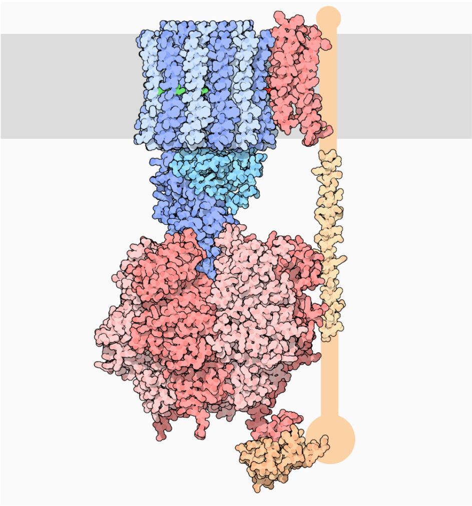
Microscopic Advances with Large-Scale Learning: Stochastic Optimization for Cryo-EM

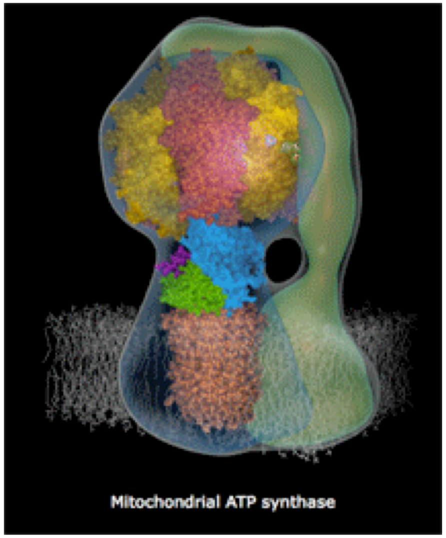
Ali Punjani, Marcus Brubaker
University of Toronto Department of Computer Science

Structure Determination

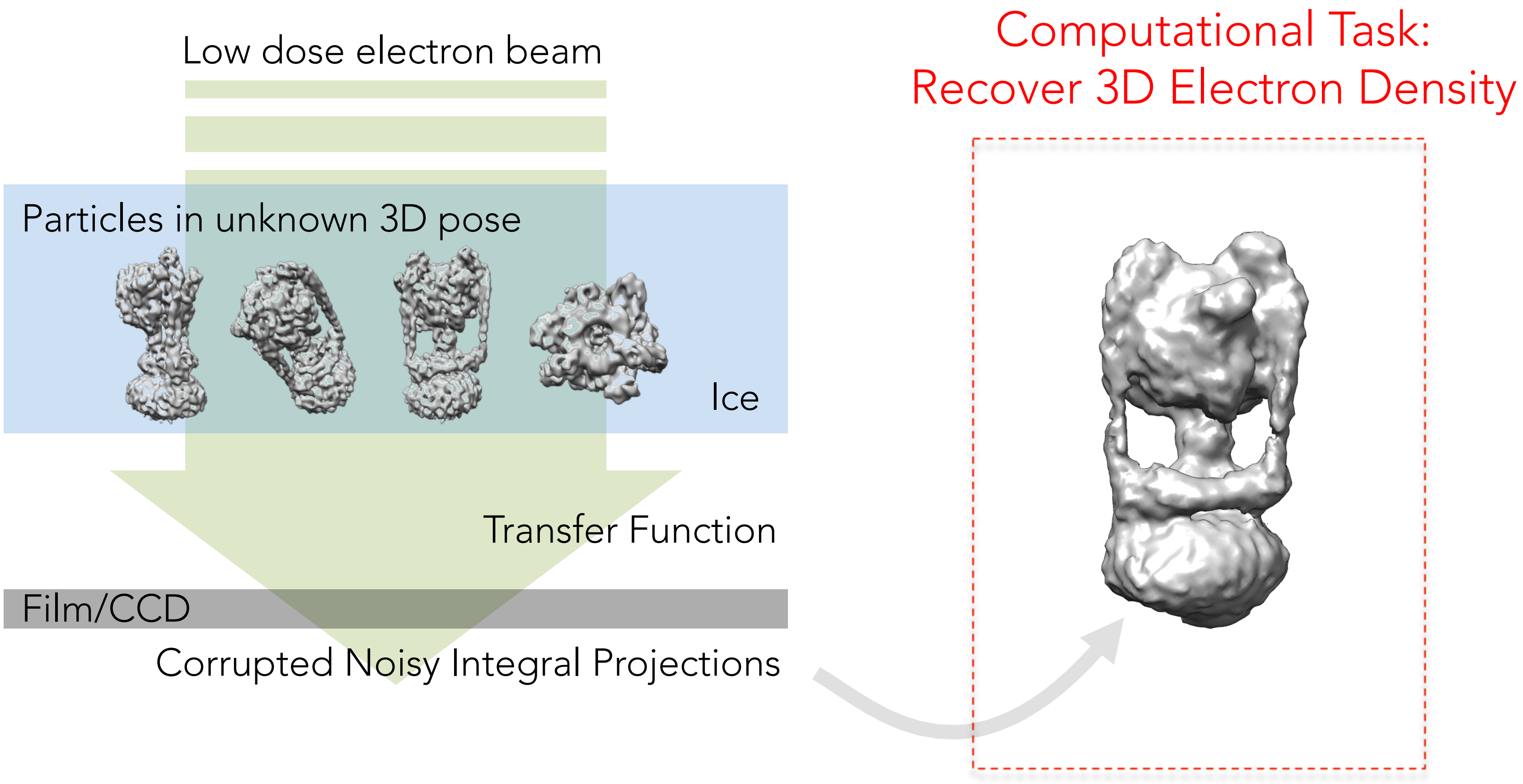
- Macromolecules
- Protein structure determines function

