CRYOSPLAT2: GAUSSIAN SPLATTING FOR CRYO-EM AB INITIO RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Cryogenic electron microscopy (cryo-EM) is a widely used structural biology technique for determining the 3D structures of biomolecules from noisy 2D particle images. In single-particle cryo-EM, *ab initio* reconstruction jointly estimates 3D density maps and unknown particle poses without an initial 3D reference, making the problem severely ill-posed due to the strong coupling between pose estimation and structure reconstruction. Gaussian representations provide an explicit, continuous, and differentiable model of molecular density, with adaptive degrees of freedom and efficient projection, but most existing Gaussian-based cryo-EM methods assume known particle poses. We introduce cryoSplat2, a Gaussian-based framework for particle-level *ab initio* reconstruction that jointly optimizes particle poses and anisotropic Gaussian representations from random initialization. In particular, cryoSplat2 combines frequency marching for coordinated coarse-to-fine pose and structure refinement, adaptive Gaussian densification for progressively increasing representation capacity, and an optimized differentiable splatting operator for efficient pose inference. Experiments on multiple real cryo-EM datasets show that cryoSplat2 robustly reconstructs 3D structures without an initial 3D reference or externally provided particle poses, including challenging cases where cryoSPARC or cryoDRGN-AI fails to converge to the correct structure. The code will be released upon paper publication.

1 INTRODUCTION

Single-particle cryogenic electron microscopy (cryo-EM) determines three-dimensional (3D) molecular structures from large collections of two-dimensional (2D) particle images. Each 2D particle image corresponds to a projection of the 3D structure at an unknown viewing orientation and an unknown in-plane translation. The task of recovering an initial 3D density directly from noisy particle images without known poses or an initial 3D reference is referred to as *ab initio* reconstruction. This problem is severely ill-posed because pose estimation and density reconstruction are tightly coupled: inaccurate poses degrade the reconstructed density, while an inaccurate density in turn makes pose inference unreliable. The difficulty is further compounded by the extremely low signal-to-noise ratio of individual particle images and by potential structural heterogeneity in experimental datasets (Singer & Sigworth, 2020; Bendory et al., 2020).

Conventional cryo-EM reconstruction methods typically represent the 3D density on a regular voxel grid, where density values are defined at fixed spatial locations throughout the reconstruction volume. Gaussian representations provide an alternative continuous representation in which the molecular density is modeled by a collection of spatial Gaussian primitives with optimizable positions, shapes, and amplitudes. Such representations have been increasingly explored in cryo-EM because they are differentiable, geometrically interpretable, and can concentrate model parameters in regions occupied by molecular density (Chen & Ludtke, 2021; Chen et al., 2024; Li et al., 2024). In particular, cryoSplat (Chen & Ling, 2026) shows that anisotropic Gaussian primitives, together with a differentiable splatting operator adapted to the cryo-EM imaging model, can be optimized stably from random initialization when particle poses are known. However, these Gaussian-based approaches largely assume that particle poses are known or have been estimated beforehand by a separate reconstruction procedure (Chen & Ludtke, 2021; Chen et al., 2023; Li et al., 2024; Chen & Ling, 2026). This assumption removes a central difficulty of *ab initio* reconstruction: the strong coupling between pose estimation and structure recovery. While these studies establish Gaussian

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

primitives as an effective representation for cryo-EM density, it remains unclear whether they can support joint pose and structure estimation required for *ab initio* reconstruction.

Extending Gaussian reconstruction from known poses to *ab initio* reconstruction introduces two central challenges. First, poses and structure must be refined together from highly uncertain initial estimates. The reconstruction should initially be restricted to coarse spatial frequencies and a correspondingly coarse pose search, then progressively incorporate higher spatial frequencies, finer pose sampling, and additional representation capacity as optimization stabilizes (Barnett et al., 2017). For Gaussian representations, this also requires increasing the number of primitives without substantially perturbing the partially reconstructed density. Second, pose estimation requires evaluating and marginalizing over many candidate poses for each particle (Scheres, 2012; Punjani et al., 2017), making the associated forward projections and gradient computations a major computational bottleneck compared with known-pose reconstruction.

To close the gap above, we introduce cryoSplat2, a Gaussian-based framework for cryo-EM *ab initio* reconstruction that jointly estimates particle poses and reconstructs anisotropic Gaussian representations from random initialization. CryoSplat2 combines three main components: a frequency marching scheduler that coordinates pose-search precision and reconstruction bandwidth, an adaptive Gaussian densification that progressively increases representation capacity while controlling structural perturbations, and an optimized differentiable Gaussian splatting implementation for the large projection workload induced by pose inference. Together, these components enable coarse-to-fine joint optimization of particle poses and Gaussian structure without an initial 3D reference or externally provided poses.

Our contributions are summarized as follows:

- To the best of our knowledge, cryoSplat2 is the first Gaussian-based method to perform cryo-EM *ab initio* reconstruction directly from raw particle images, jointly estimating structure and poses without an initial 3D reference.
- We develop adaptive Gaussian densification that increases representation capacity according to the current reconstruction resolution while preserving coarse-scale structure during Gaussian splitting.
- We introduce a frequency marching scheduler that coordinates reconstruction bandwidth, pose-search resolution, and Gaussian densification, allowing pose and structure resolution to progress together during optimization.
- We adapt and optimize differentiable Gaussian splatting for the large batched projection workload required by pose marginalization, substantially reducing the computational cost of joint pose and structure optimization.

Experiments on multiple real cryo-EM datasets demonstrate robust *ab initio* reconstruction directly from raw particle images, without externally provided particle poses or an initial 3D reference, including challenging cases where existing methods fail to converge to the correct structure.

2 Related Work

***Ab initio* reconstruction** seeks to estimate initial 3D density maps together with unknown particle poses directly from particle images, without an initial 3D reference. Early work explored a variety of specialized strategies, including orientation recovery based on common lines, global orientation optimization, and statistically weighted initial volume estimation (Sanz-García et al., 2010; Elmlund & Elmlund, 2012; Wang et al., 2013; Elmlund et al., 2013; Vargas et al., 2014; Sorzano et al., 2015). More recently, cryoSPARC (Punjani et al., 2017) brought *ab initio* reconstruction into a practical framework by combining stochastic gradient descent, Bayesian marginalization over latent poses and classes, and efficient coarse-to-fine optimization. More recent work has primarily improved individual components of this paradigm, such as pose inference and search efficiency, while retaining an explicit volumetric representation (Nashed et al., 2021; Shekarforoush et al., 2024). Modeling continuous conformational heterogeneity, however, is less direct with independent voxelized volumes and typically requires additional deformation or latent transformation models. Neural approaches such as cryoDRGN (Zhong et al., 2021a) instead parameterize a family of 3D densities with a latent-conditioned decoder, making continuous structural variability substantially easier to represent. This flexibility comes with a different set of tradeoffs. The density is encoded implicitly in network parameters, and evaluating the decoder over a large pose space is computationally ex-

pensive. As a result, neural implicit reconstruction methods in the cryoDRGN lineage and related approaches (Zhong et al., 2021b; Levy et al., 2022a;b) typically avoid explicit marginalization over the full pose posterior and instead infer point estimates of particle poses through grid based or amortized inference, with later methods such as cryoDRGN-AI (Levy et al., 2025) further introducing direct optimization of individual poses. While computationally attractive, point estimates discard posterior uncertainty and multimodality, which can increase sensitivity to incorrect pose modes and local optima, particularly in the early stages of *ab initio* reconstruction.

Gaussian-based cryo-EM methods represent 3D density as a weighted mixture of Gaussian primitives, providing an explicit and continuous structural representation. This representation has been explored for heterogeneous structural analysis (Chen & Ludtke, 2021; Chen et al., 2023; 2024; Li et al., 2024). Prior Gaussian-based cryo-EM methods in this line typically use isotropic primitives with a single spatial scale per Gaussian. Optimizing Gaussian representations, however, often relies on careful initialization, regularization, or additional structural constraints. CryoSplat (Chen & Ling, 2026) showed that increasing the representation with anisotropic Gaussians enables stable optimization while keeping large-scale Gaussian projection computationally practical.

Despite the progress, existing Gaussian-based methods largely rely on known or externally estimated particle poses. While known-pose reconstruction evaluates the forward model only at the given poses, pose estimation requires evaluating a large set of candidate poses for every particle, and Bayesian reconstruction additionally requires aggregating information over the pose posterior. Even coarse angular sampling can require thousands of projections per particle, making projection a major computational and memory bottleneck for Gaussian-based *ab initio* reconstruction.

Early work also explored Gaussian or pseudo-atom representations for initial-model reconstruction (Joubert & Habeck, 2015; Vakili & Habeck, 2021), but avoided the particle-level computational challenge by operating on preprocessed, high-SNR 2D class averages. Class averaging simplifies the pose-search problem and greatly increases signal-to-noise ratio, but collapses many individual observations into a much smaller set of averages and can obscure fine structural details and within-class heterogeneity. These approaches therefore did not address the modern particle-level *ab initio* reconstruction, where structure and poses must be jointly estimated directly from noisy, individual particle images corrupted by the contrast transfer function (CTF).

3 METHOD

3.1 OVERVIEW

Our goal is to achieve stable and efficient cryo-EM *ab initio* reconstruction using an explicit Gaussian representation. Given N observed particle images $\boldsymbol{Y}_i \in \mathbb{R}^{D\times D}$, $i = 1, \ldots, N$, the general *ab initio* task seeks to recover K discretized 3D density maps $\mathbf{V}_k \in \mathbb{R}^{D\times D\times D}$, $k = 1, \ldots, K$, while marginalizing over unknown particle poses and structural class assignments. For clarity, we assume $K = 1$ in the main text; the multi-class maximum a posteriori (MAP) formulation is in App. B.

