Cryo-EM Reconstruction of Continuous Heterogeneity by Laplacian Spectral Volumes

Group Meeting

Yu-Hsiang Lien 連昱翔

References

IOP Publishing Inverse Problems

Inverse Problems **36** (2020) 024003 (31pp) https://doi.org/10.1088/1361-6420/ab4f55

Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes

Amit Moscovich[1,4], **Amit Halevi**[1,4], **Joakim Andén**[2] **and Amit Singer**[1,3]

[1] Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, United States of America
[2] Center for Computational Mathematics, Flatiron Institute, New York, NY, United States of America
[3] Department of Mathematics, Princeton University, Princeton, NJ, United States of America

E-mail: amit@moscovich.org, ahalevi@princeton.edu, janden@flatironinstitute.org and amits@math.princeton.edu

Received 30 June 2019, revised 14 October 2019
Accepted for publication 18 October 2019
Published 28 January 2020

CrossMark

Reconstruction of Non-rigid Molecules

3D Molecule Reconstruction

Homogeneous Reconstruction → *Recover mean molecular volume*

3D ab-initio model

Heterogeneous Reconstruction → *Recover multiple molecular volumes*

Discrete Heterogeneity (a.k.a. 3D Classification)

→ *Recover K distinct molecular volumes*

RELION, cryoSPARC, EMAN2, …

Continuous Heterogeneity

→ *Recover the manifold of molecular volumes*

This manifold is the range of a smooth function that maps a set of conformation parameters to a volume

- PCA -> Eigenvolumes -> low-resolution reconstruction
- RELION -> multi-body refinement
- Normal Mode Analysis (NMA) -> harmonic oscillator model
- Manifold learning

Problem Formulation — Forward Model

- The individual particle images are formed by:

$$\mathbf{y}_s = P_s\, \mathbf{x}_s + \epsilon_s \qquad \forall s = 1,2,\cdots,n$$

Volume rotation operator R_s convolution with a point spread function $\mathbf{h}_s$

- Molecular volumes: $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{R}^{N^3}$
- Particle images: $\mathbf{y}_1, \cdots, \mathbf{y}_n \in \mathbb{R}^{N^2}$
- Noise: $\epsilon \sim \mathcal{N}(0,\, \sigma^2 \mathrm{I}_{N\times N})$
- Linear imaging operators: $P_1, \cdots, P_n \in \mathbb{R}^{N^2 \times N^3}$

*To define the imaging operator, one must incorporate an **interpolation** scheme since the volumes lie on a discrete grid → express tomographic projection in the **Fourier domain***

- Discrete Fourier transform is given by:

$$\left(\mathcal{F}_d \mathbf{s}\right)(\mathbf{k}) := \sum_{\mathbf{u} \in M_N^d} e^{-2\pi i \langle \mathbf{k}, \mathbf{u} \rangle} \mathbf{s}[\mathbf{u}] \qquad \forall \mathbf{k} \in \mathbb{R}^d$$

- $\mathbf{k}$: wave vector
- $\mathbf{u}$: voxel index
- $\mathbf{s}$: d-dimensional signal

- Using the **Fourier slice theorem** to express the projection image in the Fourier domain as follows:

$$\left(\mathcal{F}_2 P_s \mathbf{x}_s\right)\left([k_1, k_2]^{\mathrm{T}}\right) = \left(\mathcal{F}_3 \mathbf{x}_s\right)\left(R_s^{-1}[k_1, k_2, 0]^{\mathrm{T}}\right) \cdot \boxed{\left(\mathcal{F}_2 \mathbf{h}_s\right)}\left([k_1, k_2]^{\mathrm{T}}\right)$$

Contrast transfer function (CTF)

Problem Formulation — Inverse Problem

- Homogeneous case (consider mean volume $\boldsymbol{\mu} \in \mathbb{R}^{N^3}$):

$$\mathbf{y}_s = P_s\boldsymbol{\mu} + \epsilon_s \qquad \forall s = 1,2,\cdots,n$$

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \sum_{s=1}^{n} \| \mathbf{y}_s - P_s\boldsymbol{\mu} \|^2$$

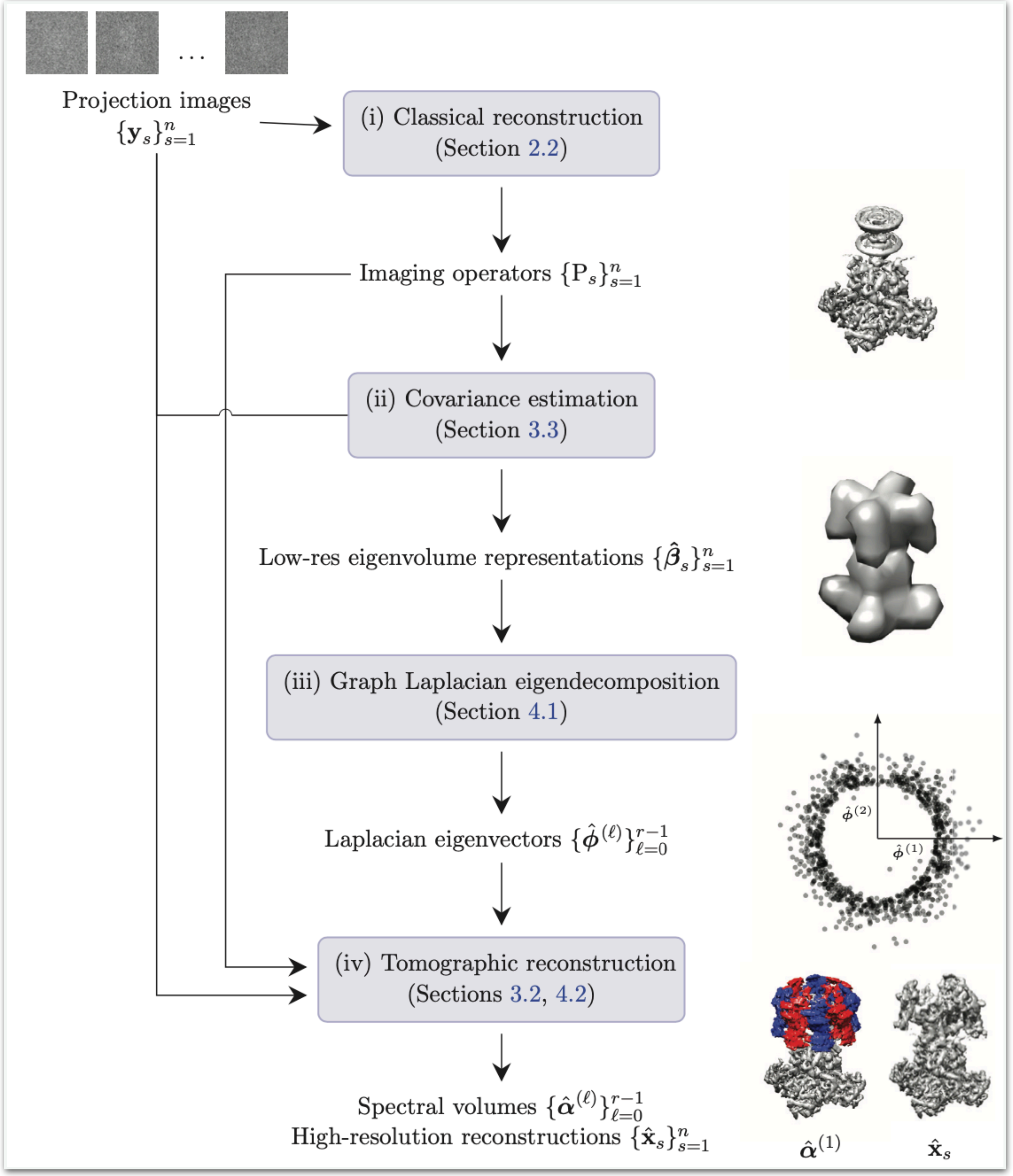
- Continuous heterogeneity:

