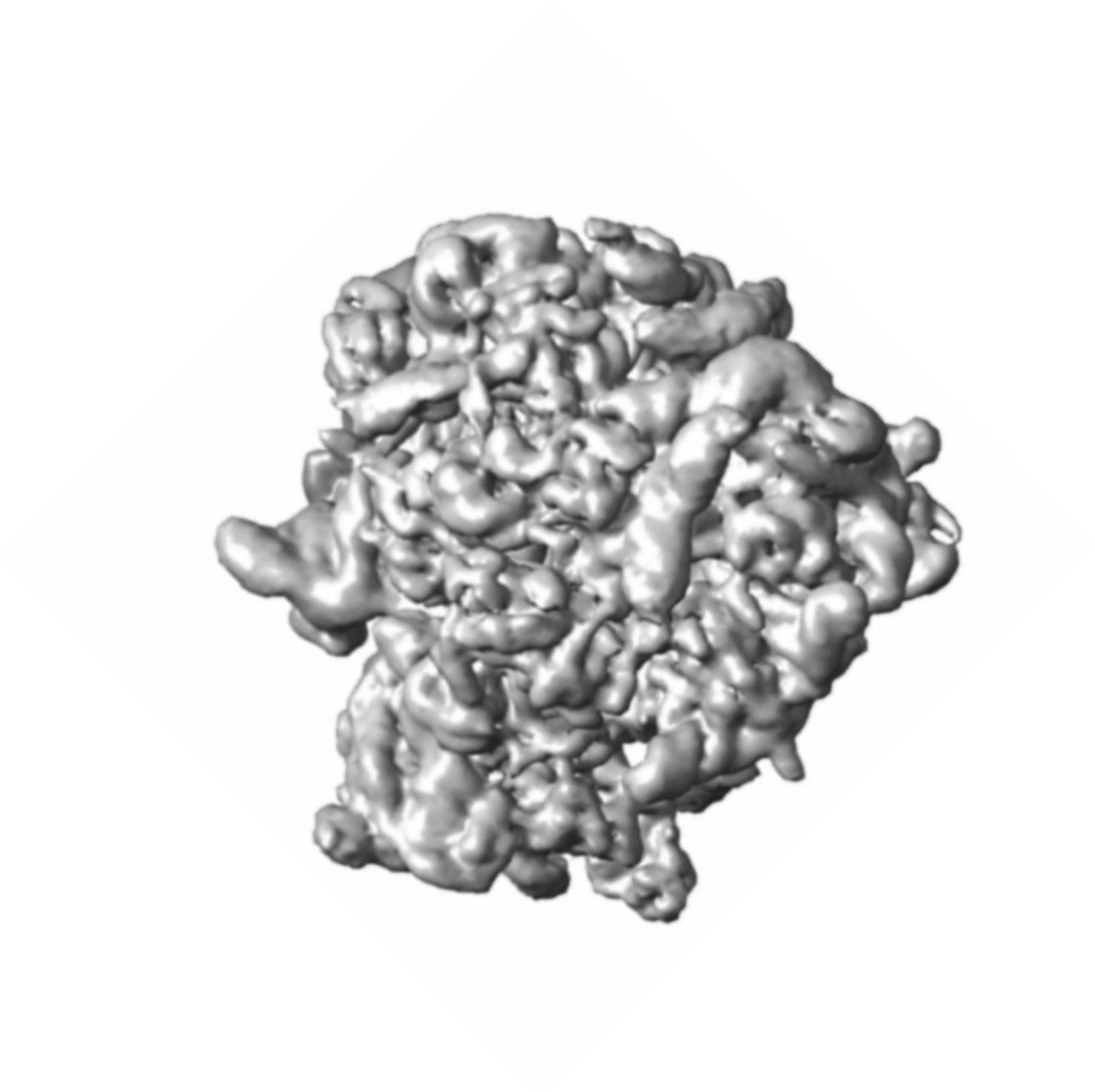
UNIVERSITY OF TORONTO
VECTOR INSTITUTE
YORK UNIVERSITÉ UNIVERSITY

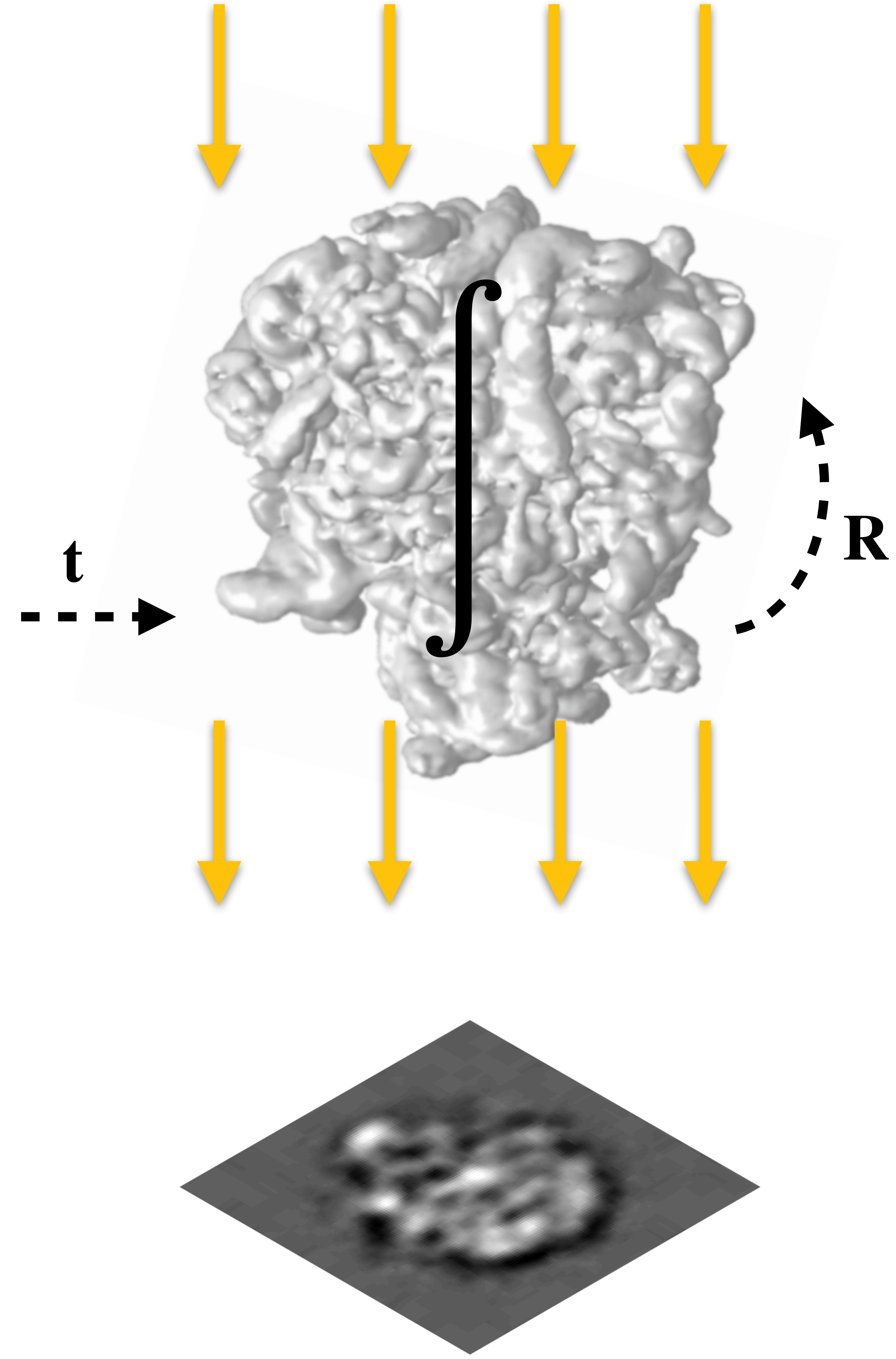
CryoSPIRE

Reconstructing Heterogeneous Biomolecules via Hierarchical Gaussian Mixtures and Part Discovery

Project Webpage

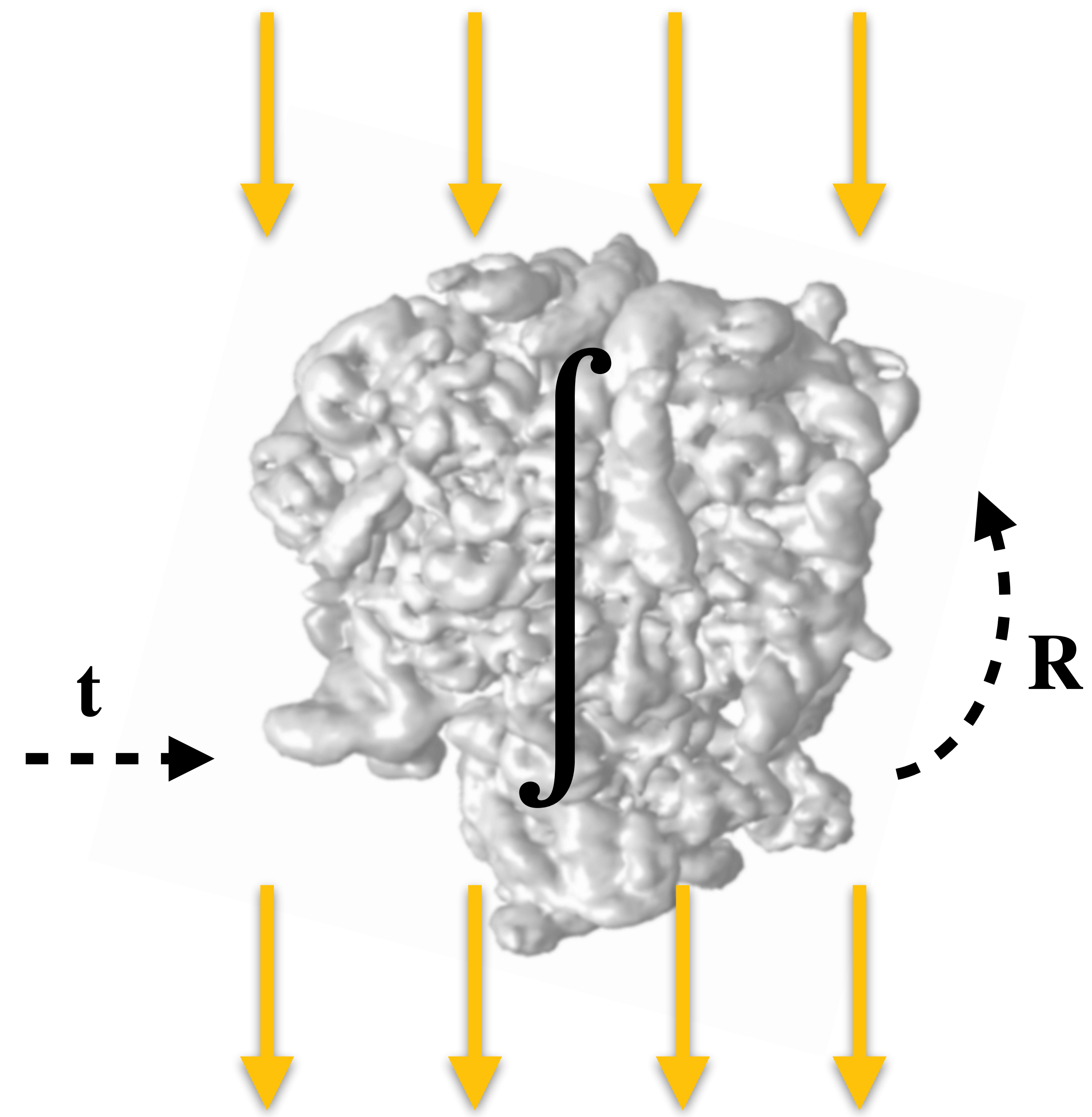


$$V(\quad \boldsymbol{p} \quad)$$

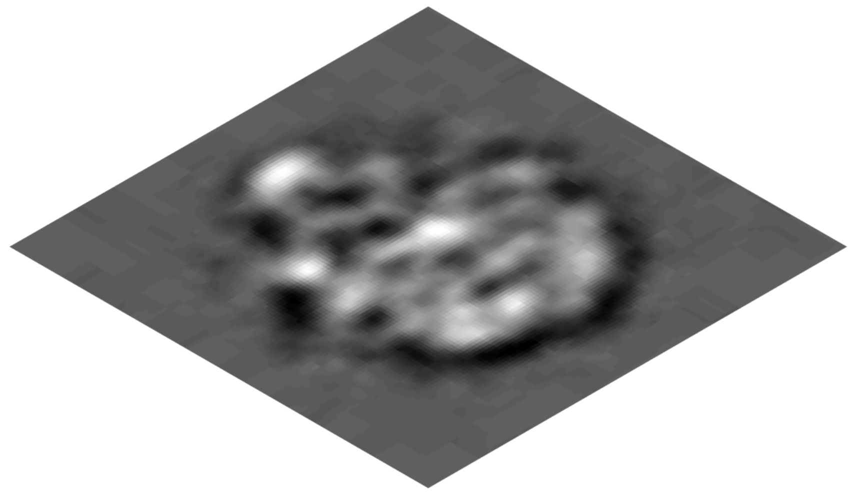


Clean Projection

$$\int_{\mathbb{R}} V(\boldsymbol{R}^T \boldsymbol{p} + \boldsymbol{t})\ dz$$



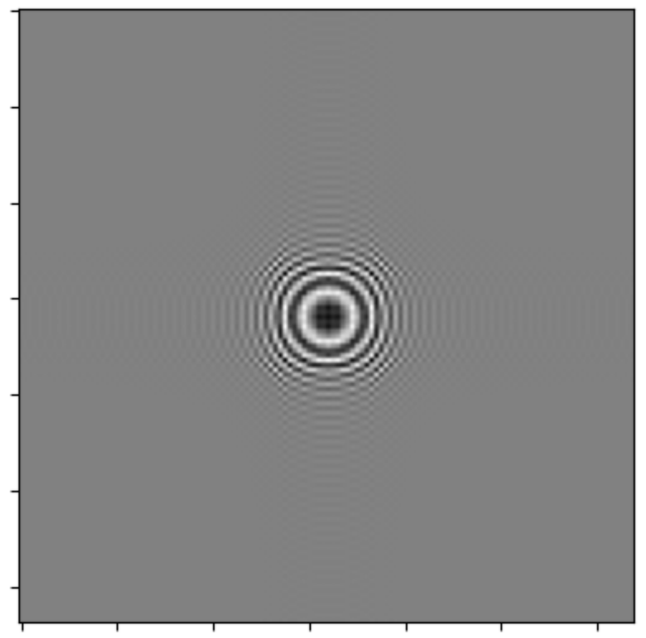
$$I(x, y) = h \star \int_{\mathbb{R}} V(\boldsymbol{R}^T \boldsymbol{p} + \boldsymbol{t})\; dz \; + \; n$$