Following the standard Bayesian formulation of cryo-EM reconstruction (Scheres, 2012; Punjani et al., 2017), and assuming a uniform pose prior, the homogeneous *ab initio* objective takes the form

$$\hat{\mathbf{V}} = \arg\max_{\mathbf{V}} \sum_{i=1}^{N} \log \int p(\boldsymbol{Y}_i \mid \mathbf{V}, \boldsymbol{\phi})\, d\boldsymbol{\phi}, \tag{1}$$

where $\boldsymbol{\phi} = (\boldsymbol{W}, \boldsymbol{t})$ denotes a particle pose consisting of a viewing rotation $\boldsymbol{W} \in SO(3)$ and an in-plane translation $\boldsymbol{t} \in \mathbb{R}^2$.

We represent the continuous density field $V : \mathbb{R}^3 \to \mathbb{R}$ using anisotropic Gaussian primitives. The corresponding discrete volume $\mathbf{V}$ is obtained by evaluating V on the voxel grid. Specifically,

$$V(\boldsymbol{r}) = \sum_{j=1}^{M} A_j G(\boldsymbol{r} \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), \tag{2}$$

where M denotes the number of Gaussians, and A_j, $\boldsymbol{\mu}_j$, and $\boldsymbol{\Sigma}_j$ denote the amplitude, center, and covariance of Gaussian j, respectively.

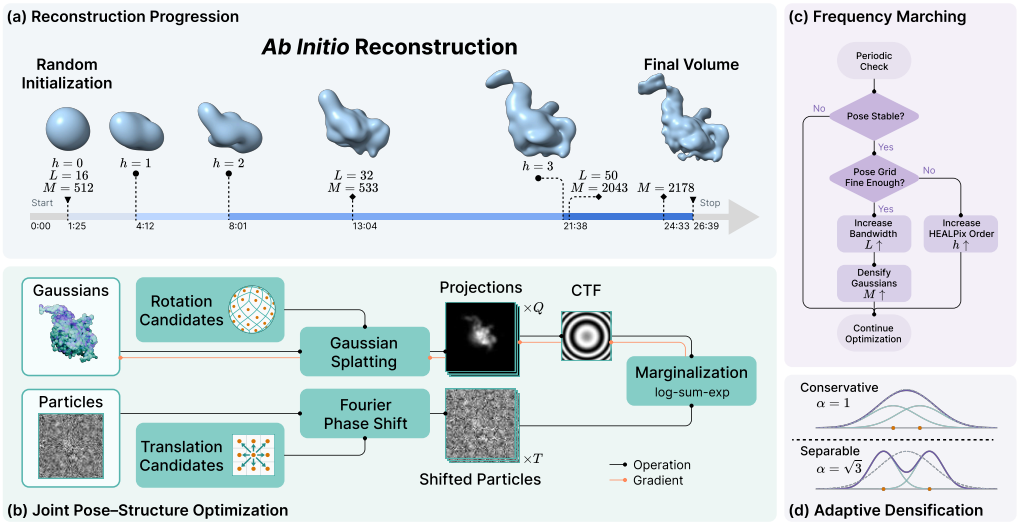

Figure 1: *Ab initio* reconstruction recovers a 3D structure from particle images without known poses or a prior 3D reference. (a) Reconstruction proceeds coarse-to-fine from random initialization, with key changes in pose-search resolution, reconstruction bandwidth, and Gaussian count highlighted along wall-clock time. (b) Joint pose-structure optimization evaluates candidate rotations through batched Gaussian splatting and candidate translations through Fourier phase shifts. (c) Frequency marching couples pose and structure resolution: it increases HEALPix order when pose sampling is insufficient, and increases reconstruction bandwidth once pose estimates are stable and sufficiently resolved, triggering Gaussian densification. (d) Adaptive densification reallocates representation capacity by pruning unsupported Gaussians and splitting coarse primitives.

This formulation introduces three key challenges for Gaussian-based *ab initio* reconstruction: (i) early-stage pose and structure estimates are both unreliable and tightly coupled, making stable joint optimization difficult under extremely noisy observations; (ii) the Gaussian representation must progressively increase its capacity as reconstruction resolution improves without substantially perturbing the partially recovered structure and its projections; and (iii) marginalization over a large pose space requires a massive number of forward and backward projection evaluations.

We address these challenges with three corresponding components: (i) a frequency-marching scheduler that progressively increases spatial-frequency bandwidth and pose-search precision for stable joint optimization; (ii) an adaptive densification strategy that increases representation capacity according to the current reconstruction resolution while controlling perturbations to the current reconstruction during splitting; and (iii) a high-throughput differentiable Gaussian splatting implementation tailored to the large-scale projection workload required by pose inference. Together, these components form a coarse-to-fine reconstruction framework that jointly improves pose estimates and the Gaussian representation while controlling optimization stability and computational cost (Fig. 1).

3.2 Joint Pose–Structure Optimization

As illustrated in Fig. 1(b), the pose space is discretized into candidate viewing rotations $\boldsymbol{W}_q$, $q = 1, \ldots, Q$, and in-plane translations $\boldsymbol{t}_\ell$, $\ell = 1, \ldots, T$. Likelihood evaluation is performed in Fourier space. Each rotation candidate defines a projection of the current Gaussian representation, followed by particle-specific CTF modulation to obtain the predicted Fourier image $\widehat{\boldsymbol{X}}_{iq}$. In parallel, each translation candidate is applied to the observed image through a Fourier phase shift, yielding $\widehat{\boldsymbol{Y}}_{i\ell}$. Let $\boldsymbol{f} = [f_x, f_y, 0]^\top$ denote a spatial-frequency coordinate on the central Fourier plane. Using the current radially averaged noise spectrum $\sigma^2(\|\boldsymbol{f}\|)$ (see App. F for its estimation), each pose hypothesis is evaluated by the noise-weighted residual energy

$$E_{iq\ell} = \sum_{\boldsymbol{f}} \frac{\left|\widehat{\boldsymbol{X}}_{iq}(\boldsymbol{f}) - \widehat{\boldsymbol{Y}}_{i\ell}(\boldsymbol{f})\right|^2}{2\sigma^2(\|\boldsymbol{f}\|)}. \tag{3}$$

Discretizing the pose integral in Eq. (1) over the current rotation and translation candidates, the posterior probability of each pose candidate becomes

$$p_{iq\ell} = \frac{\exp(-E_{iq\ell})}{\sum_{q'=1}^{Q}\sum_{\ell'=1}^{T}\exp(-E_{iq'\ell'})}. \tag{4}$$

Rather than retaining only the maximum-posterior pose, we marginalize over all pose candidates. For a batch of B particles, the negative log marginal likelihood is

$$\mathcal{L} = \frac{1}{B}\sum_{i=1}^{B}\left[\log(QT) - \log\sum_{q=1}^{Q}\sum_{\ell=1}^{T}\exp(-E_{iq\ell})\right]. \tag{5}$$

We optimize the Gaussian parameters by backpropagating through the differentiable projection operator. The resulting structure gradient combines contributions from all pose candidates.

3.3 Frequency marching scheduler

Single-particle cryo-EM images contain substantially stronger signal at low spatial frequencies, while the signal-to-noise ratio decreases rapidly toward higher frequencies. Frequency marching (Barnett et al., 2017) exploits this property by beginning reconstruction at low spatial resolution and progressively introducing higher spatial frequencies as the structure and particle poses improve. Thus, the reconstruction bandwidth and pose-search resolution should progress together.

We explicitly coordinate structure and pose resolution through a unified scheduler. In our setting, the Fourier crop side length L controls the reconstruction bandwidth; the HEALPix order h and translation sampling determine the pose-search granularity; and the spatial scales and number of Gaussian primitives control the resolution and capacity of the structural representation. At each stage, a common nominal resolution coordinates these quantities, allowing the reconstruction bandwidth, pose-search granularity, and Gaussian representation capacity to progress together.

As shown in Fig. 1(c), progression between stages is determined by the stability of the estimated poses. We periodically measure pose changes using the root-mean-square (RMS) variation of the estimated orientations and translations. If the current pose sampling is insufficient for the current reconstruction bandwidth, the HEALPix order h is increased. Once the poses stabilize at the current resolution, the scheduler increases L to incorporate higher spatial frequencies and simultaneously triggers Gaussian densification to increase representation capacity. The size of each bandwidth update is further modulated by posterior confidence, allowing faster progression when the pose posterior is concentrated and more conservative updates when pose uncertainty remains high. More details about frequency marching are in App. C.

3.4 Adaptive Gaussian densification

The Gaussian representation must adapt as the reconstruction bandwidth increases. Coarse Gaussians cannot represent newly introduced high-frequency structure, whereas overly small primitives yield unstable geometric gradients. We therefore couple Gaussian initialization, pruning, and densification to the reconstruction bandwidth.

We parameterize each Gaussian by its center $\boldsymbol{\mu}$, orientation $\boldsymbol{R} \in SO(3)$, principal-axis scales $\boldsymbol{s} = [s_1, s_2, s_3]^\top$, and amplitude A, with

$$\boldsymbol{\Sigma} = \boldsymbol{R}\,\mathrm{diag}(s_1^2, s_2^2, s_3^2)\,\boldsymbol{R}^\top. \tag{6}$$

Since G is normalized, A represents its integrated density.

Resolution-aware random initialization. Gaussian centers are randomly initialized within the estimated particle support, while their initial scales are matched to the starting reconstruction bandwidth. A bandwidth-dependent lower scale bound prevents primitives from collapsing below the current pixel grid spacing. The overall amplitude is calibrated from the low-frequency power of the observed particles. Detailed initialization is given in App. D.

Density-guided pruning. Before each densification step, we remove Gaussian primitives that are unlikely to belong to the reconstructed signal region. The Gaussian representation typically yields a

low-density background outside molecular support. Therefore, Gaussians whose centers fall outside this support are pruned before splitting, preventing repeated refinement of inactive primitives and concentrating representation capacity on signal regions.

