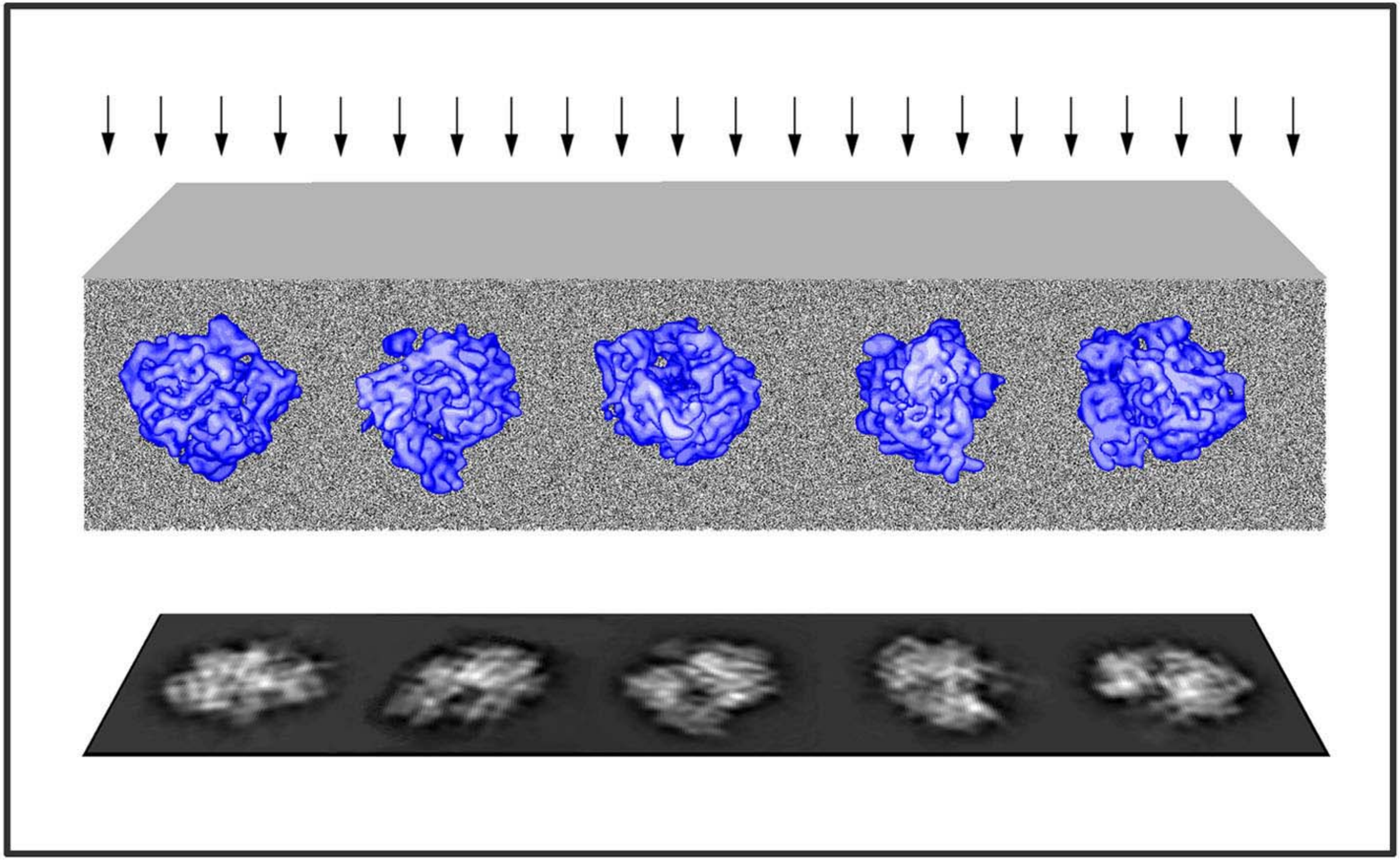
This material is provided for educational use only. The information in these slides including all data, images and related materials are the property of :

Joachim Frank
Wadsworth Center
Empire State Plaza
P.O. Box 509
Albany, New York 12201-0509
Tel: (518) 474-7002

No part of this material may be reproduced without explicit written permission.



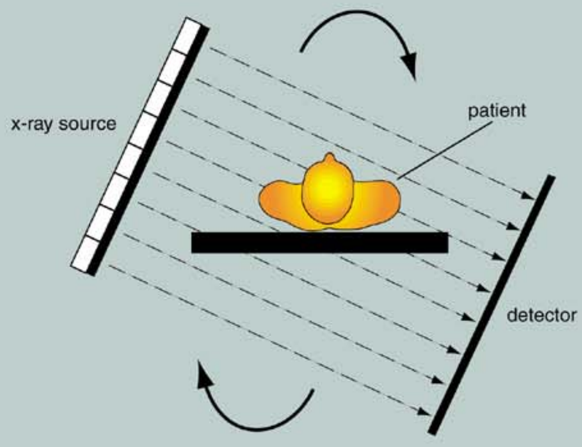
Single-particle reconstruction

Main assumptions:

1) All particles in the specimen have identical structure
2) All are linked by 3D rigid body transformations (rotations, translations)
3) Particle images are interpreted as a “signal” part (= the projection of the common structure) plus “noise”

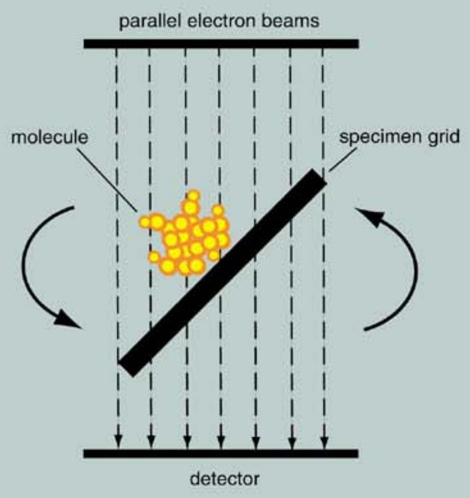
Important requirement:

even angular coverage, without major gaps.



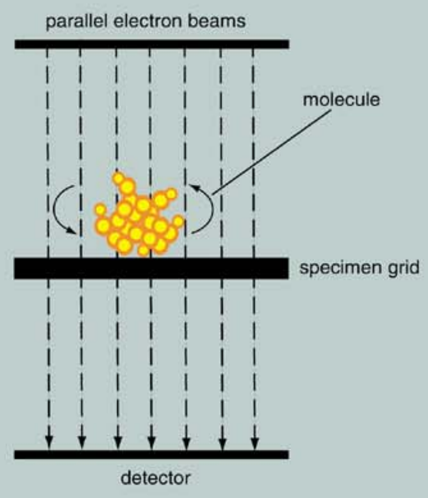
CAT - scan

- beam rotating
- patient stationary



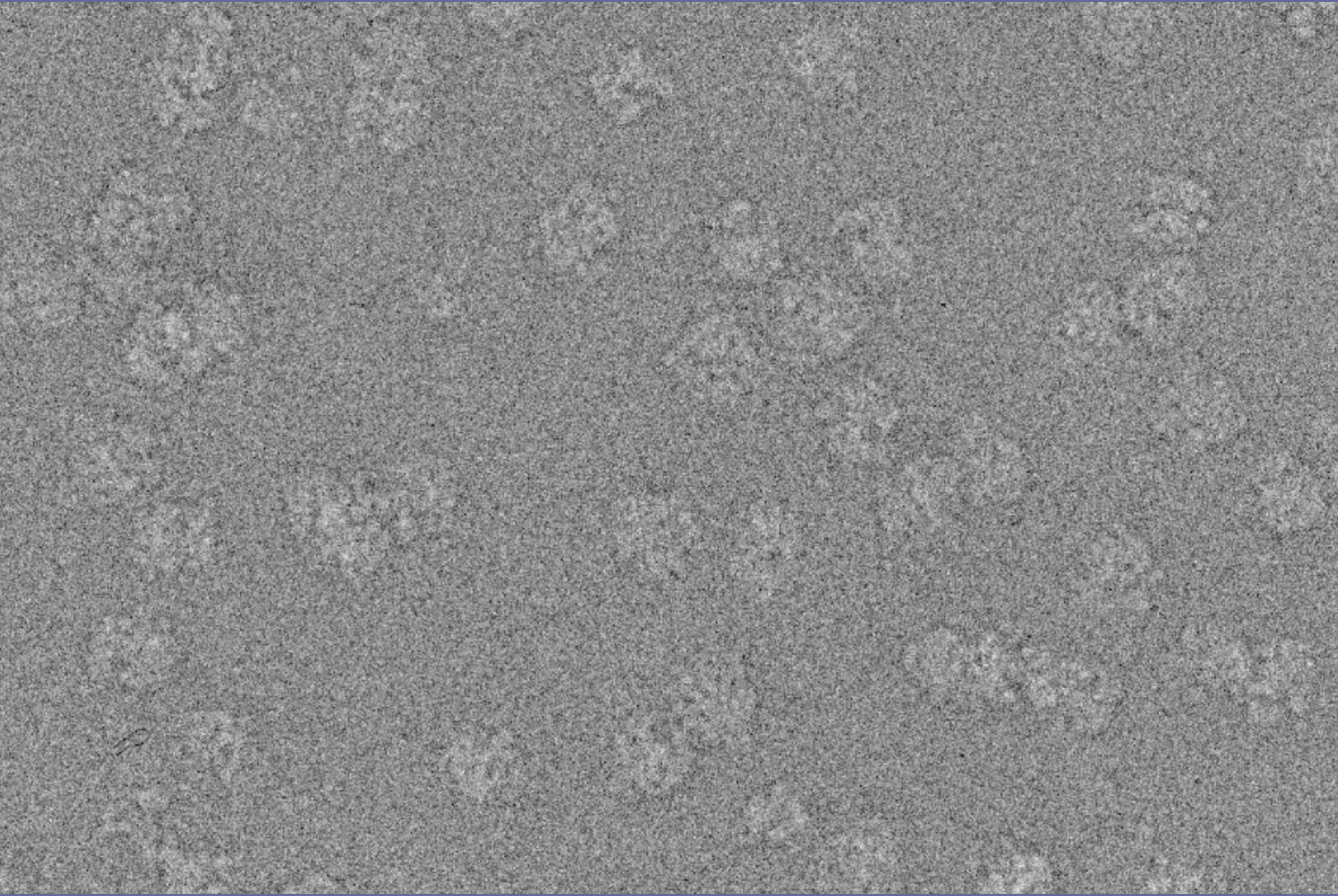
Electron Tomography

- molecule rotating
- beam stationary

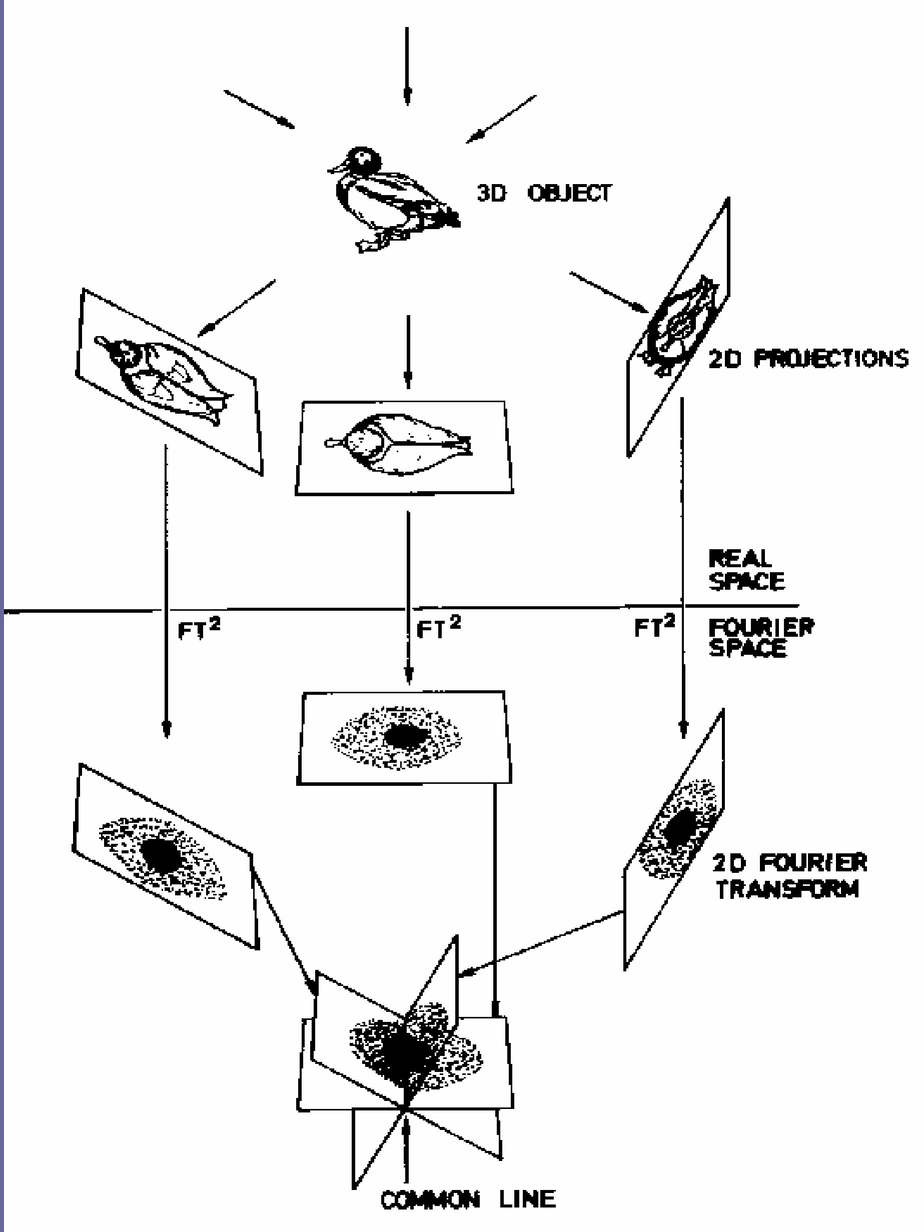


Single particle reconstruction

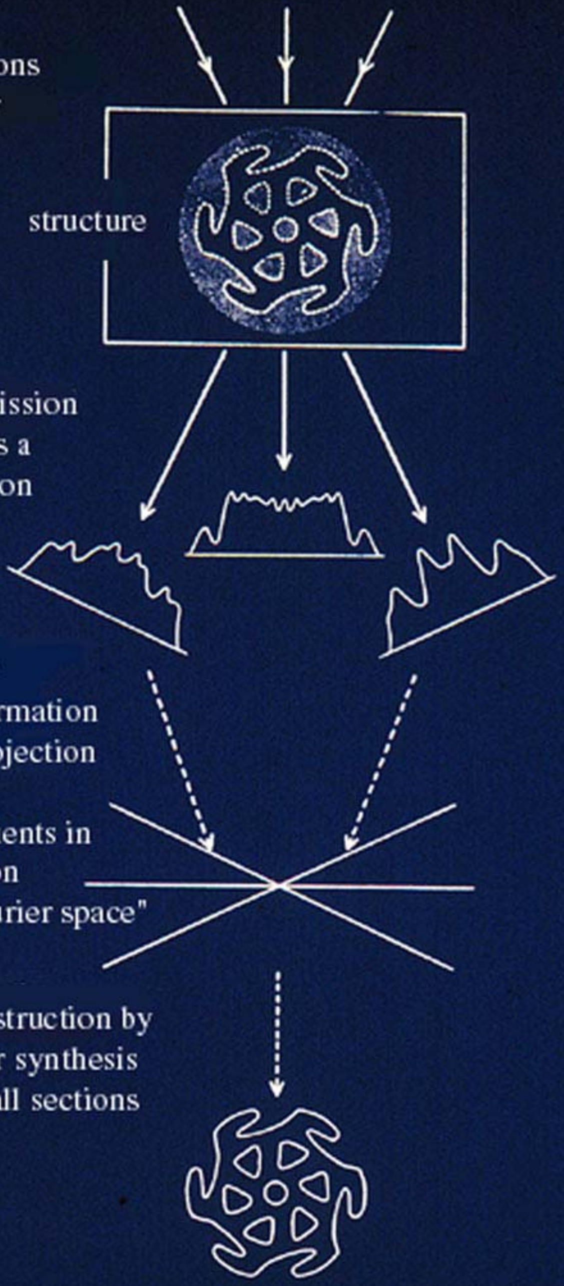
- molecule "rotating"
- beam stationary



Projection Theorem



The two-dimensional Fourier transform of the projection of a three-dimensional density is a central section of three-dimensional Fourier transform of the density perpendicular to the direction of projection.



Directions of view

structure

Transmission image is a projection

Fourier transformation of a projection gives coefficients in a section of "Fourier space"

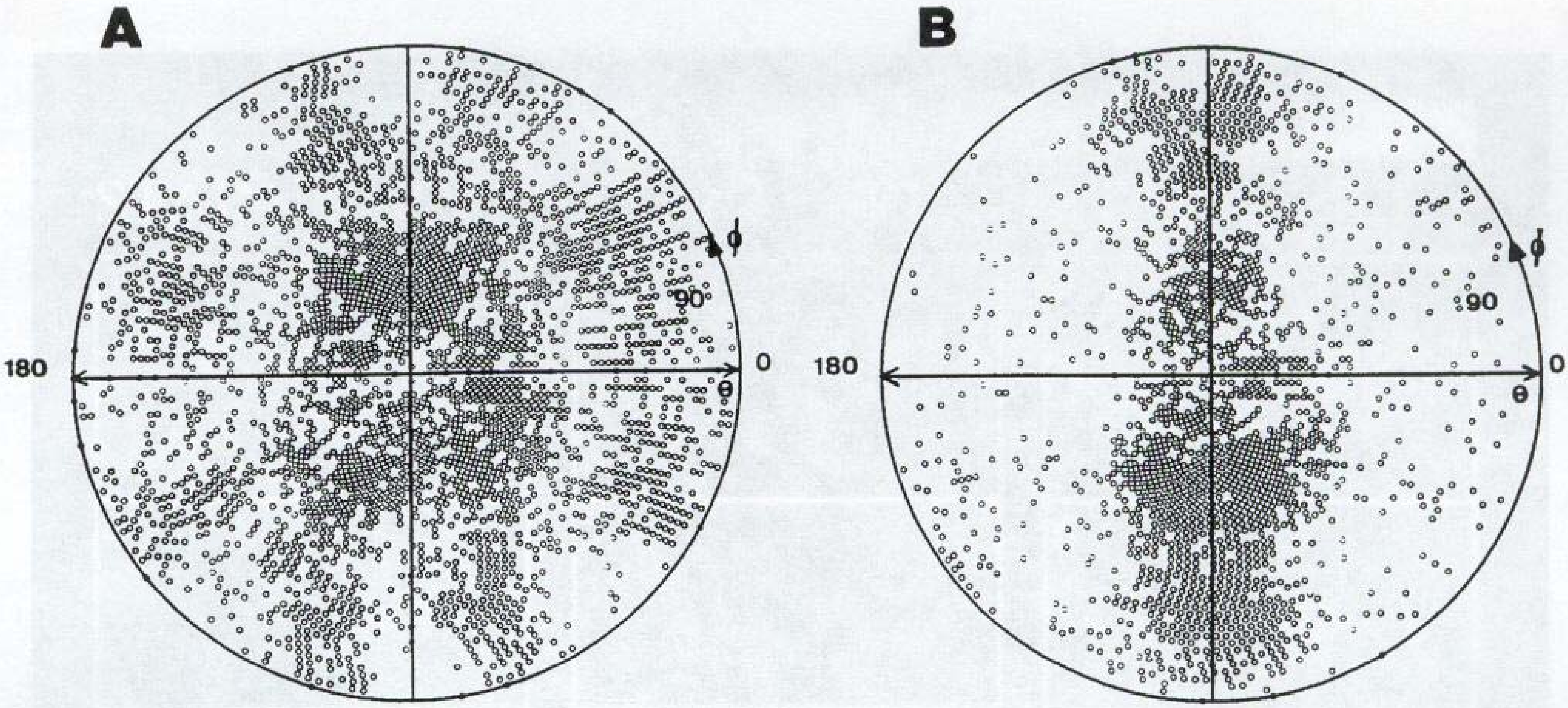
Reconstruction by Fourier synthesis using all sections

DeRosier & Klug, Nature 217 (1968) 133

Angular coverage

good

bad



Overview: the necessary steps of a single-particle reconstruction

1) Optical diffraction: quality control, defocus inventory of micrograph batch
2) Scanning of micrograph batch
3) Determine defoci, and define defocus groups
4) Pick particles
5) Determine particle orientation
6) 3D reconstruction by defocus groups

---Steve----Steve----Steve----Steve----Steve--

7) Refinement
8) CTF correction
9) Validation
10) Interpretation: segmentation, docking, etc.