$$\mathbf{y}_s = P_s\mathbf{x}_s + \epsilon_s \qquad \forall s = 1,2,\cdots,n$$

$$\hat{\mathbf{x}}_s = ?$$

- Two main assumptions made in this paper:
  - The molecular volumes in the sample lie near **a low-dimensional manifold**.
  - The imaging operators can be **accurately estimated** using standard cryo-EM reconstruction tools.

Pipeline of Method from this Paper



Molecular volumes sampled on a 3D voxel grid of dimension N^3

Continuous heterogeneity model — Volumes embedding on manifold of $\mathbb{R}^{N^3}$

Manifold Learning — Series expansion in Laplacian eigenfunctions

Estimation of Laplacian eigenfunctions (*requires the distribution of 3D volumes)	Estimation of expansion coefficient vectors (spectral volumes)
$\hat{\boldsymbol{\phi}}^{(0)}, \cdots, \hat{\boldsymbol{\phi}}^{(r-1)} \in \mathbb{R}^n$	$\hat{\boldsymbol{\alpha}}^{(0)}, \cdots, \hat{\boldsymbol{\alpha}}^{(r-1)} \in \mathbb{R}^{N^3}$

Define a high-resolution 3D reconstruction $\hat{\mathbf{x}}_s$ for each projection image

$$\hat{\mathbf{x}}_s := \sum_{\ell=0}^{r-1} \hat{\boldsymbol{\alpha}}^{(\ell)} \hat{\phi}_s^{(\ell)}$$

Low-Resolution Reconstruction

- Covariance estimation: each reconstructed volume is a linear combination of q **PCA eigenvolumes** $\Rightarrow \hat{\boldsymbol{\mu}} + \hat{V}_q \hat{\boldsymbol{\beta}}_s$

- Defines some mapping $(\mathbf{y}_s, P_s) \mapsto \boldsymbol{\beta}_s$ where $\boldsymbol{\beta}_s \in \mathbb{R}^q$ is the vector of eigenvolume coefficients corresponding to a low-dimensional representation of $\mathbf{x}_s$

- Ignore potential ambiguities due to the projection and consider the low-resolution reconstruction as a linear dimensionality reduction of the underlying volume $\mathbf{x}_s \mapsto \boldsymbol{\beta}_s$

Manifold Spectral Representation

- Approximation of molecular volumes using an orthogonal basis expansion of first r Laplacian eigenfunctions:

$$\mathbf{x} \approx \sum_{\ell=0}^{r-1} \underbrace{\boldsymbol{\alpha}^{(\ell)}}_{\text{Spectral volumes}} \underbrace{\phi^{(\ell)}\left(\boldsymbol{\beta}(\mathbf{x})\right)}_{\text{Laplacian eigenfunctions: } \mathbf{B} \mapsto \mathbb{R}} \qquad \forall \boldsymbol{\beta} \in \mathbf{B}$$

- Spectral volumes: $\boldsymbol{\alpha}^{(0)}, \cdots, \boldsymbol{\alpha}^{(r-1)} \in \mathbb{R}^{N^3}$
- Image of $\mathbf{x}$ in PCA coordinates: $\boldsymbol{\beta}(\mathbf{x}) \in \mathbf{B}$

- Employ **Laplacian eigenmap** from the field of manifold learning to obtain estimates:

$$\hat{\boldsymbol{\phi}}^{(0)}, \cdots, \hat{\boldsymbol{\phi}}^{(r-1)} \in \mathbb{R}^n$$

→ Build a **weighted undirected graph**, where the vertices correspond to the projection images $\mathbf{y}_1, \cdots, \mathbf{y}_n$ and the edge weights W_{ij} are estimates of the affinity between the underlying molecular conformations.

Gaussian kernel weight: $W_{ij} = e^{\frac{-\left\| \hat{\beta}_i - \hat{\beta}_j \right\|^2}{2\sigma^2}}$

- A data-driven variant of the spectral expansion:

$$\mathbf{x}_s \approx \sqrt{n} \sum_{\ell=0}^{r-1} \boldsymbol{\alpha}^{(\ell)} \hat{\phi}_s^{(\ell)} \qquad \forall s = 1, 2, \cdots, n$$

$\sqrt{n}$ factor is needed for normalization: $\sum_{s=1}^{n} \left(\hat{\phi}_s^{(\ell)} \right)^2 = 1$

Generalized Tomographic Reconstruction

- Applying the imaging matrix P_s for the spectral expansion and plugging in the forward model:

$$\mathbf{y}_s \approx \sqrt{n} \sum_{\ell=0}^{r-1} \left(P_s \boldsymbol{\alpha}^{(\ell)}\right) \hat{\phi}_s^{(\ell)} , \quad \forall s = 1,2,\cdots,n$$

