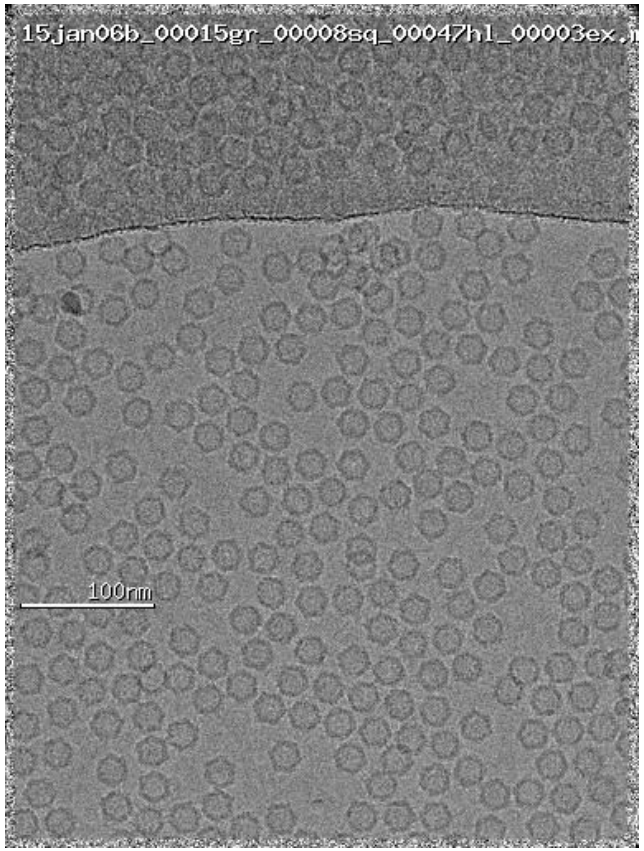
SECM^4 Workshop – Day 3

Cryo-EM Processing

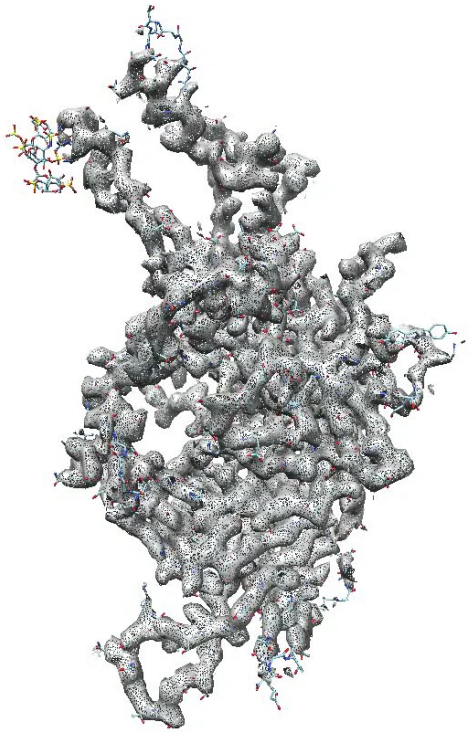
Scott Stagg

Single particle analysis



- ~1000 images (inherently atomic resolution)
- 10^4-10^6 particles

Alignment,
Classification,
Angular assignment

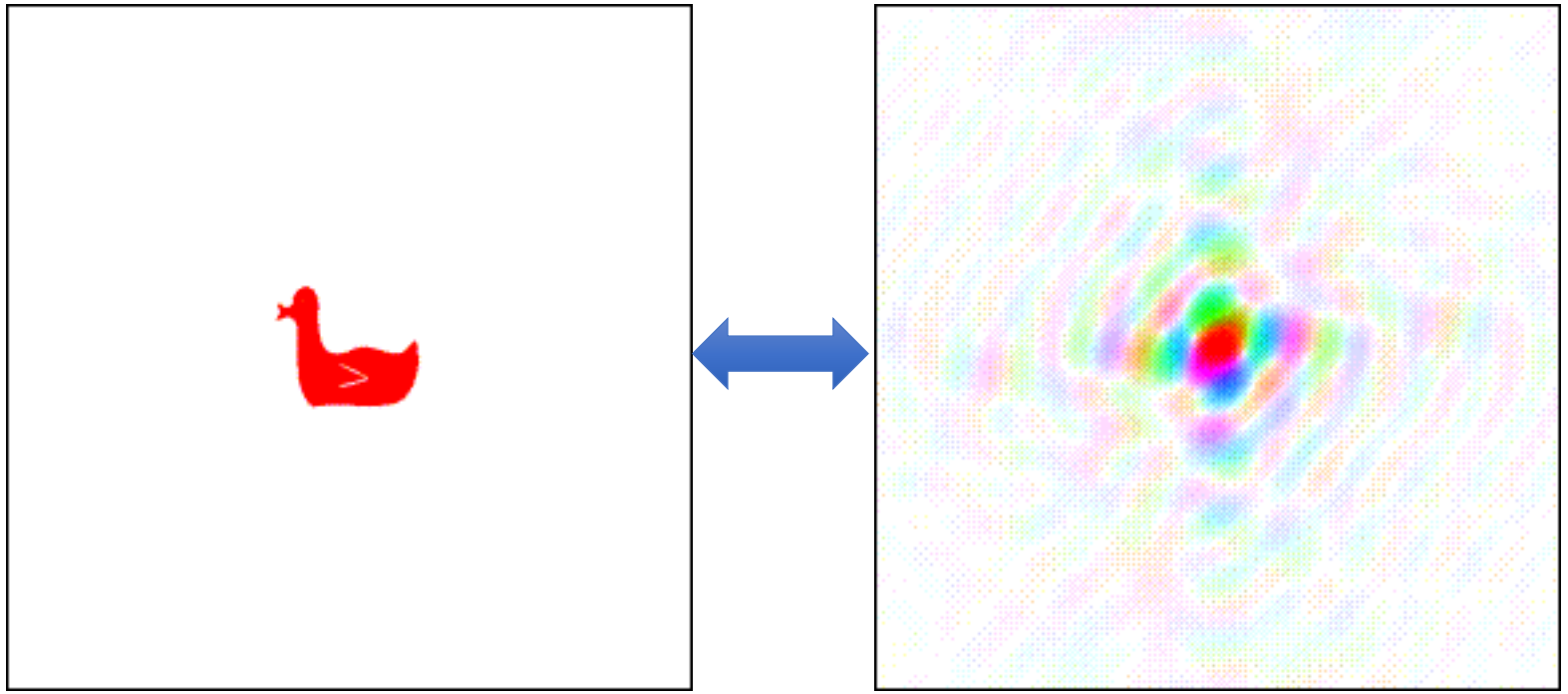


3D
reconstruction

Steps in a single particle reconstruction

- Collect data
- Align frames
- Estimate CTF
- Pick particles
- 2D Classification
- Generate initial model
- Refine data against initial model
- Estimate resolution

Fourier Stuff



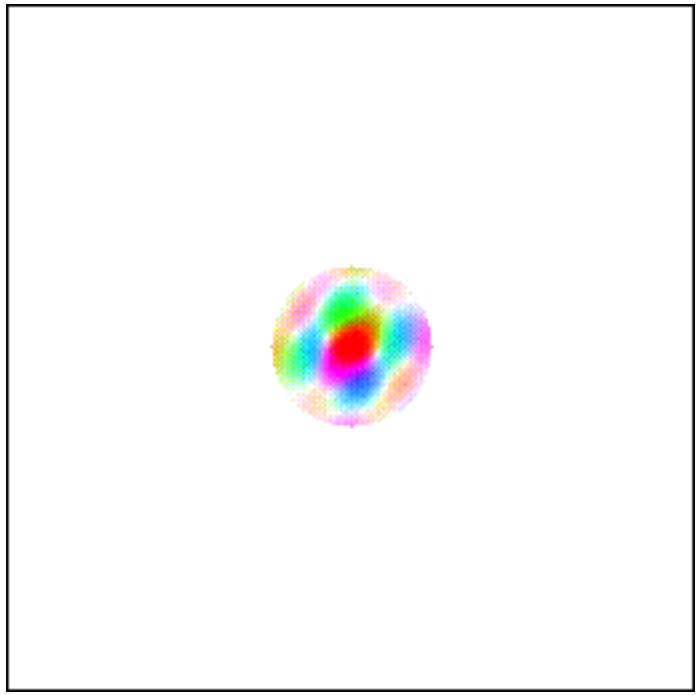
Real space image

Fourier transform

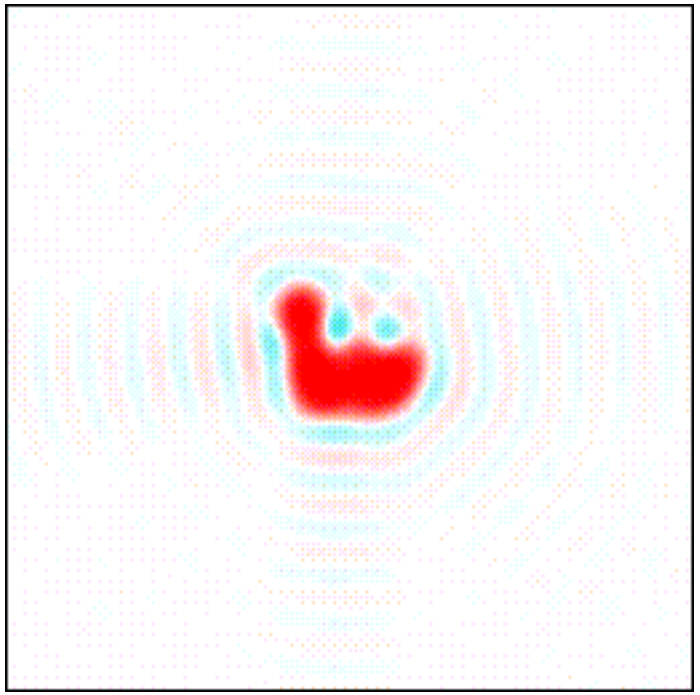
Brightness is amplitude,
colors are phases

