

In introduction to molecular electron microscopy

- Imaging macromolecular assemblies

Yifan Cheng and Adam Frost
Department of Biochemistry & Biophysics

office: GH-S472D (Yifan) or GH-S472F (Adam);
email: ycheng@ucsf.edu or adam.Frost@ucsf.edu

“You just *look at the thing!*”

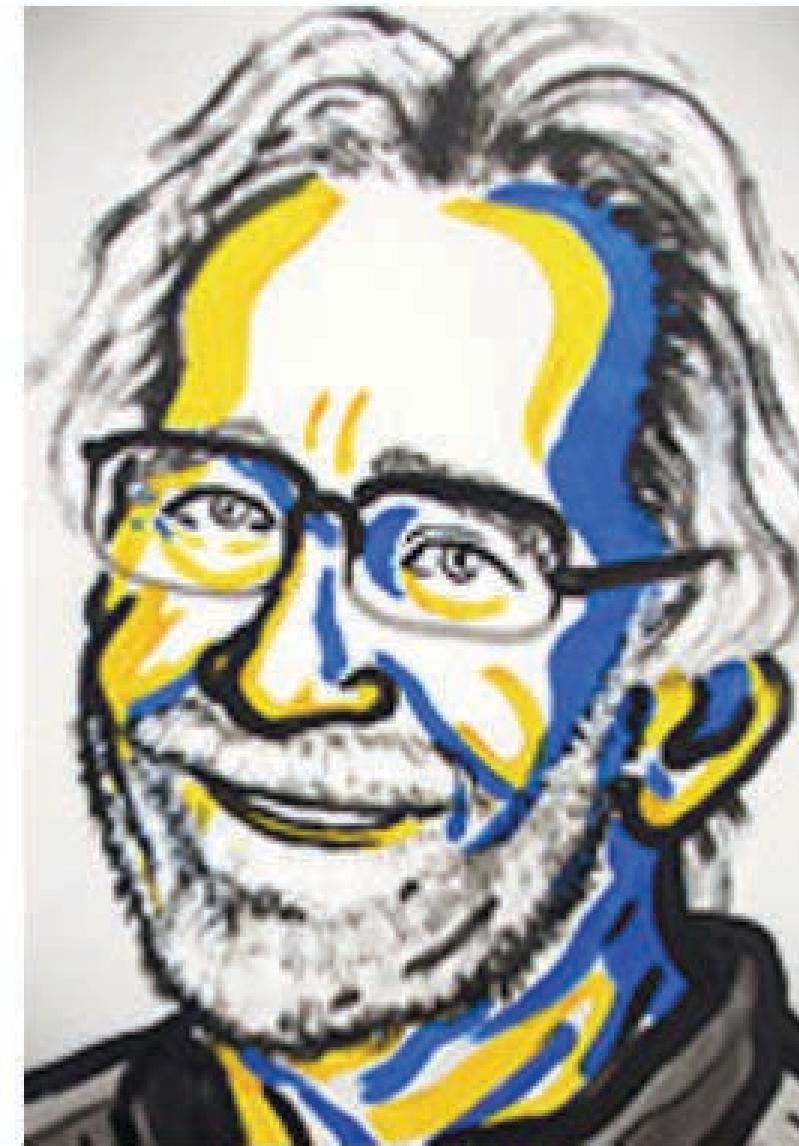
Richard Feynman: There's plenty of room at the bottom
(December 29, 1959, lecture to American Physical Society):

“It is very easy to answer many of these fundamental biological questions: you just *look at the thing!*”

“Unfortunately, the present microscope sees at a scale which is just a bit too crude. Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier.”

“... the biologists would surely be very thankful to you”

The Nobel Prize in Chemistry 2017



© Nobel Media. Ill. N.
Elmehed
Jacques Dubochet
Prize share: 1/3



© Nobel Media. Ill. N.
Elmehed
Joachim Frank
Prize share: 1/3



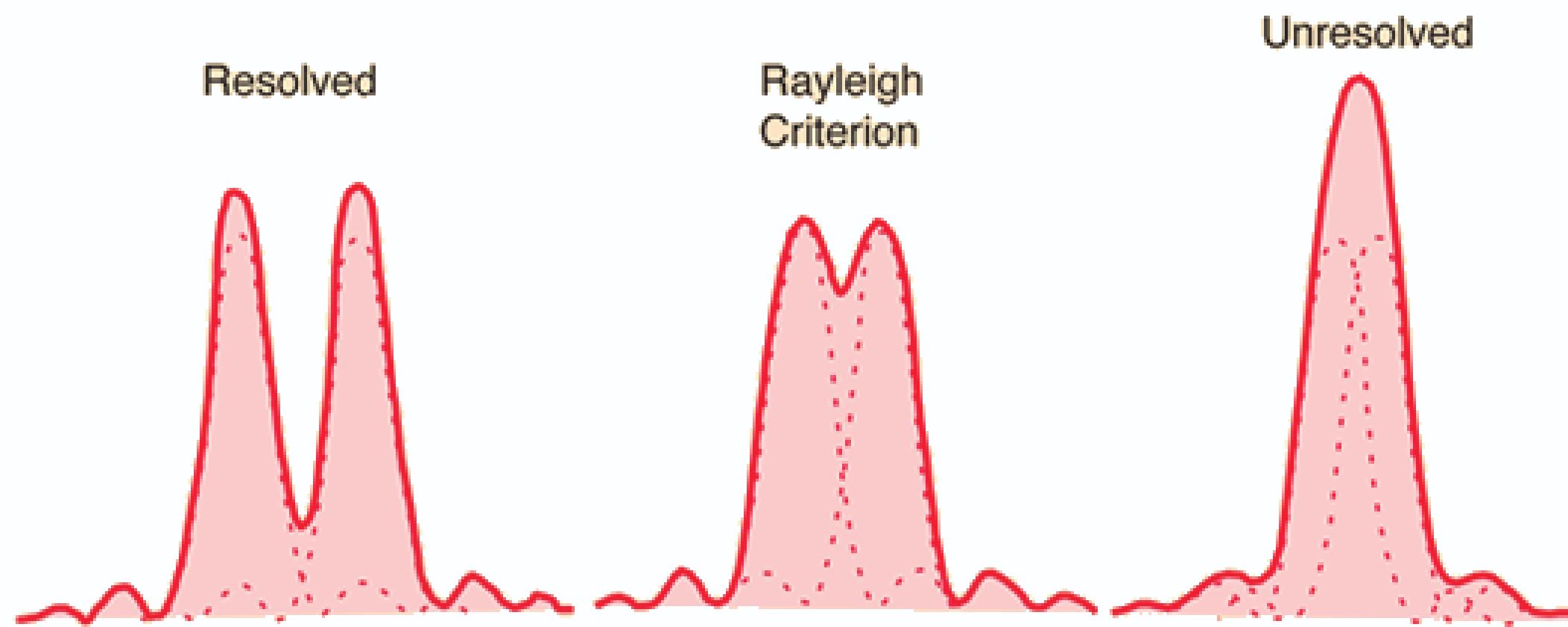
© Nobel Media. Ill. N.
Elmehed
Richard Henderson
Prize share: 1/3

The Nobel Prize in Chemistry 2017 was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution".

Resolution limit of an optical microscope system

Rayleigh criterion:

$$\sin \theta = 1.22 \frac{\lambda}{D} \quad \text{or} \quad \Delta l = 1.22 \frac{f\lambda}{D}$$



θ is the angular resolution, λ is the wavelength and D is the diameter of the lens aperture.

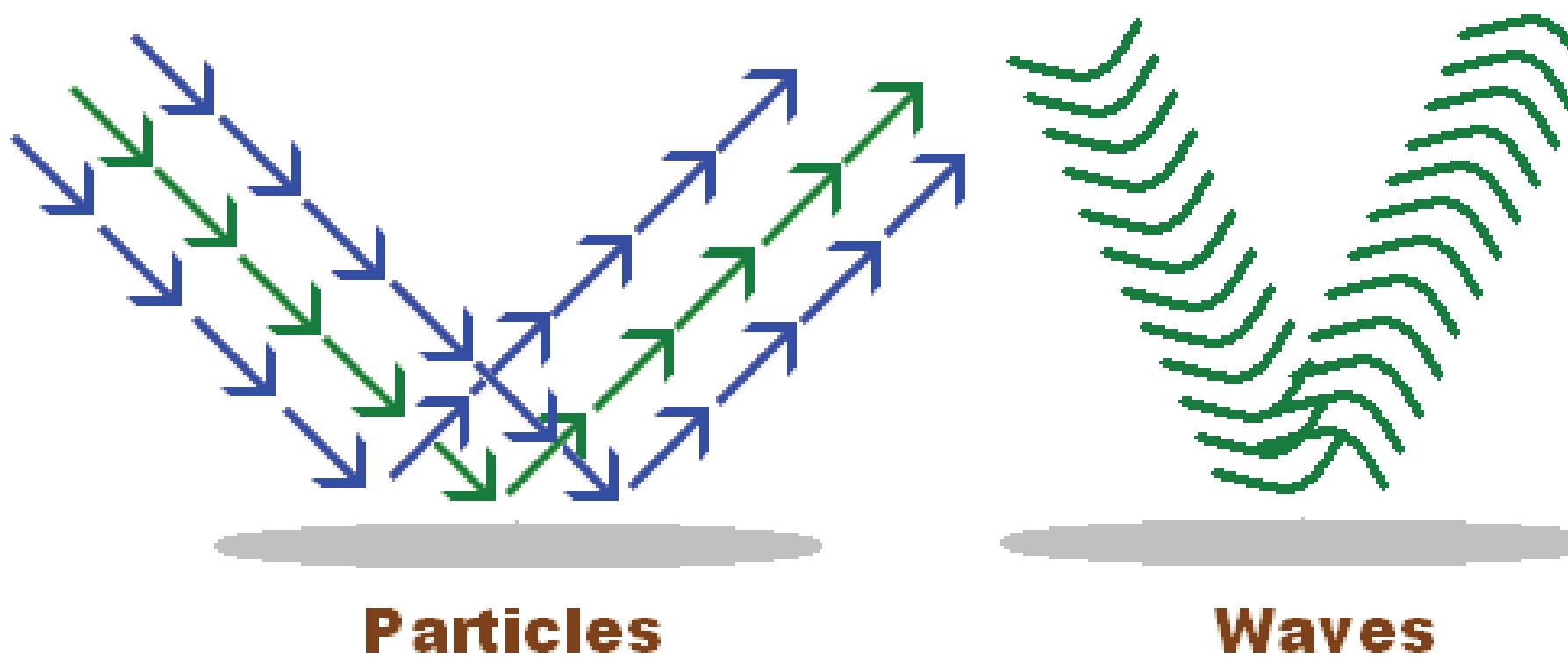
or: Δl is the spatial resolution, f is the focal length of an ideal lens.

Thus: The resolution of an light microscope system is limited by the wavelength of the light used. One of the ways to improve the resolution is to use light with shorter wavelength or use larger lens aperture.

Wave-particle duality of electron

It all started with the De Broglie's hypothesis:

$$\lambda = \frac{h}{p}$$



Particles and wave Reflected by a Mirror

λ is wavelength, h is Planck's constant, and p is momentum.

The original motivation of building an electron microscope came from the shorter wavelength of the electron.

Electron wavelength

Applying the principle of energy conservation to an electron (-e) traveled in voltage E_0 :

$$eE_0 = \frac{h^2}{2m\lambda^2}$$

$$\lambda = \frac{h}{\sqrt{2meE_0}}$$

E_0 = acceleration voltage
 λ = wavelength
 h = Planck's constant

m = electron mass
 e = electron charge

Electron wavelength

Take the relativity into consideration, the wave length is:

$$\lambda = \frac{h}{\sqrt{2m_0eE_r}} \quad E_r = E_0 + \left(\frac{e}{2m_0c^2} \right) E_0^2$$

$$\lambda = \frac{1.22639}{\sqrt{E_0 + 0.97845 \times 10^6 E_0^2}}$$

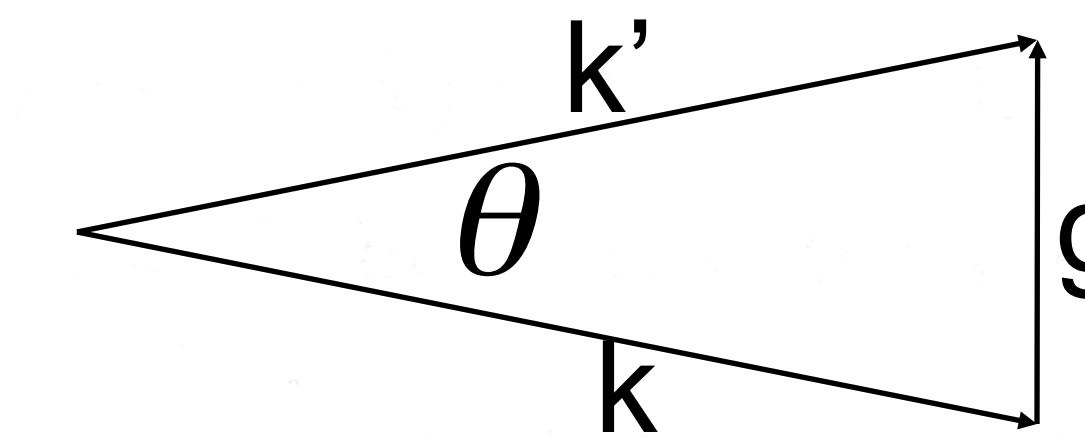
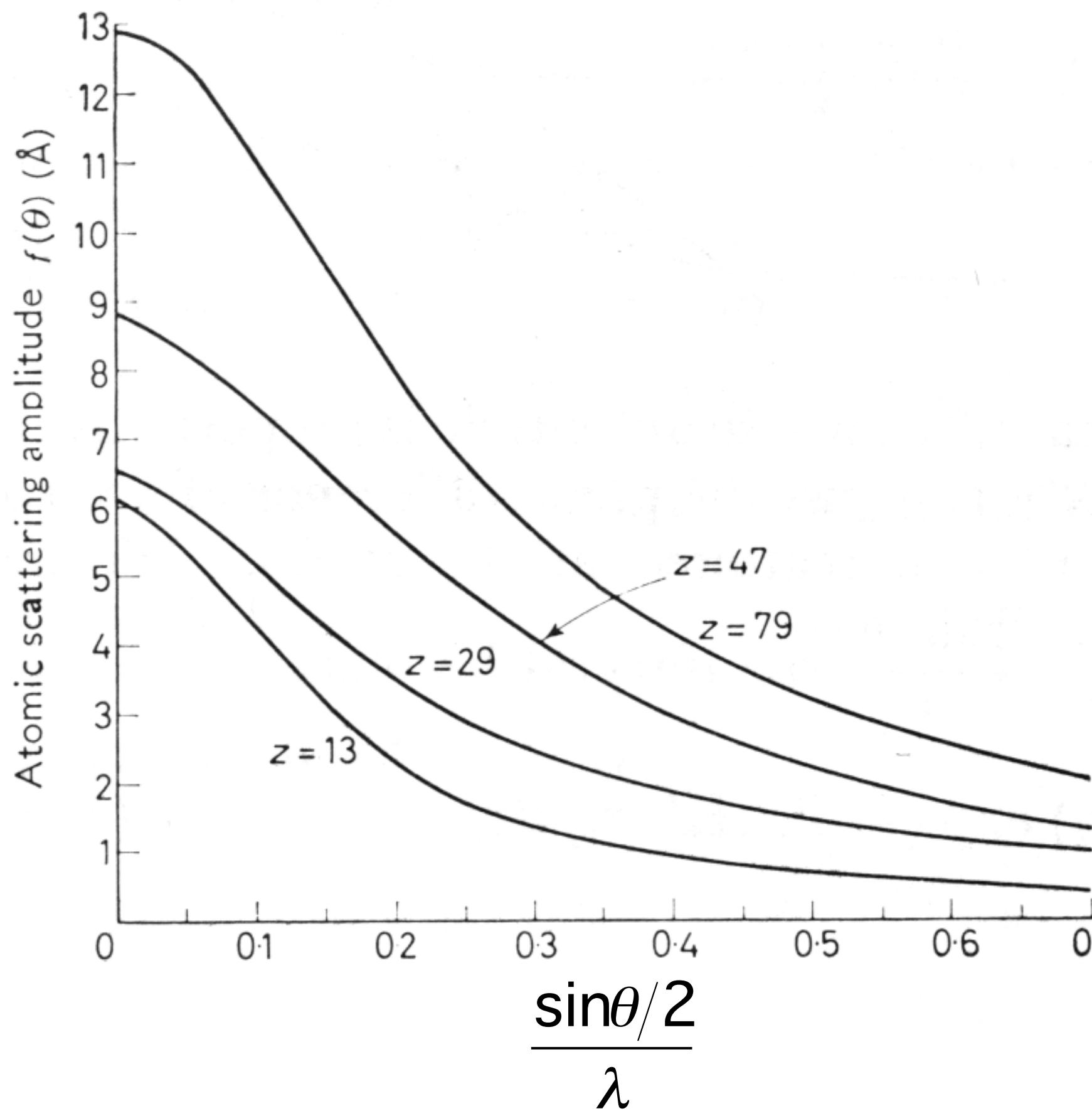
120kV $\lambda=0.033\text{\AA}$; 200kV $\lambda=0.025\text{\AA}$; 300kV $\lambda=0.020\text{\AA}$;

Note that these wavelength is considerably shorter than that used in X-ray crystallography, $\sim\text{\AA}$.

Atomic Scattering Factor for Electrons

Mott formula:

$$f_e(\theta) = \frac{m\epsilon^2}{2h^2} \left(\frac{\lambda}{\sin\theta/2} \right)^2 [Z - f_x(\theta)]$$



$$|\vec{g}| = \frac{2\sin\frac{\theta}{2}}{\lambda} = 2|\vec{k}|\sin\frac{\theta}{2}$$

Figure 4.6. Atomic scattering amplitudes as a function of $\sin\frac{1}{2}\theta/\lambda$ for Al ($Z=13$), Cu ($Z=29$), Ag ($Z=47$) and Au ($Z=79$)

Electron v.s. X-ray

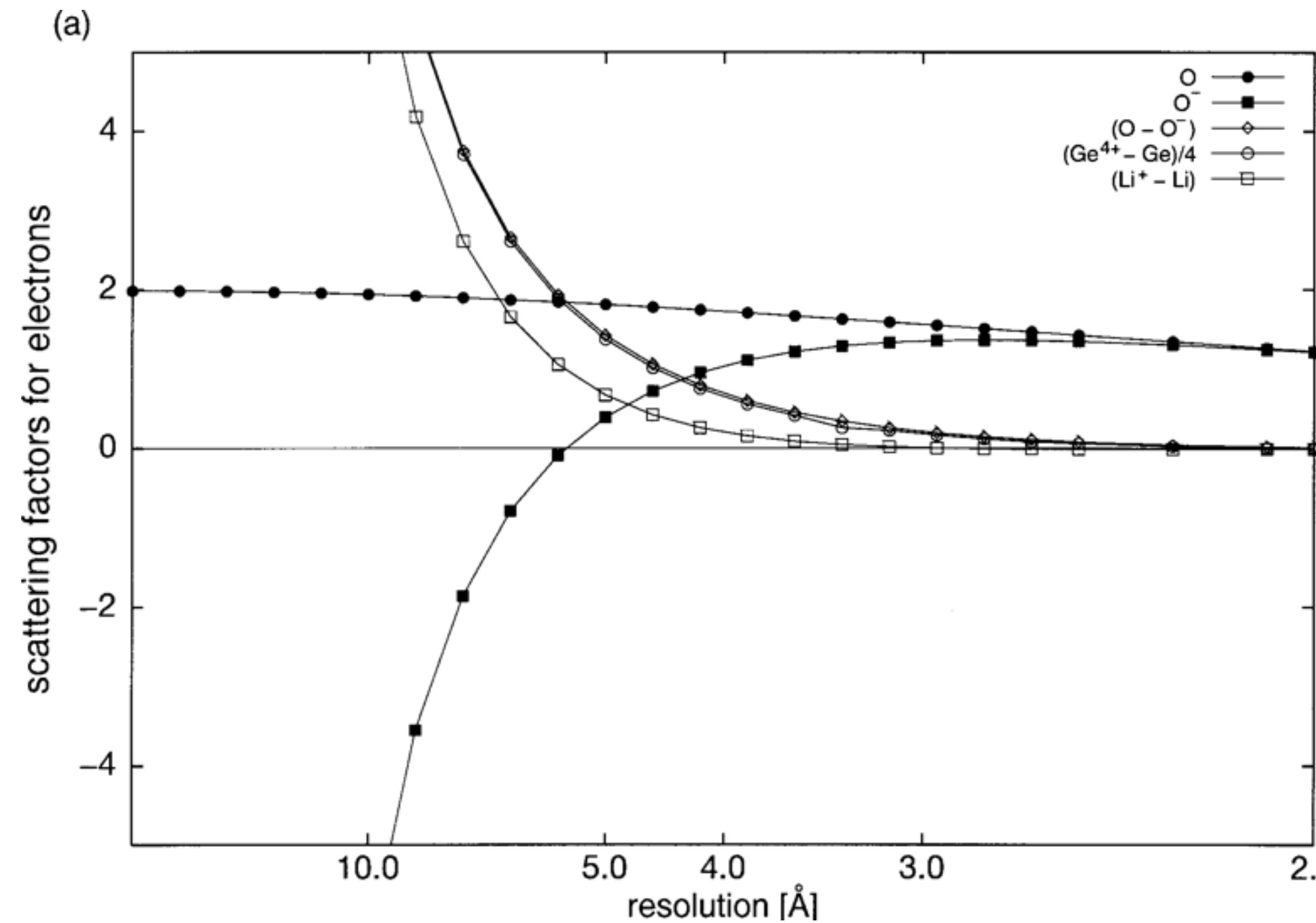
As particles:

- Electrons interact with the potential field of an atom, including shell electrons and nucleus, X-rays interact with only shell electrons;
- Electrons have much larger scattering cross-sections than X-rays; multiple scattering is severer in electron scatterings than in X-ray diffraction; For biological sample, radiation damage is also severer than X-ray diffraction.

As wave:

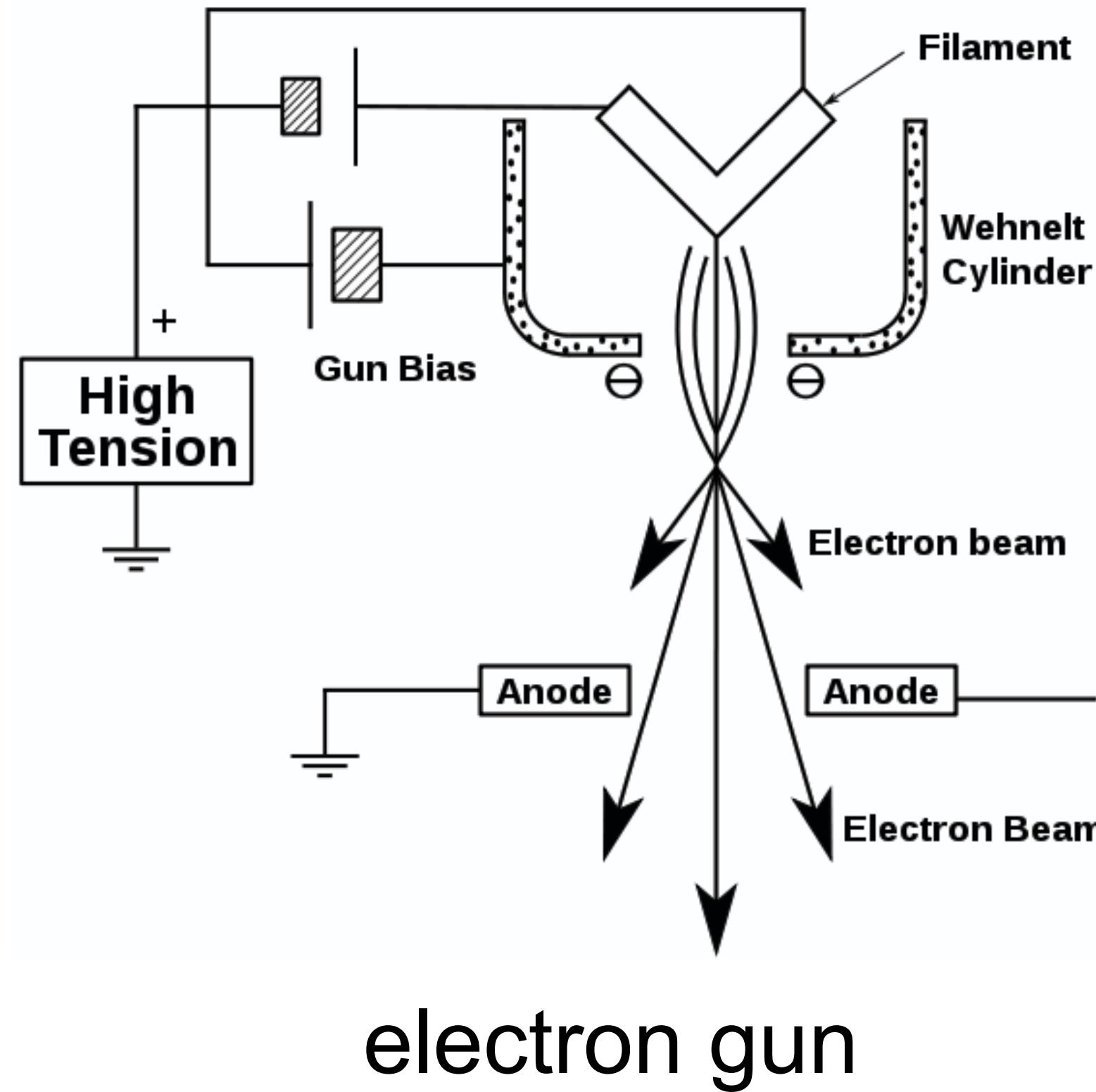
- Electrons can be focused by electromagnetic lens, X-ray can not be focused by lens;

Electron v.s. X-ray

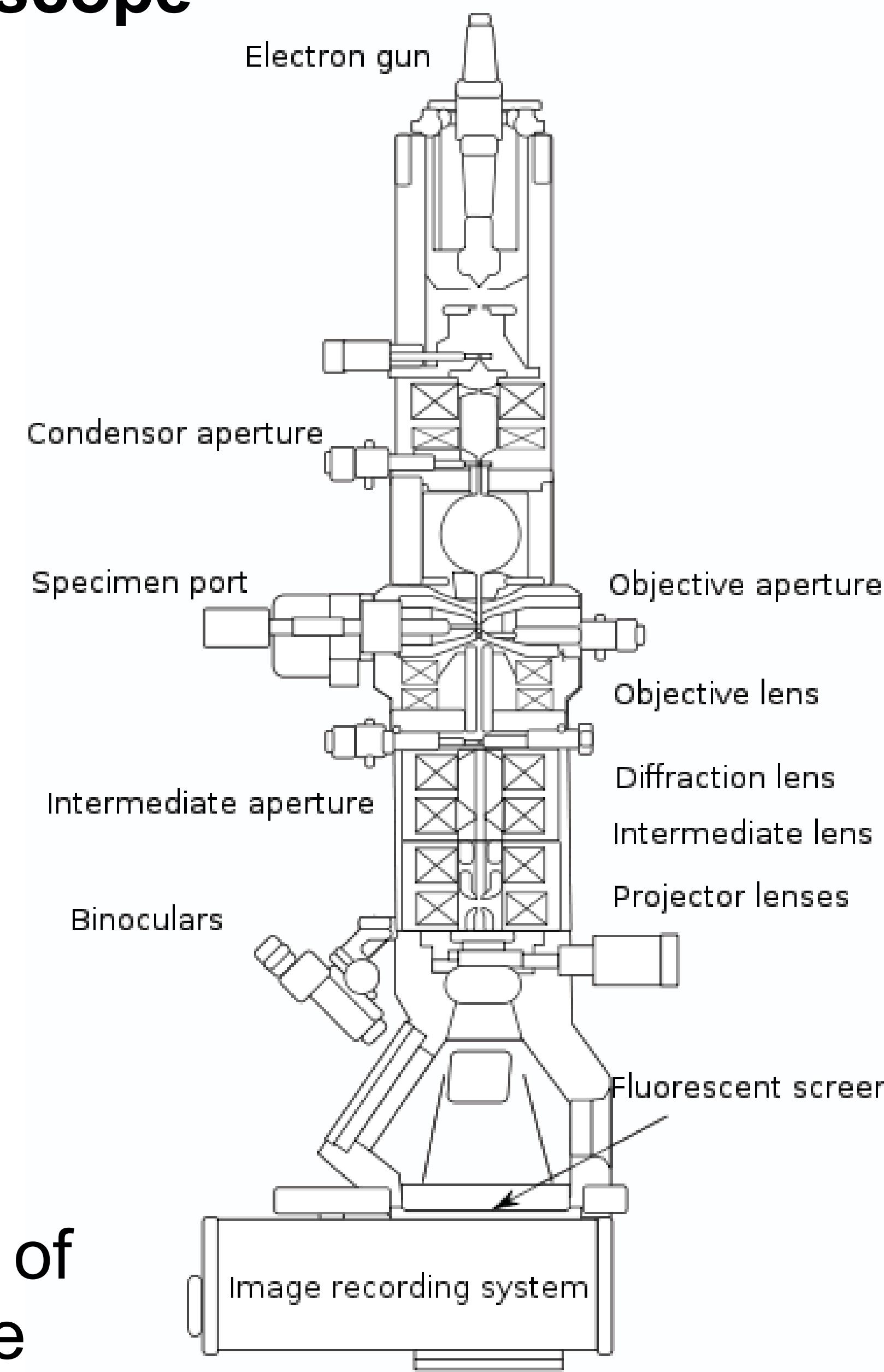


Mitsuoka et al. (1999) JMB, **286**, 861-882.

Electron Microscope

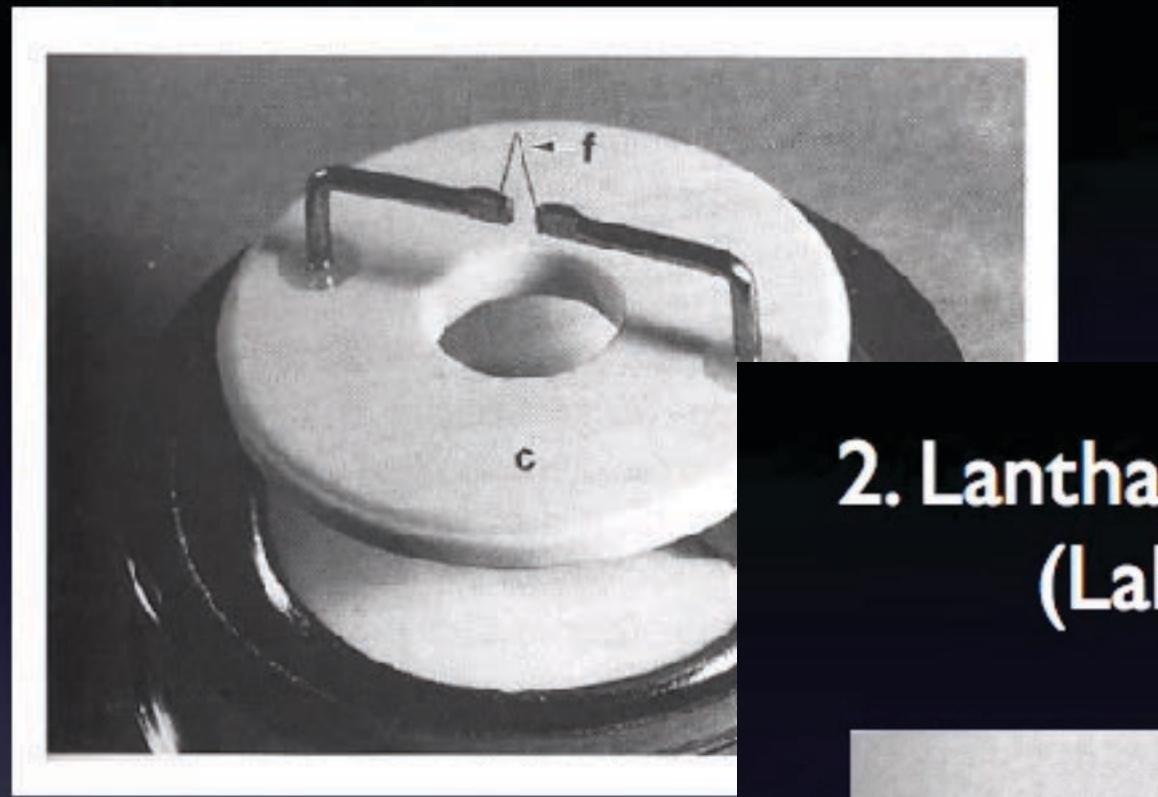


a schematic drawing of
electron microscope



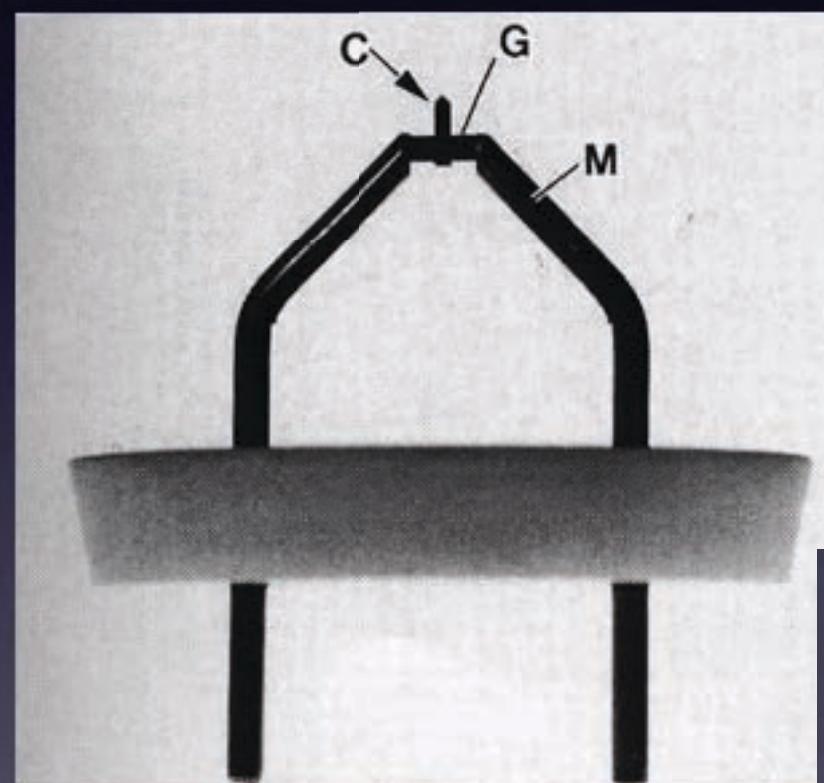
Electron source

1. Tungsten filaments

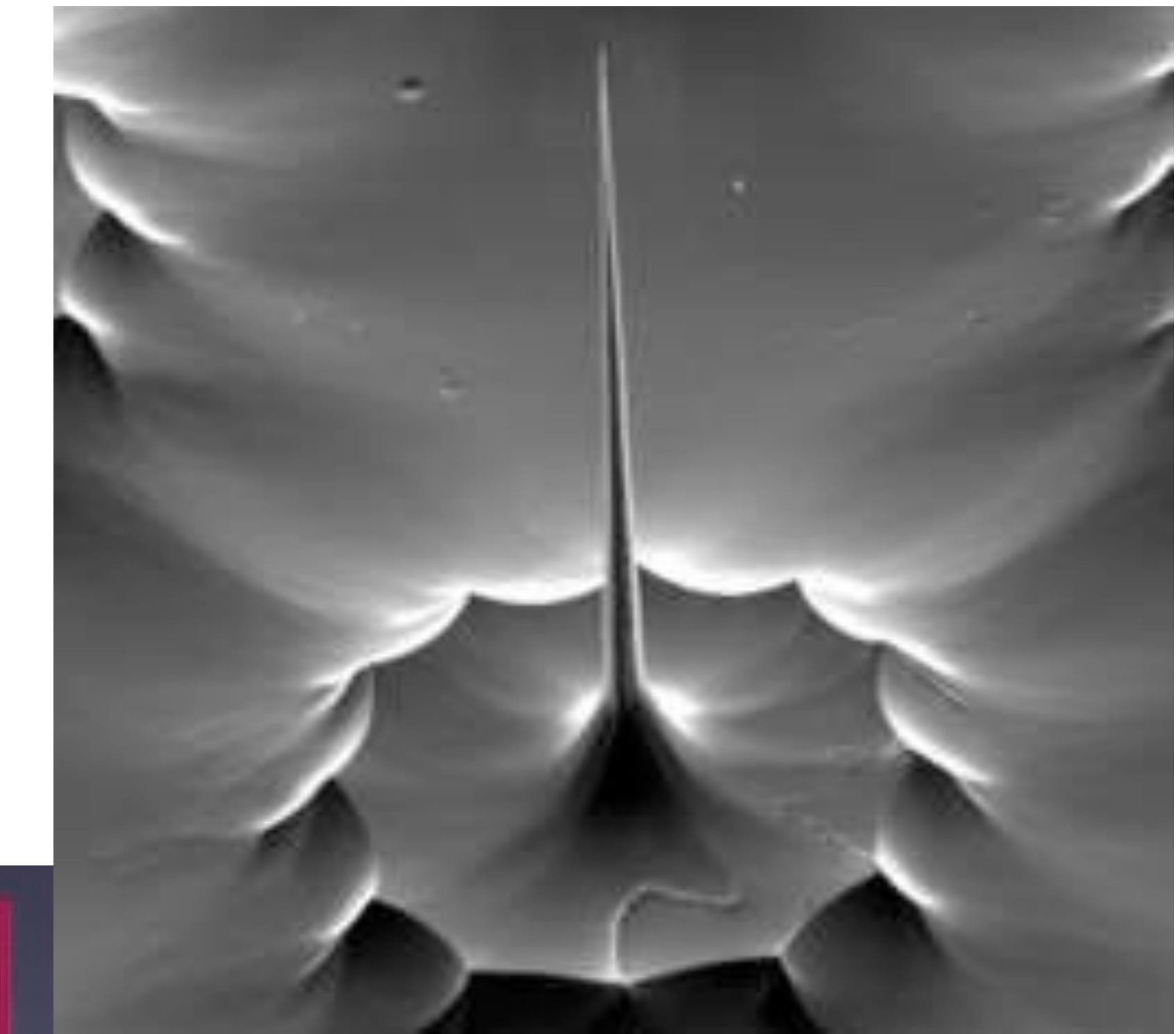


Bozzola and Russel

2. Lanthanum hexaboride
(LaB_6) crystals



Bozzola and Russell, Fig. 6.1

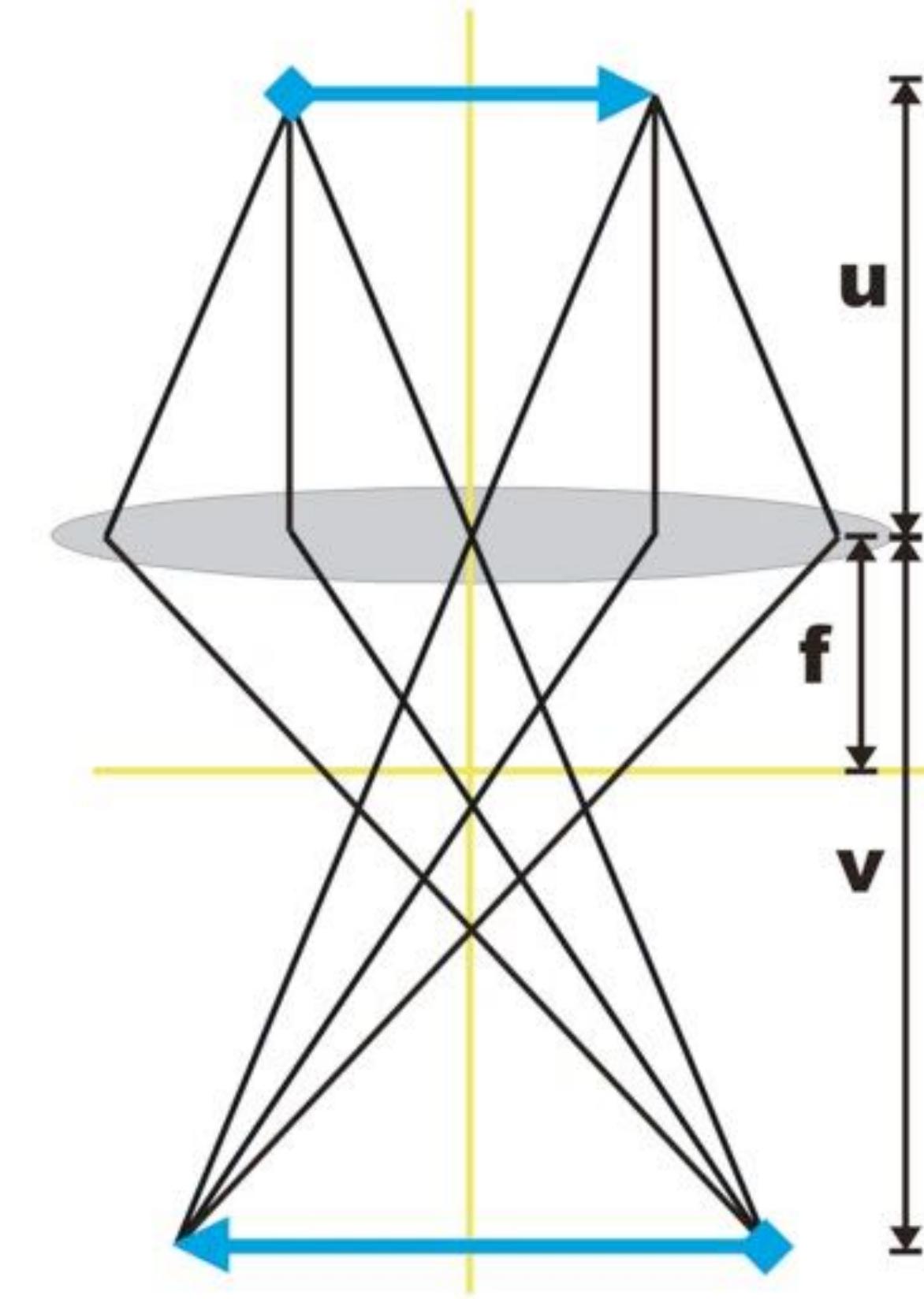
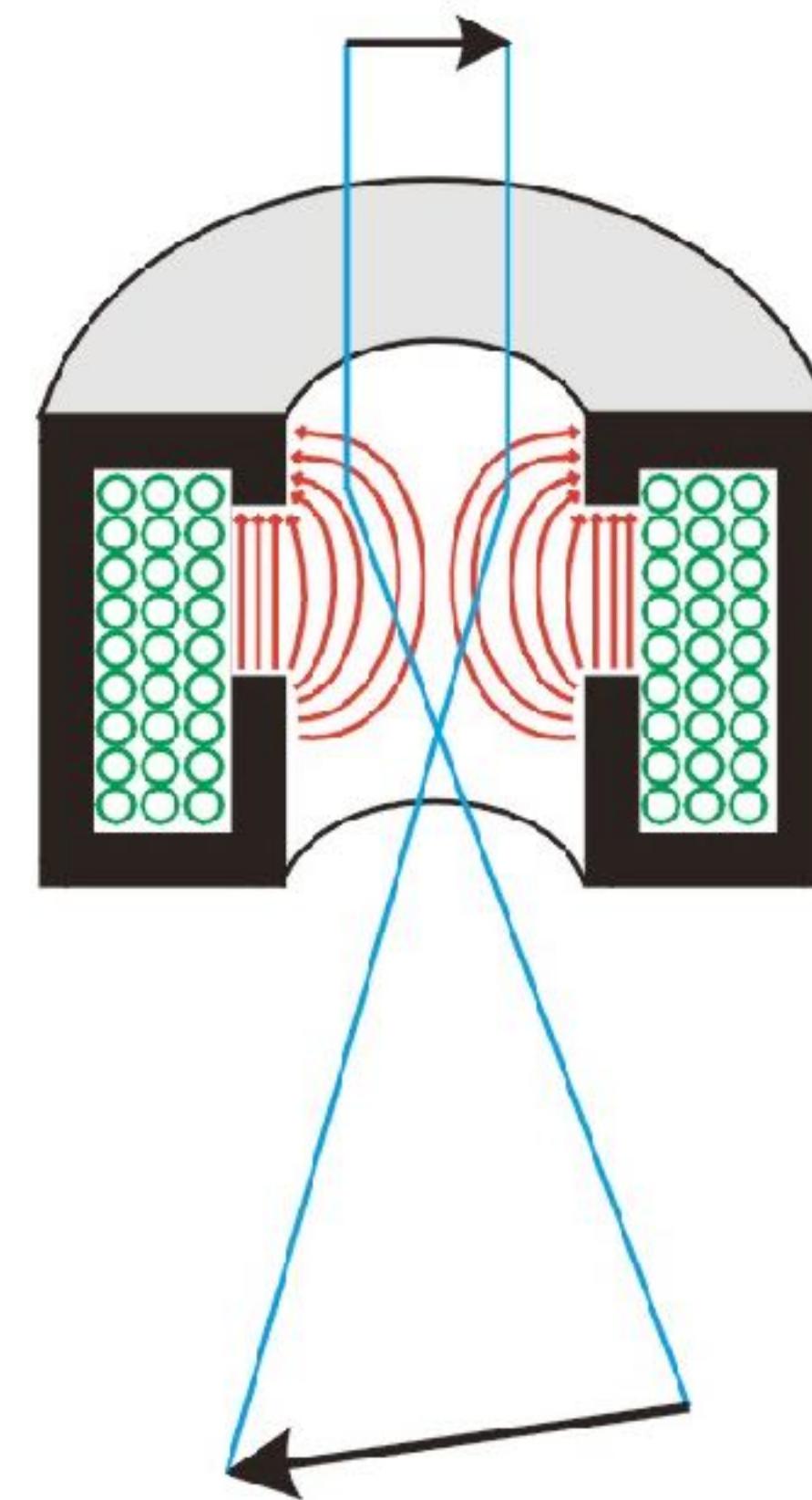


3. Field emission gun



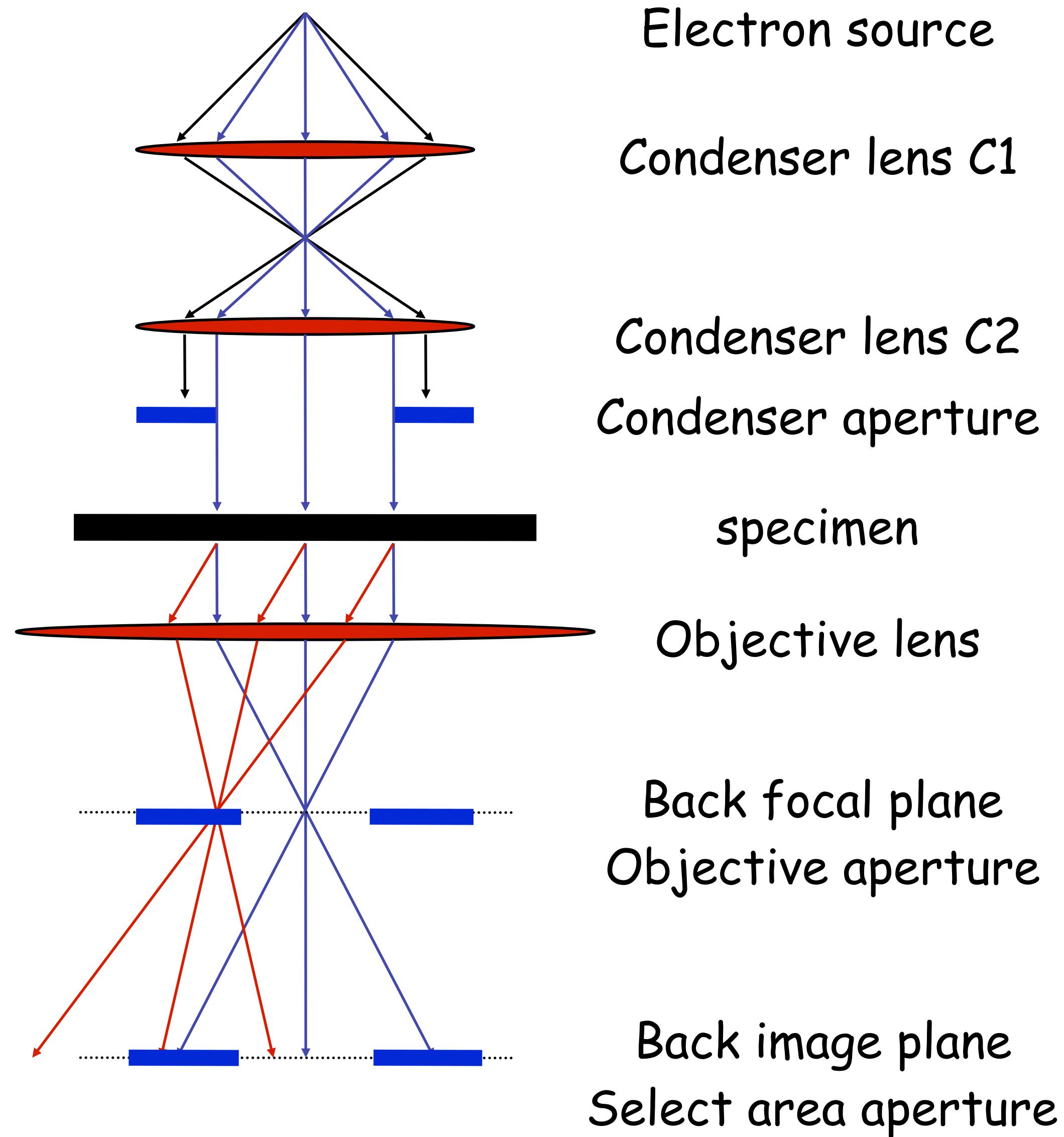
<http://www.fisica.unige.it/~rocca/Didattica/Laboratorio>

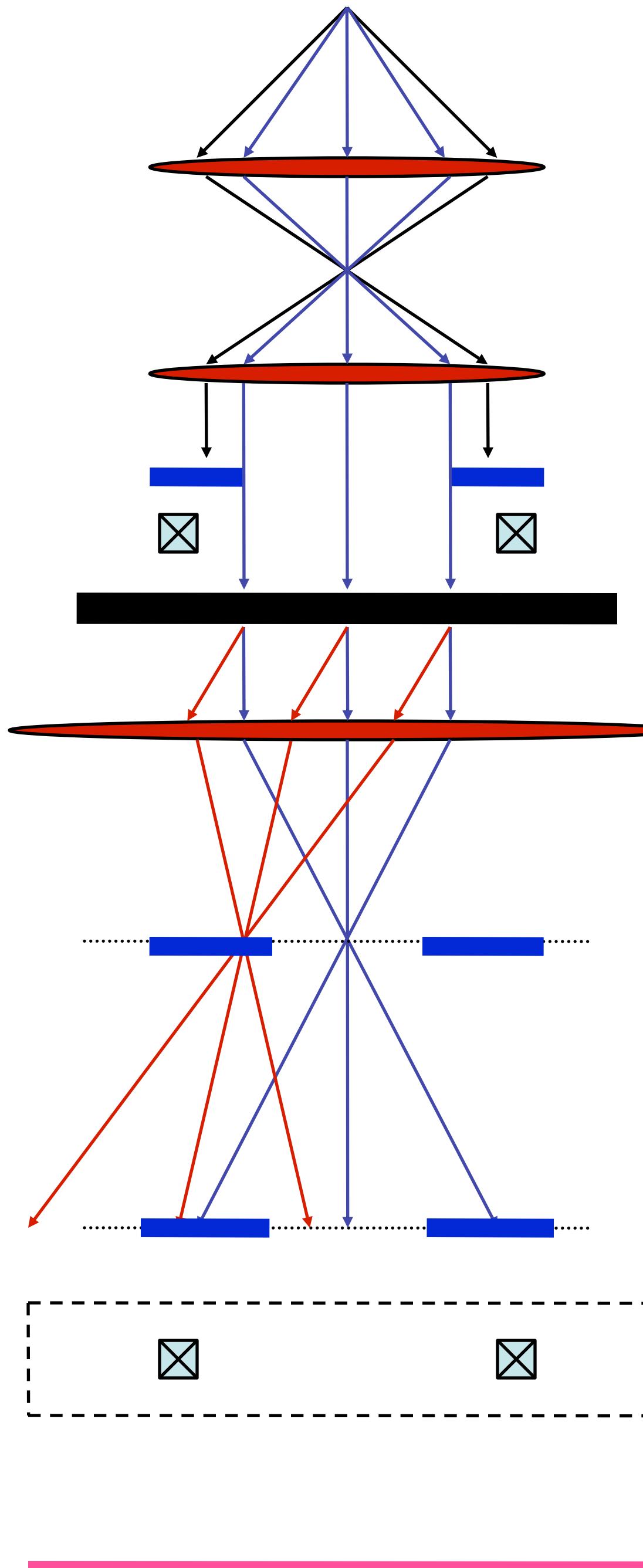
Electromagnetic lens



* The focal length of a electromagnetic lens can be easily adjusted by changing the lens current.

