THE UNIVERSITY *of* TEXAS

SCHOOL OF HEALTH INFORMATION SCIENCES AT HOUSTON

Single Particle Reconstruction Techniques

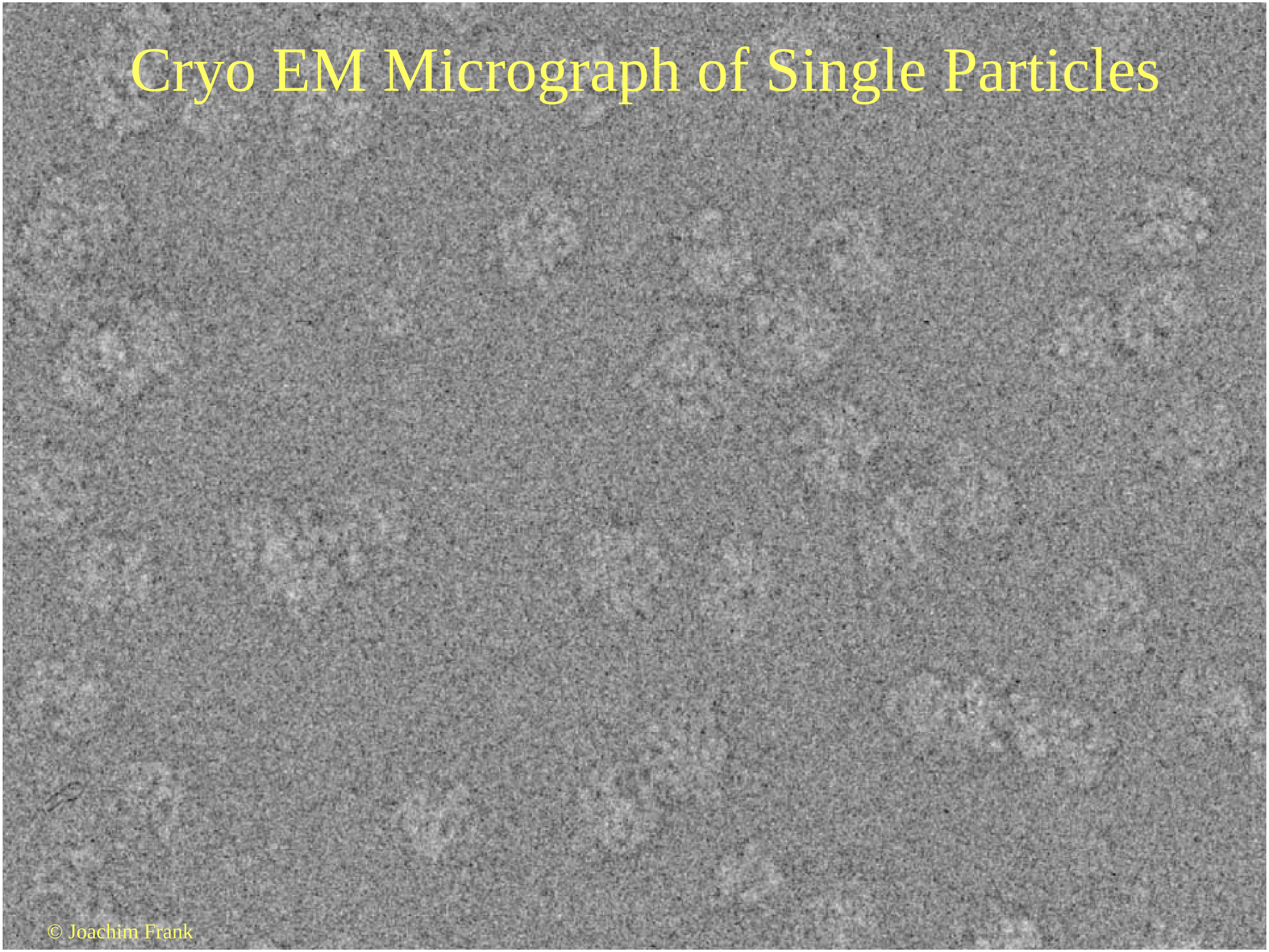
For students of HI 6001-125

"Computational Structural Biology"

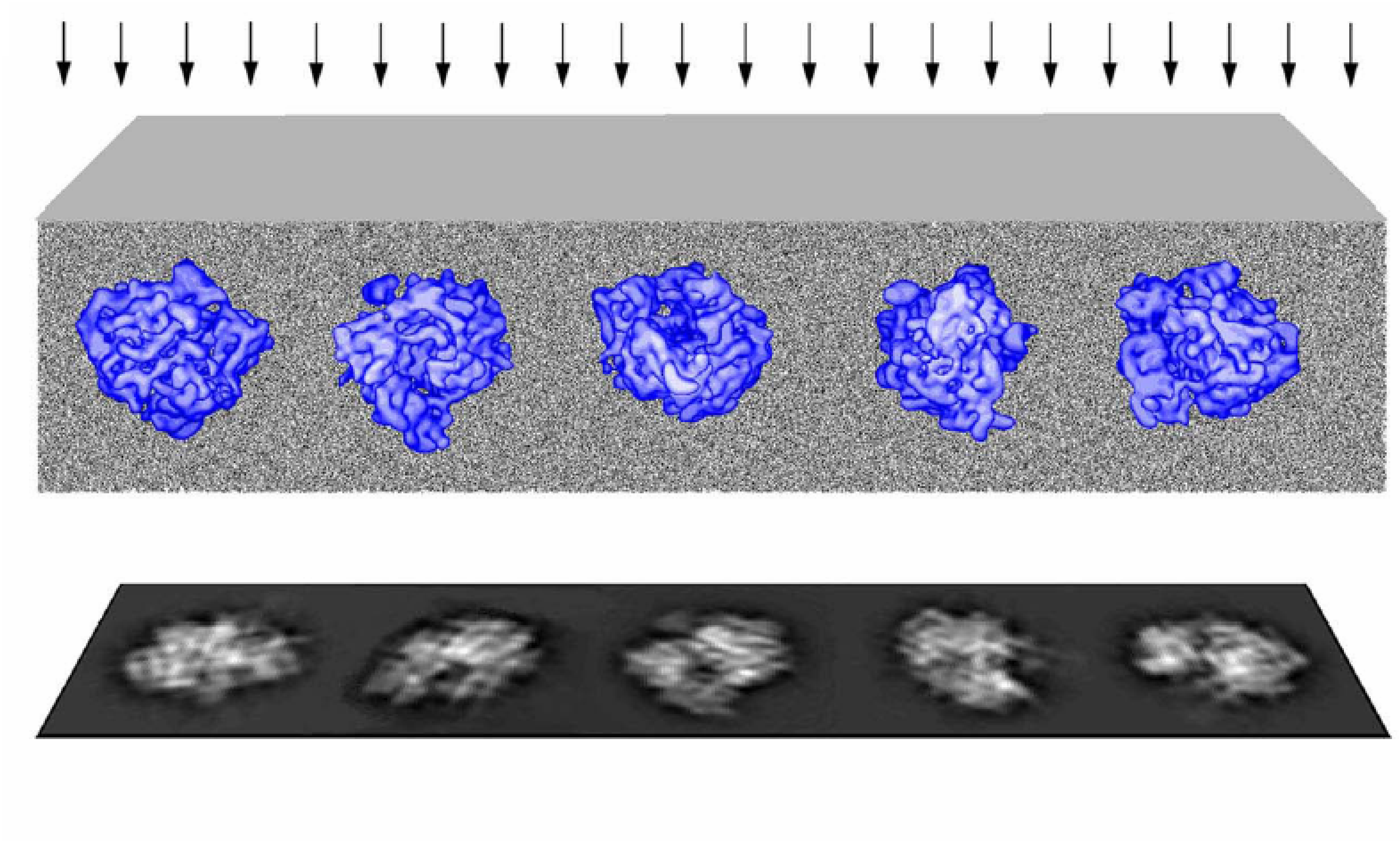
Willy Wriggers, Ph.D.
School of Health Information Sciences

http://biomachina.org/courses/structures/09.html

Cryo EM Micrograph of Single Particles



What is Observed



Main Assumptions

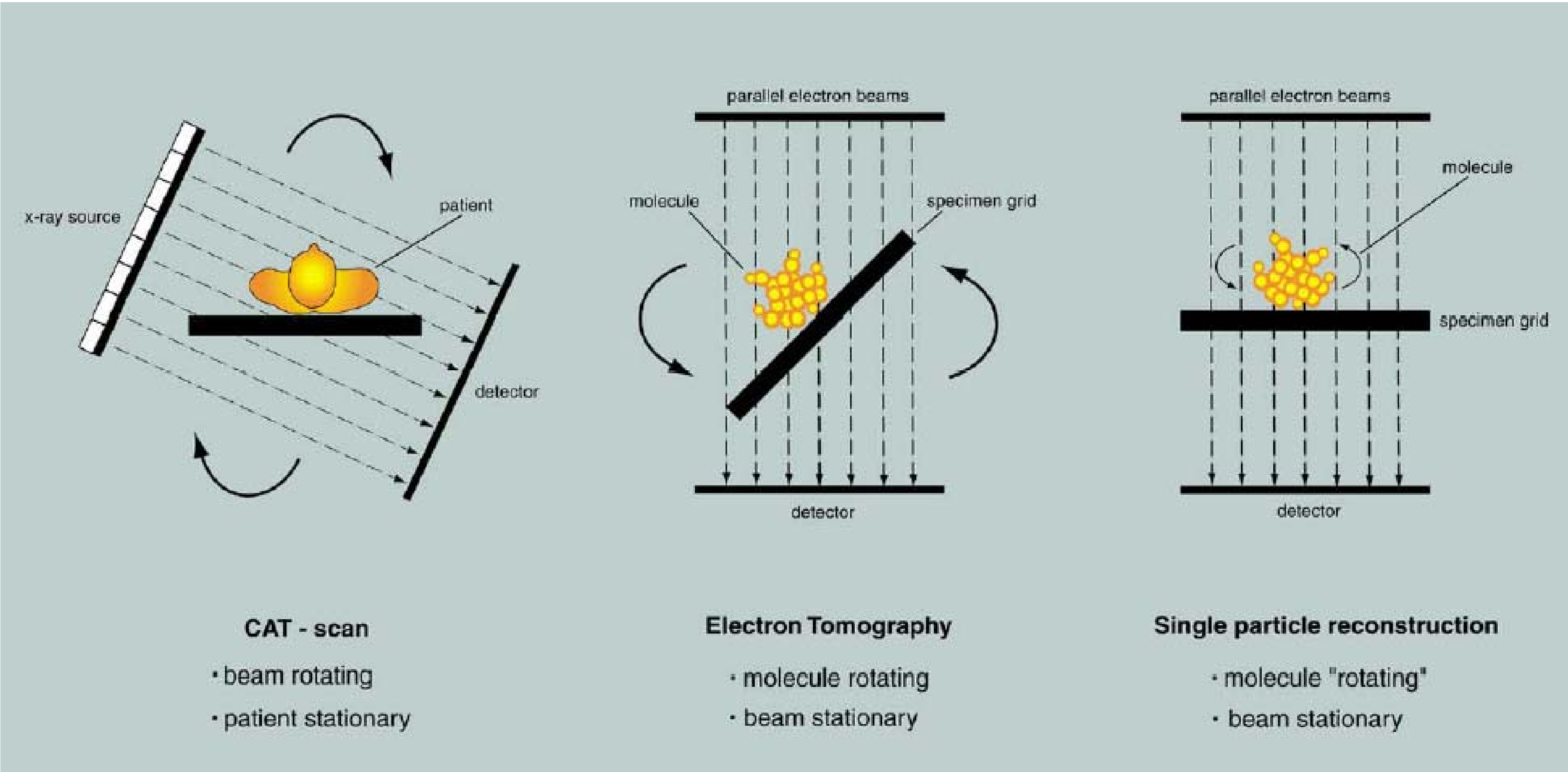
1) All particles in the specimen have identical structure

2) All are linked by 3D rigid body transformations (rotations, translations)

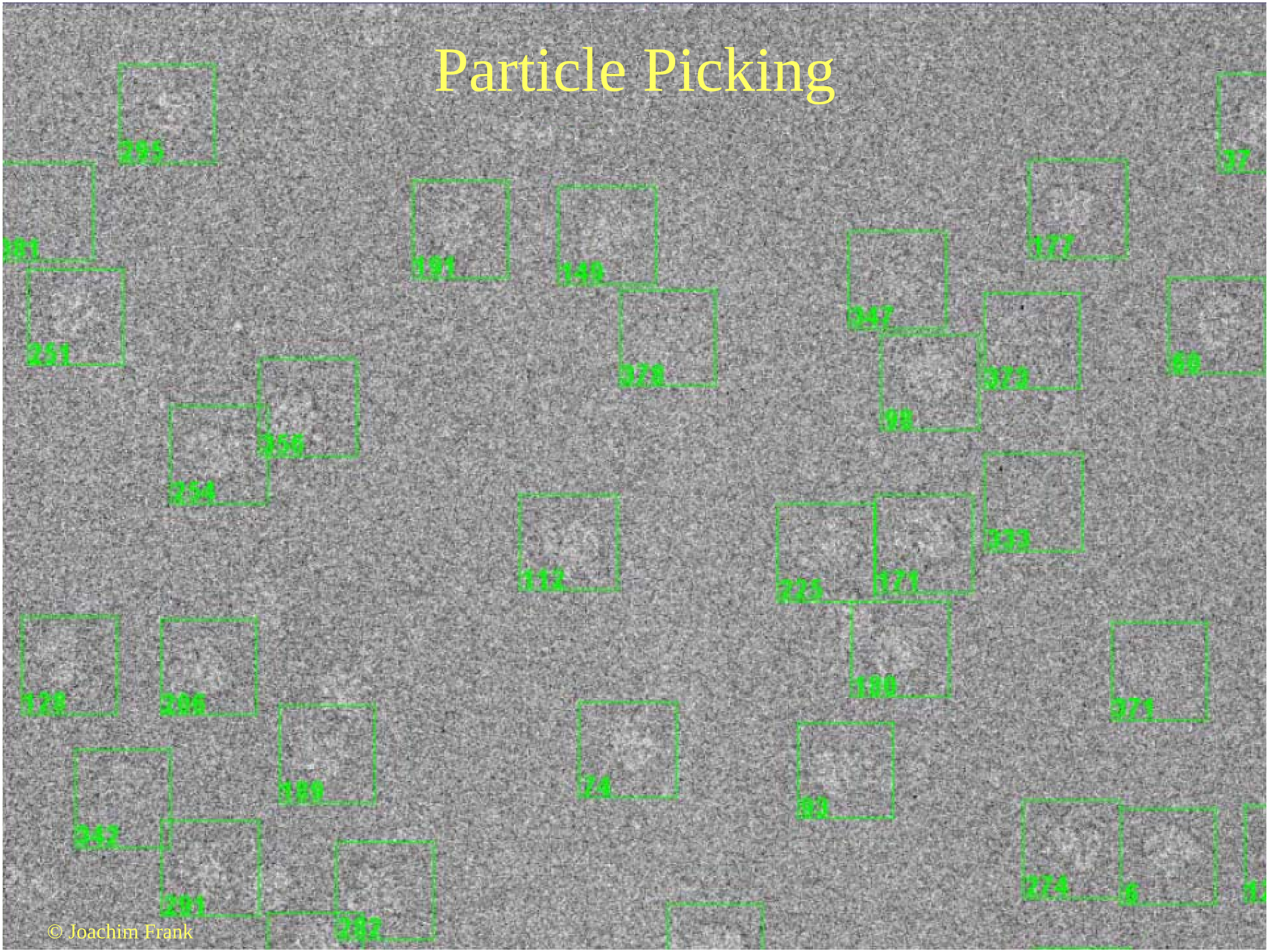
3) Particle images are interpreted as a “signal” part (= the projection of the common structure) plus “noise”

Important requirement:
even angular coverage, without major gaps.

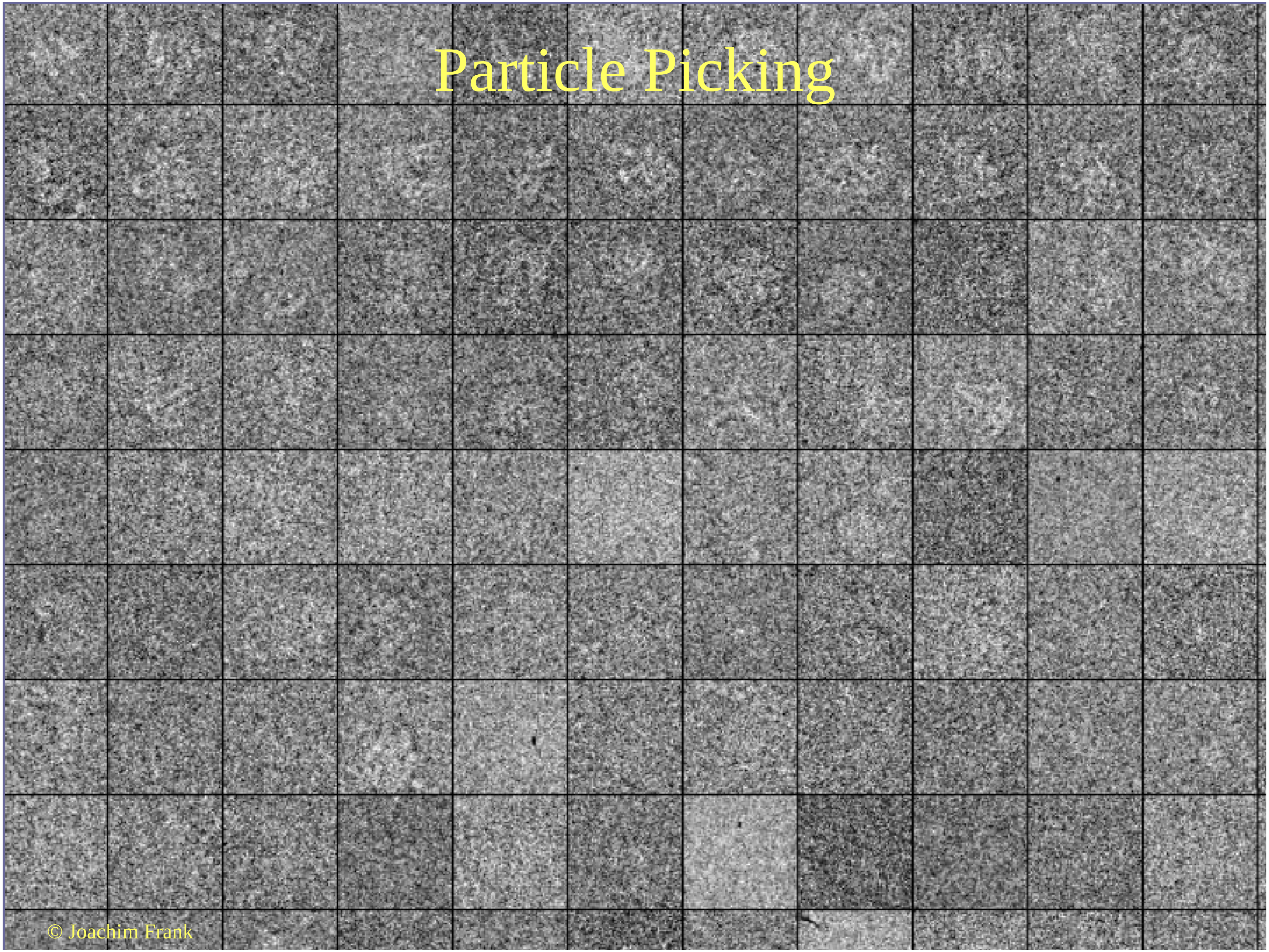
How to Get Even Angular Coverage



Particle Picking



Particle Picking

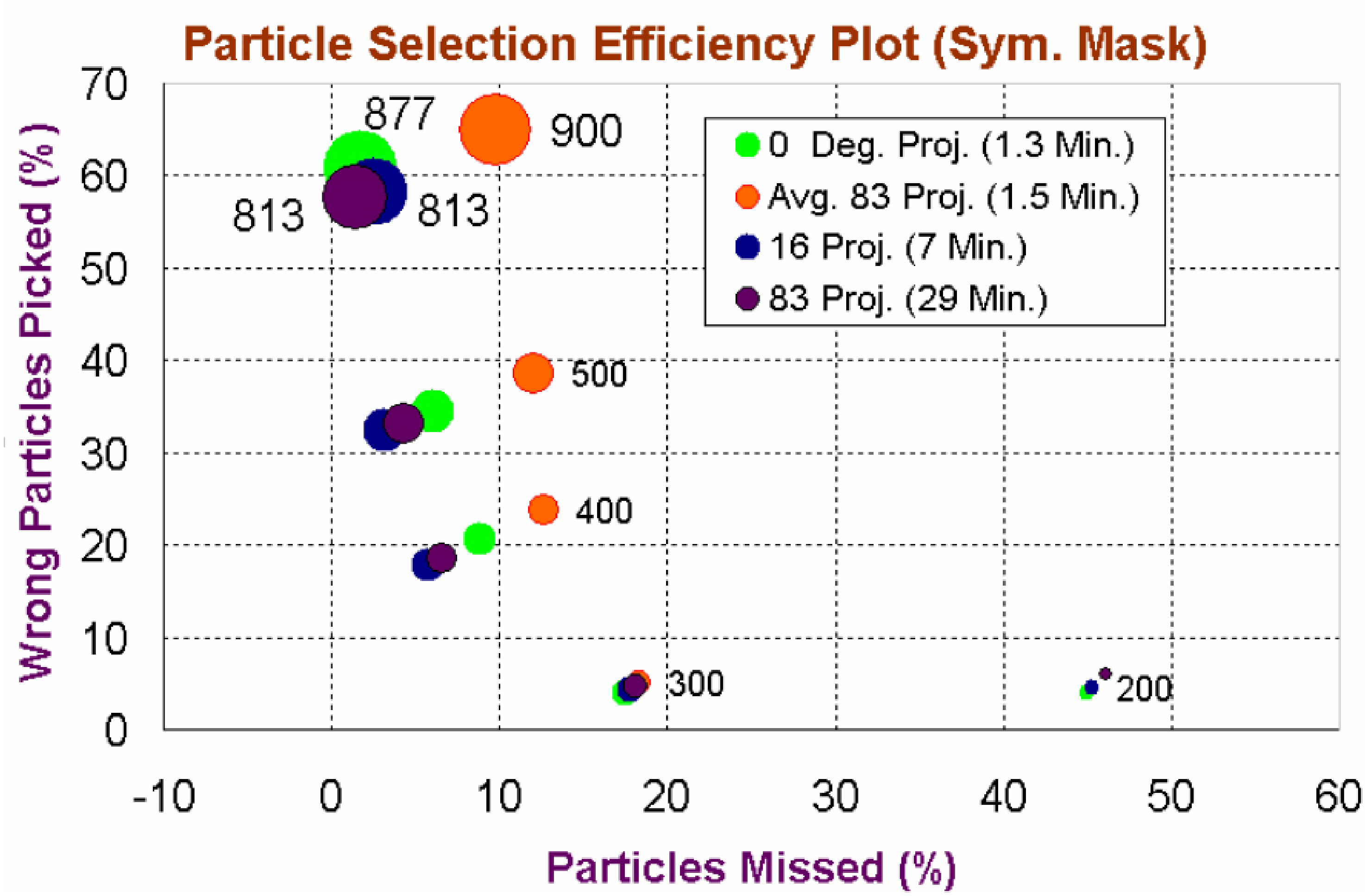


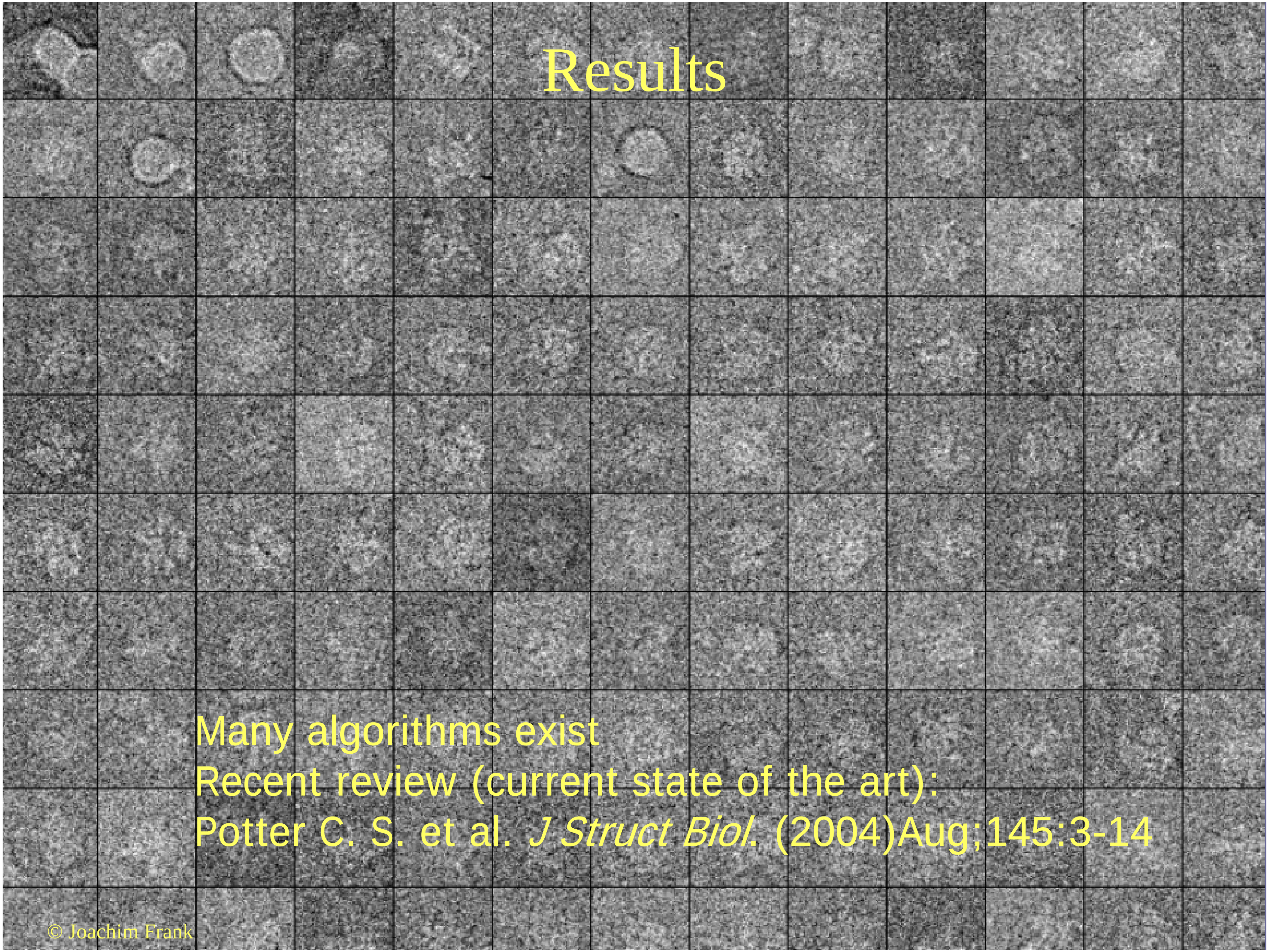
Automated Particle Picking

Example: 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

Performance





Results

Many algorithms exist

Recent review (current state of the art):

Potter C. S. et al. *J Struct Biol.* (2004)Aug;145:3-14

Classification of Images

• EM images of protein are very noisy and, therefore, the primary process of single-particle analysis is the classification of images according to their Euler angles, the images in each classified group then being averaged to reduce the noise level (Frank et al., 1978; van Heel and Frank, 1981).

