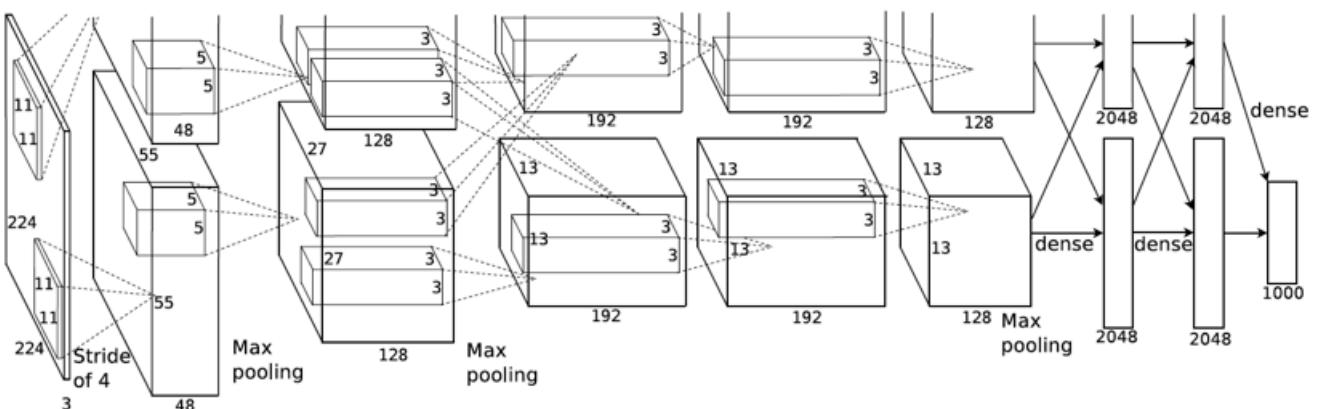
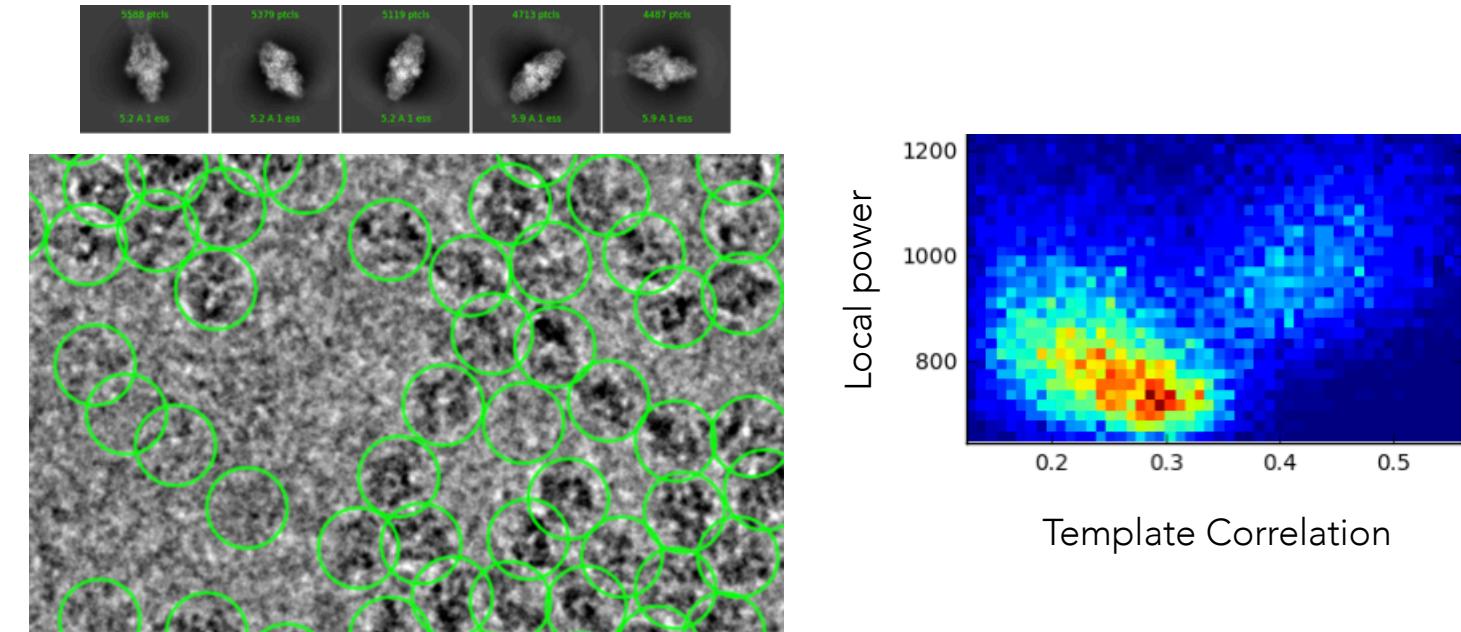


Particle Picking

Semi-automatic

- Template-free
 - “blob” picking
 - Unbiased, least discriminative
- Template based
 - 2D classes or 3D views
 - False-positives, bias
- Deep Neural Networks
 - Require manual picking and training per-dataset
 - Human bias



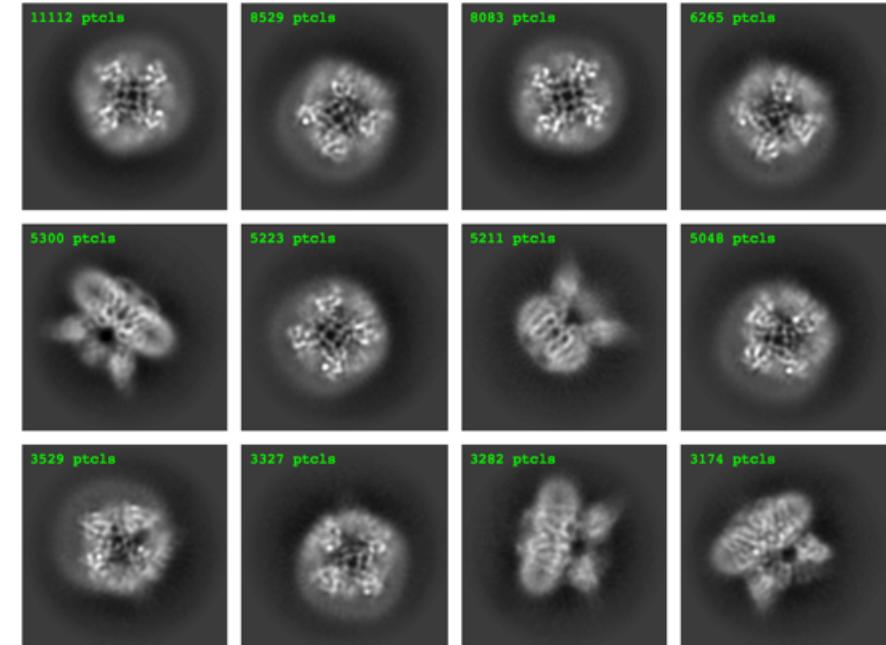
Practical: Particle picking

1. Create a blob picking job
 1. Only 5 micrographs ; 100 Å minimum ; 200 Å maximum
2. Inspect particle picks
 1. Threshold NCC score ~0.22 ; power score 600 – 1000
3. Extract from micrographs
 1. Box size 440; Fourier-crop size 256
4. 2D classification with 20 classes
5. Select 2D top and side view
6. Template picking job
 1. 200 Å particle diameter
7. Inspect particle picks
 1. Threshold NCC score ~0.34 ; power score 1100 – 1700
8. Extract from micrographs
 1. Box size 440; Fourier-crop size 256

2D Classification

Inspect and curate massive datasets

- Remove junk particles and improve homogeneity of particle stacks
- Online-EM algorithm
- Branch-and-bound alignment
- Probabilistic classification
- Scales well with dataset size
- New implementation improvements and tuning: large performance boost