Optic system in an electron microscope





Additional lens in the electron microscope

Beam shift coils

Projection lens system
Image shift/diffraction shift coils

screen

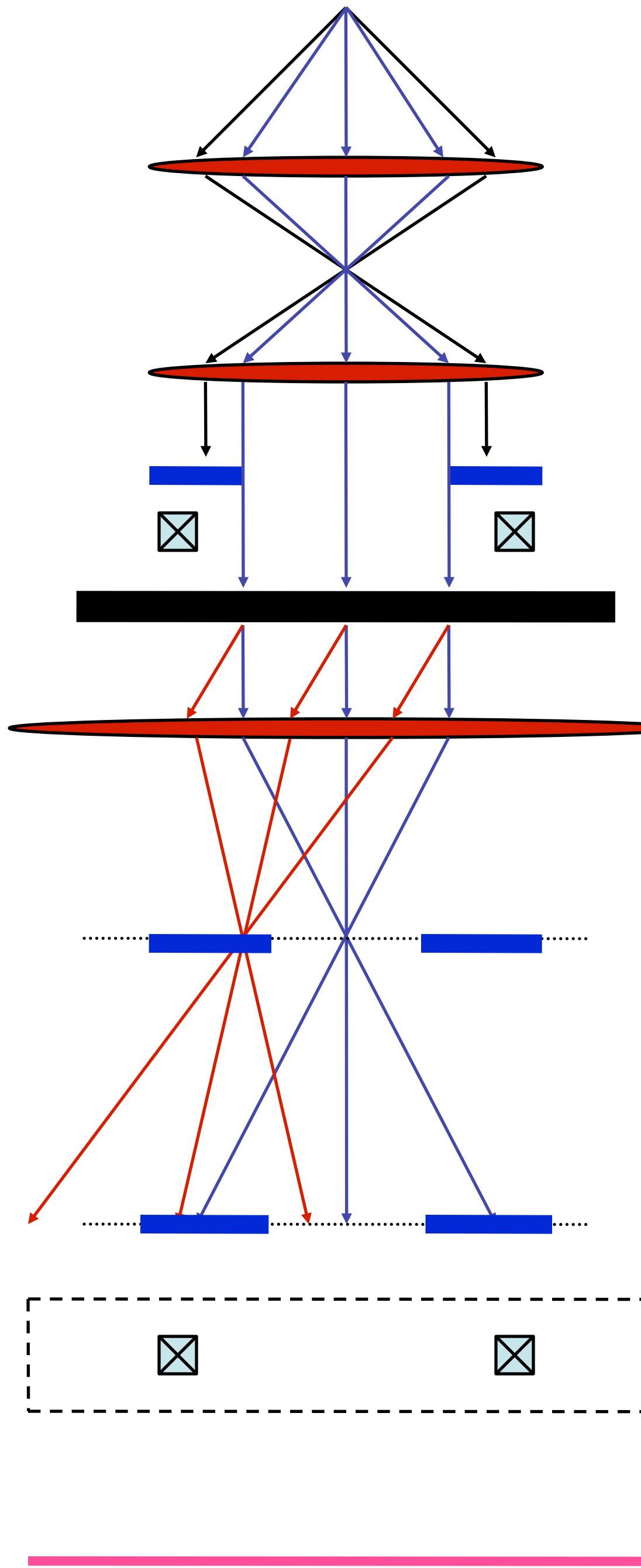


Image mode

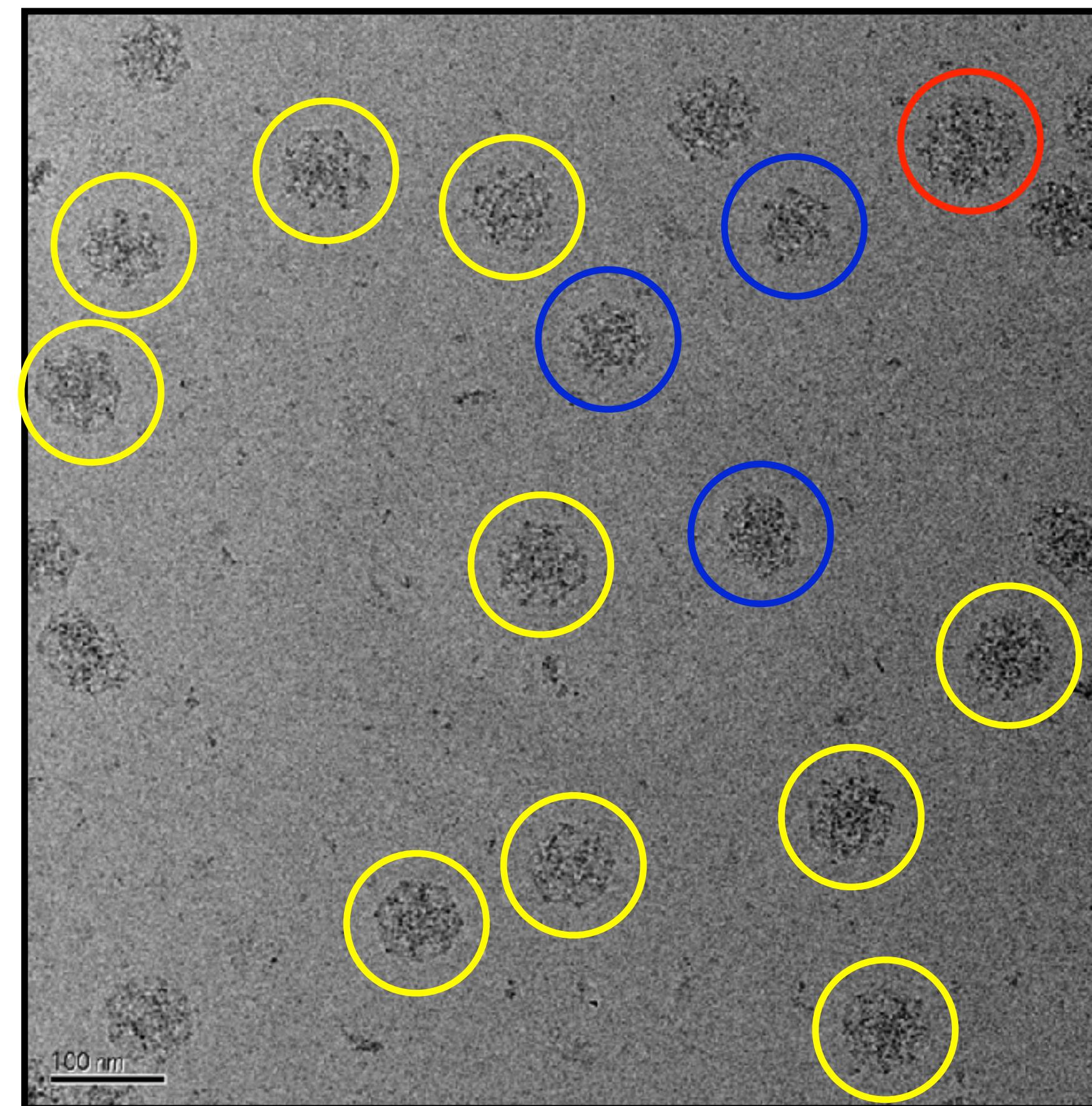
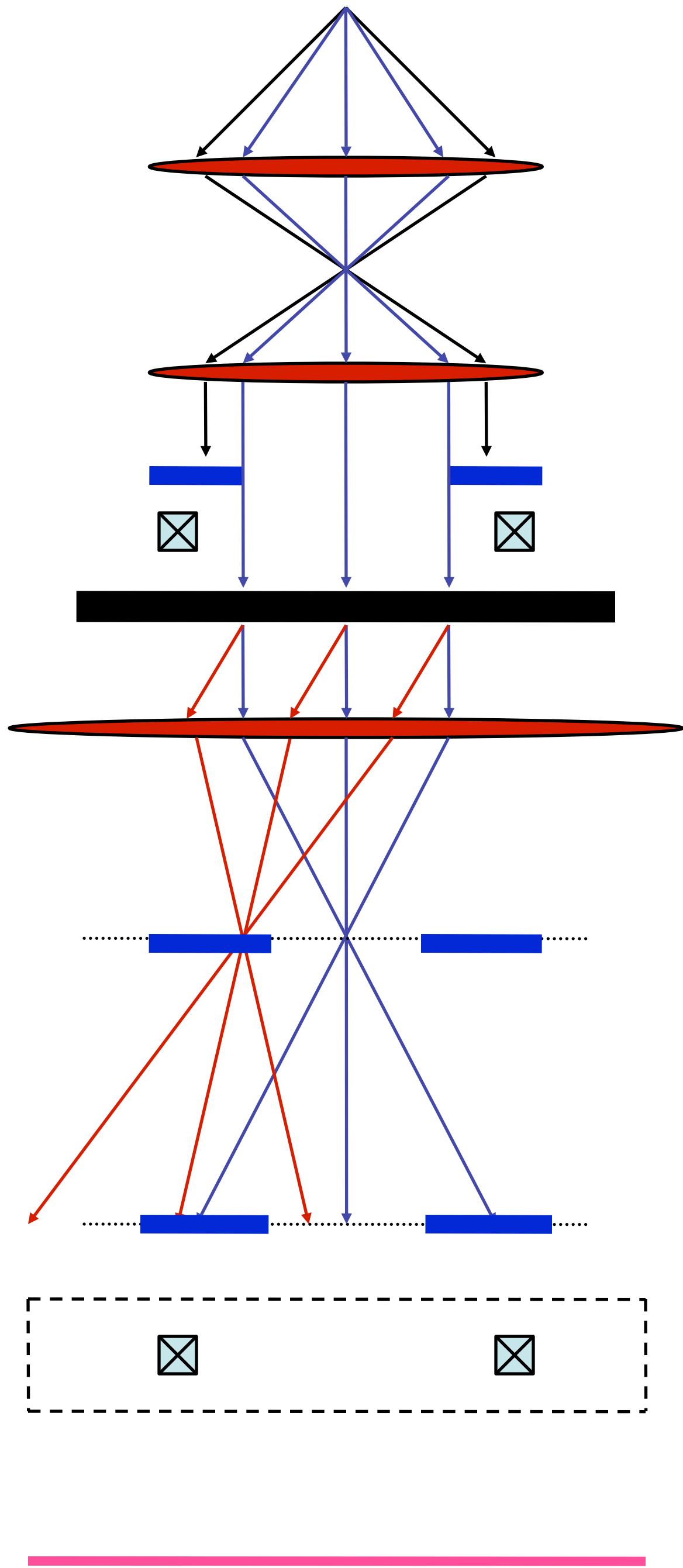
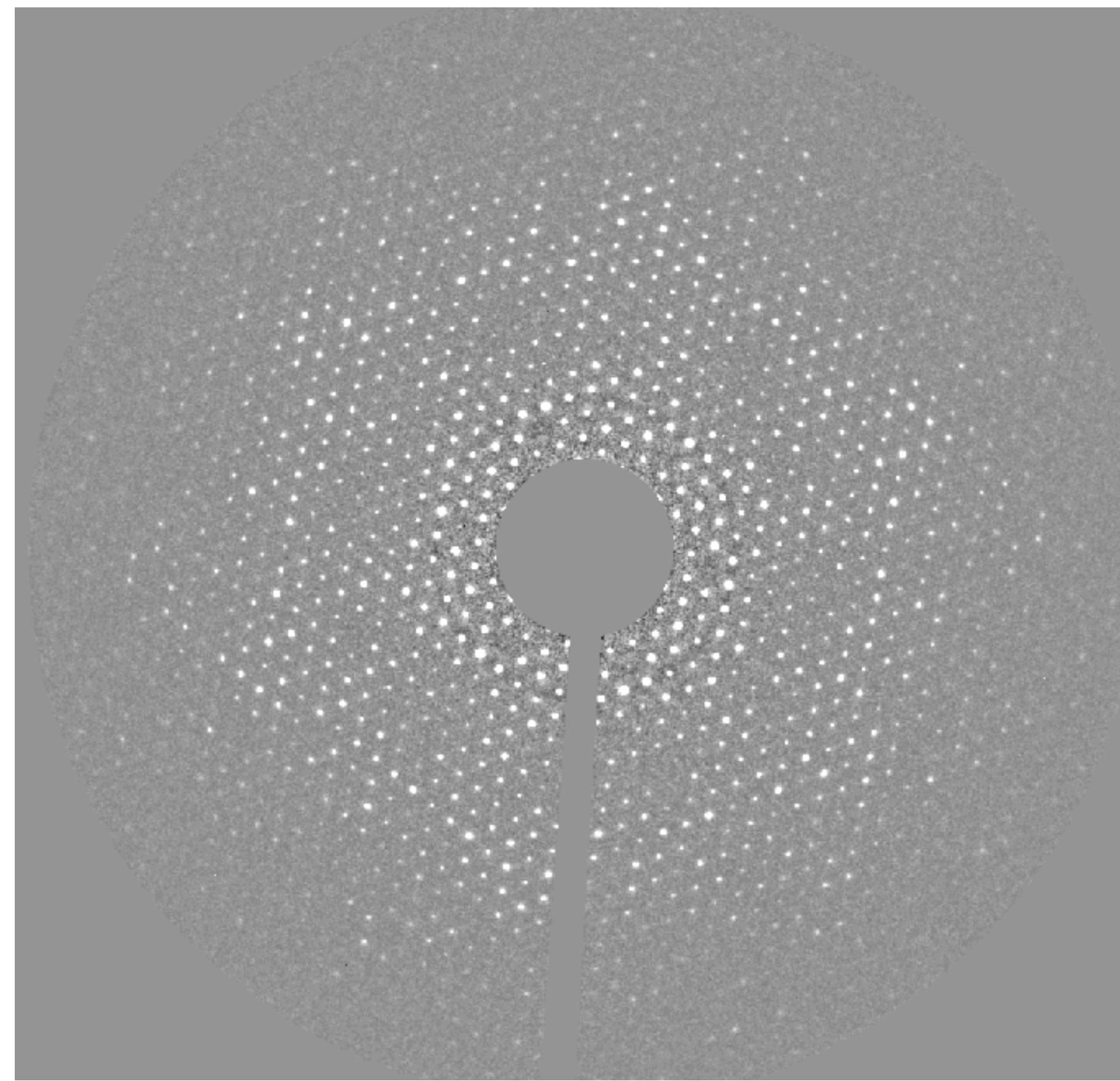


Image mode

Clathrin coat



Diffraction mode



Diffraction mode

EM images are projections



Image formation

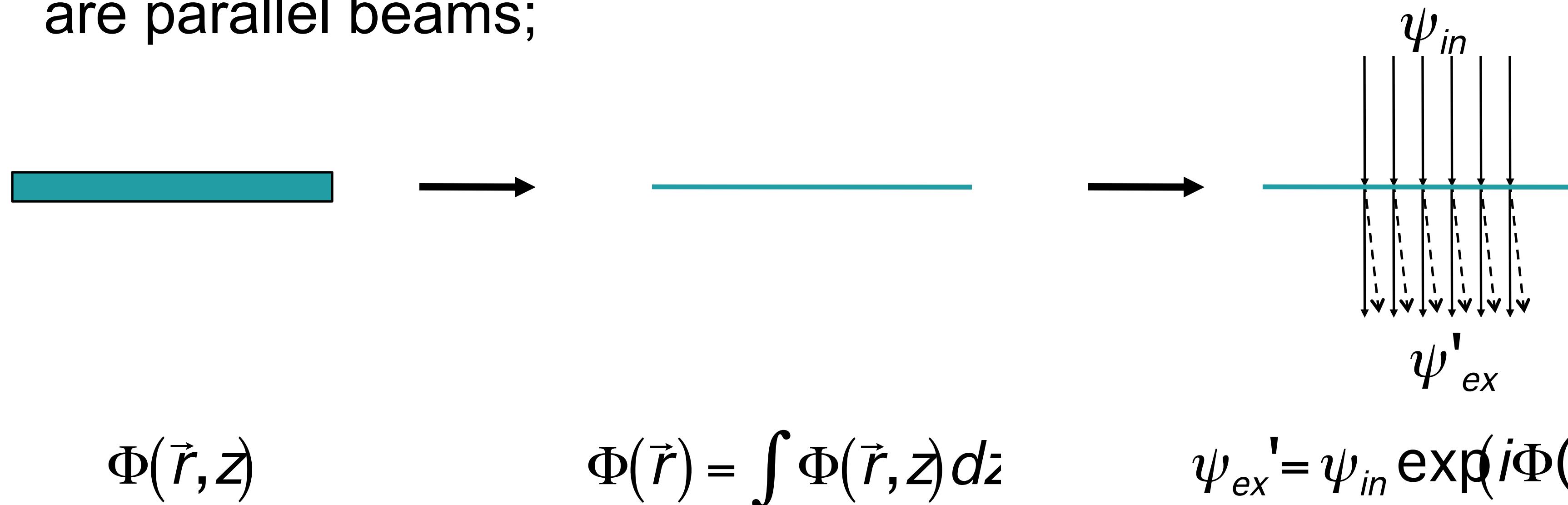
The image formation in the electron microscope can be treated as two separate processes:

- 1) The interaction of the incident beam with the specimen, described by the weak-phase object approximation, which is the theory used mostly to describe the image formation of thin specimen with light elements, such as a biological sample.
- 2) The propagation of the electron beam from exit plane of the specimen to the back image plane of the focus lens.

Weak-phase object approximation

This is a highly simplified theory based on the so-call weak-phase object, which is a very thin specimen formed mostly by low- and medium-weight molecules.

Suppose: 1) the specimen is very thin so that $\Phi(\vec{r}, z)$ can be approximated by $\Phi(\vec{r})$; 2) both in-coming and exiting beams are parallel beams;



$$\psi_{ex}' = \psi_{in} \exp(i\Phi(\vec{r})) \quad (2)$$

$$\psi_{ex}' = \psi_{in} \left[1 + i\Phi(\vec{r}) - \frac{1}{2}\Phi^2(\vec{r}) + \dots \right] \quad (3)$$

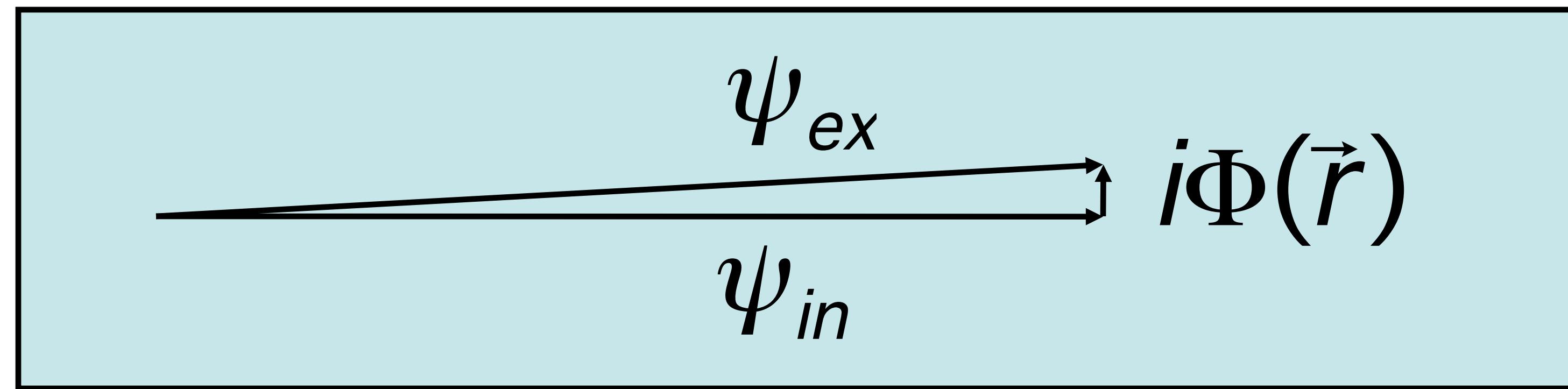
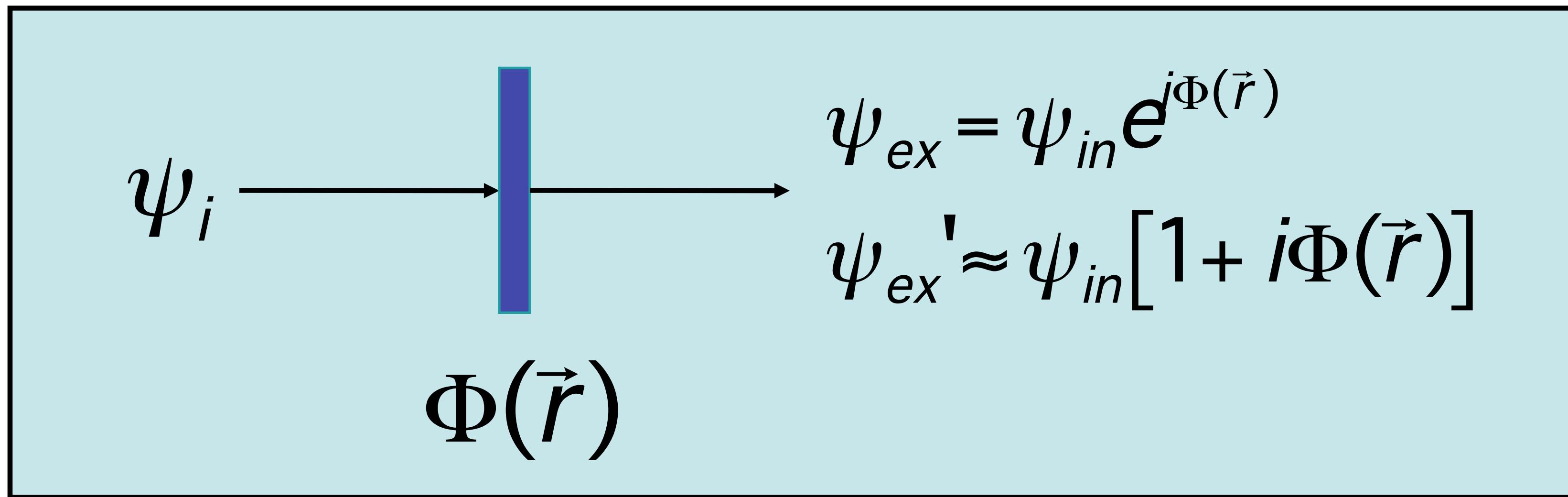
The first term in (3) represents the central unscattered beam, the second term the kinematically scattered beam and the higher terms are for the dynamical scattering. The weak phase object approximation assumes that $\Phi(r) \ll 1$, the phase shift is so small that the following approximation will work:

$$\psi_{ex}' \approx \psi_{in} [1 + i\Phi(\vec{r})] \quad (4)$$

Taking absorption into consideration:

$$\psi_{ex}' = \psi_{in} \exp(i\Phi(\vec{r}) - \mu(\vec{r})) \quad (5)$$

$$\psi_{ex}' \approx \psi_{in} [1 - \mu(\vec{r}) + i\Phi(\vec{r})] \quad (6)$$



$$I_{ex} = |\psi_{ex}|^2 = |\psi_{in}|^2 = I_{in}$$

Image formation

At exit plane of specimen:

$$\psi'_{ex}(\vec{r}) \approx \psi_{in}[1 + i\Phi(\vec{r})]$$

At back focal plane:

$$\Psi_{bf}(\vec{k}) = F[\psi'_{ex}(\vec{r})]$$

At back image plane:

$$\psi_{im}(\vec{r}) = F^{-1}[\Psi_{bf}(\vec{k})]$$

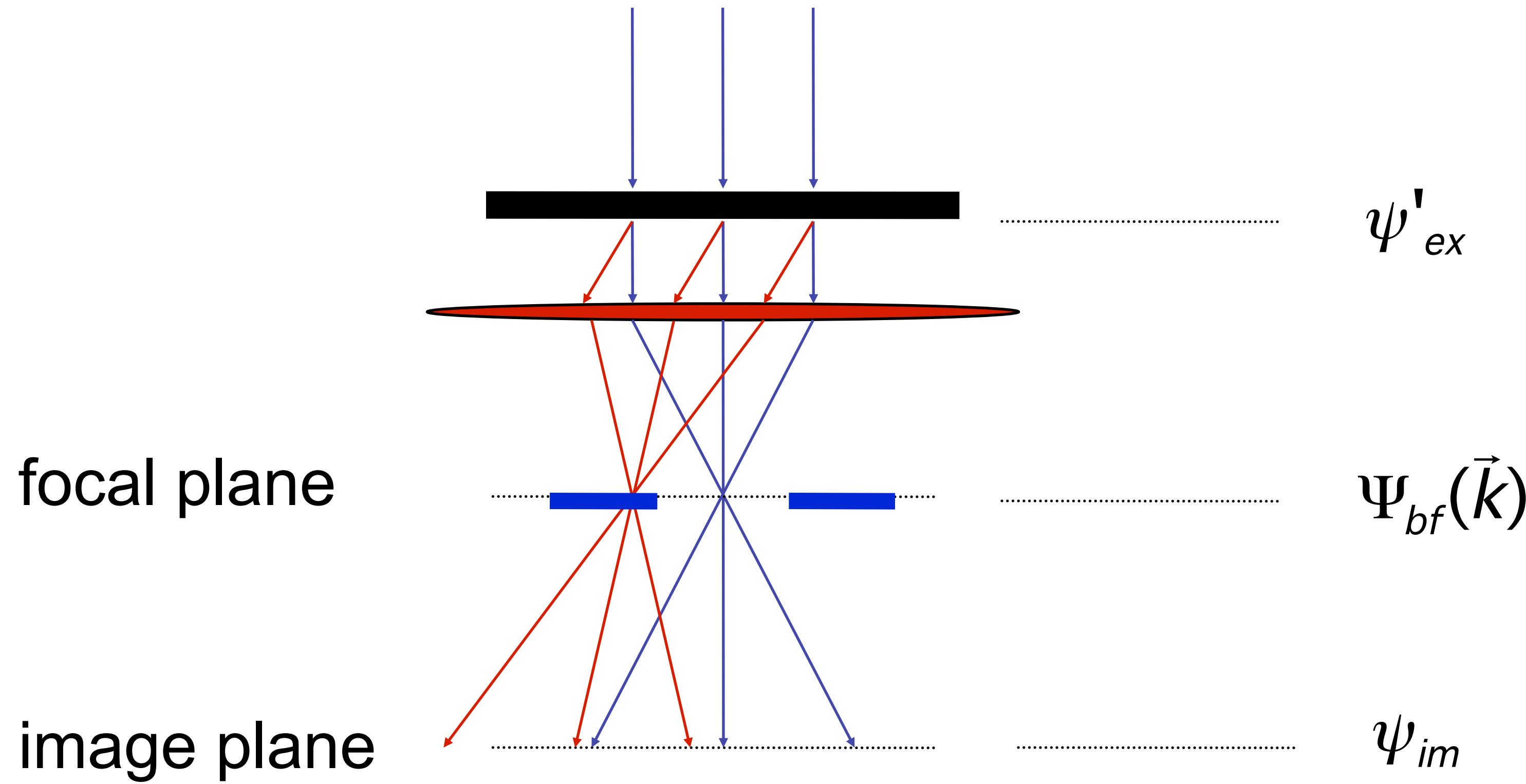


Image formation

The plane wave ψ' of exit-beam travel through objective lens to the back focal plane. The wave function at back focal plane of the objective lens is the Fourier transform of the exit wave:

$$\Psi_{bf}(\vec{k}) = F(\psi'_{ex}(\vec{r})) = F[1 + i\Phi(\vec{r})] = \delta(\vec{k}) + iF(\Phi(\vec{r})) \quad (7)$$

However the lens aberration and defocusing generate an extra phase shift to the scattered beam:

$$\begin{aligned} \gamma(\vec{k}) &= 2\pi\chi\vec{k} \\ \chi(k, \varphi) &= \frac{1}{2}\lambda\left[\Delta z + \frac{1}{2}\sin 2(\varphi - \varphi_0)\right]k^2 + \frac{1}{2}\lambda^3 C_s k^4 \end{aligned} \quad (8)$$

Together with the aperture function $A(\vec{k})$ the wave function at back focal plane will become:

$$\Psi_{bf}(\vec{k}) = F(\psi'_{ex}) A(\vec{k}) \exp(\varrho \pi i \chi \vec{k}) \quad (9)$$

Then, the wave function in the back image plane of the lens is the reverse Fourier transform of the wave function at back focal plane (\otimes is for convolution):

$$\begin{aligned} \psi_{im}(\vec{r}) &= F^{-1}\{F(\psi'_{ex}) A(\vec{k}) \exp(\varrho \pi i \chi \vec{k})\} \\ &= 1 + i\Phi(-\vec{r}) \otimes J_0(\vec{r}) \otimes F^{-1}[\exp(\varrho \pi i \chi \vec{k})] \end{aligned} \quad (10)$$

The observed intensity in the image is then:

$$\begin{aligned} I_i(\vec{r}) &= \psi_i(\vec{r}) \psi_i^* \\ &= 1 + 2\Phi(-\vec{r}) \otimes J_0(\vec{r}) \otimes F^{-1}[\sin(\varrho \pi \chi \vec{k})] \end{aligned} \quad (11)$$

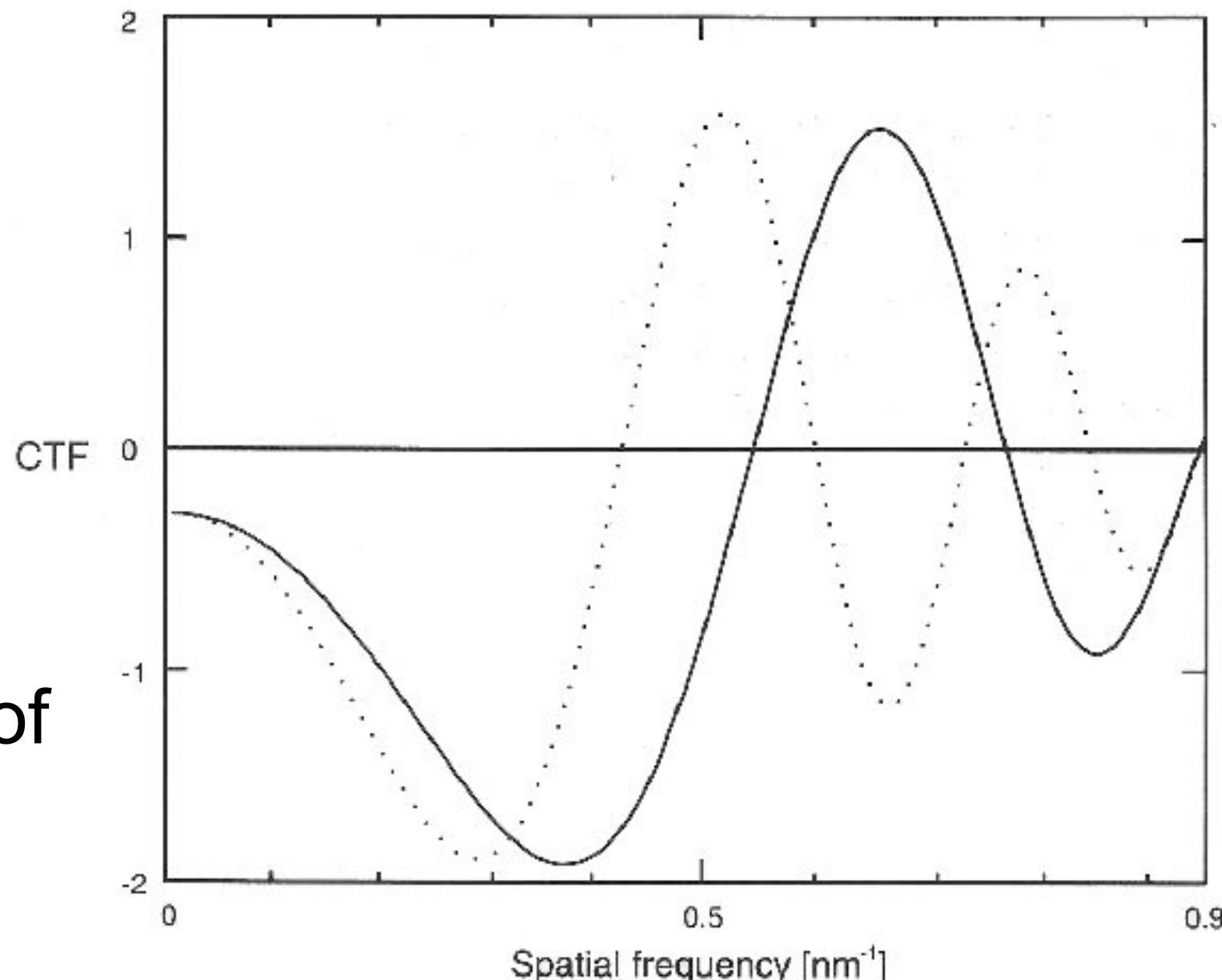
Contrast Transfer Function (CTF)

$$CTF = \sin(2\pi\chi k)$$

The intensity of a recorded image is directly related to the projection of specimen (good!) but modified by the FT of CTF (bad!).

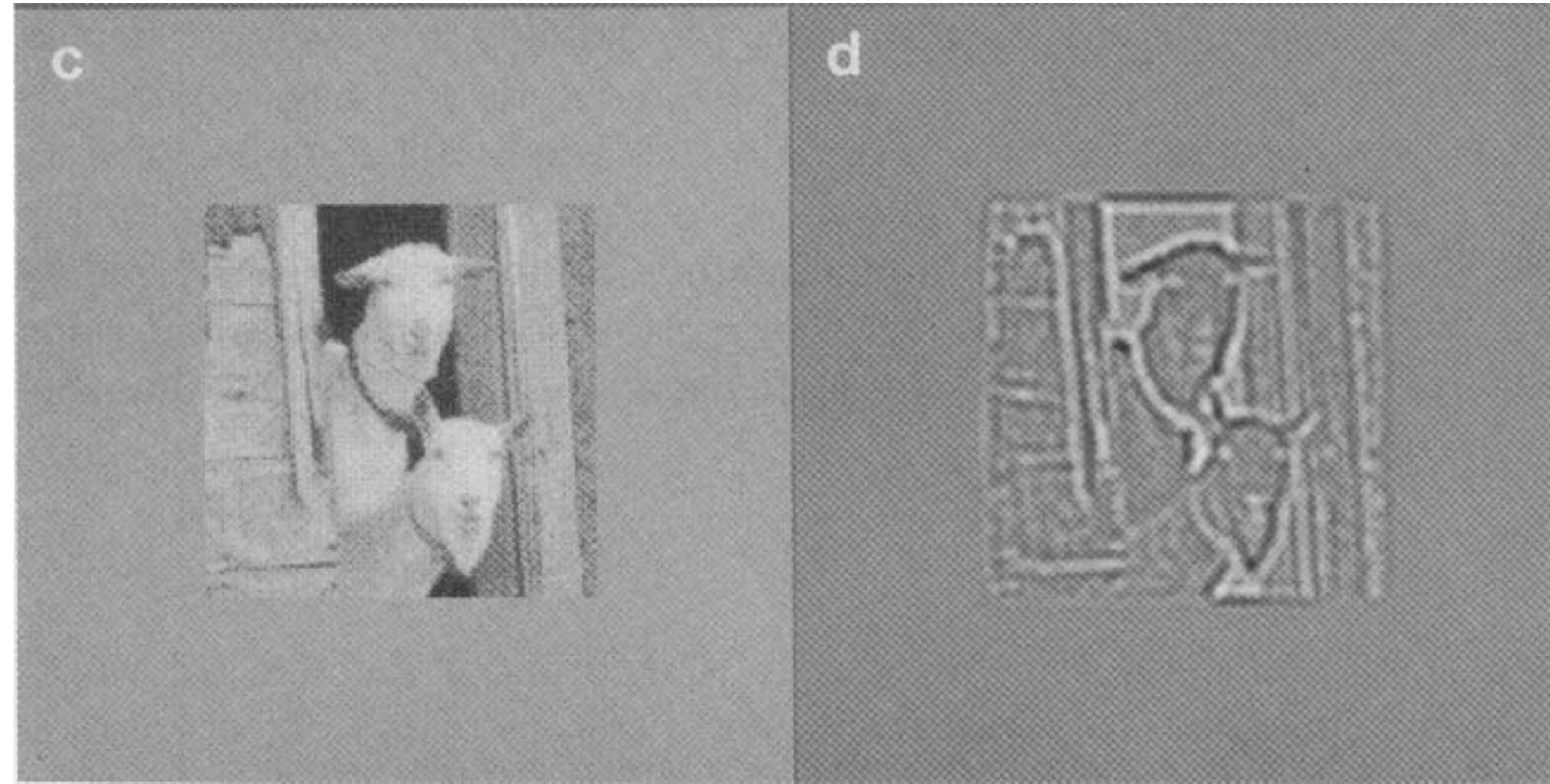
$$I_i(\vec{r}) = \psi_i(\vec{r})\psi_i^*$$

$$= 1 + 2\Phi(-\vec{r}) \otimes J_0(\vec{r}) \otimes F^{-1}(CTF) \quad (12)$$



WebCTF: <http://jiang.bio.purdue.edu/ctfsimu>

What is this CTF thing anyway and why do I care?

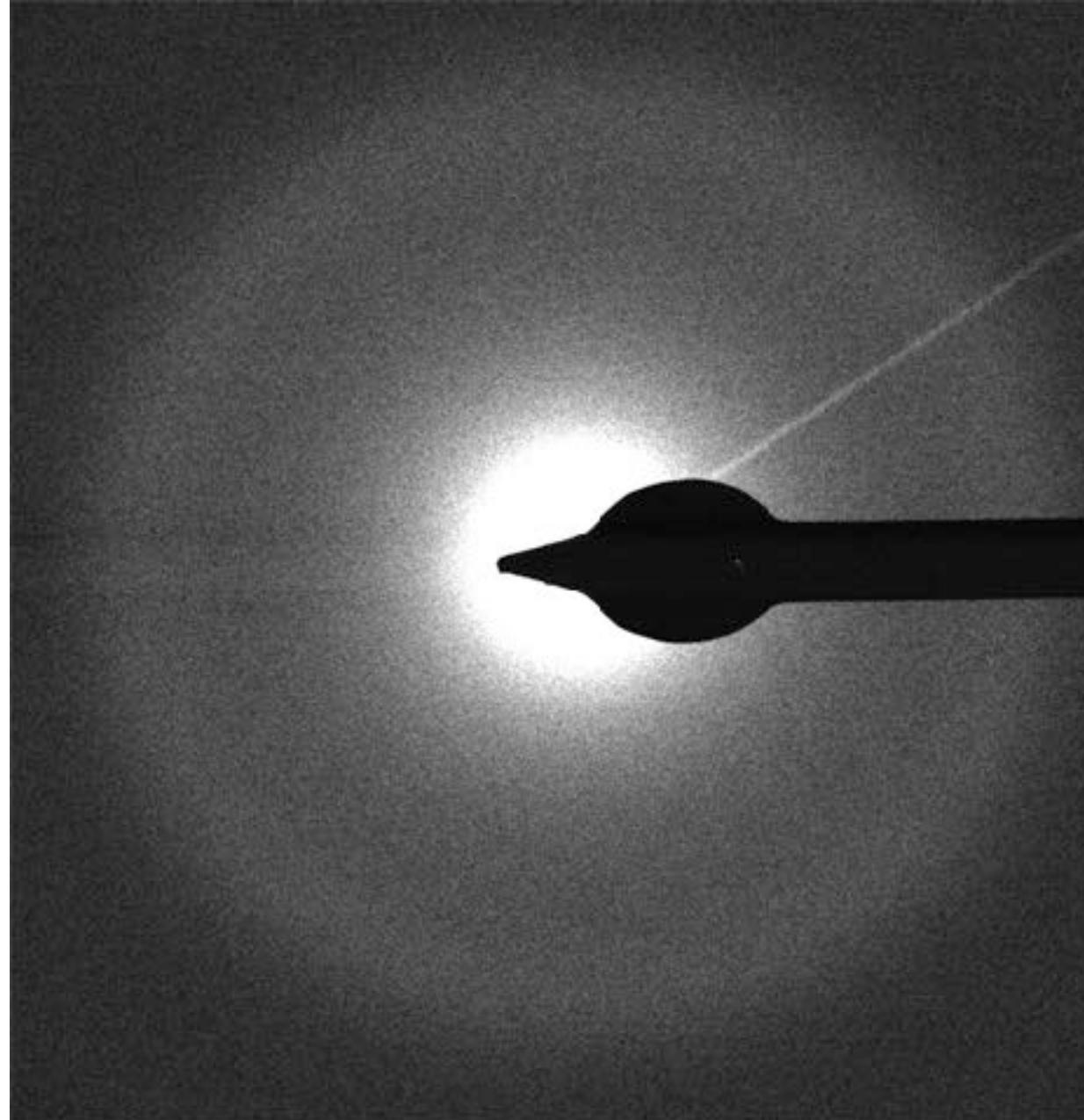


Distortions of CTF to the image are:

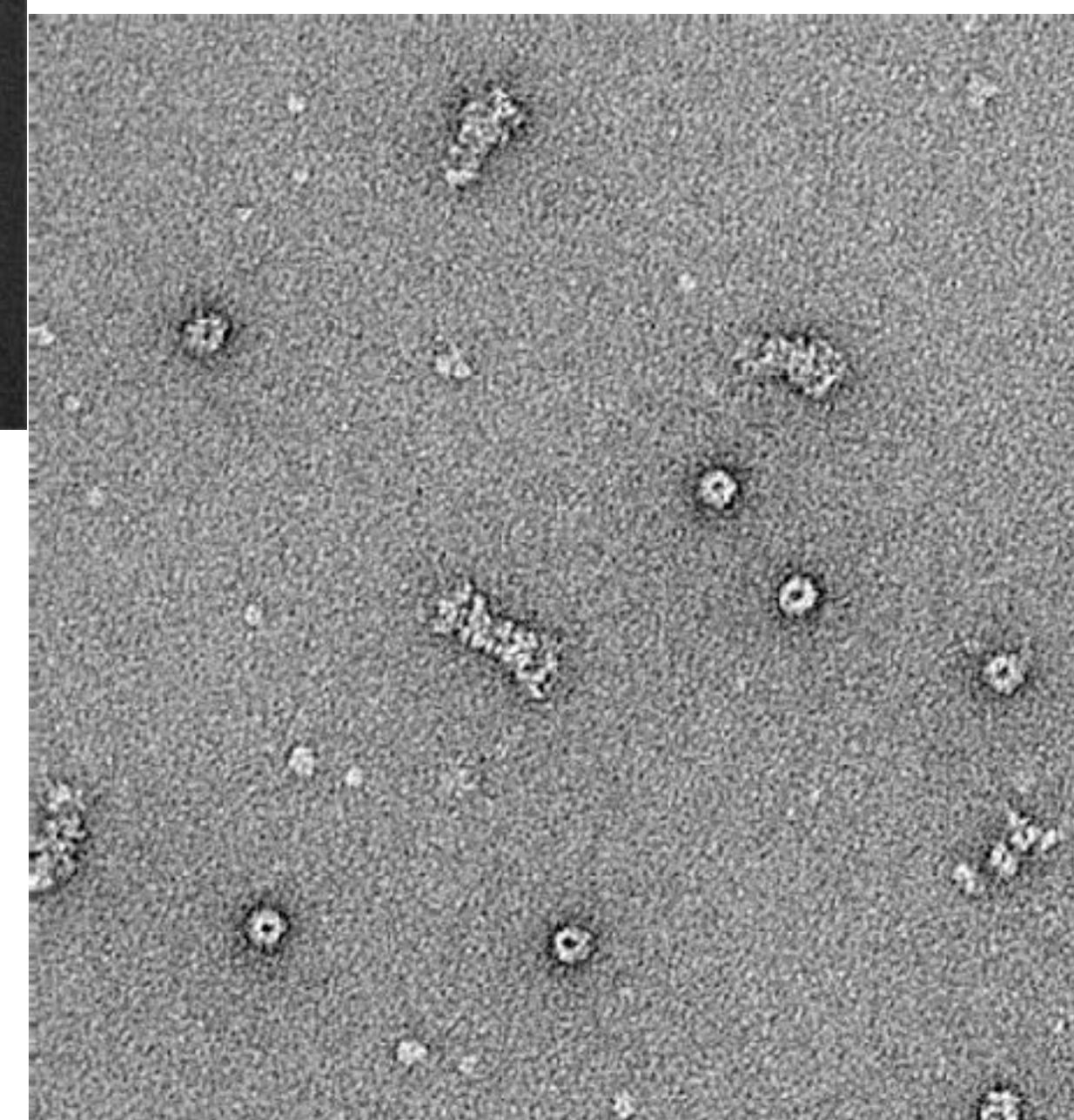
- 1) Contrast reverve of large area; 2) diminished contrast in large area; 3) edge enhancement and 4) appearance of fringes along the borders.

From Joachim Frank

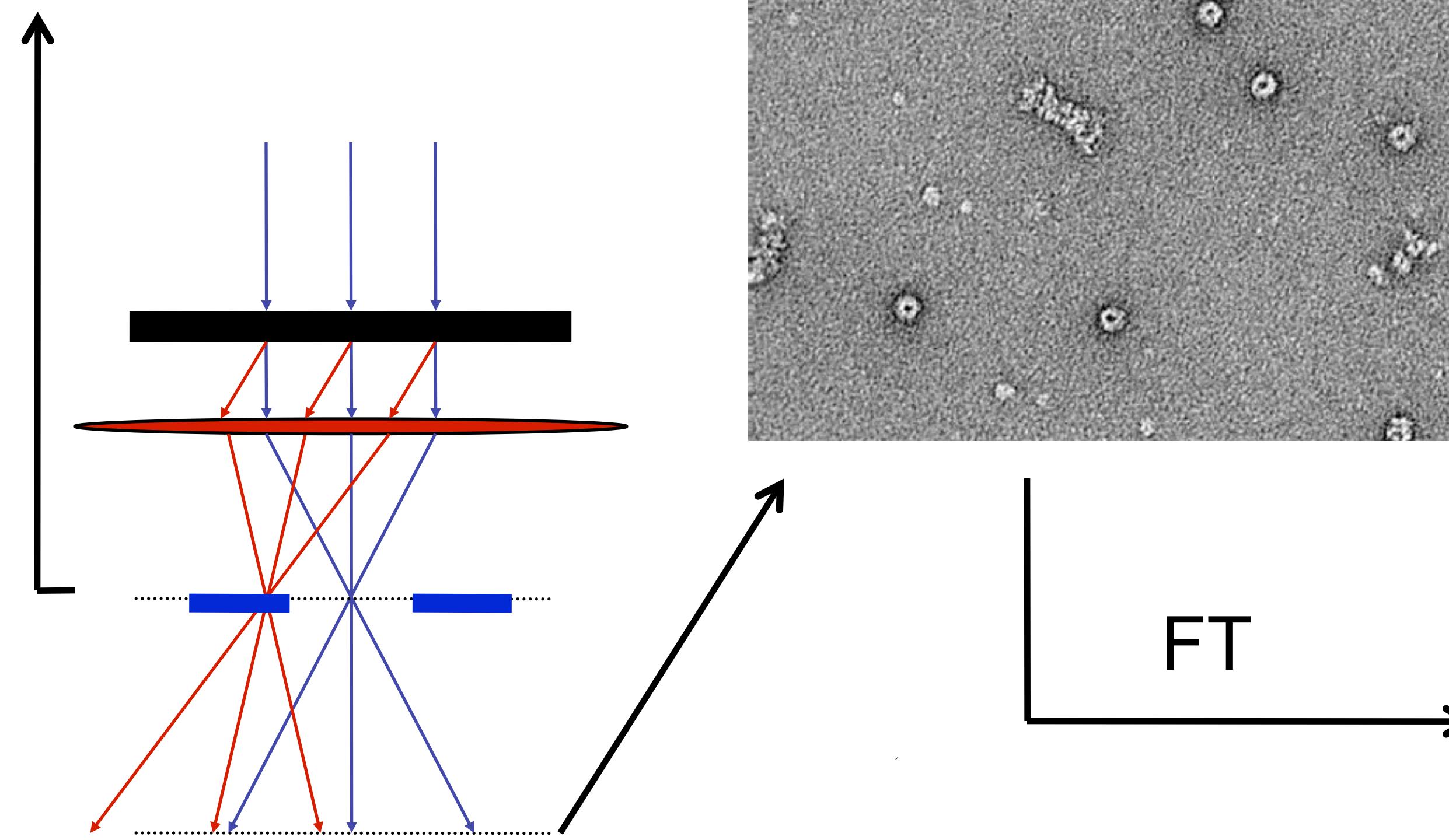
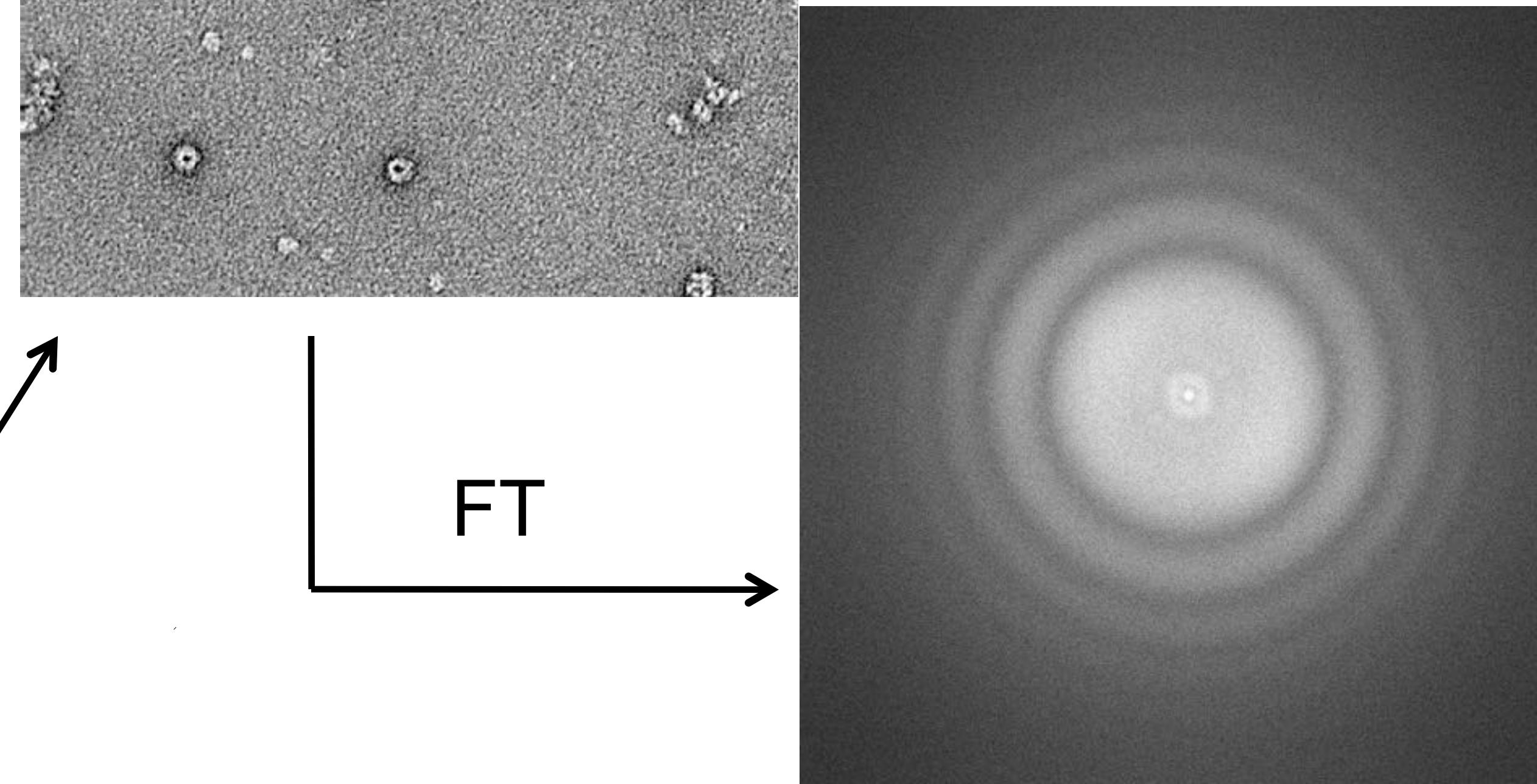
Diffraction, image and power spectra



