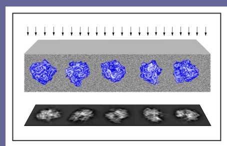
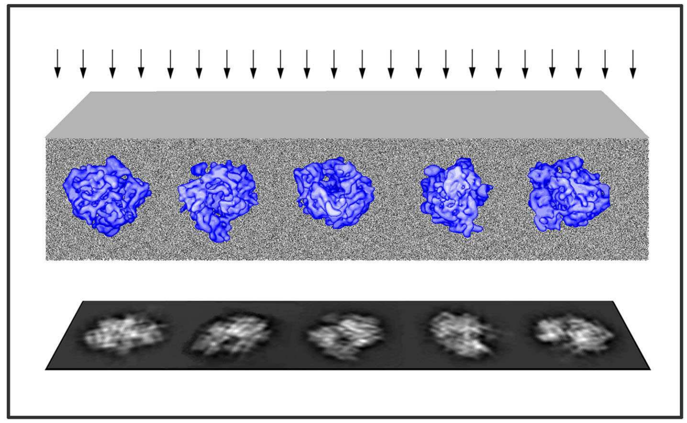
Single-Particle Reconstruction



Joachim Frank

Howard Hughes Medical Institute, Health Research, Inc., Wadsworth Center, Albany, New York & Department of Biomedical Sciences, State University of New York at Albany

Supported by HHMI, NIH/NCRR, NIH R01 GM55440, and NIH R37 GM29169



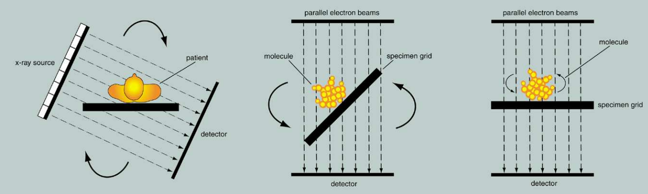
Single-particle reconstruction
Main initial assumptions:

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.

Data collection geometries for 3D reconstruction

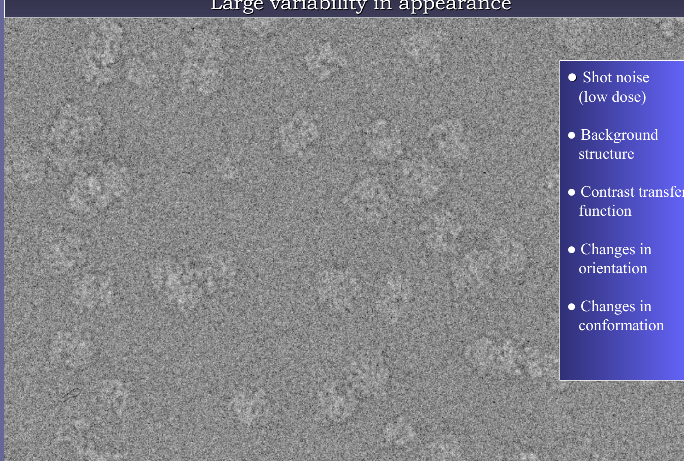


CAT - scan
- beam rotating
- patient stationary

Electron Tomography
- molecule rotating
- beam stationary

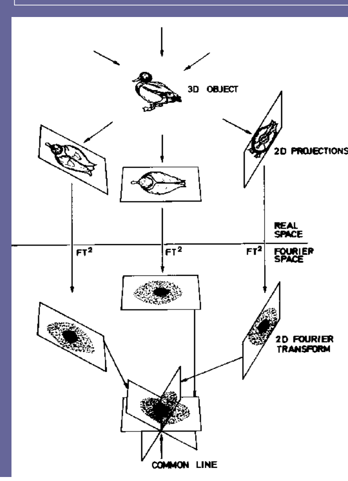
Single particle reconstruction
- molecule "rotating"
- beam stationary

Electron Micrographs of Single Molecules:
Large variability in appearance



- Shot noise (low dose)
- Background structure
- Contrast transfer function
- Changes in orientation
- Changes in conformation

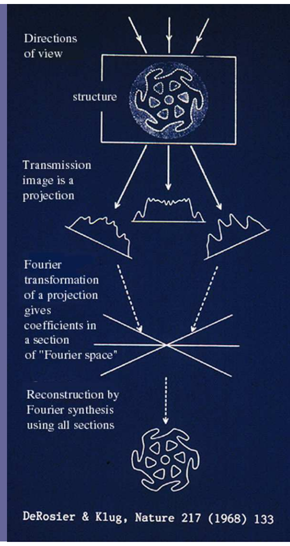
Projection Theorem



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

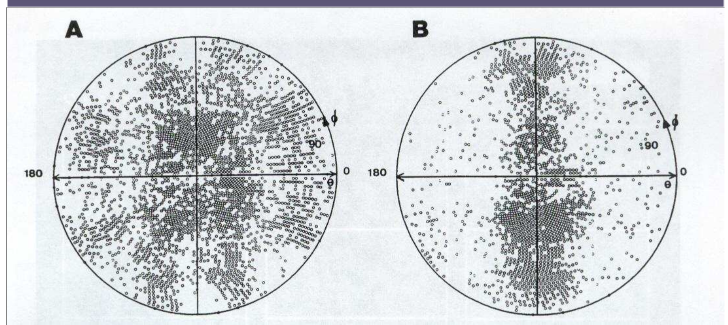
The Projection Theorem

(from the pioneering paper by DeRosier and Klug)



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 batch of micrographs
3) Determine defoci, and define defocus groups
4) Pick particles
5) Determine particle orientation
6) 3D reconstruction by defocus groups
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
     - (iii) Self-organized maps (SOMs)
8) Determine resolution
   - (a) phase residual
   - (b) Fourier shell correlation
   - (c) Spectral signal-to-noise ratio (SSNR)
12) Low-pass filtration
13) Amplitude correction (filter tailored acc. to experimental data)

