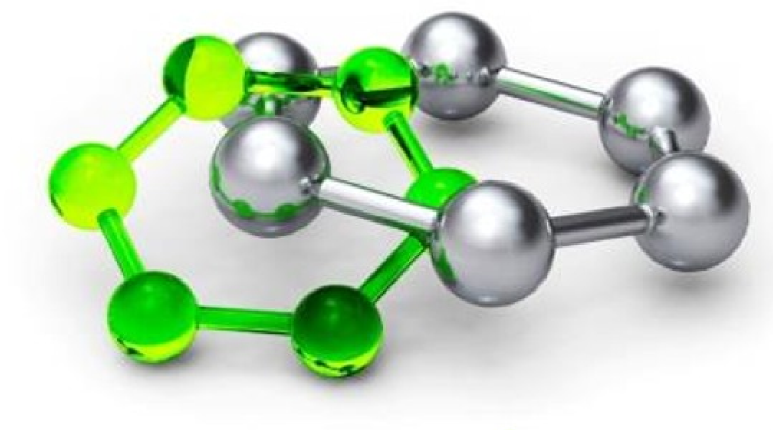
CEITEC

Central European Institute of Technology
BRNO | CZECH REPUBLIC

Image analysis IV

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

April 24, 2017



EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

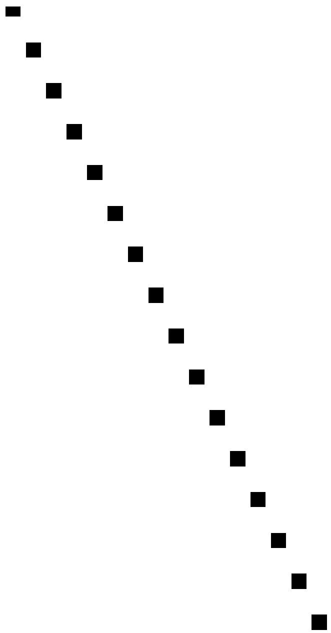
2007–13
OP Research and
Development for Innovation

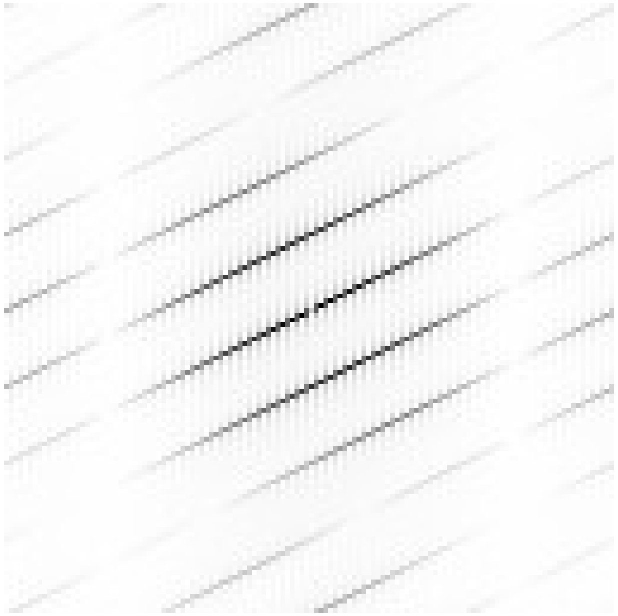
Outline

Image analysis III

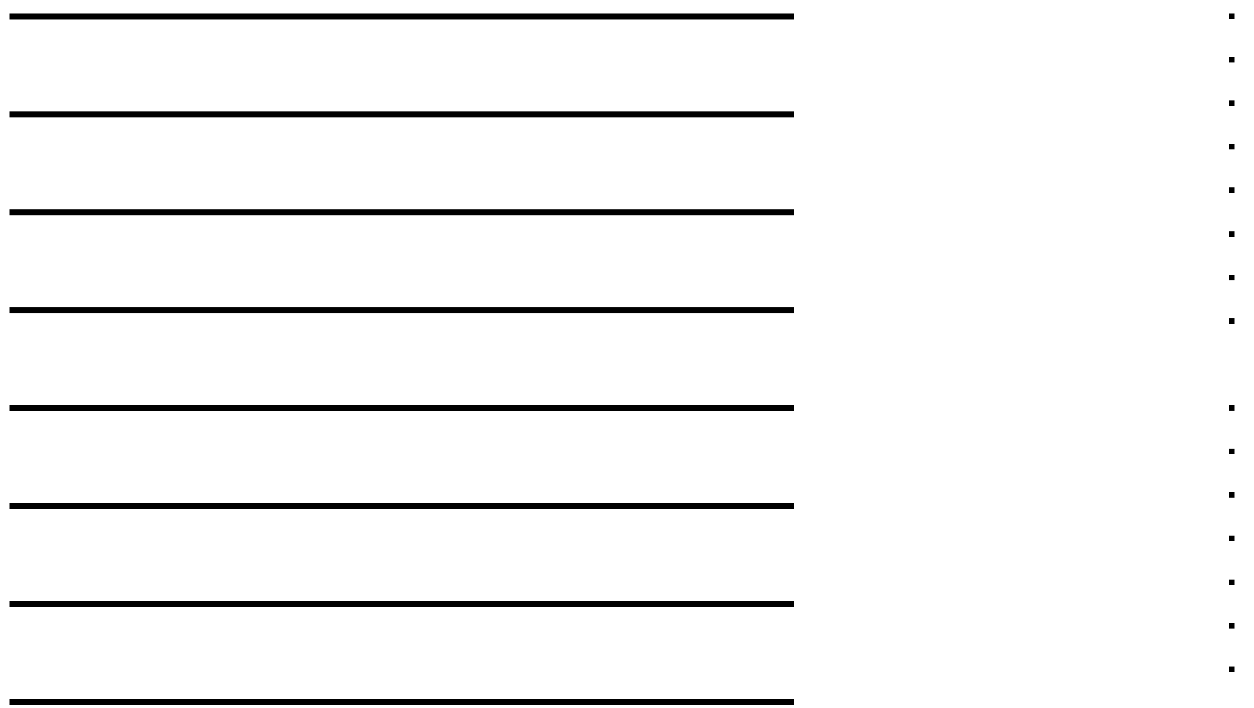
- More on FFTs
- Classification
    - Review of multivariate data analysis
    - Classification in 2D
    - Classification in 3D
- Resolution estimation
    - Fourier Shell Correlation
    - Expectation value of noise
    - “Gold standard” resolution

Some simple 2D Fourier transforms: a row of points

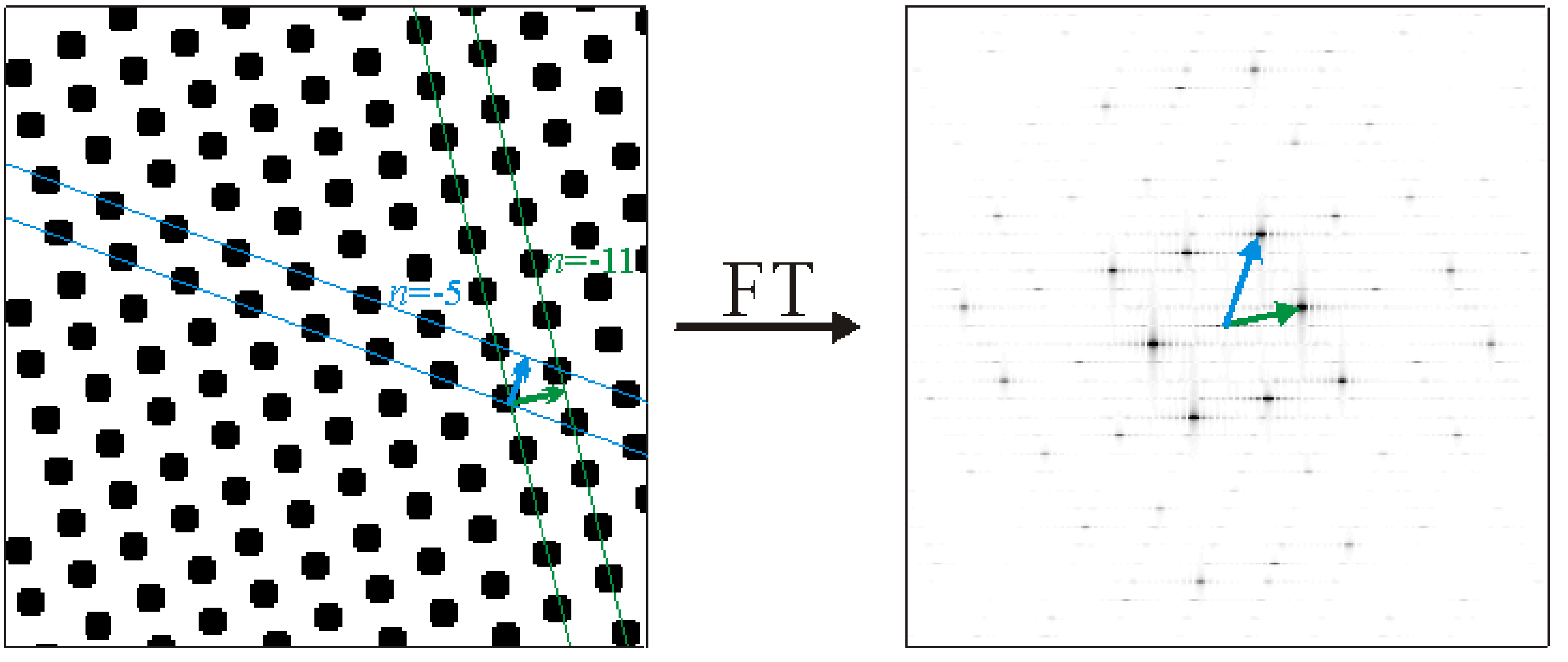




Some simple 2D Fourier transforms: a series of lines

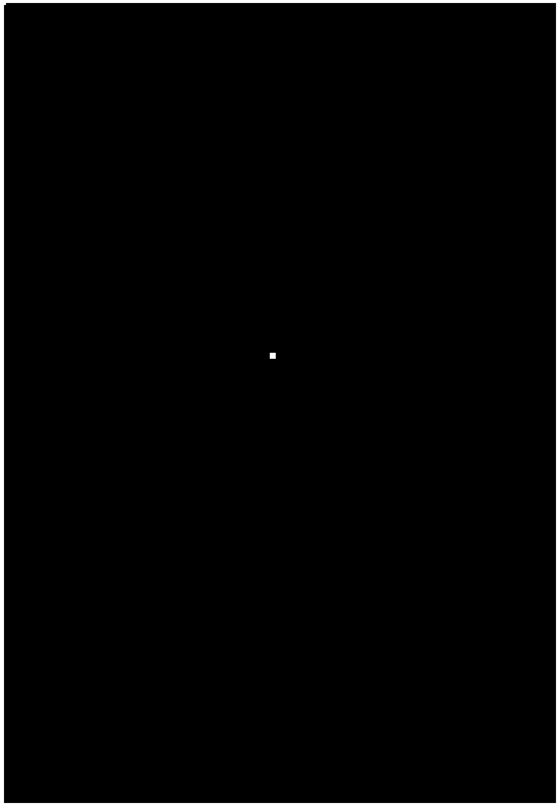


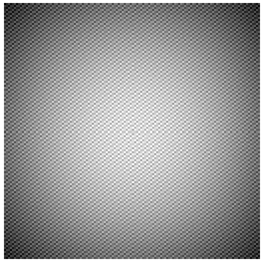
Some simple 2D Fourier transforms: a 2D lattice



CEITEC

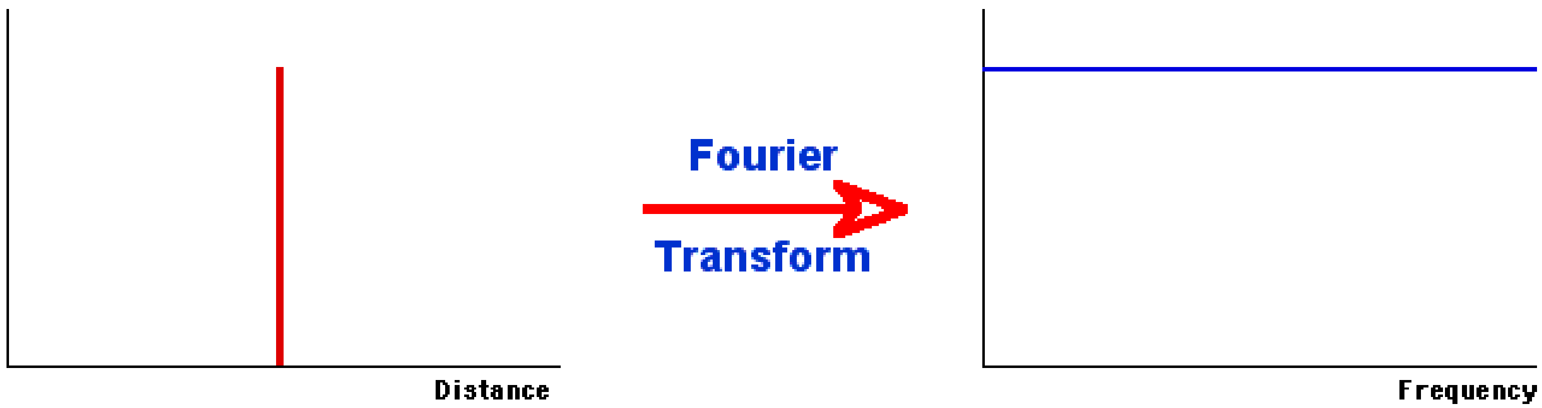
Single point





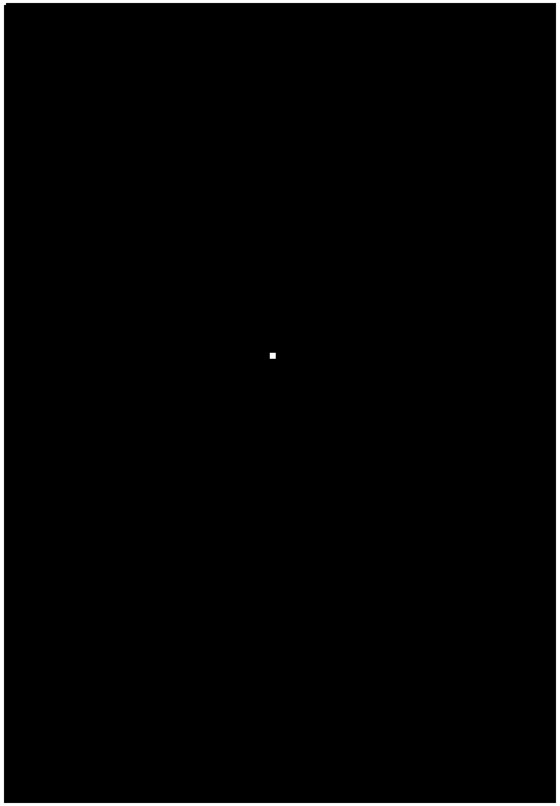
If the point was infinitely sharp,
the FFT would be flat.

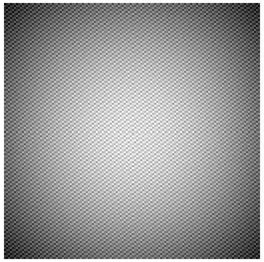
Some simple 1D transforms:
a sharp point (Dirac delta function)



http://en.labs.wikimedia.org/wiki/Basic_Physics_of_Nuclear_Medicine/Fourier_Methods

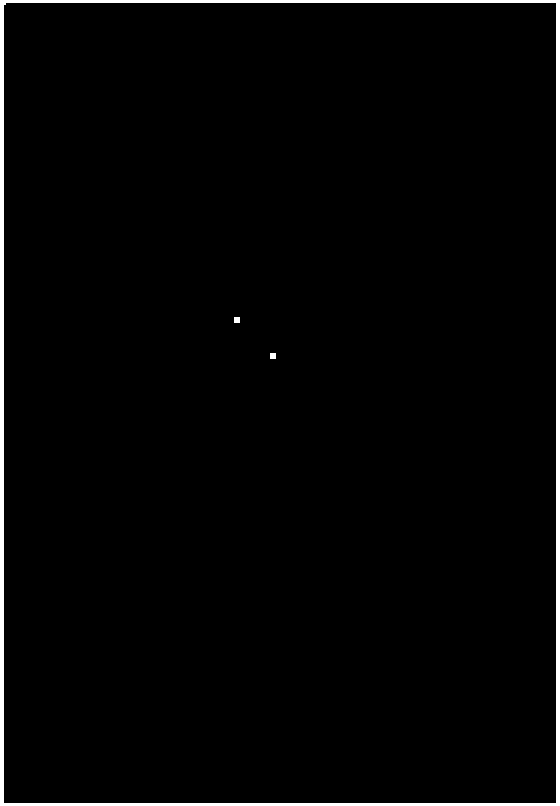
Single point

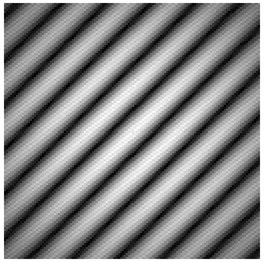




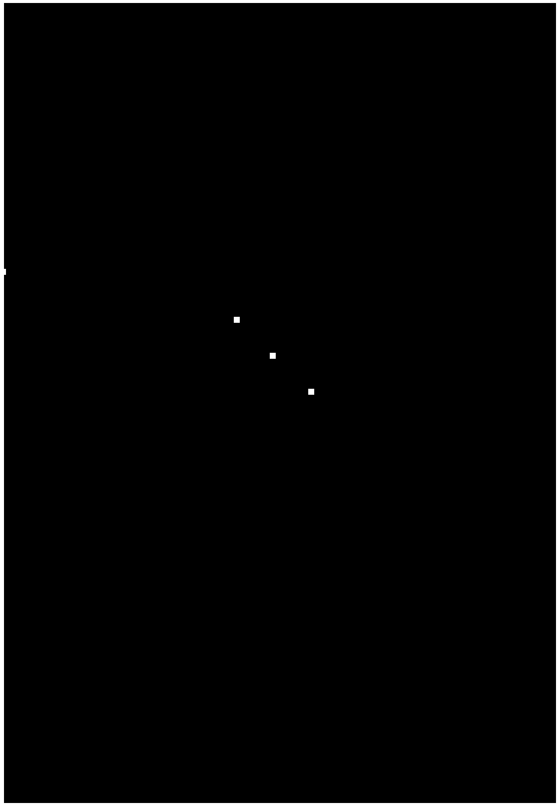
If the point was infinitely sharp,
the FFT would be flat.

