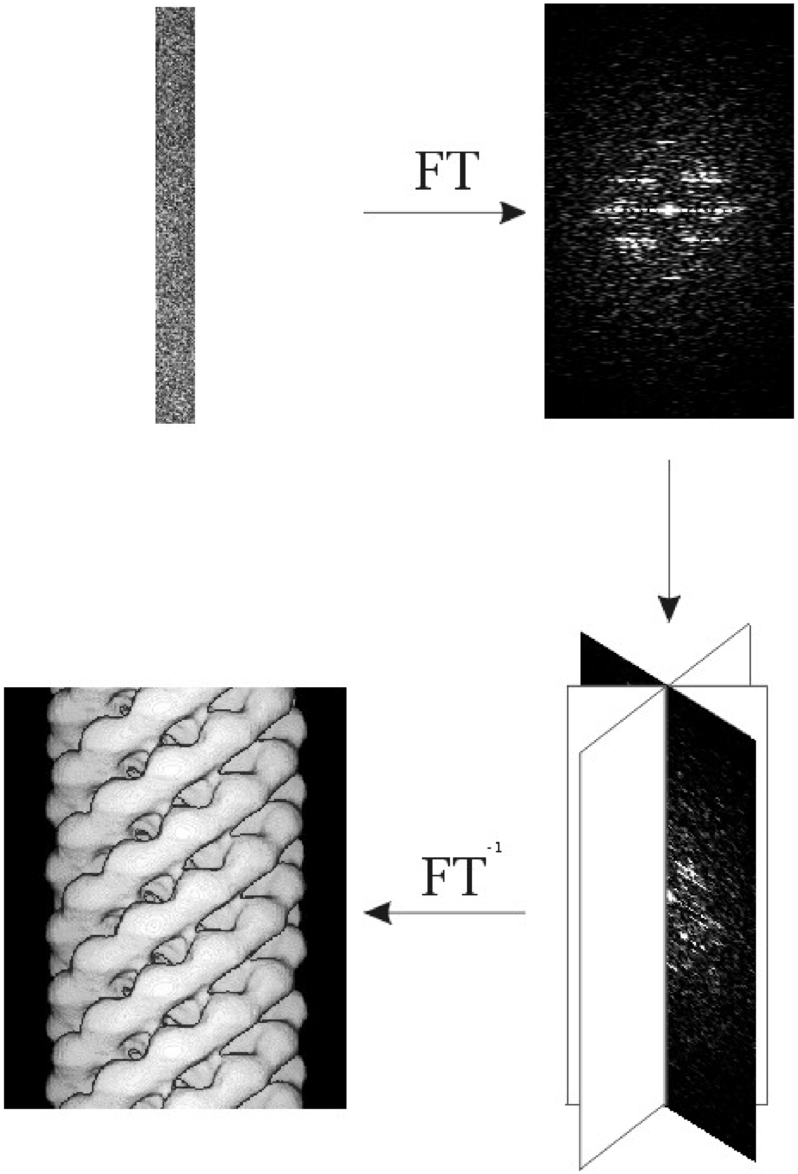
If we didn't down-weight by λ, we would overcompensate, and would amplify noise.

Reconstruction in Fourier space



Projection theorem
(or Central Section Theorem)

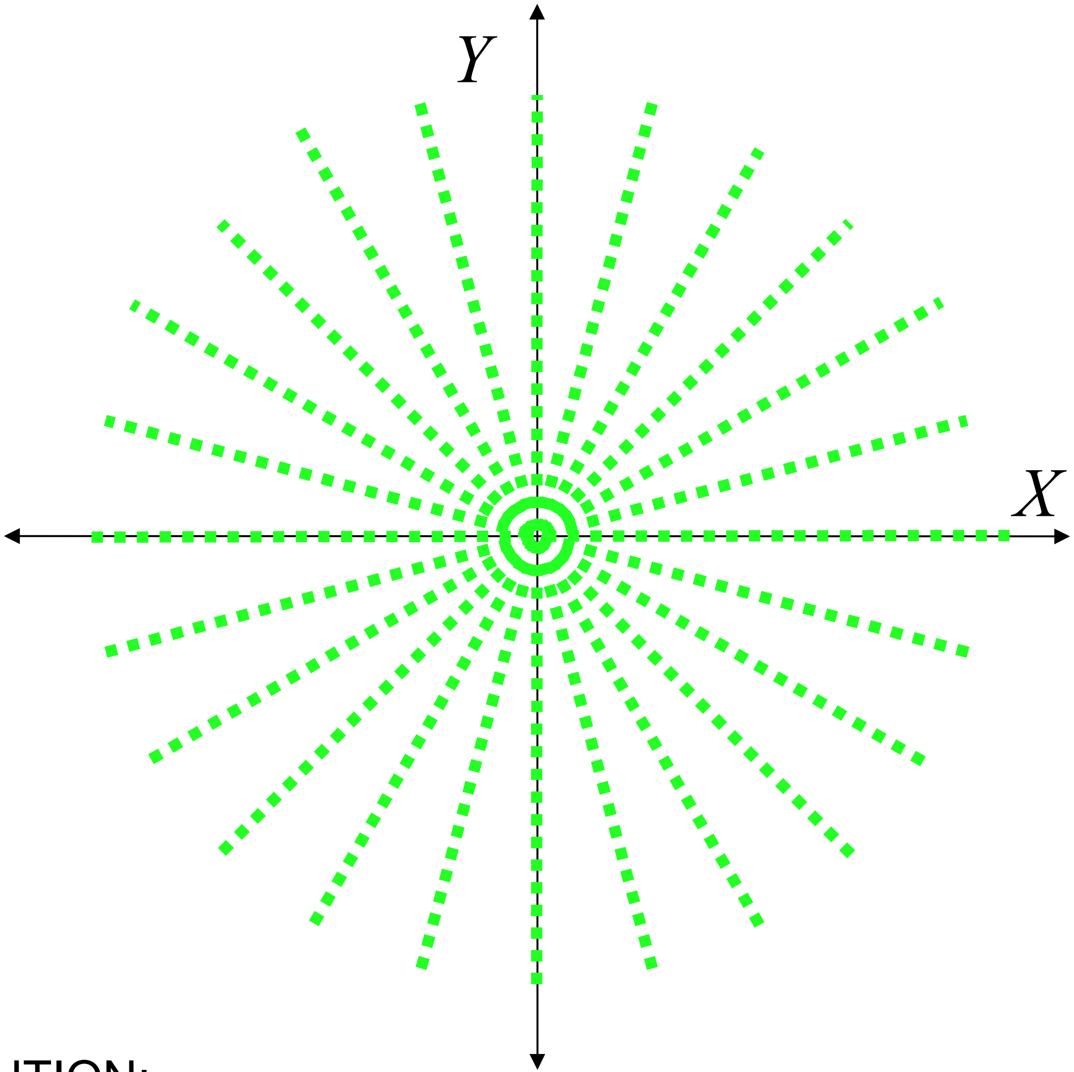
A central section through the 3D Fourier transform is the Fourier transform of the projection in that direction.



Projection theorem
(or Central Section Theorem)

The disadvantage is that you have To resample your central sections from polar coordinates to Cartesian space, i.e. interpolate. There are new methods to better Interpolate in Fourier space.

Converting from polar to Cartesian coordinates



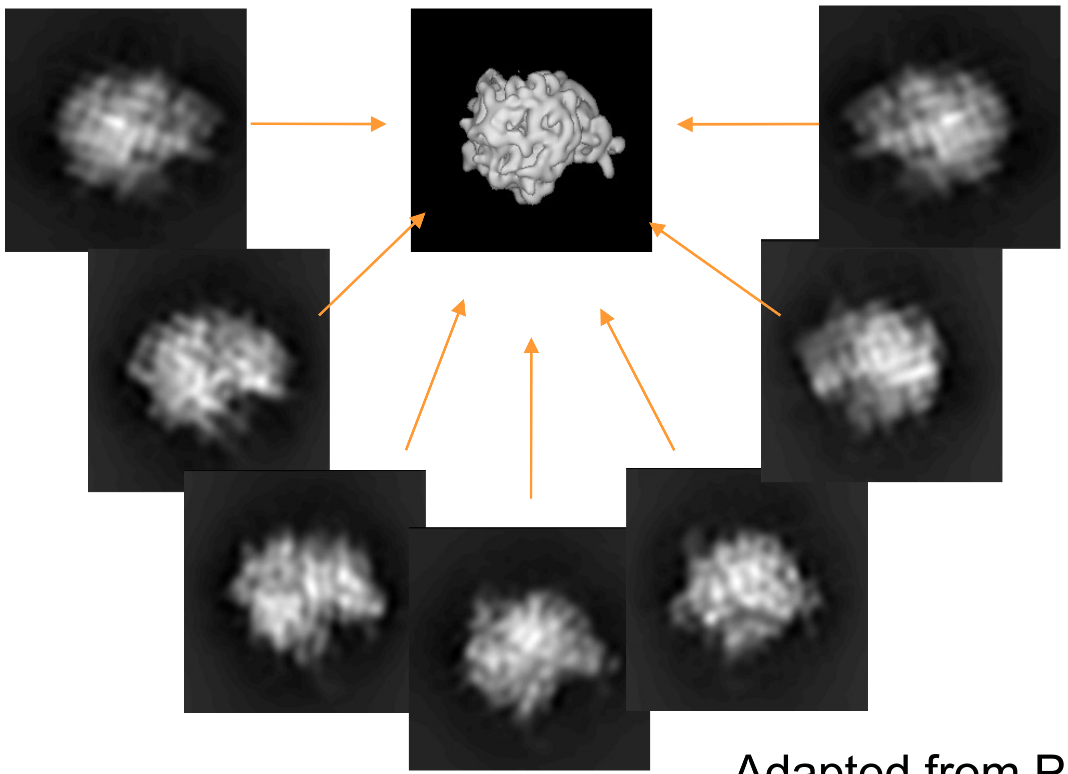
SOLUTION:

A simple weighting scheme is to divide the weight by the radius:

r^* weighting, or "r-weighted backprojection"

Going from 2D to 3D

If you know the orientation angles for each image,
you can compute a back-projection.



Adapted from Pawel Penczek

How do we determine the last two Euler angles?

Parameters required for 3D reconstruction

Two translational:

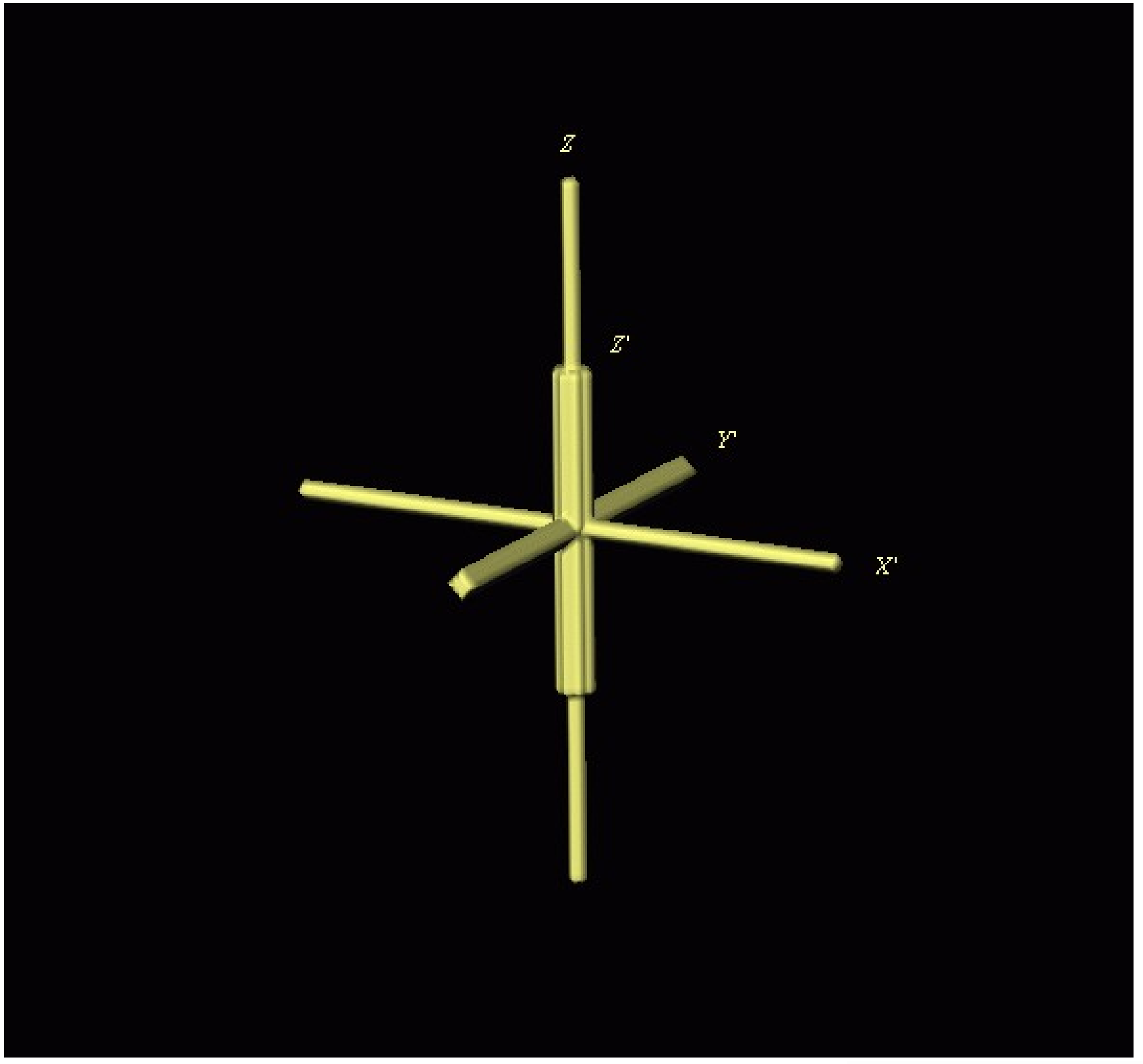
- ✓ Δx
- ✓ Δy

Three orientational (Euler angles):

- ✓ phi (about *z* axis)
- ✓ theta (about *y*)
- ✓ psi (about new *z*)

These are determined in 2D.

These are determined in 3D.

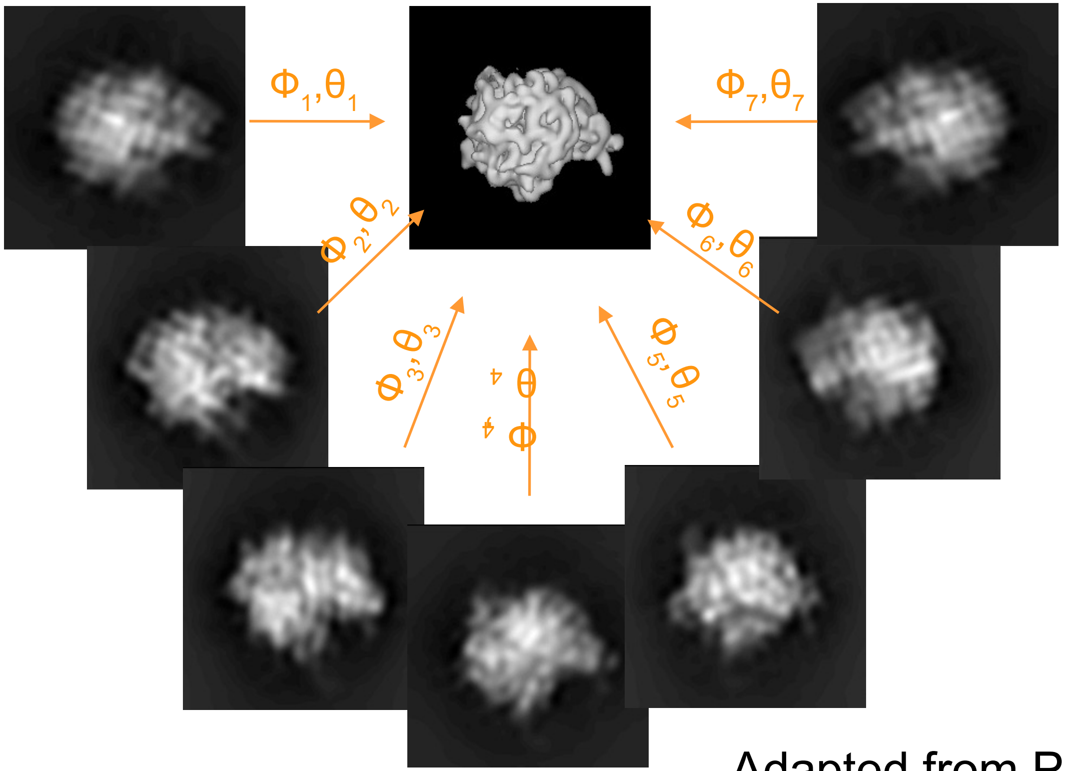


http://www.wadsworth.org

CEITEC

Going from 2D to 3D

If you know the orientation angles for each image,
you can compute a back-projection.