Overview: tools

1) 2D alignment
    usually by cross-correlation (translational, rotational)
    (a) reference-based
    (b) reference-free
2) Classification
    (a) supervised (multi-reference, 3D projection matching)
    (b) unsupervised
        (i) K-means
        (ii) Hierarchical ascendant
    (c) self-organized maps (SOM)
3) Determine resolution
    (a) phase residual
    (b) Fourier shell correlation
    (c) Spectral signal-to-noise ratio (SSNR)
3) Low-pass filtration
4) Amplitude correction (filter tailored acc. to experimental data)

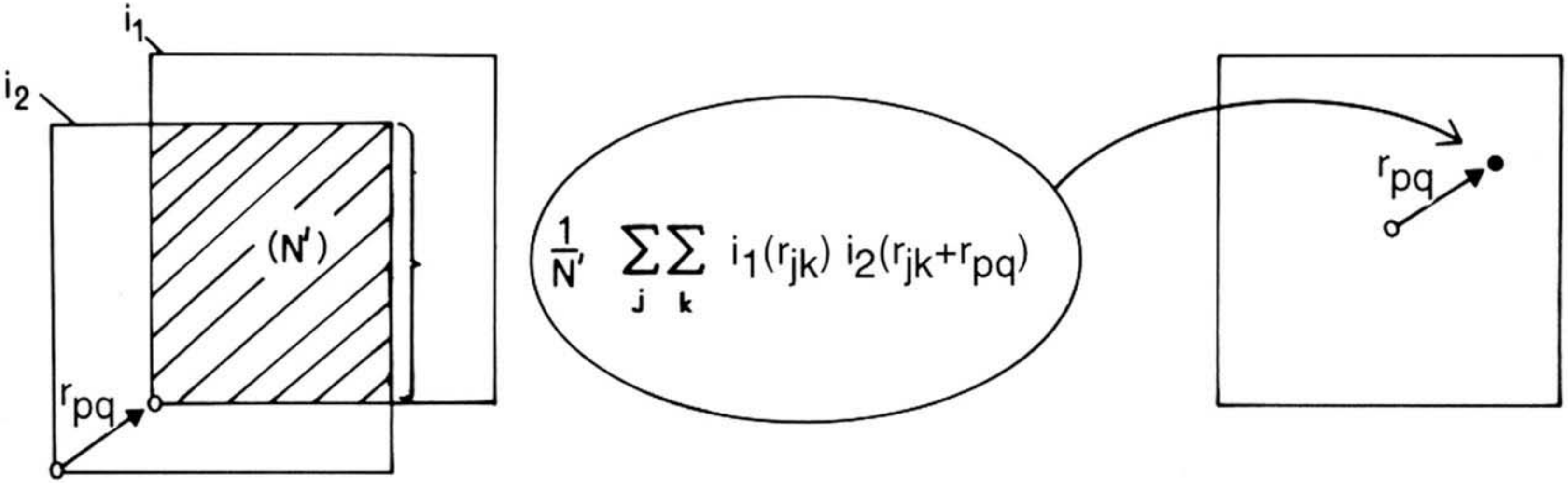


Fig. 3.8. Definition of the cross-correlation function. Image 1 is shifted with respect to image 2 by vector $\mathbf{r}_{pq}$. In this shifted position, the scalar product of the two images arrays is formed and put into the CCF matrix at position (p, q). The vector $\mathbf{r}_{pq}$ is now allowed to assume all positions on the sampling grid. In the end, the CCF matrix has an entry in each position. From Frank (1980). Reproduced with permission of Springer-Verlag, New York.

Alignment methods designed to minimize the influence of the reference

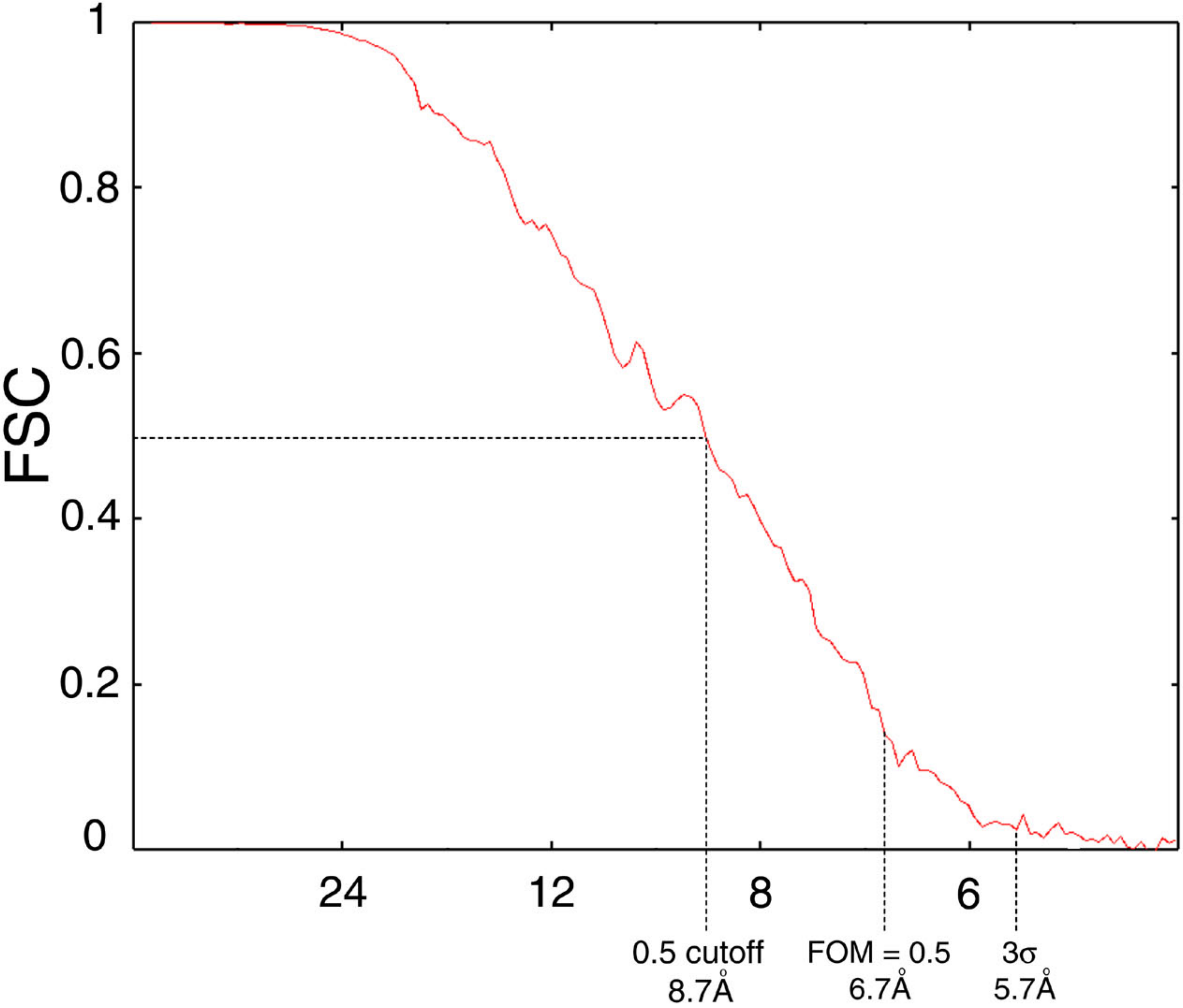
"Reference free" iterative alignment (Penczek *et al*., 1992) **:**
Two images are randomly picked, aligned, and added.
Then, a third image is aligned and added to the previous two. The process is repeated until all images are aligned.

To minimize the influence of the order in which images are picked, the first image is realigned to the [total average - image 1]. Then the second image is realigned to the [total average - image 2], etc …

The whole process is started again until no improvement is found between on alignment cycle and the next.

Resolution measures & criteria: Fourier shell correlation

$$FSC(k,\Delta k) = \frac{Re \left| \sum_{[k,\Delta k]} F_1(\mathrm{k}) F_2^*(\mathrm{k}) \right|}{\left[\sum_{[k,\Delta k]} | F_1(\mathrm{k}) |^2 | F_2(\mathrm{k}) |^2 \right]^{1/2}}$$


1
0.8
0.6
0.4
0.2
0
FSC
24
12
8
6
0.5 cutoff
8.7Å
FOM = 0.5
6.7Å
3σ
5.7Å

Classification

Classification methods are divided into those that are “supervised” and those that are “unsupervised”:

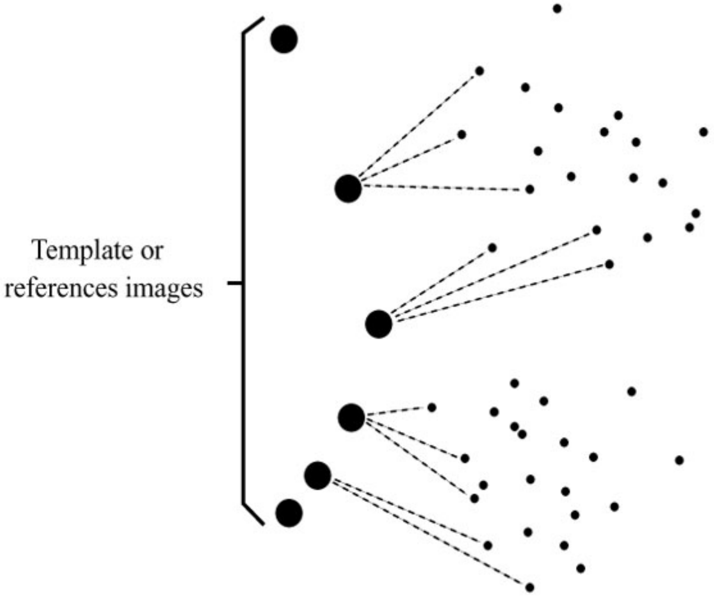
- Supervised: divide or categorize according to similarity with “template” or “reference”.

  Example for application: projection matching

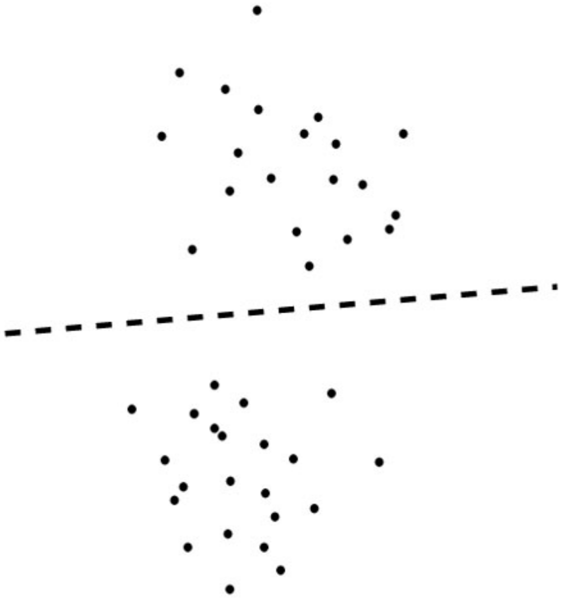
- Unsupervised: divide according to intrinsic properties

  Example for application: find classes of projections presenting the same view

Supervised Classification



Unsupervised Classification



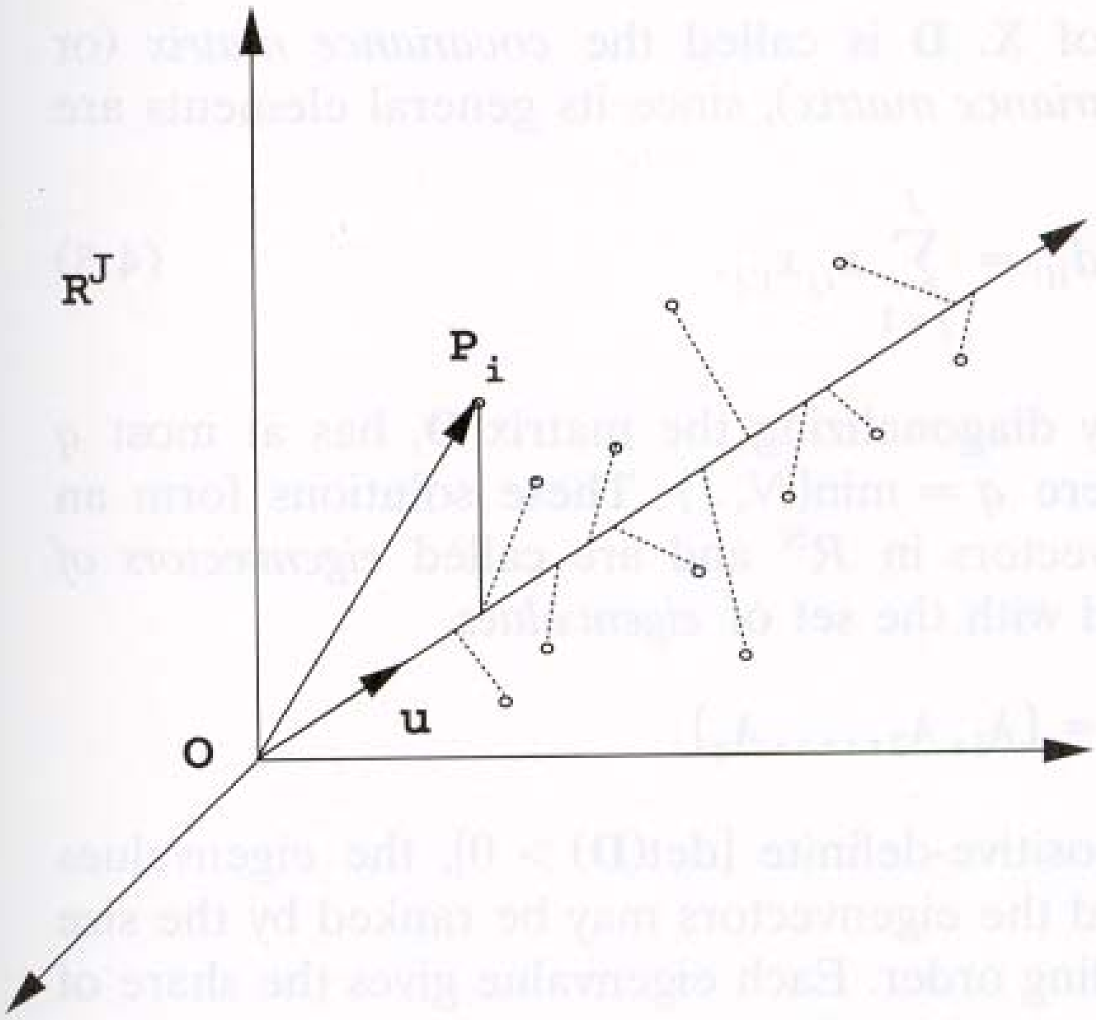
Classification, and the Role of MSA

- Classification deals with “objects” in the space in which they are represented.
- For instance, a 64x64 image is an “object” in a 4096-dimensional space since in principle each of its pixels can vary independently.

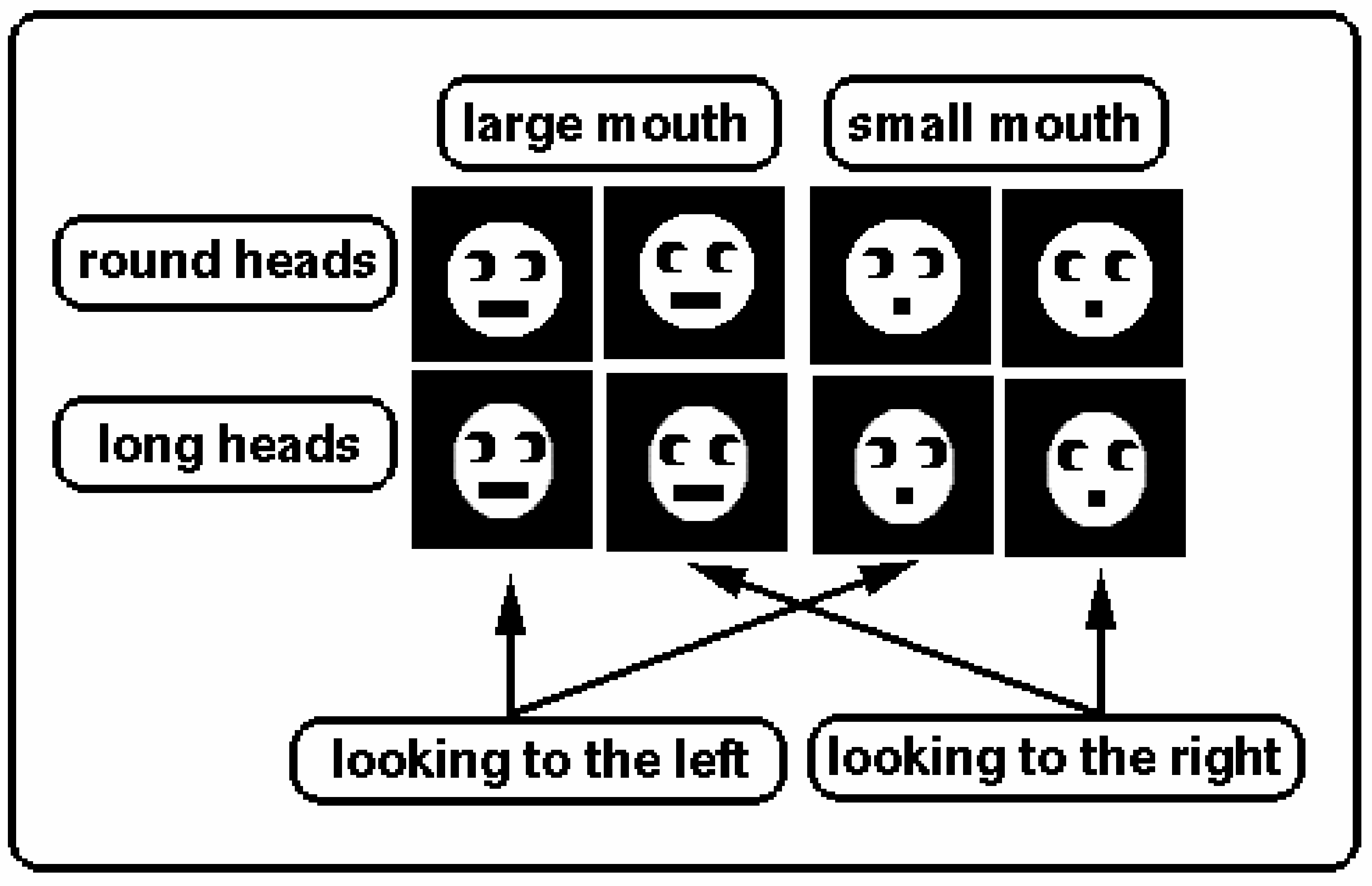
  Let’s say we have 8000 such images. They would form a cloud with 8000 points in this space.
- Unsupervised classification is a method that is designed to find clusters (regions of cohesiveness) in such a point cloud.
- Role of Multivariate Statistical Analysis (MSA): find a space (“factor space”) with reduced dimensionality for the representation of the “objects”. This greatly simplifies classification.
- Reason for the fact that the space of representation can be much smaller than the original space: resolution limitation (neighborhoods behave the same), and correlations due to the physical origin of the variations (e.g., movement of a structural component is represented by correlated additions and subtractions at the leading and trailing boundaries).

