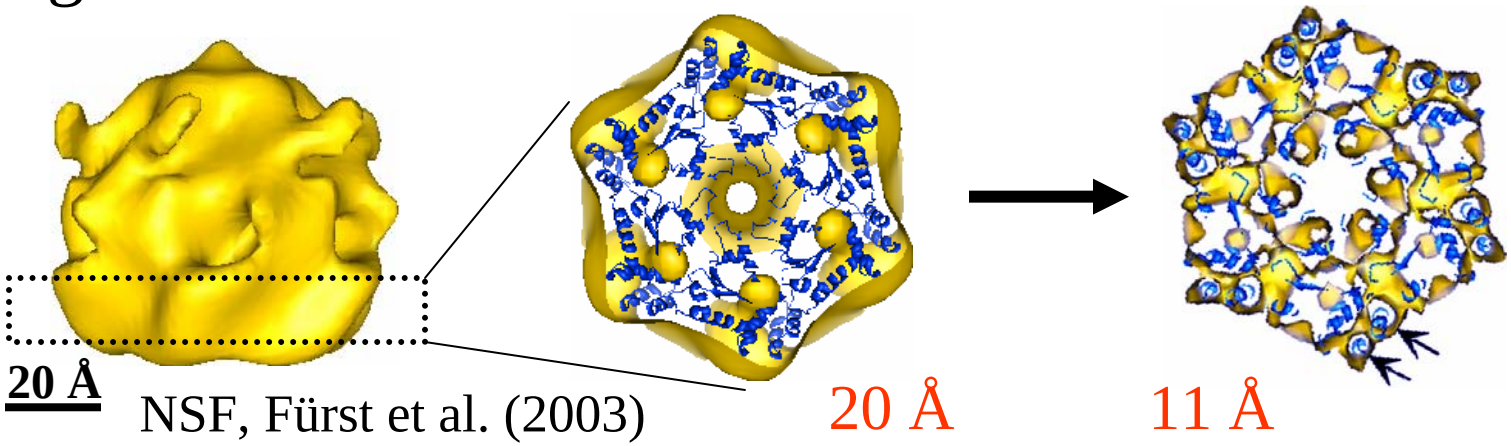
Refinement Strategies for Single Particle Structure Determination

N. Grigorieff

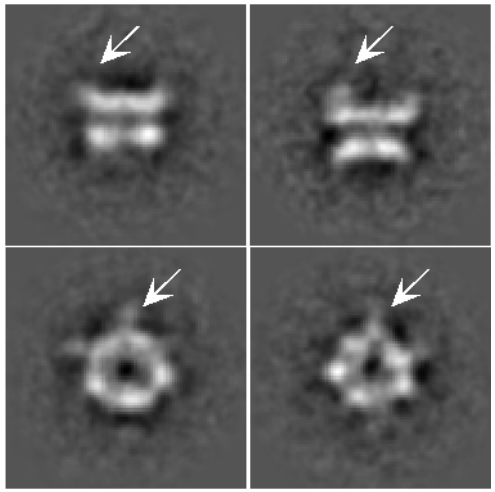
Goals

- Higher resolution



NSF, Fürst et al. (2003)

- Sorting of structural heterogeneity



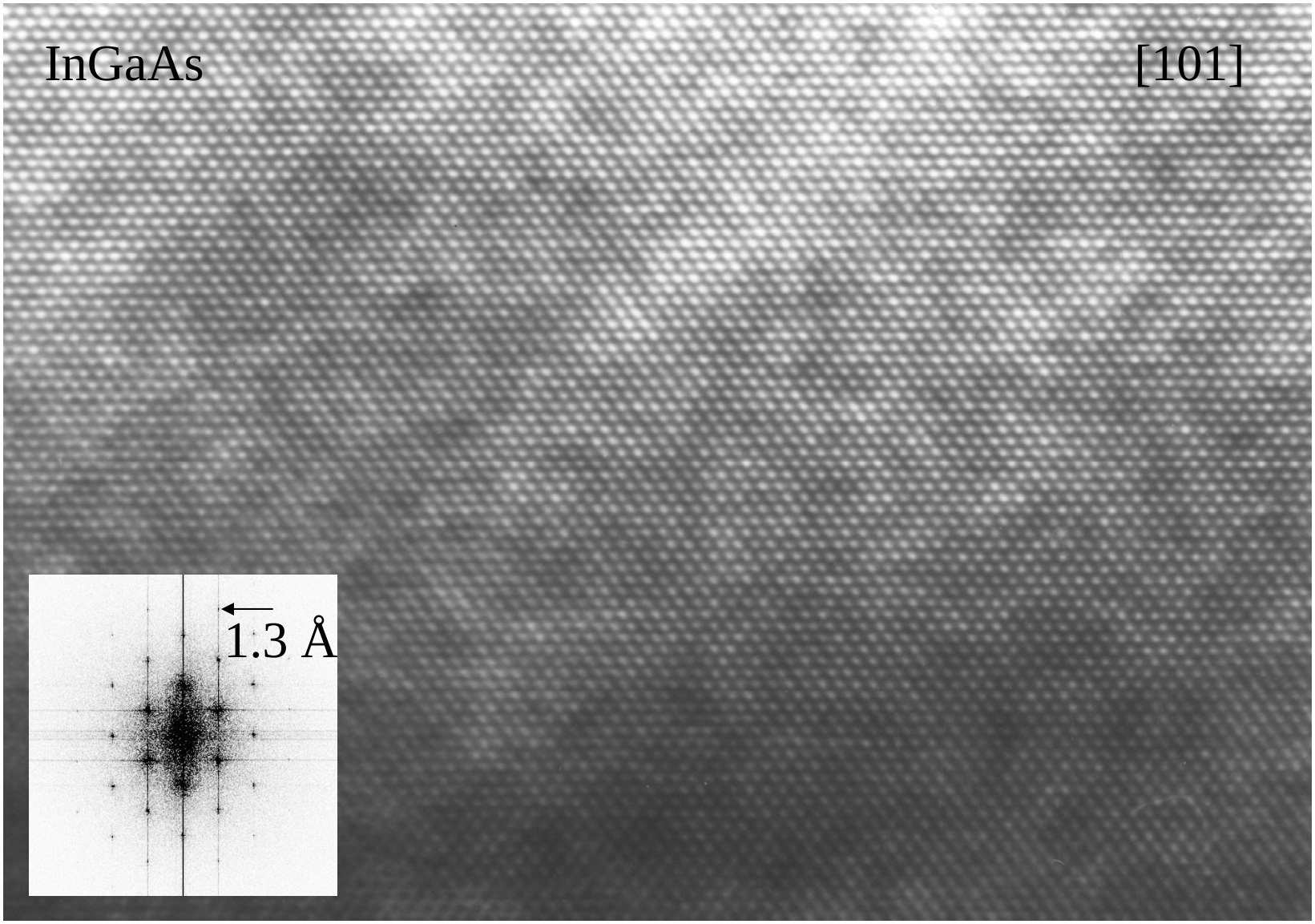
The Prophecy

King Richard hath decreed... (QRB, 1995)

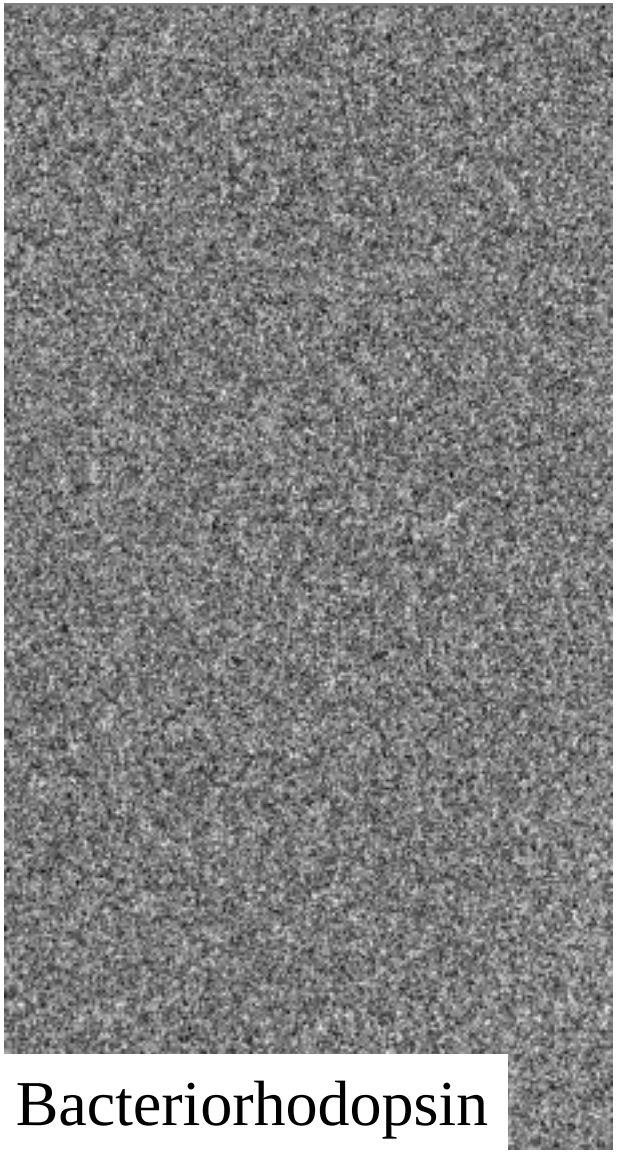
- Use 5 e^- per $Å^2$
- Demand a signal-to-noise ratio of 9 or better
- Aim for 3 Å resolution

- Thou shall need to image 13,000 molecules
- For 6 Å, thou shall need only 7,000 images