- Traditional approaches:
  - X-ray Crystallography
  - NMR Spectroscopy



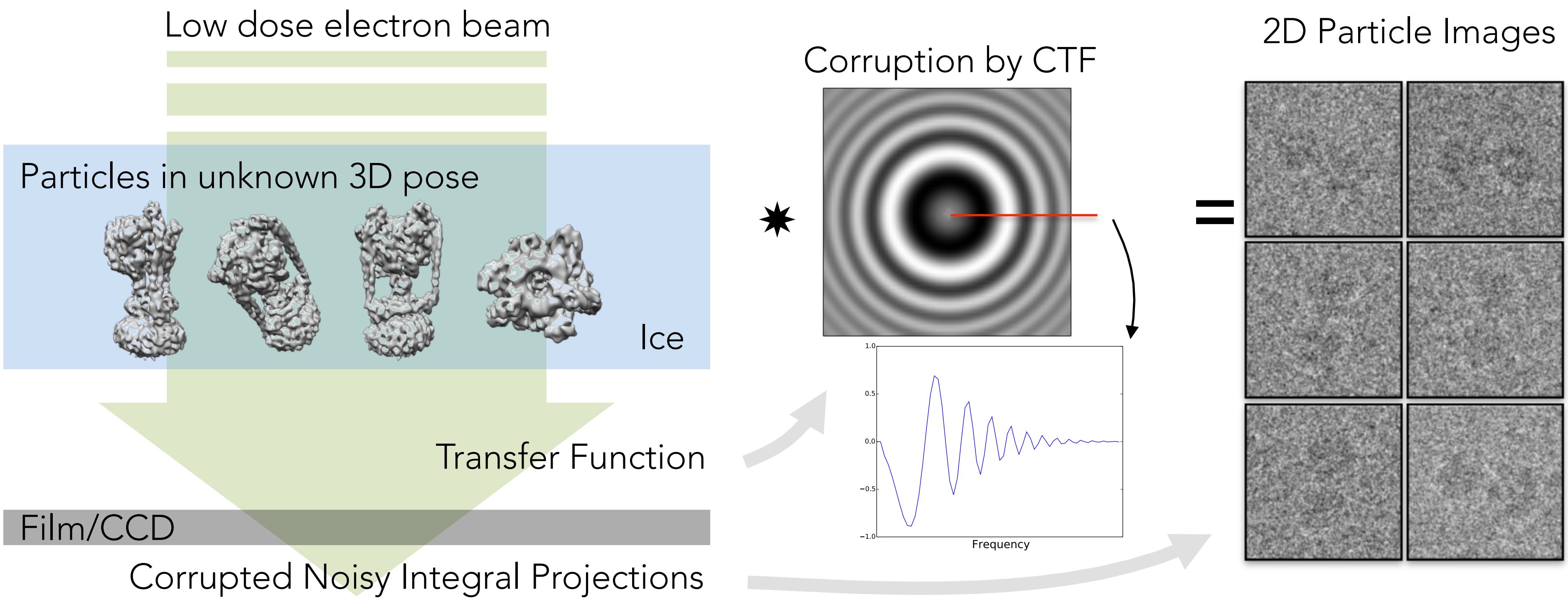


Electron Cryo-Microscopy (Cryo-EM)



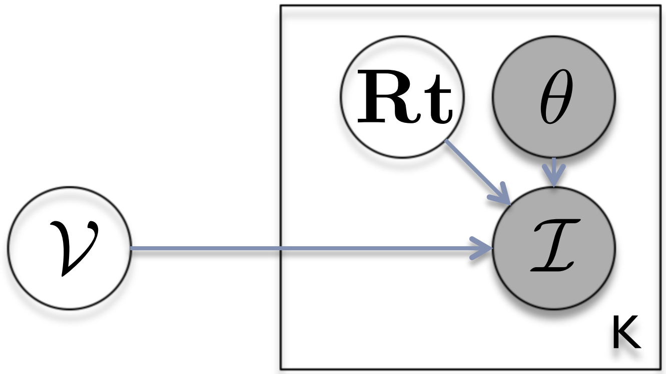
- No crystals needed, large molecules and complexes

Cryo-EM Image Formation



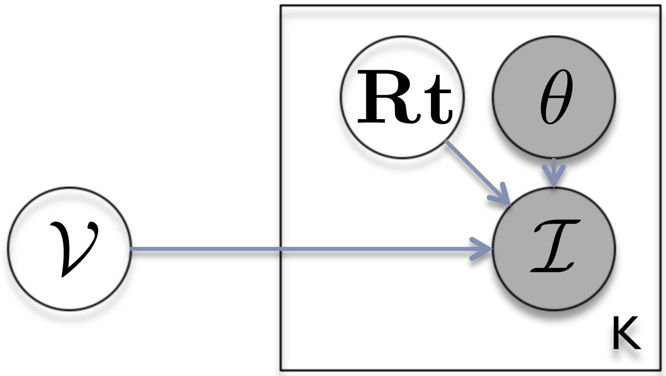
- Challenges for reconstruction:
  - Destructive CTF
  - Low SNR
  - Unknown pose

Cryo-EM Image Formation



$$p(\mathcal{I}|\theta, \mathbf{R}, \mathbf{t}, \mathcal{V}) = \mathcal{N}(\mathcal{I}|\mathbf{S_t}\mathbf{C}_\theta\mathbf{P_R}\mathcal{V}, \sigma^2\mathbf{I})$$

Cryo-EM Image Formation



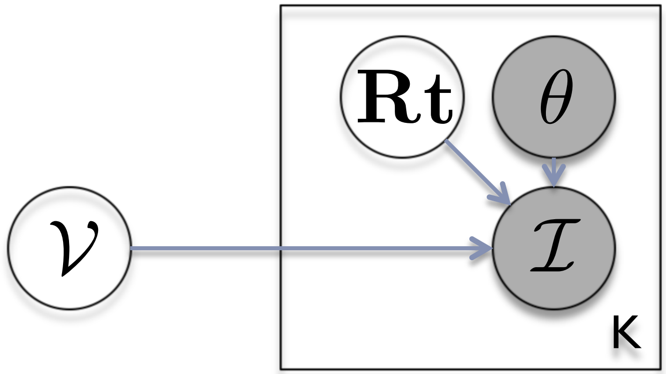
Voxels

$$p(\mathcal{I}|\theta, \mathbf{R}, \mathbf{t}, \mathcal{V}) = \mathcal{N}(\mathcal{I}|\underline{\mathbf{S}_{\mathbf{t}}\mathbf{C}_{\theta}}\mathbf{P}_{\mathbf{R}}\mathcal{V}, \sigma^2\mathbf{I})$$