Principle of MSA:

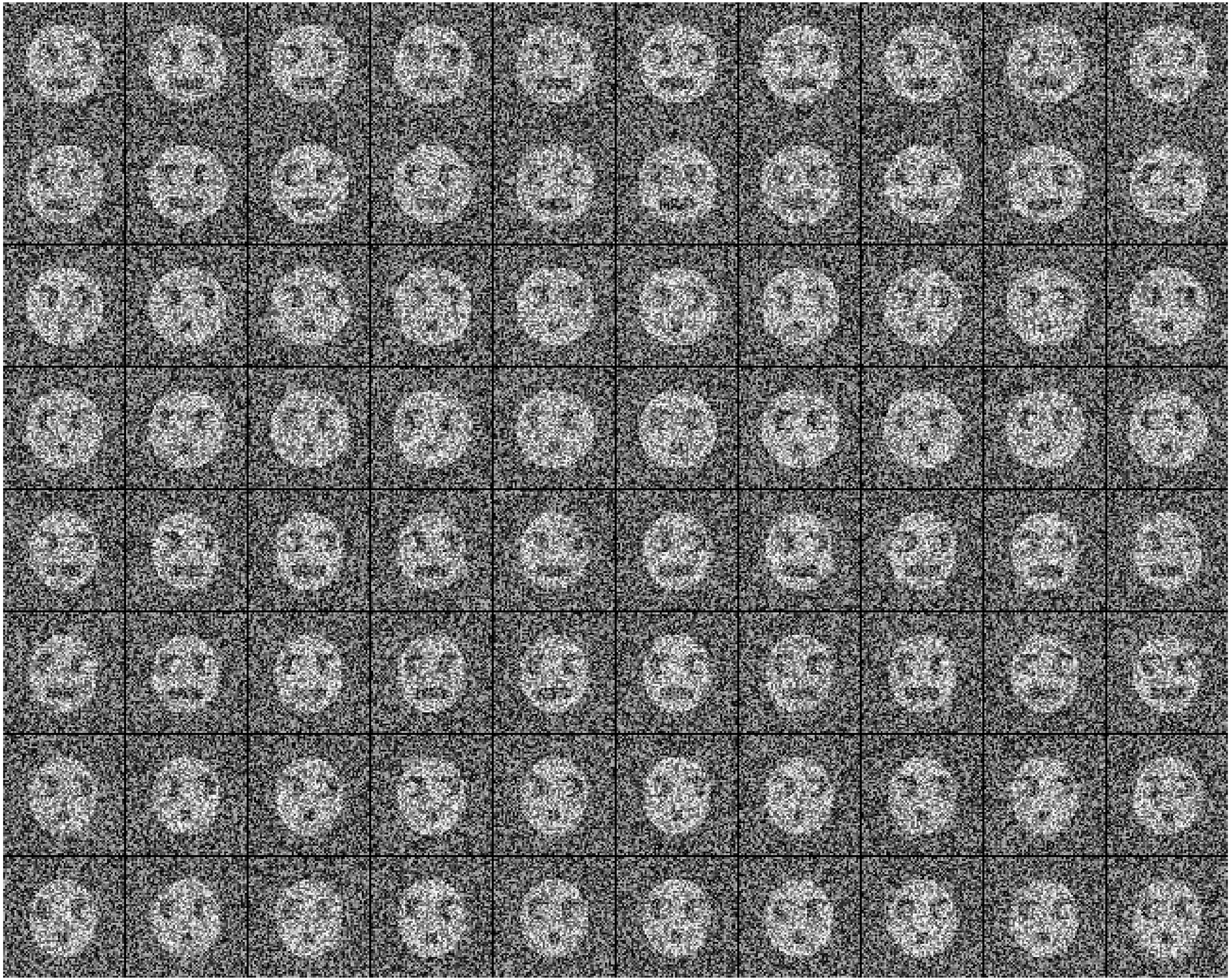
Find new coordinate system, tailored to the data



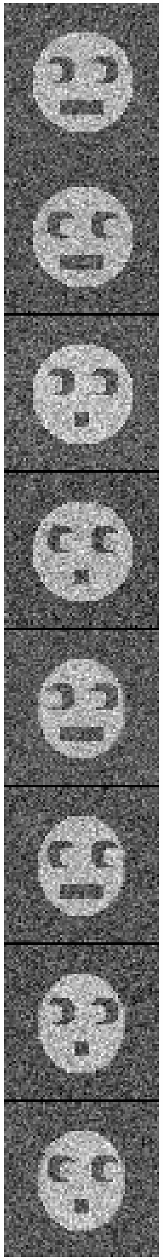
Brétaudière JP and Frank J (1986) Reconstitution of molecule images analyzed by correspondence analysis: A tool for structural interpretation. *J. Microsc.* **144**, 1-14.



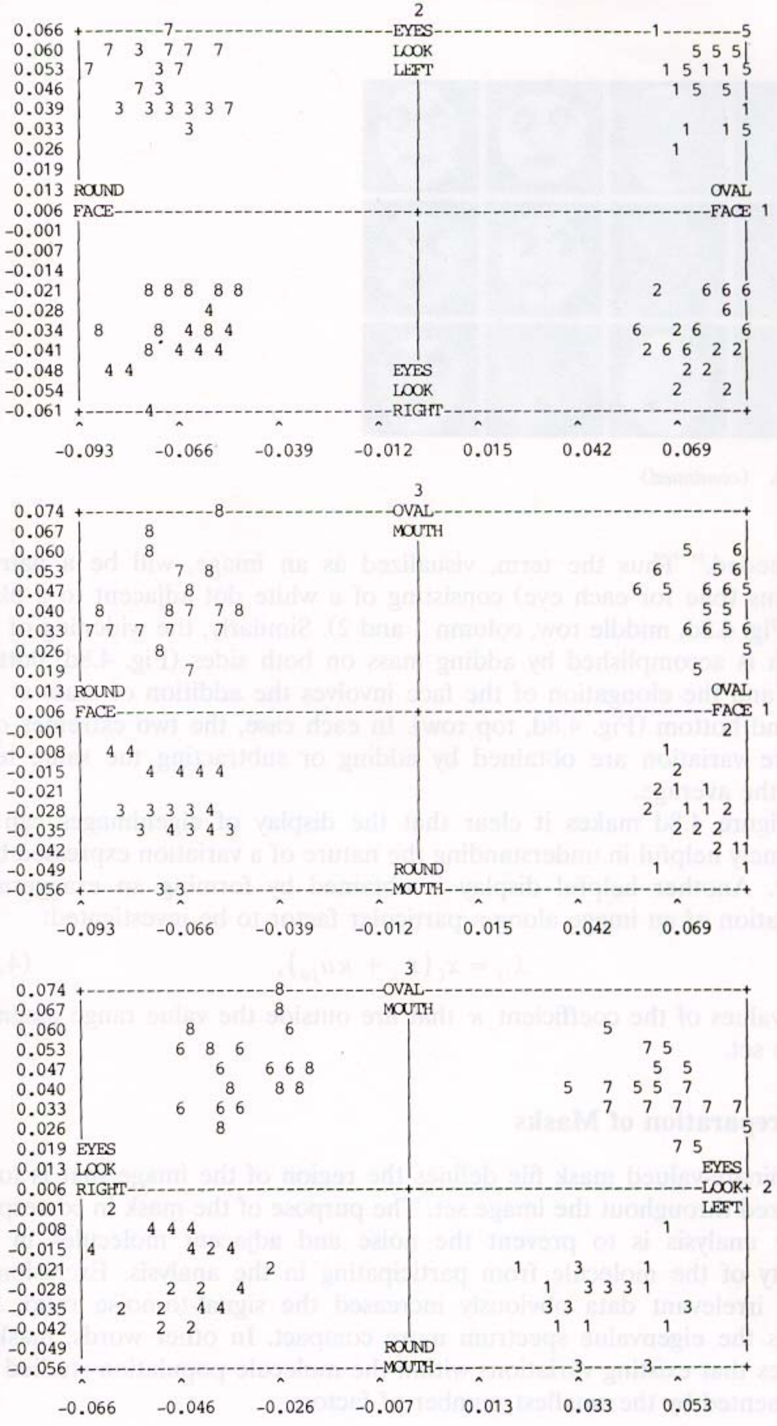
10 copies of the 8 types of heads + random noise



Averages



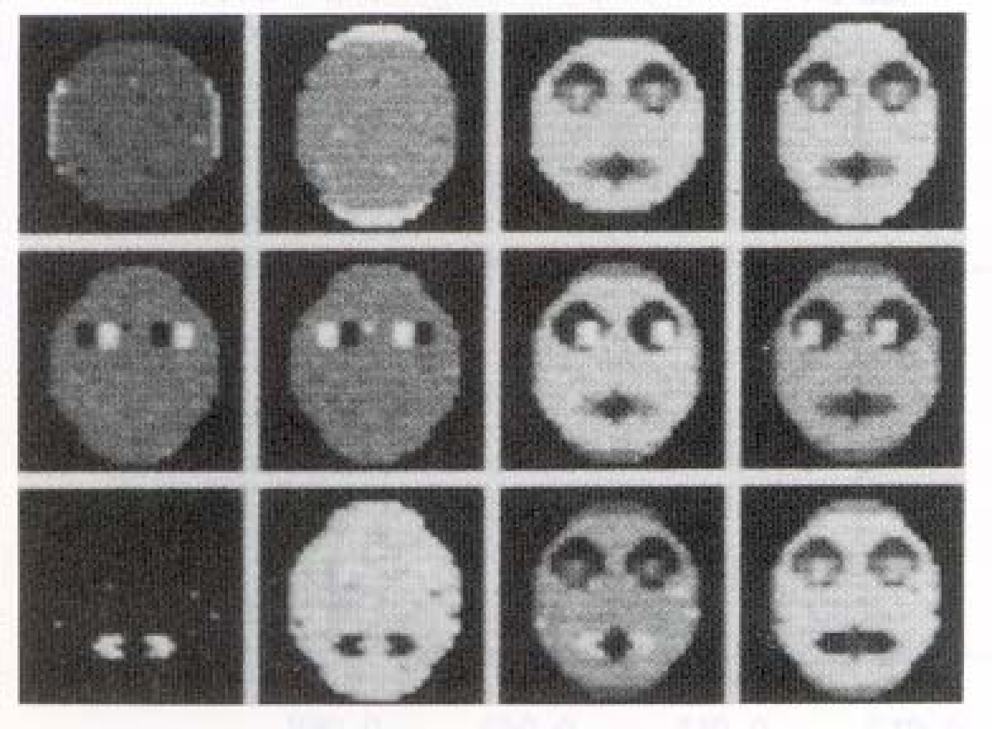
C



EYES LOOK LEFT
ROUND FACE
OVAL FACE
EYES LOOK RIGHT
0.066
0.060
0.053
0.046
0.039
0.033
0.026
0.019
0.013
0.006
-0.001
-0.007
-0.014
-0.021
-0.028
-0.034
-0.041
-0.048
-0.054
-0.061
-0.093
-0.066
-0.039
-0.012
0.015
0.042
0.069
OVAL MOUTH
ROUND FACE
OVAL FACE
ROUND MOUTH
0.074
0.067
0.060
0.053
0.047
0.040
0.033
0.026
0.019
0.013
0.006
-0.001
-0.008
-0.015
-0.021
-0.028
-0.035
-0.042
-0.049
-0.056
-0.093
-0.066
-0.039
-0.012
0.015
0.042
0.069
OVAL MOUTH
EYES LOOK RIGHT
EYES LOOK LEFT
ROUND MOUTH
0.074
0.067
0.060
0.053
0.047
0.040
0.033
0.026
0.019
0.013
0.006
-0.001
-0.008
-0.015
-0.021
-0.028
-0.035
-0.042
-0.049
-0.056
-0.066
-0.046
-0.026
-0.007
0.013
0.033
0.053

MSA: eigenimages

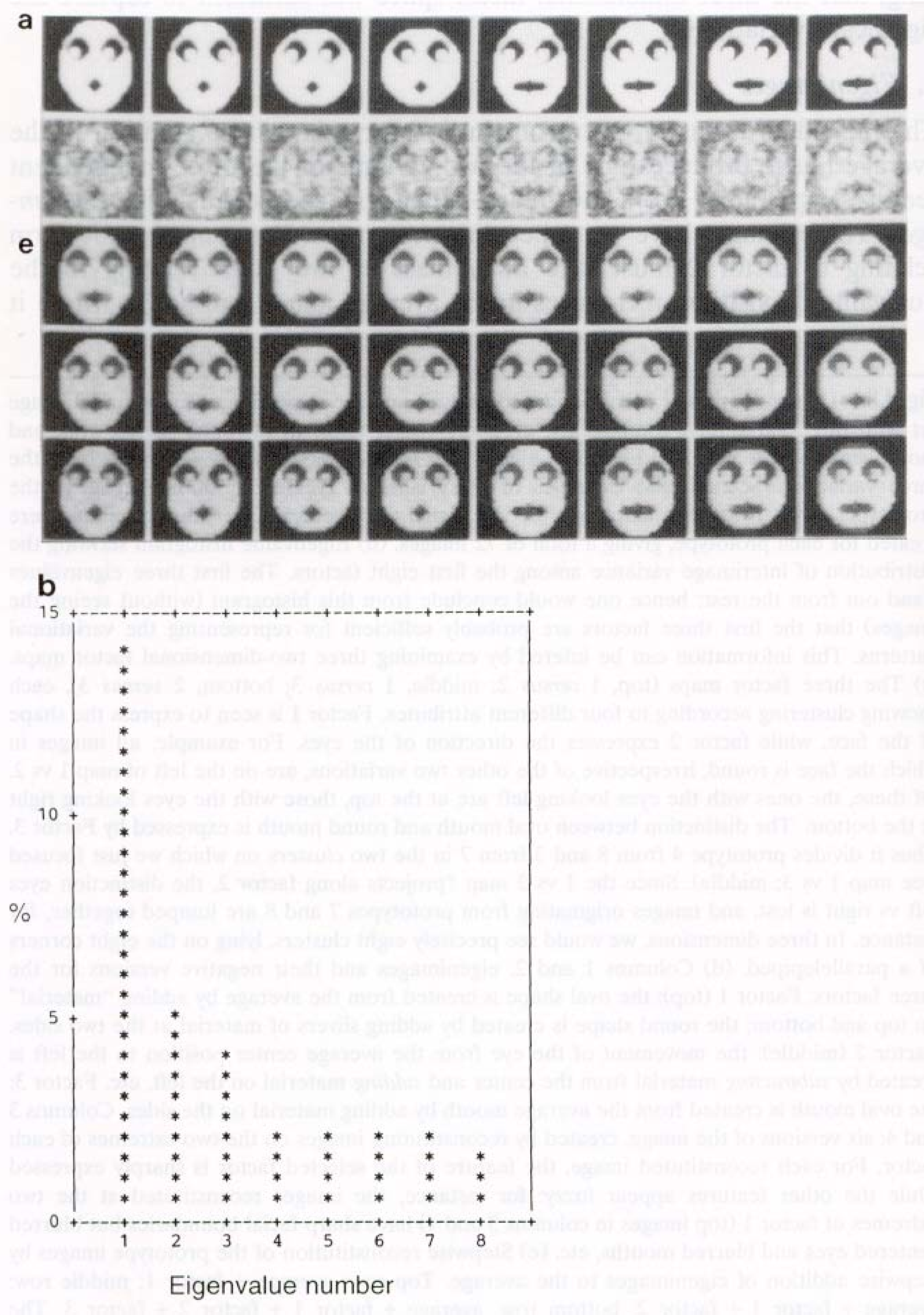
- + - rec + rec -
- Factor 1
- Factor 2
- Factor 3