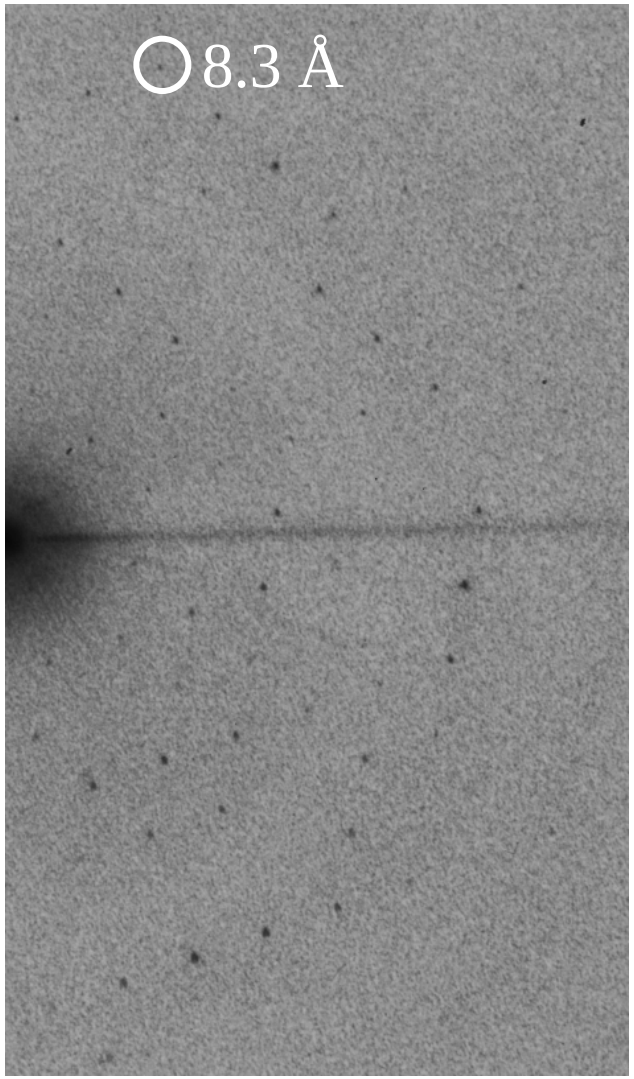
Resolving Power

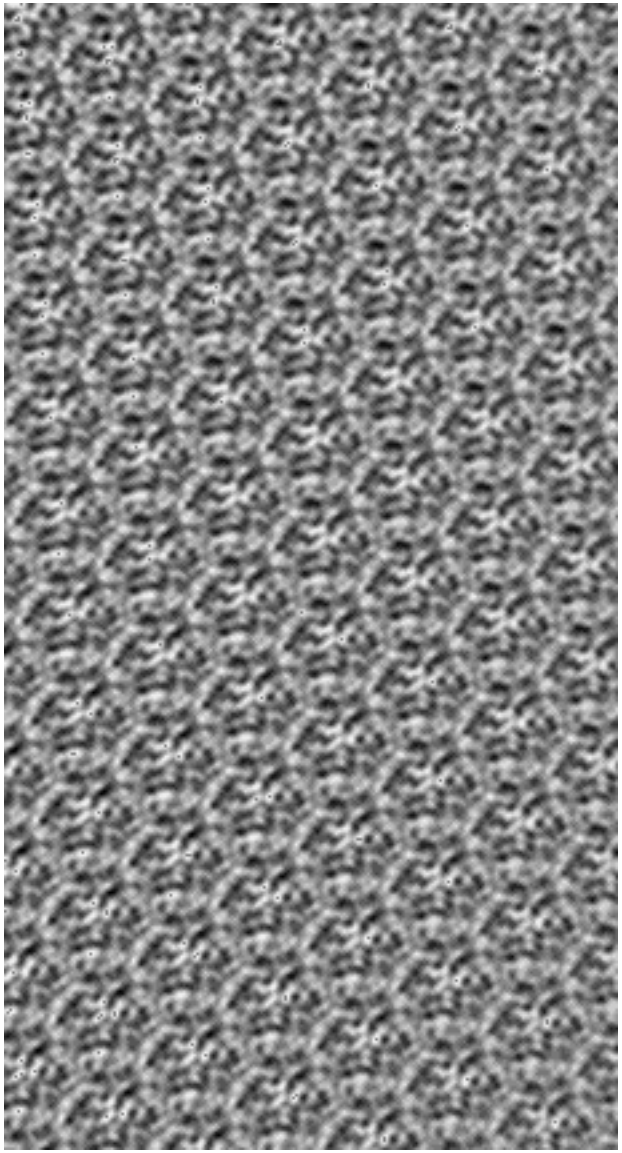


Protein Crystals

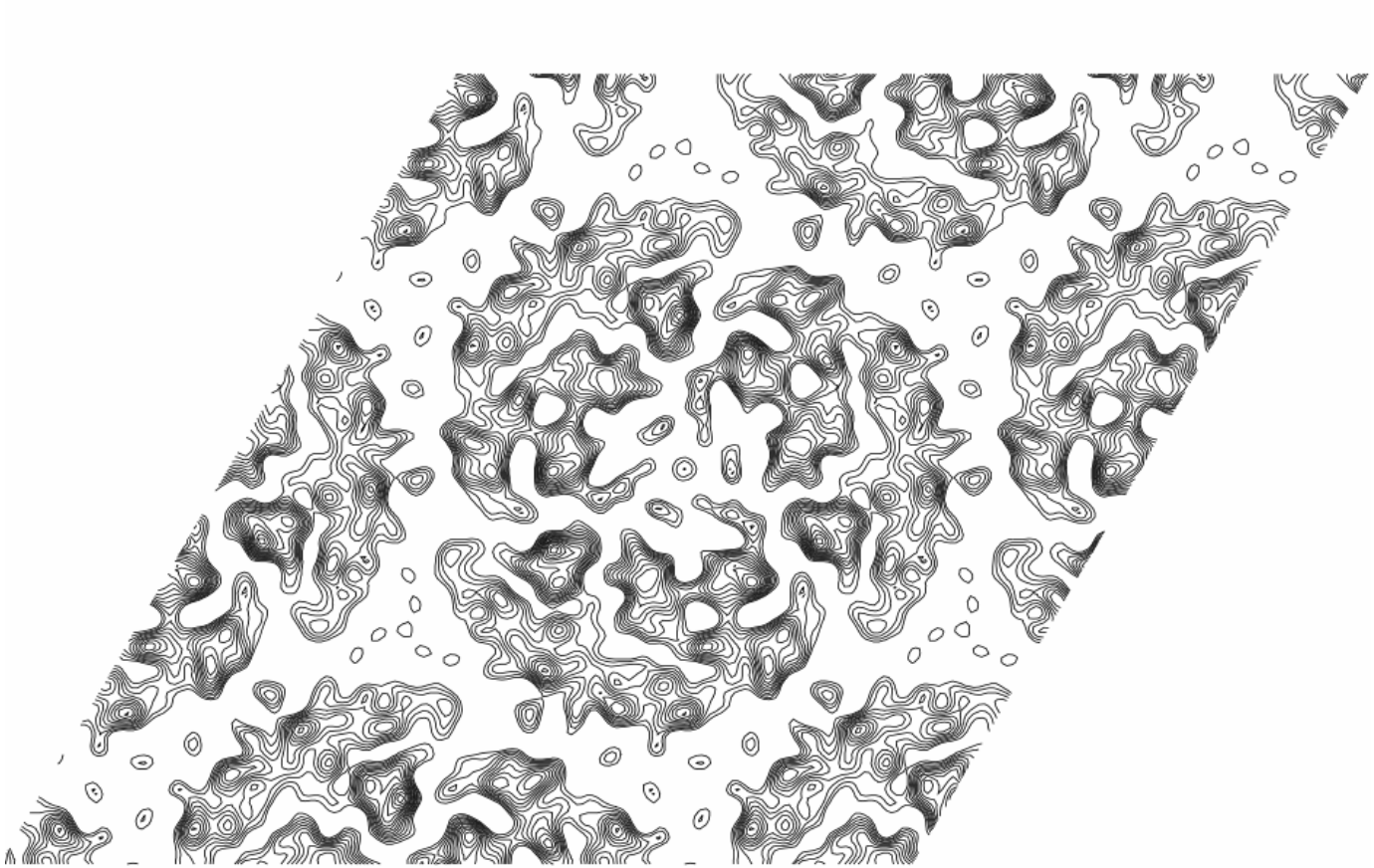


Bacteriorhodopsin



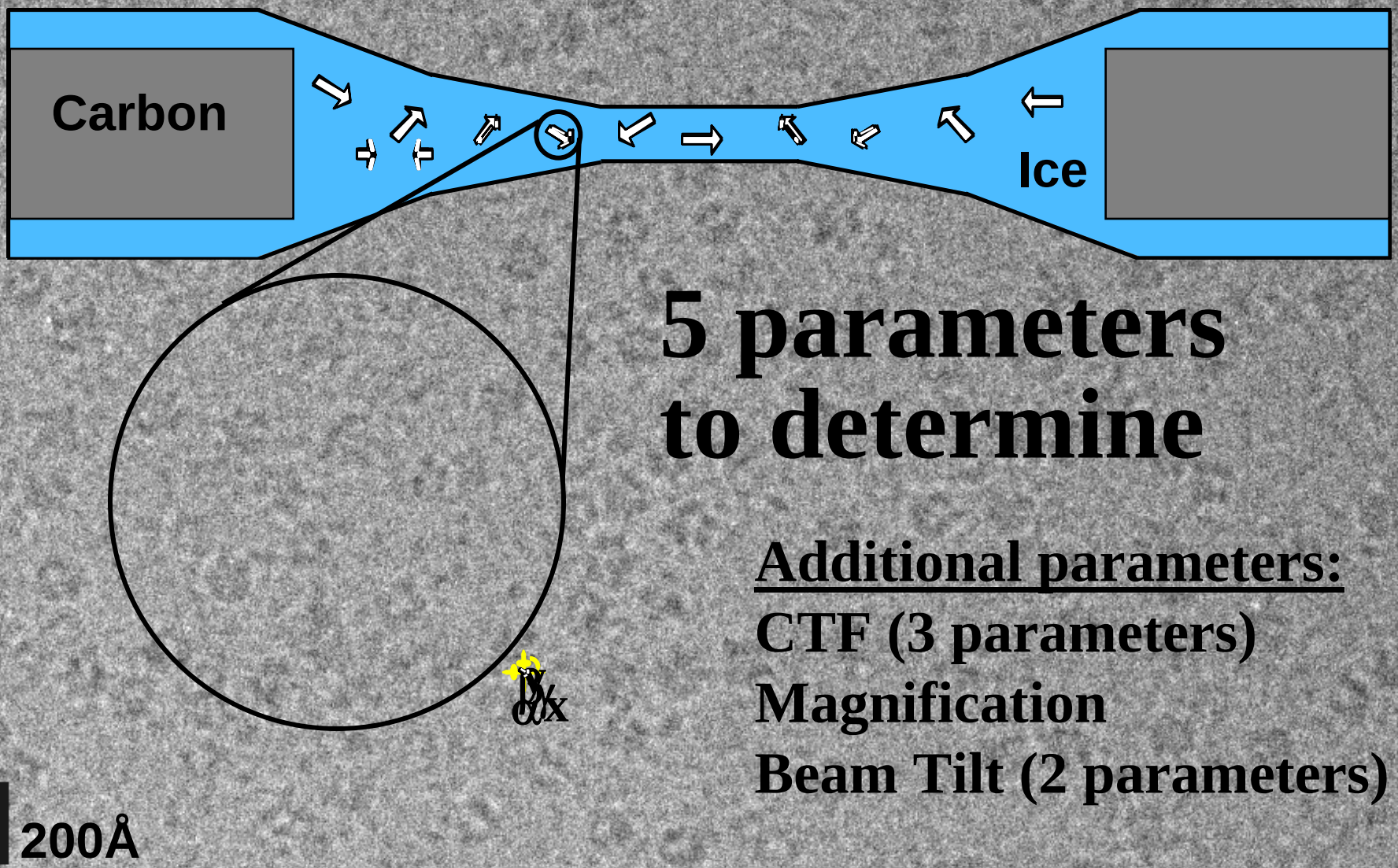


Purple Membrane



2.6 Å resolution

The Puzzle



5 parameters to determine

<u>Additional parameters:</u>

CTF (3 parameters)

Magnification

Beam Tilt (2 parameters)