Definition of the cross-correlation function (CCF)

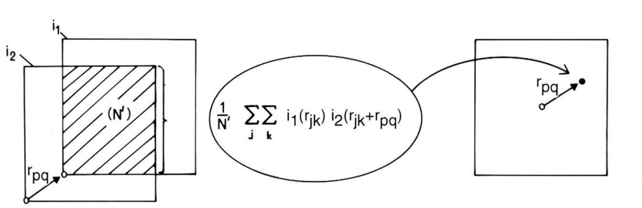


Fig. 3.8. Definition of the cross-correlation function. Image 1 is shifted with respect to image 2 by vector $\mathbf{r}_{pq}$. In this shifted position, the scalar product of the two images arrays is formed and put into the CCF matrix at position (p, q). The vector $\mathbf{r}_{pq}$ is now allowed to 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

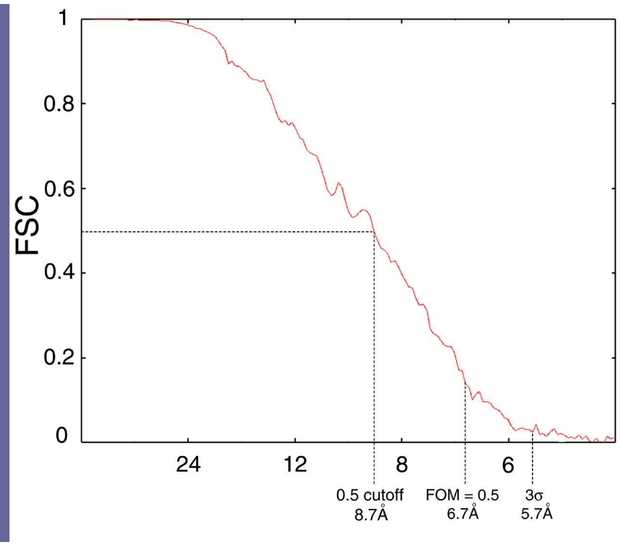
"Reference free" iterative alignment (Penczek *et al*., 1992) **:**
Two images are randomly picked, aligned, and added.
Then, a third image is aligned and added to the previous two. The process is repeated until all images are aligned.

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

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

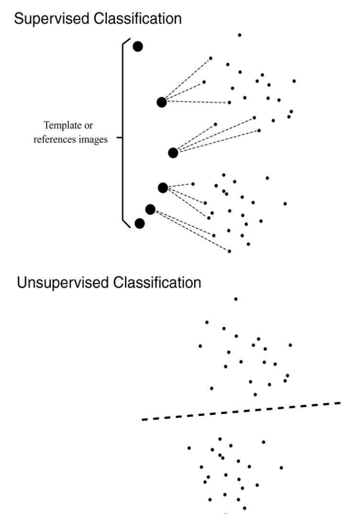


Classification

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

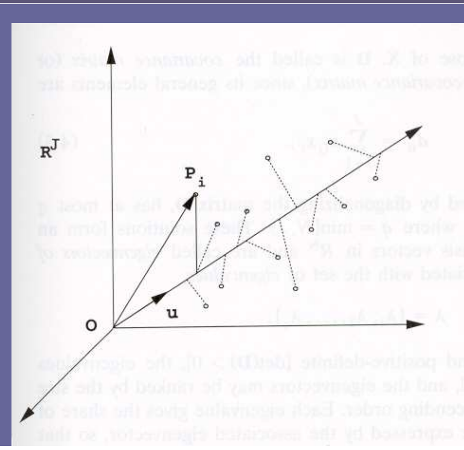
(folks, we are in Hilbert space)



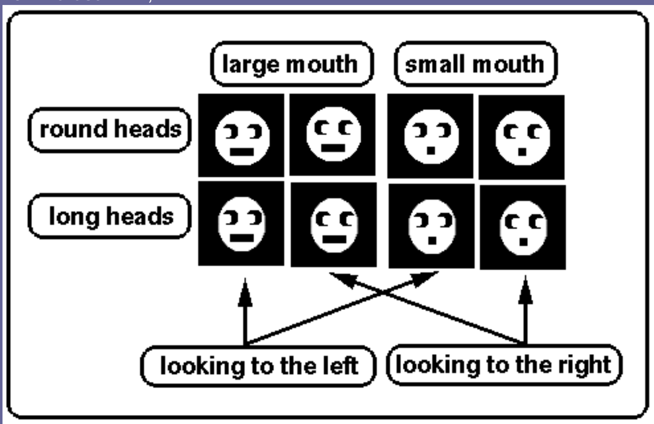
Classification, and the Role of MSA

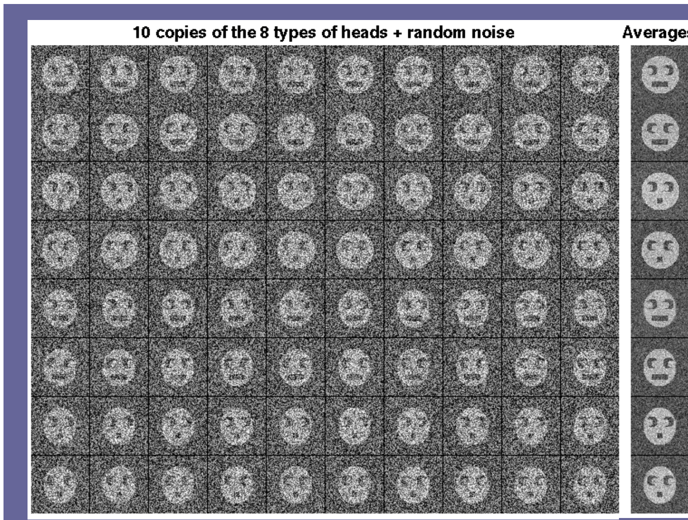
- Classification deals with "objects" in the space in which they are represented.
- For instance, a 64x64 image is an "object" in a 4096-dimensional space since, in principle, each of its pixels can vary independently.
  Let's say we have 8000 such images. They would form a cloud with 8000 points in 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.
- Reasons 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 of the component).