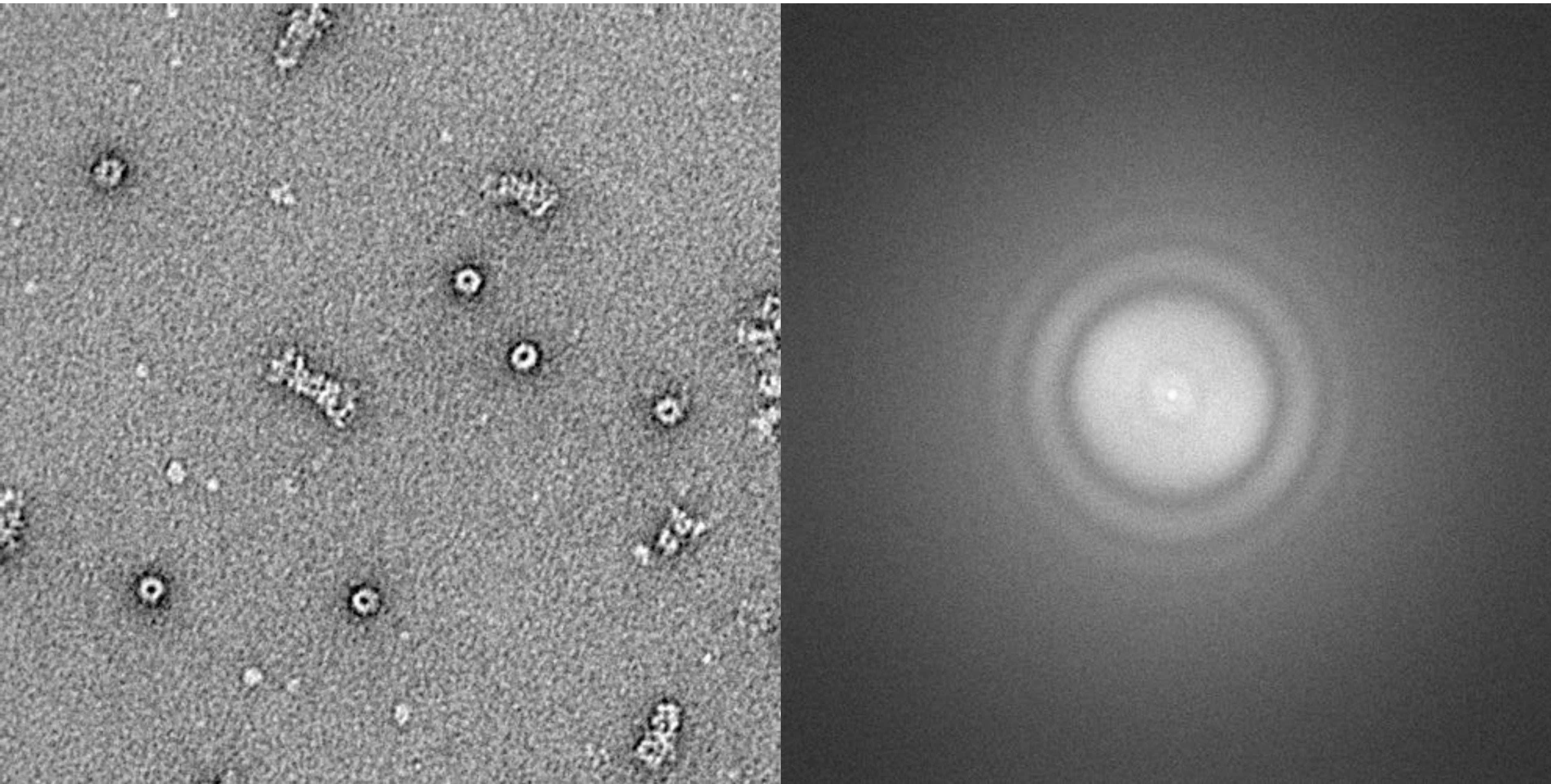
diffraction



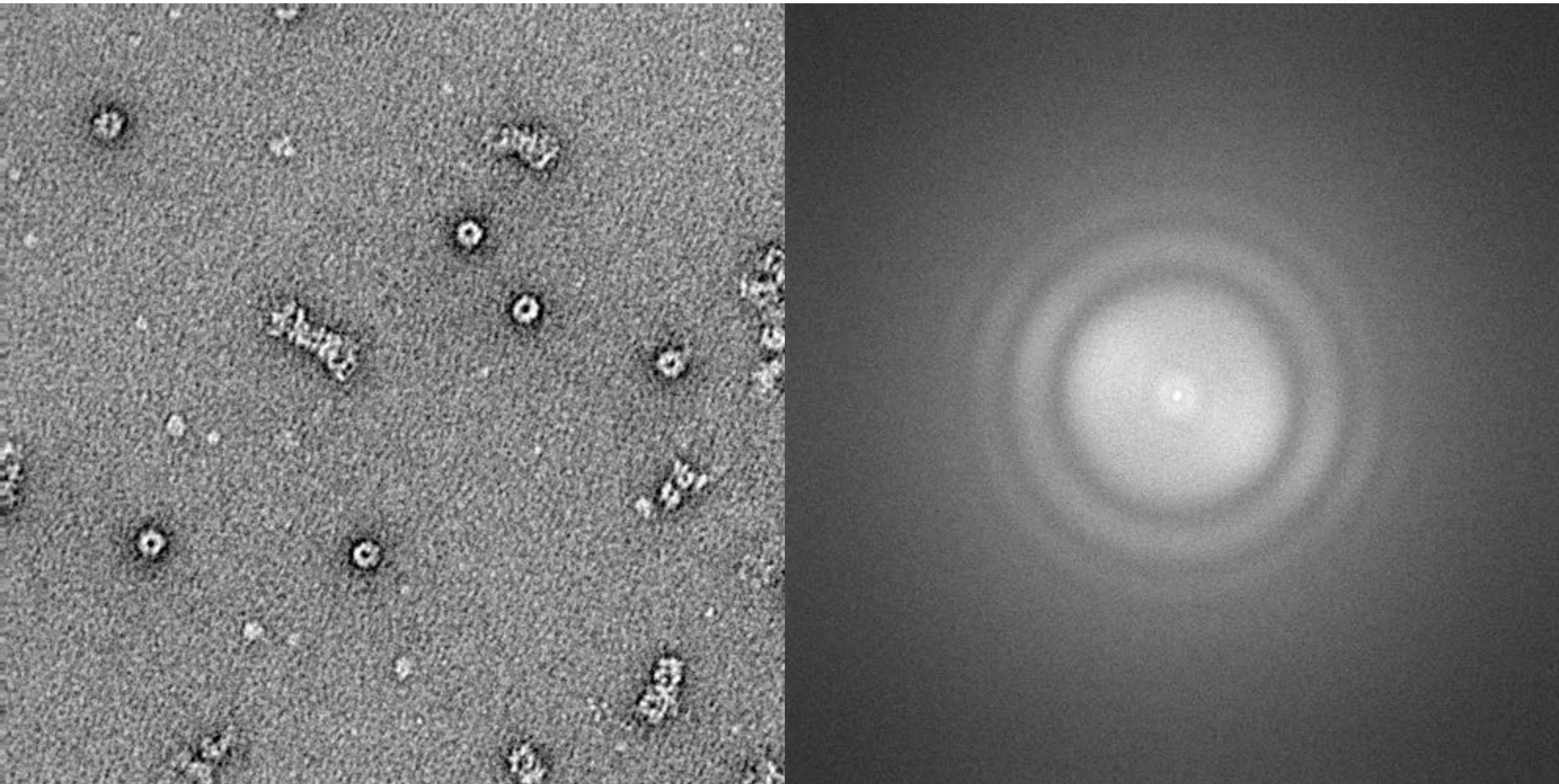
image



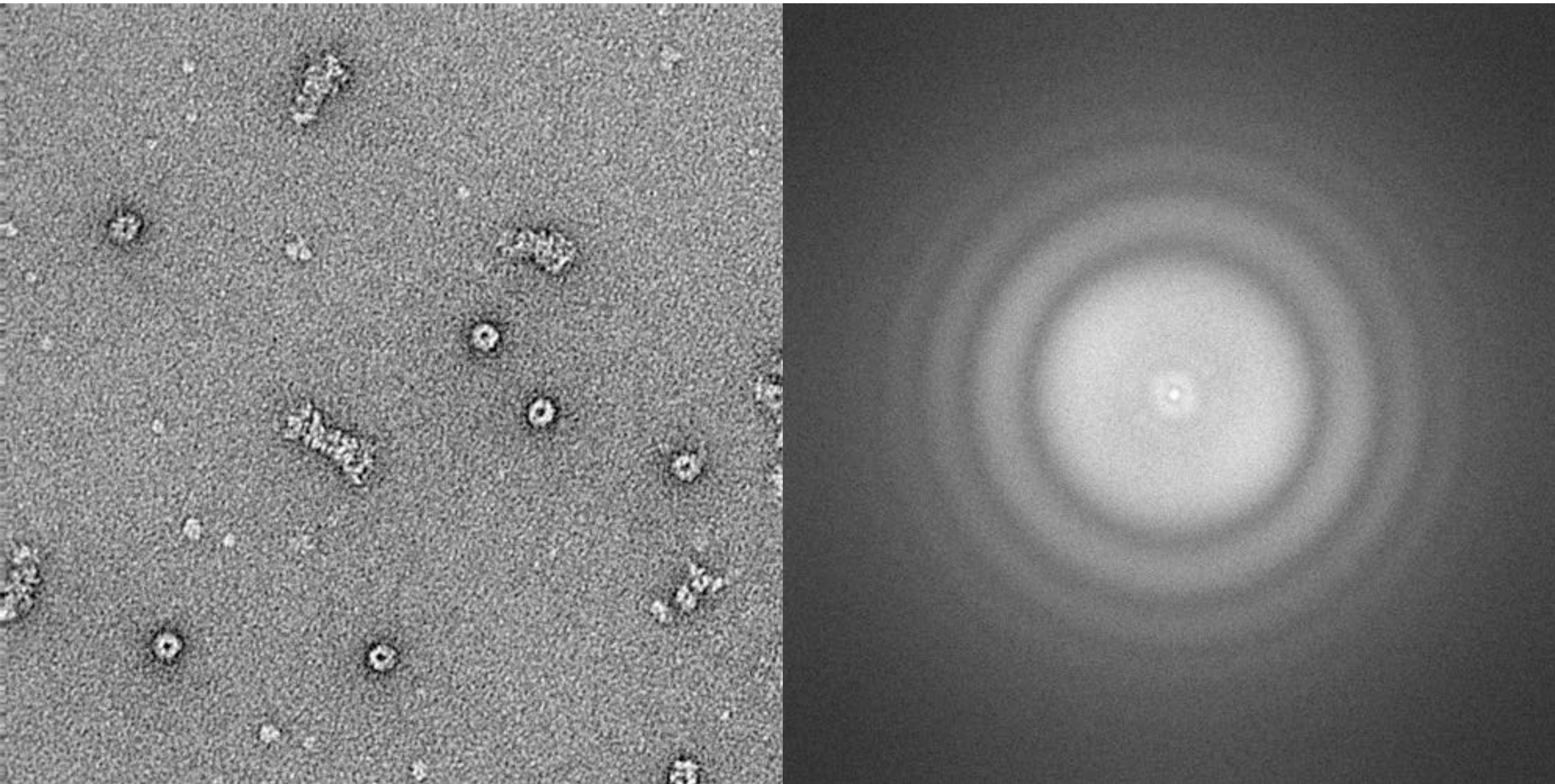
Defocus -2 μ m



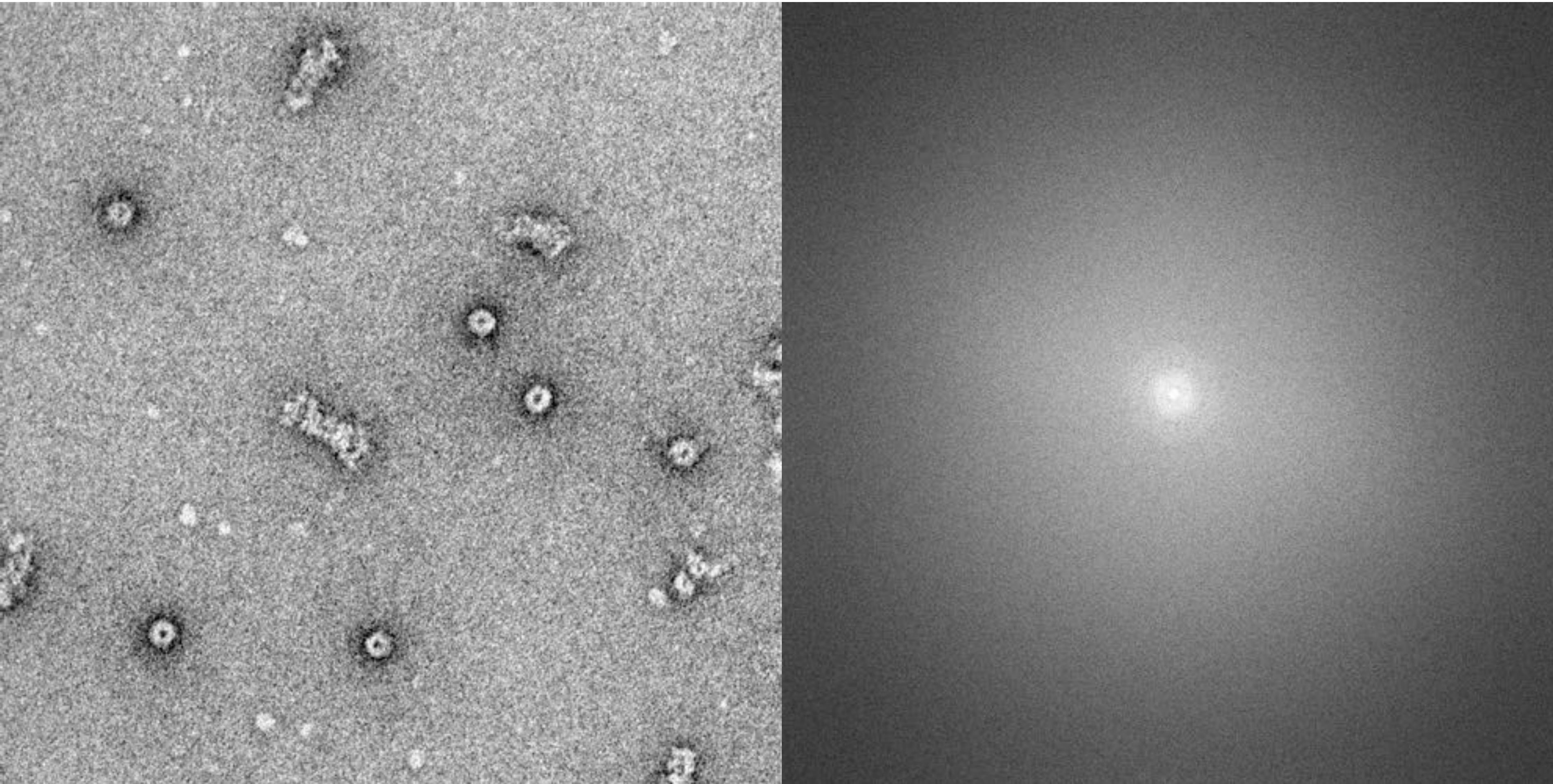
Defocus -1.5 μ m



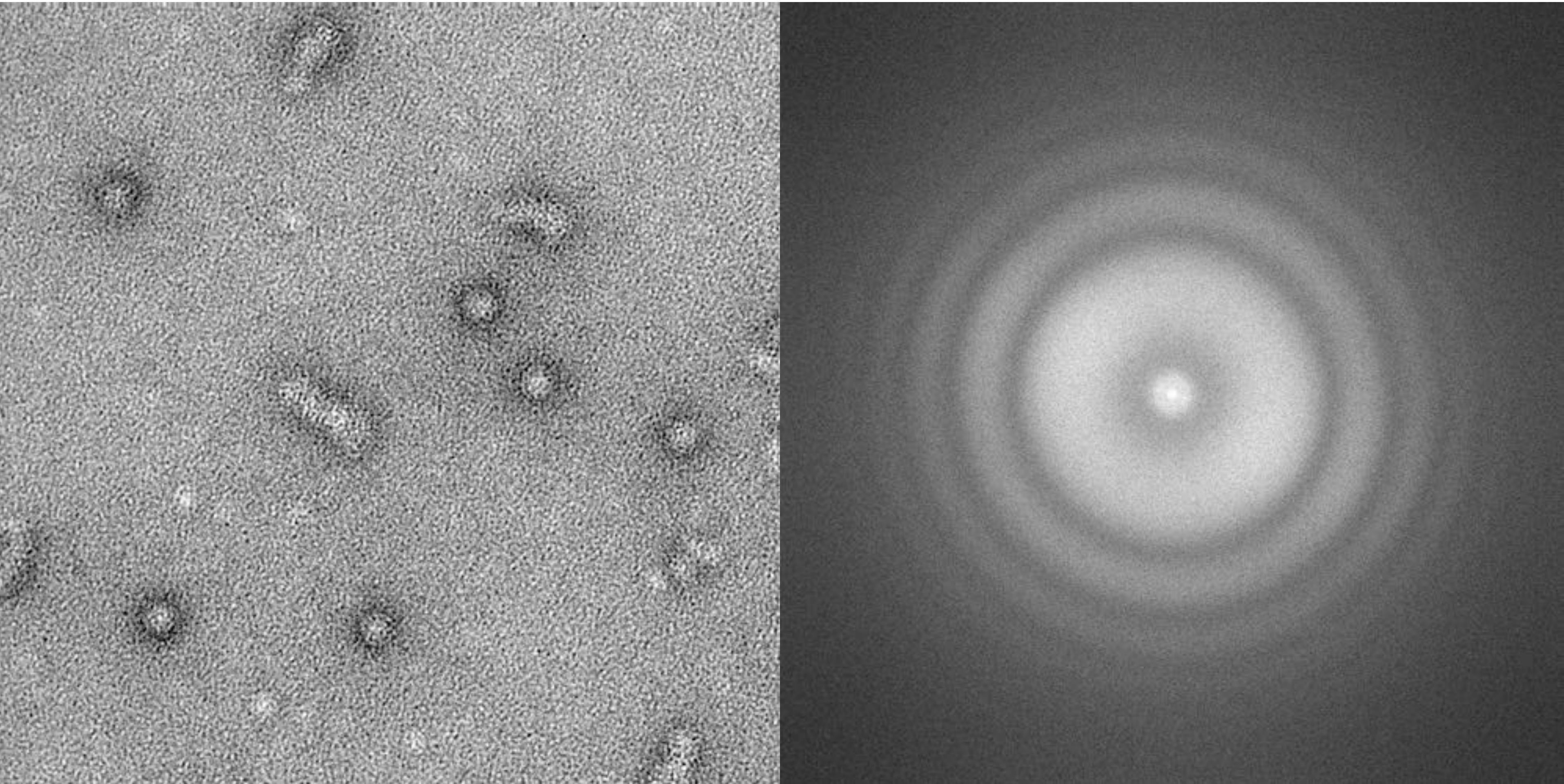
Defocus -1 μ m



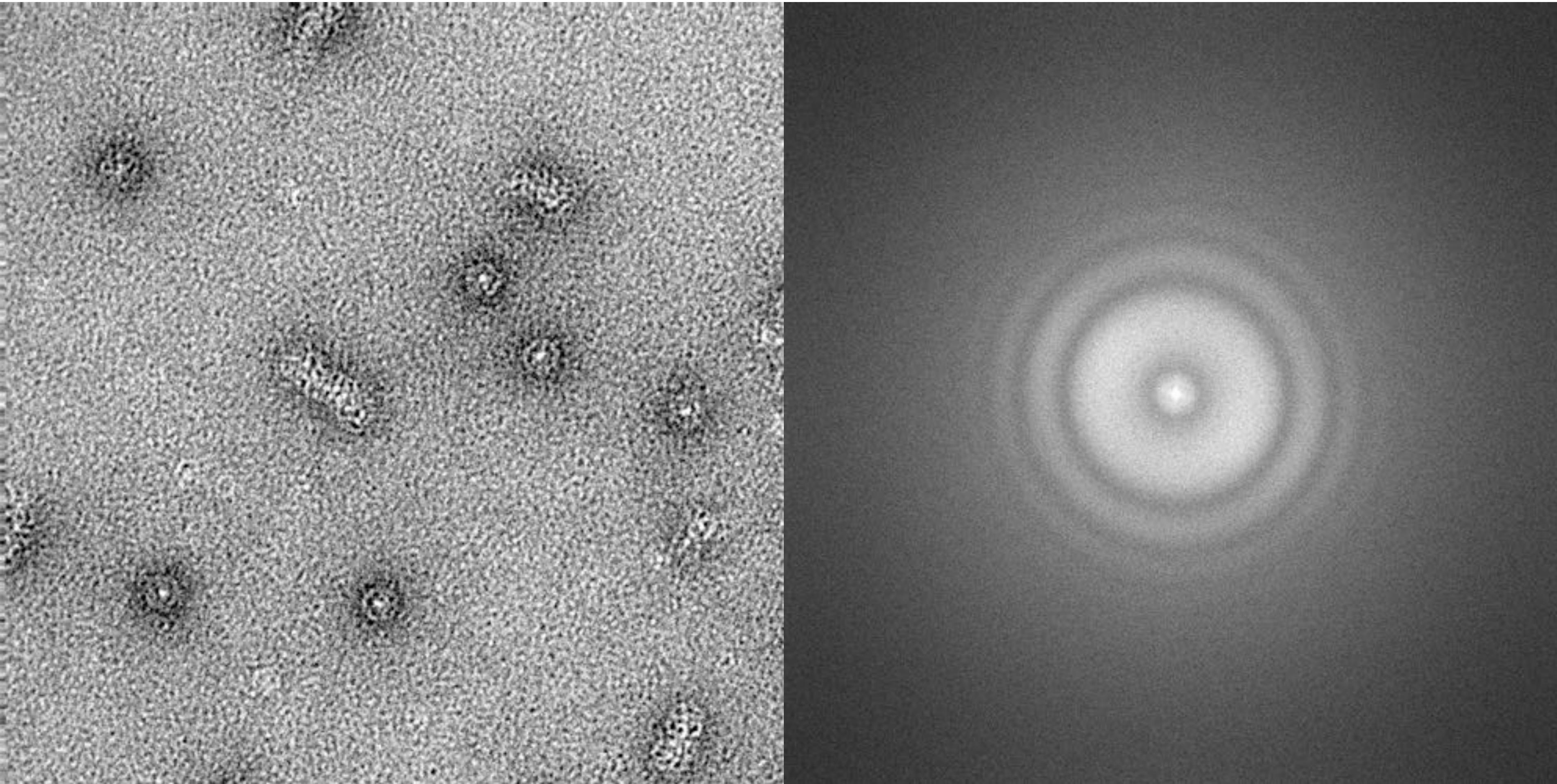
Defocus ~0 μ m



Defocus +1 μ m



Defocus +2 μ m



Determine CTF

Model

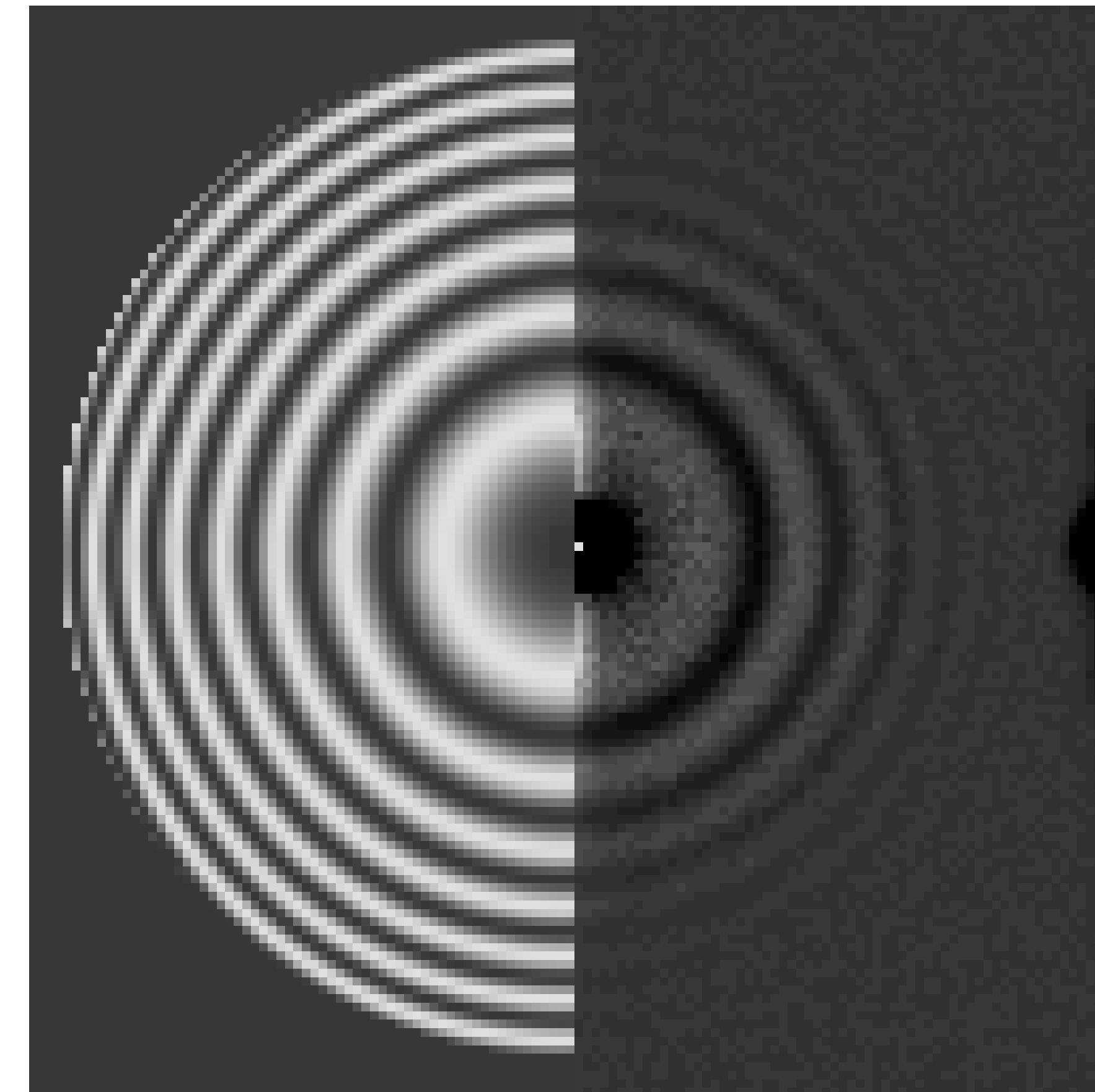


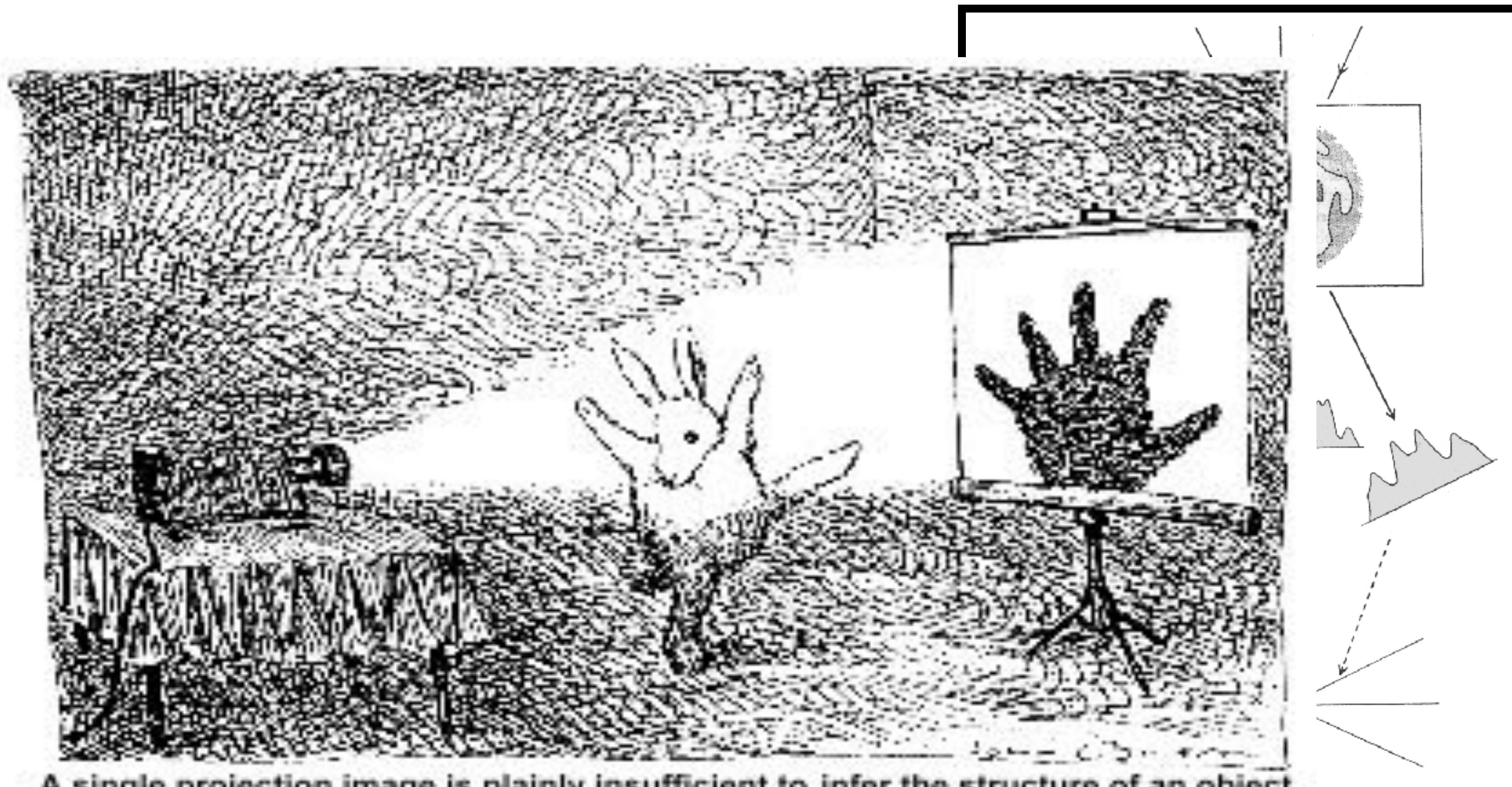
Image power
spectrum

Experiment

$$E = 120 \text{ kV}, \Delta f = 21000 \text{ \AA}, C_s = 2 \text{ mm}, A = 0.15$$

Reconstructing 3D object from 2D projection images

Central Section
Fourier transform
section through
the direction



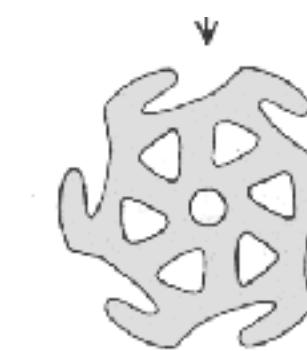
A single projection image is plainly insufficient to infer the structure of an object.

John O'Brien; © 1991 The New Yorker Magazine

DeRosier, D. and Klug, AEM (1968) “Reconstruction of a 3D object dimensional structures from electron micrographs” *Nature* 217 130-134

DeRoiser and Klug (1968)

Fig. 6. Scheme for the general process of reconstruction of a structure from its transmission images.



Molecular electron microscopy of biological sample

Strong electron scattering power means two things:

- 1 high vacuum of microscope column;
- 2 strong scattering with protein sample;

Problems:

- 1 dehydration of biological sample;
- 2 radiation damage by high energy beam;

Molecular electron microscopy of biological sample

Strong scattering by high-energy electrons imposes two challenges to biological samples:

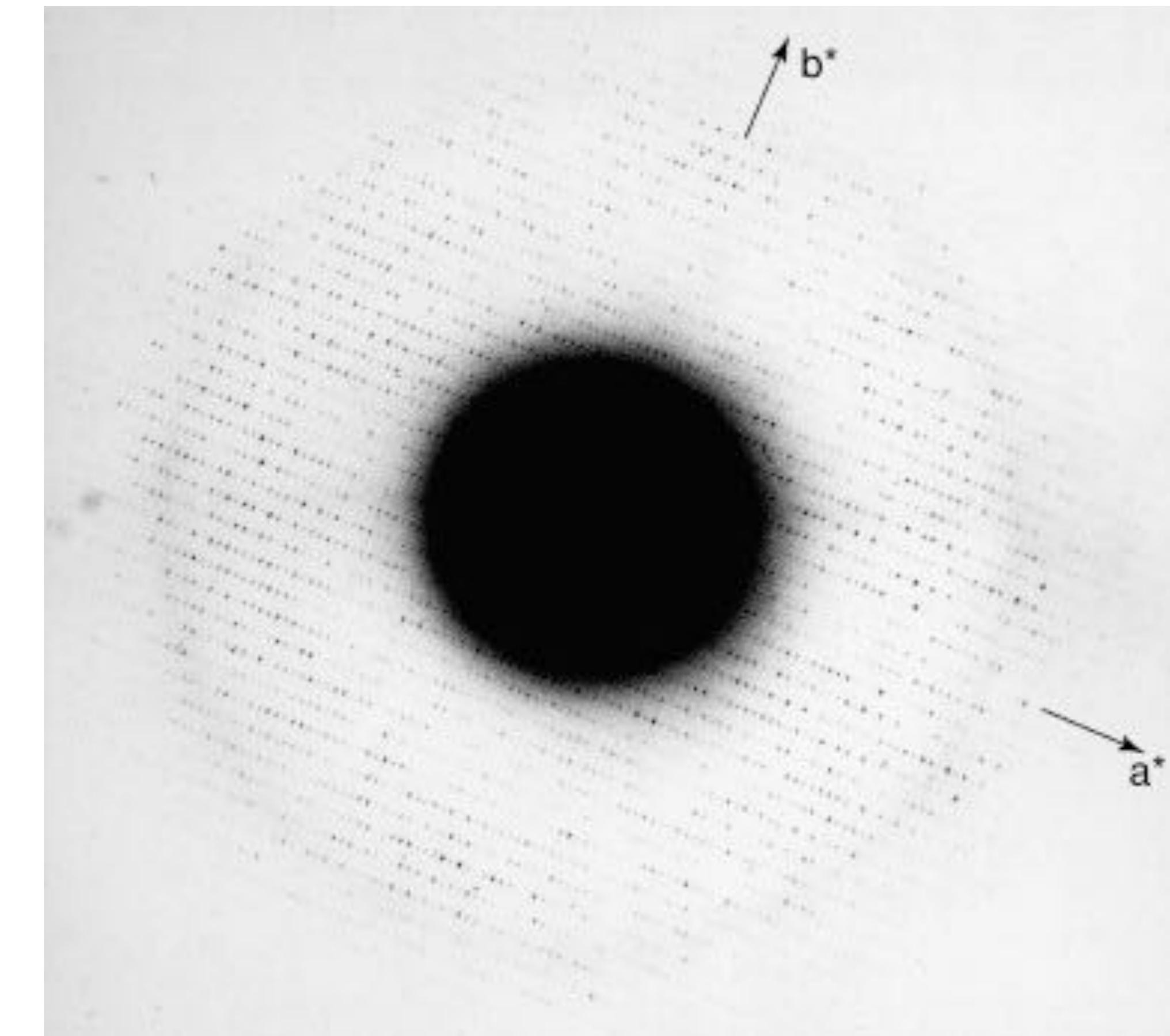
- dehydration caused by high vacuum within electron microscope column destroys biological samples;
- severe radiation damage caused by high-energy electron beam destroys biological samples;

- * Shadow casting (Williams & Wycoff, 1945);
- * Positive staining (Pease & Baker, 1948);
- * Glass knives for microtomy (Hartmann & Latta, 1950);
- * Diamond knives (Fernandez-Moran, 1953);
- * Negative staining (Hall, 1955);

Frozen hydration preserve structural integrity to atomic level.

Taylor K and Glaeser RM (1974) “Electron diffraction of frozen, hydrated protein crystals”
Science 186, 1036-1037

Taylor and Glaeser (2008) “Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future” *Journal of Structural Biology*



Cryo-electron microscopy

Against dehydration:

glucose/trehalose embedding: using glucose to substitute water, thus maintain hydration in the high vacuum. Only used for 2D crystal;

Frozen hydration: using plunge freezing to avoid crystal ice. Mostly for single particle;

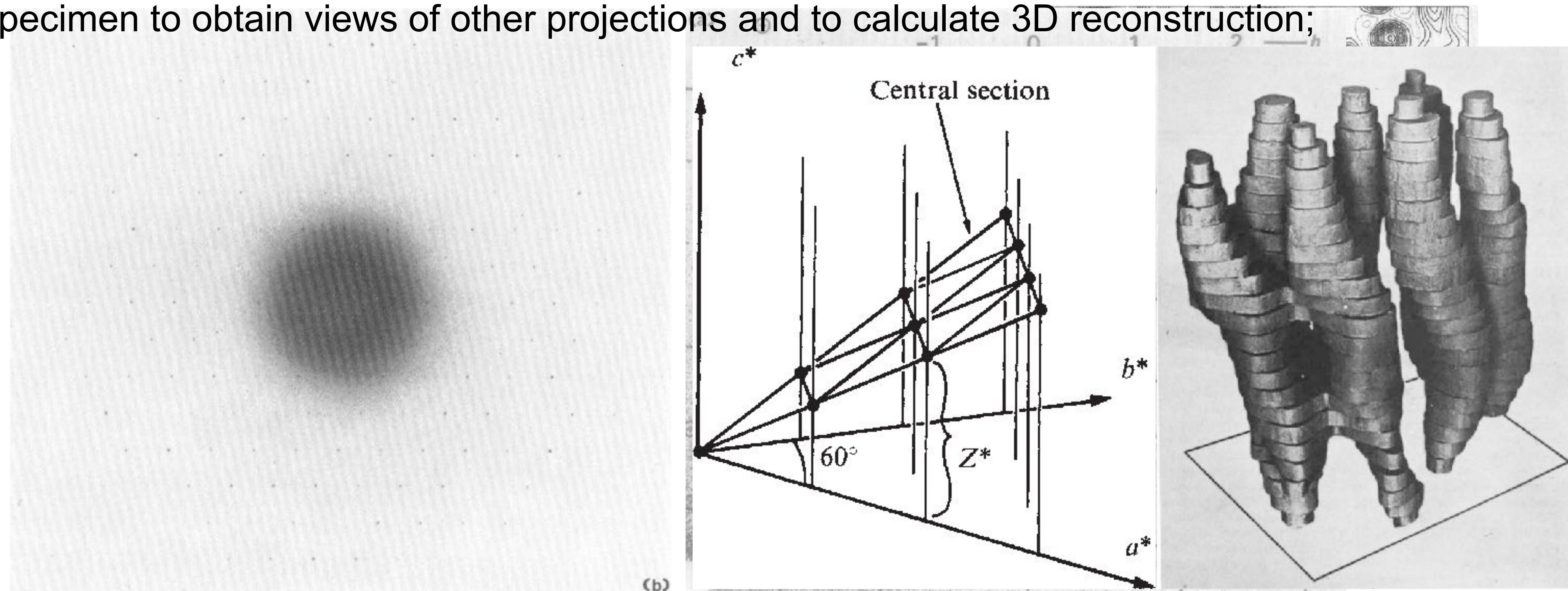
- Against radiation damage:

Low-temperature: LN2 (~80K) or LHe (~10K); Challenges to the instrumentations;

Low-electron dose: Low-dose imaging; Results in extremely noisy images, challenges for the data processing;

Structure of unstained crystalline specimen by electron microscopy

- Substituting water with sugar to prevent dehydration;
- Using crystalline samples to obtain sufficient signals from images recorded with low electron dose;
- Tilting specimen to obtain views of other projections and to calculate 3D reconstruction;



Henderson R and Unwin N (1975) "Three-dimensional model of purple membrane obtained by electron microscopy" *Nature* **257**, 28-32.

Unwin N and Henderson R (1975) "Molecular structure determination by electron microscopy of unstained crystalline specimens" *Journal of Molecular Biology* **94**, 425-440.

Single particle EM: averaging of low dose image of non-periodic objects

J Frank (1975) "Averaging of low exposure electron micrographs of non-periodic objects" *ultramicroscopy* **1**, 159.

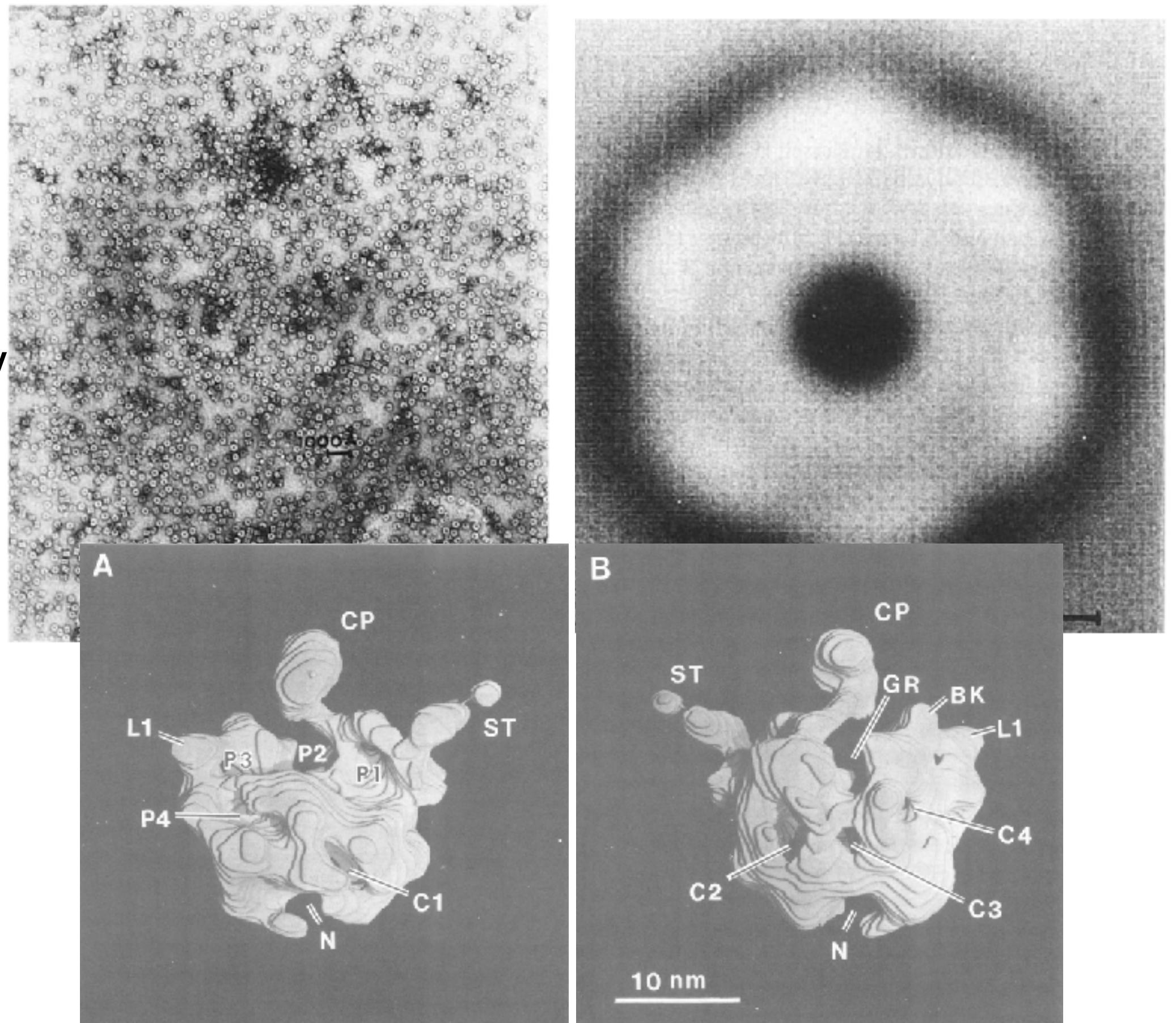
"We will investigate how the average techniques could be extended to this general case. Of all the possible regular specimen, we are interested in those which form identical particles, sufficiently well separated on the microscope grid so as not to overlap.".

Frank, J. Goldfarb, W, Eisenberg, D. and Baker, T.S.
(1978) "Reconstruction of glutamine synthetase using
computer averaging" *ultramicroscopy* **3**, 283-290.

"A single low-dose micrograph of a maximally tilted specimen will supply
all the Fourier information contained in a cone up to that tilt angle".

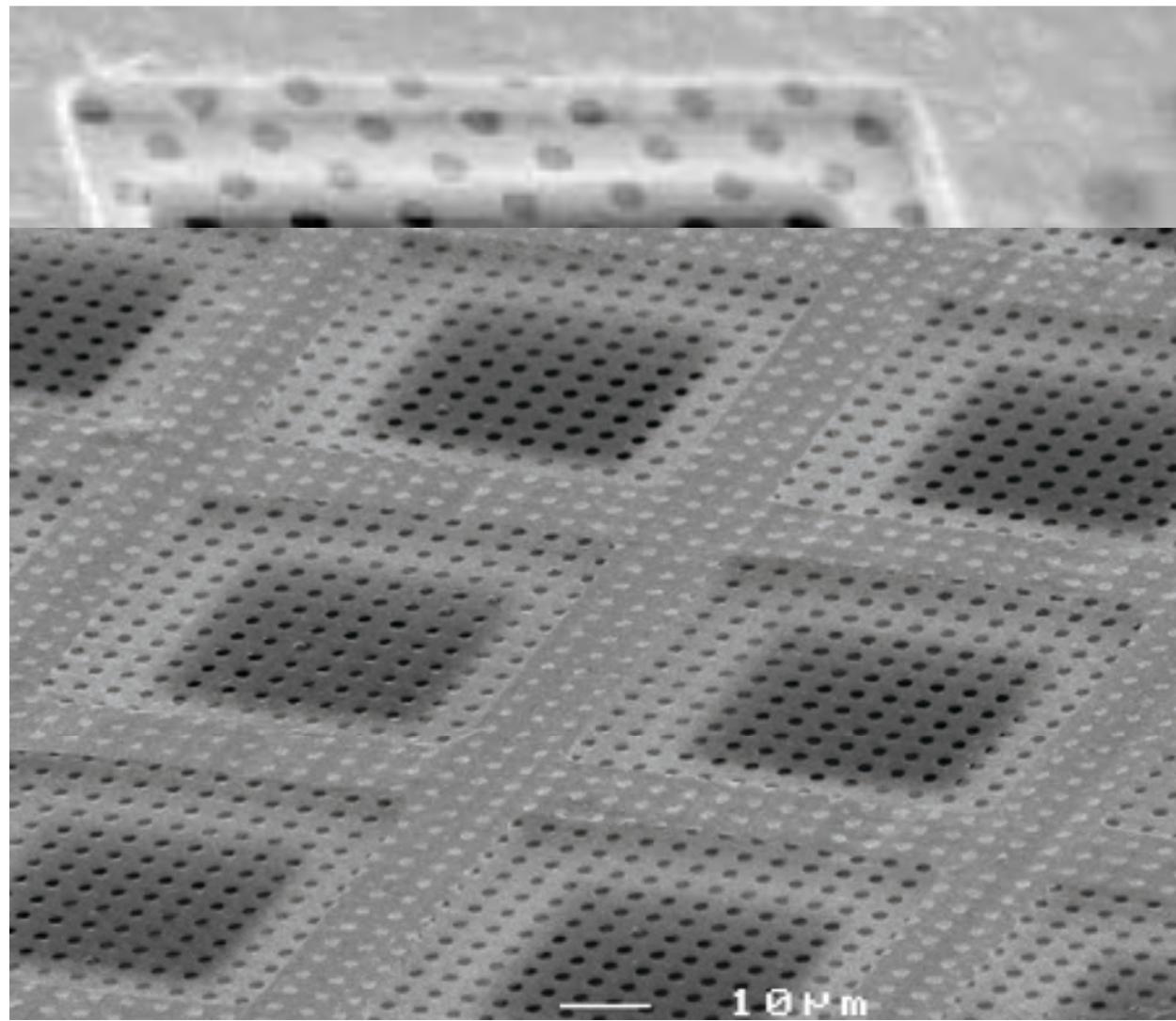
Radermacher, M., Wagenknecht, T., Verschoor, A.,
and Frank, J. (1987) "Three-dimensional Structure of
large ribosomal subunit from *Escherichia coli*" The
EMBO Journal **6**, 1107-1114.

E. coli 50S ribosome by random conical tilt (RCT)

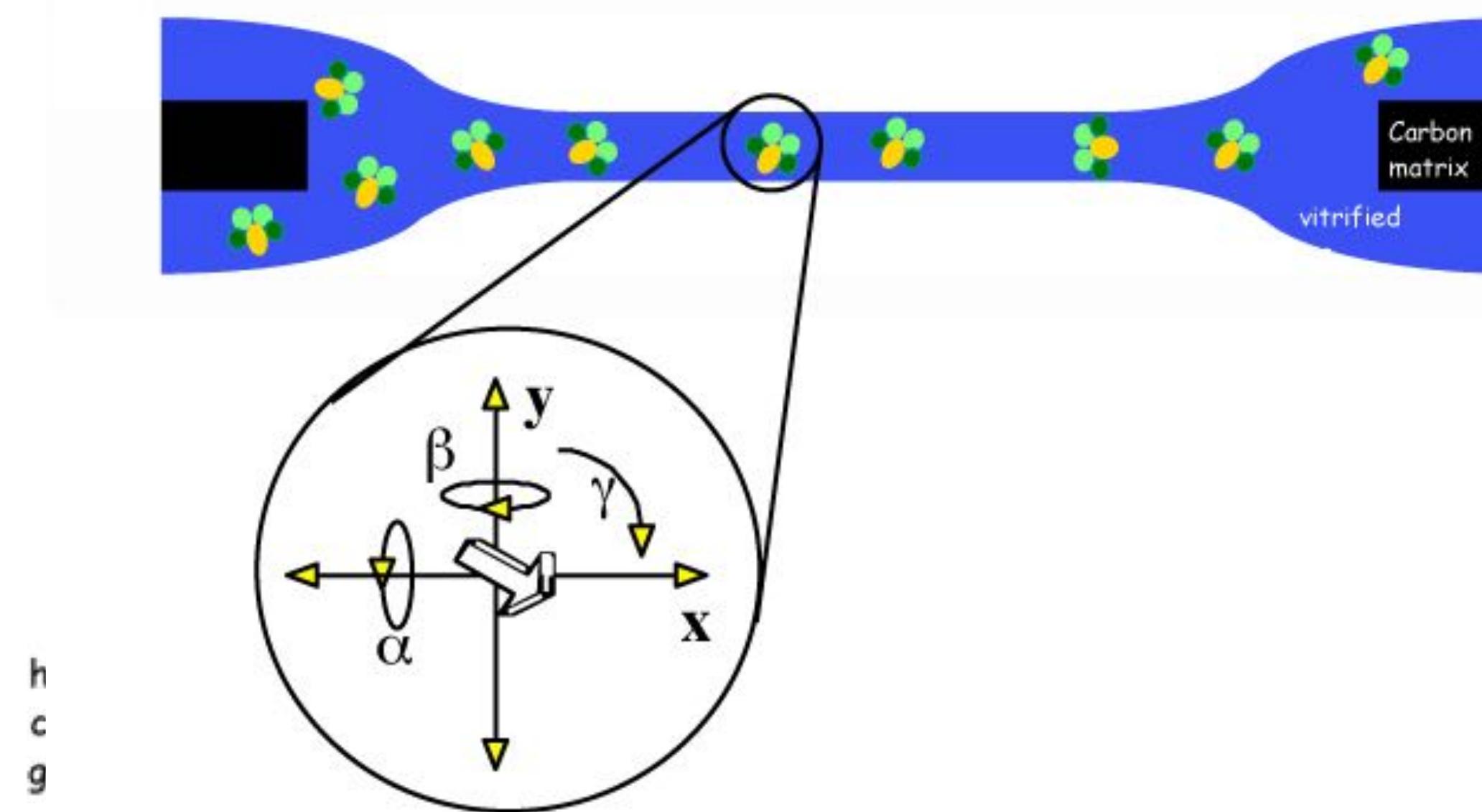


Frozen hydrated specimen preparation for single particle cryo-EM

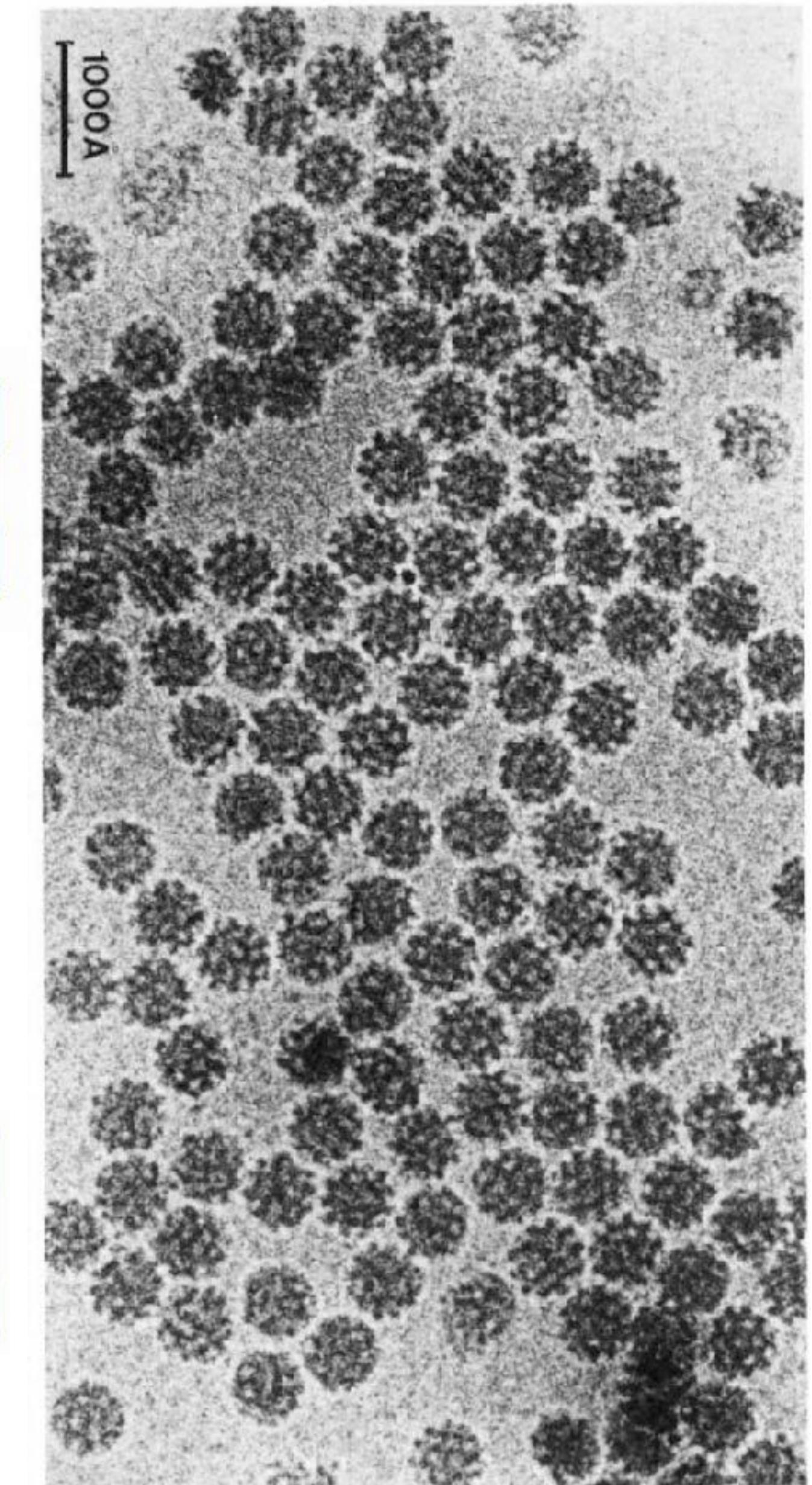
Adrian M, Dubochet J, Lepault J & McDowall AW (1984)
Cryo-electron microscopy of viruses. *Nature* **308**, 32-36.



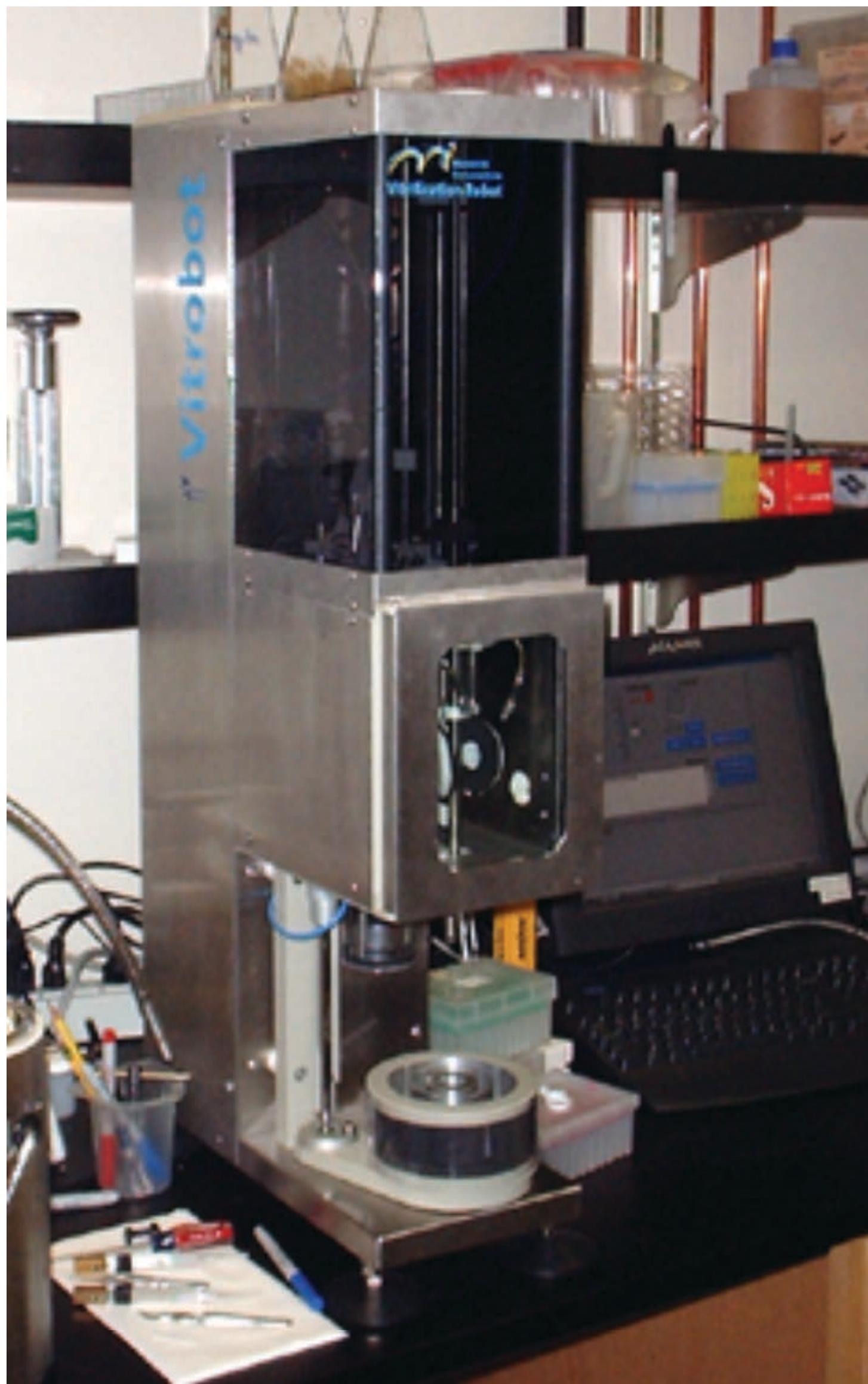
Quantifoil grid



The geometry of each particles is defined by 5+1 parameters: three Euler angles, two in-plane positions (x , y) and defocus (z). First 5 are determined and refined against a reference model iteratively. Defocus is determined separately.