From Kevin Cowtan's Picture Book of Fourier Transforms

Fourier Stuff



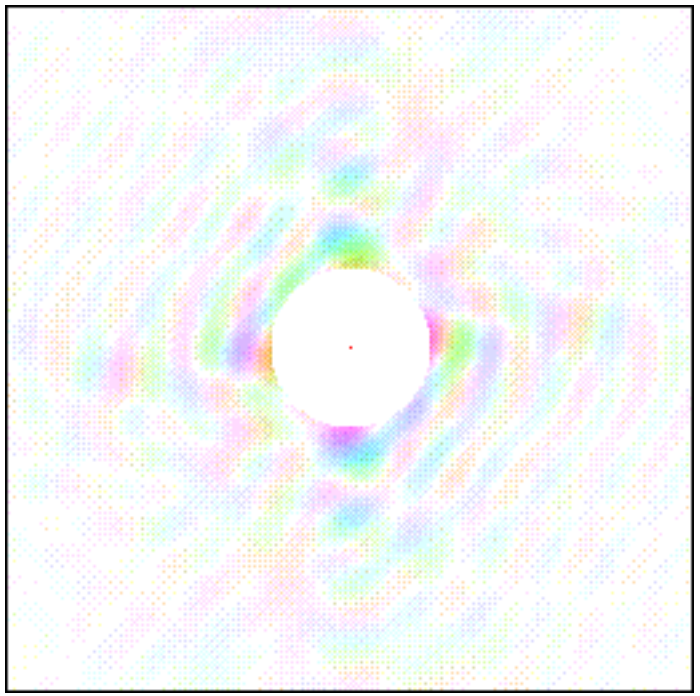
Low pass filter



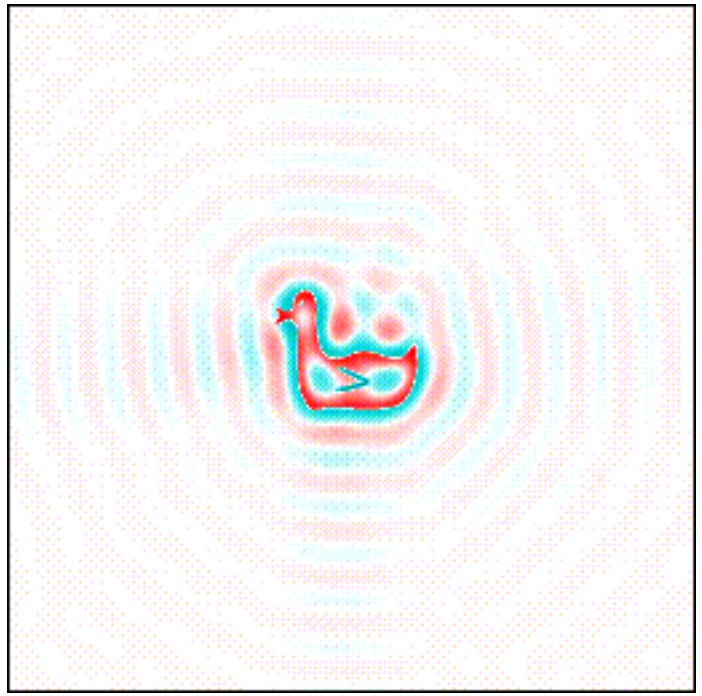
Reverse transform

From Kevin Cowtan's Picture Book of Fourier Transforms

Fourier Stuff



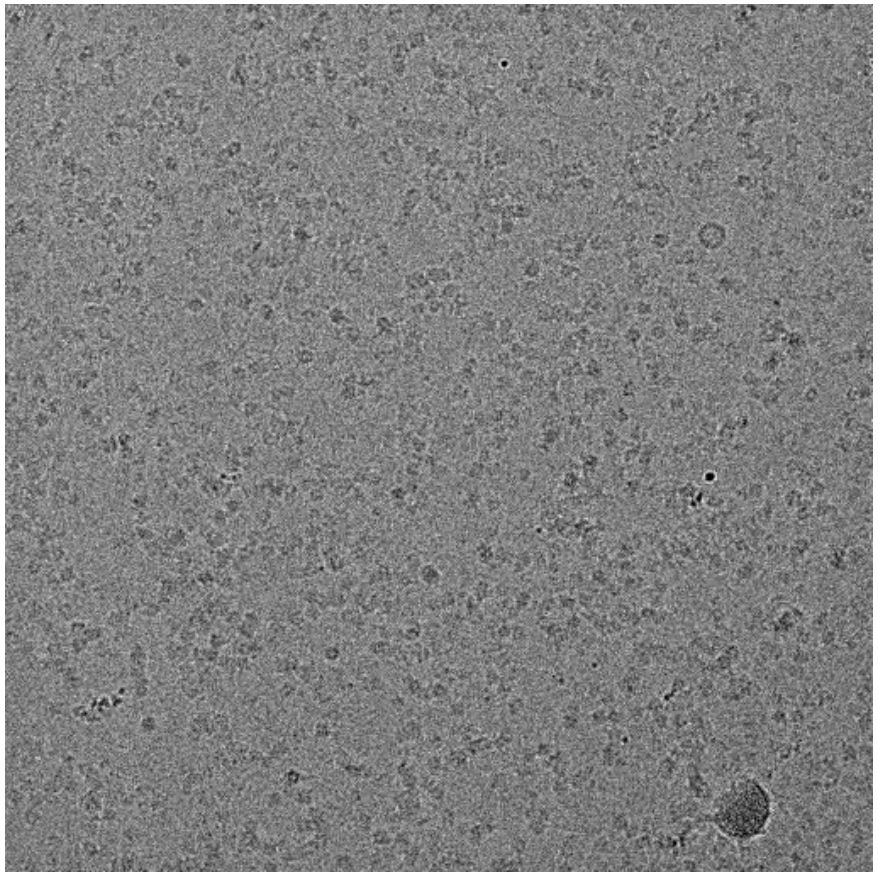
High-pass filter

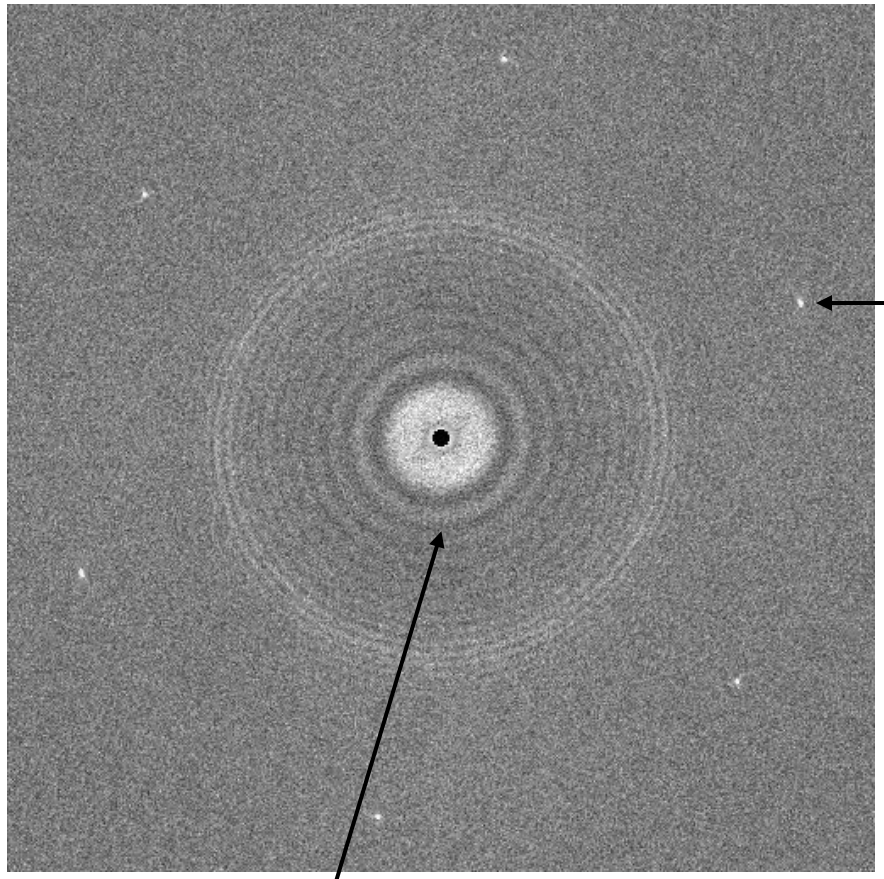


Reverse transform

From Kevin Cowtan's Picture Book of Fourier Transforms

Fourier transform of a micrograph



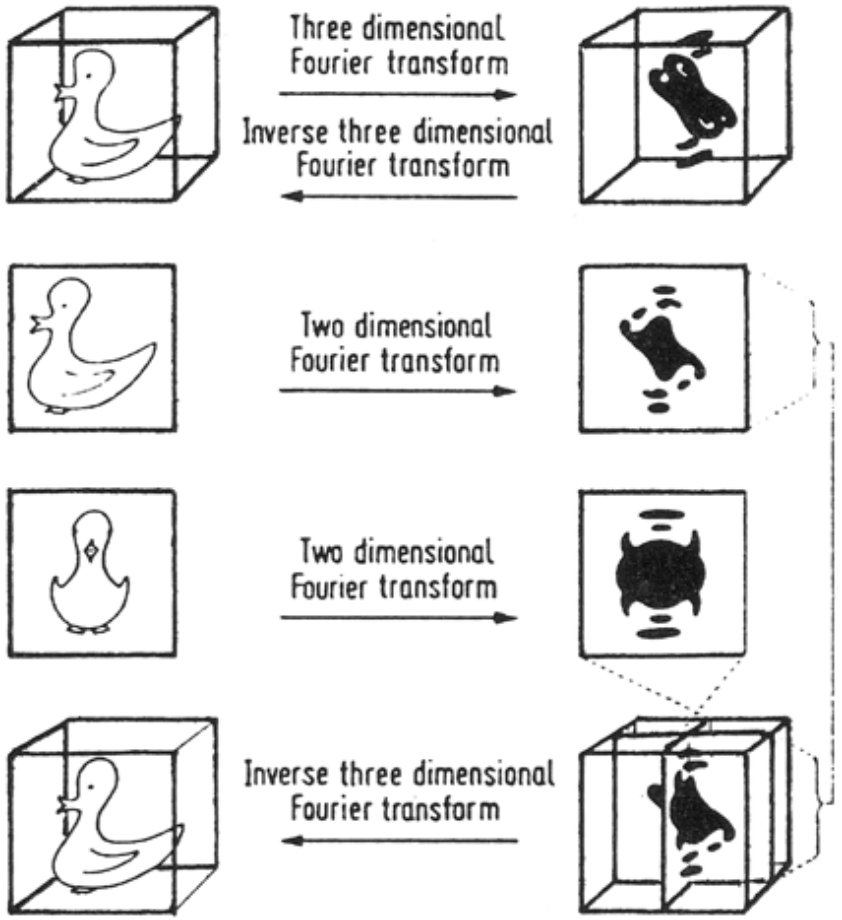


What is this?

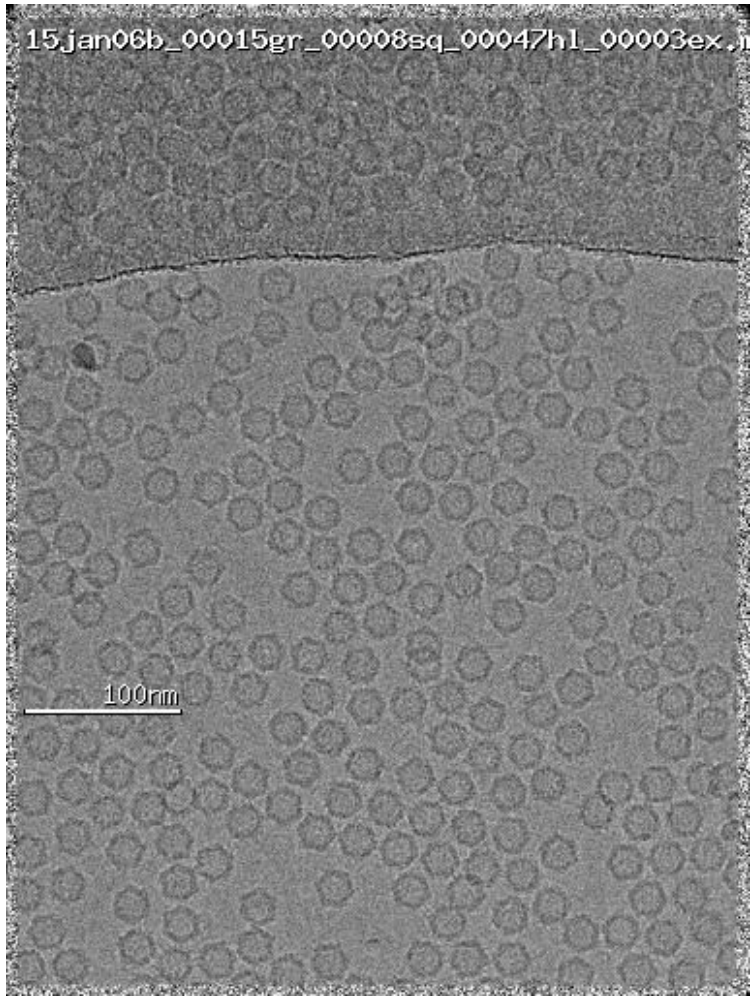
Thon rings that show the CTF

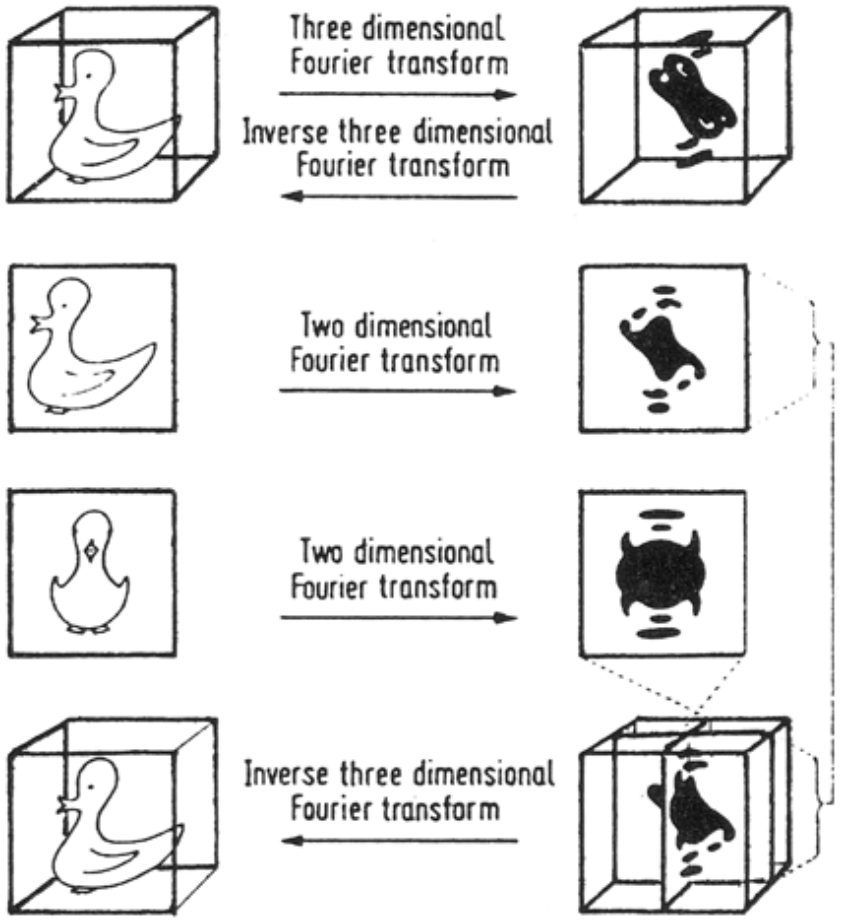
Principle of 3D reconstruction

- The projection theorem says that each projection of an object is a central section in Fourier space
- A 3-D reconstruction can be obtained by measuring a sufficiently large number of these projections covering as much of 3-D Fourier space as possible.