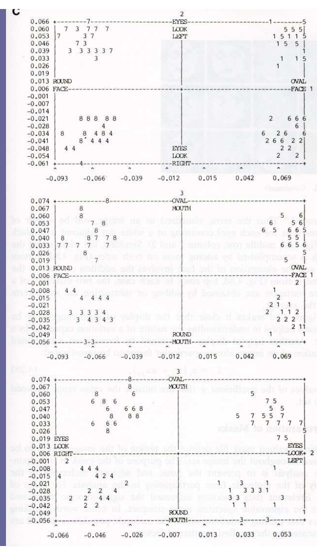
Principle of MSA:
Find new coordinate system, tailored to the data



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

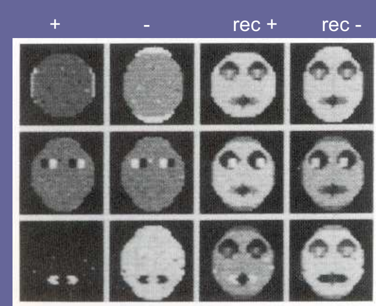






MSA: eigenimages

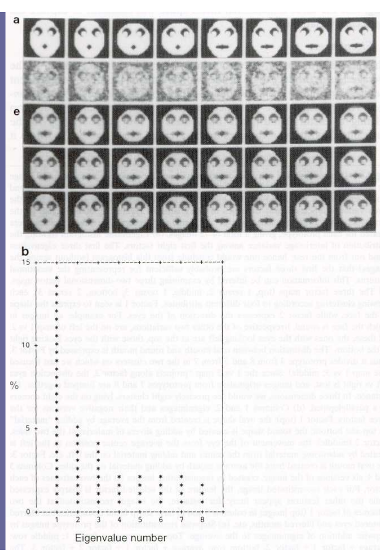
-
- Factor 1
- Factor 2
- Factor 3



Avrg + F1

Avrg + F1+F2

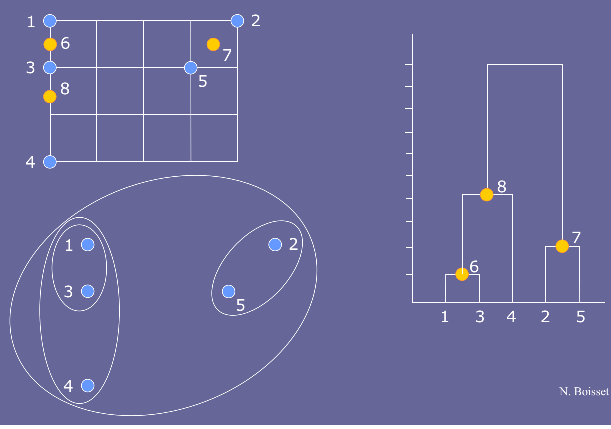
Avrg + F1+F2+F3



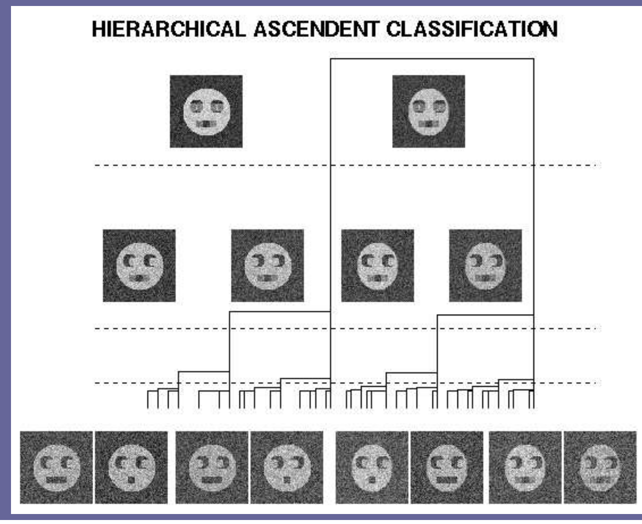
Unsupervised Classification

- Hierarchical ascendant classification (HAC): find links between objects, and group these hierarchically, in ascendant order.
- Partitional methods: divide objects into a given number of clusters. Example: K-means.
- Self-organized maps (SOMs): create a 2D similarity order among objects, by a process of “negotiation”, usually by means of a neural network.

Hierarchical Ascendant Classification

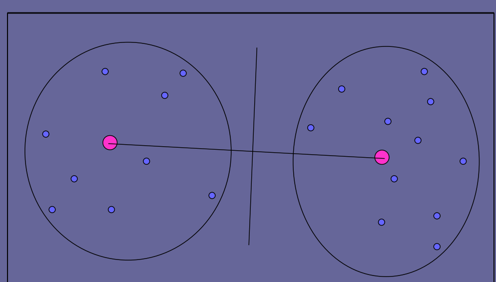


N. Boisset



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

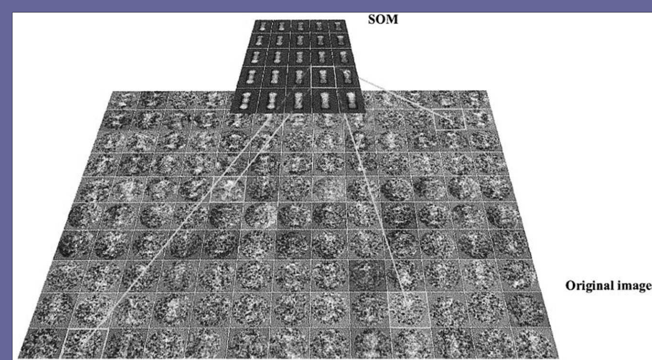
Diday E (1971) La methode des nuèes dynamiques. *Rev. Stat. Appl*. **19**, 19-34.