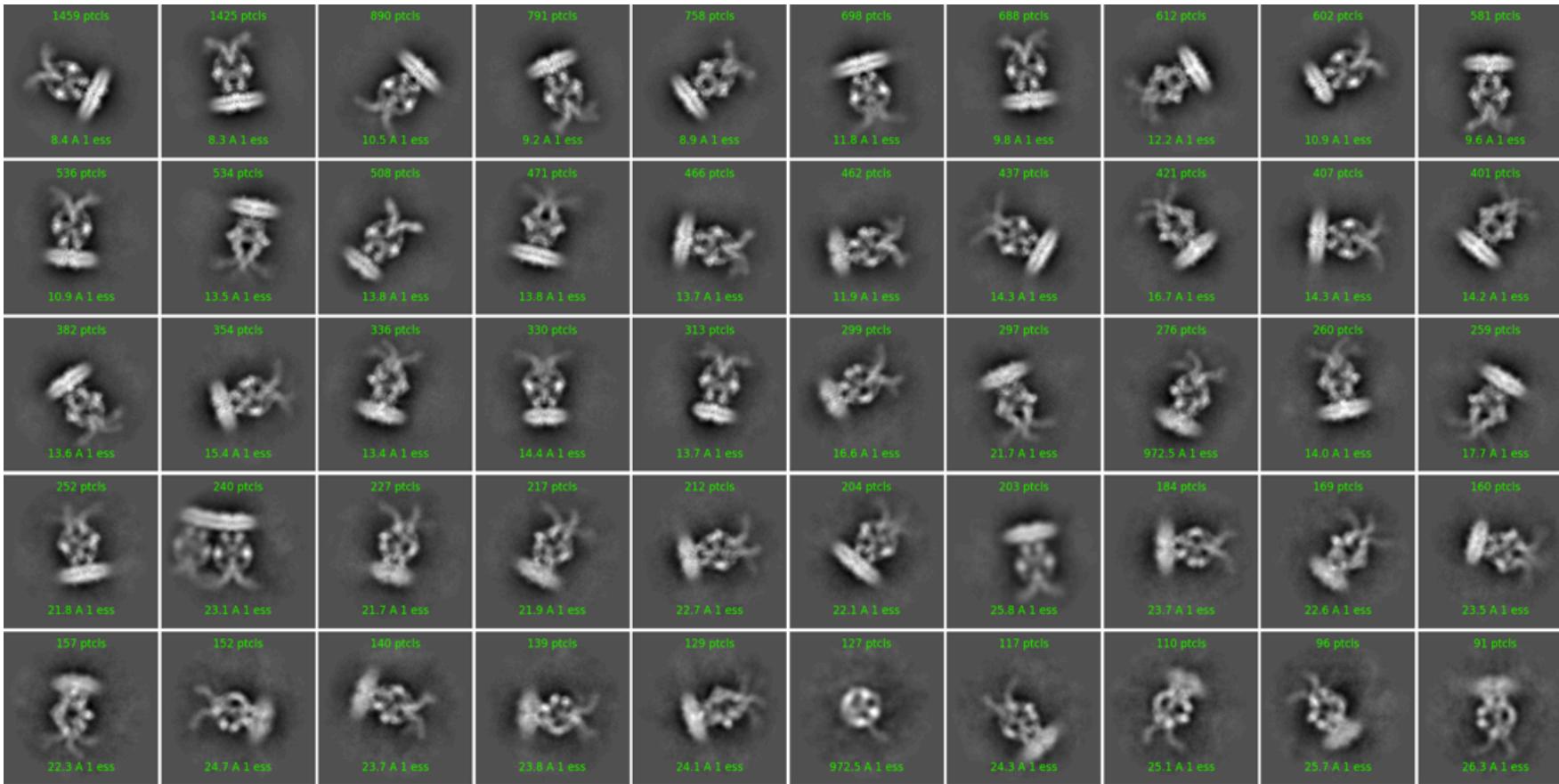


Practical: 2D classification

1. Create a 2D classification job with 50 classes
2. Create a 2D classification job with 100 classes

2D Classification

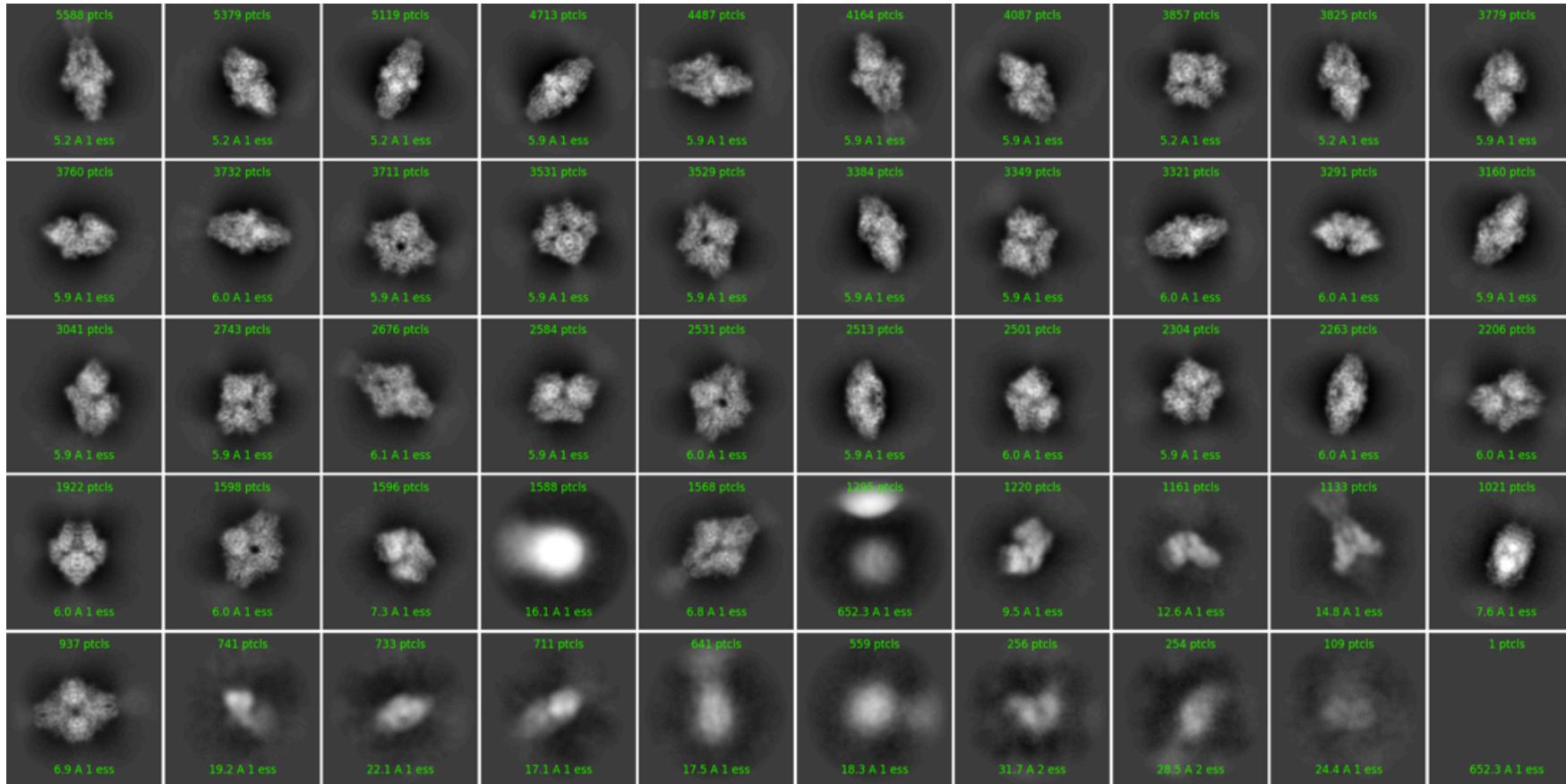
125,000 particles with 50 classes in **8 minutes** on 1x GPU



2D Classification

125,000 particles with 50 classes in **8 minutes** on 1x GPU

1,000,000 particles with 50 classes in **1 hour** on 1x GPU

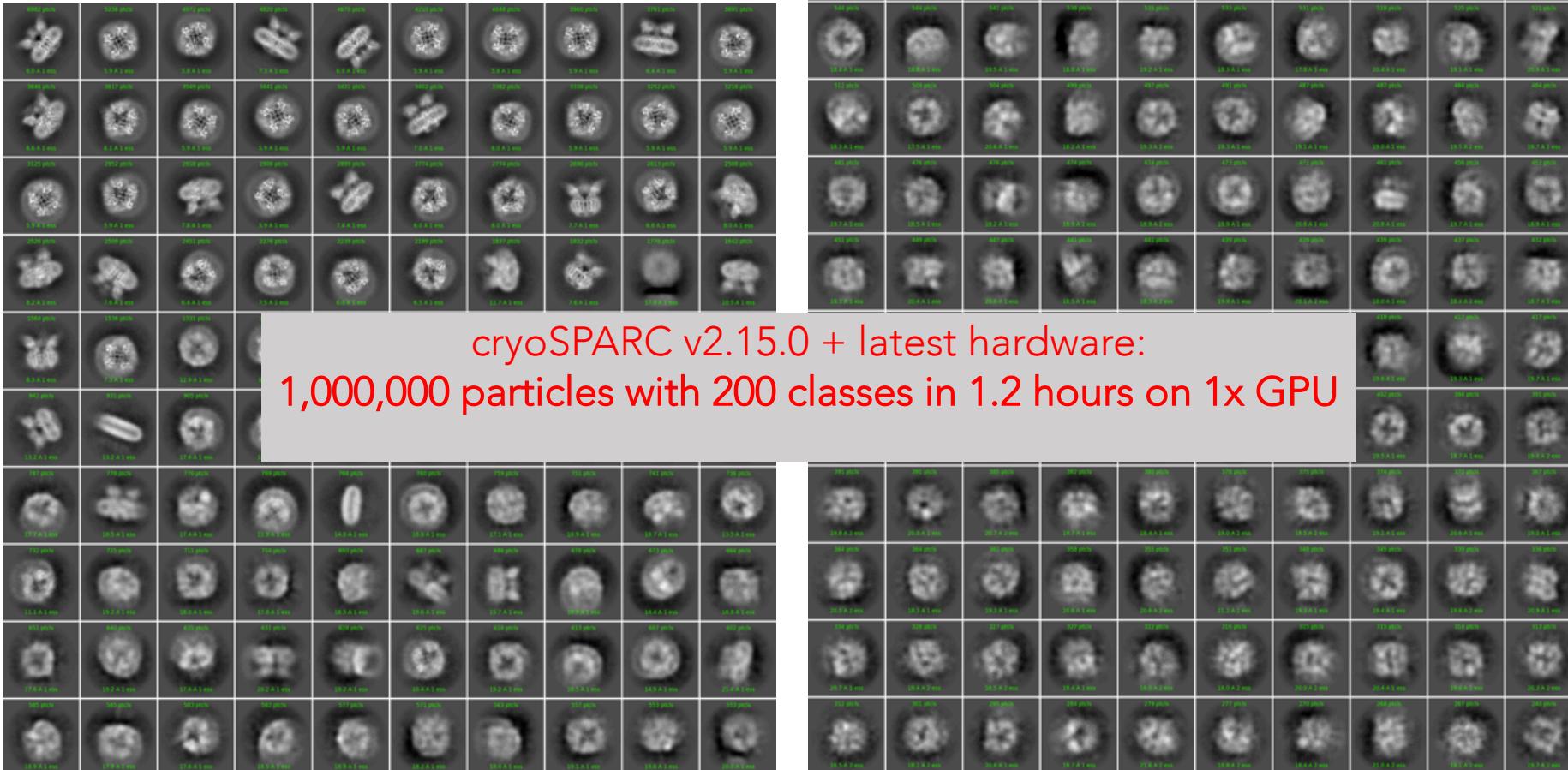


2D Classification

125,000 particles with 50 classes in **8 minutes** on 1x GPU

1,000,000 particles with 50 classes in **1 hour** on 1x GPU

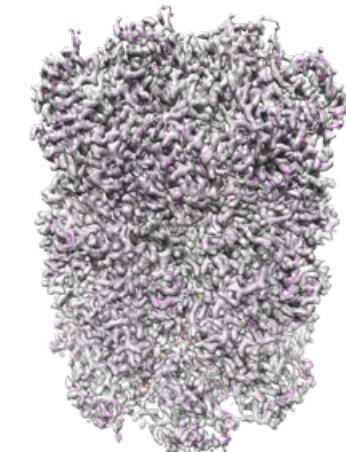
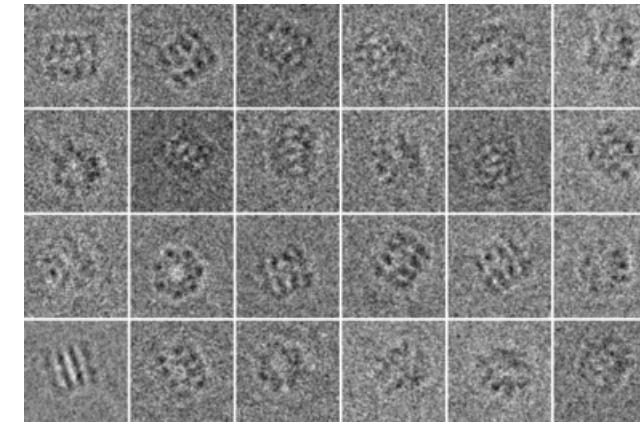
1,000,000 particles with 200 classes in **2 hours** on 1x GPU



3D reconstruction in single particle cryo-EM

What makes the problem difficult

- Unknown pose of each particle
 - 3D Orientation + 2D Shift
- High noise level
 - Irreducible due to beam damage
- Many particles (100,000+)
- Corruption by microscope contrast transfer function
- Multiple conformational states or distinct particles
- Flexibility and disorder



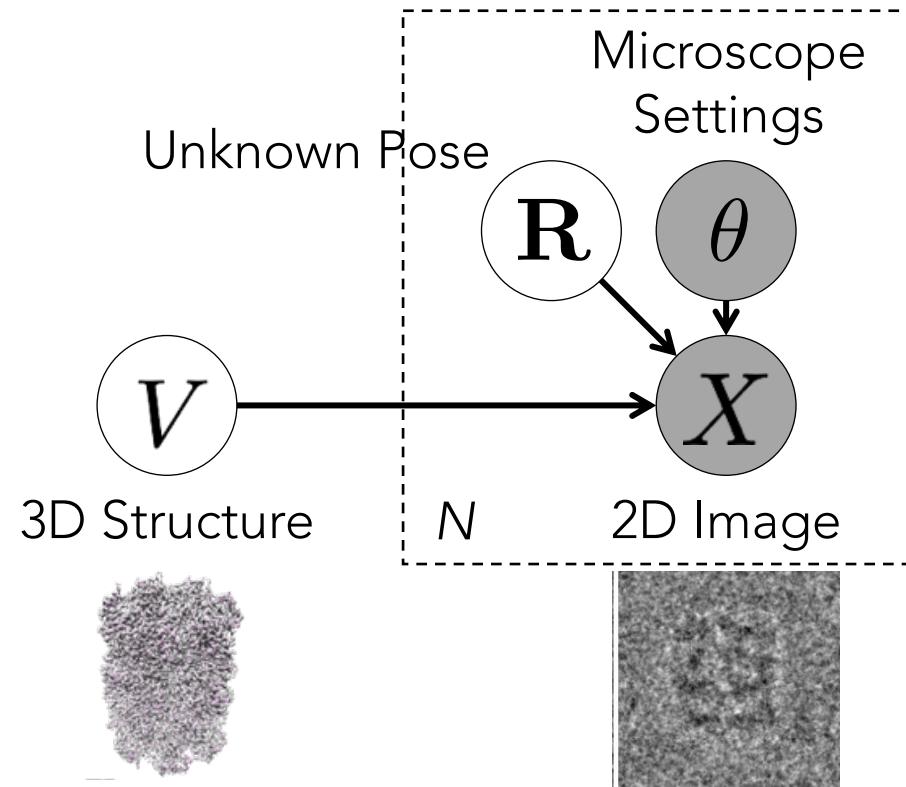
Probabilistic foundations

Probabilistic graphical model

- Fully “Bayesian” treatment:

$$p(V|X_1, X_2, \dots X_N)$$

- Meaning we want to know all the possible 3D structures that could explain the images
- Arrow from V to X is the image formation model



