
Reconstructing Heterogeneous Biomolecules via Hierarchical Gaussian Mixtures and Part Discovery

Shayan Shekarforoush^{1,2} David B Lindell^{1,2} Marcus A Brubaker^{1,2,3} David J Fleet^{1,2}

¹University of Toronto ²Vector Institute ³York University
{shayan,lindell,fleet}@cs.toronto.edu mab@eecs.yorku.ca

Abstract

Cryo-EM is a transformational paradigm in molecular biology where computational methods are used to infer 3D molecular structure at atomic resolution from extremely noisy 2D electron microscope images. At the forefront of research is how to model the structure when the imaged particles exhibit non-rigid conformational flexibility and compositional variation where parts are sometimes missing. We introduce a novel 3D reconstruction framework with a hierarchical Gaussian mixture model, inspired in part by Gaussian Splatting for 4D scene reconstruction. In particular, the structure of the model is grounded in an initial process that infers a part-based segmentation of the particle, providing essential inductive bias in order to handle both conformational and compositional variability. The framework, called CryoSPIRE, is shown to reveal biologically meaningful structures on complex experimental datasets, and establishes a new state-of-the-art on CryoBench, a benchmark for cryo-EM heterogeneity methods. [Project Webpage](#).

1 Introduction

Single-particle cryo-electron microscopy (cryo-EM) is a computationally driven experimental paradigm that is transforming molecular biology by enabling 3D structure determination of biomolecules, such as proteins and viruses, at near-atomic resolutions [3, 18, 38]. The core computational task is estimating a 3D structure from 2D images with unknown orientation and position, under extremely low signal-to-noise conditions. Essential to their biological function, biomolecules exhibit varying degrees of *conformational flexibility*, where structures deform non-rigidly, and *compositional variation*, where parts of a structure may be present in some images and absent in others (see Fig. 1). Accordingly, a major challenge in cryo-EM is the estimation of 3D structures from such heterogeneous data and, to that end, how to infer meaningful representations of structures such as parts that capture their heterogeneity. The crux of this challenge is how to effectively represent and regularize this variability without overfitting to the noise in cryo-EM images. Existing methods, while encouraging, are generally limited in either expressiveness, interpretability, or efficiency.

Here, we propose CryoSPIRE, a new method for heterogeneous reconstruction. We leverage a part-based Gaussian mixture model (GMM) of 3D density that enables CryoSPIRE to represent both conformational and compositional heterogeneity, unlike some existing deformation-based methods [13, 33]. Further, it provides a naturally interpretable and physically plausible, part-based structure in contrast to existing latent variable methods based on linear density subspaces [10, 32] or neural field models [19, 20, 47]. A key challenge with part-based GMMs concerns initialization and the discovery of parts. We propose a novel method for part discovery which estimates a coarse-grained GMM with per-Gaussian learnable features (c.f., [2]) and an MLP which defines Gaussian locations and amplitudes. We show that these learned features naturally encode characteristics of structural heterogeneity, which we leverage to infer a part-based segmentation of the structure. Inspired in part by Scaffold-GS [21], we define CryoSPIRE (Scaffold Part-Aware Mixture of Gaussians), a hierarchical model which estimates a Gaussian mixture wherein the composition of components

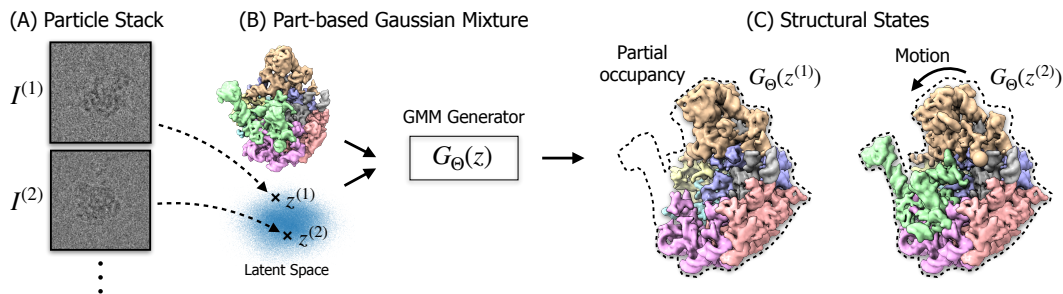


Figure 1: (A) Based on a stack of noisy particle images, (B) CryoSPIRE learns a part-based Gaussian mixture, with parameters Θ , and a latent space representing structural heterogeneity. Given a latent code z , a generator produces a 3D density map. (C) The model supports compositional variability (e.g., $G_{\Theta}(z^{(1)})$ with a missing part), and conformational flexibility (e.g., $G_{\Theta}(z^{(2)})$ with part deformation).

and their deformation are defined in terms of a set of anchors, corresponding to parts. The resulting model naturally allows for the arbitrary combination of parts which can both rigidly move and locally deform as a function of an input heterogeneity latent code (see Fig. 1).

To our knowledge, this is the first GMM-based model to be successfully benchmarked on CryoBench [15], a standardized benchmark for cryo-EM heterogeneity with ground-truth labels. In particular, CryoSPIRE outperforms widely used and state-of-the-art methods [10, 19, 32, 33, 47], sometimes by a wide margin. Through ablations, we also validate key design choices, demonstrating the benefits of Gaussian features over positional encoding as in DynaMight [40], and highlighting the benefits of hierarchical motion modeling. Finally, on experimental data, CryoSPIRE automatically discovers representations of 3D density maps that correspond to biologically meaningful parts.

To summarize our contributions: we propose a new method enabling part-discovery on 3D biomolecular structures based on a coarse-grained GMM. This part-based structure is used to initialize a novel, hierarchical GMM-based model for heterogeneous reconstruction with compositional and conformational variability. The resulting framework, CryoSPIRE, establishes a new state-of-the-art on quantitative benchmarks and qualitative experimental datasets.

2 Background and Related Work

Latent Variable Models. Heterogeneous cryo-EM reconstruction methods typically introduce latent variables to represent structural variability of the 3D density map. 3DVA [32] and RECOVER [10] learn a linear subspace to represent variation in 3D density maps, with clever numerical and regularization techniques to optimize high-dimensional basis maps at high spatial resolutions. Nevertheless, to model large-scale continuous motion with a high dimensional subspace, memory requirements are prohibitive. Much current work has shifted to nonlinear latent models and deep learning [14, 19, 47], with Cryo-DRGN [47] and DRGN-AI [19] using auto-encoders to obtain latent codes and conditional coordinate networks [24] to generate density maps. Such latent-variable models are hard to interpret, however, as conformational and compositional heterogeneity are not decoupled, and they provide no explicit model of motion between conformational states. By contrast, the latents in 3DFlex [33] encode flow fields that model the conformational deformation of a canonical structure. While resolving detailed motion and improving the quality of density maps, 3DFlex cannot handle compositional heterogeneity, and it is highly sensitive to regularization, often requiring substantial trial and error.