Single particle analysis



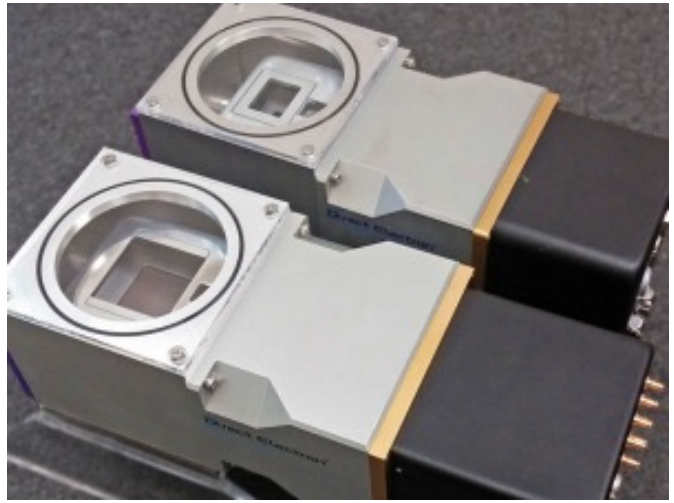


Steps in a single particle reconstruction

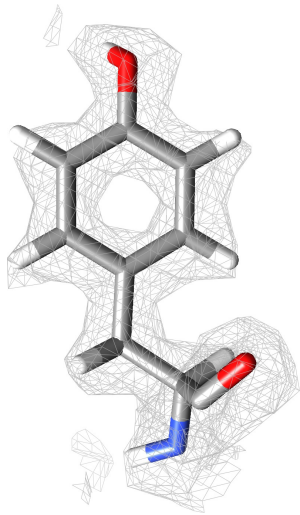
- Collect data
- Align frames
- Estimate CTF
- Pick particles
- 2D Classification
- Generate initial model
- Refine data against initial model
- Estimate resolution

The revolution in cryo-EM

+



=



The Titan Krios
- Stable
- Automated
- Aberration corrected
- Bright coherent beam

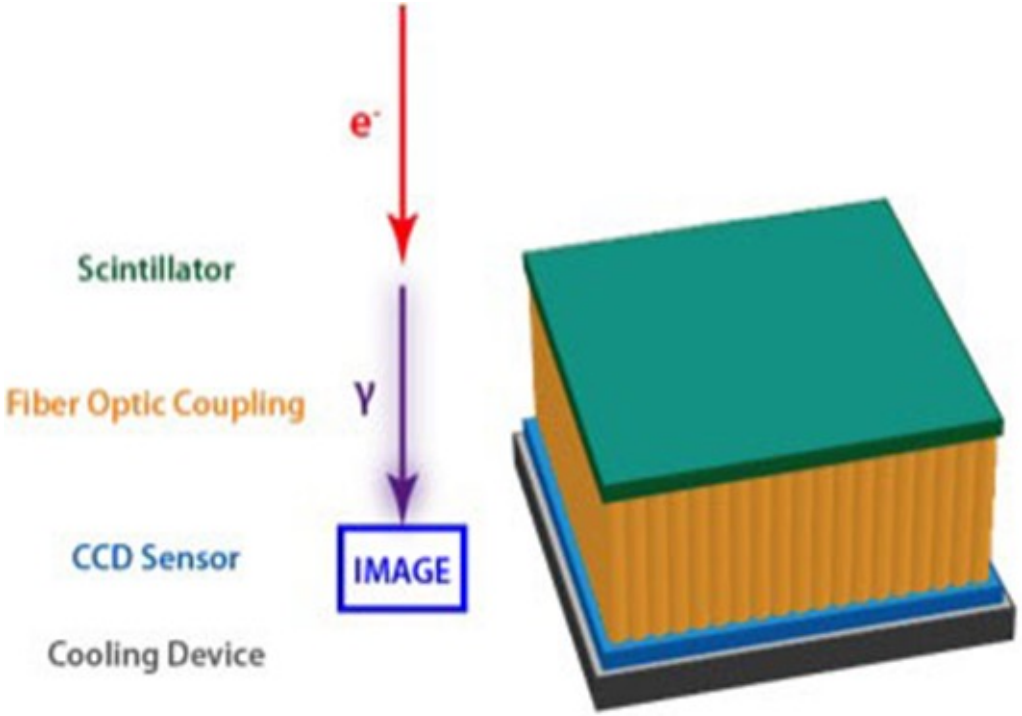
Direct electron detectors
- High DQE
- High frame rate

- HIgh resolution

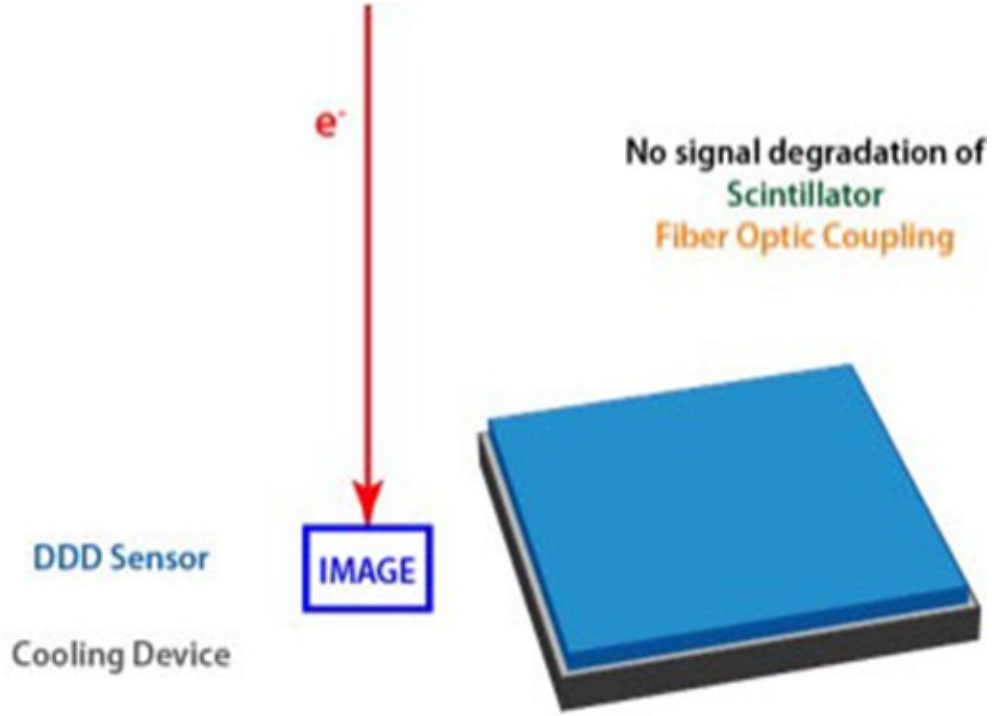
History of Cameras in Cryo-EM

- **Film Cameras** (pre-1990s)
  - First used in electron microscopy.
  - Low sensitivity and limited dynamic range.
  - Cumbersome processing and low throughput.
- **Charge-Coupled Devices (CCDs)** (1990s-early 2000s)
  - Improved sensitivity over film.
  - Digital processing, enabling easier data handling and analysis.
  - Still had limitations in terms of resolution due to indirect electron detection.
- **Direct Electron Detectors** (2010s onwards)
  - Marked a revolution in cryo-EM.
  - Detect individual electrons – facilitated large improvement in resolution.
  - Facilitated many of the advancements we'll discuss, such as movie alignment, dose compensation, and DQE improvements.

CCD vs DED



CCD Camera



Direct Detection Camera

Understanding DQE (Detective Quantum Efficiency)

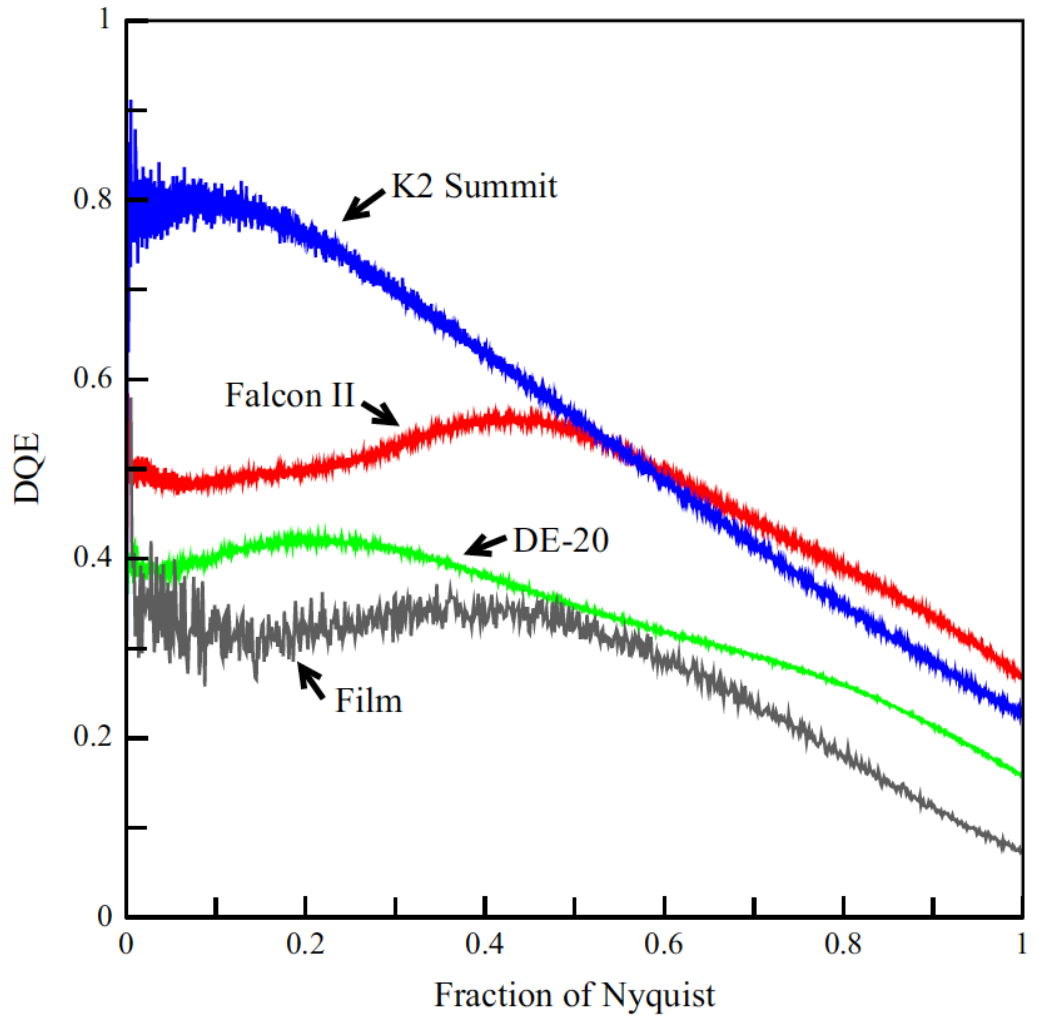
- **What is DQE?**
  - A metric for an imaging system's capability to capture information with high fidelity as a function of resolution.
- **Why DQE Matters:**
  - A higher DQE translates to clearer images with reduced noise.
  - Crucial for achieving high-resolution images in cryo-EM.
- **Evolving Tech & DQE:**
  - Direct electron detectors (DEDs) have dramatically improved DQE compared to older CCDs.
  - Result: Enhanced image clarity and information retention.

$$DQE(\omega) = \frac{SNR_{out}^2(\omega)}{SNR_{in}^2(\omega)}$$

DQE of 1 is good DQE of 0 is bad

DQE (Detective Quantum Efficiency)

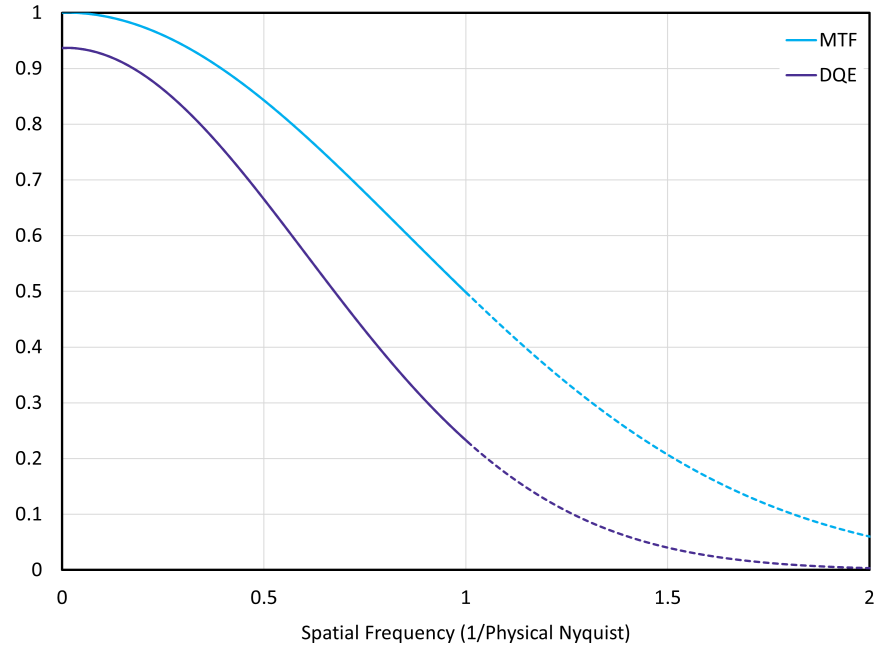
- **DEDs have higher DQEs than other detection methods**
  - With the advent of Direct Electron Detectors, there was a significant improvement in DQE over CCDs.
  - This meant that these newer cameras were better at capturing and representing the information from the electrons