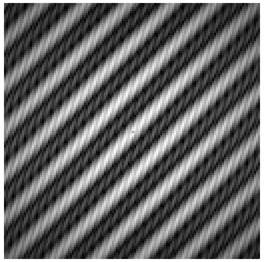
Two points



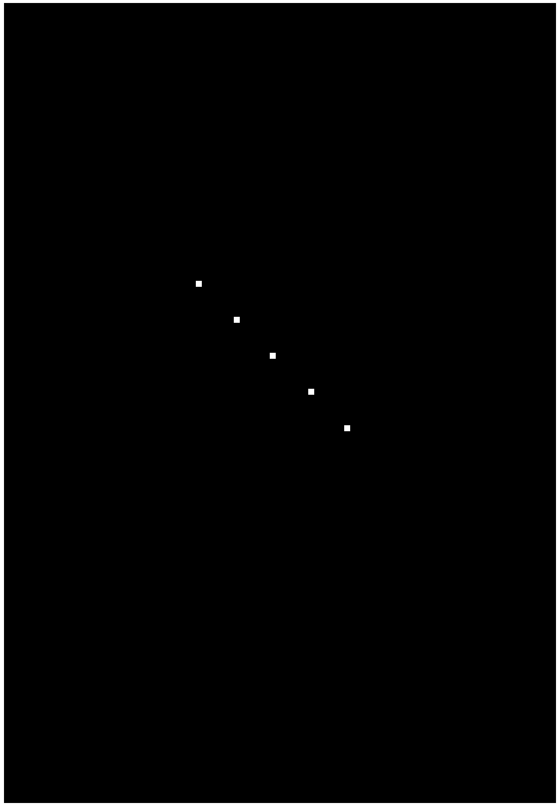


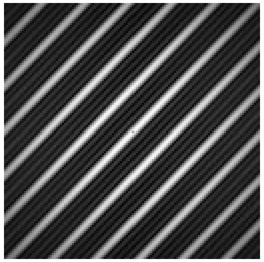
Three points



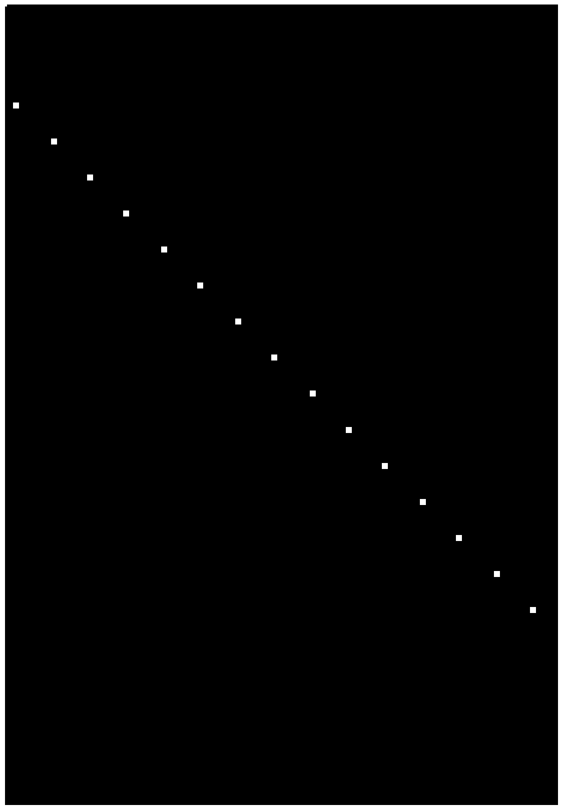


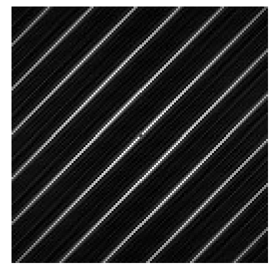
Five points



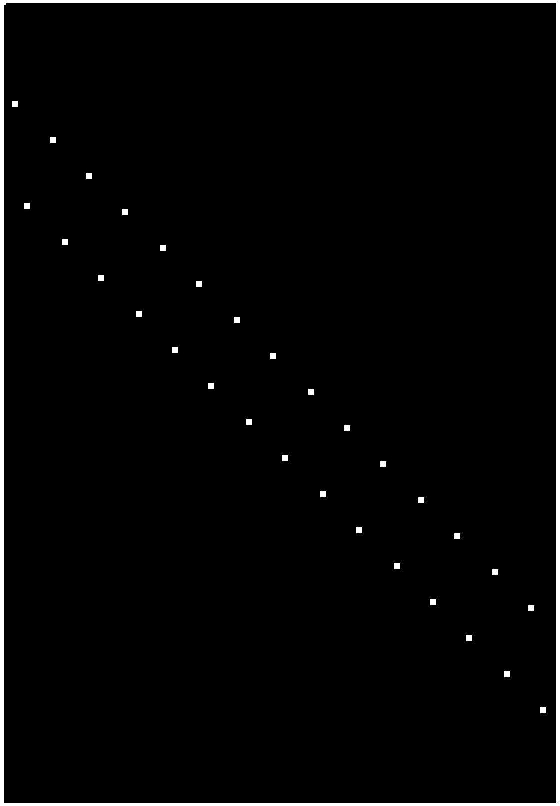


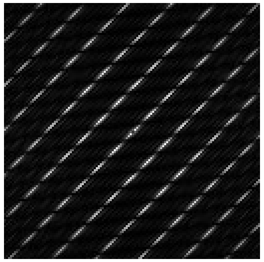
One row



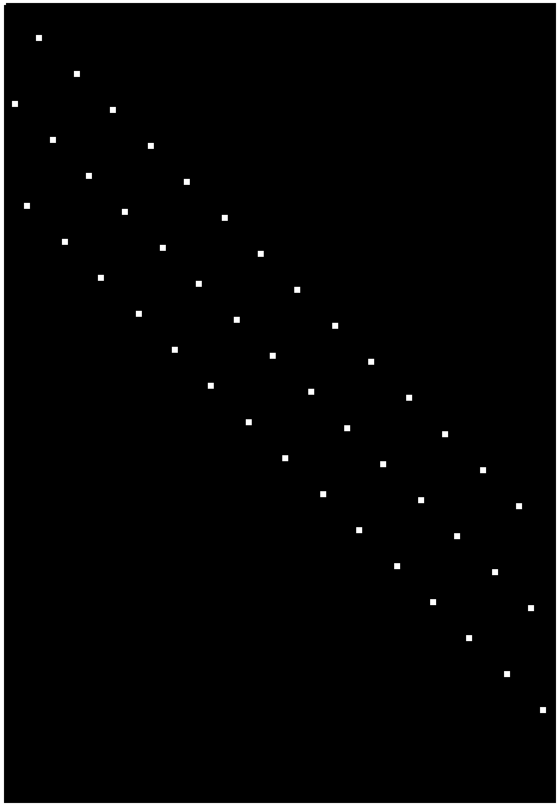


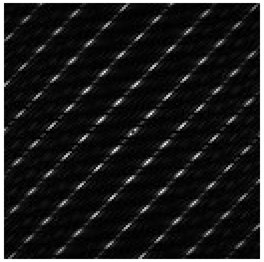
Two rows



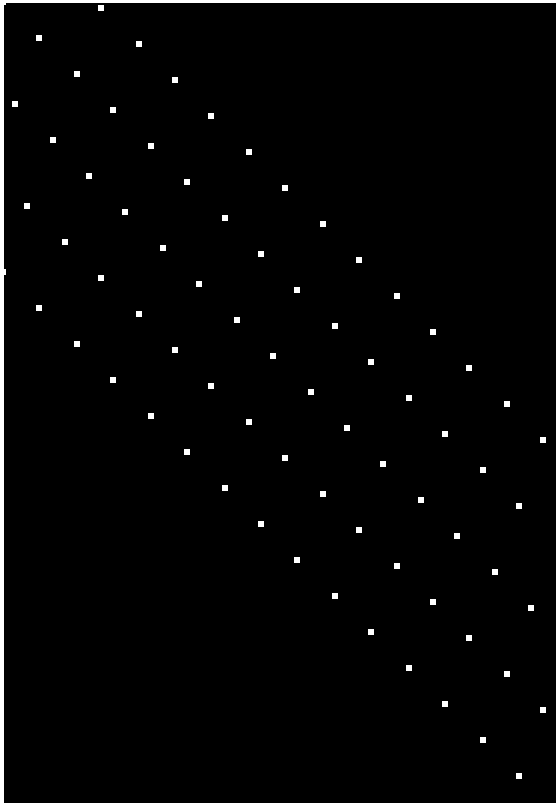


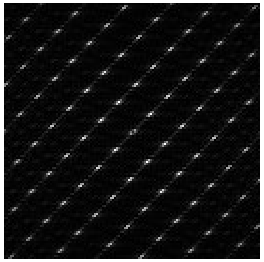
Three rows



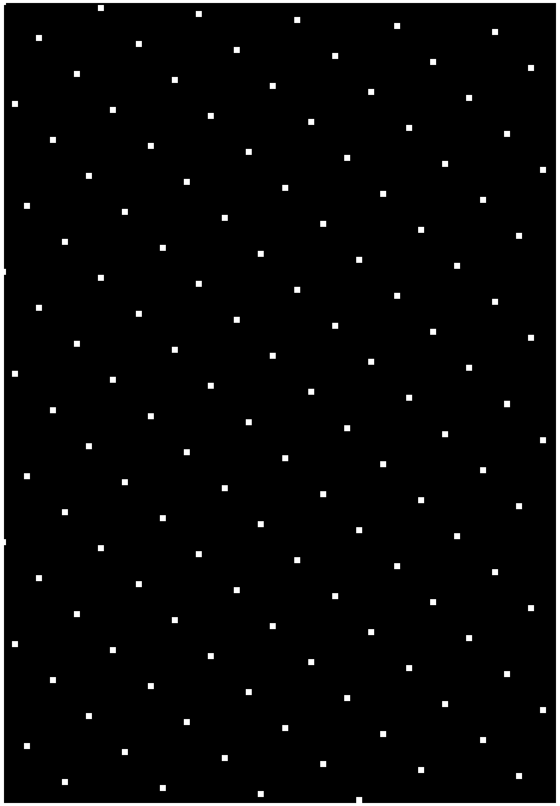


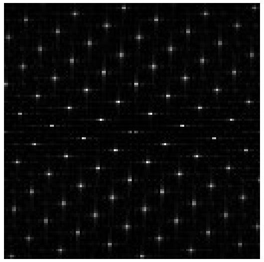
Five rows



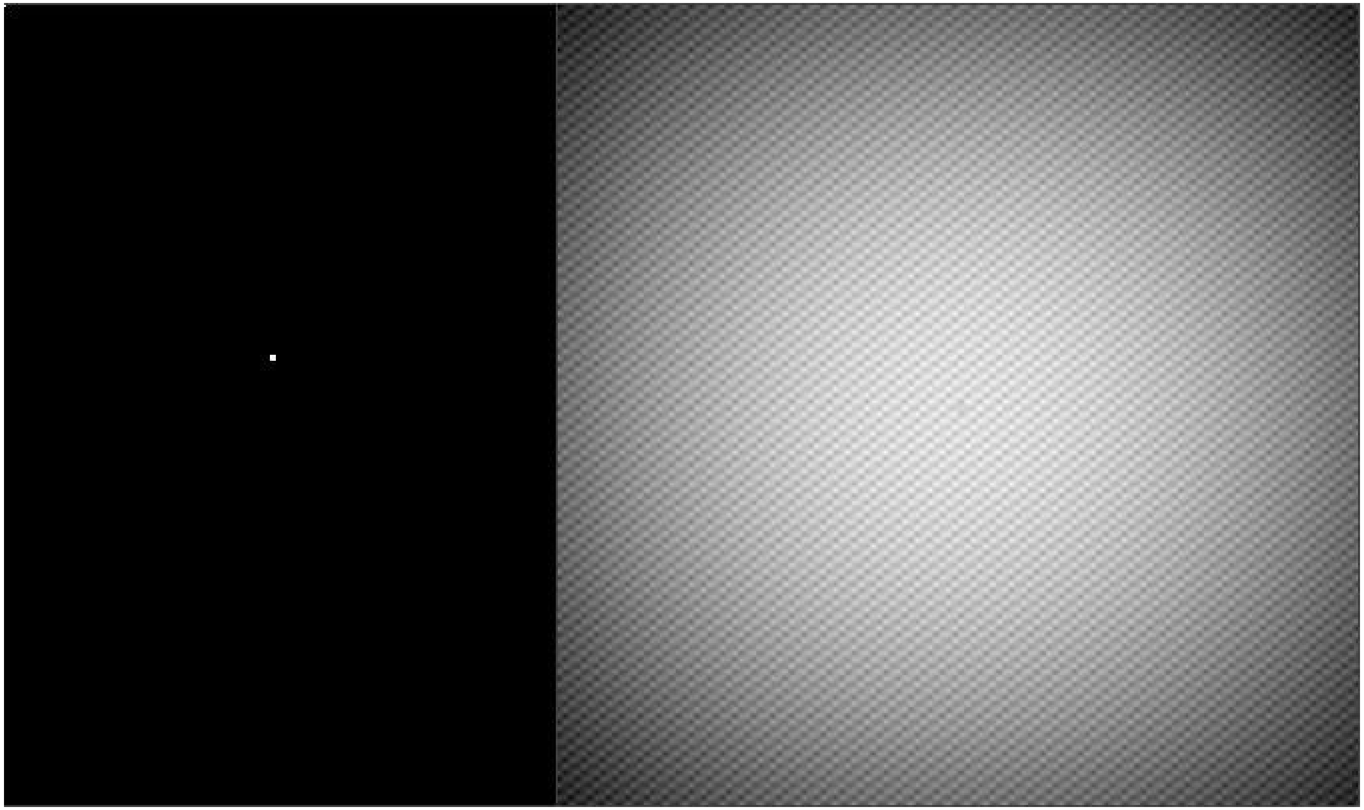


Full lattice

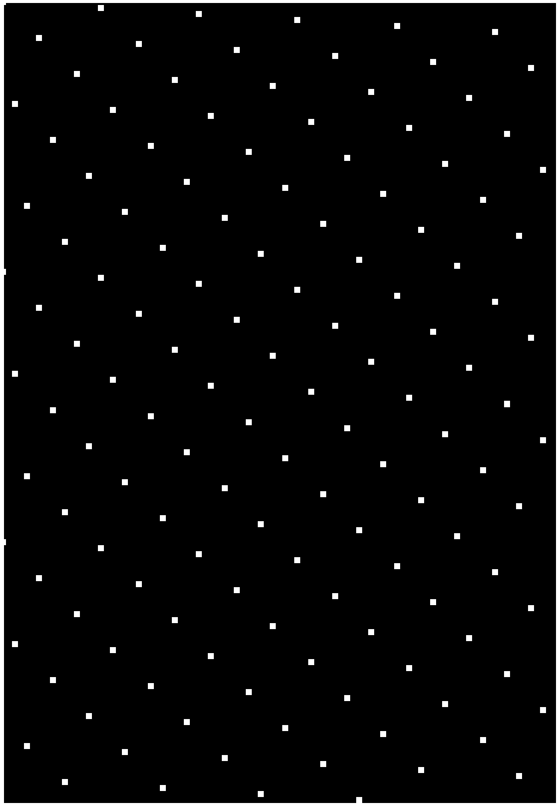


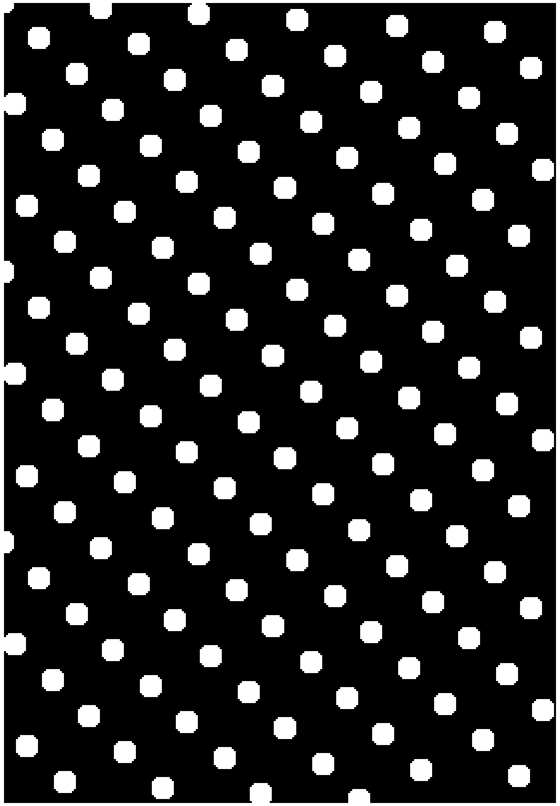


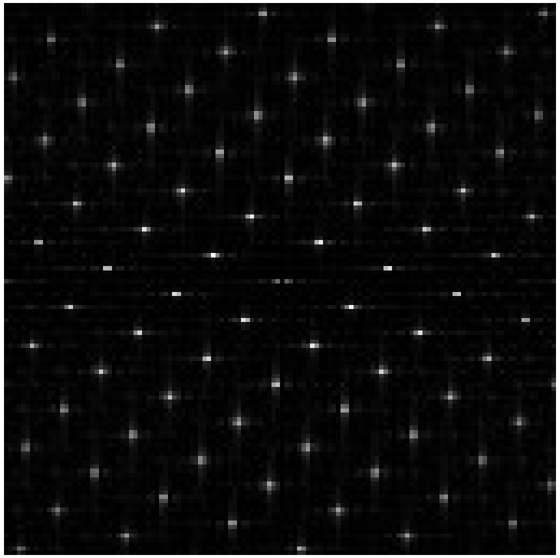
Animation



What if?



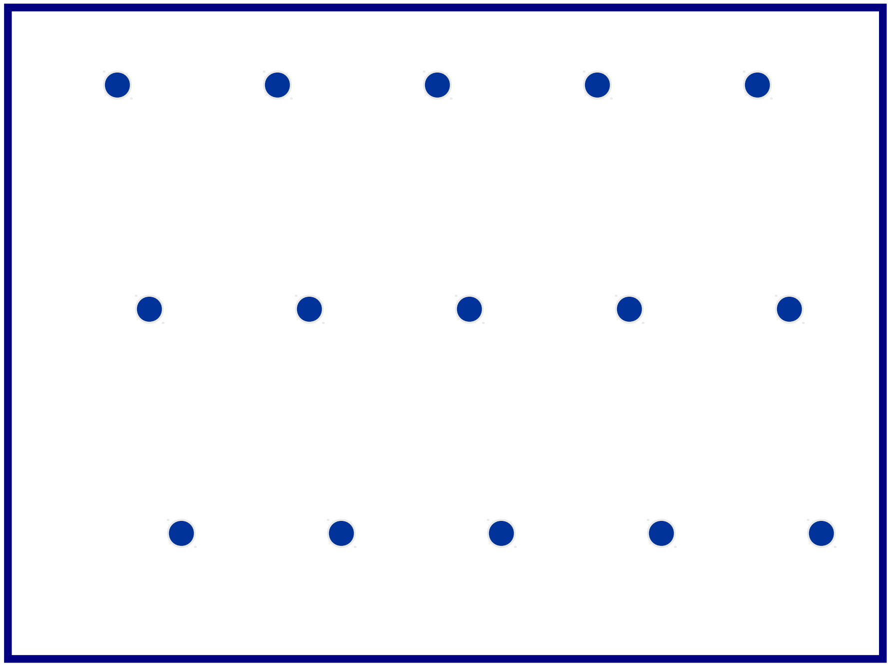




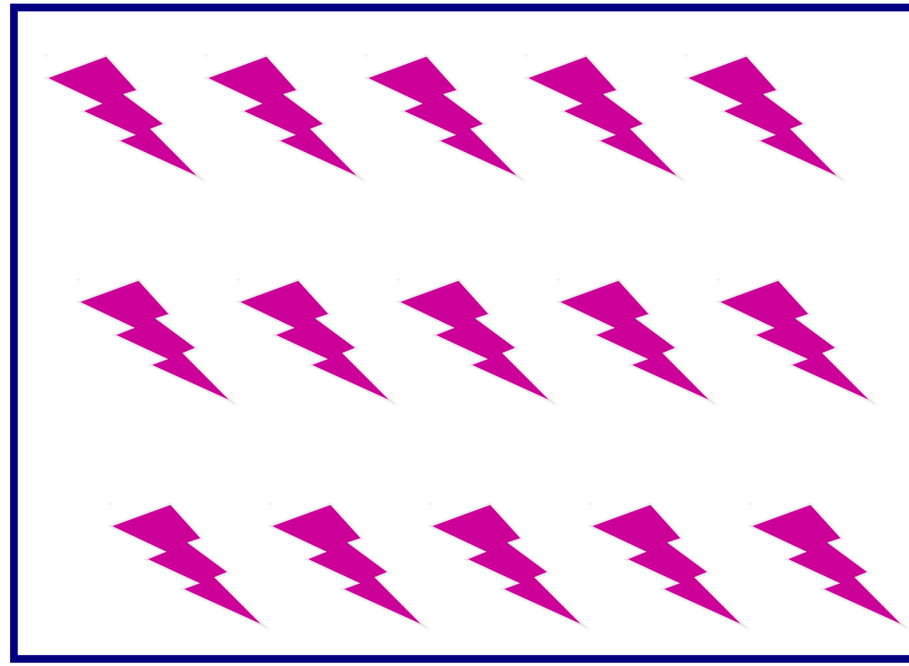
?

Convolution: a review

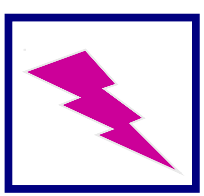
Adapted from David DeRosier



lattice: $f(x)$



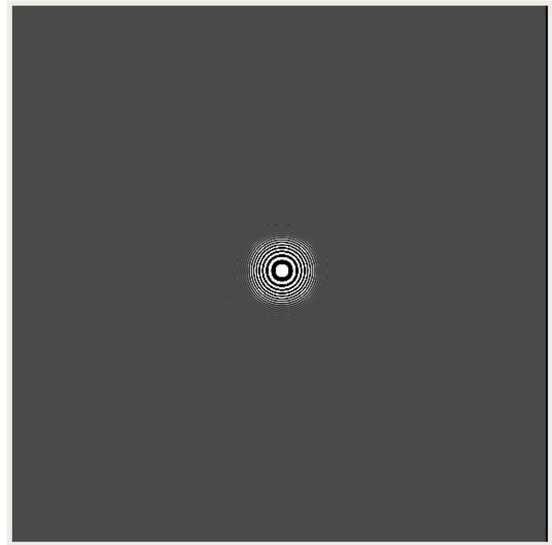
Set a molecule down at every lattice point.



