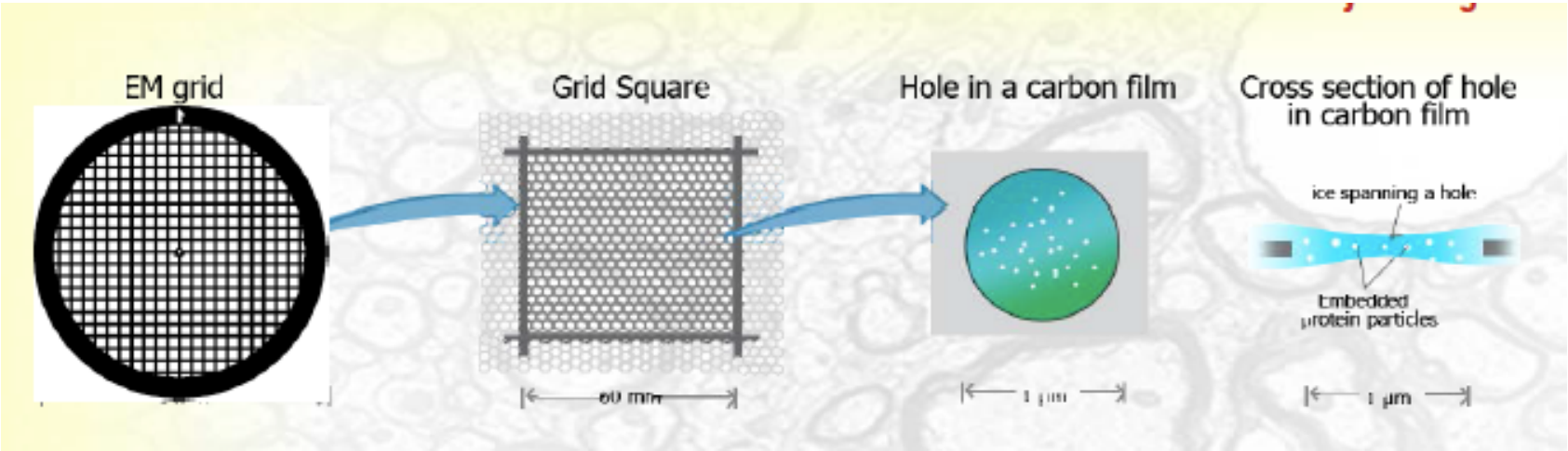
A Fast Proximal Gradient Algorithm For Single Particle Reconstruction of Cryo-EM

You-Wei WEN (文有为)

Hunan Normal University (湖南师范大学), Hunan, China

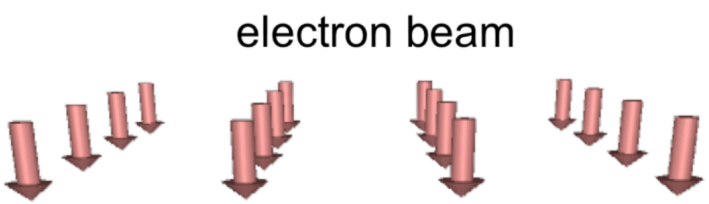
Jointed with Huan Pan and Tieyong Zeng
南京邮电大学
2018-4-14

目录

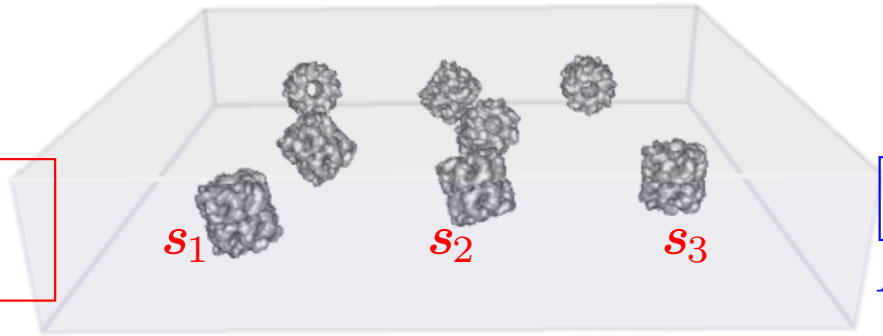


Sample preparation

- The particles of interest are embedded in the holes on this grid in vitreous ice
- a thin layer of sample suspended in water is rapidly frozen in liquid ethane to form vitreous ice
- The specimen is maintained continuously below -160°C during storage and also during imaging in the electron microscope to prevent the formation of ice crystals



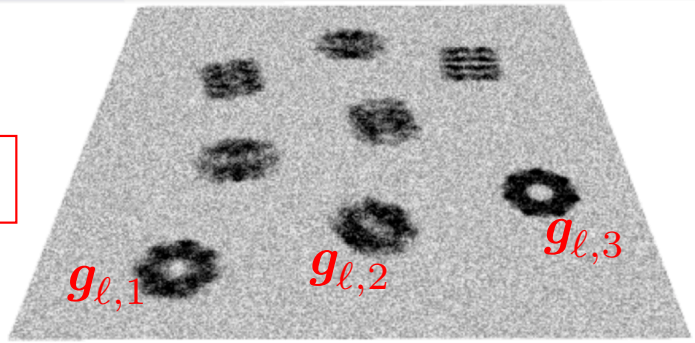
particles in vitreous ice



$s_i = R_i s$

R_i : transformation

projection images



$g_{\ell,i} = \mathcal{C}_{\ell,i}\mathcal{P}\, s_i$

$\mathcal{P}$: Projection

$\mathcal{C}_{\ell,i}$: ℓ-th contrast transform function

图: 2D projection images of 3D object are recorded with an EM. The rotation parameters are unknown! The goal of Cryo-EM is to reconstruct s from $g_{\ell,i}$.