- Seek spectral volumes $\hat{\boldsymbol{\alpha}}^{(0)}, \cdots, \hat{\boldsymbol{\alpha}}^{(r-1)}$ that minimize the squared error:

$$\left(\hat{\boldsymbol{\alpha}}^{(0)}, \cdots, \hat{\boldsymbol{\alpha}}^{(r-1)}\right) := \arg\min_{\boldsymbol{\alpha}} \sum_{s=1}^{n} \left\| \mathbf{y}_s - \sqrt{n} \sum_{\ell=0}^{r-1} \left(P_s \boldsymbol{\alpha}^{(\ell)}\right) \hat{\phi}_s^{(\ell)} \right\|^2$$

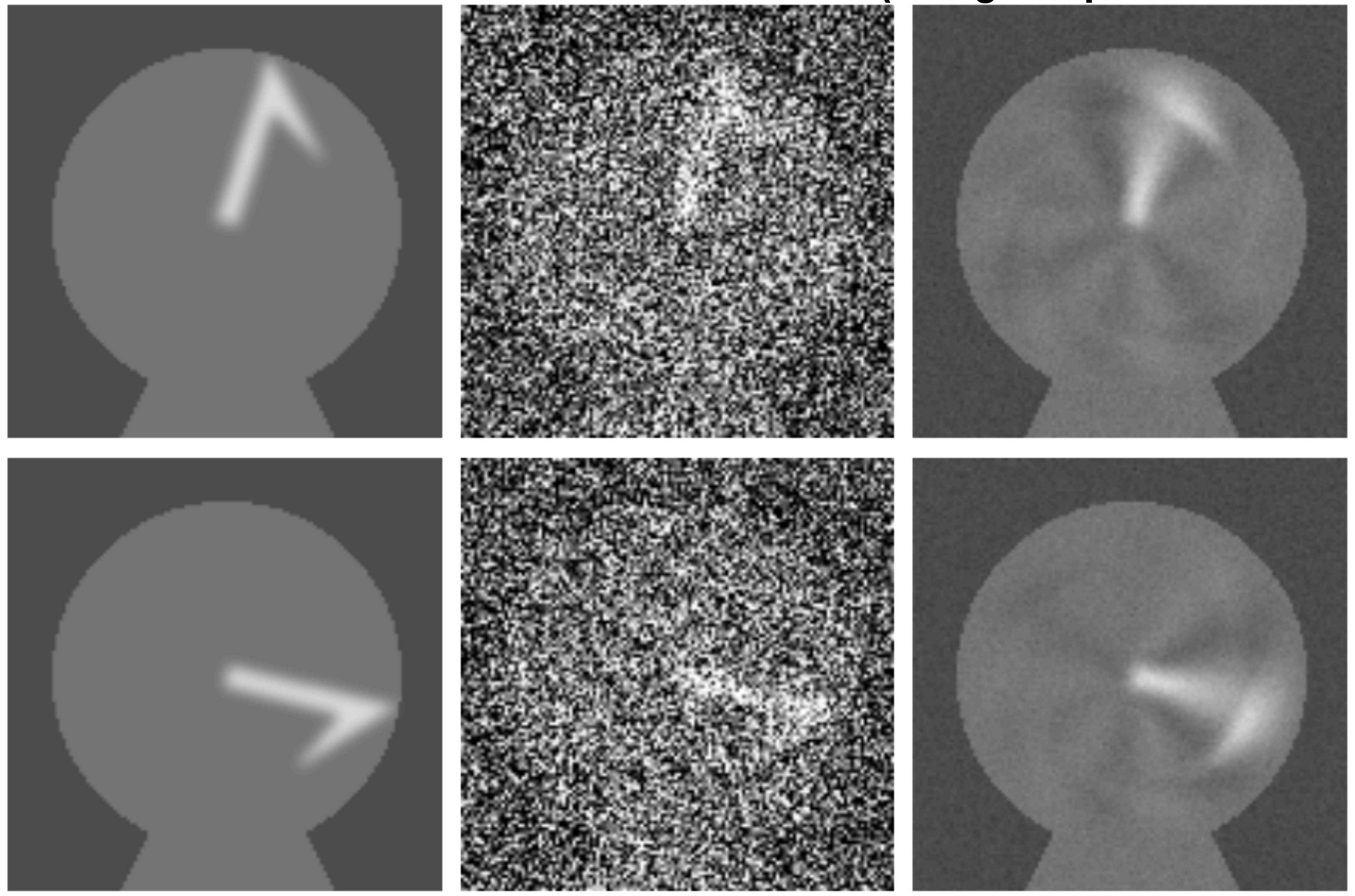
- The high-resolution reconstructions of the molecular volumes are now given by:

$$\hat{\mathbf{x}}_s = \sqrt{n} \sum_{\ell=0}^{r-1} \hat{\boldsymbol{\alpha}}^{(\ell)} \hat{\phi}_s^{(\ell)} , \quad \forall s = 1,2,\cdots,n$$

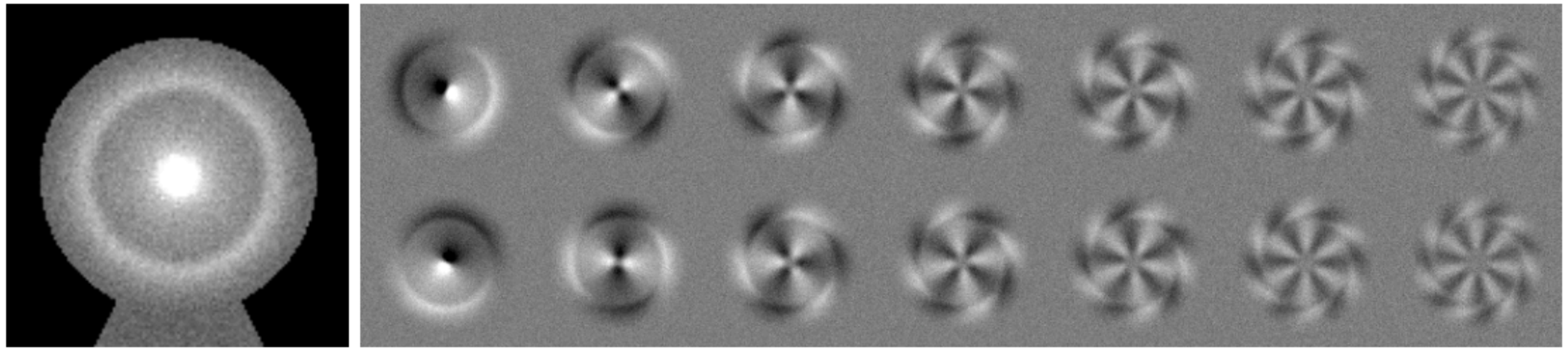
This estimator generalizes the least-squares estimator form ***a single mean volume*** $\hat{\mu}$ *to* ***multiple volumes*** $\hat{\alpha}^{(0)}, \cdots, \hat{\alpha}^{(r-1)}$ *whose contribution to the reconstructed volumes is given by the* ***Laplacian eigenvectors*** $\hat{\phi}^{(0)}, \cdots, \hat{\phi}^{(r-1)}$