A Crazy Idea

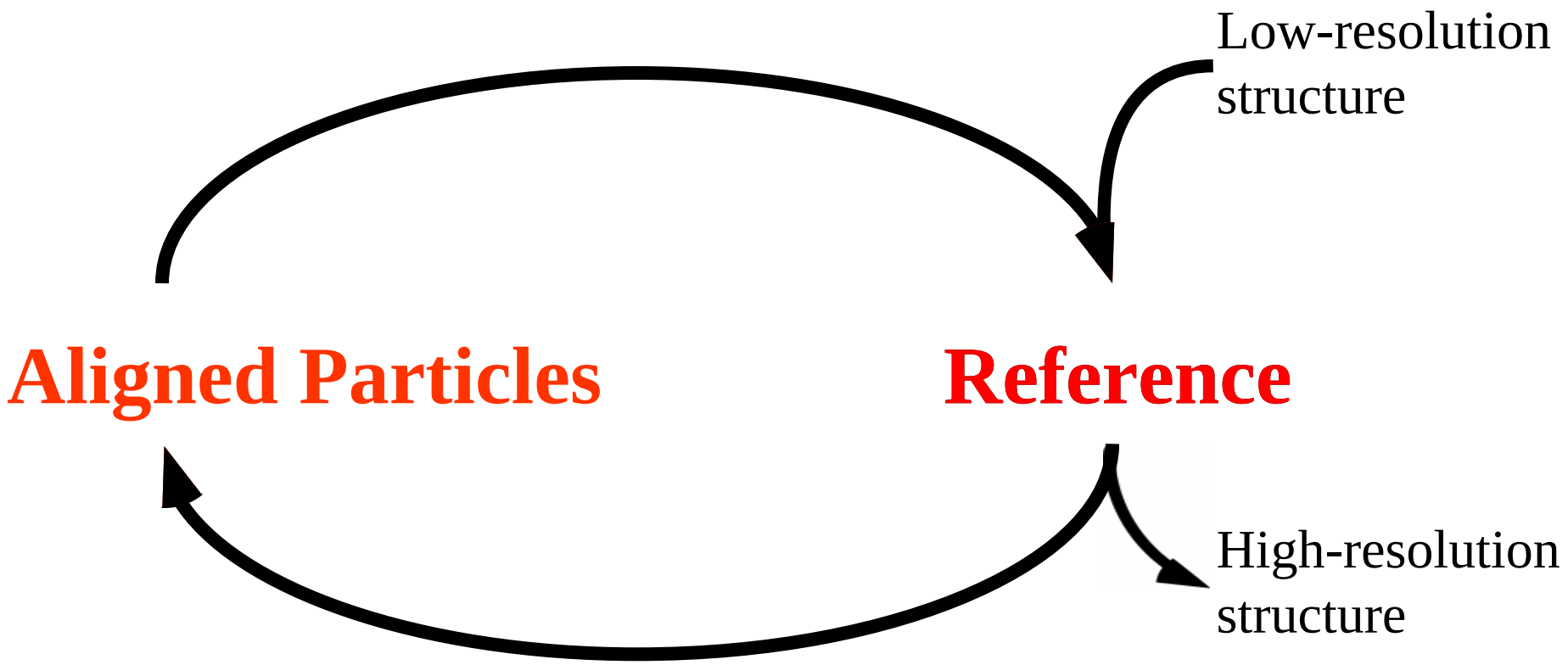
- Assume reliable resolution measure
- Search entire parameter space for highest resolution
- Given enough images, atomic resolution is reached
- Example:

  3 angles, 1 deg step; two coordinates, 1 pixel step:

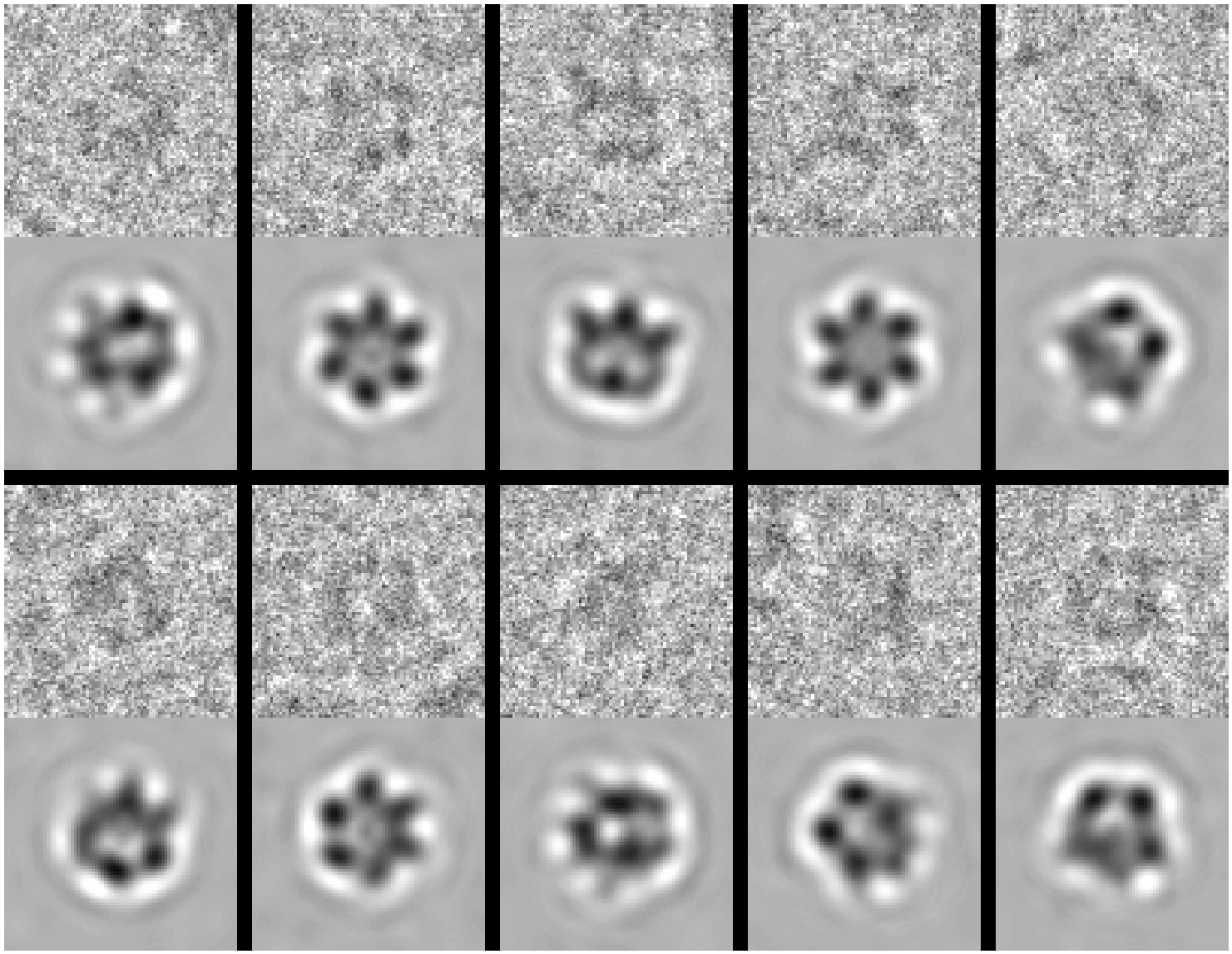
  $360 \times 360 \times 360 \times 100 \times 100 = 5 \times 10^{11}$

  13000 particles: $(5 \times 10^{11})^{13000}$ structures to search
- This is a big number!

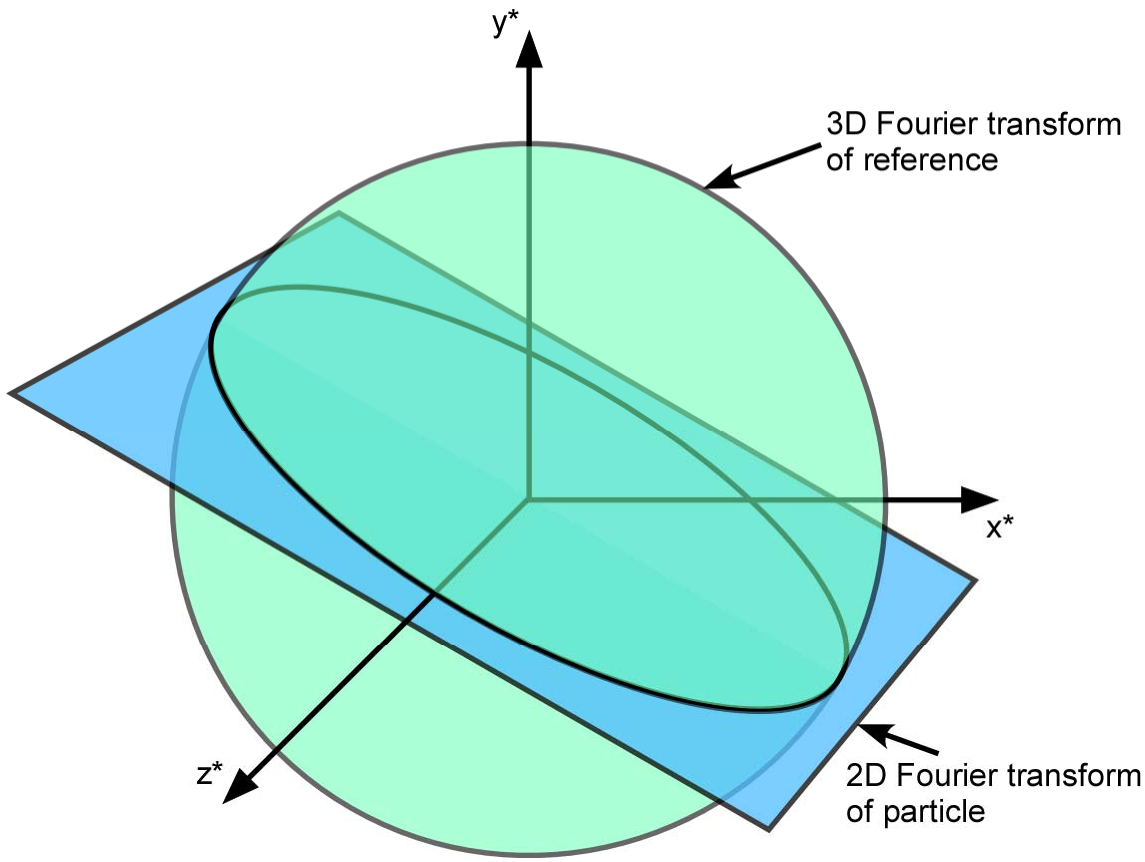
Refinement


Low-resolution structure
Aligned Particles
Reference
High-resolution structure

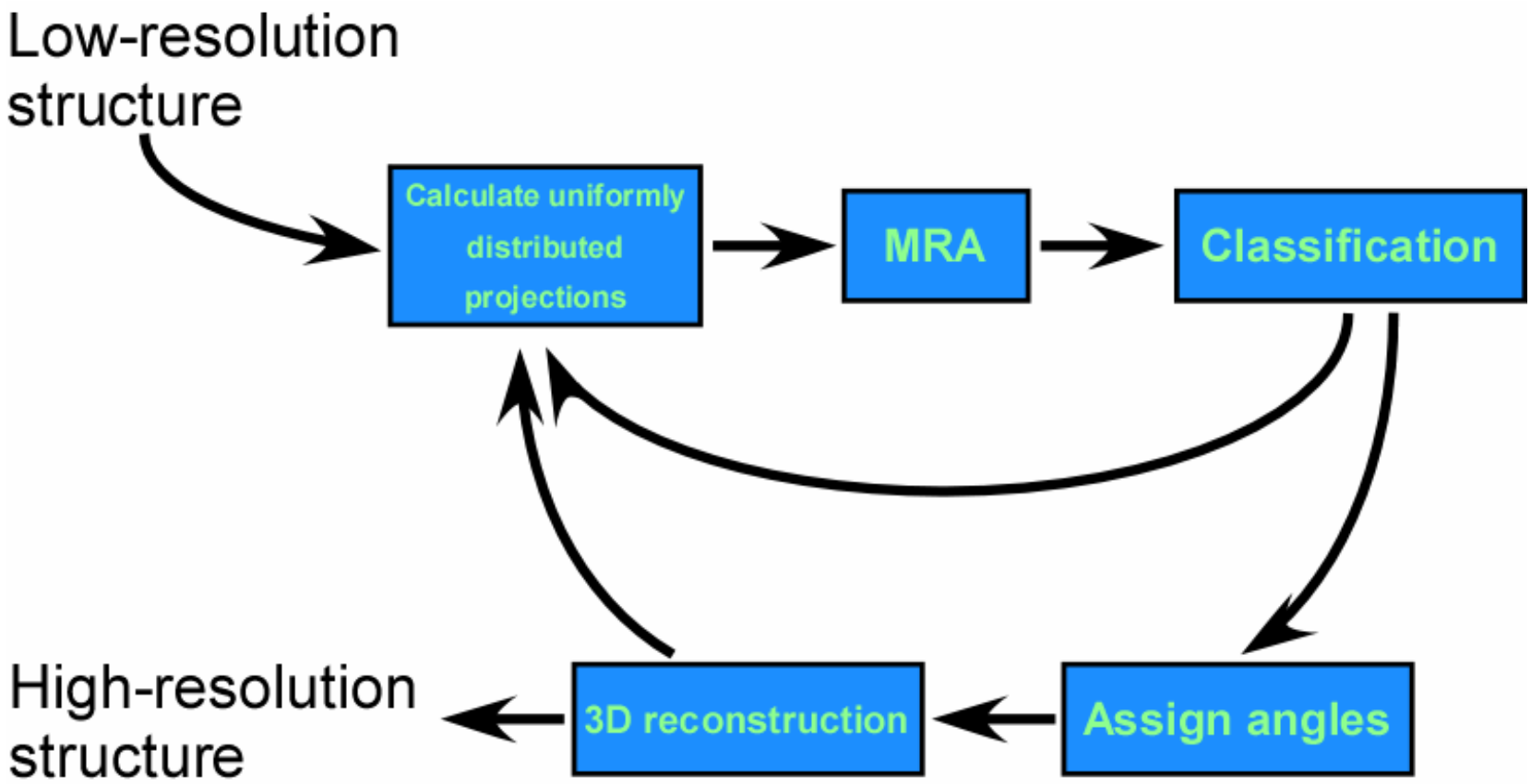
Strategy 1: Projection Matching



Strategy 2: Alignment in Reciprocal Space



Strategy 3: MRA and Classification



Strategy 4: Maximum Likelihood

Structure for *n*+1 iteration

$$A^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\int X_i(\phi) p_i\left(\phi, \Theta^{(n)}\right) \mathrm{d}\phi}{\int p_i\left(\phi, \Theta^{(n)}\right) \mathrm{d}\phi}$$