Moment-preserving Gaussian splitting. As the reconstruction bandwidth increases, Gaussians that remain too coarse are recursively split. While densification in 3DGS (Kerbl et al., 2023) is primarily designed to improve rendering quality and efficiency, cryo-EM reconstruction places a stronger requirement on preserving the underlying 3D density. We therefore use a binary split whose child parameters preserve the integrated amplitude, mean, and covariance of the parent, limiting structural perturbations introduced by densification. For a parent Gaussian with scale s along principal direction $\boldsymbol{u}$, the two children are defined as $q_1 = -\frac{\alpha}{2}$, $q_2 = \frac{\alpha}{2}$, $\boldsymbol{\mu}_c = \boldsymbol{\mu} + sq_c\boldsymbol{u}$, $A_c = \frac{A}{2}$, and $s_c = s\sqrt{1 - \frac{\alpha^2}{4}}$, while the remaining principal-axis scales and orientation are unchanged.

The split strength α controls a trade-off between structural consistency and child separability, as illustrated in Fig. 1(d). With $\alpha = 1$, the children strongly overlap and their summed density remains close to the parent, yielding a conservative split. Increasing the separation gives the children greater freedom to evolve independently; in particular, $\alpha = \sqrt{3}$ halves the split-axis scale and produces a more separable refinement. The general n-way construction and relevant derivation are in App. D.

Bandwidth-aware densification. Each reconstruction bandwidth defines a target spatial scale for the Gaussian representation. When the scheduler increases L, pruning is first applied to remove unsupported primitives, after which Gaussians that remain too coarse are recursively split until their scales are compatible with the new bandwidth or the maximum Gaussian count is reached. In this way, higher-frequency stages progressively release additional degrees of freedom only where supported by the reconstructed signal.

Overall, densification reallocates representation capacity by pruning unsupported Gaussians and recursively splitting coarse primitives as the reconstruction bandwidth increases.

3.5 Efficient Gaussian splatting

The differentiable Gaussian splatting operator in cryoSplat is designed for known-pose reconstruction, where each particle requires a single projection. In *ab initio* reconstruction, pose estimation instead evaluates Q candidate rotations for every particle. Although each projection can be reused across translation candidates, repeatedly evaluating and differentiating these views makes projection throughput a major computational bottleneck.

We retain the same cryo-EM Gaussian projection operator but redesign its implementation for batched pose inference. Our implementation builds on batched rasterization from gsplat (Ye et al., 2025), per-Gaussian backward parallelization related to Taming 3DGS (Mallick et al., 2024), and accurate Gaussian–tile overlap culling from Speedy-Splat (Hanson et al., 2025). We adapt and integrate these techniques for the large batched projection workload induced by pose marginalization.

These optimizations leave the cryo-EM forward model and marginal objective unchanged, but substantially reduce the cost of pose estimation. Implementation details are provided in App. E.

4 Experiments

4.1 Experimental Settings

Datasets. We evaluate cryoSplat2 on four publicly available experimental single-particle cryo-EM datasets from the Electron Microscopy Public Image Archive (EMPIAR) (Iudin et al., 2023): EMPIAR-10028 (Pf 80S ribosome), EMPIAR-10180 (pre-catalytic spliceosome), EMPIAR-11762 (TFIIIC τA complex), and EMPIAR-11910 (CCAN–CENP-A complex). These datasets span a range of molecular sizes, image quality, and structural complexity, providing diverse benchmarks for evaluating *ab initio* reconstruction from raw particle images. More details are in App. H.

Evaluation Metrics. The primary goal of *ab initio* reconstruction is to provide a reliable initial model for subsequent refinement. We therefore evaluate whether each *ab initio* reconstruction provides a valid initialization by using its output as the starting volume for cryoSPARC homogeneous

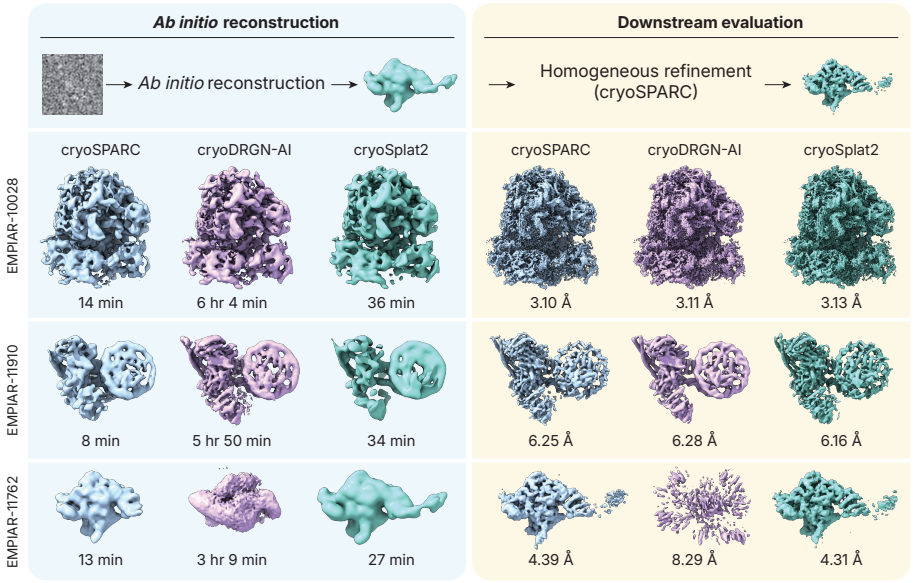

Figure 2: Comparison of *ab initio* reconstruction methods across cryo-EM datasets, with runtime shown below each reconstruction. Homogeneous refinement in cryoSPARC is used for downstream evaluation, with the resulting resolutions indicated.

refinement; successful refinement is treated as evidence that the *ab initio* reconstruction is correct. Moreover, because *ab initio* reconstruction does not naturally produce two independently refined half maps, gold-standard Fourier shell correlation (FSC) cannot be directly used to evaluate the initial reconstruction. For reconstructions optimized to higher target resolutions, such as 5 Å, we additionally compute map-to-reference FSC against the corresponding map deposited in EMDB to quantify agreement with the reference structure and report the FSC-based resolution.

Implementation Details. Unless otherwise specified, all reconstructions use a single structural class ($K = 1$) and a batch size of 32. Reconstruction starts from the coarsest global pose search with HEALPix order $h = 0$ and progressively increases the reconstruction bandwidth and pose-search resolution through frequency marching. The same configuration is used across all datasets without dataset-specific tuning. All cryoSPARC and cryoDRGN-AI baseline experiments use their default settings unless otherwise specified. For direct comparison with cryoSPARC, we set the initial and target reconstruction bandwidths to correspond to 35 Å and 12 Å, respectively, matching cryoSPARC's default *ab initio* resolution schedule. All experiments are run on a single NVIDIA RTX 3090. Additional details are provided in Appendix F.

4.2 Ab Initio Reconstruction on Experimental Datasets

We first evaluate cryoSplat2 on several experimental cryo-EM datasets and compare its *ab initio* reconstructions with cryoSPARC and cryoDRGN-AI. As shown in Fig. 2, cryoSplat2 consistently produces valid initial reconstructions across the tested datasets, as confirmed by subsequent homogeneous refinement in cryoSPARC. cryoSplat2 is moderately slower than cryoSPARC for *ab initio* reconstruction, but substantially faster than cryoDRGN-AI, reducing reconstruction time from several hours to tens of minutes. Notably, for

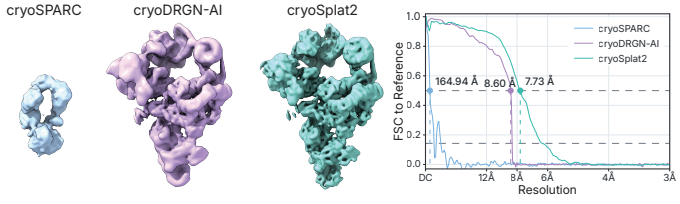

Figure 3: *Ab initio* reconstruction for EMPIAR-10180. Results from cryoSPARC, cryoDRGN-AI, and cryoSplat2 are shown together with their FSC curves against the reference map.

EMPIAR-11762, which contains a relatively small 330-kDa protein complex, cryoDRGN-AI fails to converge to the correct structure, whereas cryoSplat2 produces a valid reconstruction that can be successfully refined. cryoSPARC also occasionally fails to converge to the correct structure on this dataset; Fig. 2 shows a successful run.

We next compare the reconstruction results on EMPIAR-10180. In this case, cryoSPARC fails to converge to the correct structure during *ab initio* reconstruction (Fig. 3). Using the same target resolution for frequency marching as cryoDRGN-AI, cryoSplat2 achieves a higher-resolution reconstruction and shows better agreement with the deposited reference map across most spatial frequencies. The map-to-reference FSC reaches the 0.5 criterion at 7.73 Å for cryoSplat2, compared with 8.60 Å for cryoDRGN-AI, while the incorrect cryoSPARC reconstruction exhibits negligible agreement with the reference.

Finally, we evaluate whether cryoSplat2 can support higher-resolution *ab initio* reconstruction. We set the frequency-marching target resolution to 5 Å, and evaluate the resulting reconstructions directly against the corresponding maps deposited in EMDB (Fig. 4). cryoSplat2 remains stable under this higher-resolution setting across datasets of different sizes and complexities. The resulting map-to-reference FSC curves yield FSC=0.5 resolutions of 5.78 Å, 6.93 Å, and 6.89 Å for EMPIAR-10028, EMPIAR-11762, and EMPIAR-10180, respectively. These results demonstrate that cryoSplat2 can not only provide reliable initial densities for downstream refinement, but also support higher-resolution reconstruction within the same *ab initio* framework.