Forward Projection

- $\boldsymbol{s}(\boldsymbol{x})$ be the true particle with $\boldsymbol{x} = (x_1, x_2, x_3)^T \in \mathbb{R}^3$.
- R be the rotation of each particle $\boldsymbol{s}_R(\boldsymbol{x}) = \boldsymbol{s}(R^{-1}\boldsymbol{x})$
- Projection $(\mathcal{P}\boldsymbol{s}_R)(x_1, x_2) = \int_{-\infty}^{\infty} \boldsymbol{s}_R(x_1, x_2, x_3)\, dx_3$.
- Contrast Transfer function determination and correction

$$\mathrm{CTF}(r) = \sin\left(-\pi\left(\frac{\mathrm{defocus}}{2}\cdot r^2 - Cs\cdot\lambda^3\cdot r^4\right) - A\right)\exp\left(-\frac{r^2}{4B\mathrm{factor}}\right).$$

  - r : the magnitude of the frequency
  - Cs : the spherical aberration constant in mm
  - λ : the electron wavelength in picometers
  - A : amplitude contrast.

Challenge in Cryo-EM

- Low electron dose vs. noise. Low electron dose to minimize radiation damage, but more noise in images.

- Number of Particles Required. Thousands or millions of particles to improve the quality and resolution of 3D reconstruction, but thousands or millions unknown parameters in transformation.

Challenge in Cryo-EM

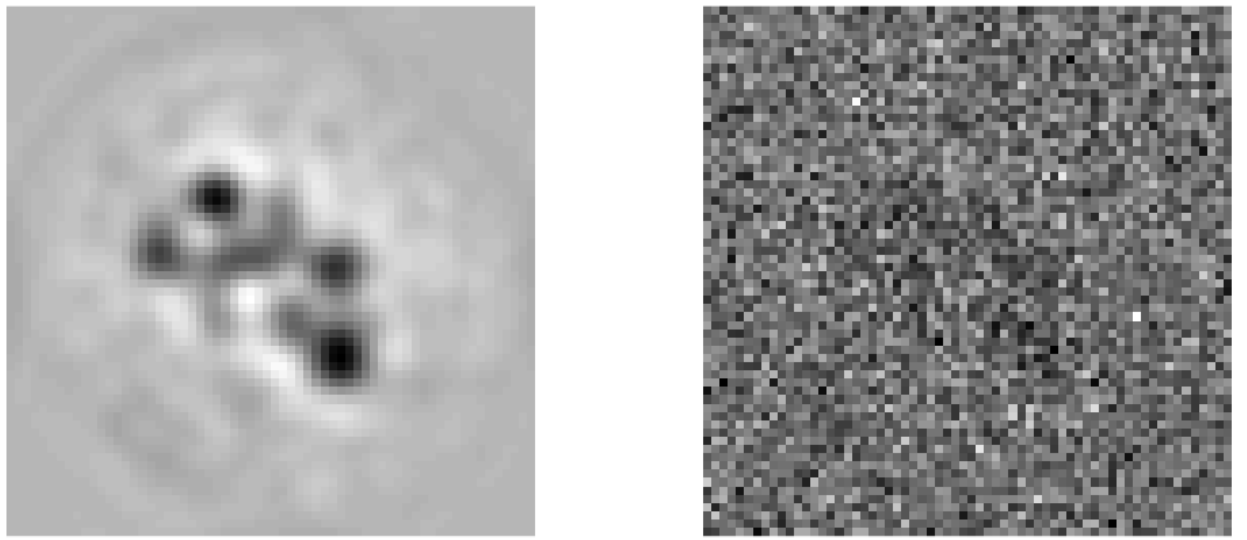
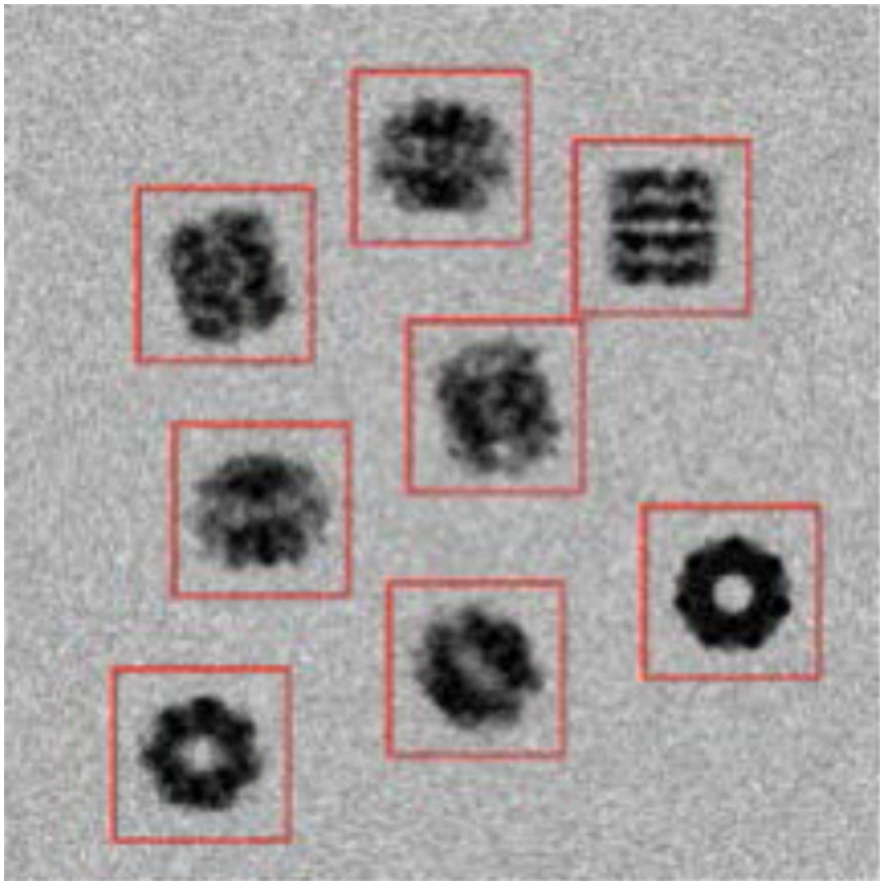