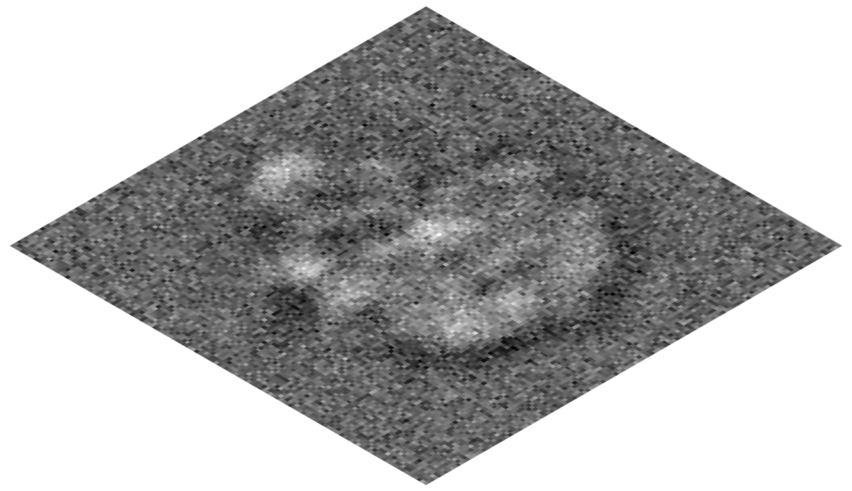
Clean Projection

$\star$

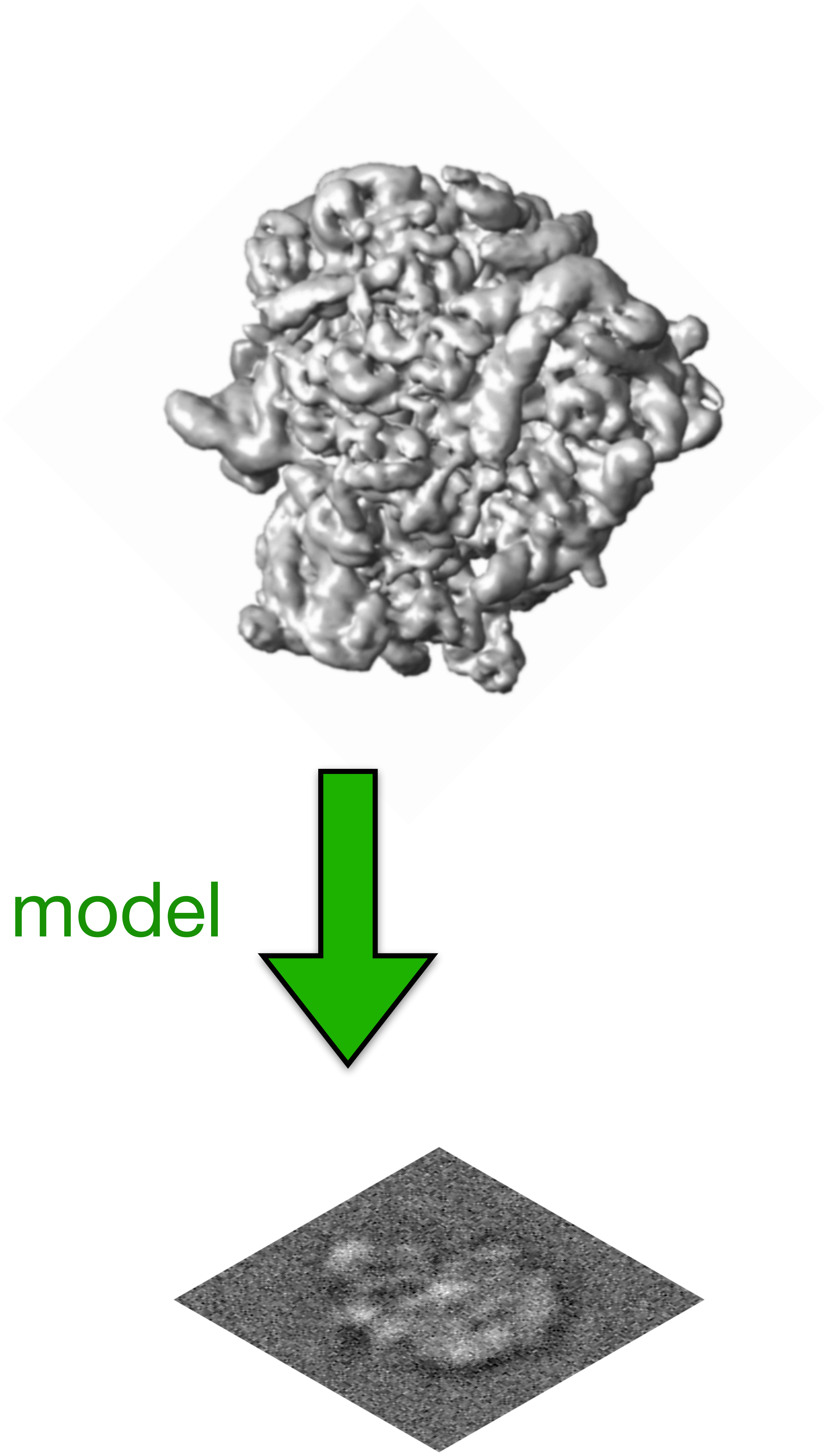
PSF



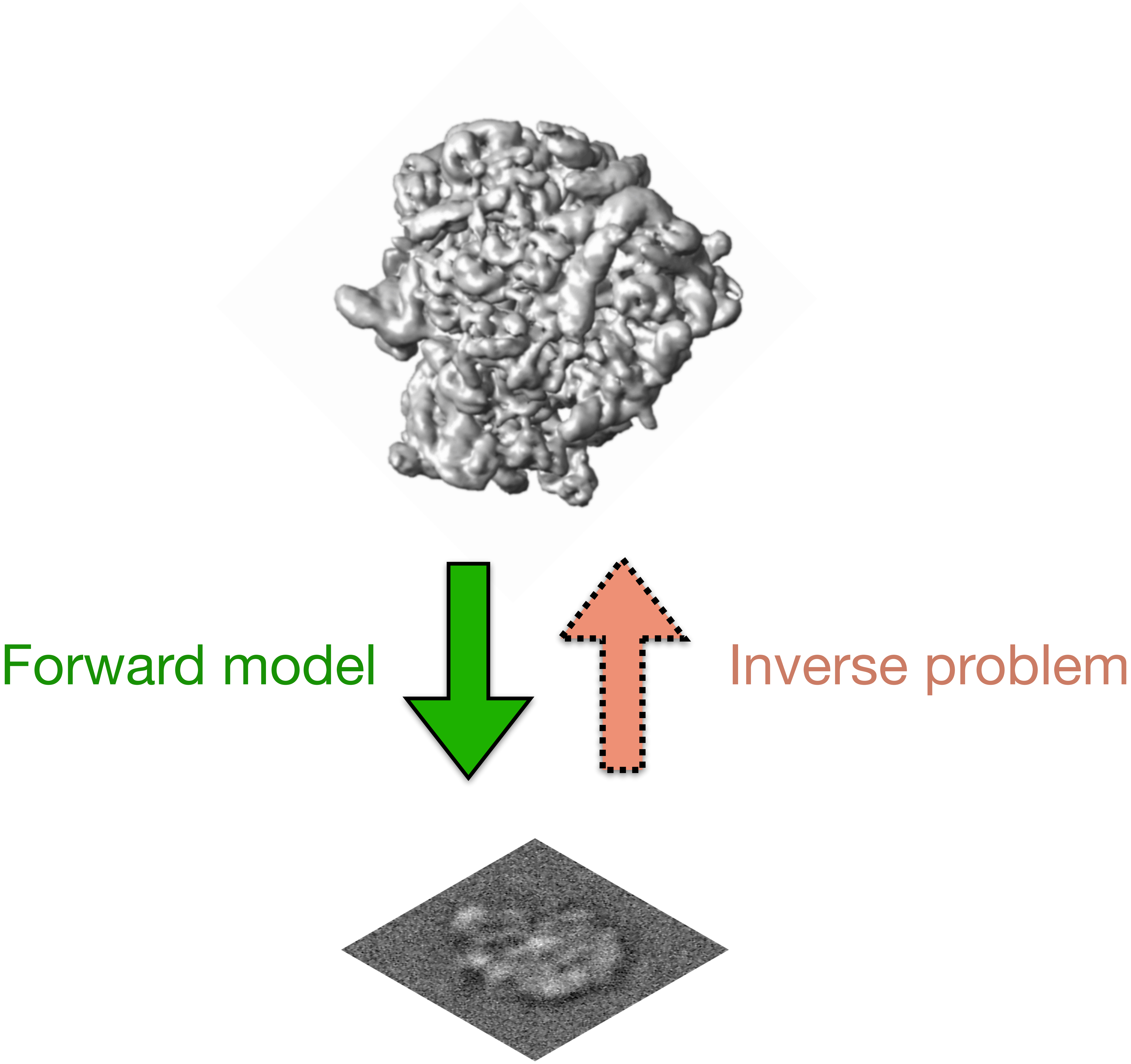
$+ \; n \; =$



Noisy Projection

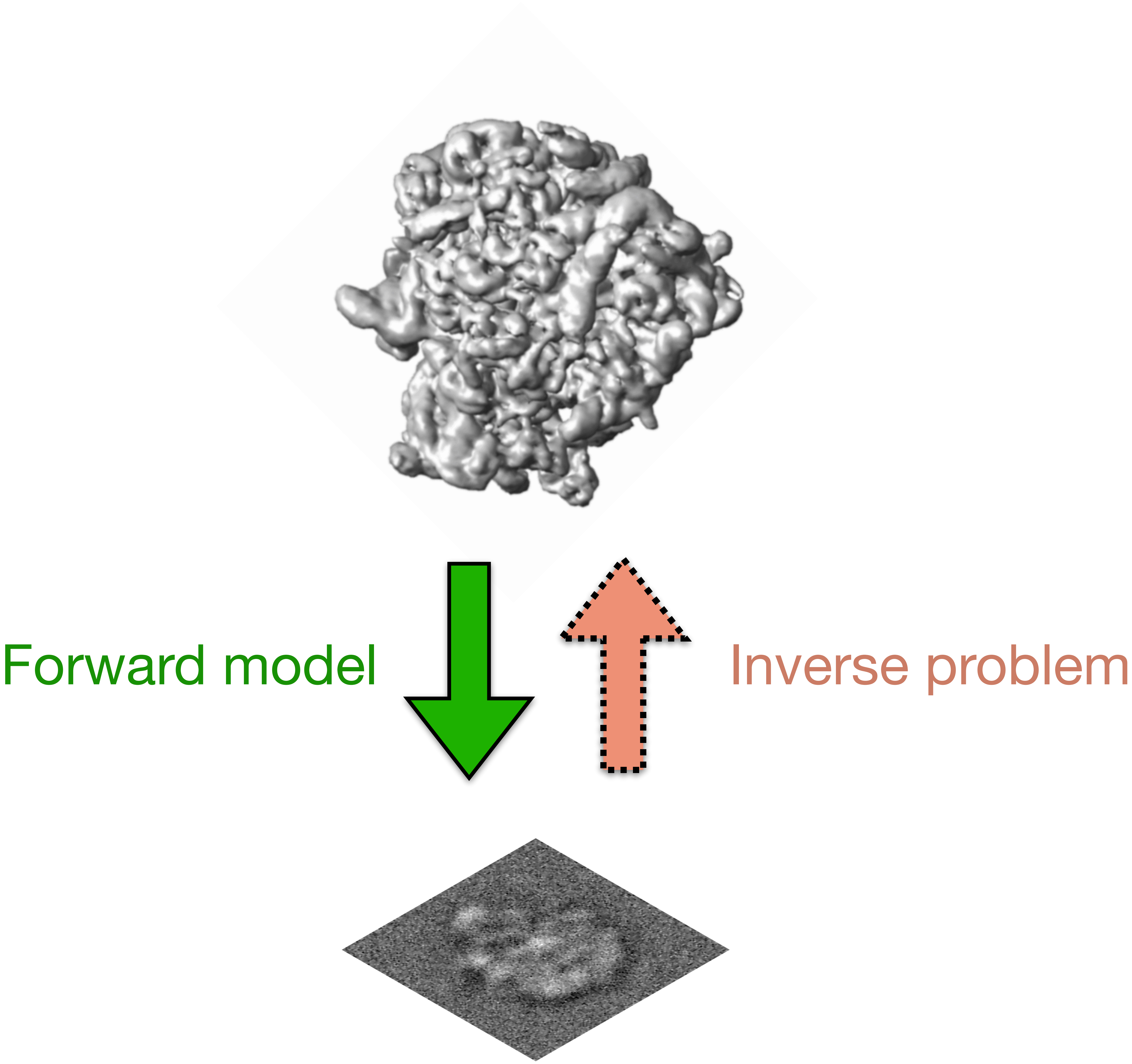


Forward model



Forward model

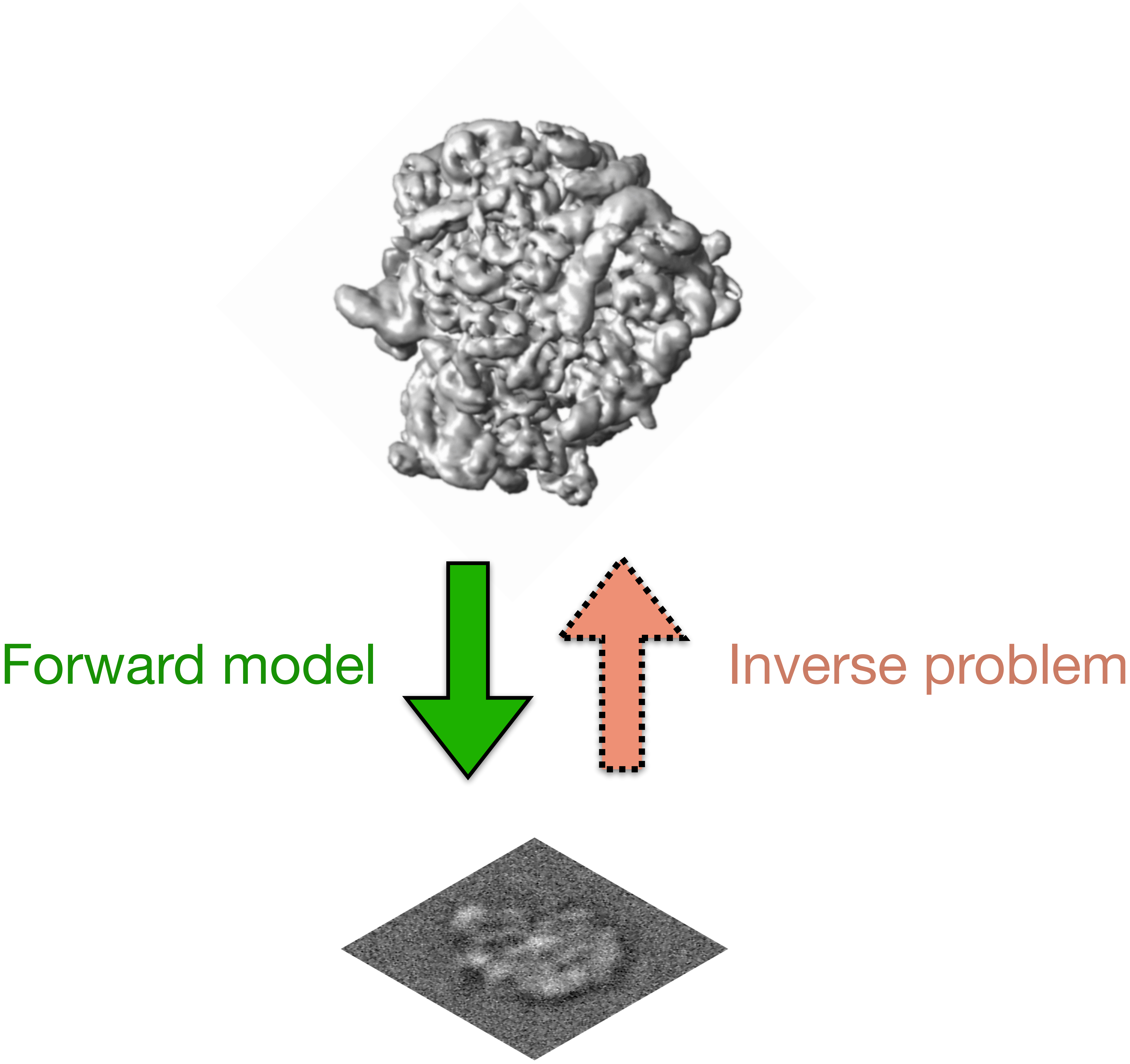
Inverse problem



Forward model

Inverse problem

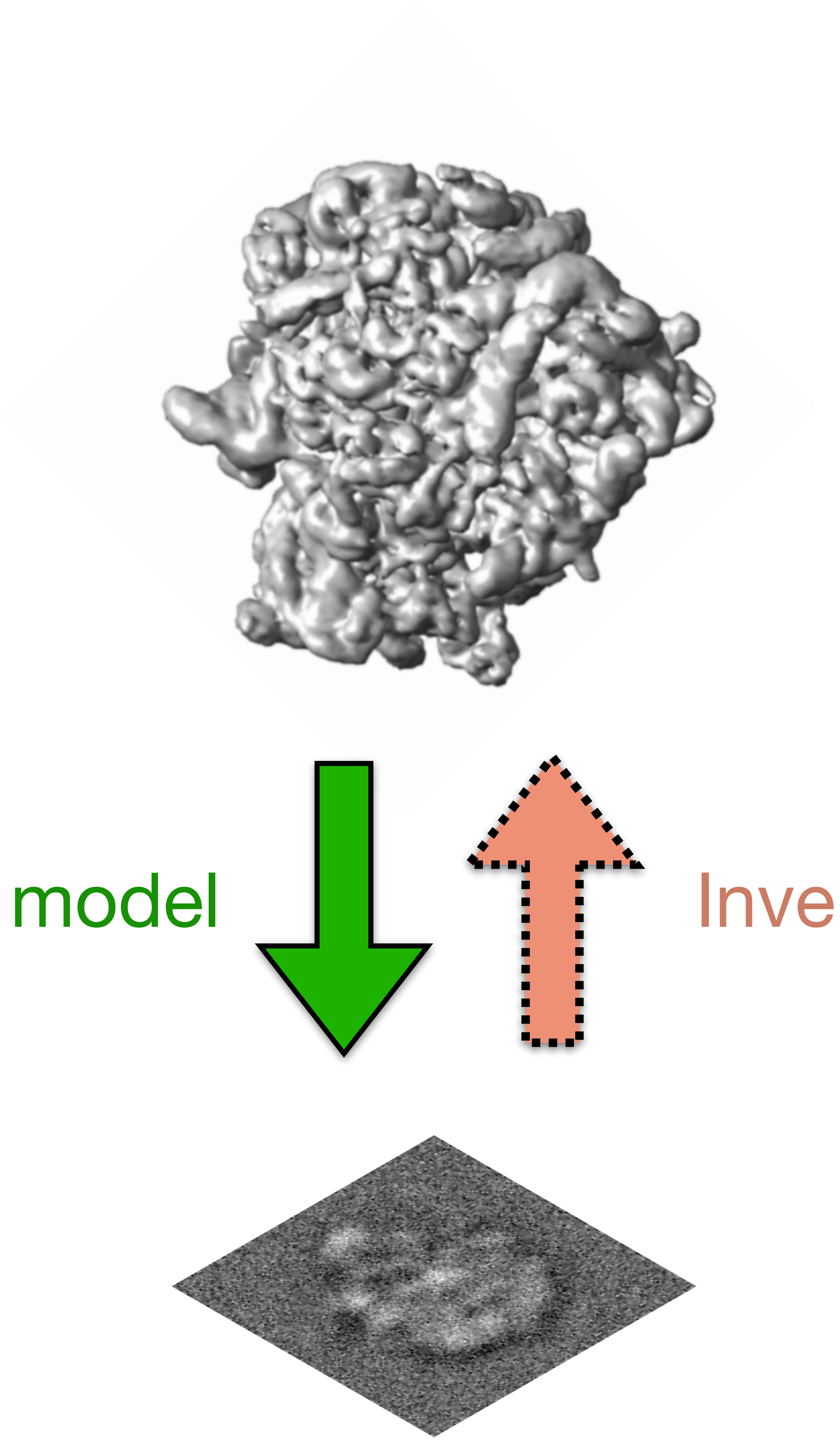
⚠ Unknown poses (R, t)