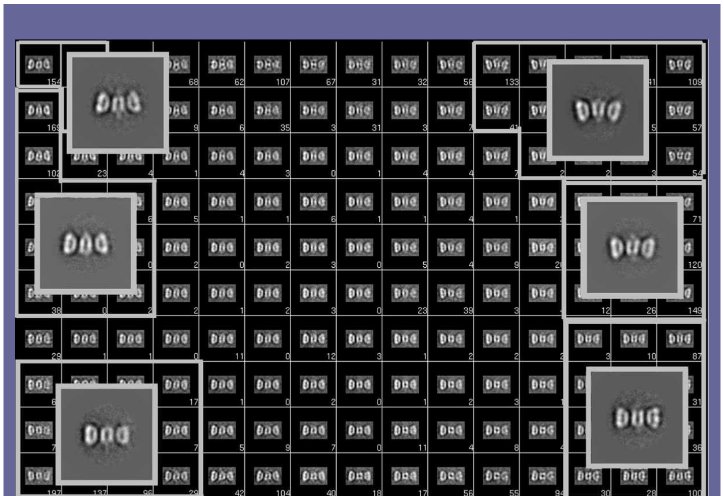
stops when centers don't move from one step to the next or after a given a selected number of iterations

N. Boisset

Self-Organized Maps



J.M. Carazo



J.M. Carazo

Overview: the necessary steps of a single-particle reconstruction

1) Optical diffraction: quality control, defocus inventory of micrograph batch
2) Scanning of batch of micrographs
3) Determine defoci, and define defocus groups
4) Pick particles
5) Determine particle orientation
6) 3D reconstruction by defocus groups
7) Angular refinement
8) CTF correction
9) Validation/determine resolution
10) Interpretation: segmentation, docking, etc.

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

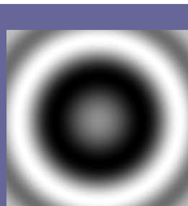
1) Optical diffraction: quality control, defocus inventory of micrograph batch
2) Scanning of micrograph batch *[I will skip both]*
3) Determine defoci, and define defocus groups
4) Pick particles
   (a) manual
   (b) automated
5) Determine particle orientation
   (a) unknown structure -- bootstrap
      (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

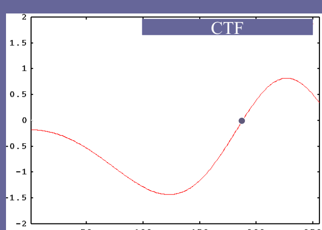

original object

x


CTF for Δz = 0.400 µm

=

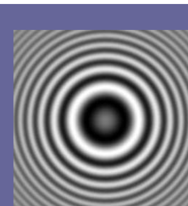

cryo-EM image



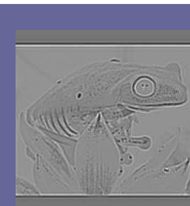

cryo-EM image, contrast-inverted

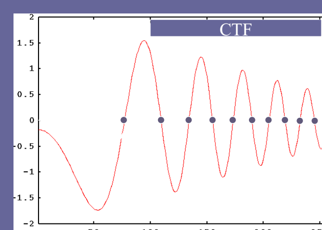

original object

x


CTF for Δz = 2.500 µm

=


cryo-EM image




cryo-EM image, contrast-inverted

N. Boisset

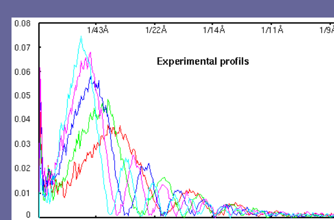
Strategy for reconstruction from multiple defocus groups

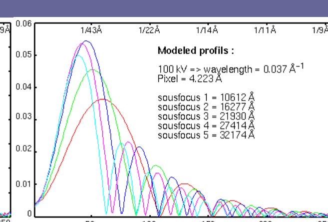
- **Coverage of large defocus range required**
- **Data collection must be geared toward covering range without major gap**
- **Characterizing all particles from the same micrograph by the same defocus is OK up to a resolution of ~1/8 A^{-1}. To get better resolution, one has to worry about different heights of the particle within the ice layer.**

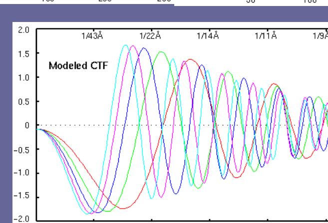
Sequence of steps:

1) Determine defocus for each micrograph
2) Define defocus groups, by creating supersets of particles from micrographs 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., by Wiener filtering.

CTF Determination







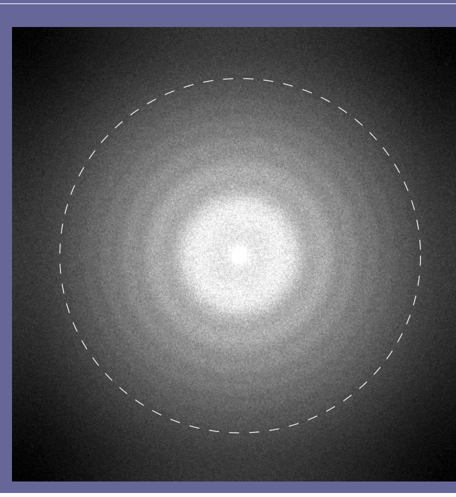
N. Boisset

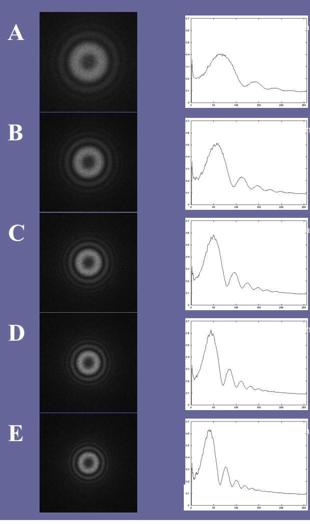
Computation of averaged power spectrum

For each micrograph …

1) Divide field into overlapping subfields of ~512 x 512
2) Compute FFT for each subfield
3) Compute $|F(k)|^2$ for each subfield
4) Form average over $|F(k)|^2$ 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