GMM-Based Methods. Gaussian mixtures have been used to model 3D density [4, 5, 6, 40]; they provide a sparse, compact representation in which conformation and compositional variability are modeled in terms of positions and amplitudes of Gaussian components. With Gaussian components viewed as atomic primitives, such models also facilitate physics-based priors [6, 40] and subsequent molecular model fitting. Nevertheless, existing GMM-based methods fall short in various ways. E2GMM [4] and related methods [5, 6] generate GMM parameters with a single network, which scales poorly to large numbers of Gaussians. Further, their multi-scale smoothness priors [6] are based on an arbitrary hierarchy which fails to capture part-based structures, thus resorting to manual part masks to resolve and estimate local motions. DynaMight [40] is similar to CryoSPIRE in defining

an explicit motion model, but it is unable to handle compositional variations, and, as we show, its positional encodings are inferior to our learnable features.

Gaussian Splatting. Beyond cryo-EM, the effectiveness of GMMs has been demonstrated in 3D Gaussian Splatting [16, 48], a technique which provides a fast approximation to the volume rendering integral [8, 23], enabling efficient high-fidelity reconstruction of 3D scenes from multi-view images [11, 17, 22, 44, 45, 46]. 3D Gaussian Splatting represents scene appearance and structure using thousands to millions of Gaussian components, each associated with parameters that control opacity and view-dependent color. CryoSPIRE is in part inspired by Gaussian Splatting [2, 21], but tailored to cryo-EM, with a different image formation model, images with signal-to-noise ratios less than 5%, and a novel method for part discovery.

GMM Image Formation. Following [4, 5, 6, 40], we parameterize the terms of a Gaussian mixture with center $\mathbf{c} \in \mathbb{R}^3$, isotropic scale $s \in \mathbb{R}$, and an amplitude $m \in \mathbb{R}$:

$$f(\mathbf{p}) = \sum_i m_i \exp\left(-\frac{\|\mathbf{p} - \mathbf{c}_i\|_2^2}{2s_i^2}\right), \quad (1)$$

for location $\mathbf{p} \in \mathbb{R}^3$. We transform the GMM into the observation space for the n -th particle image, with a rotation $\mathbf{R}^{(n)} \in SO(3)$ and translation $\mathbf{t}^{(n)} \in \mathbb{R}^3$, followed by an integral projection along the z -axis of the microscope, to obtain a noise-free 2D image, $\tilde{I}(\tilde{\mathbf{p}})$, [4]:

$$\tilde{I}^{(n)}(\tilde{\mathbf{p}}) = \sum_i \sqrt{2\pi}s_i m_i \exp\left(-\frac{\|\tilde{\mathbf{p}} - [\mathbf{R}^{(n)} \mathbf{c}_i + \mathbf{t}^{(n)}]_{xy}\|_2^2}{2s_i^2}\right), \quad (2)$$

where $\tilde{\mathbf{p}} \in \mathbb{R}^2$ and $[\cdot]_{xy}$ is an operator to discard z coordinate of the input position. Cryo-EM images are then convolved with microscope point spread function and corrupted by additive mean-zero Gaussian noise, $\hat{I}^{(n)} = g^{(n)} \star \tilde{I}^{(n)} + \epsilon^{(n)}$. Like other cryo-EM models, the parameters are typically optimized by minimizing a squared L2 reconstruction loss between model predictions and observed images. See the supplement for more details on image formation and the image likelihood.

3 CryoSPIRE

Heterogeneous cryo-EM involves non-rigid 3D reconstruction from noisy 2D images. For such an inverse problem, regularization and inductive bias are key. Local smoothness is a natural choice for regularization, however, smoothness alone is not sufficient as nearby regions can deform in somewhat independent ways [33]. Further, the presence or absence of biomolecule parts is not dictated by spatial proximity alone. Macromolecular complexes, like many objects, naturally possess a part-based structure that connects to their compositional and conformational variations. But a coherent 3D part-decomposition is unavailable *a priori*, and estimating parts from noisy 2D observations is inherently challenging. As a consequence, prior work resort to manually designed masks or meshes [25, 33].

Here, we propose a novel two-stage GMM-based framework. Given particle images with corresponding poses $\{(I^{(n)}, \mathbf{R}^{(n)}, \mathbf{t}^{(n)})\}_{n=1}^N$, and a crude initial 3D structure, we first optimize a coarse-grained GMM in which each Gaussian component is augmented with a learnable feature vector (c.f., [2]). We observe that the learned features encode meaningful information about structural regularities. In particular, Gaussian components that coherently deform or consistently appear or disappear receive similar features, facilitating the inference of a part-based segmentation of the particle. Second, based on the identified parts and inspired by Scaffold-GS [21], we define a part-aware Gaussian mixture model in terms of a set of anchors, one per part, each with a corresponding set of Gaussians. Optimizing this representation recovers a high-resolution representation of 3D density maps with compositional and conformational variability. In what follows, we describe the part-based hierarchical model, (Fig. 2B–D), followed by the part discovery method and initialization scheme (Fig. 2A).

3.1 Part-Aware Gaussian Mixture

We first specify the form of the part-aware latent-conditioned mixture model; Table 1 provides a summary of the notation used. The model is conditioned on a latent coordinate $\mathbf{z} \in \mathfrak{Z} \subset \mathbb{R}^D$ for each image, which specifies the state of the macromolecule. The density model itself comprises a set of anchors, each associated with a meaningful part of the macromolecule (Fig. 2B). We parameterize the

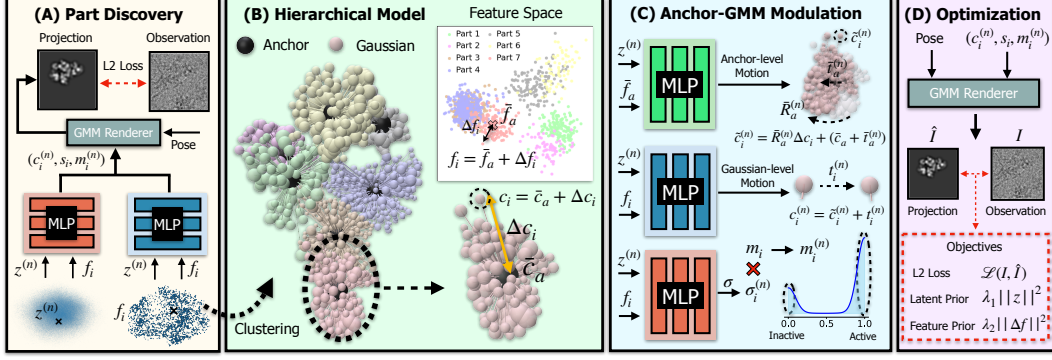


Figure 2: Overview of CryoSPIRE. **(A)** To infer parts, we optimize a coarse GMM with neural networks that generate Gaussian amplitudes and centers, conditioned on image latent codes and Gaussian features. **(B)** Clustering on learned Gaussian features provides meaningful parts. The CryoSPIRE model comprises one anchor and a set of Gaussians per part. **(C)** MLPs generate the rigid-body motion of each anchor (top), per-Gaussian displacements relative to the anchor frames (middle), and per-Gaussian activations in (0,1) to represent occupancy (bottom). **(D)** A reconstruction loss compares observed images to 2D projection of the corresponding 3D GMM. Priors encourage bounded latent code and small feature offsets.