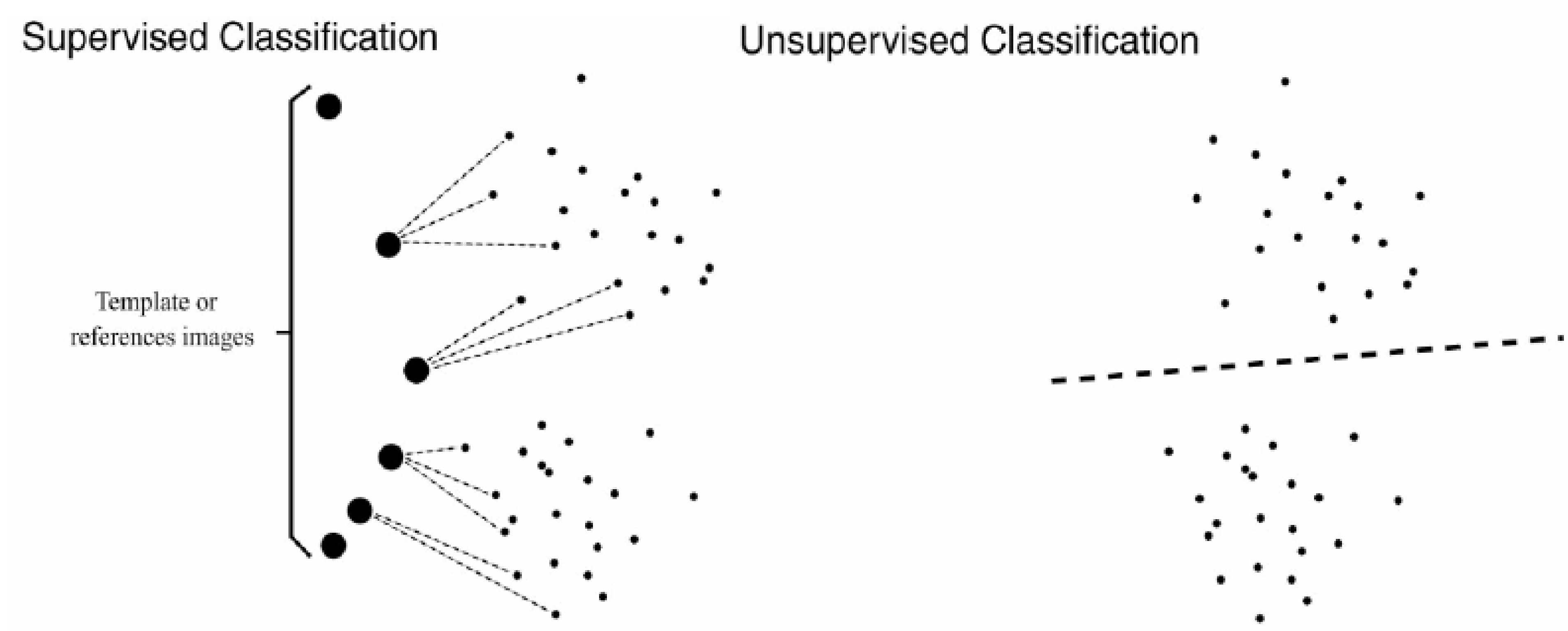
• Classification methods are divided into those that are “supervised” and those that are “unsupervised”:

• 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

Note: explained in Frank book, chapter 3.

Classification

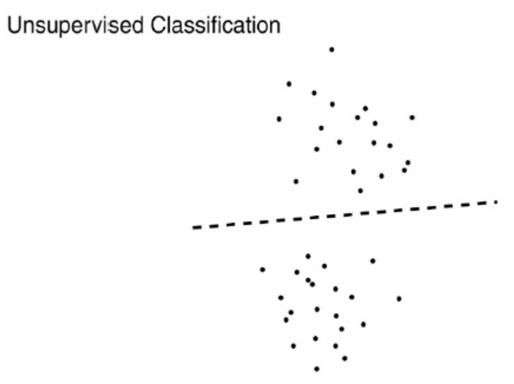


Unsupervised Classification (Clustering)

• 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.



Dimensonality Reduction

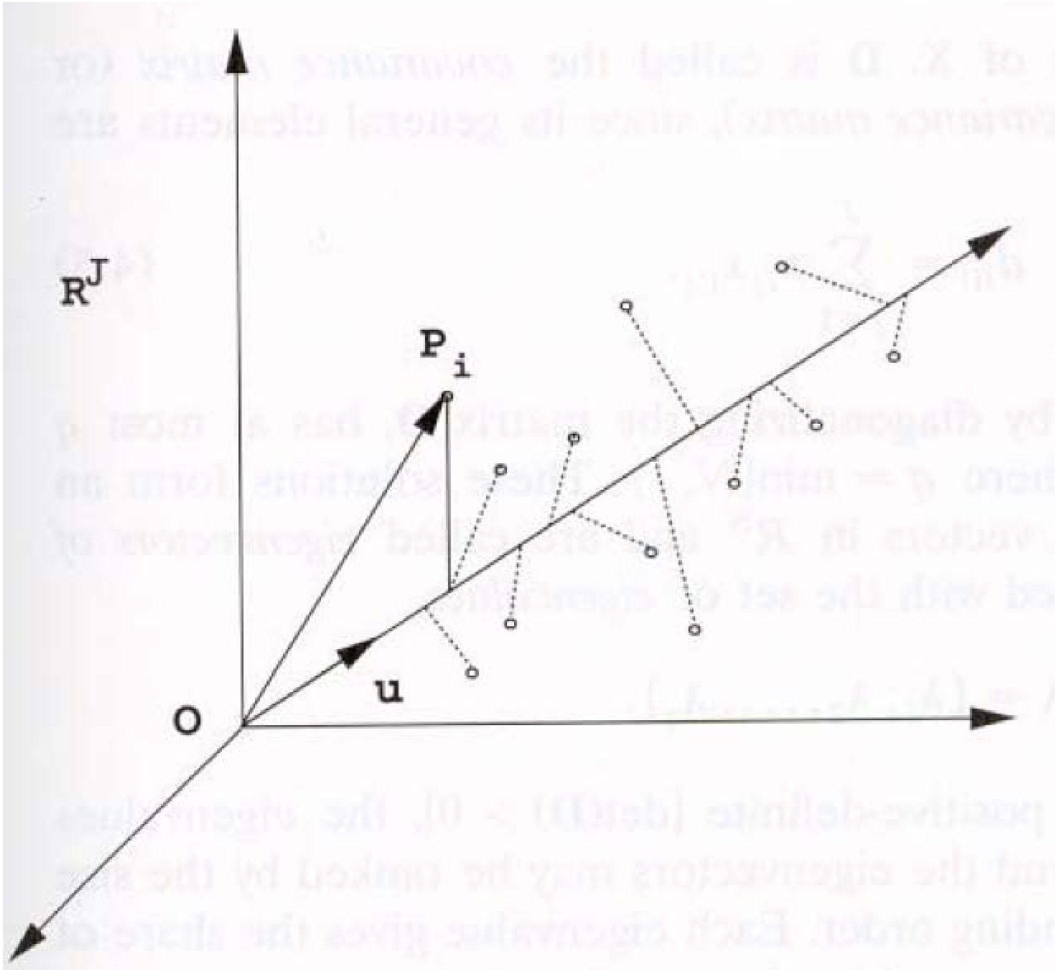
• Role of Multivariate Statistical Analysis (MSA): find a space ("factor space") with reduced dimensionality for the representation of the "objects". This greatly simplifies classification.

• 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).

• MSA very similar to Principal Component Analysis (PCA).

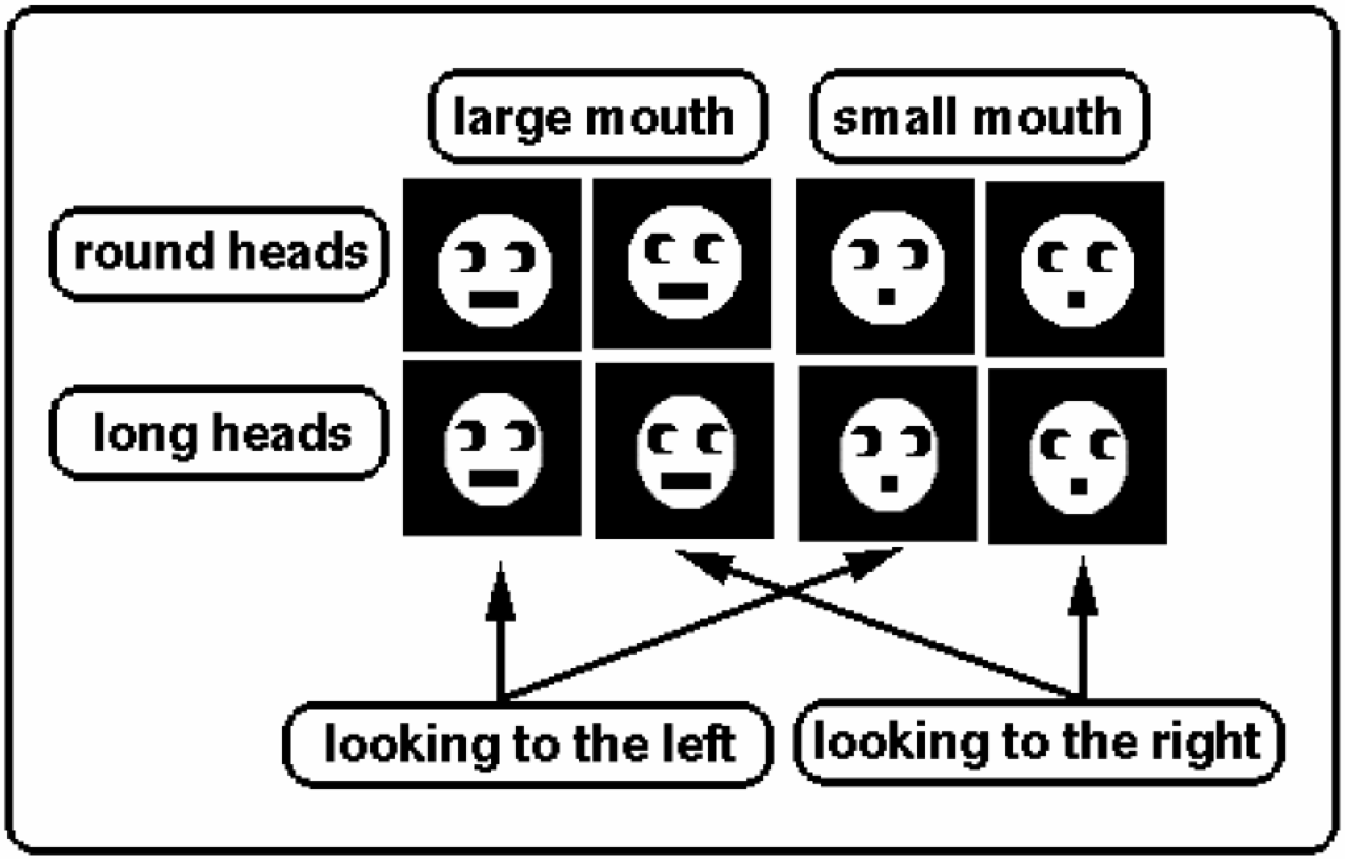
• Self-Organizing Map (SOM) and Topology-Representing Network (TRN) are neural net based approaches to dimensionality reduction.

Factor Space / Principal Components:



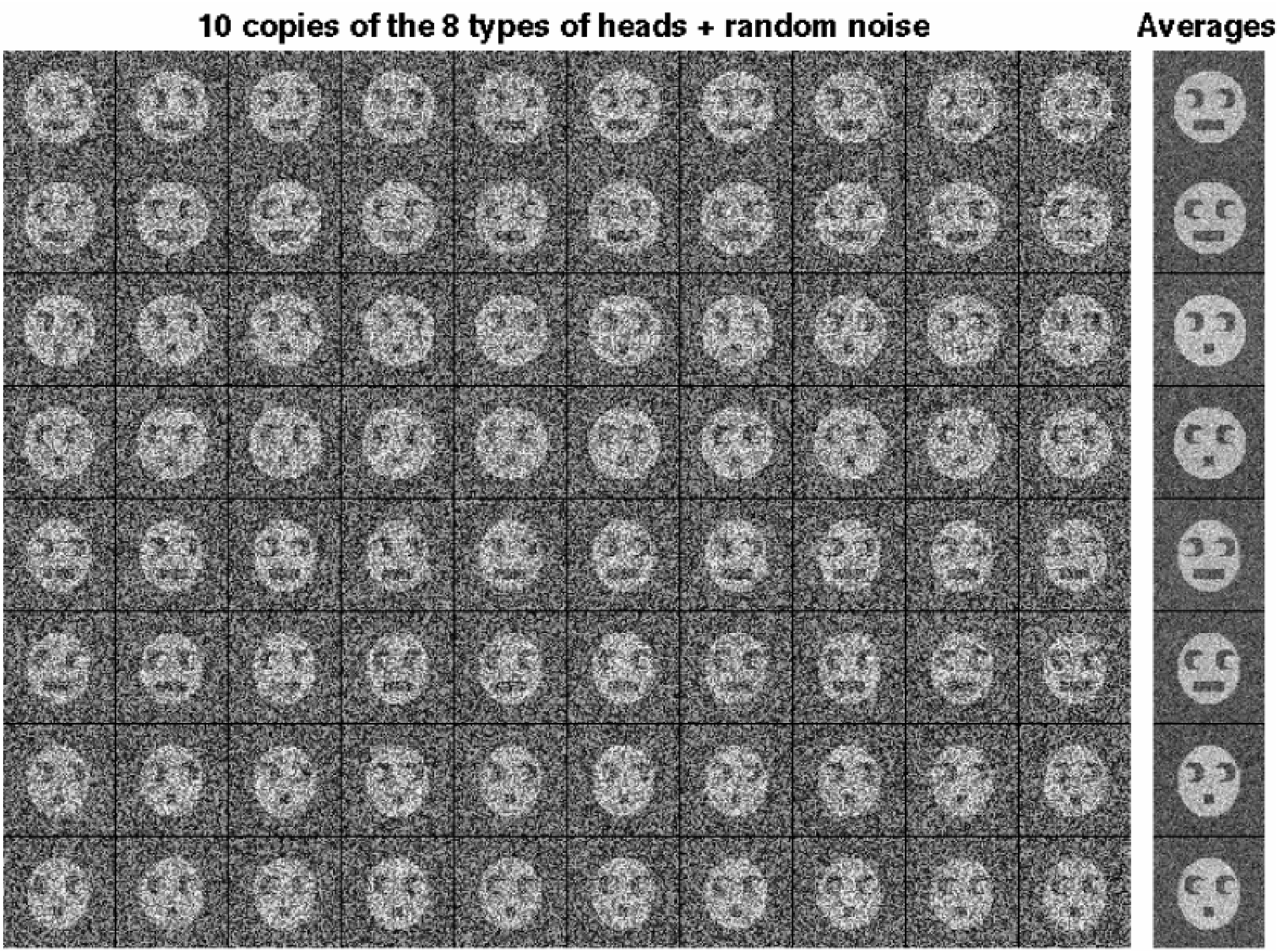
Find new coordinate system, tailored to the data

Example

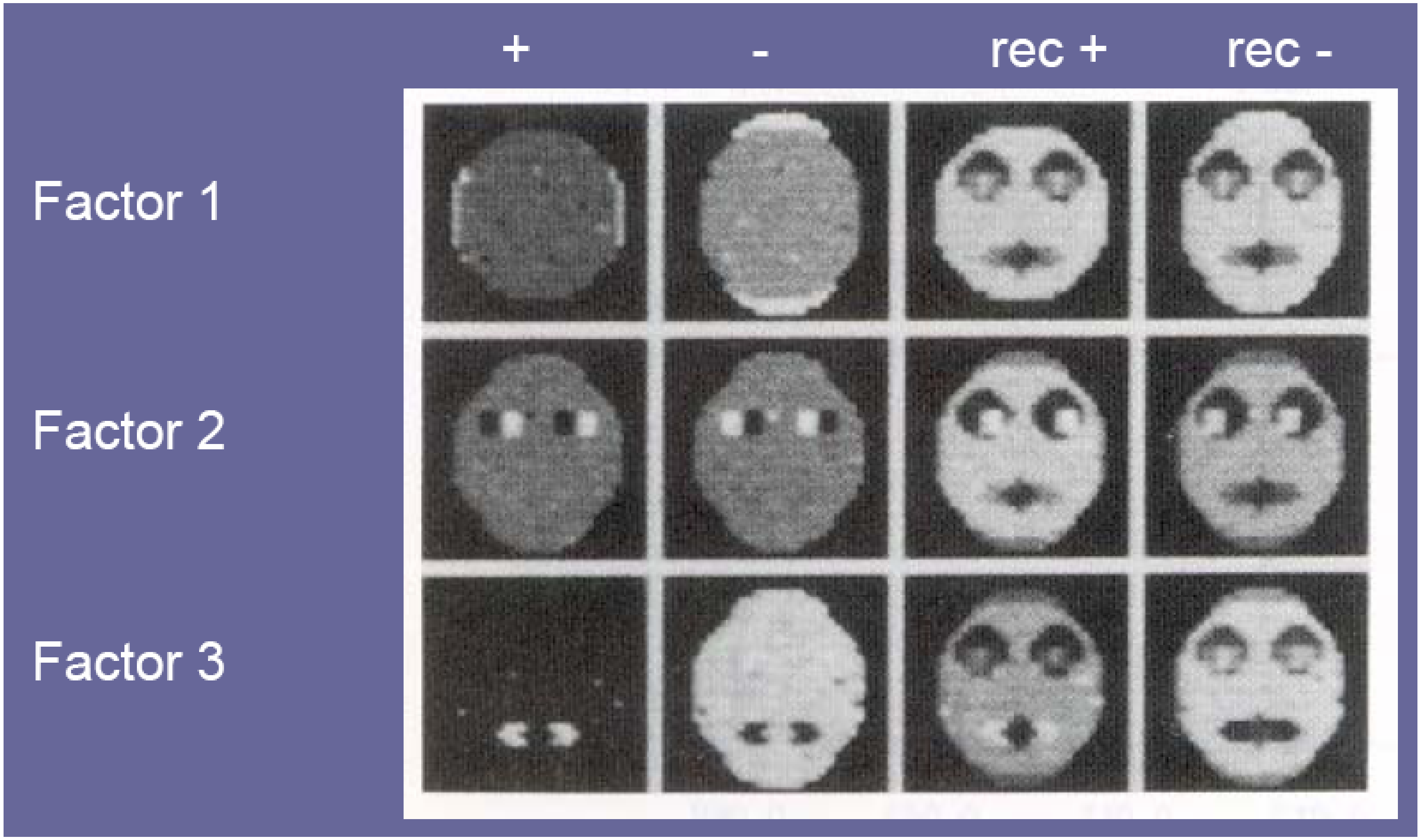


Brétaudière JP and Frank J (1986) Reconstitution of molecule images analyzed by correspondence analysis: A tool for structural interpretation. *J. Microsc.* **144**, 1-14.

Simulated Noisy Data



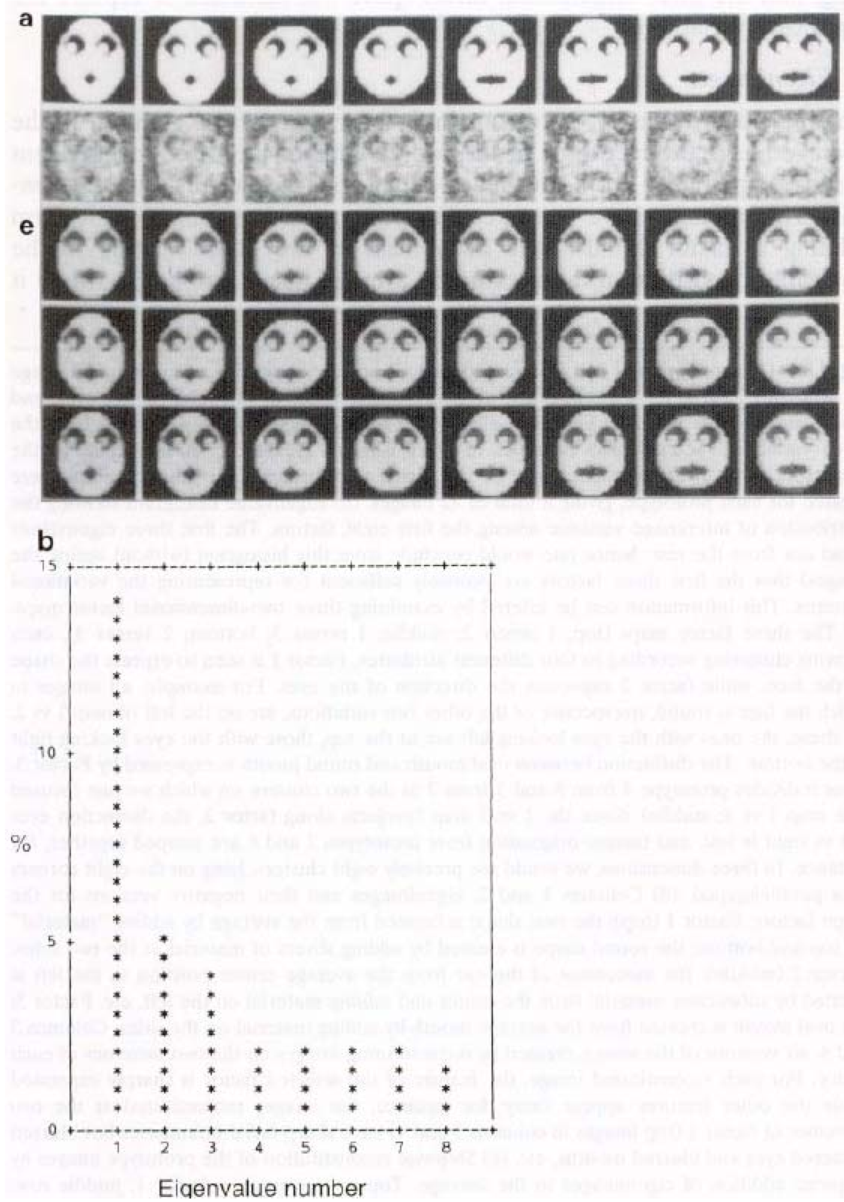
Eigenimages Extracted from Data (MSA)



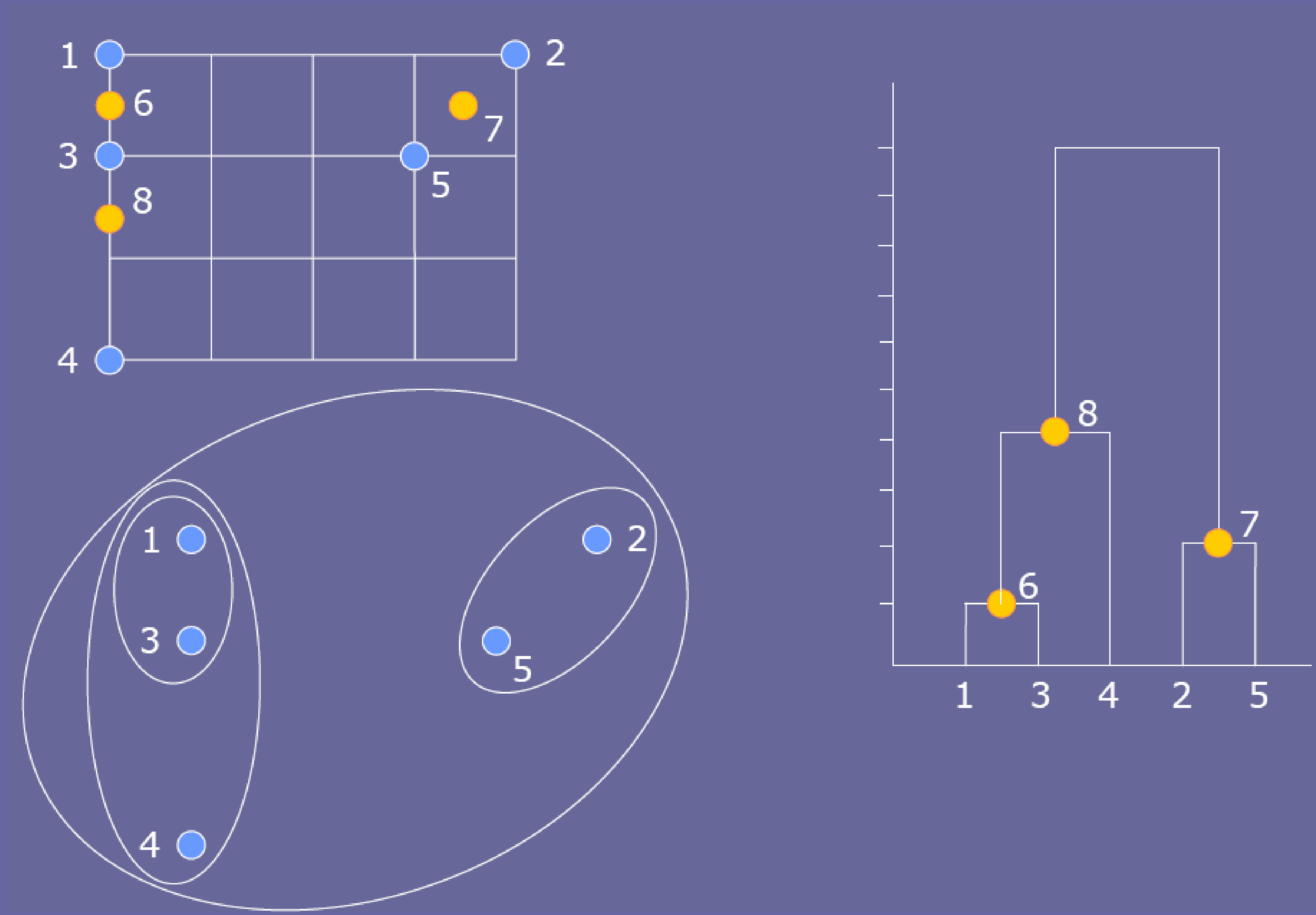
Note: explained in Frank book, chapter 4.