Result — 2D/3D Clock Face Dataset

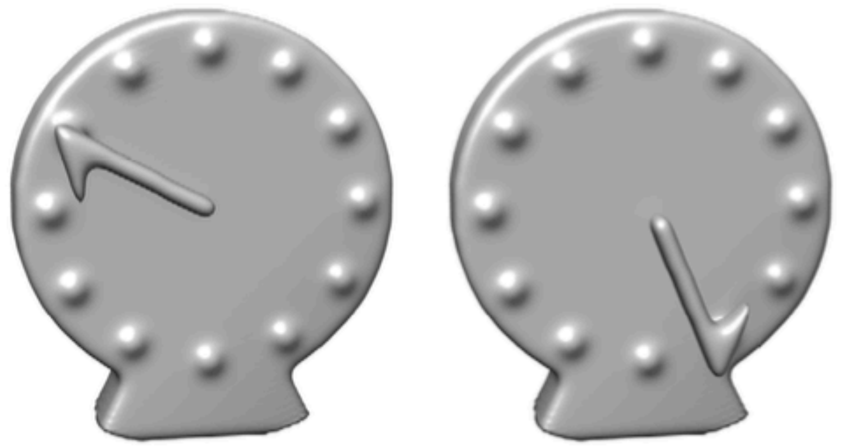
Clean clock face | Noisy clock face | Reconstruction (Using 15 spectral volumes)



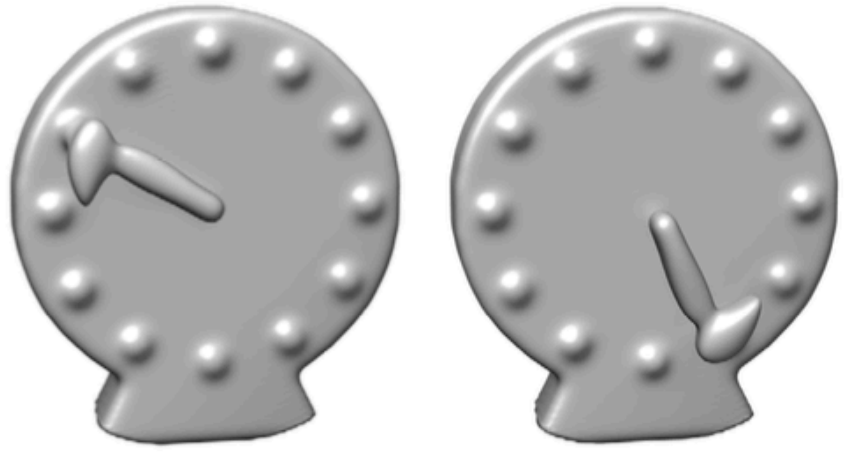
Spectral volumes



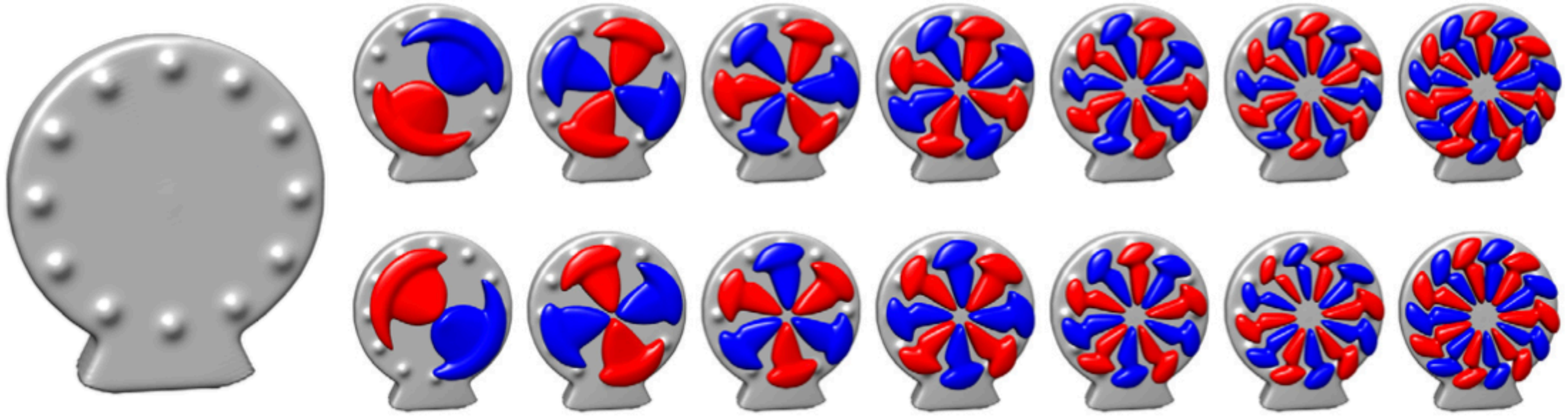
- Clock image: $\mathbf{z}_1, \cdots, \mathbf{z}_n \in \mathbb{R}^{N \times N}$
- Affinity matrix: $W_{ij} = e^{\frac{-\|\mathbf{z}_i - \mathbf{z}_j\|^2}{N^2 \sigma^2}}$



Input model



Reconstruction
(Using 15 spectral volumes)