Forward model

Inverse problem

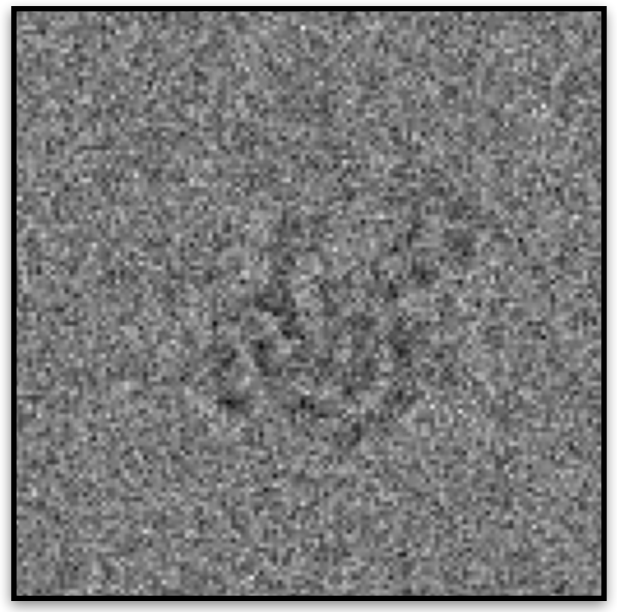
- ⚠ Unknown poses (R, t)
- ⚠ High Noise - Poor Signal

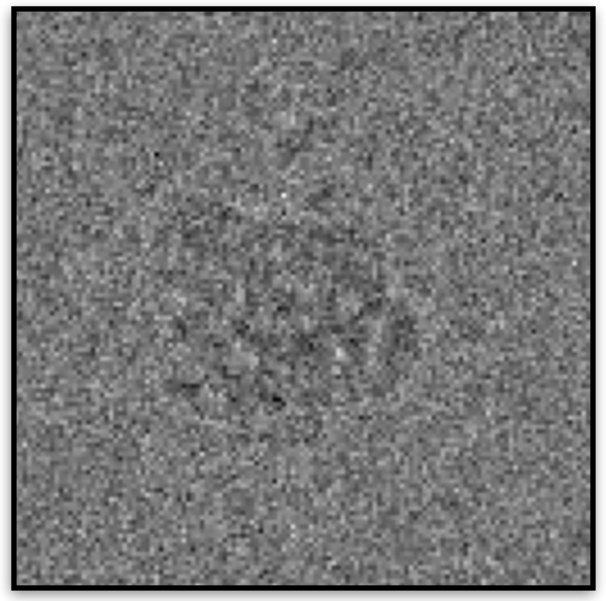


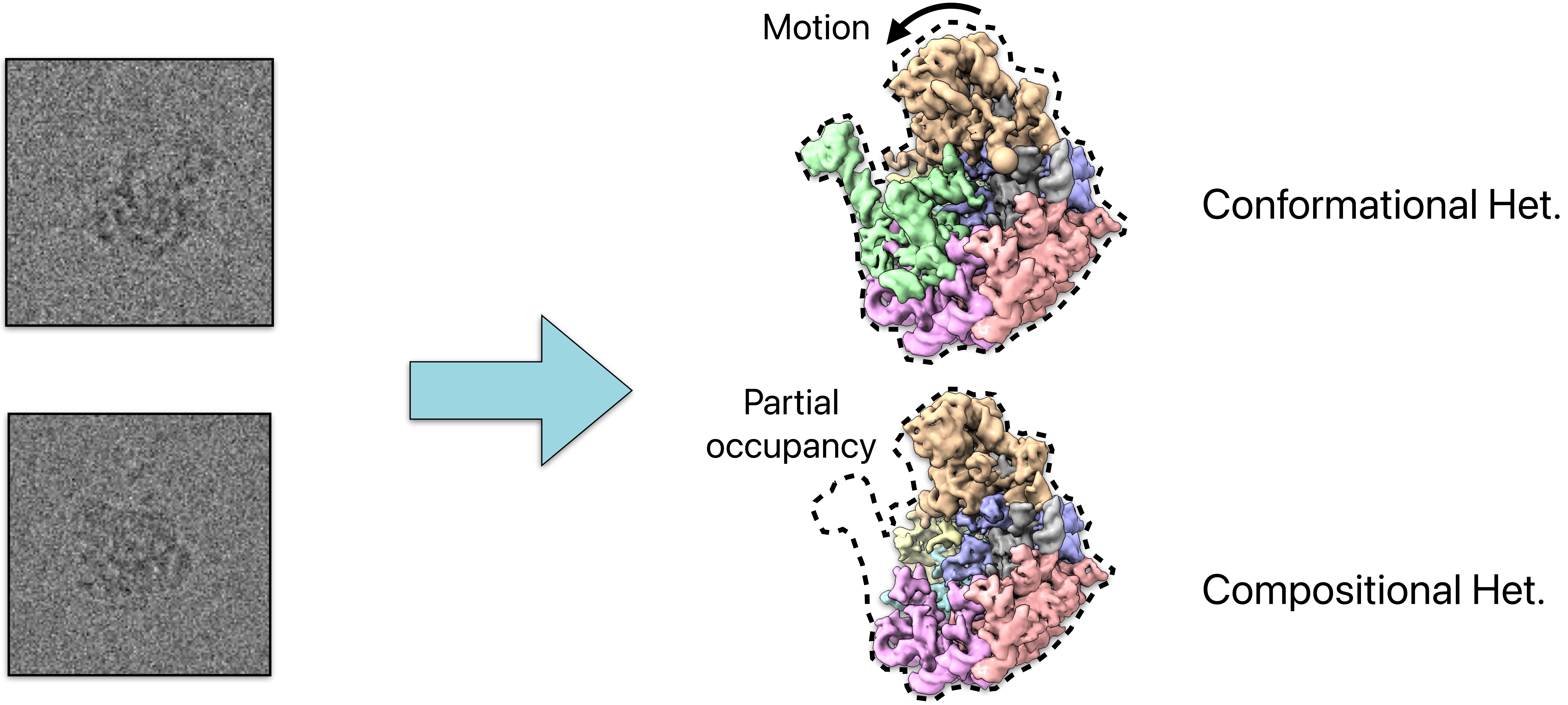
Forward model

Inverse problem

- ⚠ Unknown poses (R, t)
- ⚠ High Noise - Poor Signal
- ⚠ Structural heterogeneity





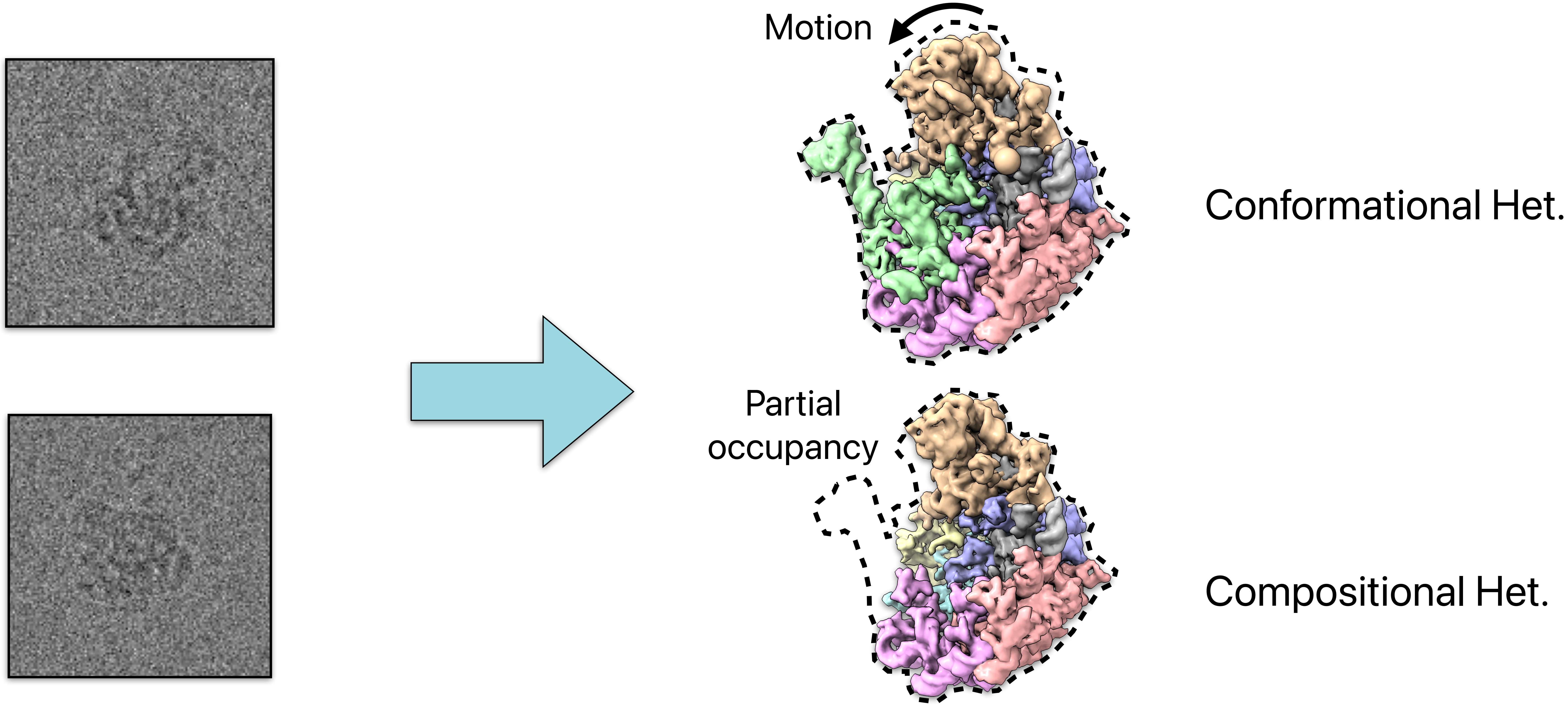


Motion

Conformational Het.