Gaussians		Anchors		Particles	
i	Gaussian index	a	Anchor index	n	Particle index
$\mathbf{c}_i$	Gaussian center	a_i	Anchor index of i -th Gaussian	$I^{(n)}$	Observed image
$\mathbf{m}_i$	Gaussian amplitude	$\bar{\mathbf{c}}_a$	Anchor center	$\hat{I}^{(n)}$	Estimated projection
s_i	Gaussian scale	$\bar{\mathbf{f}}_a$	Anchor feature	$\mathbf{z}^{(n)}$	Particle latent code
$\Delta \mathbf{c}_i$	Gaussian center offset	$\bar{\mathbf{R}}_a$	Anchor rotation	$\mathbf{R}^{(n)}$	Particle rotation
$\Delta \mathbf{f}_i$	Gaussian feature offset	$\bar{\mathbf{t}}_a$	Anchor translation	$\mathbf{t}^{(n)}$	Particle translation
$\mathbf{t}_i^{(n)}$	Gaussian translation				

Table 1: Summary of notations used to denote variables related to Gaussians, anchors or particles.

anchors as, $\mathcal{A} = \{(\bar{\mathbf{c}}_a, \bar{\mathbf{f}}_a)\}_{a=1}^A$, where $\bar{\mathbf{c}}_a \in \mathbb{R}^3$ specifies the anchor center location in a canonical frame, and $\bar{\mathbf{f}}_a \in \mathfrak{F} \subset \mathbb{R}^E$ is an associated feature vector that encodes heterogeneity information of its corresponding part. The GMM has G Gaussian components associated with anchors (Fig. 2B, left), denoted by $\mathcal{G} = \{(\mathbf{f}_i, \mathbf{c}_i, s_i, m_i, a_i)\}_{i=1}^G$ where $\mathbf{f}_i \in \mathfrak{F}$ and $a_i \in \{1, \dots, A\}$ specifies the anchor associated with the Gaussian that is set by the part discovery method below.

We parameterize the position and feature embedding of the i -th Gaussian relative to its associated anchor a_i as

$$\mathbf{c}_i = \bar{\mathbf{c}}_{a_i} + \Delta \mathbf{c}_i, \quad \mathbf{f}_i = \bar{\mathbf{f}}_{a_i} + \Delta \mathbf{f}_i, \quad (3)$$

where $\Delta \mathbf{c}_i \in \mathbb{R}^3$ and $\Delta \mathbf{f}_i \in \mathbb{R}^E$ are learnable offsets. We initially set $\Delta \mathbf{f}_i = \mathbf{0}$ so all Gaussians are initialized with the features of their corresponding anchors.

To enable conformational variability, we parameterize deformations at two levels. First, the large-scale motion of each anchor frame is parameterized as a rigid body transformation (Fig. 2C, top). Given the latent code for n -th particle image, $\mathbf{z}^{(n)} \in \mathfrak{Z}$, and the anchor feature vector $\bar{\mathbf{f}}_{a_i}$, we compute the rotated and translated center of the i -th Gaussian, $\tilde{\mathbf{c}}_i^{(n)}$, as

$$\tilde{\mathbf{c}}_i^{(n)} = \bar{\mathbf{R}}_{a_i}^{(n)} \Delta \mathbf{c}_i + (\bar{\mathbf{c}}_{a_i} + \bar{\mathbf{t}}_{a_i}^{(n)}), \quad \text{where } \bar{\mathbf{R}}_{a_i}^{(n)}, \bar{\mathbf{t}}_{a_i}^{(n)} = \text{MLP}^A([\bar{\mathbf{f}}_{a_i}, \mathbf{z}^{(n)}]; W^A), \quad (4)$$

where $[\cdot, \cdot]$ denotes concatenation, and the MLP with weights W^A returns a rotation $\bar{\mathbf{R}}_{a_i}^{(n)} \in SO(3)$ and translation vector $\bar{\mathbf{t}}_{a_i}^{(n)} \in \mathbb{R}^3$. To capture fine-scale flexibility, additional shifts are applied to individual Gaussians (Fig. 2C, middle), i.e.,

$$\mathbf{c}_i^{(n)} = \tilde{\mathbf{c}}_i^{(n)} + \mathbf{t}_i^{(n)}, \quad \text{where } \mathbf{t}_i^{(n)} = \text{MLP}_c^G([\mathbf{f}_i, \mathbf{z}^{(n)}]; W_c^G). \quad (5)$$

Here, the network MLP_c^G , with separate weights W_c^G , generates individual Gaussian displacements, $\mathbf{t}_i^{(n)} \in \mathbb{R}^3$, which are smooth as Gaussians associated with the same anchor will have similar features.

Finally, to account for compositional variability, where regions of a density map may be missing, we modulate Gaussian amplitudes (Fig. 2C, bottom), as

$$m_i^{(n)} = m_i \times \sigma_i^{(n)}, \quad \text{where } \sigma_i^{(n)} = \text{MLP}_m^G([\mathbf{f}_i, \mathbf{z}^{(n)}]; W_m^G). \quad (6)$$

Here, MLP_m^G is an MLP with a sigmoid output activation to restrict the modulation to $(0, 1)$. Values close to 0 and 1, respectively, correspond to inactive (absent) and active (present) Gaussians. Considering both modifications to centers and amplitudes, we obtain a modulated set of 3D Gaussians for n -th particle image, $\mathcal{G}^{(n)} = \{(\mathbf{c}_i^{(n)}, s_i, m_i^{(n)})\}$. Gaussian scales remain the same as they control local resolution, a factor independent of structural variability.

We jointly optimize the parameters Θ (which includes Gaussian and anchor parameters and MLP weights), and the per-image latent coordinates, $Z = \{\mathbf{z}^{(n)}\}$, by minimizing the objective (Fig. 2D)

$$L(\Theta, Z) = \frac{1}{N} \sum_{n=1}^N \mathcal{L}(I^{(n)}, \hat{I}^{(n)}) + \lambda_z \|\mathbf{z}^{(n)}\|_2^2 + \lambda_f \sum_{i=1}^G \|\Delta \mathbf{f}_i\|_2^2, \quad (7)$$

where the reconstruction loss, $\mathcal{L}$, is proportional to the negative image log-likelihood (i.e., the squared error between $I^{(n)}$ and $g^{(n)} \star \hat{I}^{(n)}$ where $g^{(n)}$ is the microscope point spread function and $\hat{I}^{(n)}$ is the 2D projection of $\mathcal{G}^{(n)}$ from Eq. 2). The second term imposes a zero-mean Gaussian prior over the per-image latent codes, ensuring latent coordinates remain bounded [26, 33], while the third term regularizes Gaussians to remain close to the anchor in the feature space. λ_z and λ_f are hyperparameters that control the relative strength of these priors.

3.2 Part Discovery for Model Initialization

The part discovery process is illustrated in Fig. 2A. We optimize a coarse-grained model without anchors and with fewer Gaussians, similarly parameterized as $\mathcal{G} = \{(\mathbf{f}_i, \mathbf{c}_i, s_i, m_i)\}_{i=1}^G$. Here, the Gaussian features, $\mathbf{f}_i$, are directly learnable parameters (and randomly initialized). We use MLP_c^G (Eq. 5), to shift Gaussian centers and MLP_m^G (Eq. 6) to modulate Gaussian amplitudes. The parameters are estimated using the L2 reconstruction loss and the latent prior, similar to the objective in Eq. 7. Once optimized, we find that the feature space naturally groups Gaussians into 3D parts that undergo consistent motion or appear and disappear together. Remarkably, this property emerges without any direct supervision on features.