McMullan et al., 2014

DQE (Detective Quantum Efficiency)

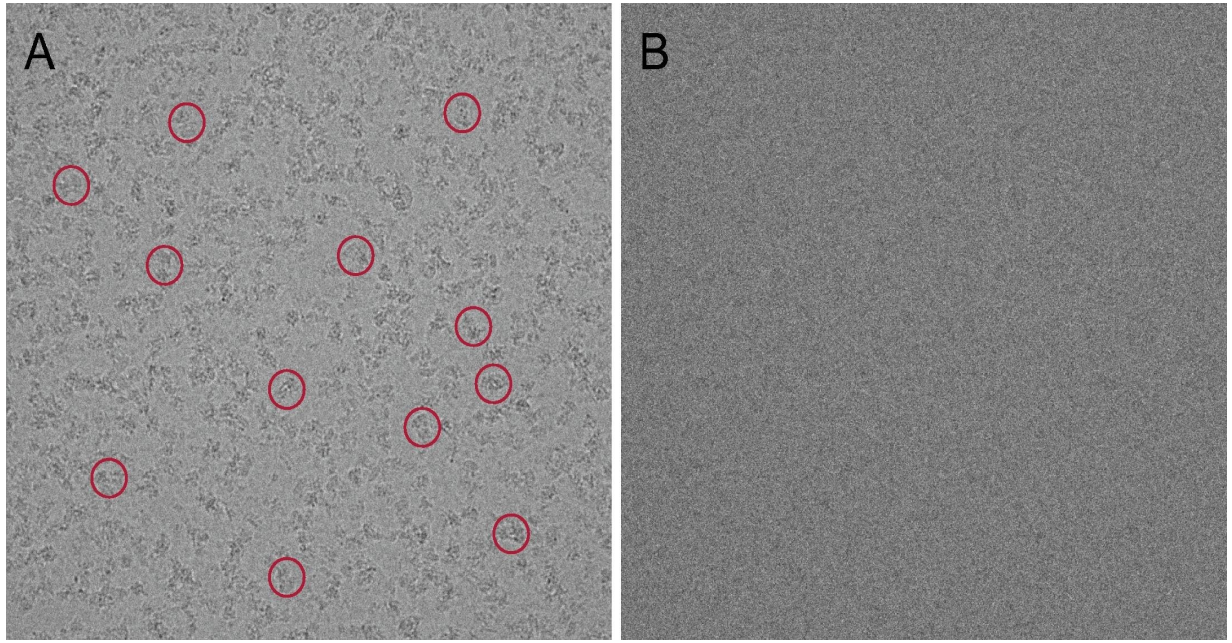
- **DQE Values and Interpretation**
  - If DQE(f)=1 or 100%: Perfect information preservation
  - If DQE(f)=0 or 0%: No preservation of information.
- **Why DQE Matters**
  - Resolution in single particle reconstruction depends on aligning particle images with high accuracy and precision
    - DQE, especially at low frequencies, is critical for image alignment accuracy and precision



DQE of the DE Apollo at FSU

Direct Detection and Frames

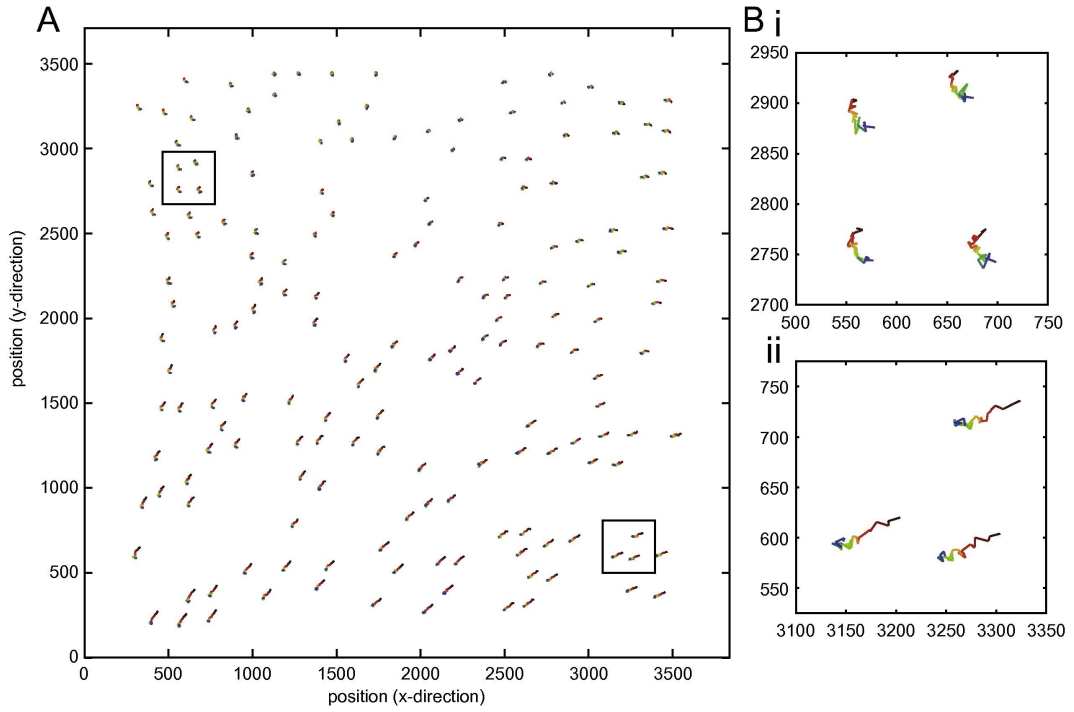
- **Movie Mode:**
  - Direct detectors capture rapid sequences of frames during exposure.



A) sum aligned high SNR

B) single frame low SNR

- **Benefits of Multiple Frames:**
  - Corrects specimen drift and movement.
  - Dose fractionation minimizes radiation damage.

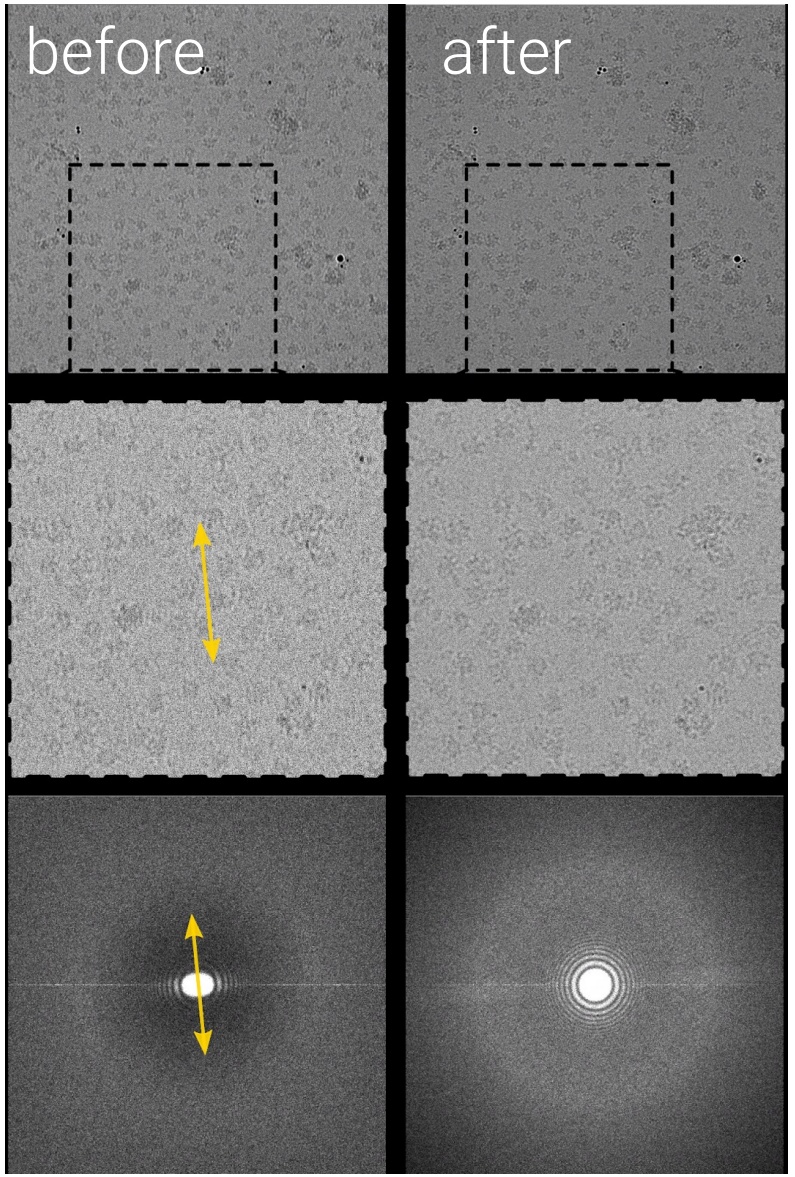


Particle motion/drift trajectories

https://doi.org/10.1016/j.jsb.2015.08.007

Motion Correction/Movie Alignment

- **Alignment Purpose:**
  - Align frames from direct electron detectors to counter specimen motion.
- **Importance:**
  - Beam induces specimen motion or stage causes drift.
    - Blurs images, reducing resolution.
  - Frame alignment enhances image clarity & signal-to-noise.
- **Techniques:**
  - **Cross-Correlation:** Aligns frames to determine shifts.
  - **Least-squares:** Minimizes frame differences.
- **Outcome:**
  - Clearer images with reduced drift.
  - Restores resolution for subsequent image processing stages.



Motion effects are especially noticeable by the directional loss of Thon rings in the power spectrum of the image

https://cryoem101.org/chapter-5/

Dose Compensation in CryoEM

Definition:

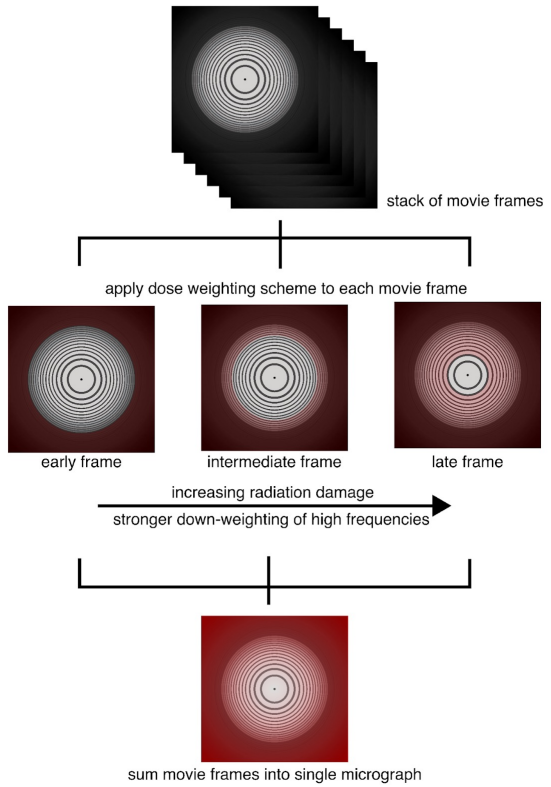
- Adjusts frame contribution based on received electron dose.
- Counters radiation damage's impact on successive frames.

Why it Matters:

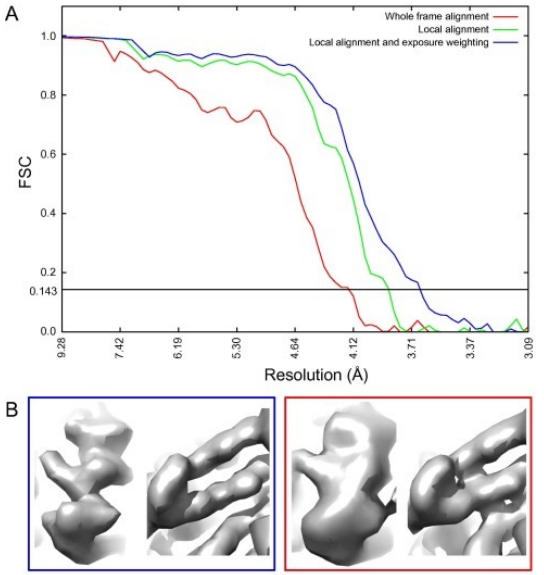
- **Radiation Damage:** Electron beams cause structural changes and information degradation in specimens.
- **Information Preservation:** Early frames have more high-res details; latter ones suffer from cumulative damage.

Impact on Image Quality:

- Blurred, artifacts, and lower resolution due to inconsistent frame weights.
- Frames weighted by reliability, improving signal-to-noise ratio, higher resolution.
- Ensures optimal data quality despite radiation damage.