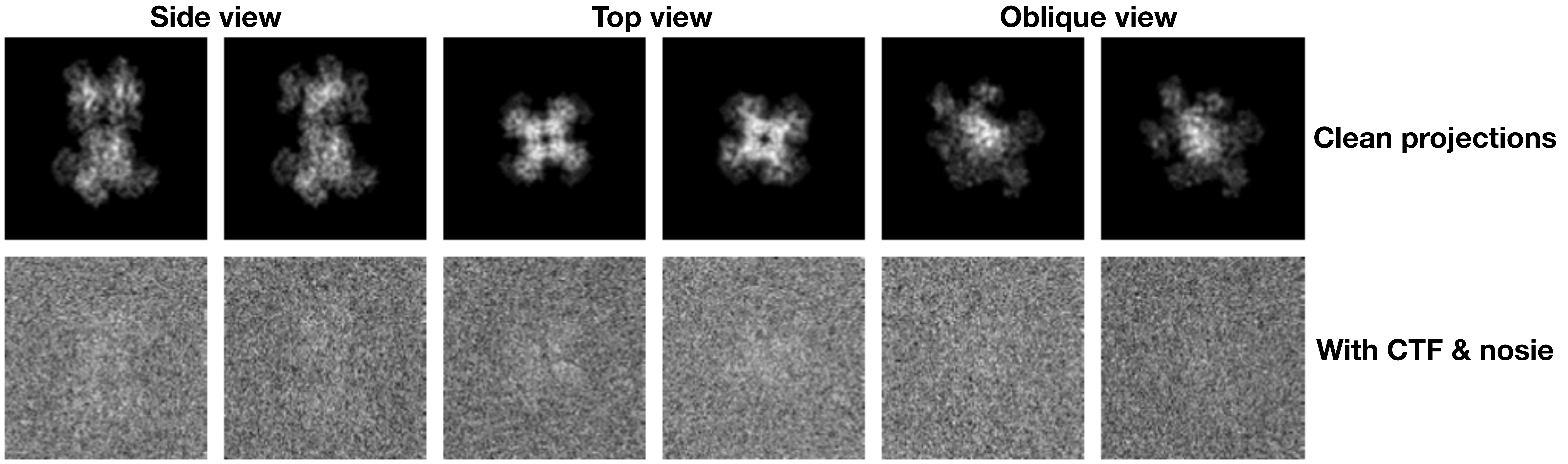


Red and blue represent negative and positive values of the higher-order spectral volume, respectively.