Probabilistic foundations

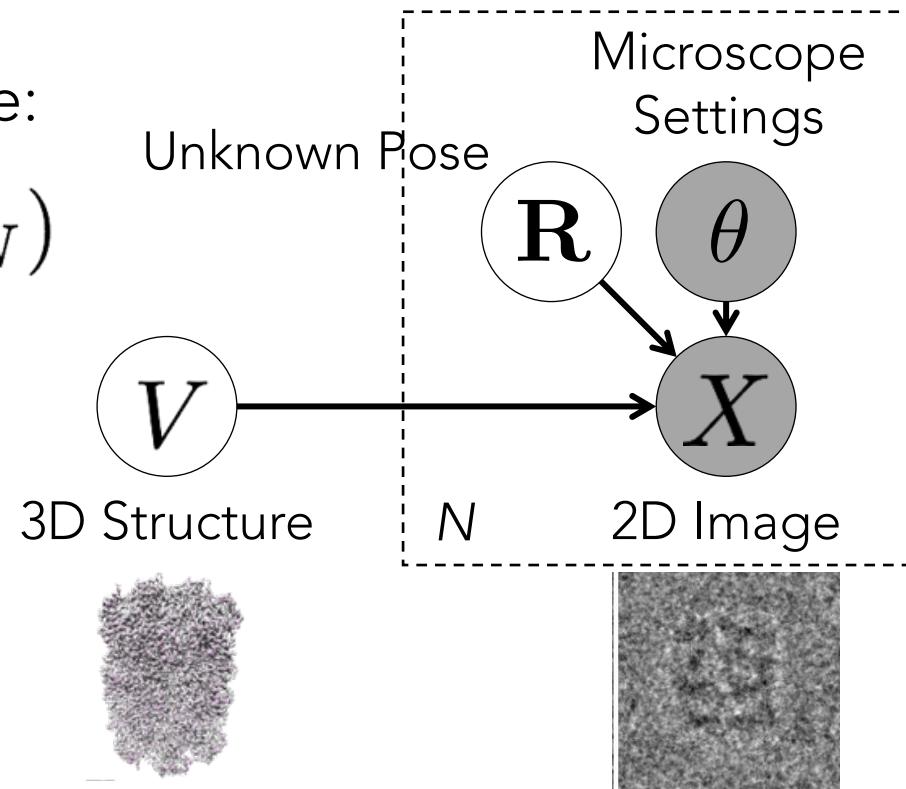
Making inference computationally feasible

- Maximum probability estimate:

$$\max_V p(V|X_1, X_2, \dots X_N)$$

- Now only looking for the single estimate of V that best explains the data

- Different choices of V yield all SPA techniques



Probabilistic foundations

Various choices of V

- Homogeneous refinement:

$$V = V$$

- Discrete Heterogeneity:

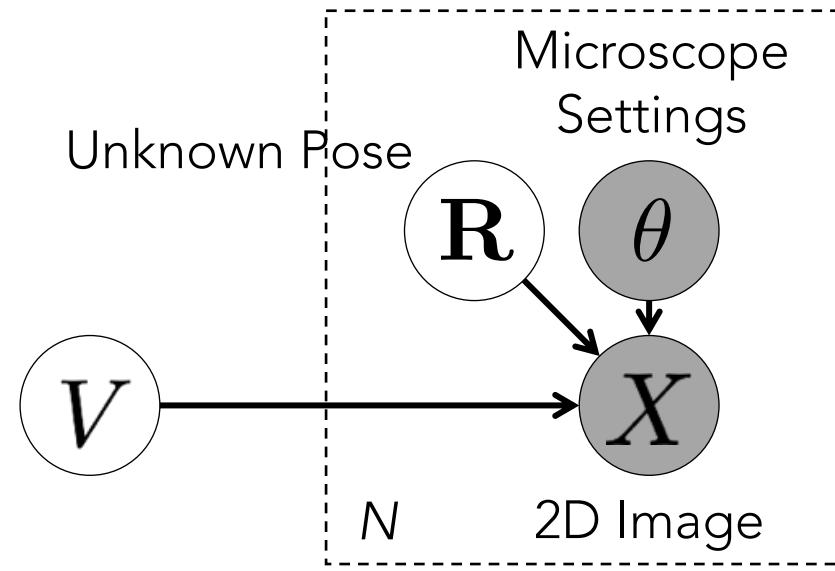
$$V = V_k | k = C_i$$

- Local refinement/multiple rigid bodies:

$$V = V_1 + R_{rel}(V_2)$$

- In all cases, need to solve an optimization problem:

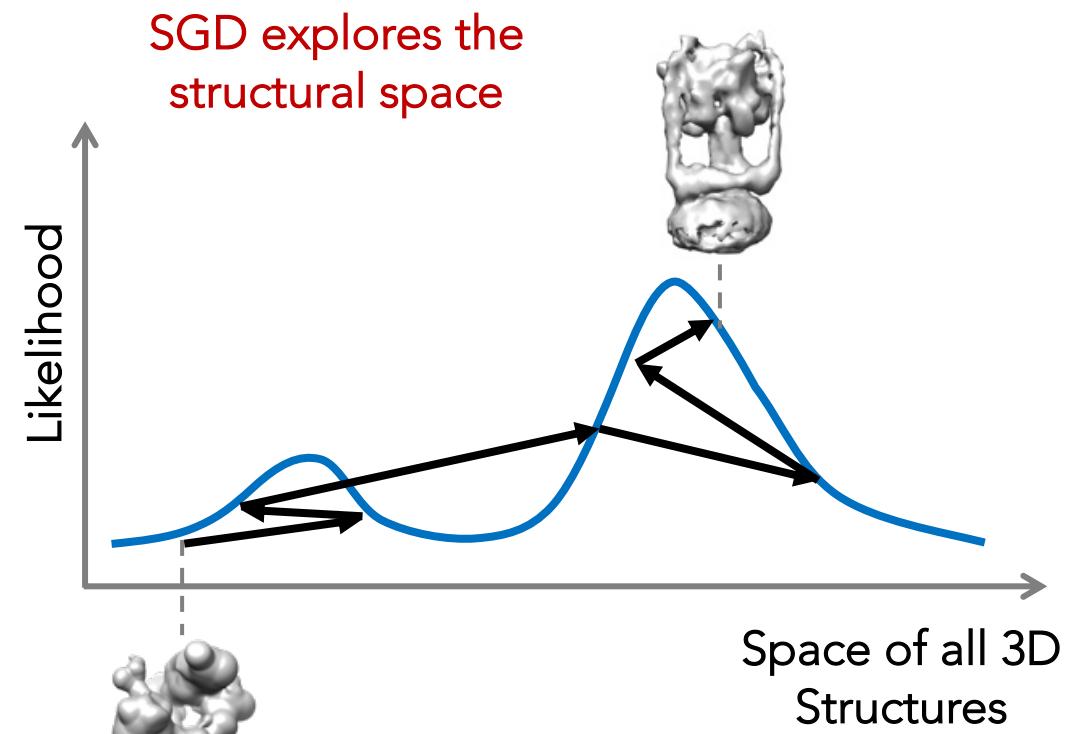
$$\max_V p(V|X_1, X_2, \dots X_N)$$



Ab-initio reconstruction: SGD

- Initial model problem
- Stochastic optimization
 - Class of modern statistical methods
- Stochastic Gradient Descent (SGD)
 - Very successful variant
 - Powers modern deep learning
- Allows unbiased reconstruction from particles only

$$\max_V p(V|X_1, X_2, \dots X_N)$$



Stochastic optimization

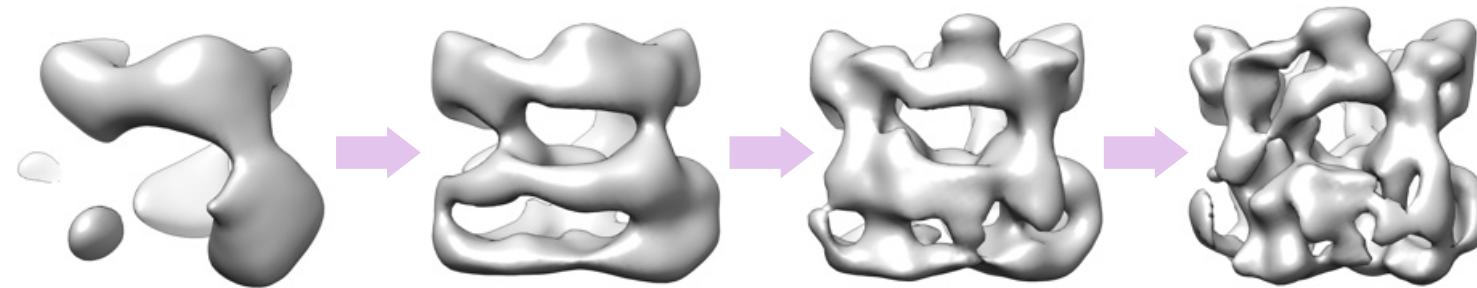
Stochastic gradient descent

$$\max_V p(V|X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, \dots X_N)$$

Iter 1 $\max_V p(V|X_2, X_4, X_5, \dots)$ Randomly select small subsets of images at each iteration

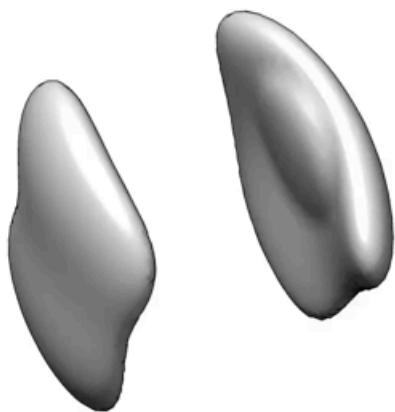
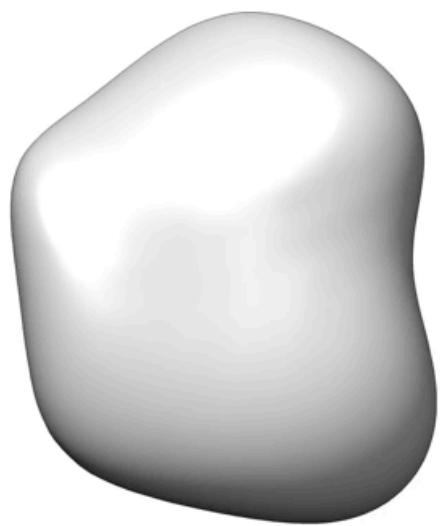
Iter 2 $\max_V p(V|X_1, X_3, \dots, X_7, \dots)$