Expansion in Factor Space

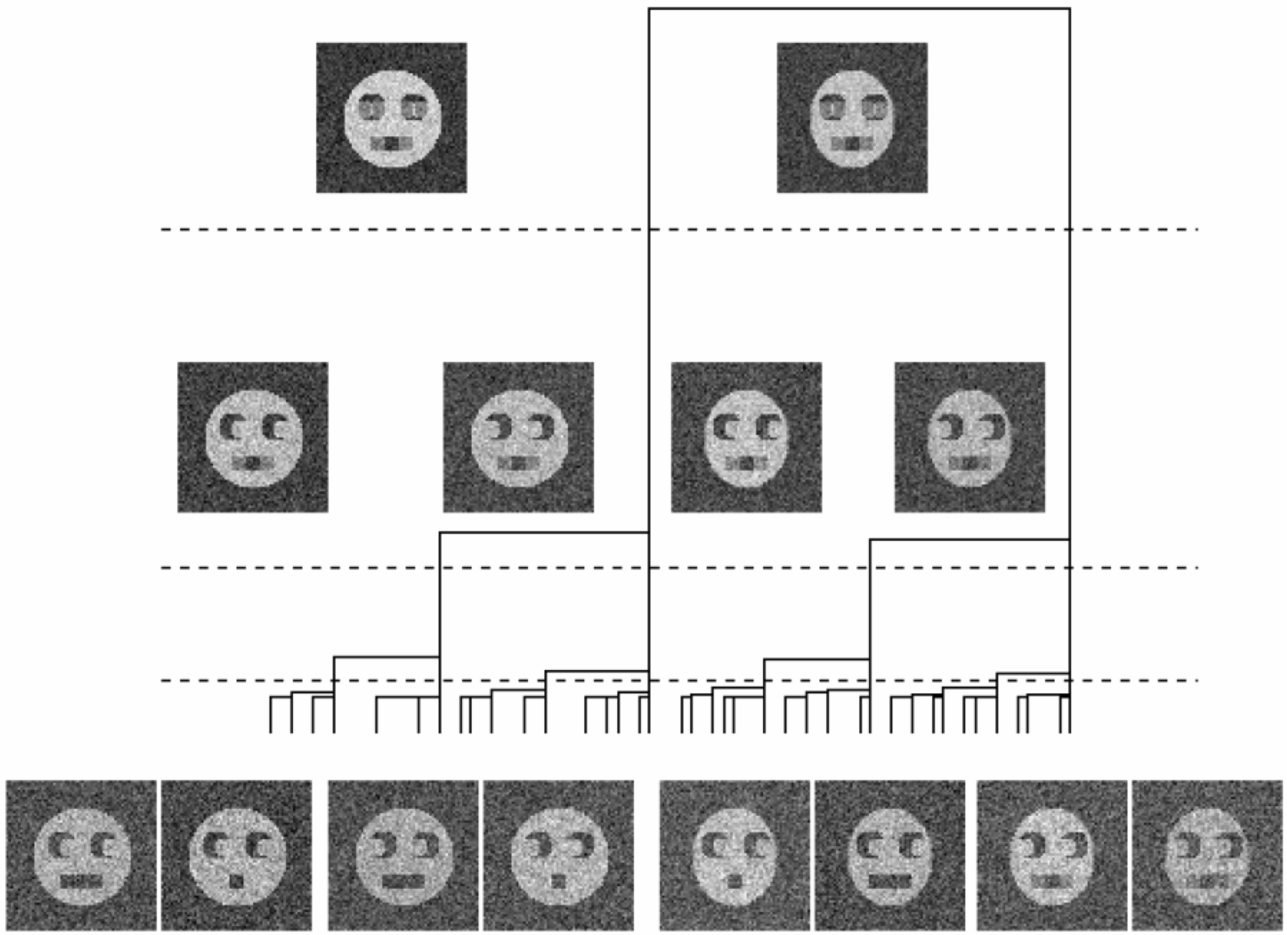
Avrg + F1
Avrg + F1+F2
Avrg + F1+F2+F3



Hierarchical Ascendant Classification

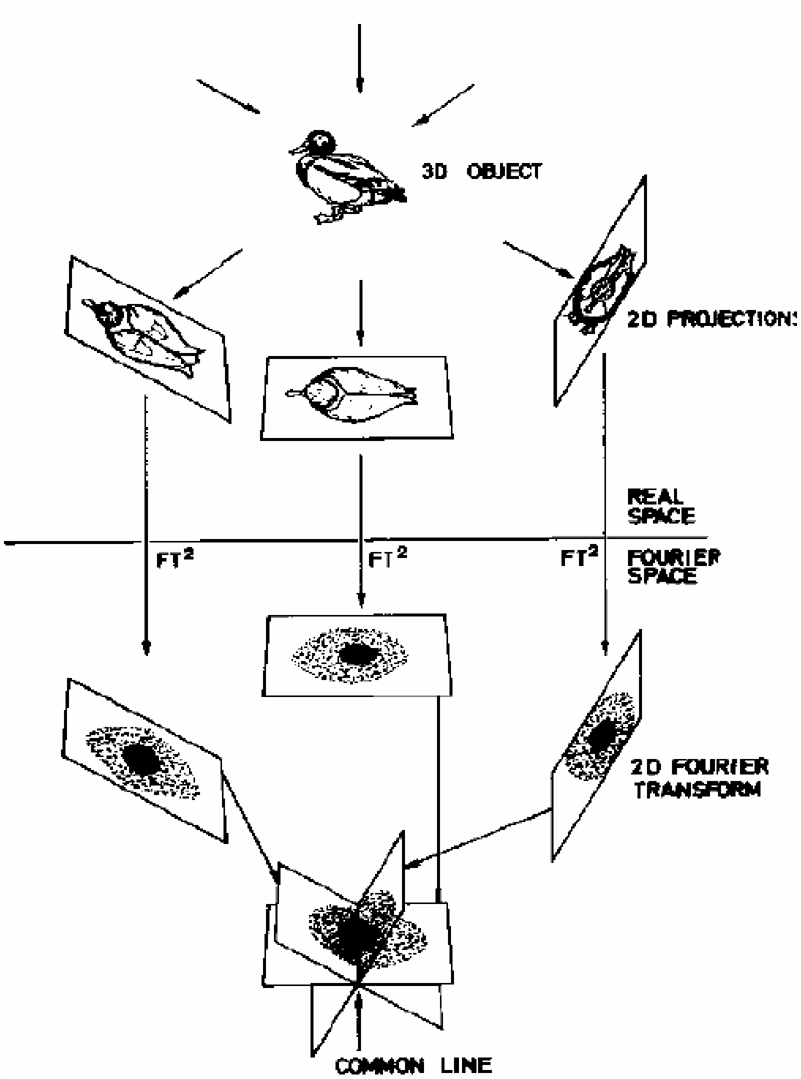


Hierarchical Ascendant Classification



Note: Many other clustering techniques exist…

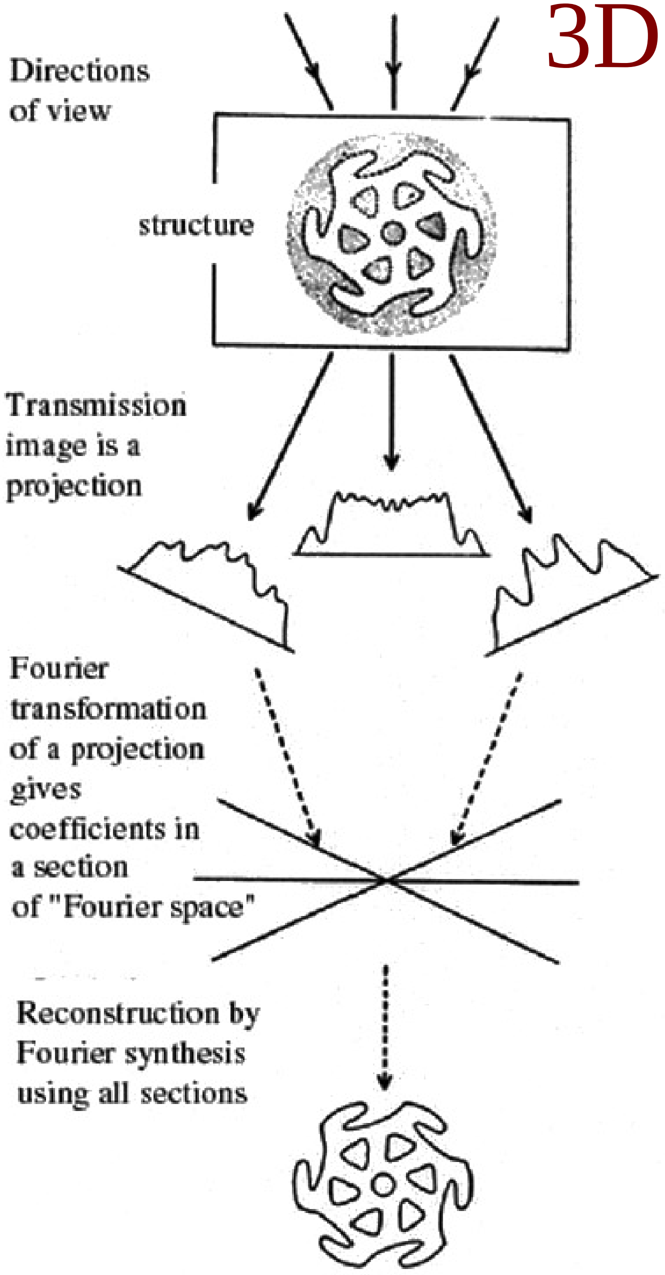
3D Reconstruction



Projection Theorem:

The 2D Fourier transform of the 2D projection of a 3D density is a central section of the 3D Fourier transform of the density perpendicular to the direction of projection.

3D Reconstruction



DeRosier & Klug, Nature 217 (1968) 133

Projection Theorem:

The 2D Fourier transform of the 2D projection of a 3D density is a central section of the 3D Fourier transform of the density perpendicular to the direction of projection.

This holds in Fourier Space.

Angular Reconstitution

Real Space:

Common Line Projection Theorem

Two different 2D projections of the same 3D object always have a 1D line projection in common.

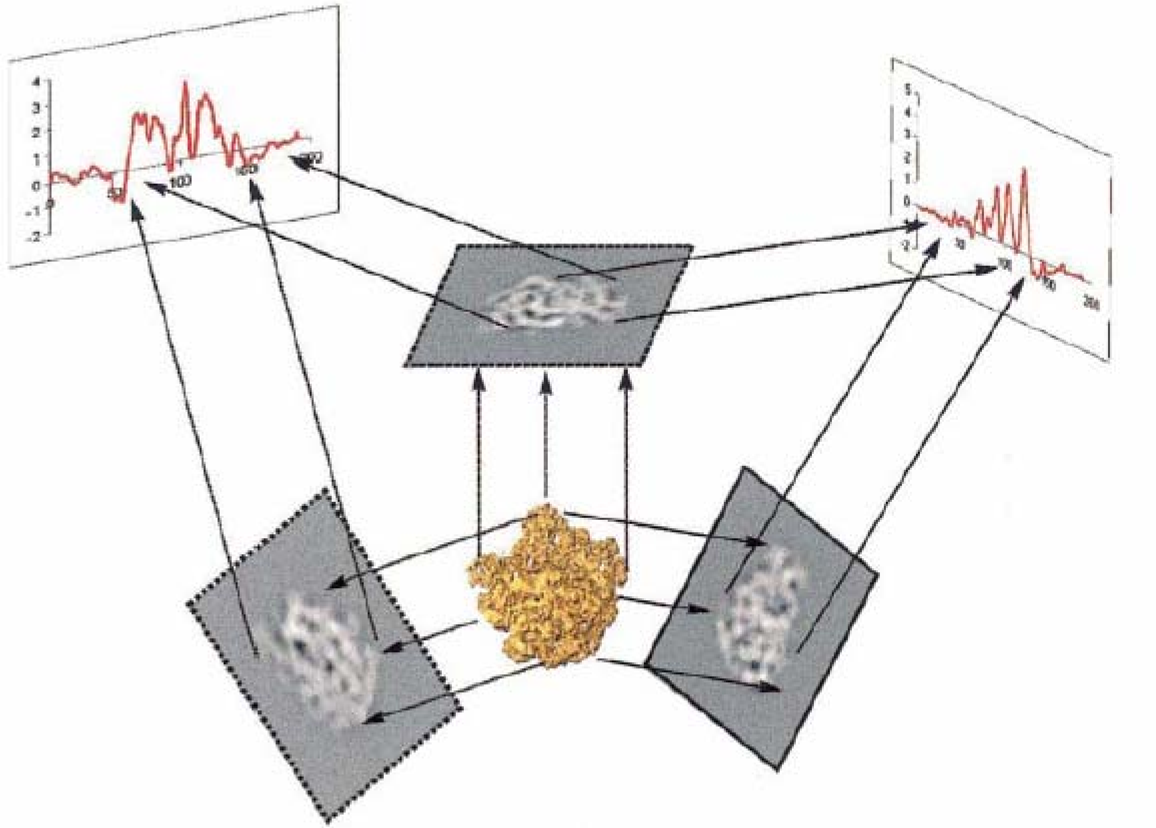


Fig. 11. The angular reconstitution technique is based on the common line projection theorem stating that two different two-dimensional (2D) projections of the same 3D object always have a one-dimensional (1D) line projection in common. From the angles between such common line projections, the relative Euler-angle orientations of set projections can be determined *a posteriori* (van Heel, 1987). For an entirely asymmetric particle like this 50S ribosomal subunit, at least three different projections are required to solve the orientation problem. For details see main text.

van Heel et al, Quarterly Reviews of Biophysics **33**, 4 (2000), pp. 307–369.

Sinograms

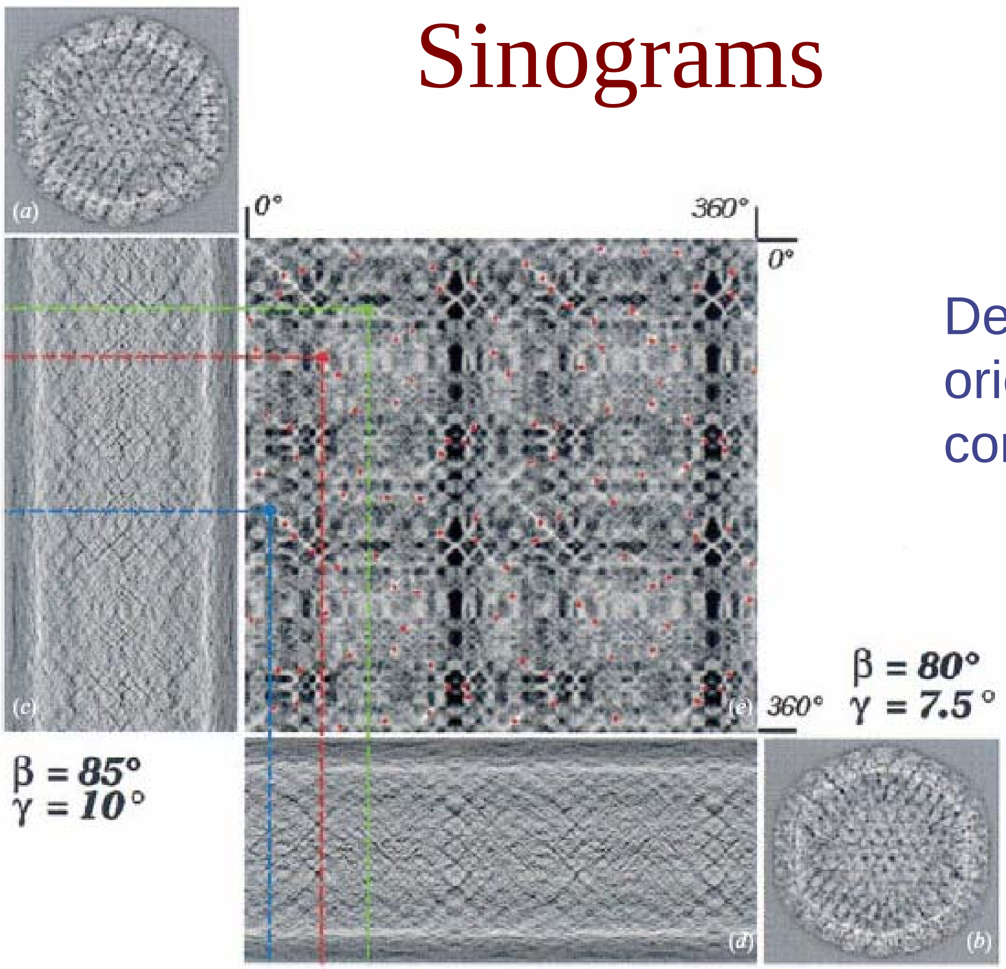


Fig. 13. Sinograms and sinogram correlation functions. This illustration provides a graphical overview of the relations between a 2D class average (noise-reduced projection images), their 'sinograms', and the sinogram correlation function between two sinograms. The images shown here (*a*, *b*) are class averages deduced a large data set of Herpes Simplex Virus Type 1 (HSV1) cryo-EM images. Each line of the sinogram images (*c*, *d*) is generated from the 2D projection image by summing all 1D lines of the 2D images, from top to bottom, after rotation of the image over angles ranging from 0° to 360°. Equivalently, the lines of the sinograms are 1D projections of the 2D images in all possible directions ranging from 0° to 360°. Each point of the sinogram correlation function contains the correlation coefficient of two lines of the two sinograms one is comparing (*e*).

Determine relative orientations with common lines!

van Heel et al, Quarterly Reviews of Biophysics **33**, 4 (2000), pp. 307–369.

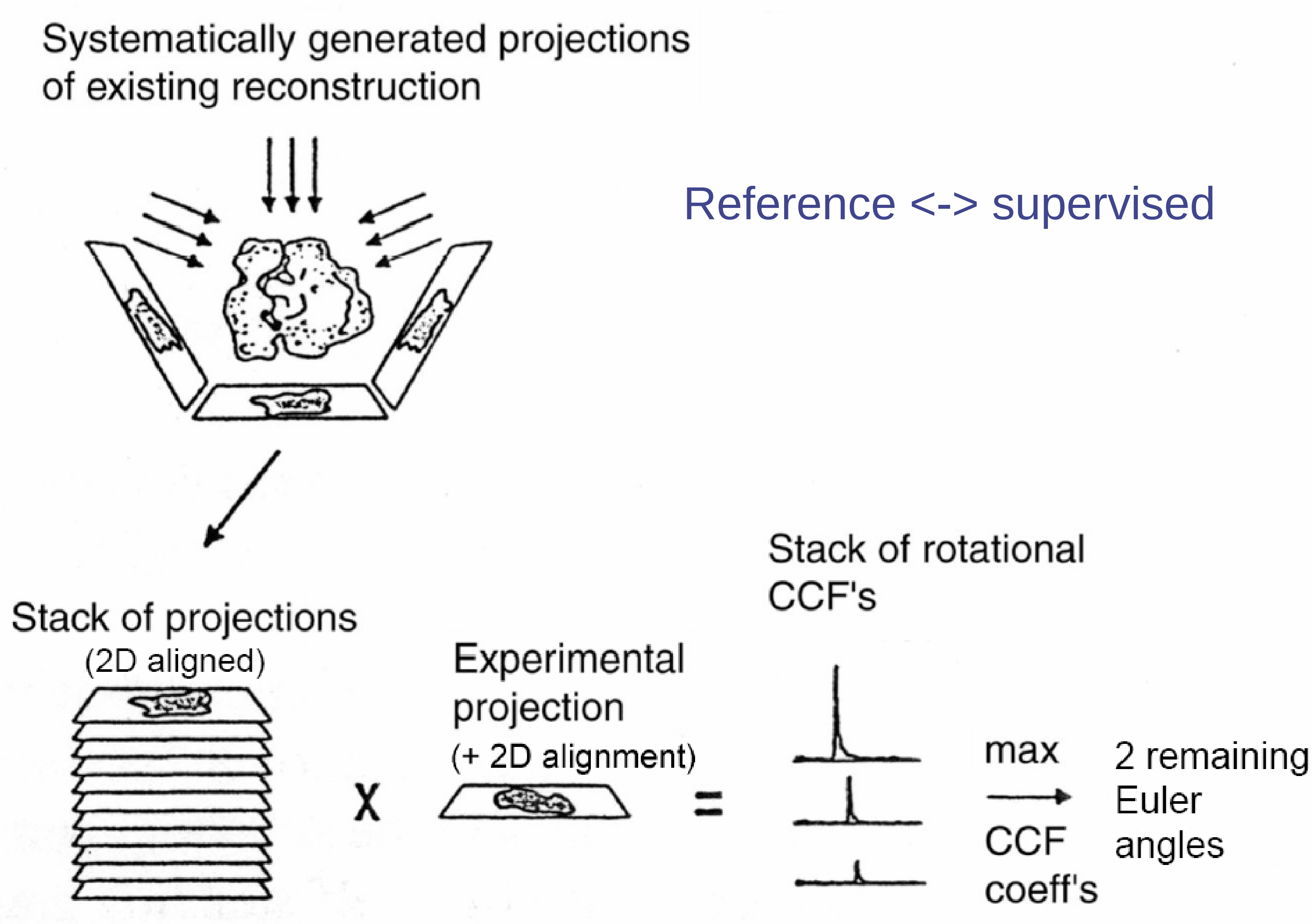
Angular Reconstitution

1) *Unsupervised* classification, to determine classes of particles exhibiting the same view
2) Unsupervised <-> Reference-free
3) Average images in each class -> class averages
4) 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