Integral Projection

Linear

Cryo-EM Image Formation



Voxels

$$p(\mathcal{I}|\theta, \mathbf{R}, \mathbf{t}, \mathcal{V}) = \mathcal{N}(\mathcal{I}|\underline{\mathbf{S}_{\mathbf{t}}\mathbf{C}_{\theta}}\mathbf{P}_{\mathbf{R}}\mathcal{V}, \sigma^2\mathbf{I})$$

Integral Projection

Linear

In Fourier Domain:

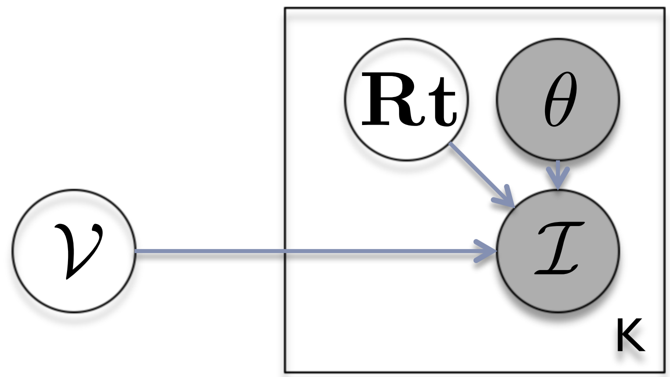
Fourier Coefficients

$$p(\tilde{\mathcal{I}}|\theta, \mathbf{R}, \mathbf{t}, \tilde{\mathcal{V}}) = \mathcal{N}(\tilde{\mathcal{I}}|\underline{\tilde{\mathbf{S}}_{\mathbf{t}}\tilde{\mathbf{C}}_{\theta}}\tilde{\mathbf{P}}_{\mathbf{R}}\tilde{\mathcal{V}}, \sigma^2\mathbf{I})$$

Slicing

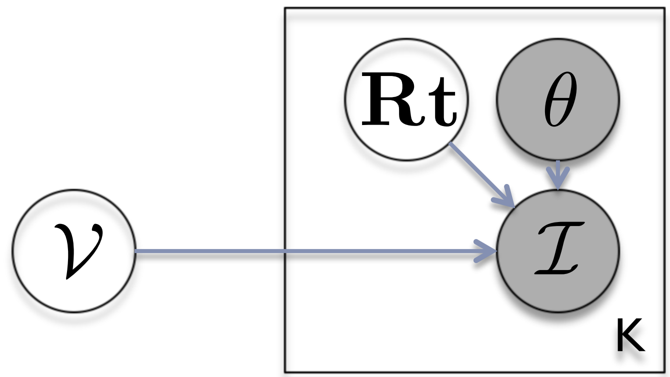
Diagonal

Marginalization for Latent Variables



$$p(\tilde{\mathcal{I}}|\theta, \tilde{\mathcal{V}}) = \int_{\mathbb{R}^2} \int_{\mathcal{SO}(3)} p(\tilde{\mathcal{I}}|\theta, \mathbf{R}, \mathbf{t}, \tilde{\mathcal{V}}) p(\mathbf{R}) p(\mathbf{t}) d\mathbf{R} d\mathbf{t}$$

Marginalization for Latent Variables

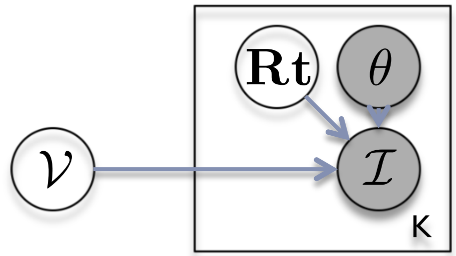


$$p(\tilde{\mathcal{I}}|\theta, \tilde{\mathcal{V}}) = \int_{\mathbb{R}^2} \int_{\mathcal{SO}(3)} p(\tilde{\mathcal{I}}|\theta, \mathbf{R}, \mathbf{t}, \tilde{\mathcal{V}}) p(\mathbf{R}) p(\mathbf{t}) d\mathbf{R} d\mathbf{t}$$

$$\approx \sum_{j=1}^{M} w_j p(\tilde{\mathcal{I}}|\theta, \mathbf{R}_j, \mathbf{t}_j, \tilde{\mathcal{V}})$$

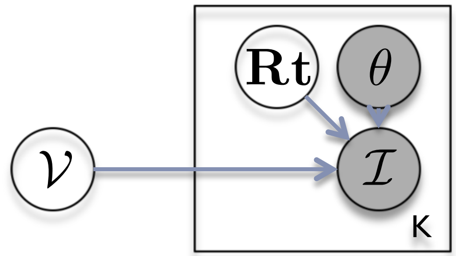
- Numerical Quadrature

Maximum-a-Posteriori Estimation



$$p(\mathcal{V}|\mathfrak{D}) \propto p(\mathcal{V}) \prod_{i=1}^{K} \underline{p(\tilde{\mathcal{I}}_i|\theta_i, \tilde{\mathcal{V}})}$$

Optimization Problem



$$p(\mathcal{V}|\mathfrak{D}) \propto p(\mathcal{V}) \prod_{i=1}^{K} p(\tilde{\mathcal{I}}_i|\theta_i, \tilde{\mathcal{V}})$$

$$\arg\min_{\mathcal{V}} -\sum_{i=1}^{K} \left(\log p(\tilde{\mathcal{I}}|\theta, \tilde{\mathcal{V}}) + K^{-1} \log p(\mathcal{V}) \right)$$

Stochastic Optimization for Cryo-EM

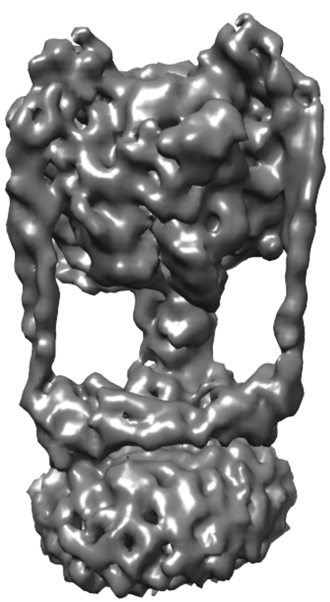
$$\arg\min_{\mathcal{V}} - \sum_{i=1}^{K} \left(\log p(\tilde{\mathcal{I}}|\theta, \tilde{\mathcal{V}}) + K^{-1} \log p(\mathcal{V}) \right)$$