图: Projection image versus noisy image (SNR=0.05dB).



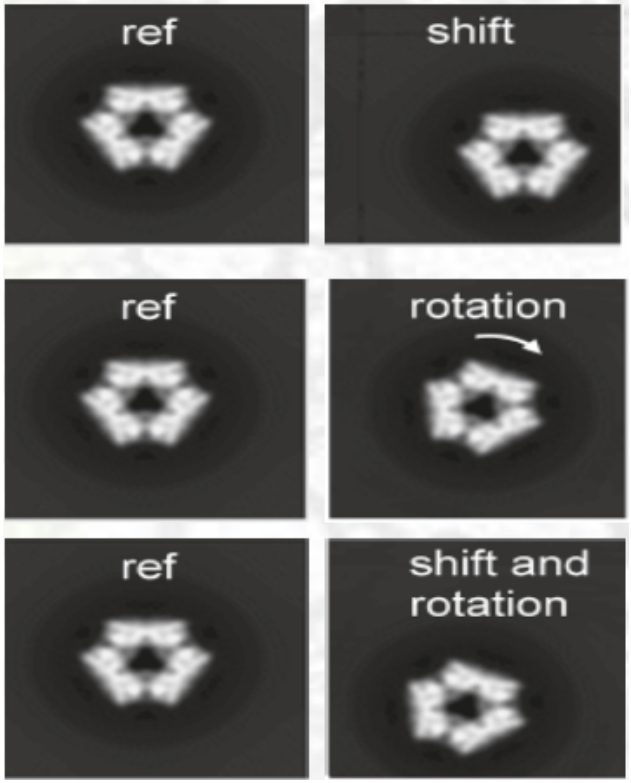


图: Particle selection and 2D classification to group similiar images.
$\boldsymbol{b}_{i,i_1} = T_{i,i_1}\boldsymbol{g}_{i,i1}$, T_{i,i_1} denotes shift and rotation.