Probability function

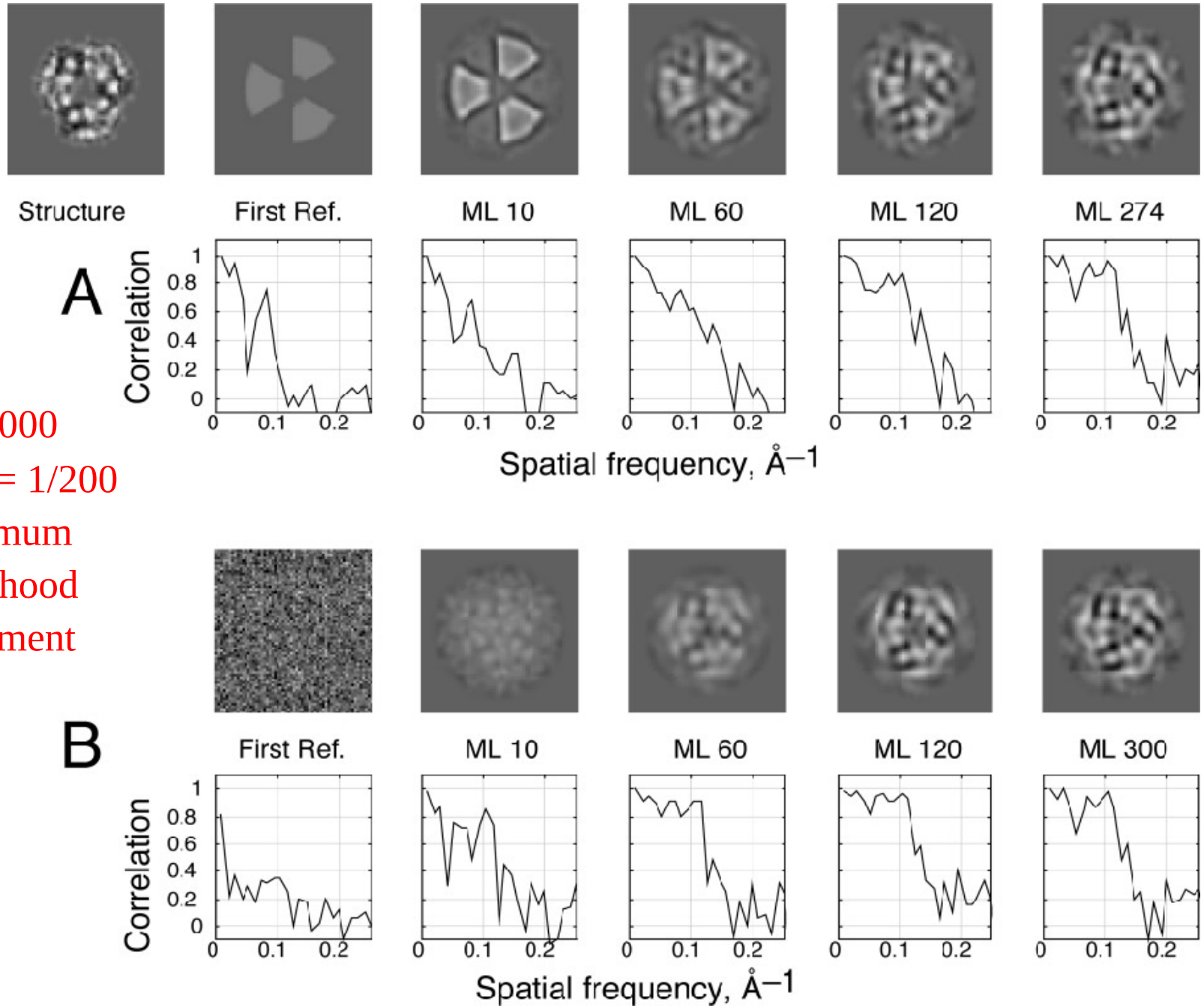
$$p_i(\phi, \Theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left[-\frac{|X_i(\phi) - A|^2}{2\sigma^2}\right] f(\phi \mid \Theta)$$

X_i: *i*th image
Θ: model parameters

N: # of images
σ: noise in images

ϕ: alignment parameters
f : positional probab.

Sigworth (1998), J. Struct. Biol. 122, 328-339



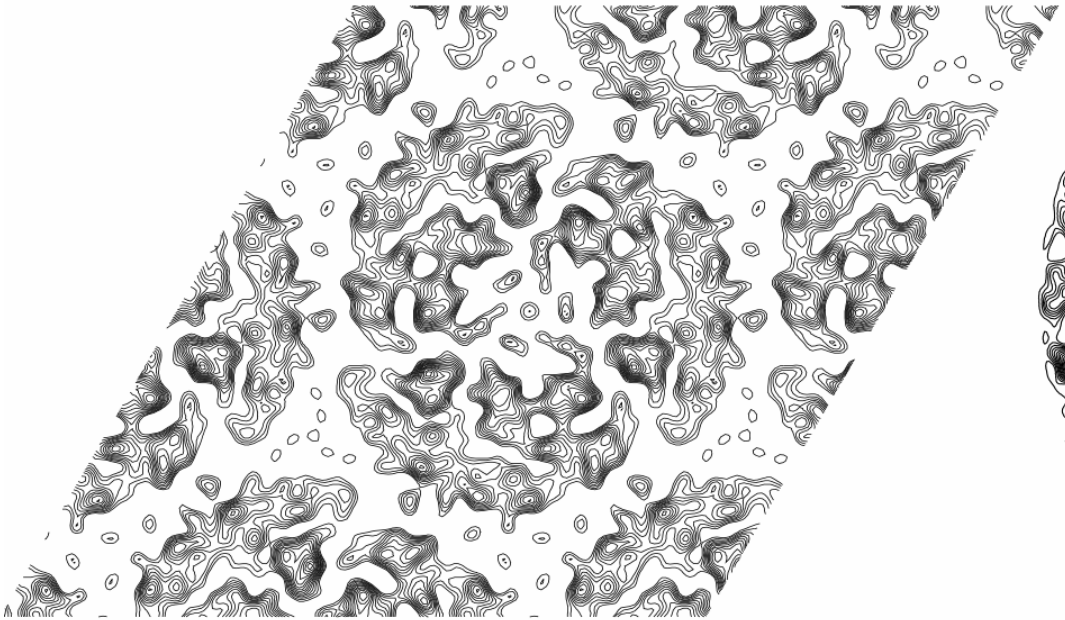
N = 4000

SNR = 1/200

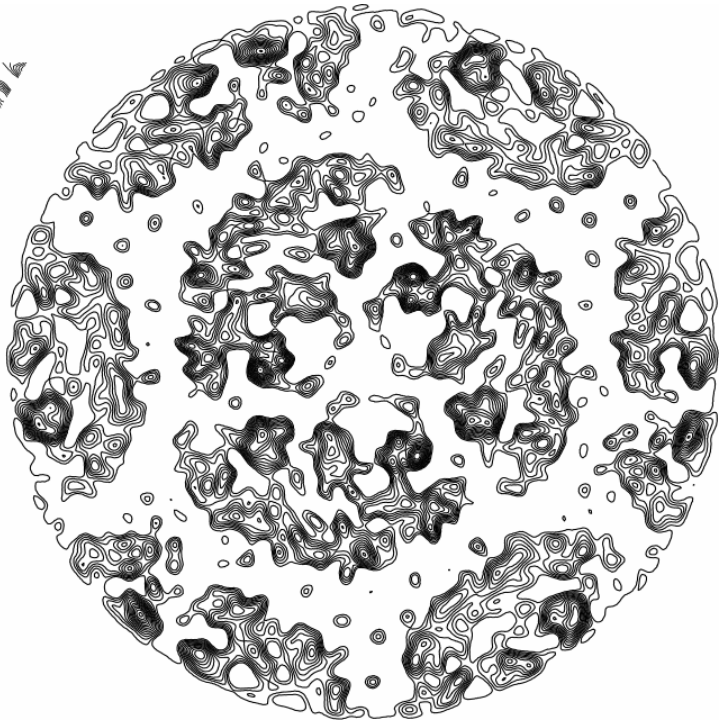
Maximum likelihood alignment

Sigworth (1998), J. Struct. Biol. 122, 328-339

ML processing of 2D crystals



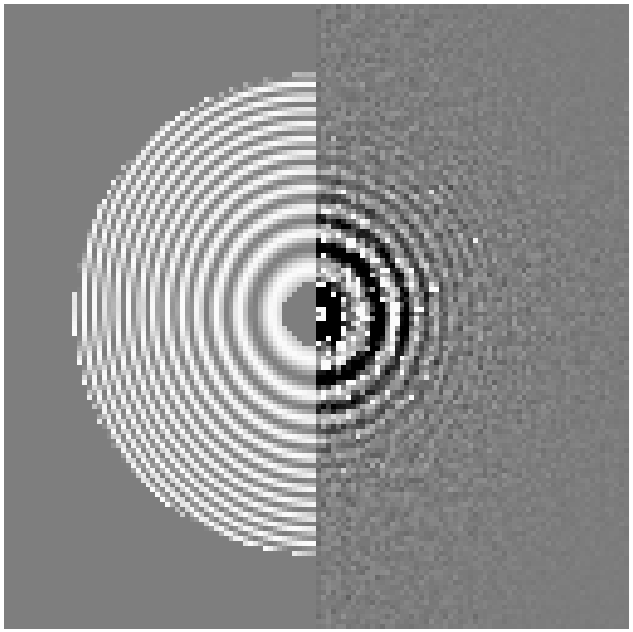
Crystallography



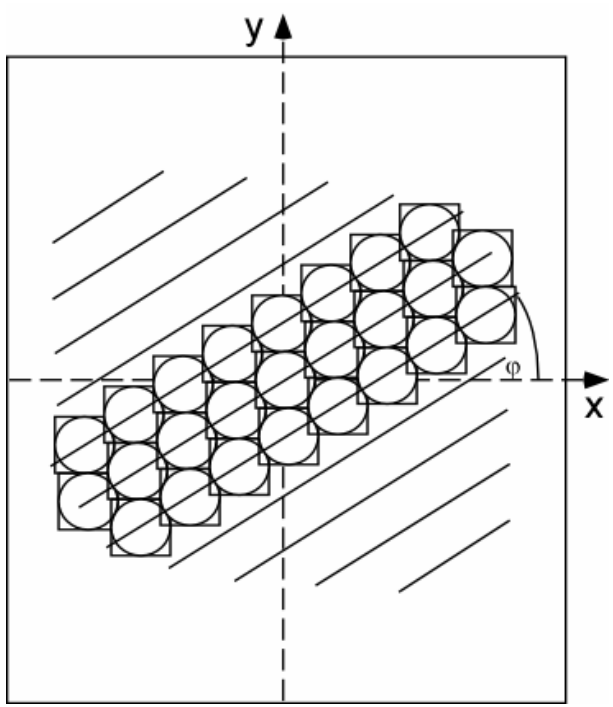
Alignment of individual unit cells using ML approach