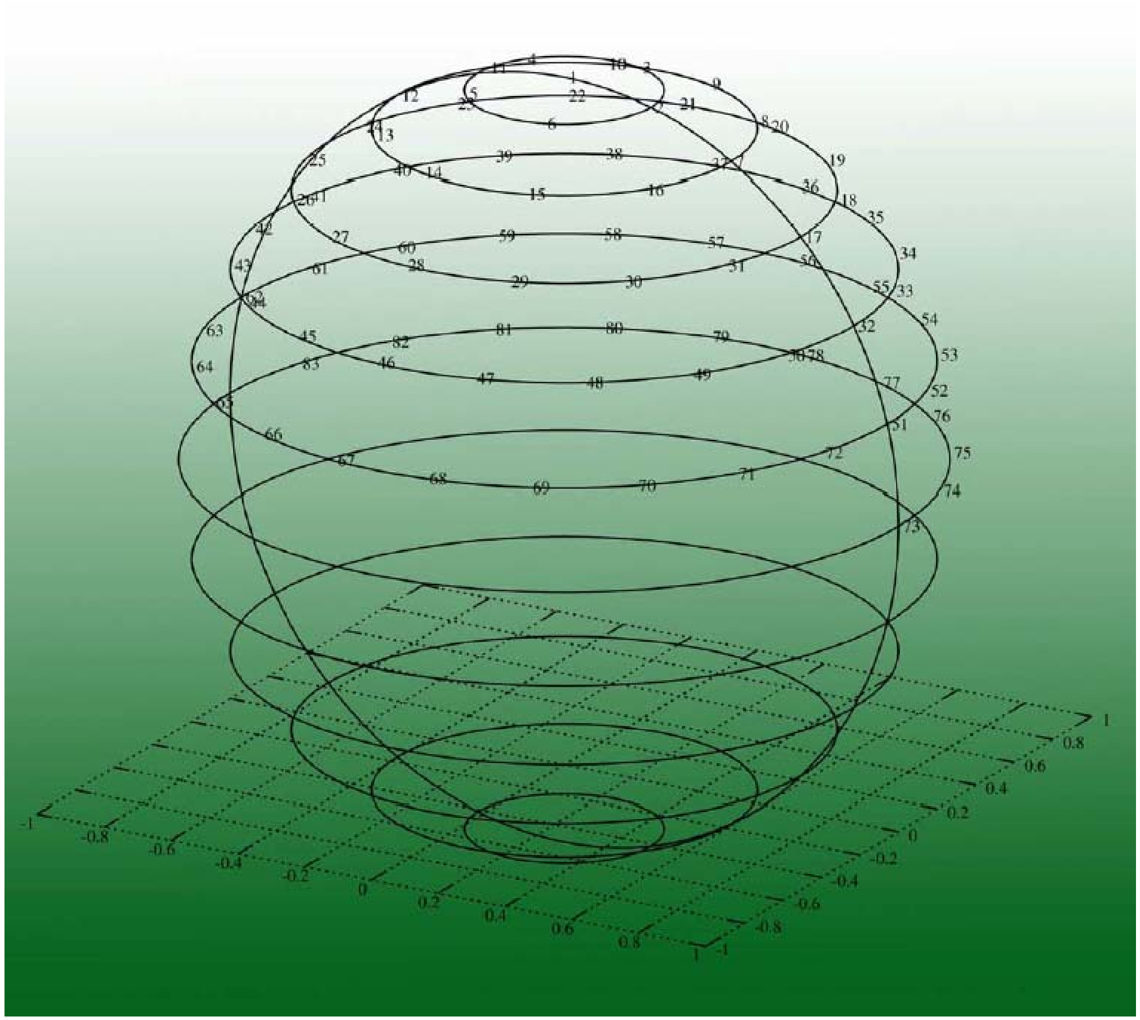
Reference-Based Projection Matching



Reference <-> supervised

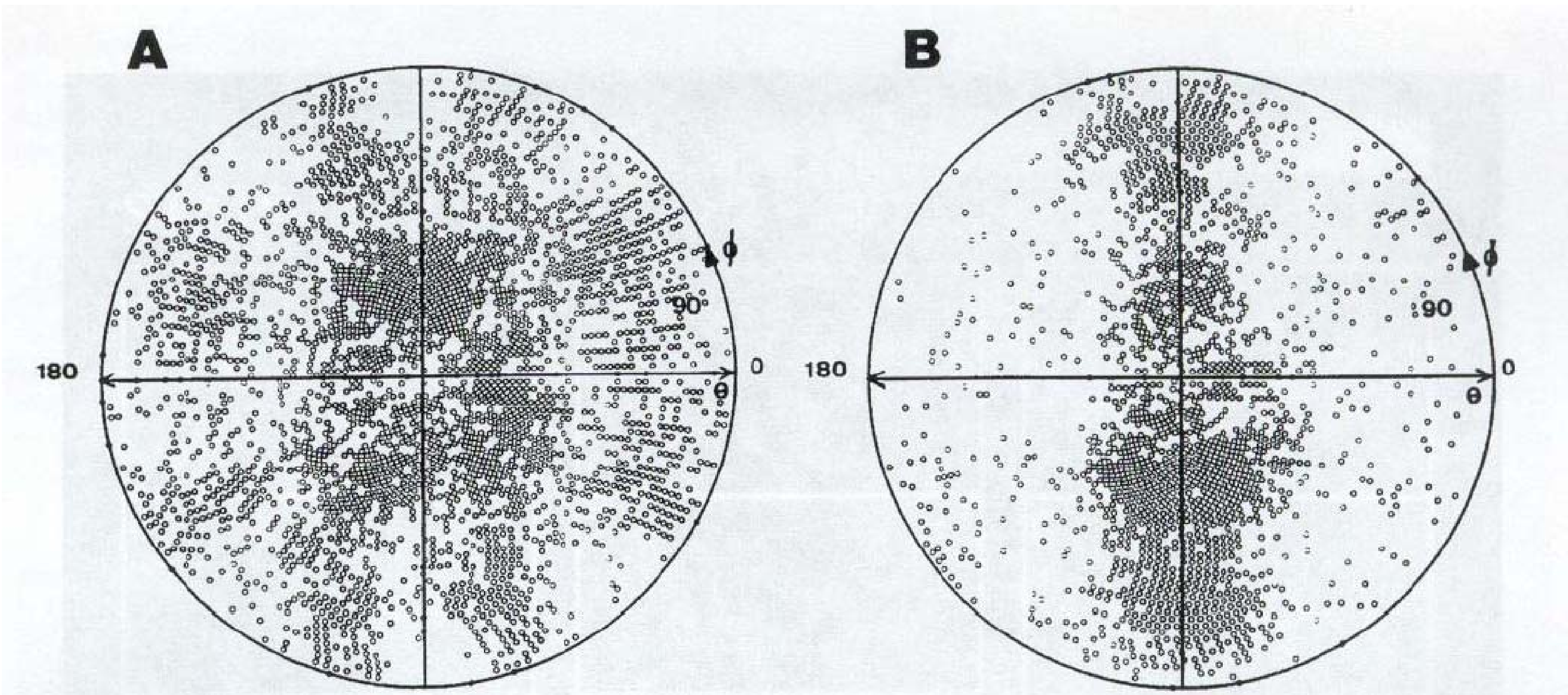
adapted from an image by Joachim Frank

Reference-Based Projection Matching



Angular Coverage

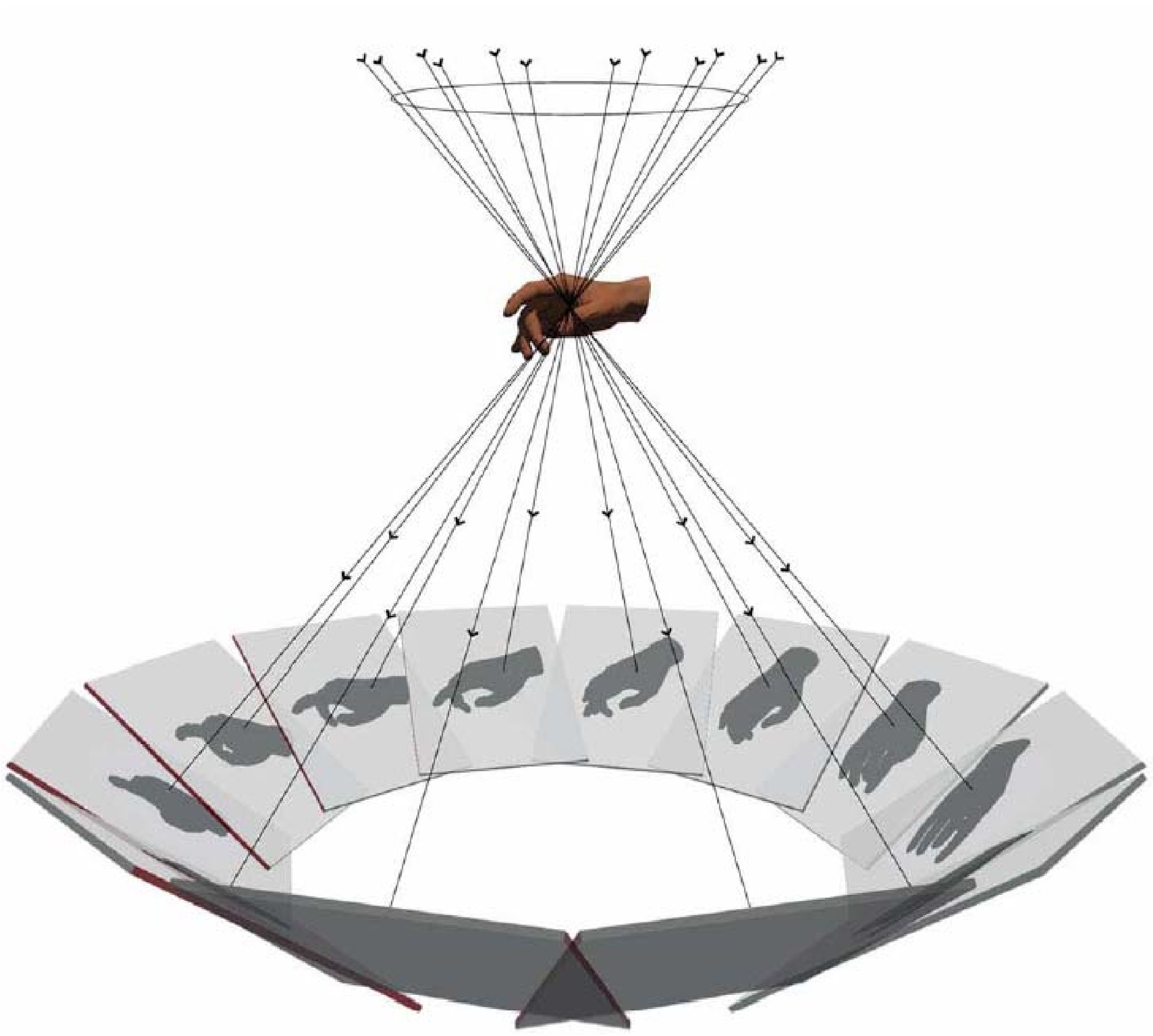
good poor



What if Particles are Aligned with Grid?

Solution: Tilt of Specimen

Random Conical Tilt



Random Conical Tilt

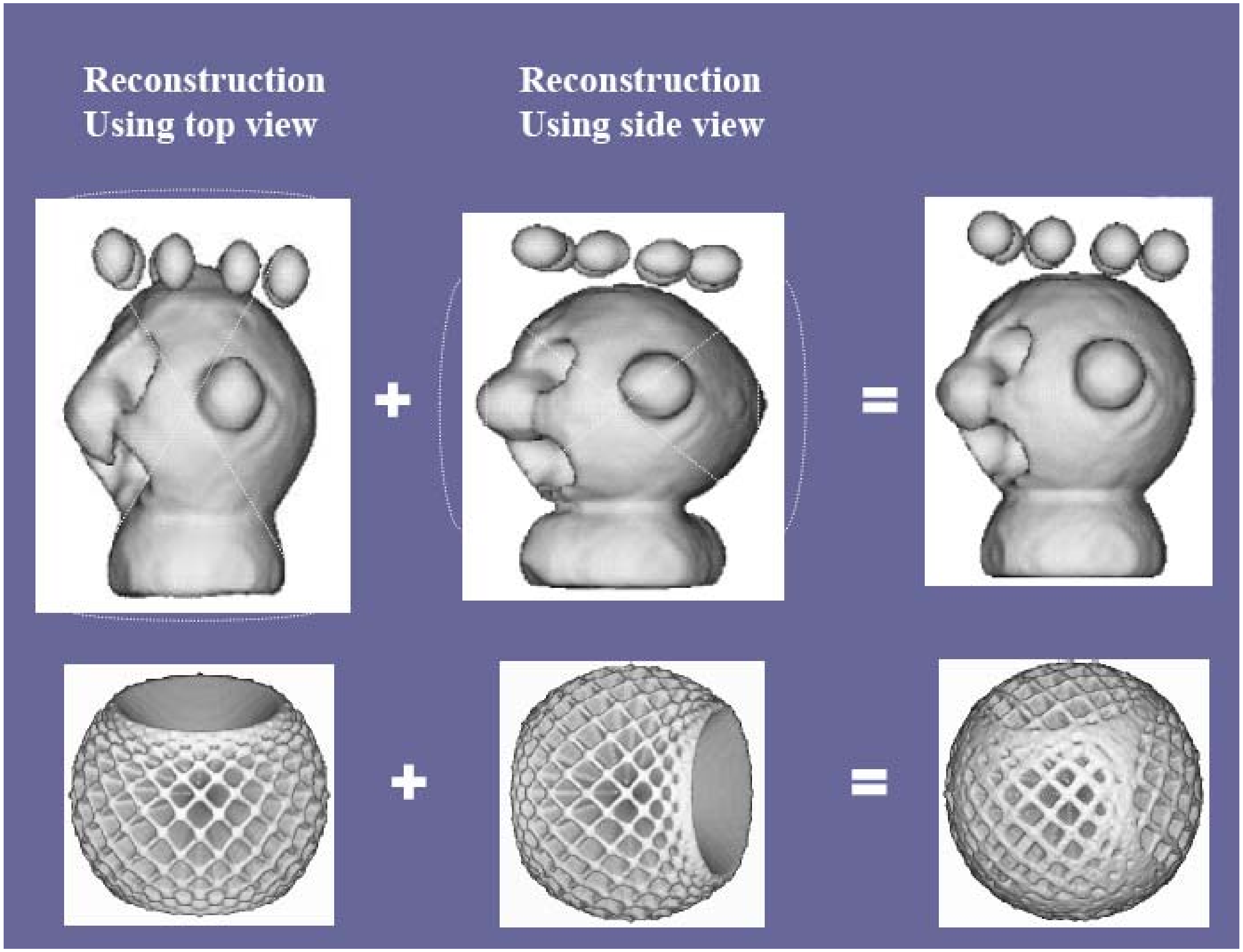
• Premise: all particle exhibit the same view
• Take same field first at theta ~50 degrees, then at 0 degrees [in this order, to minimize dose]
• Display both fields side by side
• Pick each particle in both fields
• Align particles from 0-degree field
This yields azimuths, so that data can be put into the conical geometry
• Assign azimuths and theta to the tilted particles
• Proceed with 3D reconstruction

Random Conical Tilt for Multiple Orientations

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

2) Missing cone problem: *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

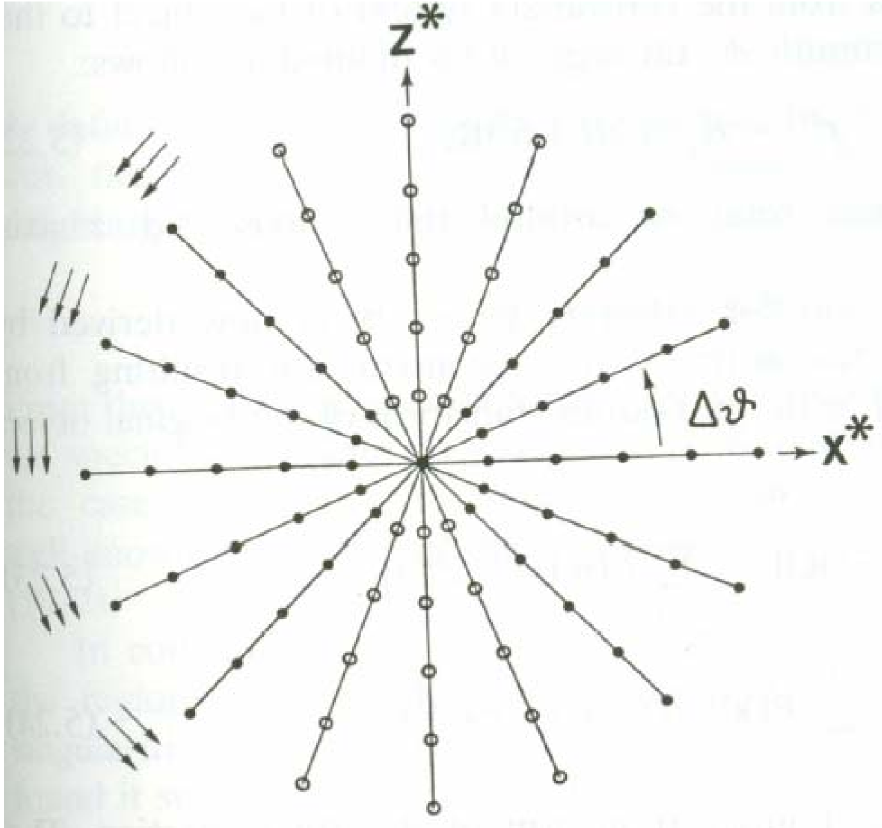


Reconstruction Algorithms

(a) Fourier interpolation
(b) Weighted back-projection
(c) Iterative algebraic reconstruction

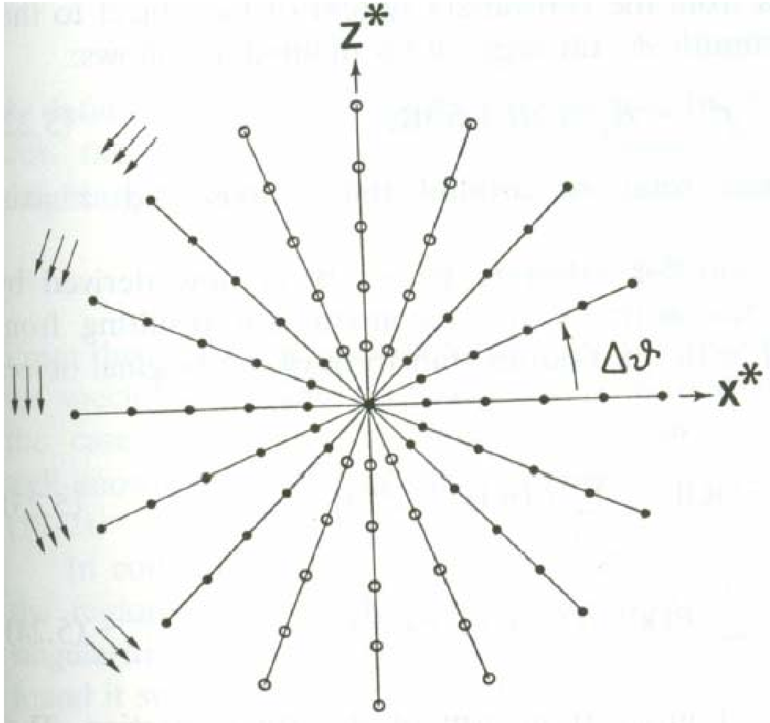
Fourier Interpolation

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.



Fourier Interpolation

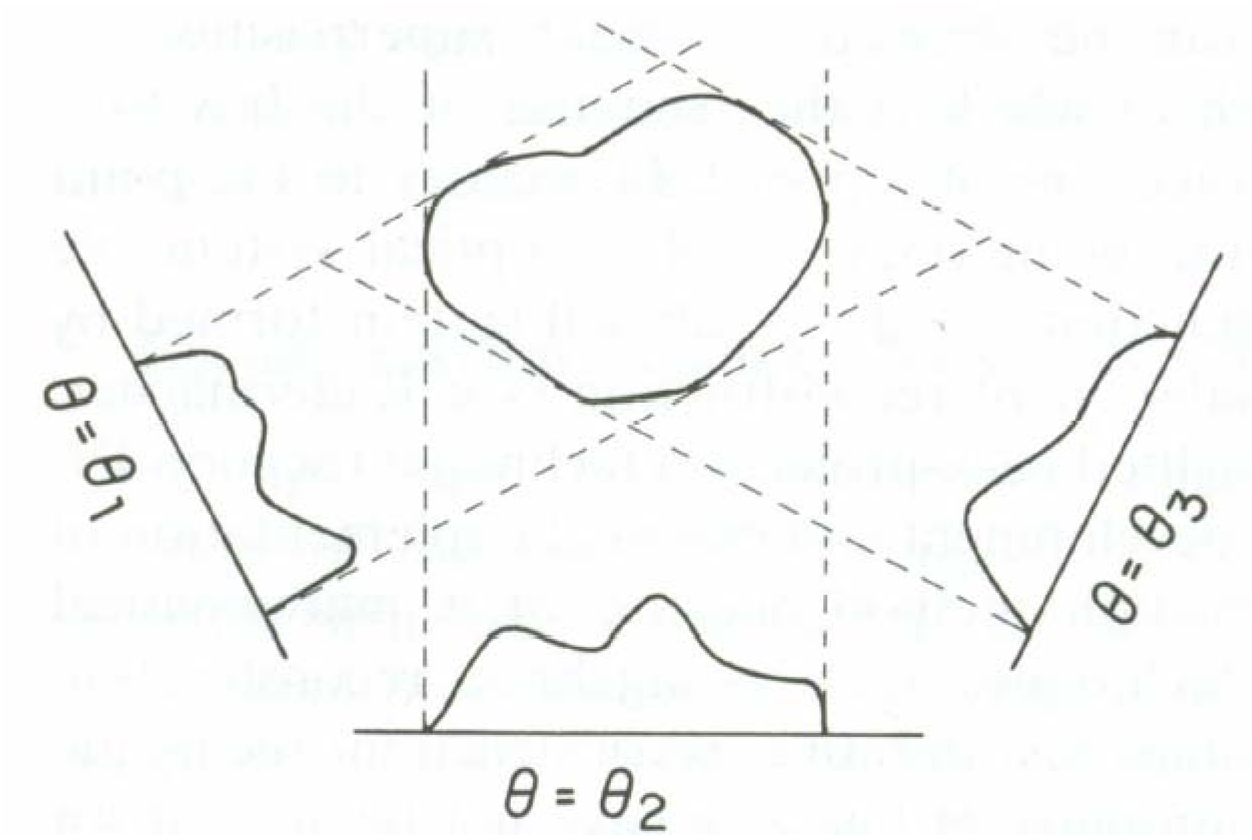
Sample points of adjacent projections are increasingly sparse as we go to higher resolution:



Speed (high) versus accuracy (low). Can be used in the beginning phases of a reconstruction project.

Back Projection

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



Weighted Back Projection

Weighted back-projection: as before, but “weight” the projections first by a function that is tailored to the angular distribution of directions (R* weighting, in X-ray terminology), then inversing the Fourier transform.

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.

Iterative Algebraic Reconstruction

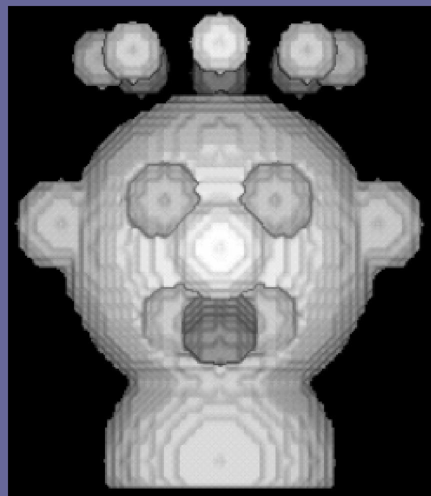
The discrete algebraic projection equation is satisfied, one angle at a time, by adjusting the densities of a starting volume. As iterations proceed, each round produces a better approximation of the object.

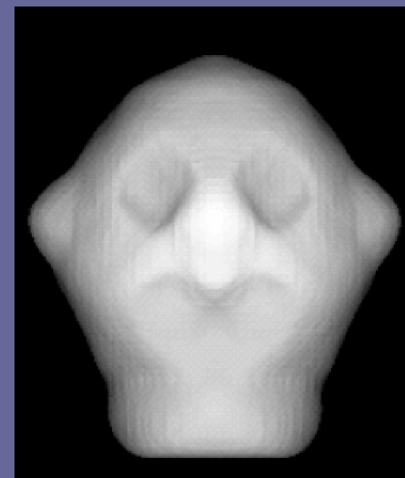
The algorithm comes in many variants. It allows constraints to be easily implemented.