Result — Simulated Ion Channel

ChannelSpin

ChannelScretch

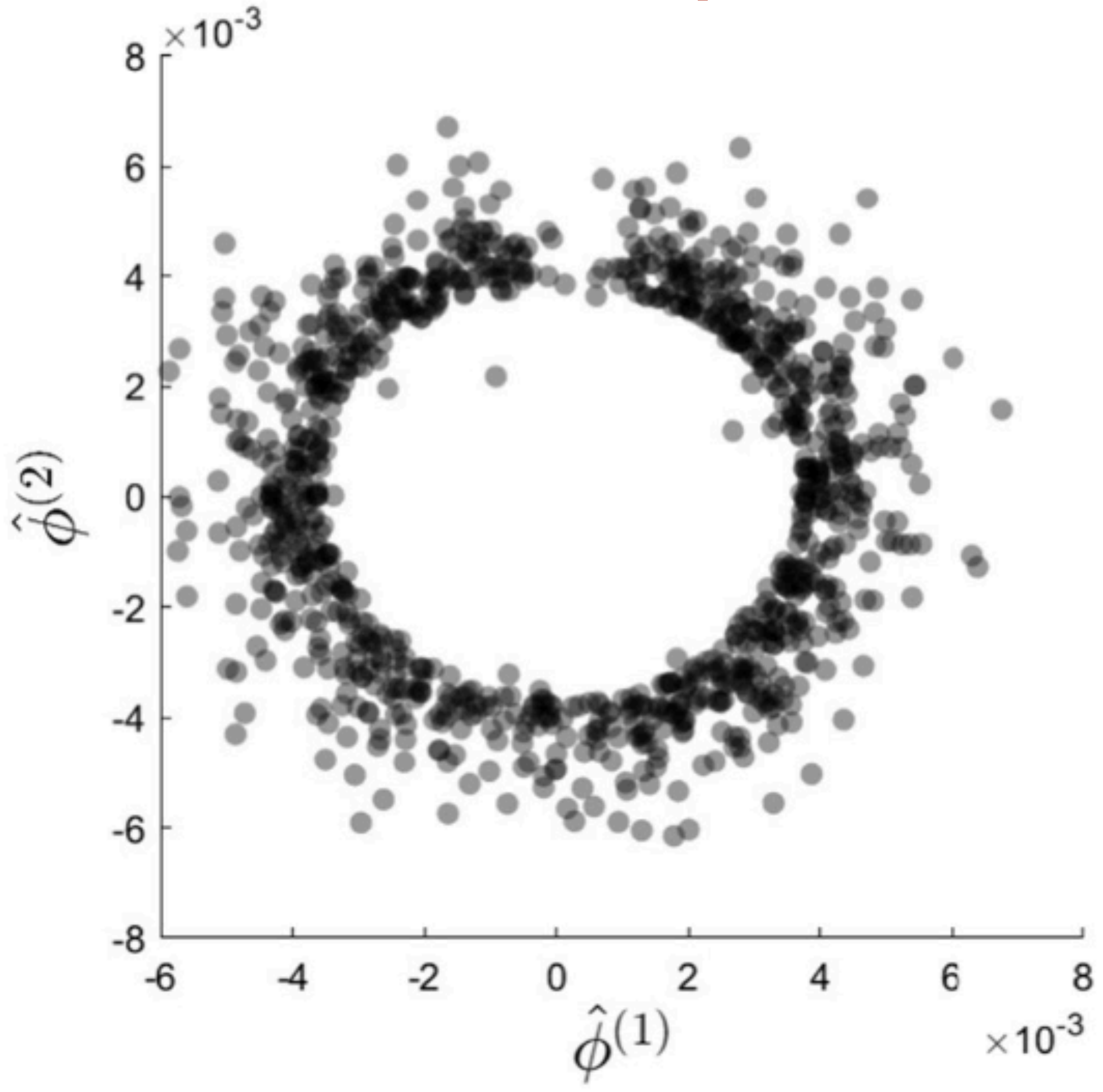


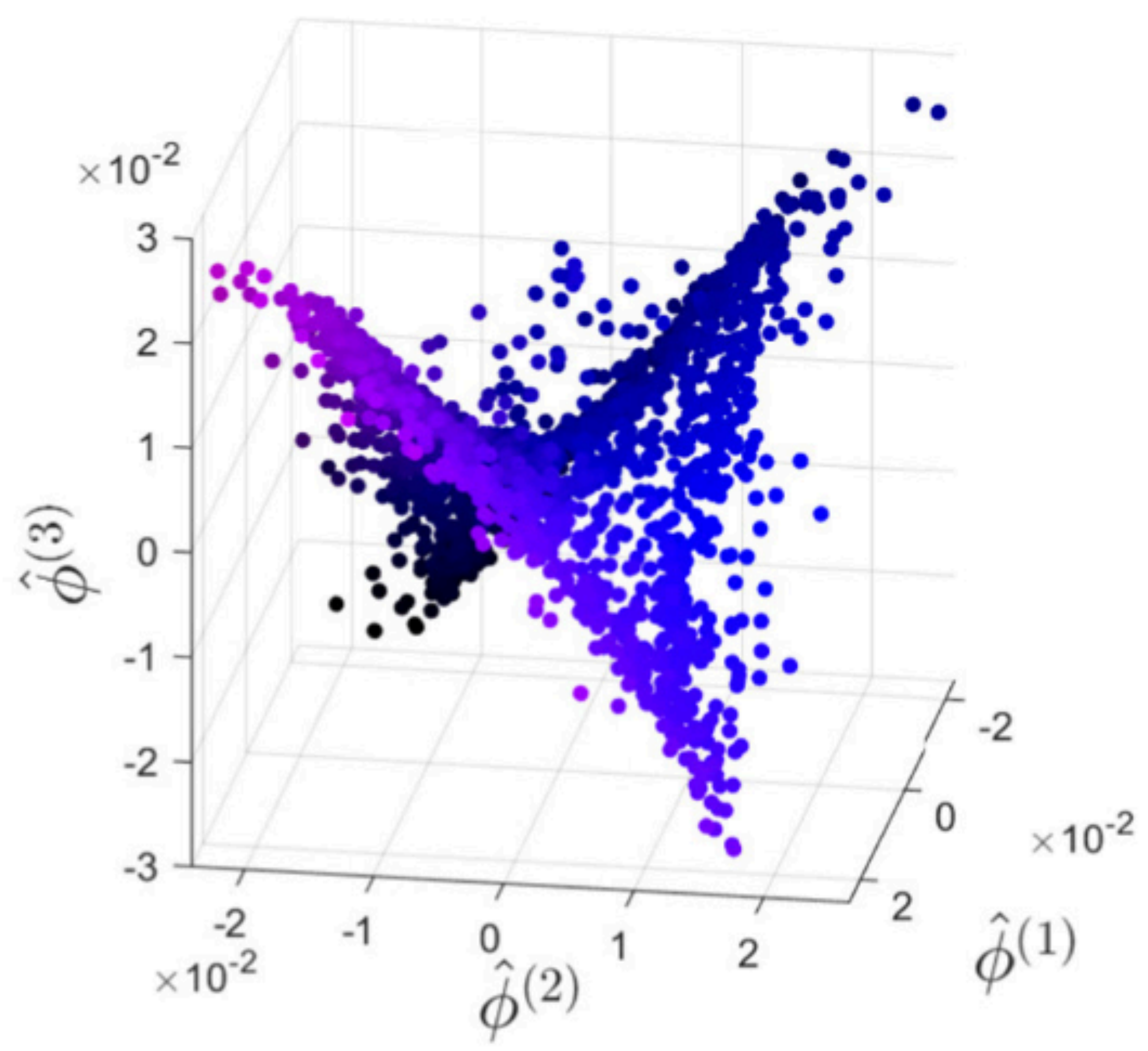
- Dataset 1 — **ChannelSpin** : a rotational motion of the top part about the z axis
- Dataset 2 — **ChannelScretch** : a nonrigid stretching of the bottom part along the $x-y$ plane

- Special resolution $N = 108$
- #volumes $n = 10000$ per dataset

- Test conditions:
  - Total energy of the noise was 30 times that of the total energy of each clean image.
  - No in-plane shift was applied.
  - Reconstructed the volumes using $r = 1,\cdots,15$ spectral volumes.
  - Used the true orientations of the projection images for both the covariance and spectral volume estimation procedures.

Examining the Laplacian Eigenmaps Embedding





- Random displacement: $\delta_x, \delta_y \in \{-16, -15, \cdots, 16\}$
- Original ion channel: $\mathbf{v} \in \mathbb{R}^{N \times N \times N}$
- The stretched ion channel $\mathbf{v}'$ is defined for every $0 \leq z \leq N/2$ by:

$$\mathbf{v}'[x, y, z] = \mathbf{v}[x + \delta_x s_z \ , \ y + \delta_y s_z \ , \ z] \quad \text{where} \quad s_z = \left(\frac{N/2 - z}{N/2 - z_0} \right)^2$$

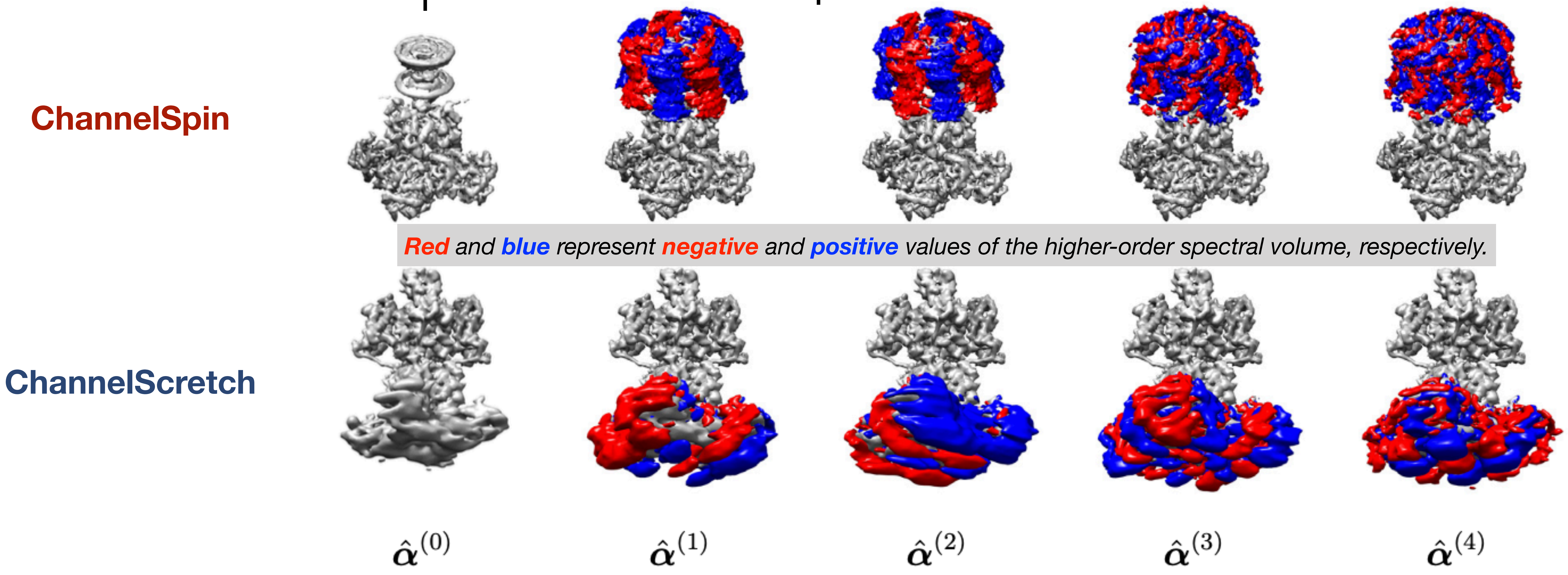
- The embedding of **ChannelSpin** clearly shows a circle.
- The embedding of the **ChannelScretch** dataset shows a 2-dimensional square in the $\hat{\phi}_s^{(1)} - \hat{\phi}_s^{(2)}$ plane that is shaped like a saddle.
- Both of these results are in accordance with the underlying motion manifold.