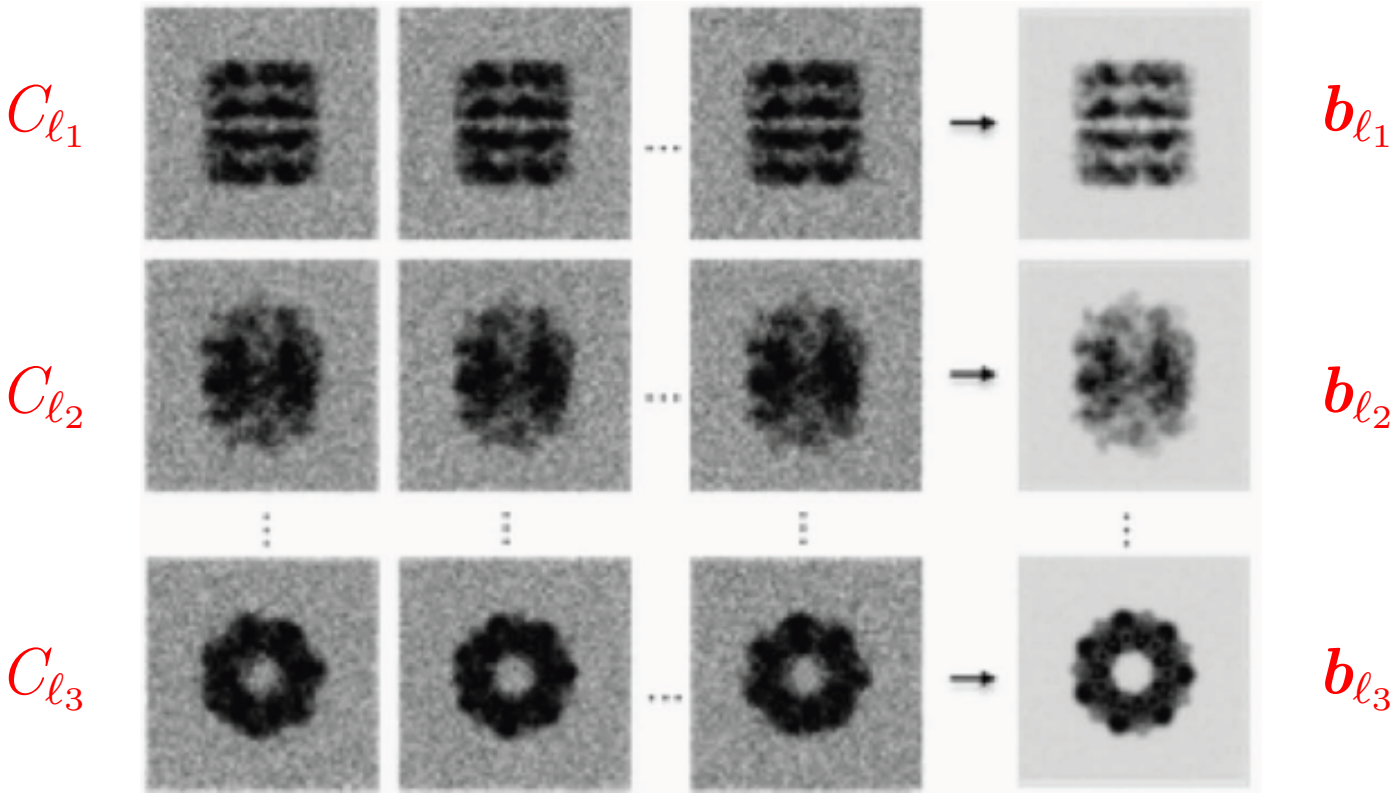


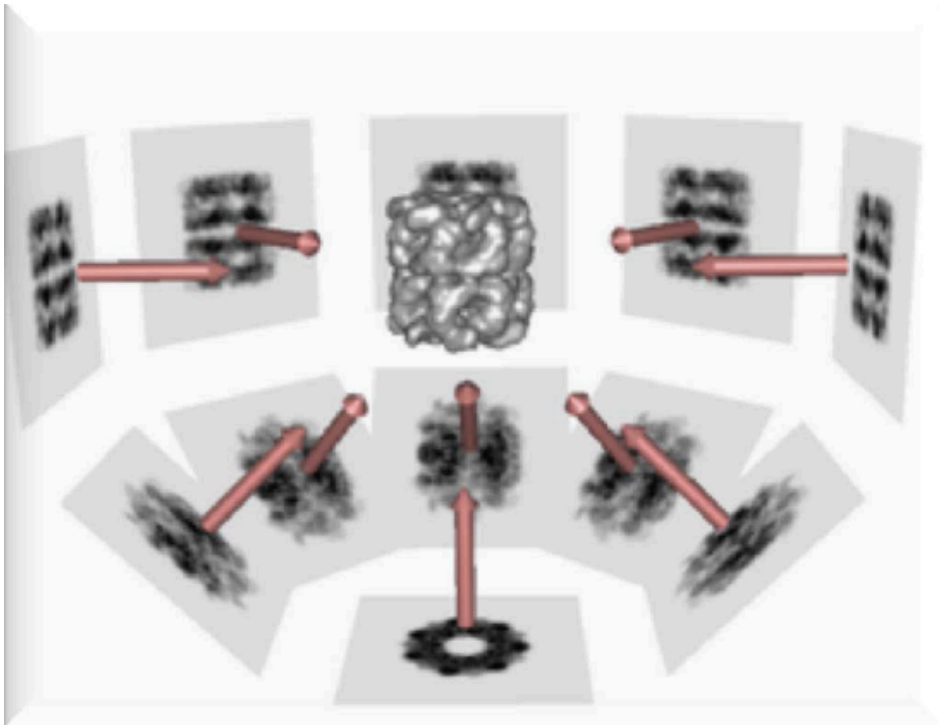
图: 使用低电子剂量需要对粒子群进行平均处理，以提高信噪比

$\min_{\boldsymbol{b}_\ell} \sum_{\boldsymbol{b}_{i,i_1} \in C_\ell} \|\boldsymbol{b}_{i,i_1} - \boldsymbol{b}_\ell\|_2^2$. Multivariate Statistical Analysis can be used to generate a class average, i.e., one single image to represent a group of images that all look the same.

Main advantages of 2D classifications

- Cryo-EM images are noisy (low contrast), averaging similar looking images enhance the contrast by increases signal to noise ratio
- identifying the features and helping in alignment during 3D reconstruction
- helping in identifying bad or unwanted images thereby improving the quality of data
- Its an unbiased process with less user interference

3D Reconstruction



$$\boldsymbol{s}_i = R_i \boldsymbol{s}$$
$$\boldsymbol{g}_{i,i_1} = C_{i,i_1} P \boldsymbol{s}_i$$
$$\boldsymbol{b}_{i,i_1} = T_{i,i_1} \boldsymbol{g}_{i,i1}$$
$$\min_{\boldsymbol{b}_\ell} \sum_{\boldsymbol{b}_{i,i_1} \in C_\ell} \|\boldsymbol{b}_{i,i_1} - \boldsymbol{b}_\ell\|_2^2$$

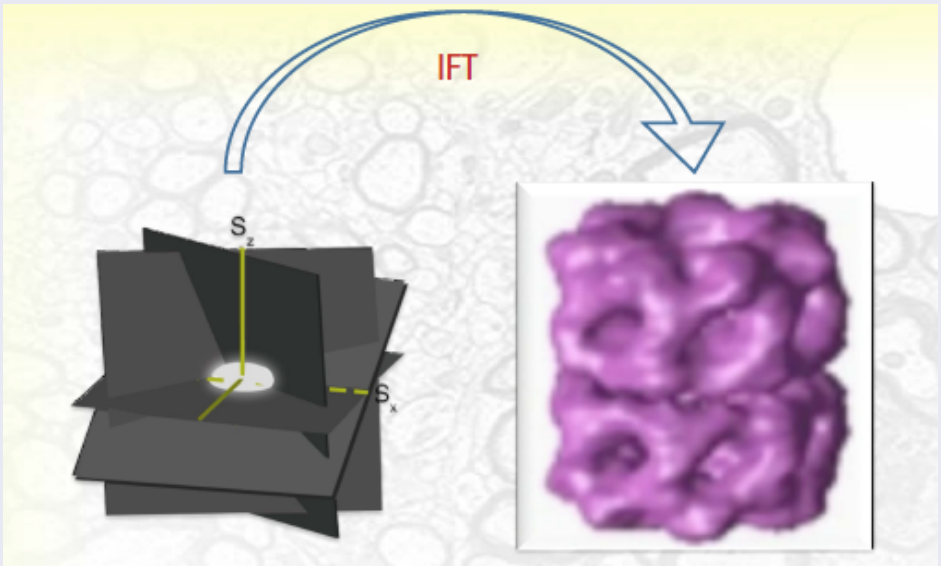
$\boldsymbol{b}_\ell$: back projection

图: 3D reconstruction.