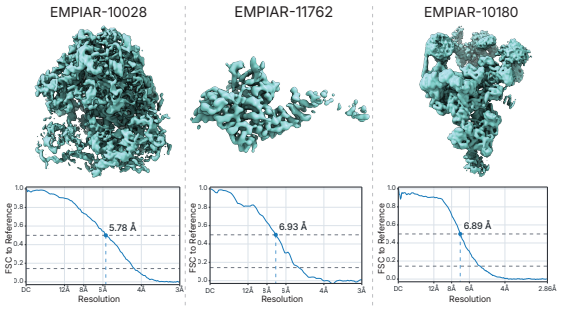

Figure 4: High-resolution *ab initio* reconstruction with cryoSplat2. Frequency marching and pose search were extended to 5 Å, with reconstructed maps and corresponding FSC curves against the reference shown for EMPIAR-10028, EMPIAR-11762, and EMPIAR-10180.

4.3 Optimization Robustness of Voxel and Gaussian Representations

To isolate the effect of the structural representation, we replace the Gaussian representation in cryoSplat2 with a voxel grid while holding the pose-search procedure, frequency-marching scheduler, optimization objective, and all other training settings fixed. We then vary only the batch size on EMPIAR-10028. Because individual cryo-EM particle images have extremely low signal-to-noise ratios, small batches produce noisier stochastic gradients, whereas larger batches improve gradient averaging at increased computational and memory cost.

As shown in Fig. 5, the Gaussian representation remains stable across the entire tested range and successfully reconstructs the global structure even with batch size 1. In contrast, the voxel representation fails to reconstruct a coherent structure at small batch sizes, with batch size 16 near the transition to stable convergence and batch size 32 producing a successful reconstruction. These results indicate that the Gaussian representation is substantially more robust to stochastic gradient noise during joint pose–structure optimization, enabling stable *ab initio* reconstruction with much smaller batches. This robustness is particularly beneficial for pose estimation, where smaller batches reduce the cost of evaluating and differentiating many candidate poses.

4.4 Computational Efficiency

We evaluate the optimized Gaussian projection operator under workloads encountered during an actual cryoSplat2 reconstruction rather than using synthetic Gaussian inputs. Specifically, we select representative states from the EMPIAR-10028 experiment with a target resolution of 5 Å and benchmark each state using its reconstructed Gaussian parameters together with the corresponding projection side length L and number of views Q. These workloads cover substantial variation in the number of Gaussians M, projection resolution, and view count encountered throughout reconstruction. The number of views Q processed by each projection call is determined by the pose-search configuration and memory-aware chunking, and therefore does not vary monotonically over training.

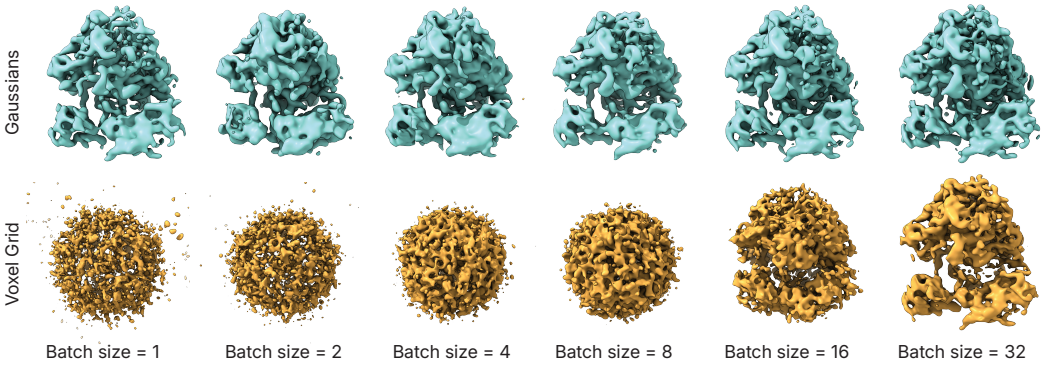

Figure 5: Qualitative comparison of *ab initio* reconstructions using Gaussians (top) and voxel grid (bottom) as representations across batch sizes from 1 to 32.

Table 1: Differentiable Gaussian projection performance on representative workloads from the EMPIAR-10028 5 Å reconstruction. Runtime measures a gradient-tracked forward pass followed by backpropagation through the Gaussian parameters. Values in parentheses are cryoSplat2 / cryoSplat.

M	L	Q	Runtime speedup (ms) ↑	Peak memory reduction (MiB) ↑
512	26	72	**68.8**× (0.645 / 44.410)	**57.5%** (2.53 / 5.96)
512	26	576	**516.0**× (0.688 / 354.714)	**57.5%** (20.16 / 47.41)
512	26	4,608	**756.6**× (3.764 / 2,848.066)	**58.4%** (157.11 / 377.96)
1,120	32	4,608	**655.3**× (7.298 / 4,782.324)	**66.0%** (346.69 / 1,021.11)
9,144	64	1,600	**503.6**× (17.426 / 8,775.963)	**62.1%** (870.09 / 2,294.49)
474,616	194	175	**125.7**× (113.561 / 14,274.132)	**60.2%** (4,982.97 / 12,522.68)
575,235	360	32	**81.4**× (39.433 / 3,210.747)	**61.1%** (1,378.13 / 3,543.95)

Table 1 reports the runtime of the differentiable Gaussian projection step, measured as a gradient-tracked forward pass followed by backpropagation through the Gaussian parameters. Across the representative real reconstruction workloads, cryoSplat2 accelerates this computation by 68.8–$756.6\times$ compared with cryoSplat. The largest gains are observed in workloads processing thousands of candidate views, while high-resolution workloads containing hundreds of thousands of Gaussians still achieve more than $80\times$ speedup. Peak allocated GPU memory is also substantially reduced across all reported workloads. The benchmark isolates the differentiable projection operator and excludes other reconstruction components such as FFTs, CTF modulation, pose scoring, posterior evaluation, optimization, and data movement. These results indicate that efficient differentiable projection is critical for making Gaussian-based *ab initio* reconstruction computationally practical.

5 Conclusion

We present cryoSplat2, a Gaussian-based framework for particle-level cryo-EM *ab initio* reconstruction from raw particle images. It jointly estimates particle poses and reconstructs the 3D density using adaptive Gaussian primitives, combining coarse-to-fine optimization with efficient Gaussian splatting. Experiments on real cryo-EM datasets demonstrate reliable reconstruction across diverse structures, including challenging cases where existing methods fail to converge correctly. Our results establish adaptive Gaussian representations as a practical basis for *ab initio* reconstruction without externally provided poses or an initial 3D reference.

Limitations and future work. The current formulation focuses on recovering low- to intermediate-resolution structures during *ab initio* reconstruction and is not currently optimized for fine-grained heterogeneous cases. Extending the framework to continuous conformational heterogeneity, incorporating structural guidance from atomic or folding models, and further reducing the computational cost of large-scale pose marginalization are promising directions for future work.

AI USE STATEMENT

We used OpenAI GPT-5.6 Sol to assist with language editing, manuscript organization, literature discovery, and selected coding tasks. For software development, the core algorithms, code structure, and overall architecture were designed by the authors. Generative AI was used primarily to assist with code reorganization, localized implementation, refactoring, debugging, and software engineering details such as naming and file organization.

Generative AI was not used to autonomously formulate the research problem, develop the core methodological contributions, design the experiments, interpret the results, or determine the conclusions. The research ideas and contributions arose from the authors' own discussions, derivations, implementation experience, and debugging. All AI-assisted code was reviewed, modified, and extensively tested by the authors before use in the reported experiments. All references, technical claims, experimental results, and AI-assisted text were independently checked by the authors. We take responsibility for the final content of this work.

ETHICS STATEMENT

This work develops computational methods for cryo-electron microscopy reconstruction and does not involve human subjects, personally identifiable information, or sensitive personal data. The experimental datasets are publicly available from established structural biology repositories such as EMPIAR. Our work is intended to improve the accuracy and efficiency of biomolecular structure determination and does not introduce foreseeable ethical risks beyond those associated with standard computational structural biology research. We are not aware of additional ethical concerns raised by this work.

REPRODUCIBILITY STATEMENT

We provide detailed methodological and implementation information throughout the main paper and appendix to support reproducibility. The main text describes the reconstruction objective, frequency marching scheduler, adaptive Gaussian densification, and differentiable Gaussian splatting framework. The appendix provides the cryo-EM image formation model, the complete multi-class probabilistic formulation, pose-search and scheduler details, Gaussian initialization, pruning and moment-preserving splitting, Gaussian splatting implementation details, and additional optimization settings including noise-spectrum estimation and posterior candidate selection. The experimental section and appendix document the datasets, preprocessing procedures, evaluation protocols, and hyperparameters used in our experiments. We will release the source code and configurations used to reproduce the reported results upon publication.

REFERENCES

Alex Barnett, Leslie Greengard, Andras Pataki, and Marina Spivak. Rapid solution of the cryo-em reconstruction problem by frequency marching. *SIAM Journal on Imaging Sciences*, 10(3): 1170–1195, 2017.

Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2):58–76, 2020.

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. URL https://doi.org/10.1038/s41592-021-01220-5.

Muyuan Chen, Bogdan Toader, and Roy Lederman. Integrating molecular models into cryoEM heterogeneity analysis using scalable high-resolution deep Gaussian mixture models. *Journal of Molecular Biology*, 435(9):168014, 2023.

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, 2024.

Suyi Chen and Haibin Ling. CryoSplat: Gaussian splatting for cryo-EM homogeneous reconstruction. In *The Fourteenth International Conference on Learning Representations*, 2026.

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.

Tom Dendooven, Ziguo Zhang, Jing Yang, Stephen H McLaughlin, Johannes Schwab, Sjors HW Scheres, Stanislau Yatskevich, and David Barford. Cryo-EM structure of the complete inner kinetochore of the budding yeast point centromere. *Science advances*, 9(30):eadg7480, 2023.

Dominika Elmlund and Hans Elmlund. SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *Journal of structural biology*, 180(3):420–427, 2012.

Hans Elmlund, Dominika Elmlund, and Samy Bengio. PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure*, 21(8):1299–1306, 2013.

Alex Hanson, Allen Tu, Geng Lin, Vasu Singla, Matthias Zwicker, and Tom Goldstein. Speedy-Splat: Fast 3D gaussian splatting with sparse pixels and sparse primitives. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21537–21546, 2025. doi: 10.1109/CVPR52734.2025.02006.