Defocus/Astigmatism and Magnification

CTF ⟷ Magnification

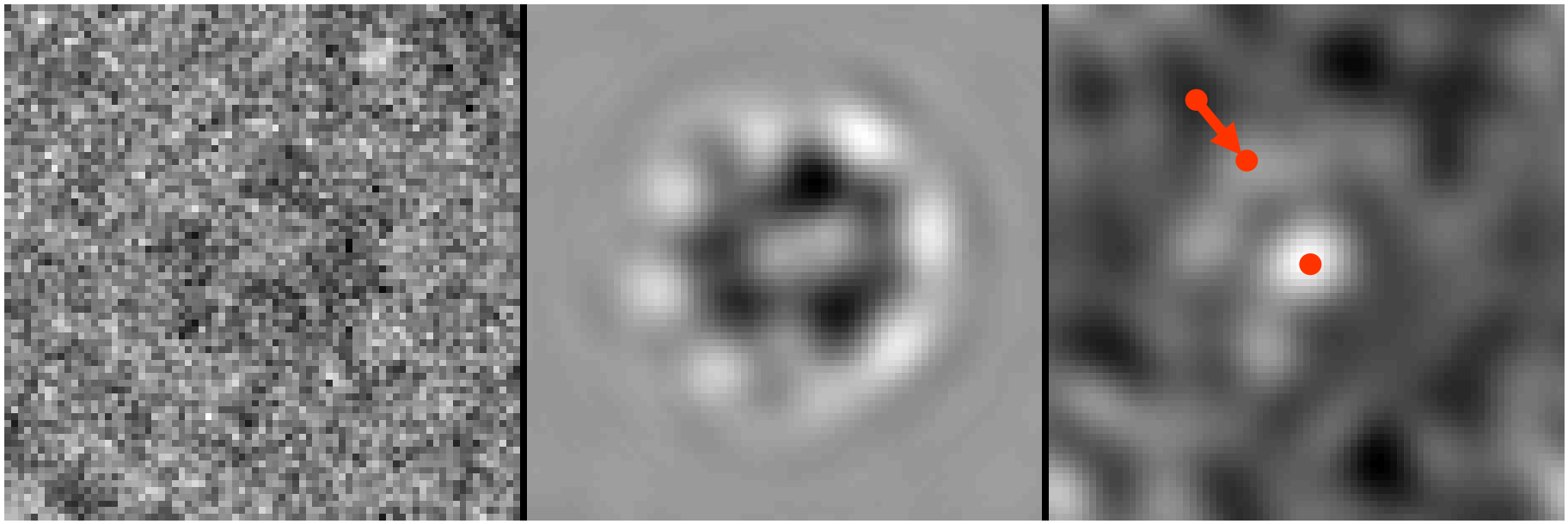


CTFFIND3



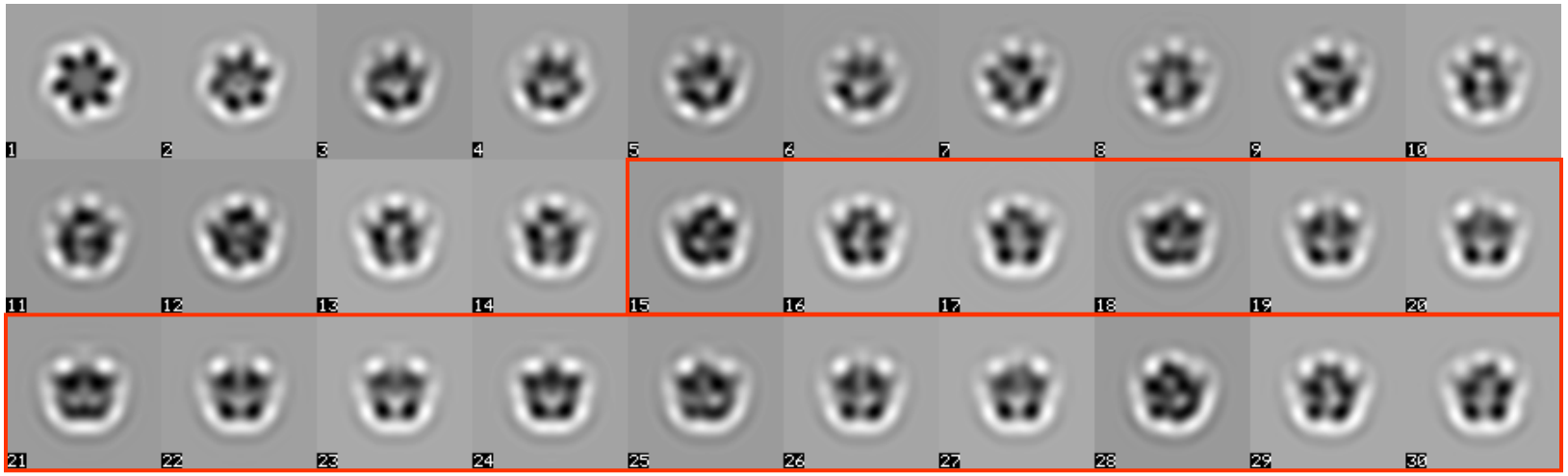
CTFTILT

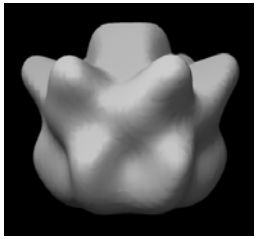
Problem 1: Local Optima



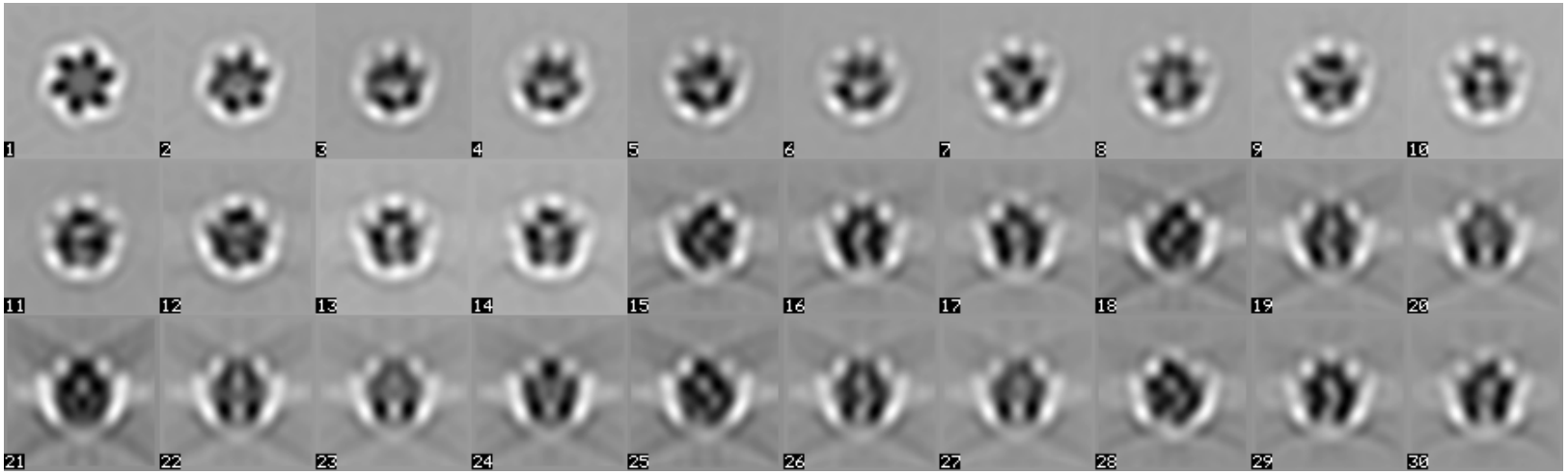
Particle Reference Correlation map

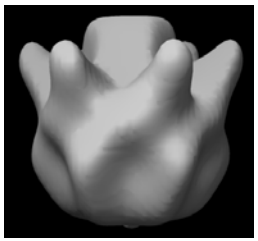
Problem 2: Missing Views



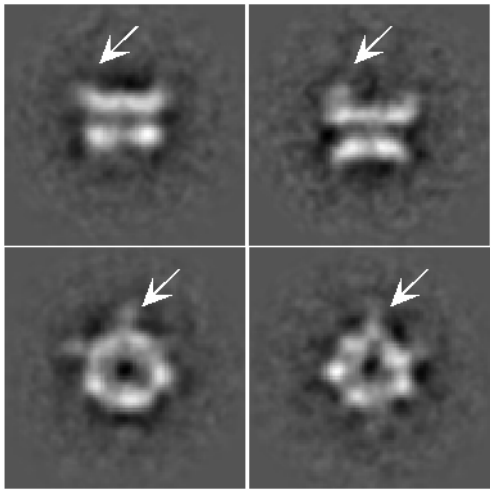


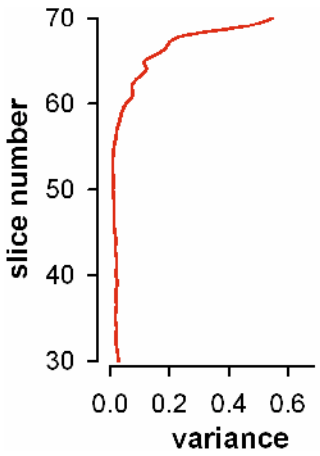
> 60º

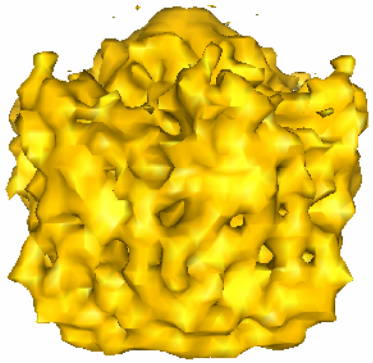




Problem 3: Heterogeneity







- Misalignment of particles
- Lower resolution in disordered regions
- Loss of features

Classification Using ML

Structure for *n*+1 iteration

$$A_k^{(n+1)} = \frac{1}{\sum_i q_i^k(\Theta)} \sum_{i=1}^{N} \frac{\int X_i(\phi) p_i^k\left(\phi, \Theta^{(n)}\right) \mathrm{d}\phi}{\sum_k \int p_i^k\left(\phi, \Theta^{(n)}\right) \mathrm{d}\phi}$$

Probability function

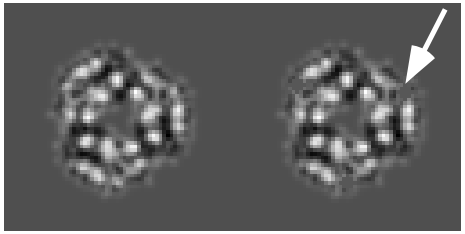
$$p_i^k(\phi, \Theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left[-\frac{|X_i(\phi) - A_k|^2}{2\sigma^2}\right] f(\phi \mid \Theta)$$