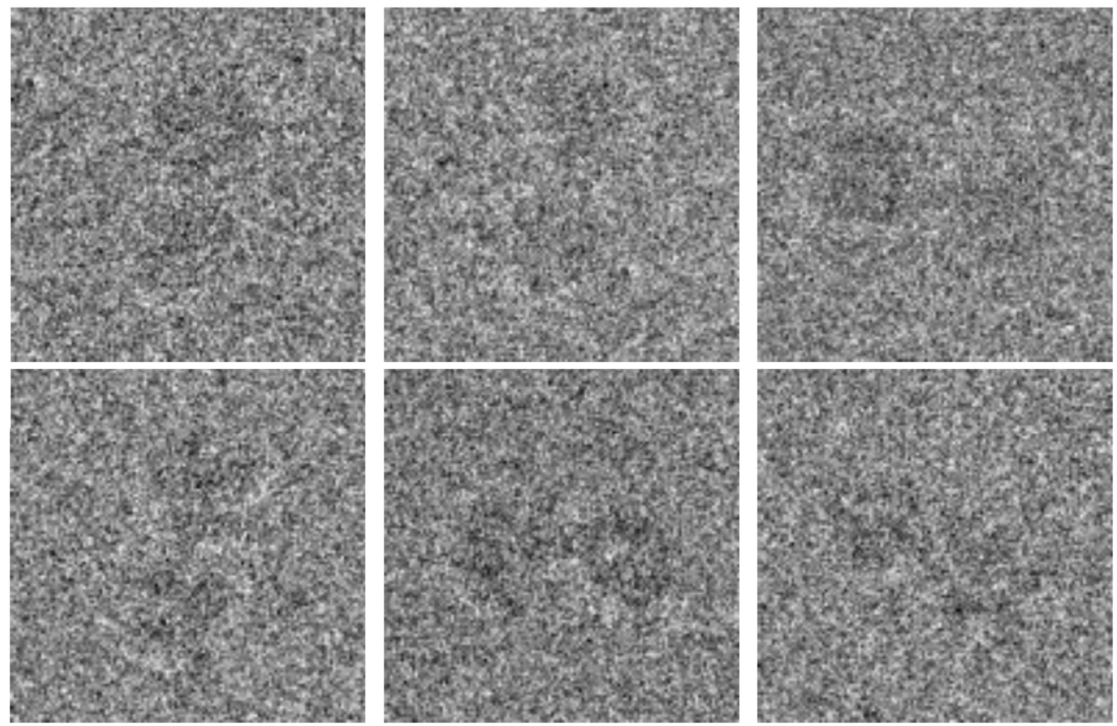
- Expensive to compute objective with large *K*
- Stochastic Optimization:
  - Approximate objective with subset of images
  - Update based on approximate gradient
- Various Algorithms (vary by update rule)

- Advantages: speed, random initialization

Experiments: Datasets

- Real Dataset:
  - 46K Images of ATP Synthase from *Thermus Thermophilius*
  - Low SNR and known CTF parameters





Experiments: Datasets

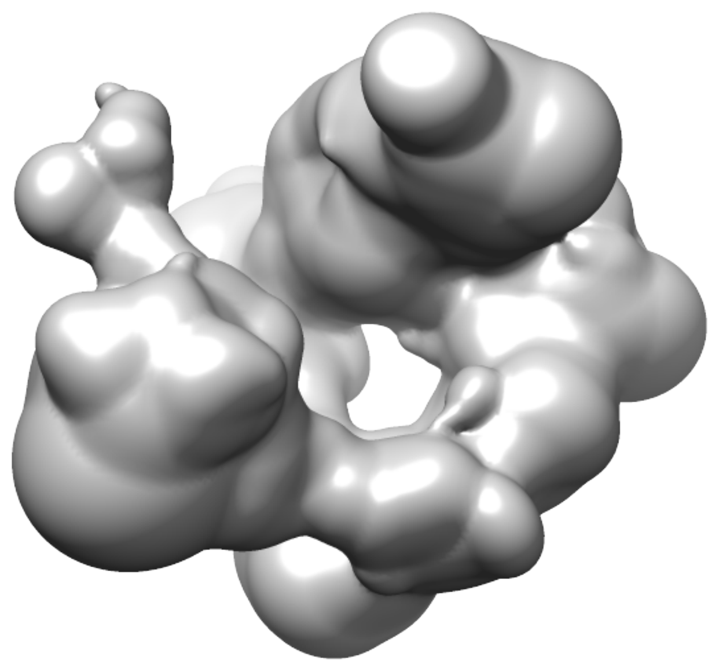
- Synthetic Dataset:
  - 50,000 Projections of known artificial density
  - Low SNR and realistic CTF parameters

Experiments: Seven Methods

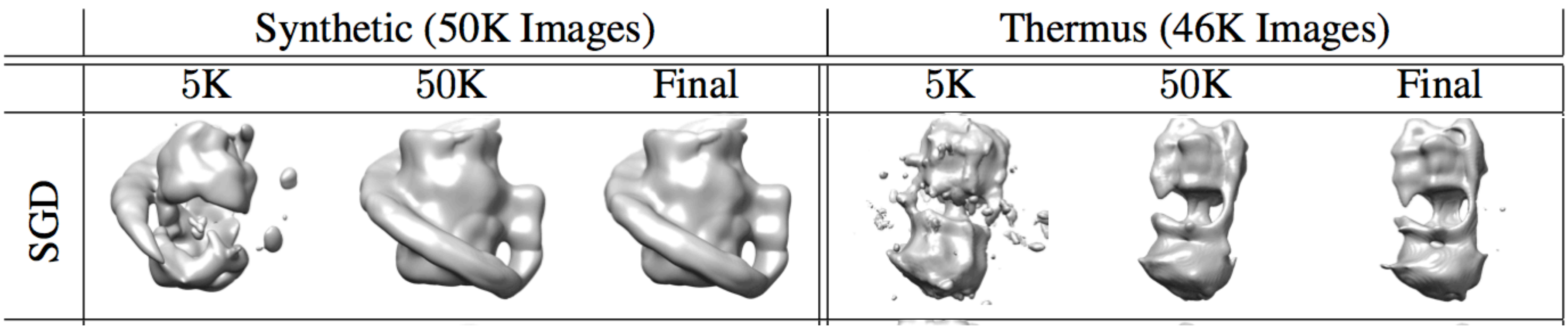
- *Vanilla* Stochastic Gradient Descent (SGD)
- Momentum Methods:
  - Classical Momentum
  - Nesterov's Accelerated Gradient
- Adaptive Methods:
  - AdaGrad
  - TONGA
- Quasi-Second Order Methods:
  - Online L-BFGS
  - Hessian Free