To obtain parts, we apply clustering on the Gaussian features, thereby finding regions with reasonably consistent motion and presence. We then further divide these clusters by clustering in 3D space to ensure large parts are well-covered with anchors. For clustering we simply use k-means++ [1]. We use the position and feature vector of the Gaussian closest to the centroid of the cluster to initialize the anchor set, $\mathcal{A} = \{(\bar{\mathbf{c}}_a, \bar{\mathbf{f}}_a)\}_{a=1}^A$. From the coarse-grained model, we also compute an improved density map which is used to seed the Gaussians of the part-aware model. This provides a more robust initialization, especially in the presence of large-scale motion which can lead to blurred or over-dispersed density. Lastly, the coarse-grained model provides a preliminary estimate of the image latent codes, which are used to initialize latent codes in the part-aware model.

Remark. Methods for 4D scene reconstruction [27, 31], and DynaMight [40] in cryo-EM, often use neural networks to output deformations or motion. However, they condition on positional encodings of input coordinates instead of learnable features. Such fixed conditioning strongly biases deformations to be spatially smooth, whereas our approach with learnable feature space enables a more flexible form of piecewise smoothness, allowing nearby parts to move quite differently. Through an ablation study, we show that positional encodings quantitatively underperform as well.

4 Experimental Setup

We quantitatively compare CryoSPIRE with the state-of-the-art methods, namely, RECOVER [10], CryoDRGN [47], DRGN-AI [19], 3DFlex [33] and 3DVA [32] using the CryoBench benchmark [15]. We also provide qualitative results on experimental datasets.

CryoBench. The sole benchmark for cryo-EM heterogeneity is CryoBench [15], a set of synthetic datasets with ground-truth labels and a protocol for quantitative evaluation. Two datasets, *IgG-ID*

Method	IgG-1D		IgG-RL		Ribosembly	
	Mean (std)	Med	Mean (std)	Med	Mean (std)	Med
3D Classification [39]	0.297 (0.019)	0.291	0.309 (0.01)	0.307	0.289 (0.081)	0.288
CryoDRGN [47]	0.366 (0.003)	0.366	0.349 (0.008)	0.348	0.415 (0.019)	0.415
CryoDRGN-AI-fixed [19]	0.366 (0.001)	0.366	0.355 (0.007)	0.354	0.372 (0.032)	0.374
3DFlex [33]	0.336 (0.002)	0.336	0.339 (0.007)	0.339	-	-
3DVA [32]	0.351 (0.003)	0.351	0.341 (0.006)	0.341	0.375 (0.038)	0.372
RECOVAR [10]	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 2: Mean (standard deviation) and median of AUC of Per-Conformation FSCs on CryoBench datasets [15]. Statistics computed across different structural states, i.e. 100 for IgG-1D and IgG-RL and 16 for Ribosembly (Best method in bold, second best underlined).

and *IgG-RL*, are based on the human immunoglobulin G (IgG) complex, simulating conformational changes by rotating the dihedral angle between the Fab domain and the IgG core (see Fig. 4D), generating 100 distinct conformations, each with 1,000 particle images. *Ribosembly* simulates compositional heterogeneity by successively adding protein subunits and ribosomal RNA, resulting in 16 discrete structural states [35]. It has 335,240 particle images, with non-uniform distribution over the 16 compositional states. All particle images have 128×128 pixels, and are simulated with realistic point spread functions and a signal-to-noise ratio (SNR) of 0.01.

Experimental Datasets. We also evaluate on two real datasets: EMPIAR-10076 is a dataset comprising assemblies of intermediates of the *Escherichia coli* large ribosomal subunit (LSU) [7], with 131,899 particle images (320×320 pixels, with pixel size 1.31 Å). In the original study, four major assembly states were identified [7], with a subset of particles labeled as unassigned (non-ribosomal impurities) or 30S subunits. We also consider EMPIAR-10180, a conformationally heterogeneous dataset of Pre-Catalytic Spliceosome [30]. A total of 327,490 particle images were collected (320×320 pixels, with pixel size 1.69 Å). Consistent with other heterogeneity methods considering this dataset [10, 47] we perform analysis on a filtered subset of 139,722 images.

Implementation Details. For part discovery, we seed $G = 2,048$ components using the rigid reconstruction and adopt lightweight MLPs with a single hidden layer of $H = 32$ units. The latent space, $\mathcal{Z}$, has dimensionality $D = 4$ and the feature space, $\mathcal{F}$, has dimensionality $E = 24$. We optimize the part discovery model for 15 and 50 epochs on synthetic and experimental datasets. The part-aware GMMs are optimized for 30 epochs, using $G = 8,192$ components, except for Ribosome synthetic and experimental datasets with $G = 16,384$, and have MLPs with three hidden layers and $H = 128$ hidden units. On experimental datasets, we perform part discovery on downsampled 128×128 images for efficiency, while the part-aware GMM is optimized on 256×256 images. We use batch size $B = 64$ and set hyperparameters $\lambda_z = 0.1$, $\lambda_f = 0.01$. The optimization runs on a single NVIDIA GeForce RTX 2080, taking 3 to 6 hours depending on the number of Gaussians in the model.

Evaluation Metrics. The quality of cryo-EM density maps are evaluated using Fourier Shell Correlation (FSC) [43], which is the normalized cross-correlation between two independently estimated density maps, as a function of frequency. Metrics for heterogeneity are less standardized, but the most common is Per-Conformation FSC (or Per-Conf FSC) [15], proposed by CryoBench [15]. Per-Conf FSC is the average FSC between the ground-truth 3D structure of a particle state, and the 3D structure corresponding to the average latent position of images associated with that state. The Per-Conf FSC requires knowledge of ground-truth 3D structures for each image which is not available for experimental data and we instead rely on qualitative evaluation of the estimated parts and structures. FSC results in a curve which can be summarized by computing the area under the curve (AUC) to more easily compare methods. See the supplement for more details on metrics.

5 Results

Quantitative comparison on the three relevant CryoBench [15] datasets are provided in Table 2 and Fig. 3. Note that CryoSPIRE outperforms 3DVA and 3DFlex which are among the most widely used methods in cryo-EM at present. As 3DFlex cannot handle compositional changes, it was not evaluated on Ribosembly. CryoSPIRE outperforms non-linear latent variable models, Cryo-DRGN

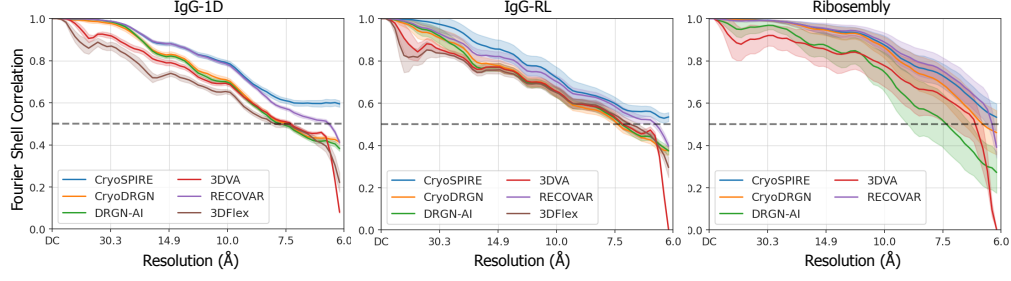


Figure 3: Per-Conformation FSC on CryoBench datasets. Error bars indicate standard deviation across different conformations. The highest possible resolution is 6 Å on these synthetic datasets.