Adapted from Pawel Penczek

Outline

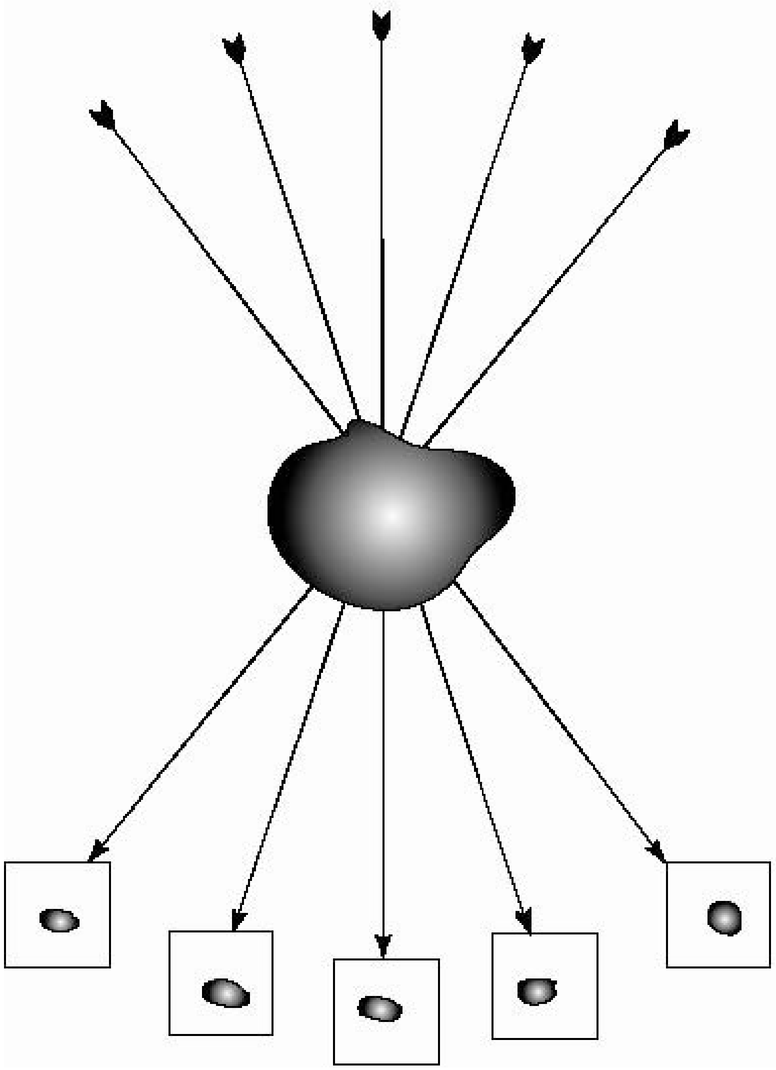
Image analysis III

- More on last week's material
  - Dependence of SNR on √N
  - Oversampling
- Classification

3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

Tomography



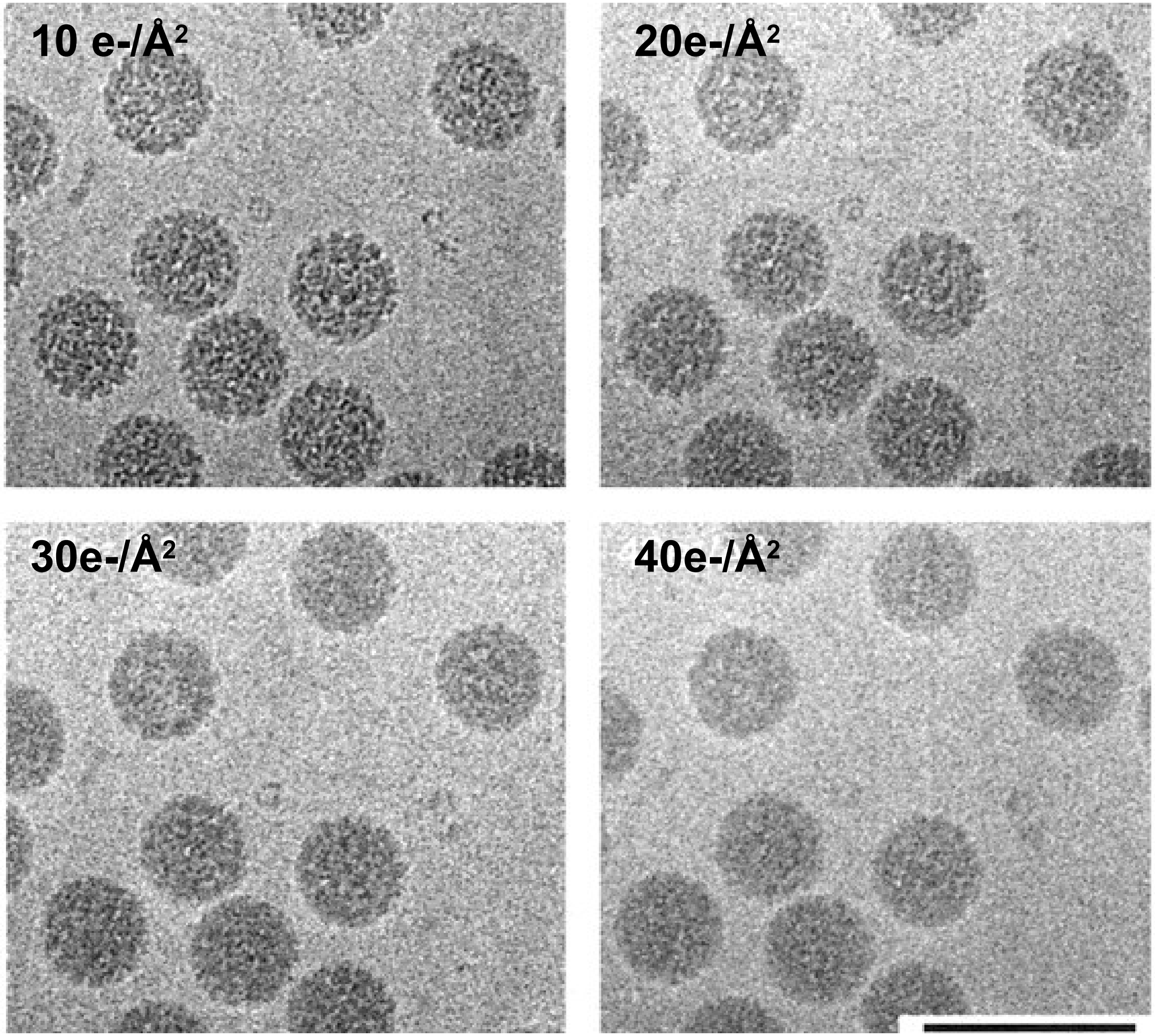
From Ken Downing

We have:

- known orientations
- different views

BUT...

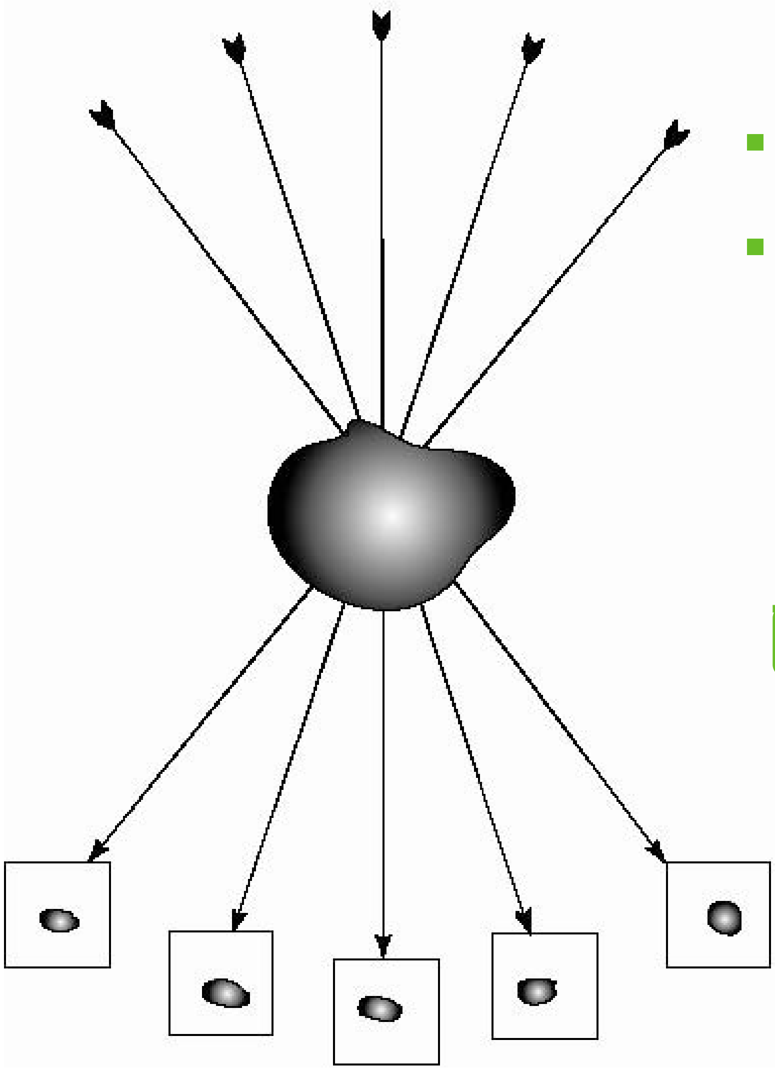
What happens when we image the sample?



Baker et al. (1999) Microbiol. Mol. Biol. Rev. **63**: 862

We are destroying the sample as we image it.

Consequences of repeated exposure



From Ken Downing

- Accumulated beam damage
- If number of views is limited, then distortions

Solution:

If we have many identical molecules,
and if we can determine the orientations,
we can use one exposure per molecule
and use these images in the reconstruction.

"Single-particle reconstruction"

If we have many identical molecules,
and if we can determine the orientations,
we can use one exposure per molecule
and use these images in the reconstruction.

BUT:
Unlike in the tomographic case,
we don't know how the orientations
between the different images are related.

Outline

Image analysis III

- More on last week's material
  - Dependence of SNR on √N
  - Oversampling
- Classification

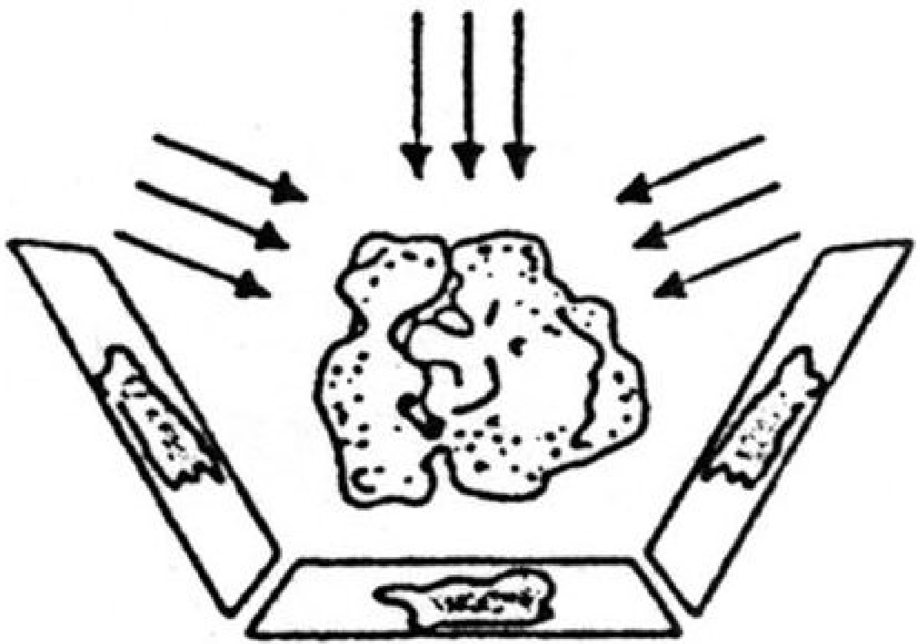
3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

Reference-based alignment

You will record the direction of projection (the Euler angles), such that
if you encounter an experimental image that resembles a reference projection,
you will assign that reference projection's Euler angles to the experimental image.

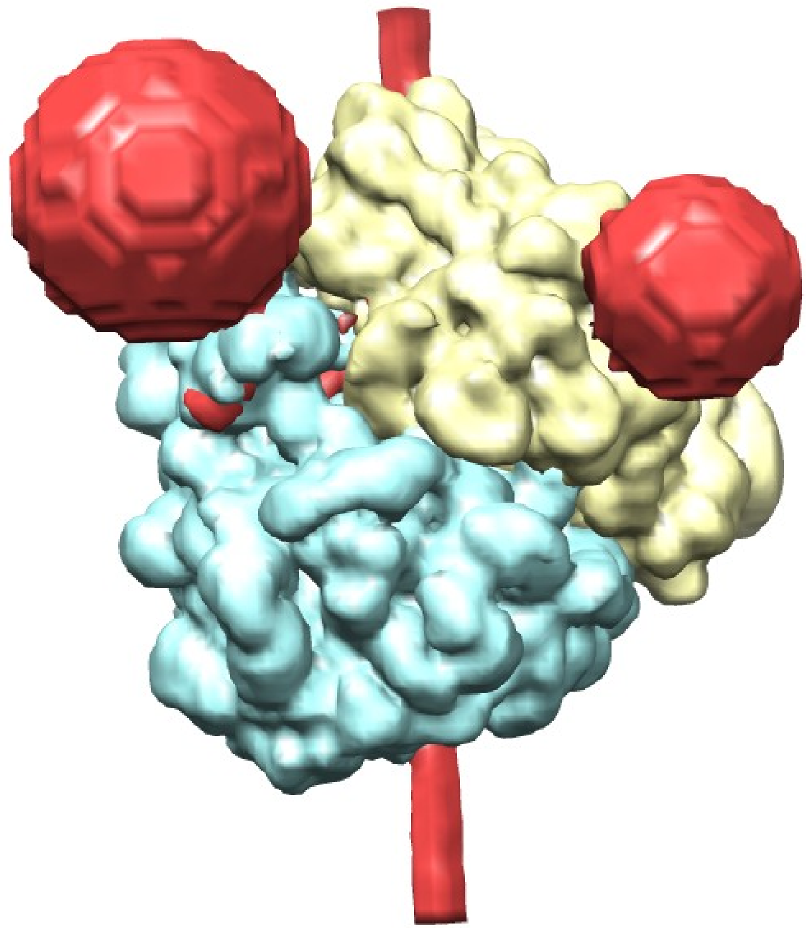
Step 1: Generation of projections of the reference.



From Penczek *et al.* (1994), *Ultramicroscopy* **53**: 251-70.

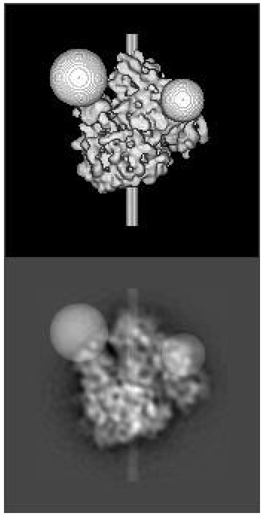
Assumption: reference is similar enough to the sample that it can be used to determine orientation.

The model

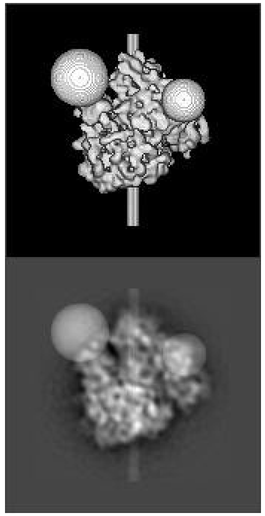


(The extra features helped determine handedness in noisy reconstructions.)