Andrii Iudin, Paul K Korir, Sriram Somasundharam, Simone Weyand, Cesare Cattavitello, Neli Fonseca, Osman Salih, Gerard J Kleywegt, and Ardan Patwardhan. Empiar: the electron microscopy public image archive. *Nucleic Acids Research*, 51(D1):D1503–D1511, 2023. doi: 10.1093/nar/gkac1062.

Paul Joubert and Michael Habeck. Bayesian inference of initial models in cryo-electron microscopy using pseudo-atoms. *Biophysical journal*, 108(5):1165–1175, 2015.

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.

Axel Levy, Frédéric Poitevin, Julien Martel, Youssef Nashed, Ariana Peck, Nina Miolane, Daniel Ratner, Mike Dunne, and Gordon Wetzstein. CryoAI: Amortized inference of poses for ab initio reconstruction of 3D molecular volumes from real cryo-em images. In *European Conference on Computer Vision*, pp. 540–557. Springer, 2022a.

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, 2022b.

Axel Levy, Rishwanth Raghu, J Ryan Feathers, Michal Grzadkowski, Frédéric Poitevin, Jake D Johnston, Francesca Vallese, Oliver Biggs Clarke, Gordon Wetzstein, and Ellen D Zhong. CryoDRGN-AI: neural ab initio reconstruction of challenging cryo-EM and cryo-ET datasets. *Nature Methods*, 22(7):1486–1494, 2025.

Yilai Li, Yi Zhou, Jing Yuan, Fei Ye, and Quanquan Gu. CryoSTAR: leveraging structural priors and constraints for cryo-EM heterogeneous reconstruction. *Nature Methods*, 21(12):2318–2326, 2024.

Saswat Subhajyoti Mallick, Rahul Goel, Bernhard Kerbl, Markus Steinberger, Francisco Vicente Carrasco, and Fernando De La Torre. Taming 3DGS: High-quality radiance fields with limited resources. In *SIGGRAPH Asia 2024 Conference Papers*, 2024. doi: 10.1145/3680528.3687694.

Youssef S. G. Nashed, Frédéric Poitevin, Harshit Gupta, Geoffrey Woollard, Michael Kagan, Chun Hong Yoon, and Daniel Ratner. CryoPoseNet: End-to-end simultaneous learning of single-particle orientation and 3D map reconstruction from cryo-electron microscopy data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp. 4066–4076, 2021.

Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. Structure of a pre-catalytic spliceosome. *Nature*, 546(7660):617–621, 2017.

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.

Eduardo Sanz-García, Aaron B Stewart, and David M Belnap. The random-model method enables ab initio 3D reconstruction of asymmetric particles and determination of particle symmetry. *Journal of structural biology*, 171(2):216–222, 2010.

Sjors H. W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519 – 530, 2012.

Wolfram Seifert-Davila, Mathias Girbig, Luis Hauptmann, Thomas Hoffmann, Sebastian Eustermann, and Christoph W Müller. Structural insights into human tfiiic promoter recognition. *Science advances*, 9(27):eadh2019, 2023.

Shayan Shekarforoush, David B Lindell, Marcus A Brubaker, and David J Fleet. CryoSPIN: Improving ab-initio cryo-EM reconstruction with semi-amortized pose inference. *Advances in Neural Information Processing Systems*, 2024.

Amit Singer and Fred J Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual review of biomedical data science*, 3(1):163–190, 2020.

Carlos Oscar Sanchez Sorzano, Javier Vargas, José Miguel de la Rosa-Trevin, Joaquin Otón, AL Álvarez-Cabrera, V Abrishami, E Sesmero, Roberto Marabini, and José-María Carazo. A statistical approach to the initial volume problem in single particle analysis by electron microscopy. *Journal of structural biology*, 189(3):213–219, 2015.

Nima Vakili and Michael Habeck. Bayesian random tomography of particle systems. *Frontiers in Molecular Biosciences*, 8:658269, 2021.

Javier Vargas, Ana-Lucia Alvarez-Cabrera, Roberto Marabini, Jose M Carazo, and Carlos Oscar Sánchez Sorzano. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics*, 30(20):2891–2898, 2014.

Lanhui Wang, Amit Singer, and Zaiwen Wen. Orientation determination of cryo-EM images using least unsquared deviations. *SIAM journal on imaging sciences*, 6(4):2450–2483, 2013.

Wilson Wong, Xiao-chen Bai, Alan Brown, Israel S Fernandez, Eric Hanssen, Melanie Condron, Yan Hong Tan, Jake Baum, and Sjors HW Scheres. Cryo-EM structure of the plasmodium falciparum 80s ribosome bound to the anti-protozoan drug emetine. *elife*, 3:e03080, 2014.

Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025.

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, 2021a.

Ellen D Zhong, Adam Lerer, Joseph H Davis, and Bonnie Berger. CryoDRGN2: Ab initio neural reconstruction of 3d protein structures from real cryo-EM images. In *IEEE International Conference on Computer Vision*, pp. 4066–4075, 2021b.

APPENDIX

A CRYO-EM IMAGE FORMATION

Under the weak-phase approximation, a cryo-EM particle image is formed by an orthographic projection of the specimen's 3D electrostatic potential along the electron-beam direction, followed by modulation by the imaging system (Singer & Sigworth, 2020). Let $V : \mathbb{R}^3 \to \mathbb{R}$ denote the continuous electrostatic potential and $Y : \mathbb{R}^2 \to \mathbb{R}$ an observed particle image. For a viewing rotation $\boldsymbol{W} \in SO(3)$ and an in-plane translation $\boldsymbol{t} = [t_x, t_y, 0]^\top$, the image formation model can be written in real space as

$$Y(r_x, r_y) = H(r_x, r_y) * \int_{\mathbb{R}} V(\boldsymbol{W}^\top \boldsymbol{r} + \boldsymbol{t})\, dr_z + \epsilon(r_x, r_y), \tag{7}$$

where $\boldsymbol{r} = [r_x, r_y, r_z]^\top$ denotes the 3D Cartesian coordinate in real space, $H : \mathbb{R}^2 \to \mathbb{R}$ is the point-spread function (PSF) of the imaging system, $*$ denotes convolution, and ϵ denotes additive noise.

By the Fourier slice theorem, the 2D Fourier transform of an orthographic projection corresponds to a central slice through the 3D Fourier transform of the volume. Let $\boldsymbol{f} = [f_x, f_y, 0]^\top$ denote the corresponding 3D coordinate on this Fourier slice. Eq. (7) can equivalently be written as

$$\widehat{Y}(f_x, f_y) = \widehat{H}(f_x, f_y)\, \widehat{V}(\boldsymbol{W}^\top \boldsymbol{f})\, e^{-2\pi i \boldsymbol{f}^\top \boldsymbol{t}} + \widehat{\epsilon}(f_x, f_y), \tag{8}$$

where $\widehat{H}$ is the contrast transfer function (CTF), and $\widehat{V}$ denotes the 3D Fourier transform of V.

Our implementation combines these equivalent real- and Fourier-space formulations. The density V is represented by anisotropic Gaussian primitives. After applying the viewing transformation, each 3D Gaussian is orthographically projected along the z-axis. As in cryoSplat (Chen & Ling, 2026), the line integral of a 3D Gaussian remains Gaussian, producing a 2D splat $\tilde{G}(\tilde{\boldsymbol{r}} \mid \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$, where $\tilde{\boldsymbol{r}} = [r_x, r_y]^\top$ denotes the 2D image-plane coordinate. The projected Gaussians are accumulated on the image plane by differentiable rasterization. We use $X, Y : \mathbb{R}^2 \to \mathbb{R}$ to denote continuous image functions and $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{D \times D}$ to denote their corresponding sampled pixel arrays.

The rasterized projection $\boldsymbol{X}$ is transformed to Fourier space and modulated by the CTF, while in-plane translations of the observed image $\boldsymbol{Y}$ are applied through the Fourier phase factor in Eq. (8). Likelihood evaluation is then performed between the resulting Fourier-domain images.

B GENERAL PROBABILISTIC FORMULATION

Cryo-EM reconstruction is commonly formulated as Bayesian inference over the unknown 3D structure together with latent particle poses and, in heterogeneous settings, structural class assignments (Scheres, 2012; Punjani et al., 2017). Given N particle images $\boldsymbol{Y}_i$, $i = 1, \ldots, N$, and K candidate structures $\{\mathbf{V}_k\}_{k=1}^K$, the general maximum a posteriori objective is

$$\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_K = \underset{\mathbf{V}_1, \ldots, \mathbf{V}_K}{\arg\max} \left[\sum_{i=1}^{N} \log \sum_{k=1}^{K} \pi_k \int p(\boldsymbol{Y}_i \mid \mathbf{V}_k, \boldsymbol{\phi})\, p(\boldsymbol{\phi})\, d\boldsymbol{\phi} + \log p(\mathbf{V}_1, \ldots, \mathbf{V}_K) \right]. \tag{9}$$

Here, π_k denotes the prior probability of structural class k, $\boldsymbol{\phi} = (\boldsymbol{W}, \boldsymbol{t})$ denotes the particle pose, $p(\boldsymbol{\phi})$ is the pose prior, and $p(\mathbf{V}_1, \ldots, \mathbf{V}_K)$ denotes a structural prior over the reconstructed volumes.

In the *ab initio* setting, only simple physical and stability constraints are imposed through the Gaussian parameterization. The structural prior therefore contributes no informative term to the objective considered here.

Neither structural class assignments nor particle poses are known initially in particle-level *ab initio* reconstruction. We therefore assume uniform priors over the current class and pose search spaces, with $\pi_k = 1/K$ and uniform $p(\boldsymbol{\phi})$. Up to constants independent of the reconstructed structures, Eq. (9) reduces to

$$\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_K = \underset{\mathbf{V}_1, \ldots, \mathbf{V}_K}{\arg\max} \sum_{i=1}^{N} \log \left[\frac{1}{K} \sum_{k=1}^{K} \int p(\boldsymbol{Y}_i \mid \mathbf{V}_k, \boldsymbol{\phi})\, d\boldsymbol{\phi} \right]. \tag{10}$$

Eq. (10) generalizes the homogeneous objective in Eq. (1), which is recovered by setting $K = 1$.