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

Comparison (w/ Missing Cone)

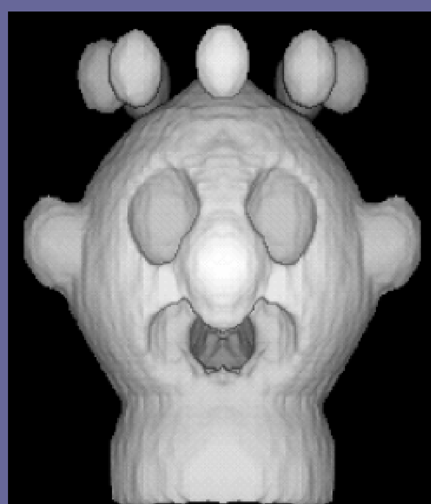
Original object

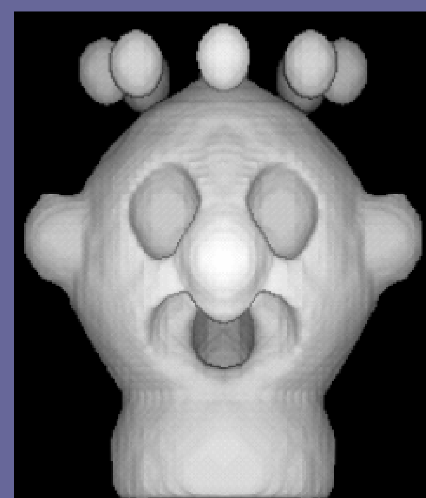




Simple back-projection

Weighted back-projection





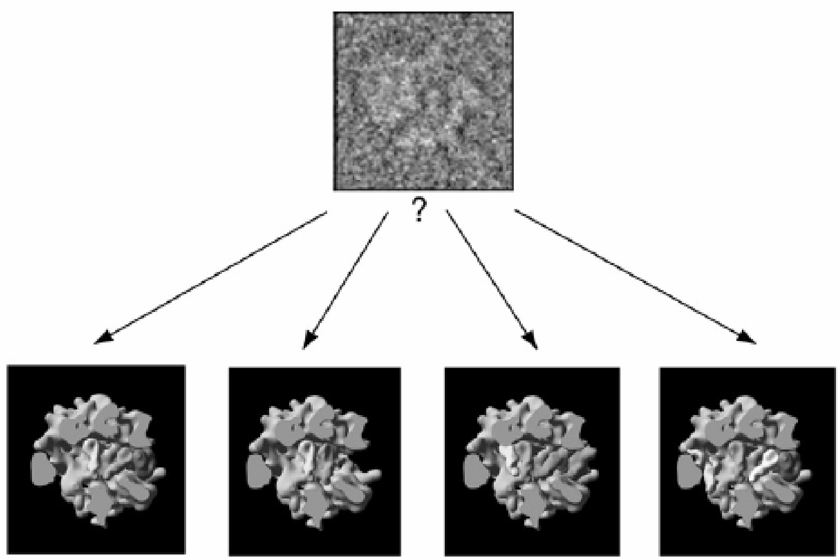
Iterative algebraic reconstruction

Sources for Limited Resolution

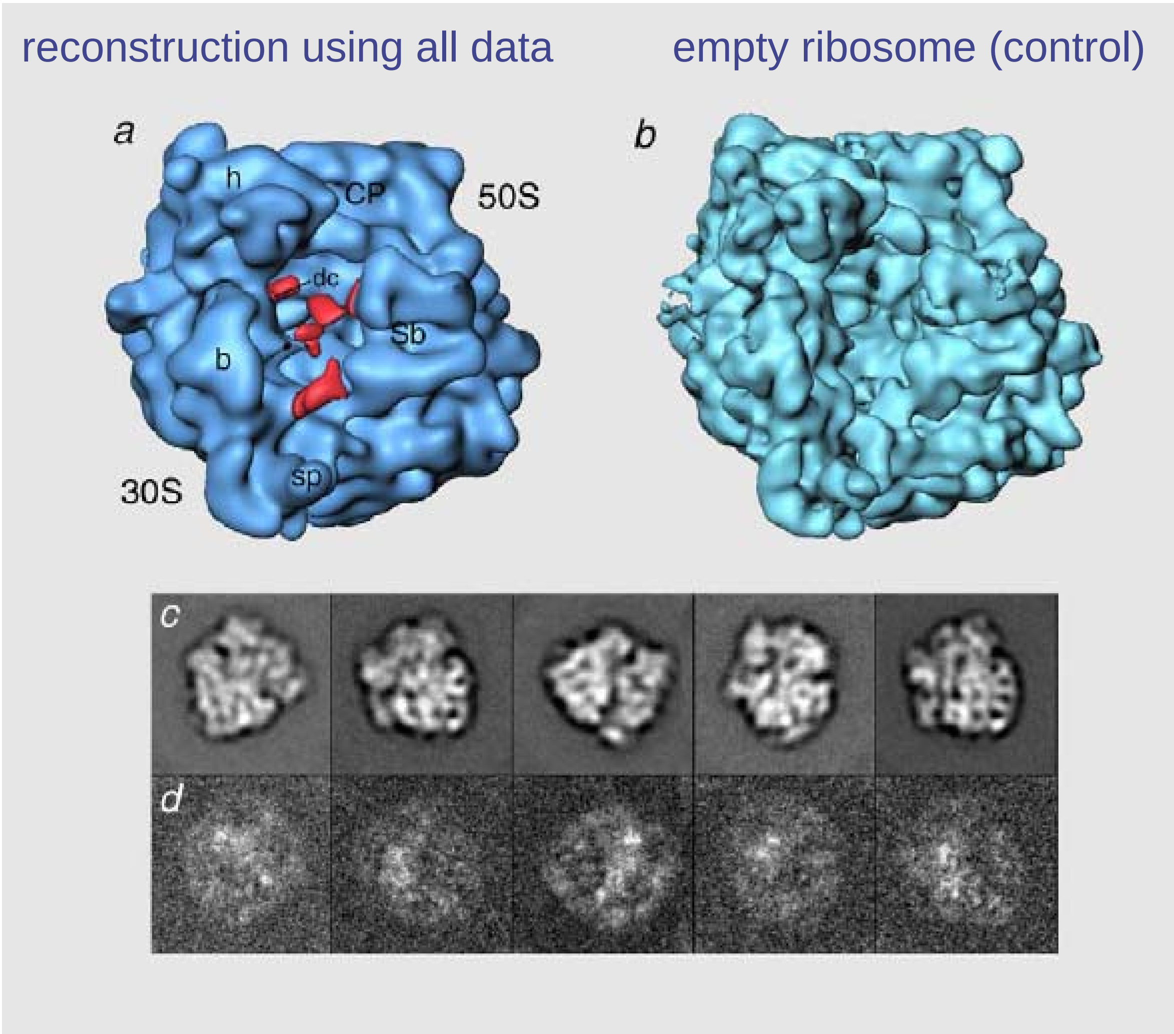
- Instrumental: partial coherence (envelope function)
- Particles with different height all considered having same defocus (envelope function)
- Numerical: interpolations, inaccuracies
- Failure to exhaust existing information
- Conformational diversity

Conformational Diversity: Heterogeneous Particle Population

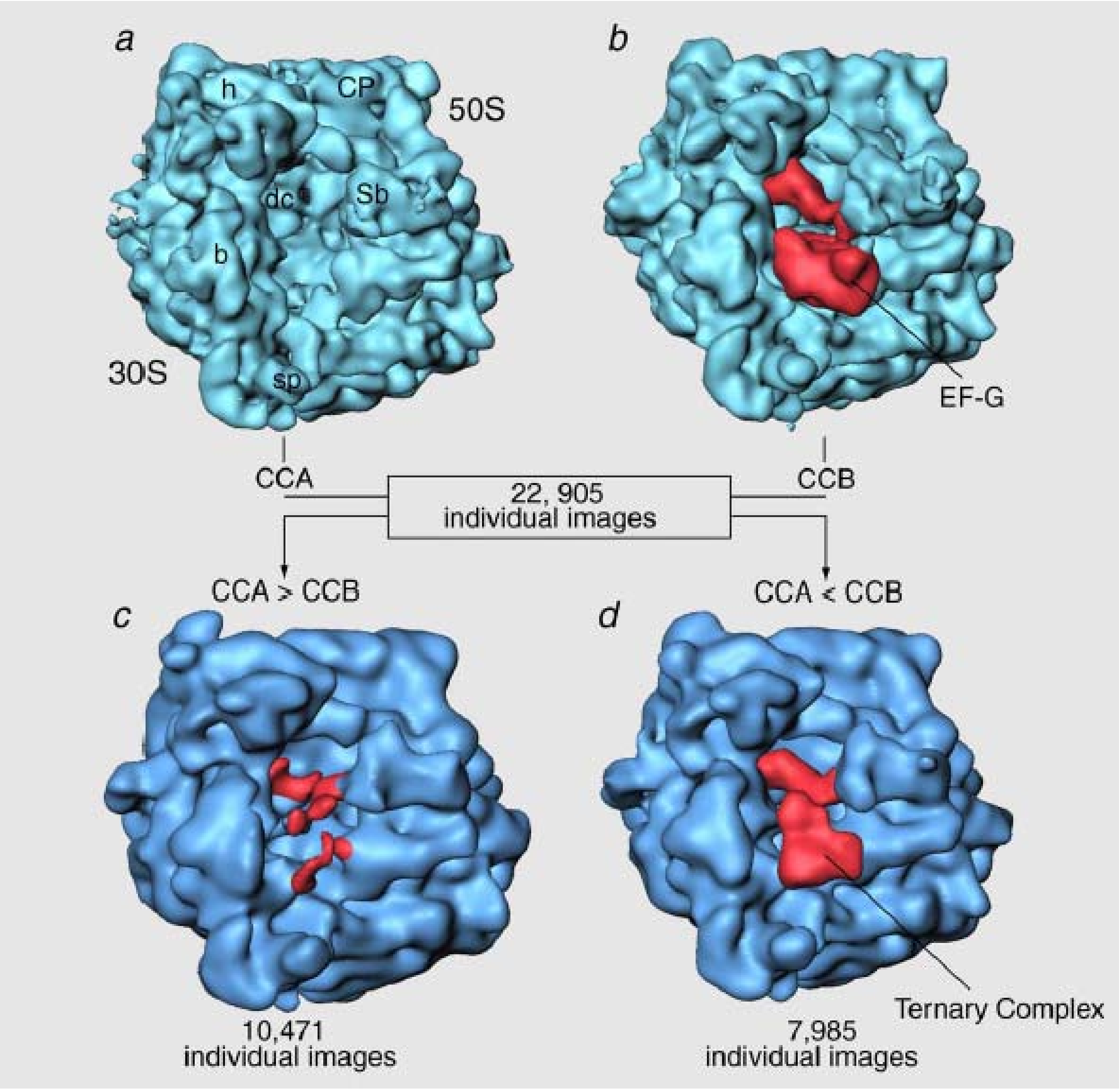
Current approach: assume all conformers are "similar". Treat problem in first approximation as a problem with a single conformer. Then try different models as references to see if population segregates.



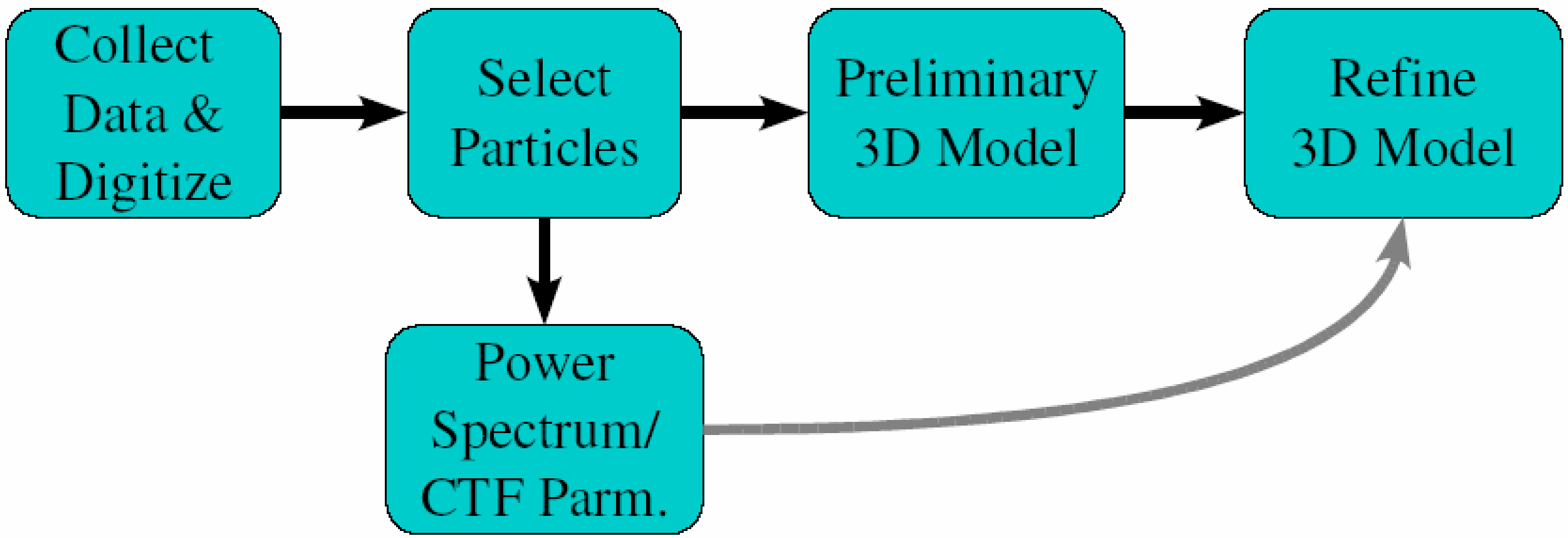
Example: Low Occupancy of Ternary Complex



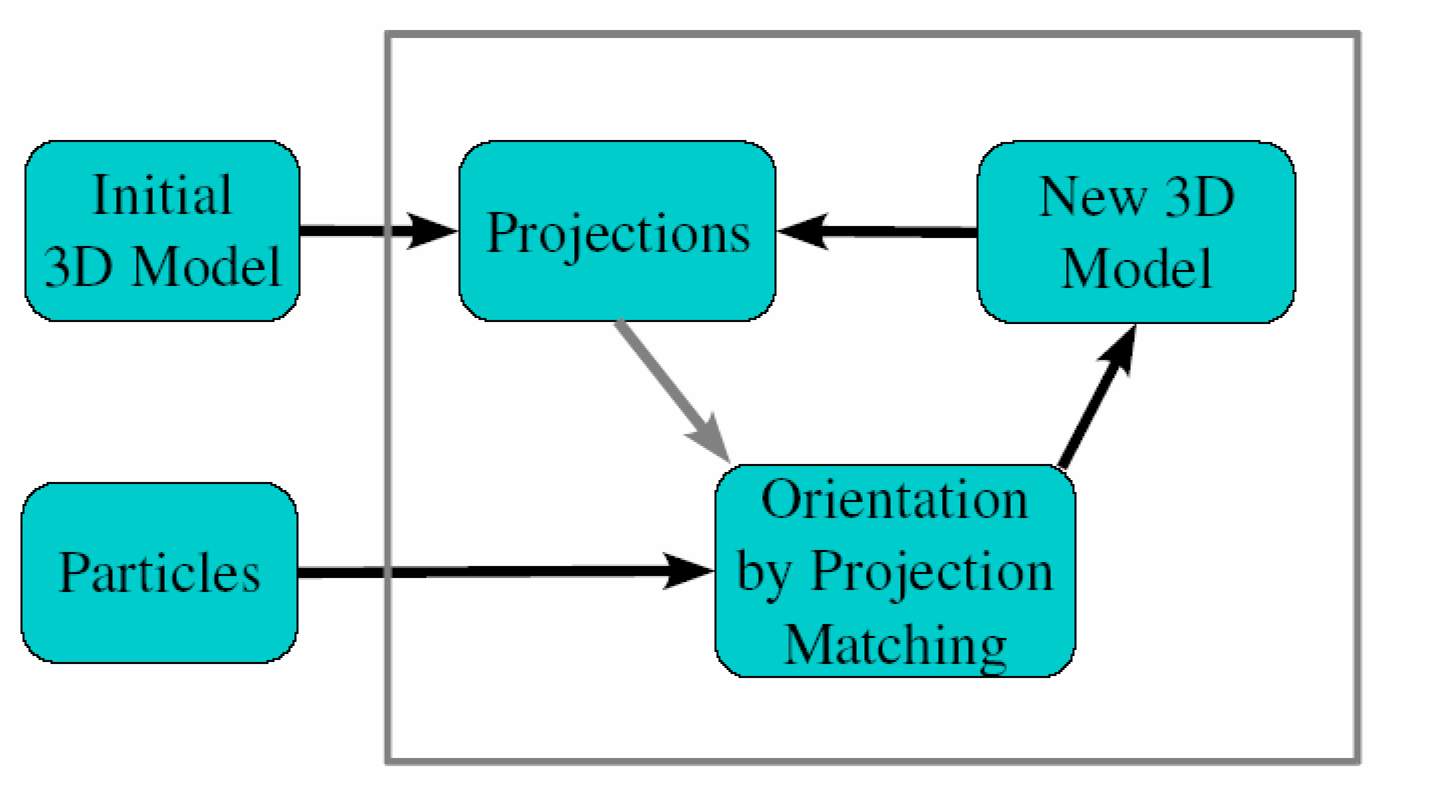
Problem Solved by Supervised Classification



