3D Gaussian Splatting for Volume Reconstruction in Cryo-EM

Nikolaus Dräger
Stanford University
draeger@stanford.edu

Abstract

In the field of cryogenic electron microscopy (cryo-EM), the reconstruction of a three-dimensional model from two-dimensional images represents a pivotal step in molecular structure analysis. This project proposes an innovative approach by employing 3D Gaussian Splatting [1], a technique primarily explored in volumetric rendering, to improve the cryo-EM reconstruction process. Current methodologies in cryo-EM, while effective, face limitations in terms of reconstruction accuracy and computational efficiency. Through the application of 3D Gaussian Splatting, we aim to address these challenges by potentially improving the speed and precision of the reconstruction process. More specifically, this work will introduce and discuss current methods used in 3D reconstruction. We will introduce a prototype for using 3D Gaussian Splatting in cryo-EM and review the potential and future directions of this project.

1 Introduction

In structural biology "*Structure determines Function*" indicates that reasoning and in-depth understanding of the inner workings of complex biological systems requires high-resolution structural maps of the underlying protein complexes.

Cryogenic electron microscopy (cryo-EM) is a popular and modern method of experimentally obtaining 2D projections of particles' electron densities with very high resolutions. The 2D images, called micrographs, can be processed in a preprocessing pipeline to yield 2D images of single particles from different angles. Passing these 2D images of particles to reconstruction algorithms enables the construction of 3D electron density maps with very high resolutions. These maps allow scientists to determine the 3D structure of proteins down to single atoms.

The idea of reconstructing 3D structures from two-dimensional electron micrographs dates back to a 1968 publication by De Rosier et al. [2] in which they exploited the helical symmetry in the tails of icosahedral bacteriophage. This symmetry made it possible to extract 3D information about the complex from few or even a single two-dimensional projection. Frank et al. [3] later came up with methods to perform single particle 3D reconstruction by computing orientations of particles. These methods base on the idea that an image of a large ensemble of complexes contains almost all possible orientations. [4]

Modern reconstruction methods such as cryoDRGN [5, 6] or CryoAI [7] largely use neural networks to represent electrostatic potentials during this reconstruction phase. While a lot of these algorithms show great reconstruction accuracy and manage to overcome challenges such as the extremely low signal-to-noise ratios in cryo-EM, the use of neural networks can be a bottleneck in the computational efficiency of these methods.

3D Gaussian Splatting [1] is a method from the 3D scene reconstruction domain. Compared to neural-based counterparts like Neural Radiance Fields (NeRFs, [8]), the representation of the electrostatic potential as a set of 3D Gaussian functions shows very promising computational performance and

large cut downs in the training time. In this work we want to explore how well the strengths of 3D Gaussian Splatting over its competitors in 3D scene reconstruction are transferable to the problem of cryo-EM reconstruction.

The main contributions of this article are summarized as follows.

1. We provide a comprehensive review of existing, modern reconstruction methods in cryo-EM volume reconstruction as well as in 3D scene reconstruction.
2. We introduce a detailed mathematical and algorithmic framework for porting 3D Gaussian Splatting to the reconstruction task in cryo-EM with a set of constraints to support the development of a proof of concept.
3. We gather and presetn first experimental results using different evaluation methods and highlight possible strengths and limitations to reason about possible future directions of this work.

The rest of this article is organized as follows. Chapter 2 reviews and discusses related work in cryo-EM and 3D scene reconstruction. Chapter 3 introduces the proposed methodology in detail. Chapter 4 discusses the experimental setups and discusses the obtained results. Chapters 5 and 6 conclude the article and discuss further directions of this project, respectively.

2 Related Work

The field of cryo-electron microscopy (cryo-EM) has witnessed remarkable progress in recent years, largely to the application of machine learning and deep learning techniques in 3D reconstruction. In recent studies, these technologies have been demonstrated to enhance the performance and resolution of cryo-EM reconstructions [9, 10].