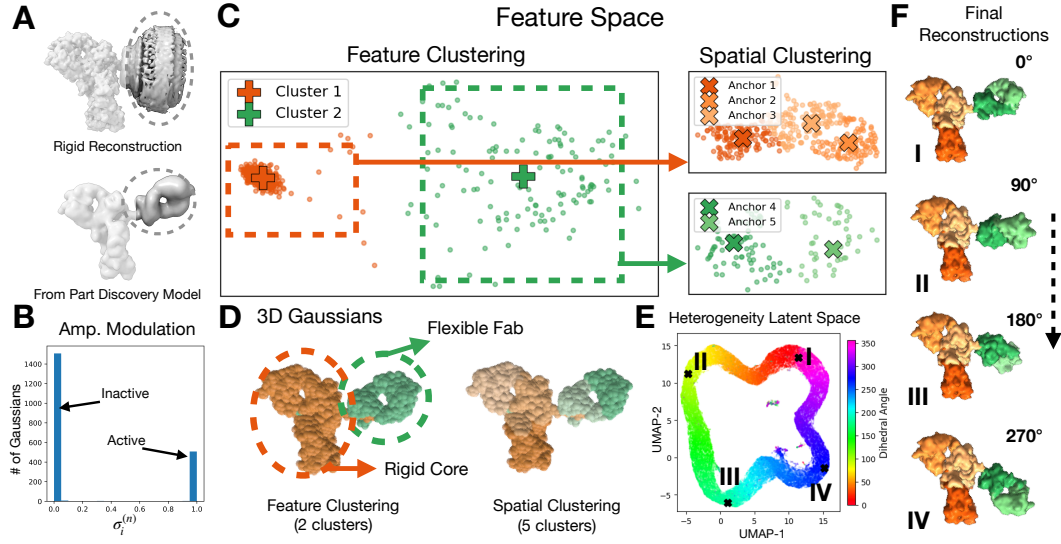


Figure 4: Results on IgG-1D [15]. (A) Due to large motion, the Fab domain (circled) is smeared out in rigid reconstruction, while our part discovery model identifies this domain and resolves its structure and motion, providing good initialization for subsequent modeling. (B) For a sample structure, the histogram of amplitude modulations indicate active and inactive Gaussians. (C) Gaussian feature space, $\mathcal{F}$, shows two distinct groups (green, orange), corresponding to the flexible Fab and the rigid core; feature clustering finds these groups and divides further based on spatial proximity, yielding 5 parts. (D) Configuration of 3D Gaussians after Level-1 and Level-2 clustering. (E) The latent space, $\mathcal{Z}$, captures conformation change (colored based on ground truth Fab orientation). (F) Sample structures from model corresponding to four latent points, showing rotation of the Fab domain (green).

and DRGN-AI, especially on IgG-1D and IgG-RL by a large margin. The most competitive method is the linear subspace model of RECOVER, which, as reported, is memory intensive due to allocation of several bases and is not as interpretable without motion modeling. While CryoSPIRE significantly outperforms RECOVER on IgG datasets, its performance on Ribosembly, where linear subspace models are more favorable by design, is not statistically different from RECOVER. Relative to the nominal FSC threshold of 0.5 for comparison to ground truth [36], the FSC curves in Fig. 3 indicate that CryoSPIRE finds higher resolution density maps. Finally, we note that CryoSPIRE is the first GMM-based method to be successfully evaluated on CryoBench.

IgG-1D & IgG-RL (CryoBench [15]). The flexible Fab domain (circled in Fig. 4A, top) in the rigid reconstruction, used as input for part discovery, is poorly resolved. However, the part discovery model learns to selectively deactivate incoherent parts, as shown in the histogram of the modulation factors $\sigma_i^{(n)}$ in Fig. 4B. This enables a more robust initialization (Fig. 4A, bottom) of the hierarchical GMM. The Gaussian feature space, $\mathcal{F}$, shows two clusters corresponding to the flexible Fab domain from the rigid core (Fig. 4C for IgG-1D and Fig. 5B for IgG-RL). Spatial clustering produces a total of five and six anchors for IgG-1D and IgG-RL, respectively. The latent heterogeneity space, $\mathcal{Z}$,

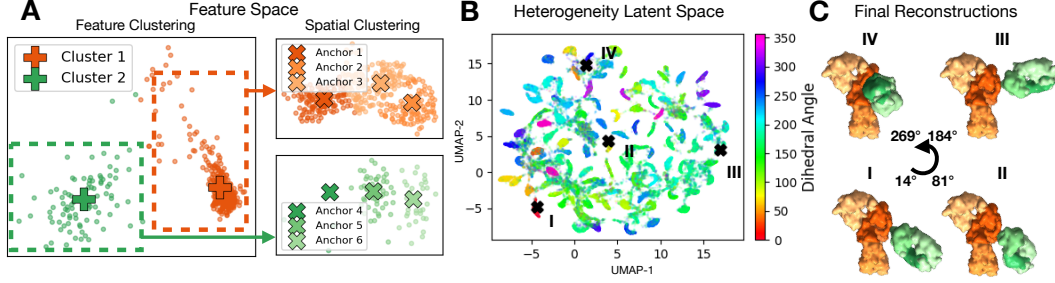


Figure 5: Results on IgG-RL [15]. (A) The feature space, $\mathcal{F}$, shows two parts (green and orange) corresponding to the flexible Fab domain and the rigid core. Subsequent spatial clustering yields six parts. (B) The latent space, $\mathcal{Z}$, is colored with Fab orientation along with four sampled latent points that capture rotation of the Fab domain (comprising three parts). The motion of the Fab domain in IgG-RL is not as regular as that in IgG-1D, as reflected in the latent space. (C) The corresponding density maps are provided with parts illustrated in different colors.

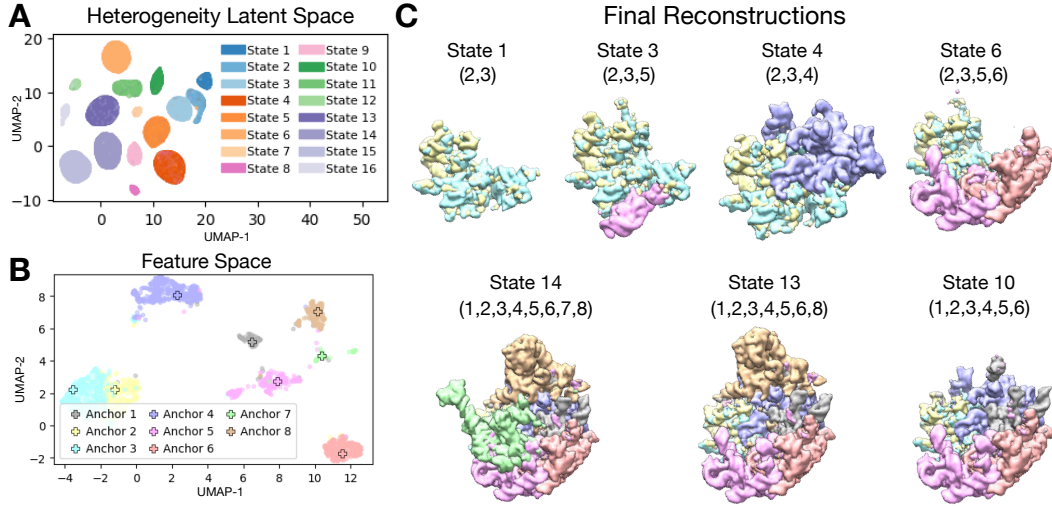


Figure 6: Results on Ribosembly [15] (A) Gaussian feature space, $\mathcal{F}$, showing eight major parts identified through clustering. (B) Heterogeneity latent space, $\mathcal{Z}$, colored coded with the ground-truth compositional state. (C) Visualizations of 3D density maps corresponding to seven points in latent space, with colors depicting parts (given in parentheses).