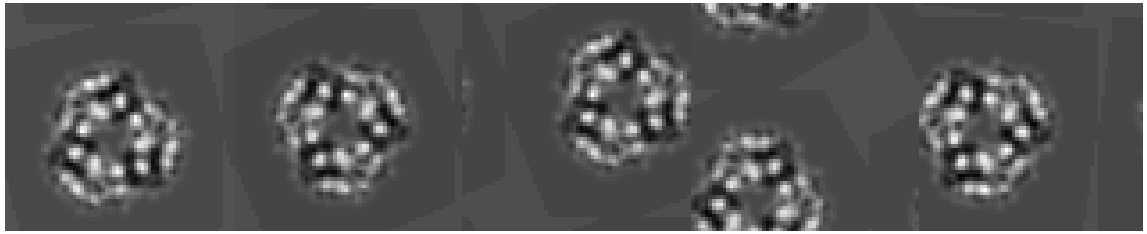
Probability for class *k*

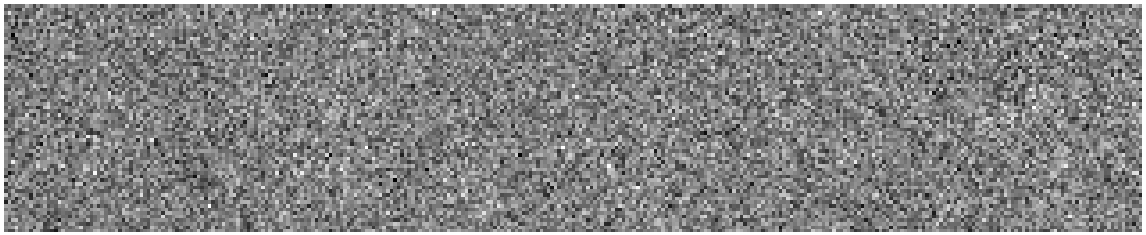
$$q_i^k(\Theta) = \int p_i^k(\phi, \Theta) \mathrm{d}\phi$$

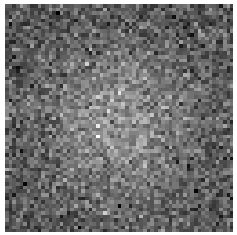
X_i: *i*th image
Θ: model parameters
N: # of images
σ: noise in images
ϕ: alignment parameters
f : positional probab.