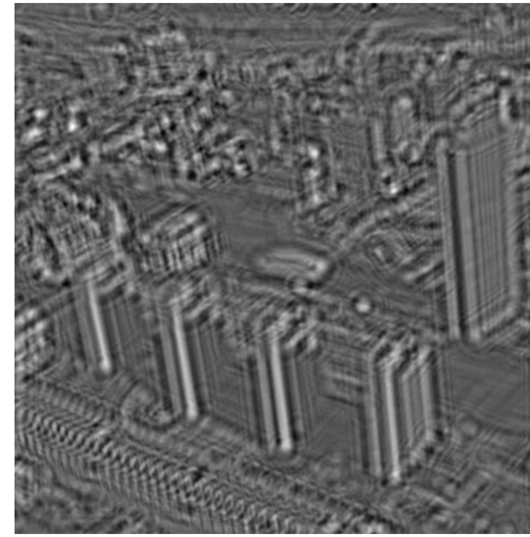
Molecule $g(x)$

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

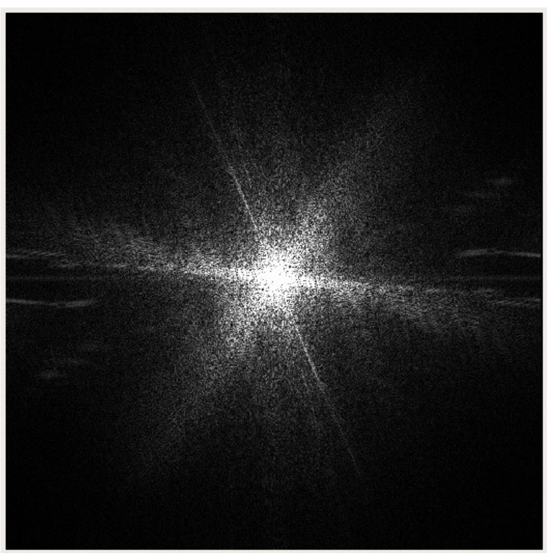
f(x)



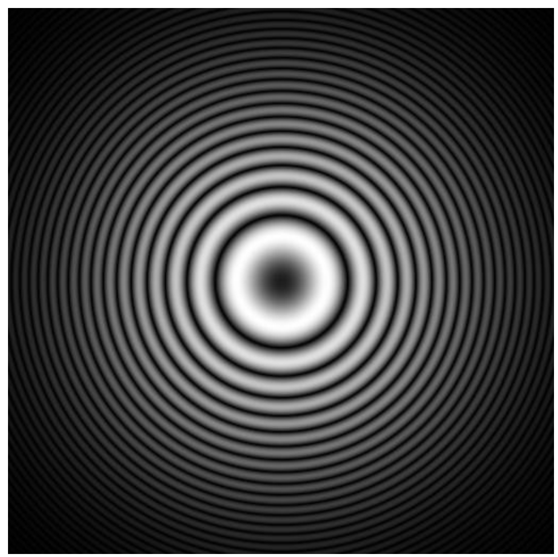
g(x)



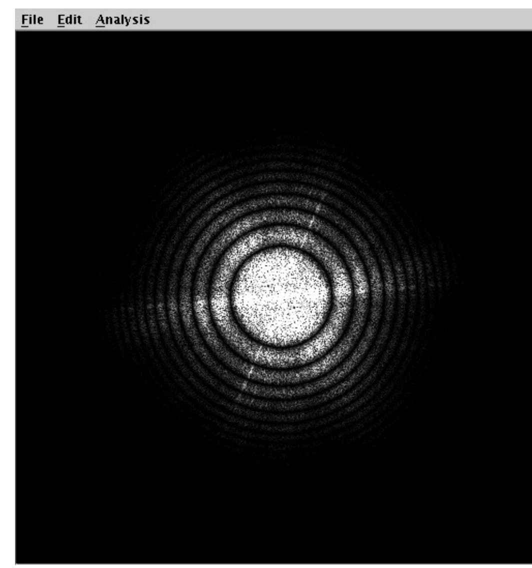
f(x)•g(x)



F(X)

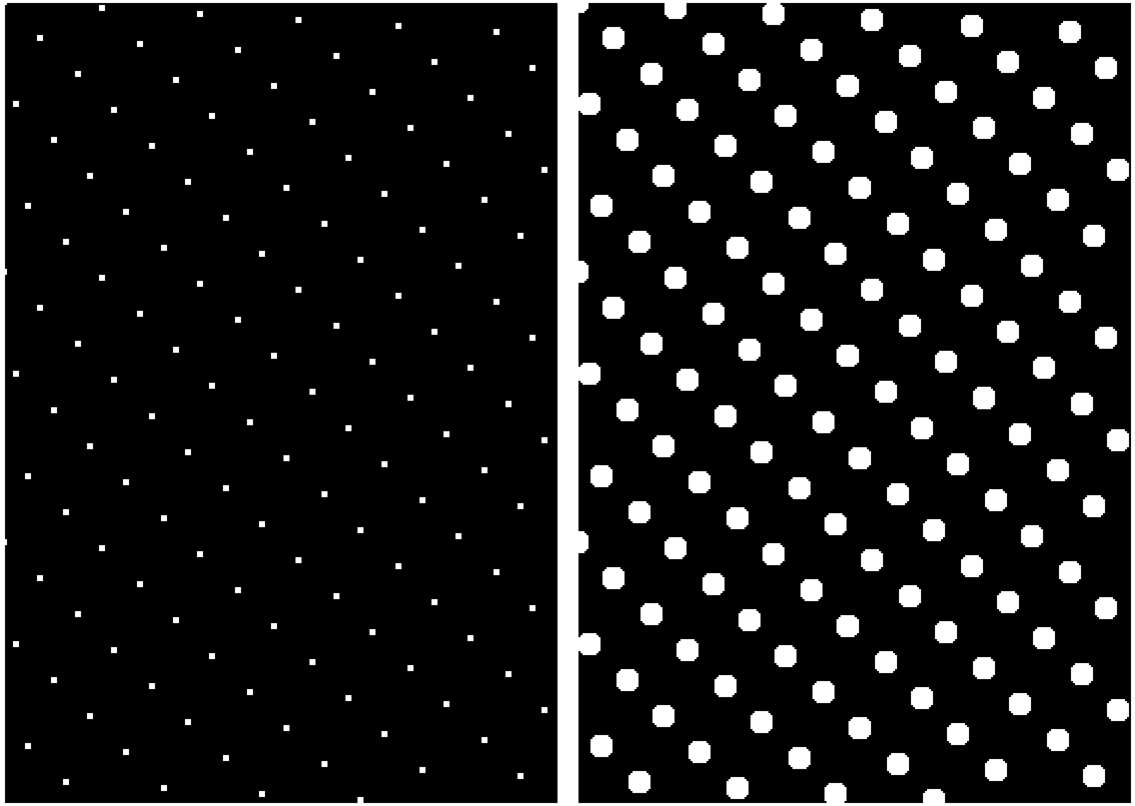


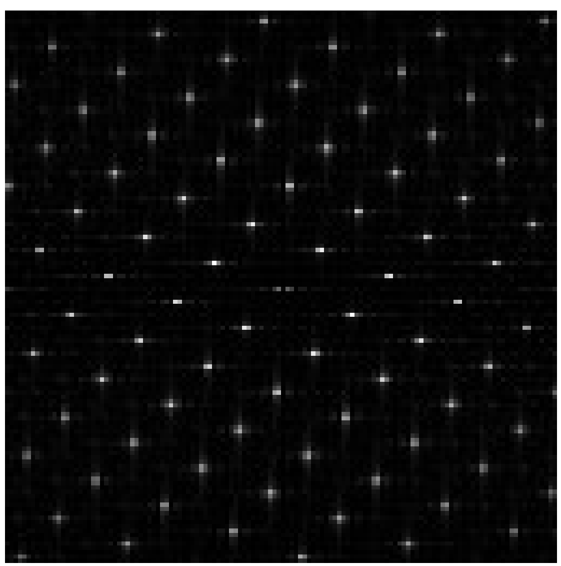
G(X)



F(X) G(X)

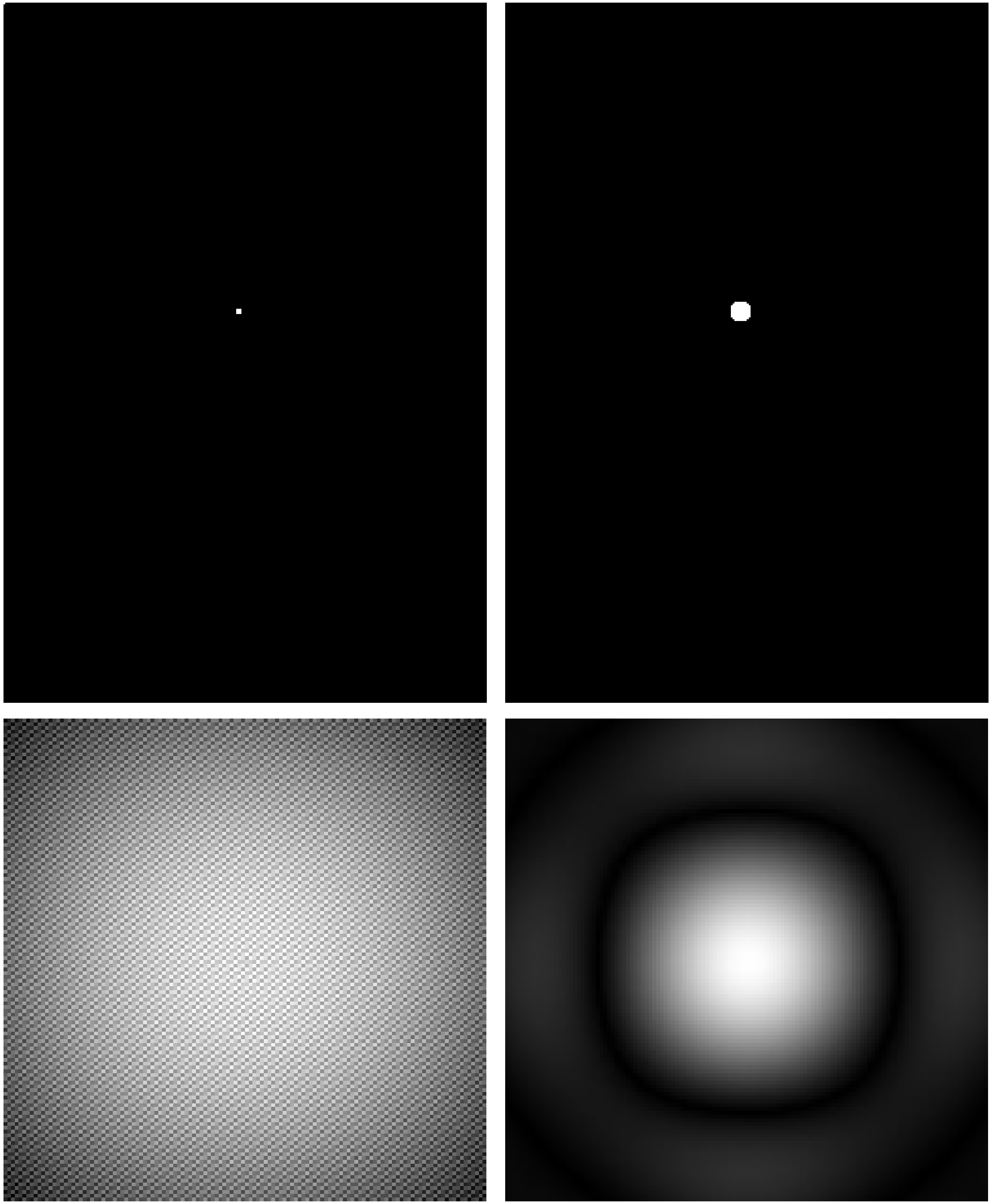
What if?



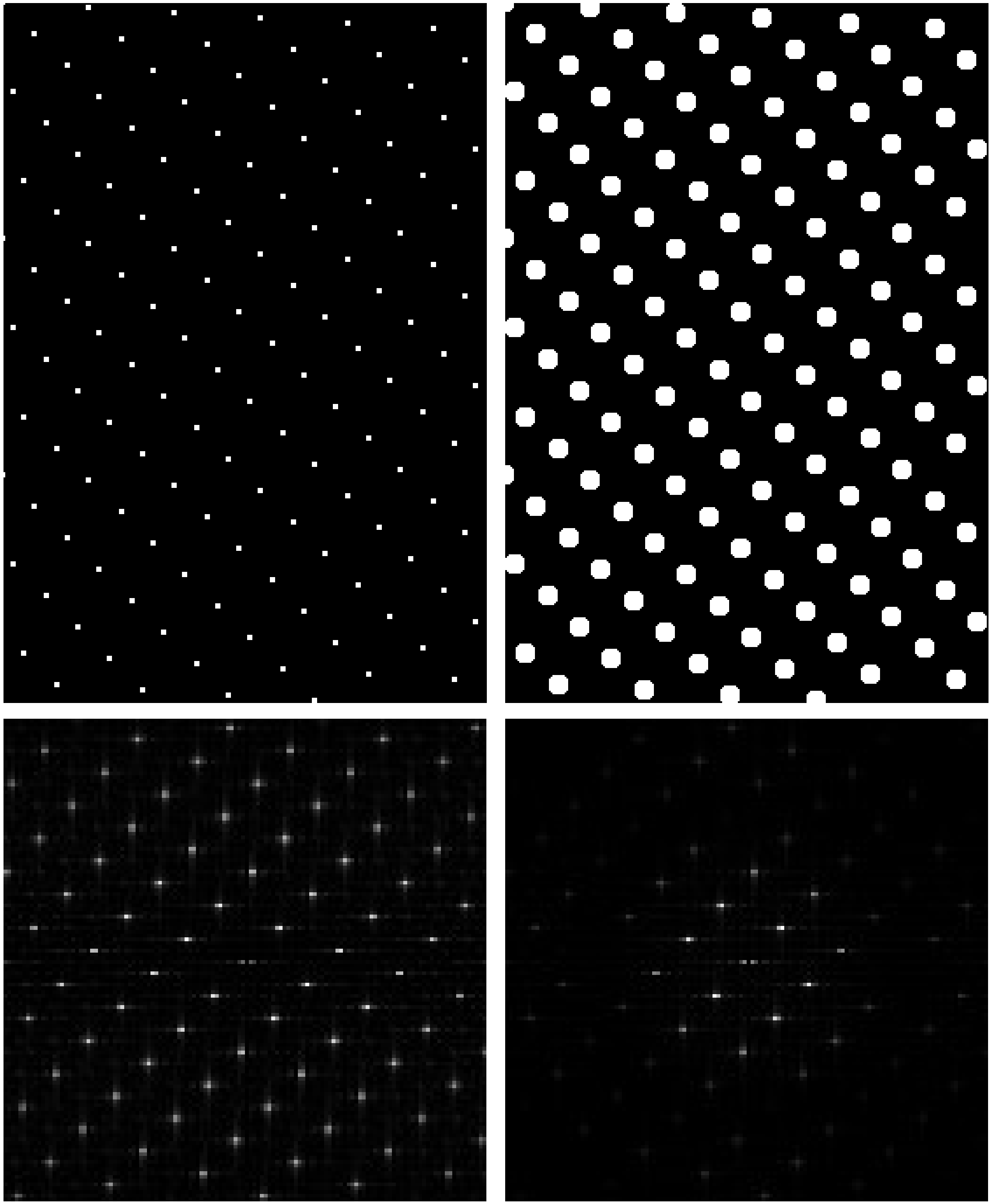


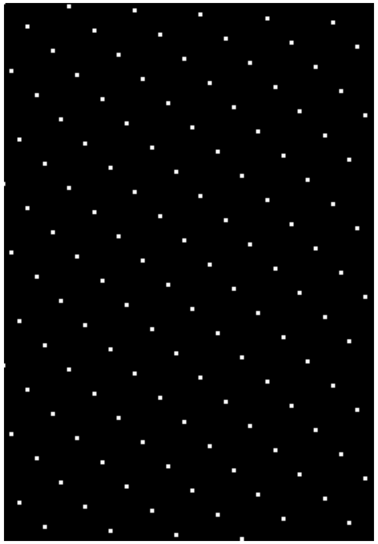
?

Hint

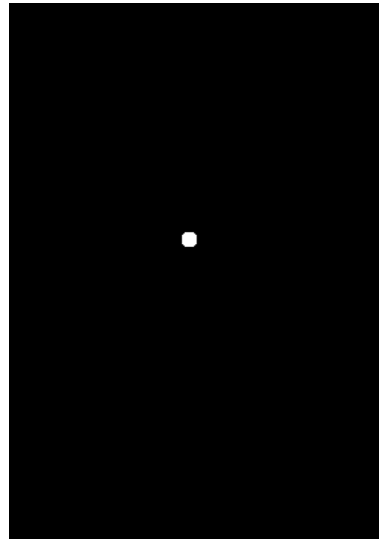


What if?

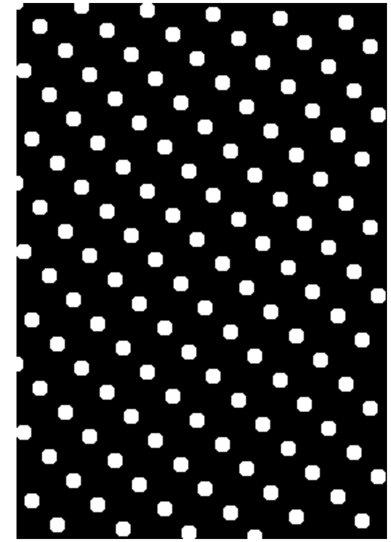




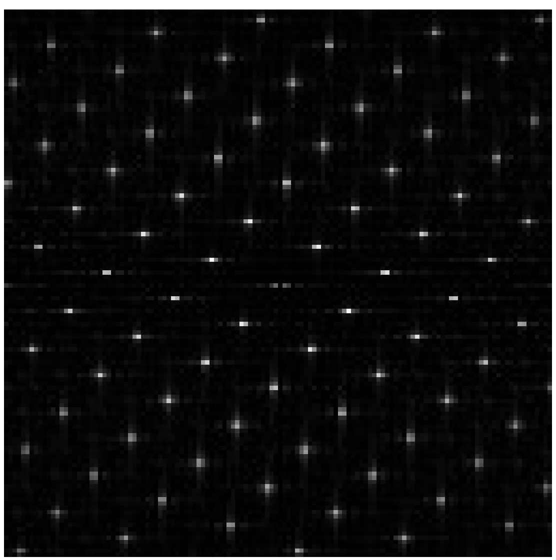
$f(x)$



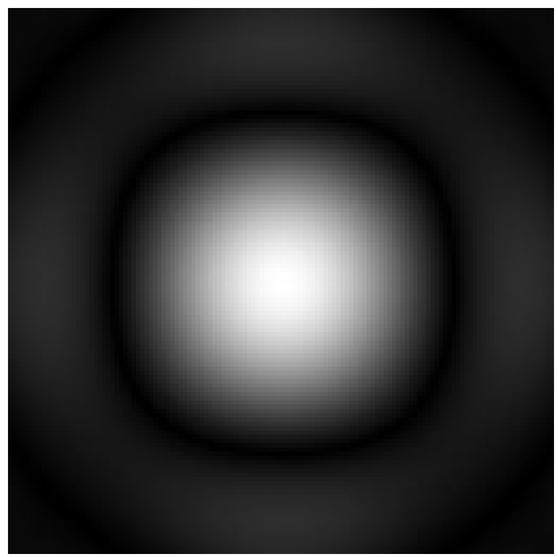
$g(x)$



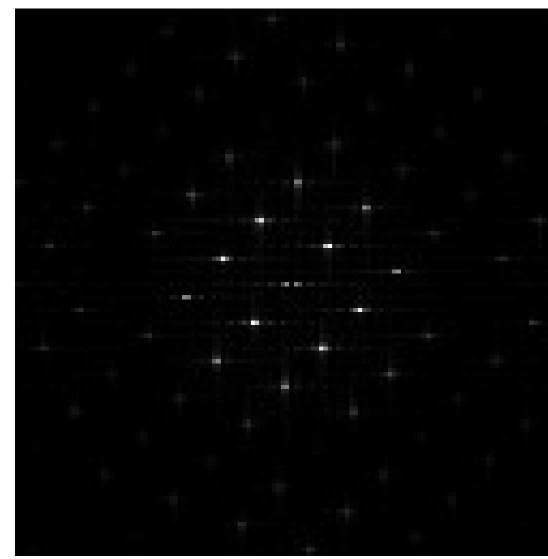
$f(x) \bullet g(x)$



$F(X)$



$G(X)$

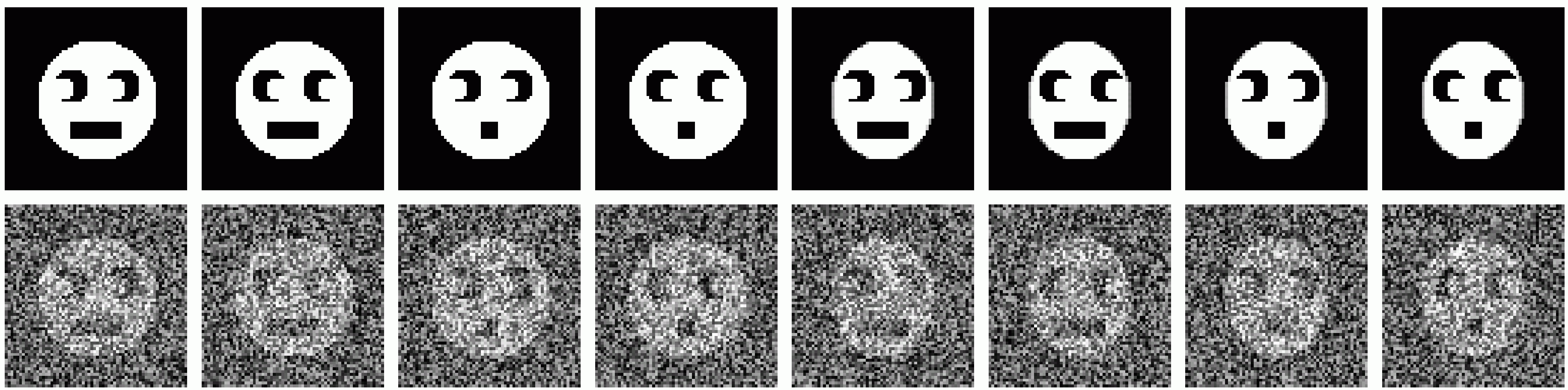


$F(X)\ G(X)$

Classification

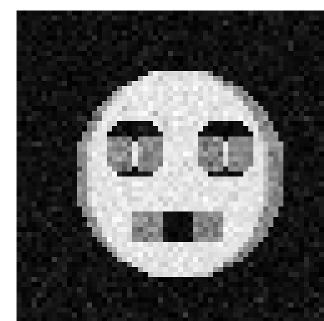
Reiteration of the problem

8 classes of faces, 64x64 pixels



With noise added

Average:

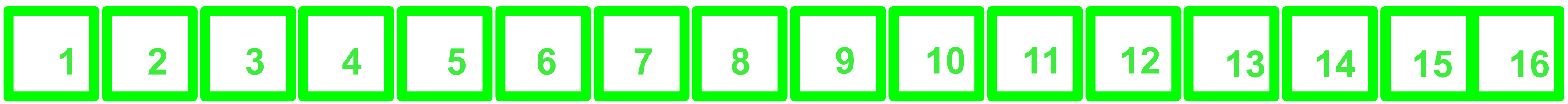


Before we can average the data, we first should find homogeneous subsets.