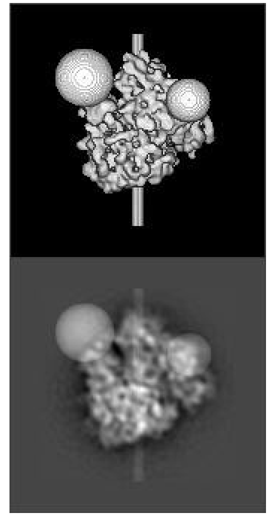
CEITEC



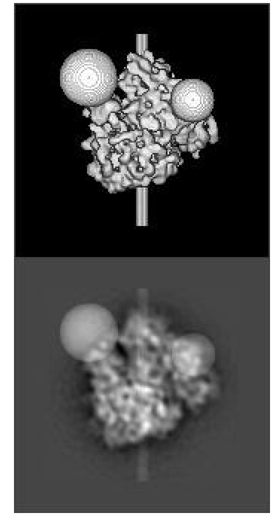
phi=000
theta=000
psi=000



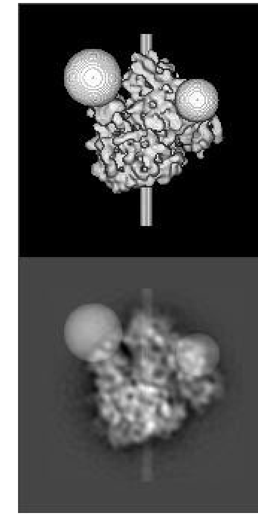
phi=000
theta=000
psi=000



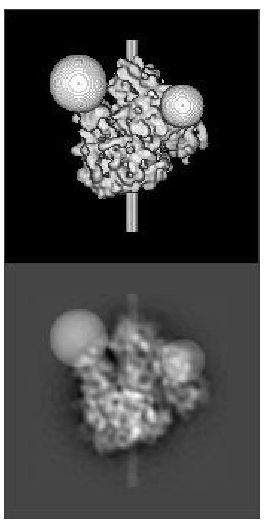
phi=000
theta=000
psi=000



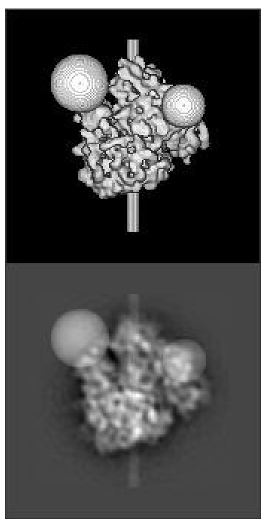
phi=000
theta=000
psi=000



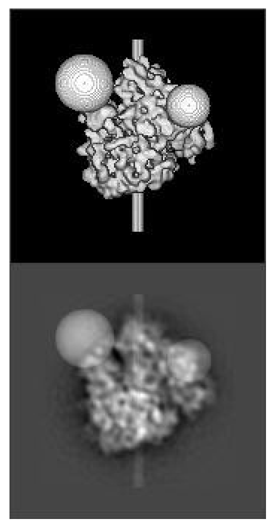
phi=000
theta=000
psi=000



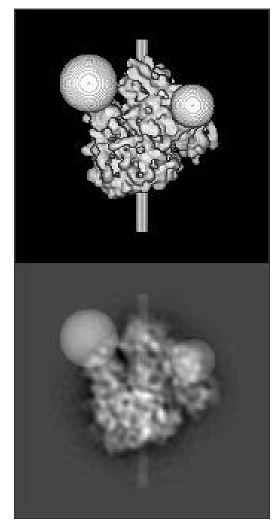
phi=000
theta=000
psi=000



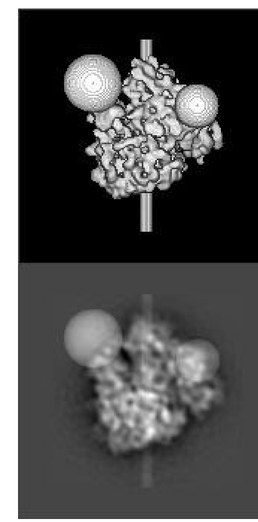
phi=000
theta=000
psi=000



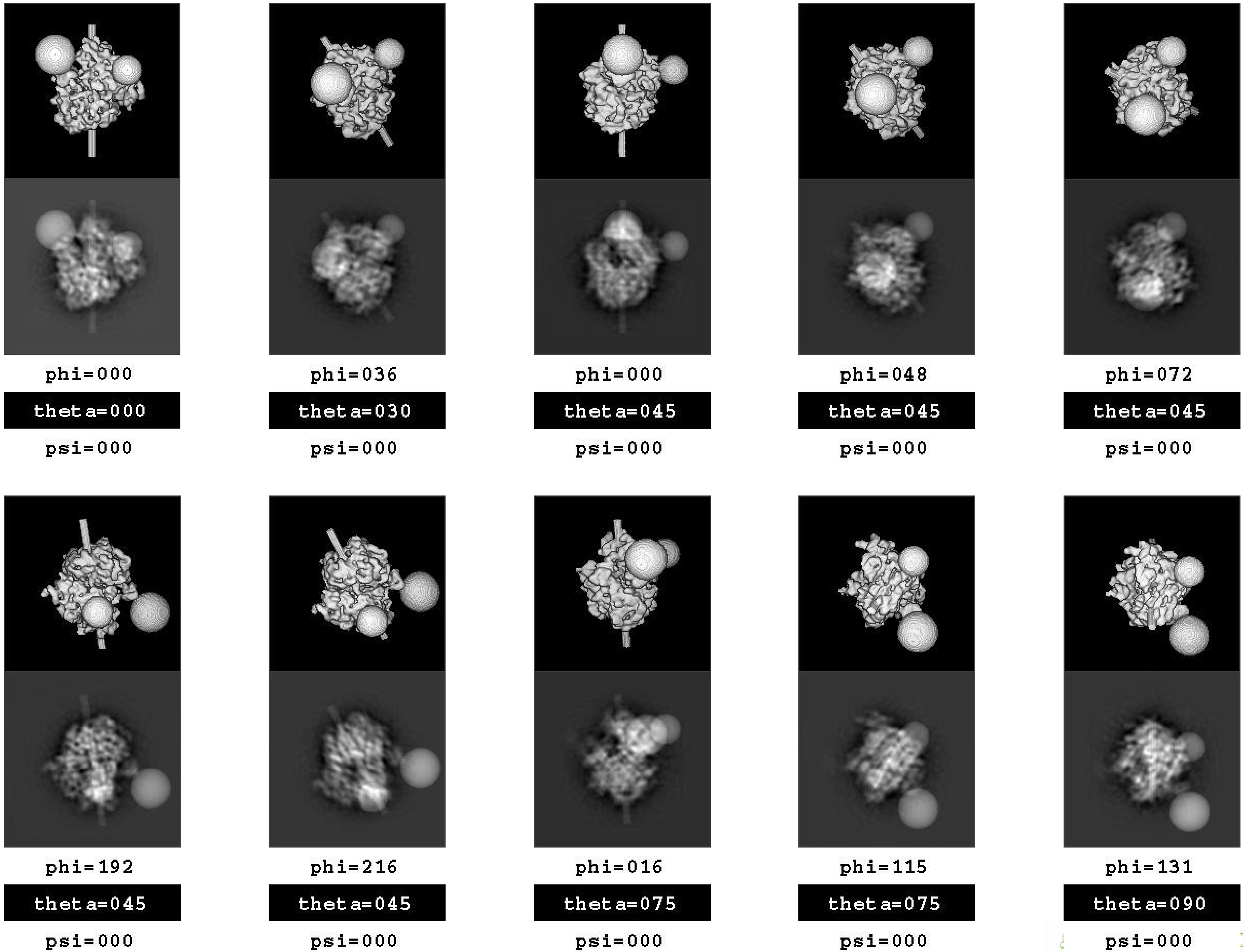
phi=000
theta=000
psi=000



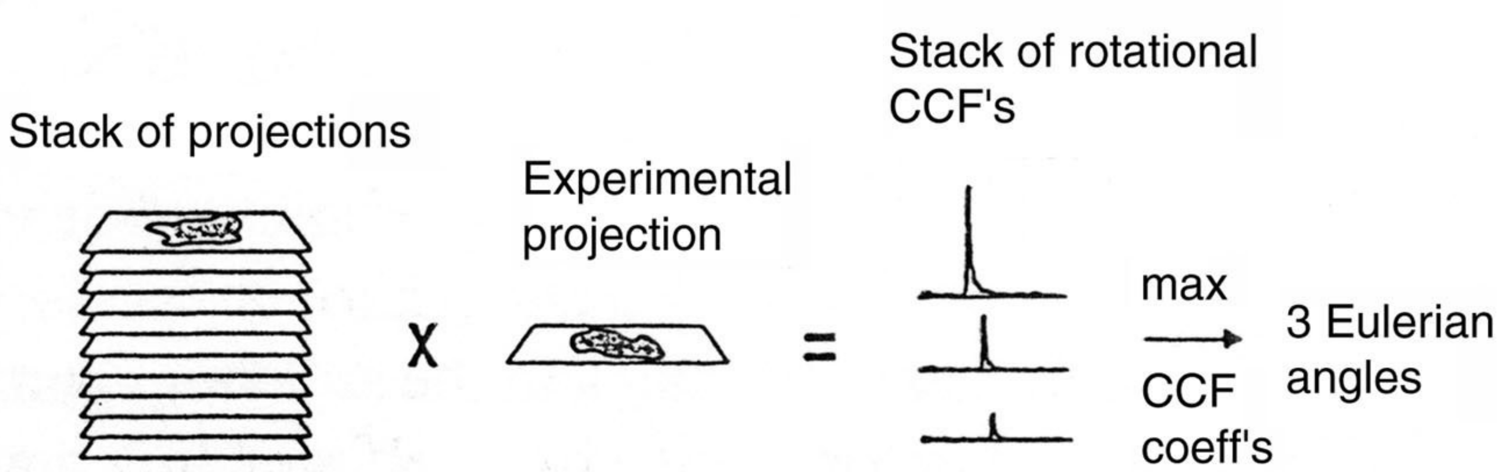
phi=000
theta=000
psi=000



phi=000
theta=000
psi=000


phi=000
theta=000
psi=000
phi=036
theta=030
psi=000
phi=000
theta=045
psi=000
phi=048
theta=045
psi=000
phi=072
theta=045
psi=000
phi=192
theta=045
psi=000
phi=216
theta=045
psi=000
phi=016
theta=075
psi=000
phi=115
theta=075
psi=000
phi=131
theta=090
psi=000

Reference-based alignment



From Penczek *et al.* (1994), *Ultramicroscopy* **53**: 251-70.

Steps:

1. Compare the experimental image to all of the reference projections.
2. Find the reference projection with which the experimental image matches best.
3. Assign the Euler angles of that reference projection to the experimental image.

Outline

Image analysis III

- More on last week's material
  - Dependence of SNR on √N
  - Oversampling
- Classification

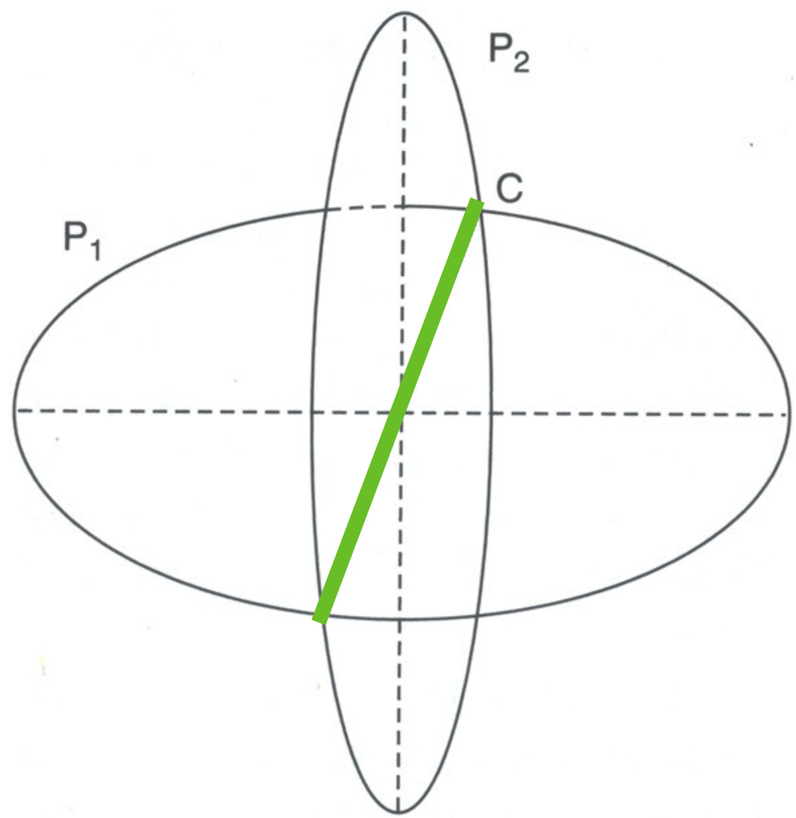
3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

Common lines
(or Angular Reconstitution)

Summary:

- A central section through the 3D Fourier transform is the Fourier transform of the projection in that direction
- Two central sections will intersect along a line through the origin of the 3D Fourier transform
- With two central sections, there is still one degree of freedom to relate the orientations, but a third projection (i.e., central section) will fix the relative orientations of all three.

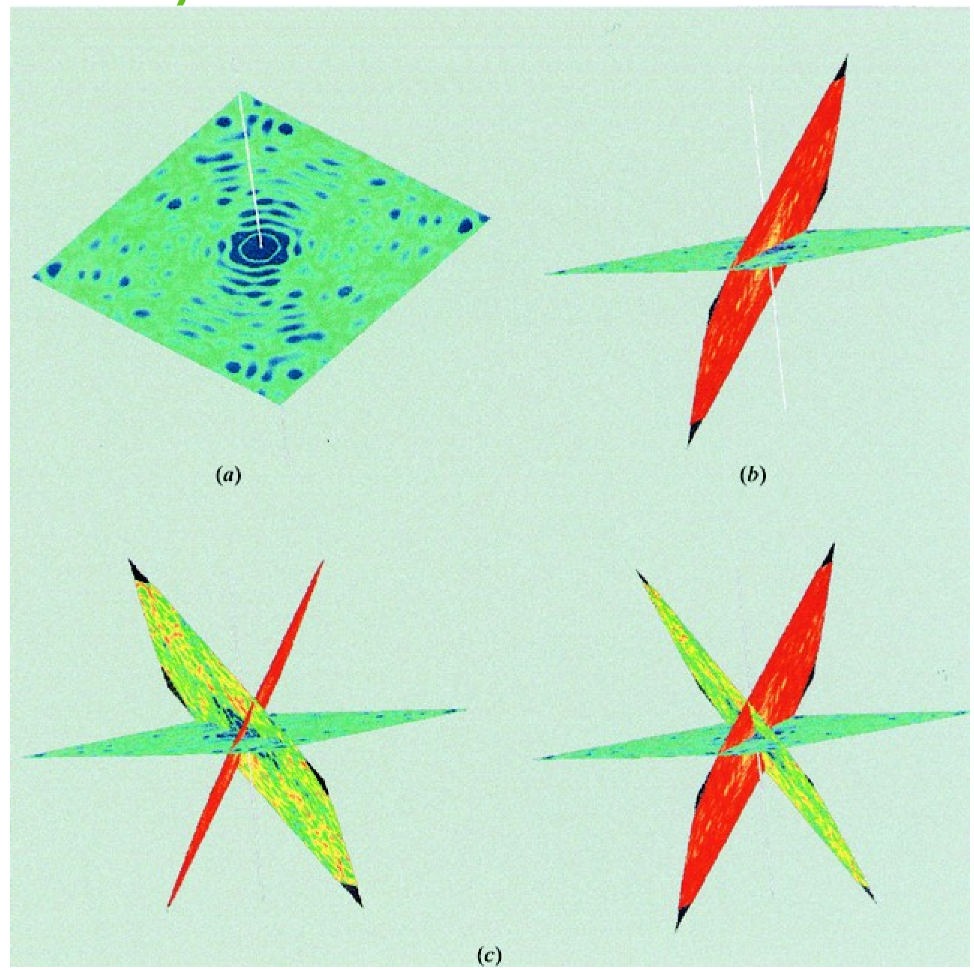


Frank, J. (2006) 3D Electron Microscopy of Macromolecular Assemblies

Common lines
(or Angular Reconstitution)

Summary:

- A central section through the 3D Fourier transform is the Fourier transform of the projection in that direction
- Two central sections will intersect along a line through the origin of the 3D Fourier transform
- With two central sections, there is still one degree of freedom to relate the orientations, but a third projection (i.e., central section) will fix the relative orientations of all three.