Partial
occupancy

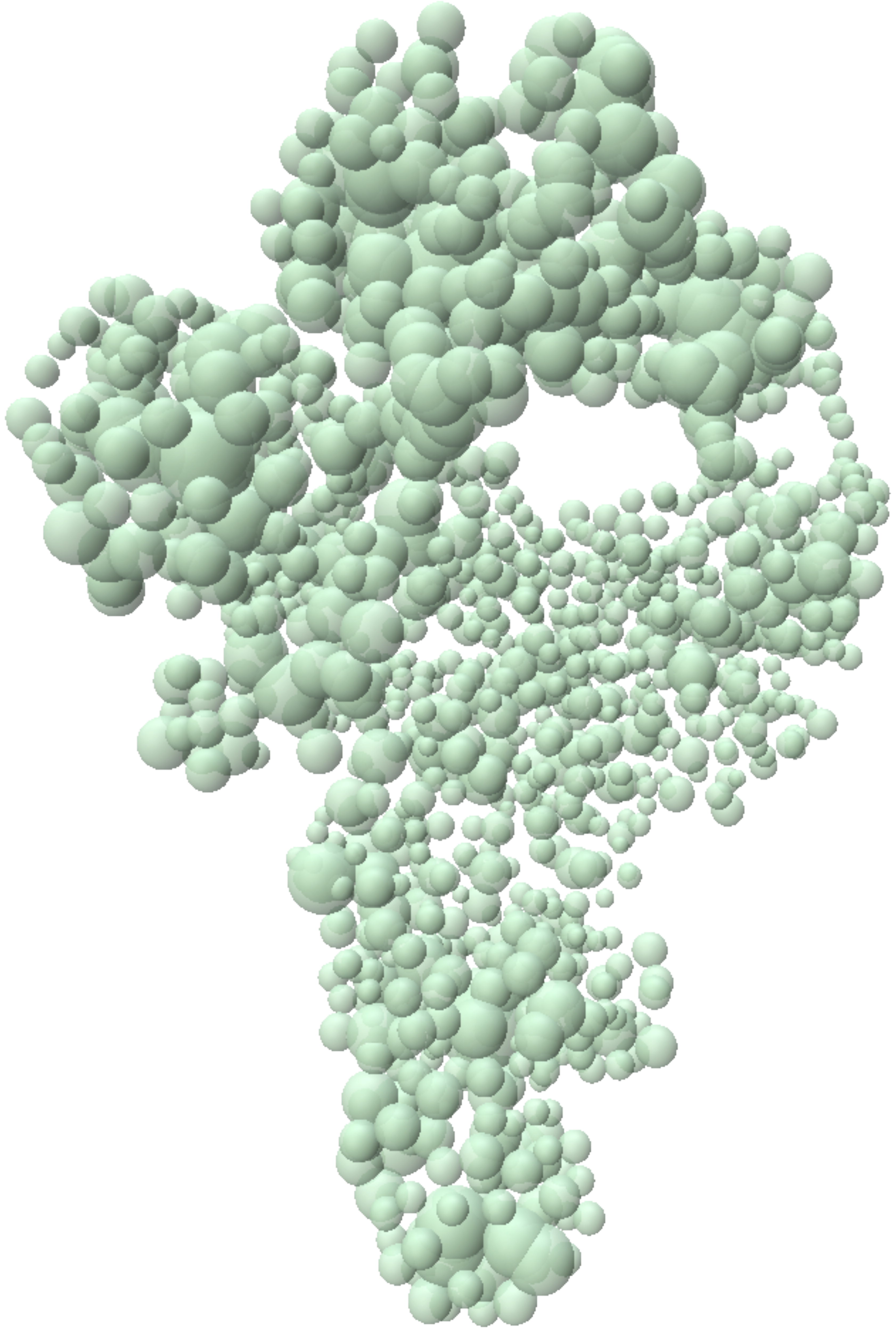
Compositional Het.



- Linear models: simple but expensive
- Neural-field: expressive but lack interpretability
- Motion-based: explicit but hard to regularize

Gaussian Mixture Model (GMM)

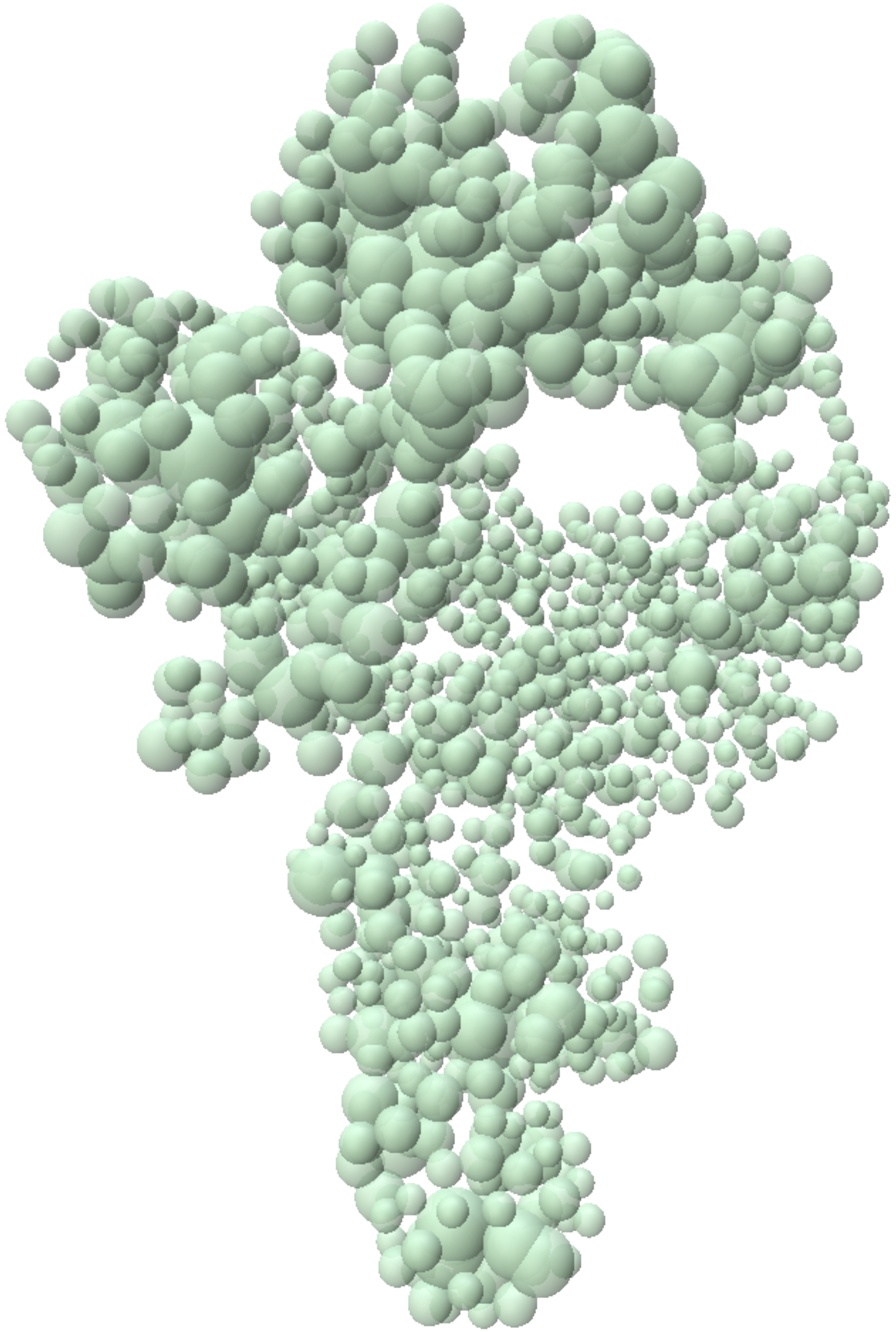
$$V(\boldsymbol{p}) = \sum_i \overset{\overset{\text{Amplitude}}{\uparrow}}{\underline{m_i}} \exp\left(-\frac{||\boldsymbol{p} - \overset{\overset{\text{Center}}{\nearrow}}{\underline{\boldsymbol{c}_i}}||_2^2}{2\underset{\underset{\text{Bandwidth}}{\searrow}}{\underline{s_i^2}}}\right)$$



Gaussian Mixture Model (GMM)

$$V(\boldsymbol{p}) = \sum_i \overset{\text{Amplitude}}{\overset{\uparrow}{\underline{m_i}}} \exp\left(-\frac{||\boldsymbol{p}-\overset{\text{Center}}{\overset{\nearrow}{\underline{\boldsymbol{c}_i}}}||_2^2}{2\underset{\underset{\text{Bandwidth}}{\searrow}}{\underline{s_i^2}}}\right)$$

👍 Compact representation

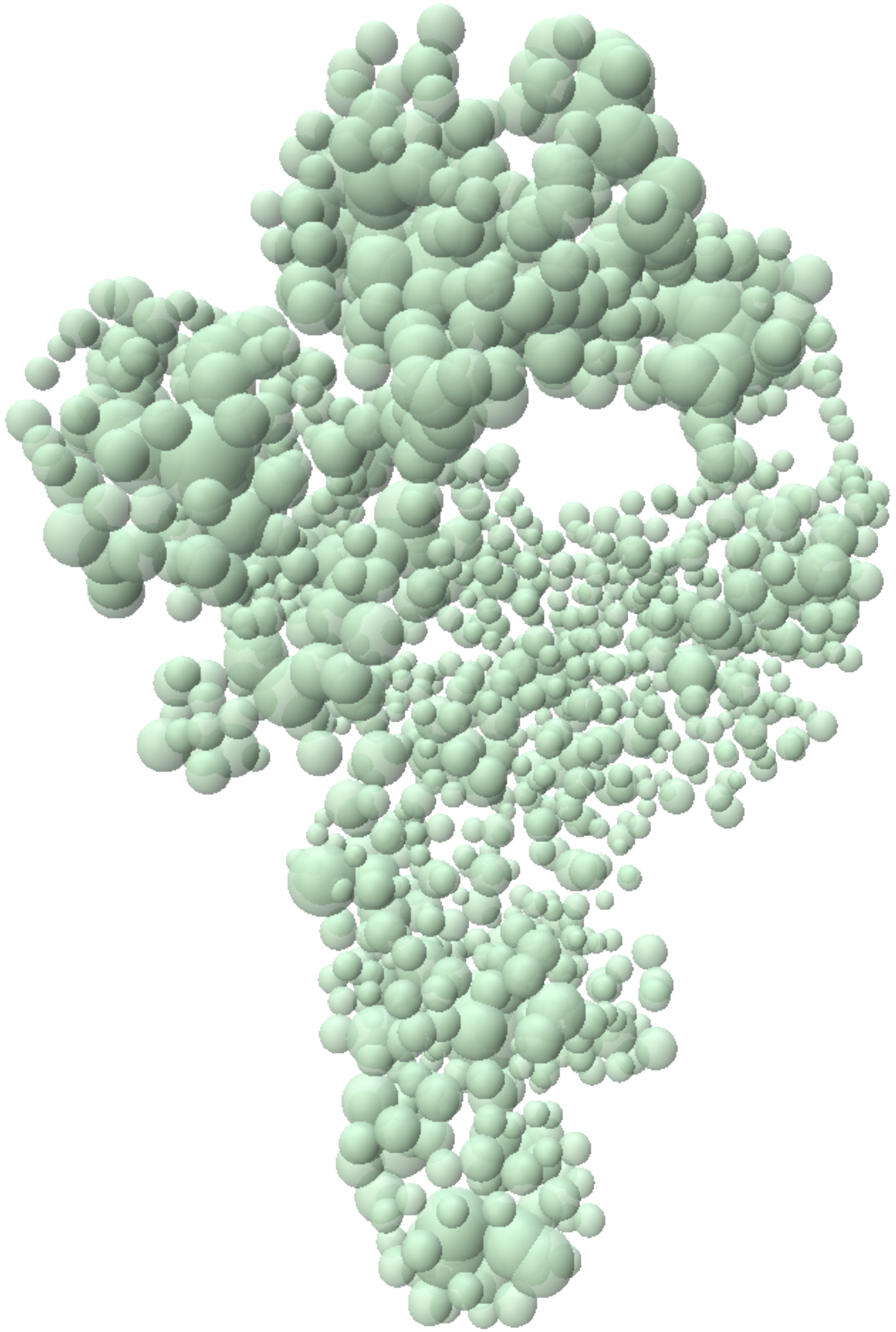


Gaussian Mixture Model (GMM)

$$V(\boldsymbol{p}) = \sum_i \overset{\text{Amplitude}}{\underline{m_i}} \exp\left(-\frac{||\boldsymbol{p} - \overset{\text{Center}}{\underline{\boldsymbol{c}_i}}||_2^2}{2\underset{\text{Bandwidth}}{\underline{s_i^2}}}\right)$$