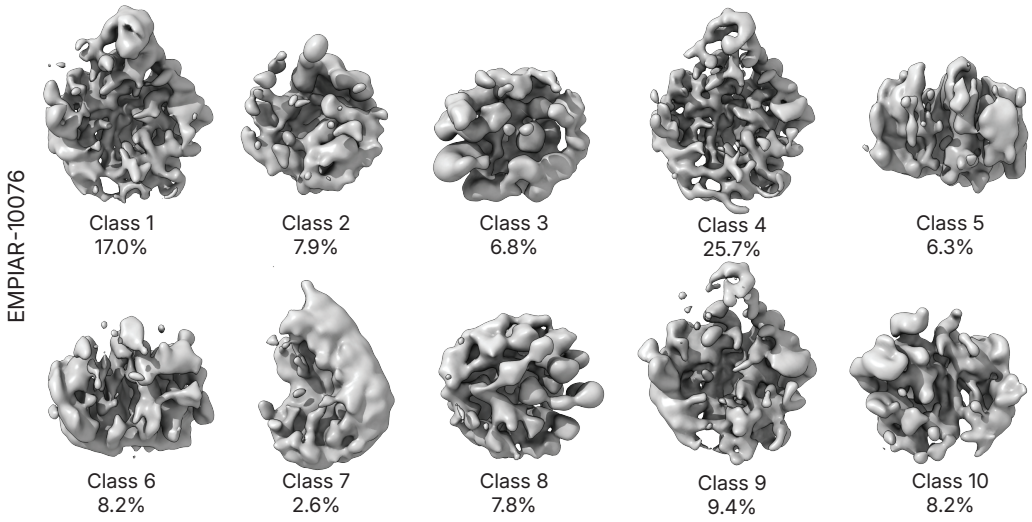

Figure 6: Multi-class *ab initio* reconstruction on EMPIAR-10076 with $K = 10$. Percentages indicate the inferred particle population of each class. The recovered maps exhibit multiple coarse structural states from random initialization without external poses, references, or class labels.

Discrete class–pose marginalization. In practice, the pose integral is approximated over discrete rotation candidates $\boldsymbol{W}_q$, $q = 1, \ldots, Q$, and translation candidates $\boldsymbol{t}_\ell$, $\ell = 1, \ldots, T$. Extending Eq. (3) to multiple structural classes, the residual energy for particle i under class k, rotation q, and translation ℓ is

$$E_{ikq\ell} = \sum_{\boldsymbol{f}} \frac{\left|\widehat{\boldsymbol{X}}_{ikq}(\boldsymbol{f}) - \widehat{\boldsymbol{Y}}_{i\ell}(\boldsymbol{f})\right|^2}{2\sigma^2(\|\boldsymbol{f}\|)}, \tag{11}$$

where $\widehat{\boldsymbol{X}}_{ikq}$ denotes the CTF-modulated projection of structure k under rotation $\boldsymbol{W}_q$, $\widehat{\boldsymbol{Y}}_{i\ell}$ denotes the observed particle after applying translation $\boldsymbol{t}_\ell$ through a Fourier phase shift, and $\sigma^2(\|\boldsymbol{f}\|)$ is the radially averaged noise spectrum.

Under the uniform class and pose priors described above, the posterior probability of each class–pose candidate is

$$p_{ikq\ell} = \frac{\exp(-E_{ikq\ell})}{\sum_{k'=1}^{K}\sum_{q'=1}^{Q}\sum_{\ell'=1}^{T}\exp(-E_{ik'q'\ell'})}. \tag{12}$$

Marginalizing over all KQT candidates gives the negative log marginal likelihood for a batch of B particles,

$$\mathcal{L} = \frac{1}{B}\sum_{i=1}^{B}\left[\log(KQT) - \log\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{\ell=1}^{T}\exp(-E_{ikq\ell})\right]. \tag{13}$$

Eq. (13) reduces to the homogeneous objective in Eq. (5) when $K = 1$.

Multi-class *ab initio* reconstruction. The formulation above naturally extends the homogeneous reconstruction setting to multiple structural classes by jointly marginalizing over class assignments and particle poses. To demonstrate this capability, we perform multi-class *ab initio* reconstruction on EMPIAR-10076 (Davis et al., 2016) using $K = 10$ classes in Fig. 6. All classes are initialized independently from random Gaussian representations and optimized jointly under Eq. (13), without using an external reference or class labels. We use a class-similarity weight of 0.3 and the same 35 Å to 12 Å frequency-marching schedule as in the main experiments.

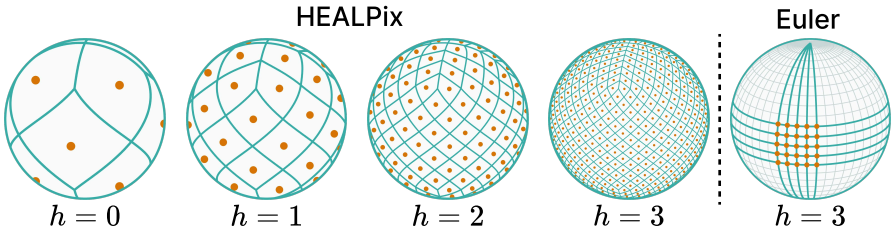

Figure 7: Illustration of global HEALPix sampling at increasing orders h and the local Euler-angle neighborhood used after the global search becomes sufficiently fine.

C Pose search and frequency marching

Spatial and angular resolution matching. The frequency-marching scheduler couples the reconstruction bandwidth to the resolution of the pose search. Let D denote the original image side length and a the physical pixel size. At a Fourier crop side length L, the highest represented spatial frequency corresponds to the nominal real-space resolution

$$r(L) = \frac{2Da}{L}. \tag{14}$$

As L increases, finer spatial frequencies become available and the reconstruction can represent progressively smaller structural features. However, these frequencies are useful only when the pose search is sufficiently fine. If the angular sampling is substantially coarser than the structural resolution, orientation errors blur the newly introduced high-frequency information; conversely, refining the pose grid beyond what can be distinguished by the current low-resolution structure only increases computational cost.

For a particle of physical diameter d, a rotation by a small angle θ produces a displacement of approximately $d\theta/2$ at the particle boundary. Requiring this displacement to remain below the current spatial resolution gives the angular precision

$$\theta_{\mathrm{req}}(L) = \frac{2r(L)}{d}. \tag{15}$$

Following the HEALPix-based coarse-to-fine orientation sampling used in RELION (Scheres, 2012), we parameterize the global rotation grid by the HEALPix order h. We approximate its characteristic angular spacing as

$$\delta(h) = \sqrt{\frac{4\pi}{12 \cdot 4^h}}. \tag{16}$$

The minimum HEALPix order required at bandwidth L is therefore the smallest h satisfying $\delta(h) \leq \theta_{\mathrm{req}}(L)$. This relation provides the central frequency-marching principle used by the scheduler: pose resolution is refined only as needed to support the current reconstruction bandwidth.

Translation search resolution. The translation search is refined according to the same spatial-resolution scale, but is controlled independently from the HEALPix order. Let the translation search span $[-e, e]$ pixels along each image axis and contain n samples per axis. The corresponding translation-grid spacing is

$$\Delta t = \frac{2e}{n}. \tag{17}$$

Because translations act directly in the image plane, their required precision is determined by the current spatial sampling rather than by the angular grid. In particular, the characteristic pixel size associated with the current Fourier crop is D/L pixels on the original image grid. The scheduler therefore updates the translation-search extent as pose estimates stabilize, while the candidate spacing is determined by the current translation-grid configuration. Increasing the HEALPix order alone does not directly refine the translation grid.

Global and local pose search. At coarse resolution, rotations are searched globally using a discretization of $SO(3)$ constructed from HEALPix viewing directions and a uniformly sampled in-plane angle. At HEALPix order h, the number of global rotation candidates is

$$Q_{\mathrm{global}}(h) = 72 \cdot 8^h. \tag{18}$$

The search starts at a low HEALPix order and increases h progressively according to Eq. (15). Once the pose estimates are sufficiently localized, the implementation switches to a local Euler-angle neighborhood centered at the current pose estimate, as illustrated in Fig. 7. This local search is an efficiency optimization: it changes only how rotation and translation candidates are generated, while the likelihood, posterior computation, and marginal objective remain unchanged.

Pose-stability statistics. Frequency marching should advance only after the current pose estimates have stabilized. Rather than comparing poses obtained from two unrelated mini-batches, we evaluate stability on a persistent probe batch. At each scheduler check, the probe particles are re-evaluated using the current structure and search configuration, and the resulting poses are compared with their previously stored estimates.

For particle i, the rotational change between normalized quaternions $\boldsymbol{q}_i^{\mathrm{old}}$ and $\boldsymbol{q}_i^{\mathrm{new}}$ is measured by the geodesic angle

$$\theta_i = 2 \arccos\left(\left|\left\langle \boldsymbol{q}_i^{\mathrm{old}}, \boldsymbol{q}_i^{\mathrm{new}} \right\rangle\right|\right), \tag{19}$$

and the aggregate rotational update is

$$\mathrm{RMS}_{\mathrm{rot}} = \sqrt{\frac{1}{B}\sum_{i=1}^{B} \theta_i^2}. \tag{20}$$

Similarly, the translational update is measured as

$$\mathrm{RMS}_{\mathrm{trans}} = \sqrt{\frac{1}{B}\sum_{i=1}^{B} \left\| \boldsymbol{t}_i^{\mathrm{new}} - \boldsymbol{t}_i^{\mathrm{old}} \right\|_2^2}. \tag{21}$$

The scheduler uses an exponential moving average (EMA) of these quantities to reduce sensitivity to short-term fluctuations.