indicates a circular manifold of dihedral angles for IgG-1D, see Fig. 4D. Four structures from the latent space in both datasets demonstrate that the Fab domain, covered by a few parts, undergo a large, predominantly rigid motion, while the rest of the complex remains fixed.

Ribosembly (CryoBench [15]). After part discovery, we obtain eight parts (see Fig. 6A) that are used to initialize eight anchors in the part-aware GMM. In Fig. 6B the learned latent space, $\mathcal{Z}$, clearly distinguishes between the different compositional states. For seven selected states, we visualize the estimated structure (Fig. 6C), colored based on the discovered parts.

Large Ribosomal Subunit (EMPIAR-10076 [7]). We find four major assembly states in the part discovery latent space (labeled as (I, II, V, VI) in Fig. 7A, left), which match classes (C, E, B, D) in the original study [7], with unassigned particles and 30S contaminants grouped in states III and IV, which are excluded when optimizing hierarchical model (See supplement for more details). The Gaussian feature space, $\mathcal{F}$, (Fig. 7B) shows four distinct parts which also align with previously reported structural blocks in the original study (cf. [7], Fig. 6). By analyzing the heterogeneity latent space, $\mathcal{Z}$, of the part-aware model (Fig. 7A, right), we show that the major states can be further divided into subpopulations; e.g., the major state I is represented with minor states (1, 2) and the major state II has branched into minor states (3, 4, 5). The associated structures, shown in Fig. 7E, are consistent with minor states reported in the original study [7].

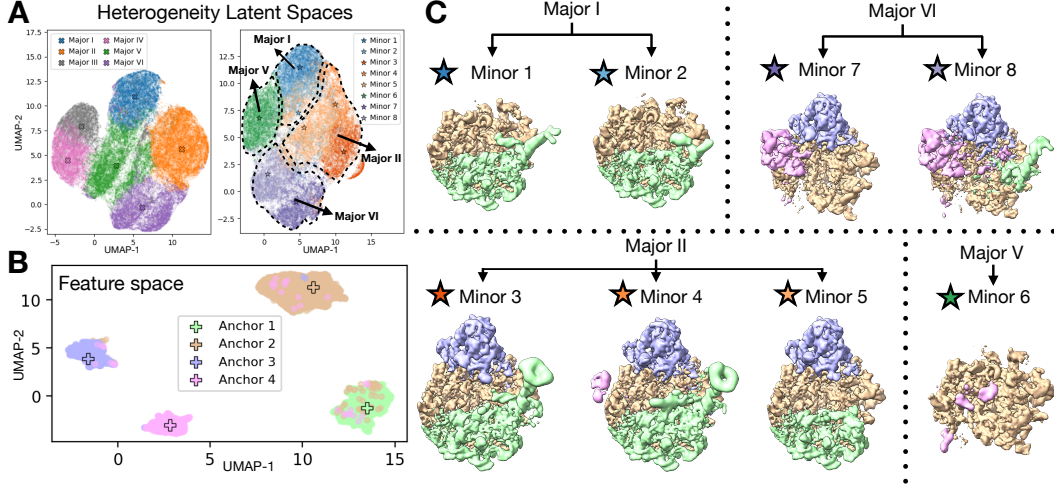


Figure 7: Results on Large Ribosomal Subunit (EMPIAR-10076 [7]). (A) The learned heterogeneity latent spaces, $\mathcal{Z}$, in part discovery (left) identifies the four major assembly states (I, II, V, VI) and two groups of impurities (III, IV). After fitting the part-aware model, the major states, with impurities excluded, can be further categorized (right) into eight color-coded minor structural states. (B) The part discovery Gaussian feature space, $\mathcal{F}$, reveals four parts which are used to construct the part-aware model. (C) The structures corresponding to different states, colored by inferred part.

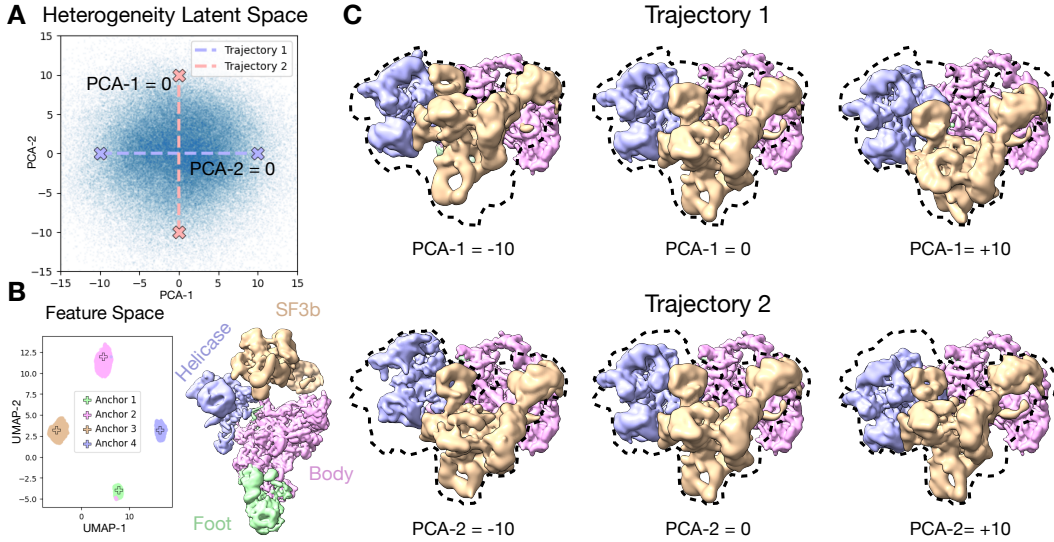


Figure 8: Results on Pre-Catalytic Spliceosome (EMPIAR-10180 [30]). (A) PCA of the latent space, $\mathcal{Z}$, is used to generate two structural trajectories. (B) The Gaussian feature space, $\mathcal{F}$, shows four parts which correspond to known helicase, SF3b, body and foot domains as shown in 3D visualization of Gaussian components configuration. (C) Three states along each trajectory. In both trajectories, body is rigid while SF3b and helicase show large-scale motion.