👍 Compact representation

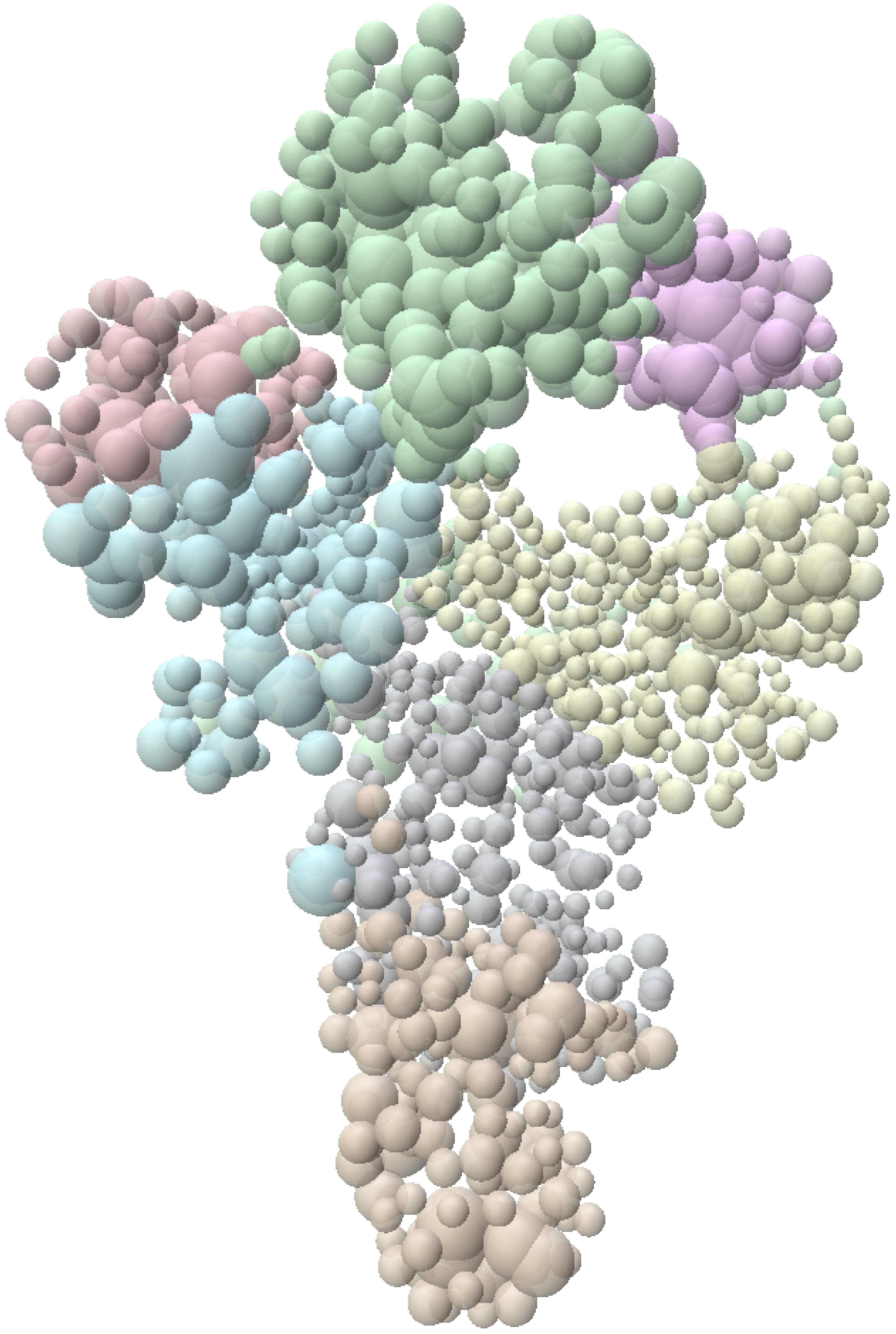
👍 Natural for heterogeneity



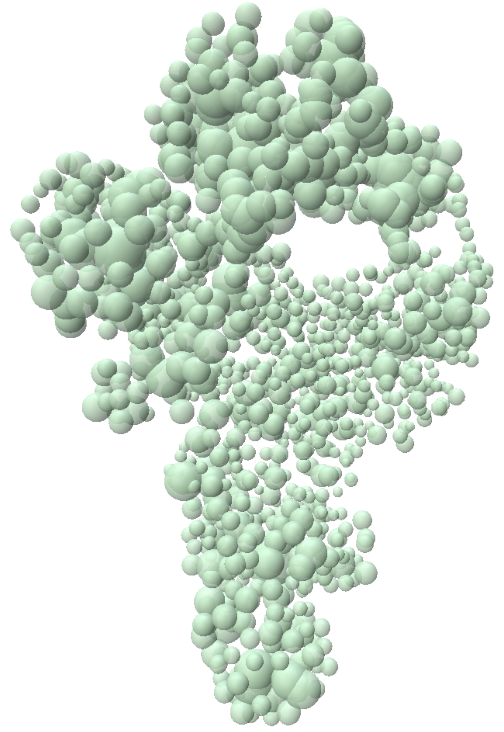
Gaussian Mixture Model (GMM)

$$V(\boldsymbol{p}) = \sum_i \overset{\text{Amplitude}}{\underline{m_i}} \exp\left(-\frac{||\boldsymbol{p} - \overset{\text{Center}}{\underline{\boldsymbol{c}_i}}||_2^2}{2\underset{\text{Bandwidth}}{\underline{s_i^2}}}\right)$$

- 👍 Compact representation
- 👍 Natural for heterogeneity
- NEW 👍 Part-based decomposition