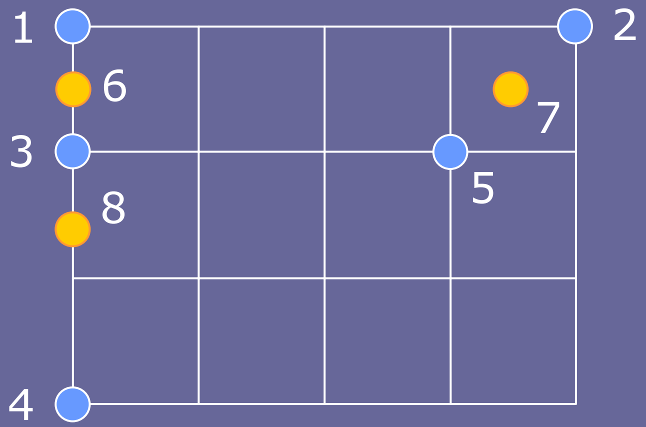
Avrg + F1

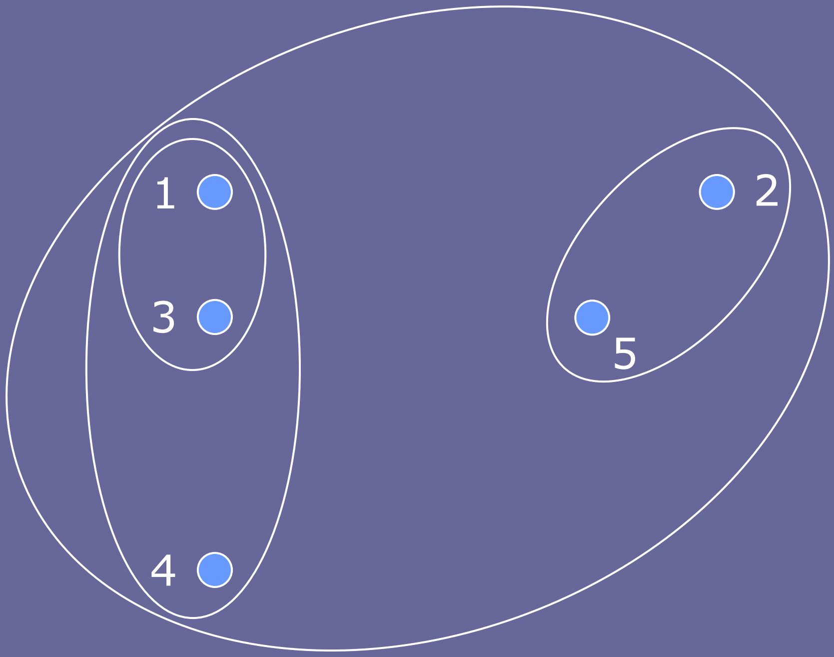
Avrg + F1+F2

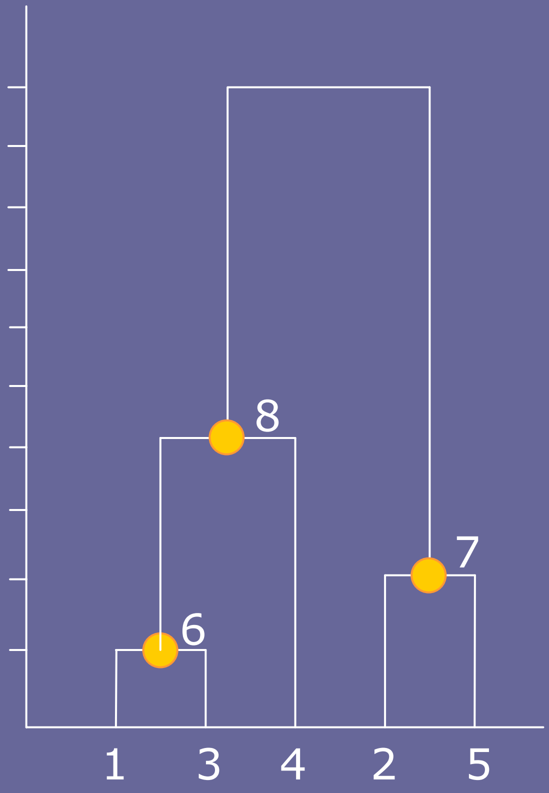
Avrg + F1+F2+F3



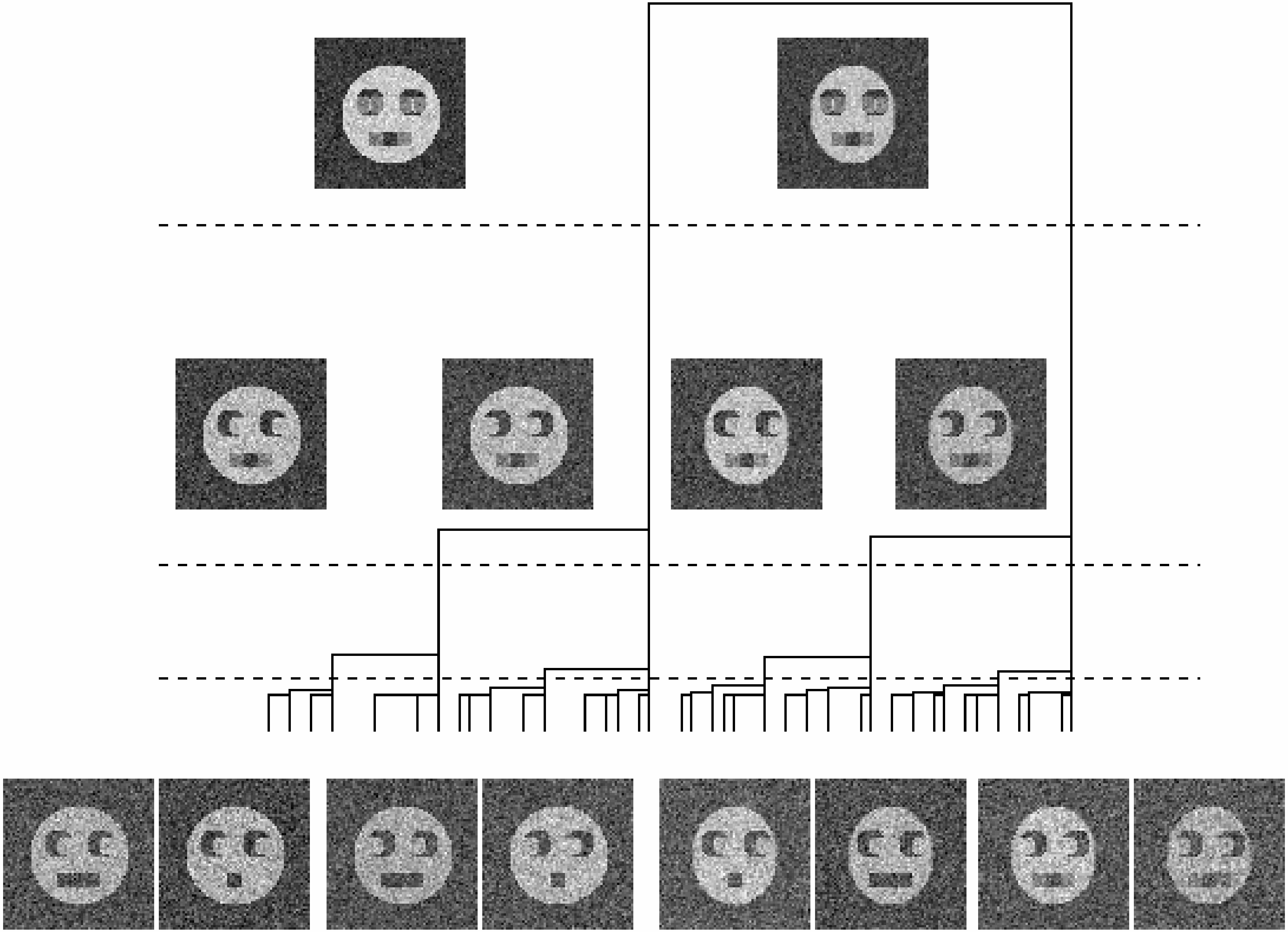
Hierarchical Ascendant Classification





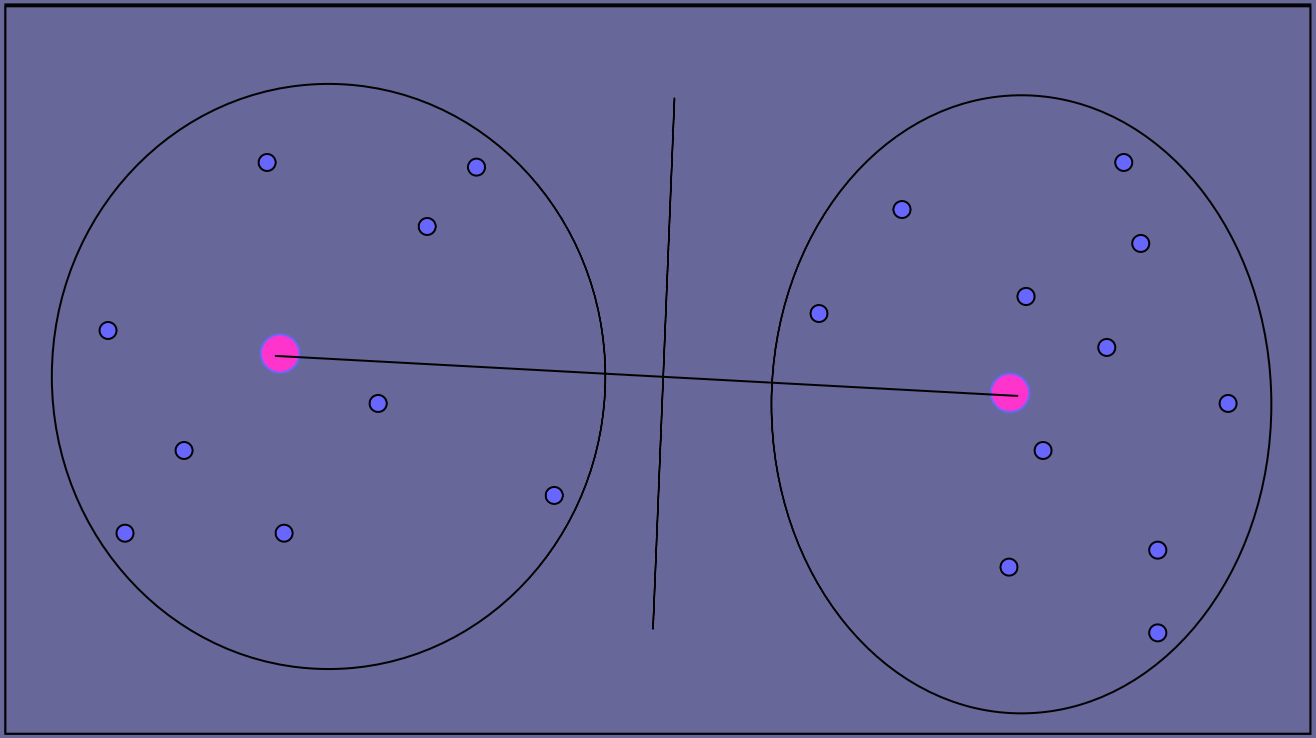


HIERARCHICAL ASCENDENT CLASSIFICATION

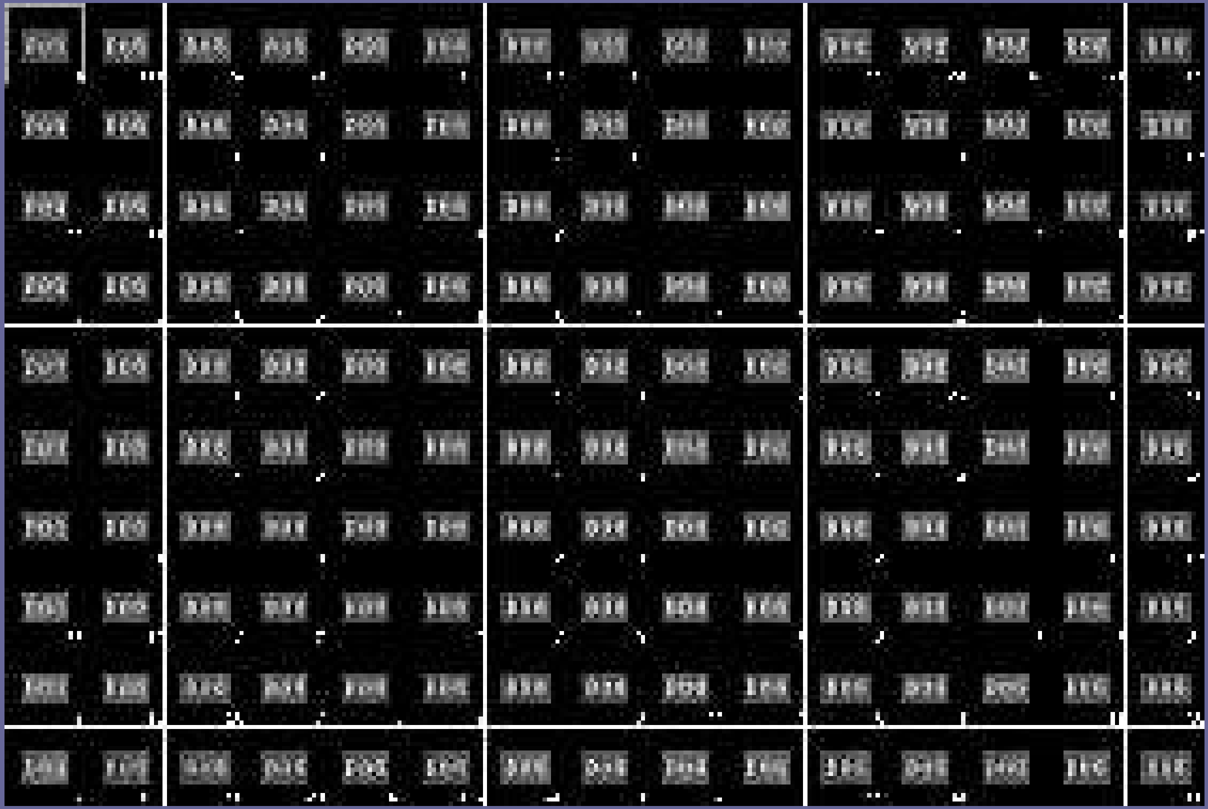


Partition methods : *e.g.* "Moving seeds" method

Diday E (1971) La methode des nuèes dynamiques. *Rev. Stat. Appl*. **19**, 19-34.



Stops when centers don't move from one step to the next or after a given a selected number of iterations.



Overview: the necessary steps of a single-particle reconstruction

1) Optical diffraction: quality control, defocus inventory of micrograph batch
2) Scanning of micrograph batch
3) Determine defoci, and define defocus groups
4) Pick particles
5) Determine particle orientation
6) 3D reconstruction by defocus groups

---Steve----Steve----Steve----Steve----Steve---

7) Refinement
8) CTF correction
9) Validation
10) Interpretation: segmentation, docking, etc.

Overview: the necessary steps of a single-particle reconstruction -- I

1) Optical diffraction: quality control, defocus inventory of micrograph batch

2) Scanning of micrograph batch

3) Determine defoci, and define defocus groups

4) Pick particles

  (a) manual

  (b) automated

5) Determine particle orientation

  (a) unknown structure -- bootstrap

    (i) random-conical (uses unsupervised classification)

    (ii) common lines/ angular reconstitution (uses unsupervised classification)

  (b) known structure

    (i) reference-based (3D projection matching = supervised classification)

    (ii) common lines/ angular reconstitution

Overview: the necessary steps of a single-particle reconstruction -- I

1) Optical diffraction: quality control, defocus inventory of micrograph batch

2) Scanning of micrograph batch

3) Determine defoci, and define defocus groups

4) Pick particles

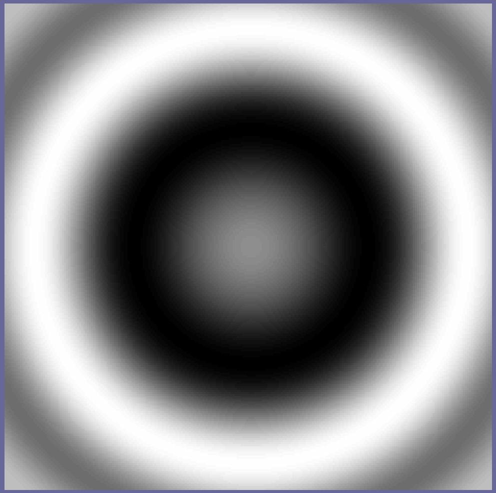
- (a) manual
- (b) automated

5) Determine particle orientation

- (a) unknown structure -- bootstrap
  - (i) random-conical (uses unsupervised classification)
  - (ii) common lines/ angular reconstitution (uses unsupervised classification)
- (b) known structure
  - (i) reference-based (3D projection matching = supervised classification)
  - (ii) common lines/ angular reconstitution

original object

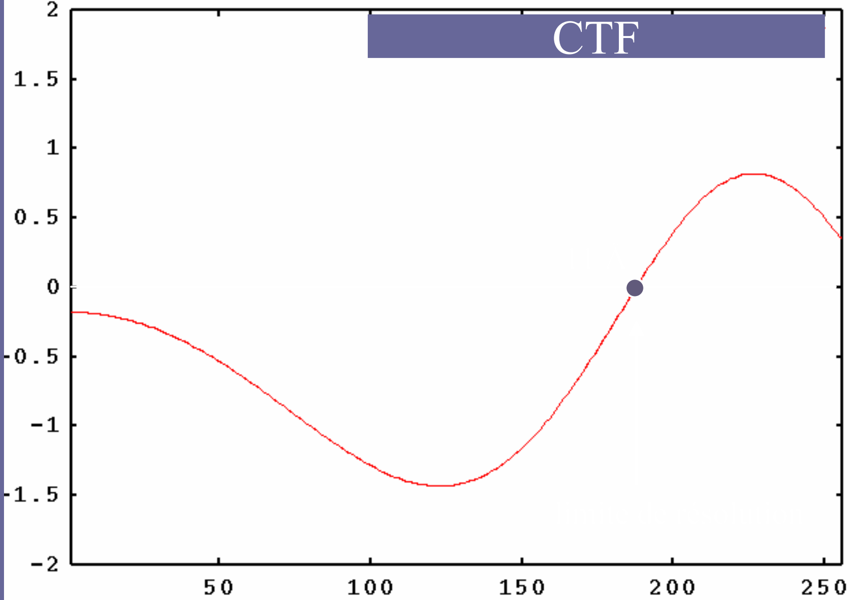
x



CTF for $\Delta z = 0.400\ \mu m$

=

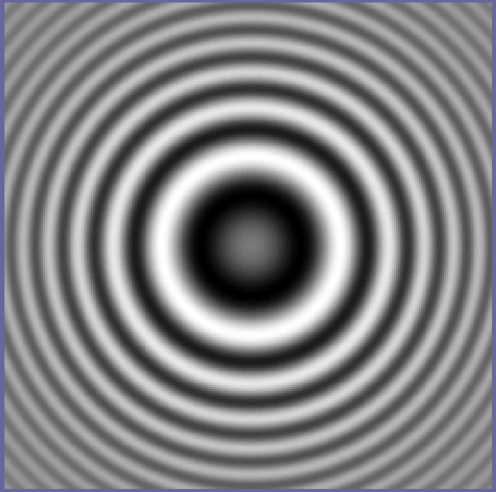
cryo-EM image



cryo-EM image,
contrast-inverted

original object

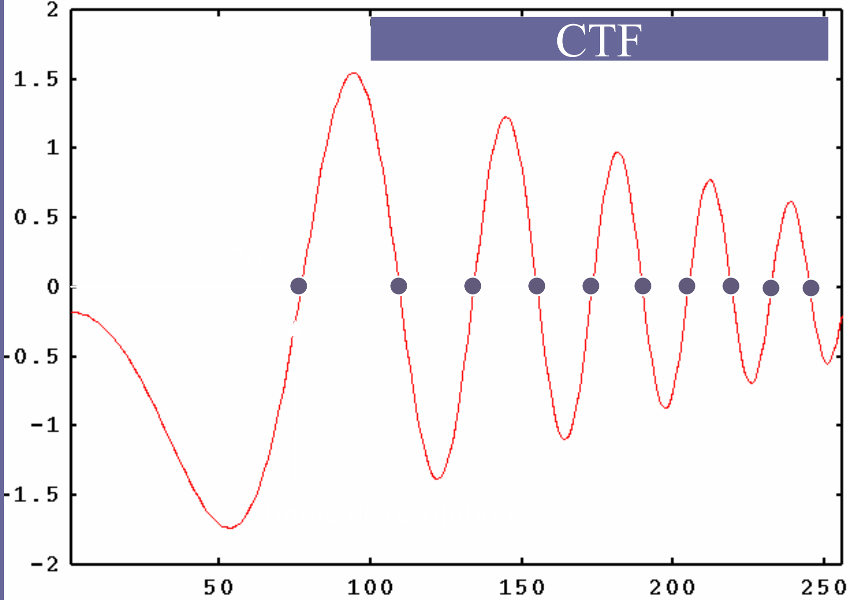
x



CTF for Δz = 2.500 µm

=

cryo-EM image



cryo-EM image,
contrast-inverted

Strategy for reconstruction from multiple defocus groups

- **Coverage of large defocus range required**
- **Data collection must be geared toward covering range without major gap**
- **Characterizing all particles from the same micrograph by the same defocus is OK up to a resolution of ~1/8 A^{-1}.**

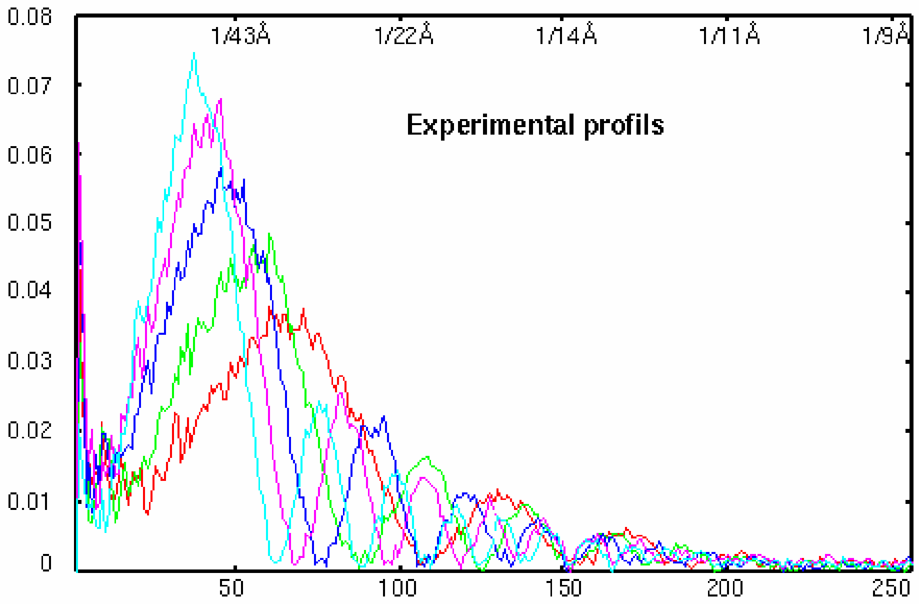
Sequence of steps:

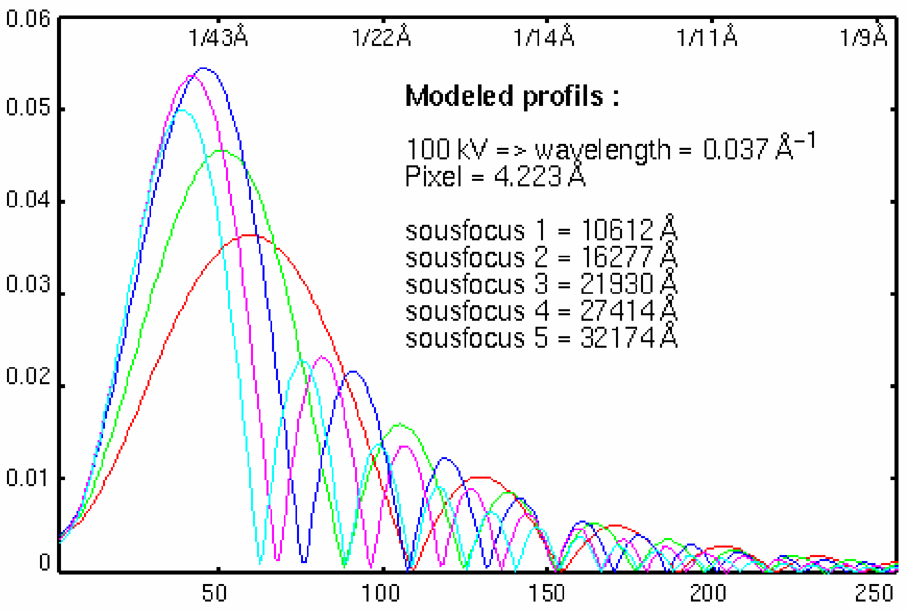
1) Determine defocus for each micrograph

2) Define defocus groups, by creating supersets of particles from micrograps in a narrow range of defoci

3) Process particles separately, by defocus group, till the very end (3D reconstruction by defocus groups)

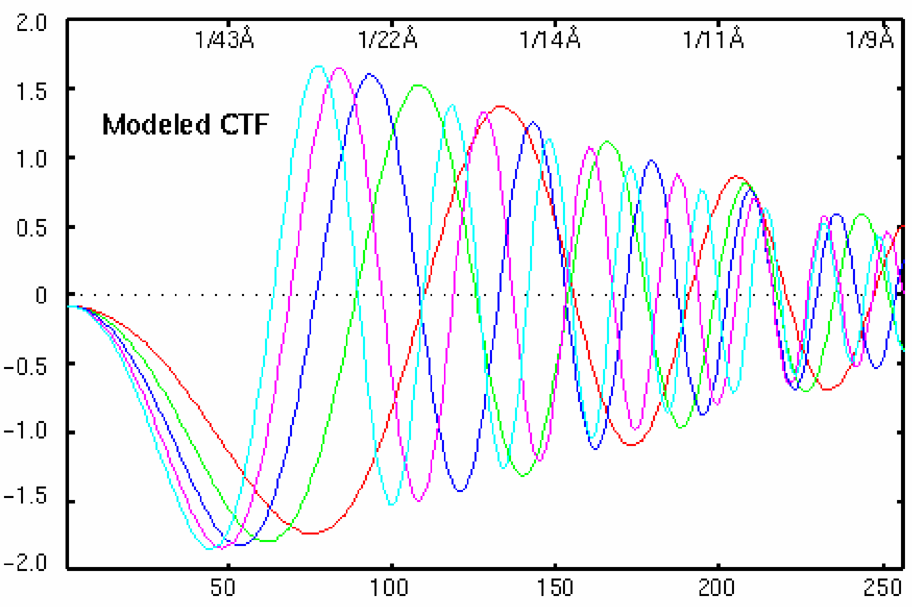
4) Compute merged, CTF-corrected reconstruction. E.g., Wiener filtering.