Multivariate data analysis (MDA)

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

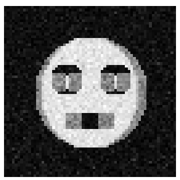
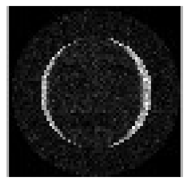
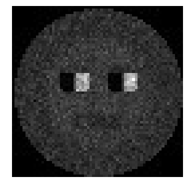
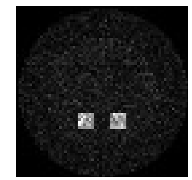
Multivariate data analysis (MDA), or Multivariate statistical analysis (MSA)



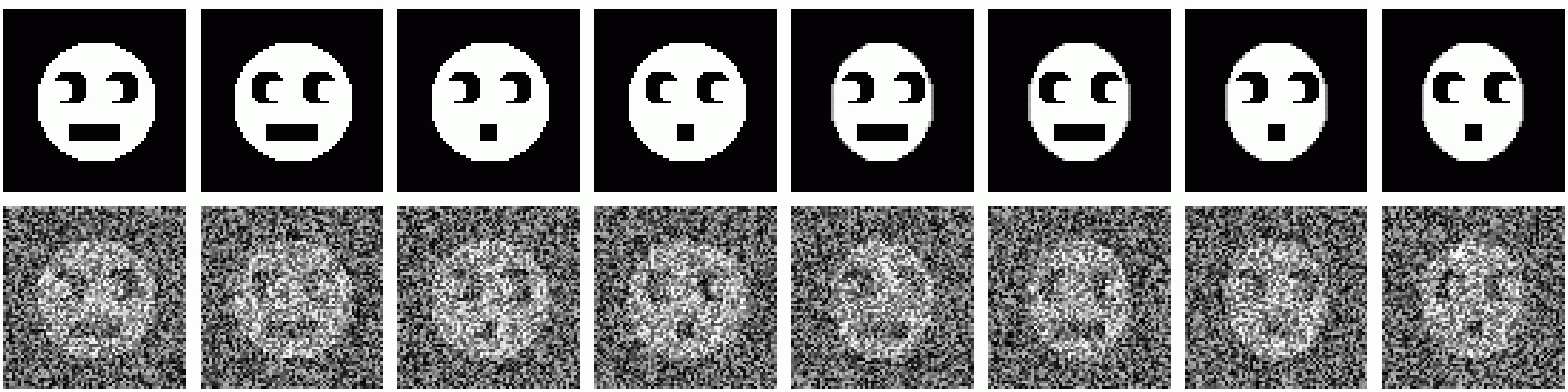
Our 16-pixel image can be reorganized into a 16-coordinate vector.

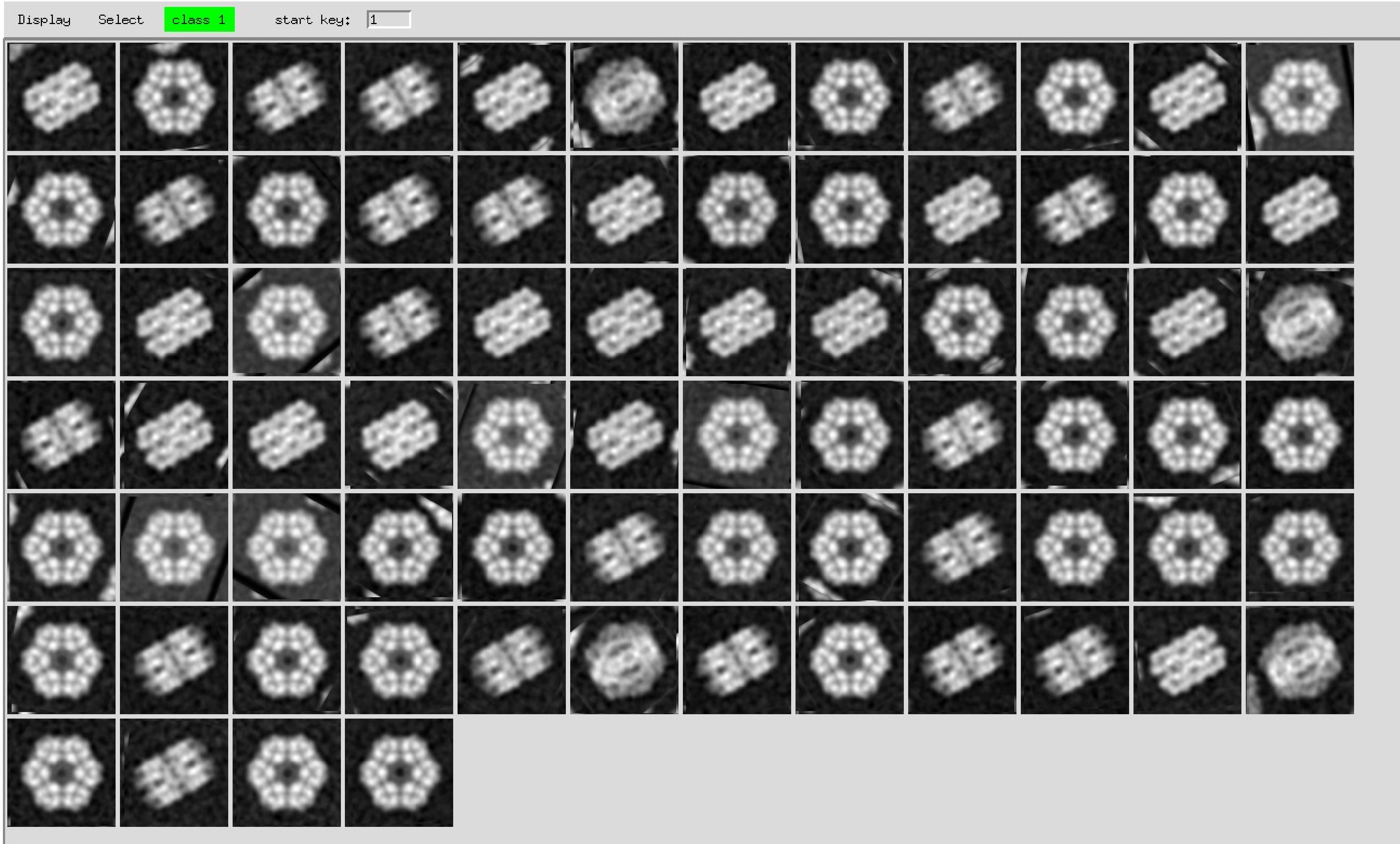
MDA: Reconstituted images

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

c_0  $+\ c_1$  $+\ c_2$  $+\ c_3$  $+ \ldots$

Average Eigenimage #1 Eigenimage #2 Eigenimage #3

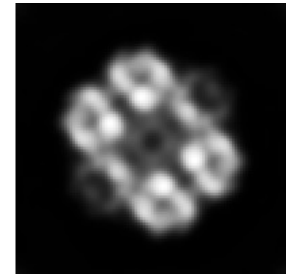


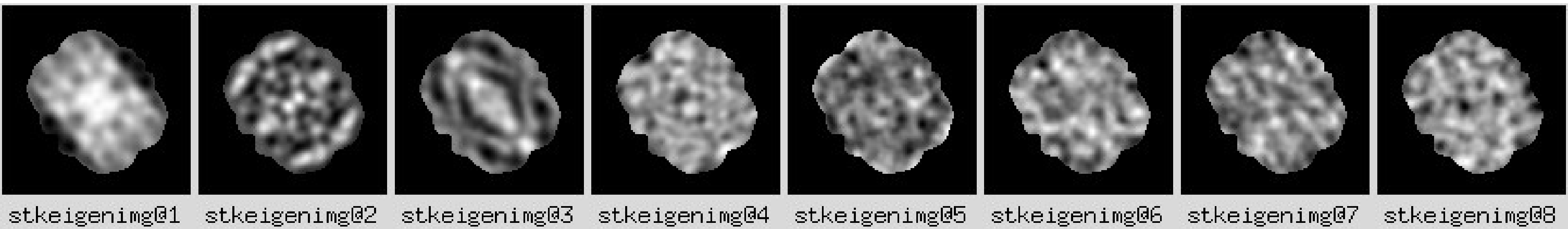


Phantom images of worm hemoglobin

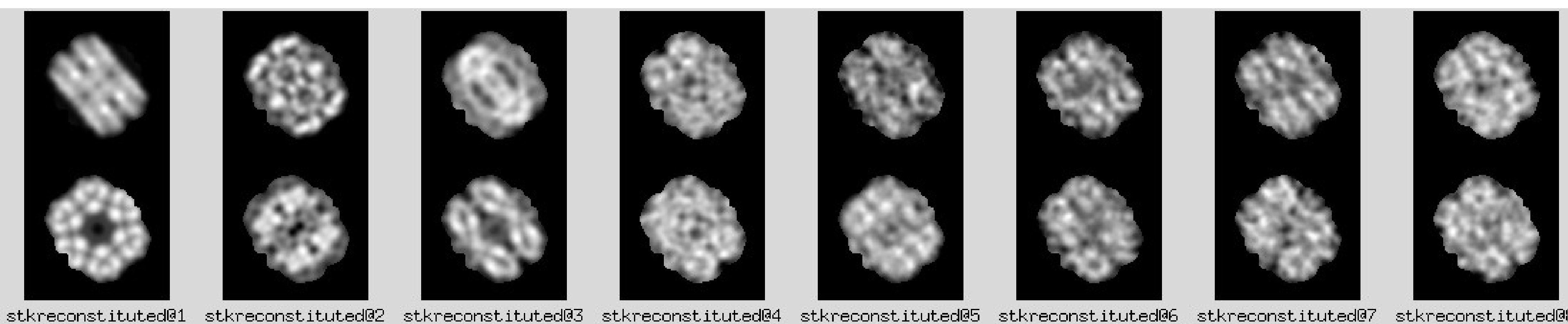
MDA of worm hemoglobin

Average:





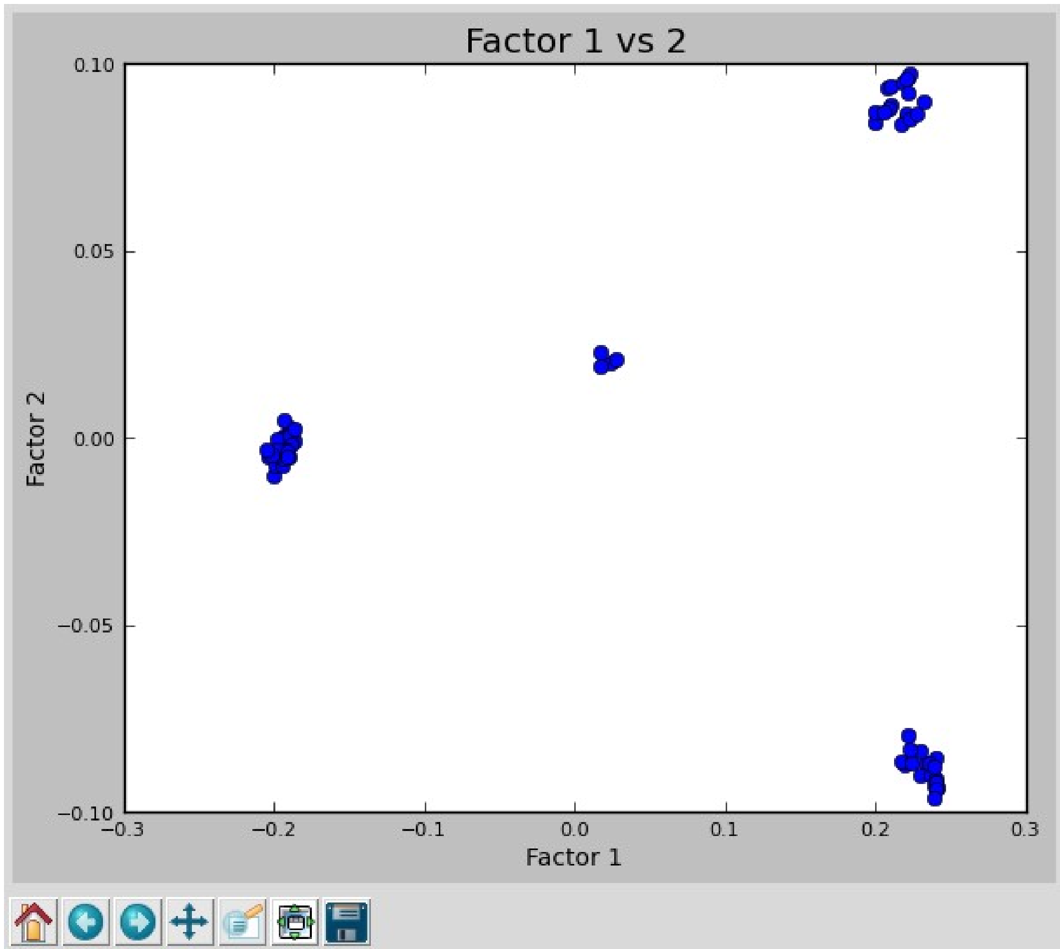
$+c_0$ $+c_1$ $+c_2$ $+c_3$ $+c_4$ $+c_5$



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

Classification

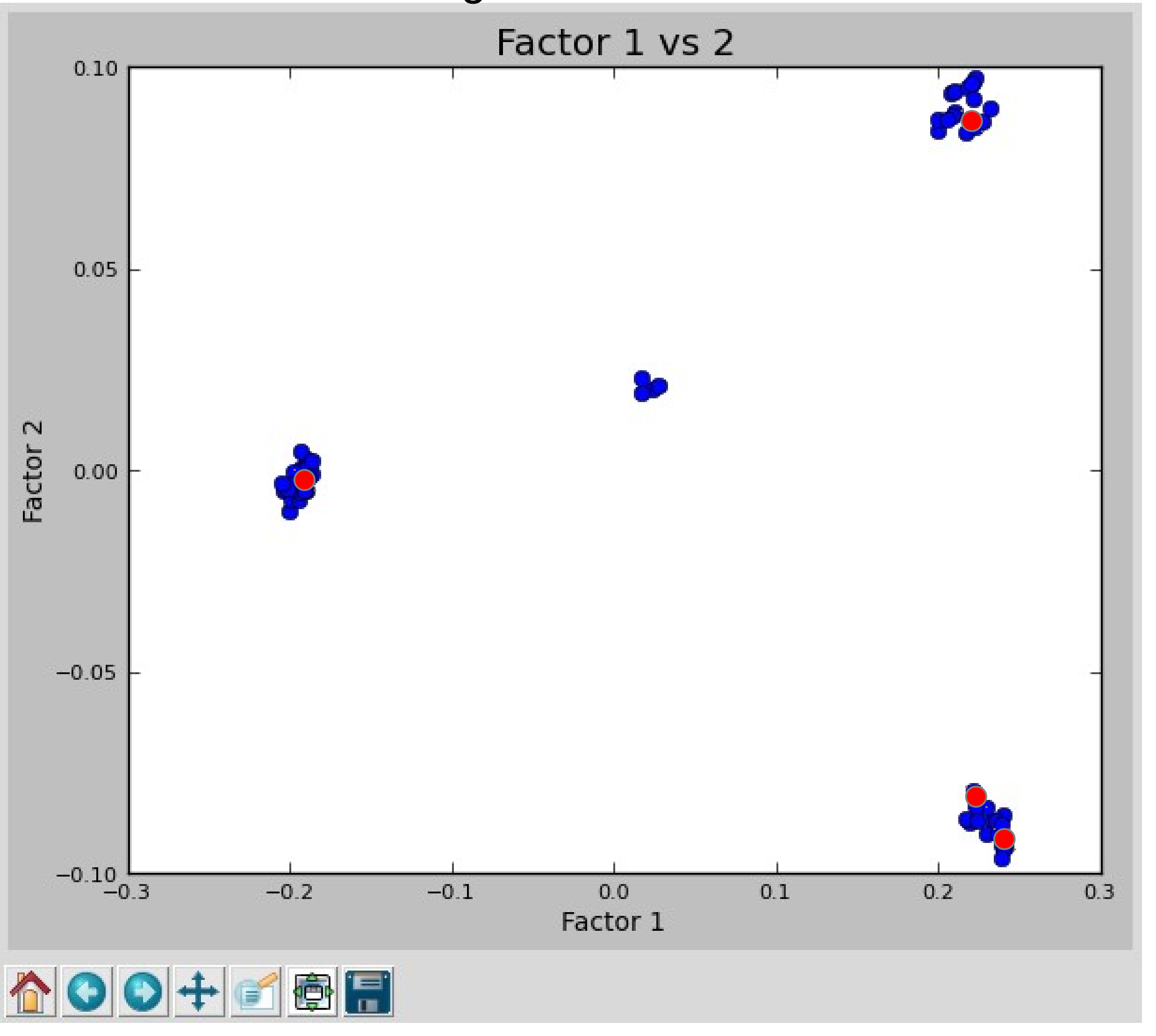
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



How do we categorize/classify the images?