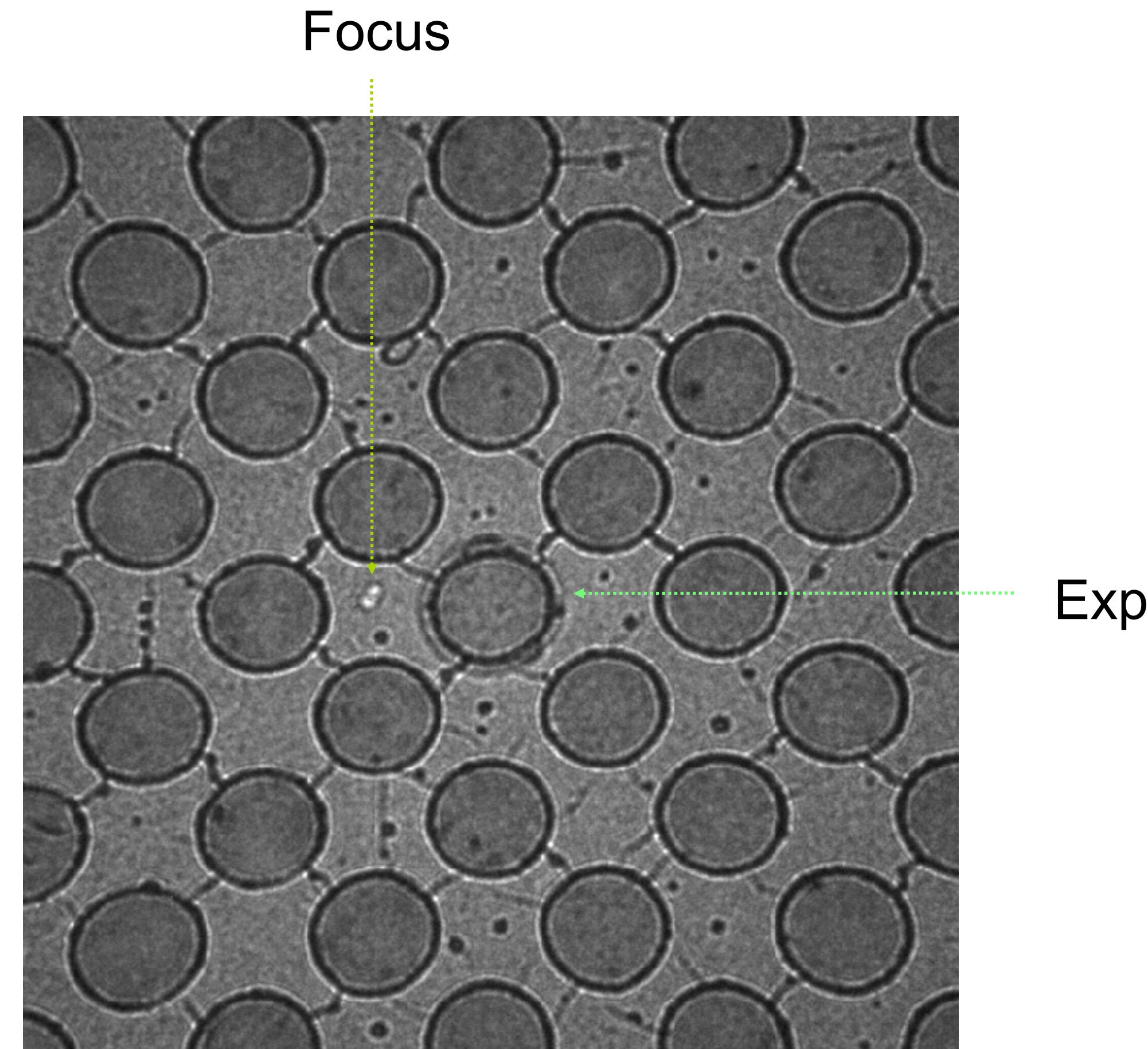


Equipment for cryo-electron microscopy

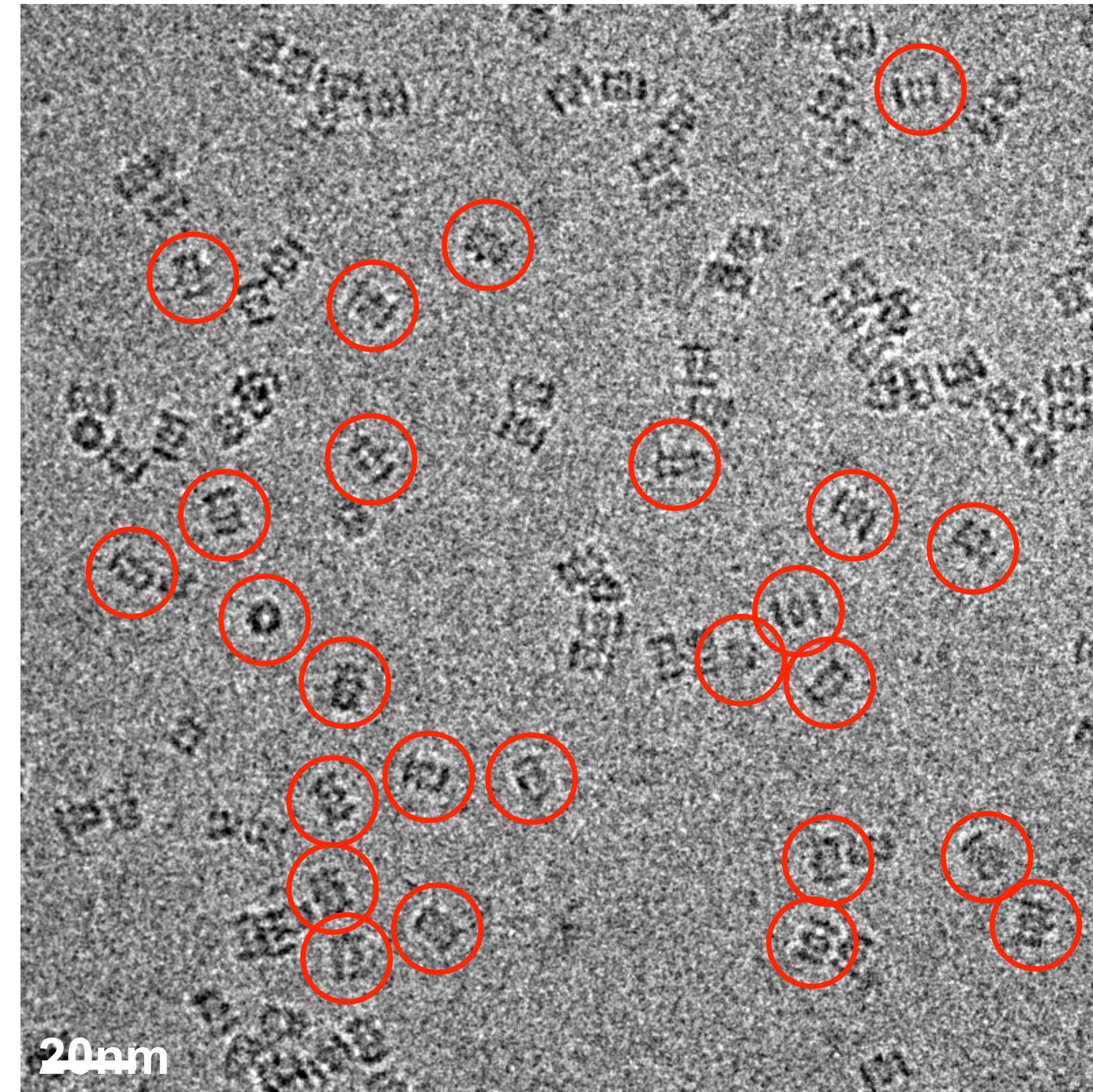


Three different modes in low dose

- * Search: lowest possible beam intensity;
- * Focus: off-exposure area, high magnification;
- * Exposure: desired magnification and beam intensity;

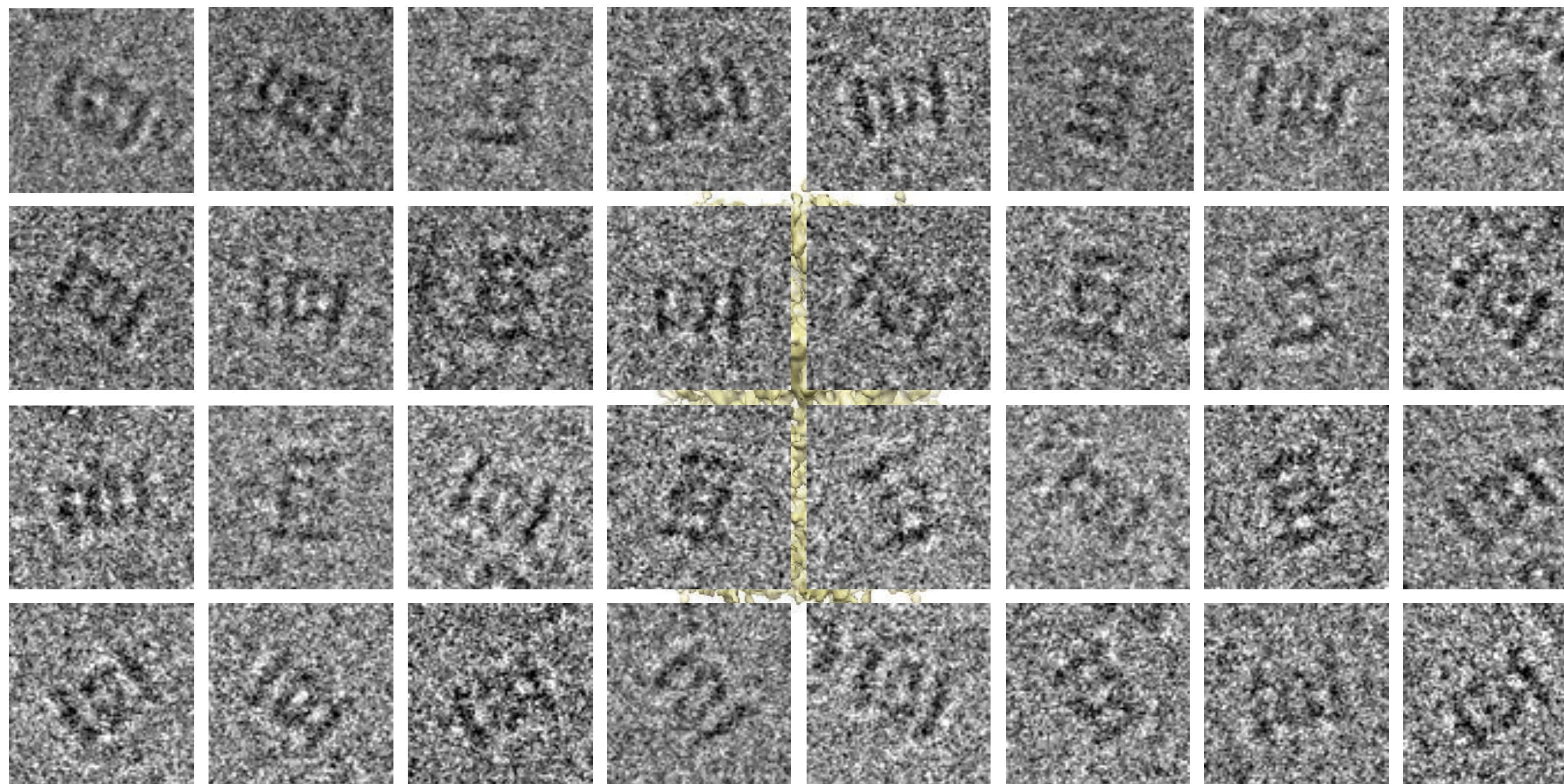


Single particle cryo-EM



Cryo-EM image of ~~S. cerevisiae~~ 20S proteasome

Single particle cryo-EM



3D reconstruction (Bokstein, Pacholska, and Agard, 2005)

Atomic resolution imaging with TEM

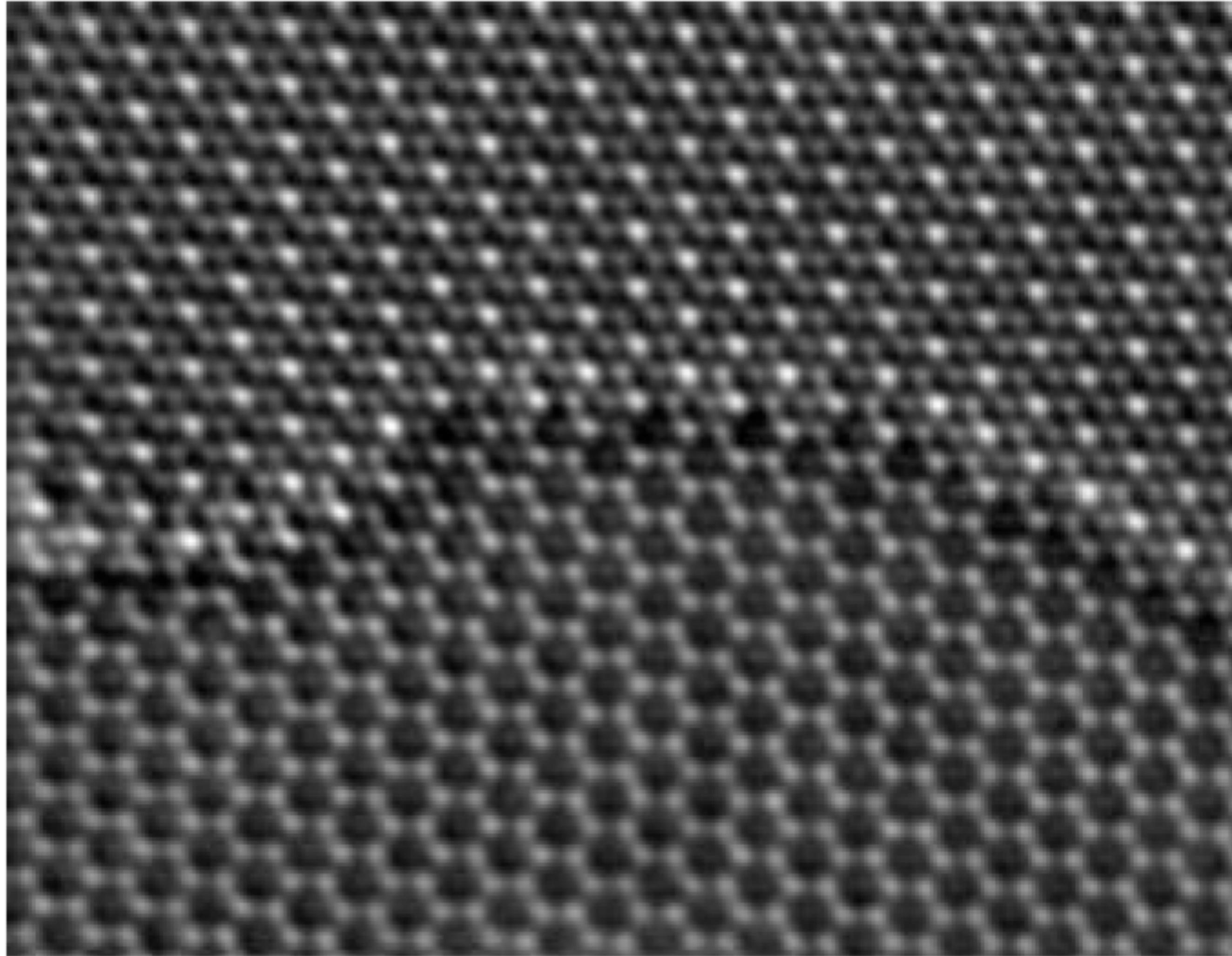


Image of graphene, Nature Mat, 2011, **10**, 165

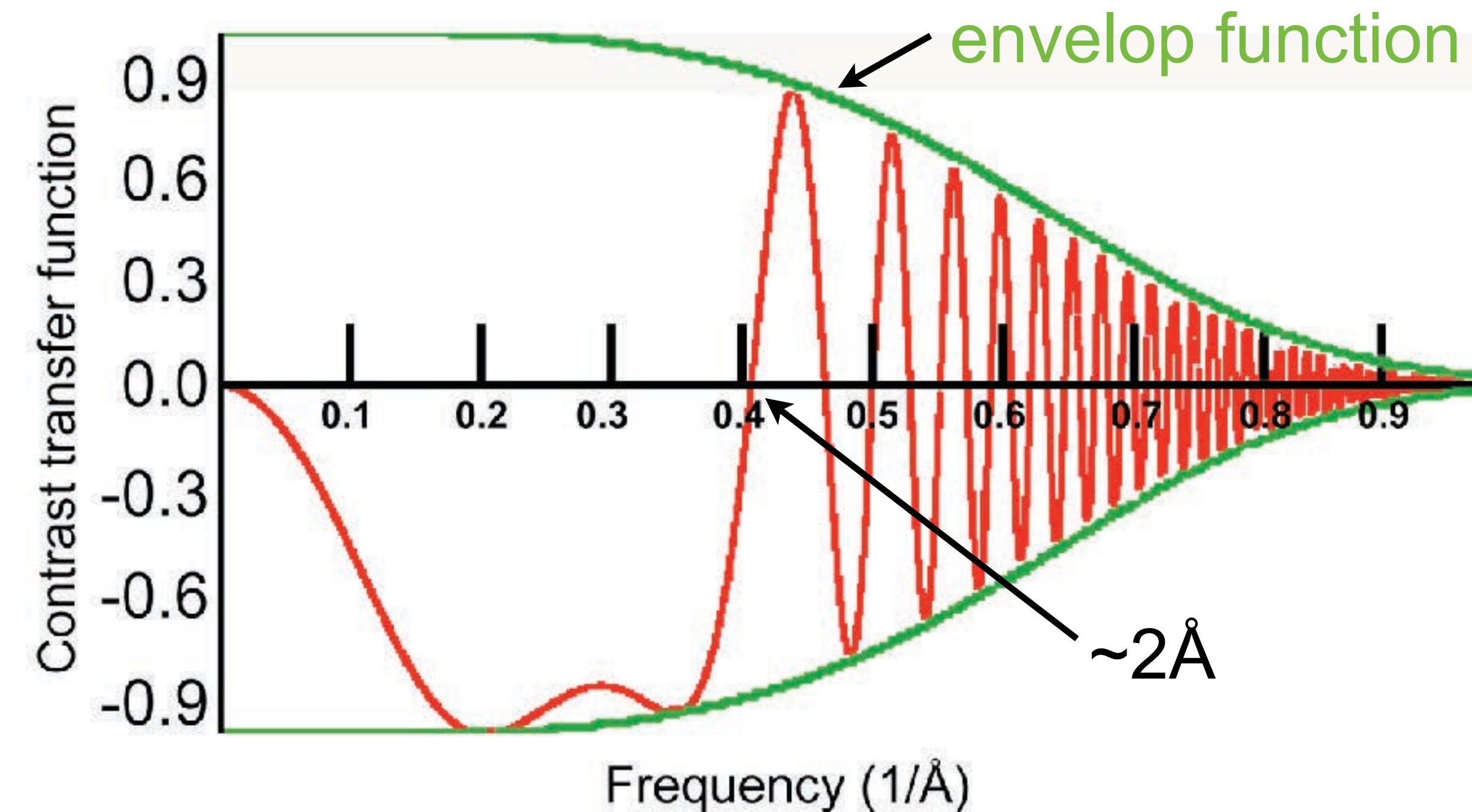
Electron optic system of a modern electron microscope is of sufficient quality to image radiation resistant material (typically inorganic) at atomic resolution ($\sim 2\text{\AA}$ or better).

Contrast Transfer Function

$$I(\vec{r}) = 1 + 2\Phi(-\vec{r}) \otimes F^{-1}(CTF)$$

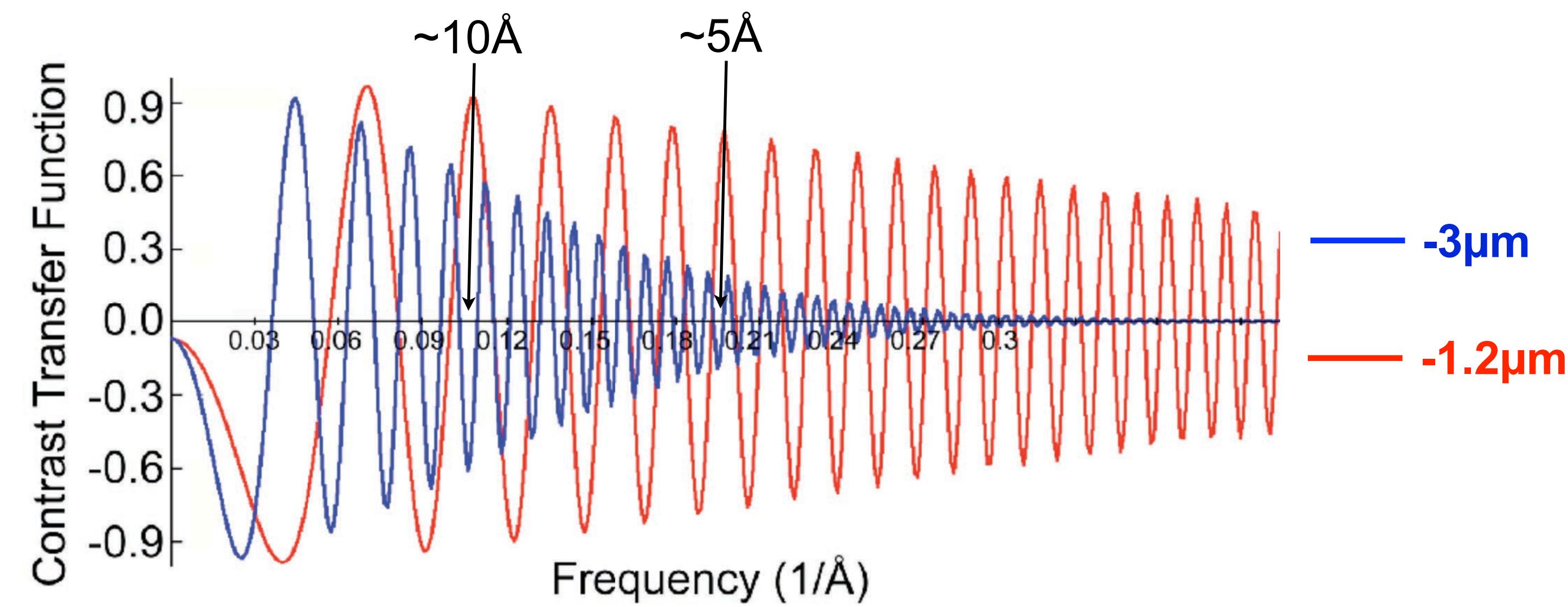
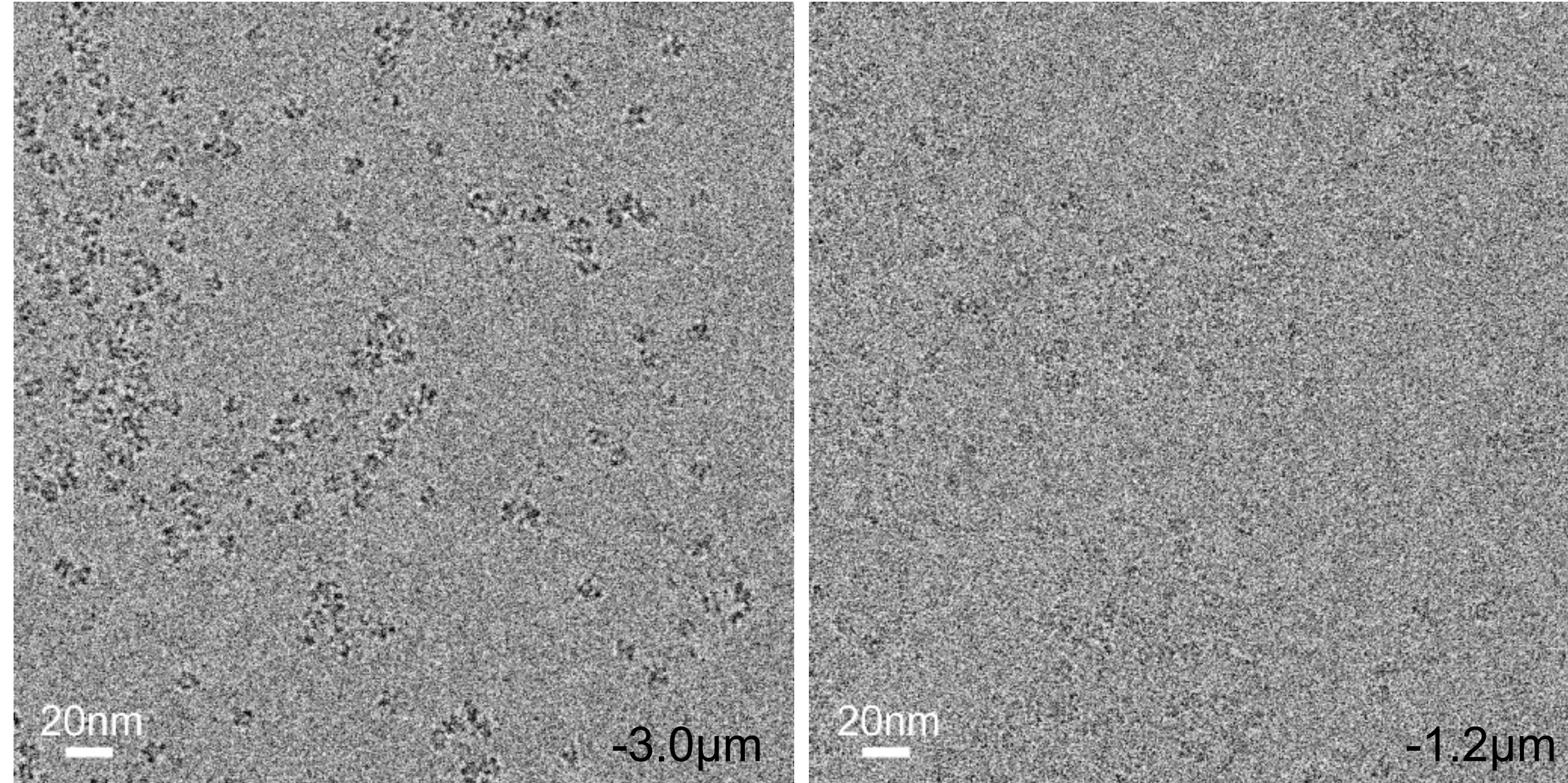
$$CTF = \sin(2\pi\chi k)$$

$I(r)$: intensity of an image;
 $\phi(r)$: projection of specimen;
 k : resolution;



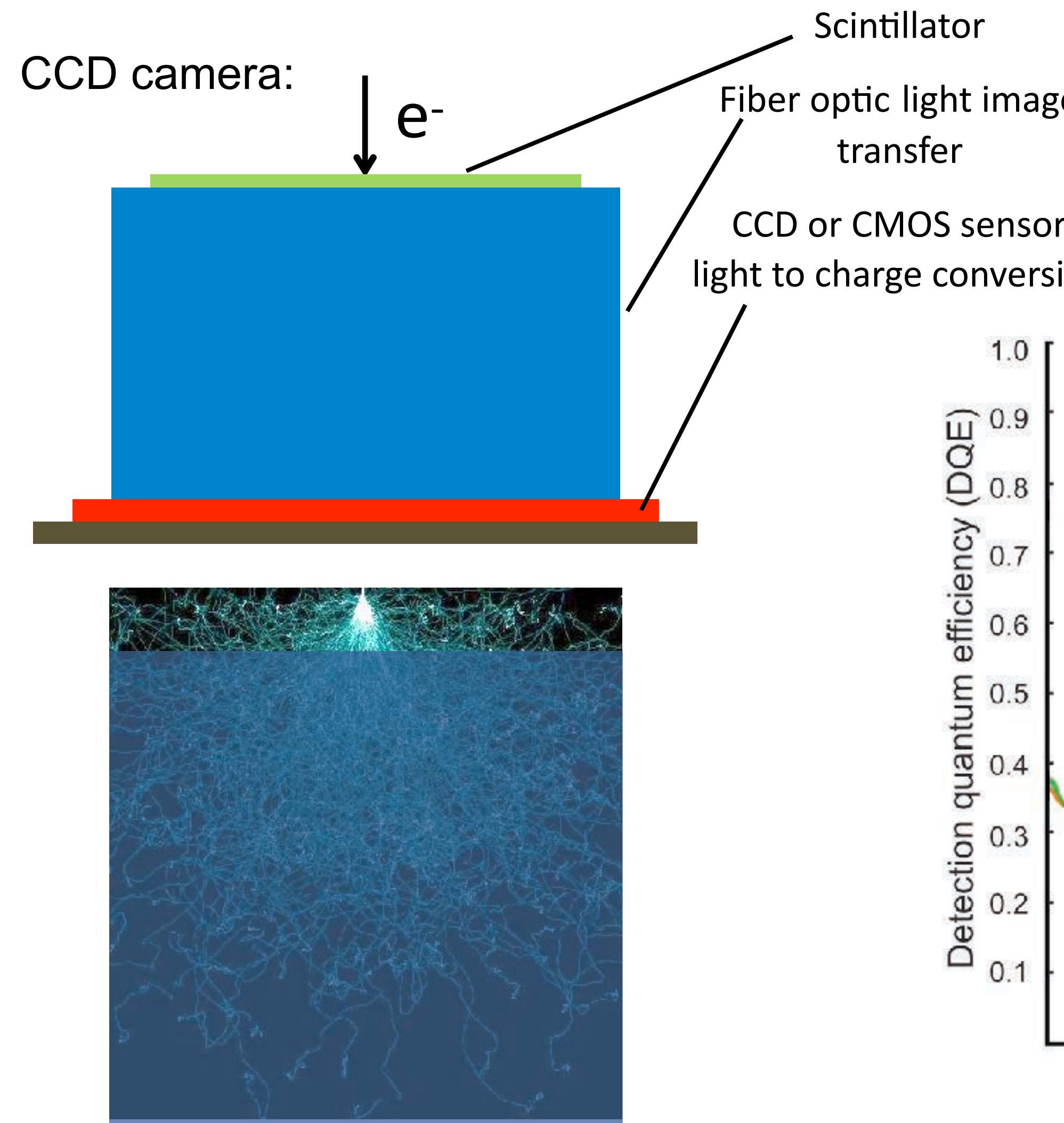
CTF modulate both amplitude and phases of an image.

Influence of CTF on image

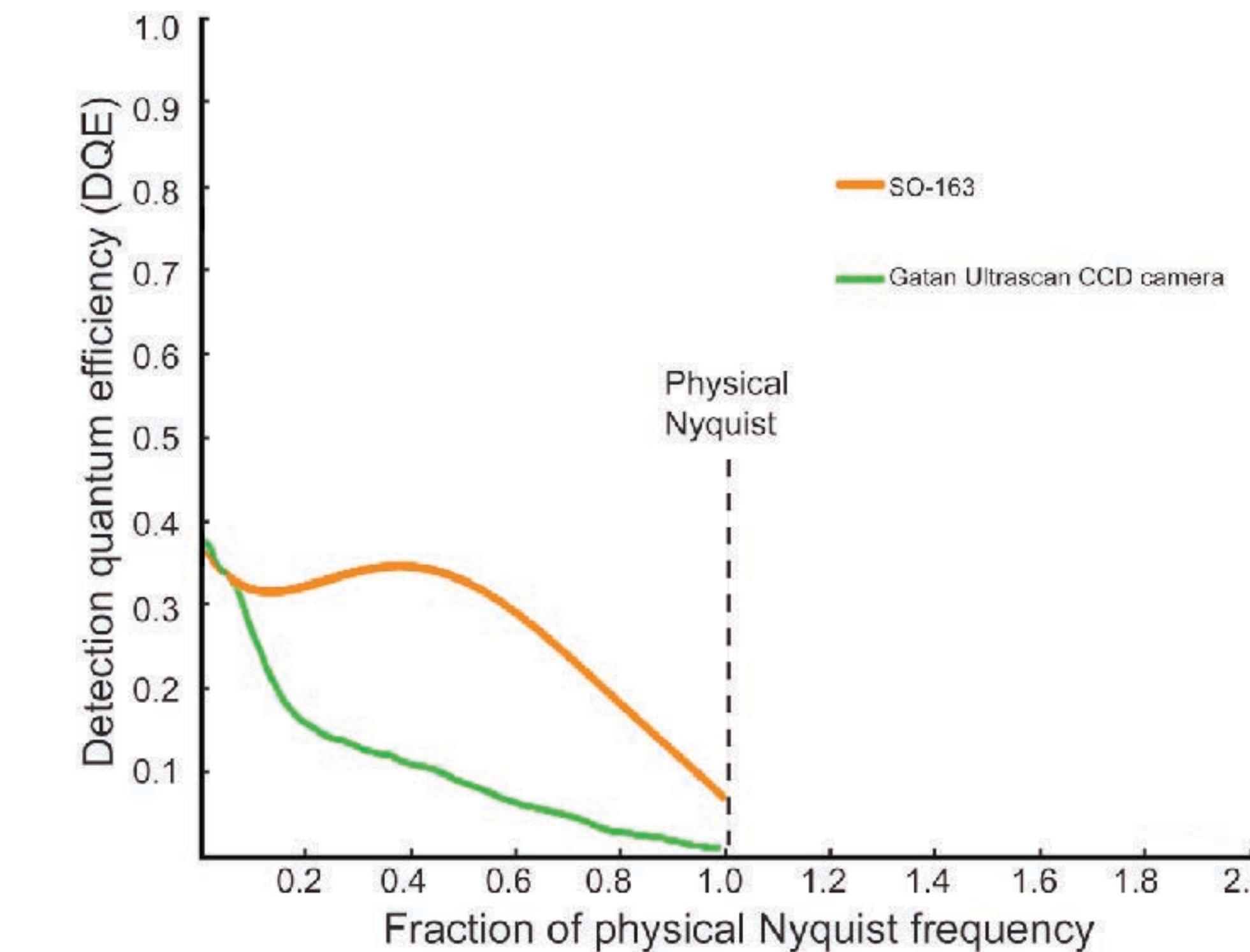


Scintillator based camera/photographic film

Scintillator based camera and photographic film are inadequate for high-resolution cryo-EM.

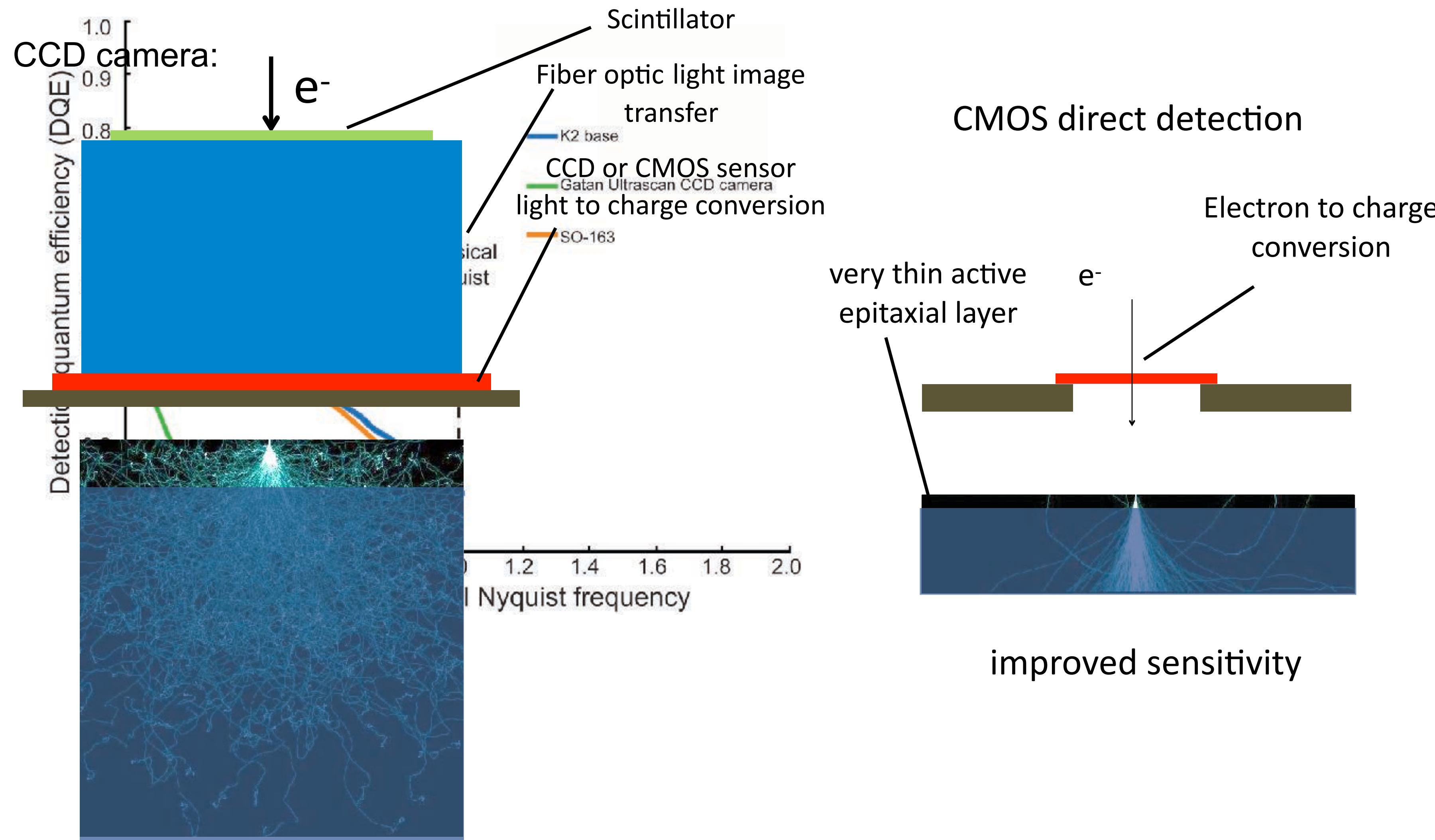


DQE:



CMOS direct detection camera

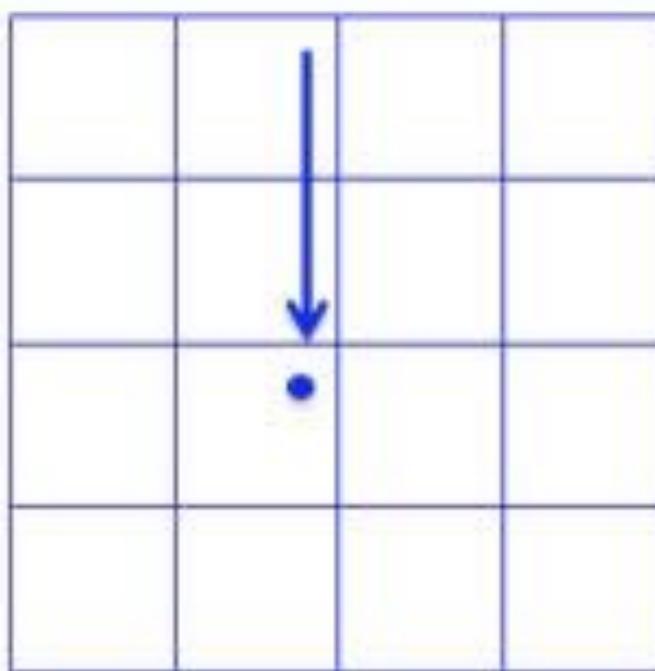
Direct detection minimizes the point spread function, and improve camera performance at both low and high resolution.



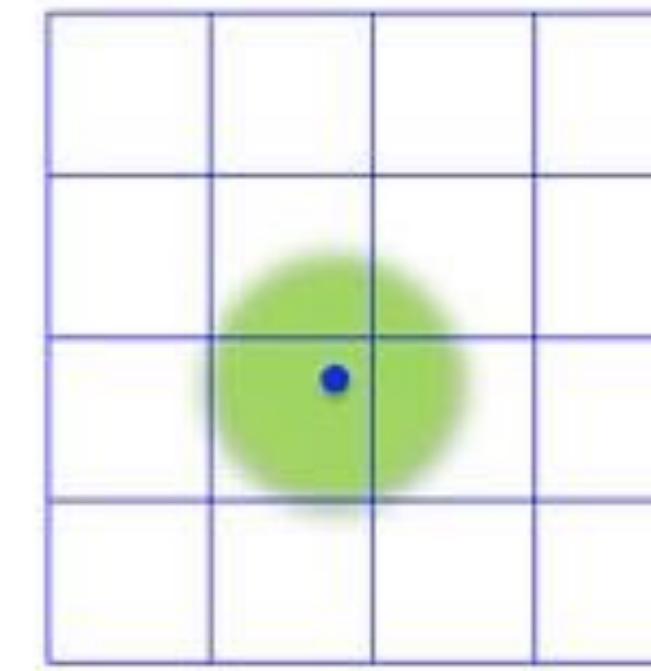
Single electron counting by the K2 Summit (UCSF, LBNL, Gatan)

- * Counting and centroiding primary electron events.
- * Counting removes Landau noise and further improves DQE;

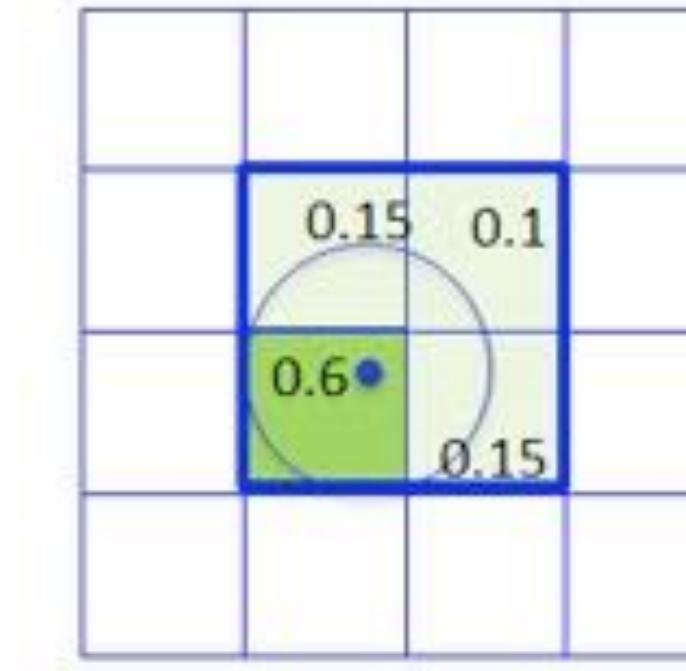
1. Electron enters detector



2. Signal Is Scattered



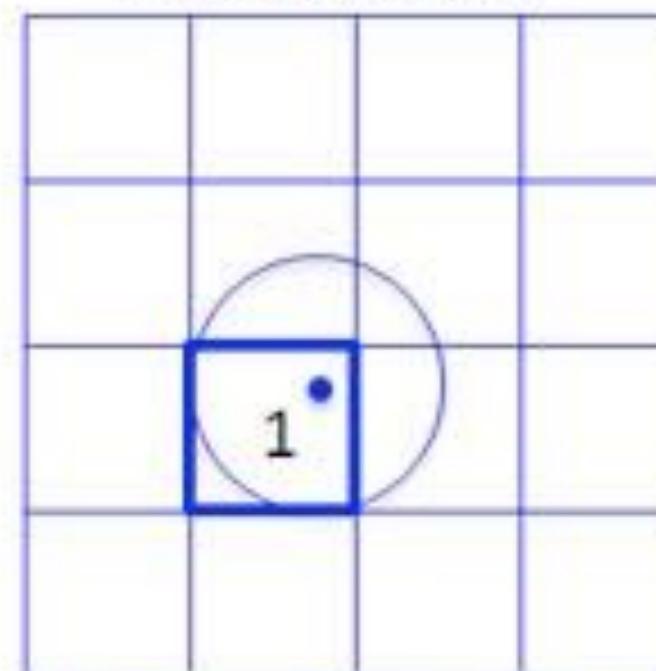
3. Charge collects in each pixel



K2 Base™: Charge Integration

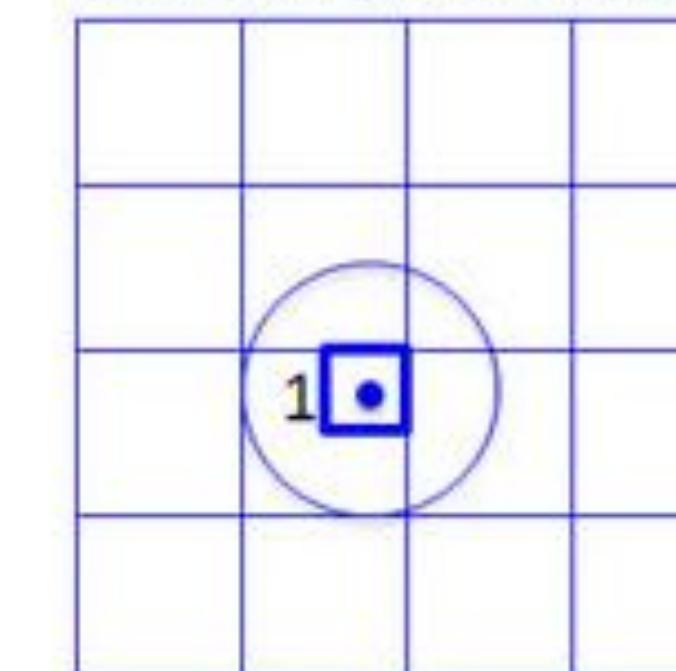
Improved DQE at high Frequency

4a. Events are reduced to the highest charge pixels

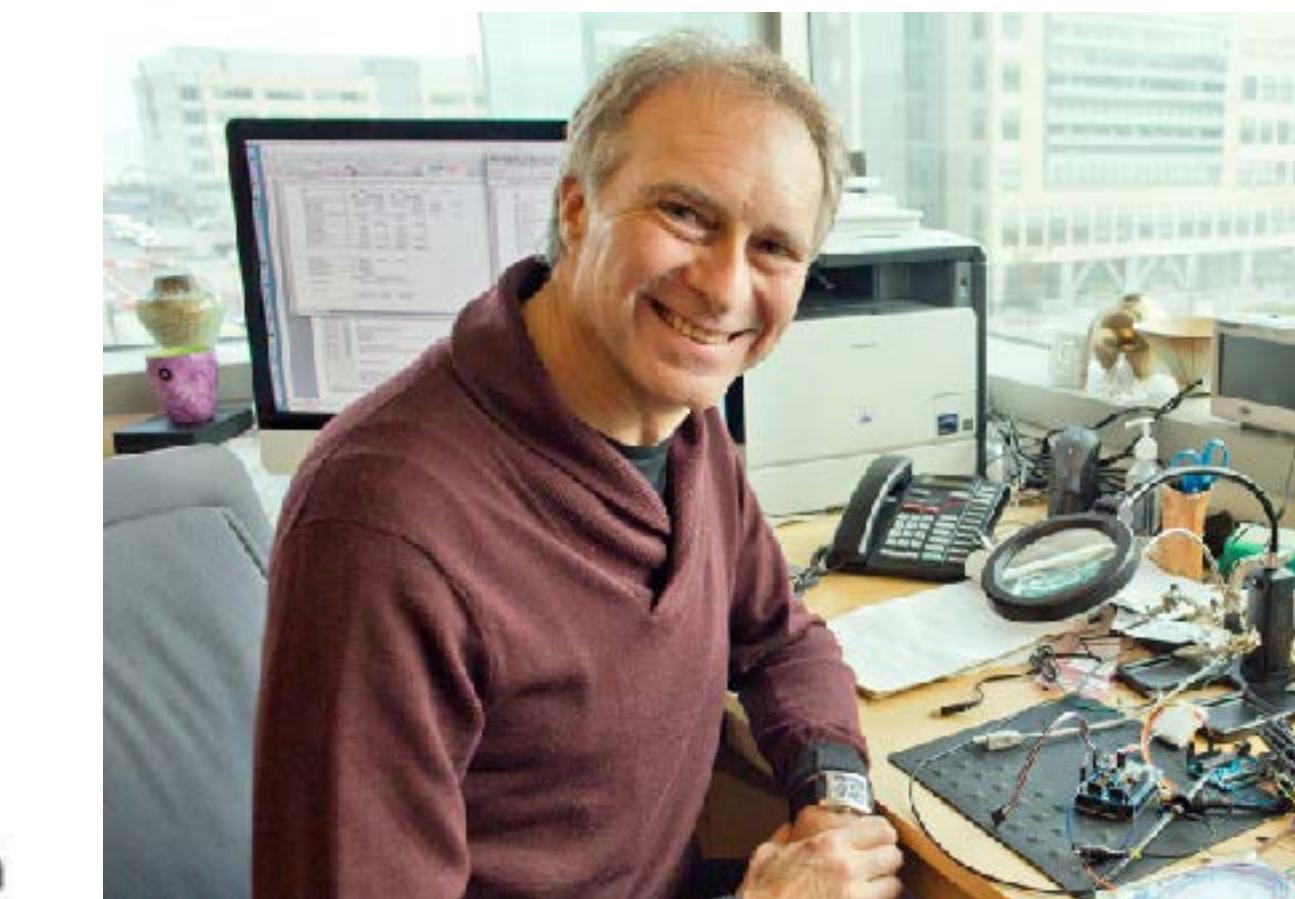


K2 Summit™ : Counting

4b. Events are localized with sub-pixel accuracy



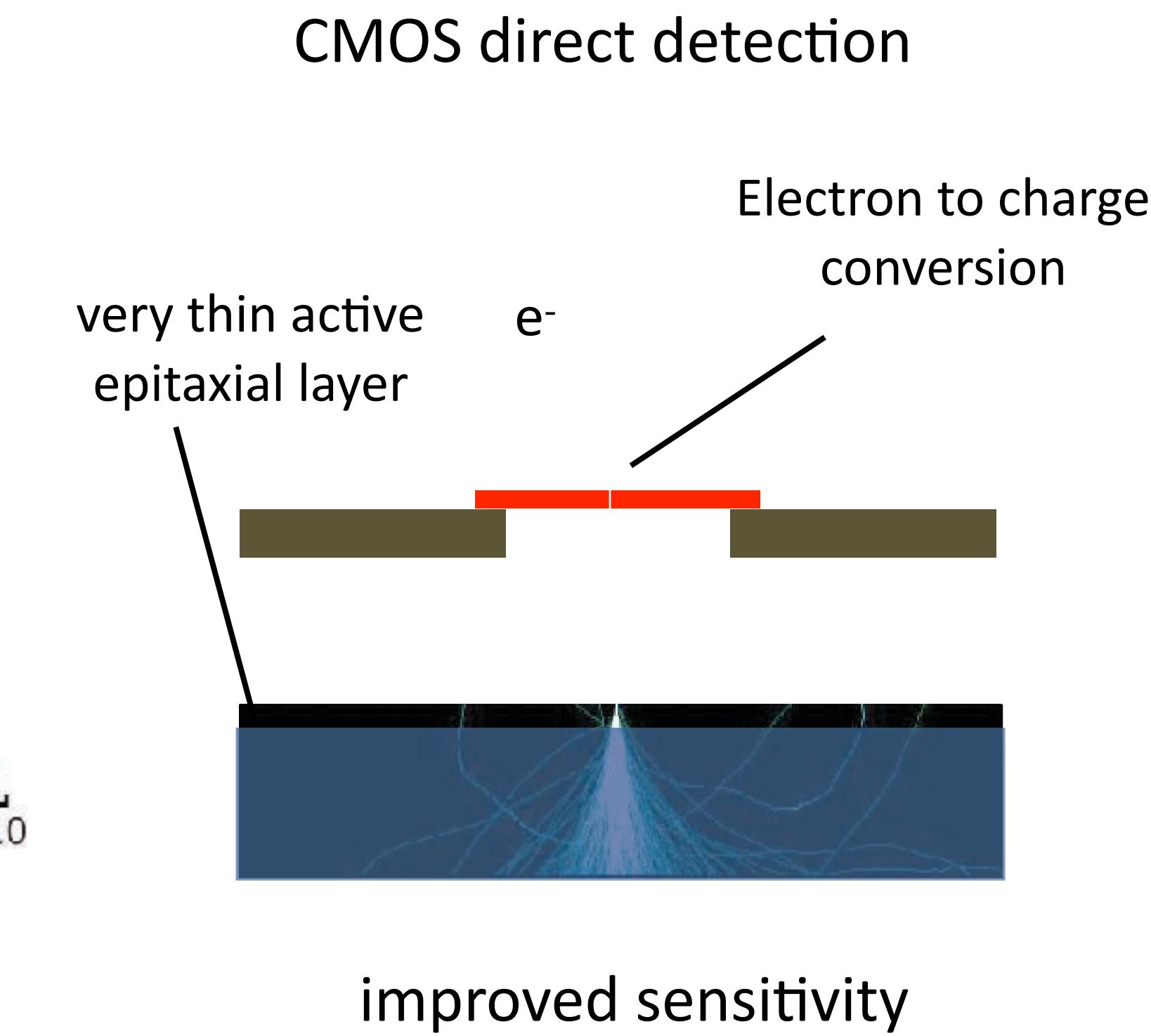
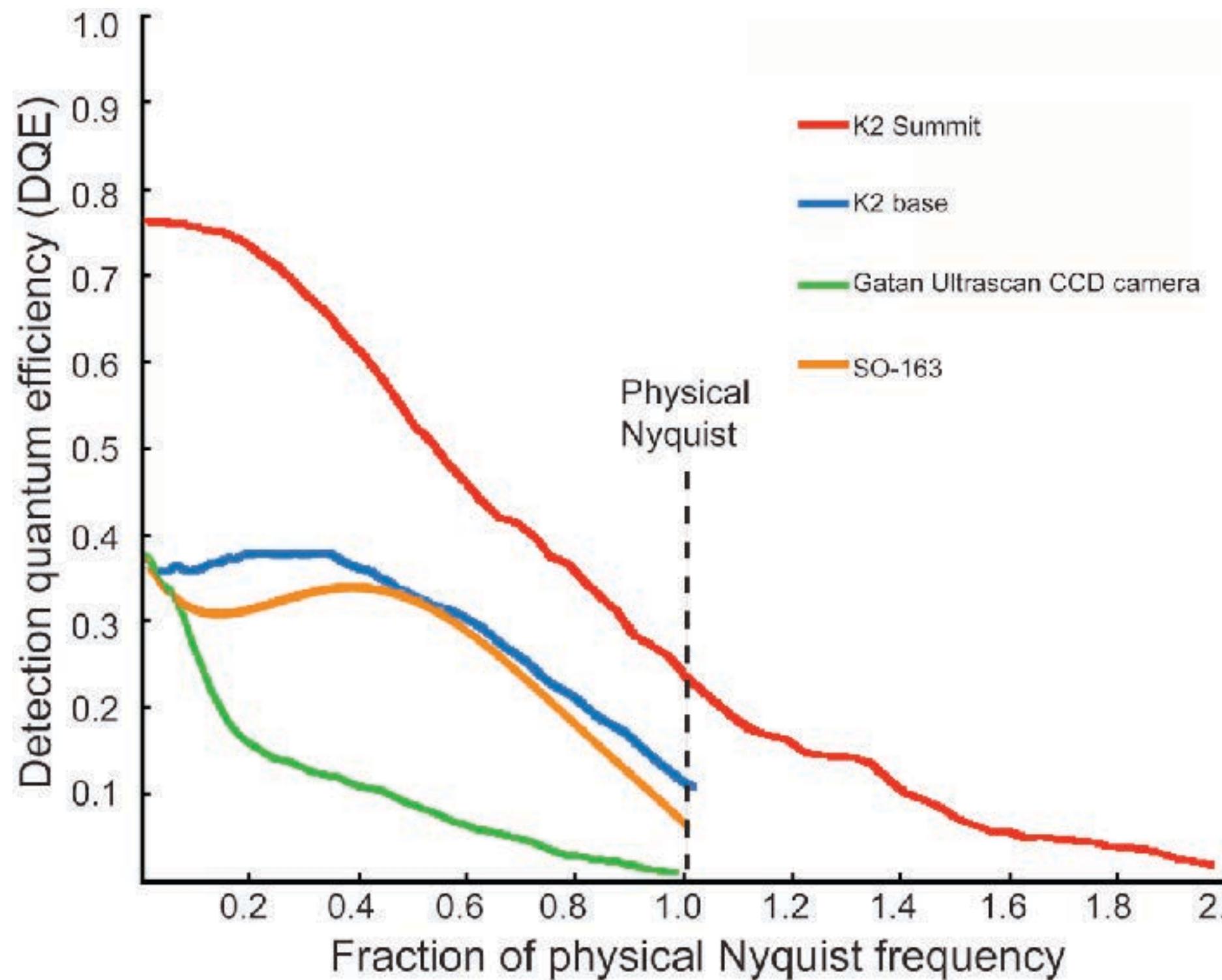
K2 Summit™: Super Resolution



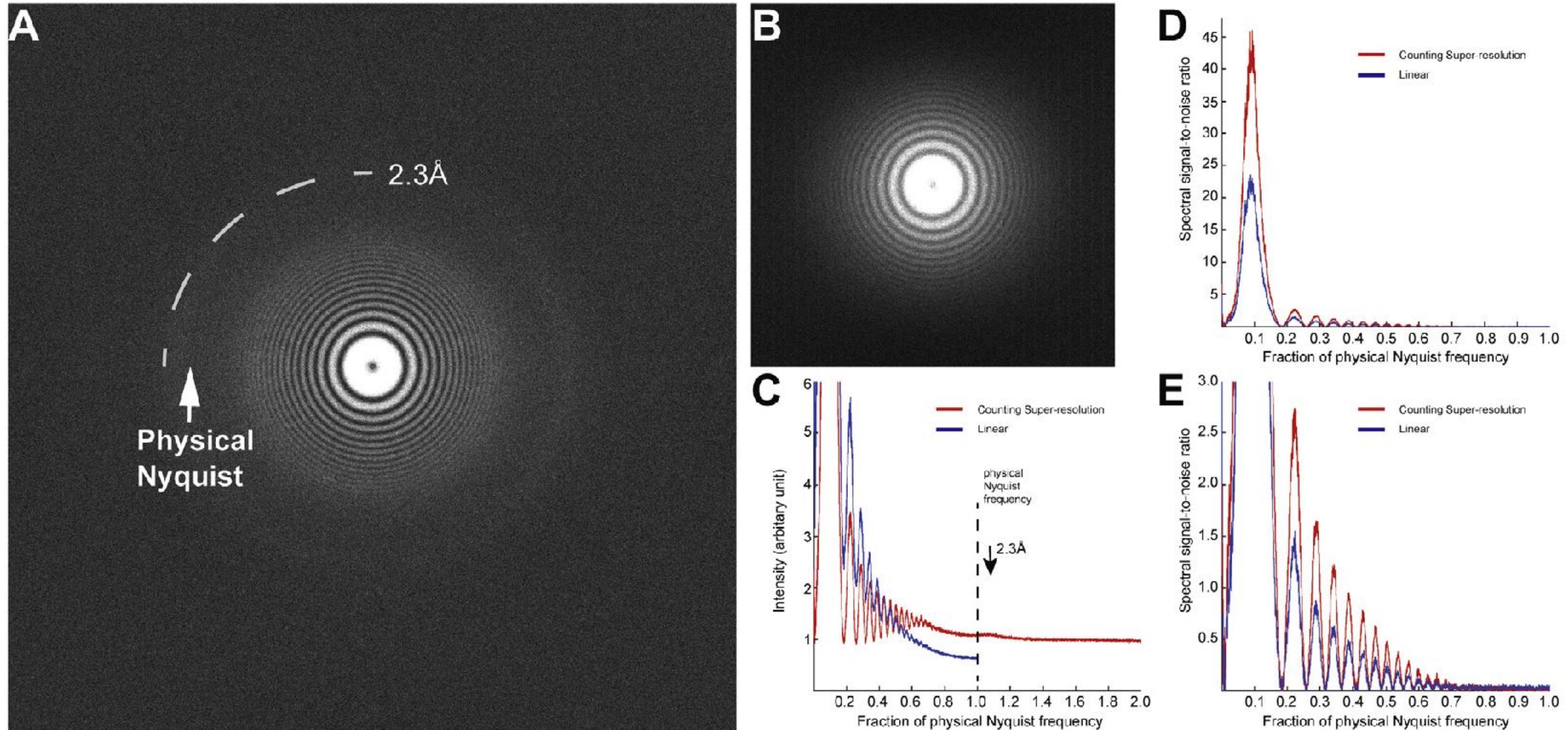
with David Agard (HHMI/UCSF)

Single electron counting improves DQE

- Direct detection of single electron remove read out noise
- Rapid read out enabled recording image as movie



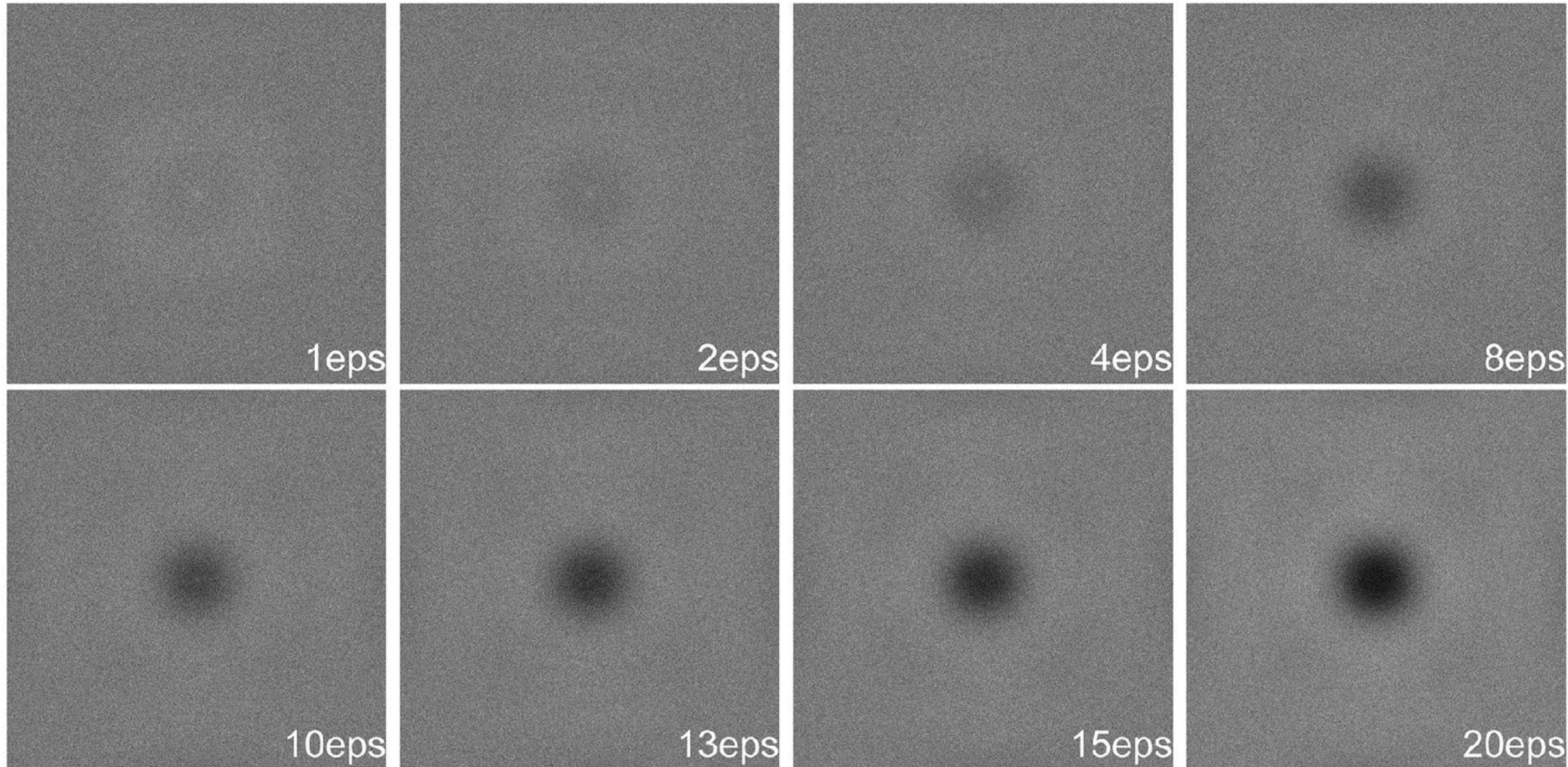
Single electron counting image and linear image