Experimental profils


Modeled profils :

100 kV => wavelength = 0.037 Å^{-1}
Pixel = 4.223 Å

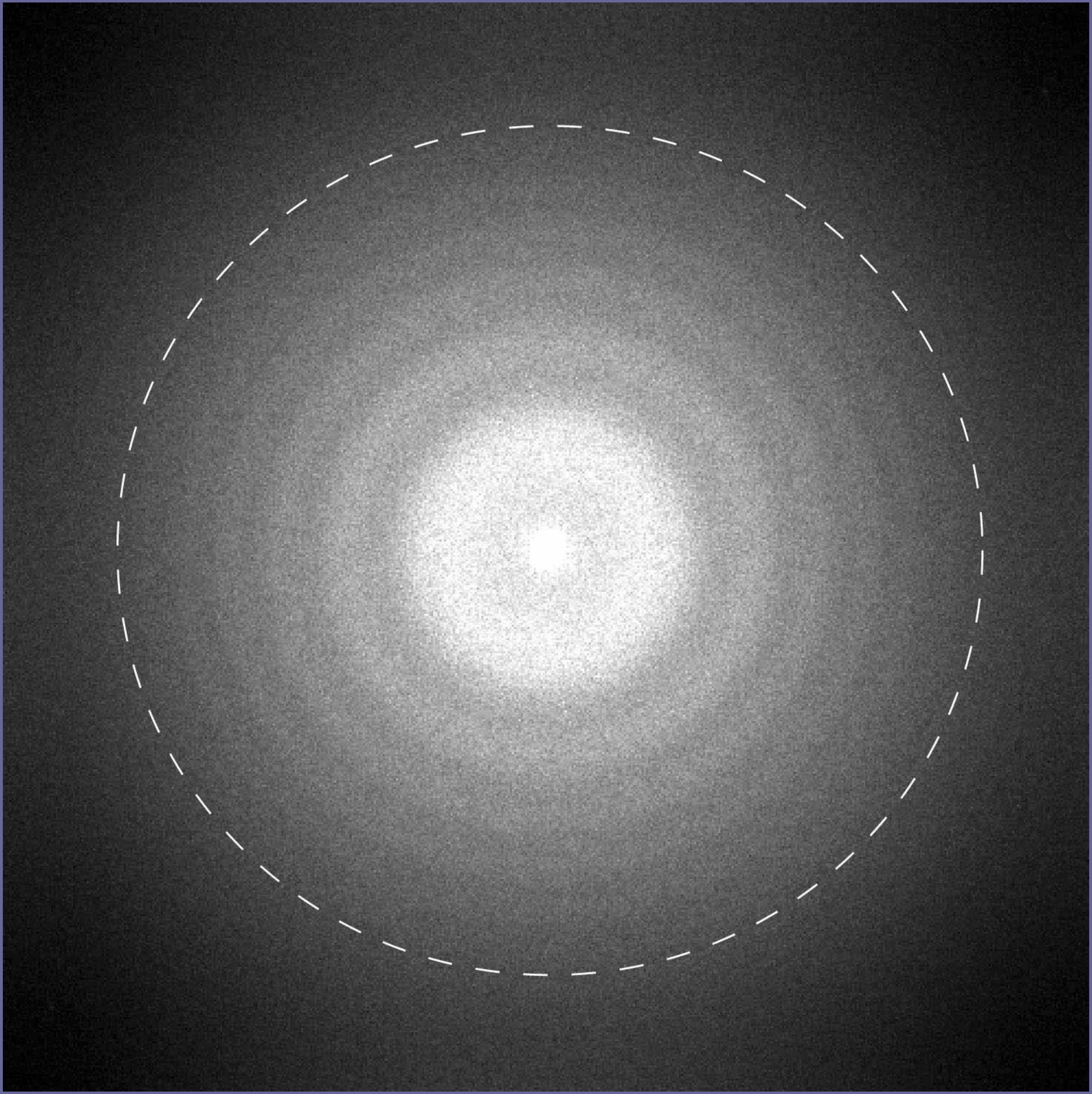
sousfocus 1 = 10612 Å
sousfocus 2 = 16277 Å
sousfocus 3 = 21930 Å
sousfocus 4 = 27414 Å
sousfocus 5 = 32174 Å


Modeled CTF

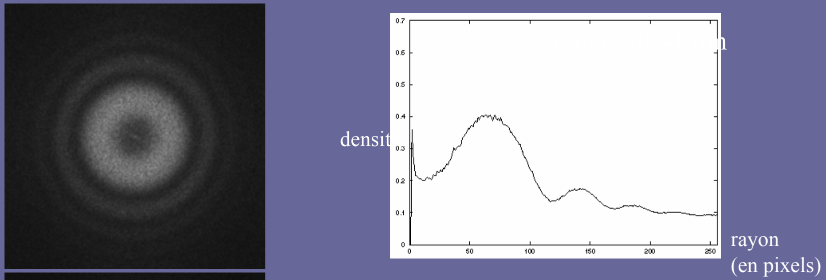
Computation of averaged power spectrum

For each micrograph …

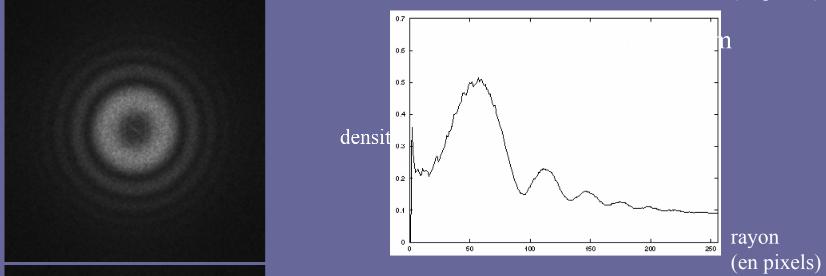
1) Divide field into overlapping subfields of ~512 x 512
2) Compute FFT for each subfield
3) Compute $|F(\mathbf{k})|^2$ for each subfield
4) Form average over $|F(\mathbf{k})|^2$ of all subfields => averaged, smoothed power spectrum
5) Take square root of result => "power spectrum" with reduced dynamic range
6) Form azimuthal average => 1D profile, characteristic for the micrograph, ready to be compared with CTF



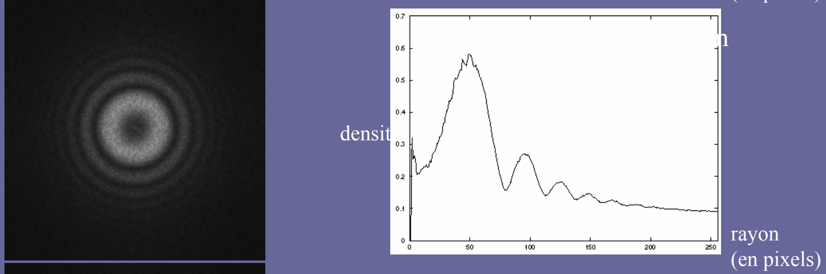
A


densit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
rayon
(en pixels)

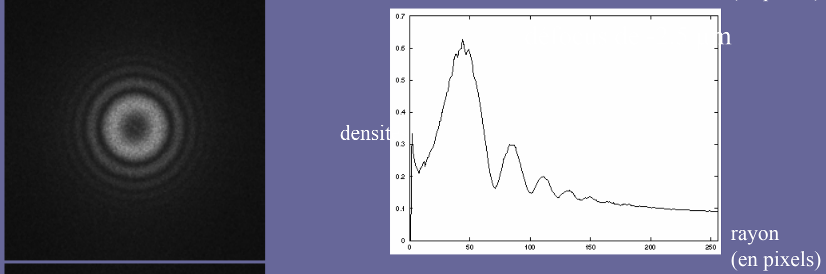
B


densit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
rayon
(en pixels)

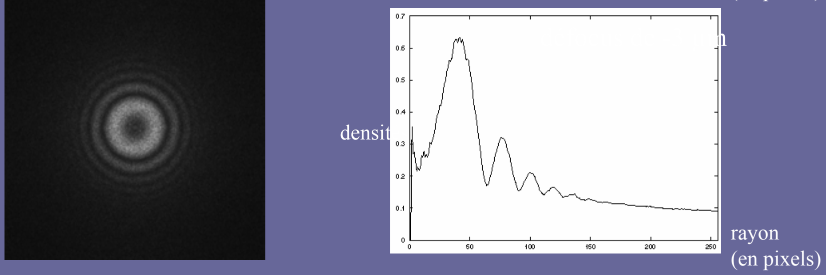
C


densit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
rayon
(en pixels)

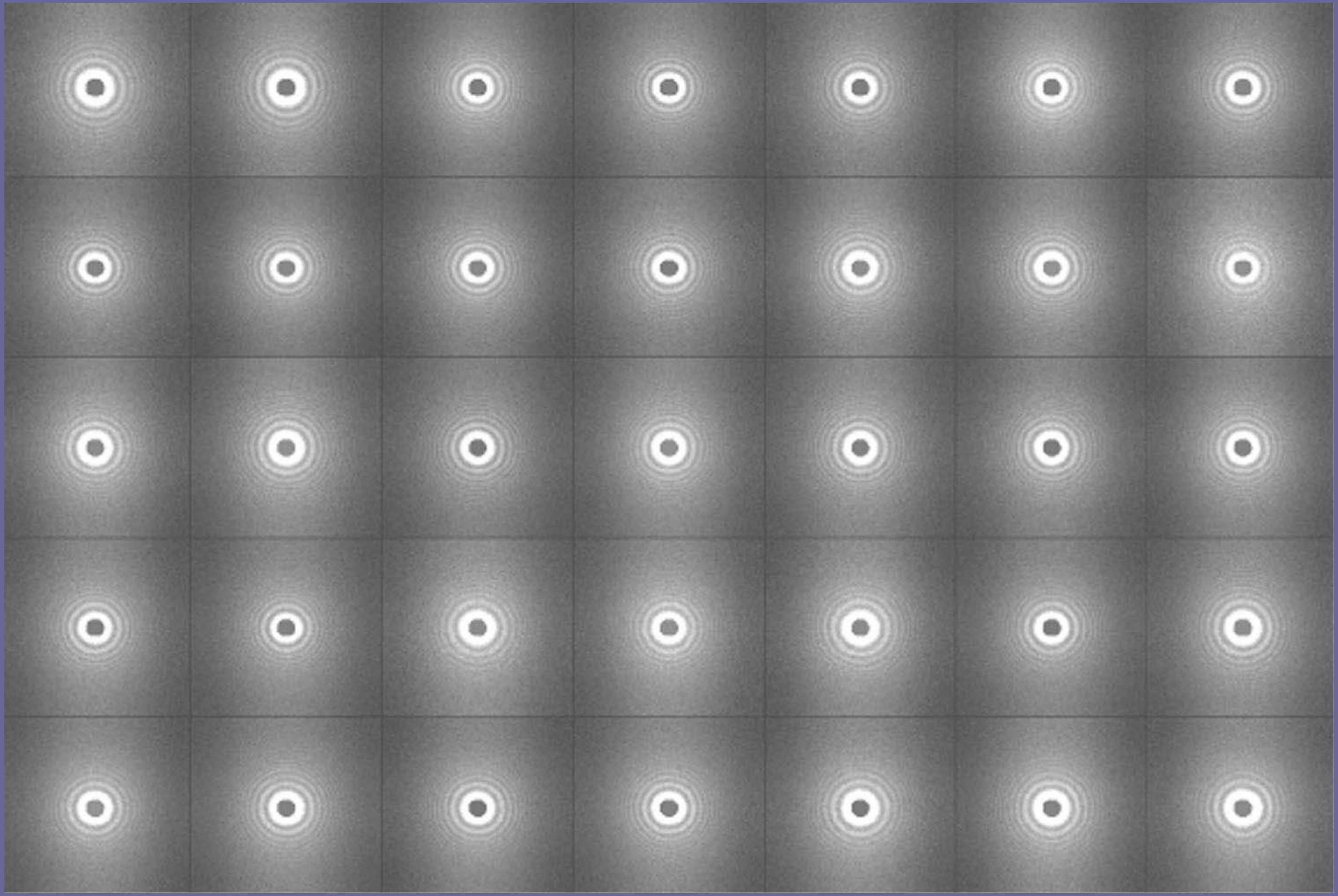
D

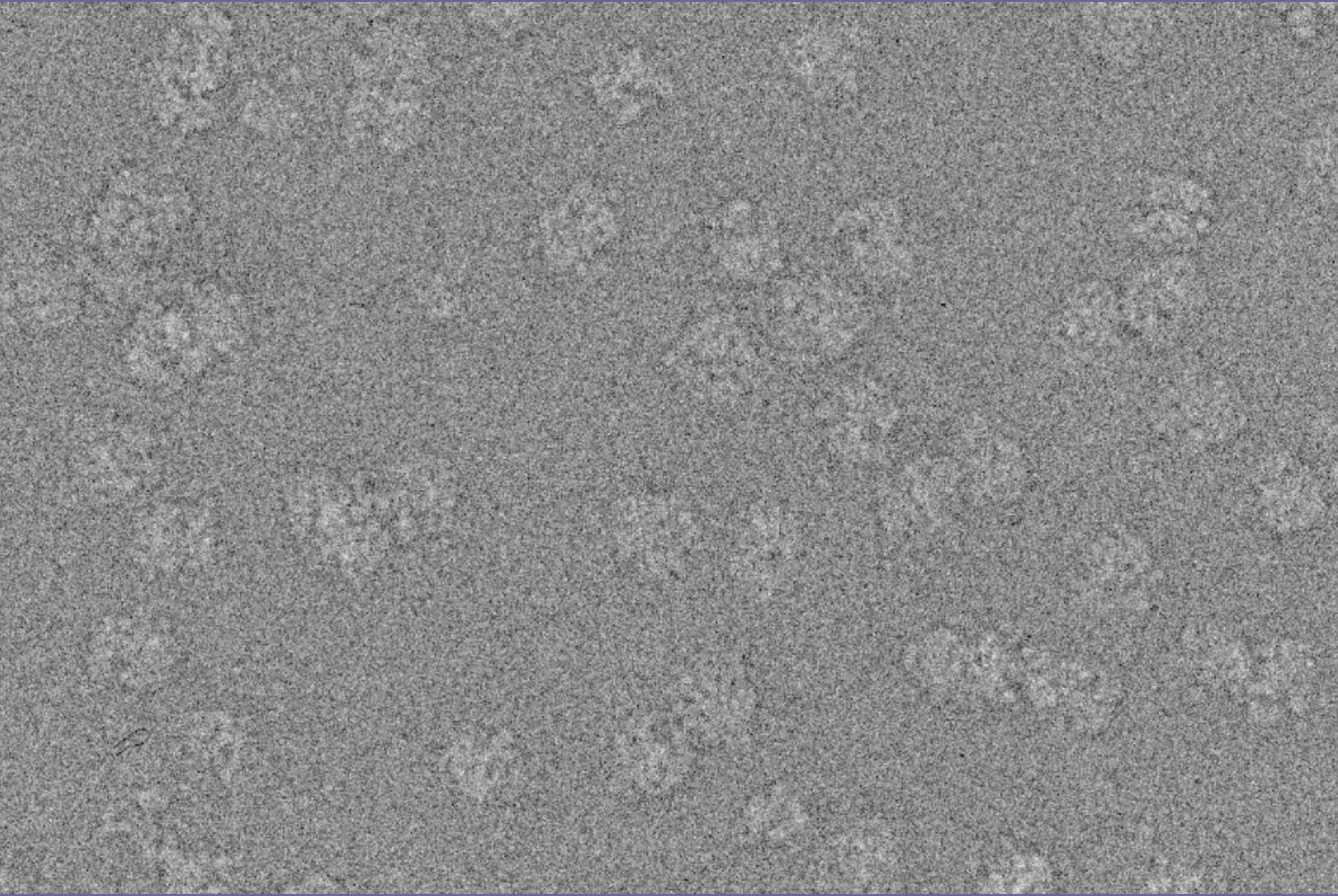

densit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
rayon
(en pixels)

E


densit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
rayon
(en pixels)

Gallery of power spectra from different micrographs





Overview: the necessary steps of a single-particle reconstruction -- I

1) Optical diffraction: quality control, defocus inventory of micrograph batch

2) Scanning of micrograph batch

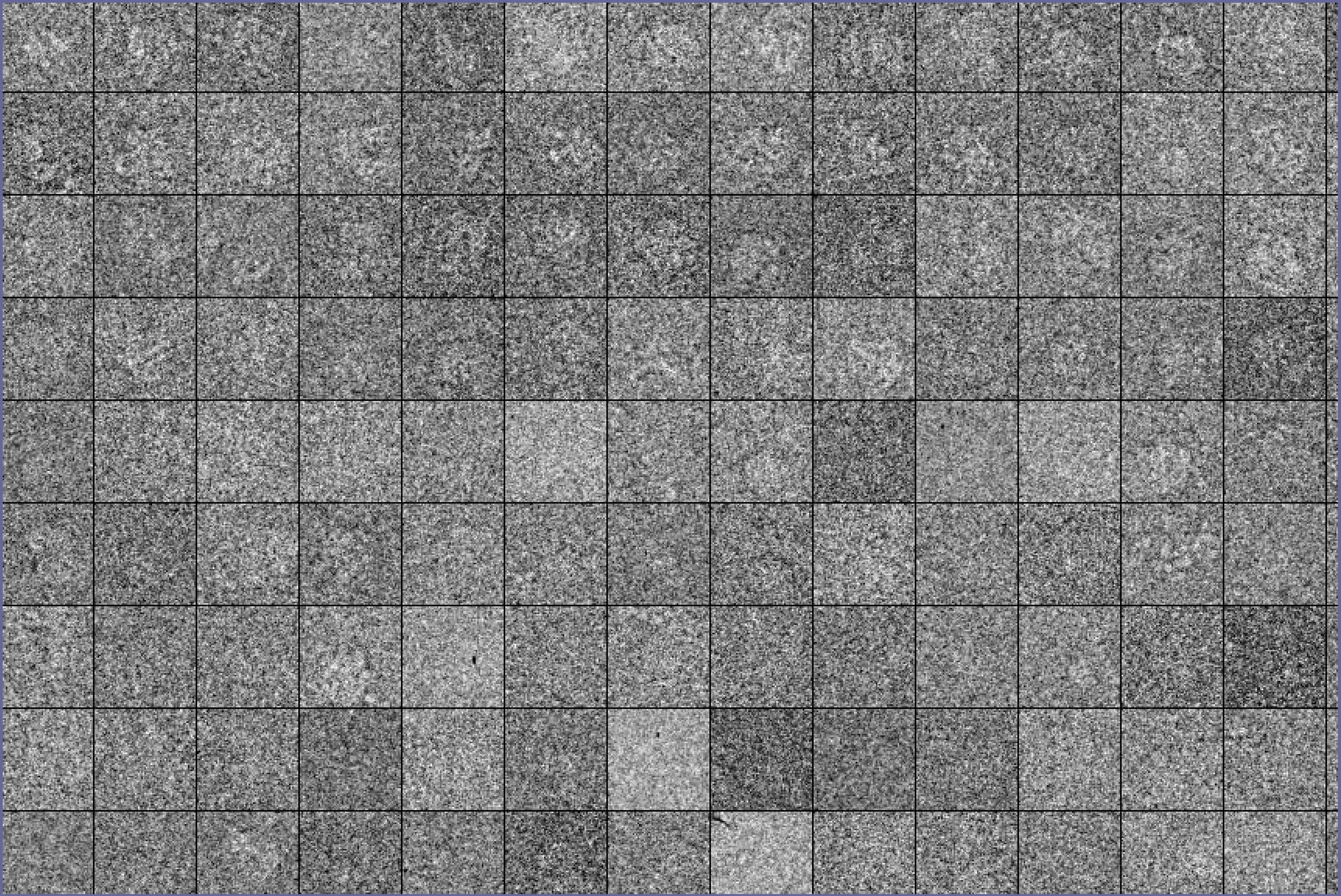
3) Determine defoci, and define defocus groups

4) Pick particles

- (a) manual
- (b) automated

5) Determine particle orientation

- (a) unknown structure -- bootstrap
  - (i) random-conical (uses unsupervised classification)
  - (ii) common lines/ angular reconstitution (uses unsupervised classification)
- (b) known structure
  - (i) reference-based (3D projection matching = supervised classification)
  - (ii) common lines/ angular reconstitution



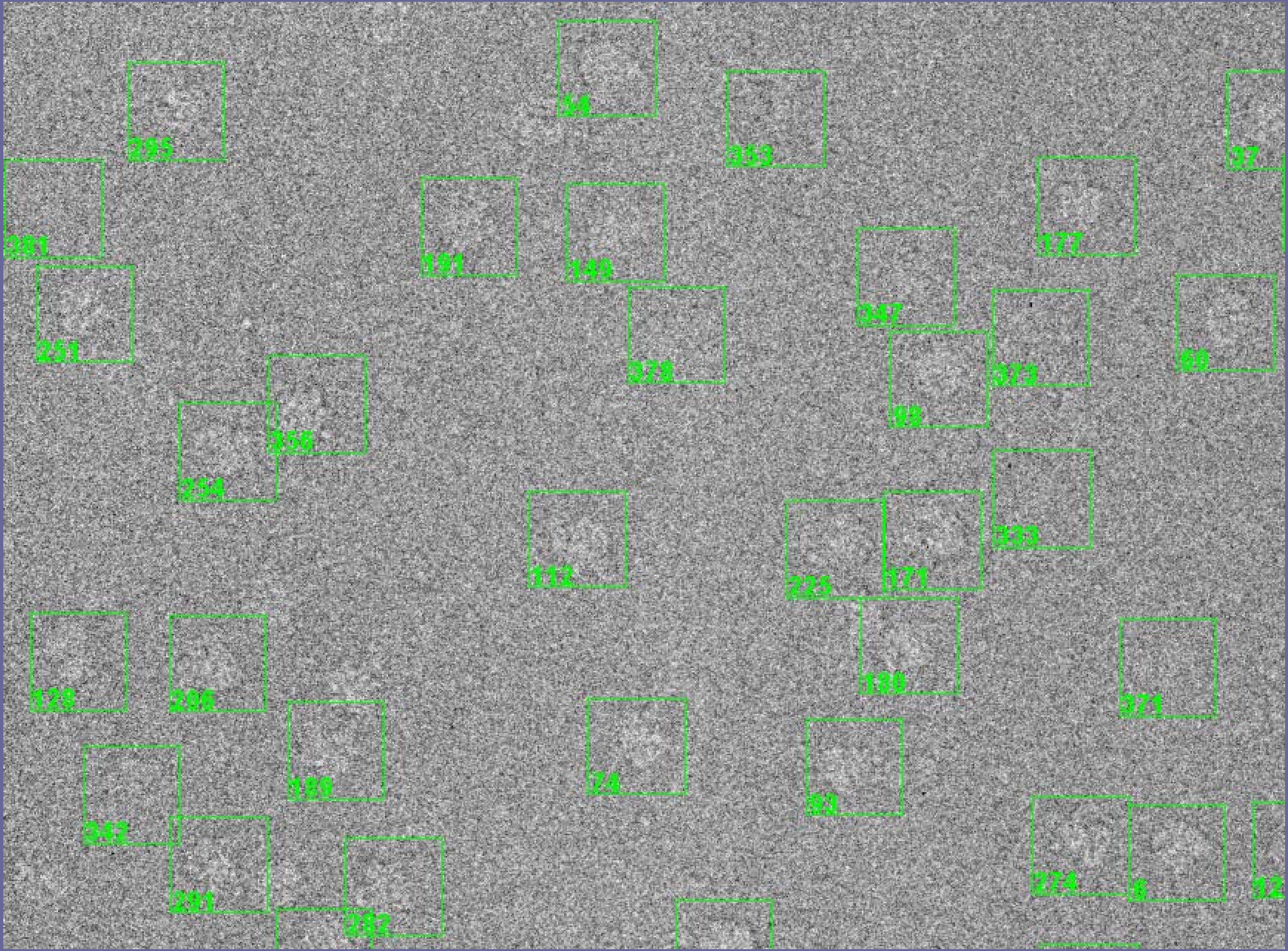
Automated particle picking, CCF-based, with local normalization