Examining the Spectral Volumes

Captures the mean over all rotations

Higher order spectral volumes have increasing angular frequency, capturing more and more detail.

Due to the C_4 symmetry of the ion channel, the lowest frequency has a period of 90 degrees.

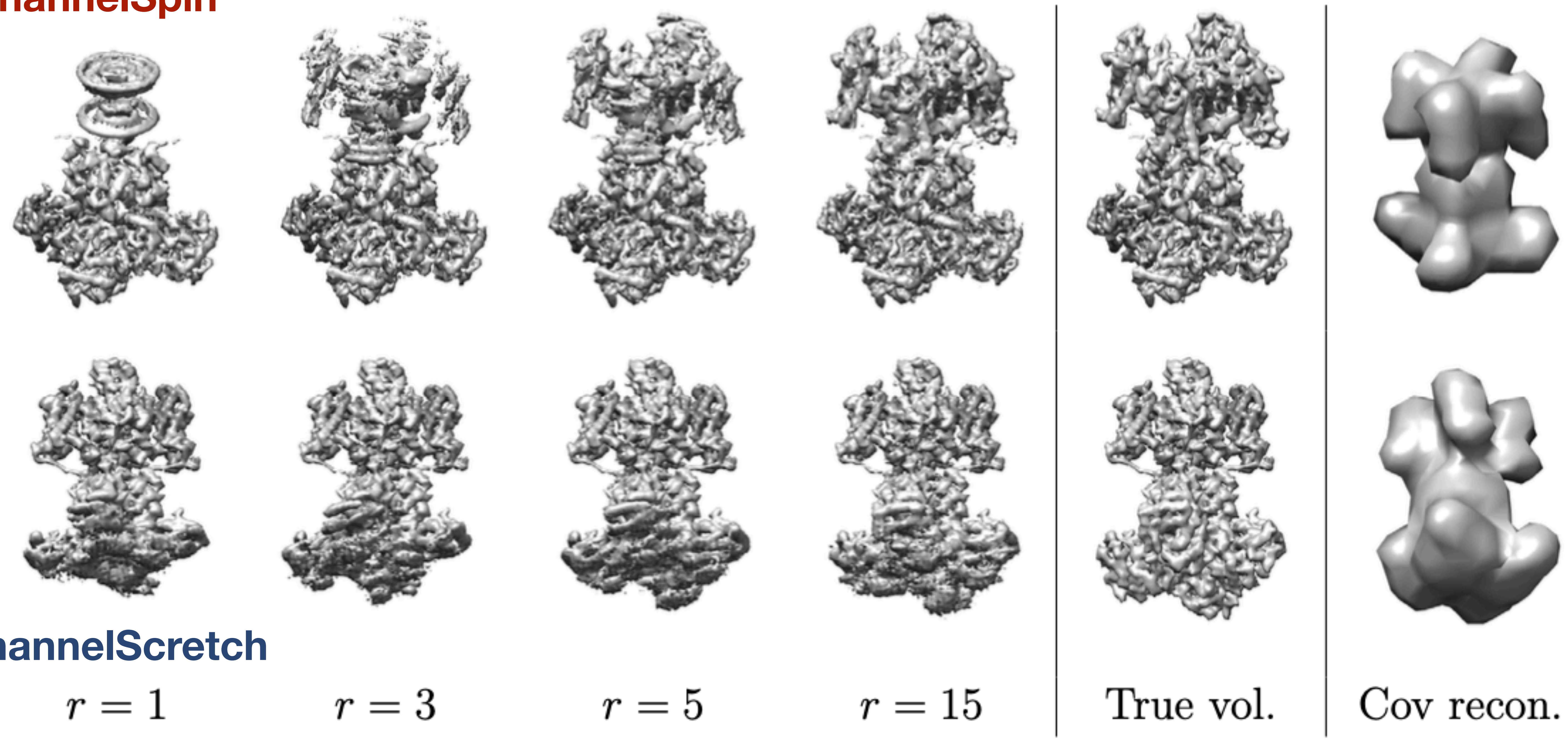


Captures the fixed part of the molecule with high resolution and shows a 'smeared' bottom portion

The first and second spectral volumes each have a low spatial frequency along the *x* and *y* axes