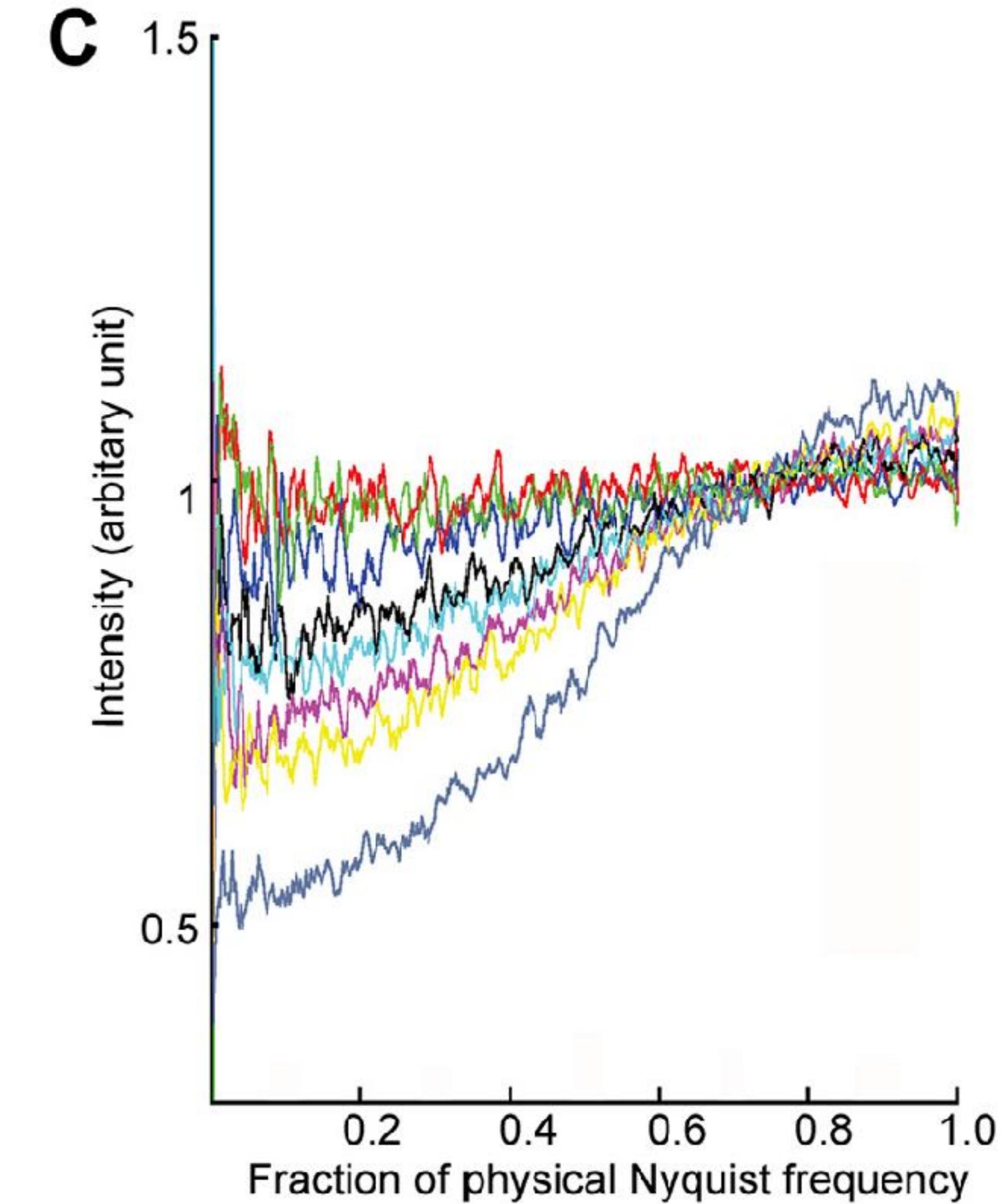
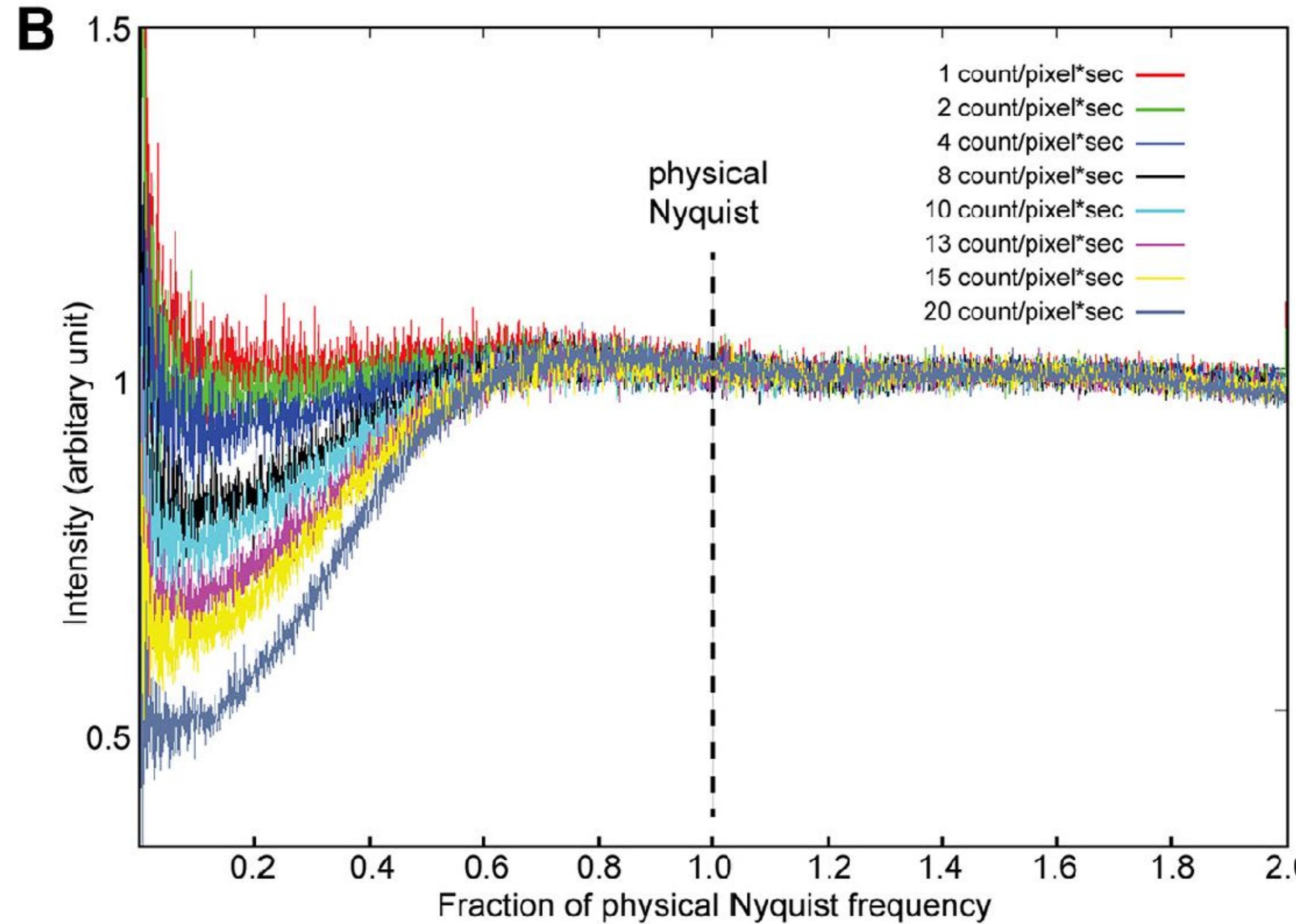
Comparison of images recorded with K2 counting and linear modes. Images of thin Pt/Ir film were recorded by the K2 Summit in both counting and linear modes. The

Coincidence loss

A

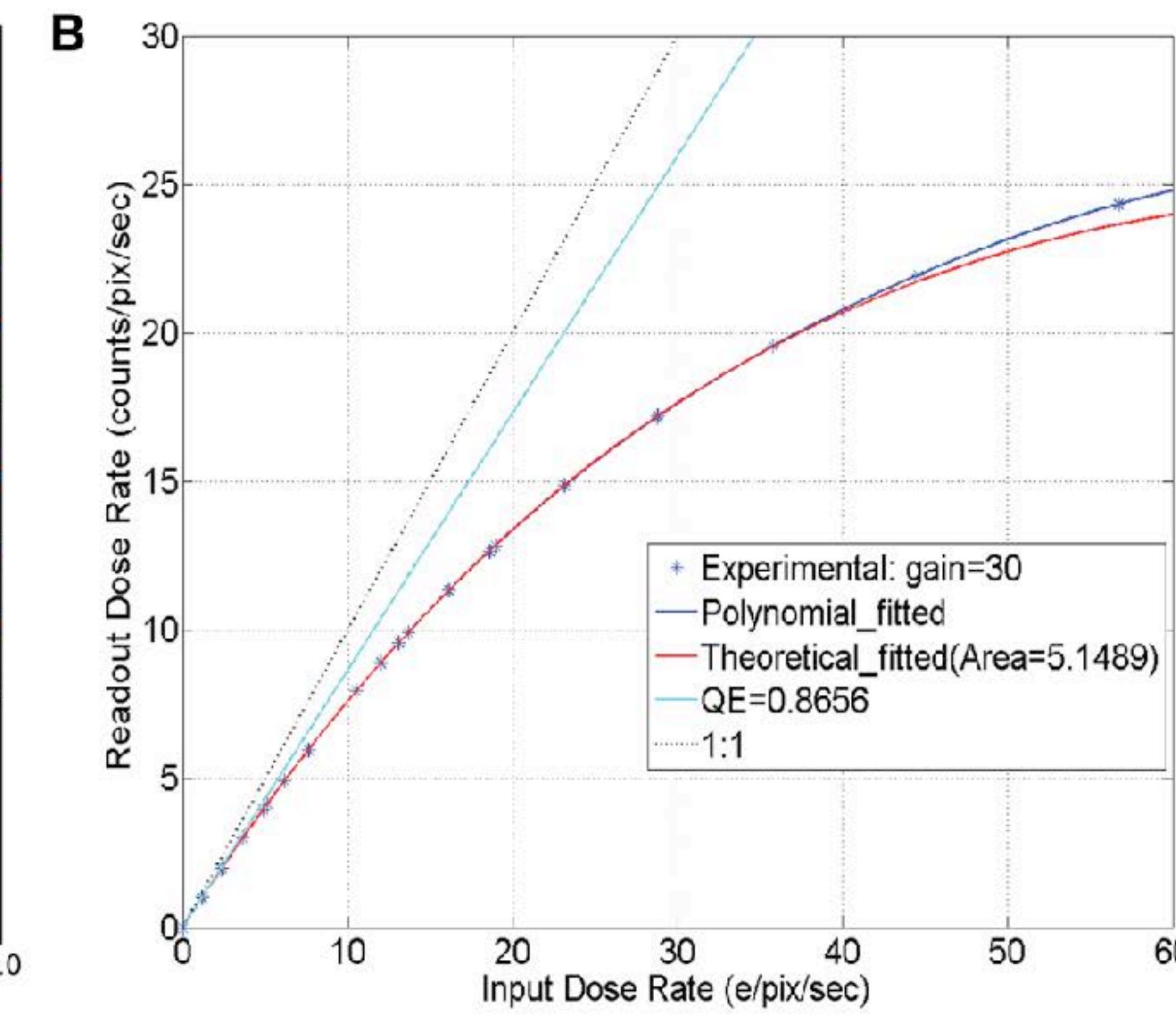
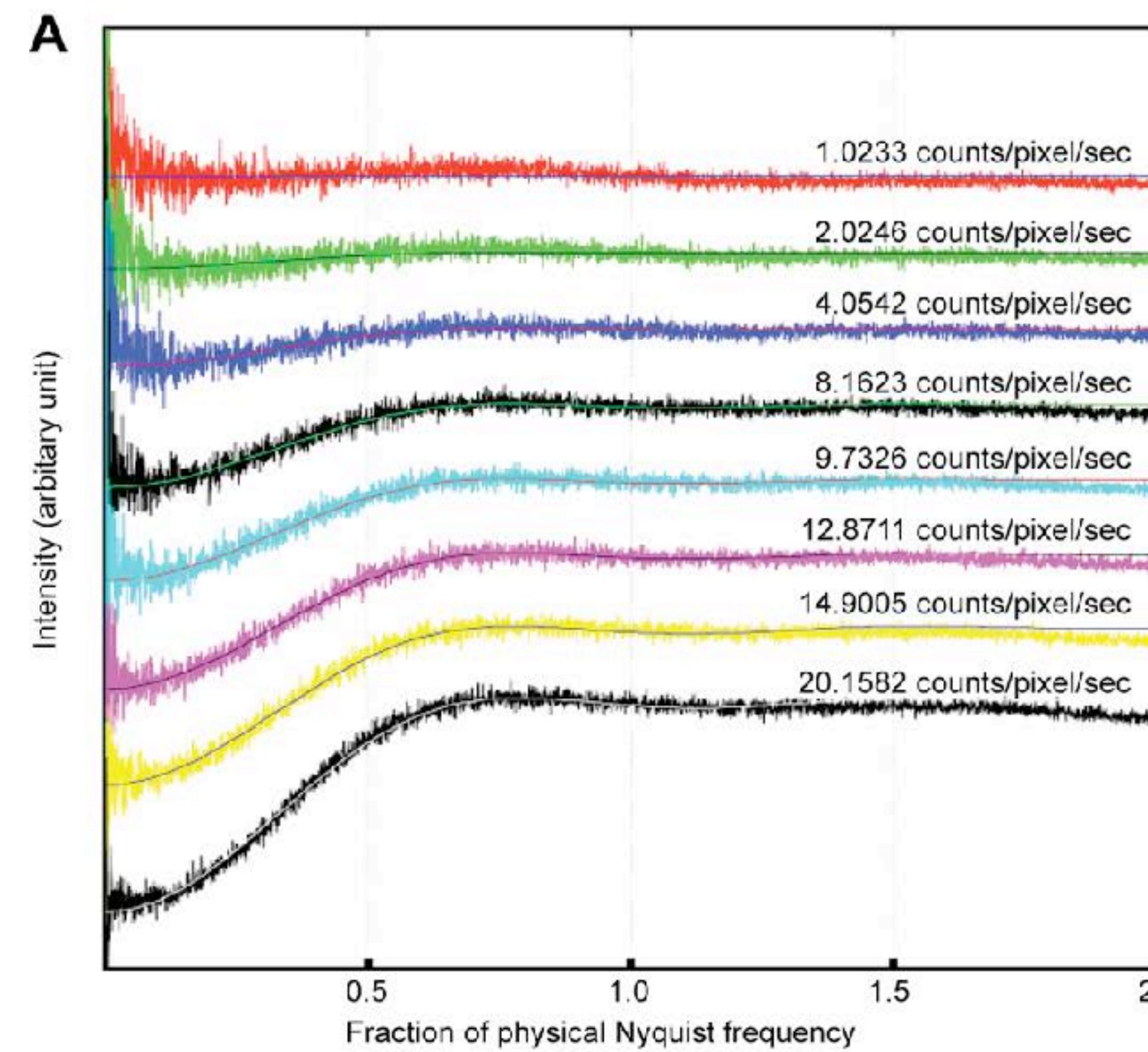


Coincidence loss



Coincidence loss

- When two electron strike the same pixel during one frame, only one is counted - coincidence loss;
- Coincidence loss deteriorate image quality, reduced linearity and DQE of camera;



Coincidence loss

- Coincidence loss is related to frame rate and camera pixel size;

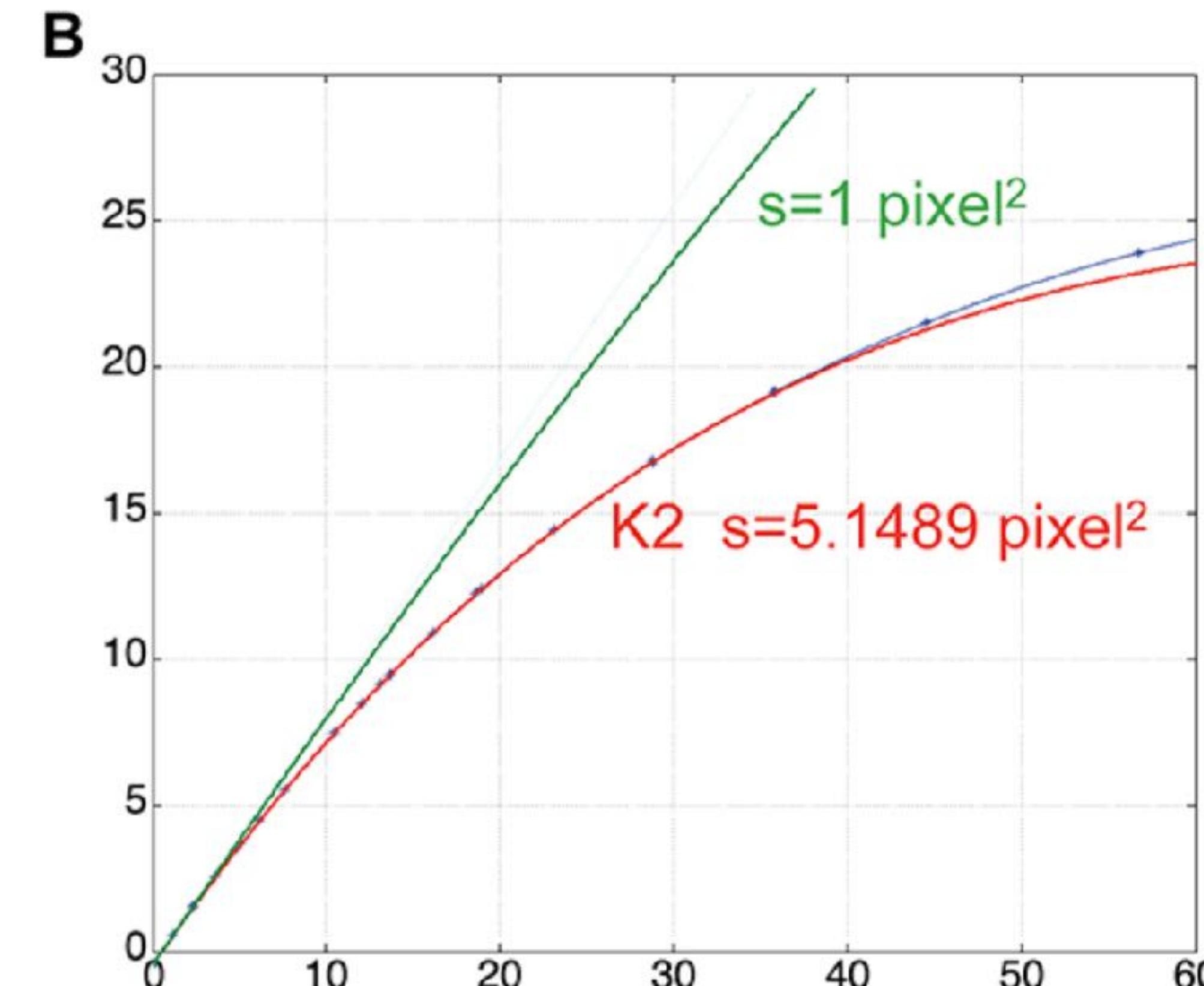
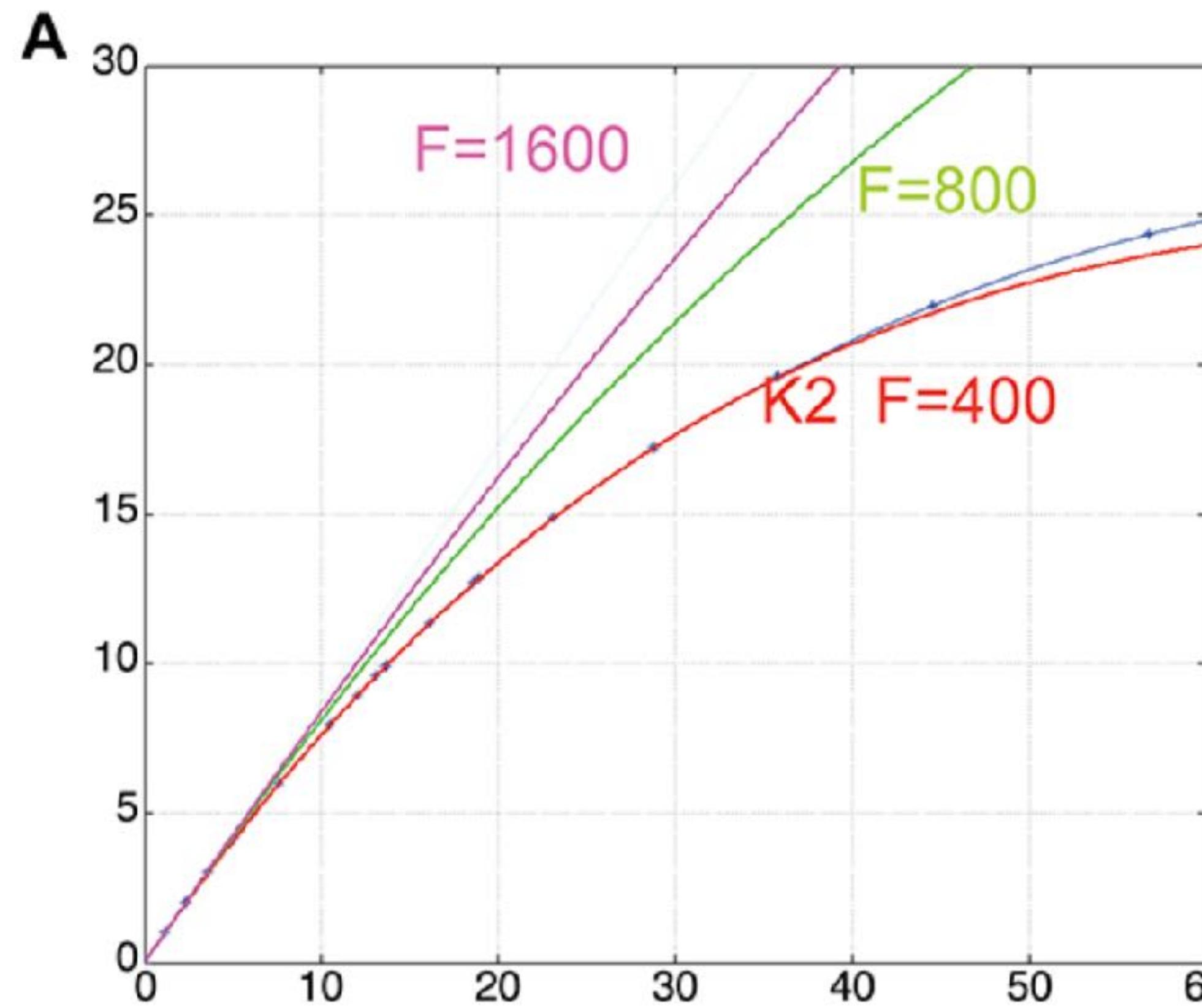
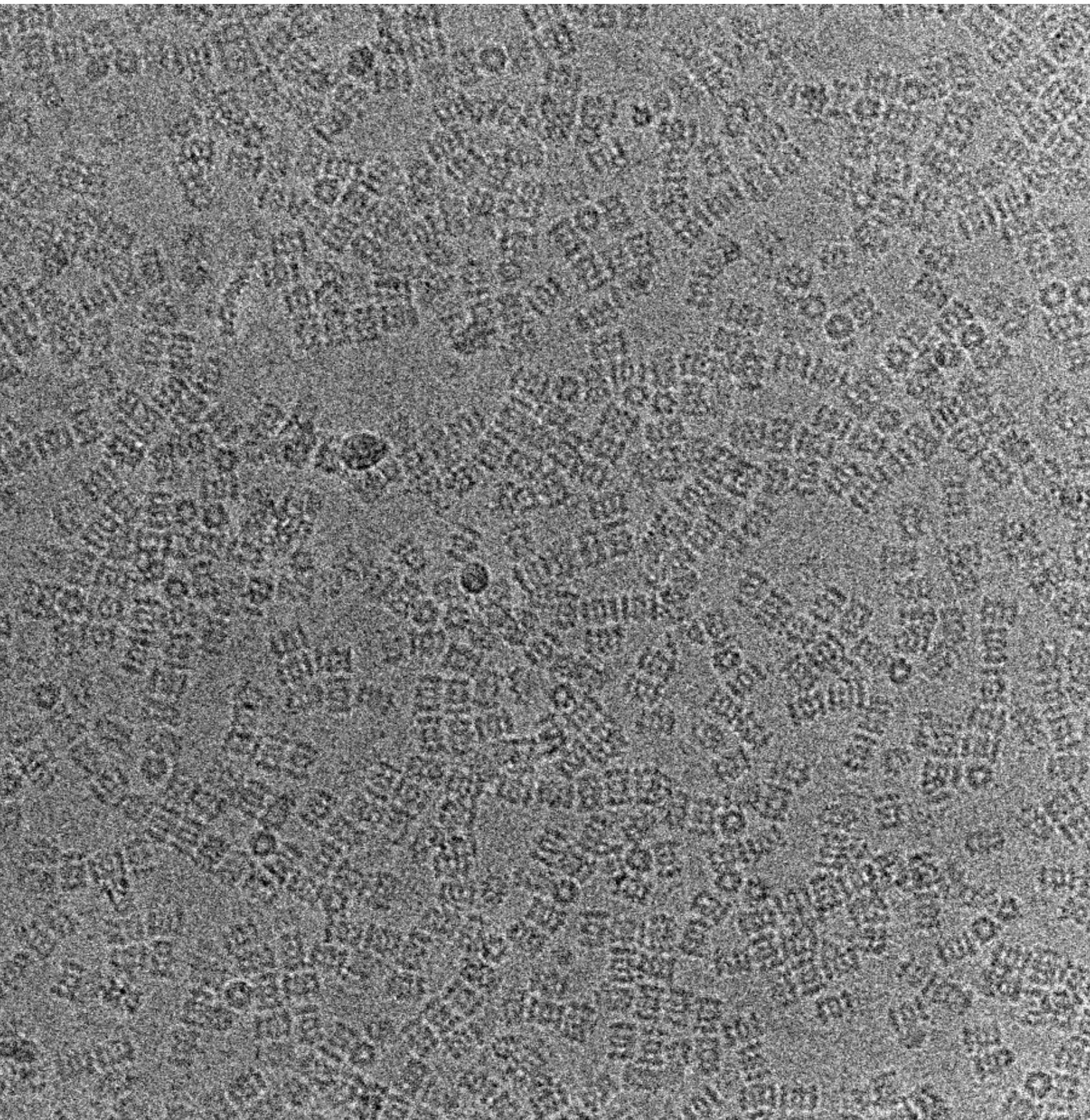
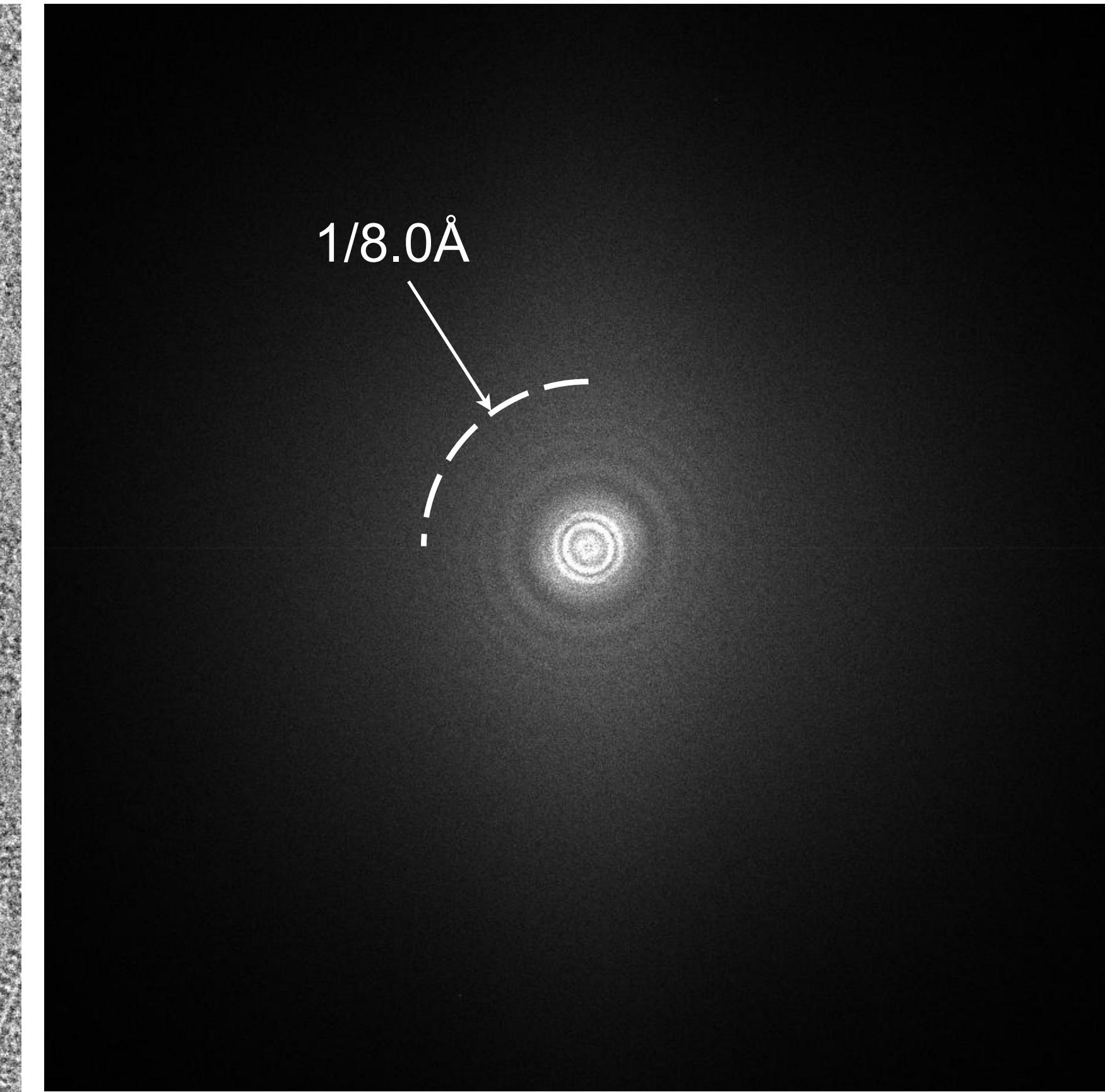


Image is further modified by recording devices

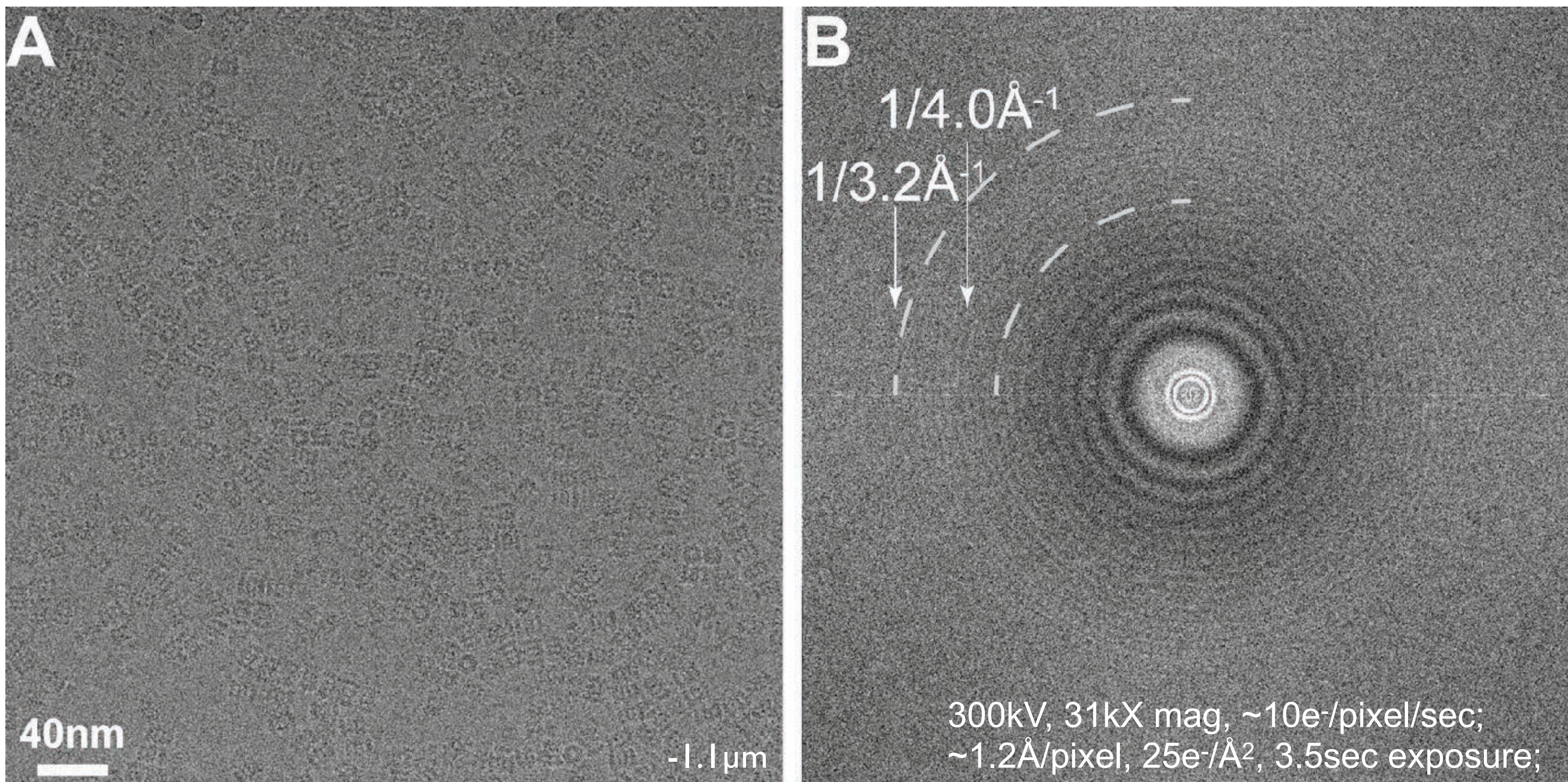


TVIPS F816 CMOS camera,
T20S proteasome, 700kDa,
200kV, -1.6 μ m (equivalent: 300kV, -2 μ m);

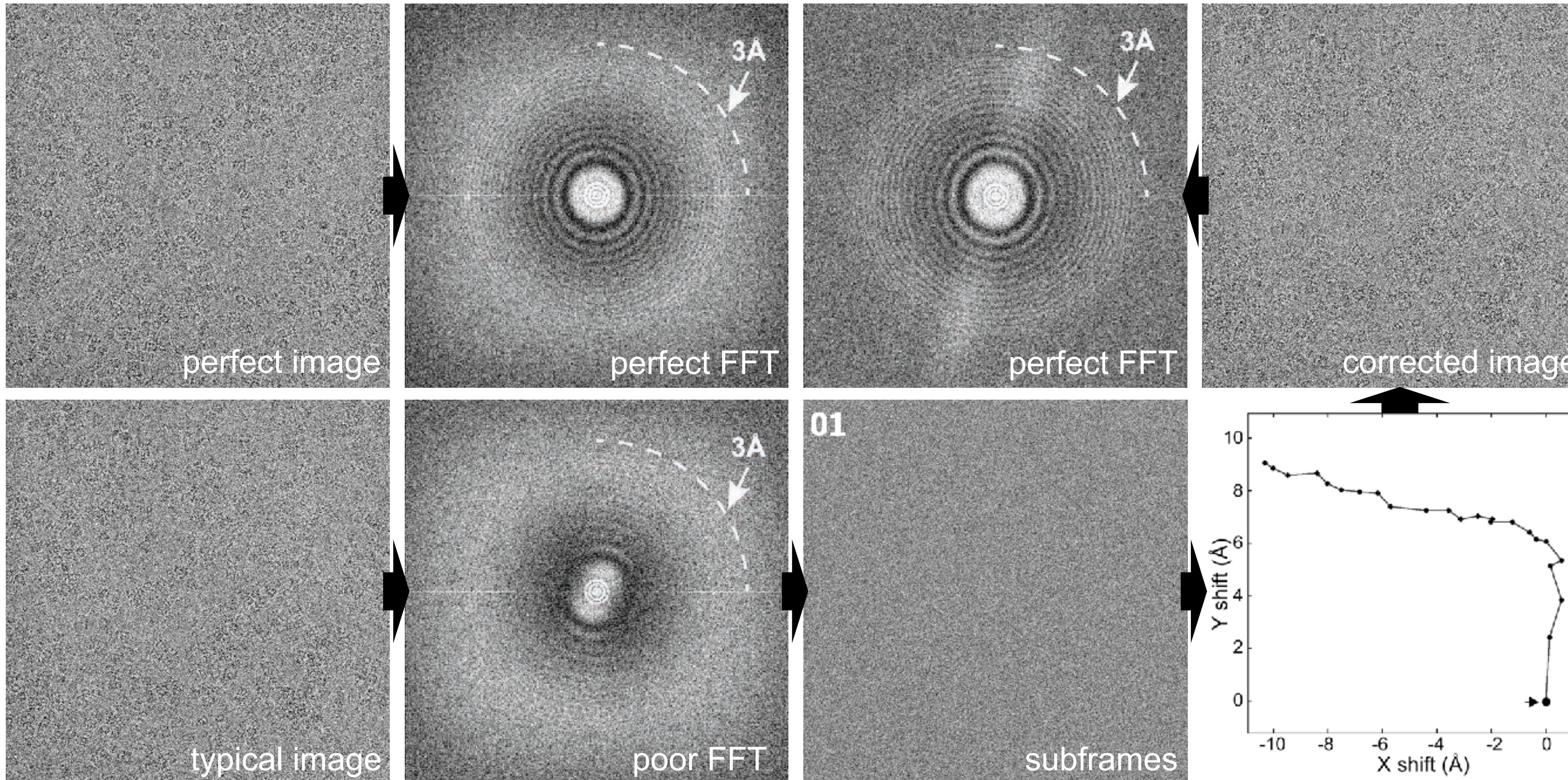


FFT: Thon ring visible to ~8.0 Å;

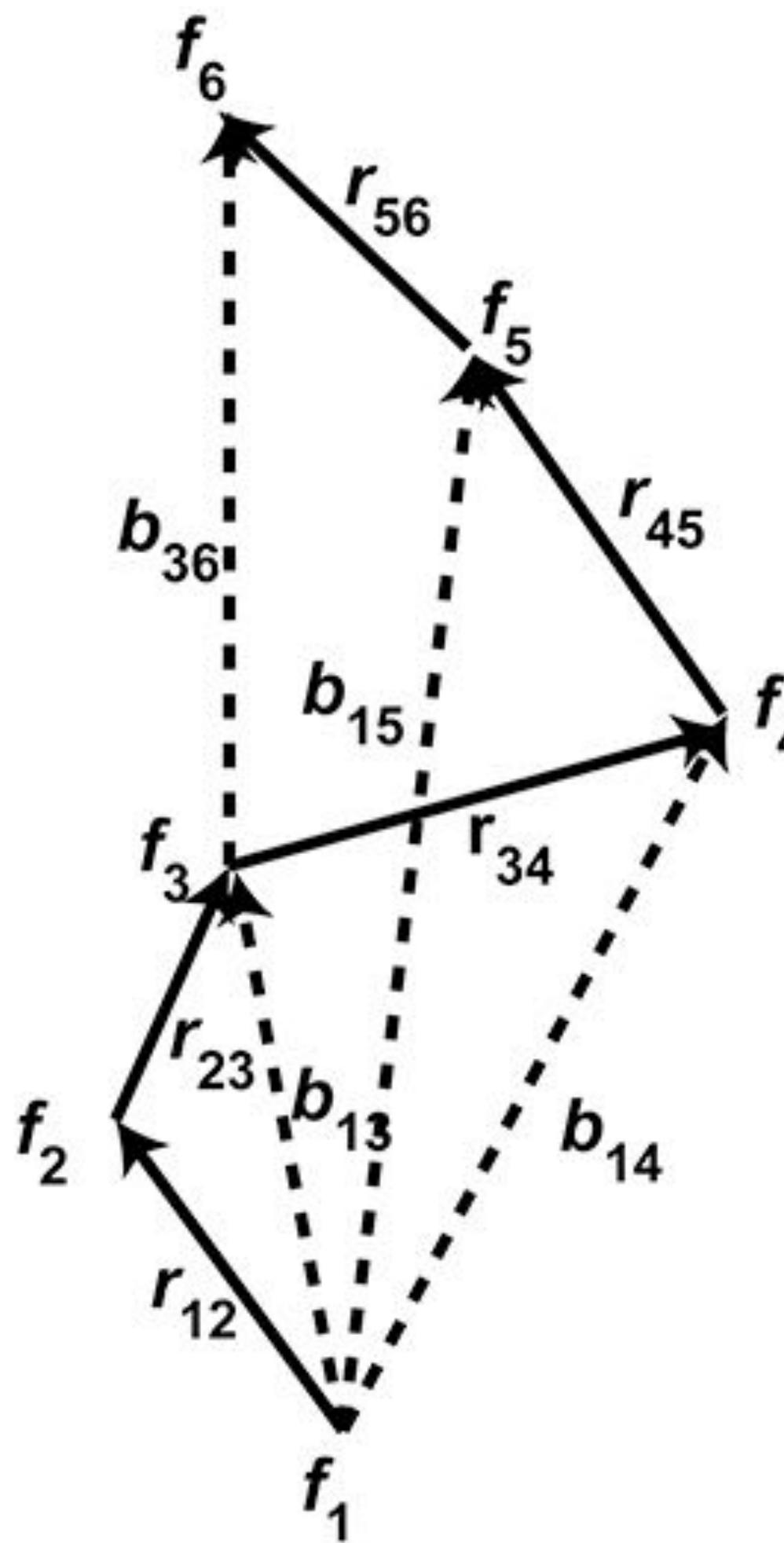
K2 image of frozen hydrated protein samples, archaeal 20S proteasome



Background subtraction recovers high-resolution quaternization



Algorithm for motion correction



$$b_{13} = r_{12} + r_{23}$$

$$b_{14} = r_{12} + r_{23} + r_{34}$$

$$b_{15} = r_{12} + r_{23} + r_{34} + r_{45}$$

$$b_{16} = r_{12} + r_{23} + r_{34} + r_{45} + r_{56}$$

.....

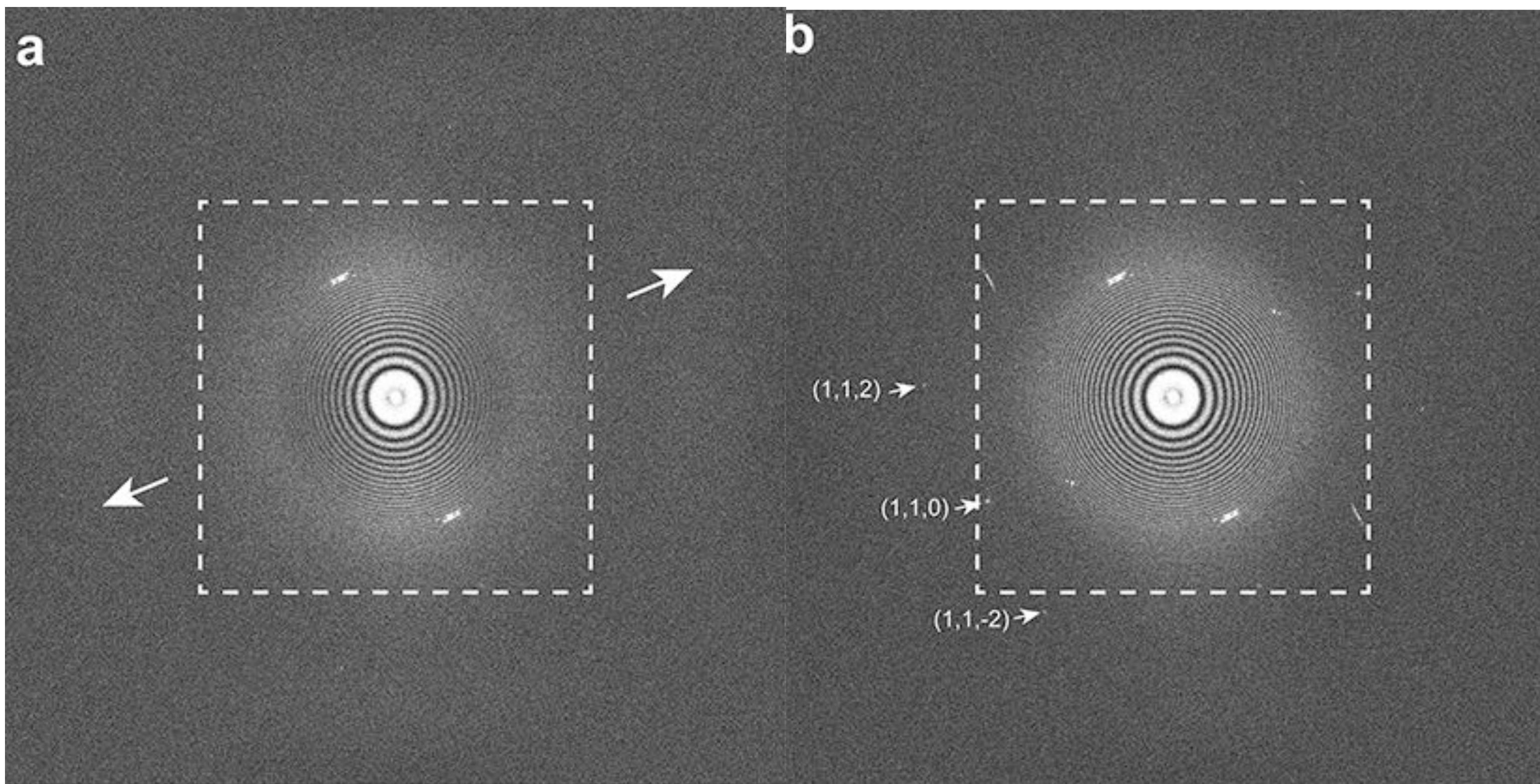
$$b_{35} = r_{34} + r_{45}$$

$$b_{36} = r_{34} + r_{45} + r_{56}$$

.....

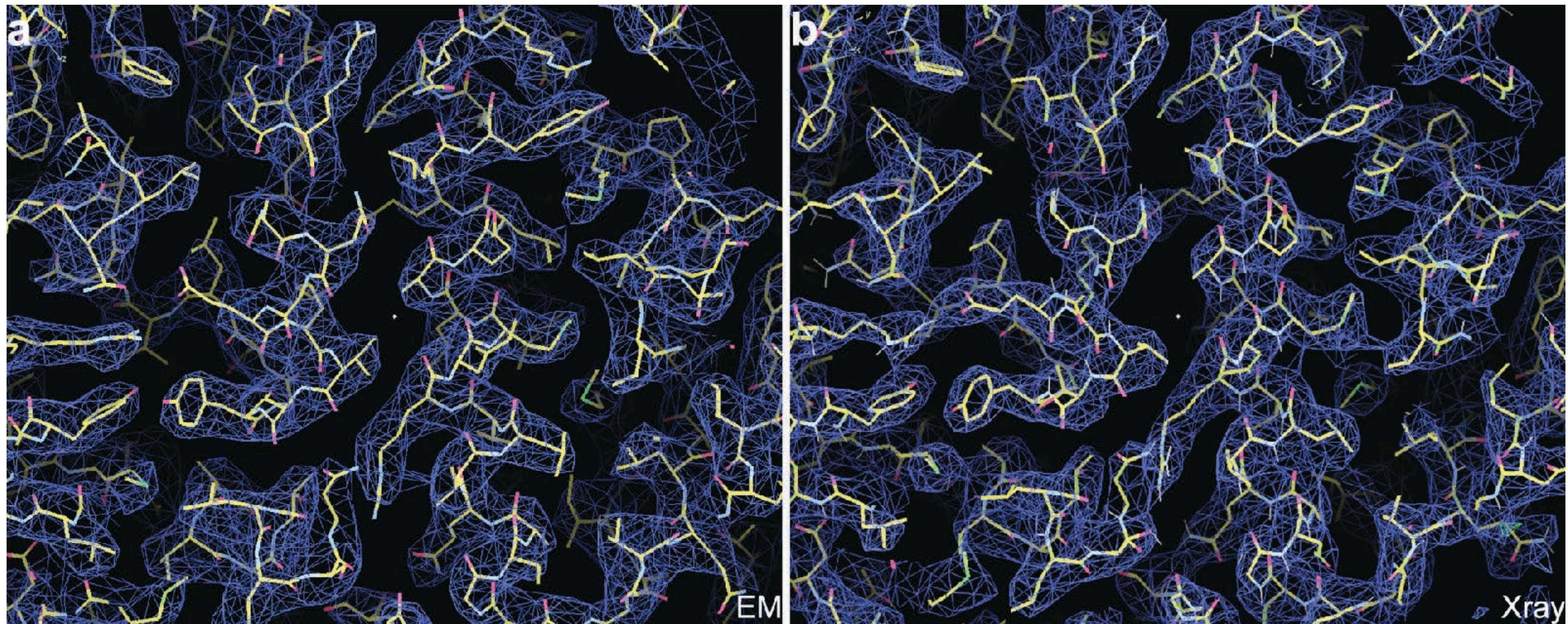
* All computations are carried out in GPU, and can be performed on-the-fly during data acquisition.

Motion correction at sub-pixel accuracy



* Motion correction restored resolution beyond physical Nyquist limit;

We achieved resolution comparable with X-ray crystallography



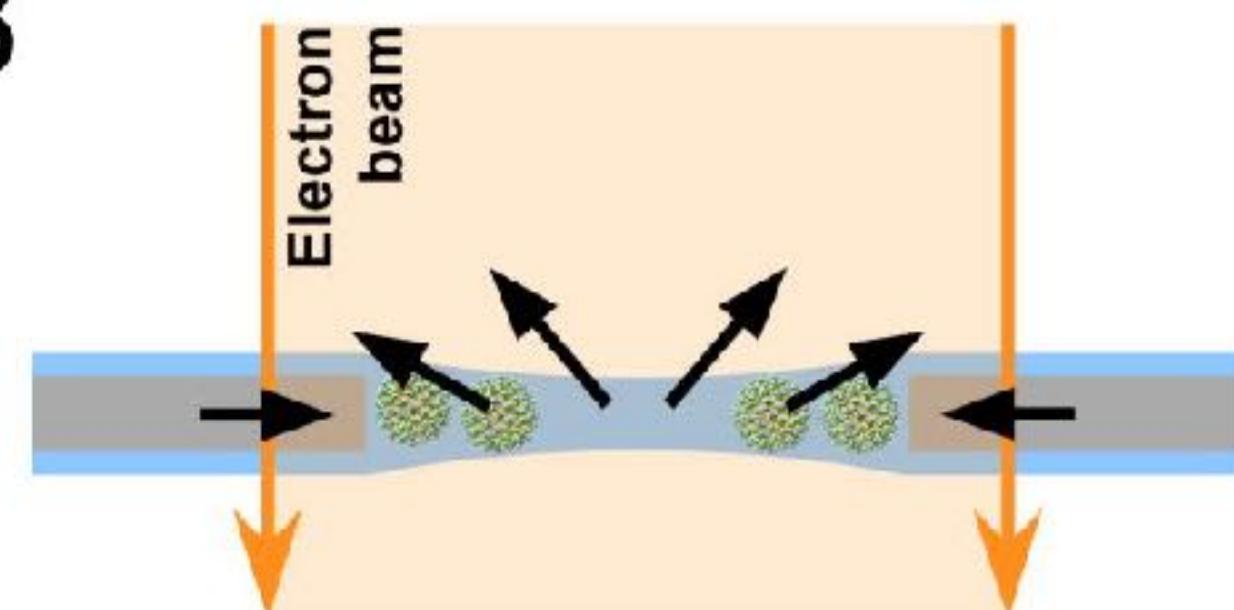
- archaeal 20S proteasome at ~3.3Å resolution, comparable to crystal map.