Central slice theorem

$$(\mathcal{F}_2 \mathcal{P} \boldsymbol{s}_R)(\omega_1, \omega_2) = (\mathcal{F}_3 \boldsymbol{s}_R)(\omega_1, \omega_2, 0).$$

$\mathcal{F}_2$: 2D-Fourier transform, $\mathcal{F}_3$: 3D-Fourier transform.



Main steps of SPR in Cryo-EM

- Sample Preparation
- Electron Microscopy Imaging,
- EM Image Processing
  - Data quality assessment
  - CTF determination and correction
  - Particle selection
  - 2D Classification: determine shift and rotation parameter T_{i,i_1}.
- 3D Classification: determine rotation parameter R_i.
- 3D Reconstruction

References

Allison Doerr.

Single-particle electron cryomicroscopy.
Nature Methods, 11:30EP, 12 2013.

Richard Gordon, Robert Bender, and Gabor T Herman.

Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography.
Journal of Theoretical Biology, 29(3):471–481, 1970.

Frank Joachim.

Three-dimensional electron microscopy of macromolecular assemblies.
Elsevier, 1996.

Alp Kucukelbir, Fred J Sigworth, and Hemant D Tagare.

A bayesian adaptive basis algorithm for single particle reconstruction.
Journal of Structural Biology, 179(1):56–67, 2012.

Sjors HW Scheres.

Relion: implementation of a bayesian approach to cryo-em structure determination.
Journal of Structural Biology, 180(3):519–530, 2012.

Lanhui Wang, Amit Singer, and Zaiwen Wen.

Orientation determination of cryo-em images using least unsquared deviations.
SIAM journal on Imaging Sciences, 6(4):2450–2483, 2013.

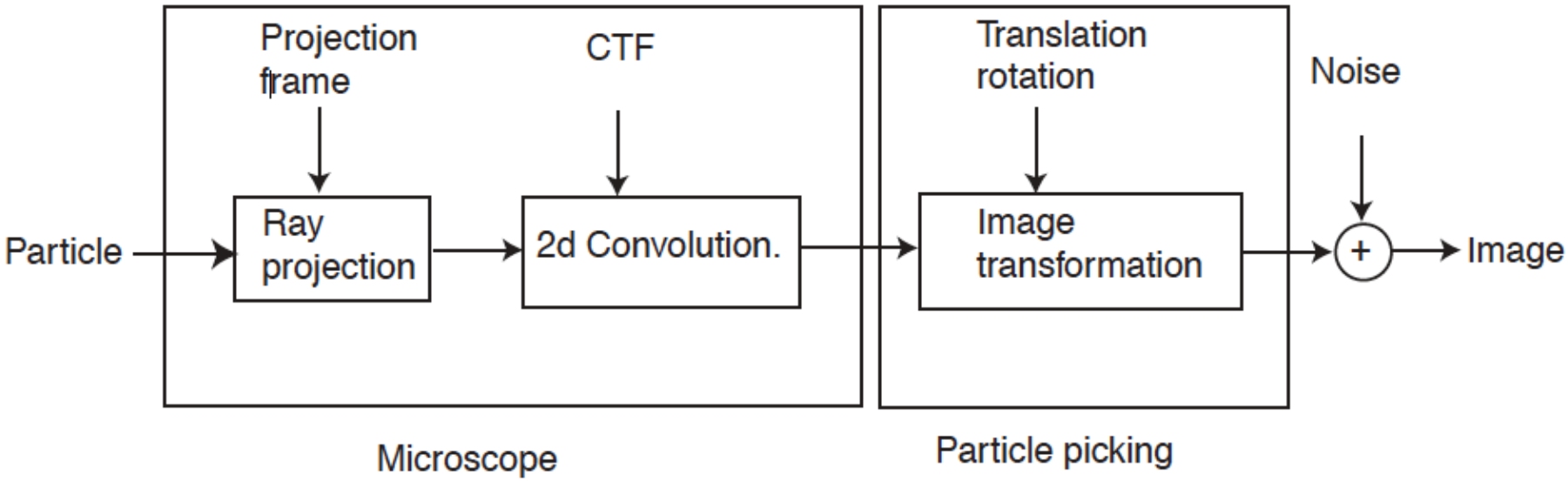


图: Signal Flow for Cryo-EM

Mathematical Model of SPR

$$\boldsymbol{b}_k = T_k C_k \Omega_k \boldsymbol{s} + \eta_k$$

- $\boldsymbol{b}_k$: the 2-D observed Cryo-EM image with size $P \times P$.
- $\boldsymbol{s}$: the 3D particle density map with size $N \times N \times N$.
- T_k: the 2-D transformation operator. and the rotation γ_k parameters.
- C_k: the contrast transfer function.
- Ω_k: the projection operator with $\Omega_k = PR_k$.
- η_k: the Gaussian noise with standard deviation σ.