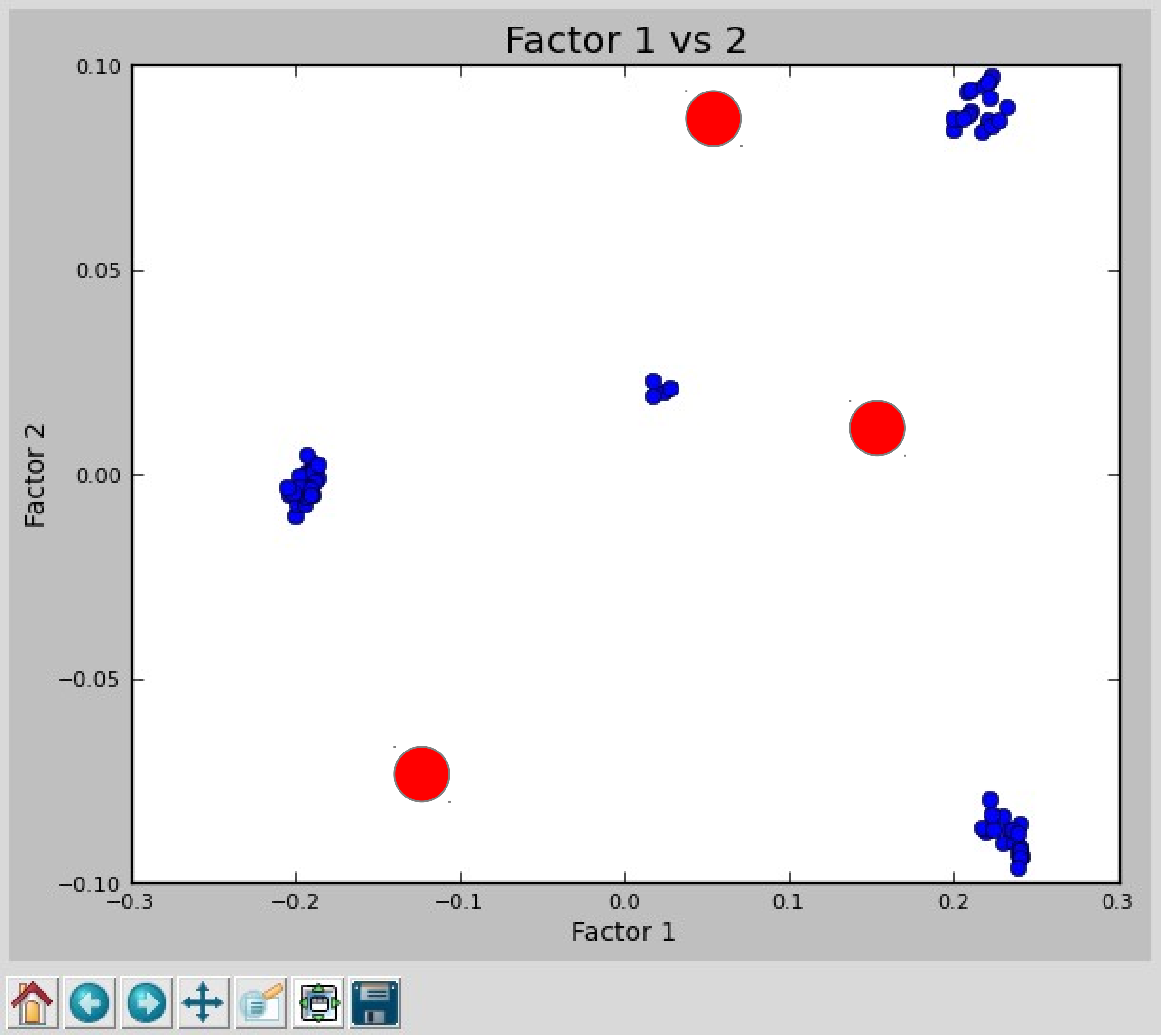
K-means classification

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

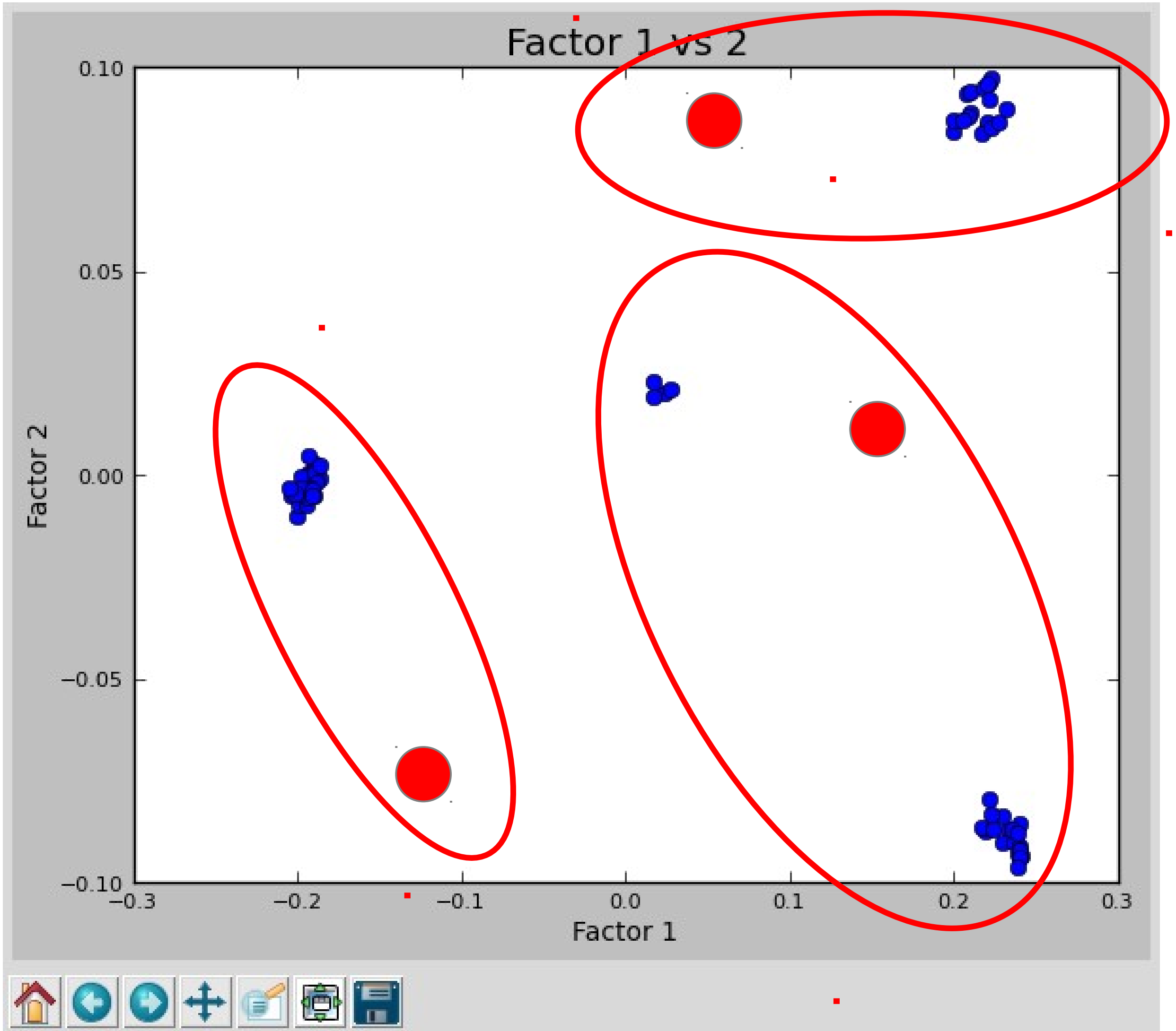


BAD: Some clusters may be overrepresented/underrepresented.

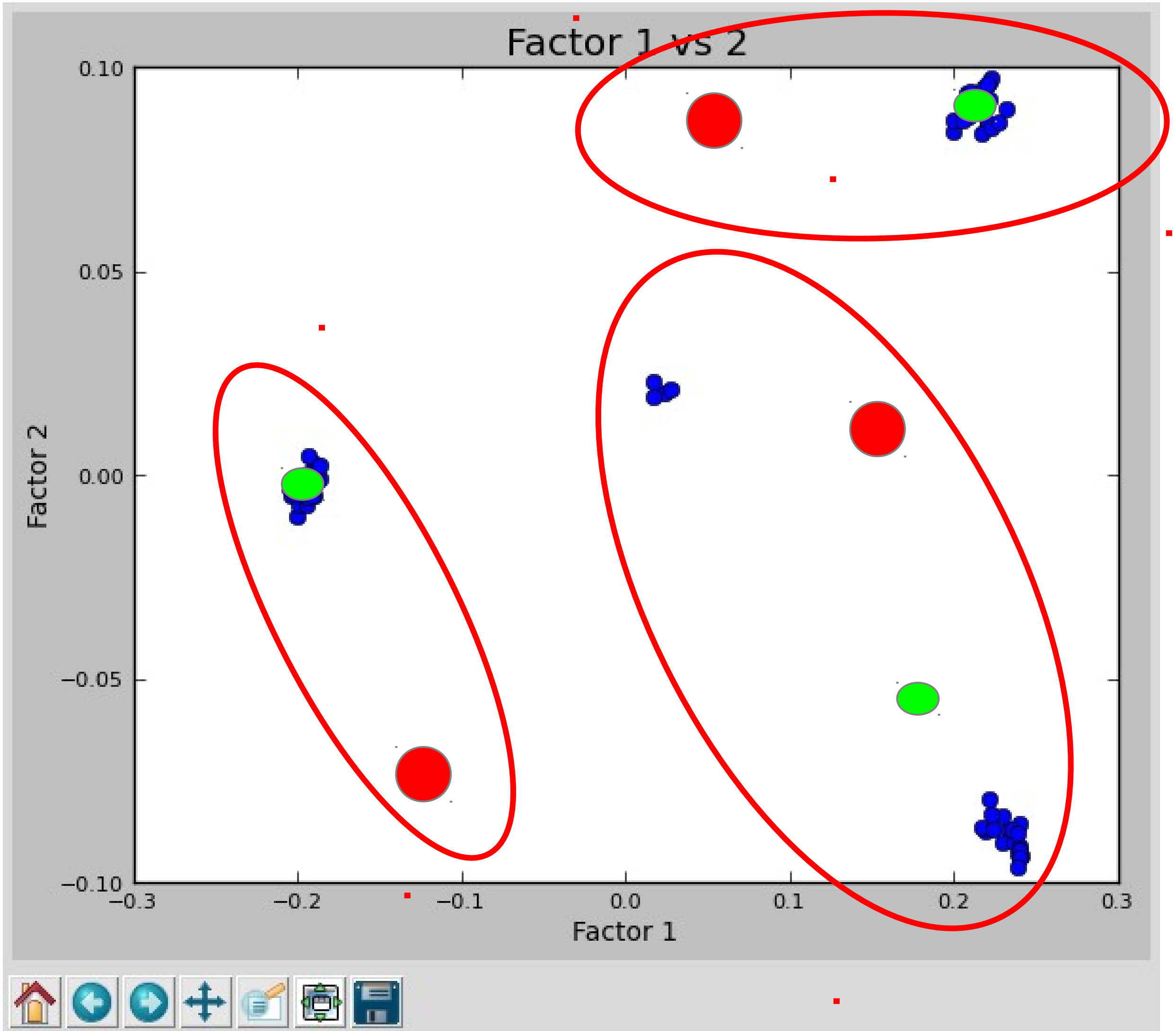
Diday's method of moving centers



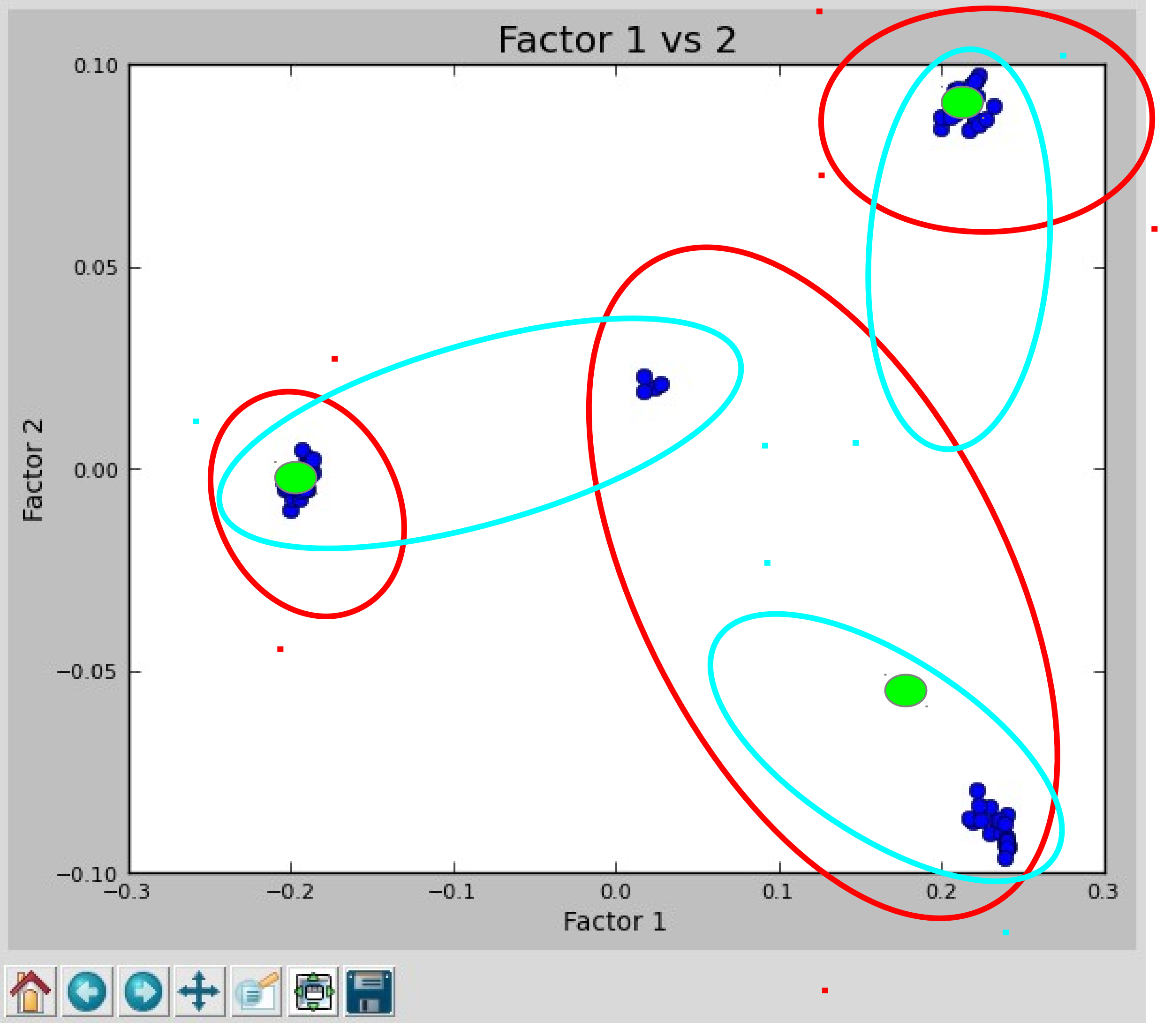
Diday's method of moving centers



Diday's method of moving centers

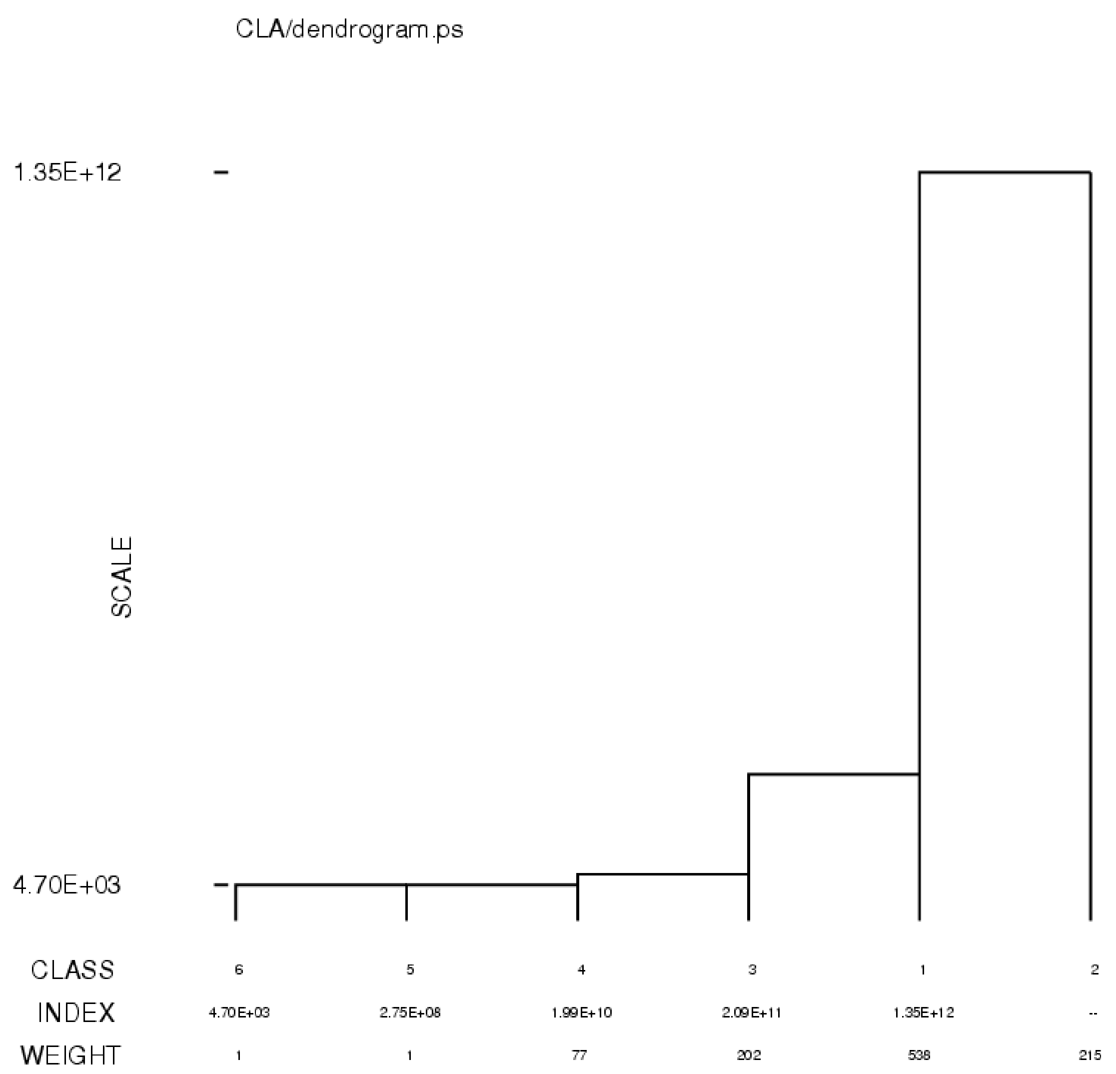


Diday's method of moving centers

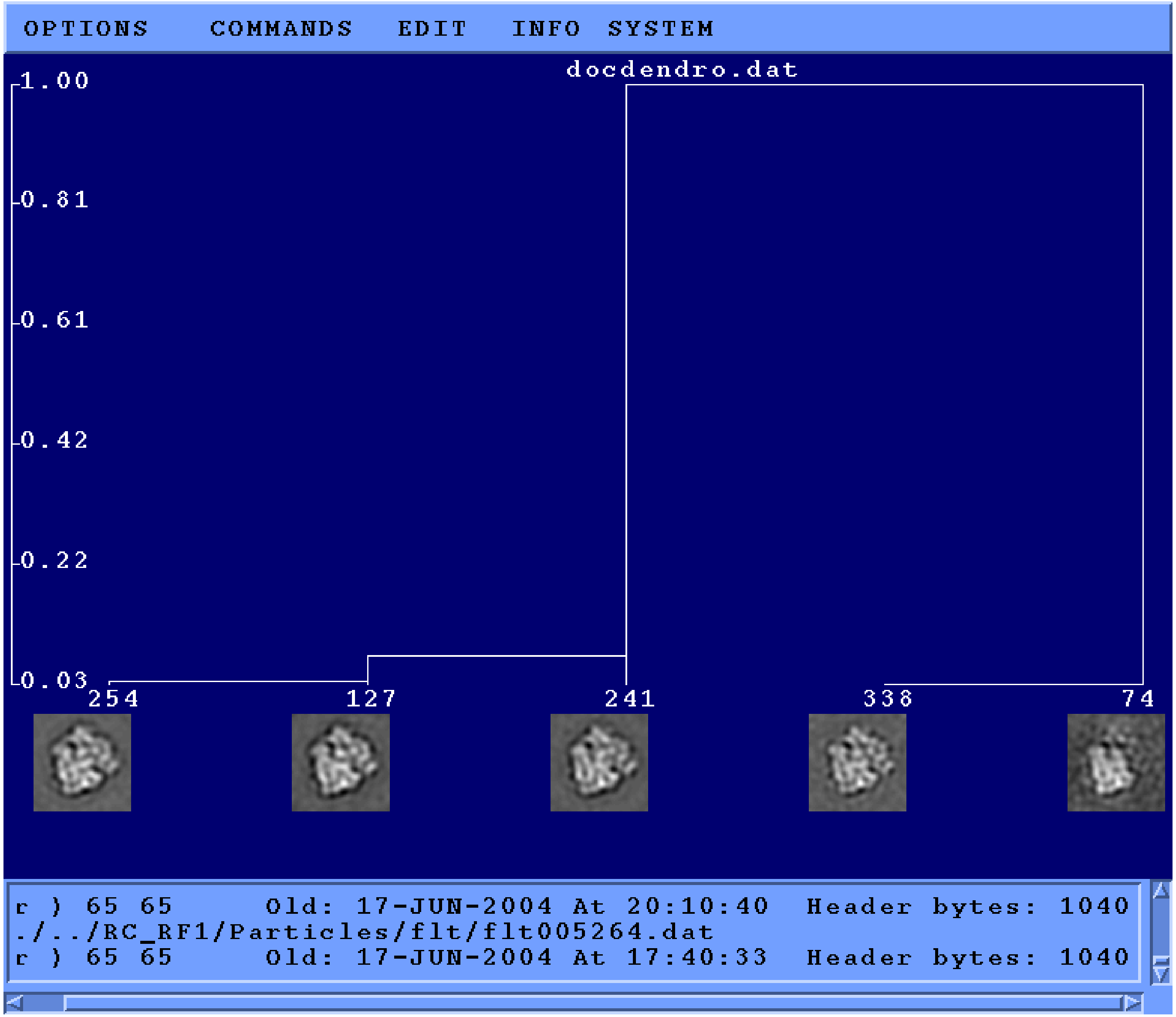


We will note the images that always “travel” together, and will call them a class.