Experiments: Results

- Identical random initialization in all experiments

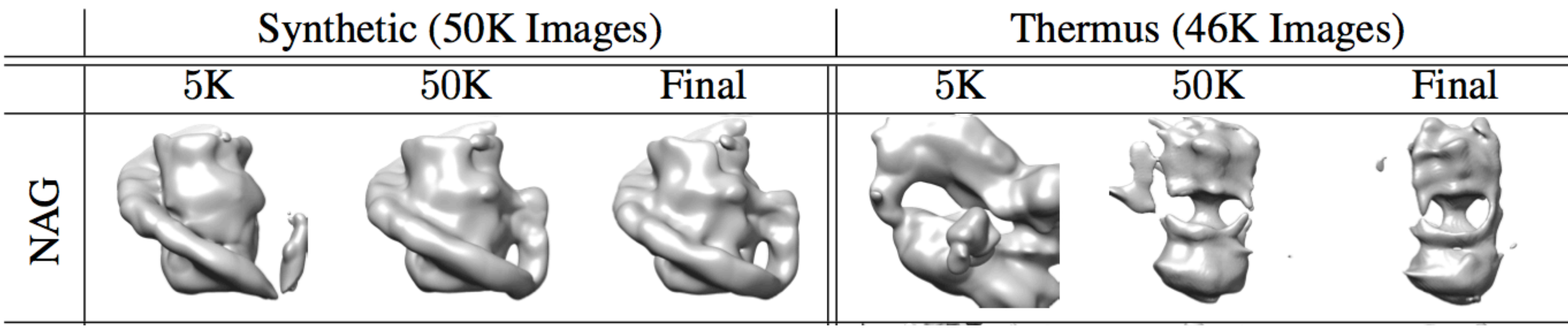


Experiments: Results



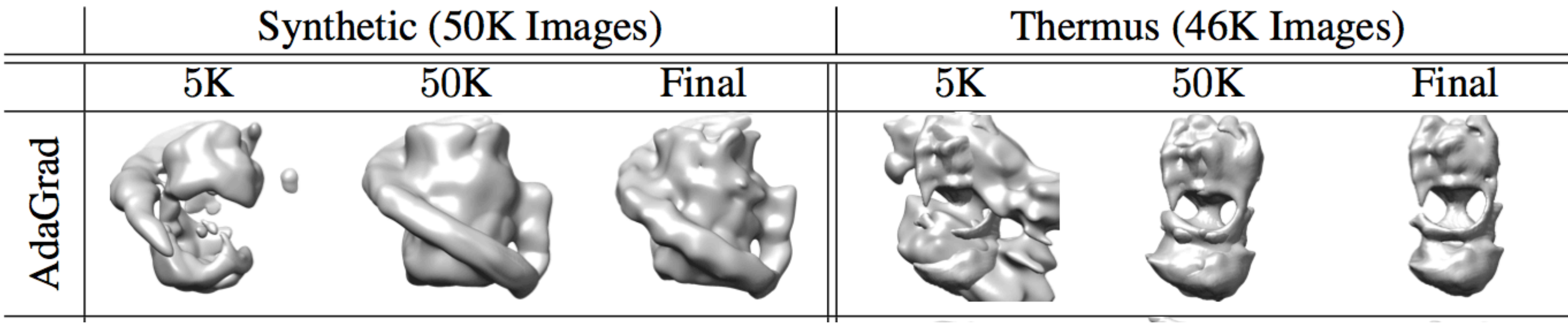
- Simplest Method

Experiments: Results



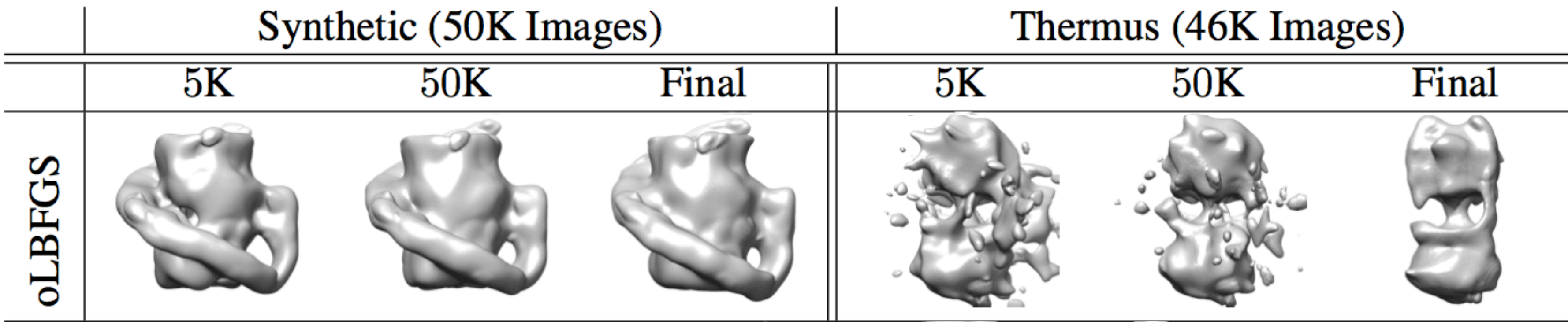
- Momentum Method

Experiments: Results



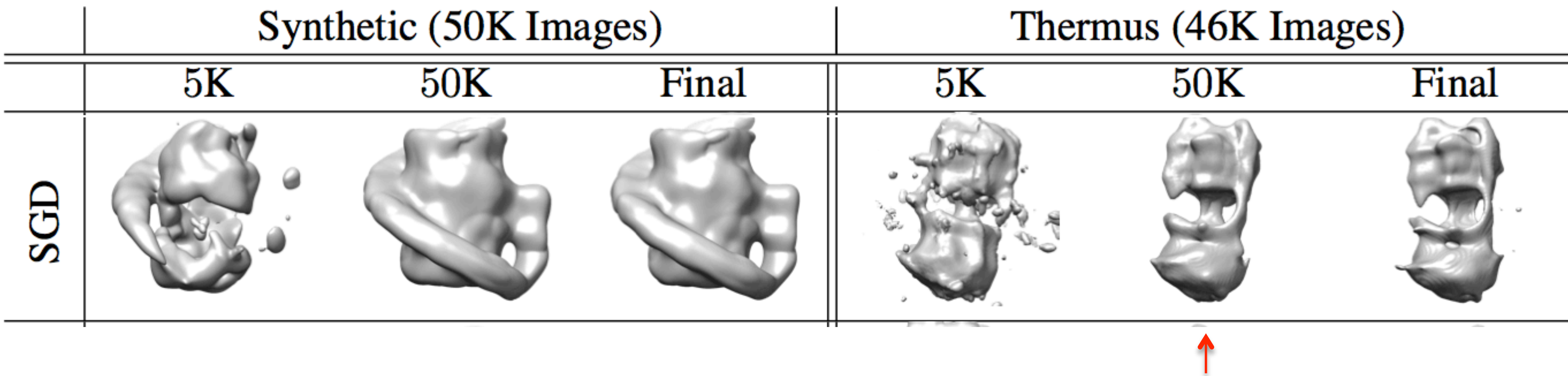