Iter 3 $\max_V p(V|X_6, X_8, X_N)$



Many **noisy** incremental changes

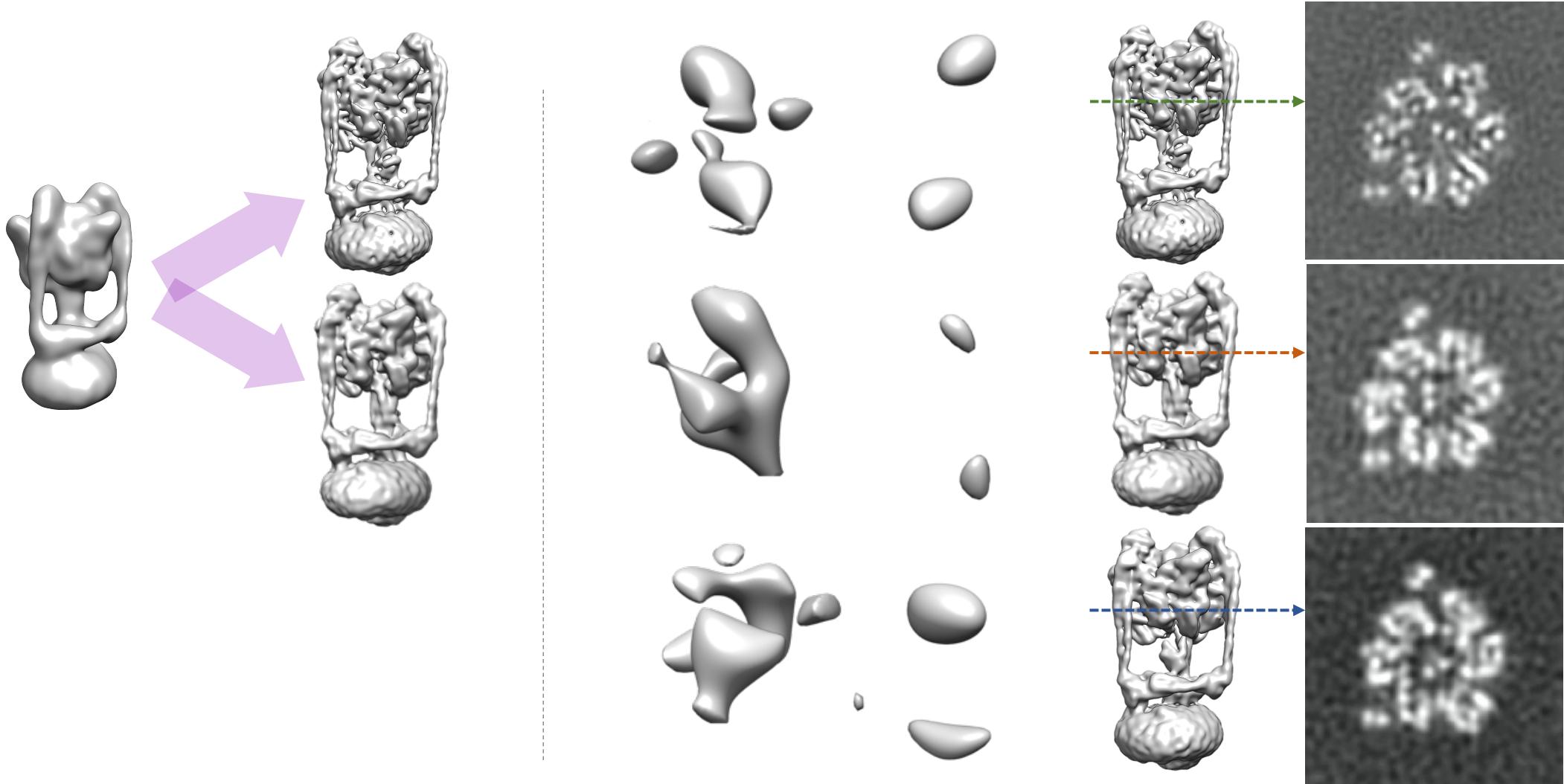
Ab-initio reconstruction: SGD



Practical: Ab-initio reconstruction

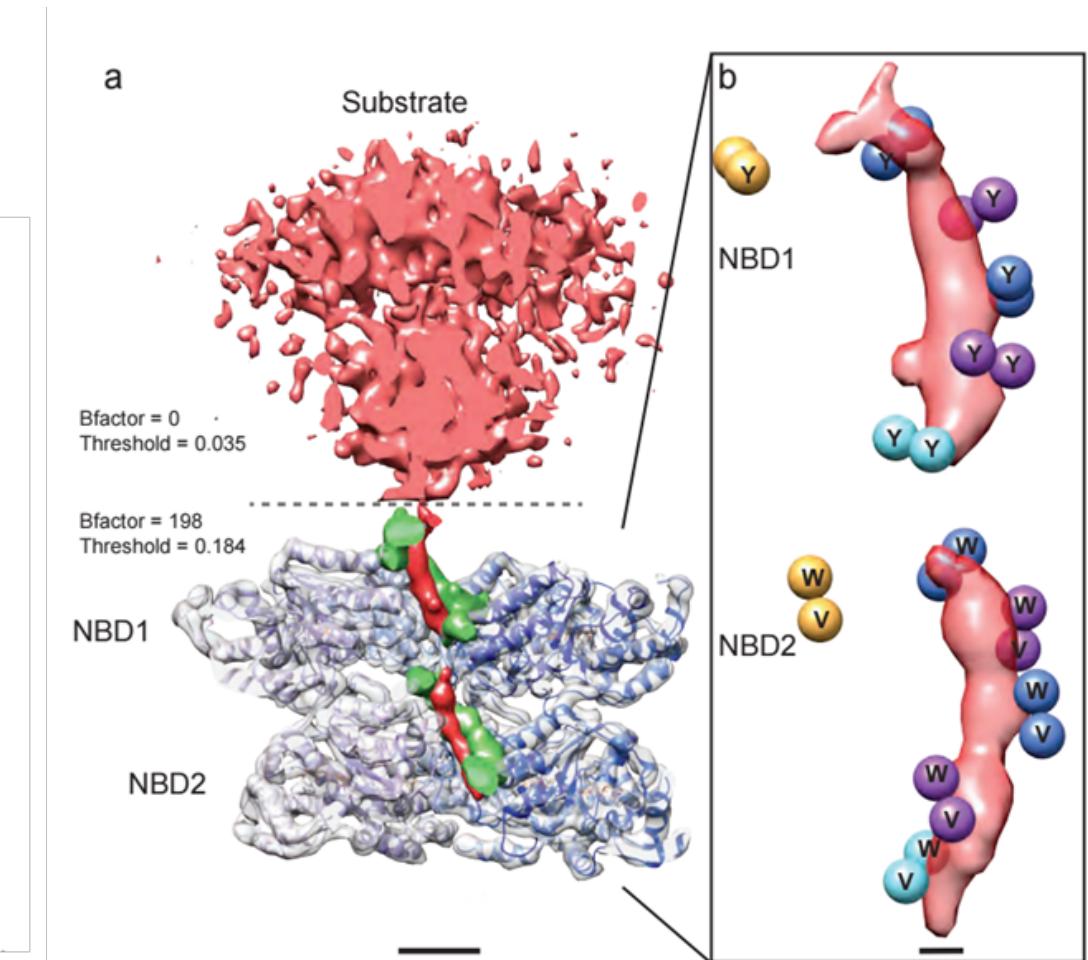
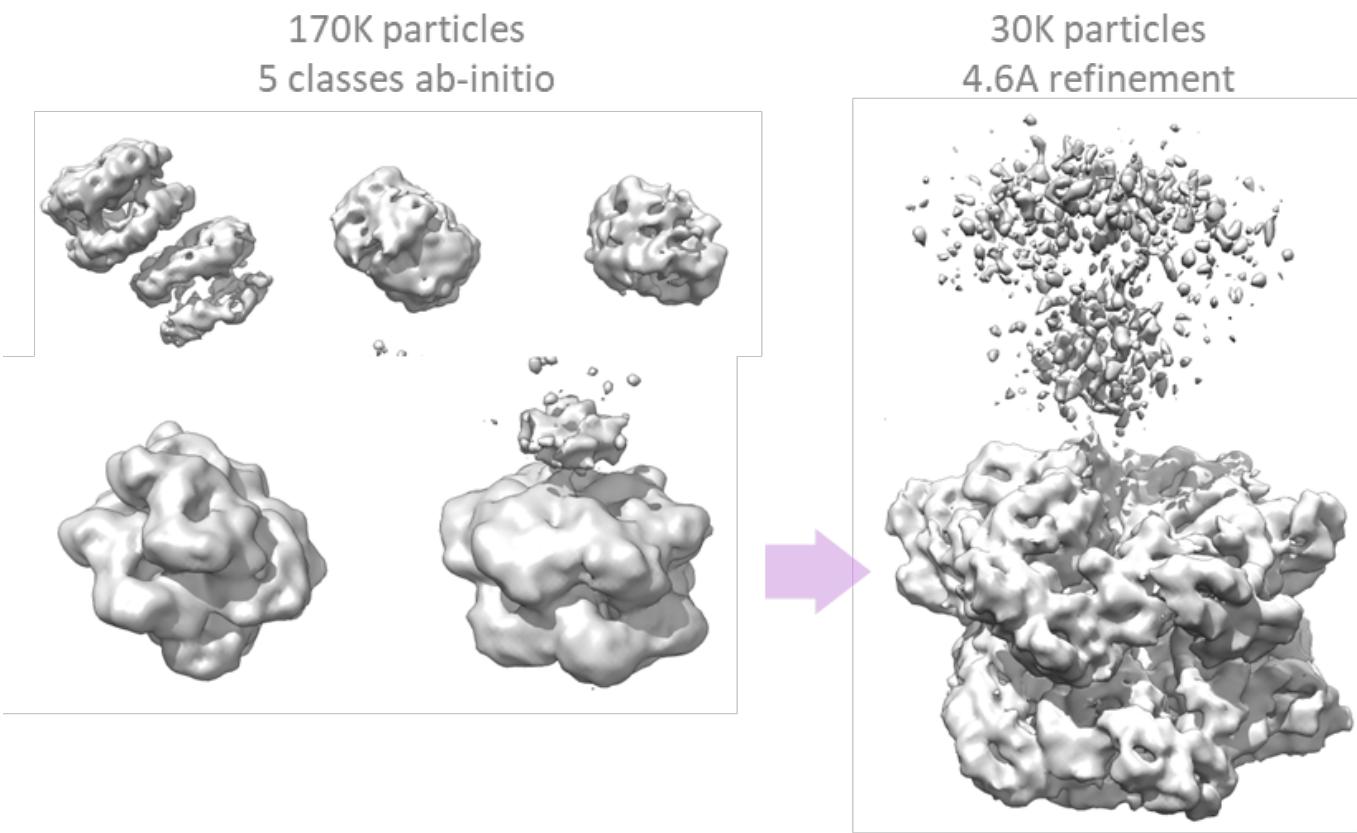
1. Create a 1 class ab-initio reconstruction job
2. Create a 2 class ab-initio reconstruction job