(i) Define a reference (e.g., by averaging projections over full Eulerian range);
(ii) Paste reference into array with size matching the size of the micrograph;
(iii) Compute CCF via FFT;
(iv) Compute locally varying variance of the micrograph via FFT (Roseman, 2003);

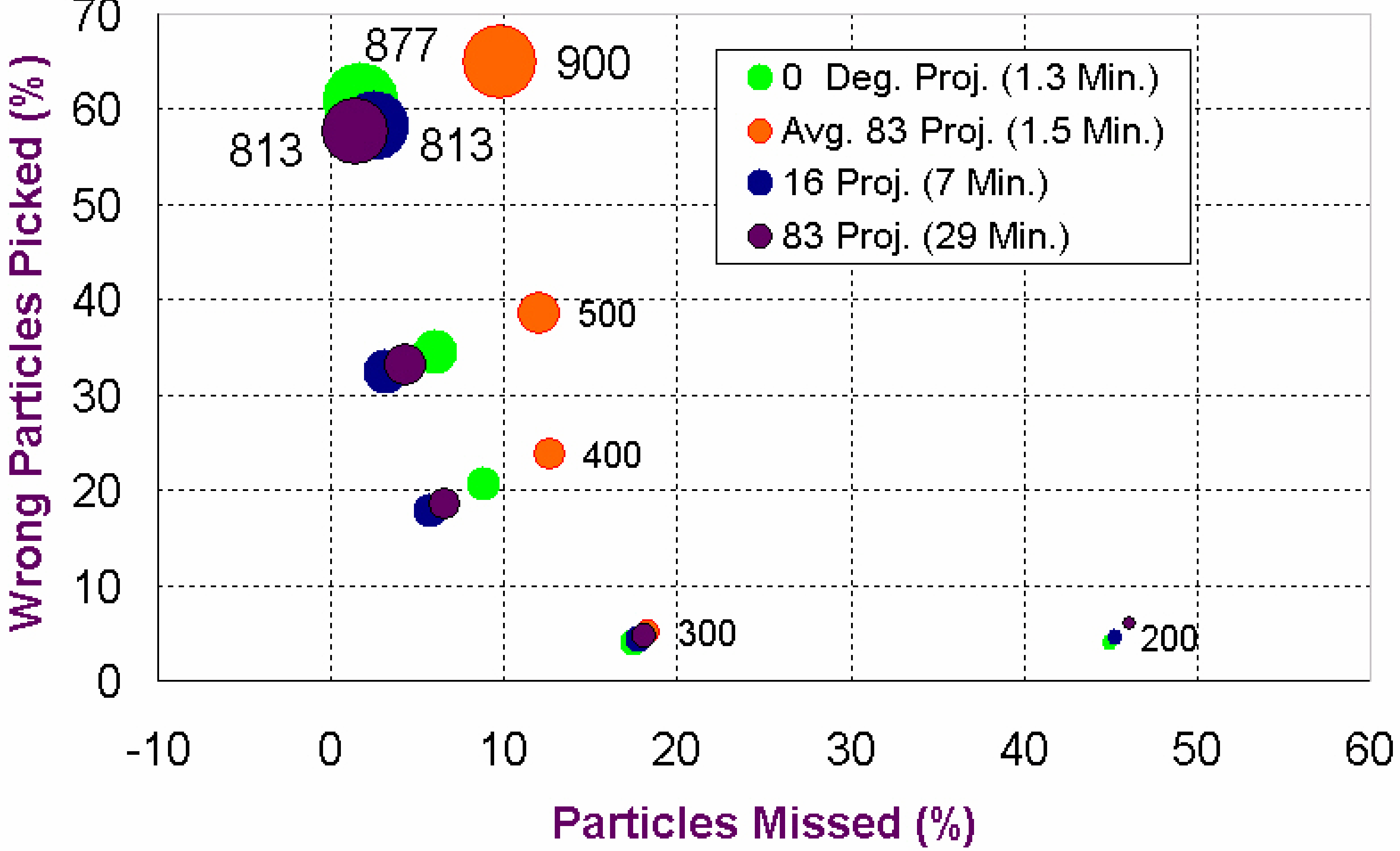
(v) ***"Local CCF" = CCF/local variance***

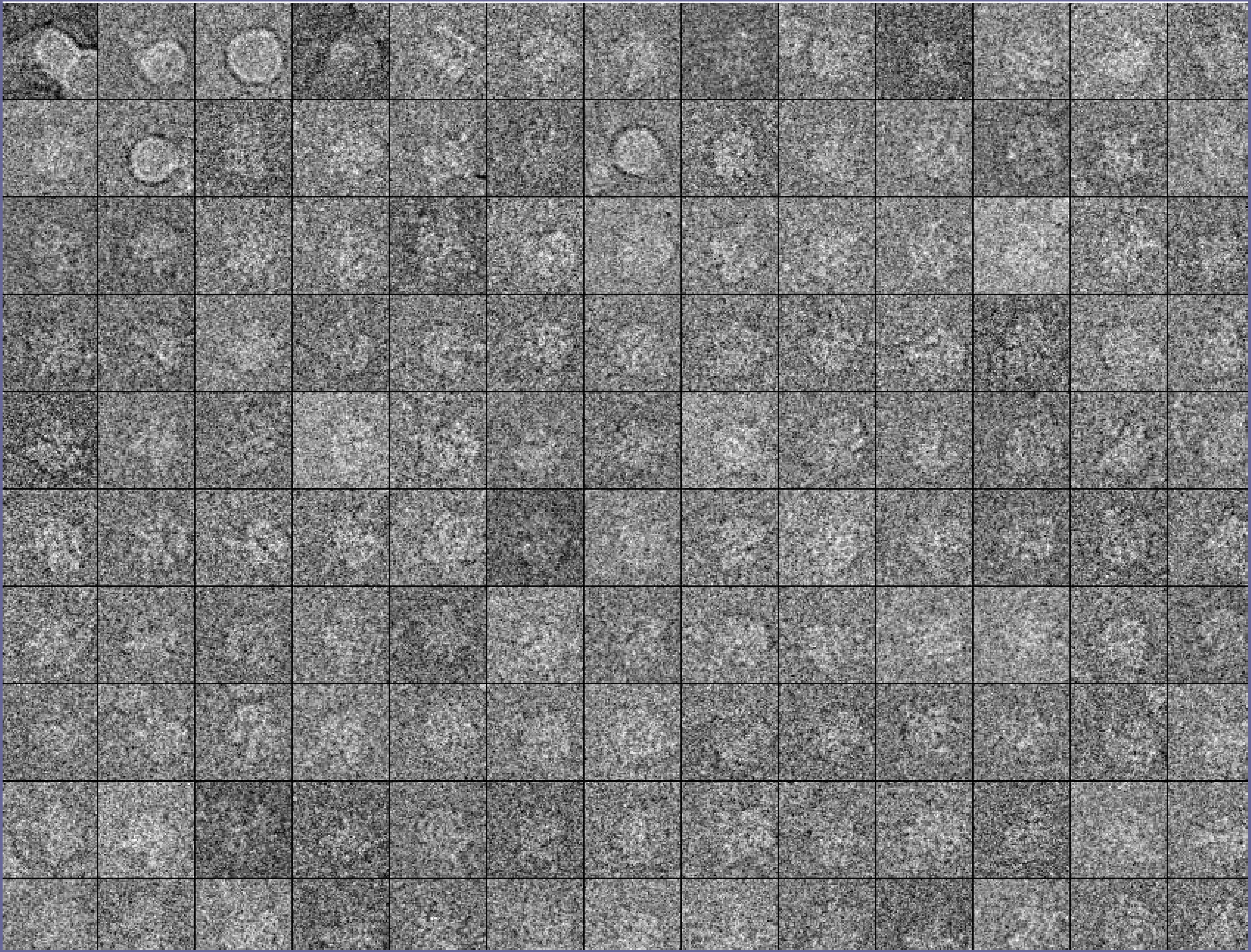
(vi) Peak search;
(vii) Window particles ranked by peak size;
(viii) Fast visual screening.

Advantage of local CCF: avoid problems from background variability


34
295
353
37
391
191
149
177
251
347
60
379
373
356
98
254
112
225
171
333
128
296
180
371
189
74
93
342
291
282
274
6
12

Particle Selection Efficiency Plot (Sym. Mask)





Overview: the necessary steps of a single-particle reconstruction -- I

1) Optical diffraction: quality control, defocus inventory of micrograph batch

2) Scanning of micrograph batch

3) Determine defoci, and define defocus groups

4) Pick particles

- (a) manual
- (b) automated

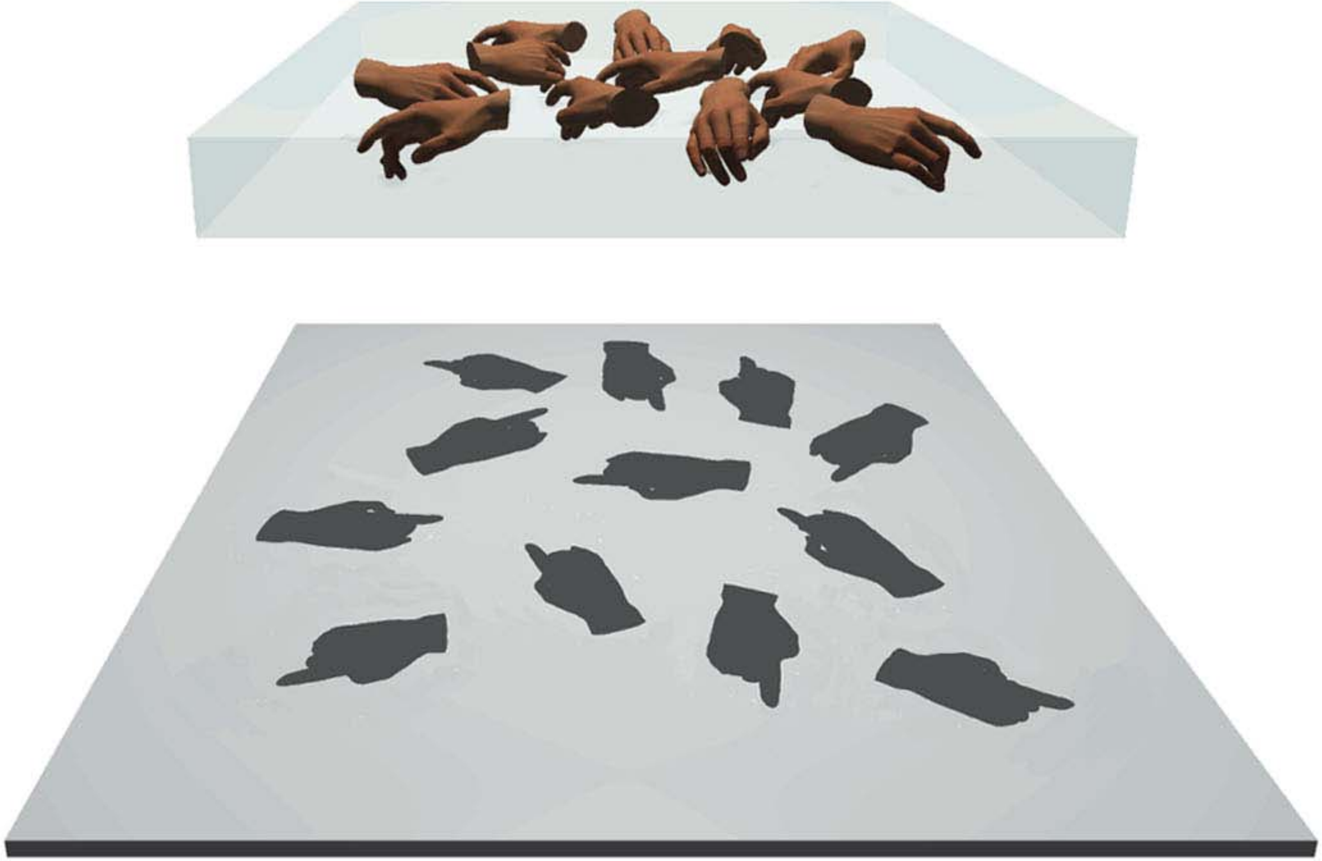
5) Determine particle orientation

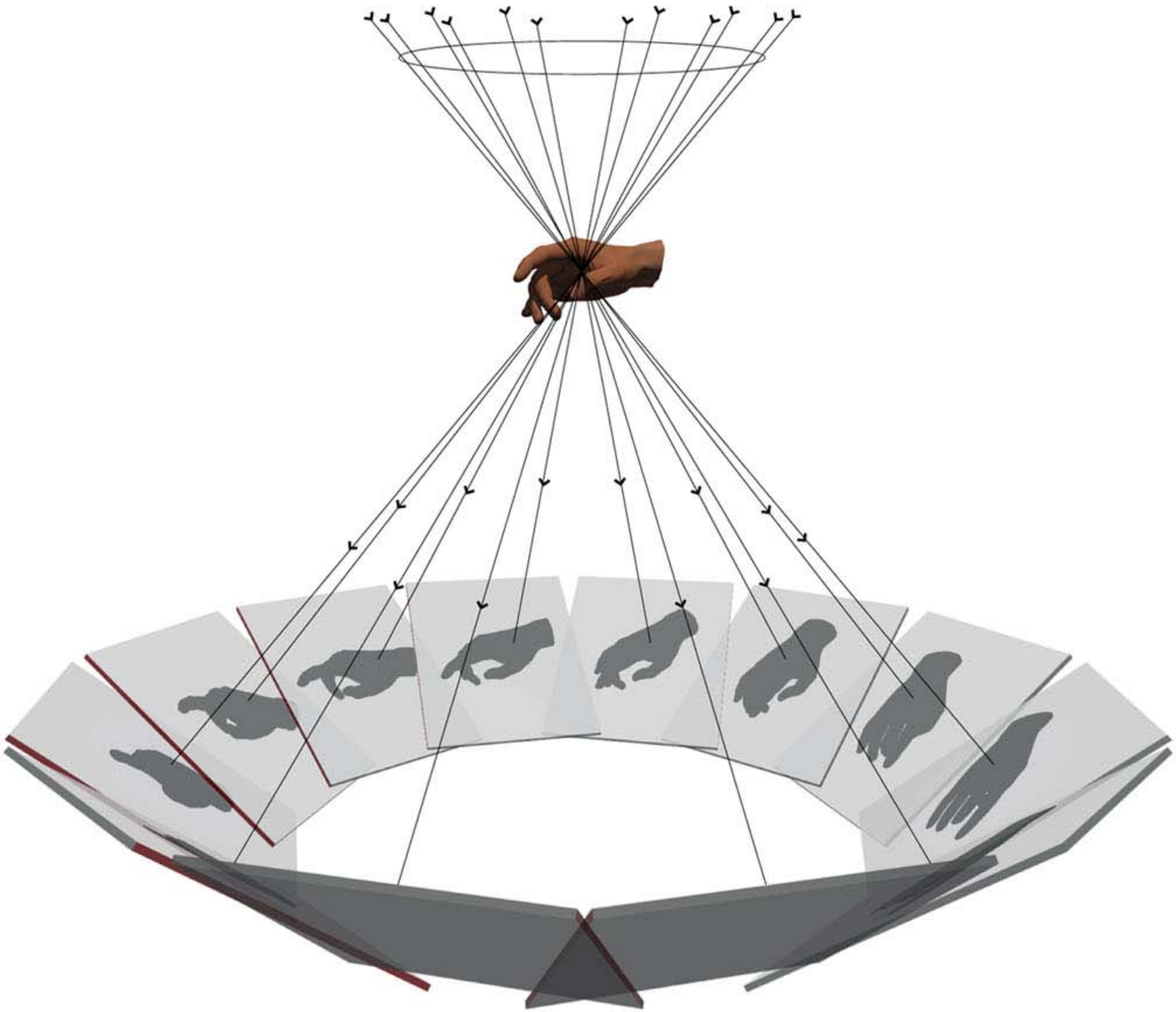
- (a) unknown structure -- bootstrap
  - (i) random-conical (uses unsupervised classification)
  - (ii) common lines/ angular reconstitution (uses unsupervised classification)
- (b) known structure
  - (i) reference-based (3D projection matching = supervised classification)
  - (ii) common lines/ angular reconstitution

Random-conical reconstruction

- Premise: all particle exhibit the same view
- Take same field first at theta ~50 degrees, then at 0 degrees [in this order, to minimize dose]
- Display both fields side by side
- Pick each particle in both fields
- Align particles from 0-degree field

  This yields azimuths, so that data can be put into the conical geometry
- Assign azimuths and theta to the tilted particles
- Proceed with 3D reconstruction