- Adaptive Step-size

Experiments: Results



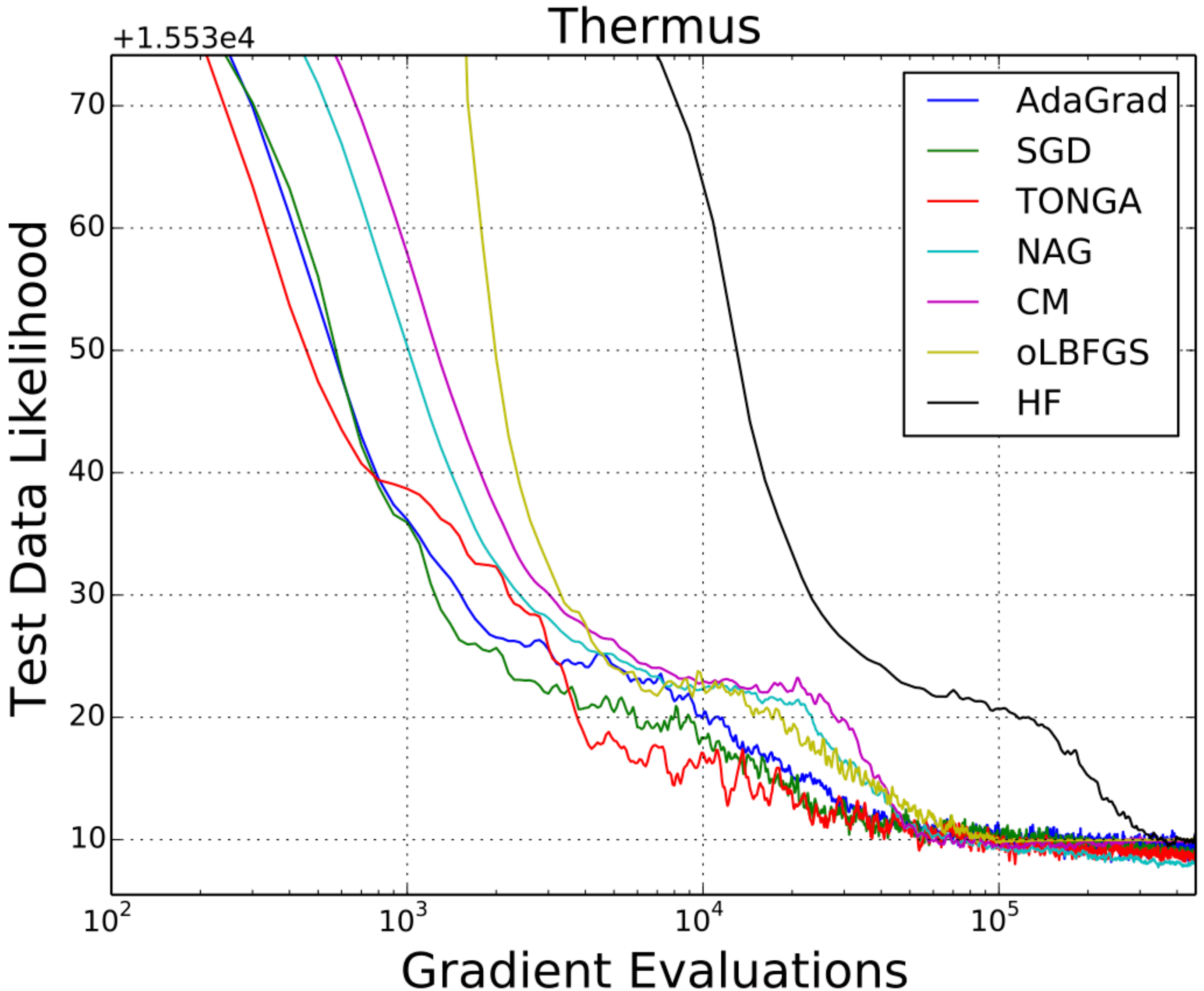
- Quasi-second order

Experiments: Results

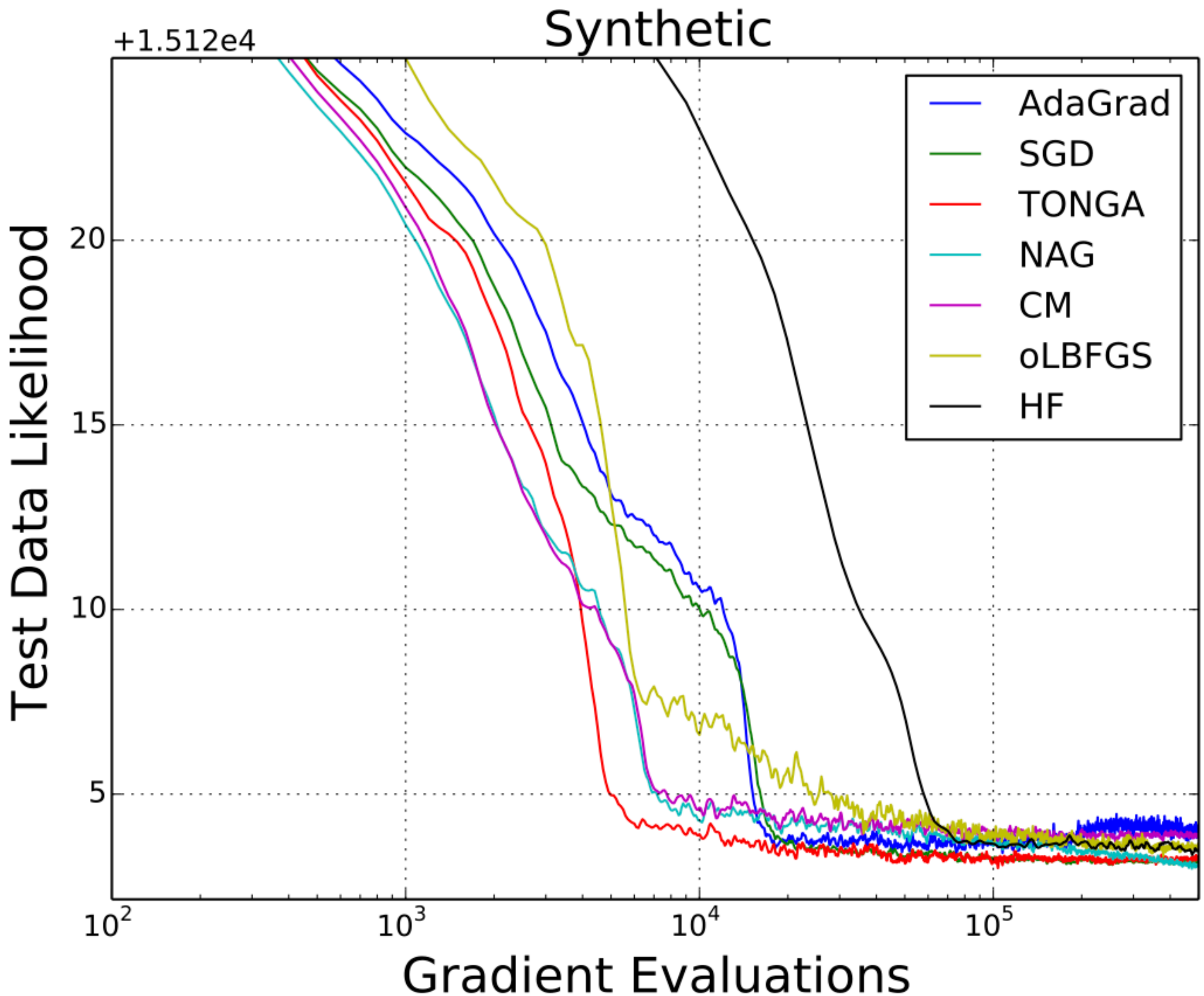


- Qualitatively Similar
- Reasonable in one pass through data

Experiments: Results



Experiments: Results



Experiments: Comparison

Projection Matching

RELION (E-M)

Proposed Approach