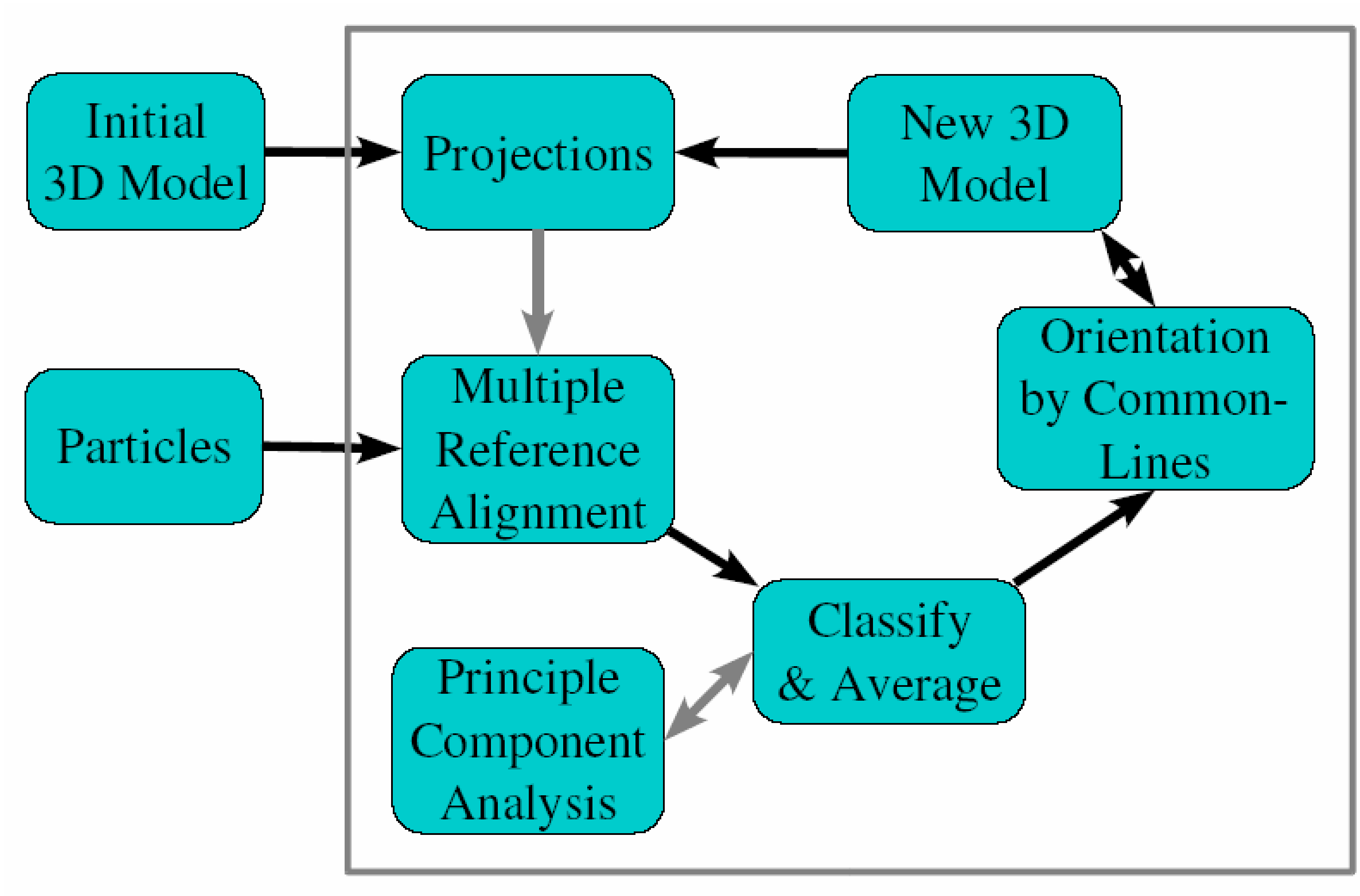
The Reconstruction Process



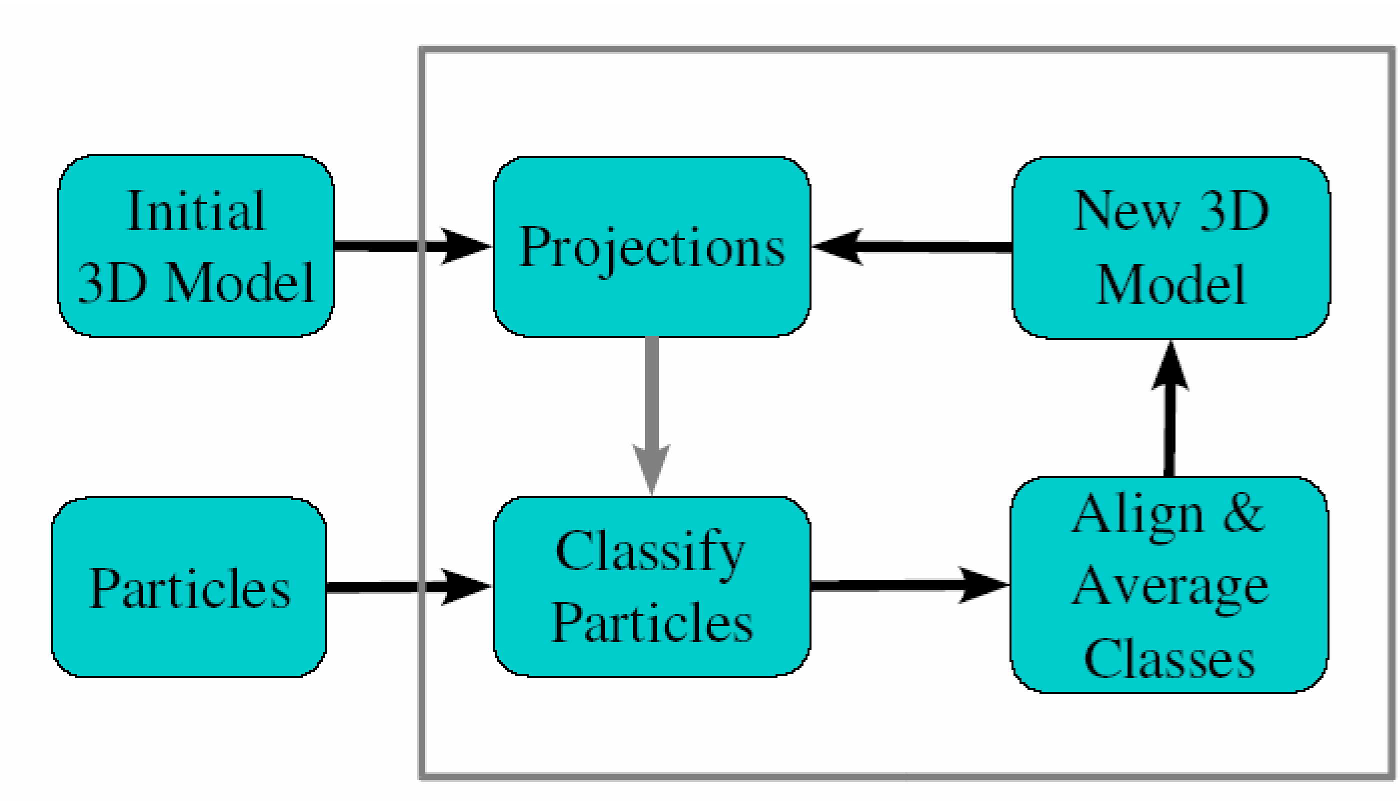
Typical Refinement - SPIDER



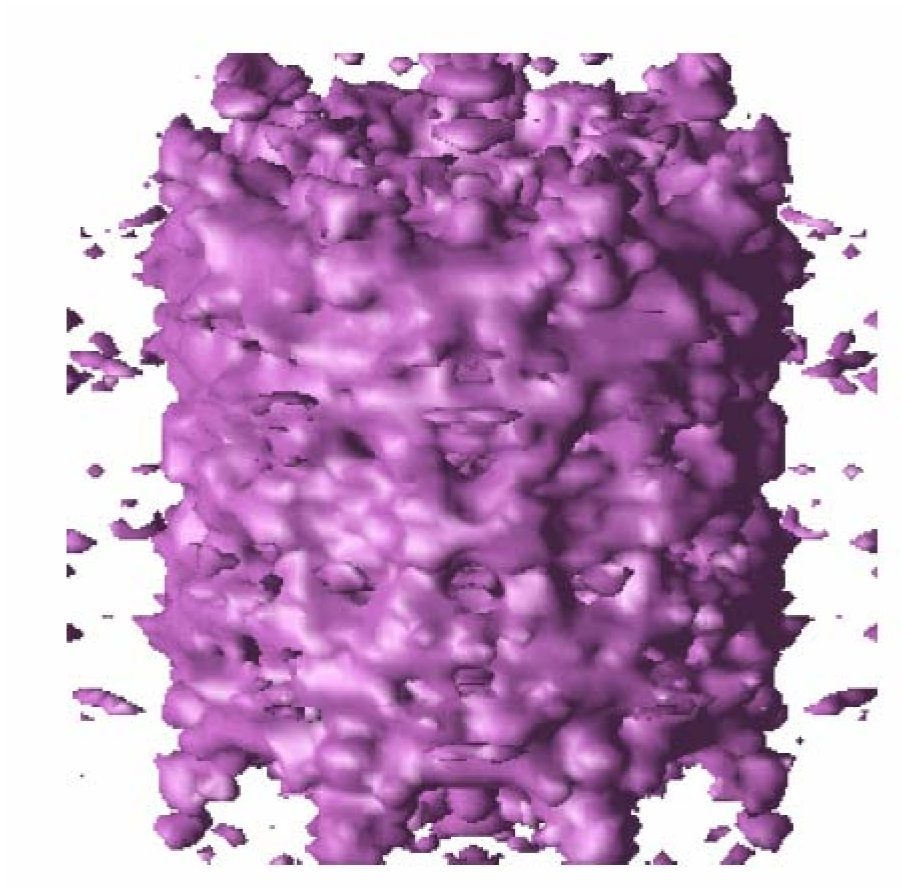
Typical Refinement - IMAGIC



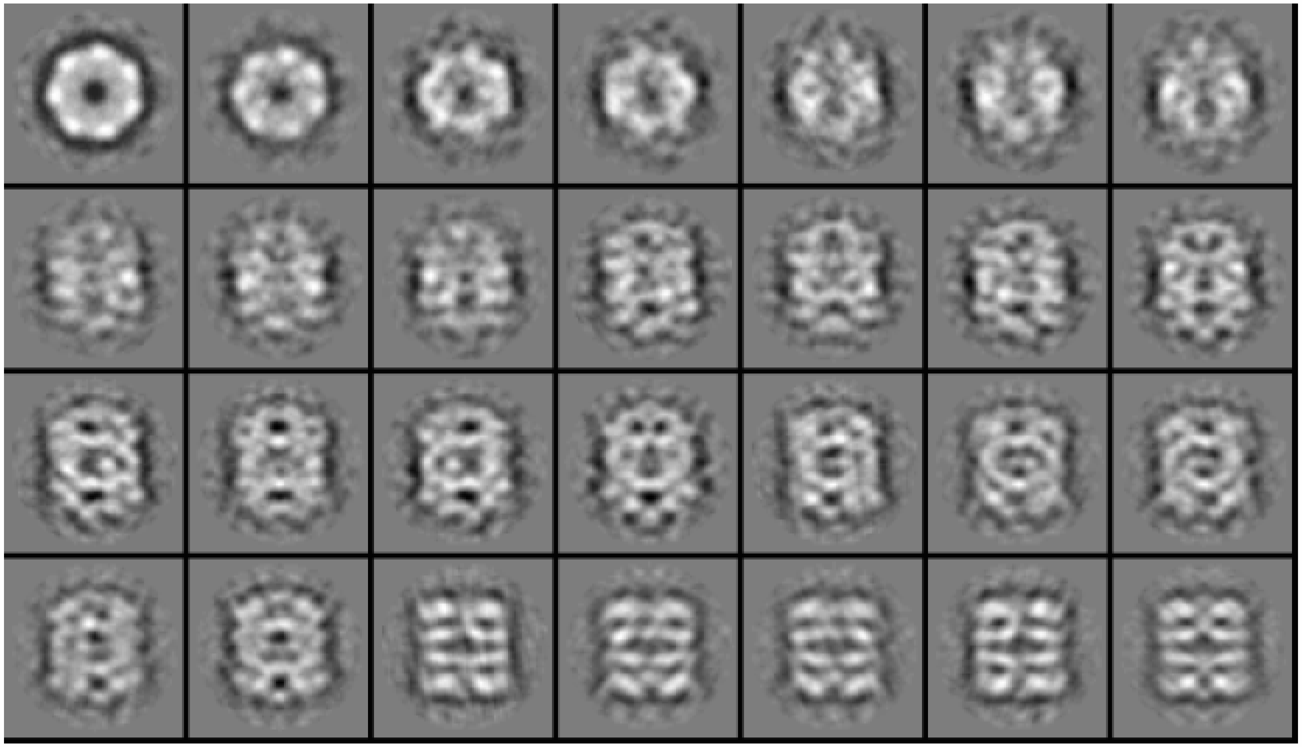
Typical Refinement - EMAN



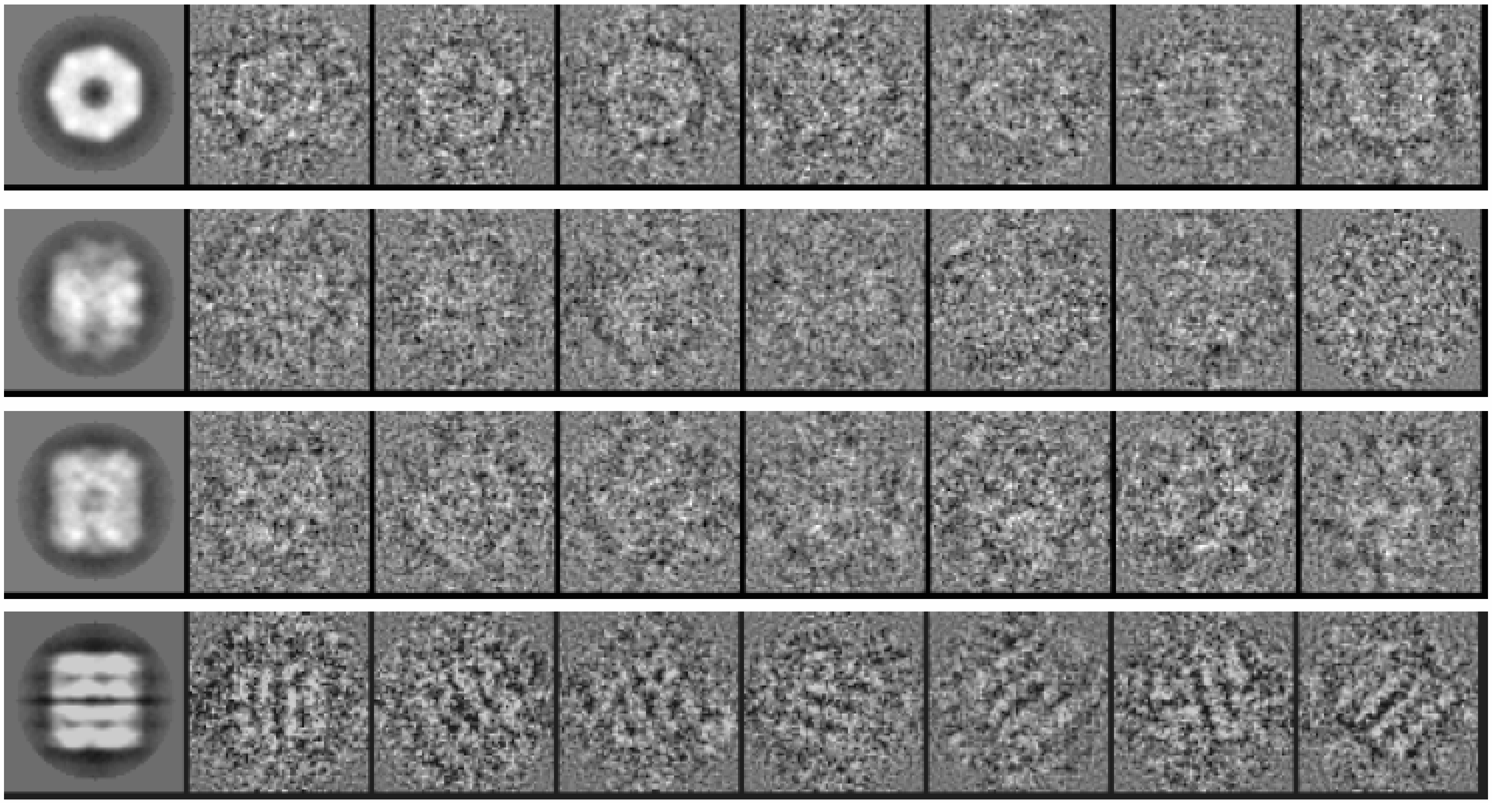
Preliminary Model



Projections



Classification



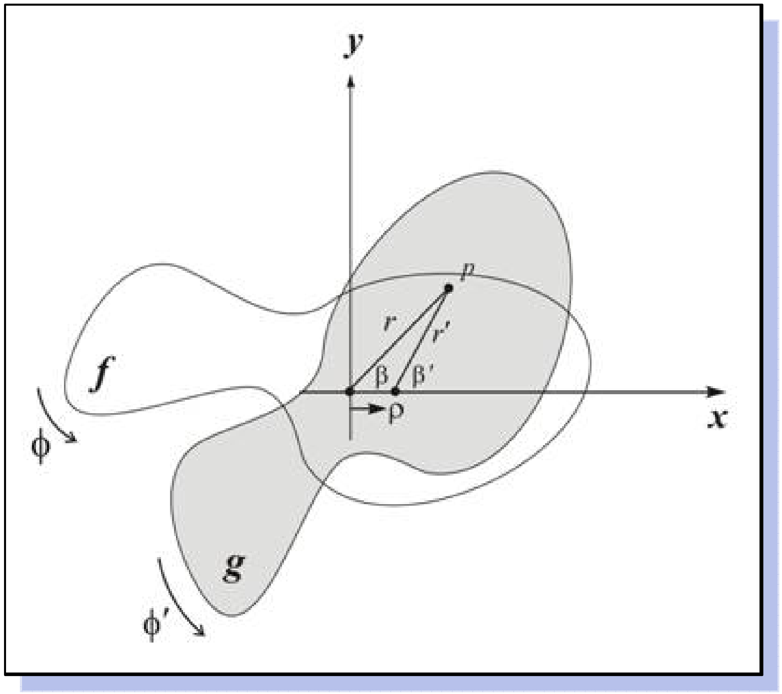
2D Alignment

Resampling the objects in Polar coordinate.

The given density objects are expanded in Fourier series:

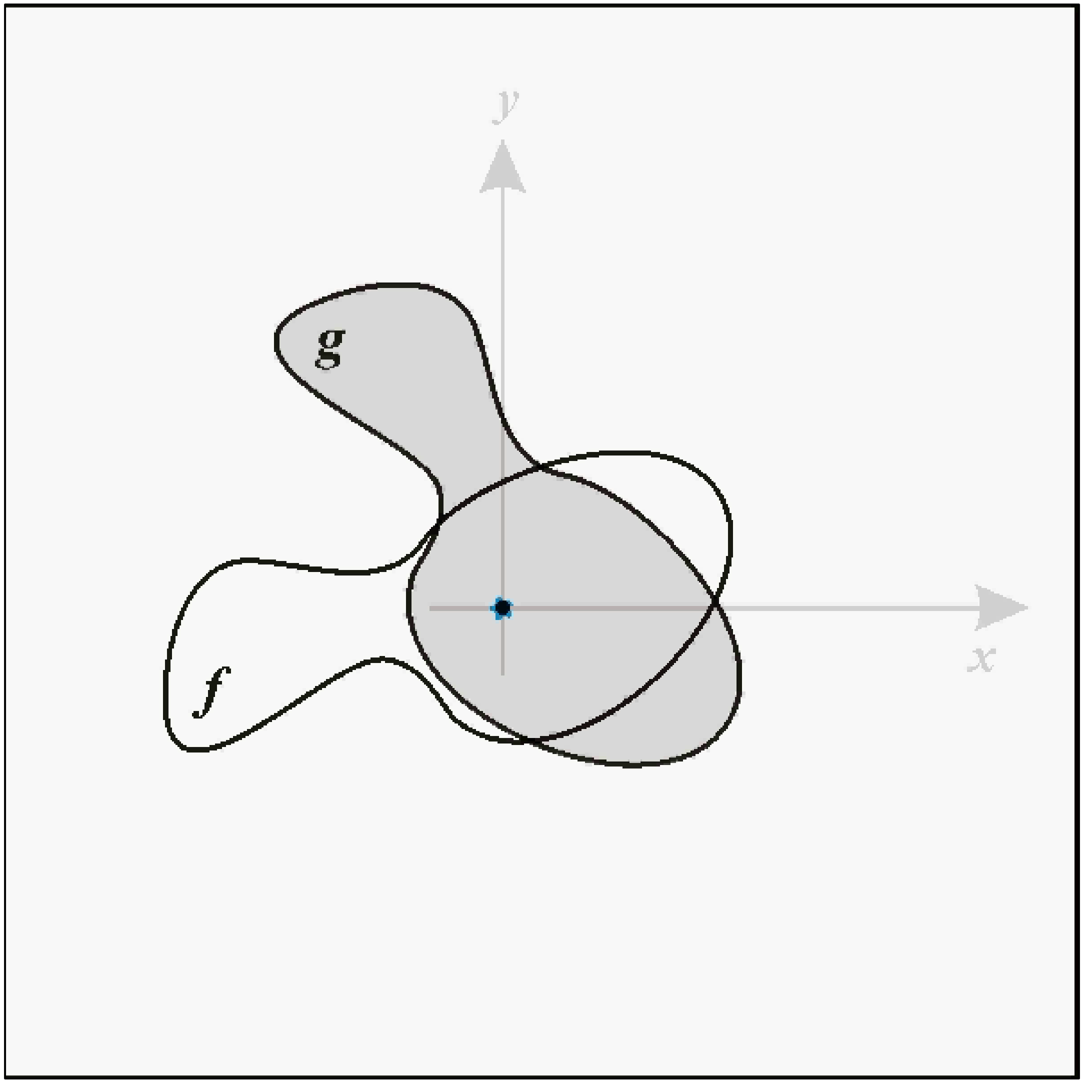
$$f(r,\beta) = \sum_m \hat{f}_m(r) e^{im\beta}$$

$$g(r,\beta) = \sum_n \hat{g}_n(r) e^{in\beta}$$



Idea: Rotate both objects, while translate one object along the positive x axis, until match is found.

2D Alignment



2D Alignment

The correlation function is a function of 2 rotations and 1 distance: $c(\phi,\phi';\rho) = \int_{\mathbf{R}^2} f(\phi)\cdot g(\phi';\rho)$

Here:

$$f(\phi)(r,\beta) = \sum_m \hat{f}_m(r) e^{im(\beta-\phi)}$$

$$g(\phi';\rho)(r,\beta) = \sum_n \hat{g}_n(r') e^{in(\beta'-\phi')}$$

The correlation function becomes:

$$c(\phi,\phi';\rho) = \sum_{m,n} e^{i(m\phi+n\phi')} I_{mn}(\rho)$$

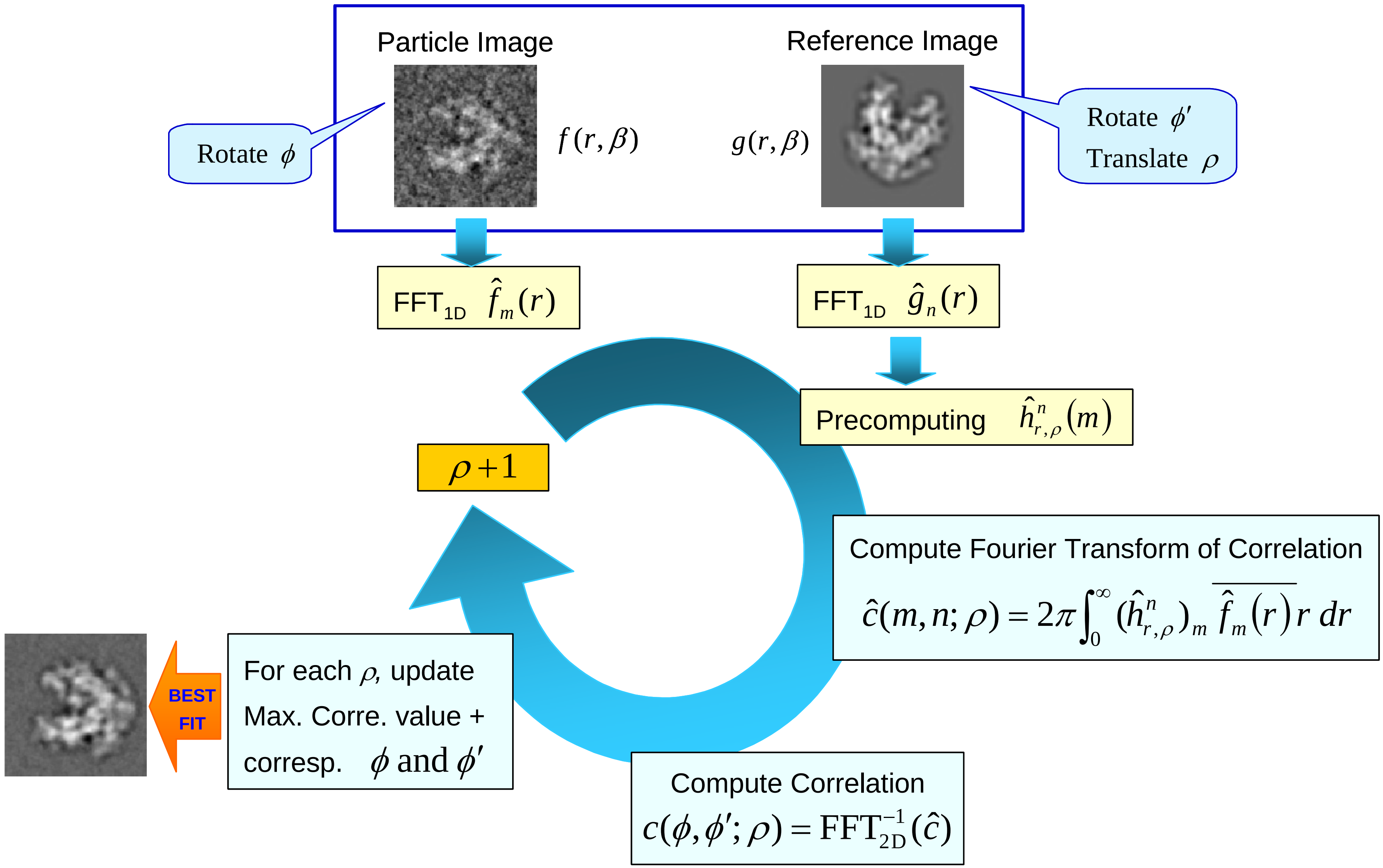
Where:

$$I_{mn}(\rho) = \int_0^\infty \left[\int_0^{2\pi} e^{-im\beta} \underbrace{\left(e^{-in\beta'} \overline{\hat{g}_n(r')} \right)}_{h^n_{r,\rho}(\beta)} d\beta \right] \cdot \overline{\hat{f}_m(r)} \cdot r dr$$

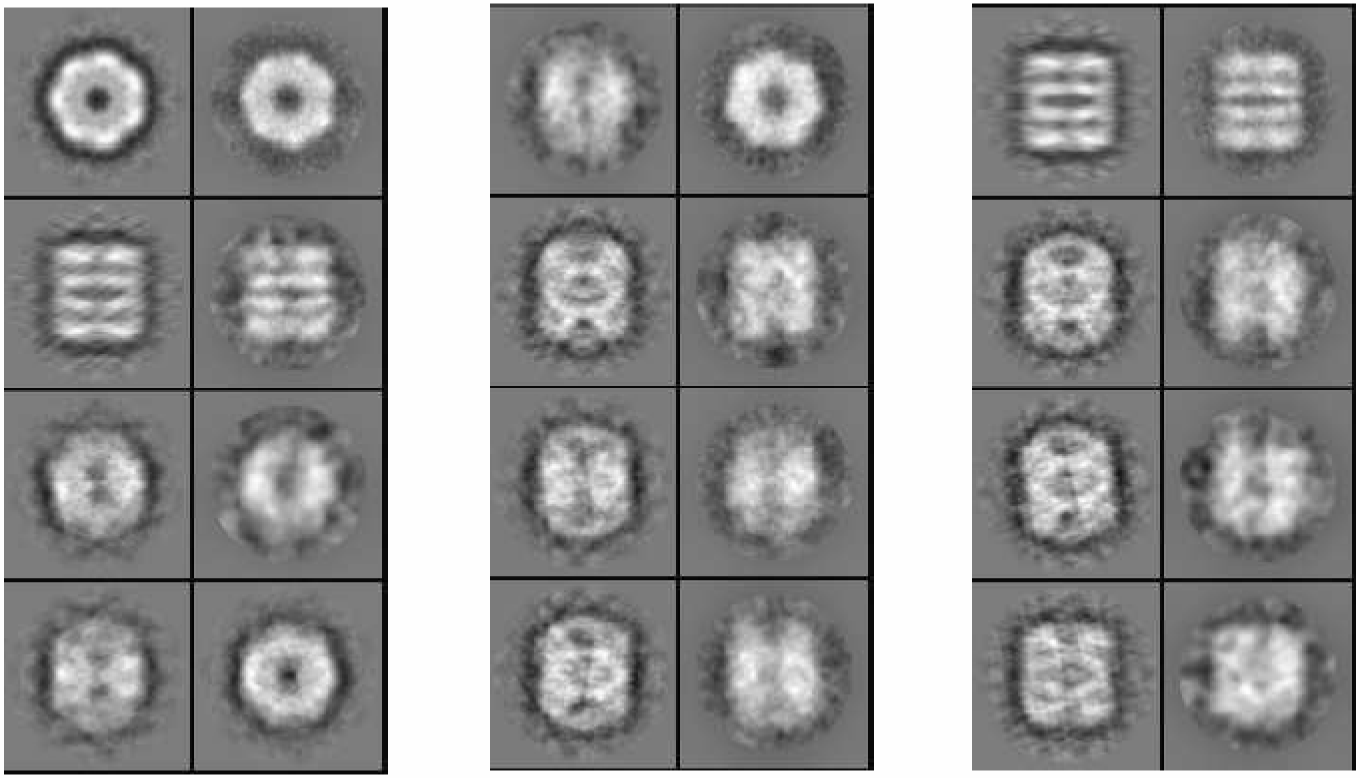
The 2D Fourier transform of correlation function:

$$\hat{c}(m,n;\rho) = I_{mn}(\rho) = 2\pi \int_0^\infty (\hat{h}^n_{r,\rho})_m \overline{\hat{f}_m(r)} r\, dr$$