Classification Using ML

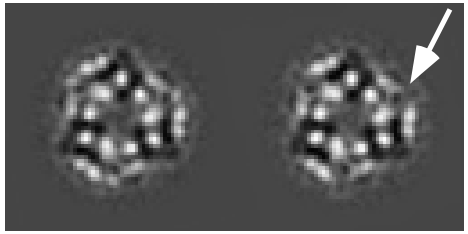


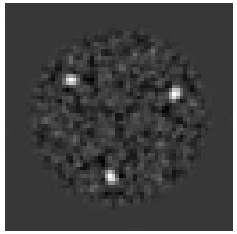






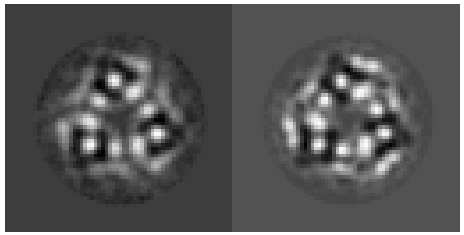
SNR = 1/50
N = 2000

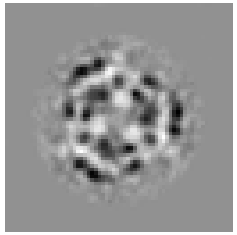




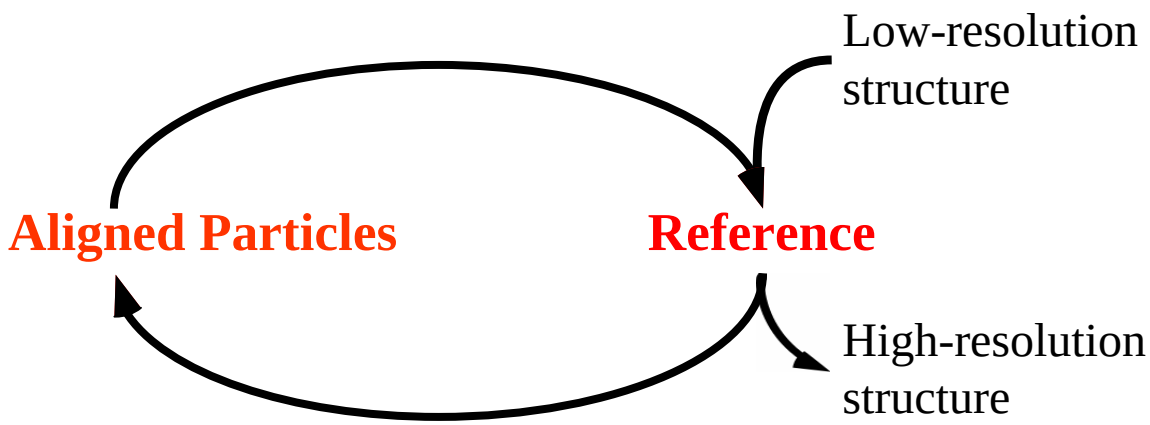
Difference map

Correlation alignment

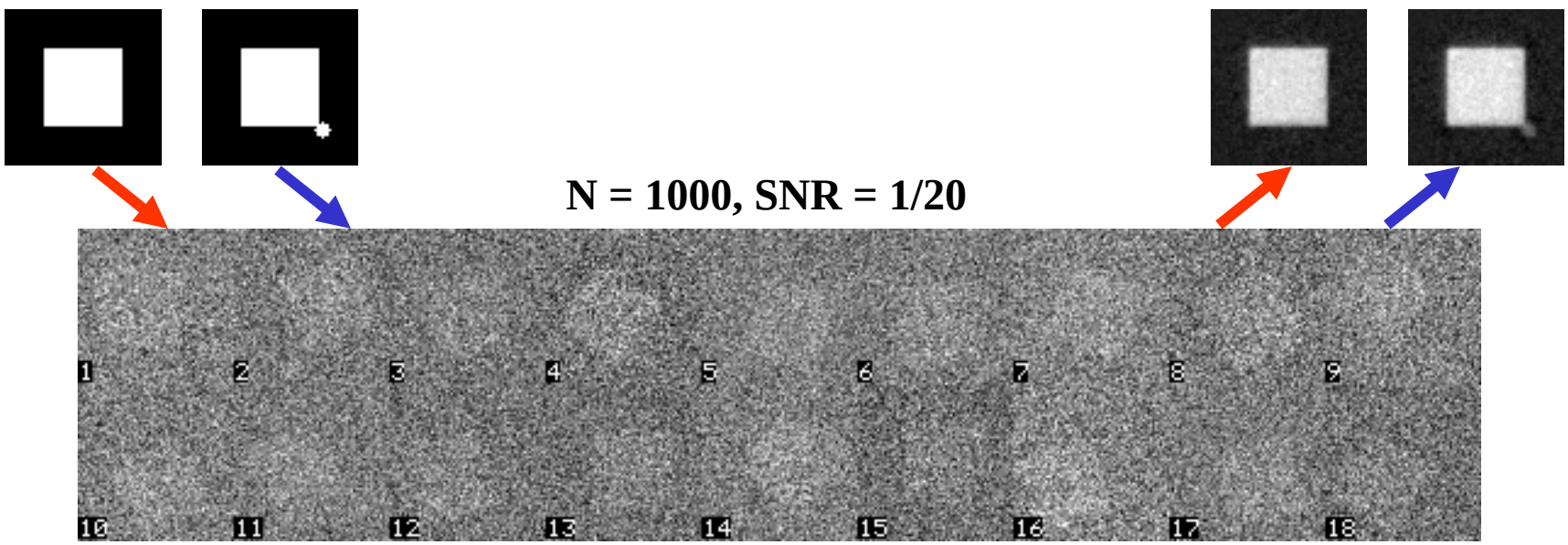




Problem 4: Processing Artifacts

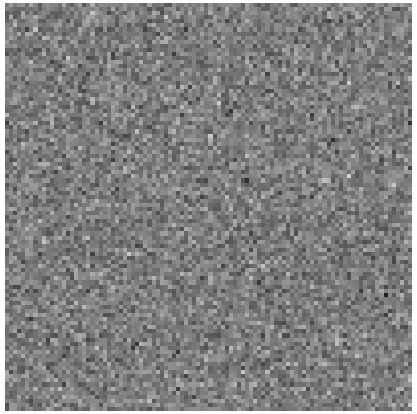


- Interpolation errors
- Masking
- Negative B-factor
- …



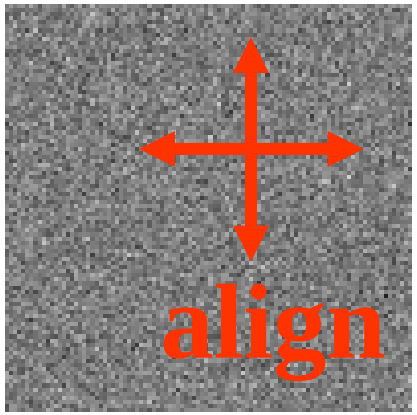
Problem 5: Noise Bias

$\otimes$



= 0 on average

$\otimes$



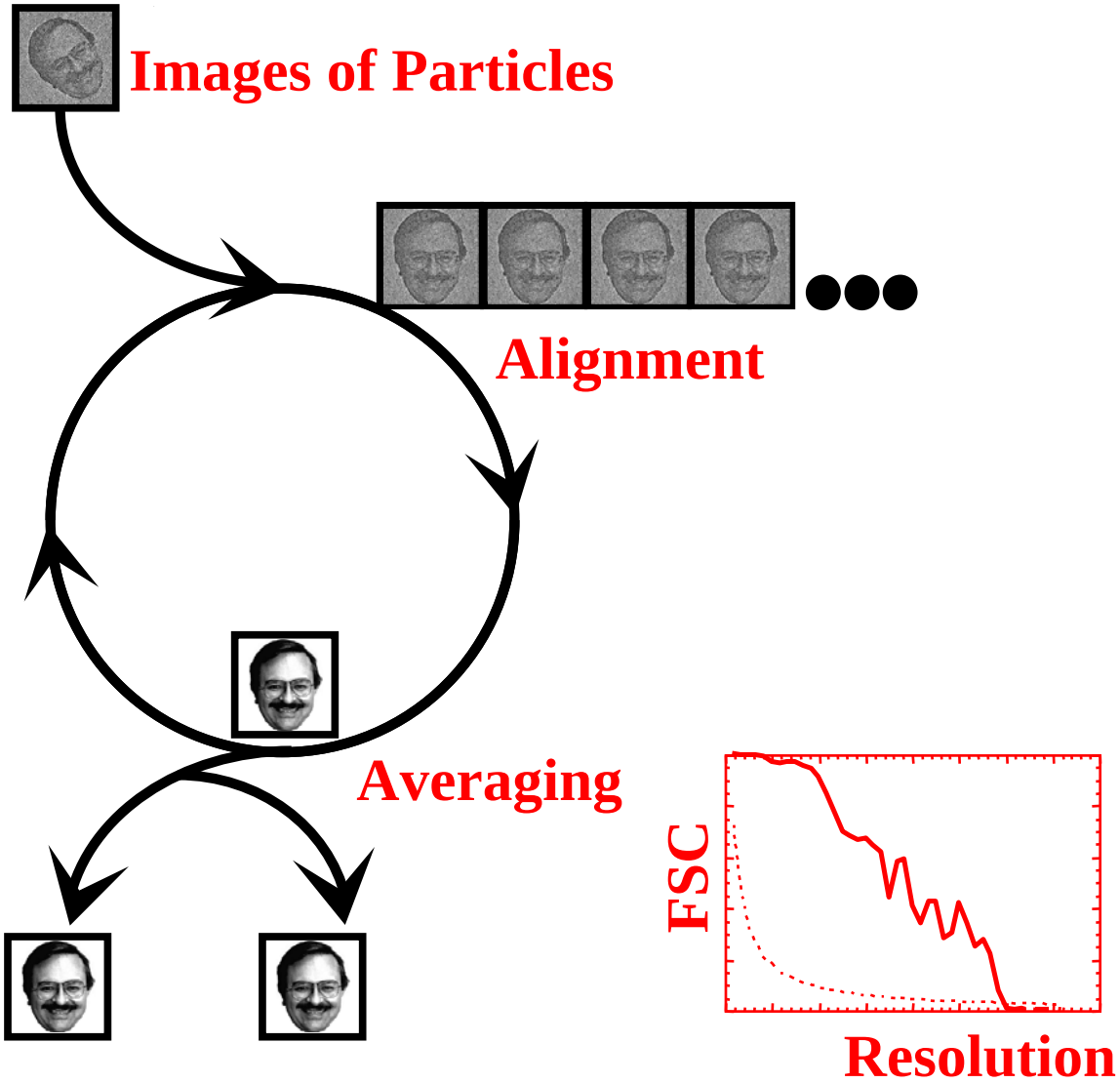
> 0

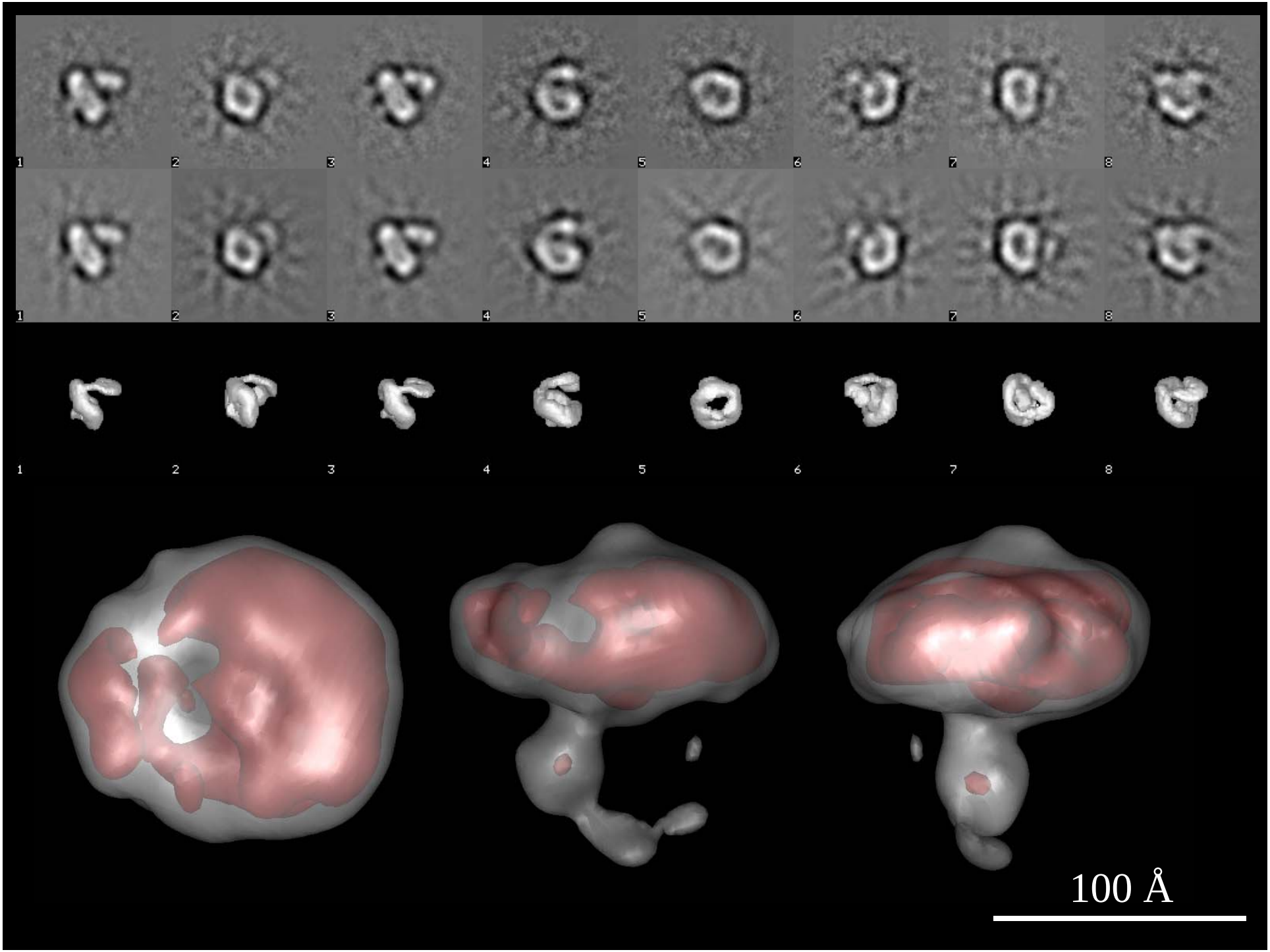
for 64x64 image:
average
correlation = 0.064