Early frames retain content while later, high-exposure frames down-weight high frequencies, all summed into the final micrograph



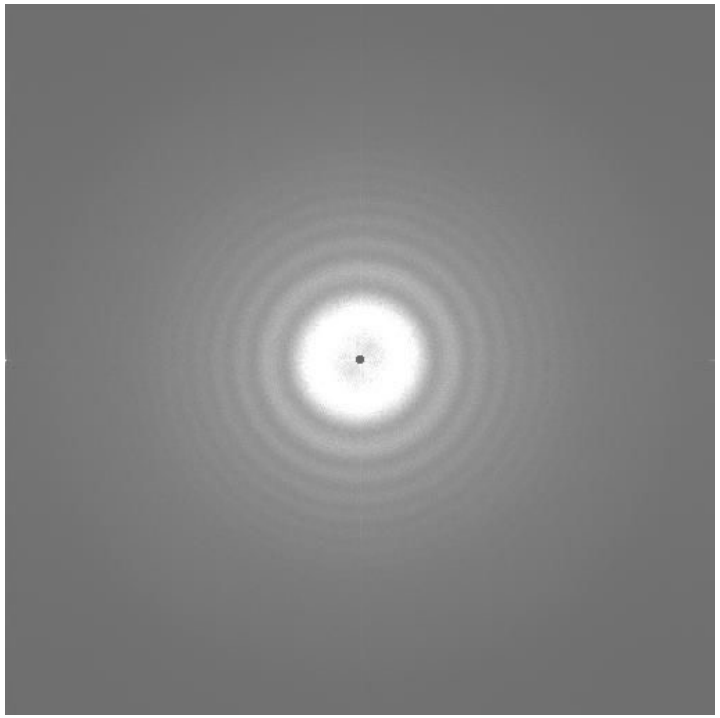
- individual particle motion corrected and exposure weighed map (blue box)
- whole frame alignment only (red box)

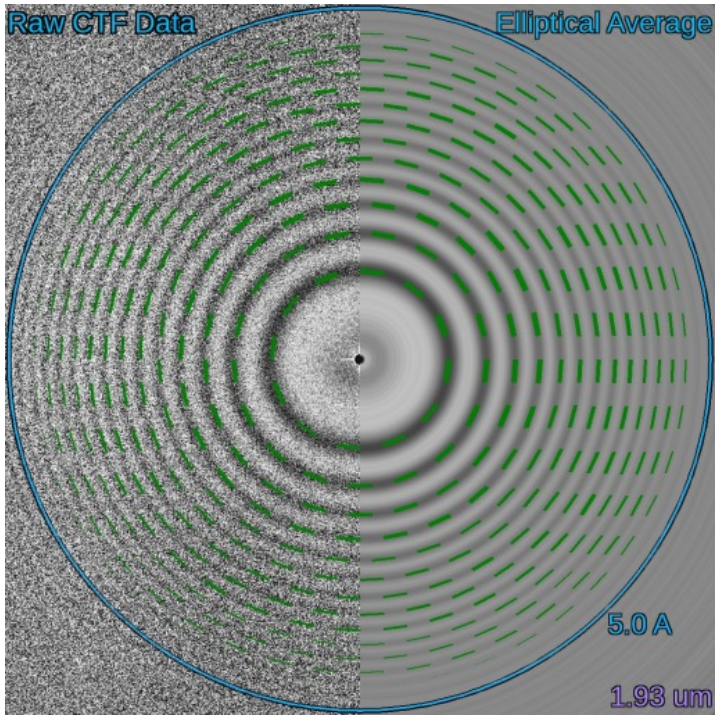
https://doi.org/10.1016/j.jsb.2015.08.007 https://cryoem101.org/chapter-5/

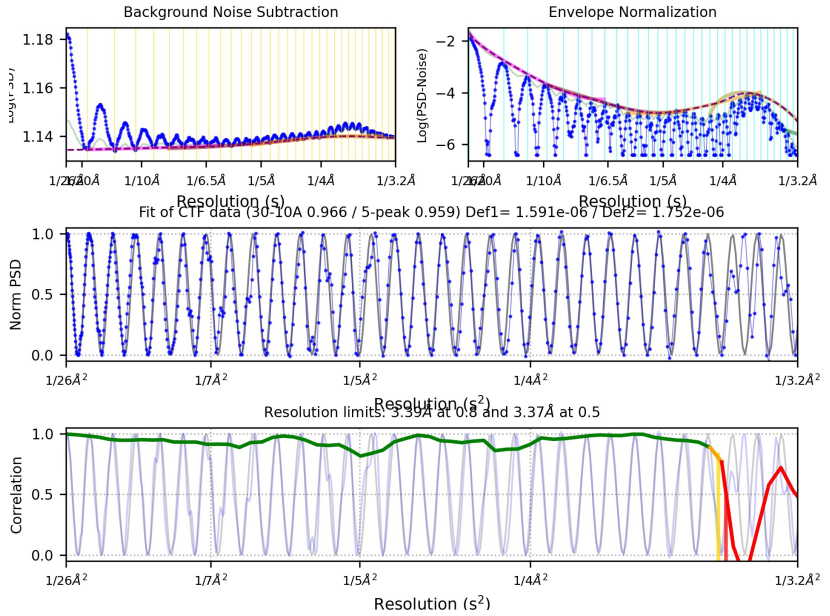
Steps in a single particle reconstruction

- Collect data
- Align frames
- Estimate CTF
- Pick particles
- 2D Classification
- Generate initial model
- Refine data against initial model
- Estimate resolution

Estimating the CTF





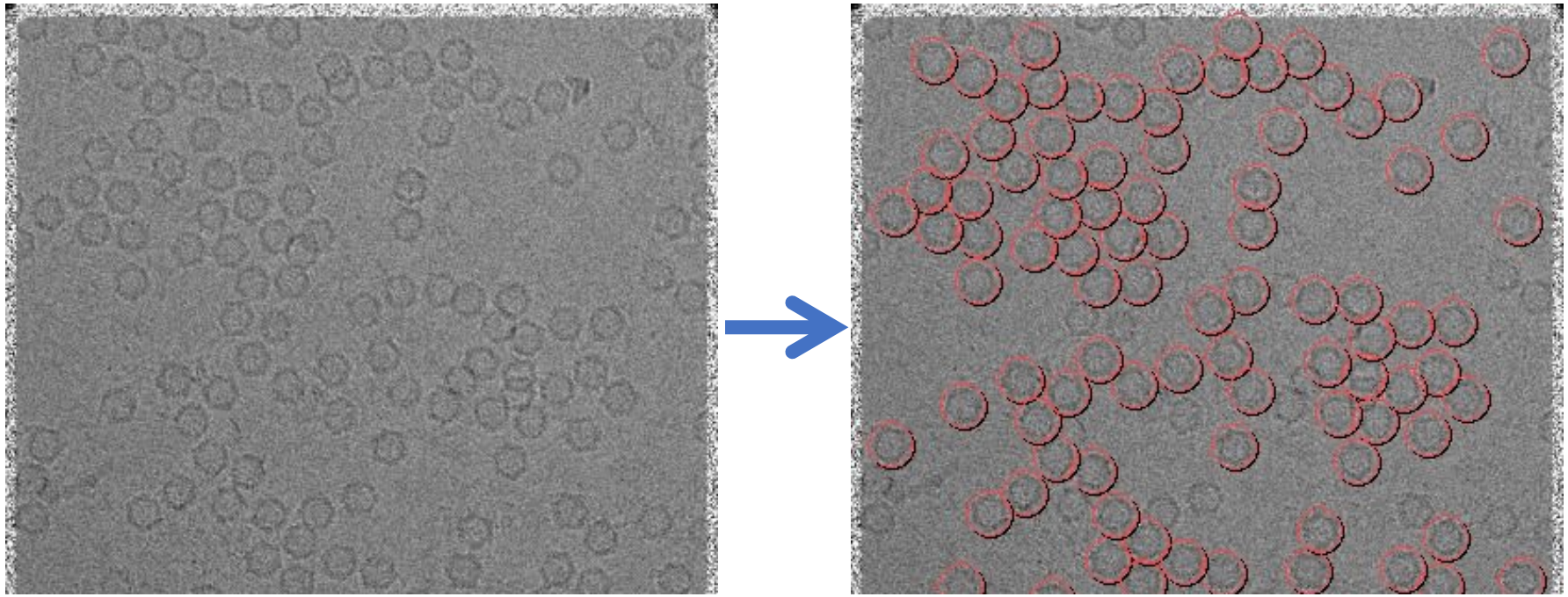


- Many tools exist for automatic CTF estimation
  - Determines defocus and astigmatism for each image
  - Resolution potential present in image can be estimated from agreement between estimate and data

Steps in a single particle reconstruction

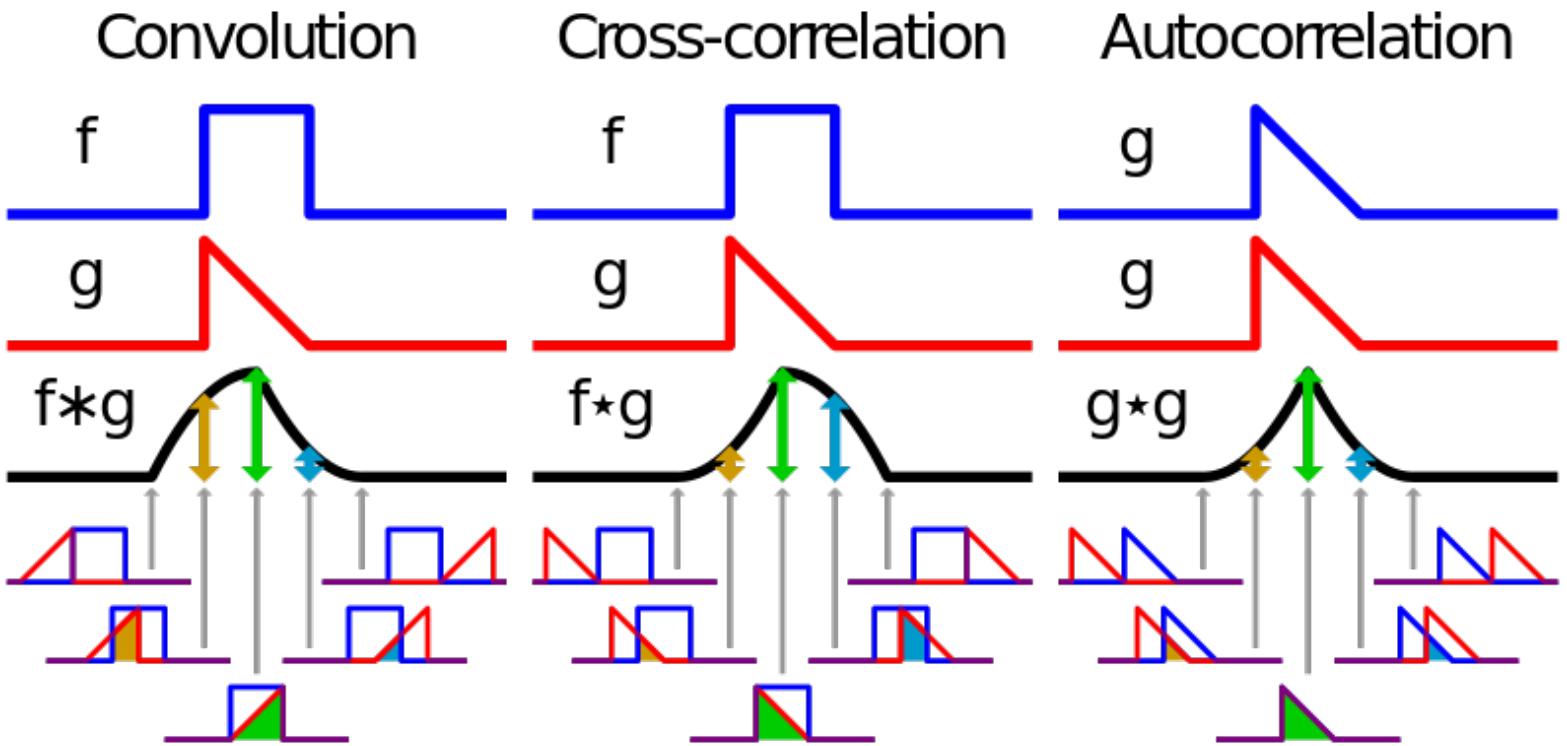
- Collect data
- Align frames
- Estimate CTF
- Pick particles
- 2D Classification
- Generate initial model
- Refine data against initial model
- Estimate resolution

Picking particles



- Mostly automated
  - Many tools exist for template-based picking
  - Many tools for machine learning based picking
  - Some tools for general picking based on size