Caveat: all significant motion is not global

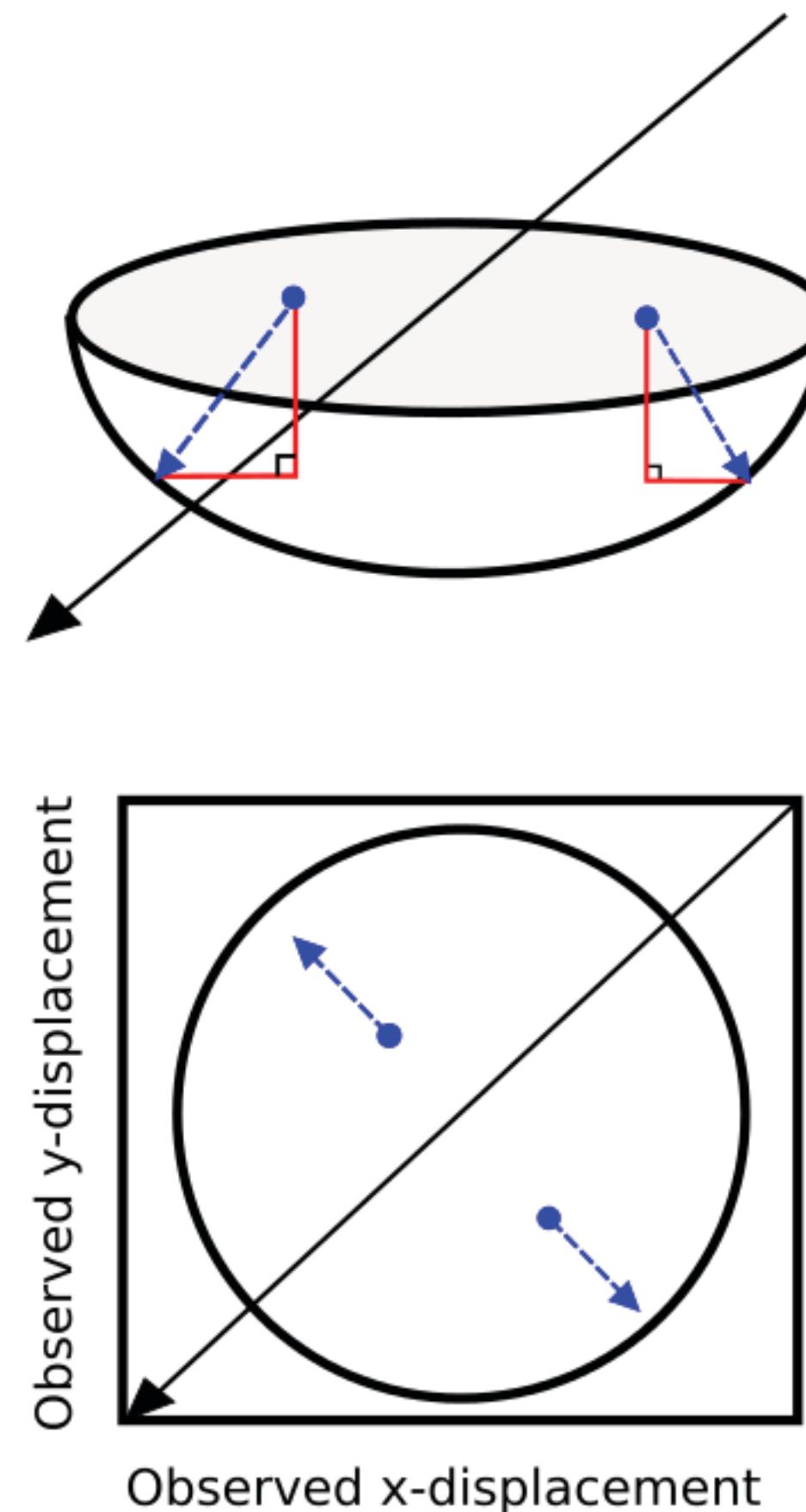
A



B



C

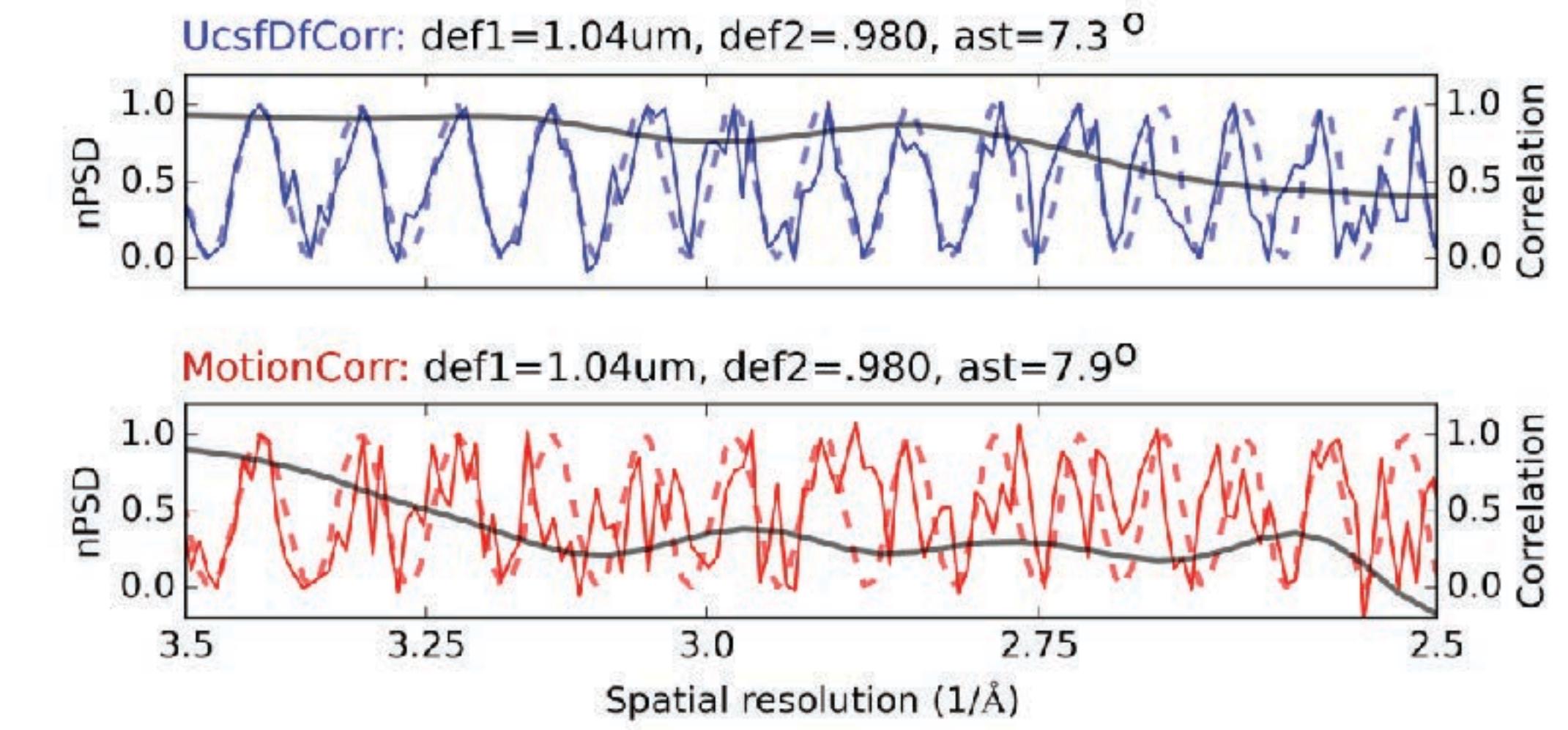
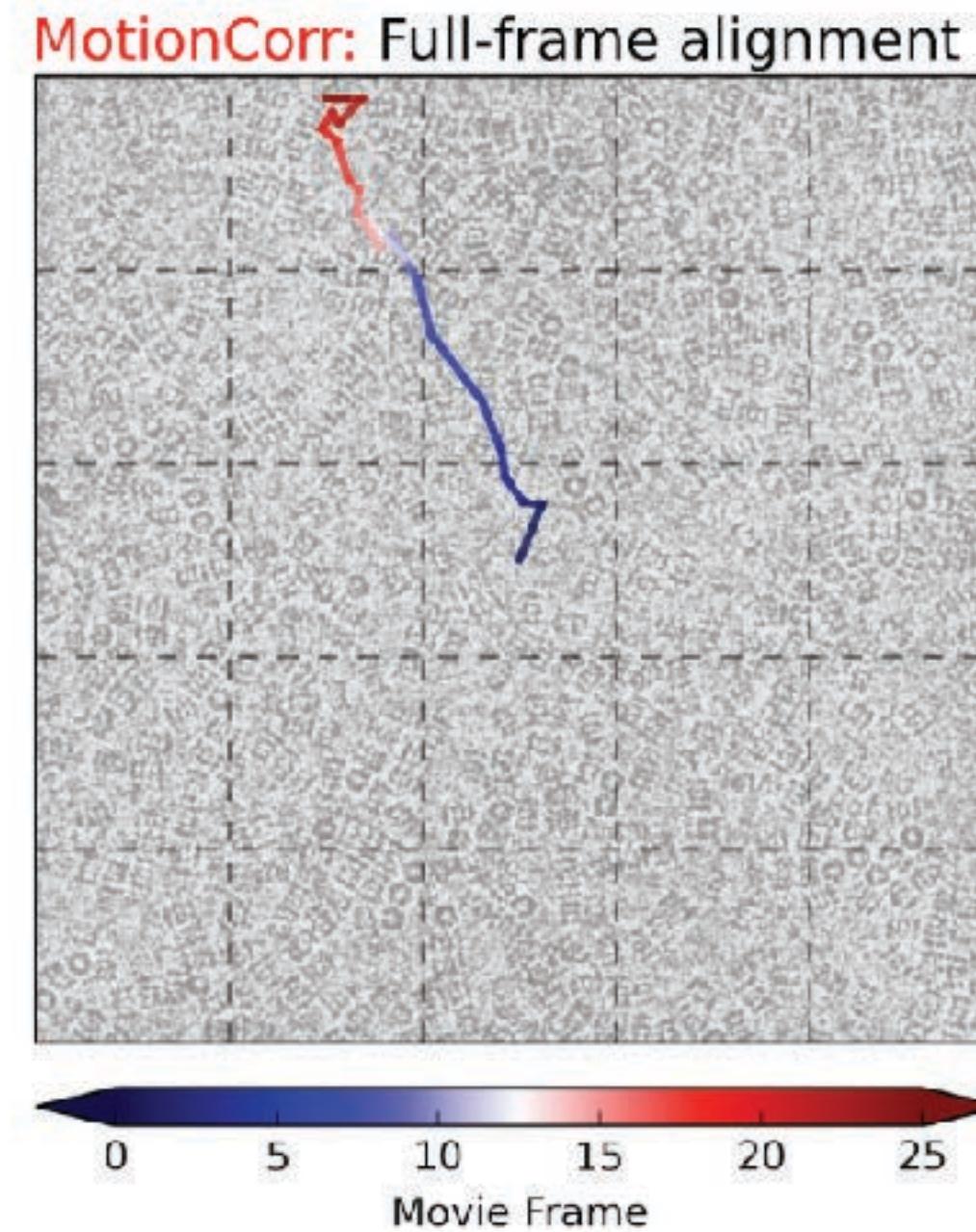
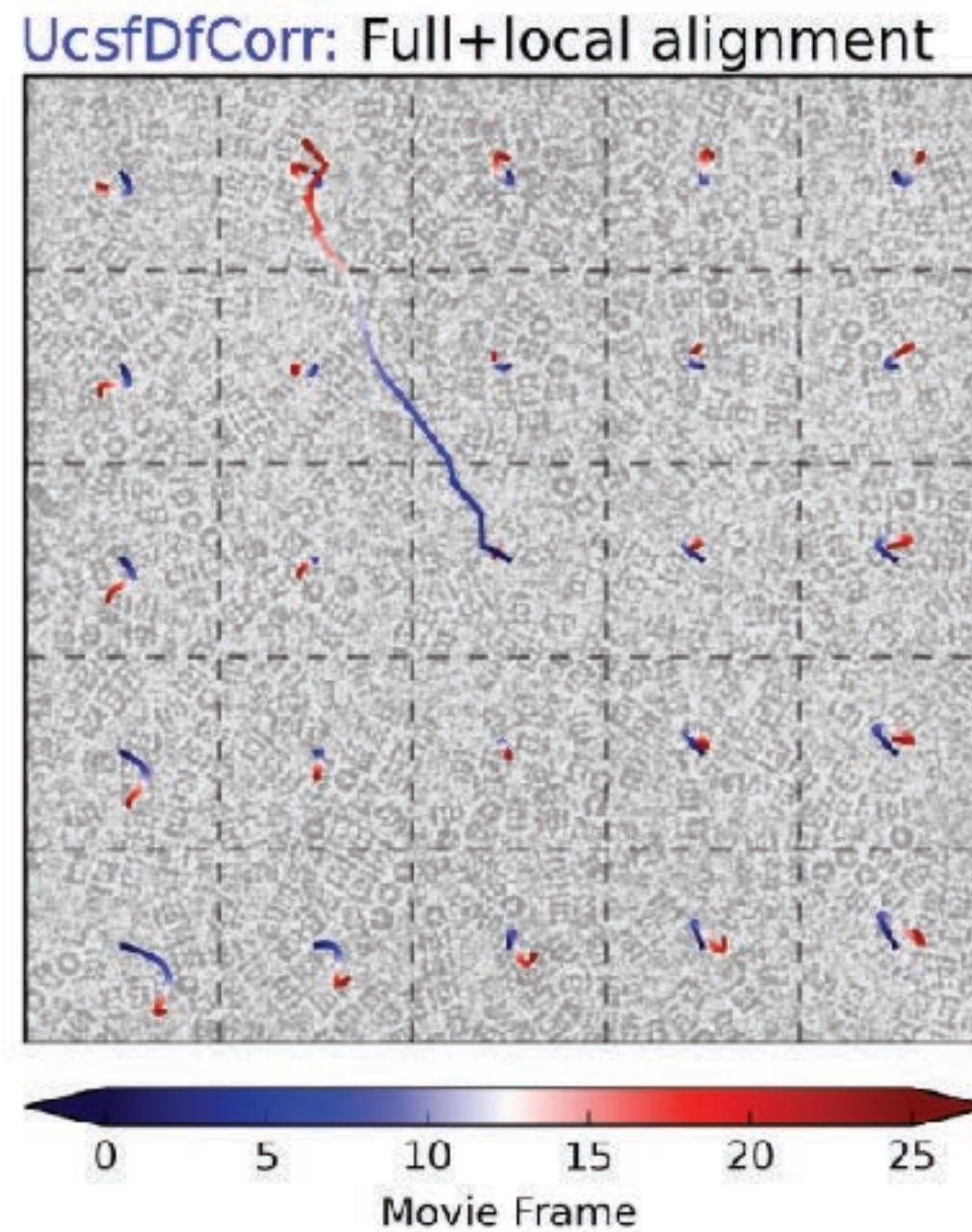


Axel Brilot (now in the Agard lab) discovered that vitrified viruses at the periphery of the sample hole move more than those in the center.

It was suggested that the electron beam causes the sample to 'dome.'

Shawn Zheng (also Agard lab) wrote a new algorithm that takes such motion into account.

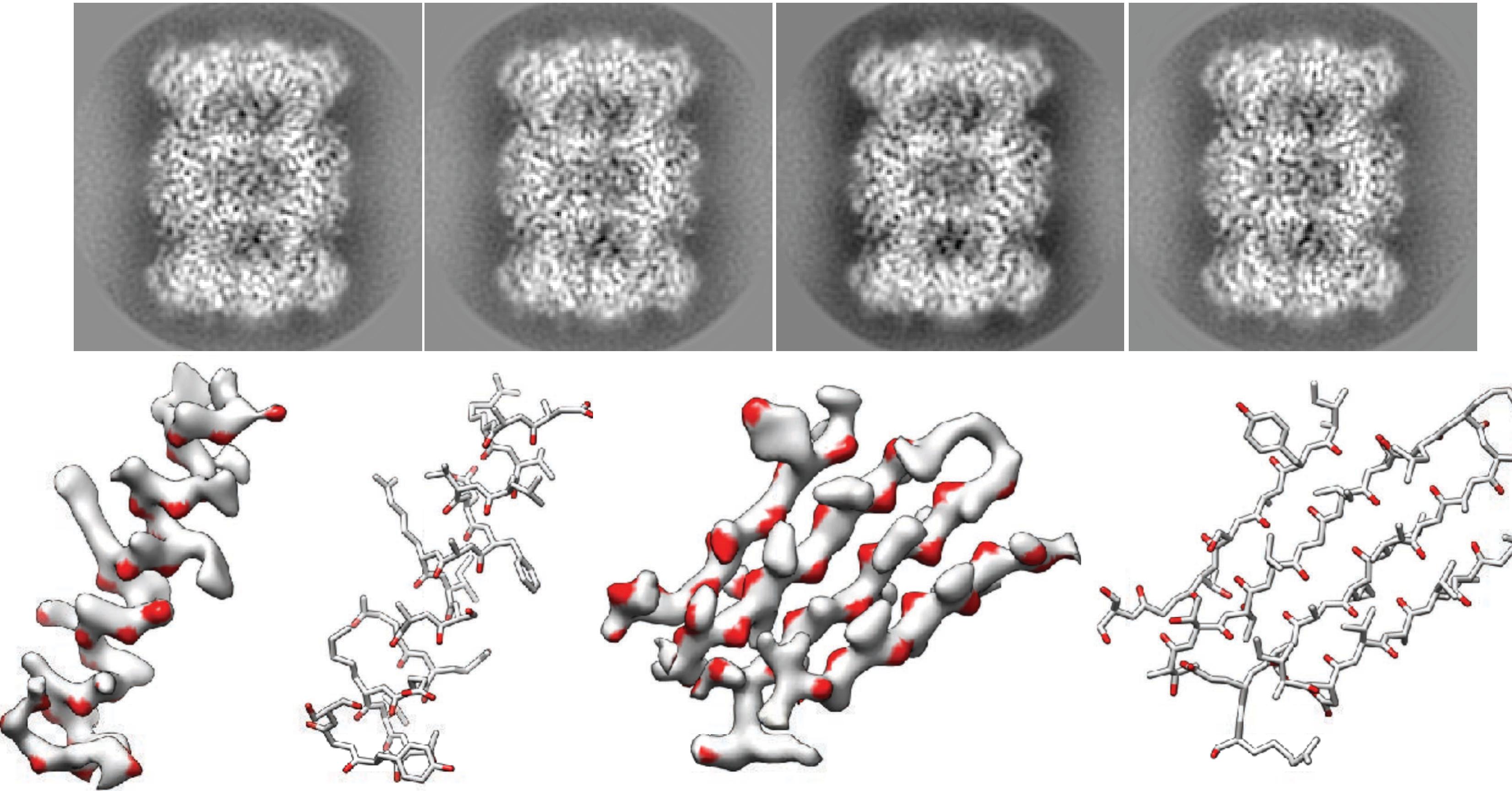
Caveat: all significant motion is not global



CTF Oscillations in the radially averaged Fourier Transform

Correcting local doming motions (interpolating each pixel of the image on the left with a time-varying vector field fitted by the trajectories in different patches of the image) improves the signal below 3A.

Re-process with the algorithm: archaeal 20S proteasome at ~2.5Å resolution

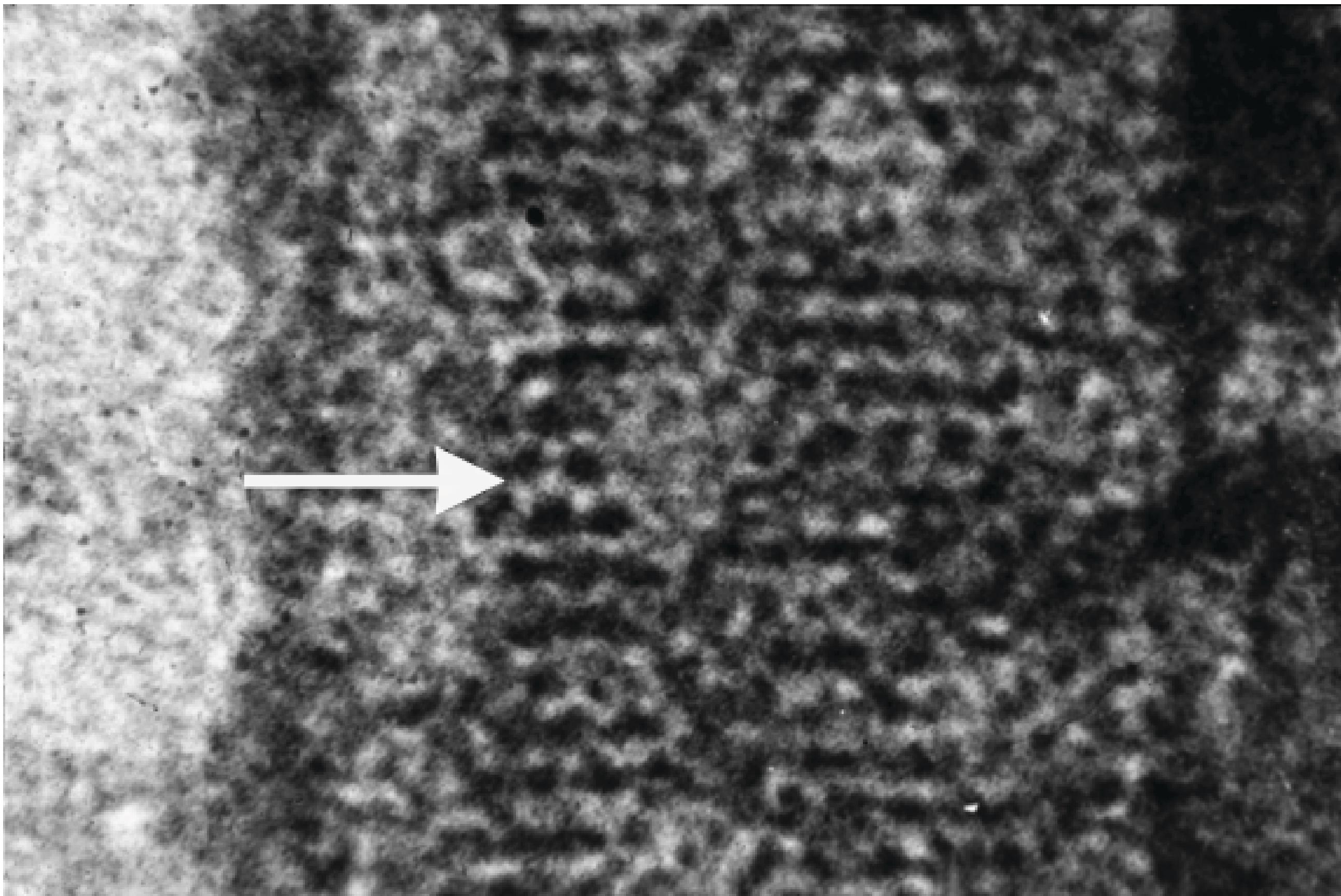


Single particle cryo-EM

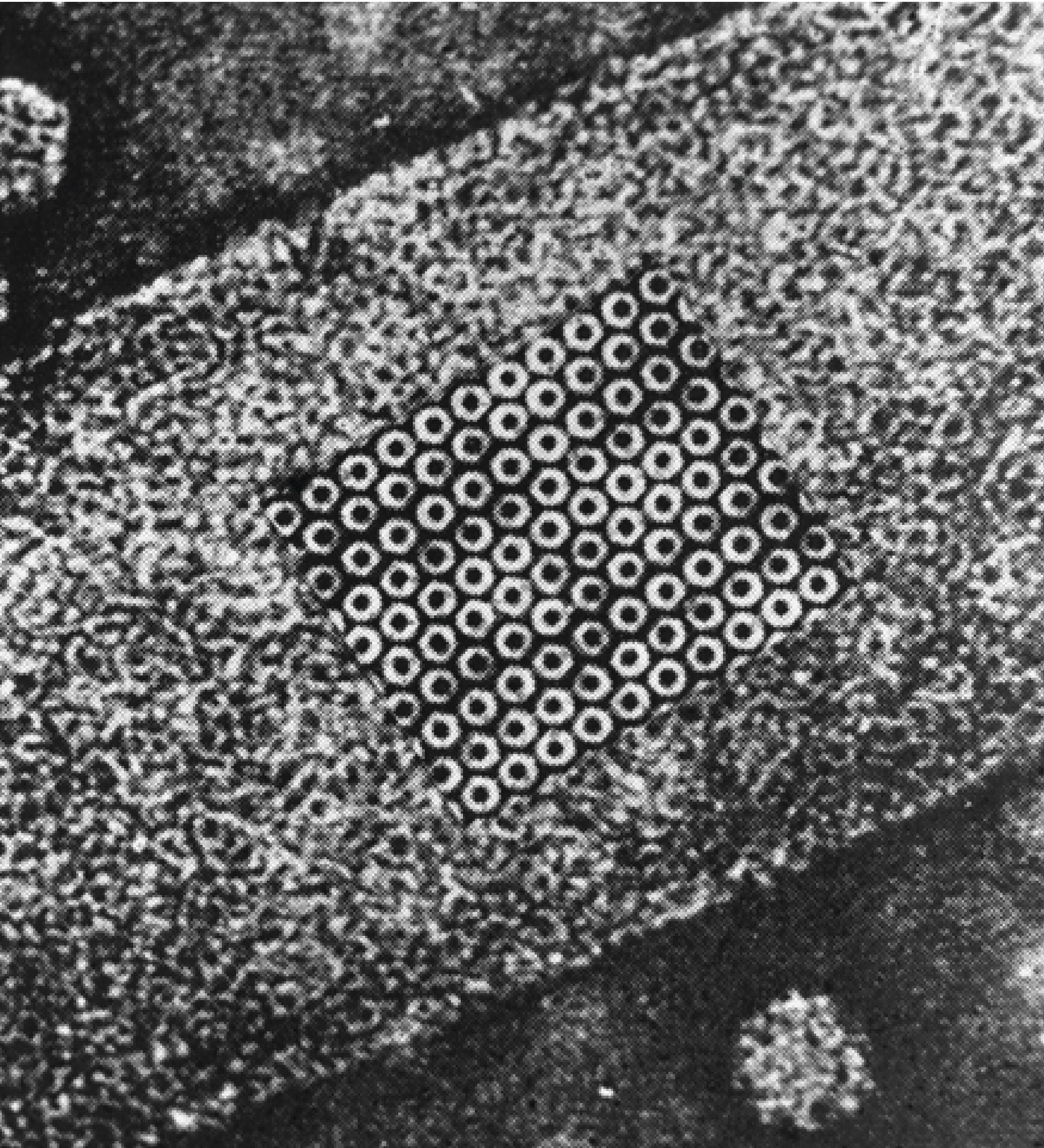
* Basic concepts of single particle cryo-EM: averaging, resolution, iterative refinement and reconstruction

Image averaging

Cryo-EM images are very noisy; have extremely low signal-to-noise ratio. Averaging of a large number of images are necessary to improve the SNR.



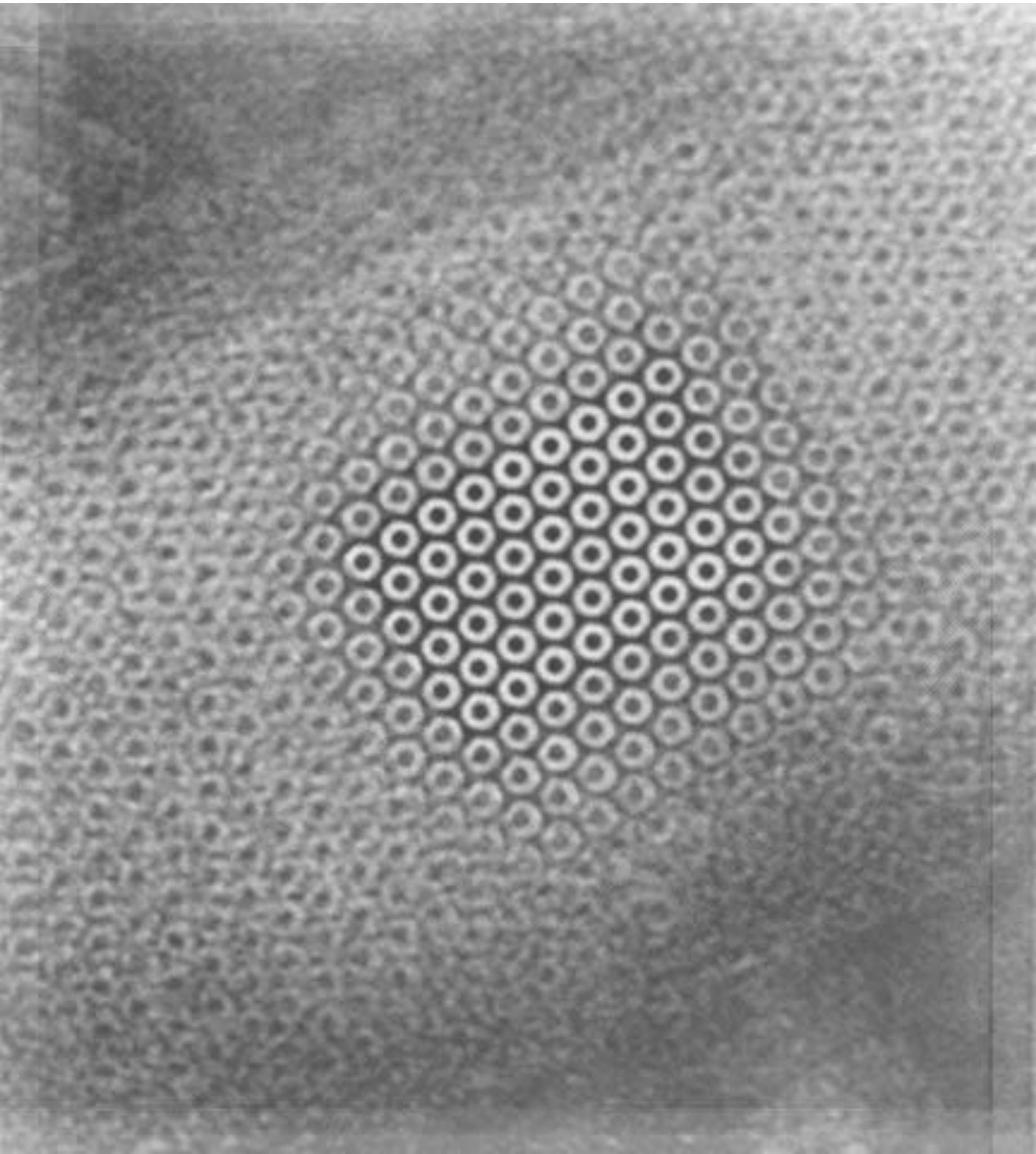
Averaging in darkroom



Photographic image superposition (averaging) by Roy Markham, who shifted image and added to the original in darkroom.

The trick is to know decide much and which direction to shift the image for superposition.

Averaging in computer.



David DeRosier used Markham's lattice to determine how much to shift, and performed averaging by using Adobe Photoshop.

Averaging in 2D crystals

How much and which direction to shift the image can be determined easily from FT of the image of a 2D crystal.

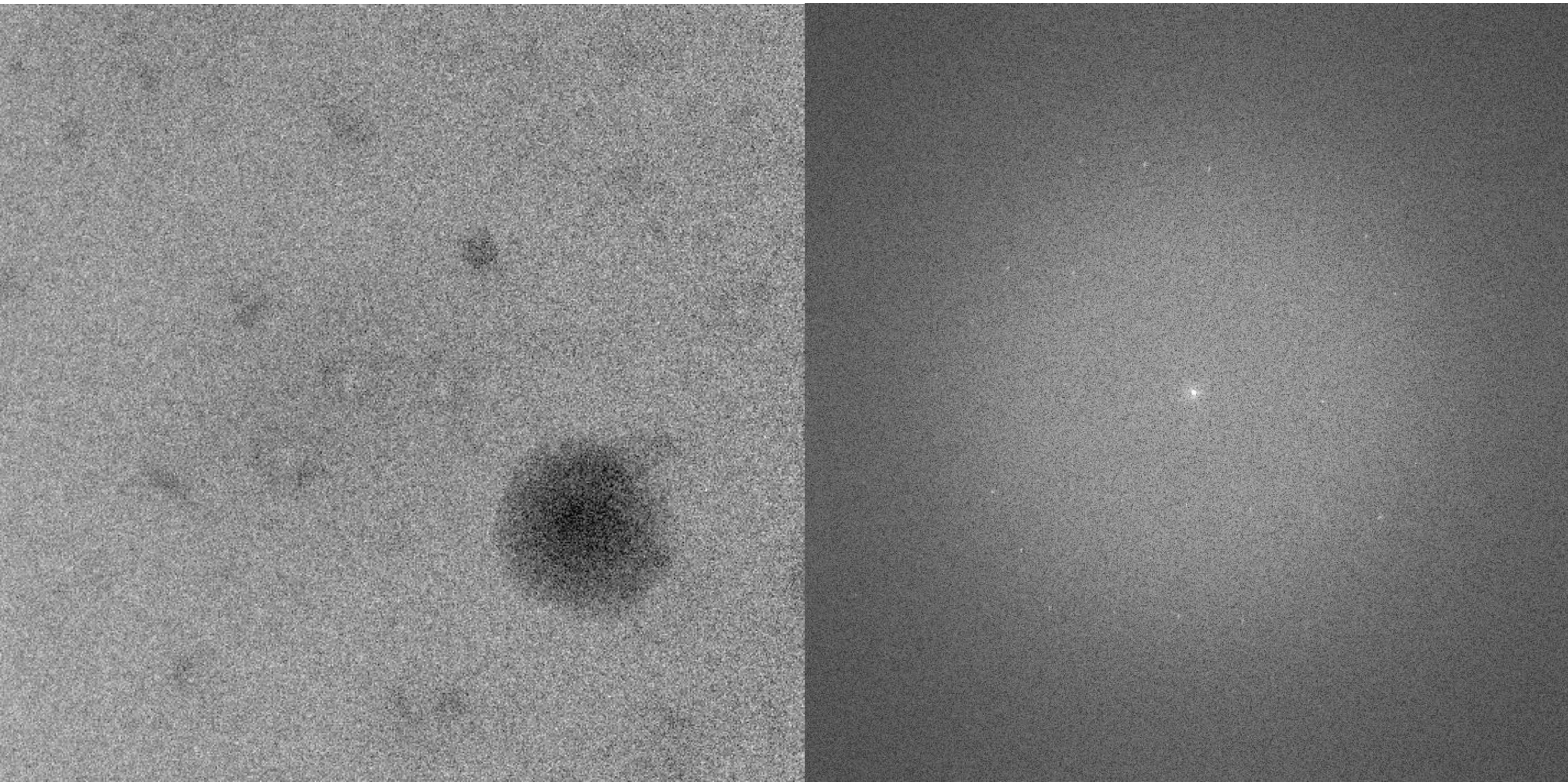
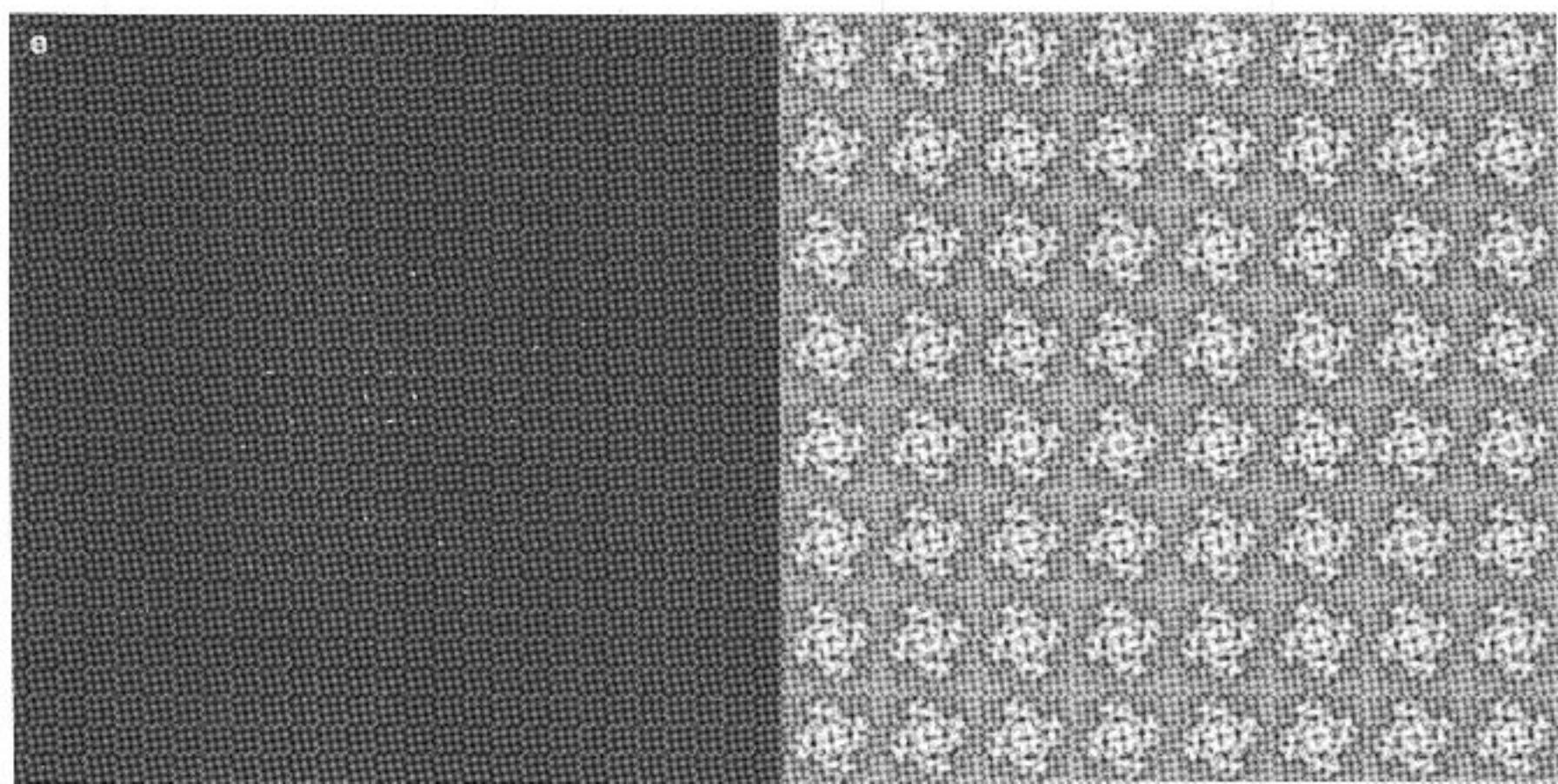
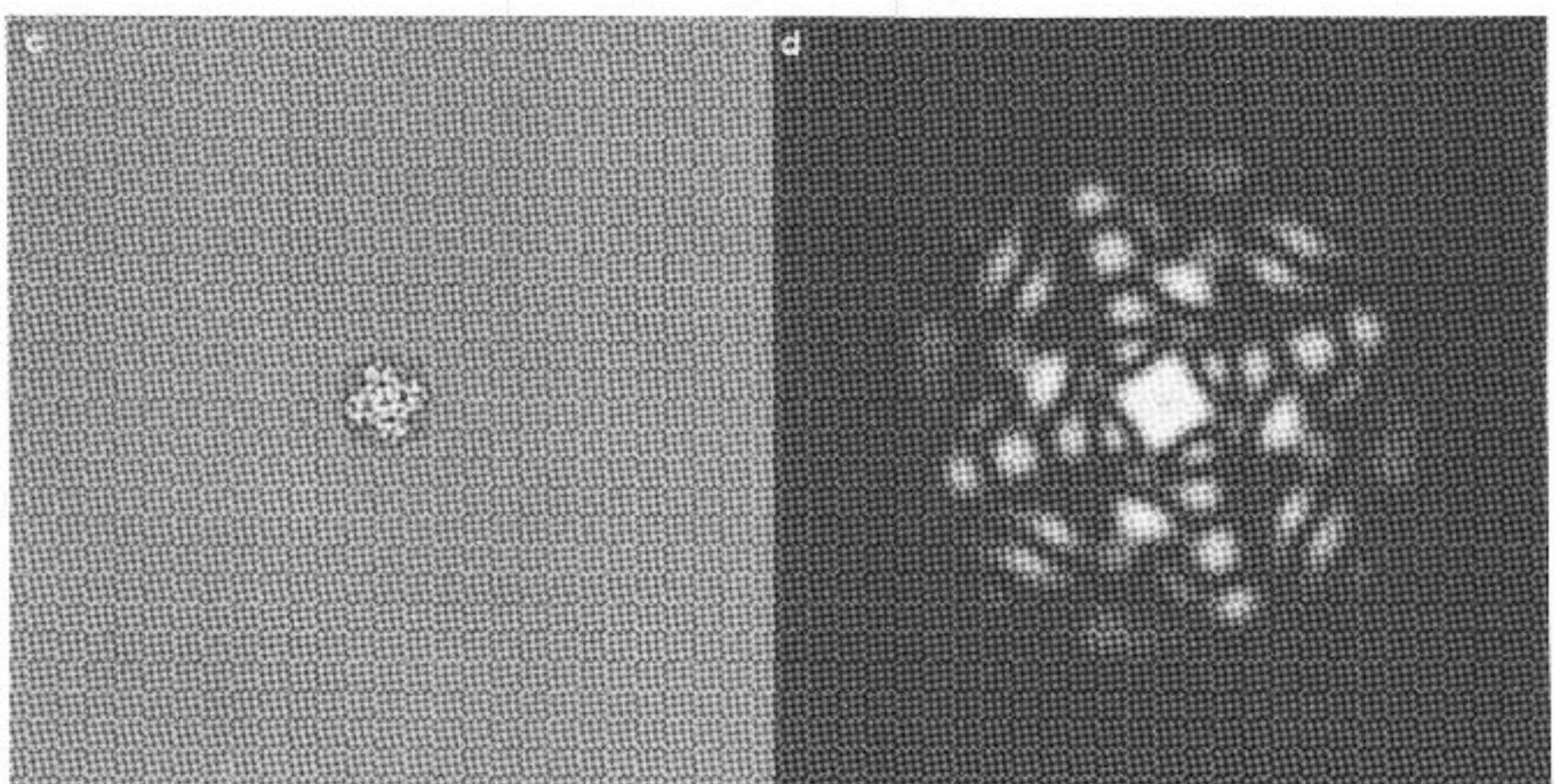
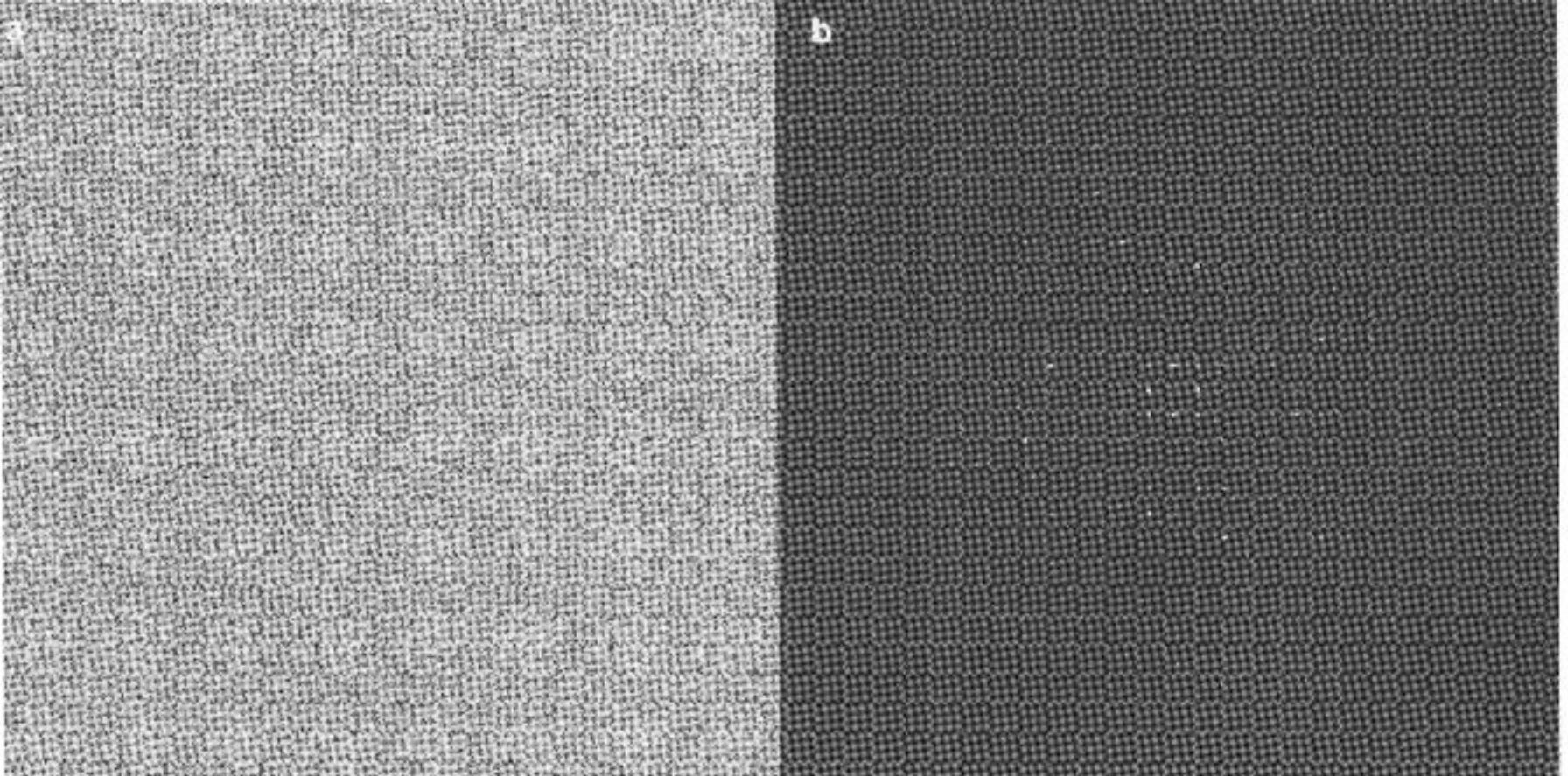


Image averaging in 2D crystal



In 2D crystal, one can extract amplitudes and phases from peaks of FT (contributed by the identical repeats of structural motif) and ignore everything in between peaks (contributed by the random noise). A reverse Fourier Transform using extracted amplitude and phases will give us an averaged features. This is equivalent to the averaging.

It is easy to perform averaging in 2D crystal. The molecules in the 2D crystal are identical in composition and orientation.

What about single molecules

A single particle image data set is a collection of images, each contains projection images of one molecules. The orientations and position of particles in all images are different. Before averaging, one needs to:

- judge how similar is the two particles: *cross-correlation coefficient*;
- shifts/rotates one particle to match another by maximizing ccc: *alignment*;
- separate different particles for averaging: *classification*;

Alignment \longleftrightarrow Classification

A digital image is collection of numbers in a grid

3	20	5	-3	4
3	5	34	45	4
0	-2	34	45	6
-1	34	2	3	1
4	5	2	2	0

$$f = \sum_{j=1}^J f(\vec{r}_j) = \sum_{j=1}^J f(m_j, n_j)$$

3	2	5	-3	4
25	2	4	2	4
0	34	45	5	6
-1	32	40	2	1
35	3	2	2	0

$$g = \sum_{j=1}^J g(\vec{r}_j) = \sum_{j=1}^J g(m_j, n_j)$$

Cross-correlation coefficient

Cross-correlation coefficient is a measure of similarity and statistical interdependence between two data sets. The mathematic definition of cross-correlation coefficient is:

$$\rho = \frac{\sum_{j=1}^J [f_1(\vec{r}_j) - \langle f_1 \rangle][f_2(\vec{r}_j) - \langle f_2 \rangle]}{\left\{ \sum_{j=1}^J [f_1(\vec{r}_j) - \langle f_1 \rangle]^2 \sum_{j=1}^J [f_2(\vec{r}_j) - \langle f_2 \rangle]^2 \right\}^{1/2}}$$

Where:

$$\langle f_i \rangle = \frac{1}{J} \sum_{j=1}^J f_i(\vec{r}_j)$$

Note that:

$$-1 < \rho < 1$$

Alignment between two images

Alignment is a process to search the grids to maximize the cross-correlation coefficient between two images. Three parameters are used to define alignment of 2D images: in-plane shift (x, y) and in-plane rotation angle.

Cross-correlation function based alignment:

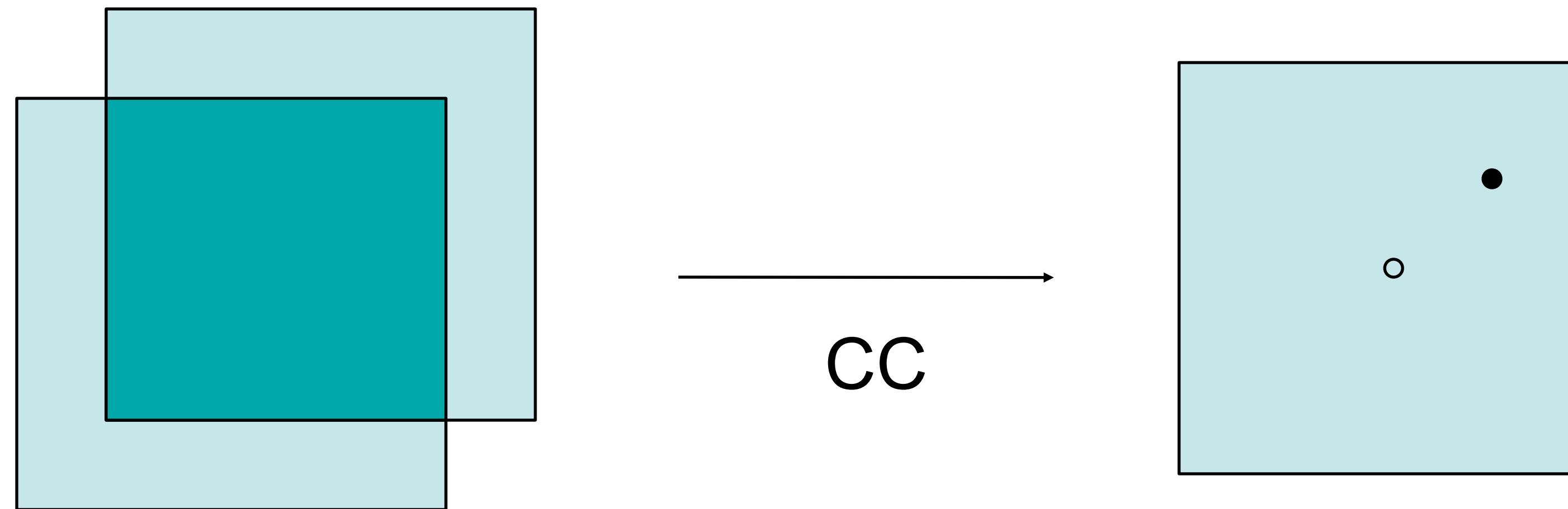
- In-plane shift can be determined by determine the peak position in the translational cross-correlation function between two images.
- Rotation can be determined by different ways: rotational cross-correlation function, Radon transform.

Cross-correlation function

The cross-correlation function is the most important tool for alignment of two images.

The mathematic definition of cross-correlation is:

$$f * g = \int_{-\infty}^{\infty} f(t)g(t - \tau)d\tau$$



Q: what happens if shift is more than half of the image size?

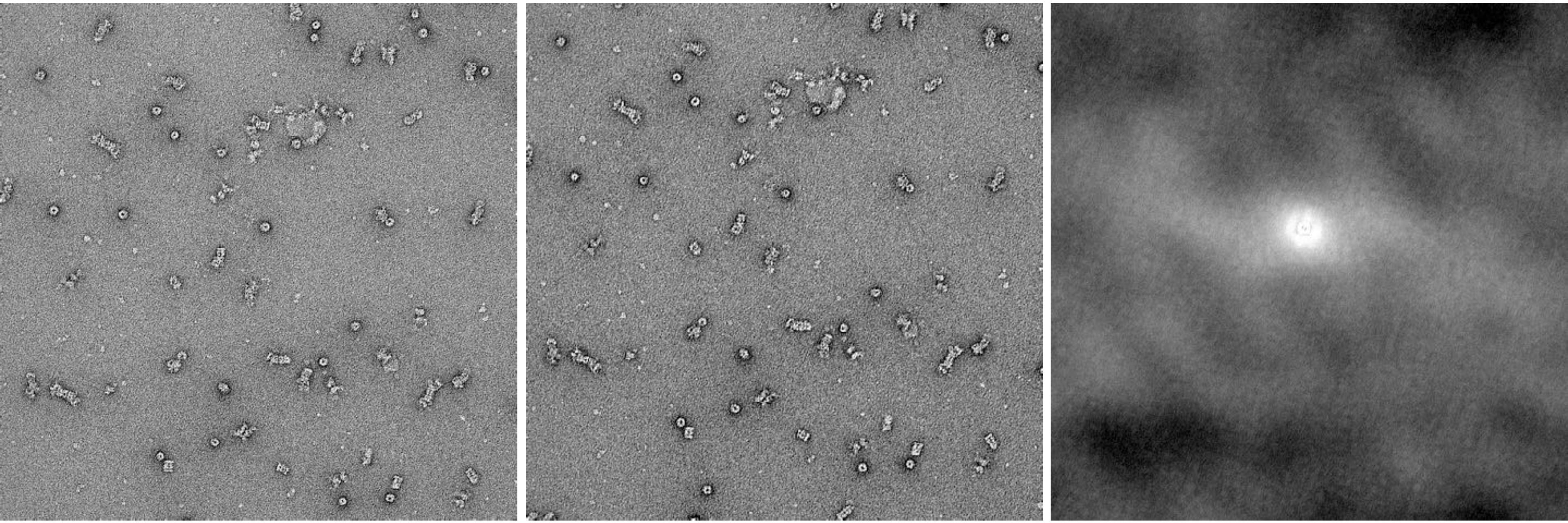
Calculating the cross-correlation

Cross-correlation theorem:

$$f * g = \int_{-\infty}^{\infty} f(t)g(t - \tau)d\tau = F\{F(f) \cdot F^{-1}(g)\}$$

This formula enable us to calculate the cross-correlation between two images easily.

How cross-correlation looks like

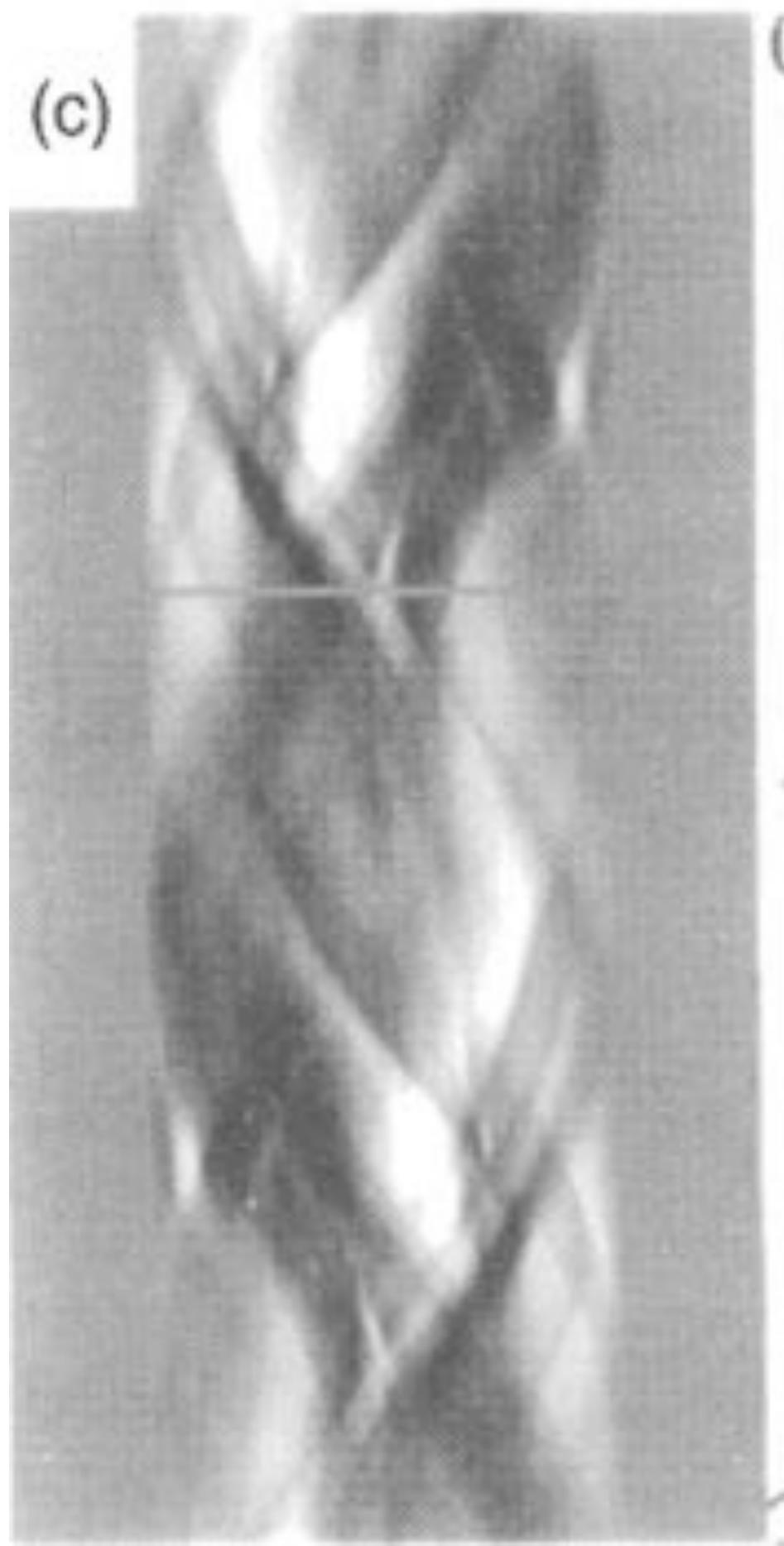
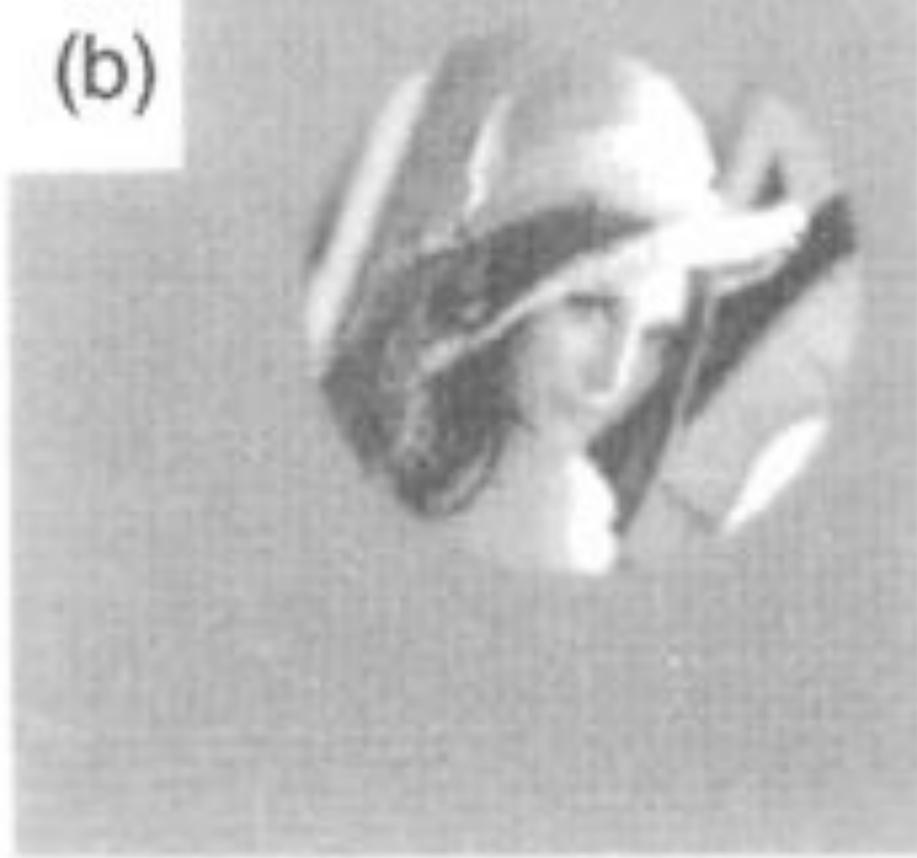


-1 μm

-1.5 μm

CCF

The image size is 1024X1024. The peak in the CCF is at (445,500). How much is the shift?



Radon transform

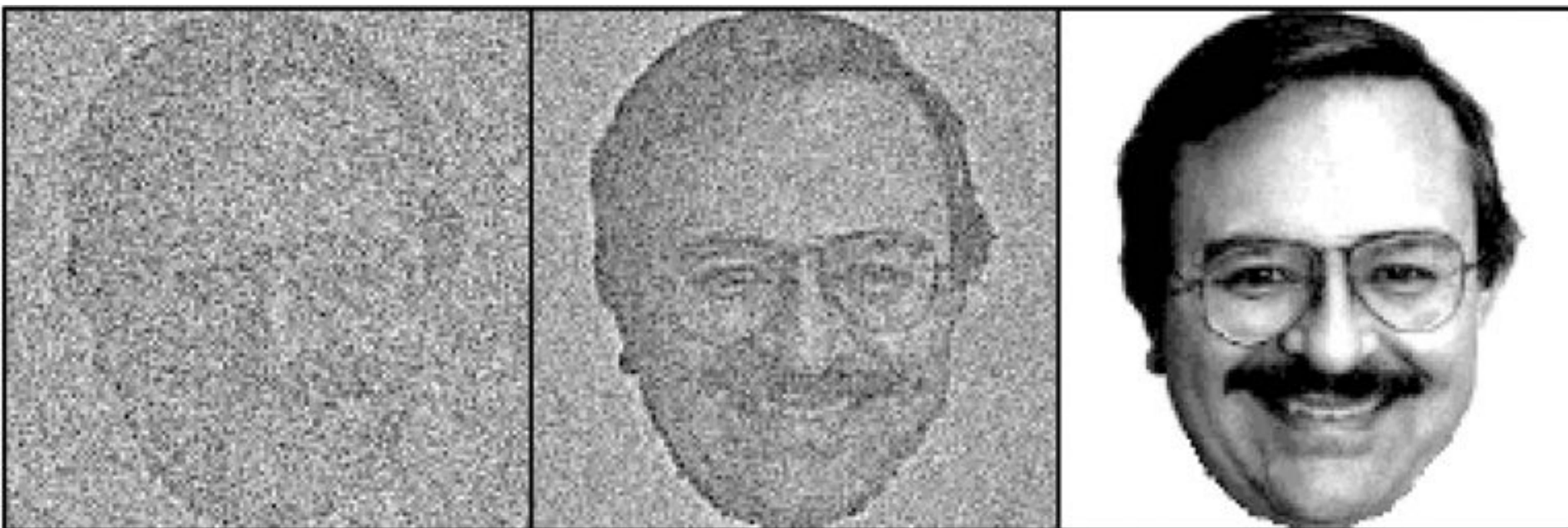
Radon transform is an efficient way for determining angular relationship between two images, but it only works well in images with high SNR.

More about the cross-correlation function

- Peak searching in the cross-correlation function;
search for a peak is not just finding the point of highest value in the CCF.
- Keep in mind that one can calculate cross correlation between any two images, and will always find a point with highest value.
- Cross-correlation based alignment and averaging always enhance the features of the reference image.

Demonstration of reference induced bias

Note: The averaged image after reference based alignment is strongly biased towards the reference.



100 images

1000 images

reference

Multi-reference alignment

For a heterogeneous data set, multiple references are used. Each images are aligned again each references, and decide which one yields highest cross-correlation coefficient.