From Steve Fuller

Common lines: Problems

- Noise can lead to incorrect angles
  - Symmetry helps
- Handedness cannot be determined without additional information
  - Tilting
  - α-helices
- Assumes conformational homogeneity

Outline

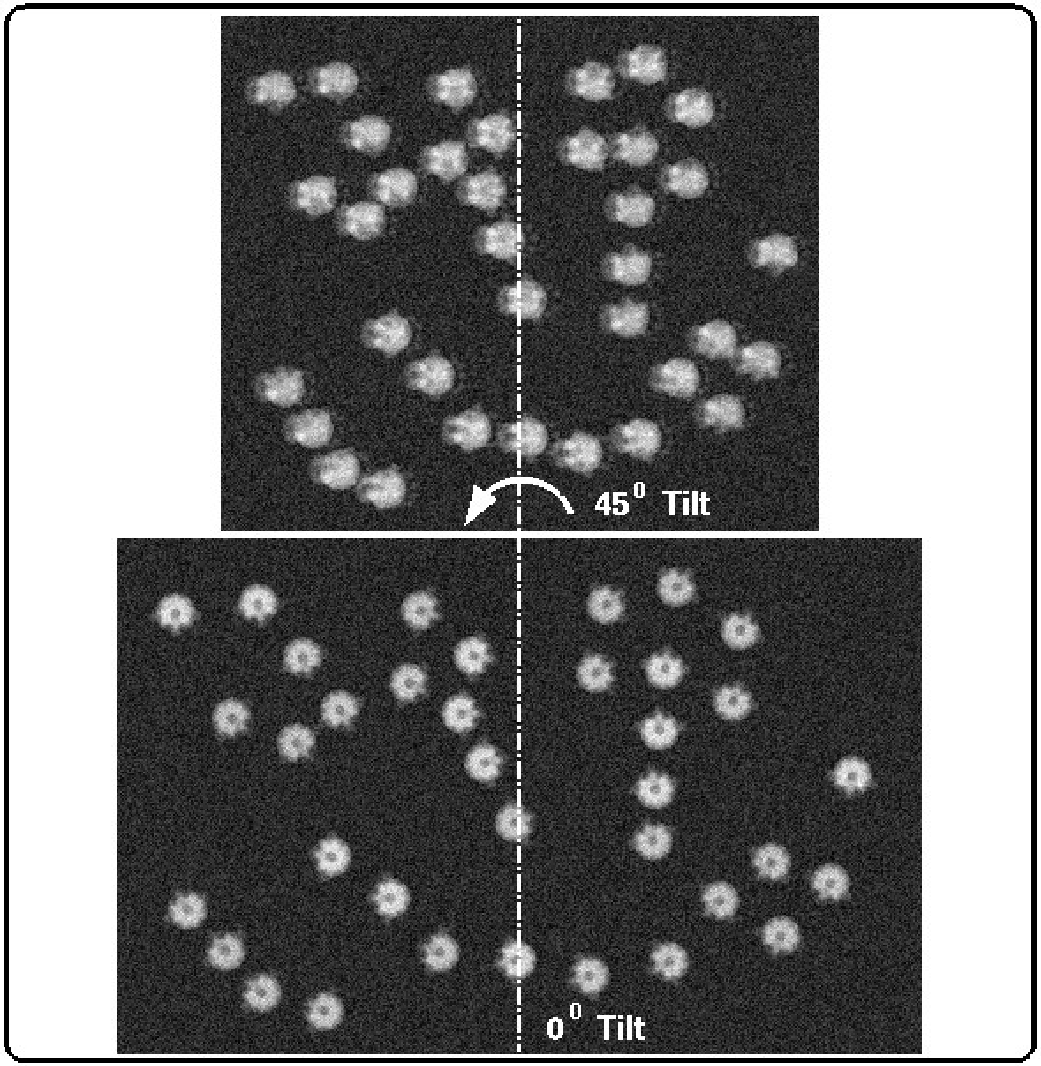
Image analysis III

- More on last week's material
    - Dependence of SNR on $\sqrt{N}$
    - Oversampling
- Classification

3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

Random-conical tilt:
Determination of Euler angles

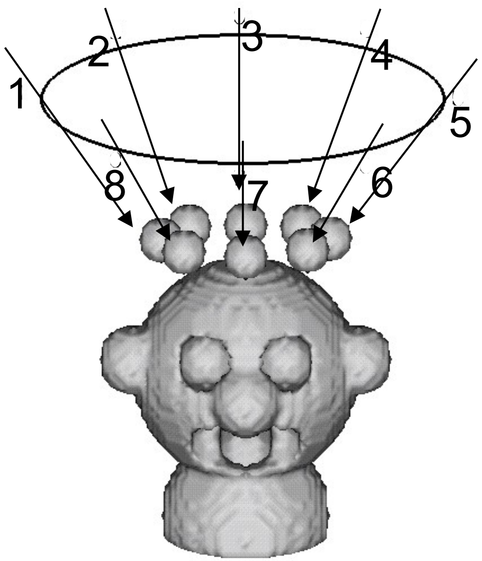


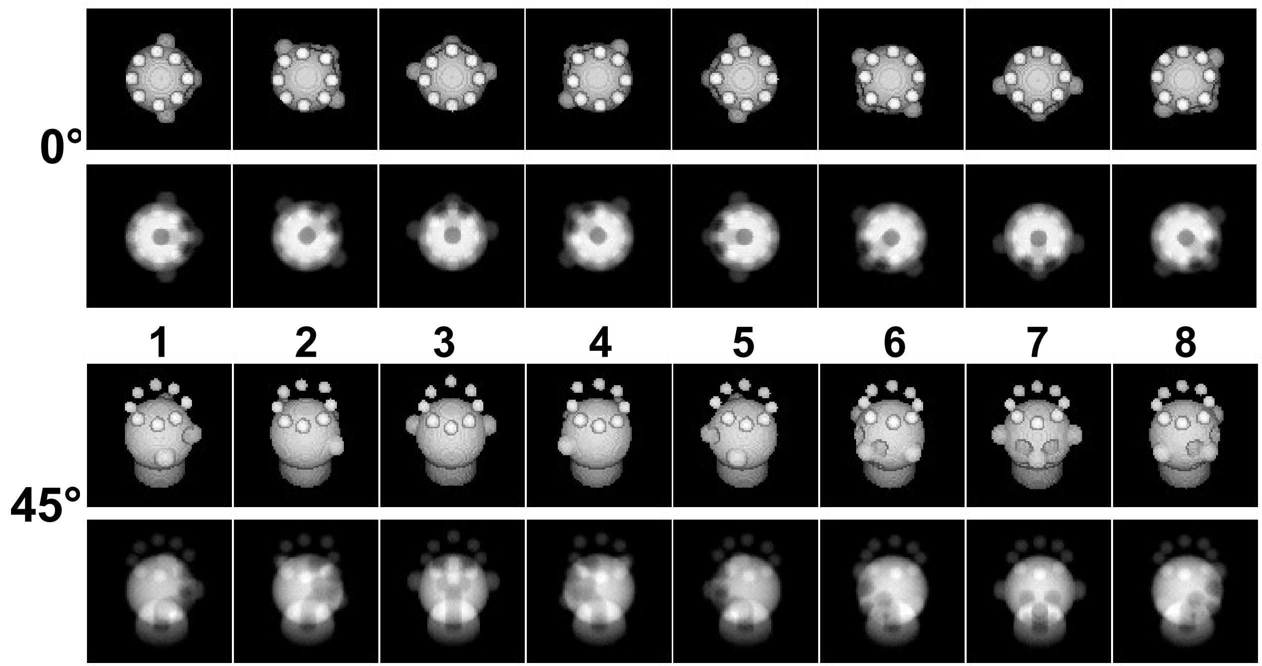
This scenario describes a worst case, when there is exactly one orientation in the 0º image. Since the in-plane angle varies, in the tilted image, we have different views available.

From Nicolas Boisset

Random-conical tilt: Geometry

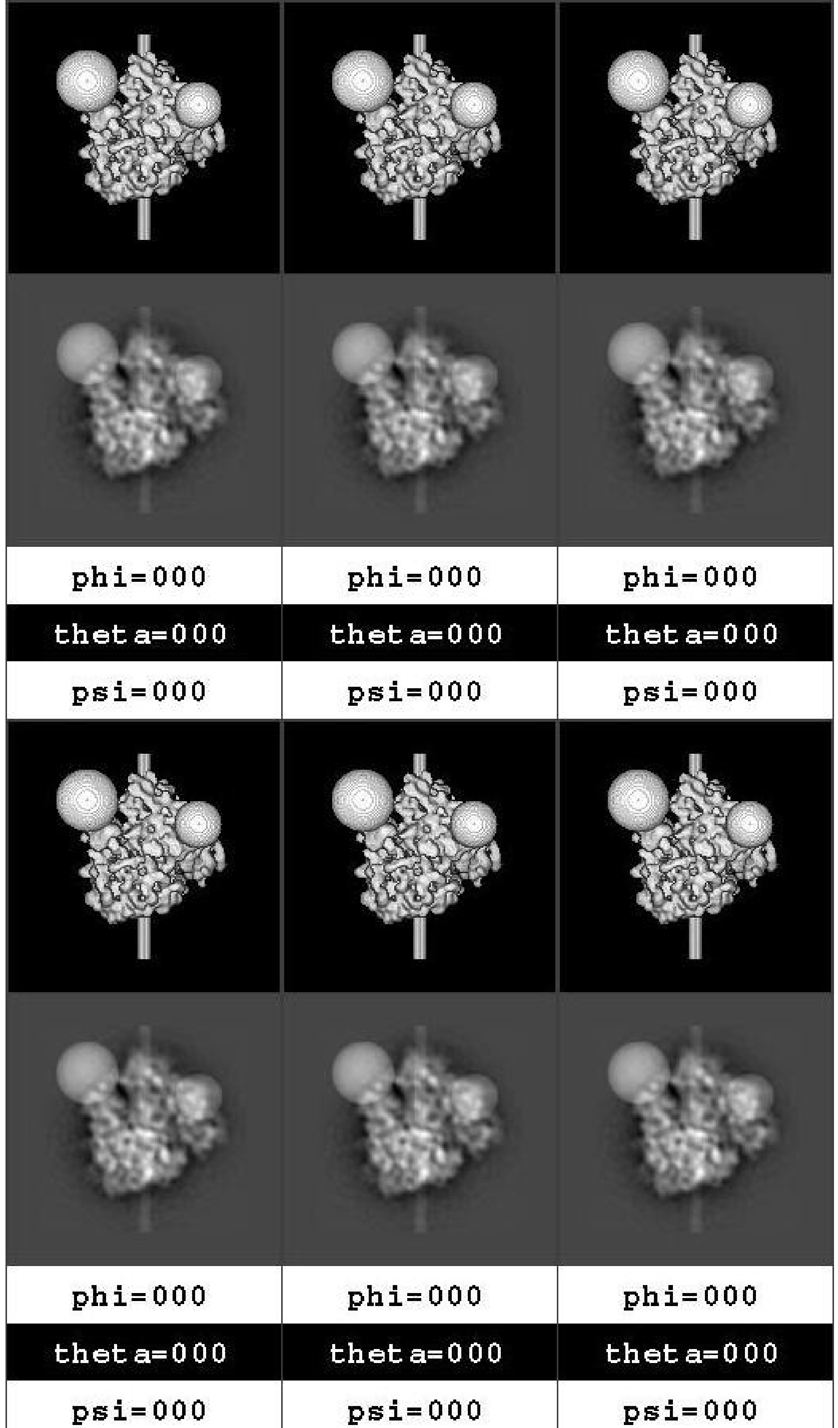
Two images are taken: one at 0° and one tilted at an angle of 45°.

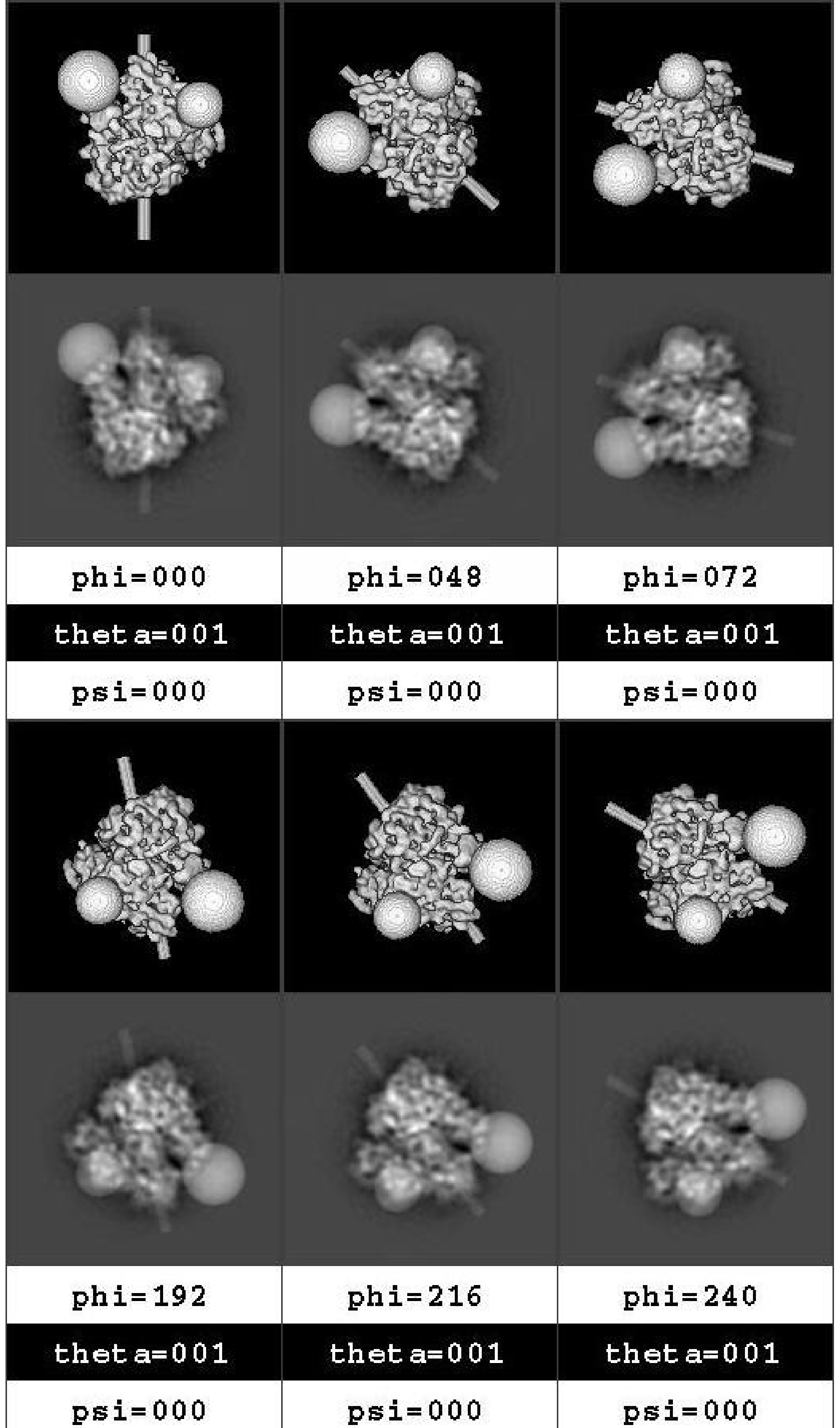


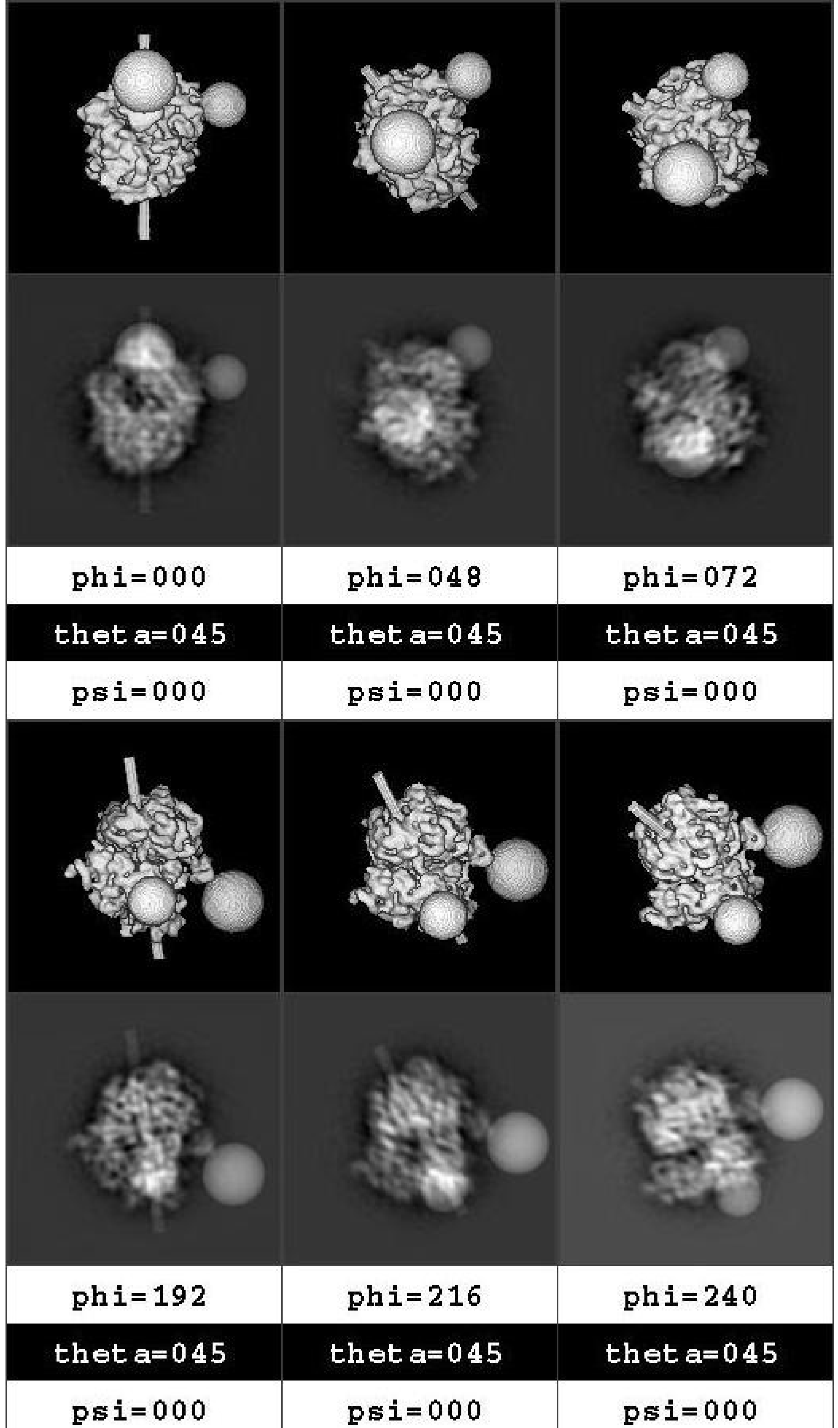


Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *J Microsc* **146**, 113-36 (1987).