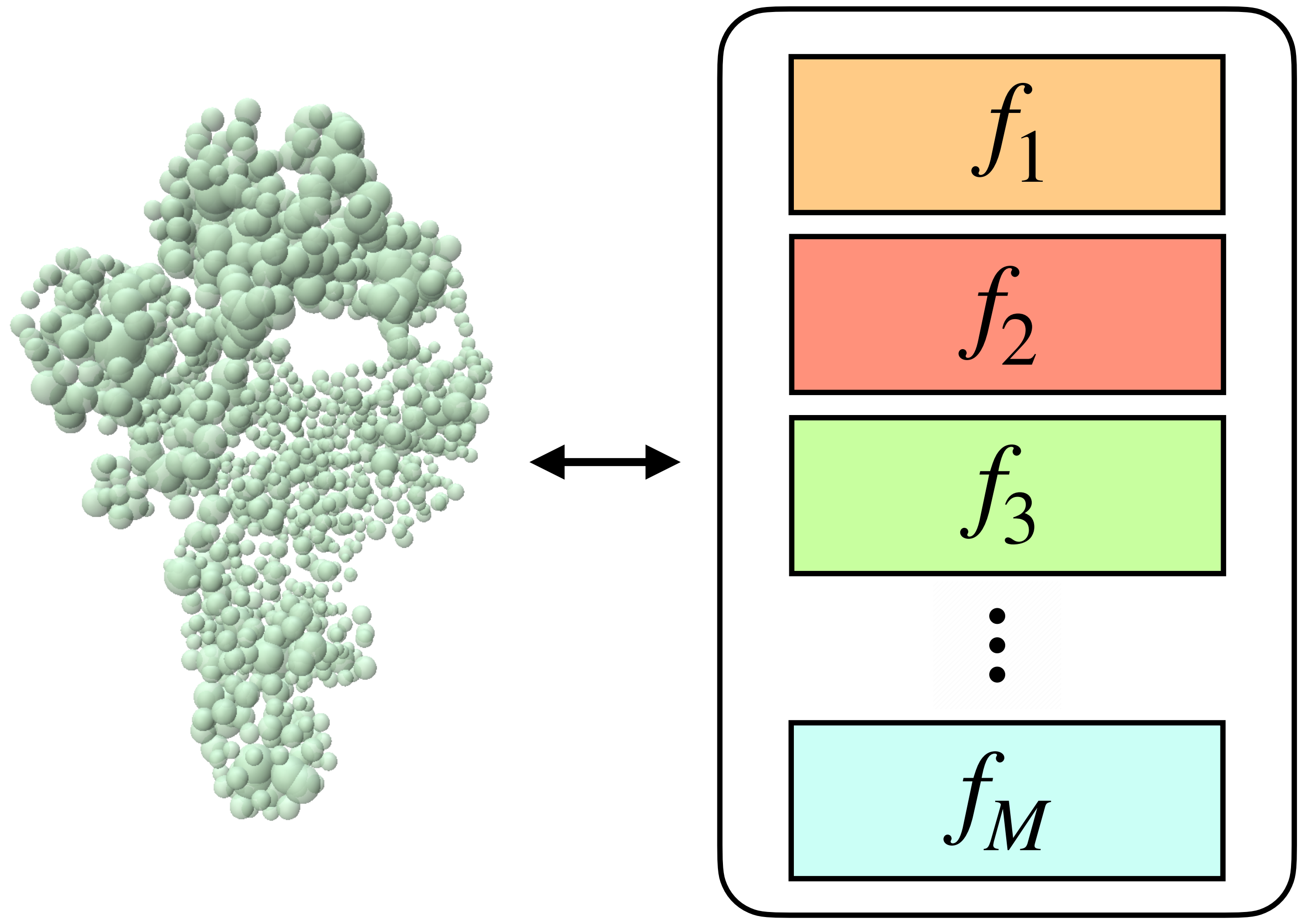


Part-Discovery

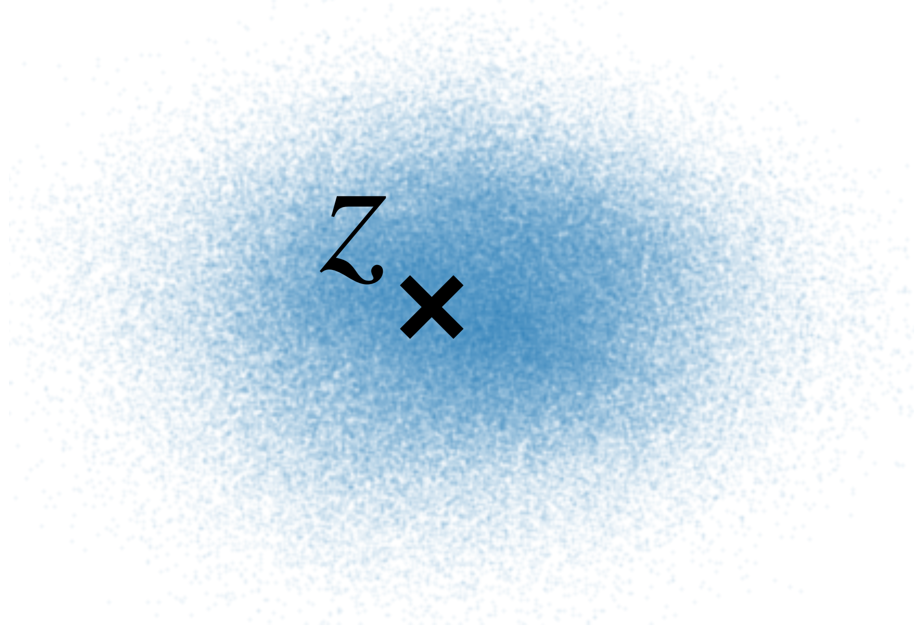


Part-Discovery

(Learnable) Gaussian Features

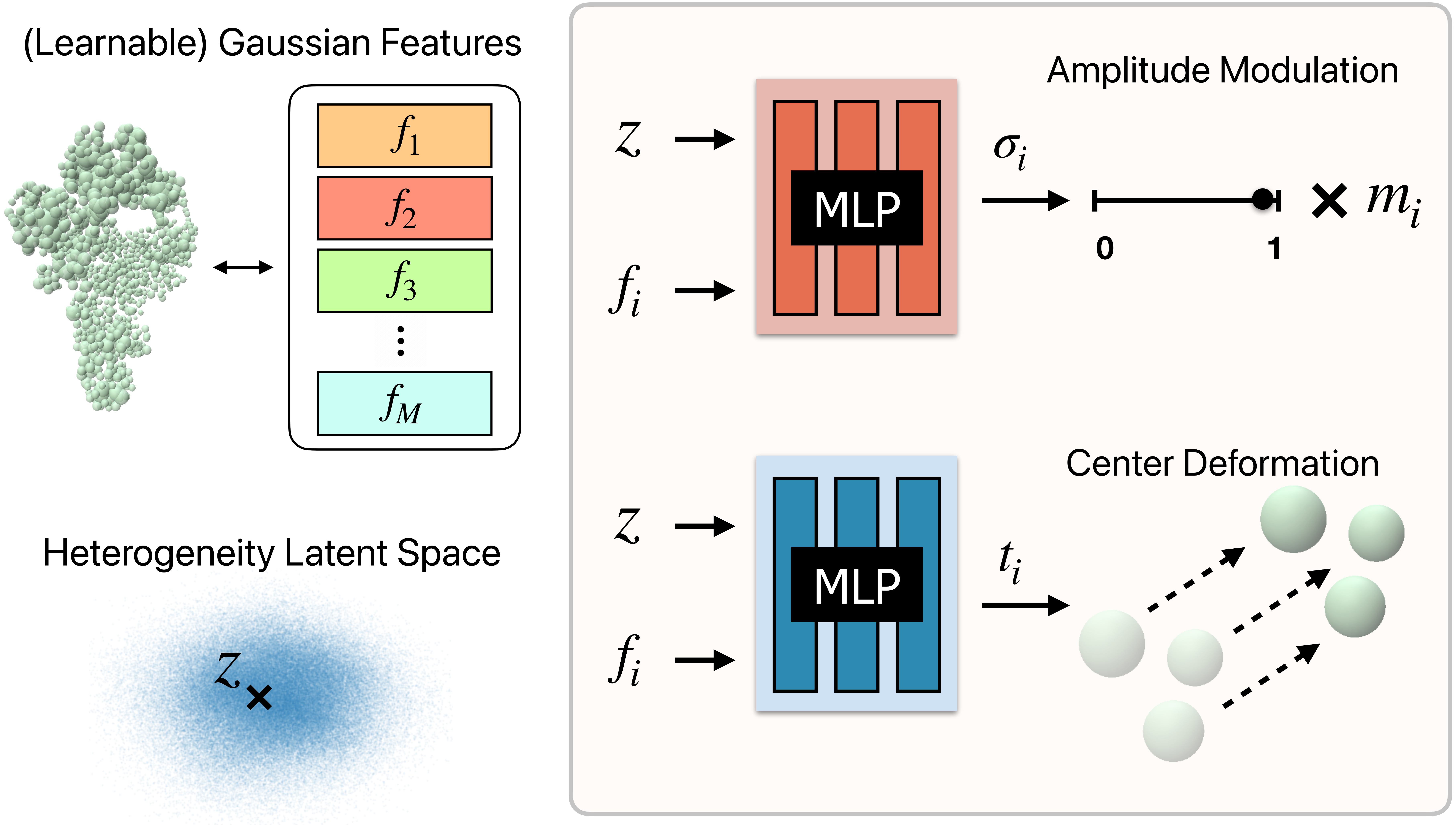


Heterogeneity Latent Space



Part-Discovery

Neural



(Learnable) Gaussian Features

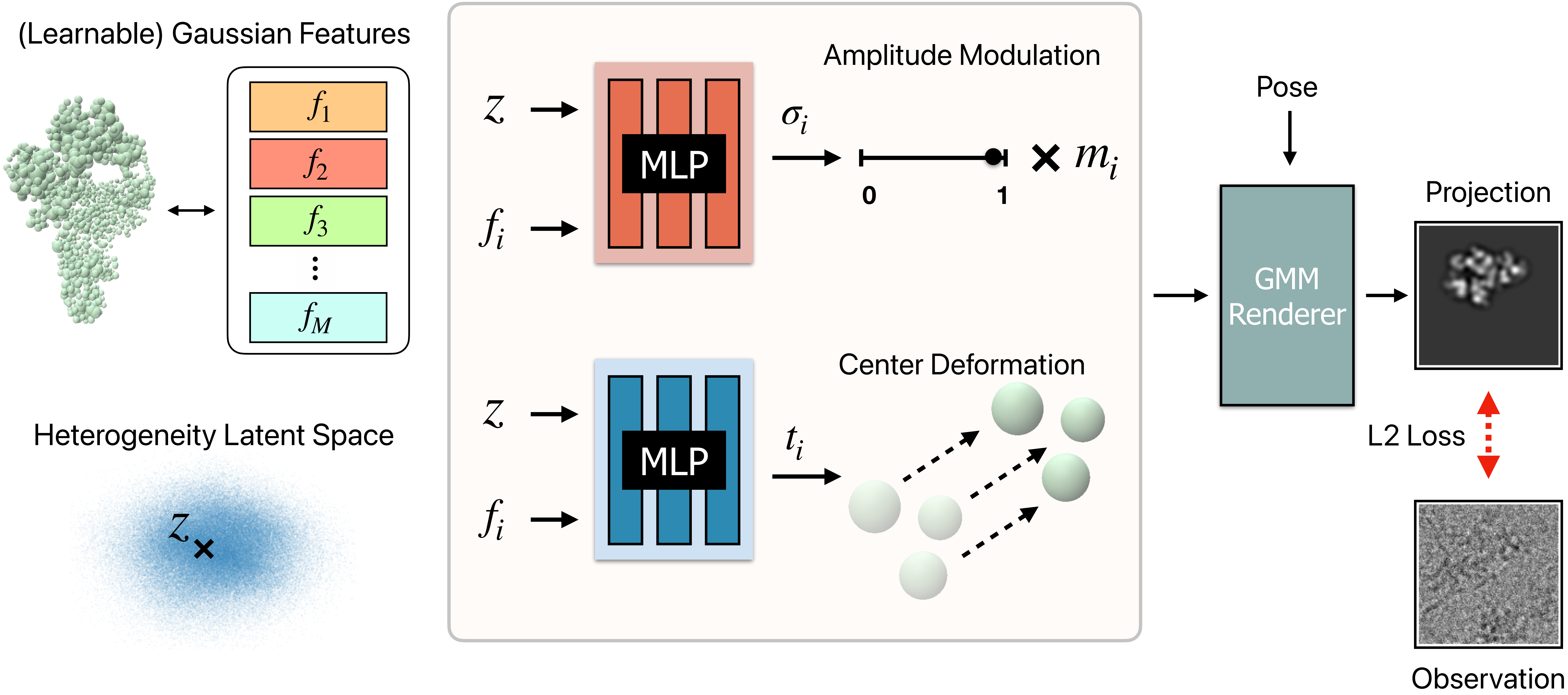
Heterogeneity Latent Space

Amplitude Modulation

Center Deformation

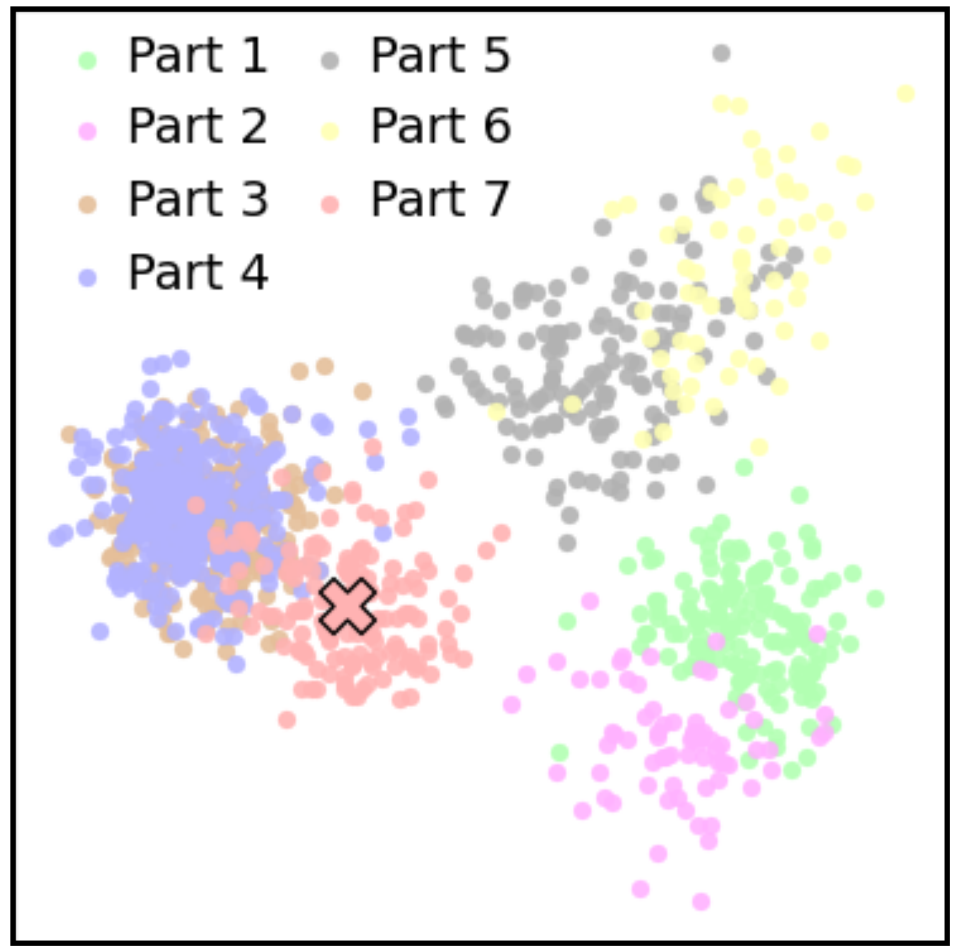
Part-Discovery

Neural



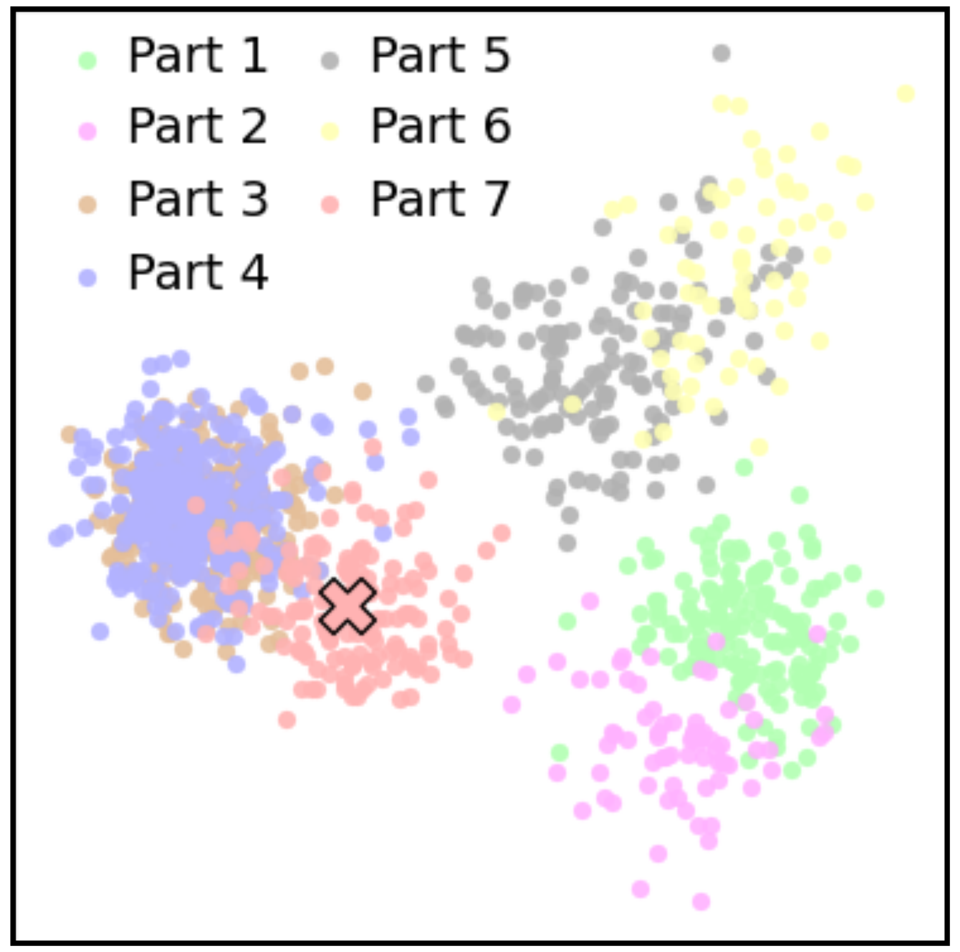
Hierarchical Part-Aware GMM

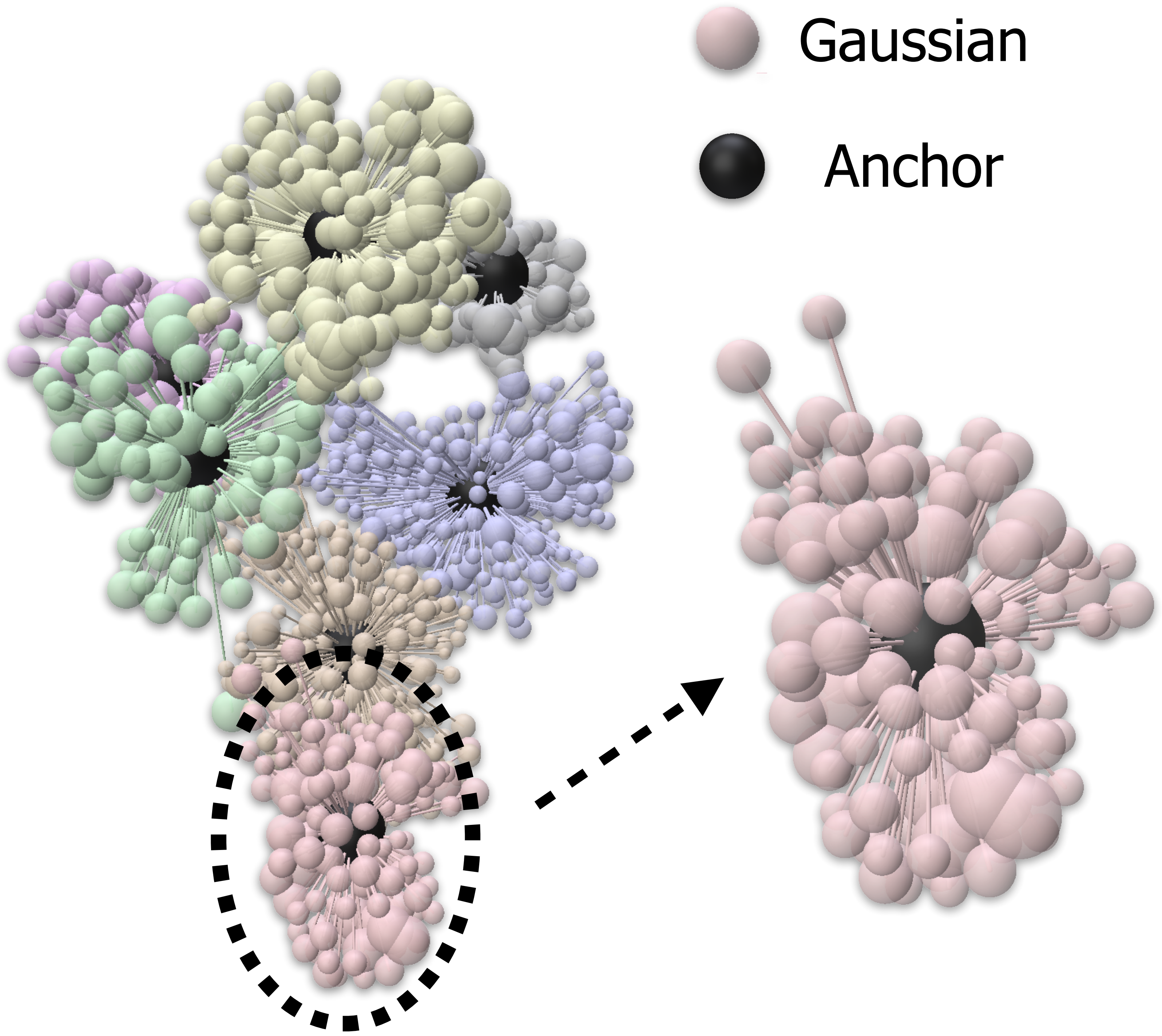
Clustering by Kmeans++



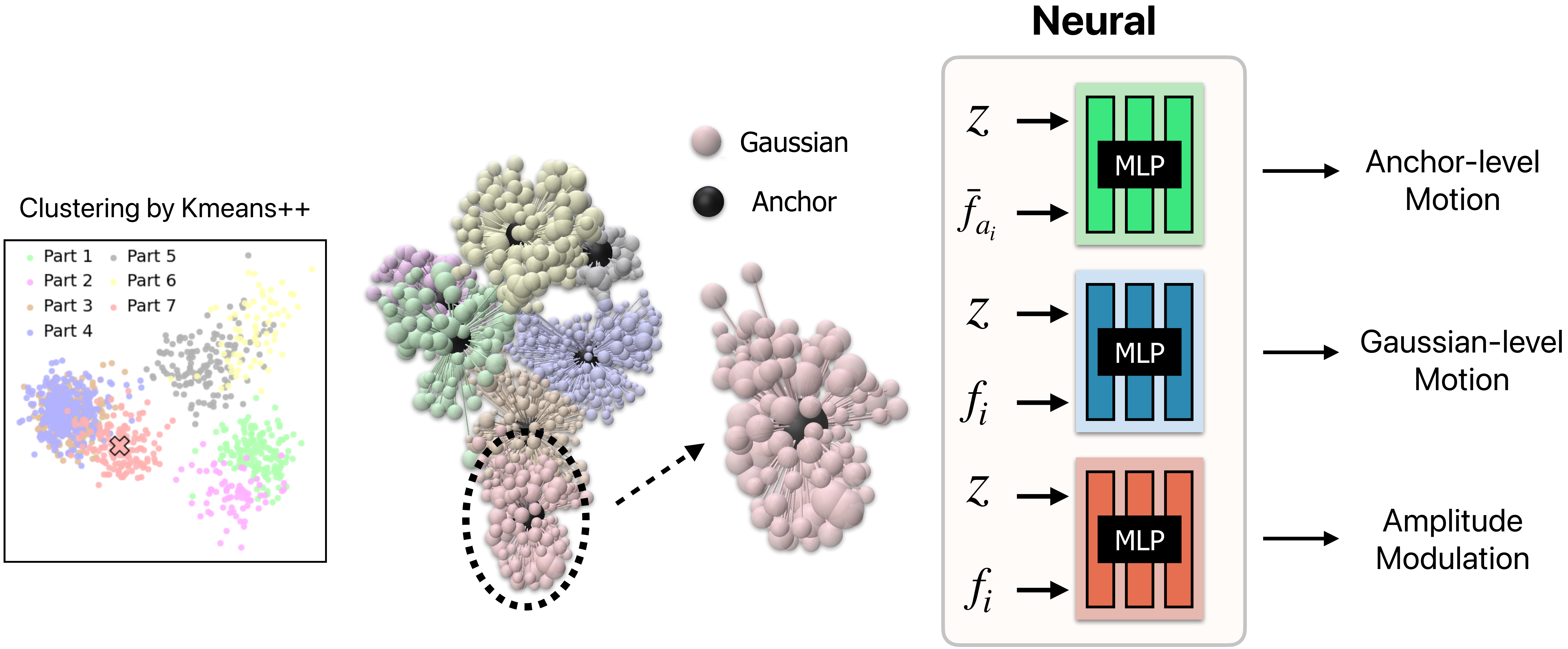
Hierarchical Part-Aware GMM

Clustering by Kmeans++





Hierarchical Part-Aware GMM



Clustering by Kmeans++

Gaussian

Anchor

Neural

z → MLP

$\bar{f}_{a_i}$ → MLP

→ Anchor-level Motion

z → MLP

f_i → MLP

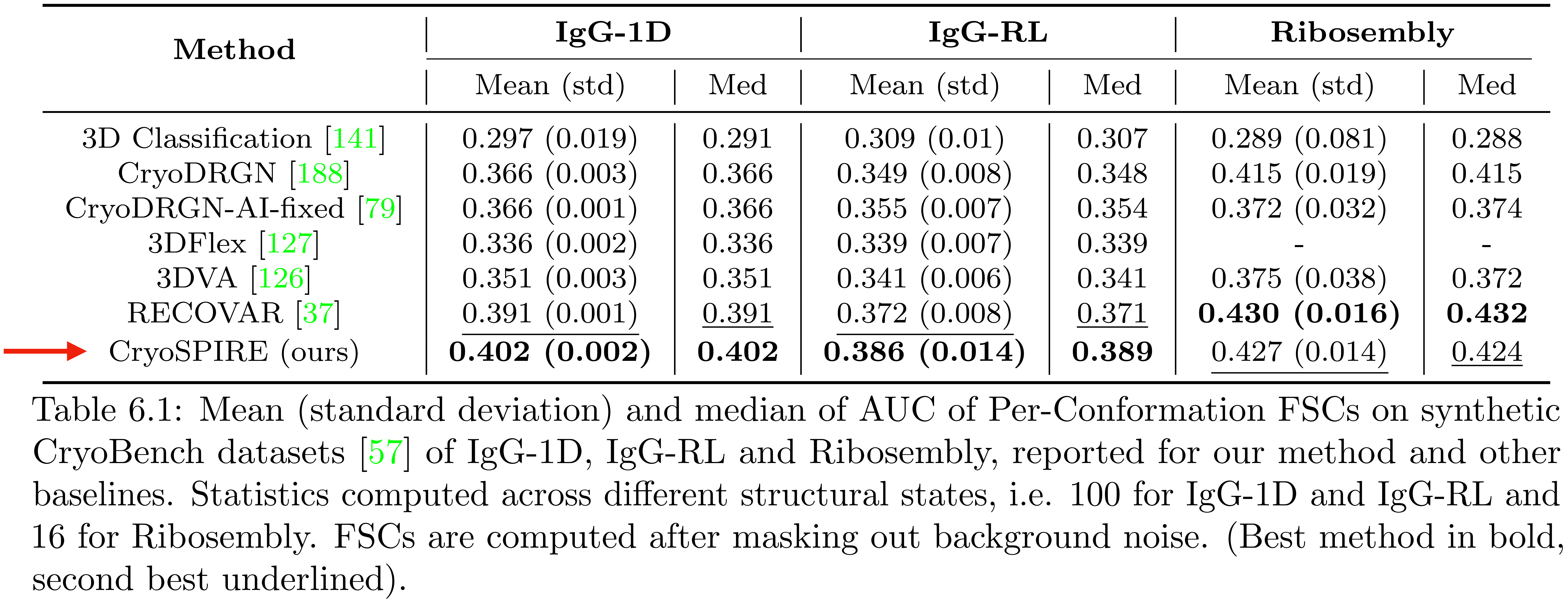
→ Gaussian-level Motion

z → MLP

f_i → MLP

→ Amplitude Modulation

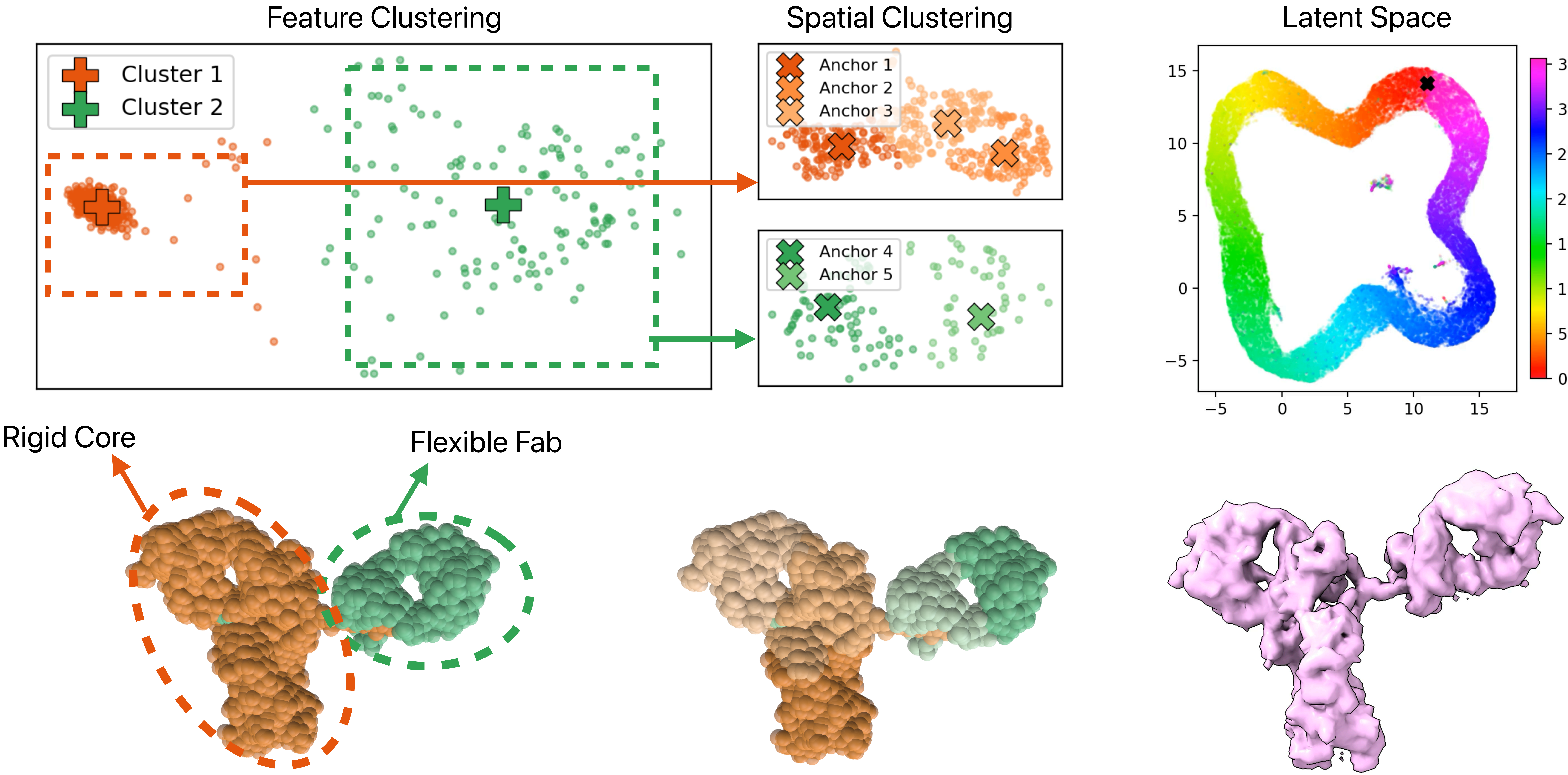
CryoBench Leaderboard