Posterior confidence. In addition to pose-update magnitudes, the scheduler tracks the concentration of the discrete class–pose posterior. For particle i, we define the posterior confidence as the probability of its most likely joint candidate,

$$c_i = \max_{k,q,\ell} p_{ikq\ell}. \tag{22}$$

The scheduler uses the average confidence over the current evaluation window,

$$\bar{c} = \frac{1}{B}\sum_{i=1}^{B} c_i, \tag{23}$$

as a measure of posterior concentration. High confidence indicates that the probability mass is concentrated around a small number of class–pose candidates, whereas low confidence indicates greater ambiguity. We use $\bar{c}$ to control the aggressiveness of bandwidth progression rather than as a hard convergence criterion.

Scheduler transition criteria. The scheduler distinguishes between refining the angular grid and increasing the reconstruction bandwidth. When the current HEALPix order is still insufficient for the spatial resolution at side length L, the angular grid may be refined once the rotational update becomes small relative to the current angular spacing,

$$\mathrm{EMA}(\mathrm{RMS}_{\mathrm{rot}}) \le c_{\mathrm{rot}}\, \delta(h), \tag{24}$$

where c_{rot} is a tolerance factor.

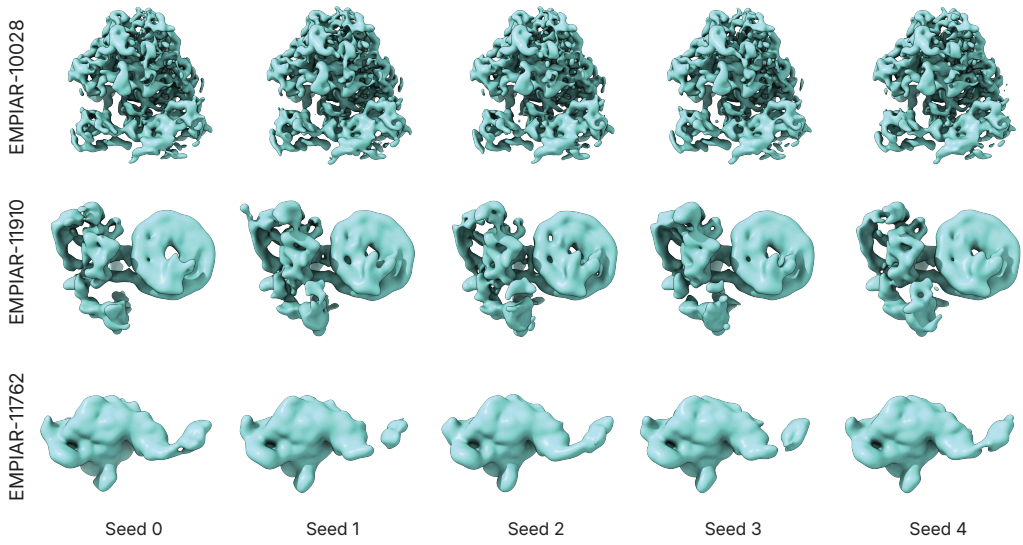

Figure 8: Robustness to random initialization. CryoSplat2 produces visually consistent *ab initio* reconstructions across five random seeds on EMPIAR-10028, EMPIAR-11910, and EMPIAR-11762.

Once $\delta(h) \le \theta_{\mathrm{req}}(L)$, the scheduler instead evaluates whether the current bandwidth has stabilized. We require

$$\mathrm{EMA}(\mathrm{RMS}_{\mathrm{rot}}) \le c_{\mathrm{rot}} \frac{2r(L)}{d}, \qquad \mathrm{EMA}(\mathrm{RMS}_{\mathrm{trans}}) \le c_{\mathrm{trans}} \frac{D}{L}, \tag{25}$$

where c_{trans} is the translation tolerance factor. The latter threshold corresponds to approximately one pixel at the current reconstruction grid spacing. A bandwidth increase is triggered only after these stability conditions are satisfied for a prescribed number of consecutive scheduler checks.

The scheduler therefore follows a fixed priority. If the angular sampling is too coarse for the current bandwidth, the HEALPix order is increased once the angular estimate is stable enough. Otherwise, optimization continues at the current resolution until both rotation and translation estimates stabilize, after which L is increased and the Gaussian representation is densified. Increasing either h or L resets the stability counter. Posterior confidence is additionally used to control the magnitude of the bandwidth increment, allowing more aggressive progression when the pose posterior is concentrated and more conservative updates when ambiguity remains.

D Adaptive Gaussian densification

Gaussian parameterization and initialization. Each Gaussian primitive is parameterized by its center $\boldsymbol{\mu} \in \mathbb{R}^3$, orientation $\boldsymbol{R} \in SO(3)$, principal-axis scales $\boldsymbol{s} = [s_1, s_2, s_3]^\top$, and integrated amplitude A. Its covariance matrix is constructed as

$$\boldsymbol{\Sigma} = \boldsymbol{R}\,\mathrm{diag}(s_1^2, s_2^2, s_3^2)\,\boldsymbol{R}^\top. \tag{26}$$

The centers are randomly initialized within the estimated particle support, while the initial scales are chosen to match the starting reconstruction bandwidth. To prevent primitives from collapsing below the spatial scale supported by the current reconstruction grid, we impose a bandwidth-dependent lower bound on the principal-axis scales. Since each Gaussian is normalized, its amplitude A corresponds to integrated density. We initialize the amplitudes by matching the low-frequency power of the Gaussian representation to that of the observed particles, providing a signal scale consistent with the input data from the beginning of optimization.

Robustness to random initialization. We further evaluate the sensitivity of cryoSplat2 to random initialization by repeating *ab initio* reconstruction with five different random seeds on EMPIAR-10028, EMPIAR-11910, and EMPIAR-11762. As shown in Fig. 8, all runs converge to visually

consistent structures across seeds, despite differences in the initial Gaussian configuration and optimization trajectory. These results indicate that cryoSplat2 is robust to random initialization and does not rely on a particular seed to obtain a successful reconstruction.

Density-guided pruning. Before each densification step, unsupported Gaussian primitives are removed to avoid repeatedly allocating representation capacity to low-density background regions. The current Gaussian representation is first evaluated on a voxel grid to obtain a density map. Because isolated high-density values can distort a relative threshold, we first suppress extreme positive-density outliers when estimating the characteristic signal range. Let μ_+ and σ_+ denote the mean and standard deviation of positive-density voxels. Values above $\mu_+ + 20\sigma_+$ are excluded when computing the reference maximum density $V_{\max}$, and the molecular support is then defined by

$$\mathcal{M} = \{\boldsymbol{r} \mid V(\boldsymbol{r}) \geq 0.1 V_{\max}\}. \tag{27}$$

Gaussian primitives whose centers fall outside $\mathcal{M}$ are pruned before splitting. This criterion concentrates subsequent refinement on regions supported by the current reconstruction while avoiding unnecessary growth of the Gaussian count in the background.

Moment-preserving Gaussian splitting. Consider a parent Gaussian with center $\boldsymbol{\mu}$, integrated amplitude A, and principal-axis scale s along a unit splitting direction $\boldsymbol{u}$. We replace it by n child Gaussians with equal amplitudes $A_i = A/n$ and symmetrically displaced centers

$$\boldsymbol{\mu}_i = \boldsymbol{\mu} + s q_i \boldsymbol{u}, \qquad q_i = \alpha \frac{2i + 1 - n}{n}, \qquad i = 0, \ldots, n-1, \tag{28}$$

where α controls the normalized separation between the children. The offsets are symmetric around the parent center and therefore satisfy

$$\frac{1}{n} \sum_{i=0}^{n-1} q_i = 0. \tag{29}$$

Consequently, the split preserves both the integrated amplitude and the first moment of the parent.

To additionally preserve the second moment along the splitting direction, let s_c denote the common child scale along $\boldsymbol{u}$. The variance of the resulting mixture around the parent mean is

$$s_c^2 + s^2 \frac{1}{n} \sum_{i=0}^{n-1} q_i^2. \tag{30}$$

For the equally spaced offsets in Eq. (28),

$$\frac{1}{n} \sum_{i=0}^{n-1} q_i^2 = \alpha^2 \frac{n^2 - 1}{3n^2}. \tag{31}$$

Matching this mixture variance to the parent variance s^2 gives

$$\frac{s_c}{s} = \sqrt{1 - \alpha^2 \frac{n^2 - 1}{3n^2}}. \tag{32}$$

The remaining principal-axis scales and the Gaussian orientation are left unchanged. The construction therefore preserves the integrated amplitude, mean, and covariance of the parent along the splitting axis, while introducing additional degrees of freedom that can subsequently adapt through optimization.

For the binary case $n = 2$, let $c \in \{1, 2\}$ index the two children and define $q_1 = -\alpha/2$ and $q_2 = \alpha/2$. The child parameters are

$$\boldsymbol{\mu}_c = \boldsymbol{\mu} + s q_c \boldsymbol{u}, \qquad A_c = \frac{A}{2}, \qquad s_c = s\sqrt{1 - \frac{\alpha^2}{4}}. \tag{33}$$

The parameter α controls the trade-off between overlap and separability. With $\alpha = 1$, the children remain strongly overlapping and retain a shape close to the parent. With $\alpha = \sqrt{3}$, the child scale becomes $s_c = s/2$, producing a more separated refinement while still preserving the parent moments.

Bandwidth-aware densification. The Gaussian representation is refined when the reconstruction bandwidth increases. For a Gaussian with principal-axis scale s, the corresponding Fourier-domain scale is $s_f = 1/(2\pi s)$. We require its Fourier-domain extent βs_f to cover the current reconstruction bandwidth $L/2$, which gives the real-space constraint $s \le \beta/(\pi L)$. A Gaussian is therefore considered too coarse when its largest principal-axis scale satisfies $s_{\max} > \beta/(\pi L)$.