Aim: reconstruct a three-dimension structure of the particle $\boldsymbol{s}$ from its two-dimension observed projections $\boldsymbol{b}_k$.

Least Square Method

$$\min g(\boldsymbol{s}) \equiv \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{b}_k - T_k C_k \Omega_k \boldsymbol{s} \right\|_F^2$$

Regularization Approach

- Assume that the particle $\boldsymbol{s}$ can be sparse under some orthogonal basis vectors: $\boldsymbol{s} = \Phi\alpha$.
- Regularization approach

$$\min_{\alpha} J(\alpha) \equiv \sum_{k=1}^{K} \frac{1}{2} \left\| b_k - T_{\tau_k} C_k \Omega_{p_k} \Phi\alpha \right\|_2^2 + \lambda \left\| \alpha \right\|_1 .$$

- Denote $h(\alpha) = \lambda\|\alpha\|_1$, we have

$$\mathcal{J}(\alpha) = g(\Phi\alpha) + h(\alpha).$$

Proximal Gradient Method

- starting with an initial guess α_0, the iterate is given by

$$\alpha_k = \text{prox}_{t_k h}\left(\alpha_k - t_k \nabla_\alpha g(\Phi\alpha_k)\right).$$

- The proximal operator of the function h is defined by

$$\text{prox}_h(\boldsymbol{v}) = \underset{\alpha}{\text{argmin}} \frac{1}{2}\|\boldsymbol{v} - \alpha\|_2^2 + h(\alpha).$$

- Notice that $h(\alpha) = \lambda\|\alpha\|_1$, we have

$$\left(\text{prox}_h(\boldsymbol{v})\right)_i = \begin{cases} \boldsymbol{v}_i - \lambda, & \boldsymbol{v}_i \geq \lambda. \\ 0, & |\boldsymbol{v}_i| \leq \lambda. \\ \boldsymbol{v}_i + \lambda, & \boldsymbol{v}_i \leq -\lambda. \end{cases}$$

Accelerated Proximal Gradient Method

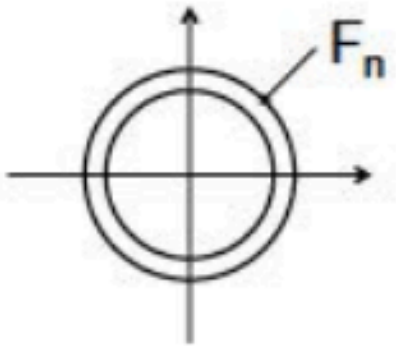
- When an extrapolation step is used in the iterate, the convergence can be improve.
- The iterate becomes to

$$\beta_k = \alpha_{k-1} + \omega^k(\alpha_{k-1} - \alpha_{k-2}).$$

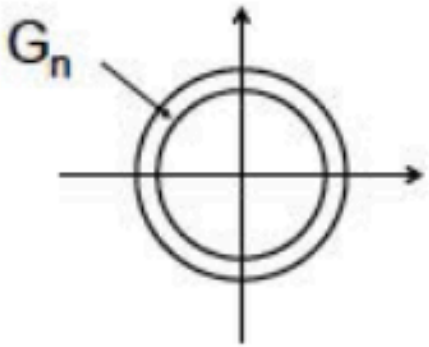
$$\alpha_k = \text{prox}_{t_k h}\left(\beta_k - t_k \nabla_\alpha g(\beta_k)\right).$$

where $\omega^k \in [0, 1)$ is an extrapolation parameter, one simple choice takes

$$\omega^k = \frac{k-1}{k+2}.$$

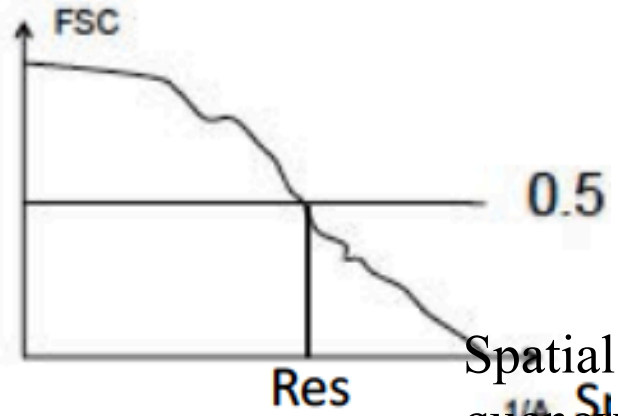


Reconstruction 1



Reconstruction 2

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



- The tested original volume is the 50S ribosomal subunit structure published(PDB ID: 8523), available online at the Protein Data Bank in Europe (PDBE) website [a].
- The PDBe volume is $64 \times 64 \times 64$ voxels with resolution of $16.9 \mathring{A}$□.
- The structure is projected along direction randomly chosen from sampling on the sphere.
- Defocus values of 1.4 μm CTFs with 120 keV electrons.
- A soft, spherical mask with a Gaussian fall-off at the edge was applied to each volume.