Method	IgG-1D		IgG-RL		Ribosembly	
	Mean (std)	Med	Mean (std)	Med	Mean (std)	Med
3D Classification [141]	0.297 (0.019)	0.291	0.309 (0.01)	0.307	0.289 (0.081)	0.288
CryoDRGN [188]	0.366 (0.003)	0.366	0.349 (0.008)	0.348	0.415 (0.019)	0.415
CryoDRGN-AI-fixed [79]	0.366 (0.001)	0.366	0.355 (0.007)	0.354	0.372 (0.032)	0.374
3DFlex [127]	0.336 (0.002)	0.336	0.339 (0.007)	0.339	-	-
3DVA [126]	0.351 (0.003)	0.351	0.341 (0.006)	0.341	0.375 (0.038)	0.372
RECOVAR [37]	0.391 (0.001)	0.391	0.372 (0.008)	0.371	**0.430 (0.016)**	**0.432**
CryoSPIRE (ours)	**0.402 (0.002)**	**0.402**	**0.386 (0.014)**	**0.389**	0.427 (0.014)	0.424

Table 6.1: Mean (standard deviation) and median of AUC of Per-Conformation FSCs on synthetic CryoBench datasets [57] of IgG-1D, IgG-RL and Ribosembly, reported for our method and other baselines. Statistics computed across different structural states, i.e. 100 for IgG-1D and IgG-RL and 16 for Ribosembly. FSCs are computed after masking out background noise. (Best method in bold, second best underlined).

IgG-1D (Synthetic)