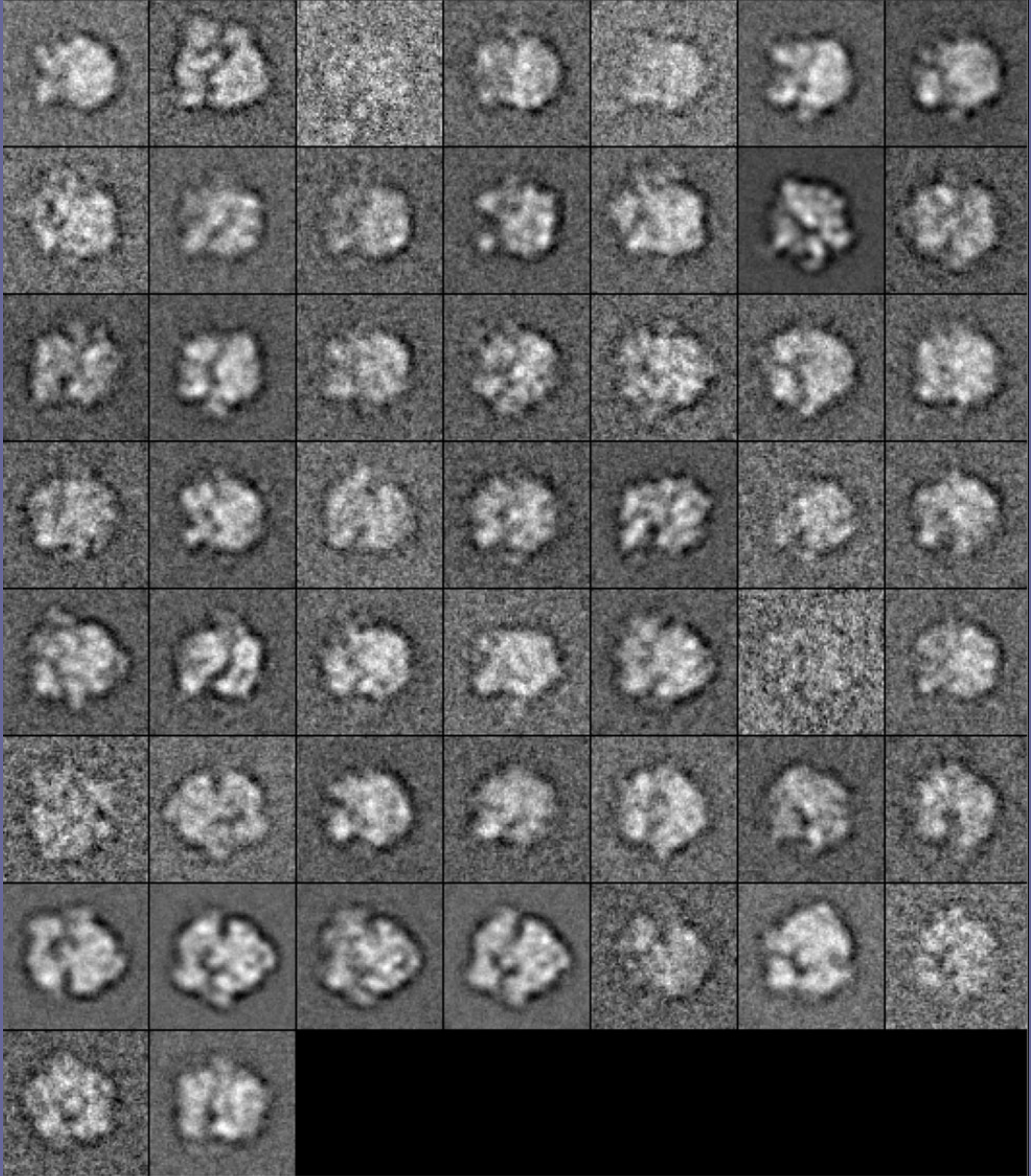




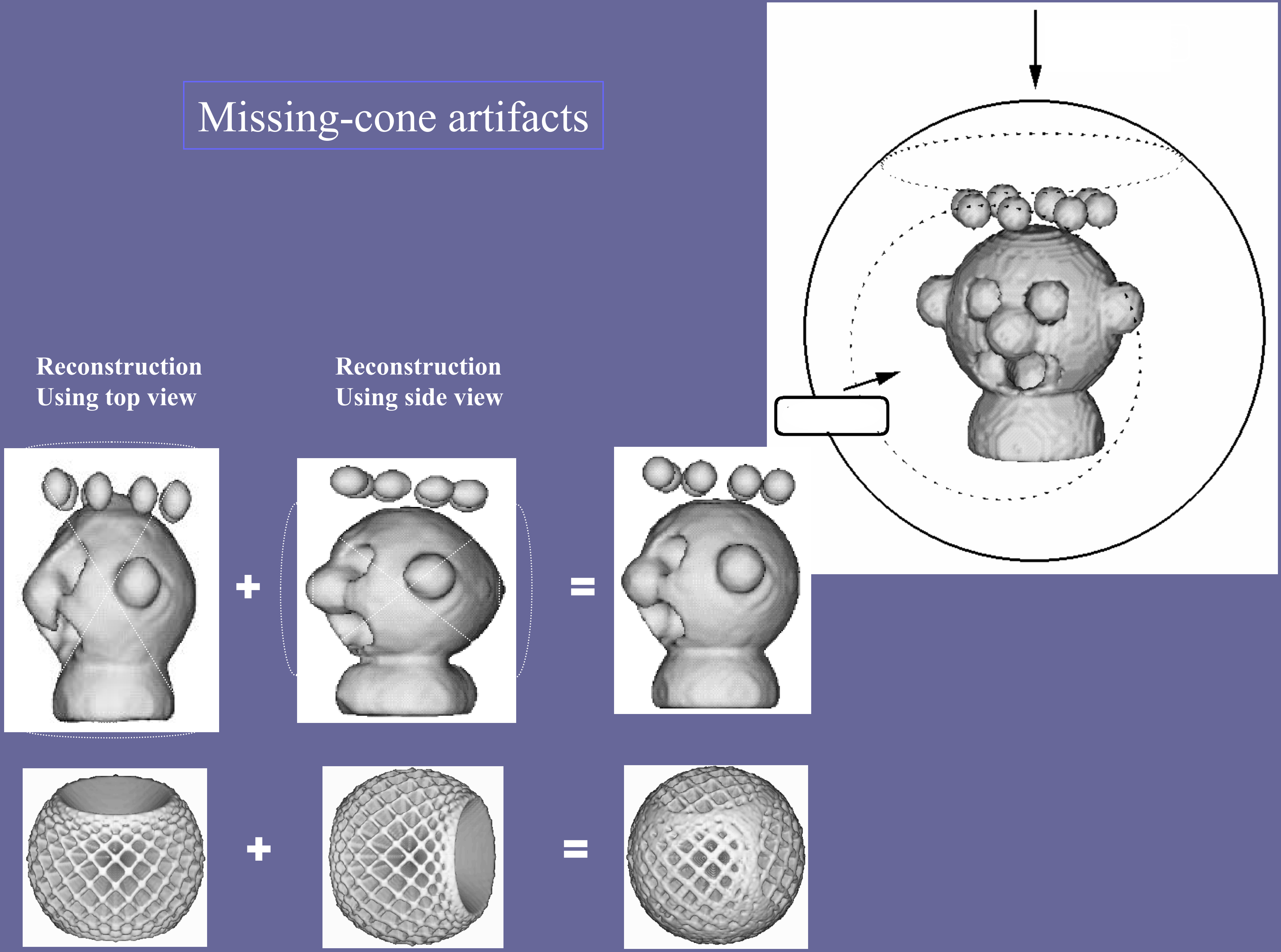
Random-conical reconstruction
Problems to be solved:

1) Find a subset (view class) of particles that lie in the same orientation on the grid
answer: unsupervised classification of 0-degree particles

2) Missing cone problem
answer: do several random conical reconstructions, each from a different subset (view class), find relative orientations, then make reconstruction from merged projections set.



Missing-cone artifacts



**Reconstruction
Using top view**

**Reconstruction
Using side view**

Overview: the necessary steps of a single-particle reconstruction -- I

1) Optical diffraction: quality control, defocus inventory of micrograph batch

2) Scanning of micrograph batch

3) Determine defoci, and define defocus groups

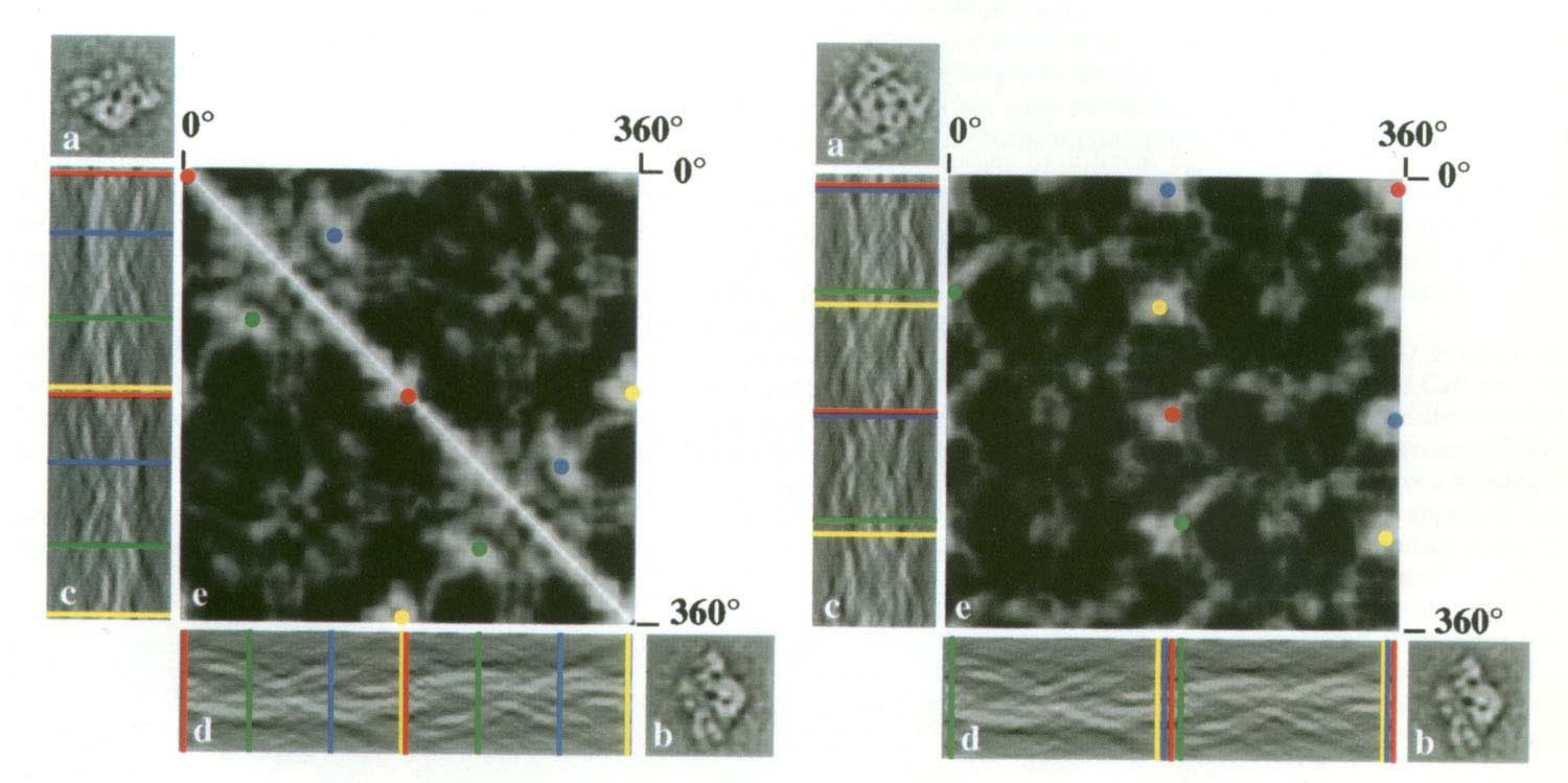
4) Pick particles

- (a) manual
- (b) automated

5) Determine particle orientation

- (a) unknown structure -- bootstrap
  - (i) random-conical (uses unsupervised classification)
  - (ii) common lines/ angular reconstitution (uses unsupervised classification)
- (b) known structure
  - (i) reference-based (3D projection matching = supervised classification)
  - (ii) common lines/ angular reconstitution

Determination of relative orientations by common lines



Serysheva et al. (1995) Nature Struct. Biol. 2: 18-24.

Common lines/ angular reconstitution

1) Unsupervised classification, to determine classes of particles exhibiting the same view

2) Average images in each class → class averages

3) Determine common lines between class averages

stepwise (van Heel, 1967)

simultaneously (Penczek et al., 1996)

Issues:

- unaveraged images are too noisy
- resolution loss due to implicit use of view range
- handedness not defined – tilt or prior knowledge needed

Overview: the necessary steps of a single-particle reconstruction -- I

1) Optical diffraction: quality control, defocus inventory of micrograph batch

2) Scanning of micrograph batch

3) Determine defoci, and define defocus groups

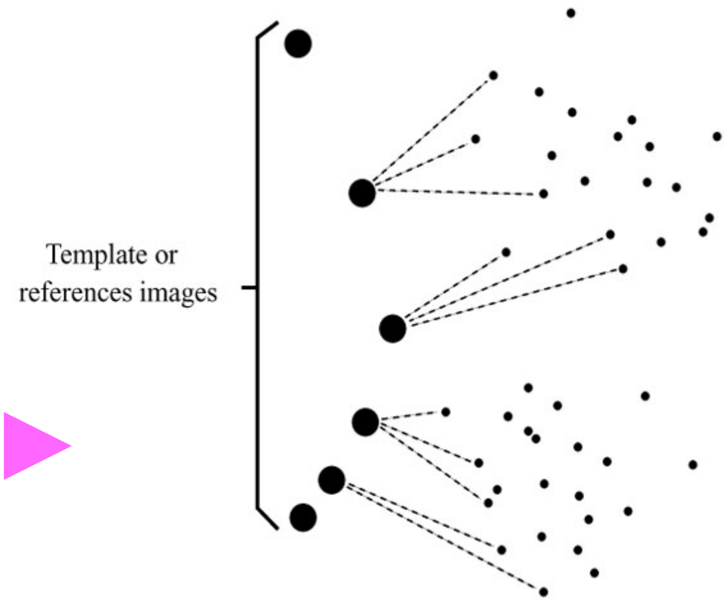
4) Pick particles

- (a) manual
- (b) automated

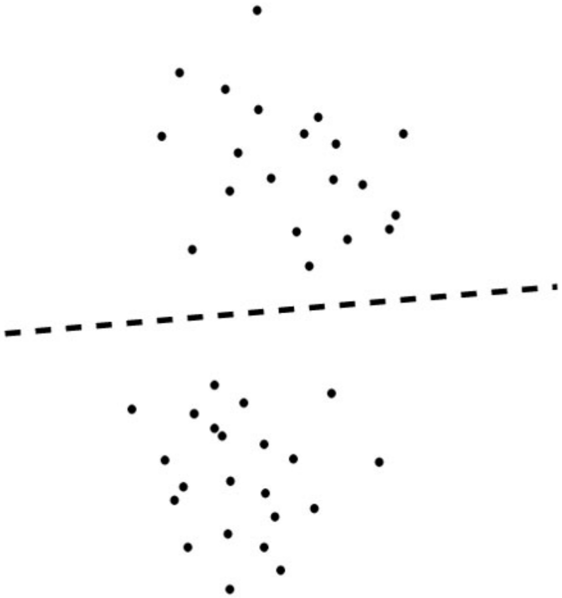
5) Determine particle orientation

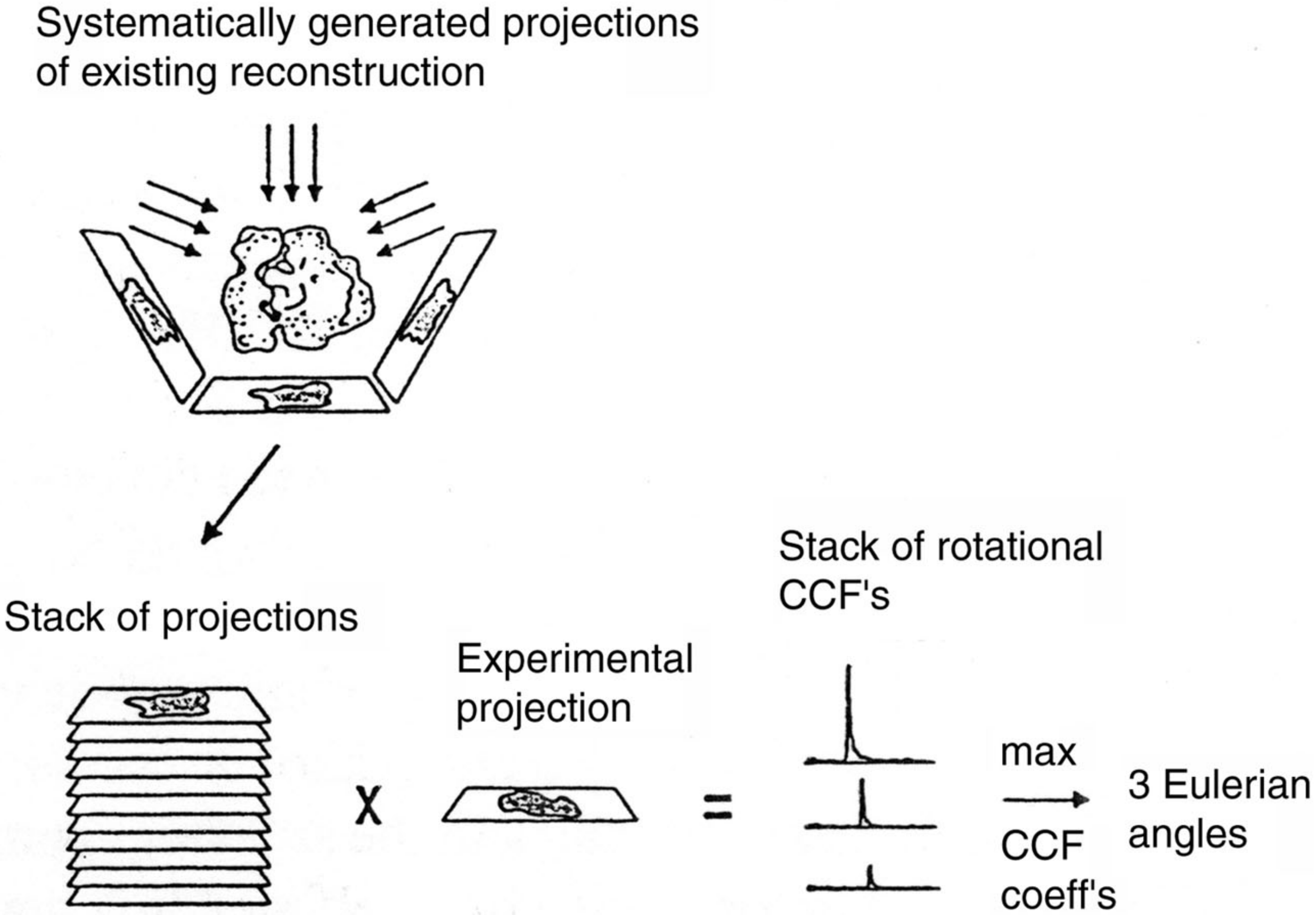
- (a) unknown structure -- bootstrap
  - (i) random-conical (uses unsupervised classification)
  - (ii) common lines/ angular reconstitution (uses unsupervised classification)
- (b) known structure
  - (i) reference-based (3D projection matching = supervised classification)
  - (ii) common lines/ angular reconstitution