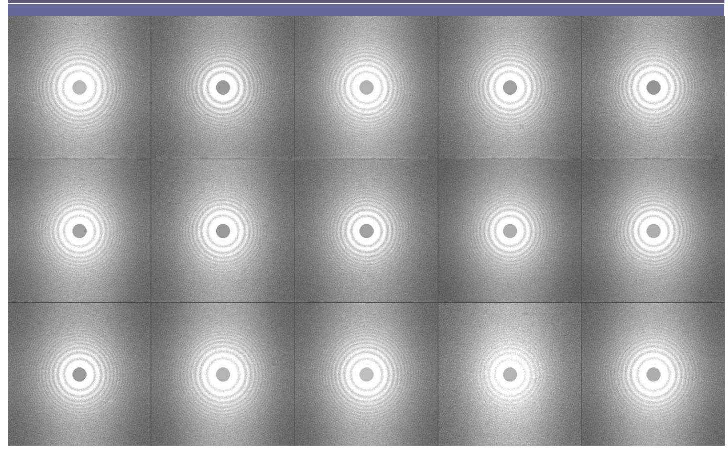
Band limit, or limit of useful information in Fourier space





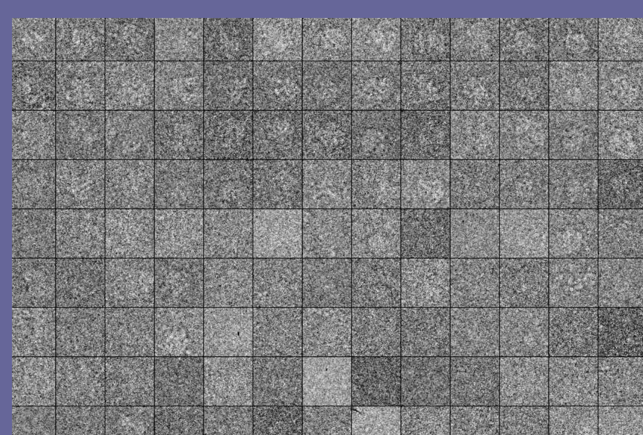
N. Boisset

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



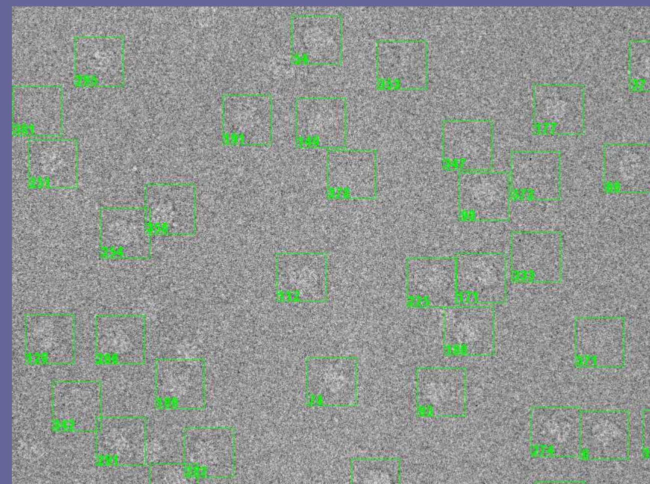
Automated particle picking, CCF-based, with local normalization

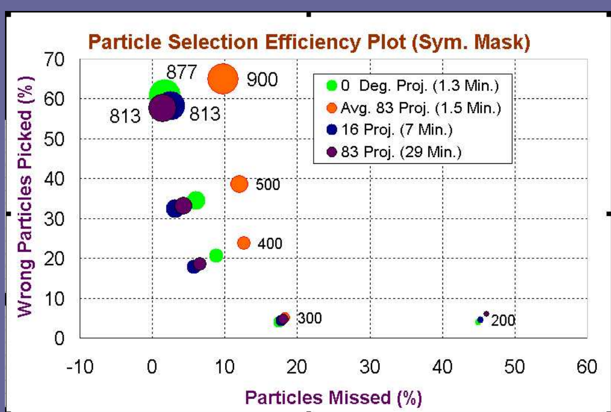
(i) Define a reference (e.g., by averaging projections over full Eulerian range);
(ii) Paste reference into array with size matching the size of the micrograph;
(iii) Compute CCF via FFT;
(iv) Compute locally varying variance of the micrograph via FFT (Roseman, 2003);

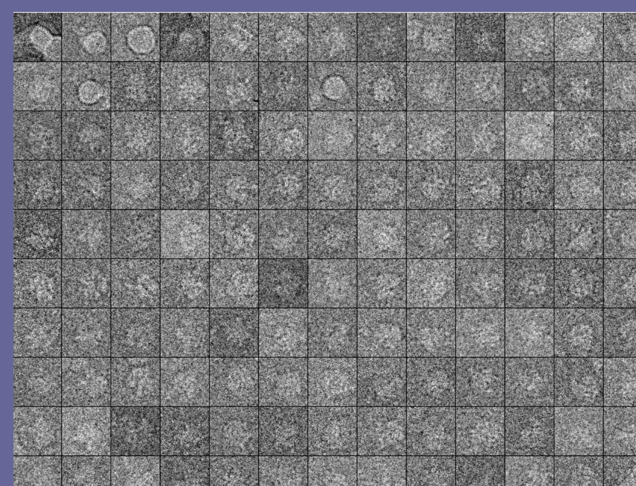
(v) ***"Local CCF" = CCF/local variance***