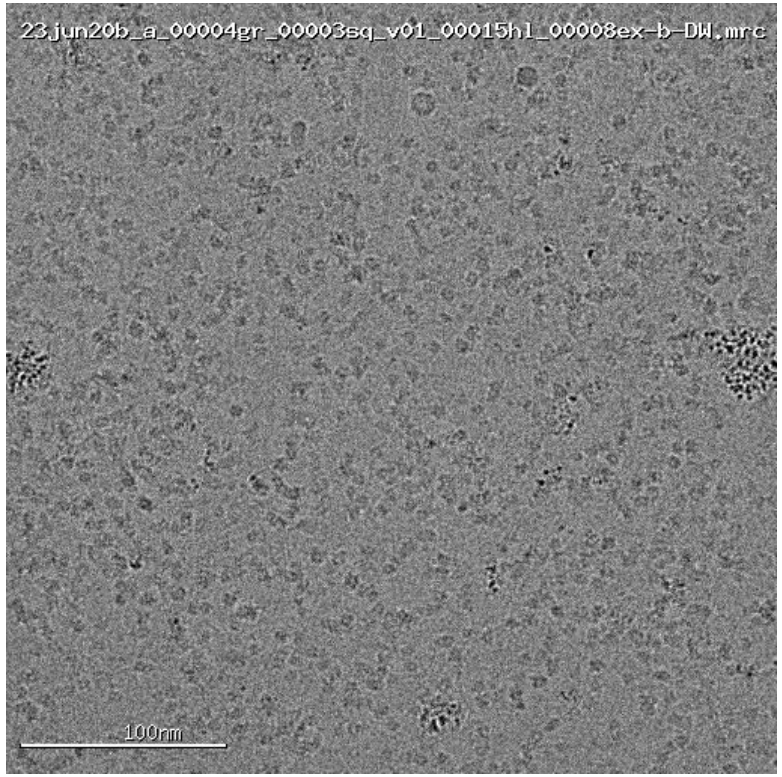
An aside about image alignment

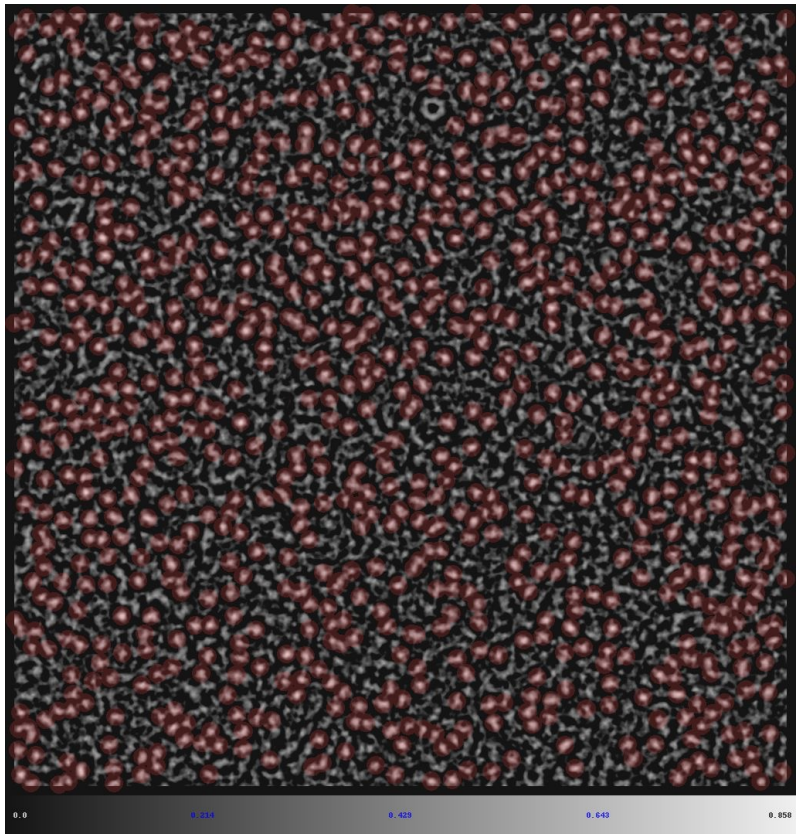


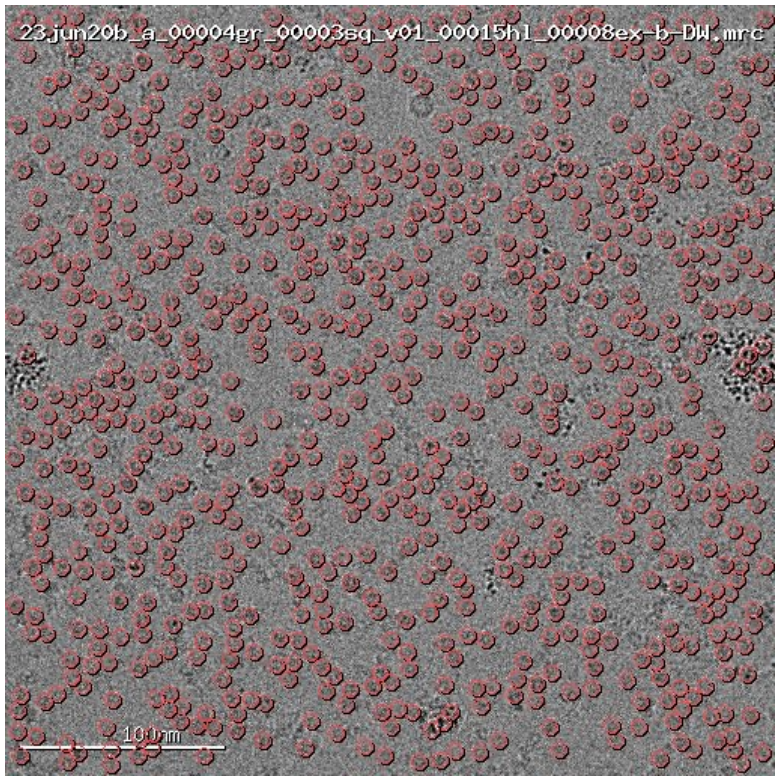
The discrete autocorrelation R at lag l for a discrete signal $y(n)$ is

$$R_{yy}(l) = \sum_{n \in Z} y(n)\,\overline{y}(n-l).$$

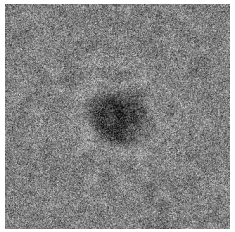
Template based picking







Cross correlated with template



Steps in a single particle reconstruction

- Collect data
- Align frames
- Estimate CTF
- Pick particles
- 2D Classification
- Generate initial model
- Refine data against initial model
- Estimate resolution

Image Alignment in cryo-EM

Importance: Particles in cryoEM have random orientations. Accurate alignment ensures high-resolution 3D reconstruction.

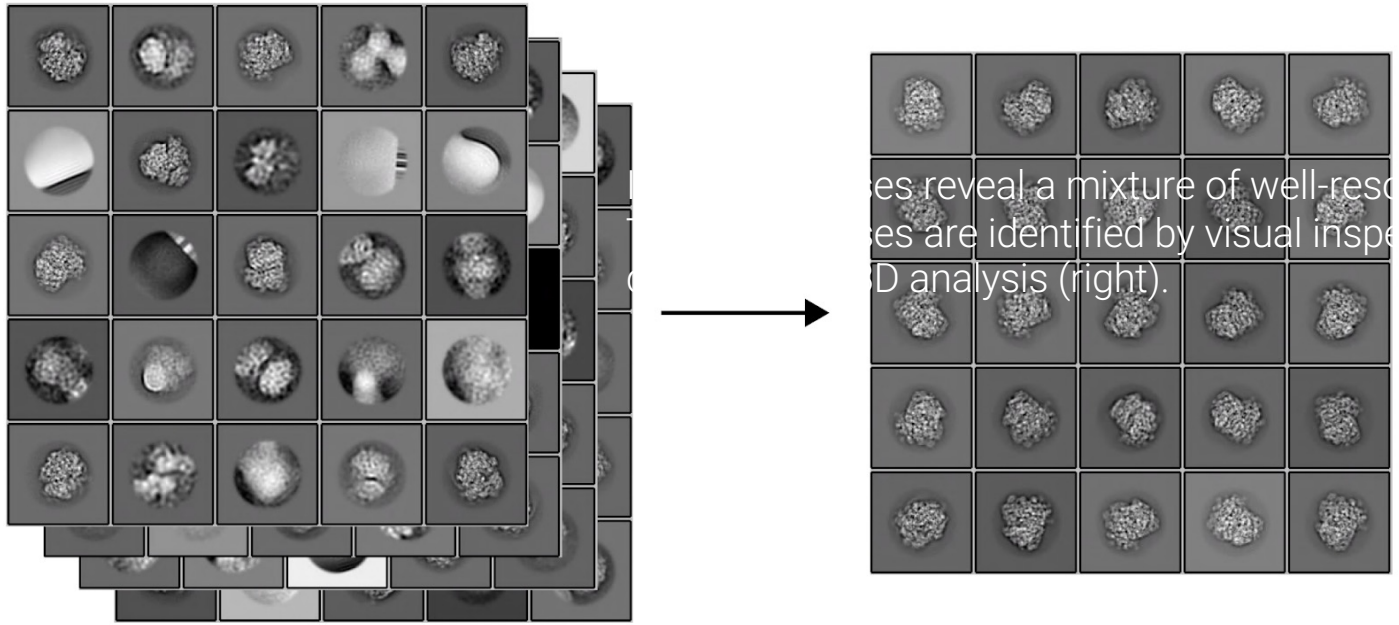
Techniques:

- **Cross-Correlation:** Shifts one image over another to measure similarity.
- **Maximum Likelihood:** Estimates the best orientation and position for particles.
- **Reference-based:** Iteratively aligns using a reference image.

Challenges:

- **Reference Bias:** Wrong references can skew results.
- **Computational Needs:** High-resolution alignment requires powerful computation.
- **SNR Issues:** Noise can hinder alignment.

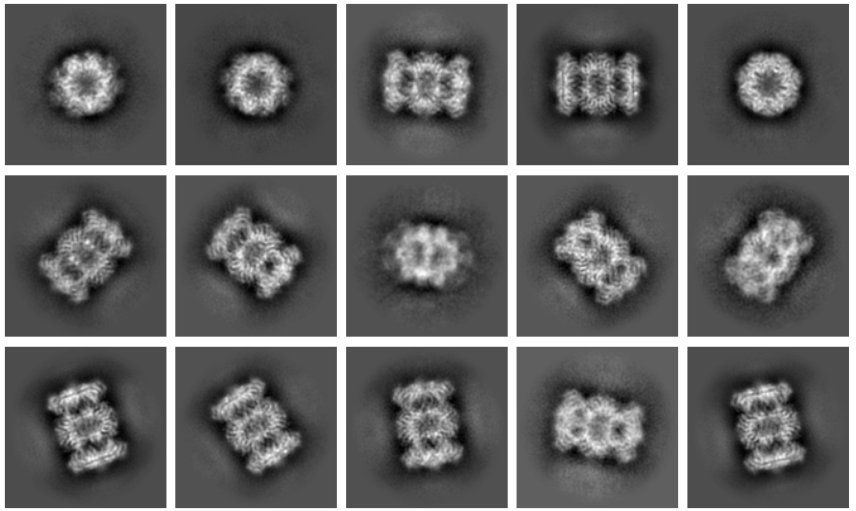
Outcome: Groups similar projections and produces accurate 2D and 3D (later) reconstructions.



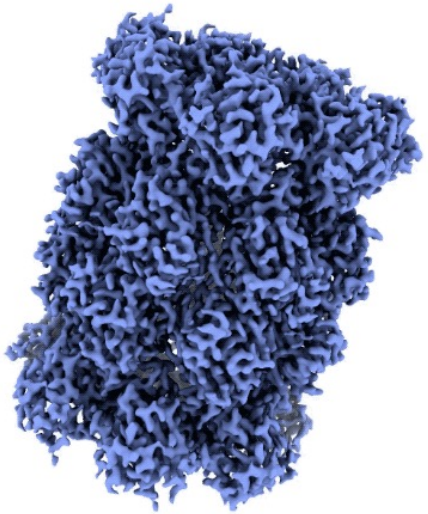
2D classification in cryo-EM

High resolution 2D classes

- Multiple orientations visible
- Well-aligned images
- High-res features visible
- Large number of ptcls per class
- All prerequisites for high-resolution 3D reconstruction



20S proteasome - 2D class averages



20S proteasome - 3D reconstruction

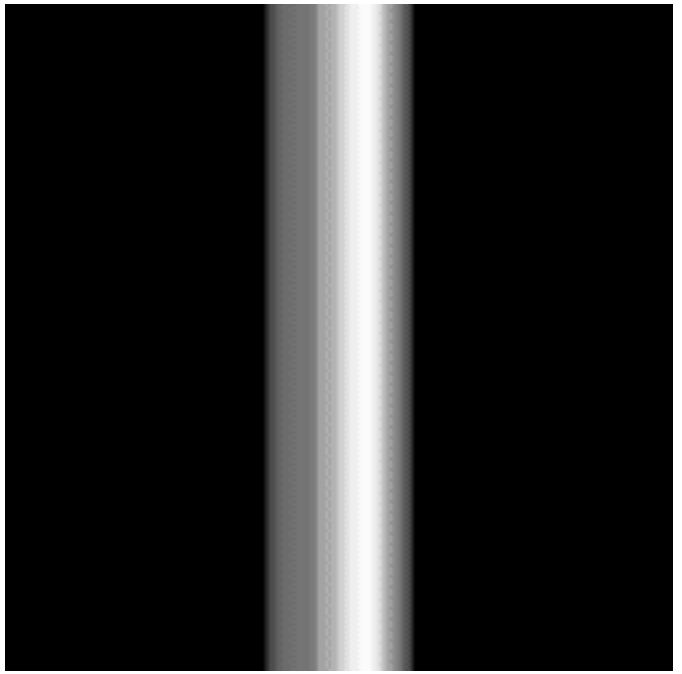
Steps in a single particle reconstruction