257 **Pre-Catalytic Spliceosome (EMPIAR-10180 [30]).** The feature space, $\mathcal{F}$, of the part discovery
 258 model (Fig. 8B), shows four distinct clusters, which correspond to coherent structural regions –
 259 foot, body, helicase, and SF3b – consistent with the original study [30]. Accordingly, we optimize
 260 the part-aware model with four anchors. To illustrate structural variability, we run PCA on the
 261 heterogeneity latent space, $\mathcal{Z}$, and extract two principal directions illustrated in Fig. 8A. Top views
 262 of density maps along the two principal directions (Fig. 8C) show two modes of conformational
 263 heterogeneity. The first direction reflects a forward–backward rotation of the SF3b and helicase
 264 regions. The second direction captures a side-to-side rotation of SF3b, and a diagonal shift of the
 265 helicase. Please see the supplement for more visualization on conformational changes.

Method	IgG-1D	IgG-RL	Ribosembly
CryoSPIRE	0.402 (0.002)	0.386 (0.014)	0.427 (0.014)
w/o hier. motion	0.388 (0.002)	0.372 (0.010)	0.425 (0.015)
over-segment	0.384 (0.002)	0.375 (0.010)	0.423 (0.016)
w/ pos. encoding	0.377 (0.002)	0.361 (0.007)	0.415 (0.023)

Table 3: Mean AUC-FSCs reported on datasets from CryoBench [15] for ablation study.

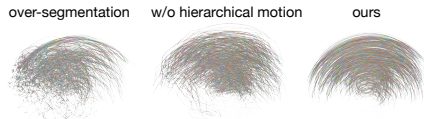


Figure 9: Estimated motion of Gaussians for 30 states of IgG-1D. The baselines fail to capture local rigidity.

5.1 Ablation Study

Here, we ablate key design decisions in our framework. To demonstrate the importance of anchor-based motion modeling in CryoSPIRE, we consider a baseline without anchors that uses an MLP to directly learn deformations of individual Gaussians. Quantitative comparison on IgG-1D and IgG-RL, as in Table 3, shows that the lack of anchor based motion leads to inferior results. This is less critical for Ribosembly with minor conformational changes. We also compare with a model that over-segments the structure by using $K = 64$ anchors, which achieves worse performance. In Fig. 9, we visualize the estimated motion of Gaussians on the IgG-1D dataset. Both baselines fail to capture the locally rigid and smooth motion. Finally, we consider a baseline where the Gaussian feature space is replaced with a positional encoding, similar to previous methods, e.g., [40]. This baseline is unable to identify meaningful parts and achieves inferior quantitative performance.

6 Conclusion

We present CryoSPIRE, a hierarchical cryo-EM density model to capture conformational and compositional heterogeneity in the 3D structure of biomolecules from 2D images. This includes a novel method for part discovery and a hierarchical Gaussian mixture model for which the parts provide meaningful inductive biases to regularize model fitting. CryoSPIRE establishes a new state-of-the-art on the CryoBench heterogeneous benchmark, and produces meaningful parts on experimental data.

While CryoSPIRE shows promising results, limitations exist. First, validation of estimated structures and variability from heterogeneous experimental data remains an open problem for all methods, including CryoSPIRE. Second, interpreting the inferred latent space remains challenging, specifically how it may relate to the biophysical energy landscape of molecular states. Third, learning per-Gaussian features is a key design choice in cryoSPIRE, as it provides the inductive bias that drives features to encode local structural heterogeneity. To that end, we have only used very simple algorithms like k-means++, which requires manual selection of the number of clusters (parts). Further research will be useful to find more effective forms of clustering, perhaps incorporating principled biophysics criteria like free energy. Finally, like other methods, we presume an initial estimate of the structure and image poses; inaccuracies in these may limit CryoSPIRE’s efficacy. A fully *ab initio* method for heterogeneous data remains an open problem.

Broader Impact

Cryo-electron microscopy (cryo-EM) has emerged as a revolutionary technique in structural biology, enabling the determination of macromolecular structures with significant societal impact. Computational methods, grounded in machine learning and computer vision have now been used to determine many thousands of biological structures. Notably, cryo-EM played a pivotal role in elucidating the structure of the SARS-CoV-2 spike protein, revealing its pre-fusion conformation and aiding in the assessment of medical countermeasures. Complementing computational methods such as AlphaFold for protein structure prediction, cryo-EM has revolutionized our understanding of cellular processes and accelerated the development of novel therapeutics, including synthetic antibodies. Nevertheless, we strongly condemn any usage of our proposed hierarchical 3D GMM representation for generating malicious data, improperly modifying signals, or spreading misinformation.

Acknowledgments and Disclosure of Funding

We thank Ali Punjani and John Rubinstein for helpful discussions. This research was supported in part by the Province of Ontario, the Government of Canada (through NSERC, CIFAR, and the Canada First Research Excellence Fund for the Vision: Science to Applications (VISTA) programme), and by the companies sponsoring the Vector Institute for Artificial Intelligence.

References

- [1] David Arthur and Sergei Vassilvitskii. K-Means++: The advantages of careful seeding. *Proc. 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, page 1027–1035, 2007. 5, 2
- [2] Jeongmin Bae, Seoha Kim, Youngsik Yun, Hahyun Lee, Gun Bang, and Youngjung Uh. Per-Gaussian embedding-based deformation for deformable 3d Gaussian splatting. In *European Conference on Computer Vision (ECCV)*, pages 321–335. Springer, 2024. 1, 3
- [3] Marcus A Brubaker, Ali Punjani, and David J Fleet. Building proteins in a day: Efficient 3d molecular reconstruction. *Proc IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015. 1
- [4] Muyuan Chen and Steven J Ludtke. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature Methods*, 18(8):930–936, 2021. 2, 3
- [5] Muyuan Chen, Michael F. Schmid, and Wah Chiu. Improving resolution and resolvability of single particle cryoem structures using Gaussian mixture models. *Nature Methods*, 21:37–40, 2023. 2, 3
- [6] Muyuan Chen, Bogdan Toader, and Lederman Roy. Integrating molecular models into cryoem heterogeneity analysis using scalable high-resolution deep Gaussian mixture models. *Journal of Molecular Biology*, 435(9):168014, 2023. 2, 3
- [7] Joseph H Davis, Yong Zi Tan, Bridget Carragher, Clinton S Potter, Dmitry Lyumkis, and James R Williamson. Modular assembly of the bacterial large ribosomal subunit. *Cell*, 167(6):1610–1622, 2016. 6, 8, 9, 2, 7
- [8] Robert A Drebin, Loren Carpenter, and Pat Hanrahan. Volume rendering. *ACM SIGGRAPH*, 22(4):65–74, 1988. 3
- [9] J. Frank and L. Al-Ali. Signal-to-noise ratio of electron micrographs obtained by cross correlation. *Nature*, 256(5516):376–379, 1975. 3
- [10] Marc Aurèle Gilles and Amit Singer. Cryo-EM heterogeneity analysis using regularized covariance estimation and kernel regression. *Proc. National Academy of Science*, 122(9):e2419140122, 2025. 1, 2, 5, 6, 4
- [11] Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned Gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering. In *Proc. CVPR*, 2024. 3
- [12] G. Harauz and M. van Heel. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Optik*, 73:146–156, 1986. 3
- [13] David Herreros, Roy R Lederman, James M Krieger, Amaya Jiménez-Moreno, Marta Martínez, David Myška, D Strelak, J Filipovic, Carlos OS Sorzano, and José M Carazo. Estimating conformational landscapes from cryo-em particles by 3d zernike polynomials. *Nature Communications*, 14(1):154, 2023. 1
- [14] Yue Huang, Chengguang Zhu, Xiaokang Yang, and Manhwa Liu. High-resolution real-space reconstruction of cryo-em structures using a neural field network. *Nature Machine Intelligence*, 6:892–903, 2024. 2
- [15] Minkyu Jeon, Rishwanth Raghu, Miro Astore, Geoffrey Woollard, Ryan Feathers, Alkin Kaz, Sonya M. Hanson, Pilar Cossio, and Ellen D. Zhong. CryoBench: Diverse and challenging datasets for the heterogeneity problem in cryo-em. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2, 5, 6, 7, 8, 10, 3, 4