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

Advantage of local CCF: avoid problems from background variability, false positives







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

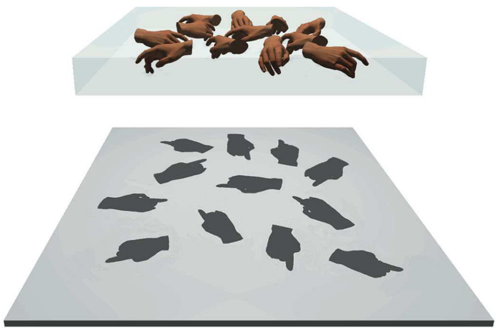
1) Optical diffraction: quality control, defocus inventory of micrograph batch
2) Scanning of micrograph batch
3) Determine defoci, and define defocus groups
4) Pick particles
   (a) manual
   (b) automated
5) Determine particle orientation
   (a) unknown structure -- bootstrap
      (i) random-conical (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

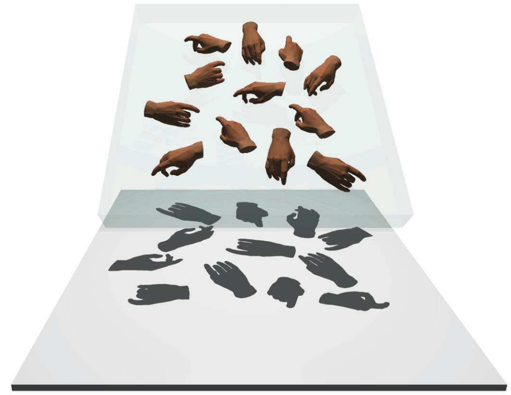
- Premise: all particles exhibit the same view (could be a subset, determined by classification)
- Take same field first at theta ~50 degrees, then at 0 degrees (in this order, to minimize dose)
- Display both fields side by side
- Pick each particle in both fields
- Align particles from 0-degree field

  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

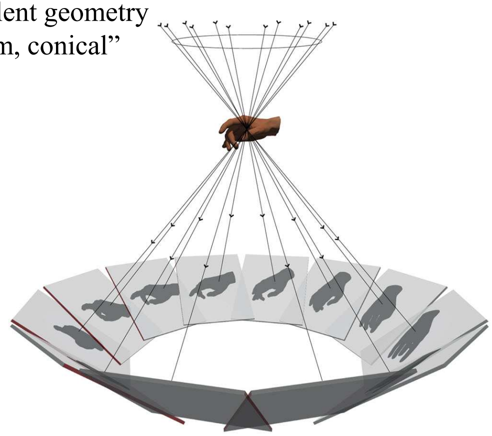
0-degree view



50-degree view



Equivalent geometry
“random, conical”

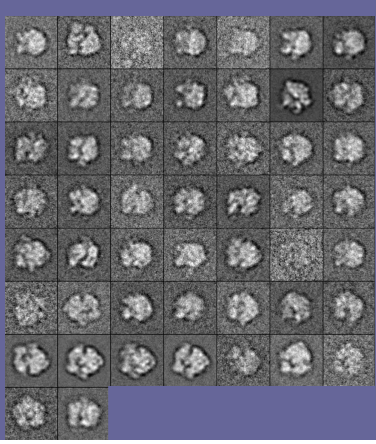


Random-conical reconstruction -- Problems to be solved:

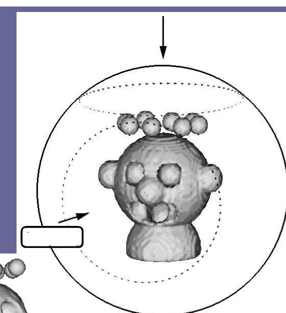
1) Find a subset (view class) of particles that lie in the same orientation on the grid
answer: unsupervised classification of 0-degree particles

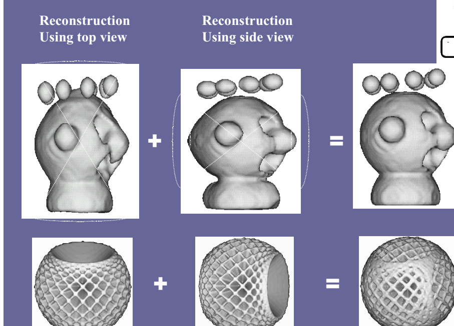
2) Missing-cone problem
answer: do several random conical reconstructions, each from a different subset (view class), find relative orientations, then make reconstruction from merged projections set.

Class averages determined by K-means



Missing-cone artifacts



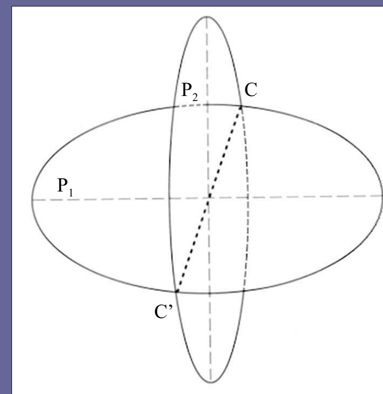


N. Boisset

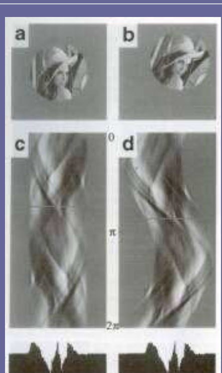
Overview: the necessary steps of a single-particle reconstruction -- I