Classification

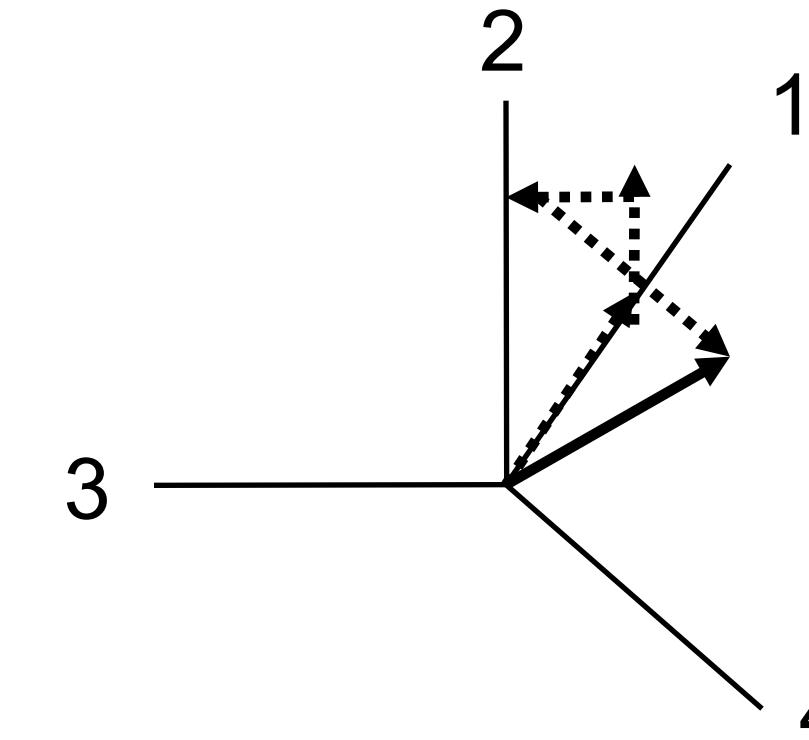
Classification - a process of dividing a set of images into subsets with similar features.

One can perform classification based on CCC to determine if the images are similar with each other; But for a very large data set of very noisy images (> 50,000 images)?

Hyperspace

An image of $m \times m$ pixels can be represented by a vector (or end point of a vector) in the hyperspace of $m \times m$ dimensions.

3	2
2	3



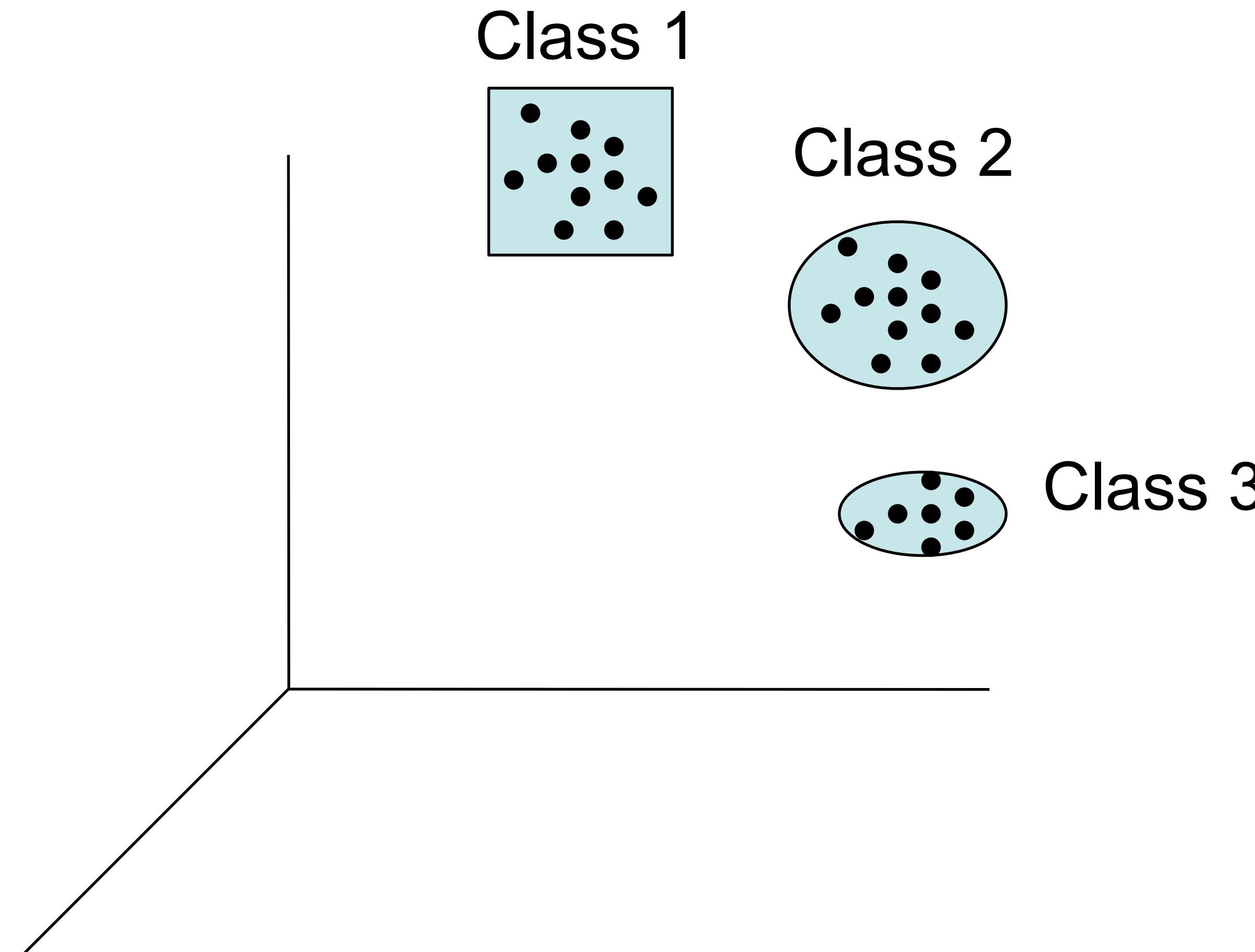
$$f = (f_1, f_2, \dots, f_m) = \sum_{i=1}^m f_i \vec{a}_i \quad \text{Where:} \quad |\vec{a}_i| = 1; \\ \vec{a}_i \perp \vec{a}_j \quad (j \neq i; j = 1, \dots, m);$$

Similar to the cross-correlation coefficient, the distance between two spots in the hyperspace represents the difference between two images.

A data set is represented as a cloud in the hyperspace. The center of the cloud is the average of the all images in the data set.

Classification

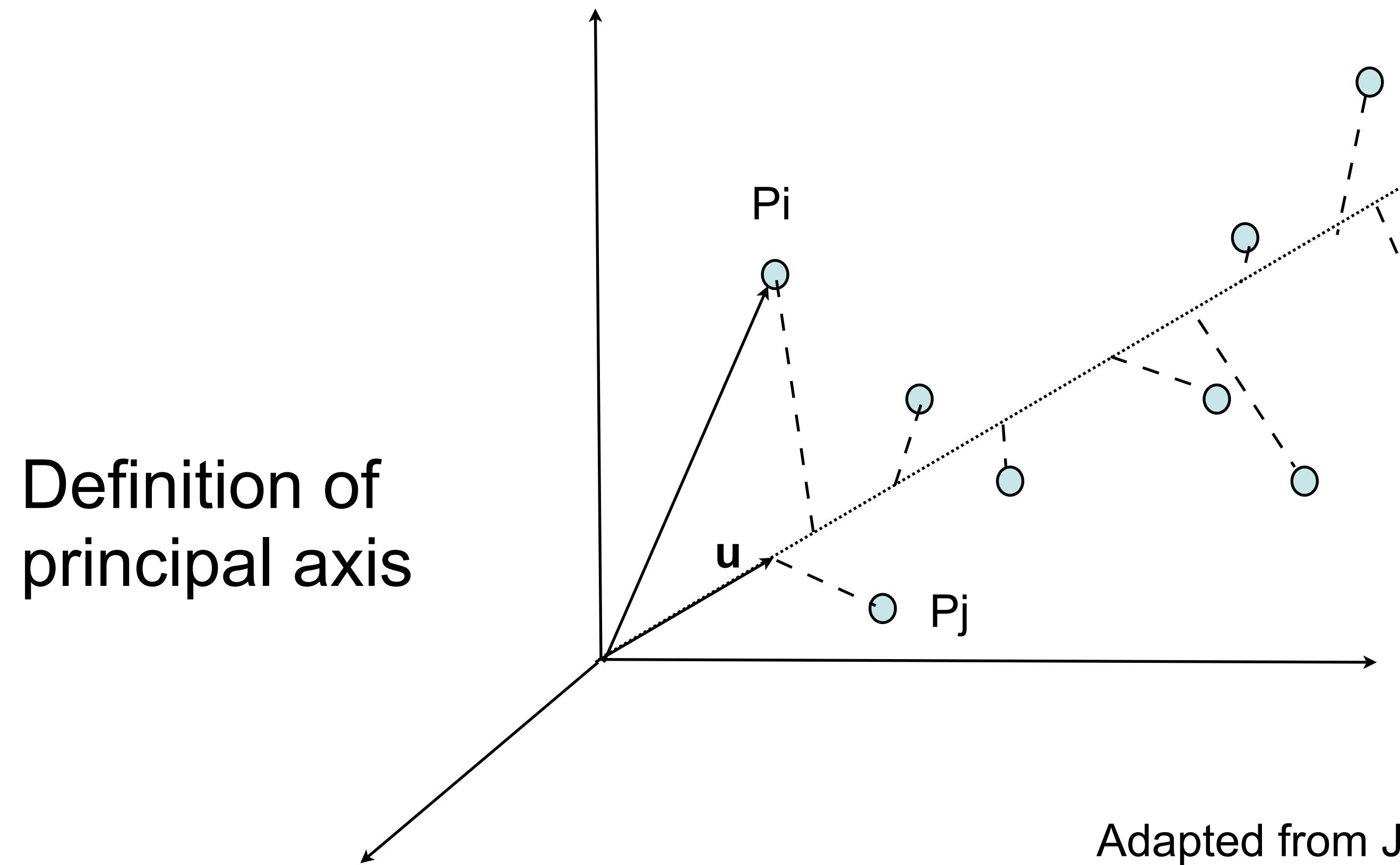
Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



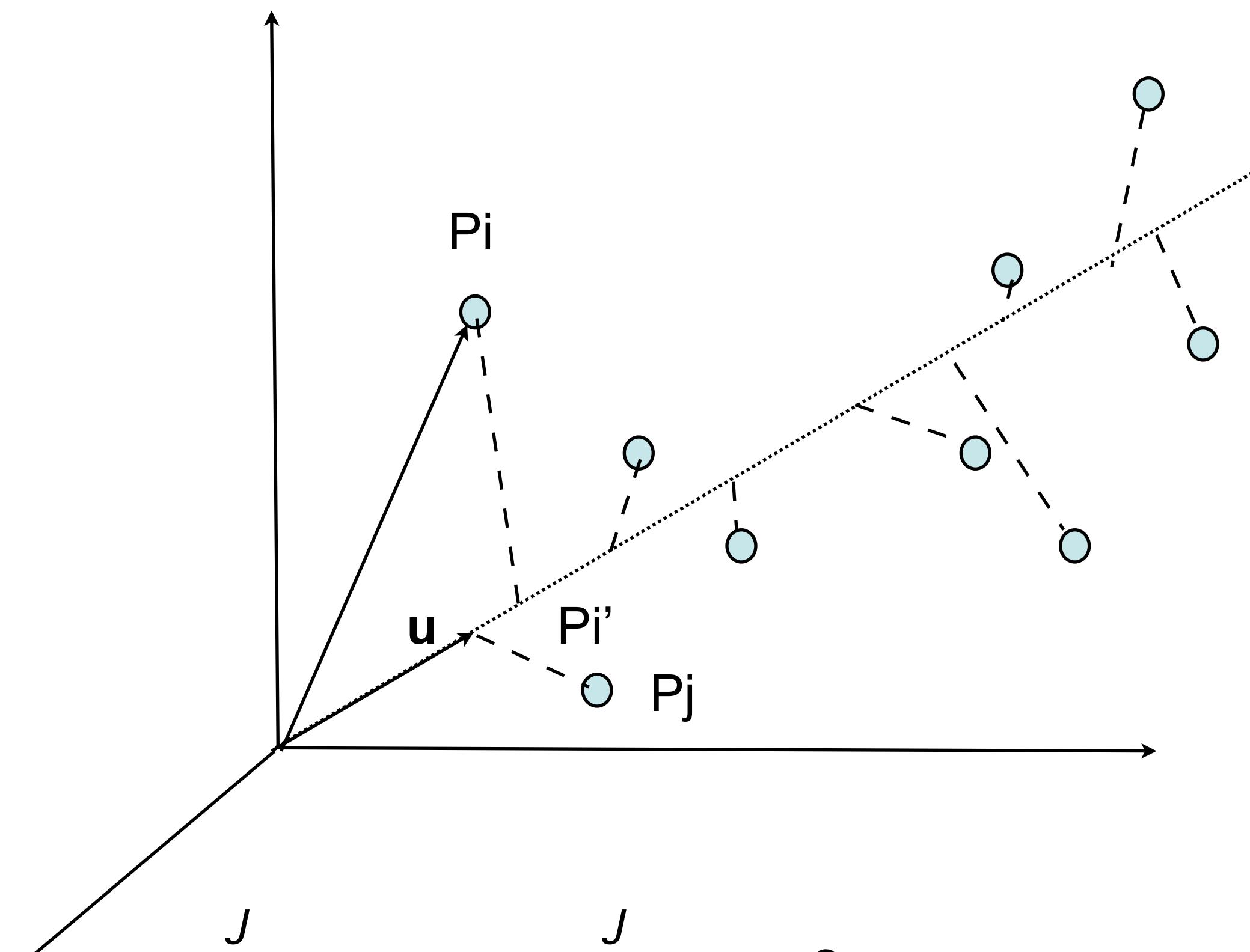
Multivariate statistical analysis

Making patterns emerge from data

Multivariate statistical analysis:
Principal Component Analysis
Correspondence Analysis



Principal component analysis (PCA)



$$\sum_{j=1}^J (OP'_i)^2 = \sum_{j=1}^J (x_i u)^2 = (Xu)' Xu = u' X' Xu \longrightarrow \max$$

with constraint: $u'u = 1$

X: coordinate matrix

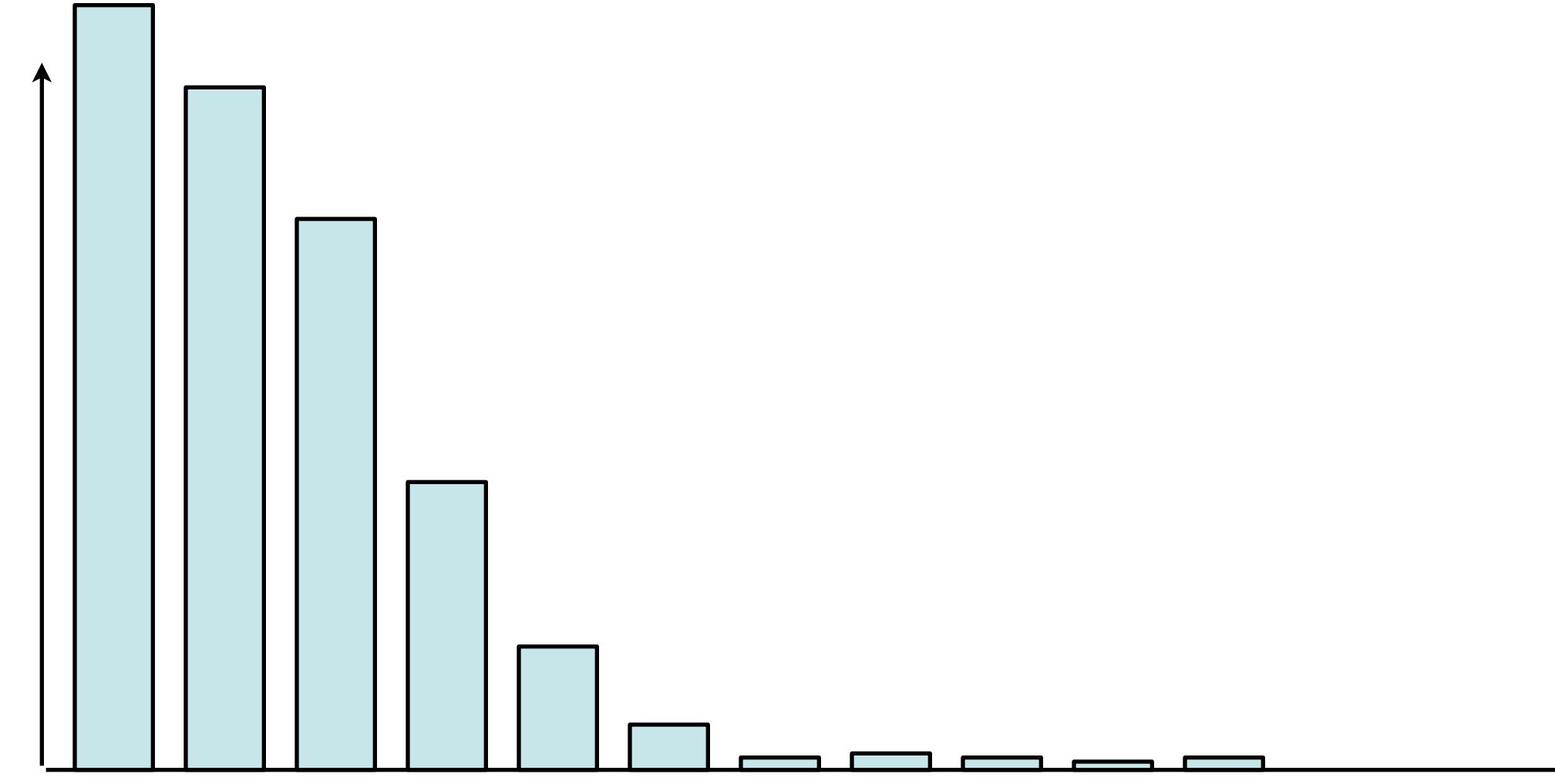
Eigenvector-eigenvalue equation

$$\mathbf{D}\mathbf{u} = \lambda\mathbf{u}$$

where $D = (X - \bar{X})'(X - \bar{X})$

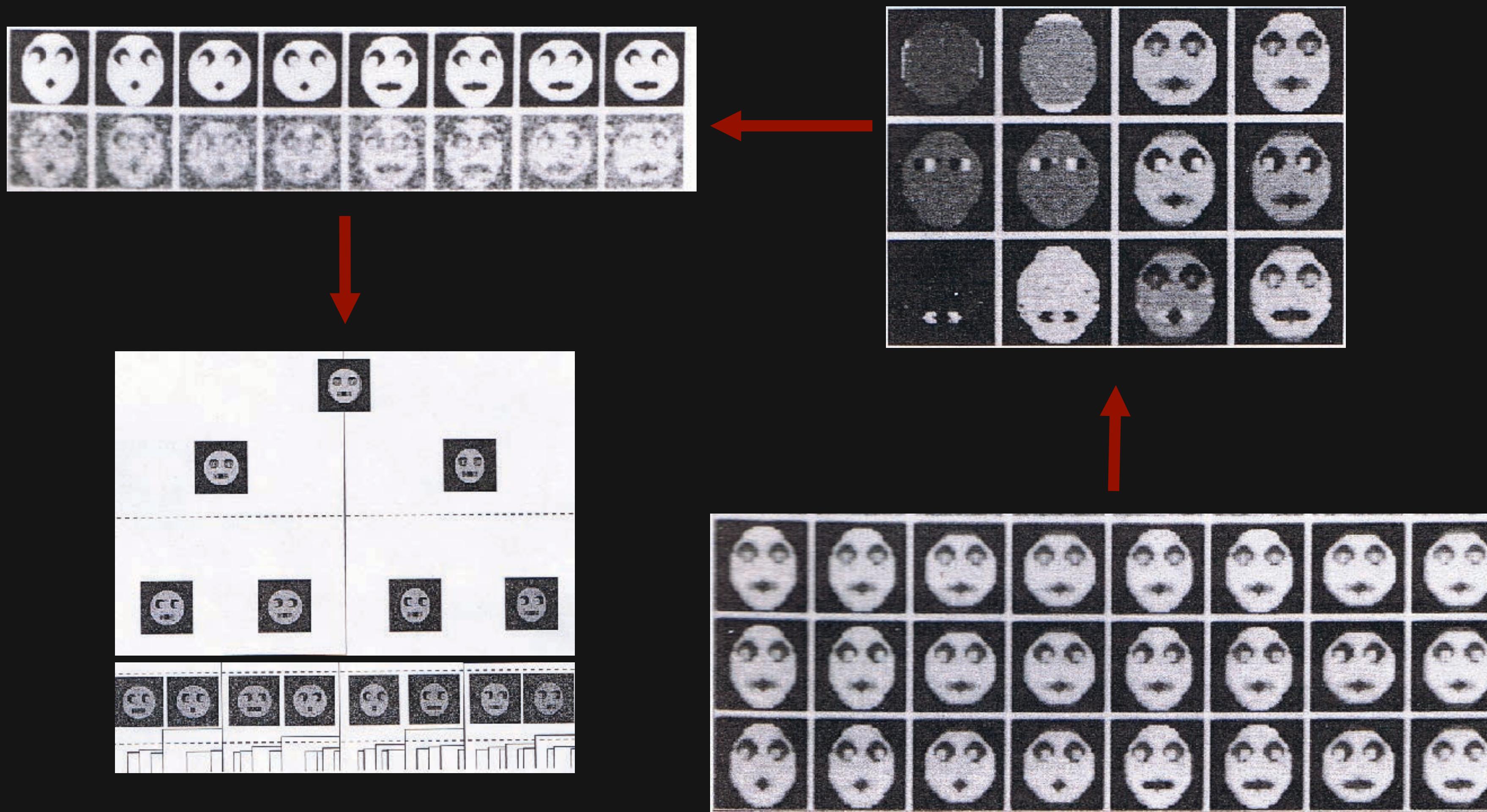
Solution of this equation generate a set of eigenvectors and eigenvalues.

Significant factors:



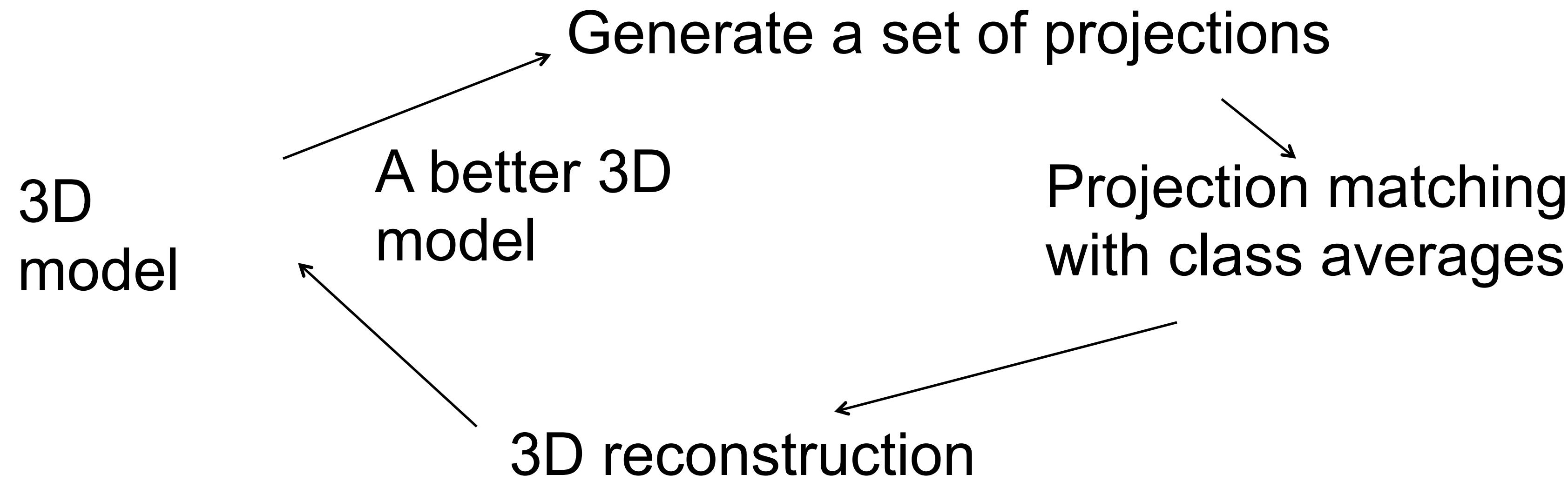
Classification based on eigenvector/eigenvalue clustering;

Principle Component Analysis



Iterative refinement procedure

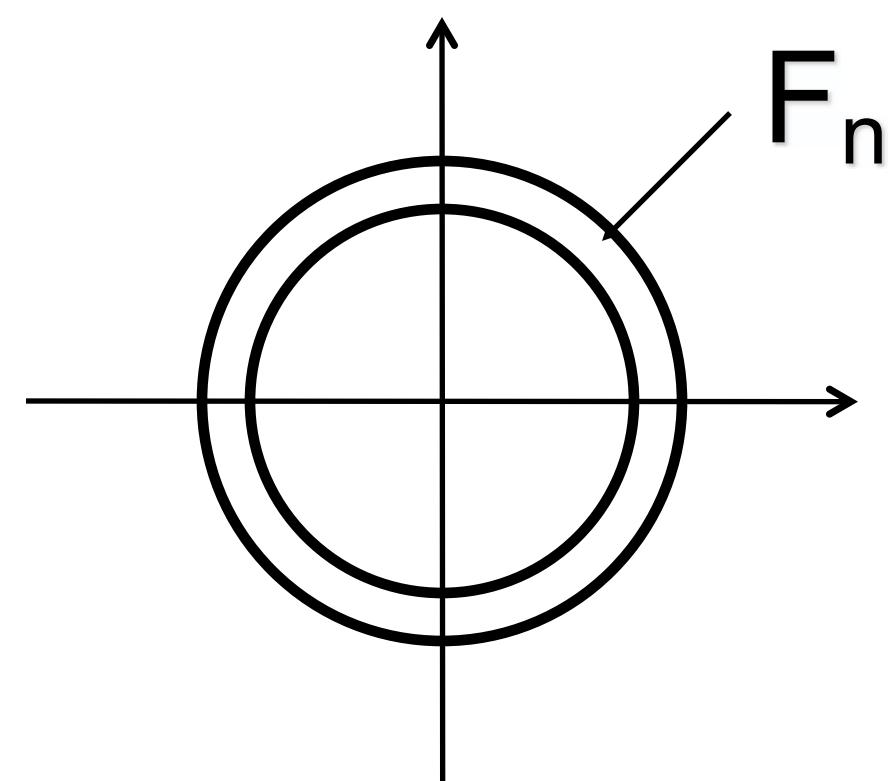
Iterative refinement procedure, using reference model based projection matching:



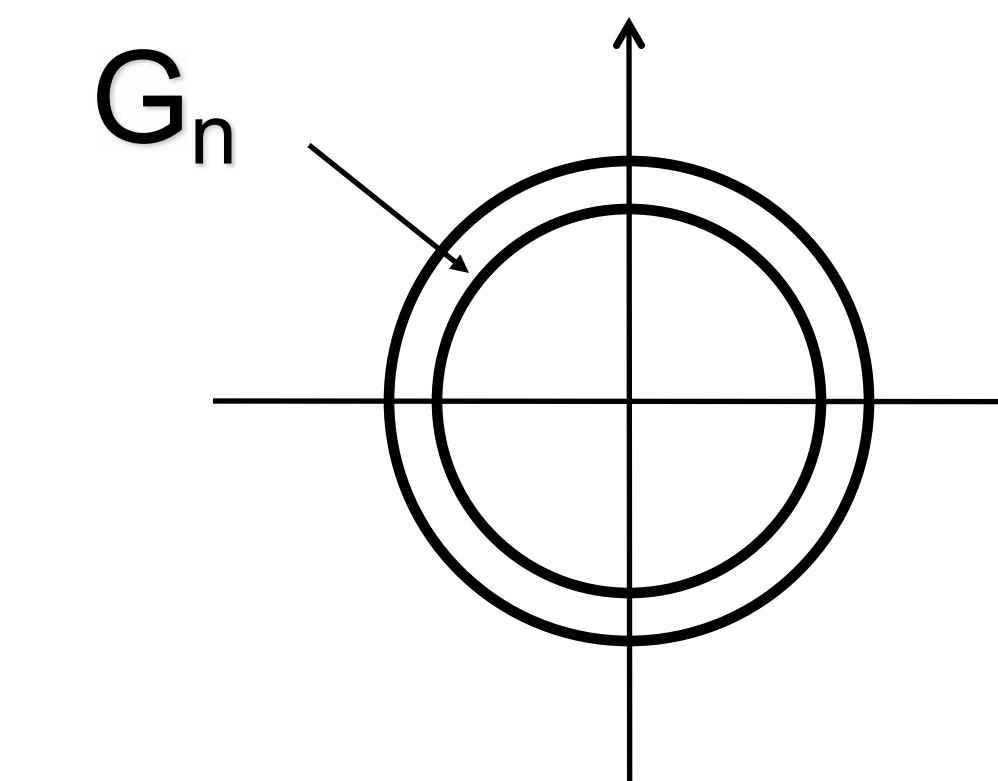
A least square approach to find the best solution that matches all data.

Resolution estimation

In single particle cryoEM the resolution is often estimated by Fourier Shell Correlation.

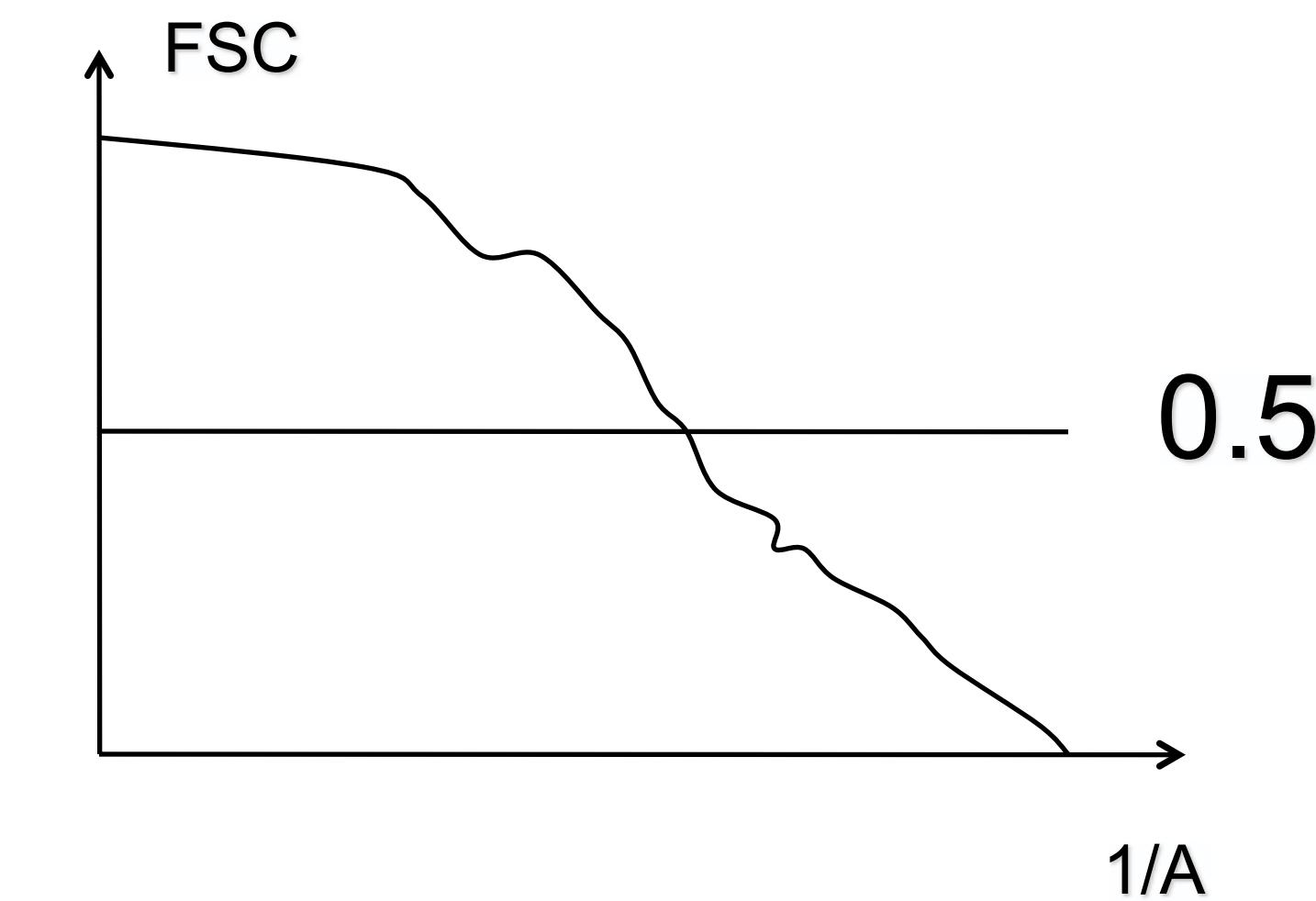


Reconstruction 1

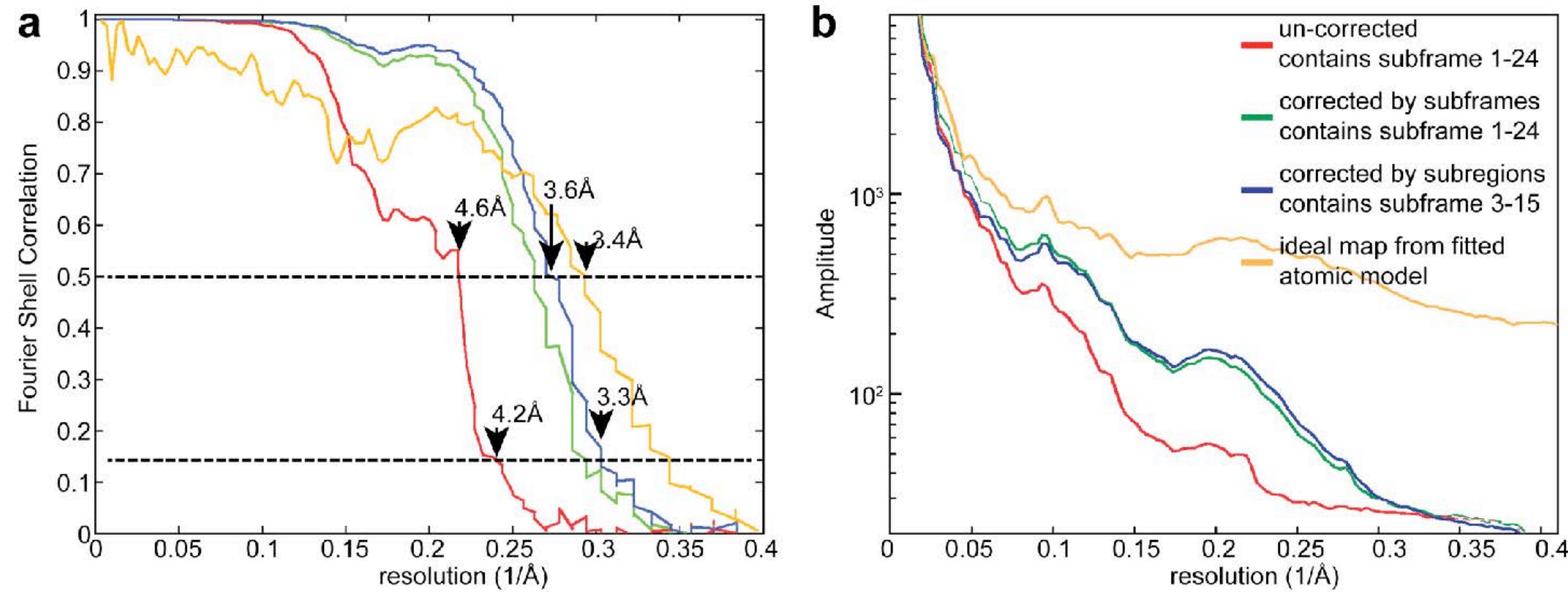


Reconstruction 2

$$FSC(R) = \frac{\sum_{n \in R} F_n G_n}{\left\{ \sum_{n \in R} |F_n|^2 \sum_{n \in R} |G_n|^2 \right\}^{1/2}}$$



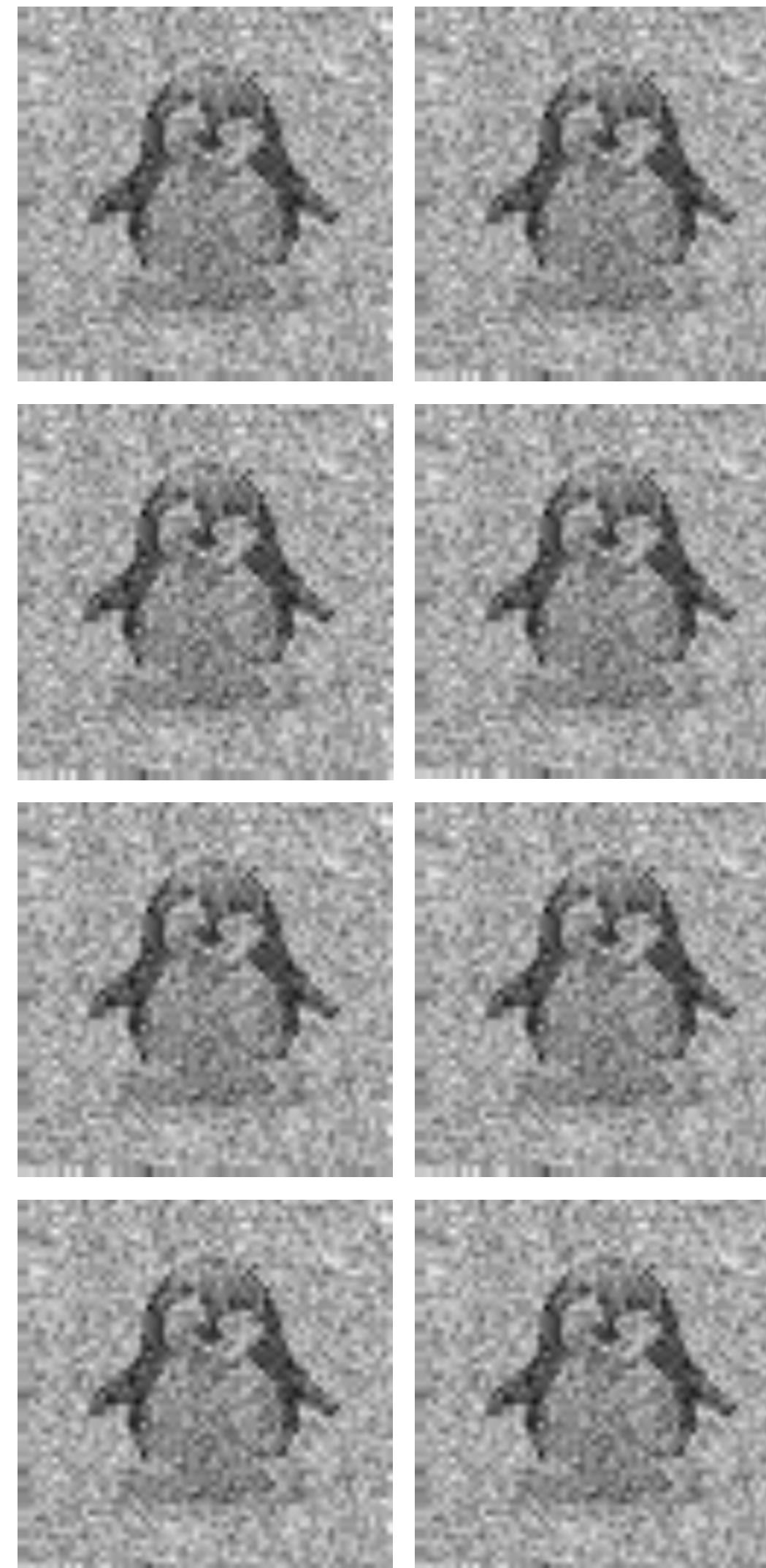
3D reconstruction of T20S proteasome



Fourier Shell Correlation curves and amplitude plot of T20S proteasome reconstruction

Image averaging

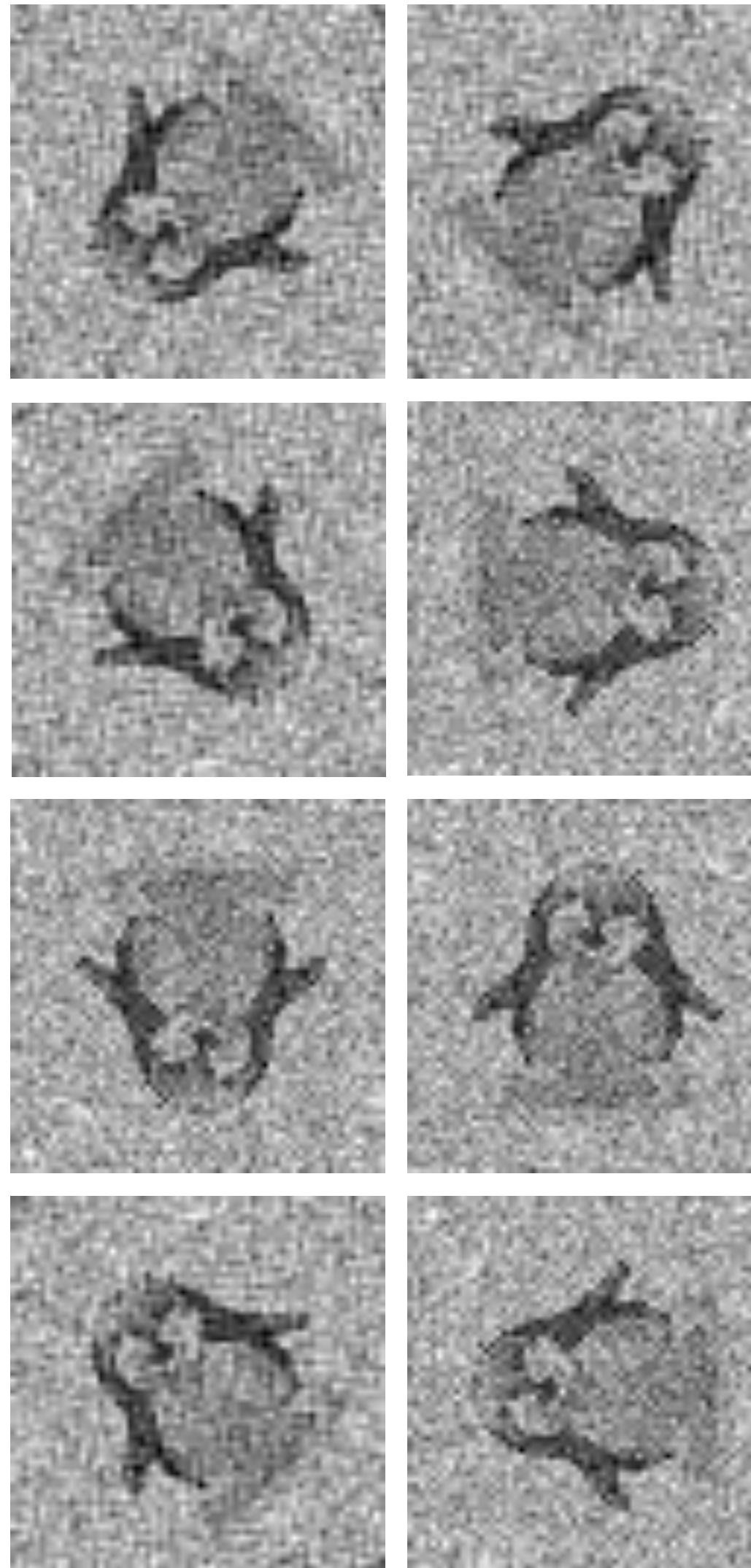
Averaging of a large number of identical images improves the SNR. A complete problem is simple to solve.



$$Average = \frac{1}{N} \sum_{i=1}^N X^i$$

Observed data (X): images

But we have an incomplete data set



Observed data (X): images

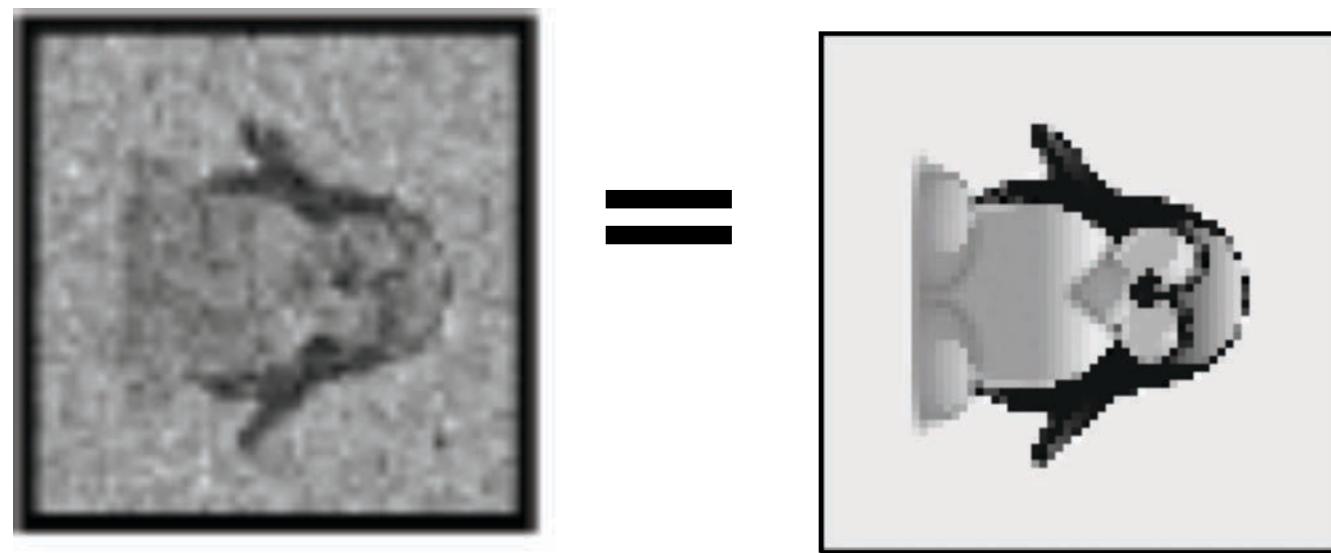
Missing data (Y):

Rotations, translations, classes & conformations

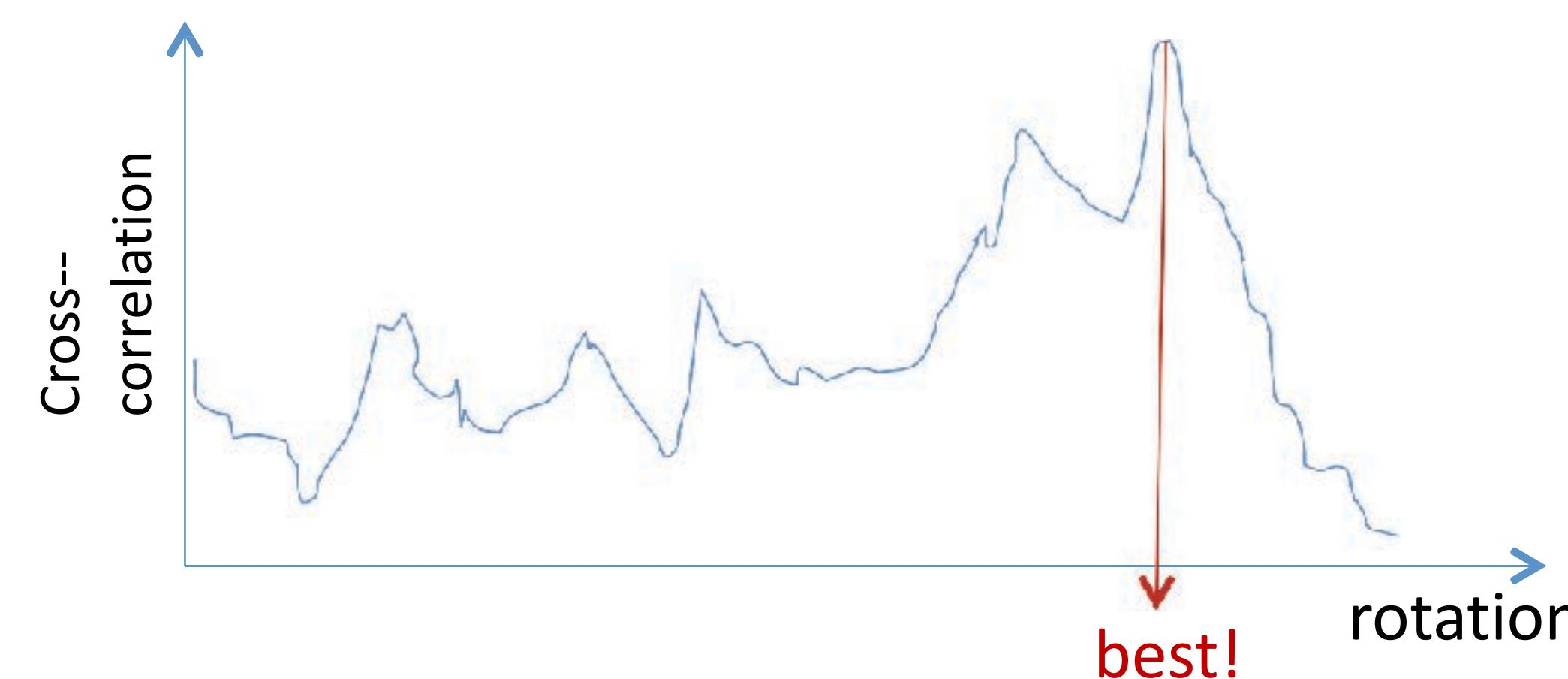
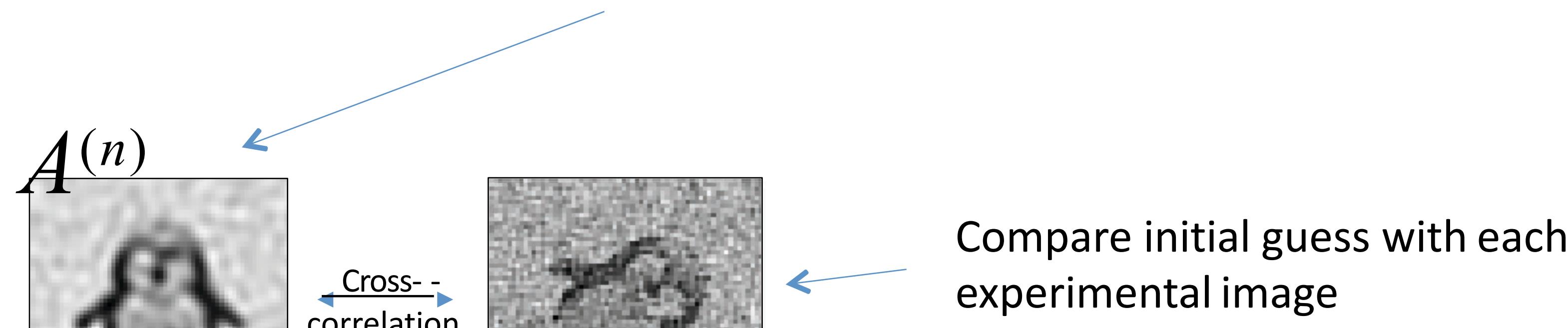
How do we find Y ?

$$X_i = P_\varphi V_k$$

X_i (Observed Projection) = P_φ (Rotations, etc) V_k (Actual Object)

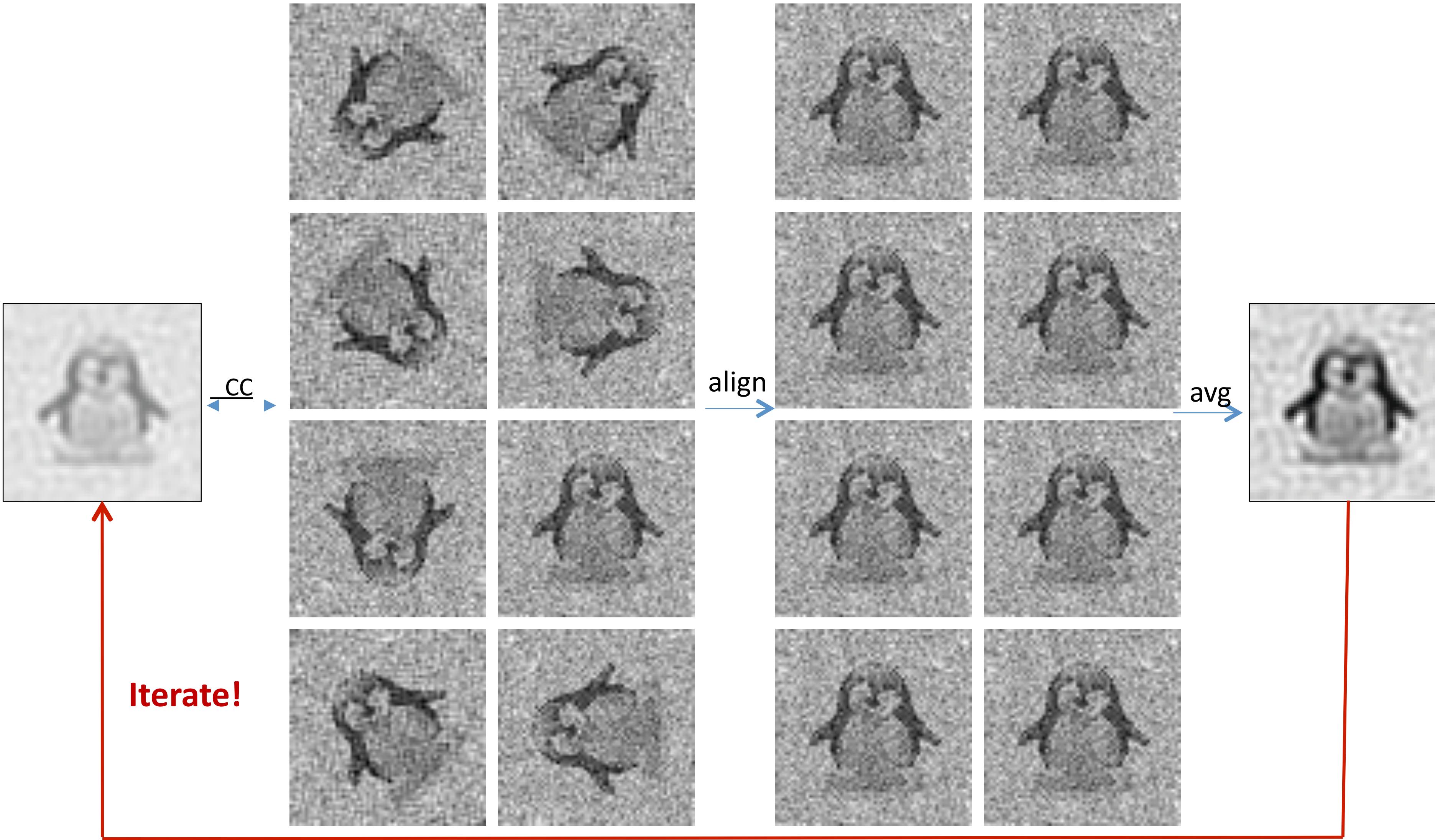


- Starts from some **initial guess** (source of model bias) about the structure



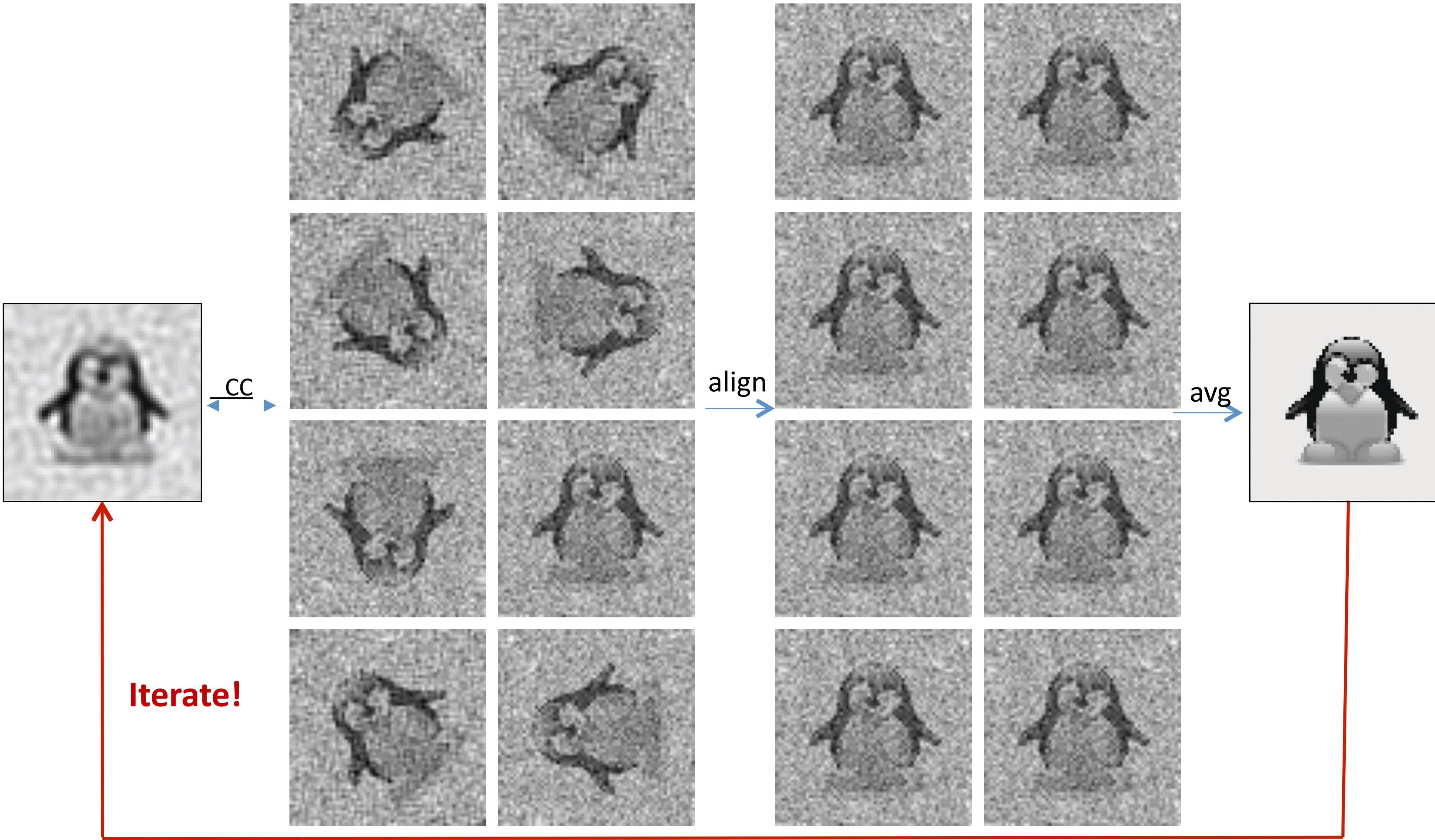
Iteratively align and average

How big is the search space?