From Nicolas Boisset


phi=000
theta=000
psi=000
phi=000
theta=000
psi=000
phi=000
theta=000
psi=000
phi=000
theta=000
psi=000
phi=000
theta=000
psi=000
phi=000
theta=000
psi=000


phi=000
theta=001
psi=000
phi=048
theta=001
psi=000
phi=072
theta=001
psi=000
phi=192
theta=001
psi=000
phi=216
theta=001
psi=000
phi=240
theta=001
psi=000

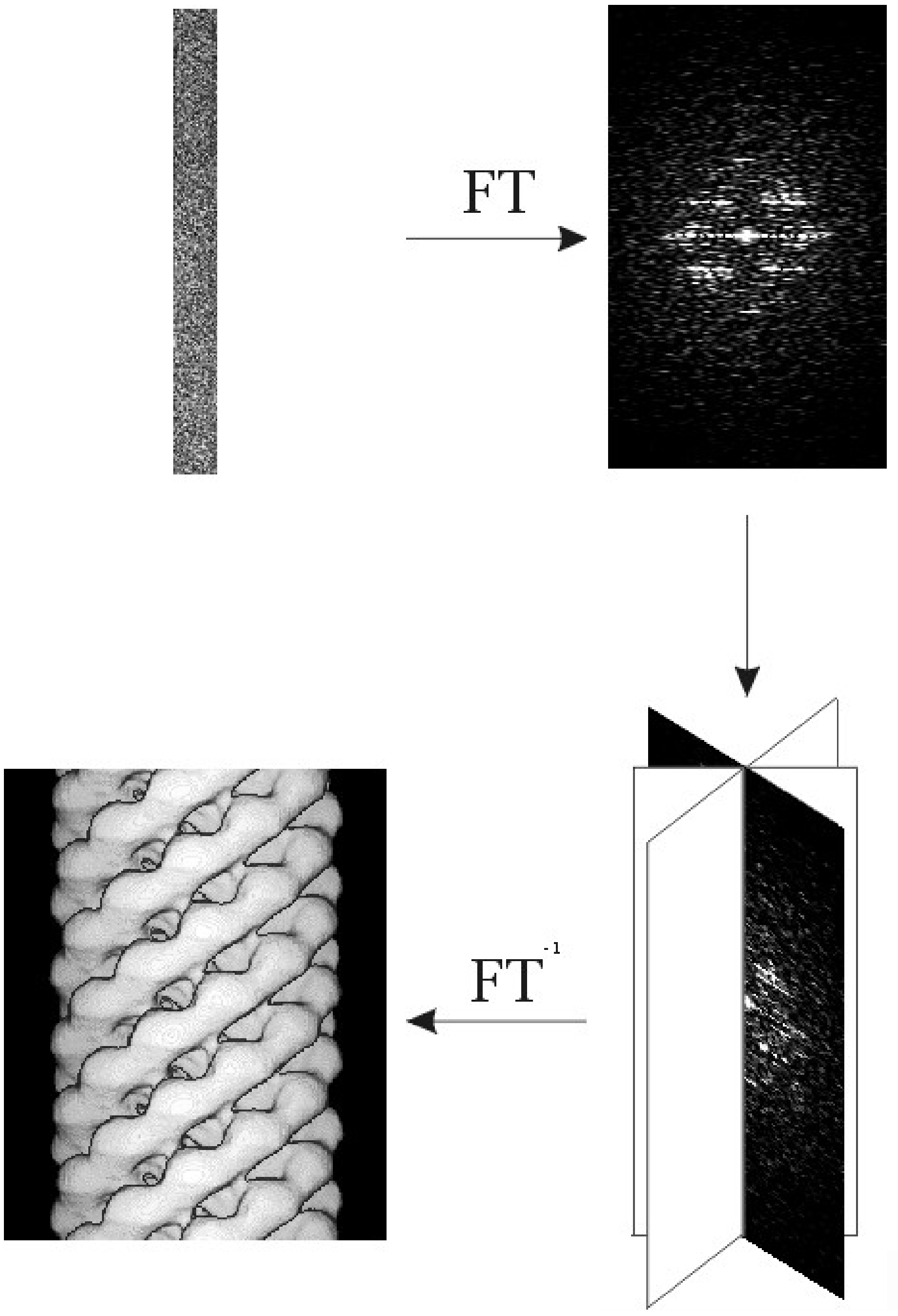

phi=000
theta=045
psi=000
phi=048
theta=045
psi=000
phi=072
theta=045
psi=000
phi=192
theta=045
psi=000
phi=216
theta=045
psi=000
phi=240
theta=045
psi=000

One problem though:

We can't tilt the stage all the way to 90 degrees.

Review:

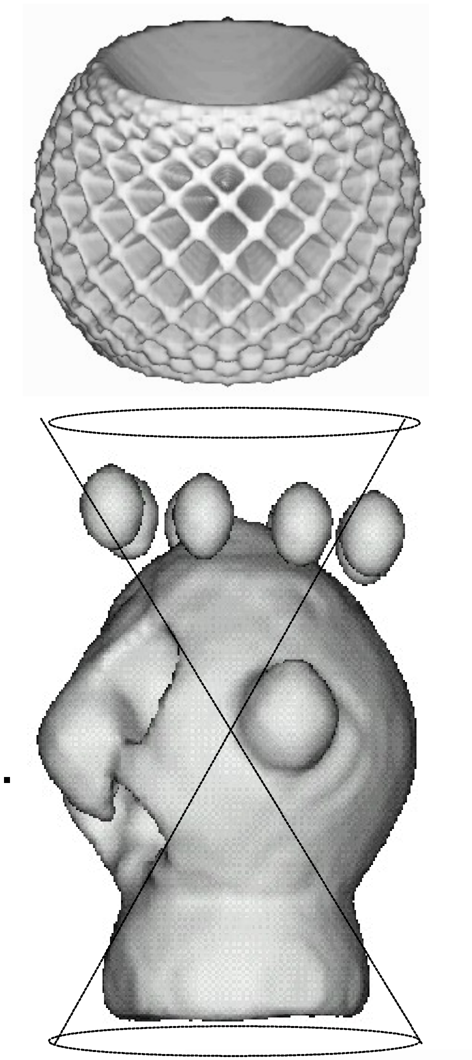
Projection theorem



Random-conical tilt:
The “missing cone”

Representation of the distribution of views, if we display a plane perpendicular to each projection direction

The missing information, in the shape of a cone, elongates features in the direction of the cone's axis.



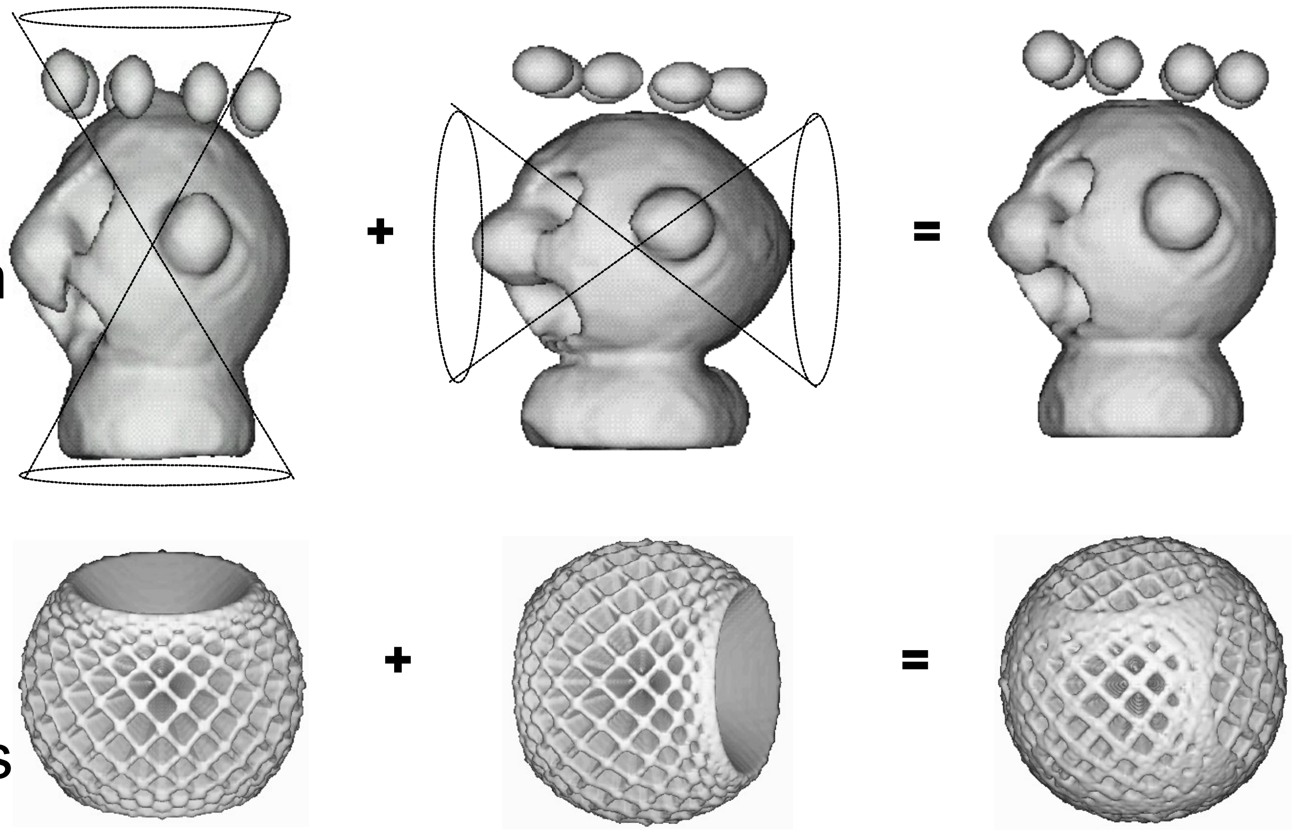
From Nicolas Boisset

Random-conical tilt:
Filling the missing cone

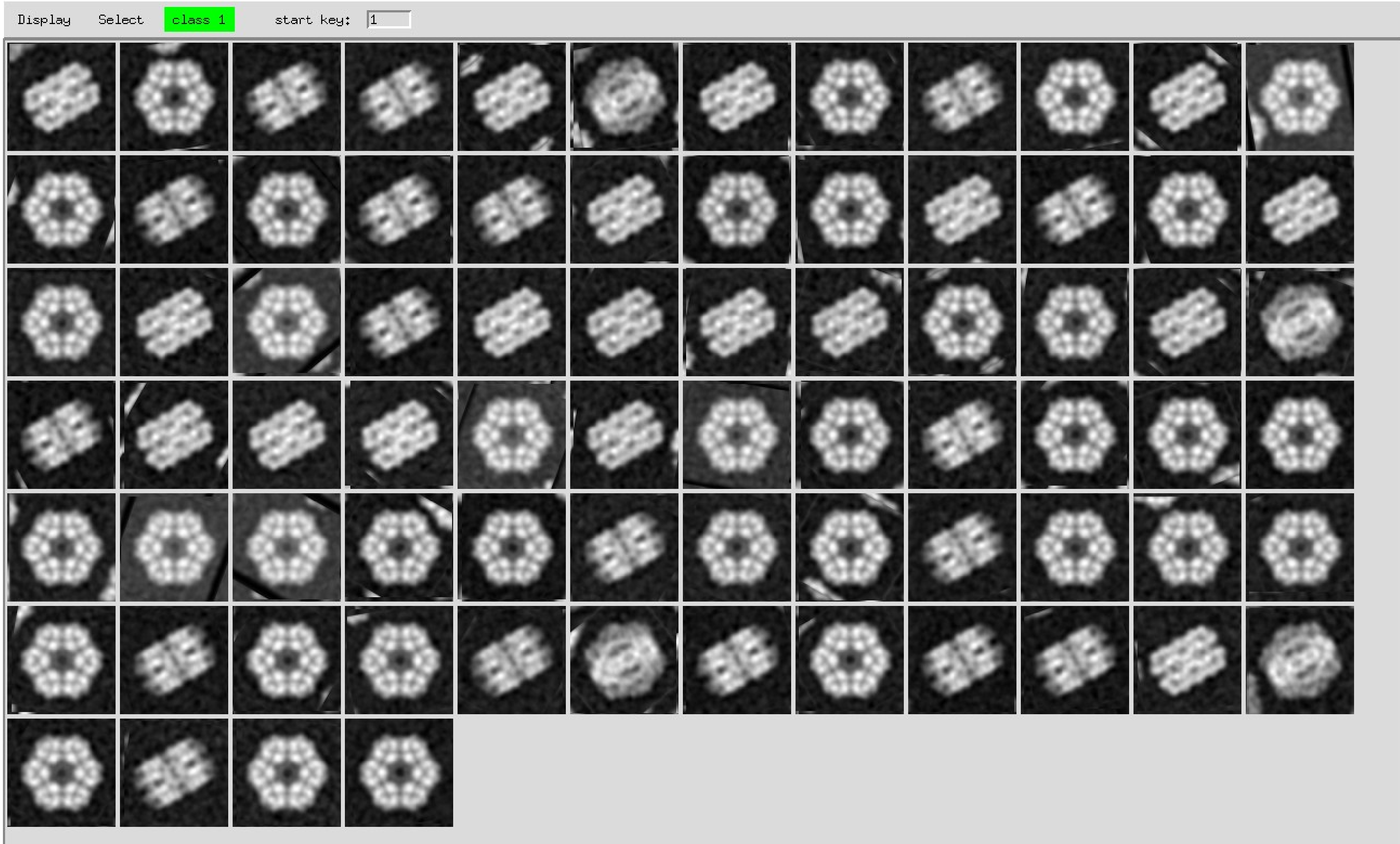
If there are multiple preferred orientations, or if there is symmetry that fills the missing cone, you can cover all orientations.