[a] http://www.ebi.ac.uk/pdbe/entry/emdb/EMD-8523/

Results

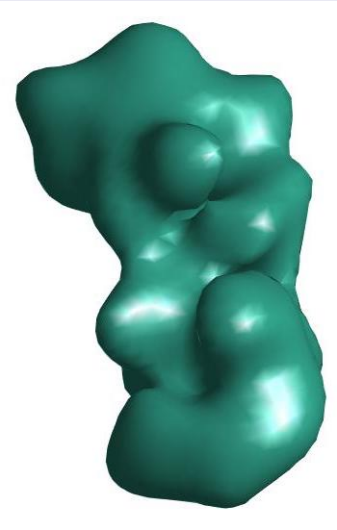


图: origional volume from EMD-8523 in PDBE

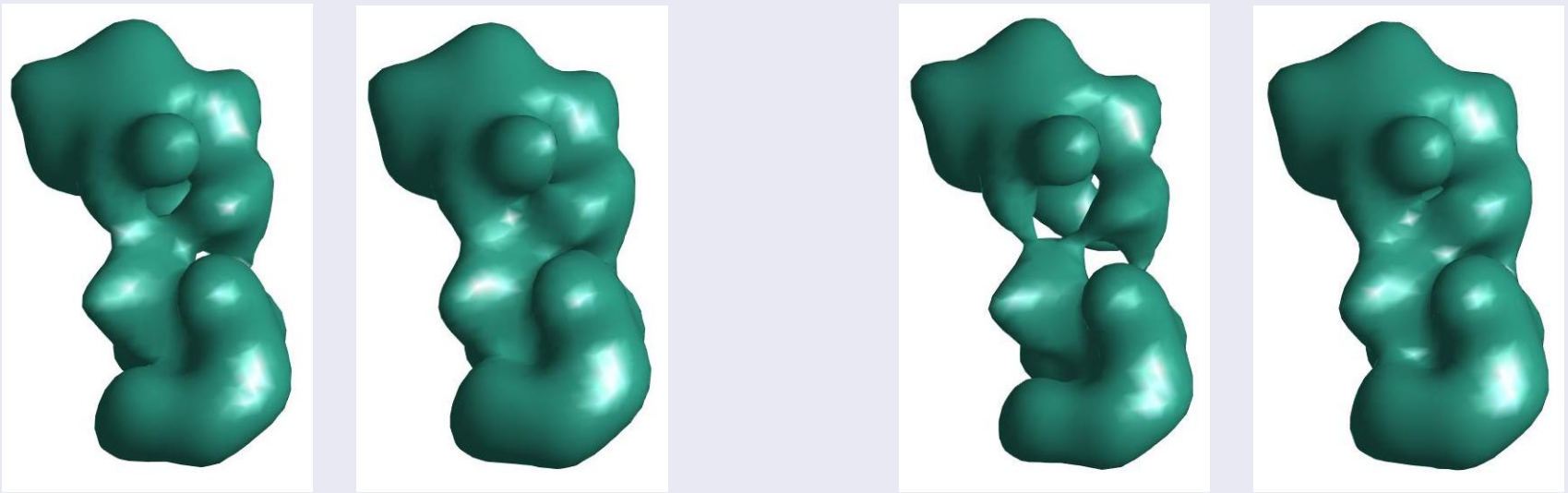


图: Left:σ=1, right: σ=2. Reconstruction volume by AB and FISTA respectively .