- [16] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2023. 3, 2
- [17] Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3D Gaussian splatting as Markov chain Monte Carlo. In *Proc. NeurIPS*, 2024. 3
- [18] Werner Kühlbrandt. The Resolution Revolution: Advances in detector technology and image processing are yielding high-resolution electron cryo-microscopy structures of biomolecules. *Science*, 343(6178):1443–1444, 2014. 1
- [19] Axel Levy, Michal Grzadkowski, Frederic Poitevin, Francesca Vallese, Oliver B Clarke, Gordon Wetzstein, and Ellen D Zhong. Revealing biomolecular structure and motion with neural ab initio cryo-em reconstruction. *bioRxiv*, 2024. 1, 2, 5, 6, 4
- [20] Axel Levy, Gordon Wetzstein, Julien NP Martel, Frederic Poitevin, and Ellen Zhong. Amortized inference for heterogeneous reconstruction in cryo-EM. *Advances in Neural Information Processing Systems*, 35:13038–13049, 2022. 1
- [21] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-gs: Structured 3d Gaussians for view-adaptive rendering. In *Proc IEEE/CVF Conf Computer Vision and Pattern Recognition (CVPR)*, pages 20654–20664, 2024. 1, 3
- [22] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3D Gaussians: Tracking by persistent dynamic view synthesis. In *Proc. 3DV*, 2024. 3
- [23] Nelson Max. Optical models for direct volume rendering. *IEEE Trans. Vis. Comput. Graph.*, 1(2):99–108, 2002. 3
- [24] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NERF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2022. 2
- [25] Takanori Nakane and Sjors HW Scheres. Multi-body refinement of cryo-em images in relion. In *CryoEM: Methods and Protocols*, pages 145–160. Springer, 2020. 3
- [26] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proc IEEE/CVF Conf Computer Vision and Pattern Recognition (CVPR)*, pages 165–174, 2019. 5
- [27] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proc IEEE/CVF International Conference on Computer Vision (CVPR)*, pages 5865–5874, 2021. 5
- [28] Pawel A. Penczek. Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. *Journal of Structural Biology*, 138(1):34–46, 2002. 3
- [29] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Elaine C Meng, Gregory S Couch, Tristan I Croll, John H Morris, and Thomas E Ferrin. Ucsf chimeraX: Structure visualization for researchers, educators, and developers. *Protein science*, 30(1):70–82, 2021. 1
- [30] Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. Structure of a pre-catalytic spliceosome. *Nature*, 546(7660):617–621, 2017. 6, 9, 2, 7, 8
- [31] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR)*, pages 10318–10327, 2021. 5
- [32] Ali Punjani and David J Fleet. 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *Journal of Structural Biology*, 213:107702, 2021. 1, 2, 5, 6, 4, 7, 8
- [33] Ali Punjani and David J Fleet. 3DFlex: Determining structure and motion of flexible proteins from cryo-em. *Nature Methods*, 20:860–870, 2023. 1, 2, 3, 5, 6, 4, 7, 8

- [34] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. CryoSPARC: Algorithms for rapid unsupervised cryo-em structure determination. *Nature Methods*, 14:290–296, 2017. 2, 7
- [35] Bo Qin, Simon M Lauer, Annika Balke, Carlos H Vieira-Vieira, Jörg Bürger, Thorsten Mielke, Matthias Selbach, Patrick Scheerer, Christian MT Spahn, and Rainer Nikolay. Cryo-em captures early ribosome assembly in action. *Nature Communications*, 14(1):898, 2023. 6
- [36] P. B. Rosenthal and R. Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Molecular Biology*, 333(4):721–745, 2003. 7, 3
- [37] Sjors H. W. Scheres. RELION: Implementation of a Bayesian approach to cryo-em structure determination. *Journal of Structural Biology*, 180(3):519 – 530, 2012. 2
- [38] Sjors H W Scheres. Processing of structurally heterogeneous cryo-em data in RELION. In R A Crowther, editor, *The Resolution Revolution: Recent Advances In cryoEM*, volume 579 of *Methods in Enzymology*, pages 125–157. Academic Press, 2016. 1
- [39] Sjors H W Scheres, Haixiao Gao, Mikel Valle, Gabor T Herman, Paul P B Eggermont, Joachim Frank, and Jose-Maria Carazo. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods*, 4(1):27–29, 2007. 6, 4
- [40] Johannes Schwab, Dari Kimanius, Alister Burt, Tom Dendooven, and Sjors H. W. Scheres. DynaMight: Estimating molecular motions with improved reconstruction from cryo-em images. *Nature Methods*, 21:1855–1862, 2024. 2, 3, 5, 10
- [41] Amit Singer and Fred J Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual review of biomedical data science*, 3(1):163–190, 2020. 1
- [42] M. Unser, B. L. Trus, and A. C. Steven. A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy*, 23(1):39–51, 1987. 3
- [43] Marin Van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3):250–262, 2005. 6, 3
- [44] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4D Gaussian splatting for real-time dynamic scene rendering. In *Proc. CVPR*, 2024. 3
- [45] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proc. CVPR*, 2024. 3
- [46] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3D Gaussian splatting. In *Proc. CVPR*, 2024. 3
- [47] Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. CryoDRGN: Reconstruction of heterogeneous cryo-em structures using neural networks. *Nature Methods*, 18:176–185, 2021. 1, 2, 5, 6, 4, 7, 8
- [48] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):223–238, 2002. 3

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: We clearly state the claims in the abstract, summarize them in a list at the end of introduction section, and provide empirical evidence in the results.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: We discuss limitations in the conclusion.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: We provide no theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: In a subsection called implementation details, we included all details needed to reproduce the results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: The datasets are publicly available and we promise to release the code in future.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We provide all the necessary information in the experimental setup section.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We provide various statistics of the quantitative metrics.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We mentioned the GPU resources used to run the proposed method.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We conform with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We discuss societal impacts in the supplement.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We mention the benchmark CryoBench throughout which the comparison with other methods is provided.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset’s creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: We don’t release any new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: No crowdsourcing or research with human subject.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: No crowdsourcing or research with human subject.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

750 Question: Does the paper describe the usage of LLMs if it is an important, original, or
751 non-standard component of the core methods in this research? Note that if the LLM is used
752 only for writing, editing, or formatting purposes and does not impact the core methodology,
753 scientific rigorousness, or originality of the research, declaration is not required.

754 Answer: [NA]

755 Justification: The core method development in this research does not involve LLMs.

756 Guidelines:

- 757 • The answer NA means that the core method development in this research does not
758 involve LLMs as any important, original, or non-standard components.
- 759 • Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>)
760 for what should or should not be described.