The CryoGAN reconstruction methods by Gupta et al. [11] is based on an unsupervised deep adversarial model. More specifically, they pair a physics simulator with a discriminator, a deep neural network designed to differentiate between real and synthetic images. To achieve this, the physics simulator computes a projection of an underlying estimated density map and the discriminator evaluates whether an image is expected to be fake or not. In this method, the estimated density map and the discriminator are jointly optimized.

Another notable development is cryoDRGN by Zhong et al. [5, 6], which utilizes Variational Autoencoders (VAEs) [12] for improved molecular structure reconstruction. CryoDRGN's VAE architecture is built upon an image encoder and a volume decoder. During training, the network compresses the two-dimensional images to a lower-dimensional latent variable $z \in \mathbb{R}^n$ from which the decoder then generates a three-dimensional map. Comparing the map to the ground truth structure provides the loss for backpropagation. Once trained, CryoDRGN's decoder network allows the generation of a density map from an arbitrary variable in the latent space.

In a related attempt, Zhong et al. [13] developed Cryofold, a method for cryo-EM recosntruction that employs Gaussian-based techniques. The main idea behind Cryofold is the representation of the density model using radial basis functions (RBFs), where each amino acid is represented by two RBFs - one for the backbone and the second for the sidechain. Consecutive backbone RBFs are physically constrained using a bond term $\mathcal{B}$ while sidechain RBFs are linked to the backbone using a side-chain constraint $\mathcal{C}$. While promising, Cryofold has encountered challenges, particularly with experimental validation and limitations in overcoming local minima when initializing the coordinates of the RBFs. Additionally, the work only considered small proteins and used synthetic data for the evaluation underscoring the need for further innovation in this space.

CryoAI by Levy et al. [7] utilizes an encoder network that outputs a pose of a particle, specified by a rotation and translation, for a given two-dimensional image of a particle. The rotation is applied to a two-dimensional grid of 3D coordinates which is then fed into a neural network representation of the volume. The contrast transfer function (CTF) is applied to the network's output yielding a noise-free estimate of the input image that can be compared to the input to compute the loss. It is important to note that the neural network represents the volume in the Fourier domain, such that the output image after applying the CTF is given in the Fourier space as well. By employing a neural network for the pose estimation, the method overcomes challenges in the computational efficiency faced by other algorithms, including cryoSPARC [14] and RELION [15].

The algorithm behind the RELION software package [15] makes use of a Bayesian approach for computing poses. Using an algorithm called expectation maximization, it computes a distribution over poses for each image making it rather costly. In comparison, cryoSPARC [14] optimizes the map using stochastic gradient descent and uses a branch-and-bound algorithm to quickly limit the space of poses for pose estimation. The branch-and-bound algorithm represents an improvement over RELION in terms of computational efficiency but is still limited by cost. [7]

Alongside this, Mildenhall et al. [8] introduced Neural Radiance Fields (NeRFs), a deep fully-connected neural network approach for general 3D scene reconstruction. Although NeRFs have shown impressive results in image quality, they tend to fall short in terms of training and inference speeds. NeRFs have shown impressive results in image quality, but they fall short in terms of training and inference speeds. This limitation has led to innovative approaches like Plenoxels by Fridovich-Keil et al. [16] and, more relevant to this project, 3D Gaussian Splatting by Kerbl et al. [1]. Both methods abandoned the neural network in favor of different methods for representing the latent space. The latter method leverages anisotropic Gaussians and stands out for its significantly faster inference and training times, making it a strong candidate for cryo-EM applications.

CryoGAN [11], cryoDRGN [5], Cryofold [13], and CryoAI [7] all operate in the Fourier domain using the Fourier slice theorem [17] for fast projection of the volume onto a two dimensional plane. The approach presented in this paper will instead use the image space as its output space, thereby not only improving the interpretability of physical constraints but also allowing for a more intuitive parameterization of certain operations like the interpolation of density maps.

3D Gaussian Splatting shows remarkable results in the performance and quality of reconstructing 3D scense compared to methods using neural representations. In this work, we want to explore how well the strengths of this method can be harnessed for the reconstruction process of the cryo-EM pipeline.