Seeing is NOT Always Believing

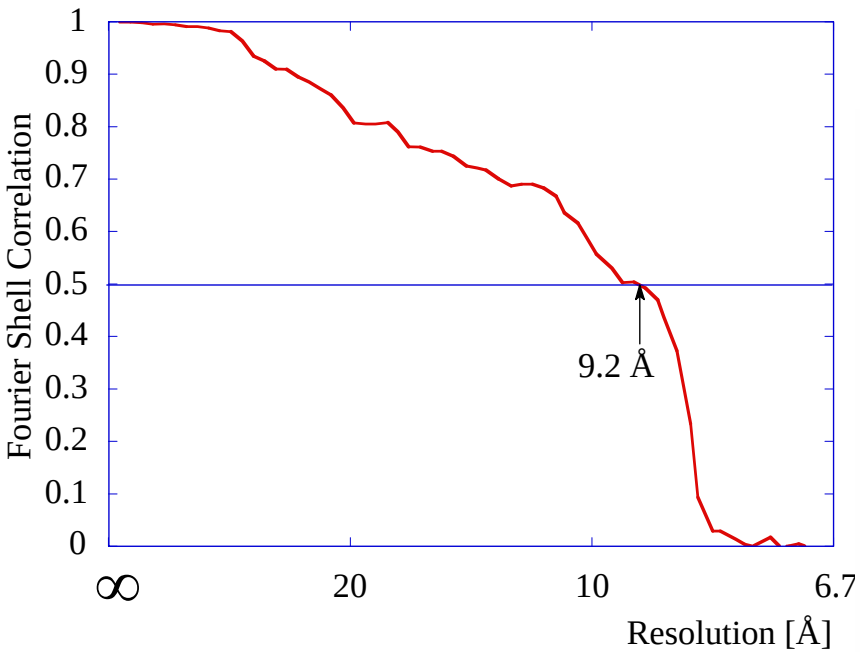
100 Images 1000 Images Reference

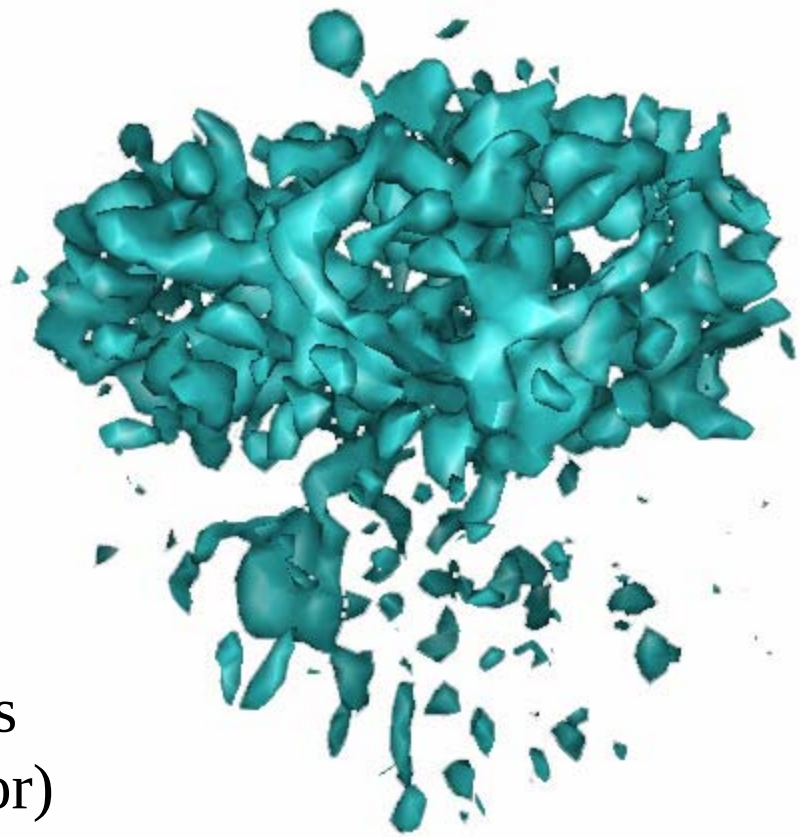
Resolution Measurement




1
2
3
4
5
6
7
8
100 Å

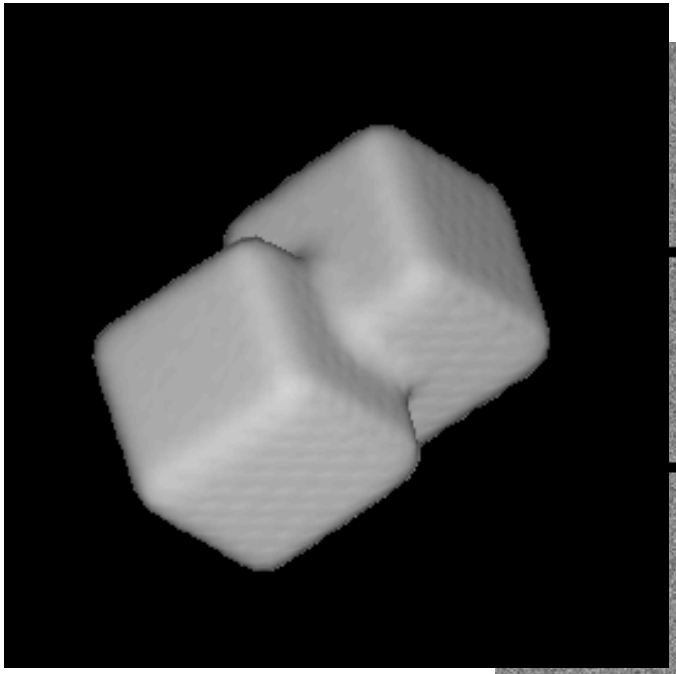
Swiss Cheese

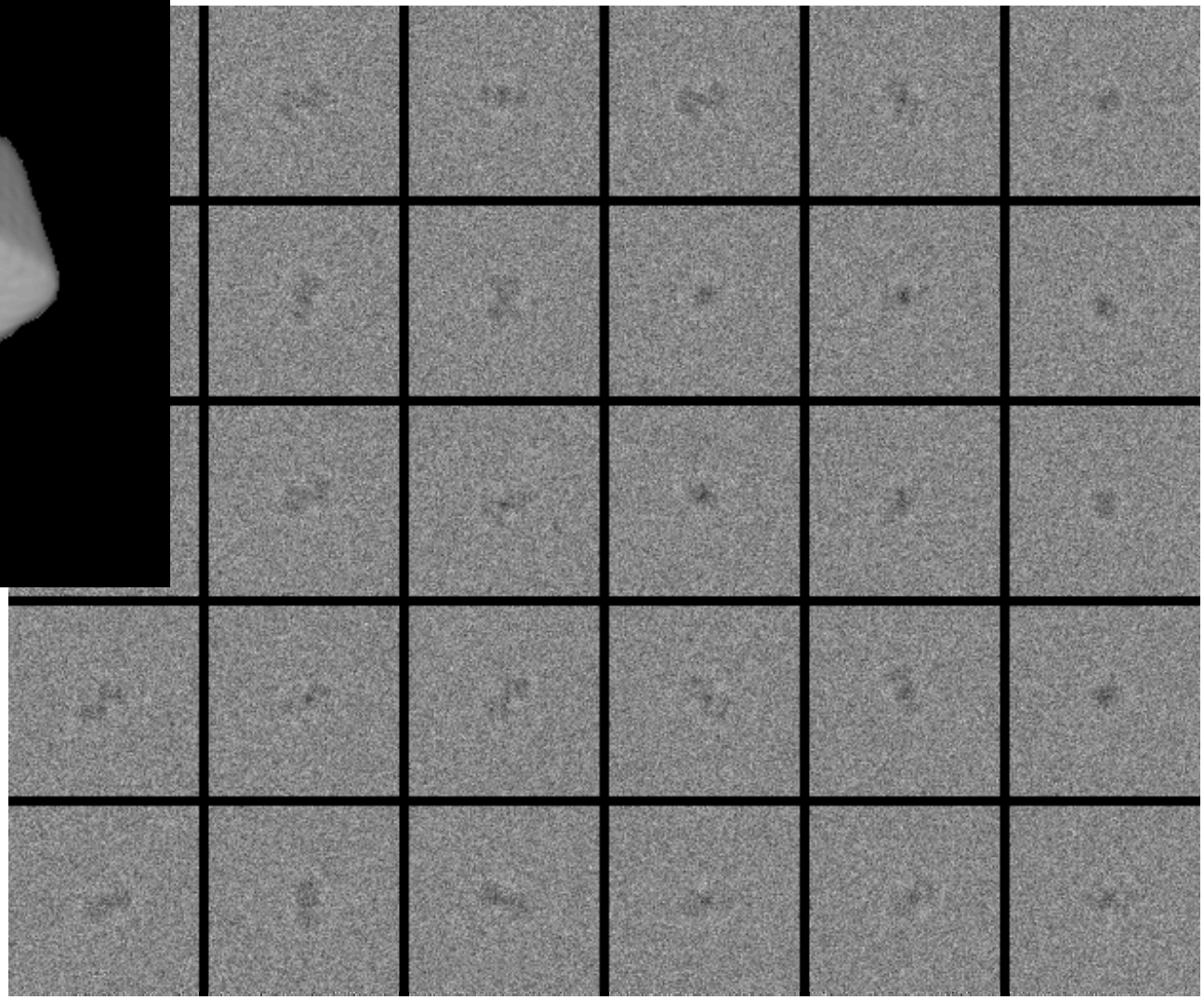




<u>Dangerous:</u>
Boosting of high-resolution terms
(application of a negative B-factor)

Gedanken Experiments

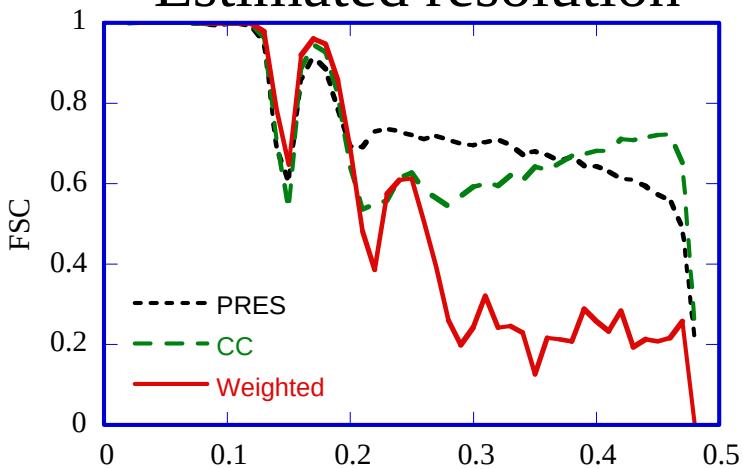




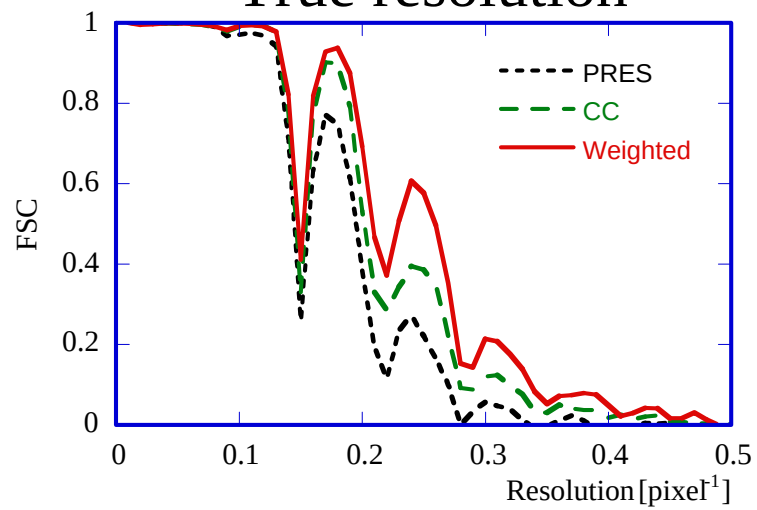
N = 30000
SNR = 1/50

Weighted Correlation

Estimated resolution



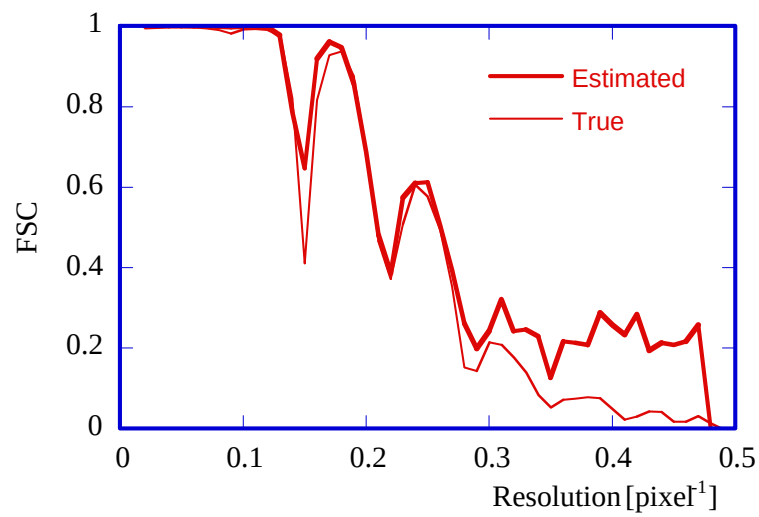
True resolution



$$\mathrm{PRES}(X,Y) = \frac{\sum_{k\in[0,0.5]} \Delta\Phi_{X,Y}(\mathbf{k})\left|F_X(\mathbf{k})\right|}{\sum_{k\in[0,0.5]} \left|F_X(\mathbf{k})\right|}$$

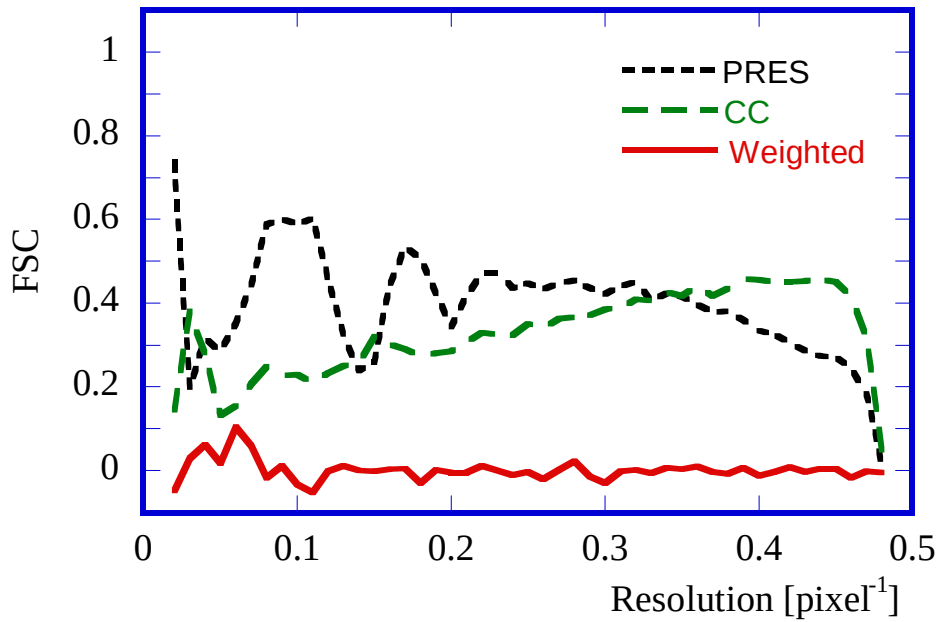
$$\mathrm{CC}(X,Y) = \frac{\sum_{k\in[0,0.5]} F_X(\mathbf{k})F_Y^{\dagger}(\mathbf{k})}{\sqrt{\sum_{k\in[0,0.5]} \left|F_X(\mathbf{k})\right|^2 \sum_{k\in[0,0.5]} \left|F_Y(\mathbf{k})\right|^2}}$$

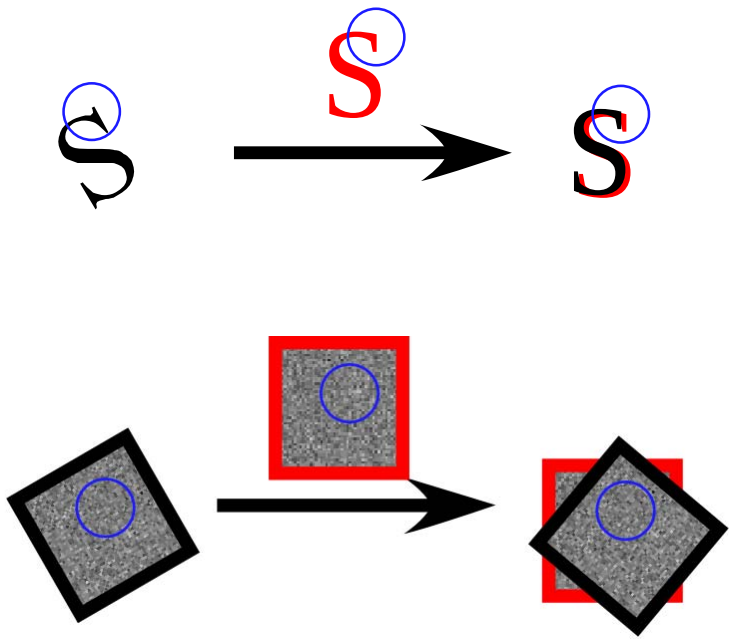
$$\mathrm{CC}_W(X,Y) = \sum_{k\in[0,0.5]} W_X(\mathbf{k})F_X(\mathbf{k})F_Y^{\dagger}(\mathbf{k})$$



Coherence Constraint

$$CC_W = \sum_i \left| CC \right|_i^3$$





Acknowledgements

- Ca Channel — Matthias Wolf (Glossmann/Striessnig)
- NSF/20S — Johannes Fürst (Axel Brünger)
- Noisy Face — David DeRosier
- **Financial Support: HHMI, NIH, NSF**