Reconstruction

Distribution of orientations



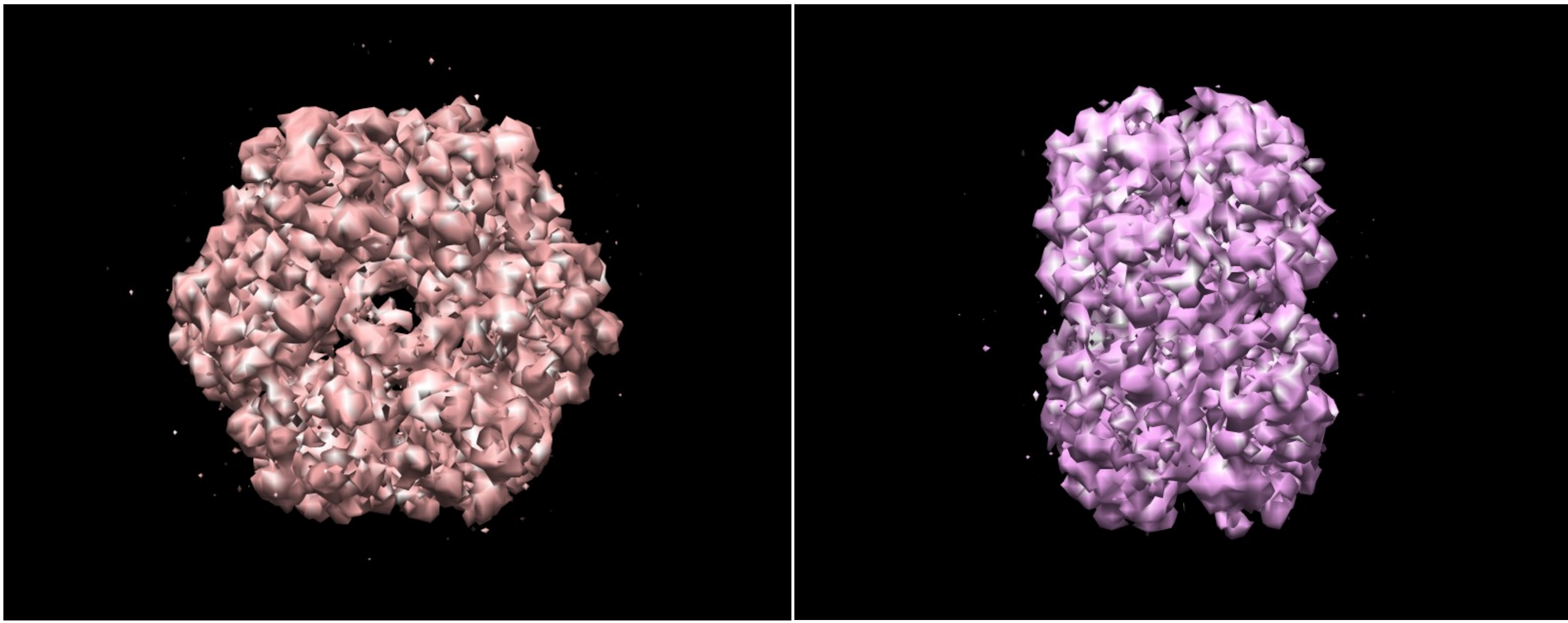
From Nicolas Boisset



Phantom images of worm hemoglobin

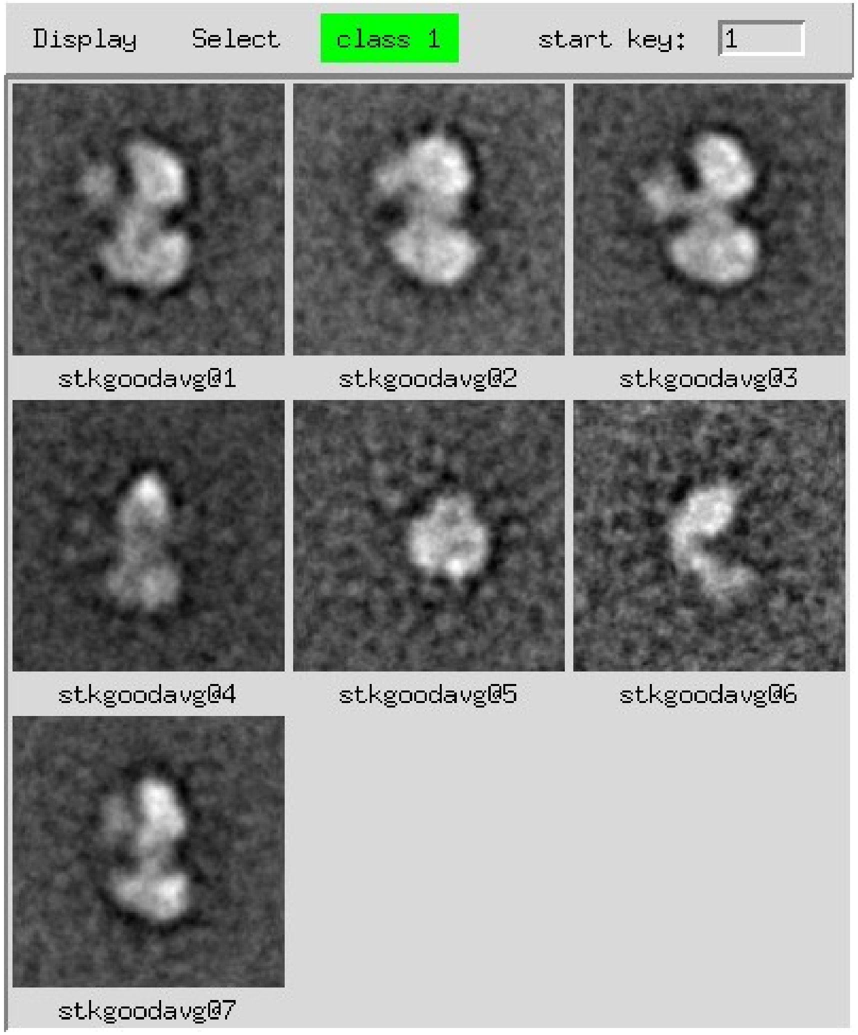
CEITEC

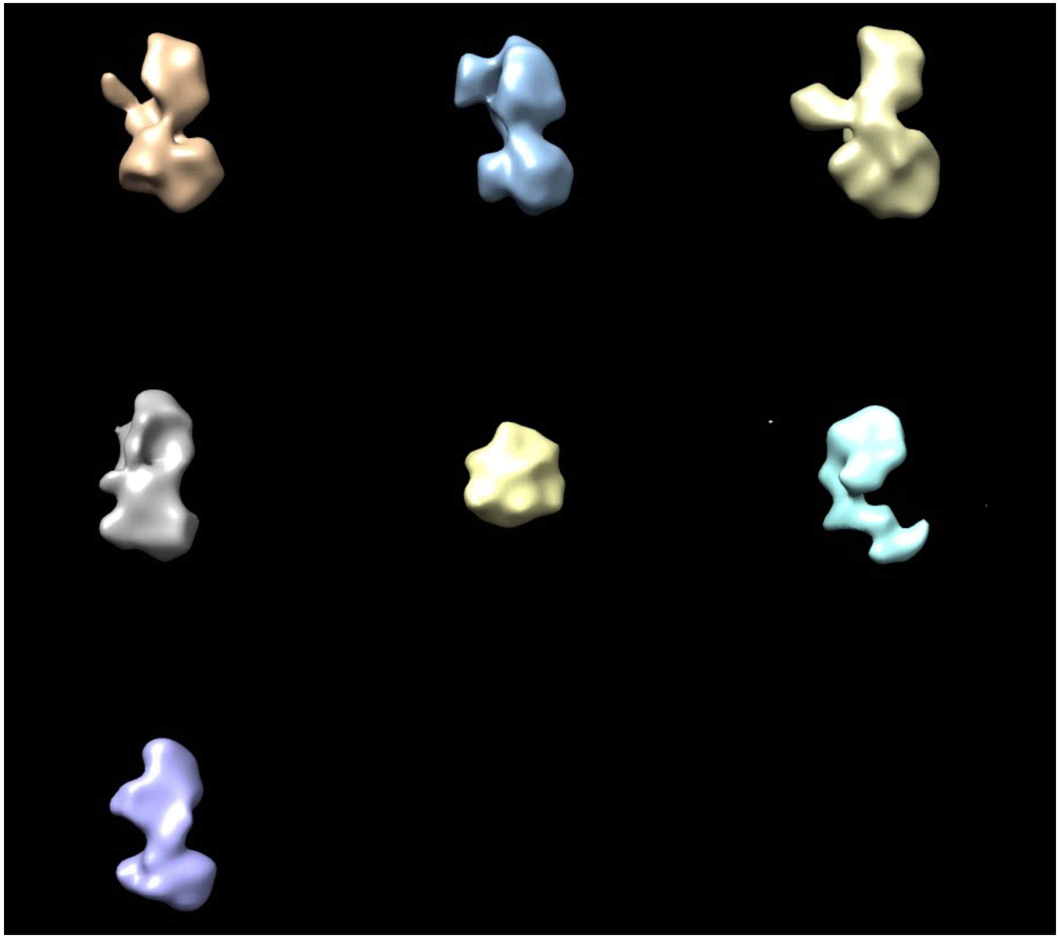
We compute a separate reconstruction for each class



IF the classes simply correspond to different orientations, you can combine them, and boost the signal-to-noise.

Helicase G40P





If the classes correspond to different conformations, then you have to keep them as separate reconstructions.

CEITEC

Outline

Image analysis III

- More on last week's material
  - Dependence of SNR on $\sqrt{N}$
  - Oversampling
- Classification

3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

More properties of Fourier transforms: Convolutions

Why might two images in a data set look different?

- different sample
- different magnification
- different illumination
- different orientations
- different defocus
- different conformations

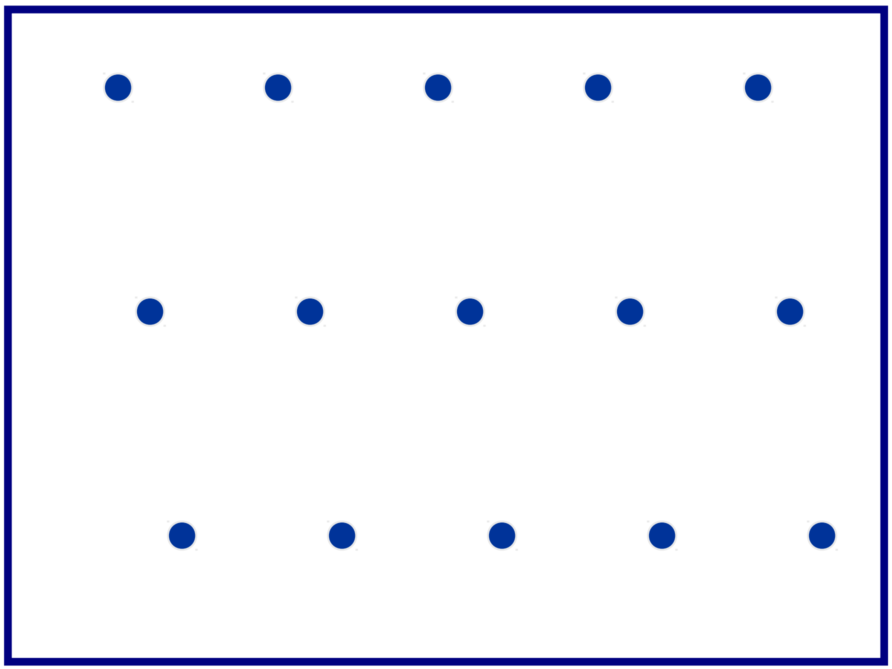
- better biochemistry
- better microscopy
- normalization
- determine angles
- CTF correction
- Classification

CEITEC

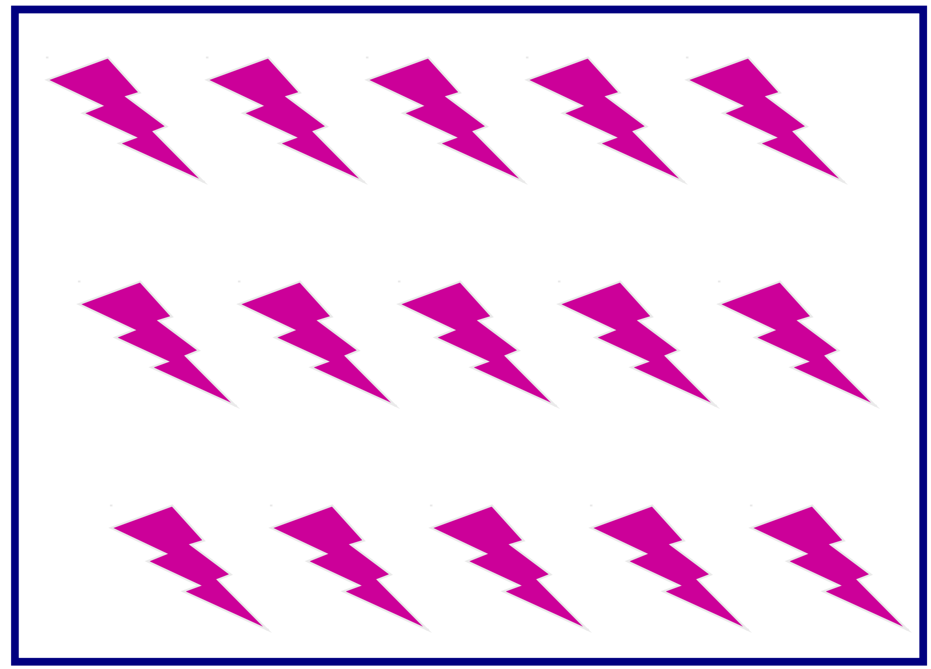
Convolution of a molecule with a lattice generates a crystal.

Notation: $f(x) \bullet g(x)$

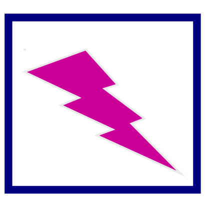
Adapted from David DeRosier



lattice: $f(x)$



Set a molecule down at every lattice point.

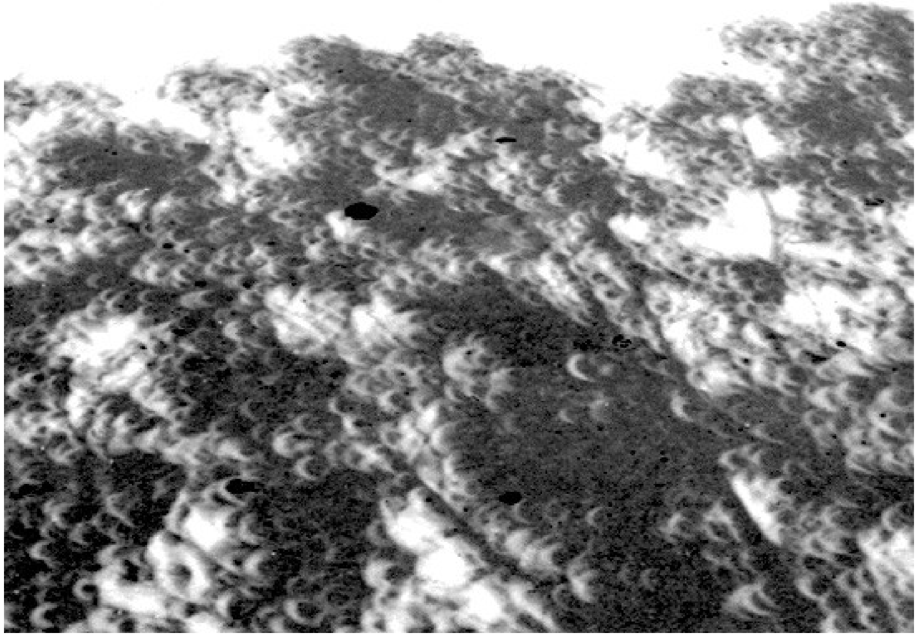


Molecule $g(x)$

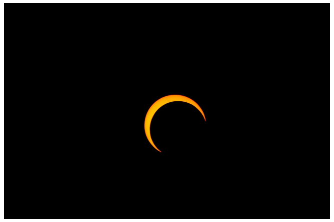
Convolution in real life

Notation: $f(x) \bullet g(x)$

lattice: $f(x)$
http://www.photos-public-domain.com



http://www.symbolicmessengers.com



Molecule: $g(x)$
http://en.wikipedia.org

Set a molecule down at every lattice point.