3 Methodology

3.1 3D Gaussian Splatting in Cryo-EM

Image Formation Model We define $V : \mathbb{R}^3 \to \mathbb{R}$ as the electrostatic potential of a protein or molecule. I_i are two-dimensional images and represent orthographic projections of the electrostatic potential onto a 2D image plane. C_i is the point spread function (PSF), the Fourier transform of which we call the contrast transfer function (CTF). The image formation model is then given as:

$$I_i(x, y) = C_i * \int_t V(\Phi_i \begin{pmatrix} x \\ y \\ t \end{pmatrix}) \, dt + \eta_i, \tag{1}$$

where $\Phi_i \in SO(3)$ is a rotation of the coordinates $(x, y, t)^T$ and η_i is a noise term [18]. Intuitively, this equation models the intensity $I_i(x, y)$ at a pixel position x and y by integrating the electrostatic density of a protein along a ray that is orthogonal to the image plain.

3D Gaussian Splatting In the Gaussian mixture model used for 3D Gaussian splatting, the potential at a coordinate $\boldsymbol{r} = (x, y, z)^T$ is modeled as the sum of normal distributions

$$V(\boldsymbol{r}) = \sum_k \mathcal{N}(\boldsymbol{r}, \boldsymbol{\mu}_k, \Sigma_k). \tag{2}$$

In this model, $\boldsymbol{\mu}_k$ defines the mean of a distribution and Σ_k the covariance matrix. We introduce $\theta = \{(\boldsymbol{\mu}_k, \Sigma_k)\}$ as the set of means and covariance matrices defining our potential such that $V \in \{V_\theta, \theta \in \Theta\}$.

The problem then becomes a highly non-convex optimization problem in which we solve for

$$\arg\min_\theta \|I_i - \Gamma(V_\theta, C_i, \boldsymbol{\phi}_i)\|_2^2, \tag{3}$$

where I_i is a ground truth image from the training dataset. Γ models the rendering function constructing a 2D image for a given potential V_θ, together with the point spread function C_i, and a set

of coordinates ϕ_i describing a rotation, according to the image formation model described in Eq. 1. Intuitively, we want to align the rendered or reconstructed images as closely as possible with the actual images from the training dataset. The point spread function will for the first stage of the project be largely ignored to reduce the complexity of the task.

3.2 Implementation

To build a first prototype of the described model, we start by reducing the complexity of the task by making several assumptions.

1. We will, for now, ignore the point spread function C_i.
2. Instead of using real, experimental cryo-EM data for training, we generate an artificial potential from an existing structure model and render a synthetic the training dataset. This allows for a noise free environment for this first proof of concept and grants access to the rotation Φ_i performed when generating a training image.
3. We assume that the angles for the rotation Φ_i of the protein complex are known during the training phase.
4. The number of Gaussians used to model the protein remains fixed during the optimization.
5. The Gaussians will be isotropic, share the same variance σ which will stay fixed during the optimization phase: $\Sigma_i = \sigma \cdot \mathbb{1}$.

Generation of Synthetic Training Dataset The generation of a training dataset starts with the initialization of an electrostatic potential. This is obtained by extracting the positions of the atoms of a protein. These positions are directly mapped to the means $\boldsymbol{\mu}_i$ of the Gaussian potential V. The variance σ is set to a shared, constant value c for all Gaussians.

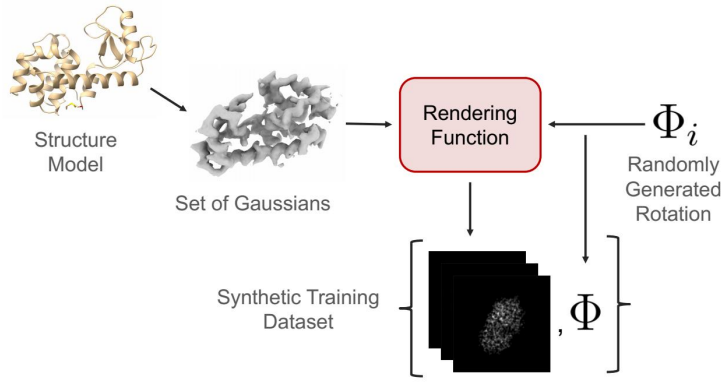


Figure 1: Flowchart highlighting the generation of a synthetic dataset given a known structure model. The model is mapped to a Gaussian potential. The potential and a randomly generated rotation are passed to the rendering function. The output image and the rotation parameters are stored in pairs in the training dataset.