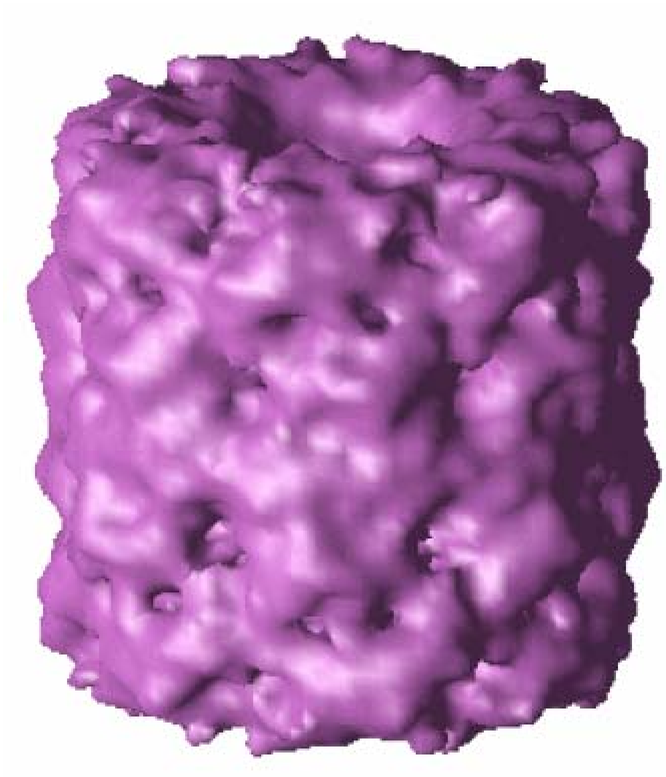
2D Alignment



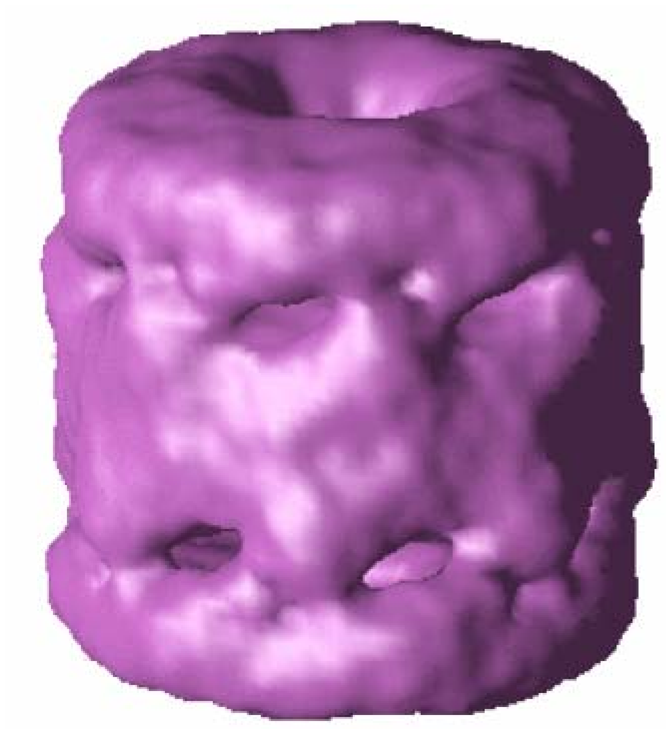
Class Averages



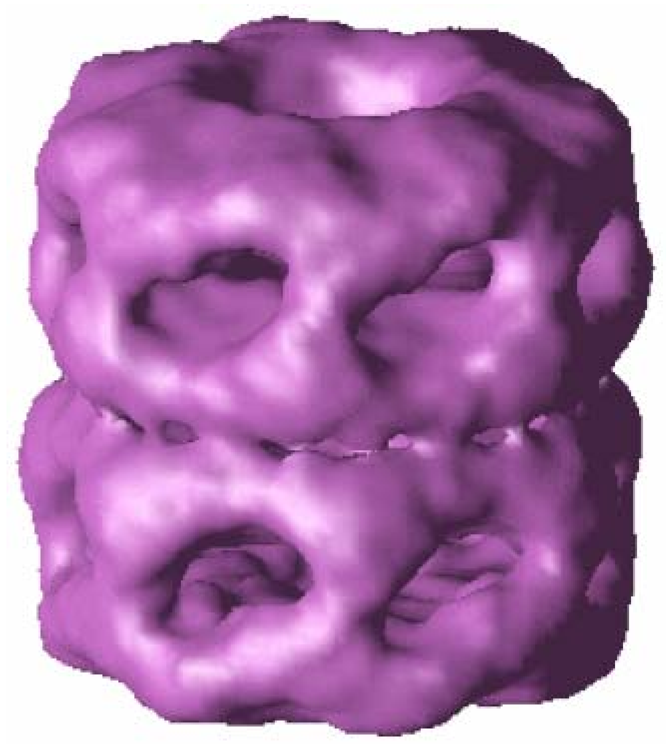
Iteration 1



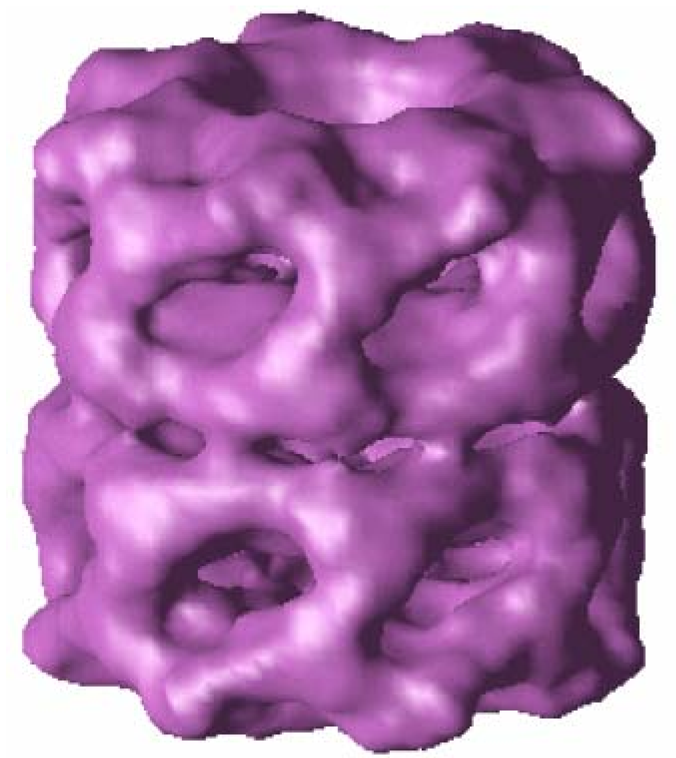
Iteration 2



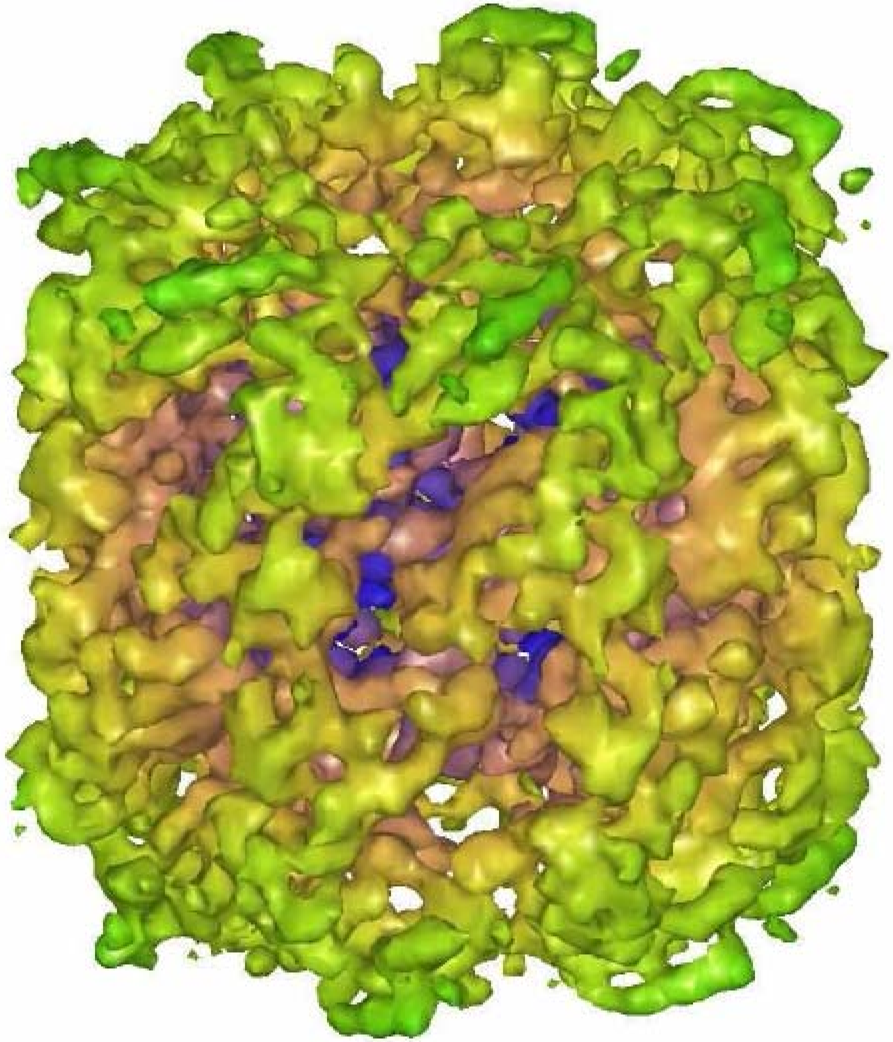
Iteration 3



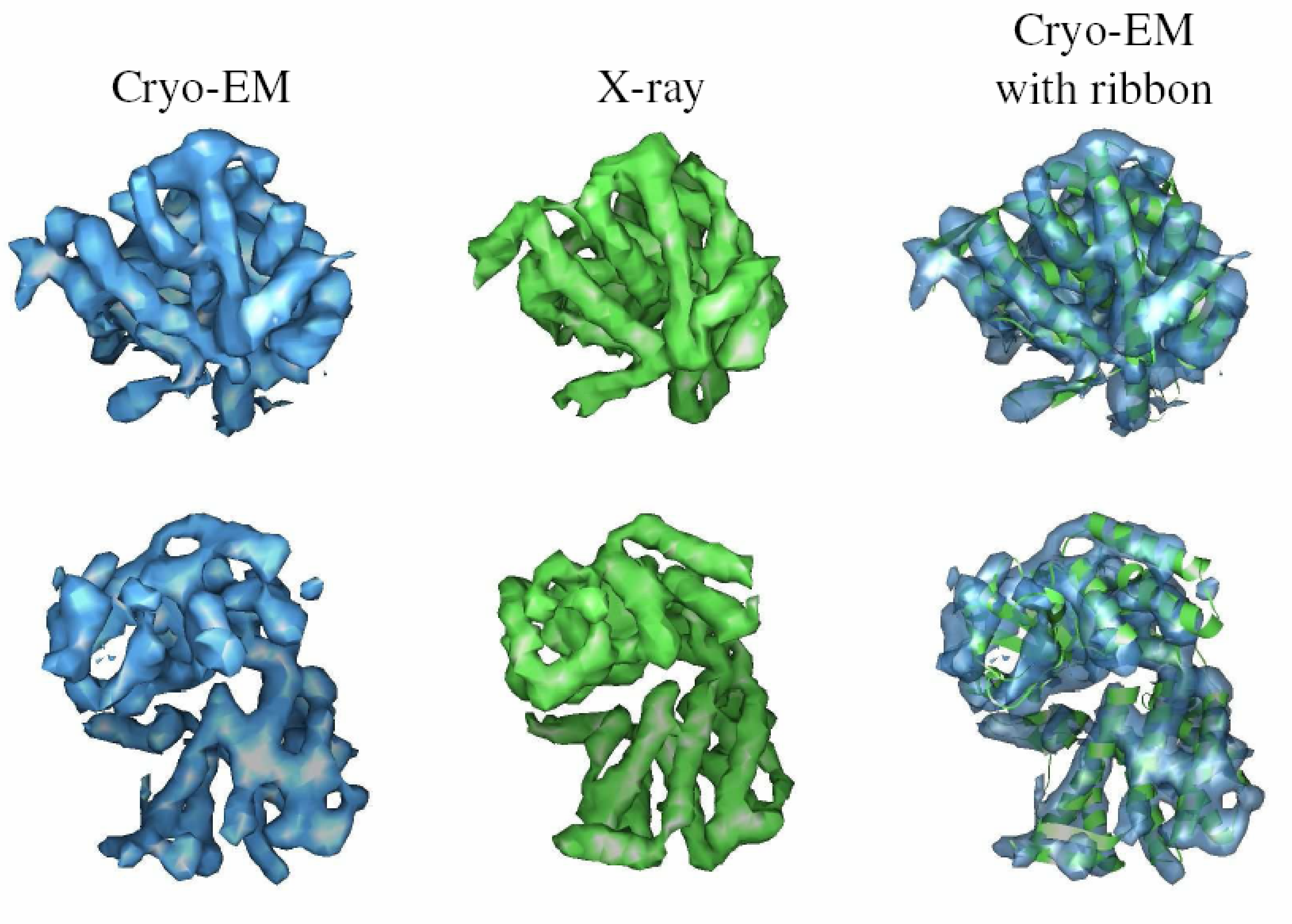
Iteration 4



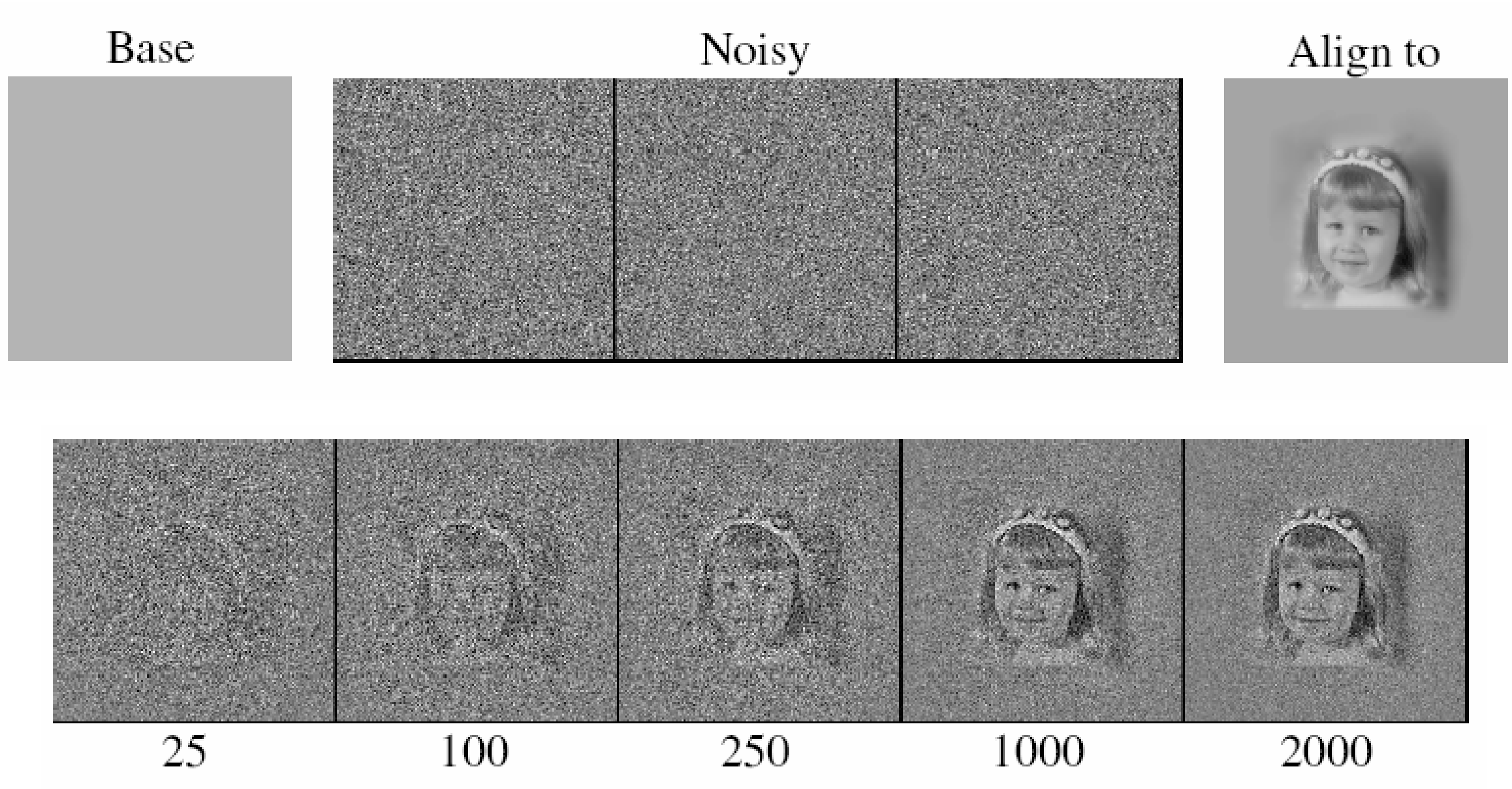
GroEL Reconstruction at 6.5Å



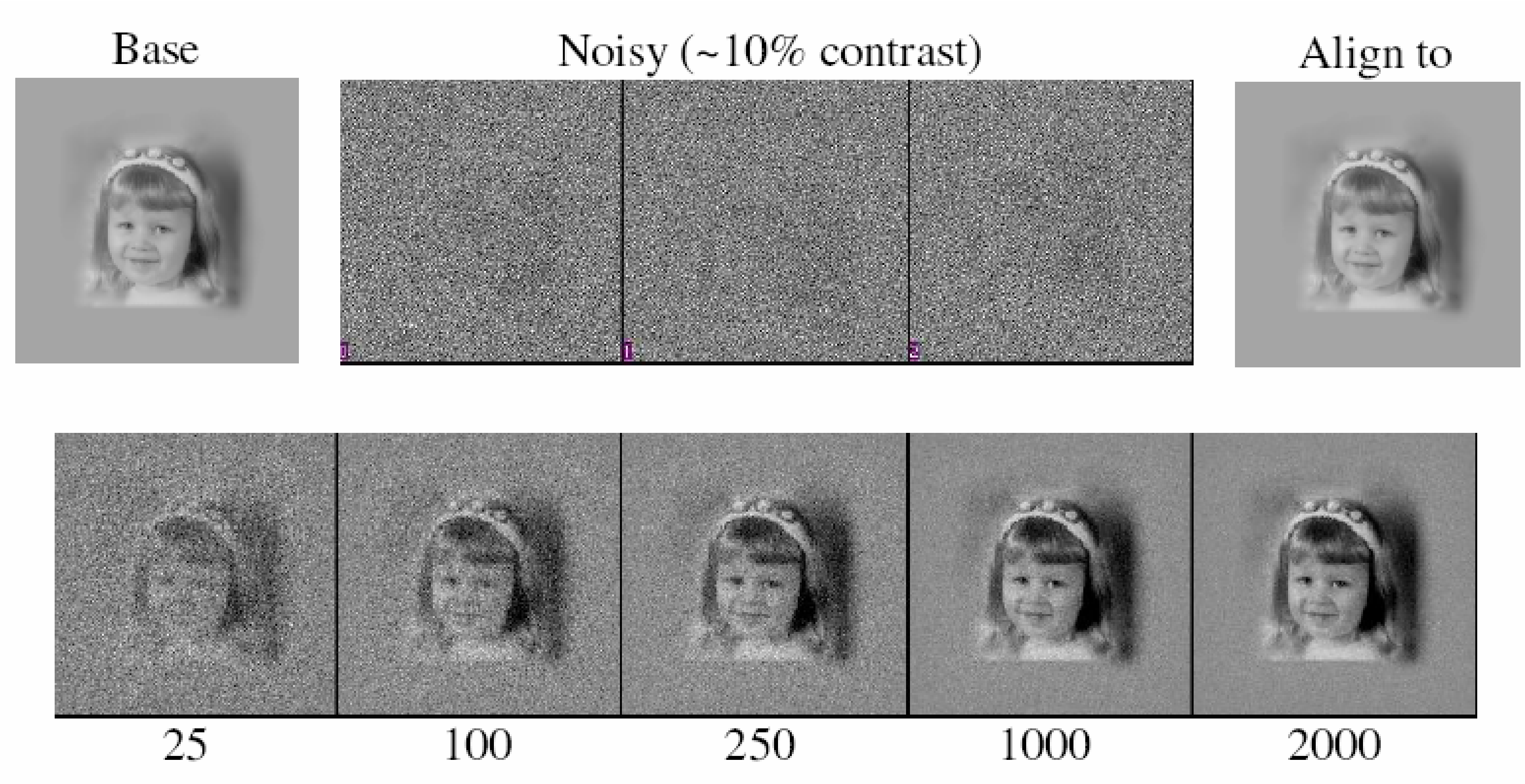
Comparison with Xtal Structure



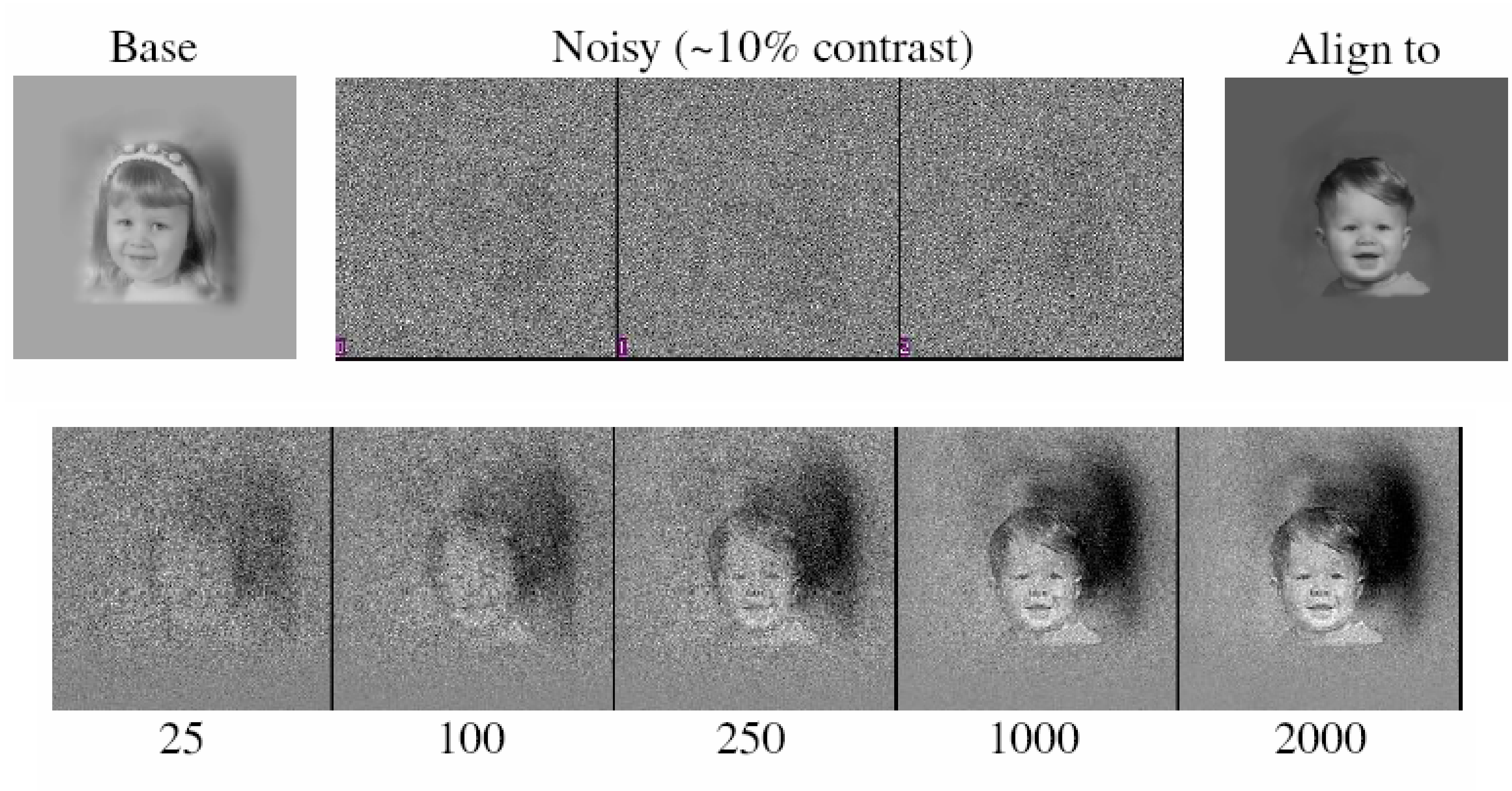
Caveat: Model Bias



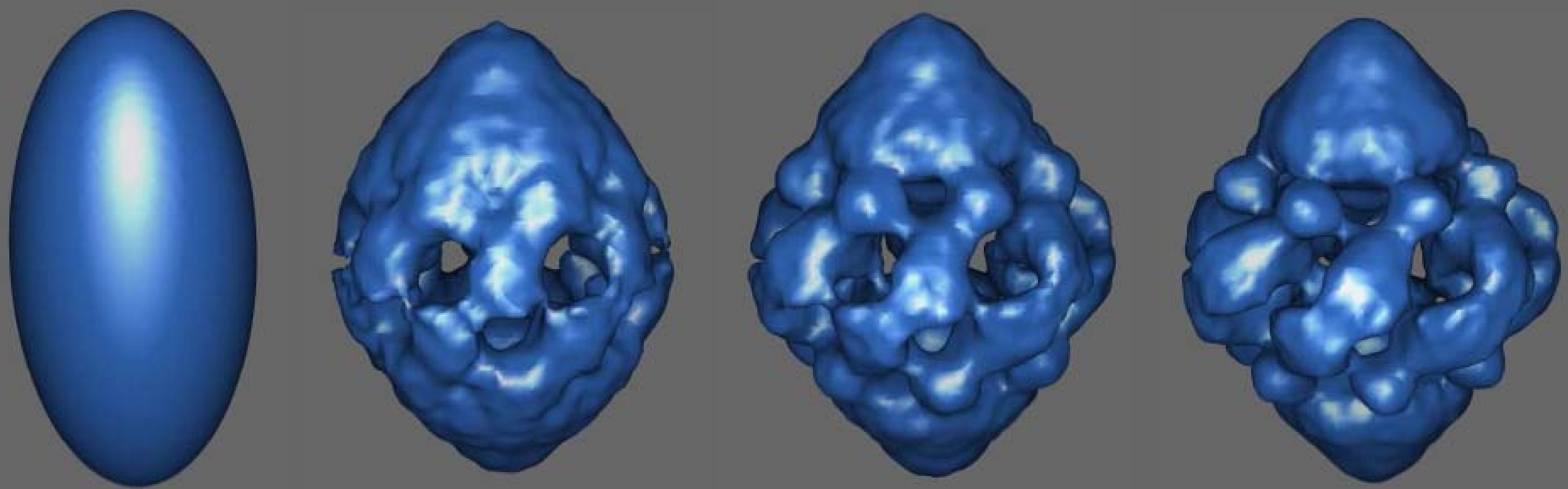
Caveat: Model Bias