Cross-correlation vs. convolution

Complex conjugate:
If a Fourier coefficient $F(X)$ has the form: $a + bi$
The complex conjugate $F^*(X)$ has the form: $a - bi$

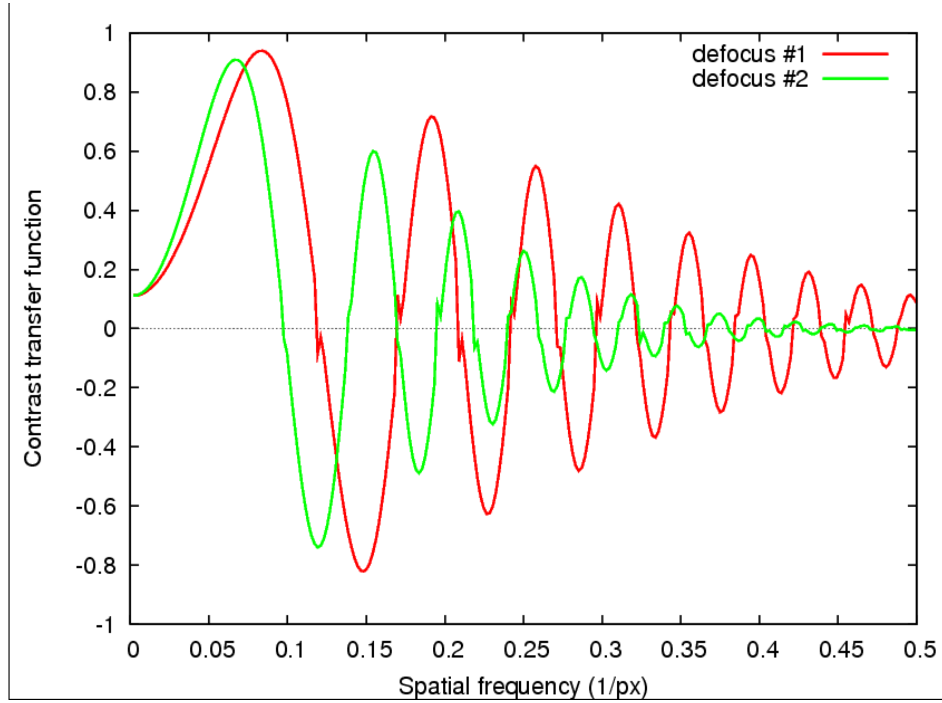
Cross-correlation: $F^*(X)\ G(X)$

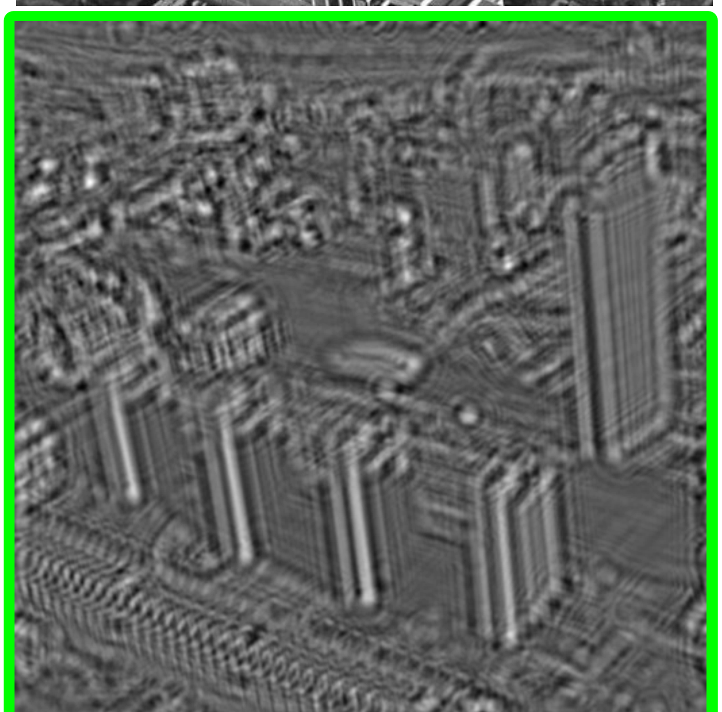
Convolution: $F(X)\ G(X)$

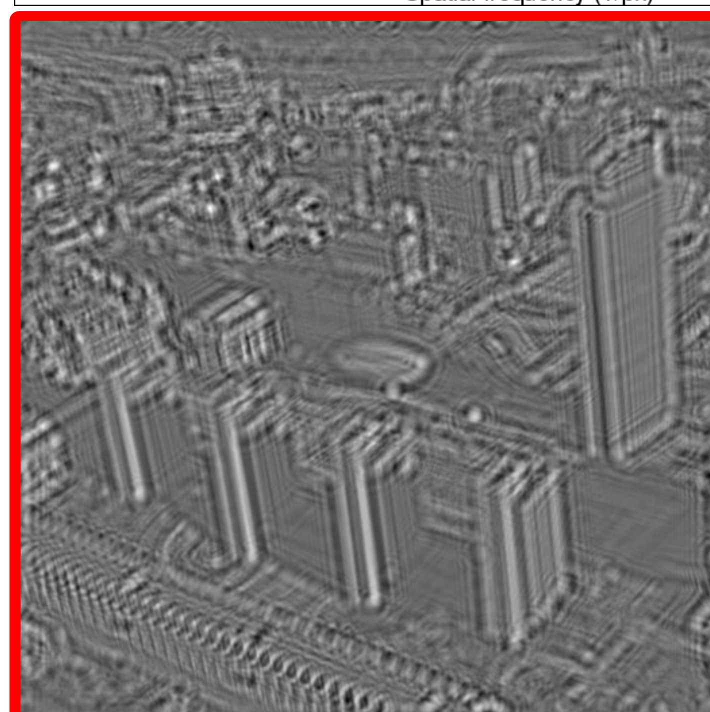
Remember:
$f(x)$, $g(x)$ are real-space functions
$F(X)$, $G(X)$ are Fourier-space functions

CEITEC

original



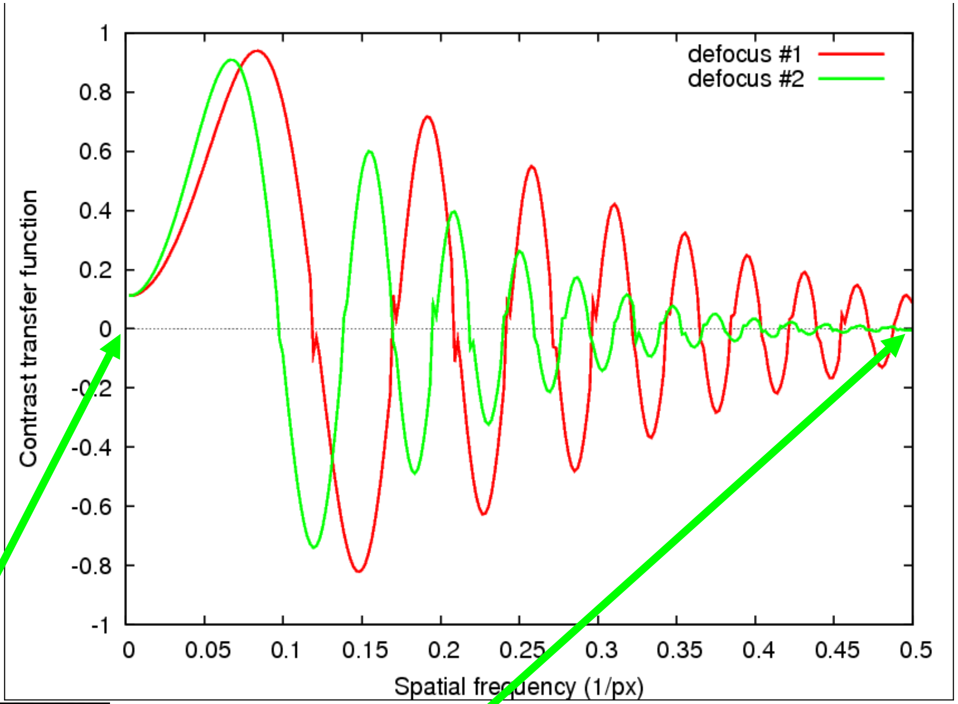


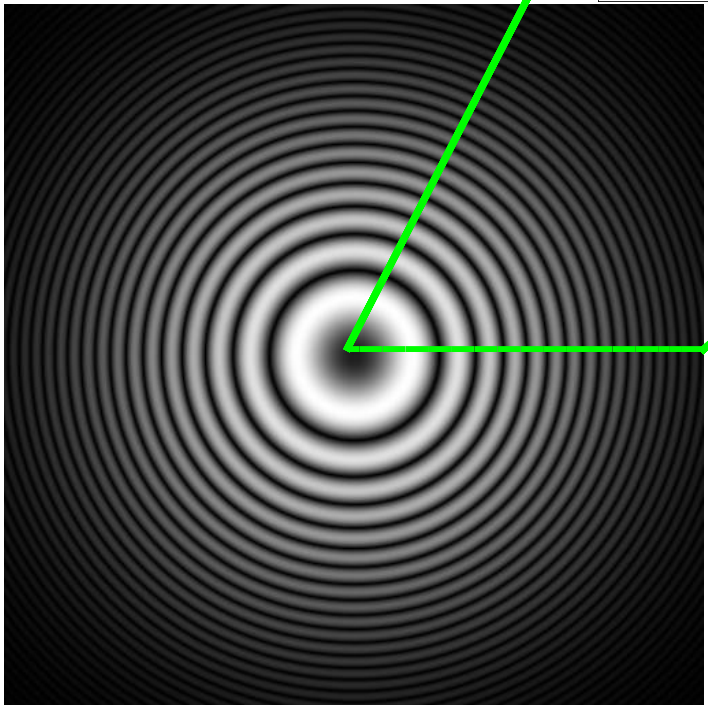


CEITEC

CTF

1D profile

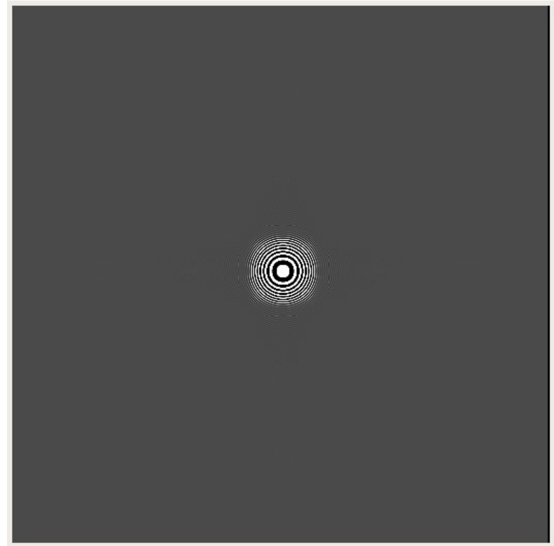




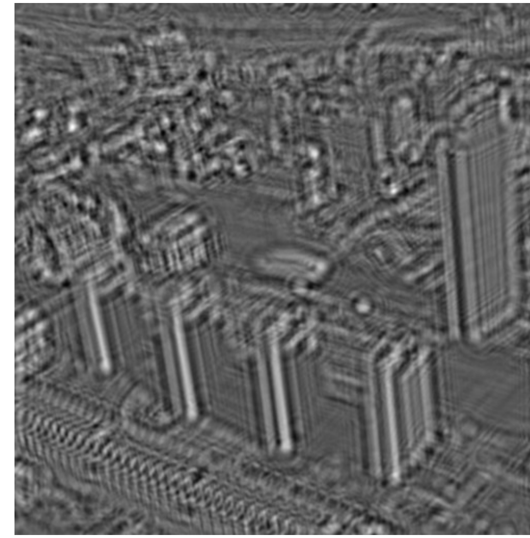
2D power spectrum

G(X)

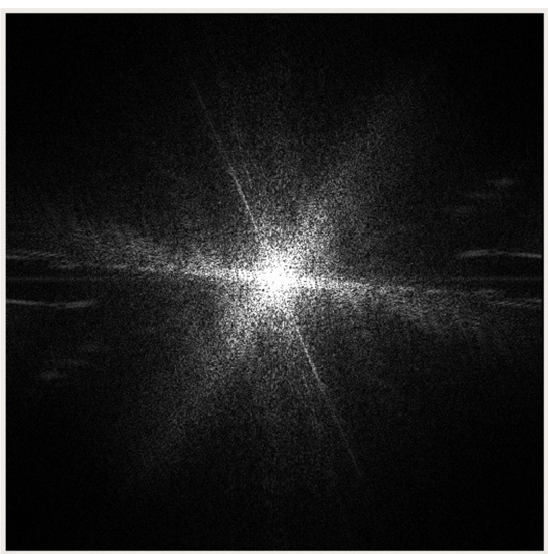
f(x)



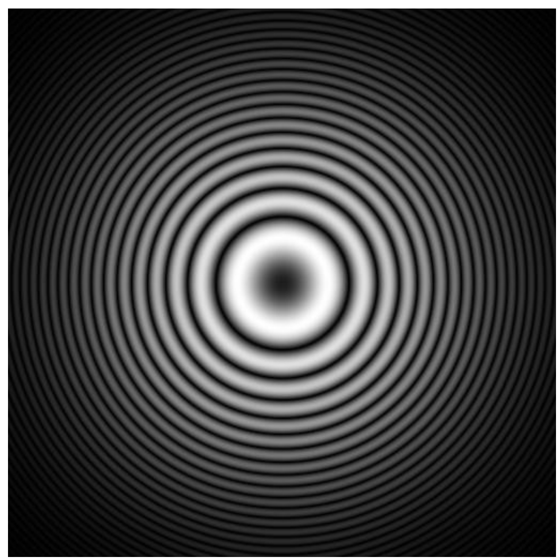
g(x)



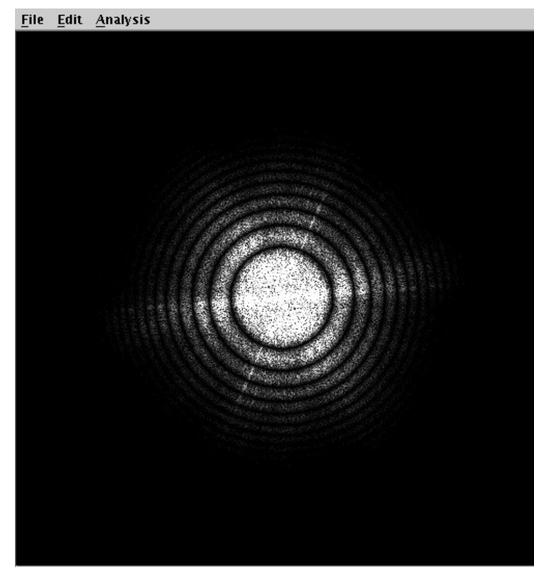
f(x)•g(x)



F(X)

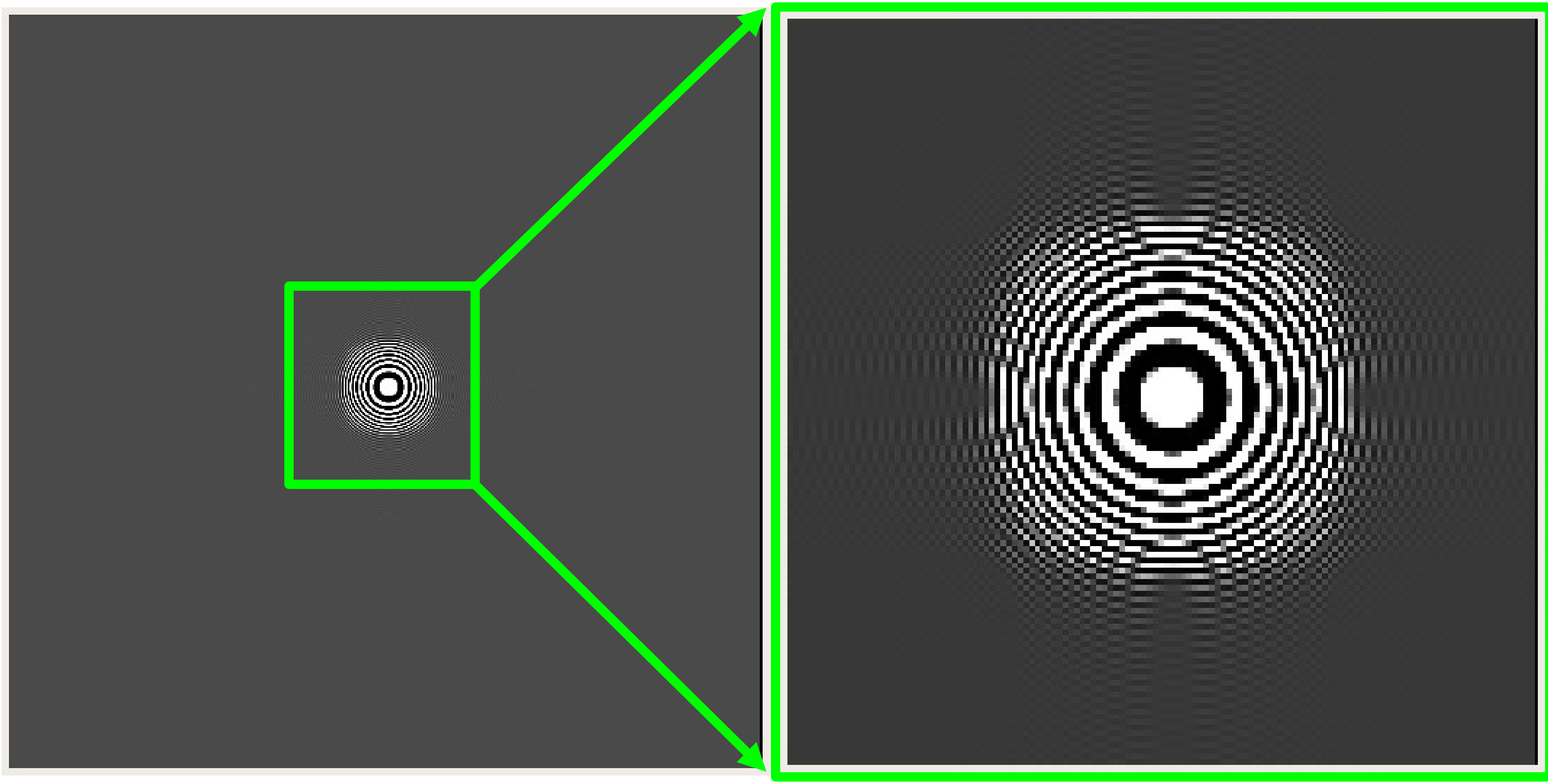


G(X)