We generate random rotations of the potential Φ_i. The potential and the rotations are passed to the rendering function to generate a 2D image. The output of the rendering function and the corresponding rotation are then stored as a tuple in the synthetic training dataset. Fig. 1 illustrates this algorithm.

Reconstruction of 3D Potential The reconstruction algorithm is based on gradient descent to update the reconstructed potential. It repeatedly samples pairs of images and the corresponding rotations from the previously generated synthetic dataset. The rotation is passed to the rendering function together with the current reconstructed potential. The output 2D image of the rendering function is compared to the ground truth image of the protein for the given rotation. This comparison yields a loss using the L2 norm with an optional regularizer. A gradient descent step updates the means

of the potential's Gaussian functions. This algorithm is illustrated in Fig. 2 and the corresponding pseudocode is listed in Algorithm 1.

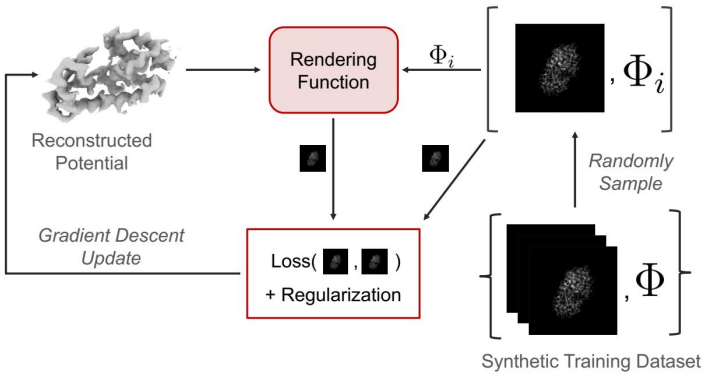


Figure 2: Flowchart illustrating the reconstruction of a 3D potential under the assumption that rotation angles are known. The rotation of a pair of a randomly sampled 2D ground truth image is passed to the rendering function. The loss is computed based on the difference between the rendered image and the ground truth image with the option of adding a regularizer. Gradient descent is used to update the means of the Gaussians of the reconstructed potential.

Algorithm 1 Optimization using noisy ground truth potential

procedure OPTIMIZE($V_0, \{I, \Phi\}, \boldsymbol{\eta}$)
  $V_r \leftarrow V_0 + \boldsymbol{\eta}$
  while not converged **do**
    $I_i, \Phi_i \leftarrow$ SAMPLE($\{I, \Phi\}$)
    $I_r \leftarrow$ RENDER(V_r, Φ_i)
    $L \leftarrow$ LOSS(I_r, I_i) + REG
    $V_r \leftarrow$ GD(∇L)
  end while
end procedure

The first statement of the optimization procedure shown in Algorithm 1 demonstrates the initialization of the reconstructed potential used to obtain first results. Instead of a random initialization, we use the ground truth potential V_0 obtained by mapping the structure model to a set of Gaussian functions. To control the complexity of the reconstruction task, the means are shifted by Gaussian noise $\boldsymbol{\eta}$ of a chosen strength.

4 Preliminary Results

We conducted first experiments using the structure model of T4 Lysozyme (PDB: 107L) consisting of 1451 atoms and image resolutions of 100×100 pixels. We initialize the means of the reconstructed potential by adding Gaussian noise to the means of the ground truth potential as modeled by

$$\boldsymbol{\mu}_0 \leftarrow \boldsymbol{\mu}^* + c \cdot \boldsymbol{\varepsilon}. \tag{4}$$

The ground truth means $\boldsymbol{\mu}^* \in [-1, 1]^3$ are shifted by a Gaussian noise $\boldsymbol{\varepsilon}$ where $\varepsilon_i \sim \mathcal{N}(0, 1)$. $\boldsymbol{\varepsilon}$ is shifted by a constant c to give $\boldsymbol{\eta}$ as shown in Algorithm 1.