- Collect data
- Align frames
- Estimate CTF
- Pick particles
- 2D Classification
- Generate initial model – magic happens (i.e. we’re skipping over this part)
- Refine data against initial model
- Estimate resolution

Basic Principles of 3D Reconstruction

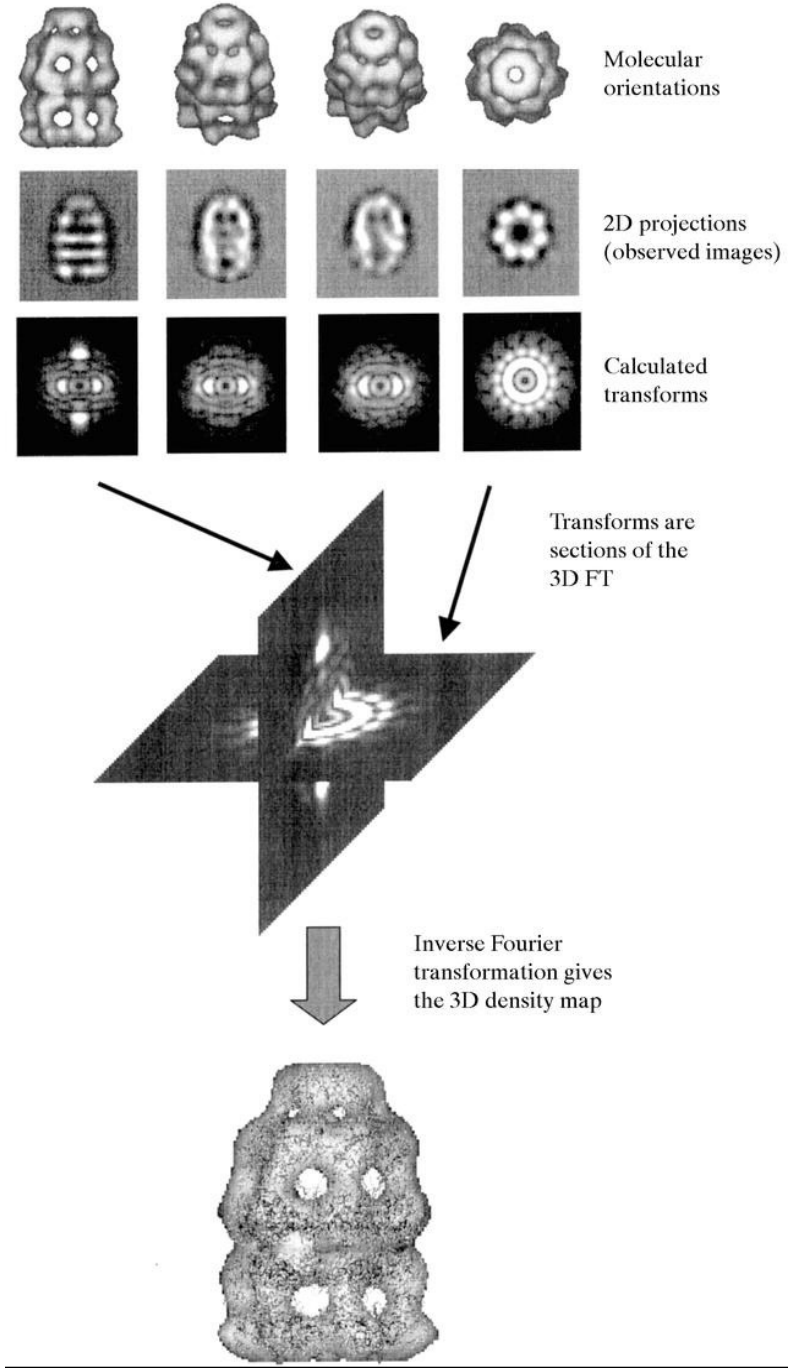
- **Definition:** Transforming 2D particle images from various orientations into a coherent 3D volume, revealing the specimen's structure.
- **Back-projection:** Uses 2D images to estimate density at each point in 3D space.
- **Fourier Space:** CryoEM uses Fourier transforms to convert spatial images into frequency space. This helps in manipulation, alignment, and averaging.
  - **Fourier Inversion:** Converts the frequency information back to real space to produce a 3D structure.
  - **Central Section Theorem:** Each 2D projection provides a slice (or section) of the 3D Fourier transform of the specimen.

Reconstructing by backprojection

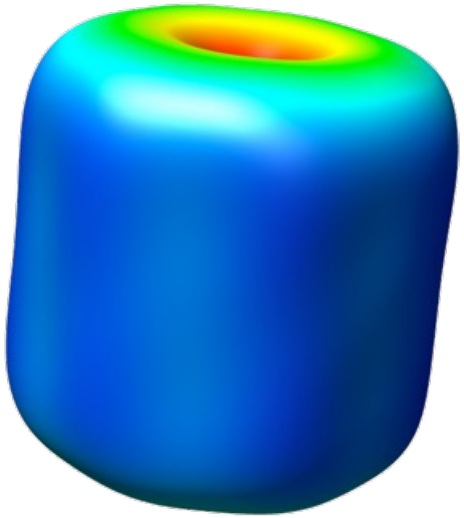


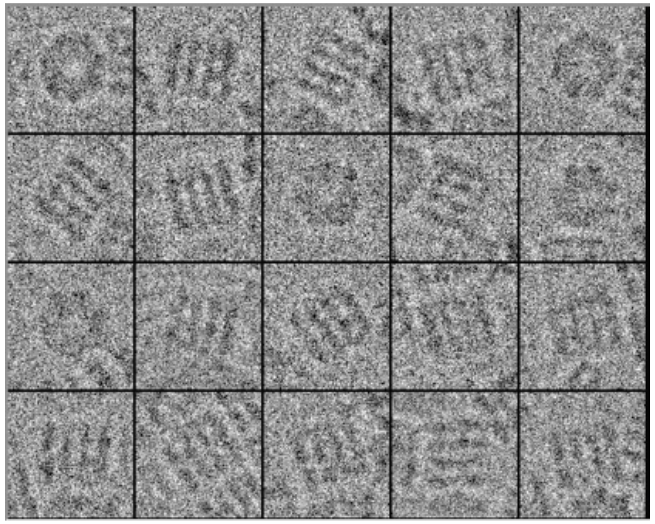
Fourier Inversion and Cryo-EM

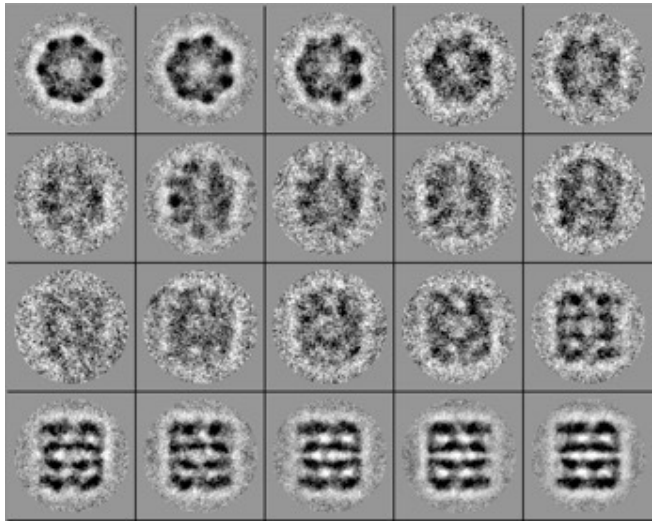
- Transforming data from spatial to frequency domain
- Essential for image processing and 3D reconstruction
- Allows for filtering and resolution enhancement
- Underpins many Cryo-EM algorithms

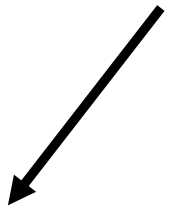


Classification and averaging

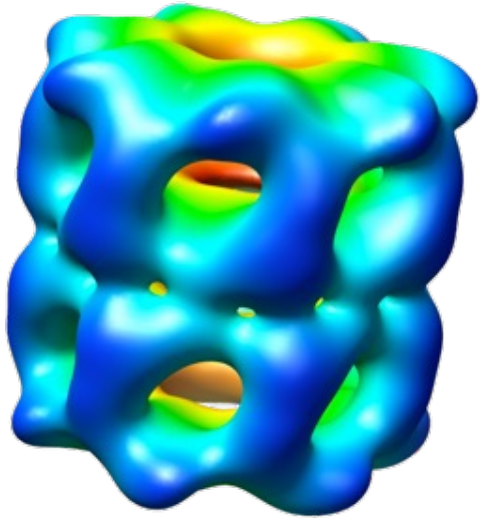


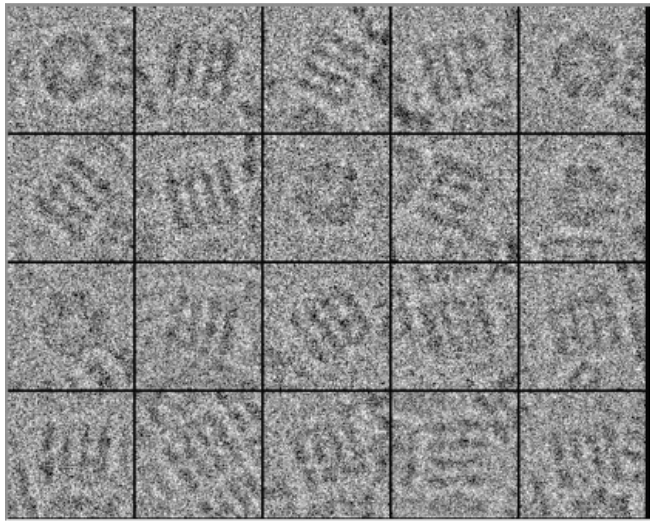


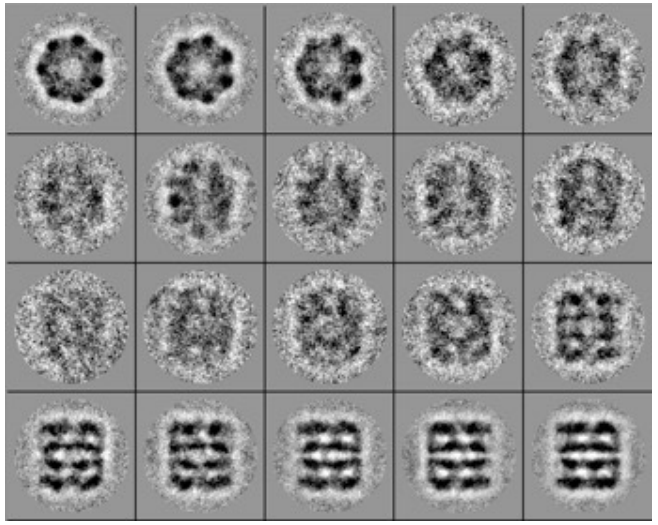




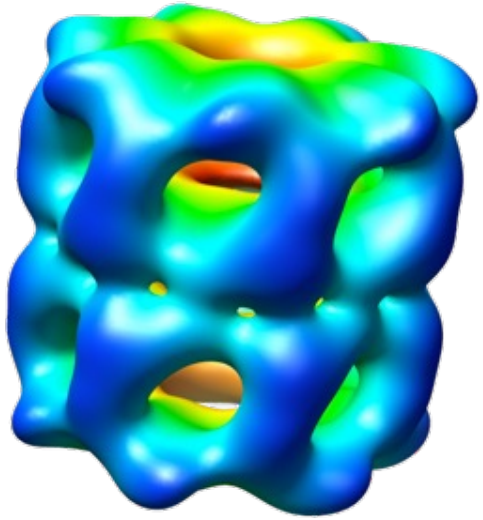
Backprojection yields new 3D model

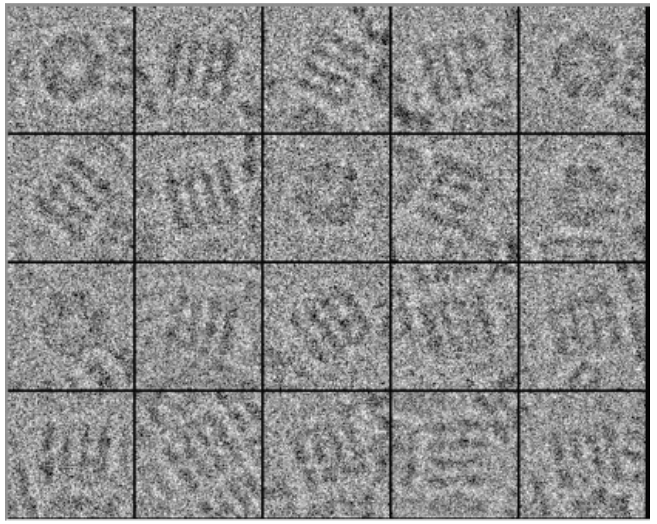


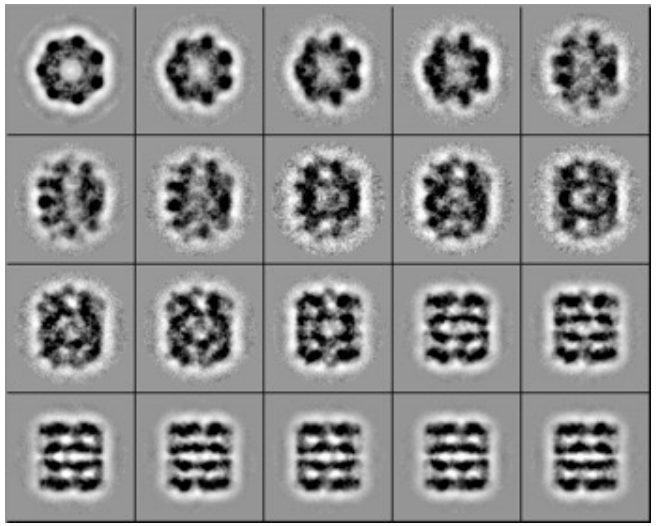




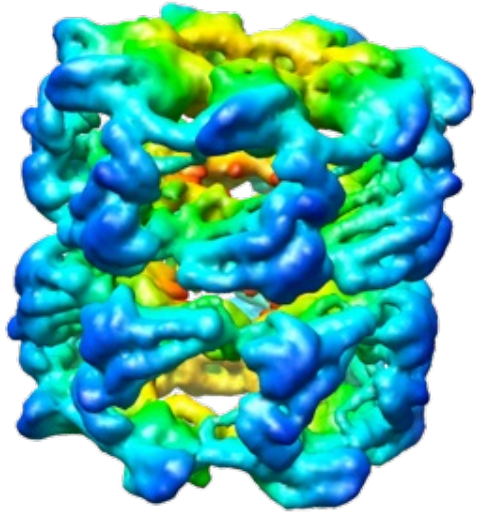
Higher resolution model yields better classification