Heterogeneous samples: ATPase

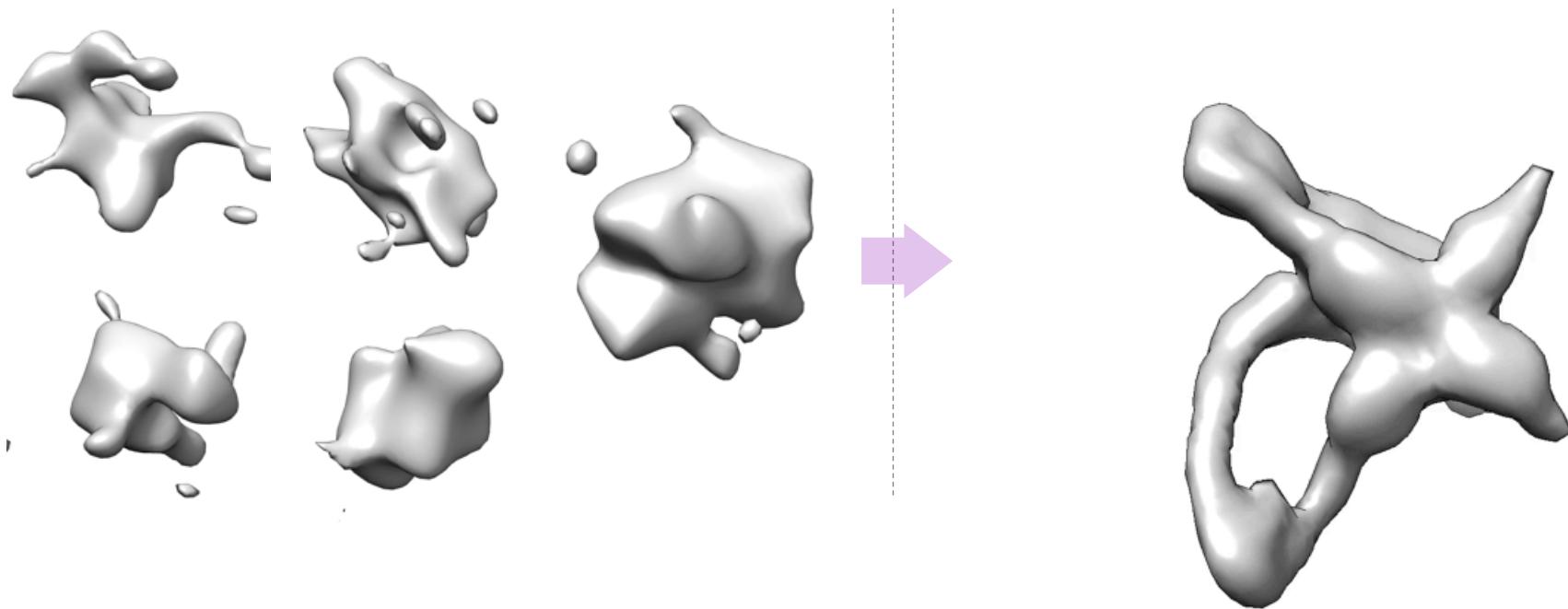


Data from Schep et al. PNAS 2016

Ab-initio reconstruction: AAA+ Unfoldase

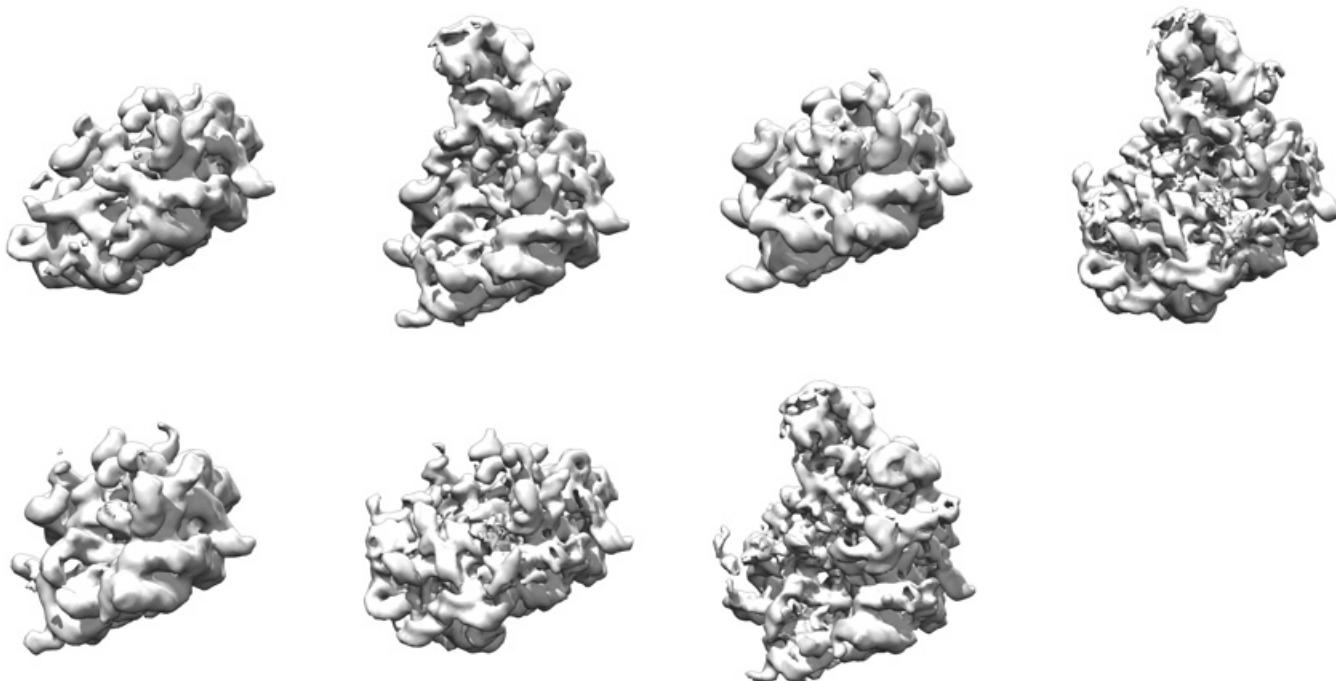


Heterogeneous samples: Holiday Junction

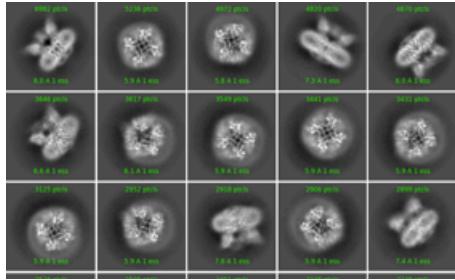


Heterogeneous Refinement (3D classification)

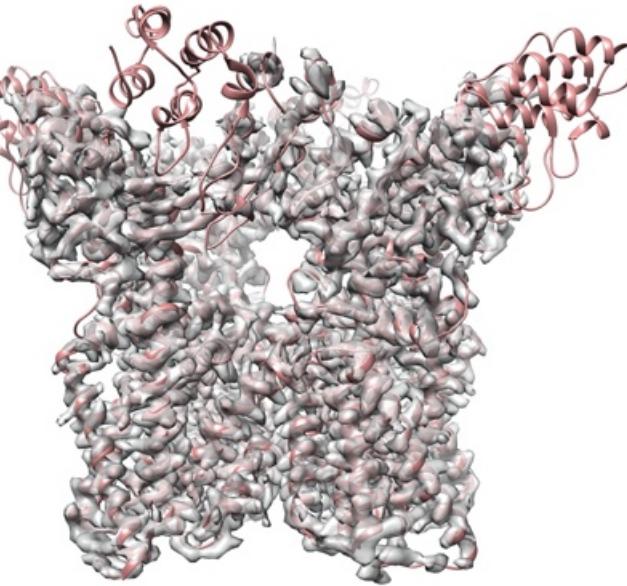
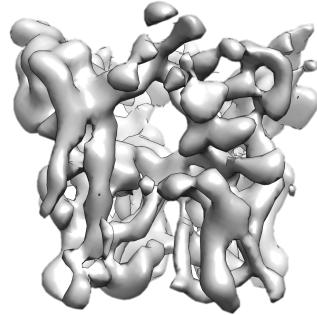
- One or more starting models
(eg. from ab-initio reconstruction)



High-resolution refinement



+

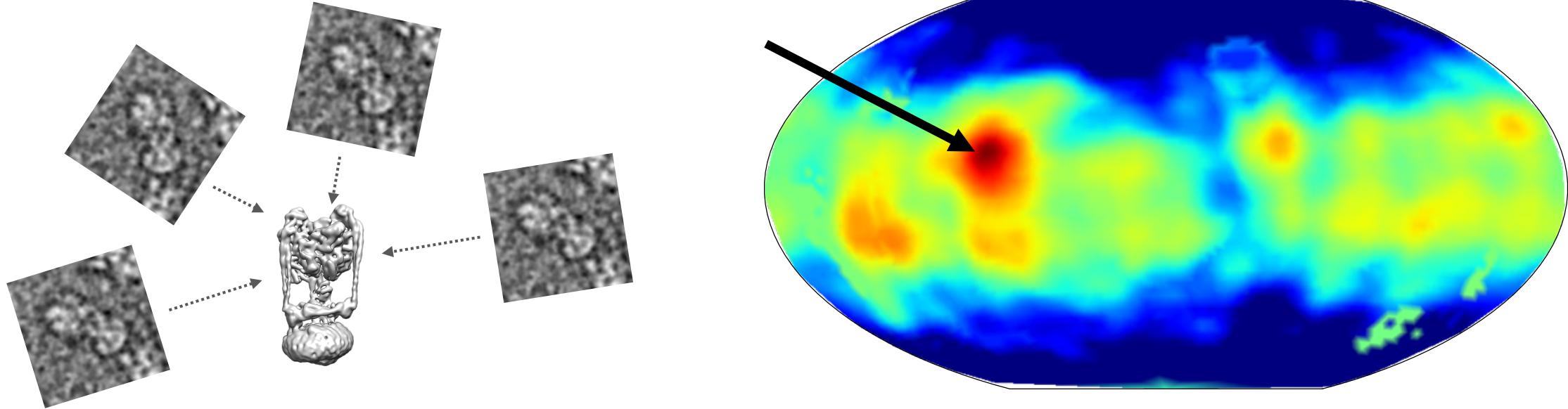


- Ab-initio reconstruction is coarse, low-resolution
- Iterative refinement can proceed to recover high-resolution detail
- Usually **very computationally expensive:** hours to days of computation
- Many refinements in a complex workflow

Practical: Refinement

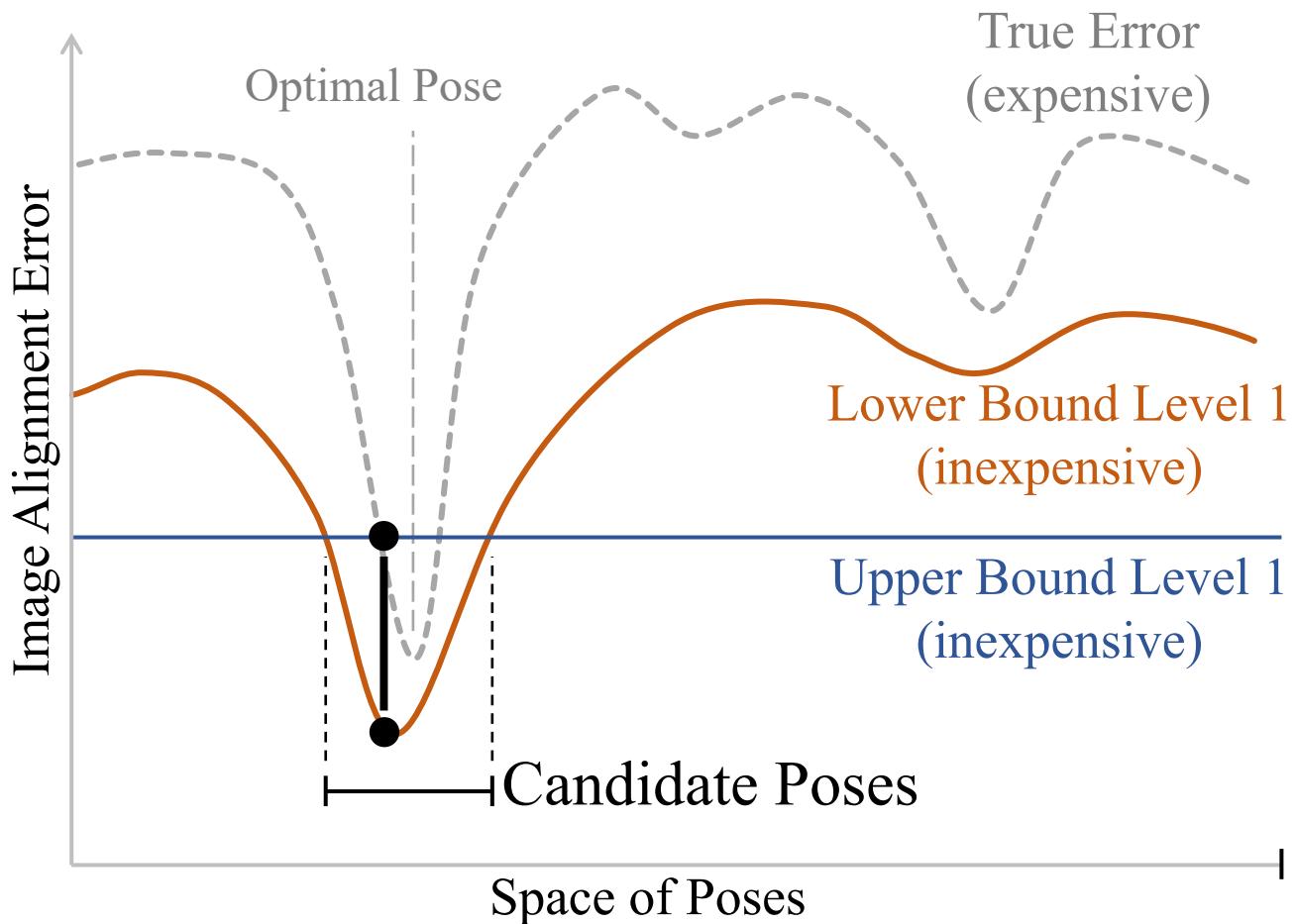
1. Kill the 2 class ab-initio reconstruction job
2. Create a “homogeneous refinement (NEW!)” job
 1. C1 symmetry
3. Create a “homogeneous refinement (NEW!)” job
 1. D7 symmetry