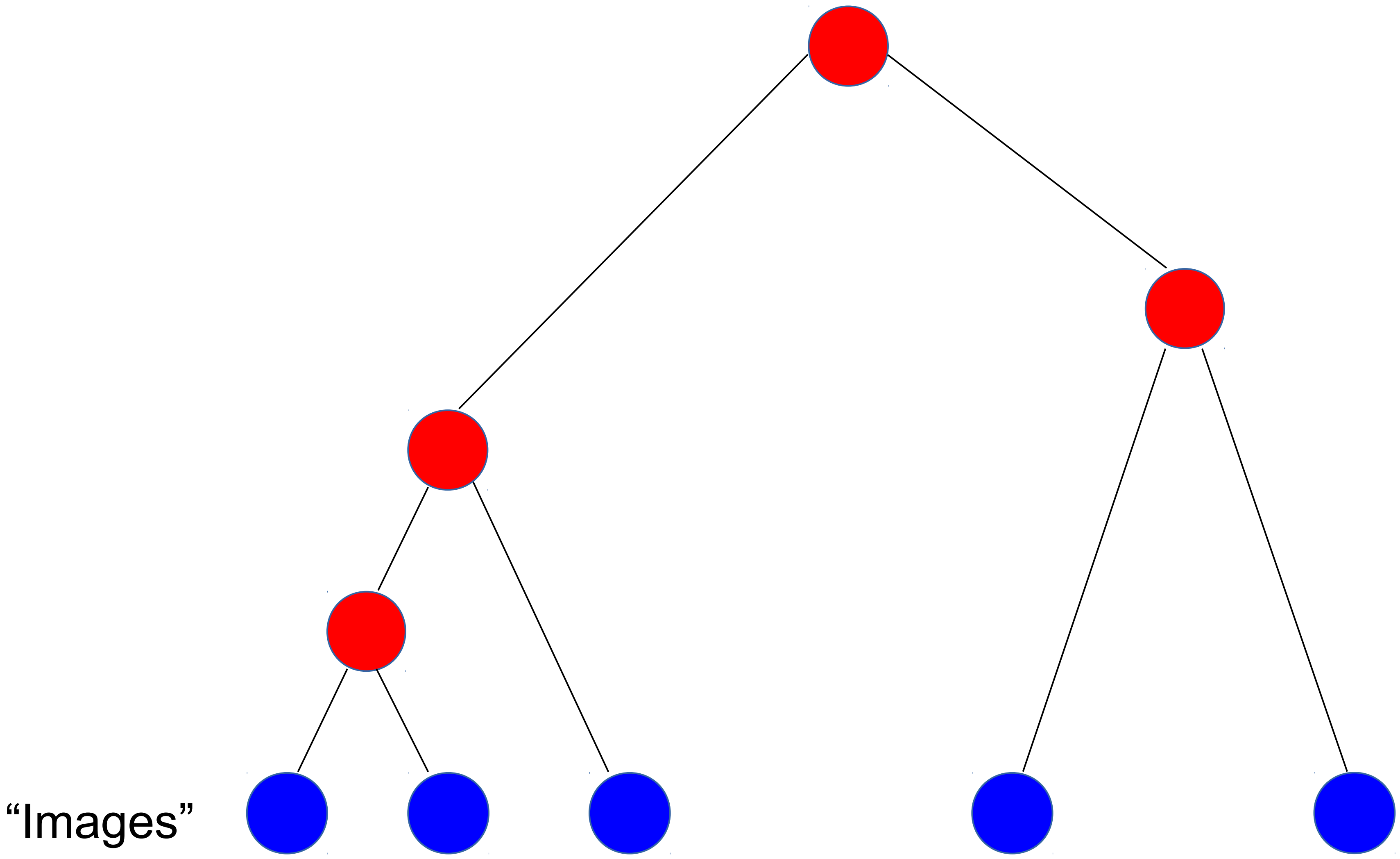
Dendrogram


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

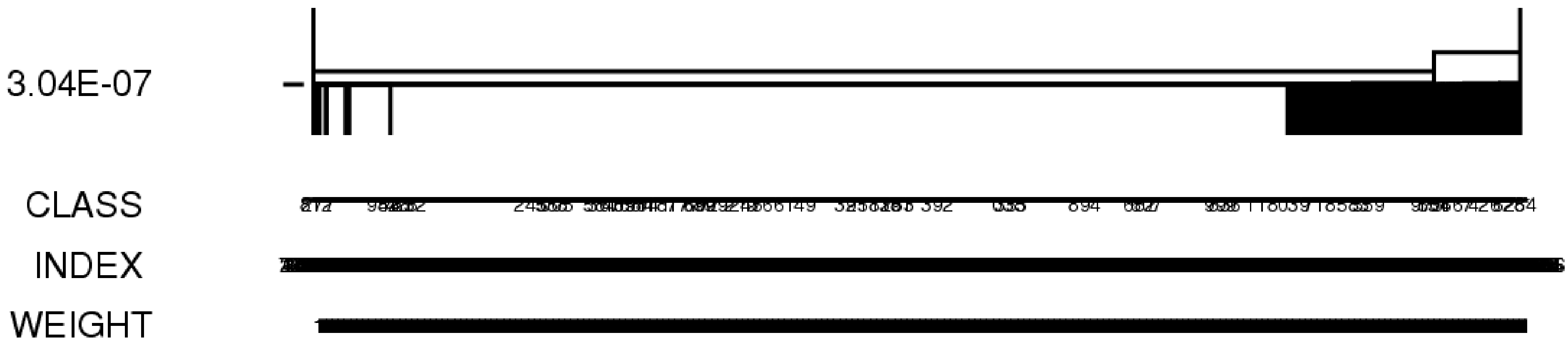
Dendrogram


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

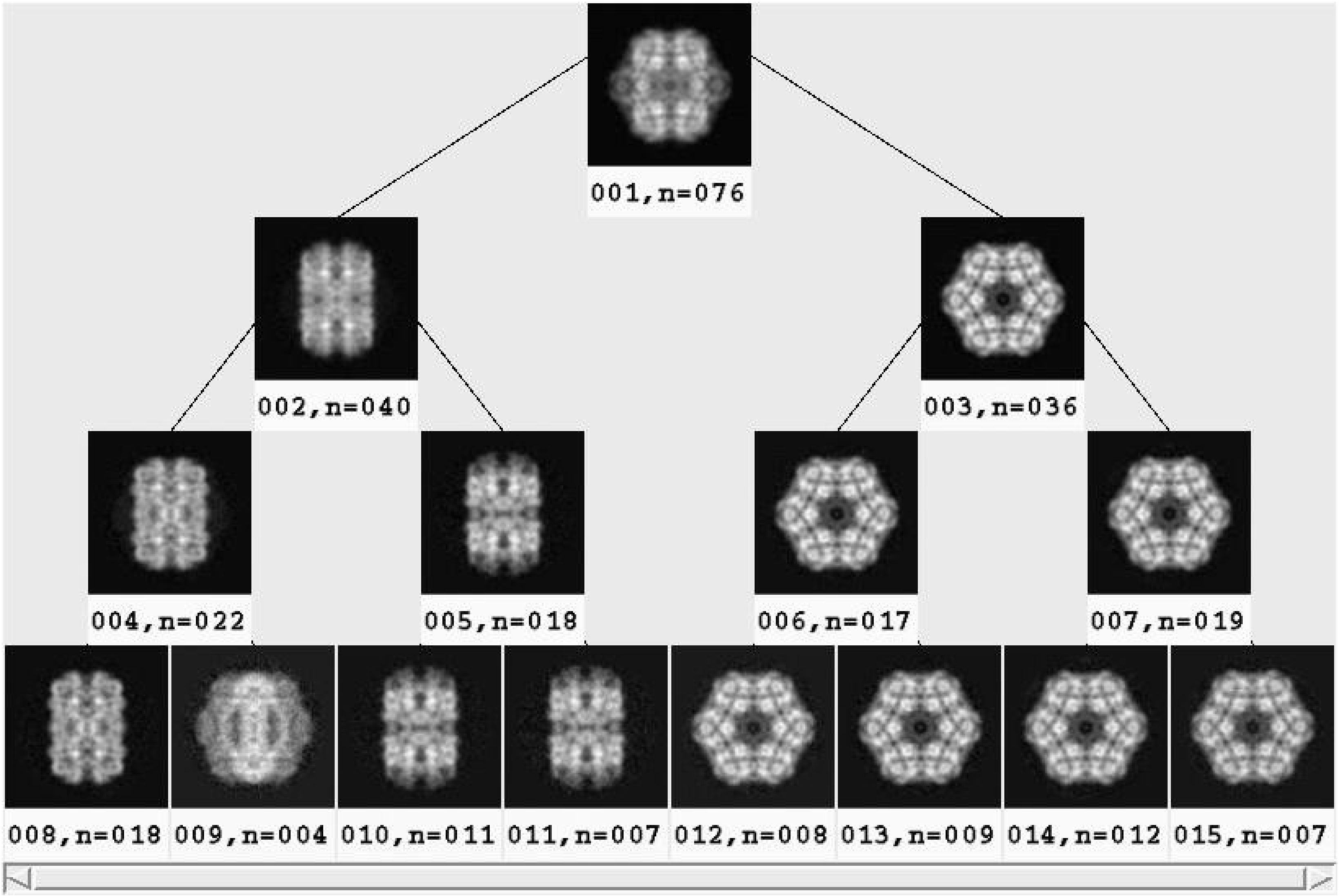


Hierarchical Ascendant Classification



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.

Classification in 3D

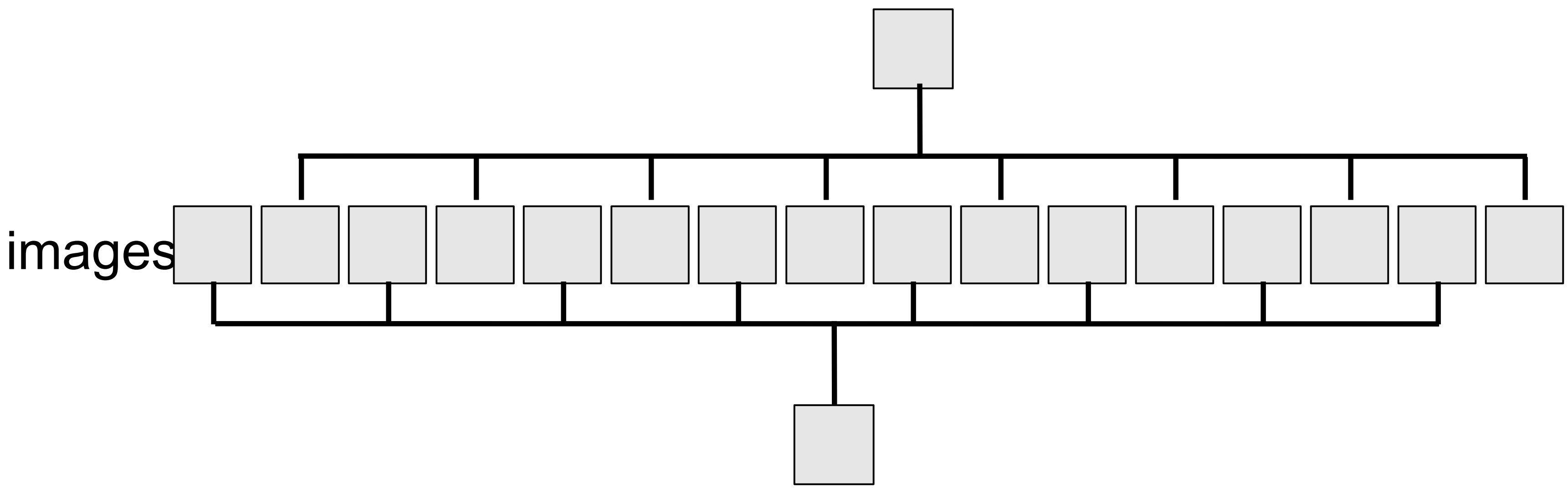
Classification: Reference-based classification vs. Maximum likelihood (ML3D)

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.

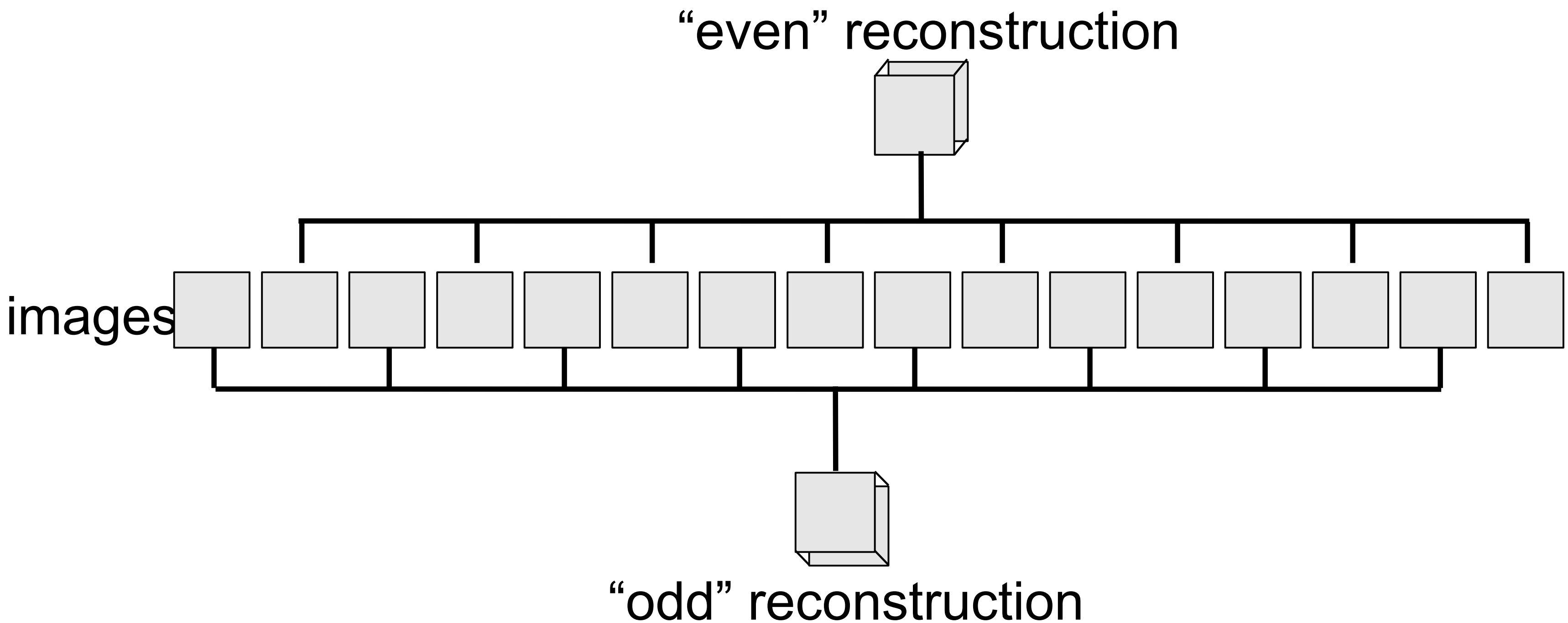


There will be slight differences in the reconstructions.
We will iteratively maximize the likelihood of a particle belonging to a particular class.

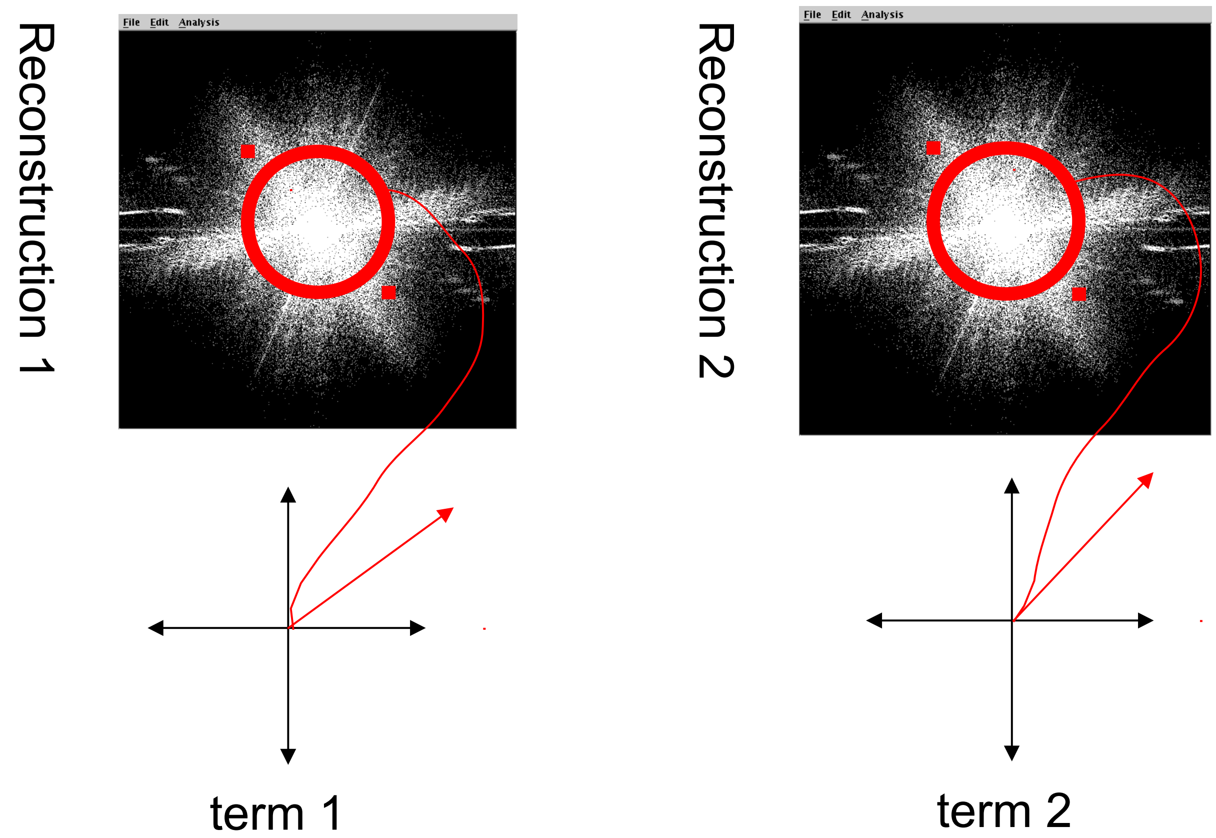
How good is our reconstruction?

How do we evaluate the quality of a reconstruction?

We split the data set into halves and compare them.