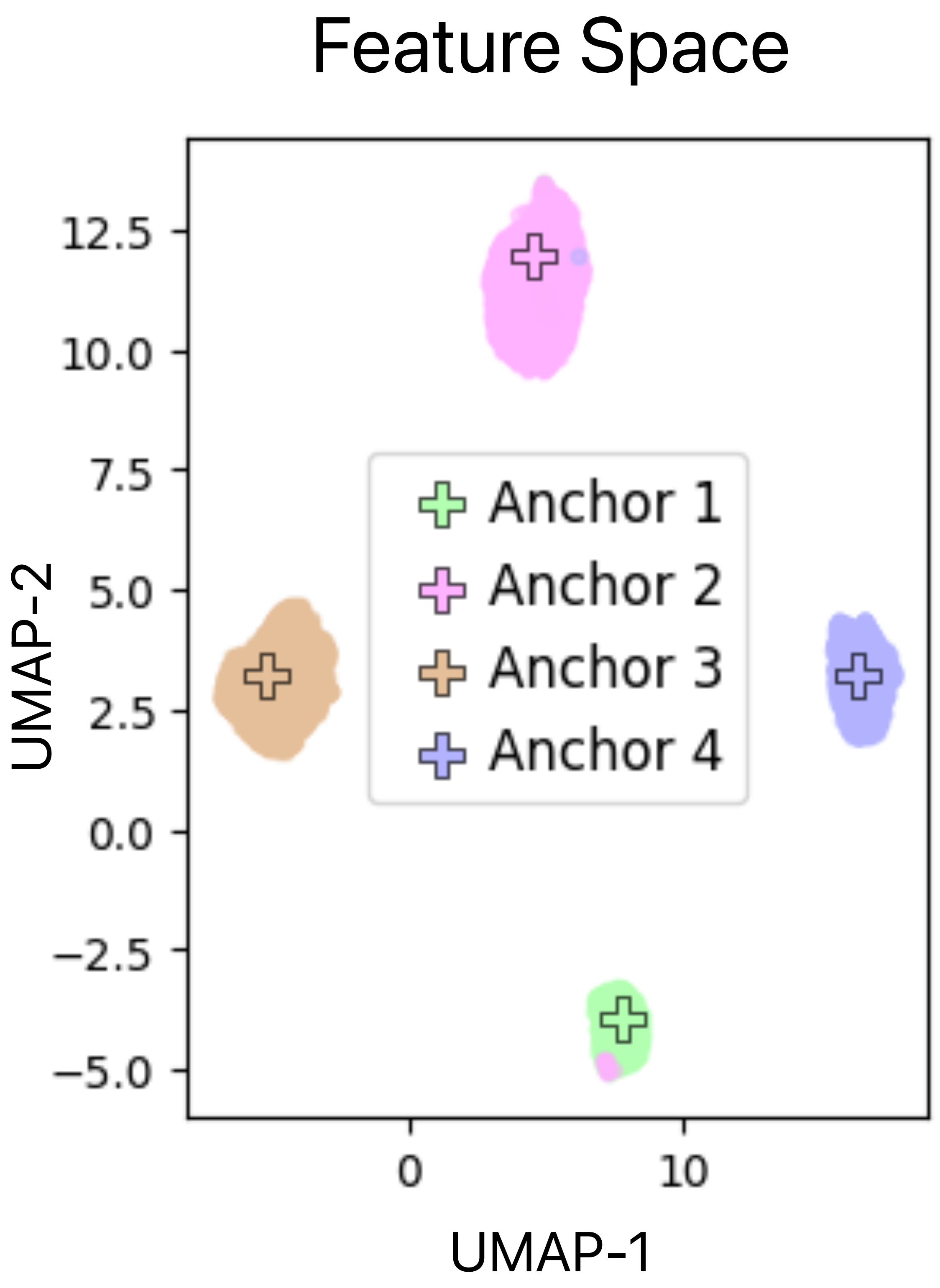


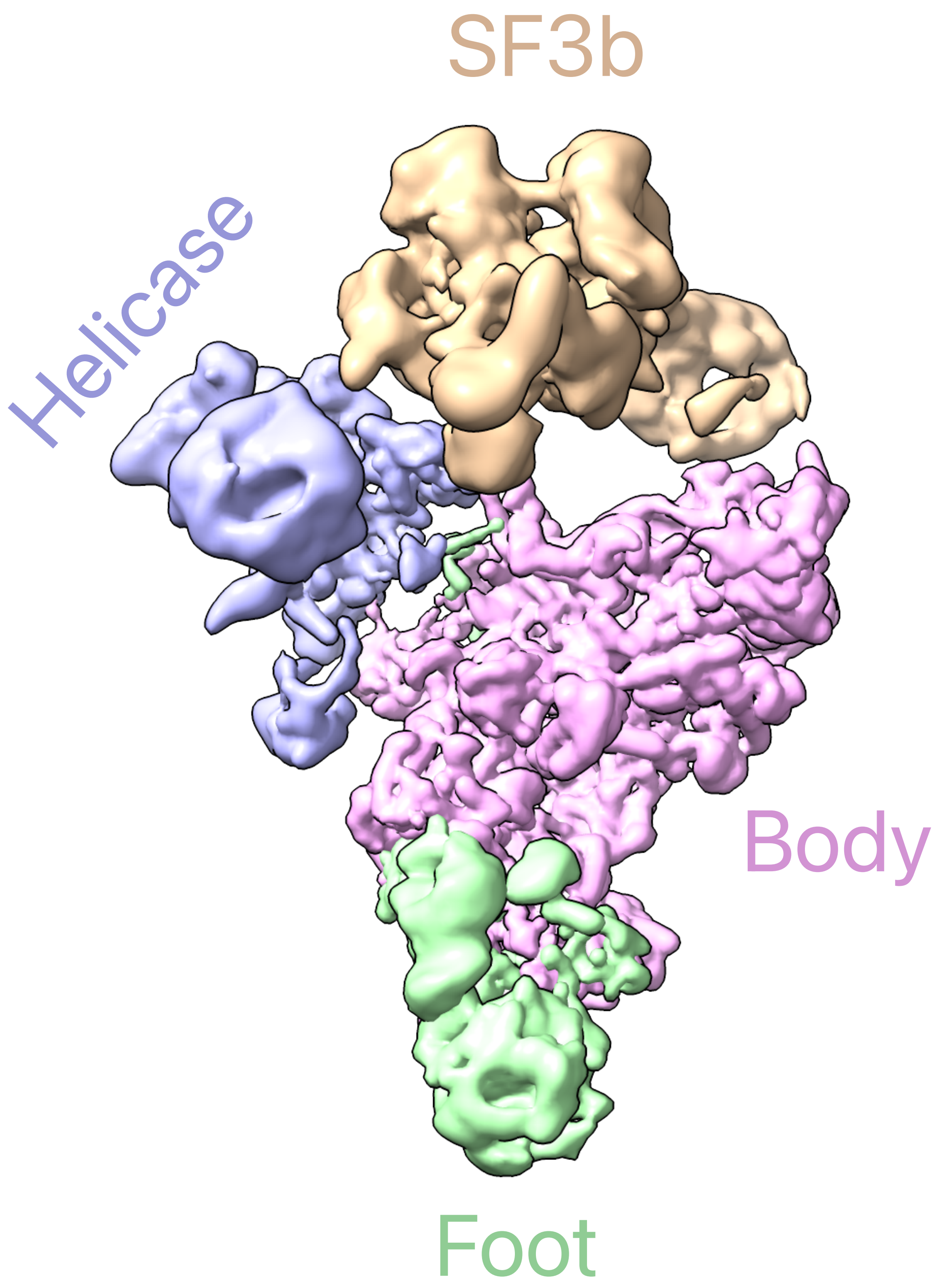
Feature Clustering

Spatial Clustering

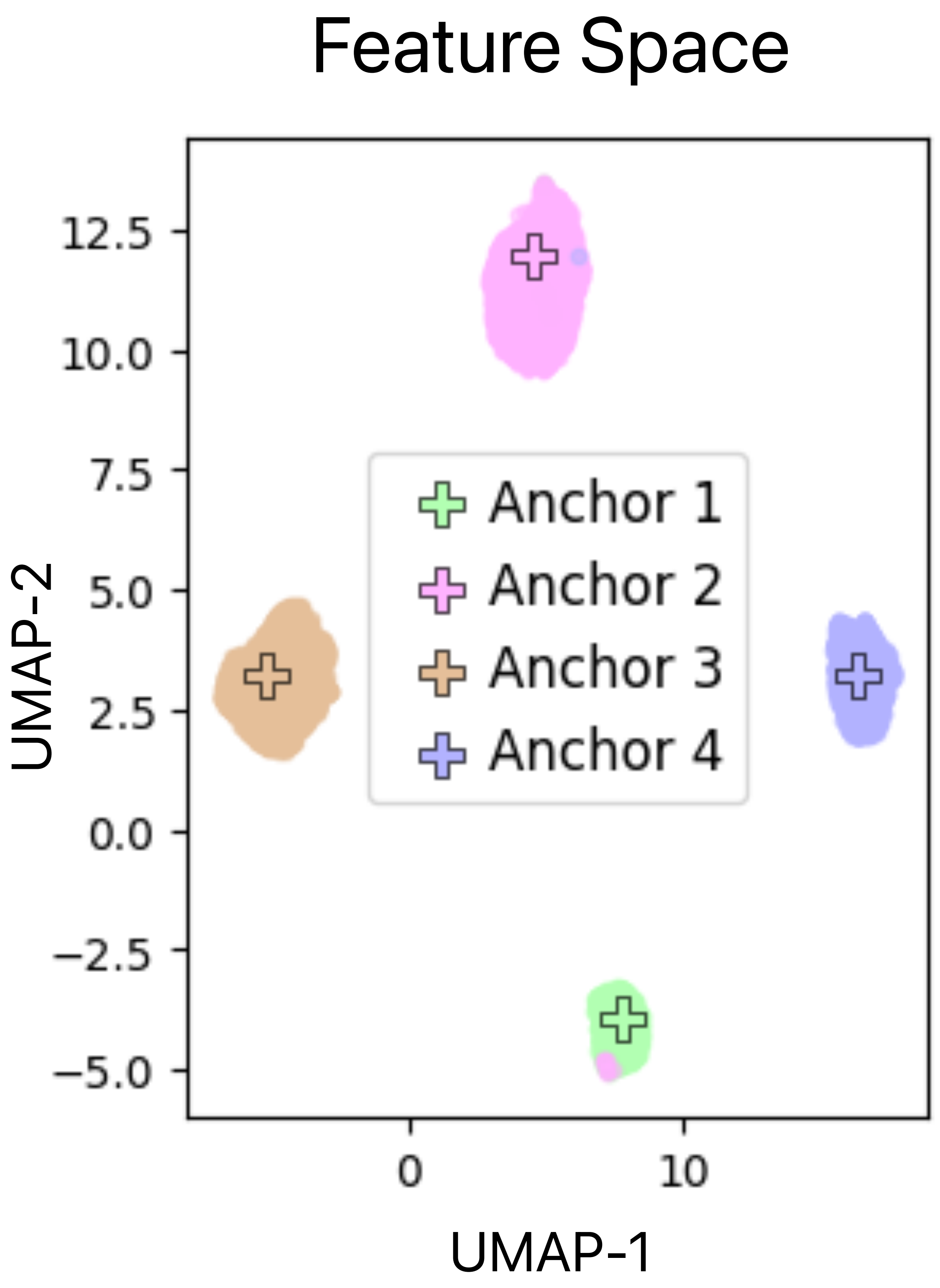
Latent Space

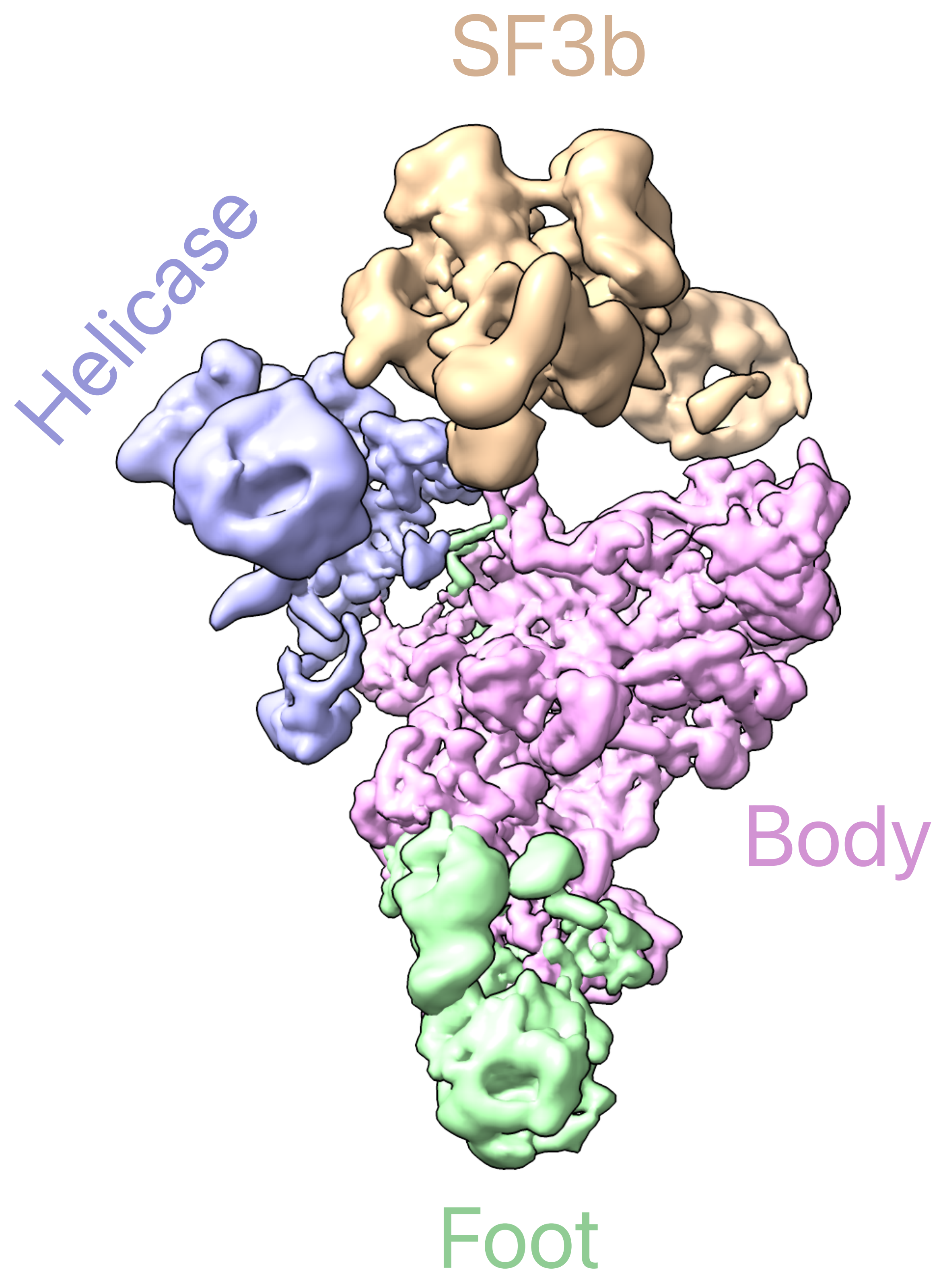
Pre-catalytic Spliceosome (Real)

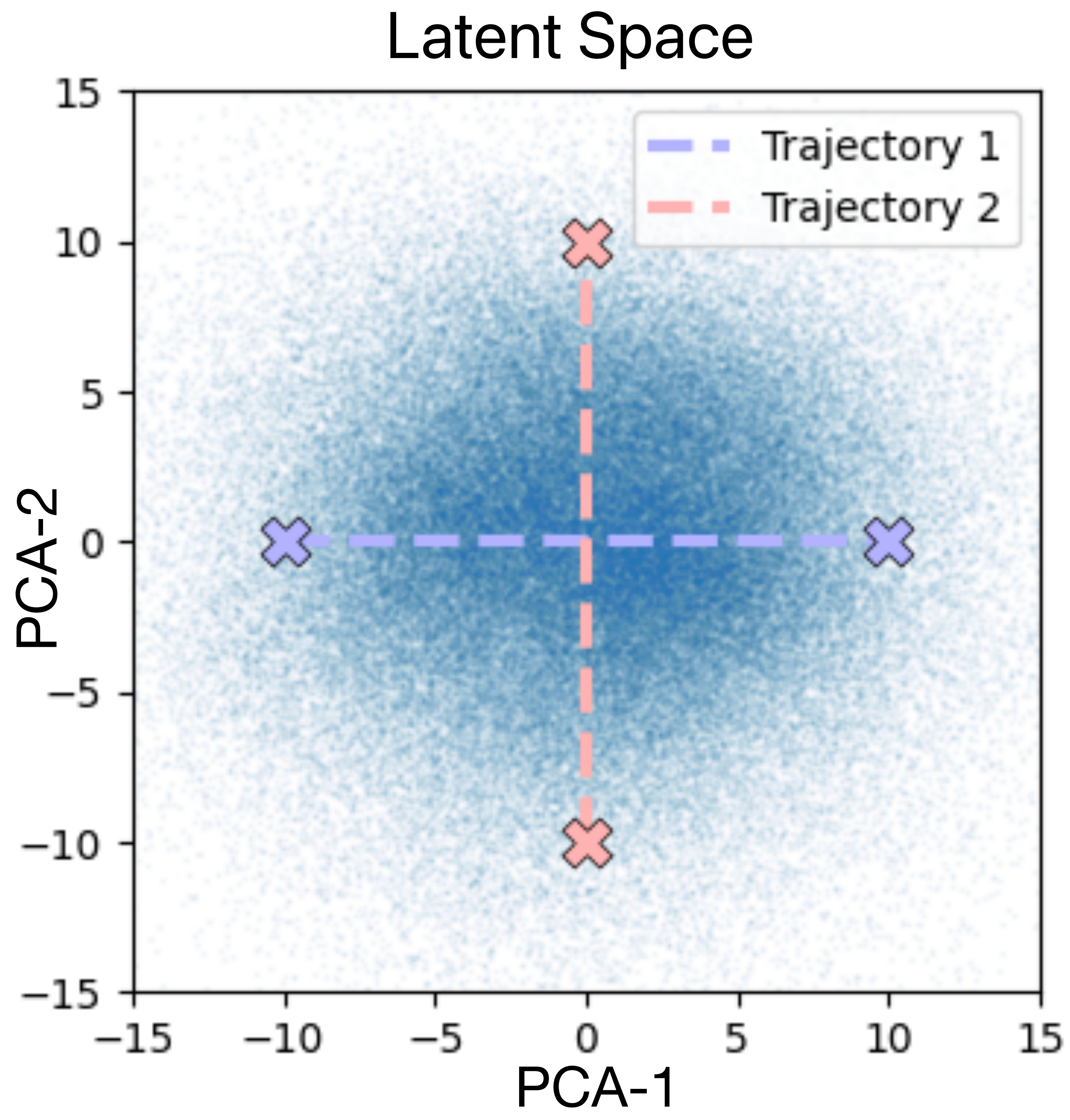


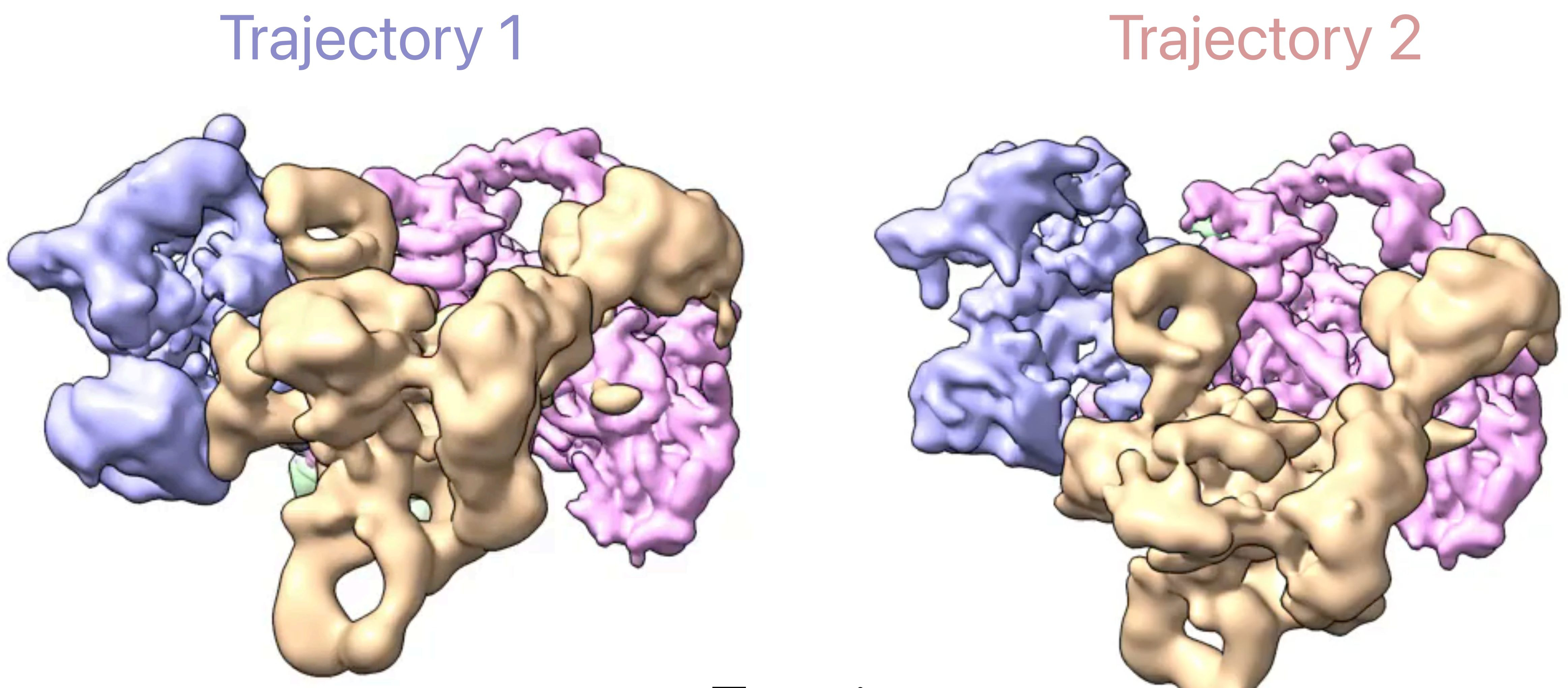


Pre-catalytic Spliceosome (Real)









Top views

Conclusion

- 👍 A novel hierarchical **part-aware GMM** for heterogeneous cryo-EM reconstruction
- 👍 Capturing both **conformational** and **compositional** variability.
- 👍 Establishing a new **state-of-the-art** on CryoBench synthetic datasets
- 👍 Obtaining **biologically meaningful parts** on experimental datasets

Project Webpage