Supervised Classification



Unsupervised Classification





Systematically generated projections of existing reconstruction

Stack of projections

Experimental projection

Stack of rotational CCF's

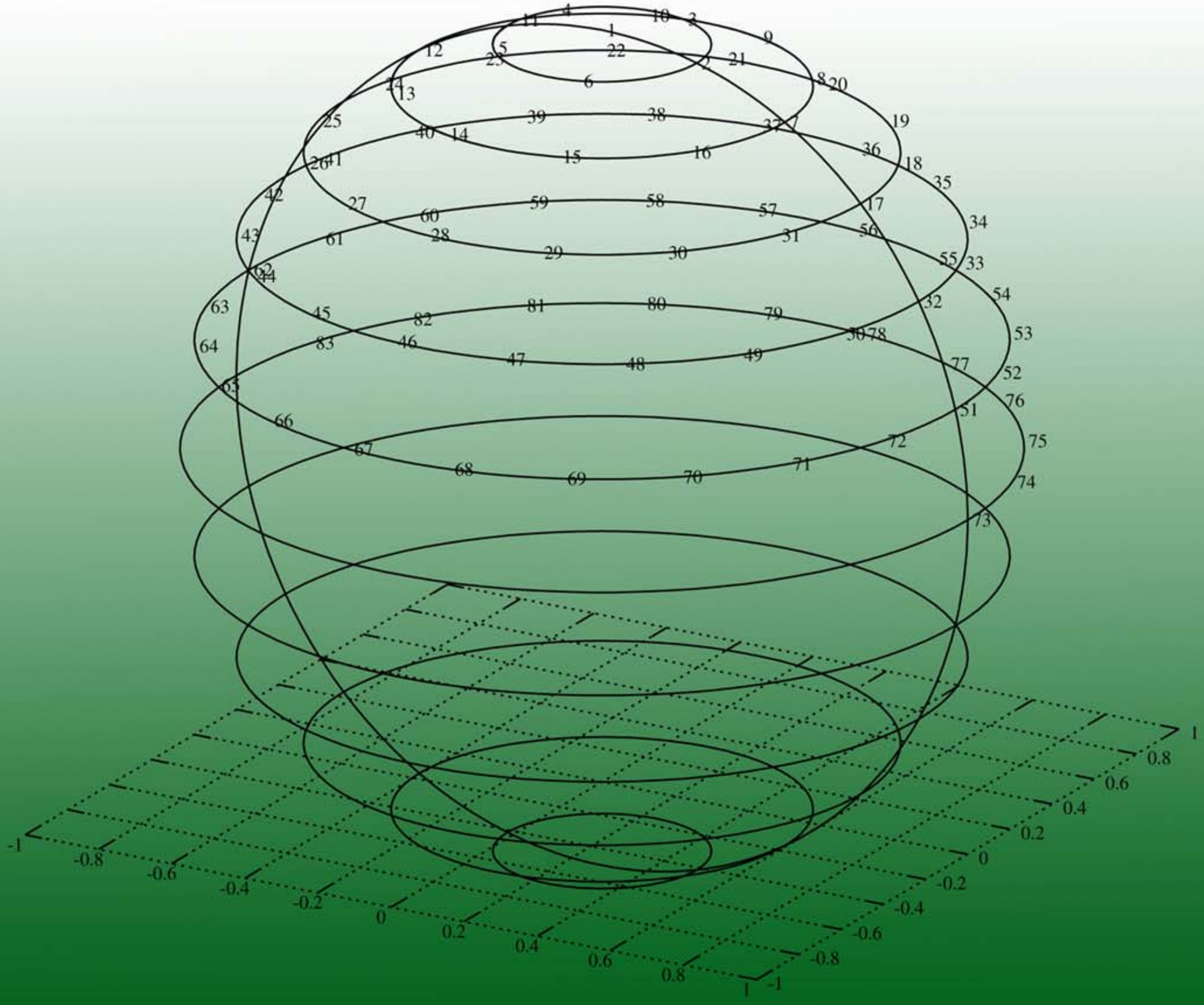
X

=

max

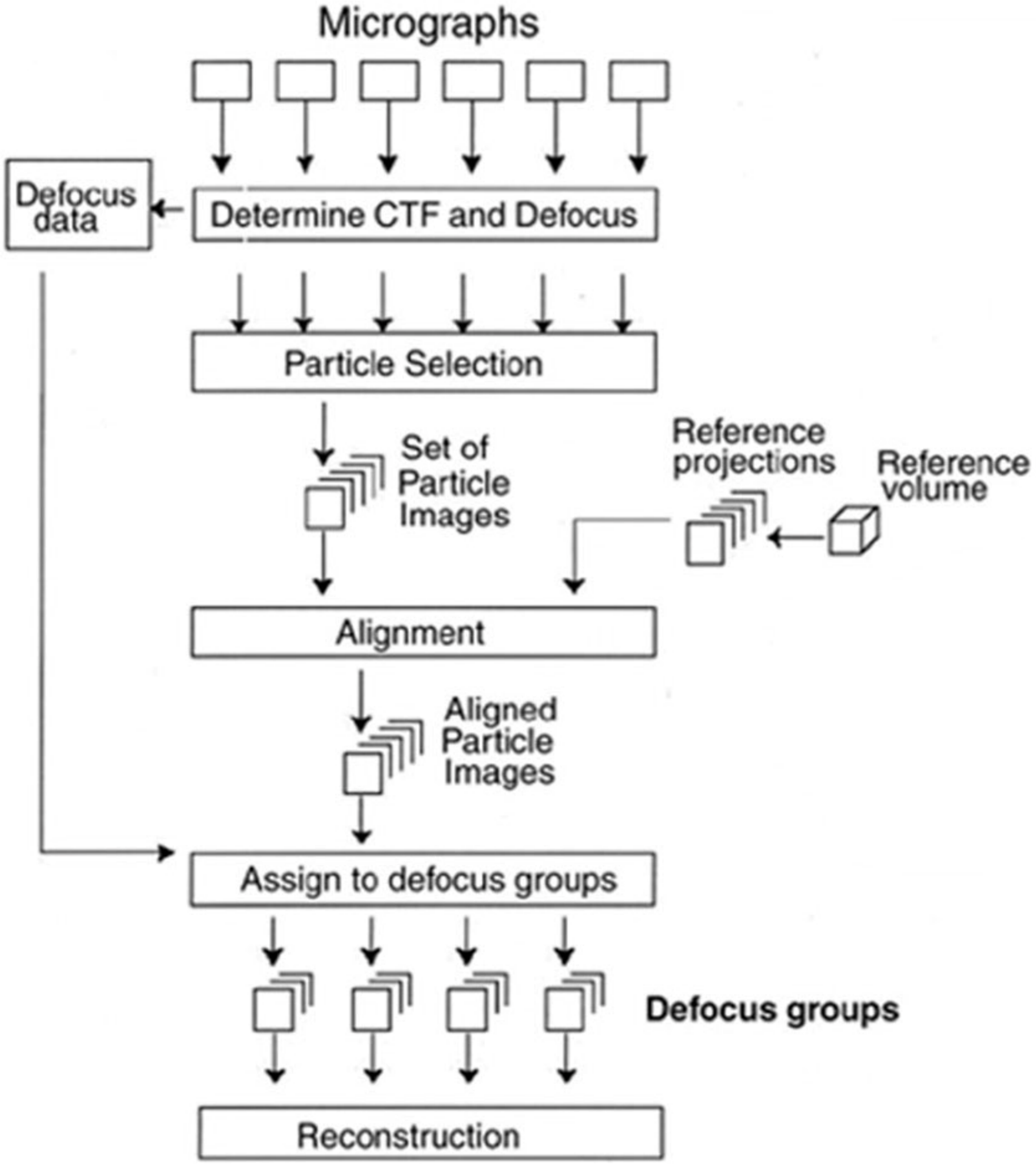
CCF coeff's

3 Eulerian angles



4
11
10
3
1
9
12
5
23
22
2
21
6
8
20
24
13
25
39
38
37
7
19
40
14
26
41
15
16
36
18
35
42
27
60
59
58
57
17
34
43
61
28
29
30
31
56
55
33
62
44
63
45
81
80
79
32
54
82
64
83
46
47
48
49
50
78
53
77
52
65
76
66
51
67
72
75
68
69
70
71
74
73
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

Reference-based Reconstruction



Overview: the necessary steps of a single-particle reconstruction -- II

6) ***3D reconstruction by defocus group***
  (a) Fourier interpolation
  (b) Weighted back-projection
  (c) Iterative algebraic reconstruction
  (d) Conjugate gradient

---Steve----Steve----Steve----Steve----Steve--

7) ***Refinement***
- given an initial 3D reference,
  iterate the steps {3D projection matching + reconstruction}
- beware of problem of reference-dependence

8) ***CTF correction***
9) ***Validation***
10) ***Interpretation:*** segmentation, docking, etc.

3D reconstruction by defocus group

(a) Fourier interpolation

(b) Weighted back-projection

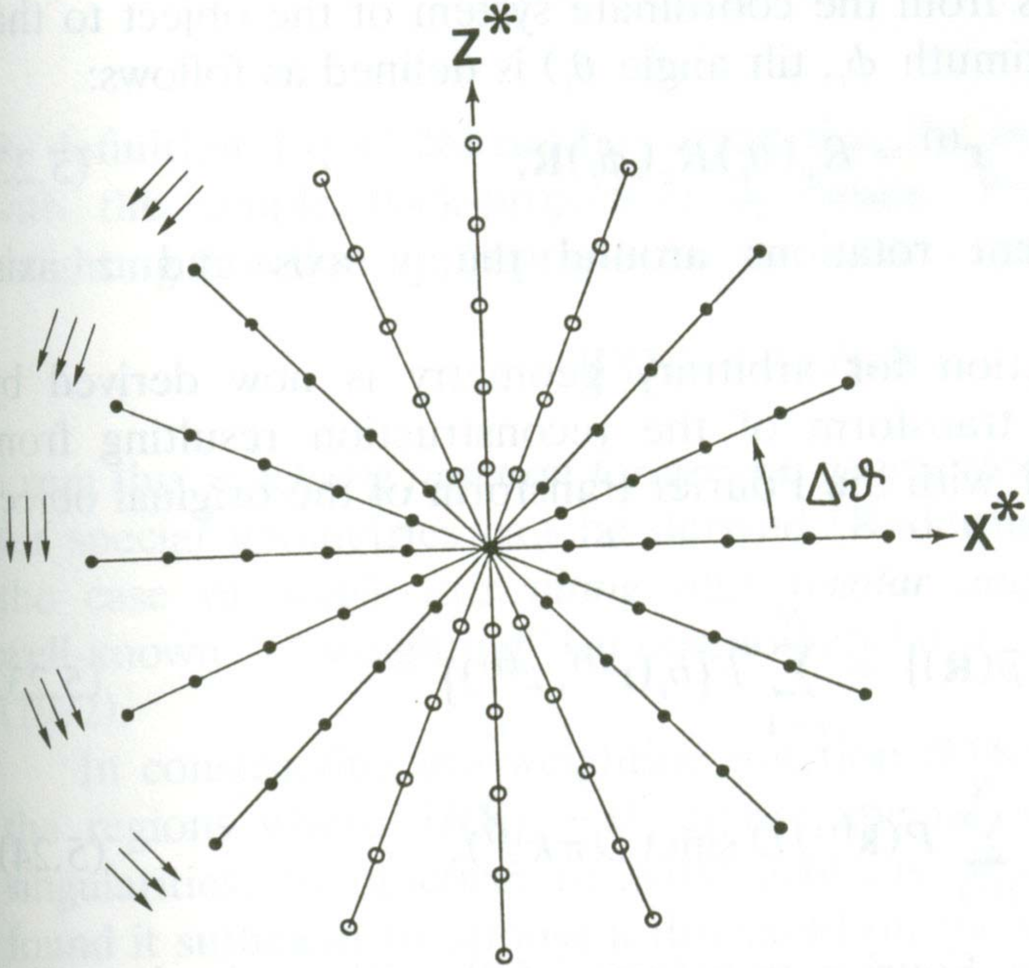
(c) Iterative algebraic reconstruction

(d) Conjugate gradient

Obtain samples on a regular Cartesian grid in 3D Fourier space by interpolation between Fourier values on oblique 2D grids (central sections) running through the origin, each grid corresponding to a projection.

Speed (high) versus accuracy (low). Can be used in the beginning phases of a reconstruction project. However, see new development by Pawel Penczek!

Sample points of adjacent projections are increasingly sparse as we go to higher resolution



3D reconstruction by defocus group

(a) Fourier interpolation

(b) Weighted back-projection

(c) Iterative algebraic reconstruction

(d) Conjugate gradient

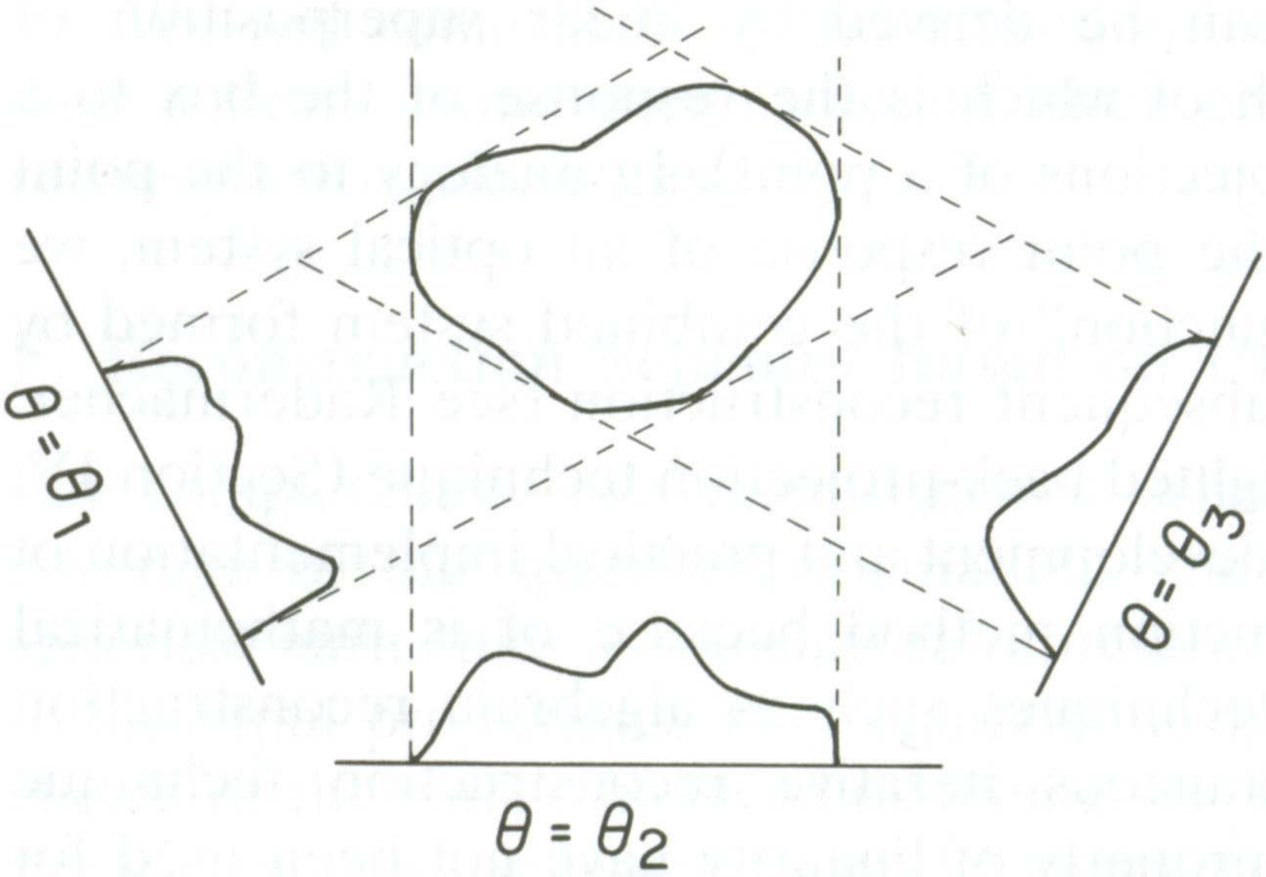
(1) Simple back-projection: Sum over “back-projection bodies”, each obtained by “smearing out” a projection in the viewing direction.

(2) Weighted back-projection: as (1), but “weight” the projections first by multiplying their Fourier transforms with |K| (R* weighting, in X-ray terminology), then inversing the Fourier transform.

(3) For general geometries, the weighting function is more complicated, and has to be computed every time.

- Weighted back-projection is fast, but does not yield the “smoothest” results. It may show strong artifacts from angular gaps.

Principle of back-projection



3D reconstruction by defocus group

(a) Fourier interpolation

(b) Weighted back-projection

(c) Iterative algebraic reconstruction

(d) Conjugate gradient

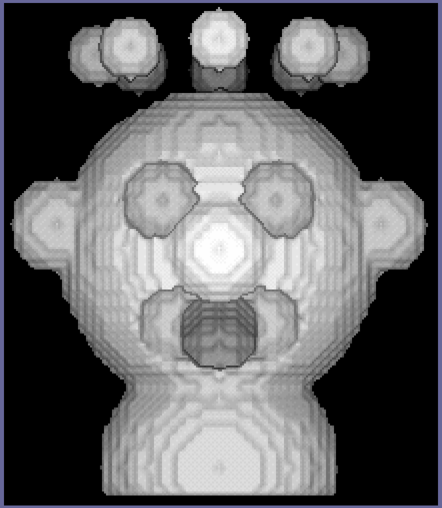
The discrete algebraic projection equation is satisfied, one angle at a time, by adjusting the densities of a starting volume. As iterations proceed, each round produces a better approximation of the object.

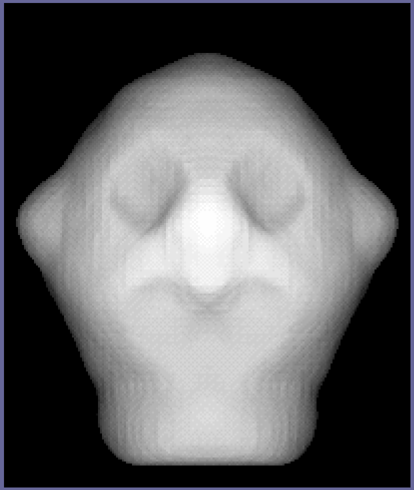
The algorithm comes in many variants. It allows constraints to be easily implemented.

It produces a very smooth reconstruction, and is less affected by angular gaps.

Comparison of some reconstruction algorithms

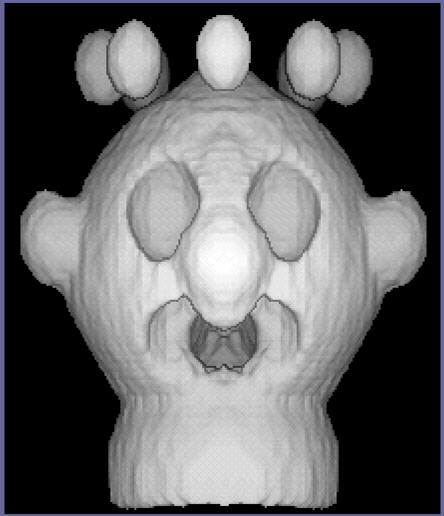
Original object

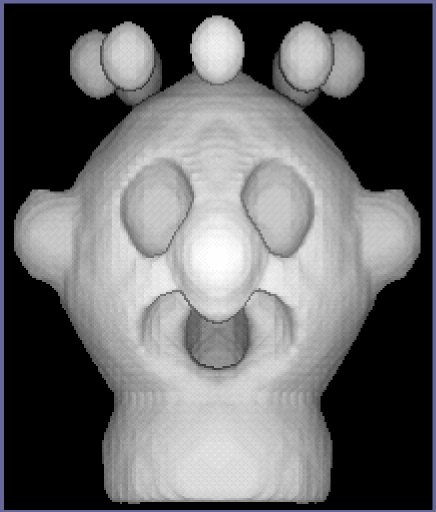




Simple back-projection

Weighted back-projection





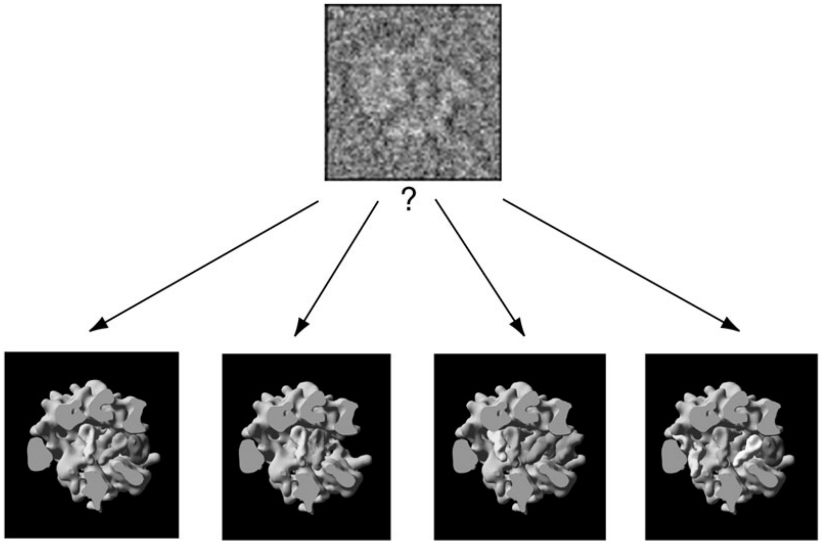
Iterative algebraic reconstruction

Sources for limited resolution

1) Instrumental: partial coherence (envelope function)
2) Particles with different height all considered having same defocus (envelope function)
3) Numerical: interpolations, inaccuracies
4) Failure to exhaust existing information
5) Conformational diversity

Conformational diversity: heterogeneous particle population

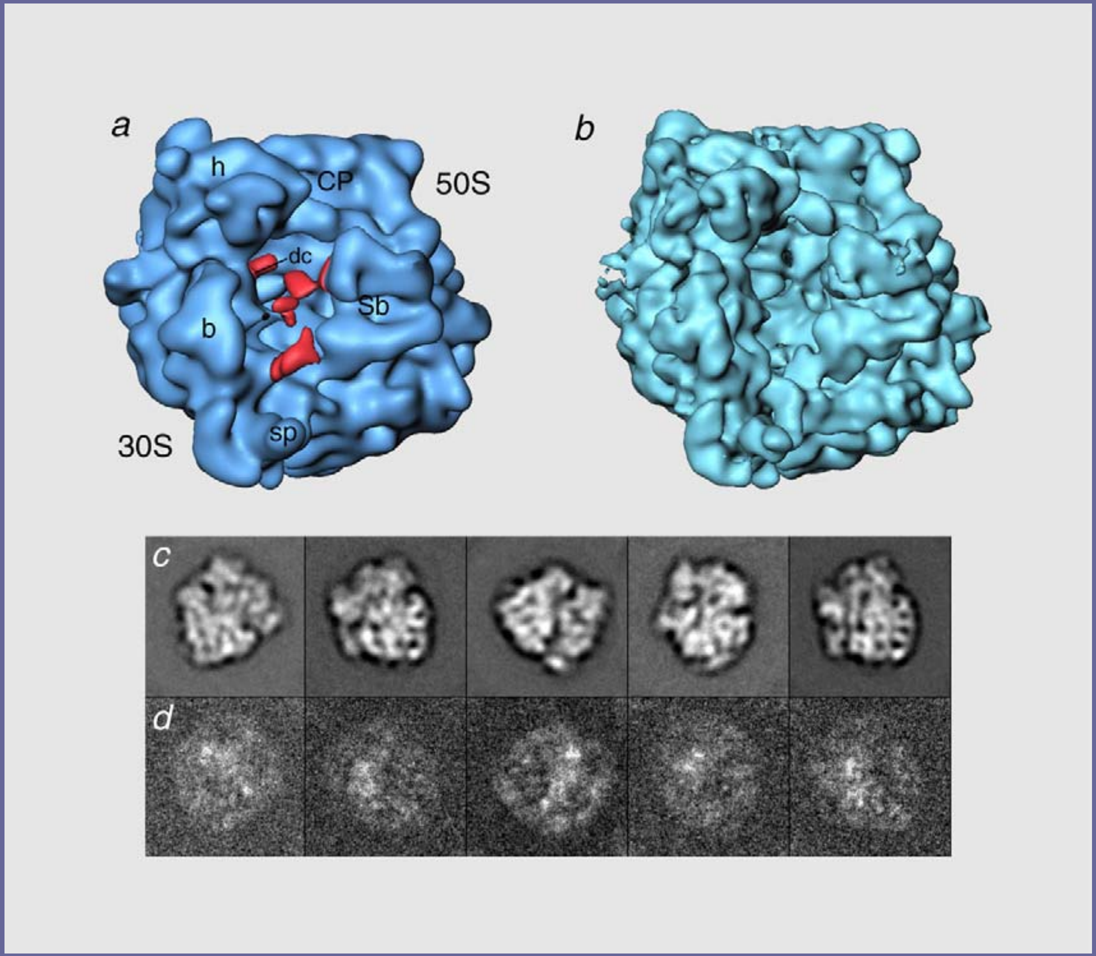
Current approach: assume all conformers are "similar". Treat problem in first approximation as a problem with a single conformer. Then try different models as references to see if population segregates.



Example: low occupancy of ternary complex

reconstruction using all data

empty ribosome (control)



Problem solved by supervised classification