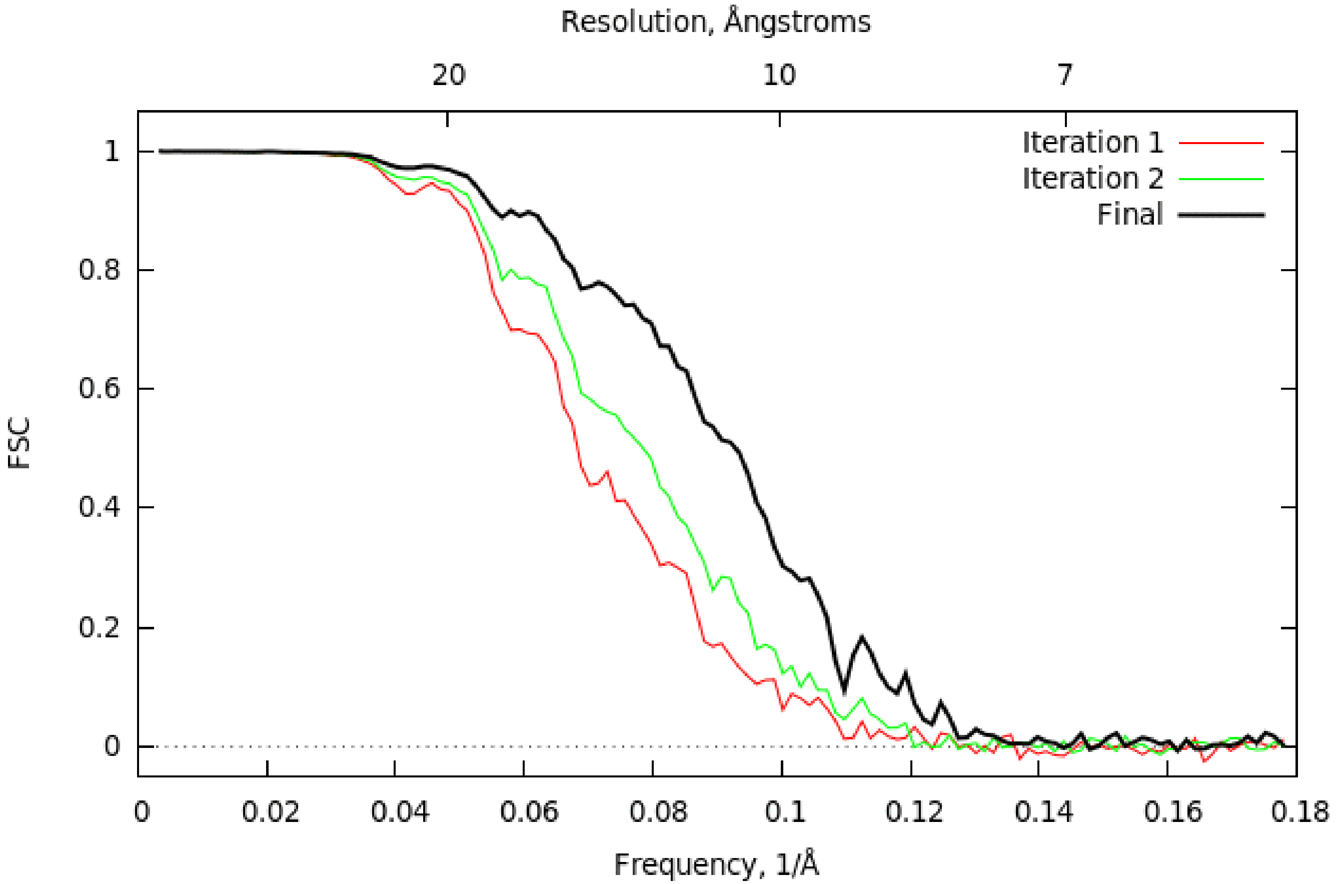


Fourier Shell Correlation (FSC)

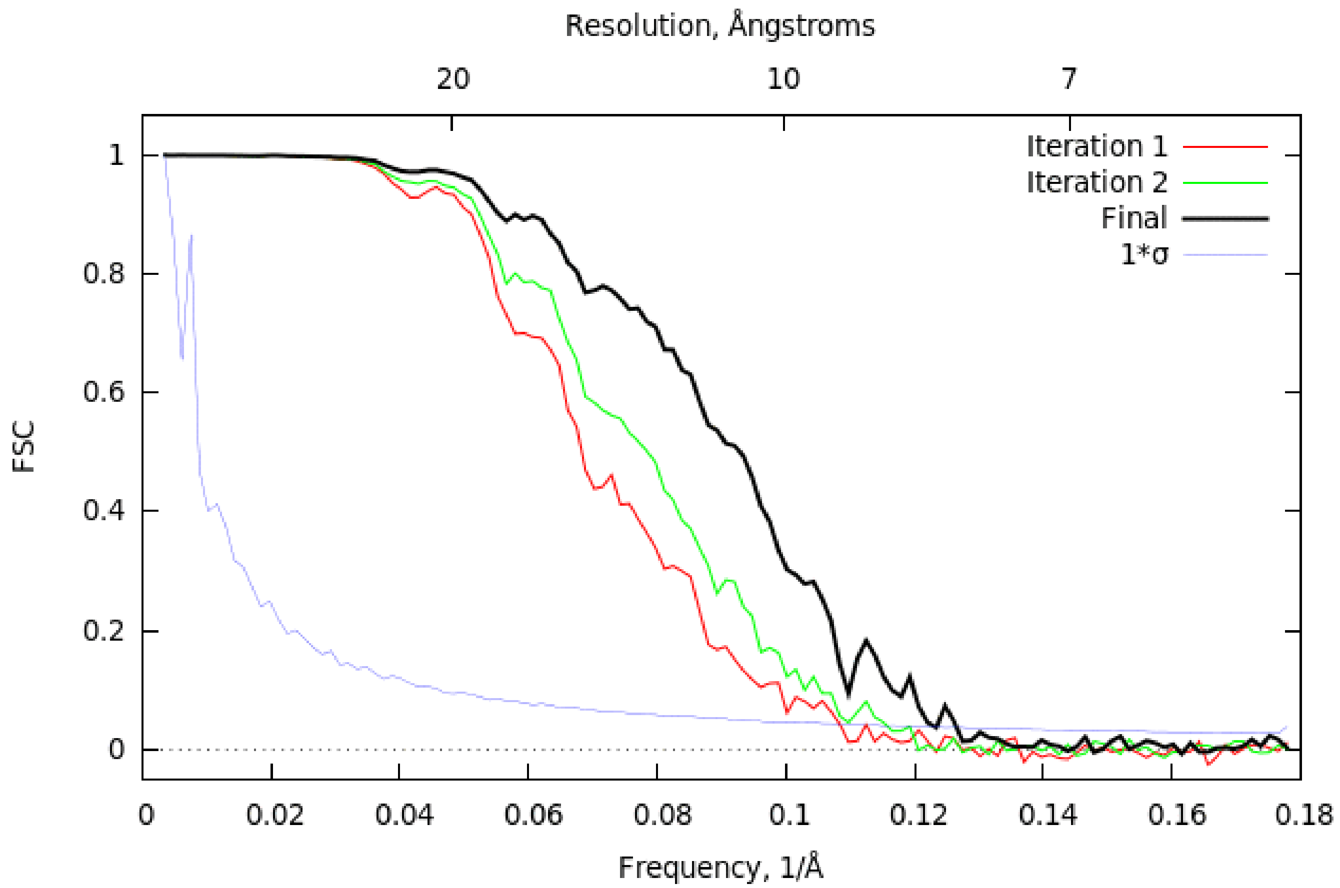


Properties:
- Fourier terms have amplitude + phase.
- Correlation values range from -1 to +1.
- Noise <u>should</u> give an average of 0.
- The comparison is done as a function of spatial frequency (or “resolution”)

Fourier Shell Correlation curve

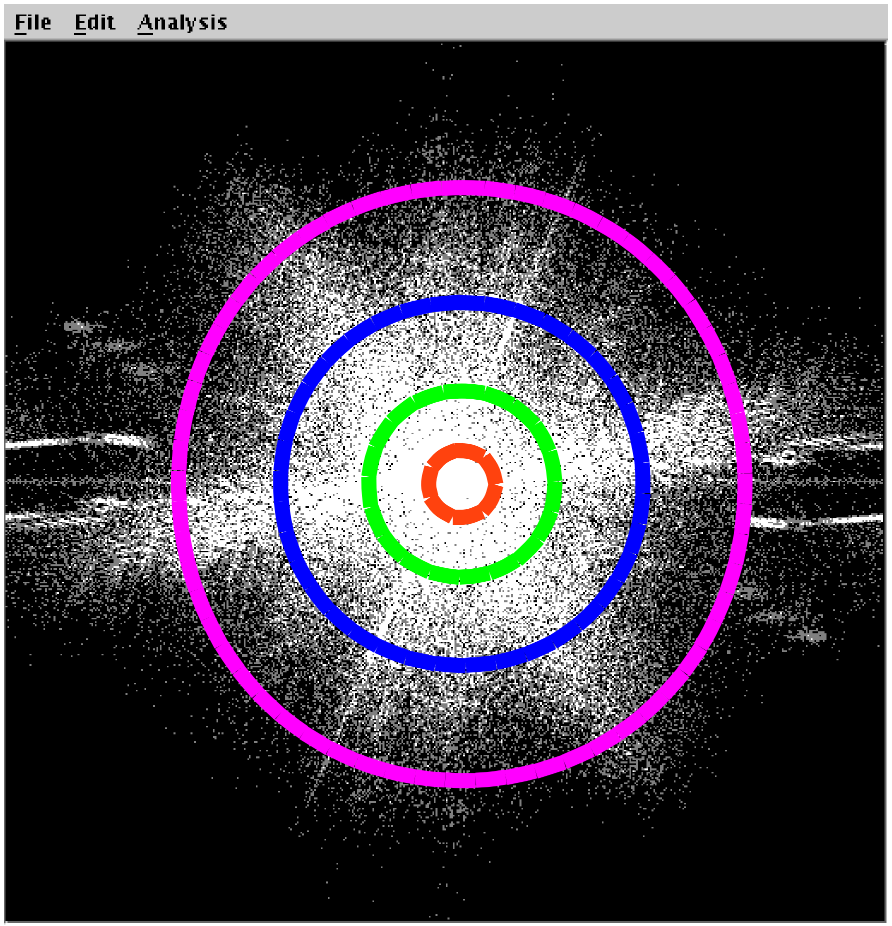


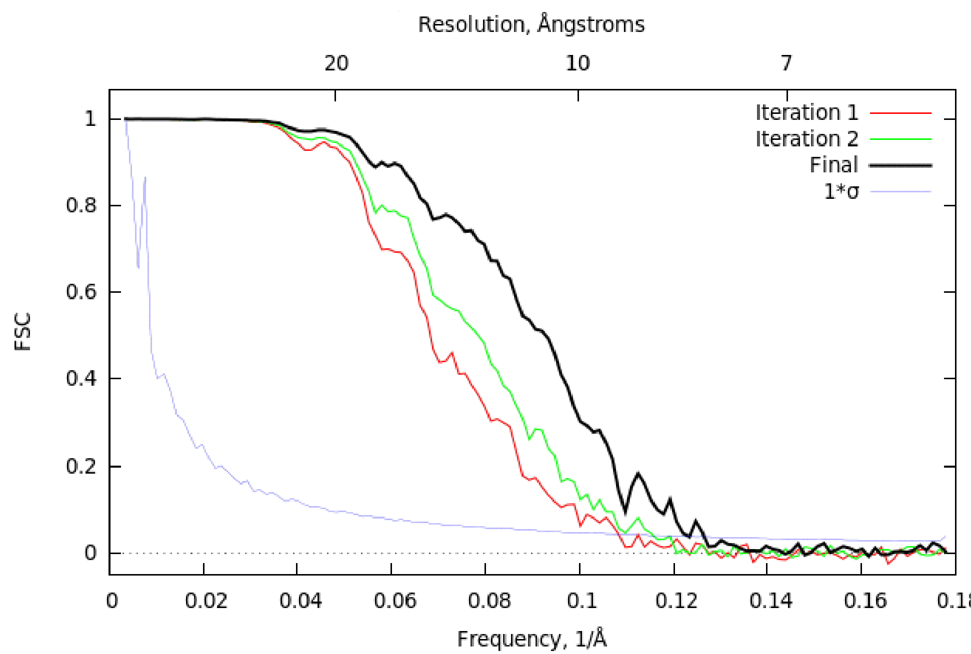
FSC curve with expectation value of noise



CEITEC

Why does σ vary with spatial frequency?



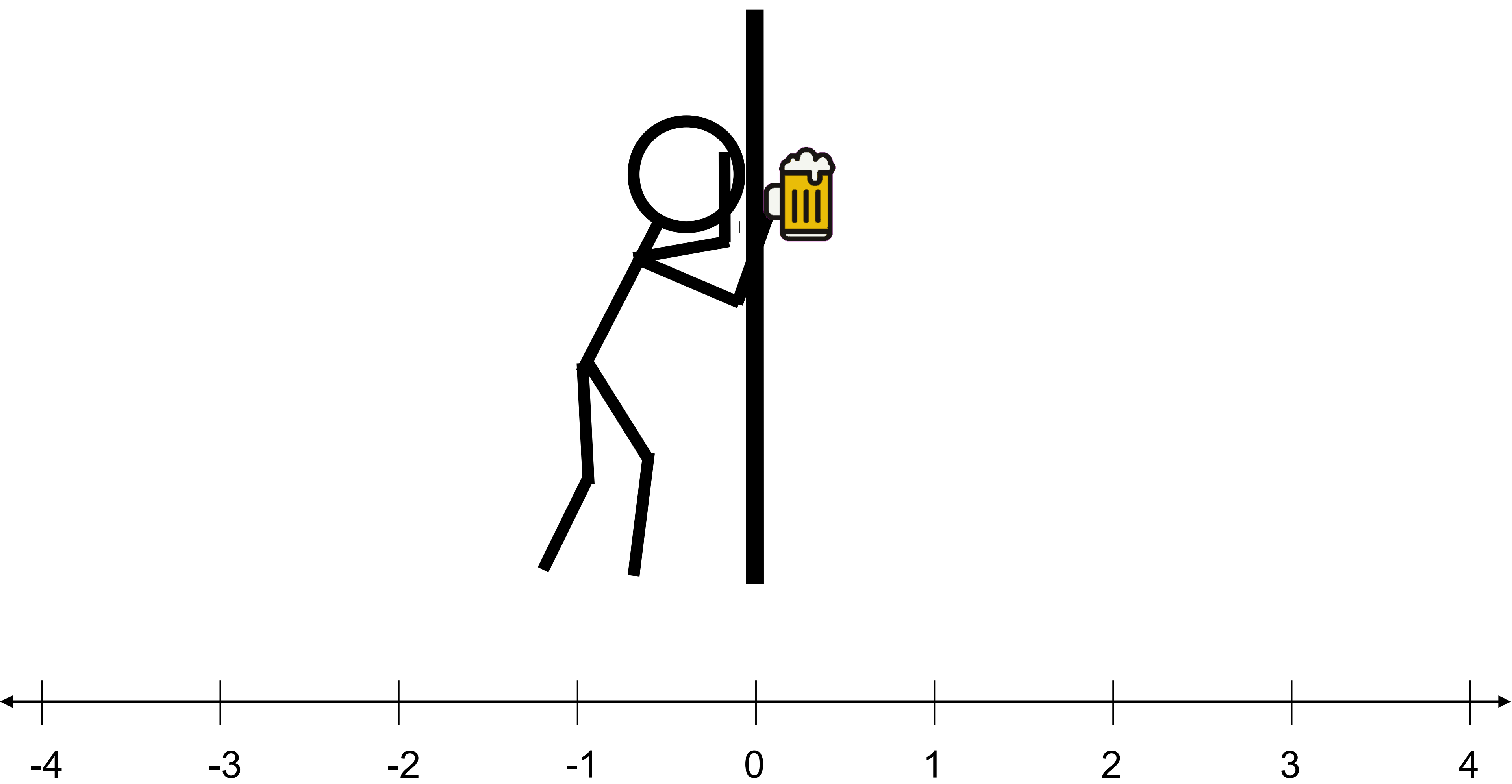


Random walks:
Why signal-to-noise improves with √N

The “Drunkard's walk”

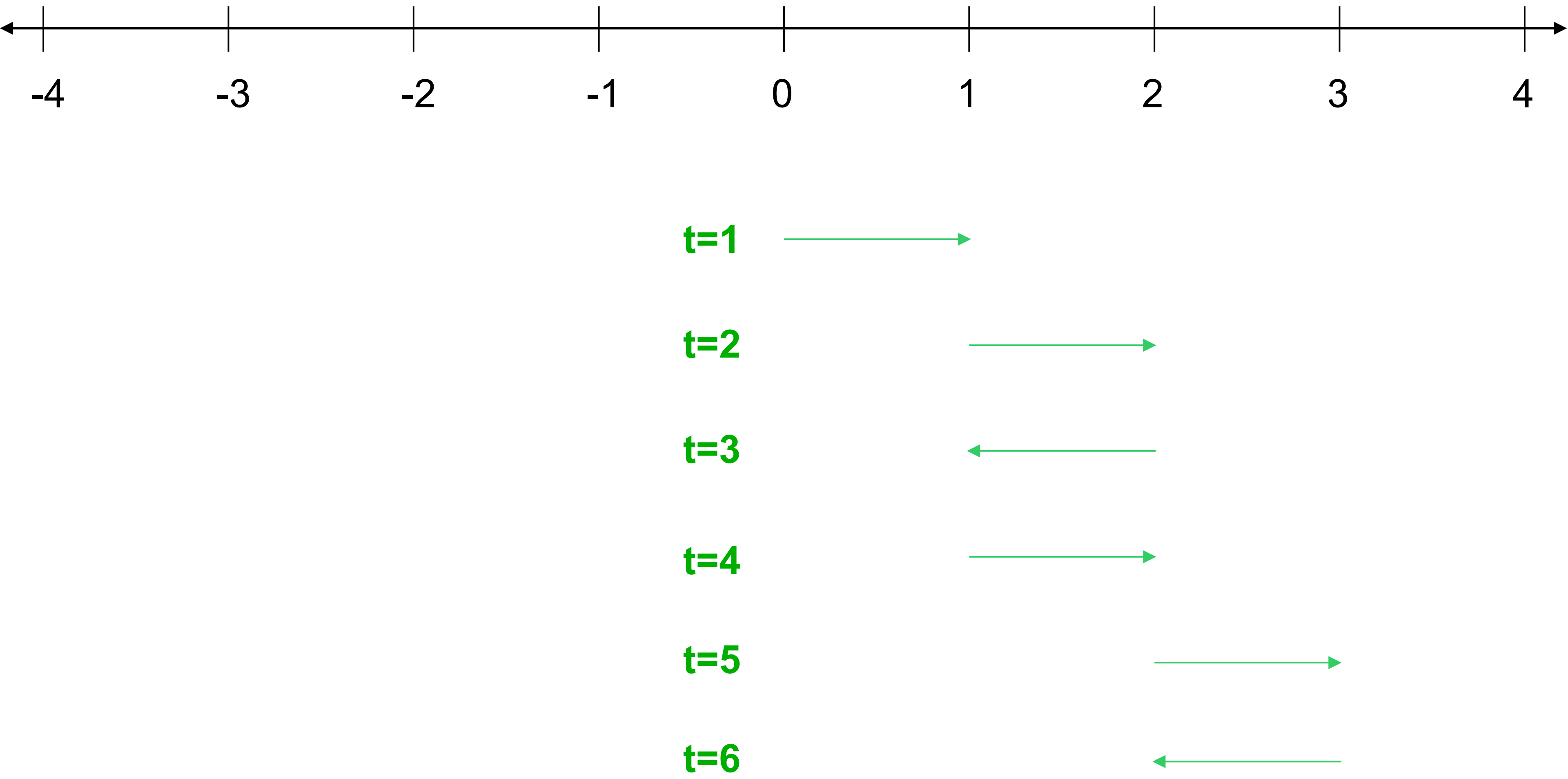
Let's conduct an experiment.

The “Drunkard's walk”

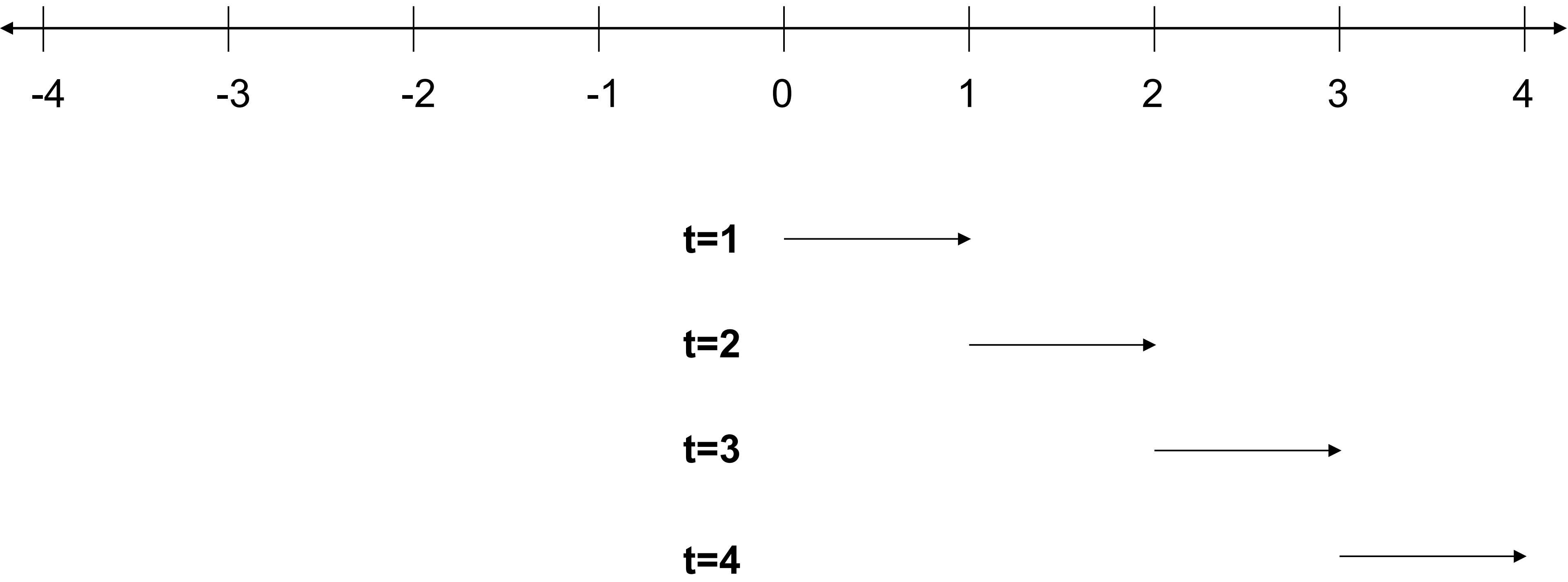


We're going to assume that each step is random and independent of previous steps.