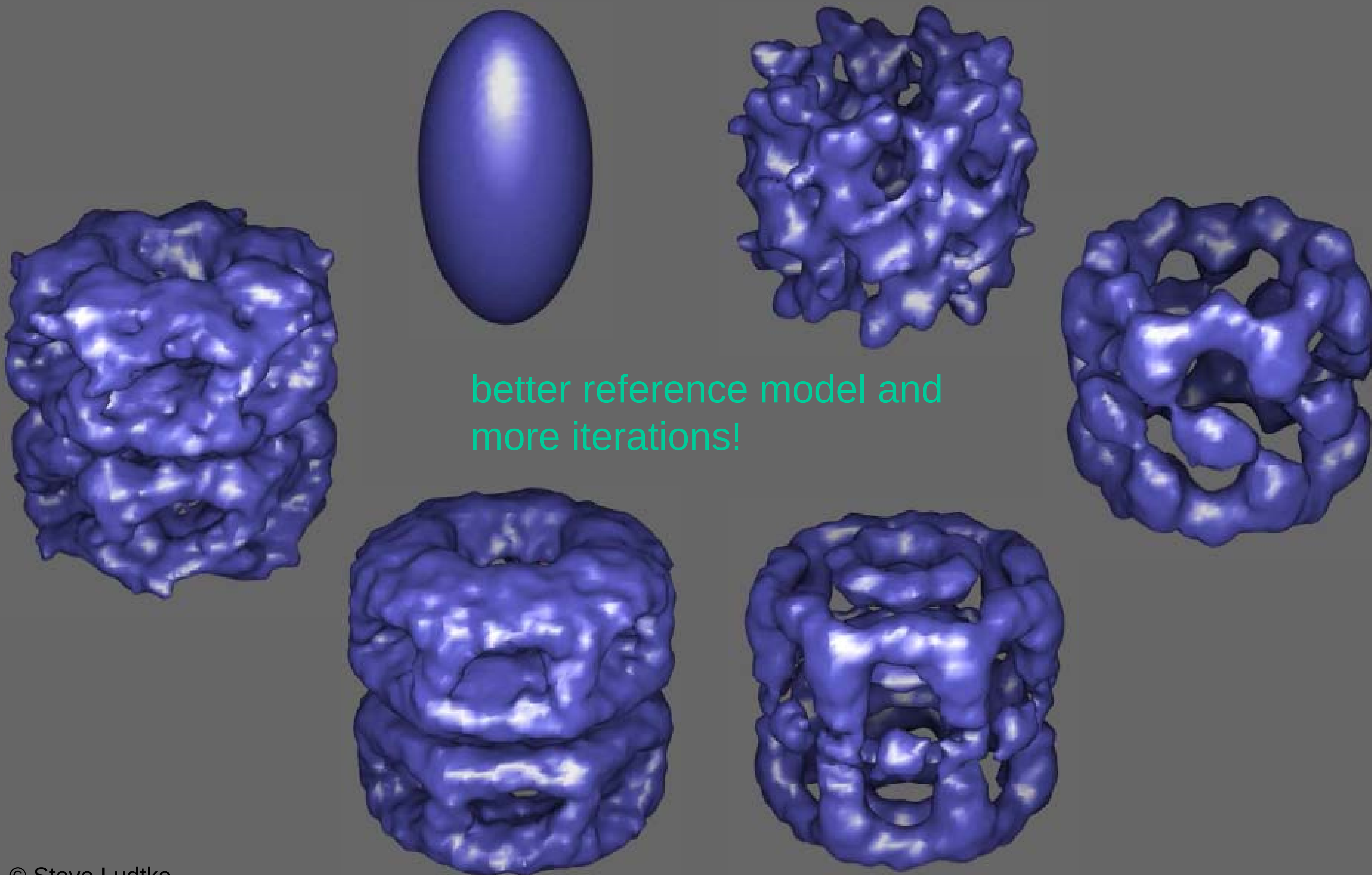
Caveat: Model Bias



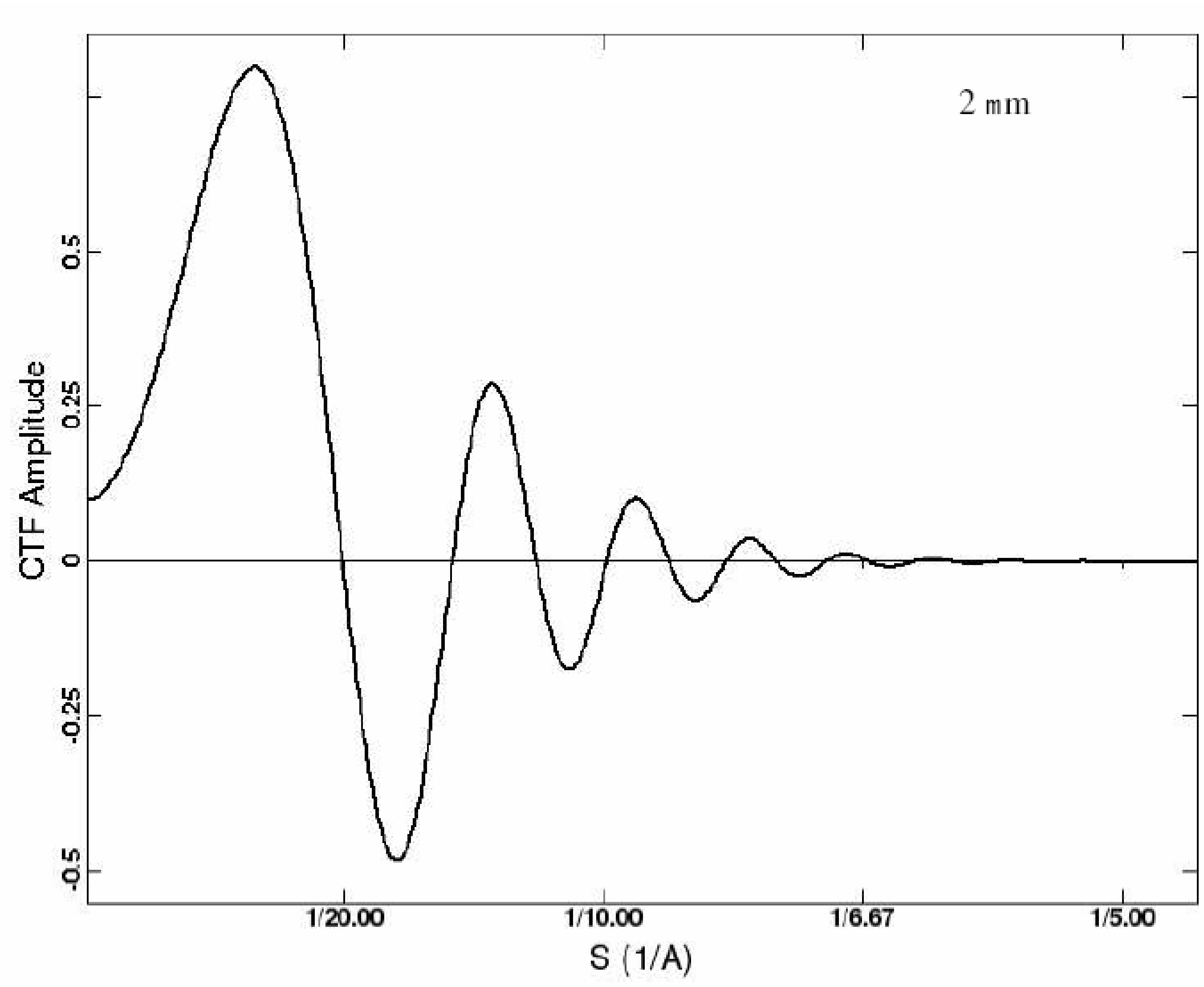
Caveat: Model Bias



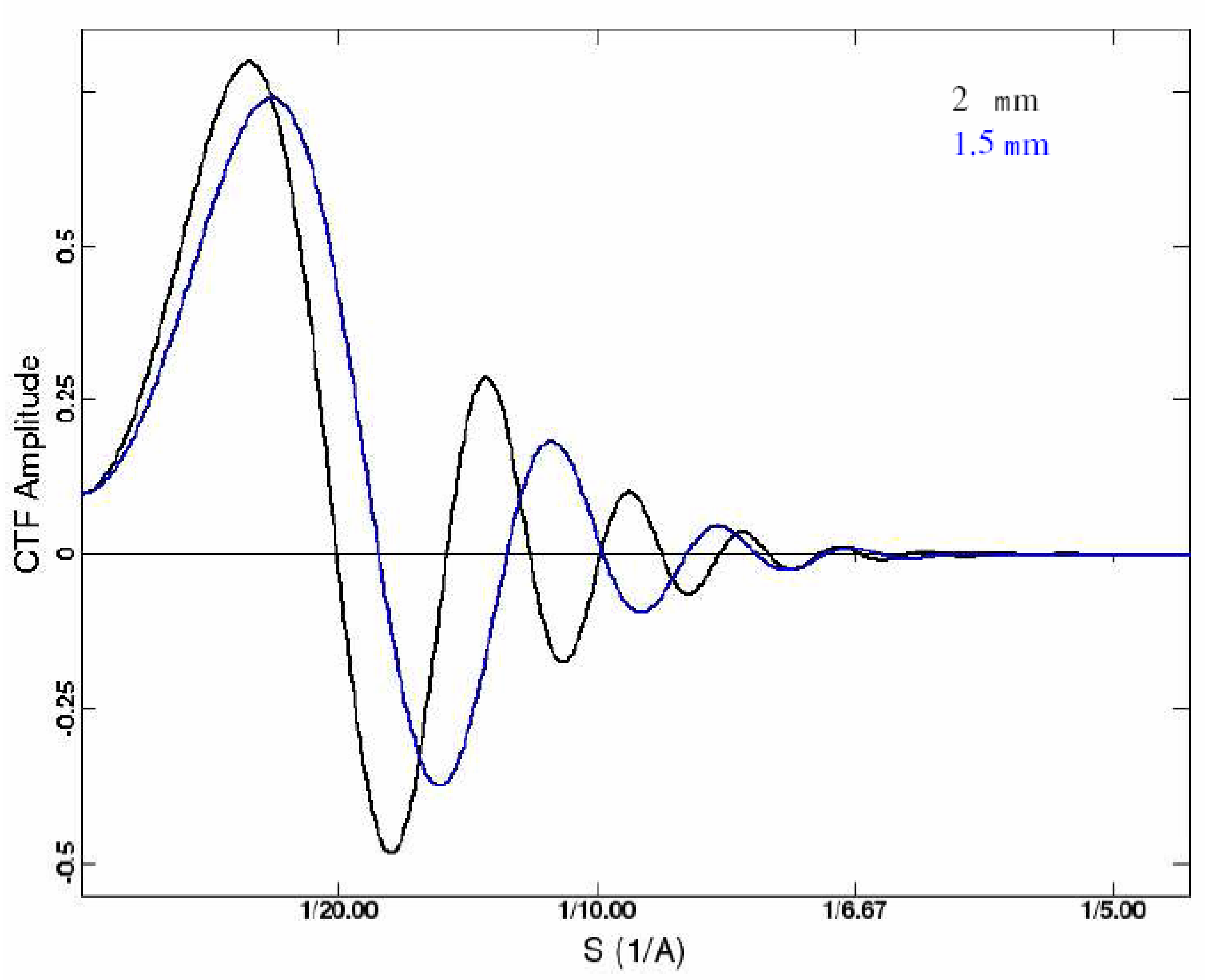
Caveat: Model Bias



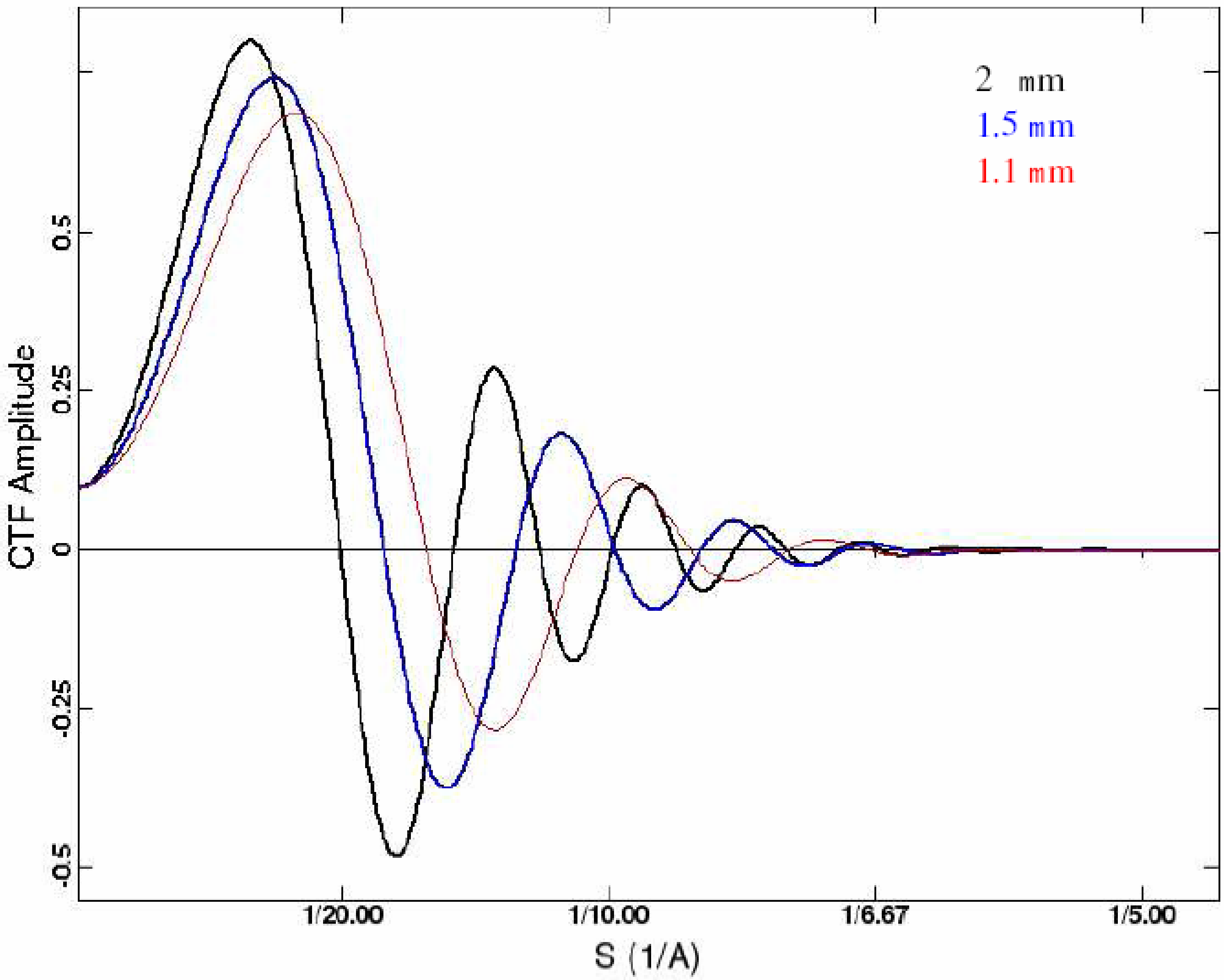
CTF Correction



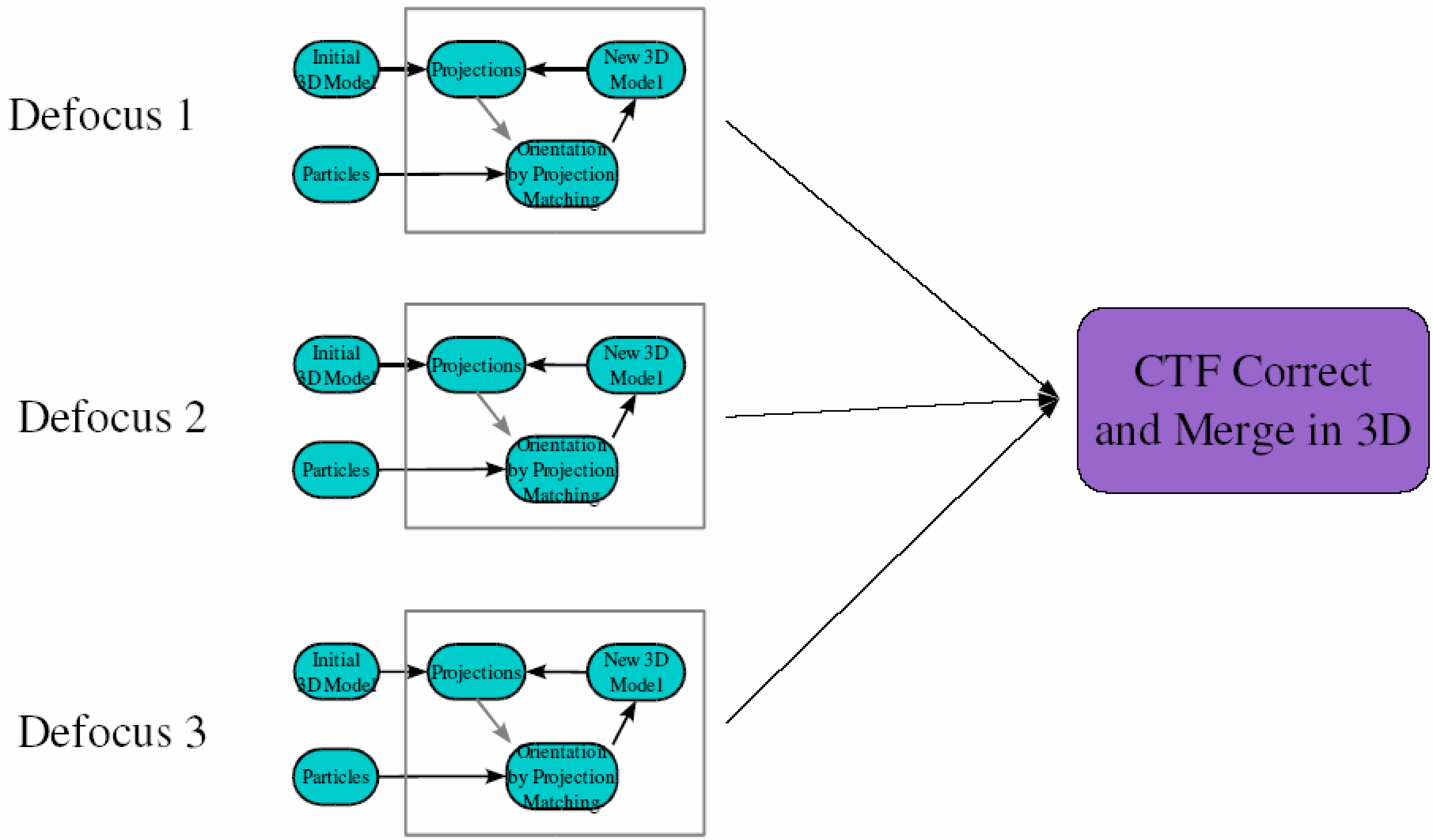
CTF Correction



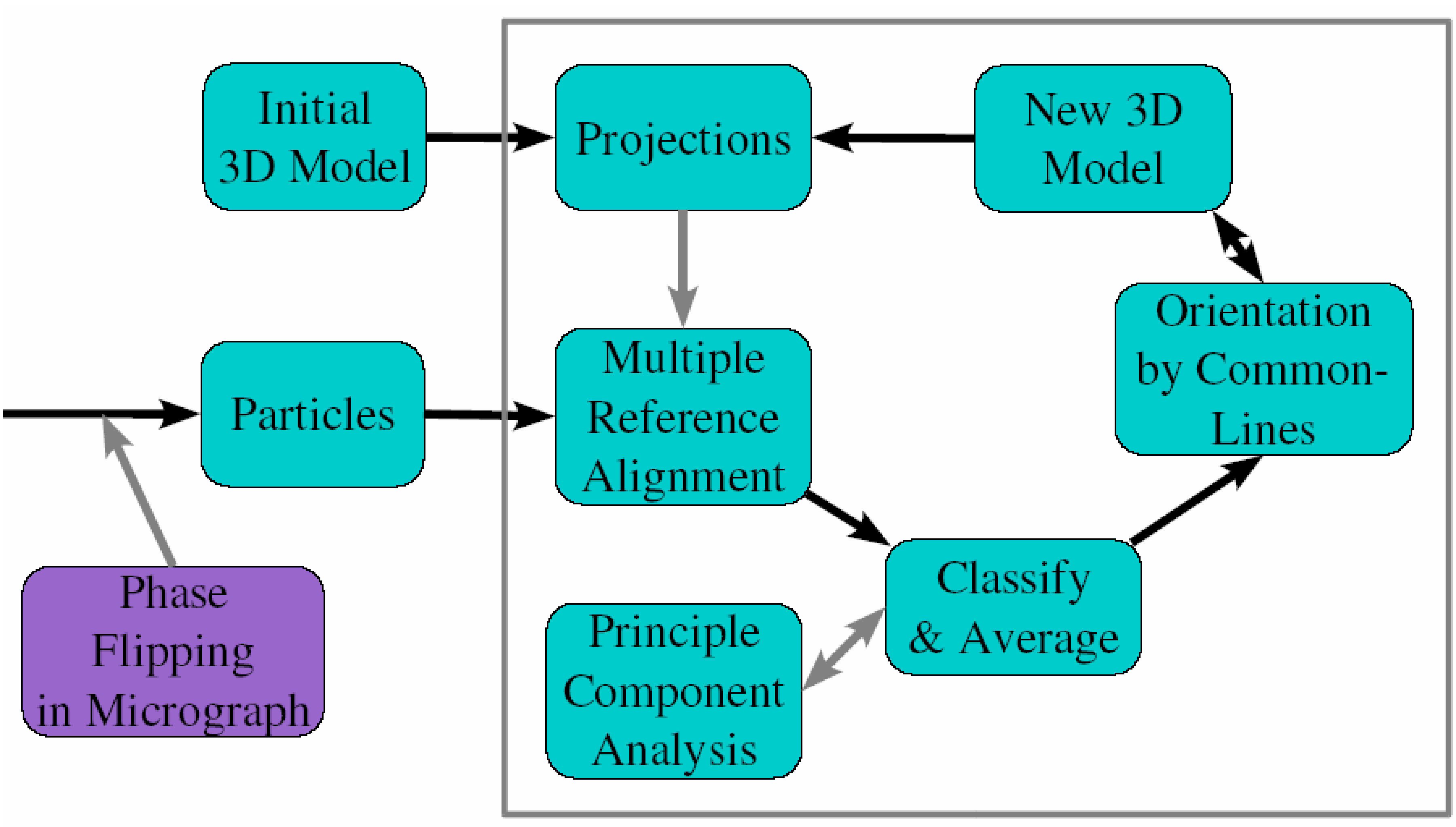
CTF Correction



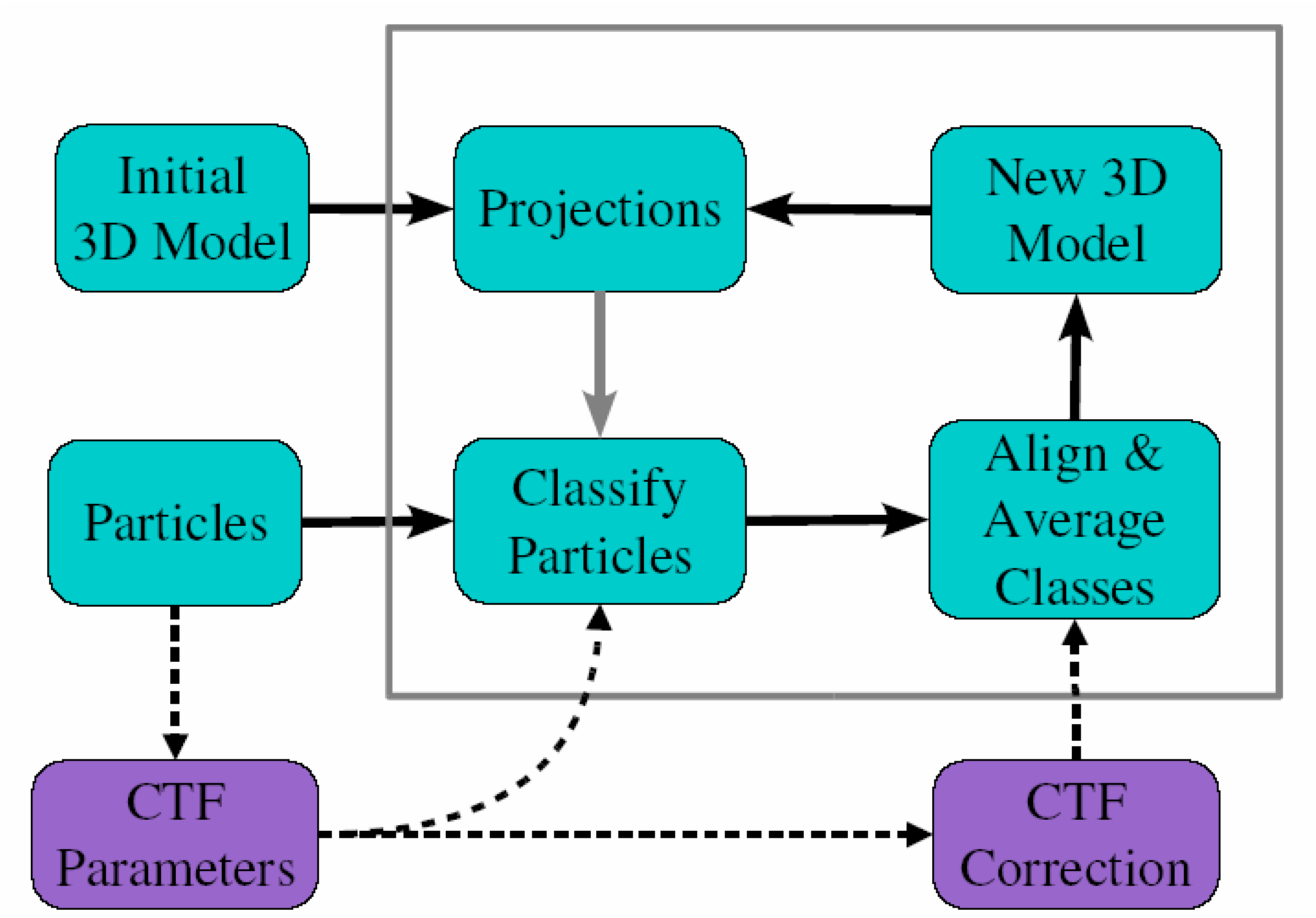
CTF Correction - SPIDER



CTF Correction - IMAGIC



CTF Correction - EMAN



Resources

Textbook:
Chapters 3,4,5 in: Joachim Frank, Three-Dimensional Electron Microscopy of Macromolecular Assemblies (1996, Academic Press)