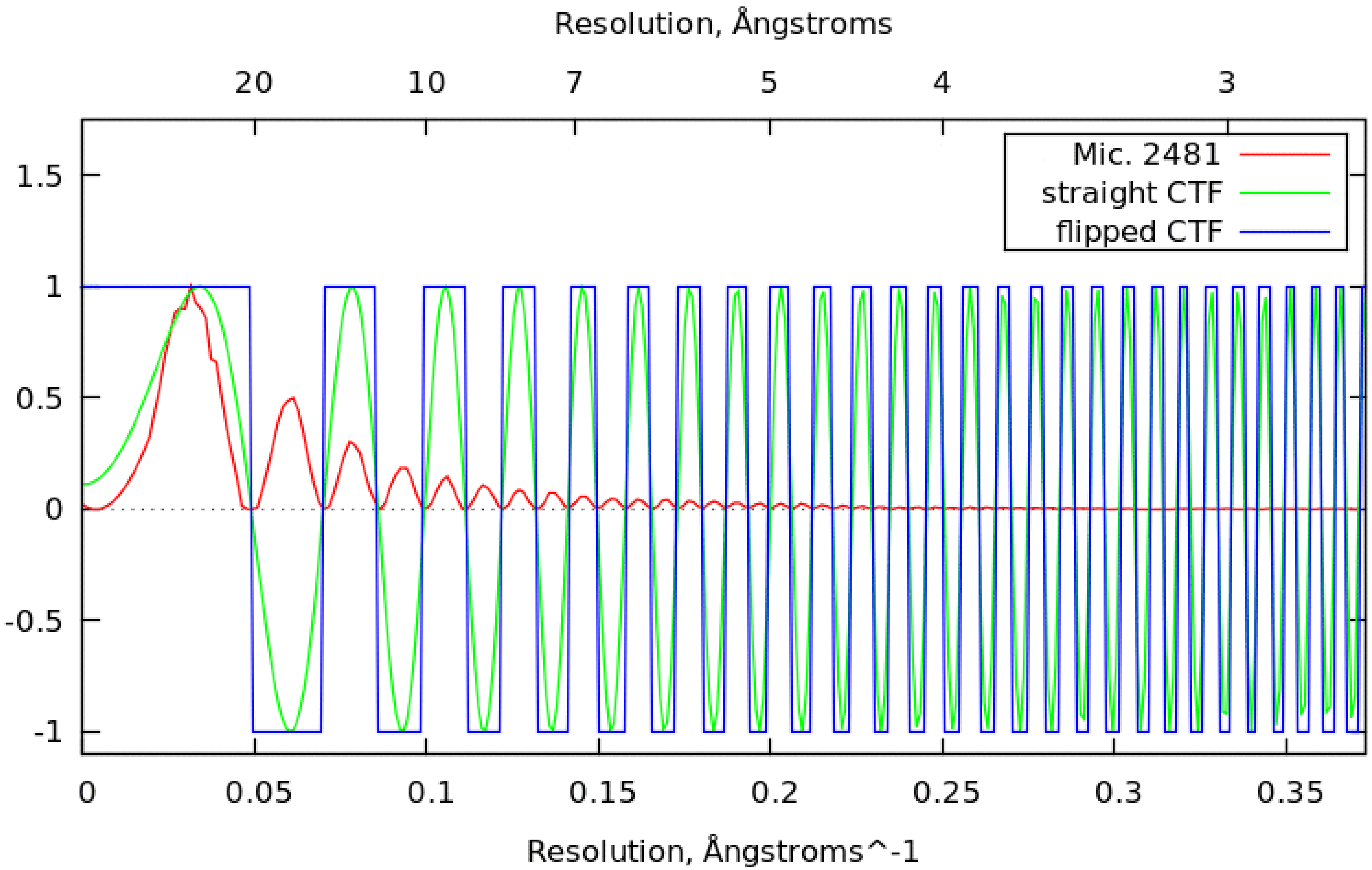
F(X) G(X)

Point spread function



g(x)   zoomed

An ideal point spread function would be an infinitely-sharp point.

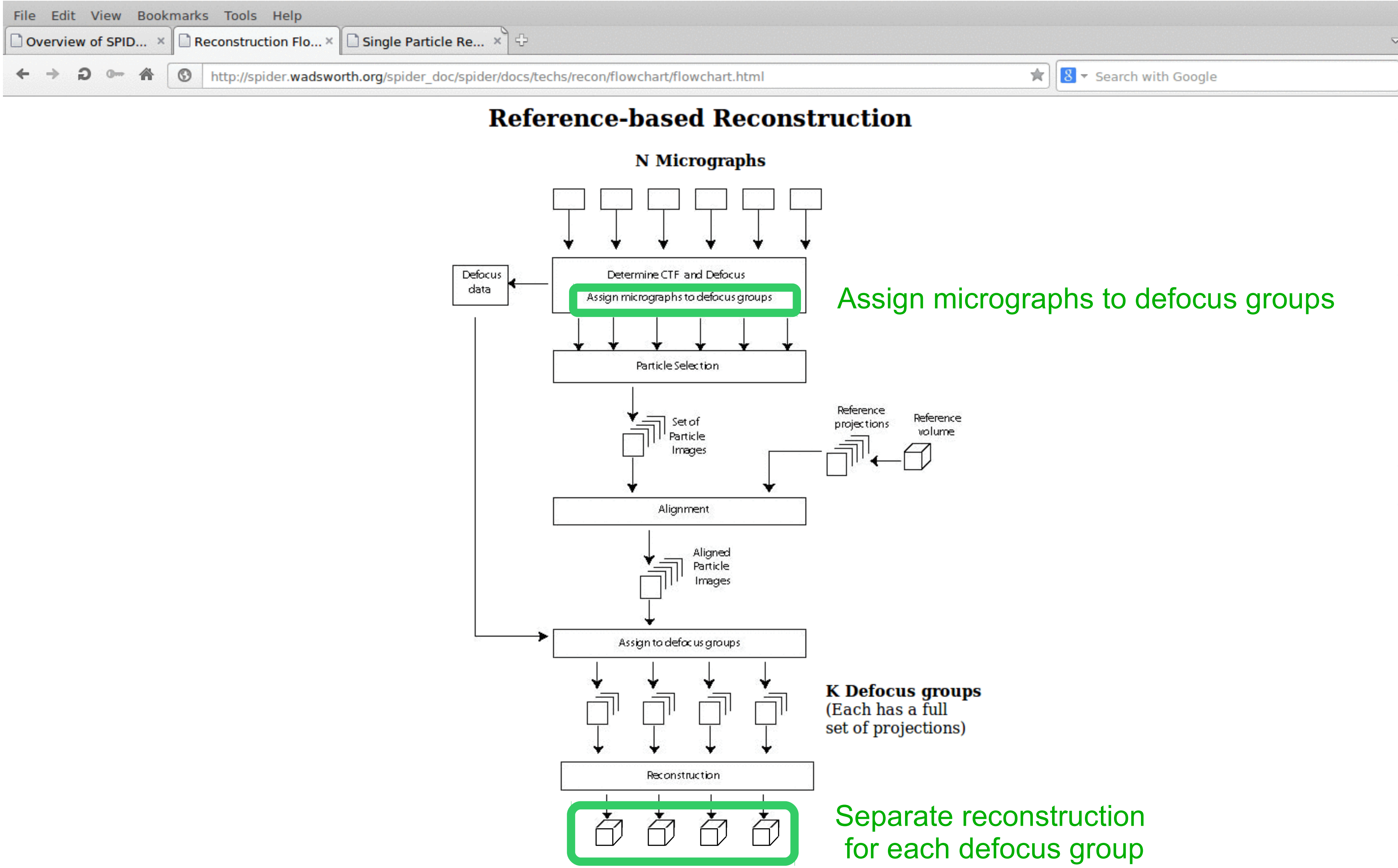


Red: Power-spectrum profile calculated from experimental image

Green: Fitted, theoretical power-spectrum profile

Blue: Phase-only correction profile

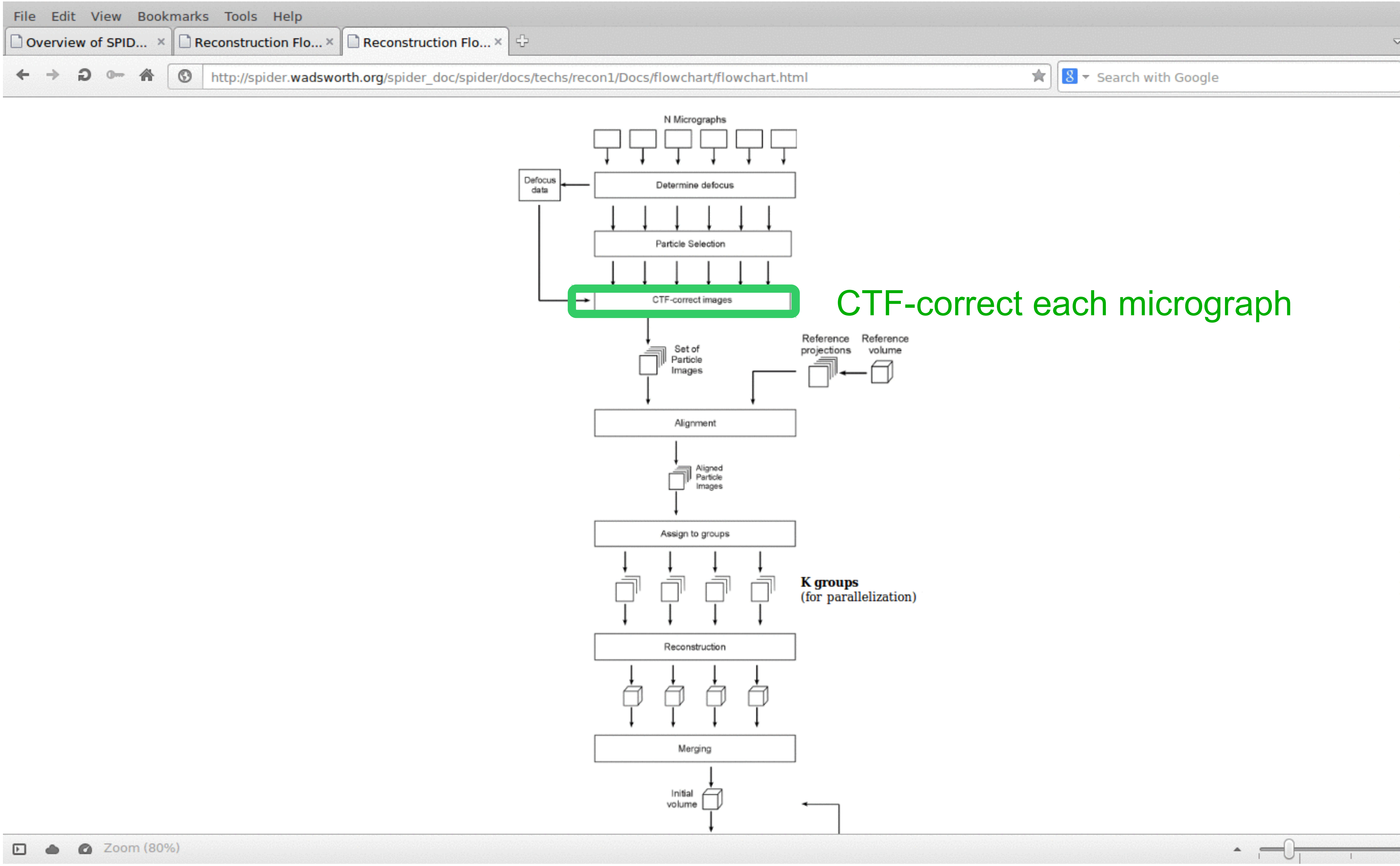
Defocus groups: CTF correction in 3D



Assign micrographs to defocus groups

Separate reconstruction for each defocus group

CTF-correction of micrographs in 2D



CTF-correct each micrograph

Outline

Image analysis III

- More on last week's material
  - Dependence of SNR on √N
  - Oversampling
- Classification

3D Reconstruction

- Principles
- Tomography
- Reference-based alignment
- Common lines
- RCT
- CTF-correction
- 3D classification

Why might two images in a data set look different?

- different molecule
- different magnification
- different illumination
- different orientations
- different defocus
- different conformations

- better biochemistry
- better microscopy
- normalization
- determine angles
- CTF correction
- Classification

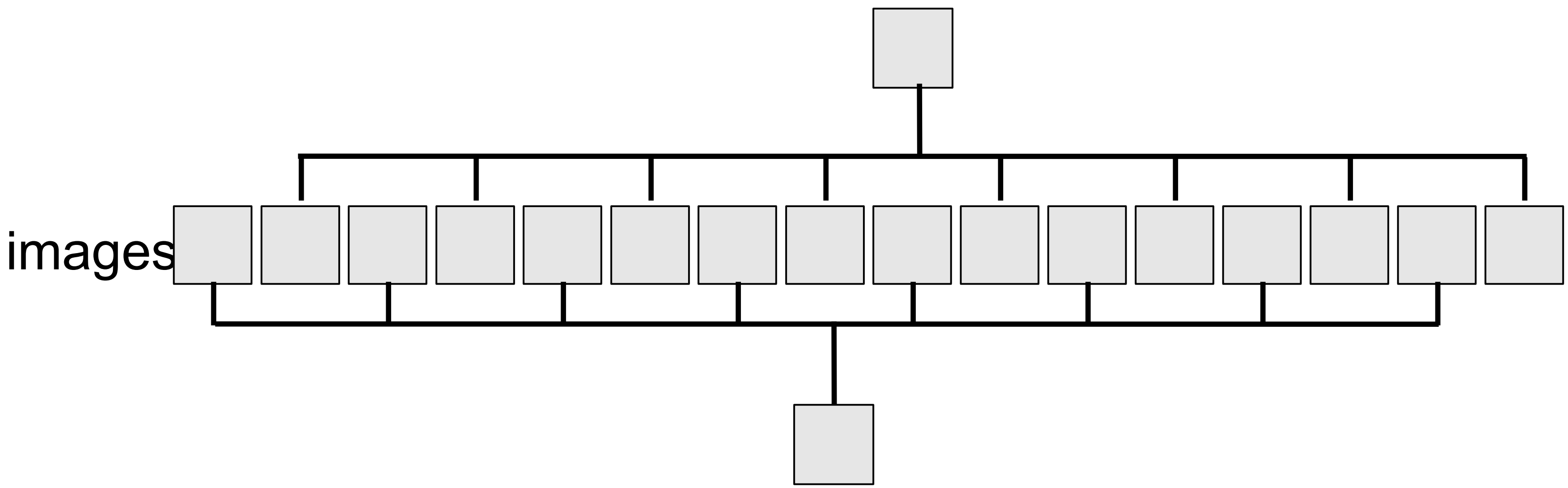
Classification: Reference-based classification vs. Maximum likelihood (ML3D)

Reference-based classification:	ML3D
• Possible conformations must be known. • The combination of parameters (shift, rotation, class) is chosen from the highest correlation value. • Possible reference bias	• Possible conformations are not known. • The probability of the occurrence of the parameters (shift, rotation, class) is maximized. • Random, data-dependent

RELION is a variation of maximum likelihood.

Seeding ML3D classification

We split the data set into *K* classes at random.



There will be slight differences in the reconstructions. We will iteratively maximize the likelihood of a particle belonging to a particular class.

Thank you for your attention

CEITEC

Central European Institute of Technology
Masaryk University
Kamenice 753/5
625 00 Brno, Czech Republic

www.ceitec.muni.cz | info@ceitec.muni.cz

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

2007–13
OP Research and
Development for Innovation