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

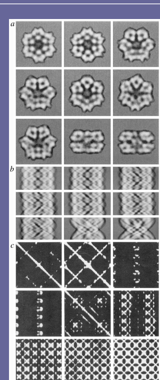
Common line C-C' of two projections represented by central sections P_1 and P_2



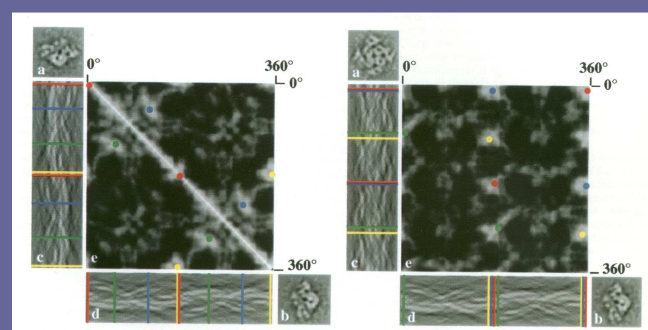
Use of sinogram/Radon transform



Lena


worm hemoglobin

Determination of relative orientations by common lines



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

Ryanodine receptor/calcium release channel

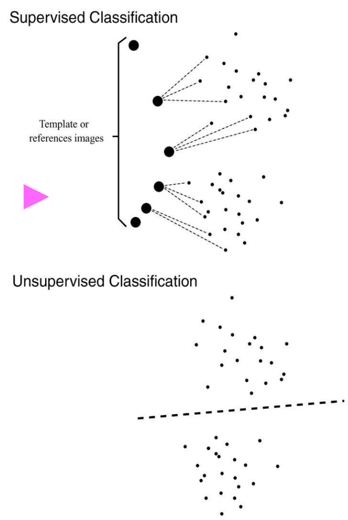
Common lines/angular reconstitution

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, 1987)
-- or -- simultaneously (Penczek et al., 1996)

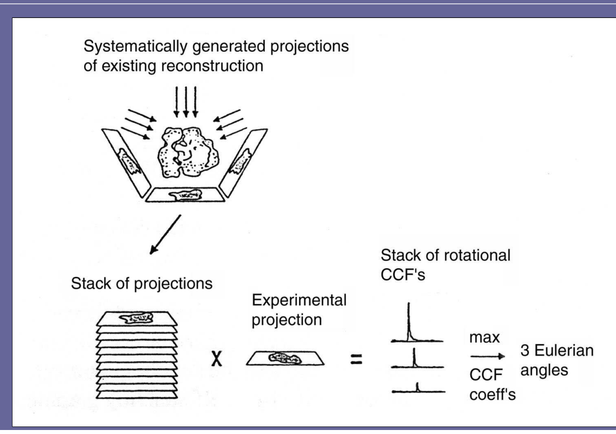
Issues:
- unaveraged images are too noisy – class averages must be used
- resolution loss due to implicit use of view range
- handedness not defined – tilt or prior knowledge needed

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

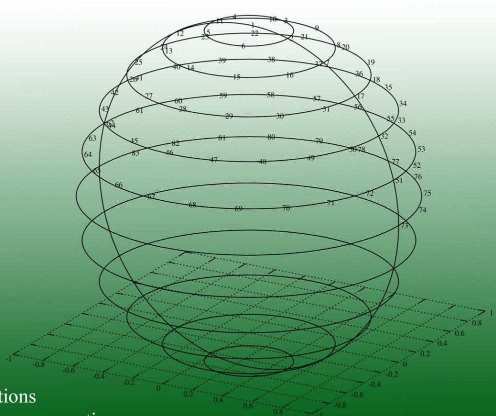
1) Optical diffraction: quality control, defocus inventory of micrograph batch
2) Scanning of micrograph batch
3) Determine defoci, and define defocus groups
4) Pick particles
    (a) manual
    (b) automated
5) Determine particle orientation
    (a) unknown structure -- bootstrap
        (i) random-conical (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



Orientation determination by reference to an existing reconstruction (supervised classification)

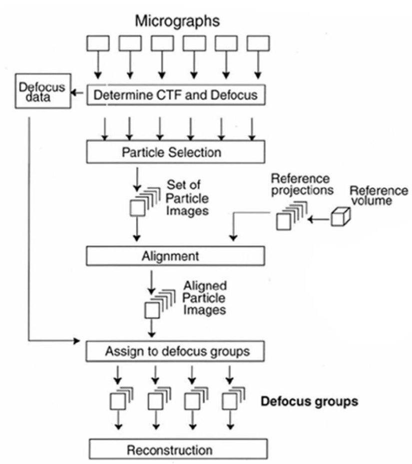


Initial Angular Grid



83 directions
~15 degrees separation

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

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

11) ***CTF correction***

12) ***Validation***

10) ***Interpretation:*** segmentation, docking, etc.

3D reconstruction by defocus group

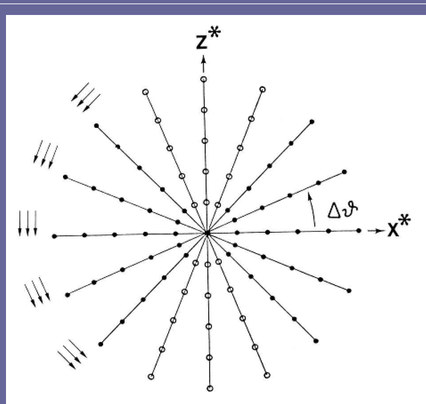
(a) Fourier interpolation
(b) Weighted back-projection
(c) Iterative algebraic reconstruction
(d) Conjugate gradient

1) 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.

2) Speed (high) versus accuracy (low).

3) Can be used in the beginning phases of a reconstruction project.