Before splitting, density-guided pruning removes unsupported primitives. The remaining Gaussians are recursively split along their largest principal axis until the bandwidth-dependent scale constraint is satisfied or the maximum Gaussian count is reached. In practice, we use the binary moment-preserving construction in Eq. (33) for repeated refinement.

E GAUSSIAN SPLATTING IMPLEMENTATION DETAILS

The optimized projection operator preserves the same cryo-EM image formation model as cryoSplat, but reorganizes the execution around the large number of structure–view candidates required by pose and class marginalization. In the original cryoSplat implementation (Chen & Ling, 2026), different classes and viewing directions are processed largely independently, with substantial host-side looping and repeated rasterization setup. We instead batch the candidate classes and rotations and process their projected Gaussians jointly, reducing per-candidate launch and scheduling overhead.

The backward pass is organized around Gaussian primitives rather than image tiles. Because cryo-EM projection combines Gaussian contributions additively, the gradient of each primitive can be accumulated directly from the pixels influenced by that primitive. Our implementation therefore gathers the relevant pixel contributions for each Gaussian and performs the corresponding reduction locally, rather than scattering gradients from tiles to shared Gaussian parameters through repeated atomic updates. To reduce saved autograd state, the backward pass does not retain the complete forward Gaussian–tile intersection list and instead reconstructs the required interactions during gradient evaluation.

The implementation further incorporates existing GPU rasterization optimizations for Gaussian–tile intersection, reduction, buffering, and elementary-function evaluation. These components follow prior Gaussian rasterization systems and are used as engineering building blocks rather than treated as independent contributions. Together, the resulting operator is tailored to the high-throughput batched projection workload induced by joint class and pose marginalization while leaving the physical forward model and marginal objective unchanged.

F ADDITIONAL IMPLEMENTATION DETAILS

Initialization and optimization. Gaussian centers and scales are initialized as described in Appendix D. The initial total amplitude is calibrated by matching the low-frequency power of randomly generated projections to that of the observed particles. Specifically, if P_{data} and P_{model} denote the corresponding low-frequency powers of the particle images and a unit-amplitude Gaussian representation, respectively, we set $A_0 = \sqrt{P_{\mathrm{data}}/P_{\mathrm{model}}}$ and initialize each of the M Gaussians with amplitude A_0/M. Gaussian parameters are optimized with Adam using a learning rate of 10^{-3} and $(\beta_1, \beta_2) = (0.9, 0.999)$. Geometry updates that change the reconstruction bandwidth or pose-search configuration restart the structure optimizer from its initial learning rate.

Gaussian densification and scale constraints. For adaptive densification, Gaussians whose largest principal-axis scale exceeds the current bandwidth-dependent threshold $4/(\pi L)$ are recursively split along that axis. We use the separable binary split with $\alpha = \sqrt{3}$, halving the child scale along the splitting axis while preserving the parent moments. Gaussian scales are constrained to remain above $1/(2L)$ throughout optimization.

Noise spectrum estimation. Following the resolution-dependent noise model used in RELION (Scheres, 2012), we represent the Fourier-domain noise variance by a radially averaged spectrum $\sigma^2(\rho)$, where $\rho = \|\boldsymbol{f}\|$. For particle i and class–pose candidate (k, q, ℓ), let

$$\Delta_{ikq\ell}(\boldsymbol{f}) = \widehat{\boldsymbol{X}}_{ikq}(\boldsymbol{f}) - \widehat{\boldsymbol{Y}}_{i\ell}(\boldsymbol{f}) \tag{34}$$

denote the Fourier residual. Let $\langle\cdot\rangle_\rho$ denote averaging over Fourier samples at radial frequency ρ. At noise-model update step τ, we accumulate

$$u_\tau(\rho) = \frac{1}{2}\sum_i \sum_{k,q,\ell} p_{ikq\ell} \left\langle |\Delta_{ikq\ell}(\boldsymbol{f})|^2 \right\rangle_\rho, \qquad n_\tau(\rho) = B_\tau, \qquad w_\tau(\rho) = \sum_i \left\langle C_i(\boldsymbol{f})^2 \right\rangle_\rho. \quad (35)$$

The factor $1/2$ accounts for the real and imaginary components of the complex Fourier noise. When posterior candidate selection is used, the retained posterior probabilities are renormalized before these statistics are accumulated.

We maintain exponentially decayed statistics $\bar{u}_\tau = \eta\bar{u}_{\tau-1} + u_\tau$, $\bar{n}_\tau = \eta\bar{n}_{\tau-1} + n_\tau$, and $\bar{w}_\tau = \eta\bar{w}_{\tau-1} + w_\tau$, giving the residual-based variance estimate $v_\tau(\rho) = \bar{u}_\tau(\rho)/\bar{n}_\tau(\rho)$.

Following cryoSPARC (Punjani et al., 2017), this estimate is regularized using a fixed background-noise prior v_0 and a decaying inflated prior v_I:

$$\sigma^2_\tau(\rho) = \frac{\bar{w}_\tau(\rho)v_\tau(\rho) + \lambda_0 v_0 + \lambda_{I,\tau} v_I}{\bar{w}_\tau(\rho) + \lambda_0 + \lambda_{I,\tau}}, \qquad \lambda_{I,\tau} = \lambda_{I,0}\gamma^\tau. \quad (36)$$

We use $\eta = \gamma = 0.9999$, $\lambda_0 = 50$, $\lambda_{I,0} = 2500$, and $v_I = 8v_0$, as in cryoSPARC (Punjani et al., 2017).

Posterior candidate selection. The full pose-search path evaluates the marginal objective over the complete set of class, rotation, and translation candidates. For auxiliary statistics and pose storage, candidates are ranked by posterior probability and the smallest prefix whose cumulative probability exceeds a prescribed threshold is retained. The retained probabilities are then renormalized. This selection is adaptive to posterior concentration rather than based on a fixed top-k rule. In the full gradient path, this candidate selection does not alter the marginal objective; the stored maximum-posterior candidate is used for pose anchoring and scheduler statistics rather than replacing marginalization with hard assignment.

For memory-intensive configurations, the implementation also supports a selected gradient path. In this mode, the complete candidate set is first searched and scored without constructing the volume-gradient graph, after which differentiable projections are recomputed only for the retained candidates. The resulting loss is therefore a truncated approximation to the full marginal likelihood rather than an exactly equivalent reformulation.

Local-search and fixed-pose alternation. After the scheduler enters the local-search regime, each local-search epoch is followed by a fixed-pose epoch. During the local-search epoch, pose candidates are evaluated within a neighborhood around the current estimates while the Gaussian representation is updated jointly. In the subsequent fixed-pose epoch, the stored rotation, translation, and structural class for each particle are held fixed, while the Gaussian representation and noise statistics continue to be updated. The fixed-pose epoch therefore provides structure-only optimization between successive pose-search epochs without introducing an additional pose update.

G Baseline Settings

CryoSPARC. We use cryoSPARC version 4.7.1 and run the "Ab-Initio Reconstruction" job with its default parameters, including the default 35 Å initial and 12 Å target resolution schedule. The resulting volume is subsequently used to initialize a "Homogeneous Refinement" job with default settings for evaluation.

CryoDRGN-AI. We use the official cryoDRGN implementation, version 4.3, in which cryoDRGN-AI is integrated as the `cryodrgn abinit` workflow. Experiments are launched using `cryodrgn abinit` with the default pose-search, network, and optimization settings unless otherwise specified. Following the official preprocessing pipeline, particle images are downsampled to 128×128 before reconstruction.

H Dataset details

We provide detailed statistics and characteristics of the cryo-EM datasets used in our experiments.

- **EMPIAR-10028 (Wong et al., 2014) (Plasmodium falciparum 80S ribosome).** EMPIAR-10028 contains 105,247 particle images with an original box size of 360×360 pixels at 1.34 Å/pixel. The Pf 80S ribosome is a relatively rigid, high-contrast complex and has been widely used as a benchmark for cryo-EM reconstruction. It therefore provides a comparatively well-conditioned setting for evaluating whether an *ab initio* method can reliably recover the correct global structure.
- **EMPIAR-11910 (Dendooven et al., 2023) (CCAN–CENP-A).** EMPIAR-11910 contains 100,000 particles at 0.831 Å/pixel. The dataset contains the monomeric constitutive centromere associated network complex bound to a CENP-A nucleosome (CCAN–CENP-A). The corresponding deposited EMDB map exhibits weak density at intermediate spatial frequencies, and in our tests we were unable to reliably align any of the *ab initio* reconstructions to this reference. We therefore do not report map-to-reference FSC for this dataset.
- **EMPIAR-11762 (Seifert-Davila et al., 2023) (human TFIIIC TauA subcomplex).** EMPIAR-11762 contains 55,079 particle images of size 352×352 pixels at 0.822 Å/pixel. The particles correspond to the TauA subcomplex of human TFIIIC bound to a tRNA gene. This dataset represents a substantially smaller complex than the other datasets, providing a more challenging setting for *ab initio* reconstruction.
- **EMPIAR-10180 (Plaschka et al., 2017) (pre-catalytic spliceosome).** EMPIAR-10180 contains 327,490 particle images with an original box size of 320×320 pixels at 1.69 Å/pixel. This dataset has been widely used as a benchmark for continuous heterogeneity analysis. Unlike the original cryoDRGN analysis (Zhong et al., 2021a), we use the complete particle stack downloaded directly from EMPIAR without filtering. The deposited particles also contain substantial in-plane translations, and previous work found that recentering the particles was necessary for successful reconstruction (Zhong et al., 2021b). Here, we use the deposited particles directly without recentering.
- **EMPIAR-10076 (Davis et al., 2016) (assembling *E. coli* 50S ribosomal subunit).** EMPIAR-10076 contains 131,899 particle images with an original box size of 320×320 pixels at 1.31 Å/pixel. The dataset contains substantial compositional and conformational heterogeneity associated with ribosome assembly, making pose and structure estimation considerably more challenging than for a rigid homogeneous complex.