Iteratively align and average

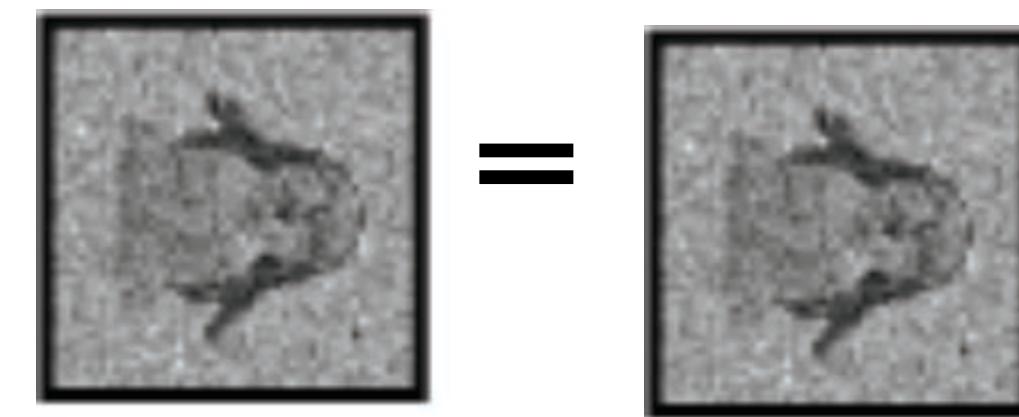
How big is the search space?



1) Uses model projections that include noise

$$X_i = P_\varphi V_k + N_i$$

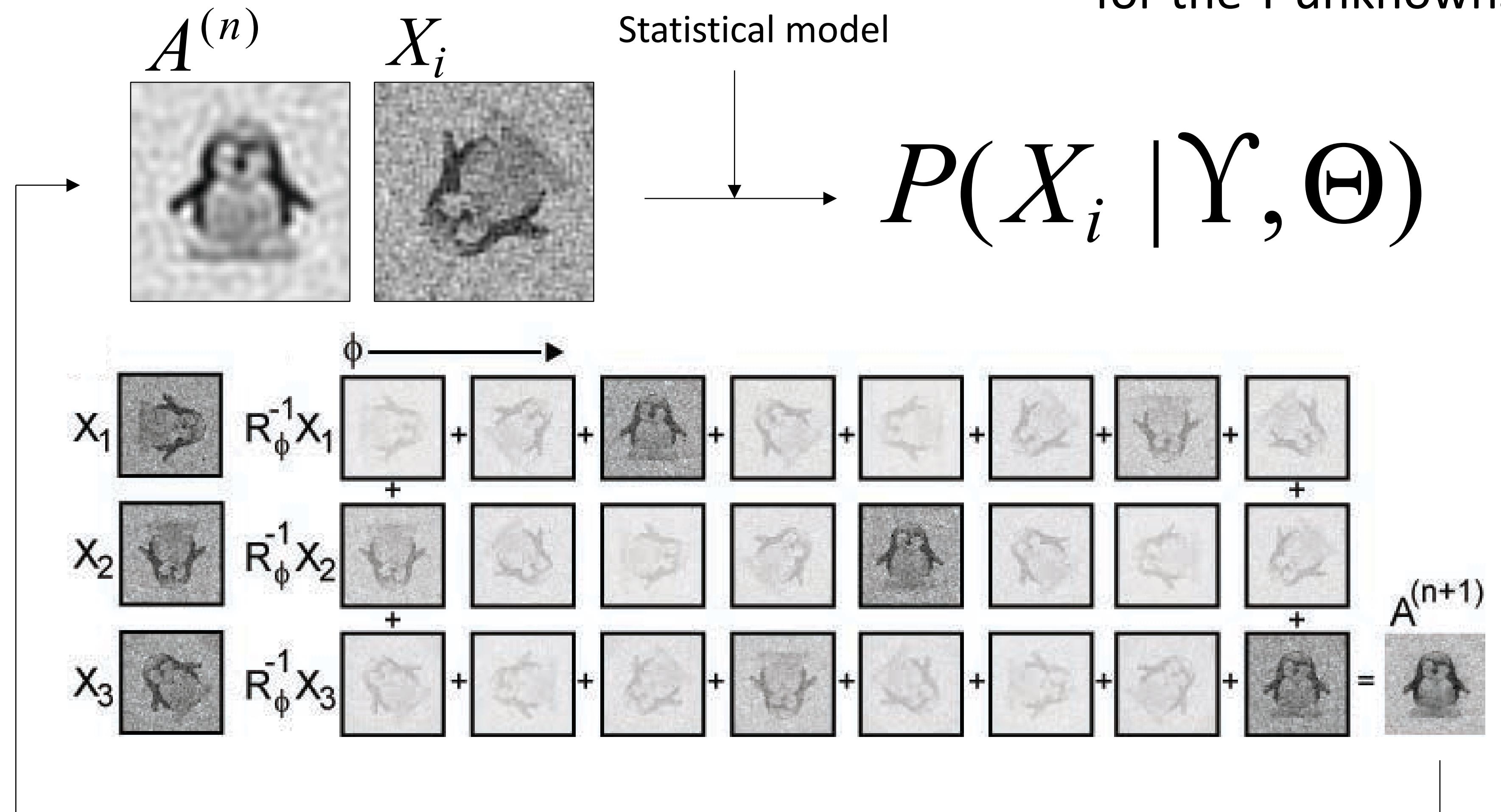
X_i (Observed Projection) = P_φ (Rotations, etc) V_k (Actual Object) + Noise



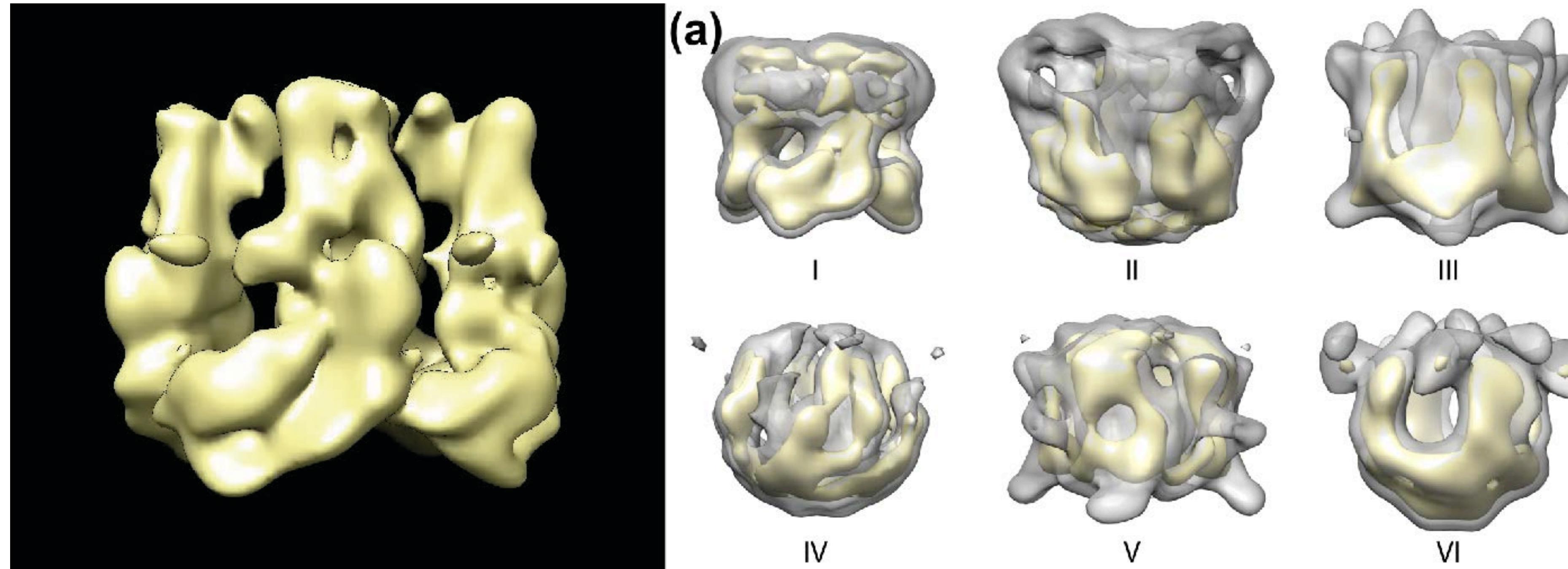
2) Maximizes likelihood with “marginalization” over Y

Need more? See *Methods in Enzymology*, 482 (2010)

Does not assign discrete “best” values for the Y unknowns !



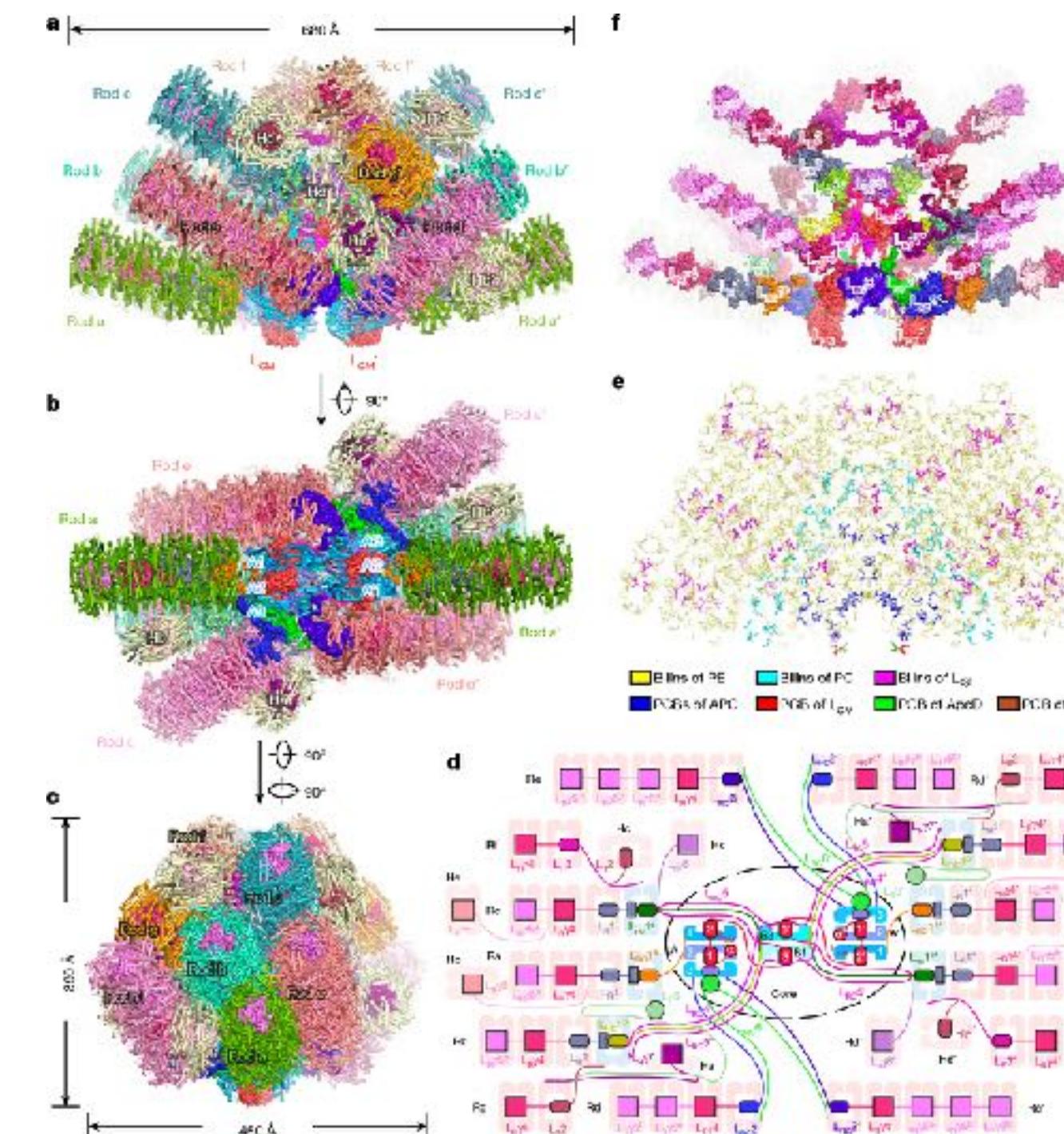
Classification of heterogeneous dataset



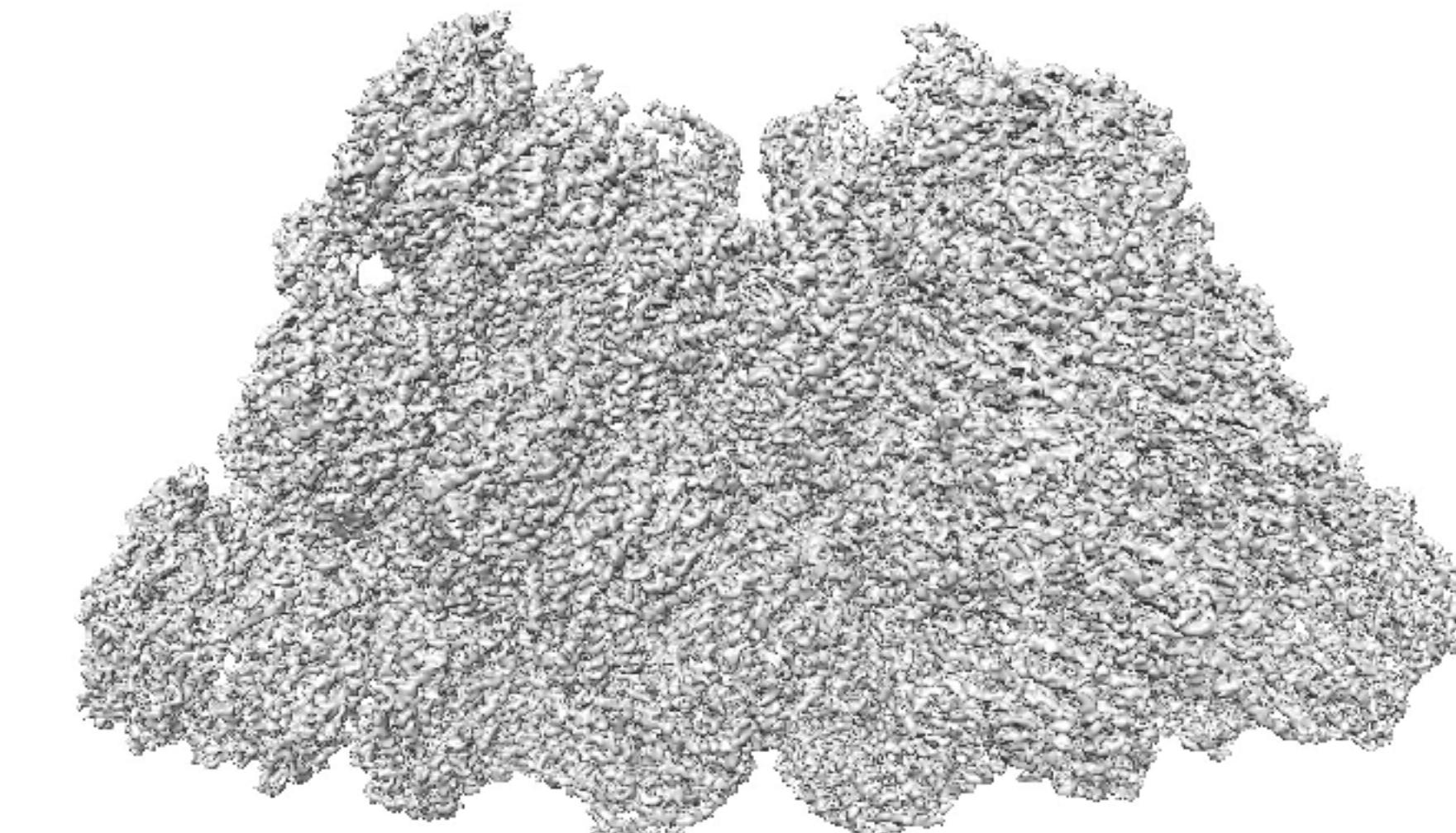
What have we accomplished since 2014

Biological applications

- Routine application to many types biological macromolecules;
- Large and dynamic assemblies: spliceosome, etc.;
- Integral membrane proteins: ion channels, transporters, etc.;
- Very difficult samples: tau filament, etc.
- Very small proteins, GPCR, or very large protein complex, picobilisome;



Picobilisome: 16.8 MDa, among the largest protein complex in the living world.



Zhang et al. (2017) Structure of phycobilisome from the red alga *Griffithsia pacifica*, *Nature*

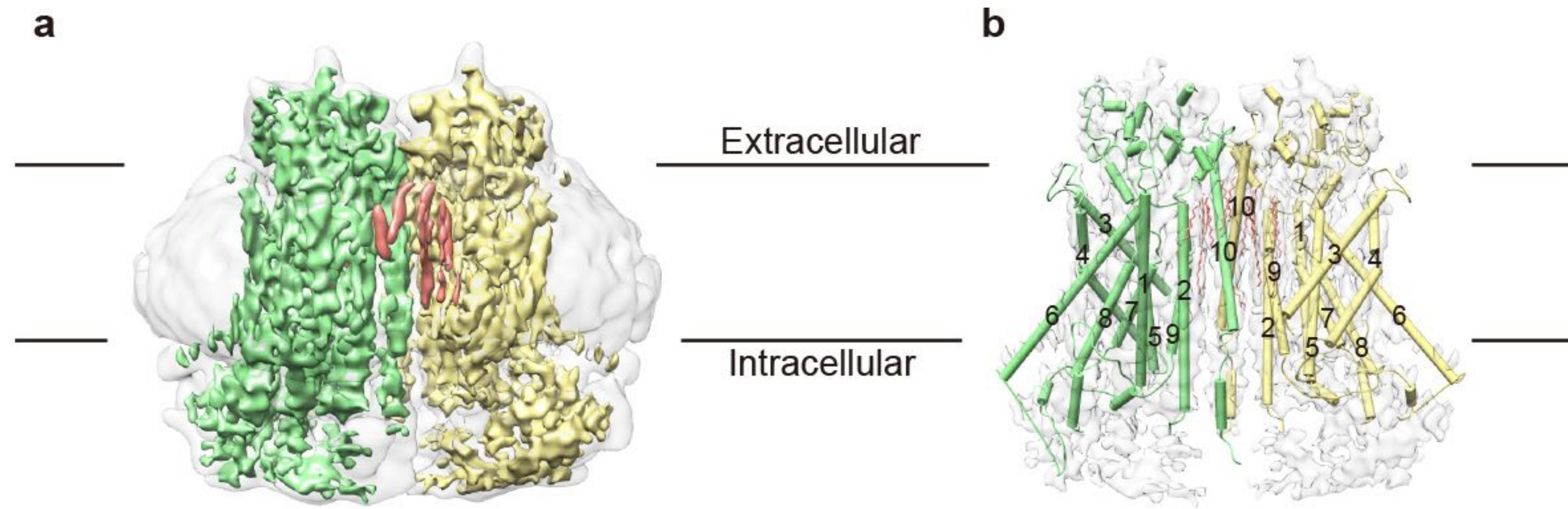
What are the common limiting factors?

From our own experience:

- Cryo-EM sample preparation is still a major obstacle;
 - ➔ Sample falling apart during plunge freezing is the most common issues;
 - possible solution: cross-linking or GraFix, Spotiton,
 - ➔ Sample preferred orientation

Example: TMEM16A channel

TMEM16A: Ca^{2+} activated Cl^{2-} channel

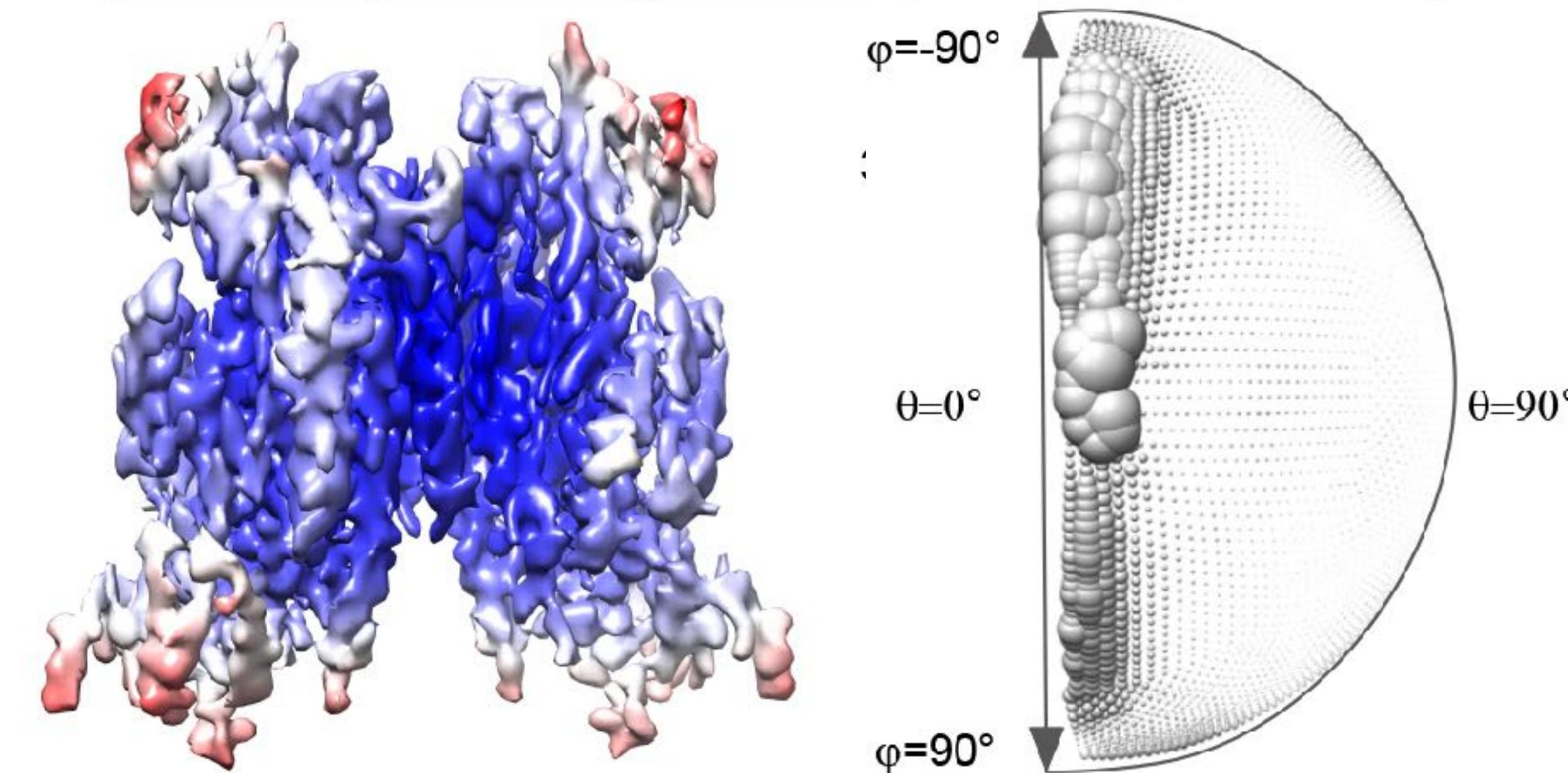
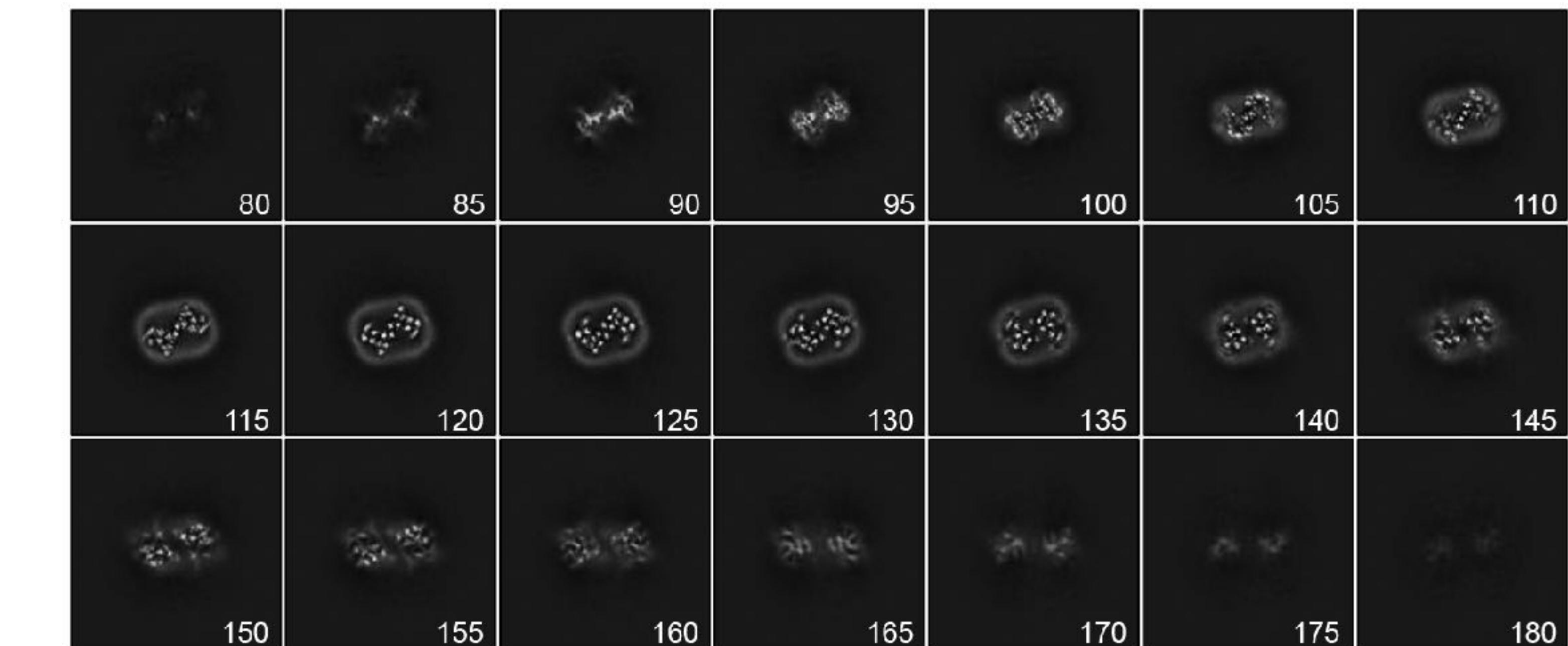
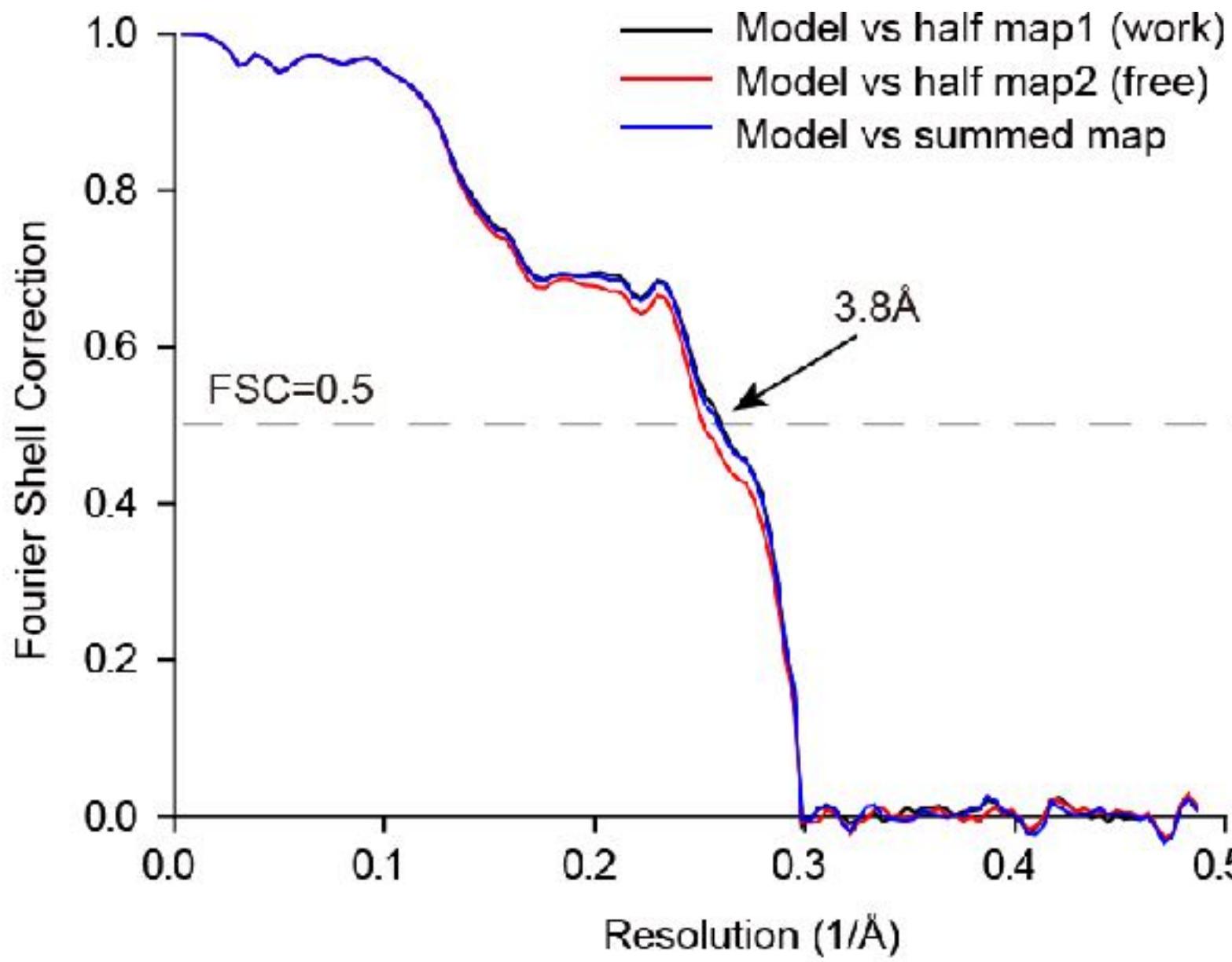
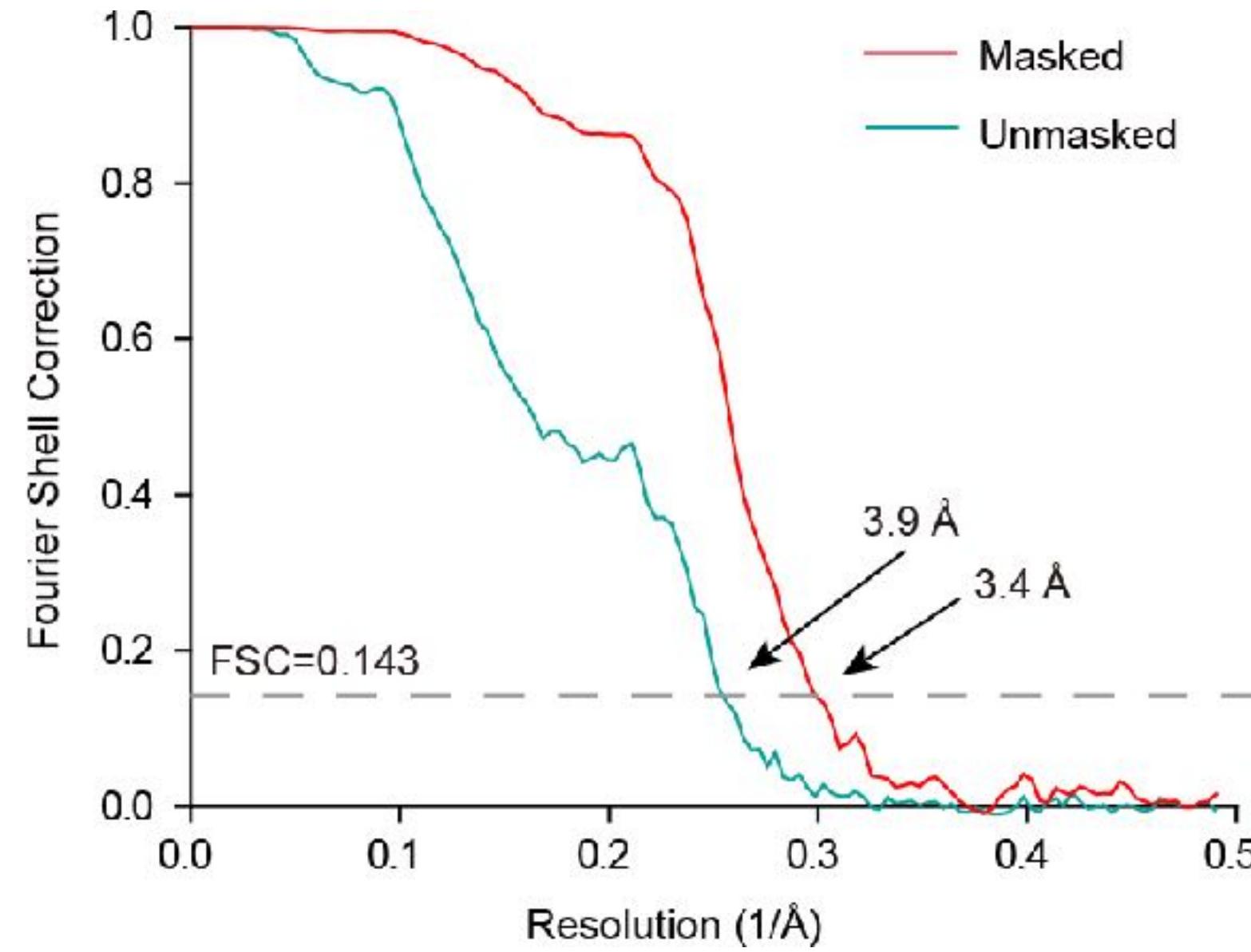


Shangyu Dang

Shengjie Feng

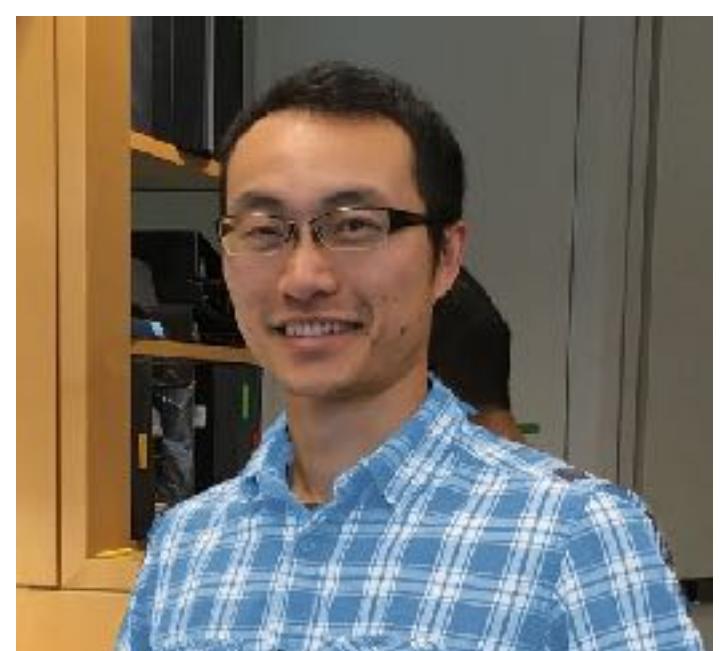
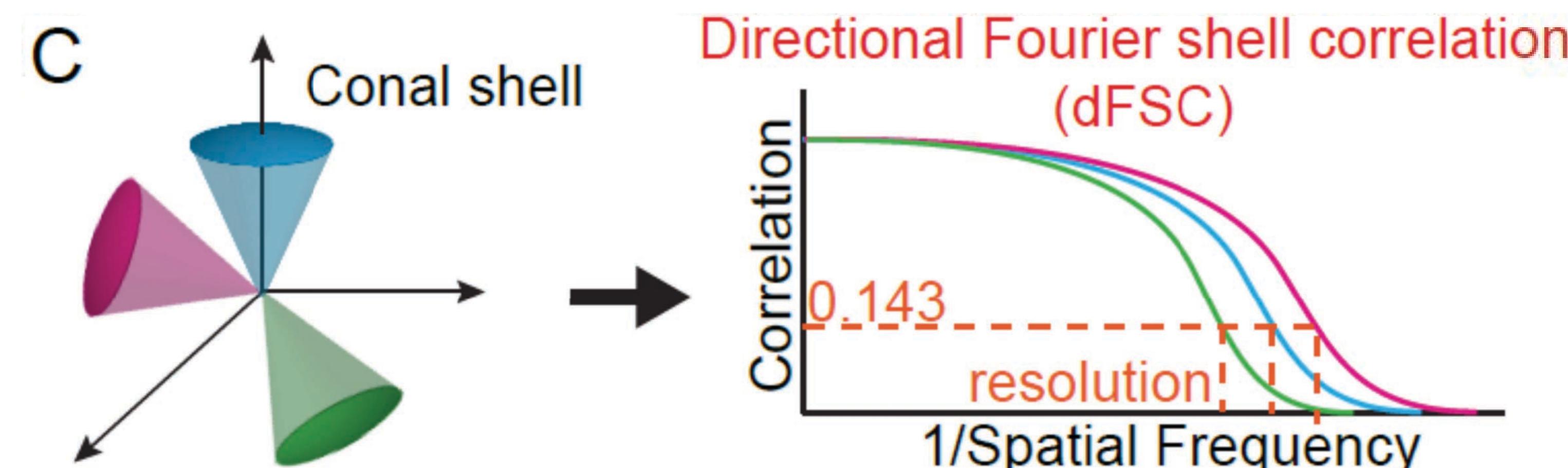
Unpublished

Structure of TMEM16A



Unpublished

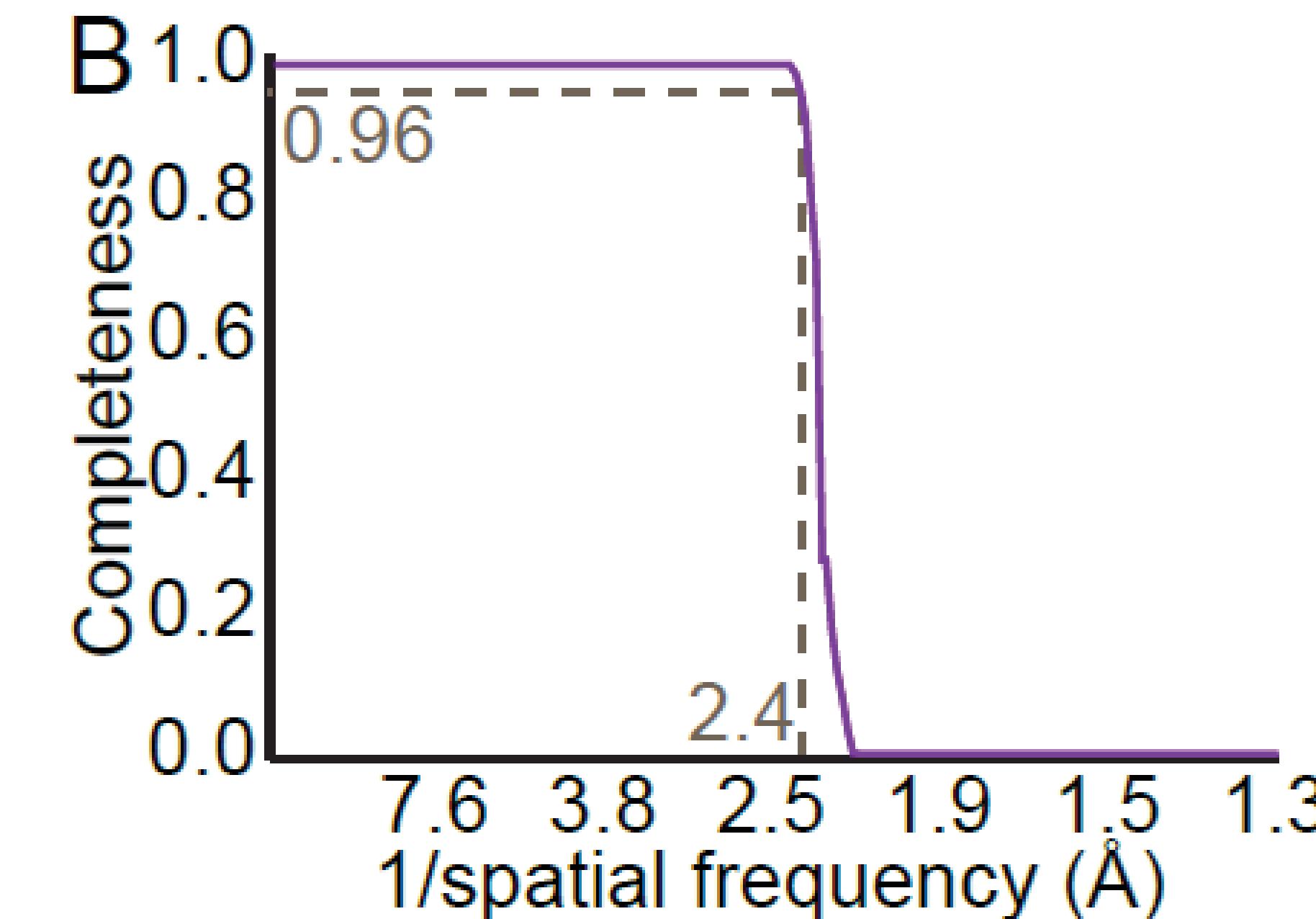
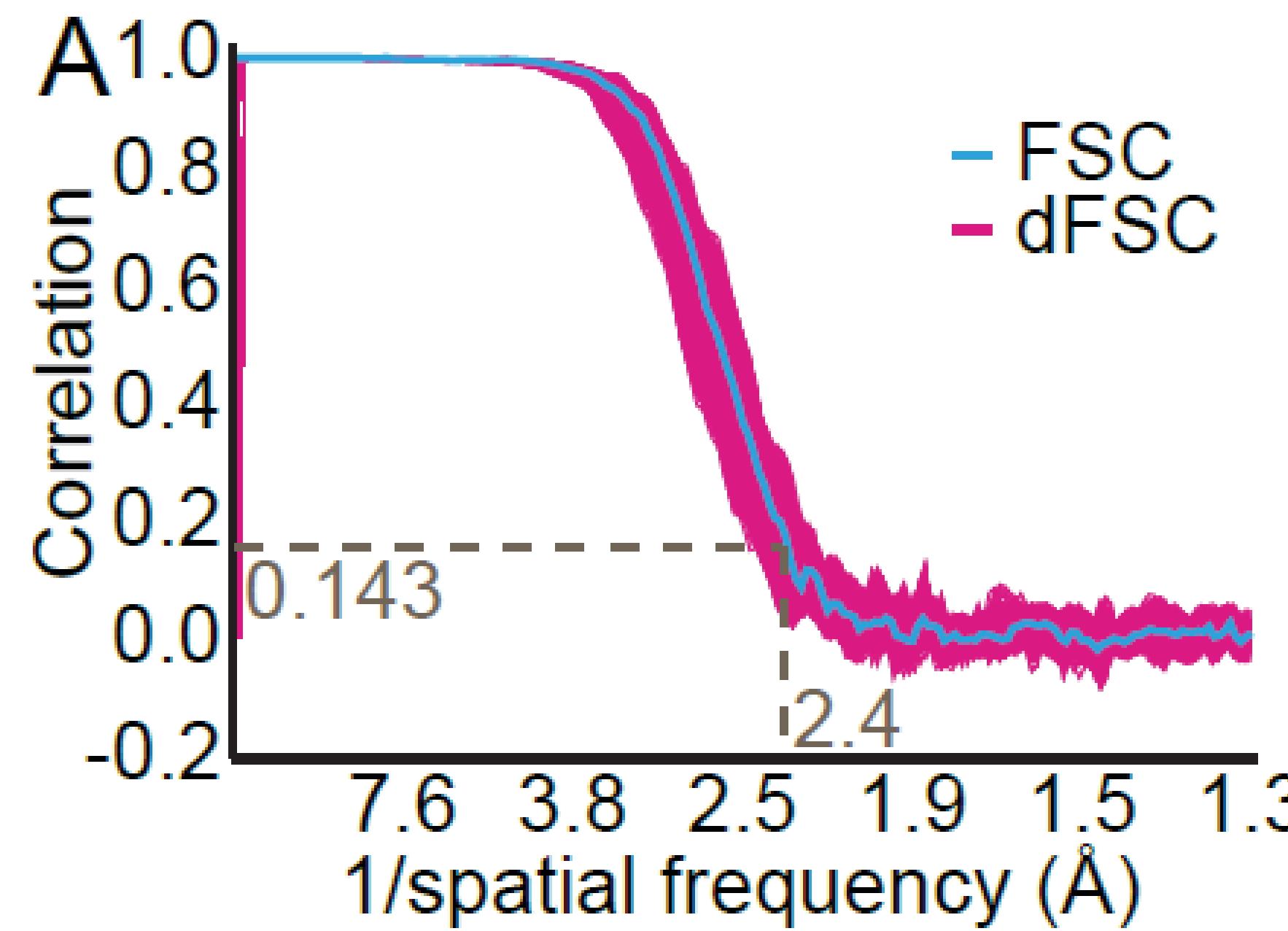
Directional FSC (dFSC)



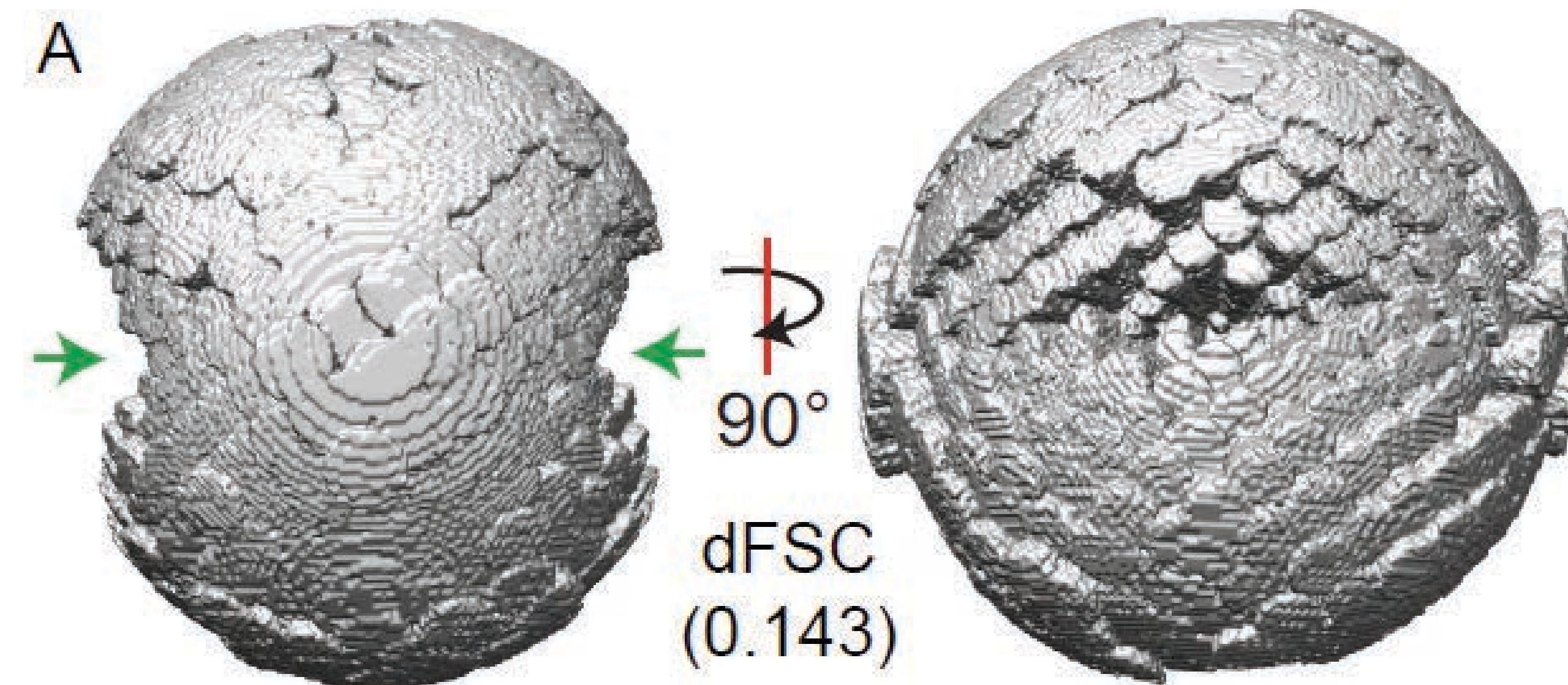
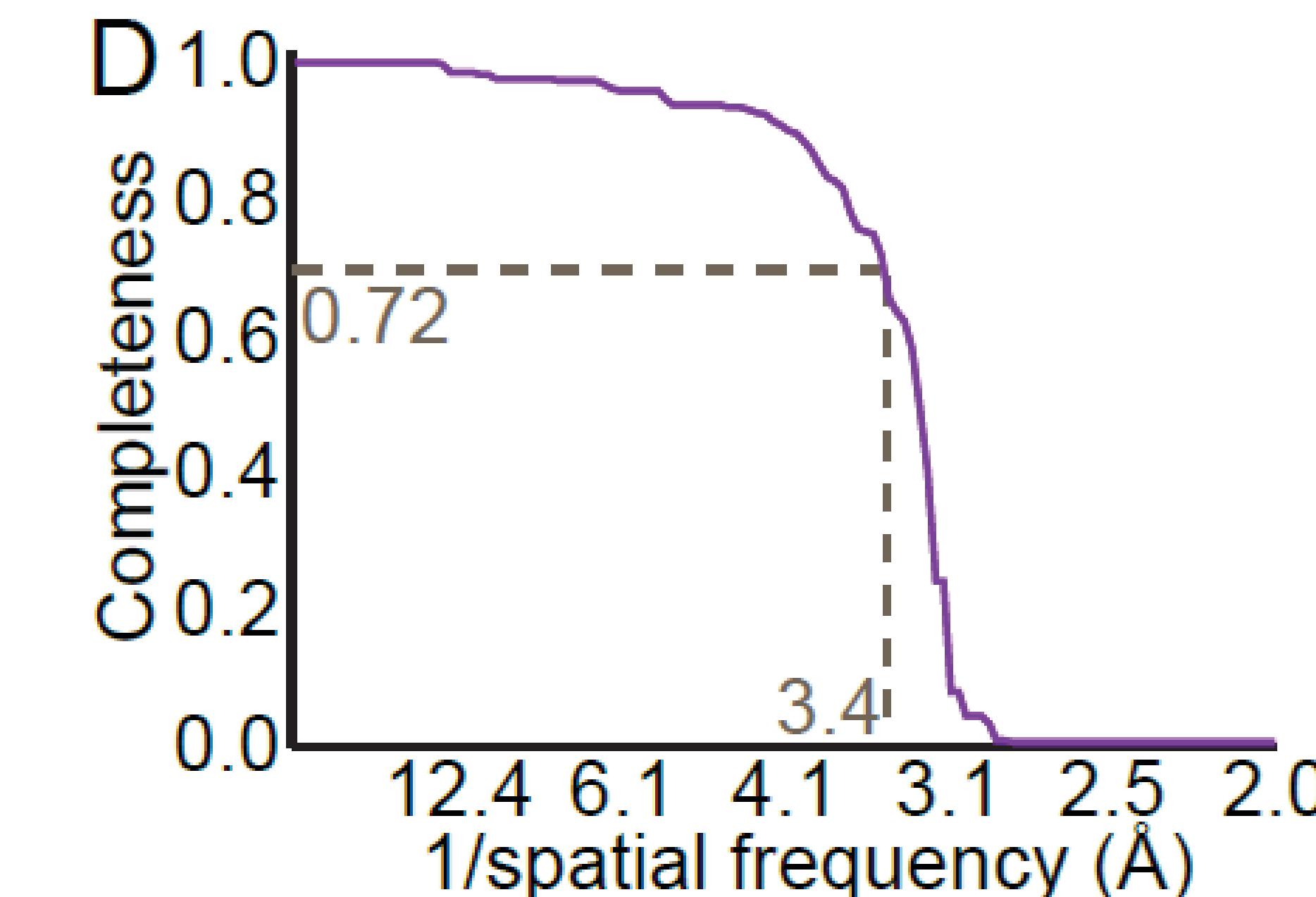
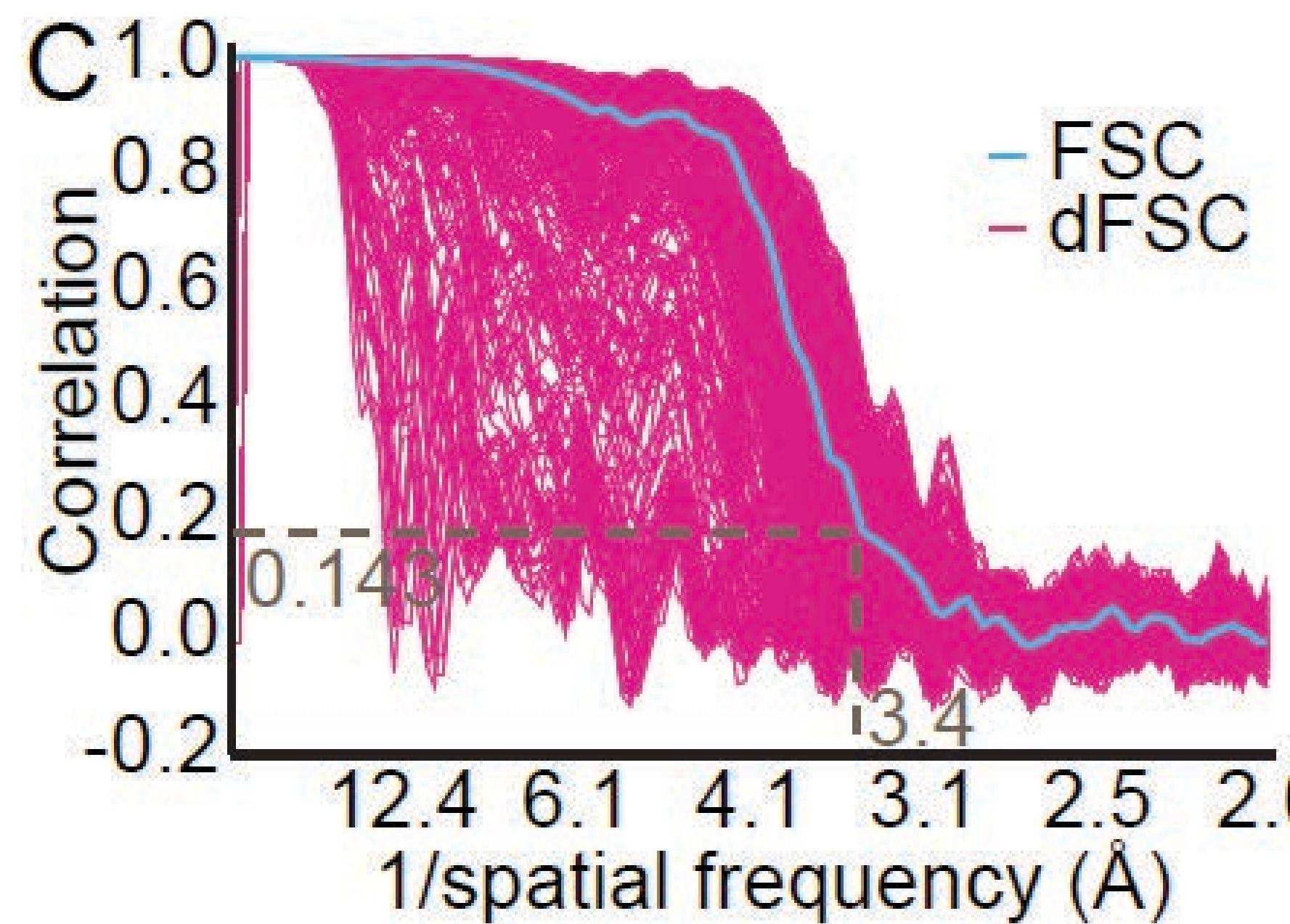
Jianhua Zhao

Unpublished

B-galactosidase



Directional FSC (dFSC) of TMEM16A



What are the common limiting factors?

From our own experience:

- Cryo-EM sample preparation is still a major obstacle;
 - ➡ Sample falling apart during plunge freezing is the most common issues;
 - possible solution: cross-linking, GraFix, Spotiton,
 - ➡ Sample preferred orientation
 - possible solution: tilting specimen, adding detergent below CMC, Fab, using substrate, etc.

Fab reduces preferred orientation of TMEM16A