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

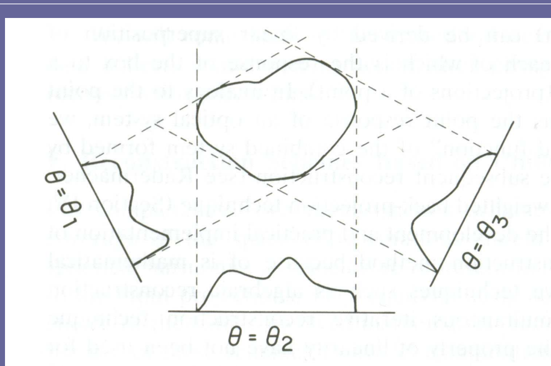
(1) Simple back-projection: Sum over “back-projection bodies”, each obtained by “smearing out” a projection in the viewing direction.

(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 related to angular gaps.

Principle of back-projection



3D reconstruction by defocus group

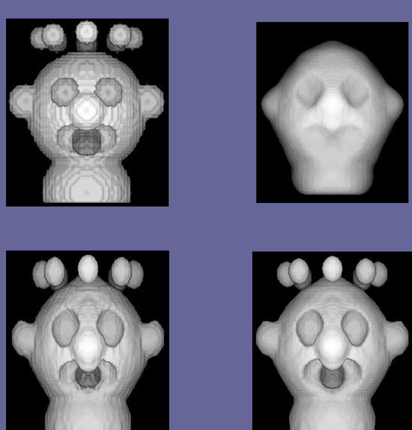
(a) Fourier interpolation
(b) Weighted back-projection
(c) Iterative algebraic reconstruction
(d) Conjugate gradient

1) 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.
2) The algorithm comes in many variants. It allows constraints to be easily implemented.
3) 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

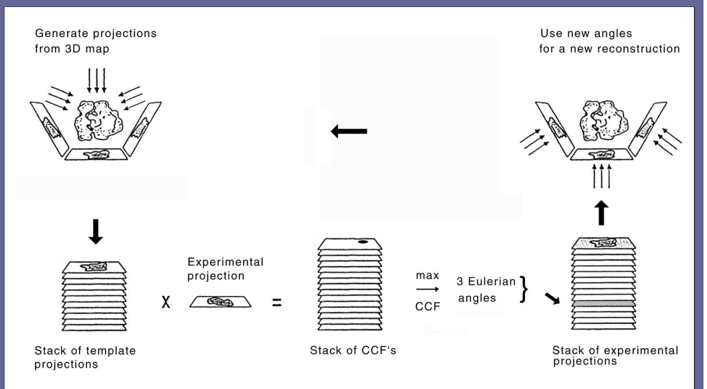
Iterative algebraic reconstruction

N. Boisset

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
7) *Refinement*
   - given an initial 3D reference, iterate the steps {3D projection matching + reconstruction}
   - beware of problem of reference-dependence
11) *CTF correction*
12) *Validation*
10) *Interpretation:* segmentation, docking, etc.

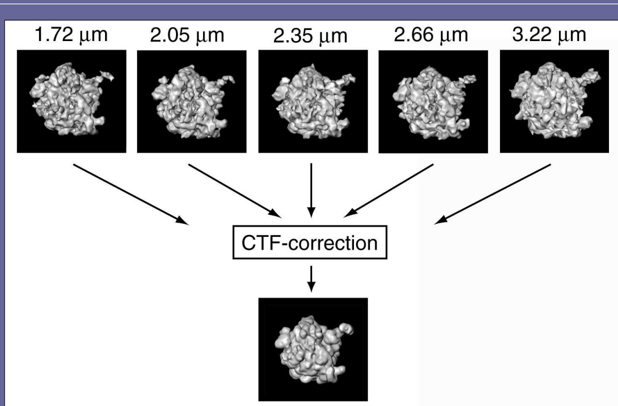
Angular Refinement, by Iterative 3D Projection Matching



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
7) ***Refinement***
•given an initial 3D reference,
iterate the steps {3D projection matching + reconstruction}
•beware of problem of reference-dependence
11) *CTF correction*
12) ***Validation***
10) ***Interpretation:*** segmentation, docking, etc.

CTF correction and merging of defocus group reconstructions by Wiener filtering

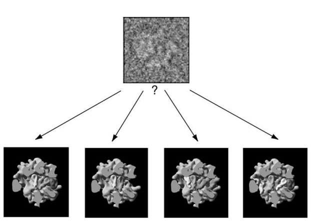


Reasons for limited resolution

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

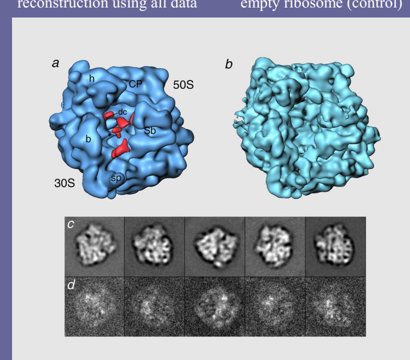
Conformational diversity: heterogeneous particle population

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

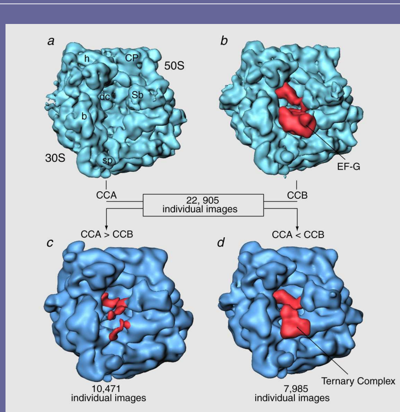
reconstruction using all data   empty ribosome (control)



averages

variance maps

Problem solved by supervised classification



Conclusions:

Many tools & strategies available now
Mix and match!
Software should accommodate mix & match, by providing interfaces and complying to certain standards and conventions
Atomic resolution is just around the corner
(but the corner for some reason moves farther and farther away)