We tested a the reconstruction using a regularized and a non-regularized loss. To regularize, we modeled the distance r between consecutive means using a harmonic potential

$$V_H(r) = k \cdot |r - r_0|, \tag{5}$$

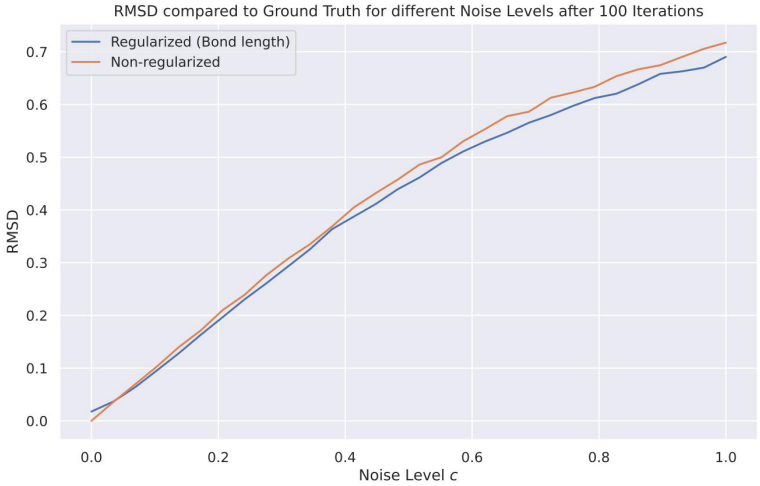


Figure 3: RMSD between reconstruction and ground truth potential after 100 iterations for different noise scalings c for regularized and non-regularized loss.

where k is a scaling factor and r_0 is the equilibrium distance that minimizes the potential. For the experiments, r_0 was chosen to be the median distance between consecutive means in the ground truth potential and $k = 0.1$.

In the first experiment, the reconstruction was run for different noise scaling values c between 0 and 1. The RMSD between the ground truth potential's means and the positions of the reconstructed was computed to give a measure for how well the ground truth potential can be reconstructed from the noisy initialization. Fig. 3 illustrates the results after 100 iterations.

The plot indicates a slightly lower RMSD and thus a better reconstruction when accounting for bond lengths between consecutive means. It also shows an almost linear correlation between the noise level and the RMSD. Thus, even for very small shifts of the means, the reconstruction does not perfectly realign with the ground truth. This behaviour could be influenced by a learning rate that is chosen too high leading to overfitting to single images. The verification of this speculation requires further experiments.

Fig. 4 shows the qualitative results of the reconstructions and demonstrates that the RMSD might not be a sufficient metric to evaluate the quality of the reconstruction. Even though the RMSD of the regularized implementation is almost consistently lower, the density map of the reconstructed potential indicates a low quality reconstruction that seems physically unplausible.

The distribution of the Gaussians when utilizing regularization compared to the non-regularized version is especially surprising considering that the regularization is based on the bond lengths. Further work is required to reason about the cause of this qualitative result.

The qualitative results between the regularized and non-regularized implementations is backed by the computation of the gold-standard Fourier shell coefficient curves (GSFSC) curves [19]. The training dataset is split in half and two separate potentials are reconstruced, one from each half of the dataset. The FSC is a measure for the correlation between the two maps at different frequencies/resolutions and can be employed to assess the resolution of the map. The resolution is chosen to be the point on the curve at which the FSC value falls below 0.5. Fig. 5 shows the resulting GSFSC curves for scaling factors $c = 0.1$ and $c = 1.0$.