3 Hours – 1 Epochs

Experiments: Comparison

Projection Matching

24 Hours – 5 Epochs

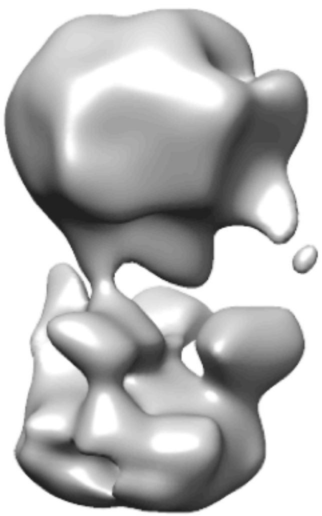
RELION (E-M)

24 Hours – 5 Epochs

Proposed Approach

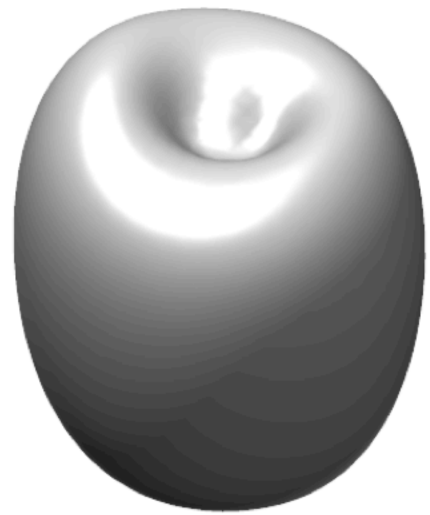
3 Hours – 1 Epochs

Experiments: Comparison



Projection Matching

24 Hours – 5 Epochs



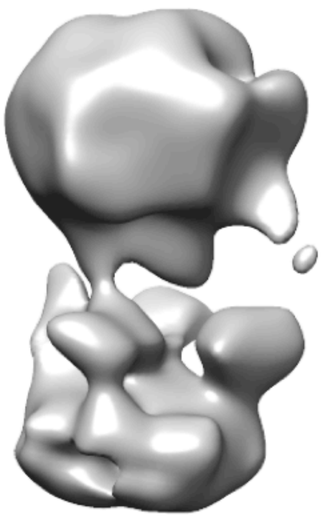
RELION (E-M)