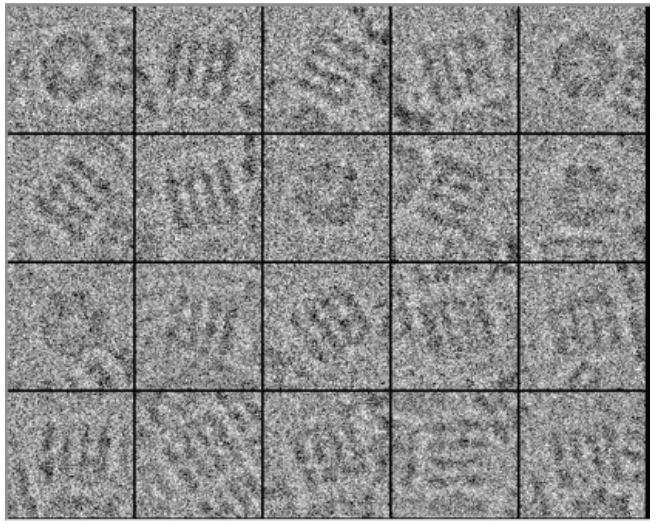


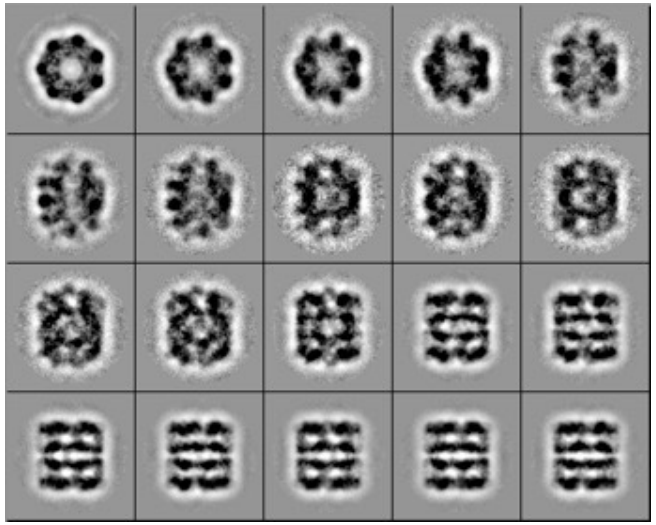




Refinement converges on high resolution reconstruction







Another approach is to use maximum likelihood (ML) methods for alignment and classification

- Statistical approach to optimize alignment and classification
- Improves accuracy and reduces overfitting
- Foundation of many modern cryo-EM software tools
- Has revolutionized high-resolution cryo-EM

Maximum Likelihood in General

Definition: In statistics, maximum likelihood estimation (MLE) is a method used to find the values of parameters that maximize the likelihood of the observed data, given a particular model.

How it works:

- **Likelihood Function:** For a given dataset and a model with certain parameters, the likelihood function calculates the probability of observing that data.
- **Optimization:** MLE aims to find the parameter values that maximize this likelihood function. This means that under the parameters determined by MLE, the observed data is the most probable (or "most likely").

Maximum Likelihood in Cryo-EM

Relevance: In cryoEM, each particle image offers a 2D projection of a 3D structure, but its exact orientation and position in 3D space are unknown. The ML approach estimates these unknown parameters.

How it works in cryoEM:

- **Projection Matching:** For each 2D particle image, the algorithm estimates which 3D orientation and position most likely resulted in that 2D projection.
- **Refinement:** Using these initial estimates, the 3D structure is refined iteratively. Each iteration involves:
  - Generating 2D projections from the current 3D model.
  - Comparing the experimental 2D images to these projections.
  - Adjusting the 3D model to increase the likelihood that it led to the observed 2D images.
- **Regularization:** To avoid overfitting (fitting noise in the data), some form of regularization is applied. This ensures the solution remains physically meaningful.

Steps in a single particle reconstruction

- Collect data
- Align frames
- Estimate CTF
- Pick particles
- 2D Classification
- Generate initial model
- Refine data against initial model
- Estimate resolution

FSC (Fourier Shell Correlation)

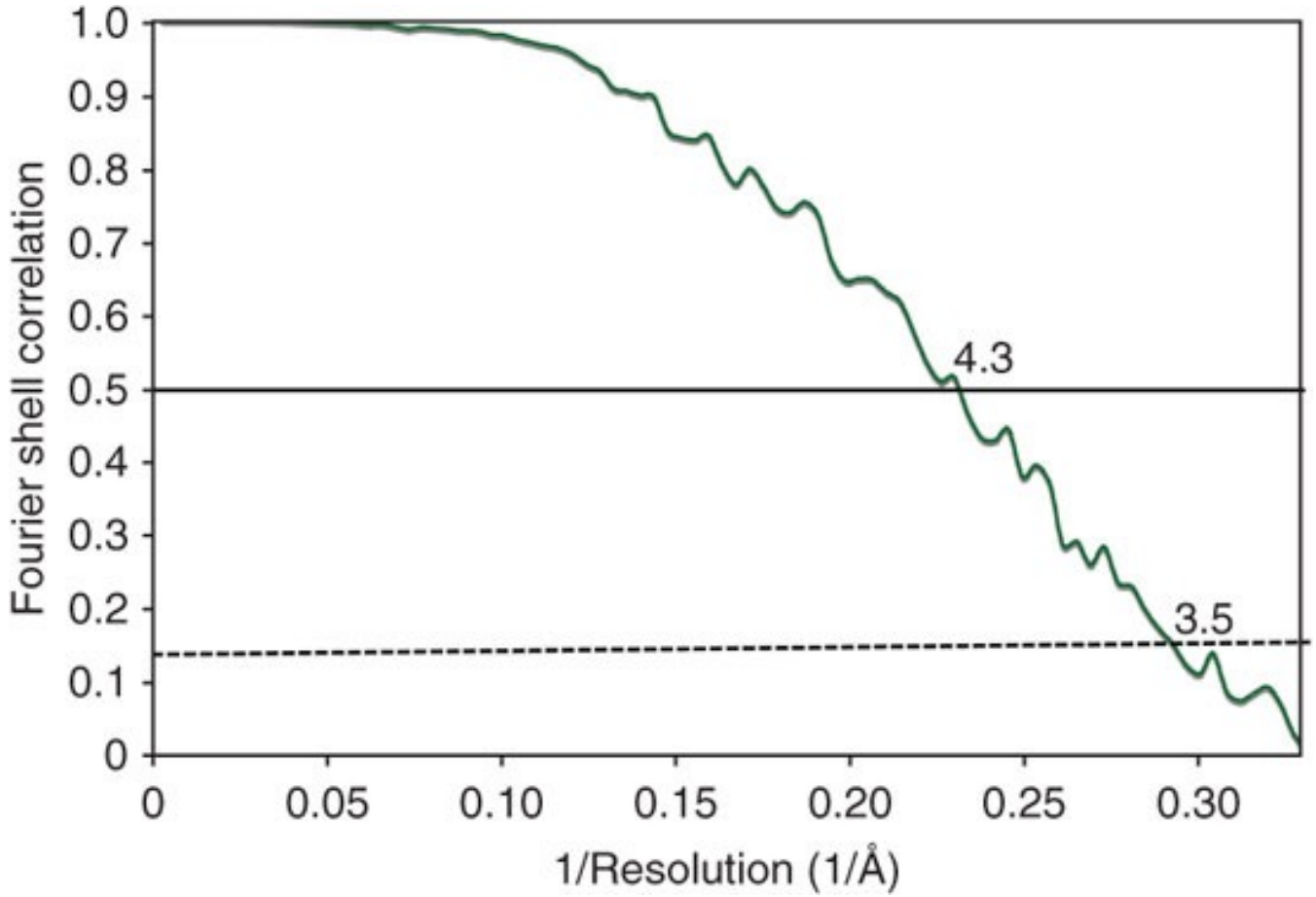
- Quantitative measure of resolution in Cryo-EM
- Compares independent half-maps for consistency
- 0.143 criterion for resolution estimation
- Essential for validating Cryo-EM structures

How to measure resolution?

- In X-ray crystallography, resolution is measured by how far diffraction goes out from origin in diffraction pattern
- No analogous criterion in 3DEM
  - Instead, we use a self-consistency criterion called the Fourier Shell Correlation (FSC)
- With FSC, we split our data into two halves, reconstruct them separately, then compute the correlation between the Fourier transforms of the two reconstructions at discrete resolution shells

An FSC curve

- Assign resolution at a cutoff correlation value
  - $FSC_{0.143}$ is the standard, but there are other metrics as well



The Southeastern Center for Microscopy of Macromolecular Machines ($SECM^4$) at FSU

SECM^4 services

- Vitrified grid screening
  - You send grids, we screen to find grids that will enable structure determination
- High-resolution data collection
  - You send *screened* grids, we collect thousands of images
- Sample preparation and screening *
  - You send *screened* sample, we freeze and screen grids
  - Use special vitrification technology
- Training
  - You come on-sight, we provide practical cryo-EM training
  - Focus on practical aspects, many other resources available for training in theory

There is no cost for using the facility

How to access the center

- Access through secm4.org website
- Two stages
  - Stage 1
    - Register on website
    - Apply for access
      - Project-focused
    - Proposal goes out for review
  - Stage 2
    - Accepted proposals can then request services

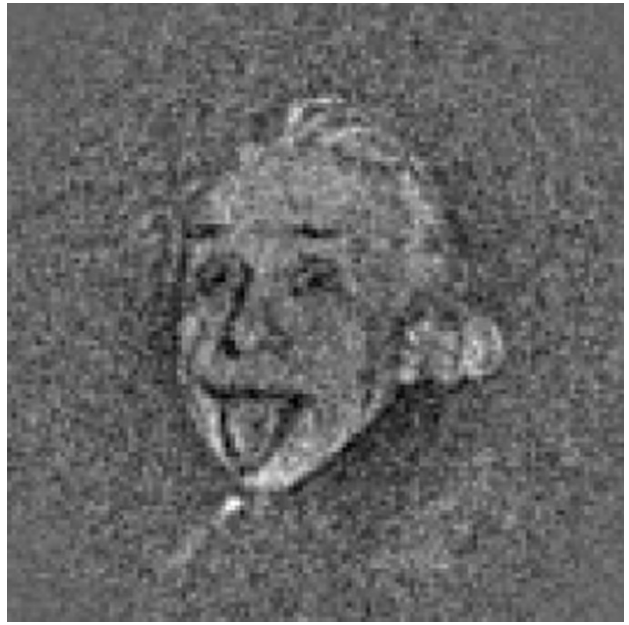
New digs


Mechanical
0119S
Specimen Prep (LH)
0119R
Corridor
0119M(N)
Specimen Prep
0119G
Gas
0119J
CryoTEM Control
0119L
Office #3
0119F
Corridor
0119M
TEM Room
0119Q
Light Microscopes
0119K
Shared Office
0119B
Corridor
0119
Corridor
0119M(S)
Office #1
0119C
Office #2
0119D
cryoFIB Room
0119N
Confocal Room
0119A
Auxiliary / Pump
0119P
Titan UPS
0116A
Wet Lab
0116
Eye Wash/
Shower

Questions?

Single particle caveats

- Initial model bias
  - Since the images are so noisy, it is possible to regenerate an image of a reference from aligning pure noise



It is possible to get a “high resolution” structure from a completely incorrect reconstruction

Figure 3