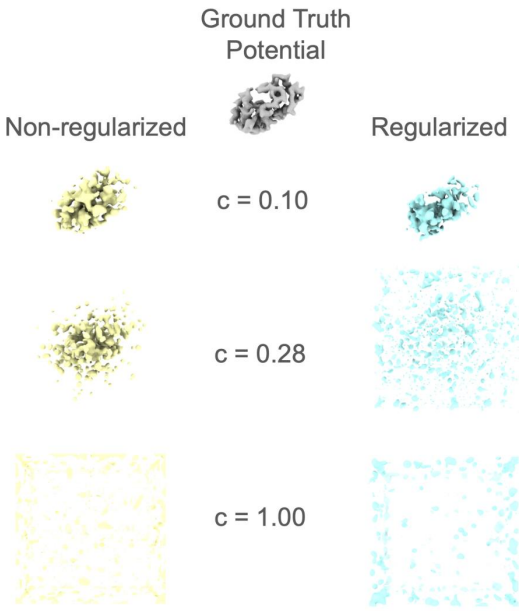
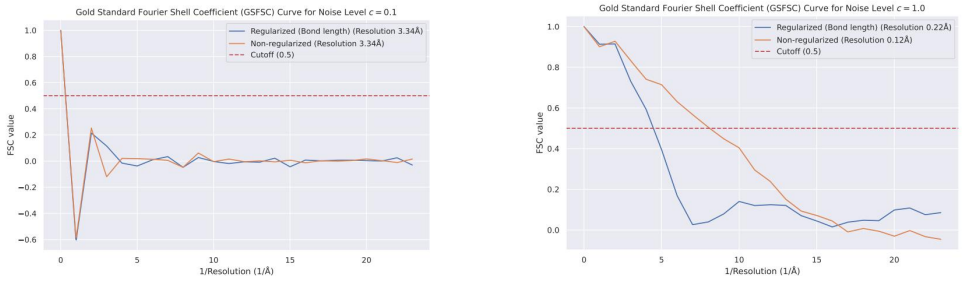


Figure 4: 3D render of the reconstructed density maps for the non-regularized (left) and regularized (right) algorithm for different initial noise scaling factors c compared to the ground truth density map.



(a) $c = 0.1$ (b) $c = 1.0$

Figure 5: GSFSC curves for regularized and non-regularized implementation for two different initial noise scaling values c.

While the curves are very similar for low noise levels, the right plot indicates a higher reconstruction quality in the non-regularized reconstruction. It must be noted that the provided resolution values in angstrom must be interpreted with great caution due to the highly synthetic setup of these experiments and are unlikely to represent plausible values for reconstructions on real, experimental datasets.

5 Conclusion

In this article, we introduce how 3D Gaussian Splatting can be applied to the ab-inition reconstruction process in cryo-EM. Specifically, we discuss and introduce a mathematical and algorithmic framework for a first proof of concept. While most existing methods rely on neural networks and neural representations for modeling the electrostatic potentials of the protein complexes, 3D Gaussian splatting can represent this information using a set of Gaussians with the potential for great computational speedup in the training phase compared to existing methods. We introduce concrete algorithms for the generation of a synthetic dataset and the following training phase in a simplified

environment. We further conduct experiments to analyze the introduced method quantitatively and qualitatively for different intitialization parameters and reconstruction losses. The results presented indicate a promising way to represent electrostatic potentials during the reconstruction phase in cryo-EM. However, further work is required to optimize the algorithm, enable the method to work with experimental data and benchmark 3D Gaussian splatting against existing methods.

6 Future

This work introduced 3D Gaussian splatting in cryo-EM in a simplified environment constrained by several assumptions presented in Chapter 3. While first proof of concept and first results are promising, there is further work to be done in several domains.

Experimental Data In this work, the training data set consisted of noise-free synthetic data generated from a structure model downloaded from the Protein Data Bank (PDB). A necessary next step is the transition from using synthetic data to experimental data. Not only will this introduce noise to the 2D images but it will eliminate any label information on the rotation of the proteins. Thus, a future version of this method will require a way to estimate the rotation Φ_i applied to the imaged protein for a given experimentally obtained image.

Variable Gaussians This work allowed for optimization on the Gaussian's means. In a future work, we want to explore the addition of backpropagation on the covariances, possibly using isotropic Gaussians, and opacity values. We also intend to make the number of Gaussians variable by employing algorithmic techniques like pruning, cloning and splitting as described in [1].