24 Hours – 5 Epochs

Proposed Approach

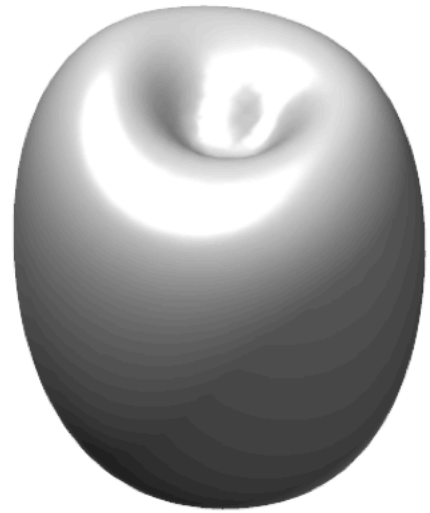
3 Hours – 1 Epochs

Experiments: Comparison



Projection Matching

24 Hours – 5 Epochs



RELION (E-M)

24 Hours – 5 Epochs

Proposed Approach

3 Hours – 1 Epochs

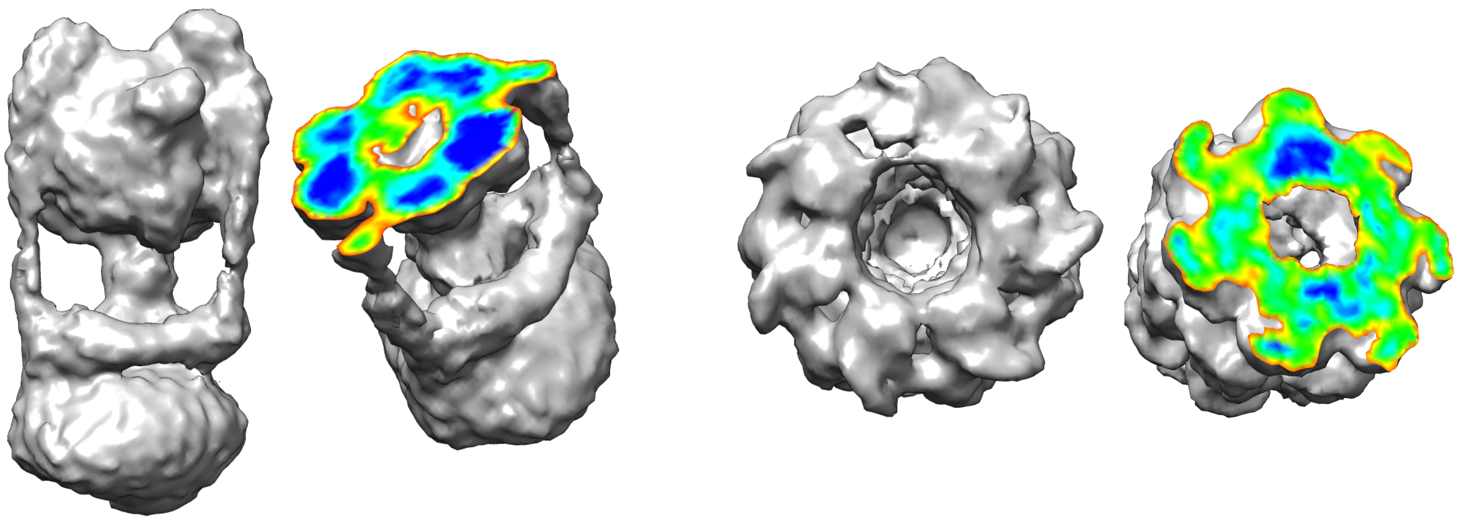
- Random Initialization is difficult for other methods

Conclusions

- Introduced Cryo-EM Structure Determination
- Stochastic Optimization solution
- Simple methods are best
- State of the art speed and robustness

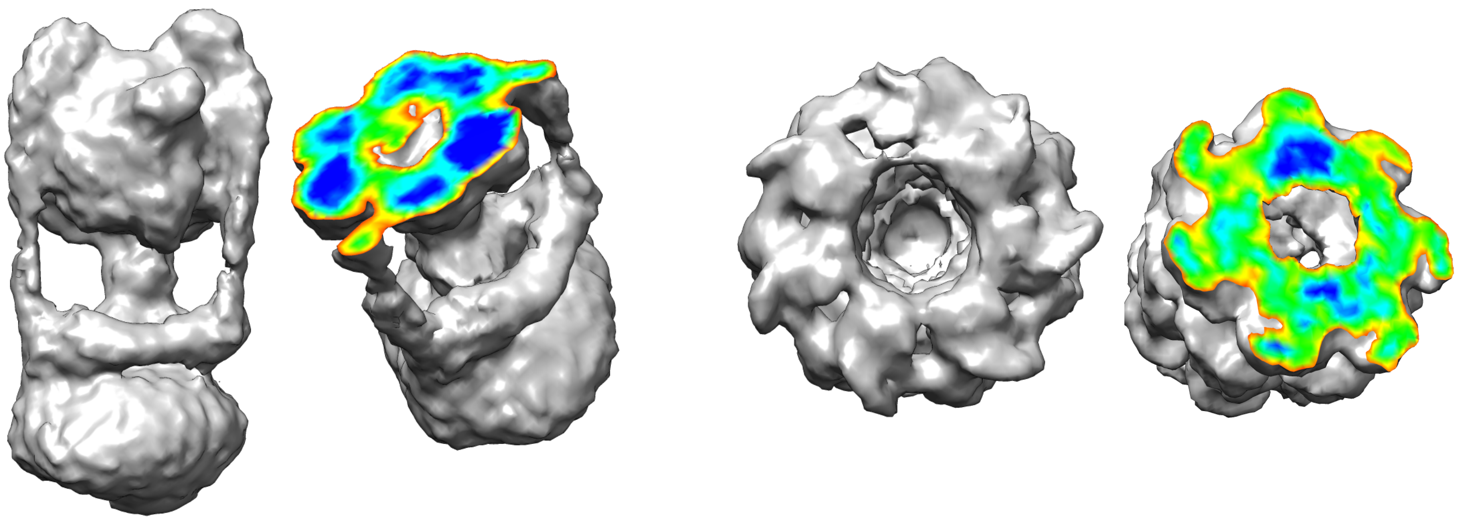
Recent Progress

- Higher resolution reconstructions
- Importance Sampling: 100,000x speedup



Recent Progress

- Higher resolution reconstructions
- Importance Sampling: 100,000x speedup



- Forward:
  - Heterogeneous mixtures of particles
  - Better priors
  - Video exposure