2D-to-3D Image alignment

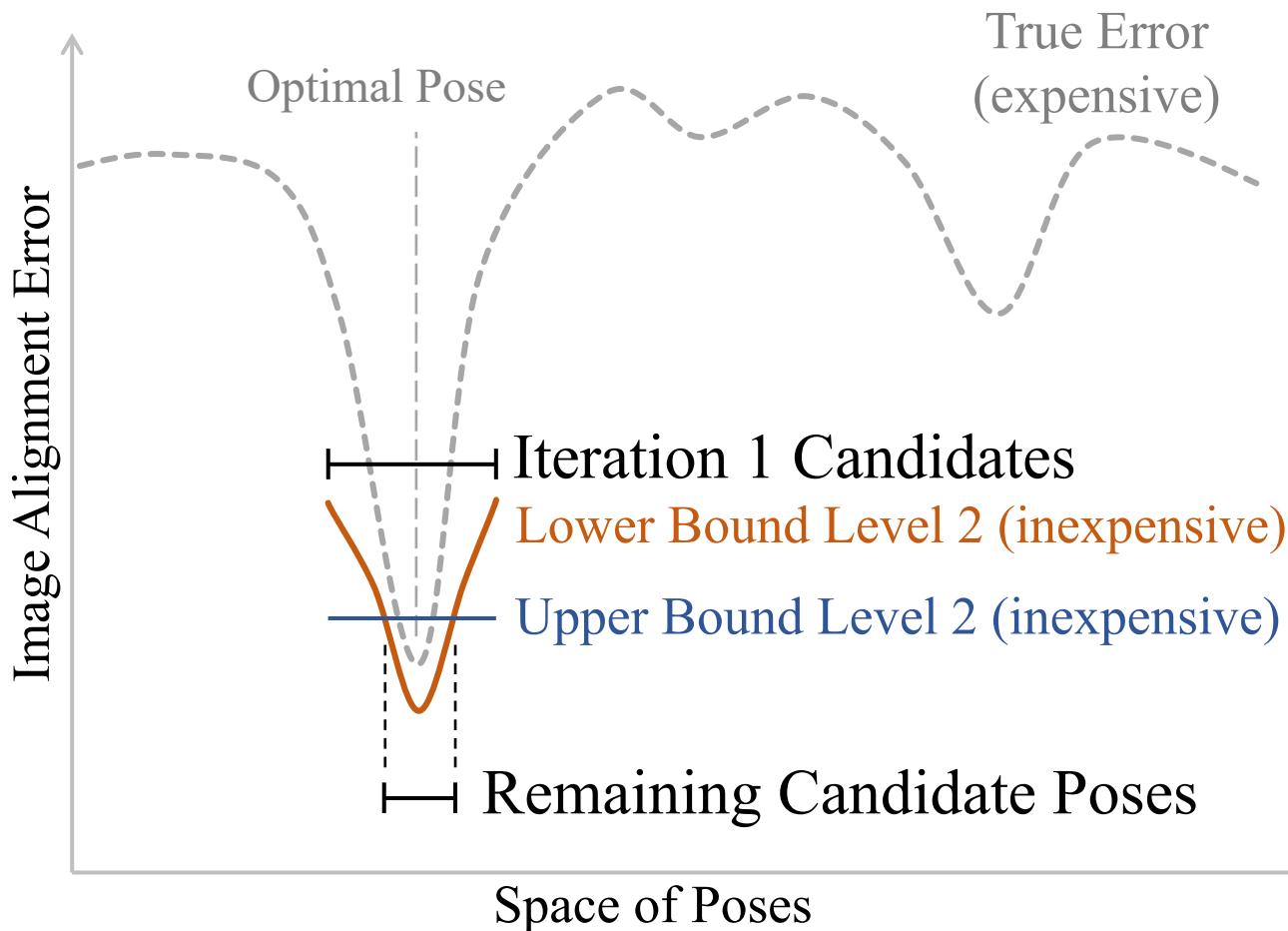


- 5D pose search is expensive, for every image
- Existing techniques search exhaustively or locally

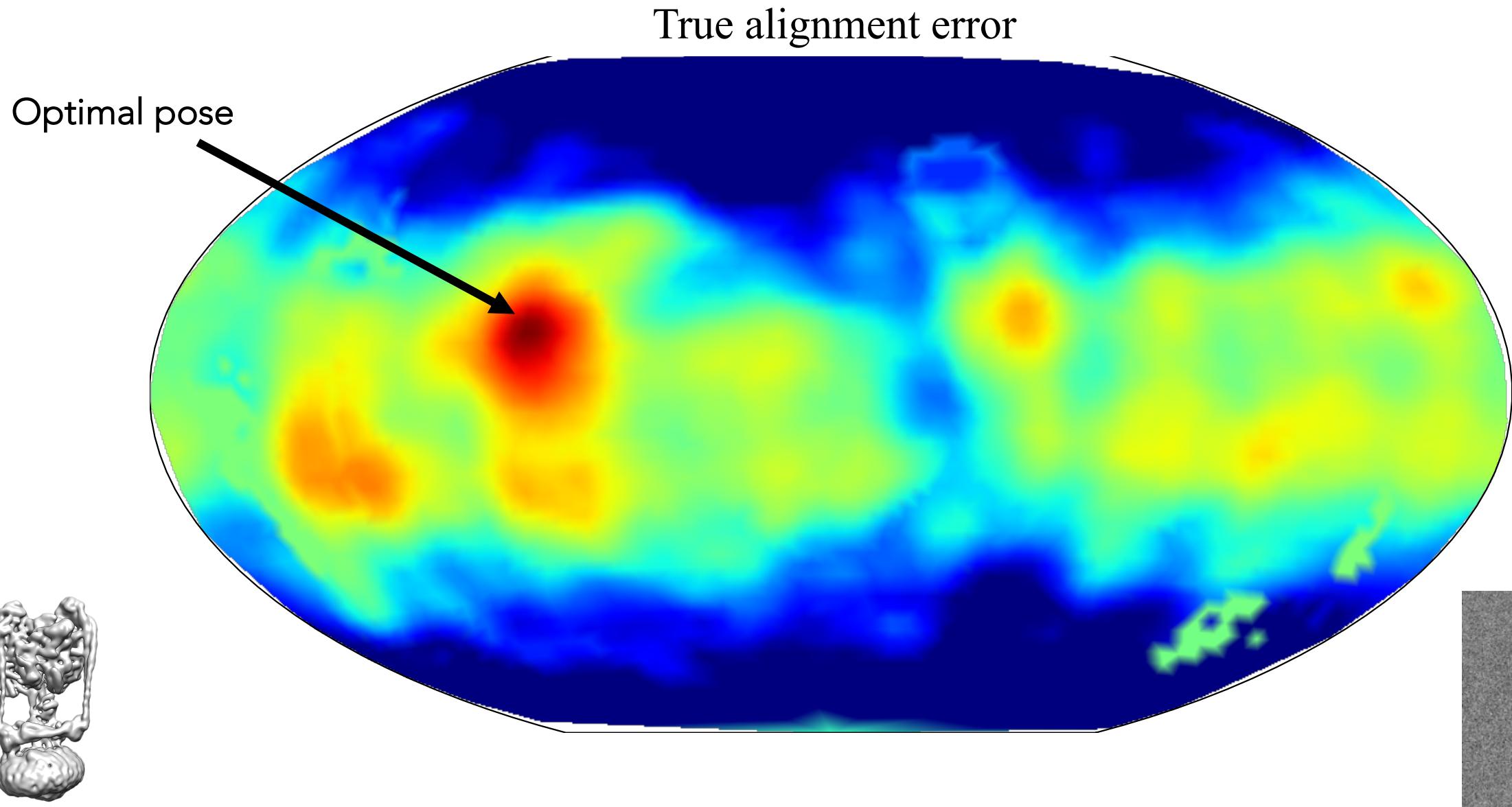
Branch and bound method



Branch and bound method

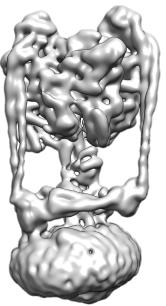
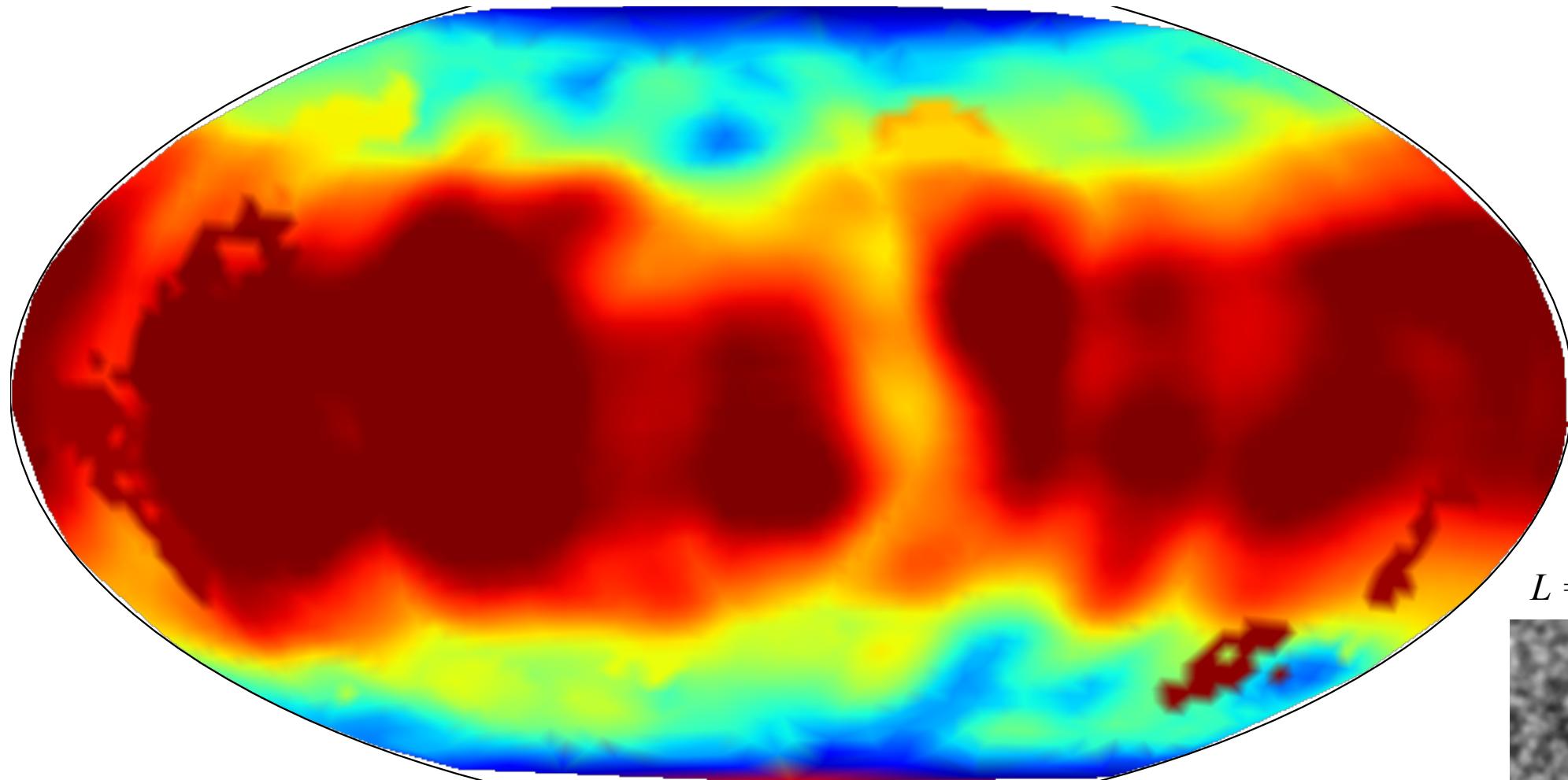