Regularization The regularizer used in this work was based on a coarse approximation of the bond length between atoms and showed unsatisfactory results. In future work we want to explore the reason for this performance and find ways to further tune the physics-based regularization. Using a well-tuned regularizer and with a possible switch from pure gradient descent to more sophisticated algorithms like half-quadratic splitting (HQS) or ADMM [20], we expect better speed convergence and accuracy.

Evaluation In future work, we intend to rely on more robust methods like the GSFSC over the RMSD for evaluation the quality of the reconstruction. The introduction of experimental data will also enable direct comparison in both accuracy and computational performance to existing methods. It will also require a change in the initialization of the reconstructed potential to using either random values or to choosing a different prior.

Challenges in Systems Engineering Major parts of the current implementation can be executed on GPUs. However, further optimizations, like batched processing, have to be implemented to improve efficiency and to overcome memory limitations, especially when using large datasets.

References

[1] Bernhard Kerbl et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". In: *ACM Transactions on Graphics* 42.4 (2023).

[2] D. J. De Rosier and A. Klug. "Reconstruction of Three Dimensional Structures from Electron Micrographs". In: *Nature* 217.5124 (Jan. 1968), pp. 130–134. ISSN: 1476-4687. DOI: 10.1038/217130a0. URL: http://dx.doi.org/10.1038/217130a0.

[3] Joachim Frank, Adriana Verschoor, and Miloslav Boublik. "Computer averaging of electron micrographs of 40 S ribosomal subunits". In: *Science* 214.4527 (1981), pp. 1353–1355.

[4] Edward H Egelman. "The current revolution in cryo-EM". In: *Biophysical journal* 110.5 (2016), pp. 1008–1012.

[5] Ellen D Zhong et al. "CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks". In: *Nature methods* 18.2 (2021), pp. 176–185.

[6] Ellen D Zhong et al. "Reconstructing continuous distributions of 3D protein structure from cryo-EM images". In: *arXiv preprint arXiv:1909.05215* (2019).

[7] Axel Levy et al. “Cryoai: Amortized inference of poses for ab initio reconstruction of 3d molecular volumes from real cryo-em images”. In: *European Conference on Computer Vision*. Springer. 2022, pp. 540–557.

[8] Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis”. In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[9] Jeong Min Chung, Clarissa L Durie, and Jinseok Lee. “Artificial intelligence in cryo-electron microscopy”. In: *Life* 12.8 (2022), p. 1267.

[10] Claire Donnat et al. “Deep generative modeling for volume reconstruction in cryo-electron microscopy”. In: *Journal of structural biology* 214.4 (2022), p. 107920.

[11] Harshit Gupta et al. “CryoGAN: A new reconstruction paradigm for single-particle cryo-EM via deep adversarial learning”. In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 759–774.

[12] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).

[13] Ellen D Zhong et al. “Exploring generative atomic models in cryo-EM reconstruction”. In: *arXiv preprint arXiv:2107.01331* (2021).

[14] Ali Punjani et al. “cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination”. In: *Nature methods* 14.3 (2017), pp. 290–296.

[15] Sjors HW Scheres. “RELION: implementation of a Bayesian approach to cryo-EM structure determination”. In: *Journal of structural biology* 180.3 (2012), pp. 519–530.

[16] Sara Fridovich-Keil et al. “Plenoxels: Radiance fields without neural networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 5501–5510.

[17] Marc Levoy. *Volume rendering using the Fourier projection-slice theorem*. Computer Systems Laboratory, Stanford University, 1992.

[18] Amit Singer. “Mathematics for cryo-electron microscopy”. In: *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific. 2018, pp. 3995–4014.

[19] Grigore Pintilie and Wah Chiu. “Validation, analysis and annotation of cryo-EM structures”. In: *Acta Crystallographica Section D: Structural Biology* 77.9 (2021), pp. 1142–1152.

[20] Stephen Boyd et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers”. In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122.