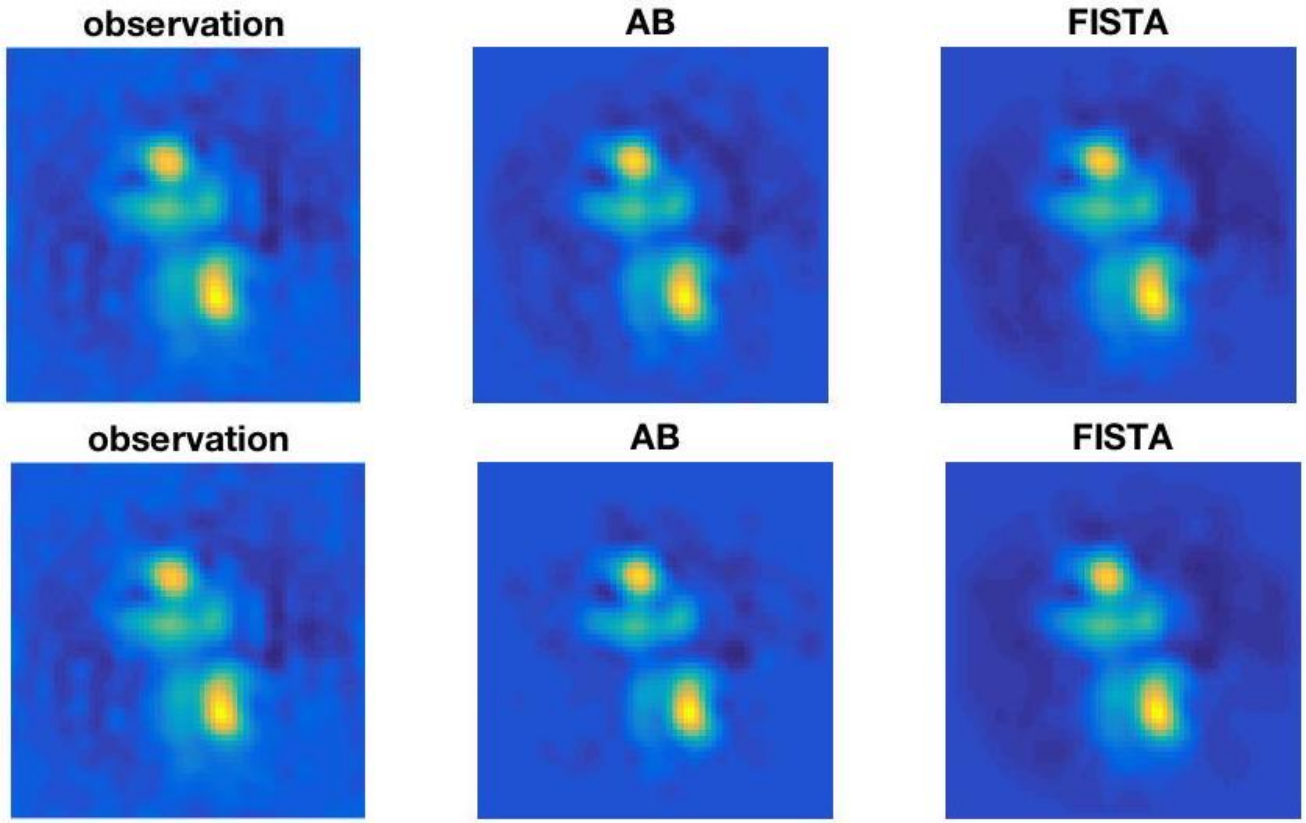


图: Central slices at $n = 32$ from the simulated data reconstructions for 50S ribosomal subunit (PDB ID: 8523). Top: σ=1, Below: σ=2.

σ	Error		CPU Time	
	AB	FISTA	AB	FISTA
1	0.2871	0.1834	178.46	82.18
2	0.3839	0.2200	164.08	74.37

Table: run time and relative error Comparison of AB algorithm and FISTA method at σ=1 and σ=2 respectively

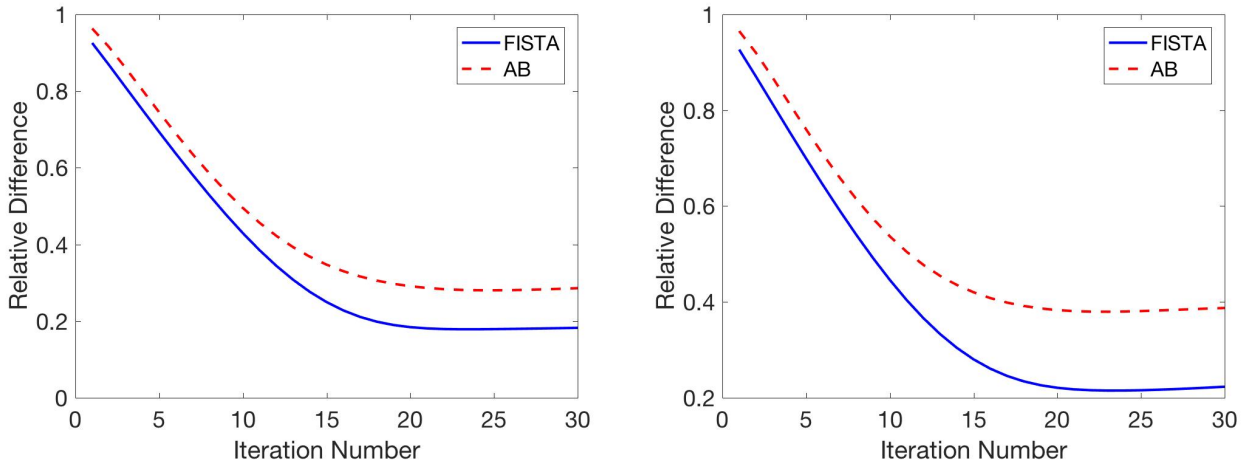


图: Relative error curves for AB and FISTA for $\sigma = 1$ (left) and $\sigma = 2$ (right) respectively.

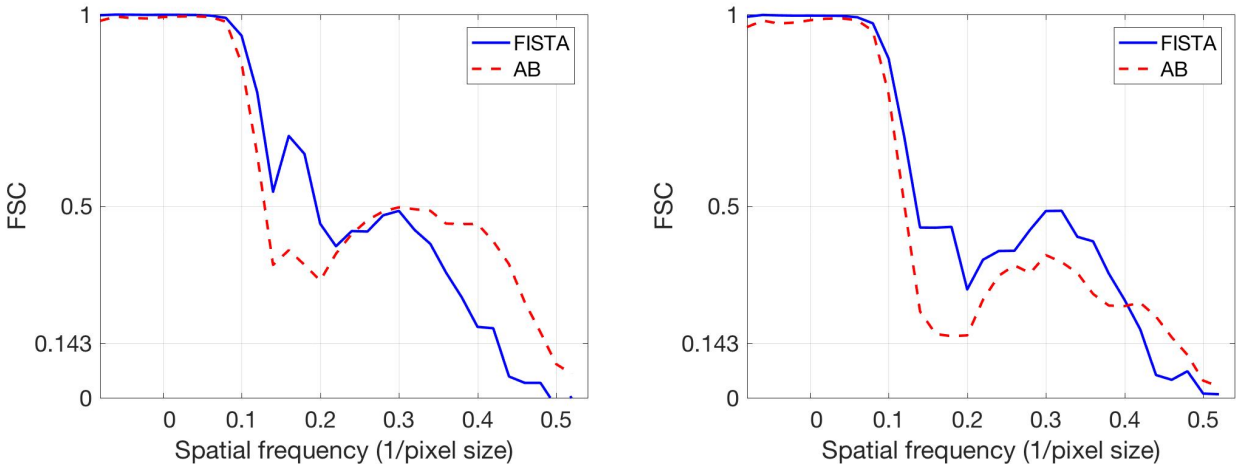


图: FSC curves for AB and FISTA $\sigma = 1$ (left) and $\sigma = 2$ (right) respectively.

Conclusion

- The relative errors obtained the proposed method is less than that obtained by AB.
- The FSCs confirm that the FISTA algorithm reconstructions contain more detail than the AB algorithm results.
- The FISTA methods shows improved performance for simulated data when compared to the adaptive basis reconstruction algorithm (AB).

Thank You! Qestions?