Higher spectral volumes show higher spatial frequencies

Reconstruction Accuracy and Runtime



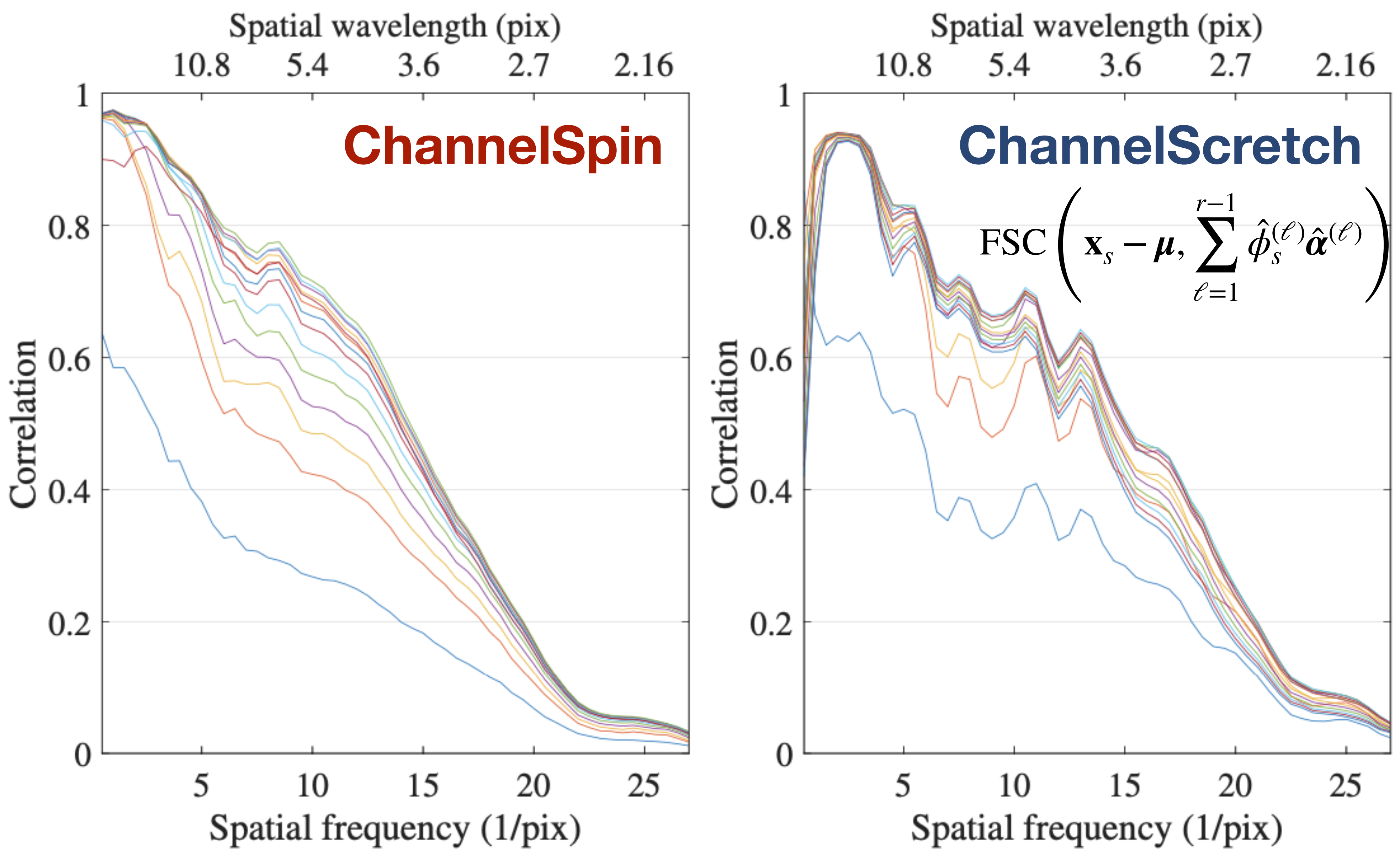


Table 2. Runtimes for the main steps of our method on the ChannelSpin dataset, with $n = 10\,000$ images of 108×108 pixels.

Procedure	Running time (s)
Calculation of $\hat{\boldsymbol{\mu}}$	624.6
Calculation of $\hat{\Sigma}$	5044.7
Calculation $\hat{V}_q$	0.8
Calculation of $\{\hat{\boldsymbol{\beta}}_s\}$	2084.8
Calculation of $\{\hat{\boldsymbol{\phi}}_s\}$	531.5
Calculation of K	12 378.0
Calculation of $\mathbf{b}$	4014.1
Estimation of $\{\hat{\boldsymbol{\alpha}}^{(\ell)}\}_{\ell=0}^{15}$	1769.9

Back Up

Name	Domain	Description
n	$\mathbb{N}$	Number of images and underlying molecular volumes
s	$1,\ldots,n$	Index to molecular image/volume
N	$\mathbb{N}$	Image/volume size
$\check{N}$	$\mathbb{N}$	Downsampled image/volume size
$\mathbf{x}, \mathbf{x}_s$	$\mathbb{R}^{N^3}$	Molecular volume
$\hat{\mathbf{x}}_s$	$\mathbb{R}^{N^3}$	Our high-resolution molecular volume estimate
$\mathbf{u}$	$\{1,\ldots,N\}^3$	Voxel index
$\mathbf{y}, \mathbf{y}_s$	$\mathbb{R}^{N^2}$	Molecular image
$\mathbf{h}, \mathbf{h}_s$	$\mathbb{R}^{N^2}$	Contrast transfer function (CTF)
R, R_s	$\mathrm{SO}(3)$	3D viewing orientation
P, P_s	$\mathbb{R}^{N^2 \times N^3}$	Imaging matrix (rotation, projection, and CTF)
$\mathcal{F}_d$		The d-dimensional discrete Fourier transform (DFT)
M_N	$\mathbb{R}^N$	Sampling grid in $[-1,+1)$ used to define the DFT
$\boldsymbol{\mu}$	$\mathbb{R}^{N^3}$ or $\mathbb{R}^{\check{N}^3}$	Mean volume (high-res or low-res)
Σ	$\mathbb{R}^{\check{N}^3 \times \check{N}^3}$	Covariance matrix of downsampled molecular volumes
q	$\mathbb{N}$	Number of PCA eigenvolumes
$\hat{V}_q$	$\mathbb{R}^{\check{N}^3 \times q}$	Eigenvolumes of the estimated covariance matrix
$\boldsymbol{\beta}(\mathbf{x}), \boldsymbol{\beta}_s$	$\mathbb{R}^q$	PCA coordinates of a molecular volume
$\mathbf{B}$	$\subseteq \mathbb{R}^q$	The domain of PCA coordinates
$\nu(\mathbf{B})$		Measure of volumes in PCA coordinate representation
$\mathbf{W}$	$\mathbb{R}^{n \times n}$	Edge weights matrix
L	$\mathbb{R}^{n \times n}$	Graph Laplacian matrix
$\mathcal{M}$	$\subset \mathbb{R}^{N^3}$	Riemannian submanifold of molecular volumes
$\phi^{(\ell)}$	$\mathbf{B} \to \mathbb{R}$	Laplace–Beltrami eigenfunction of the ℓth smallest eigenvalue
$\hat{\boldsymbol{\phi}}^{(\ell)}$	$\mathbb{R}^n$	Laplacian eigenvector of the ℓth smallest eigenvalue
r	$\mathbb{N}$	Number of spectral volumes
K	$\mathbb{R}^{rN^3 \times rN^3}$	Matrix of weighted projection-backprojections
$\mathbf{b}$	$\mathbb{R}^{rN^3}$	Concatenation of weighted back-projection images
$\boldsymbol{\alpha}^{(\ell)}$	$\mathbb{R}^{N^3}$	Spectral volumes