Branch and bound method

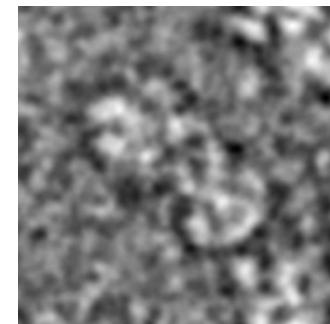


Branch and bound method

Bound on error

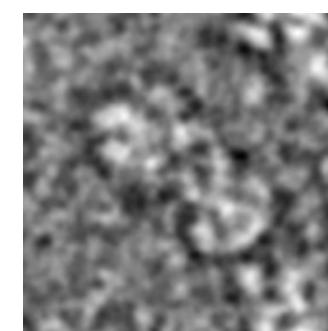
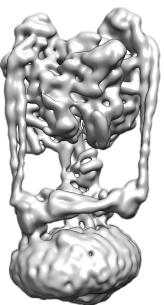
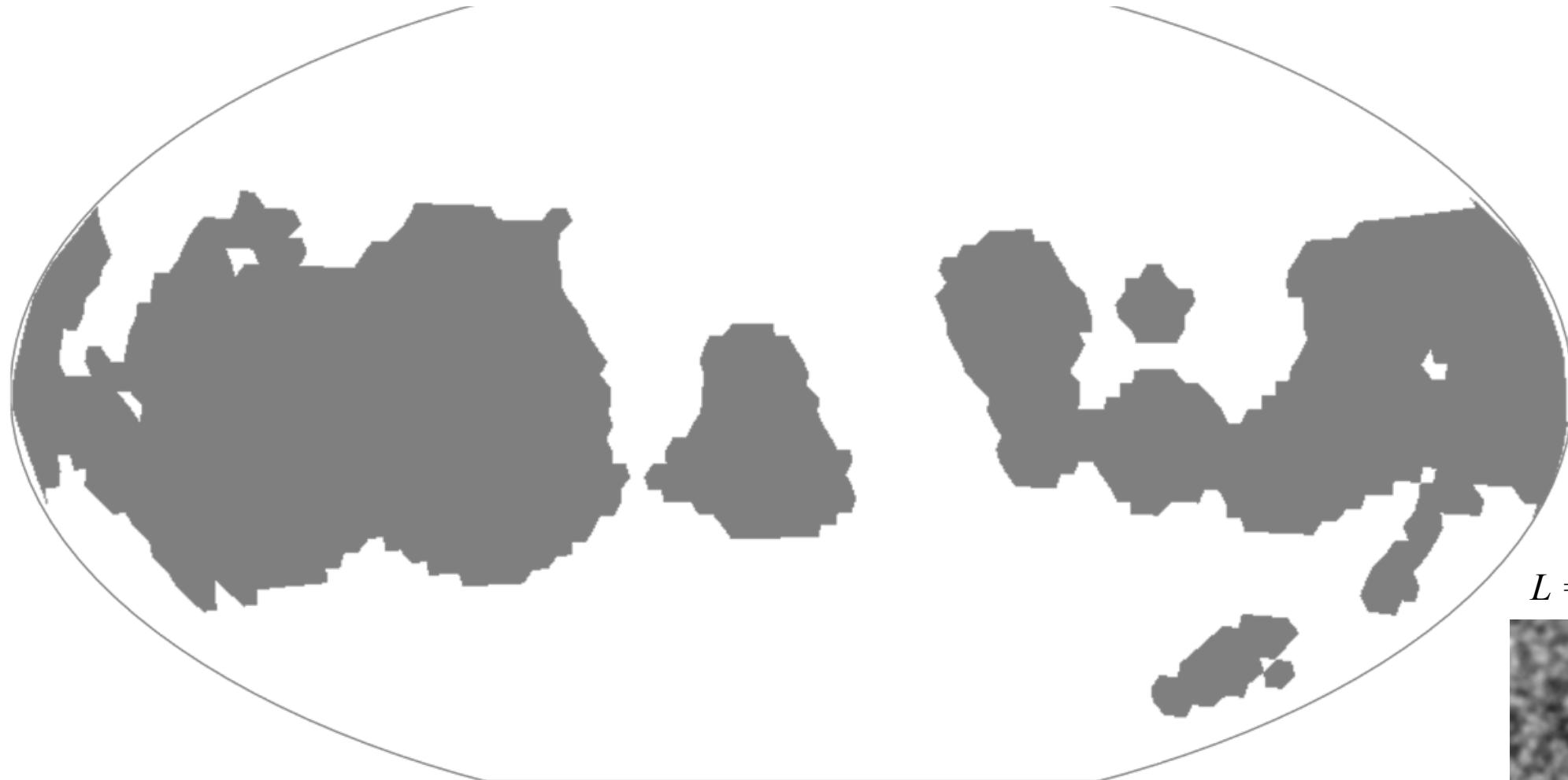


$L = 15$ (24Å)



Branch and bound method

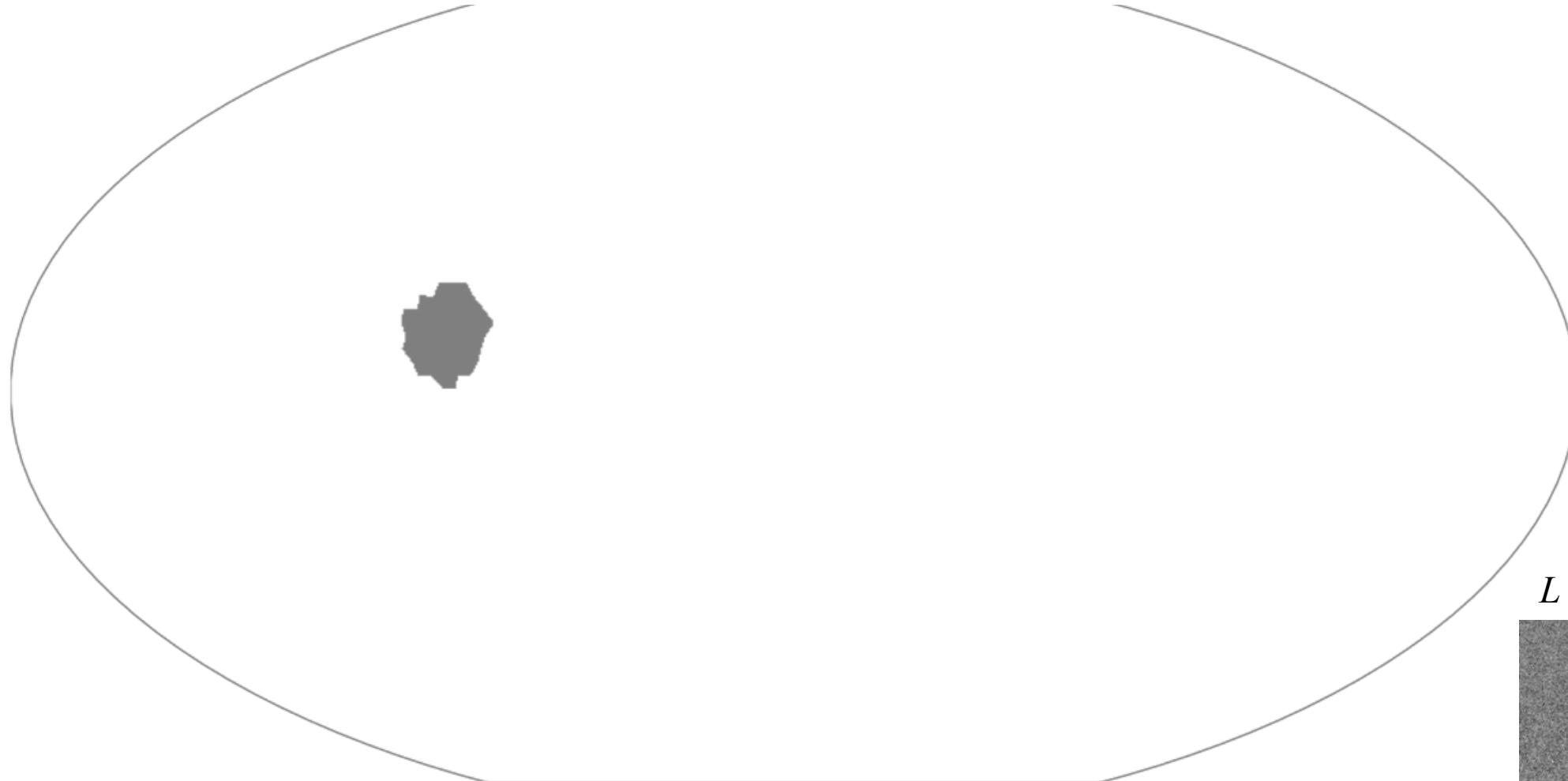
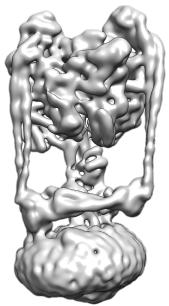
Remaining candidates at $L = 10$ (37Å)



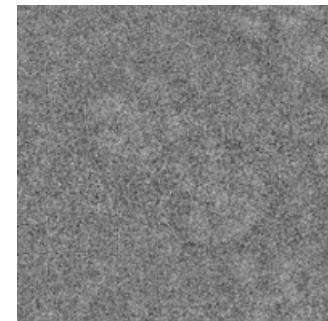
$L = 15$ (24Å)

Branch and bound method

Remaining candidates at $L = 15$ (24Å)

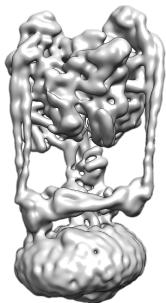


$L = 20$ (14Å)