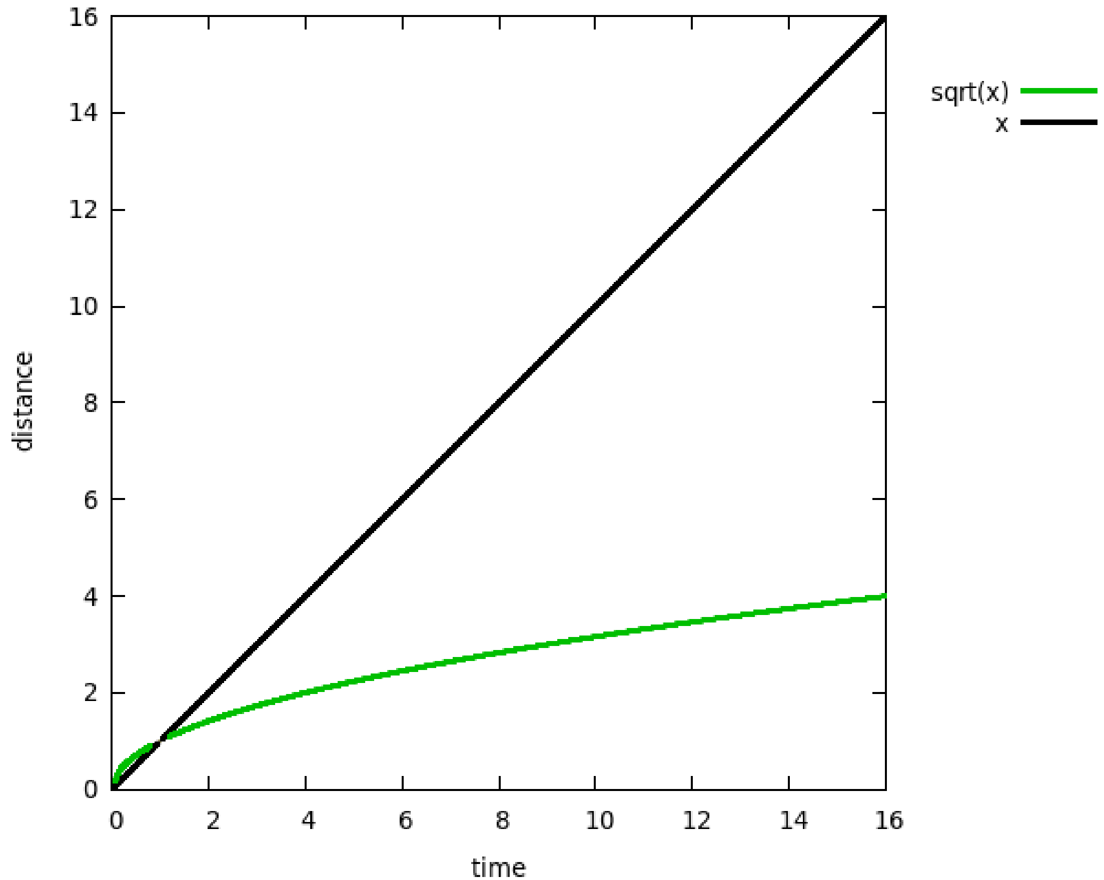
The “Drunkard's walk”



The teetotaler's walk

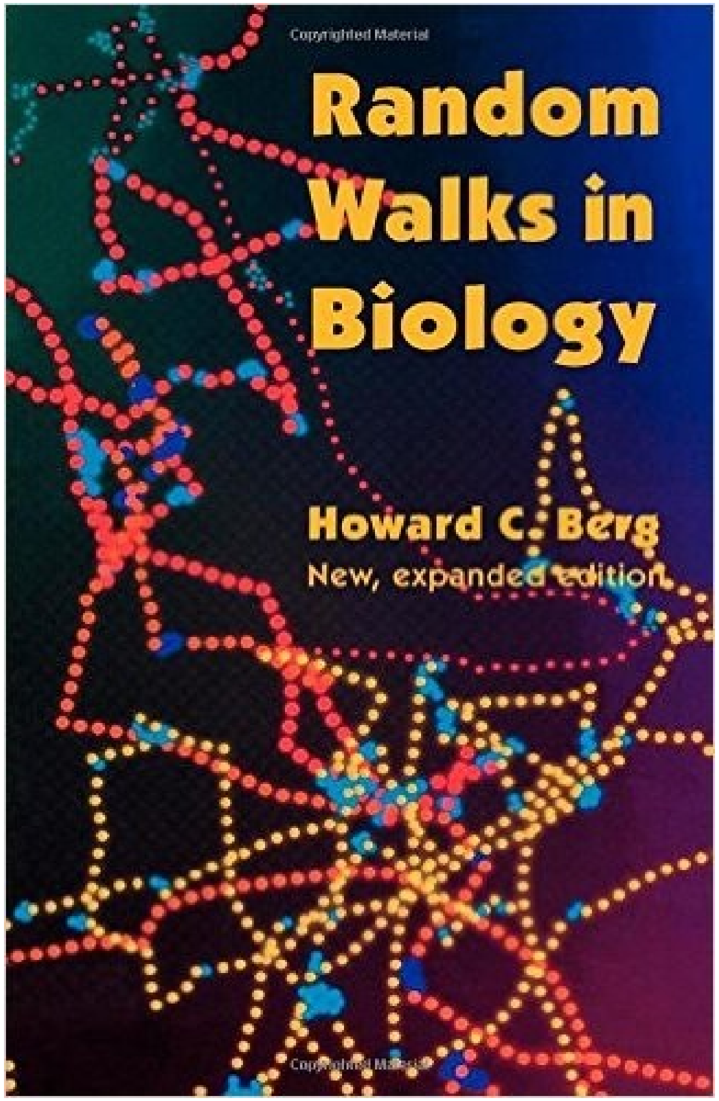


Expectation value



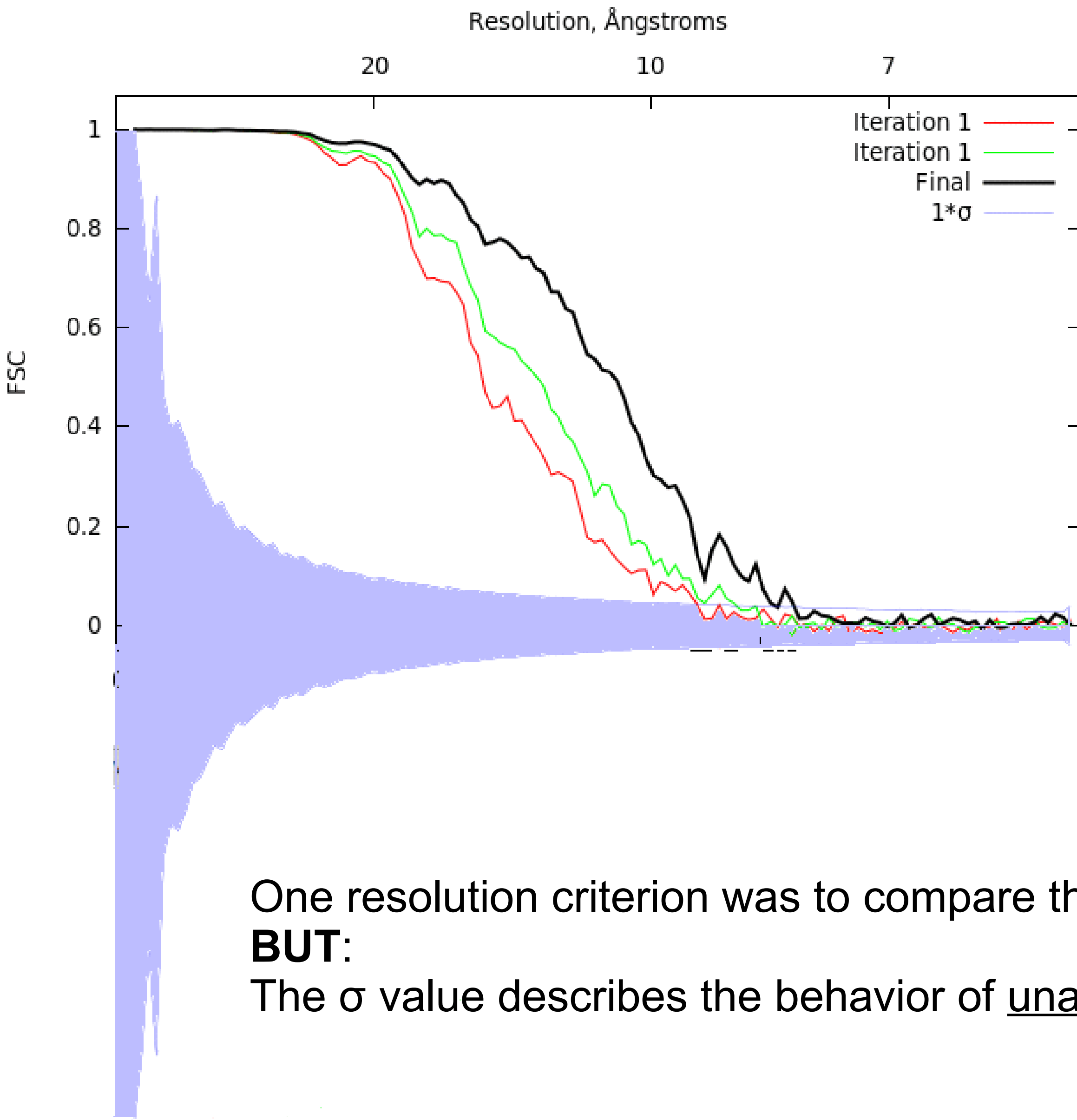
The expected distance that "noise" travels increases with √N.
However, it is not as fast as the distance that "signal" travels.
Thus, as we collect more data, the SNR increase by $N/\sqrt{N} = \sqrt{N}$

Random walks: more information



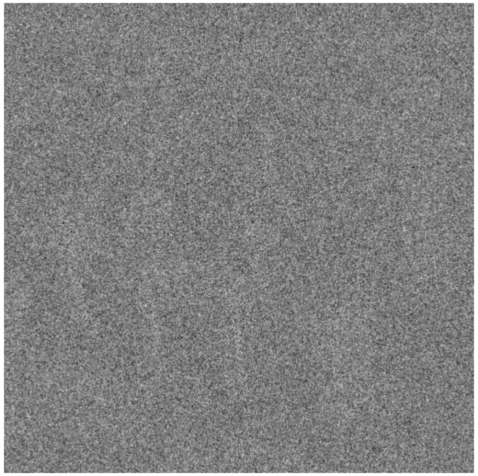
Expectation values
and how they related to resolution criteria

With small N, behavior is more unpredictable

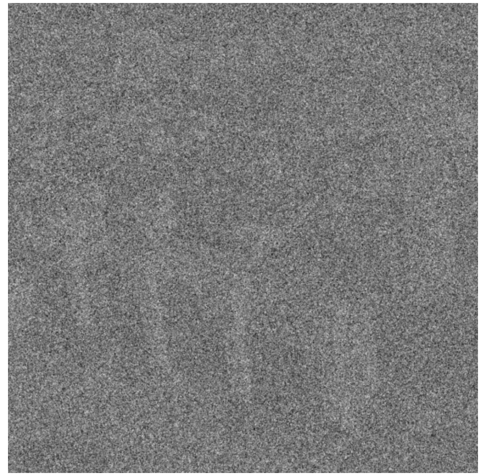


One resolution criterion was to compare the FSC to, say, 3*σ.
BUT:
The σ value describes the behavior of <u>unaligned</u> noise.

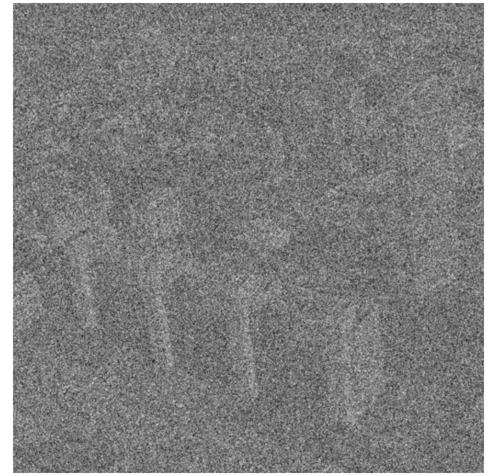
Review: model bias



N = 128

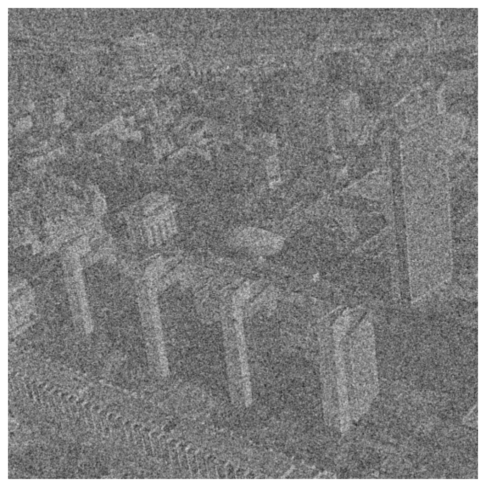


N = 256



N = 512

N = 1024

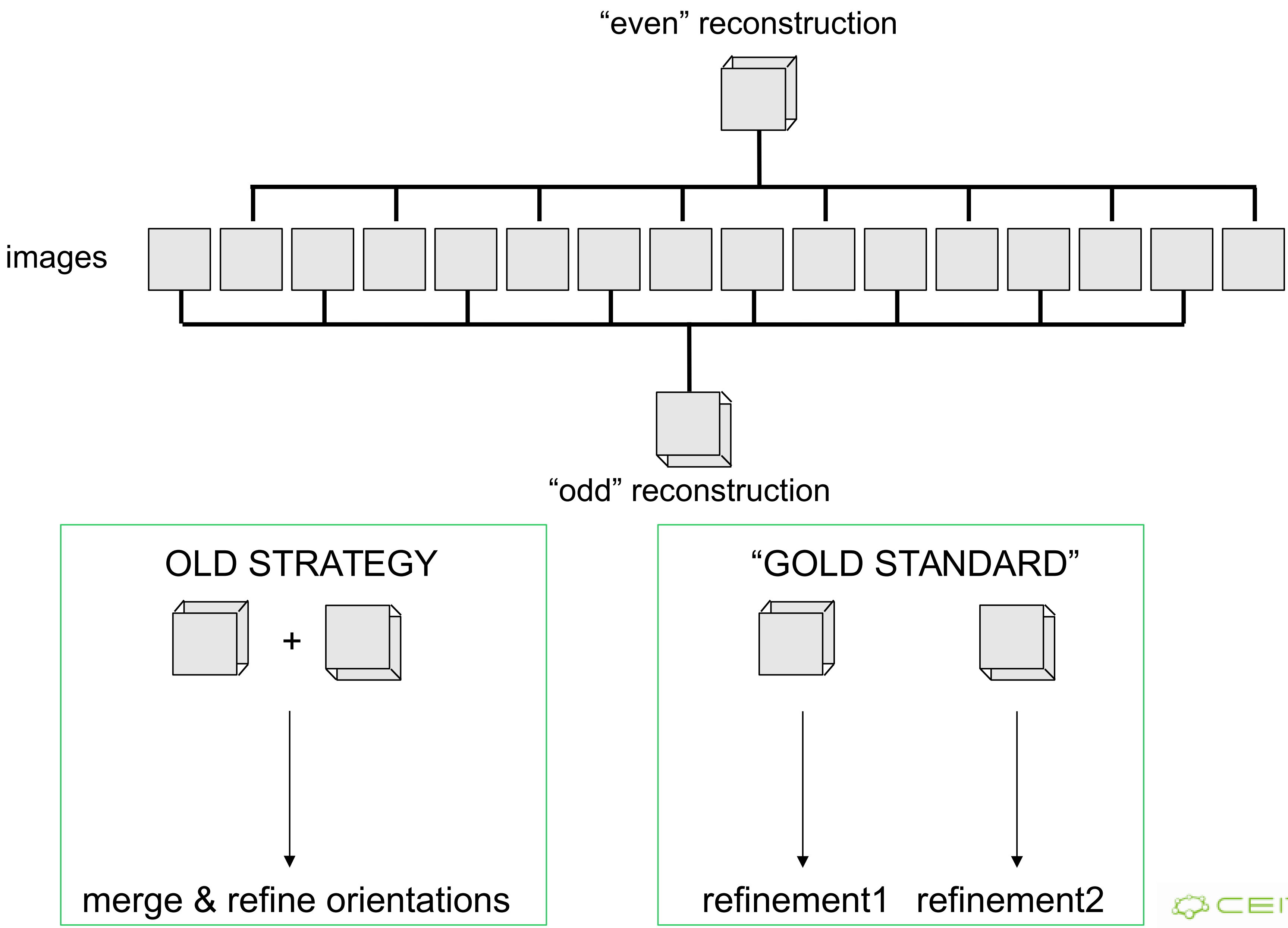


N = 2048

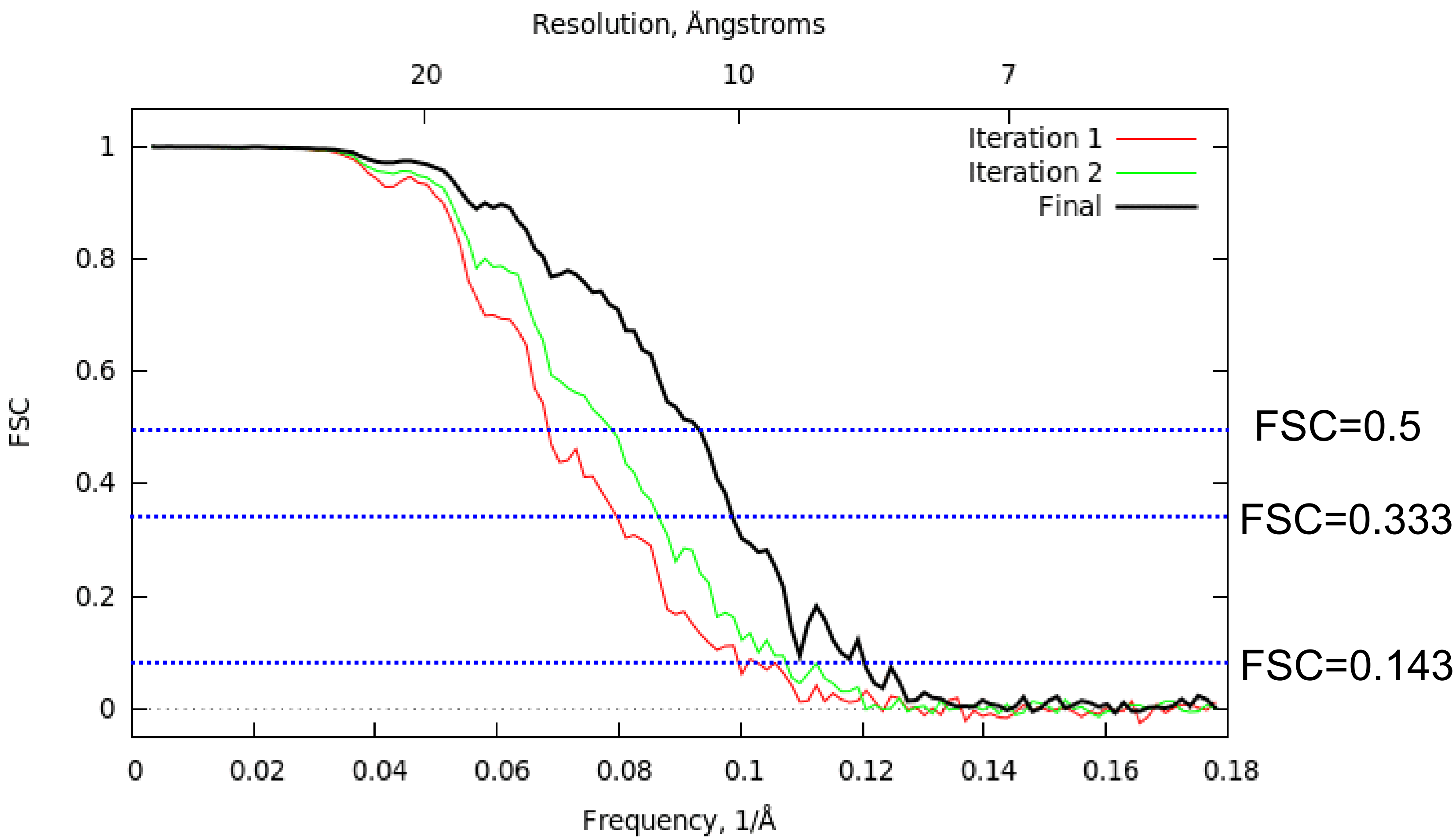
original

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

Refinement: classical and “gold standard”



Different resolution criteria

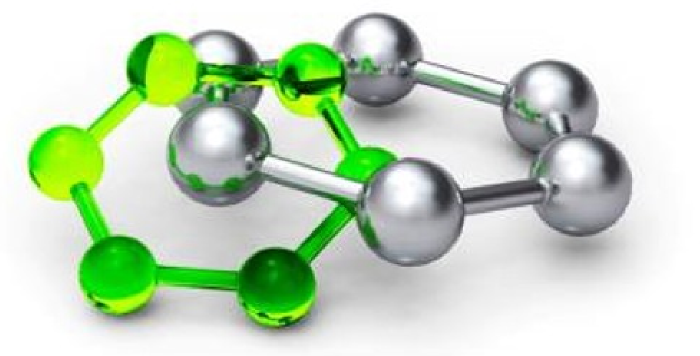


Thank you for your attention

CEITEC

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

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



UNIVERSITAS MASARYKIANA BRUNENSIS

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

2007–13
OP Research and Development for Innovation