Branch and bound method

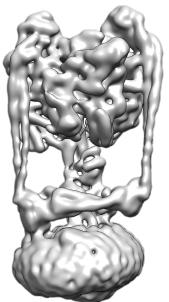
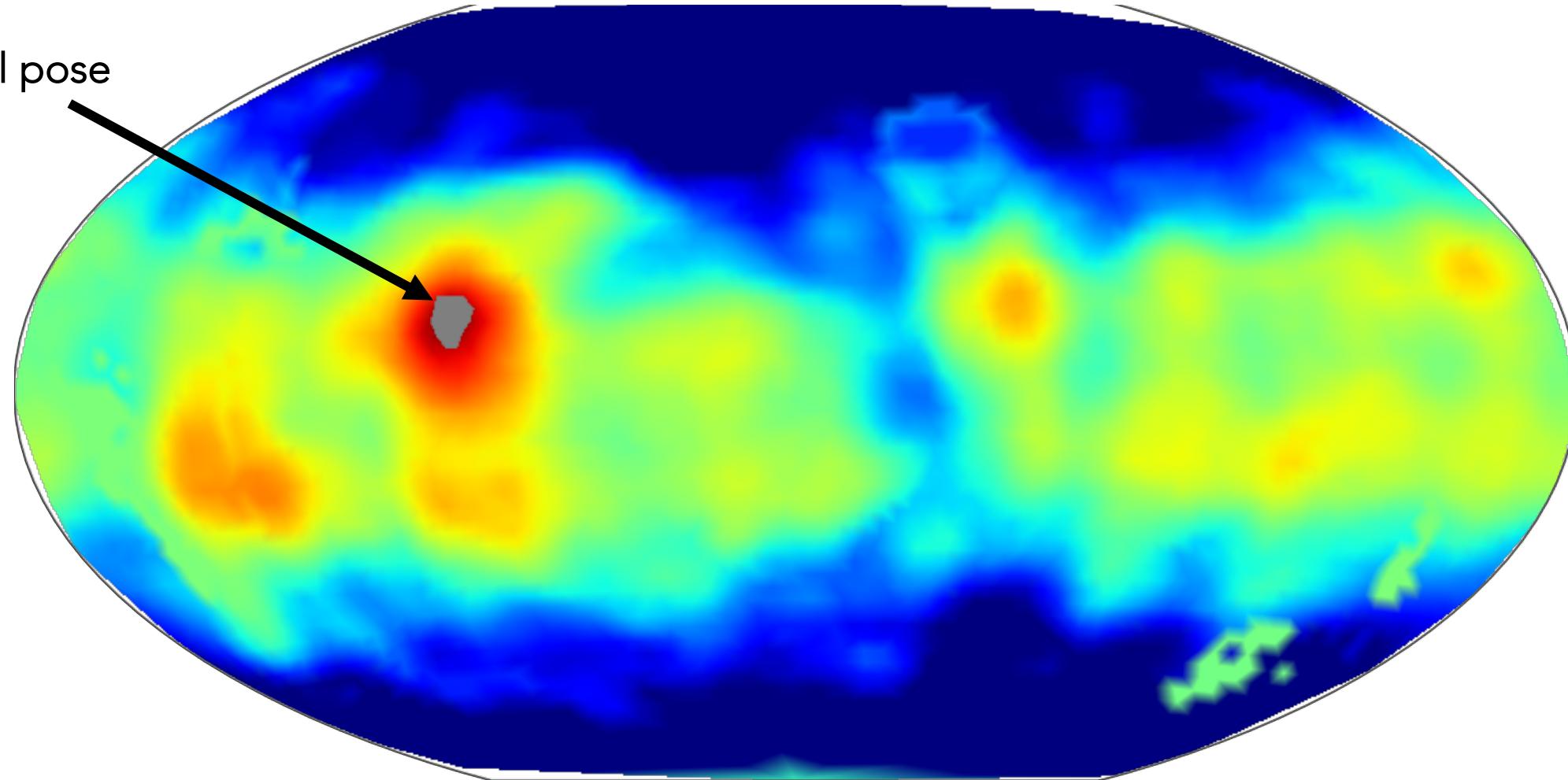
Remaining candidates at $L = 20$ (14Å)



Branch and bound method

Optimal pose is within remaining candidates

Optimal pose

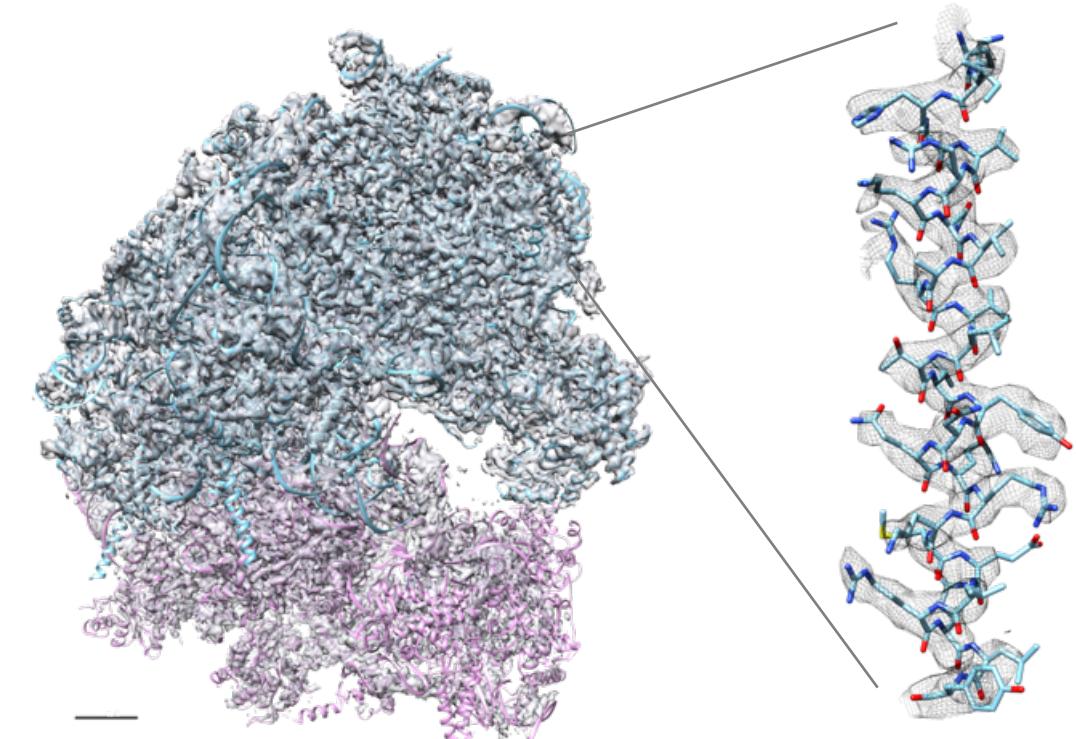


High speed, high-resolution refinement

- Iterative refinement is very expensive – multiple rounds of alignment and reconstruction required
- Branch and bound (BnB) algorithm drastically reduces computation required without any loss in quality
- High-resolution structures in minutes
- Refinements are repeated dozens of times in a real cryoEM workflow

80S ribosome dataset:
105,000 particles refined to 3.2Å

11 min on 1x NVIDIA GPU

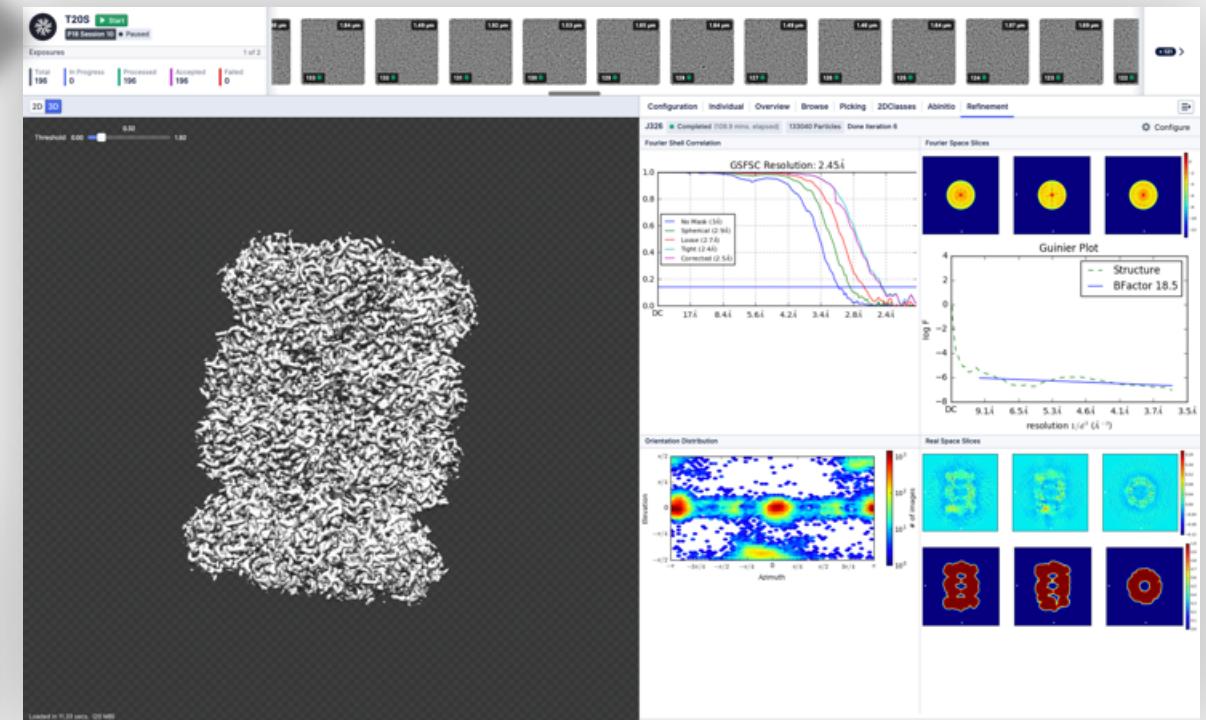
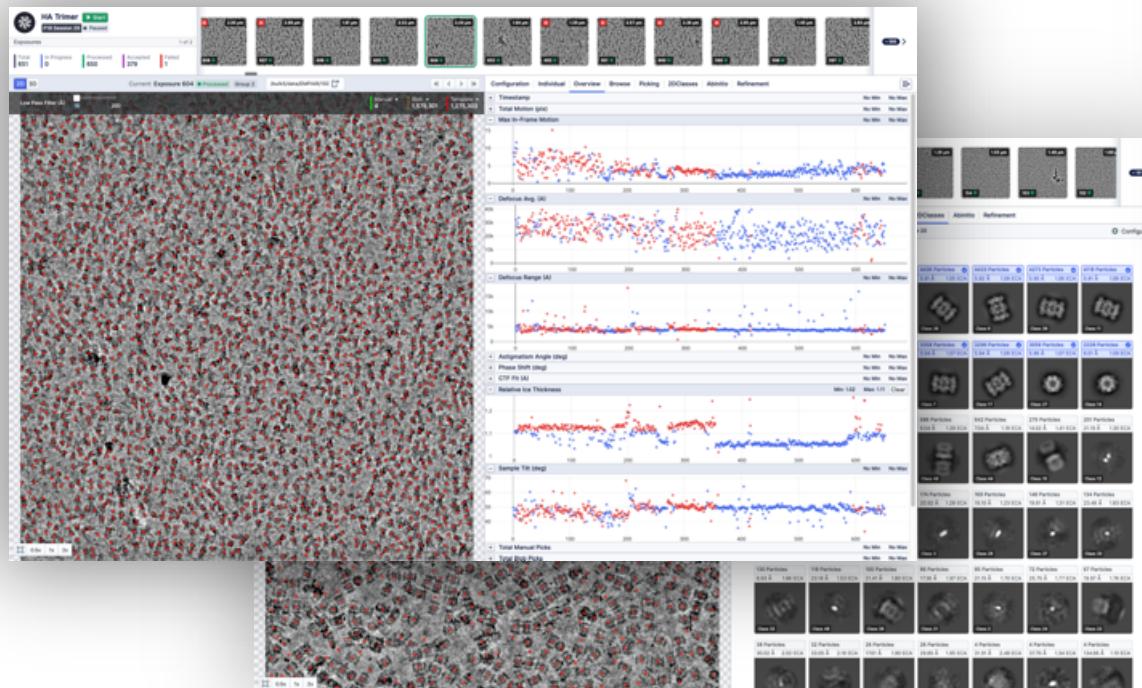


Real-time processing of single-particle EM data

CryoSPARC Live: 3D structures during data collection

Keep up with data collection at 10,000+ exposures per day

- Fully real-time processing to 3D
- Seamless workflow with user interaction
- Curate micrographs and particles



Questions?

Please ask via Zoom Q&A!

cryoSPARC newsletter:
cryosparc.com/#newsletter

Discussion forum:
discuss.cryosparc.com

Documentation:
cryosparc.com/docs