Missing Wedge Completion in Cryo-Electron Tomography - Dave Van Veen

Dave Van Veen

Introduction

Cryo-electron tomography (cryo-ET) is an increasingly popular modality for imaging cellular structure at high resolution (Fig. 1). A biological sample such as macromolecules or cells are prepared by suspending them in liquid media and subsequently flash freezing at liquid-nitrogen temperatures to maintain their native cellular structure, such as lipid membranes. The three-dimensional sample is then imaged in a transmission electron microscope, which transmits a beam of electrons through the sample to create a two-dimensional projection image captured by a detector. The sample is then tilted by a small amount (typically 1-2 degrees) to capture a second projection from a different angle; this process is repeated until projections have been captured from a range of angles (typically -60 to +60 degrees), resulting in a "tilt series" (Fig. 2). This tilt series of projection images are then reconstructed to create a three-dimensional view of the sample itself, referred to as a tomogram.

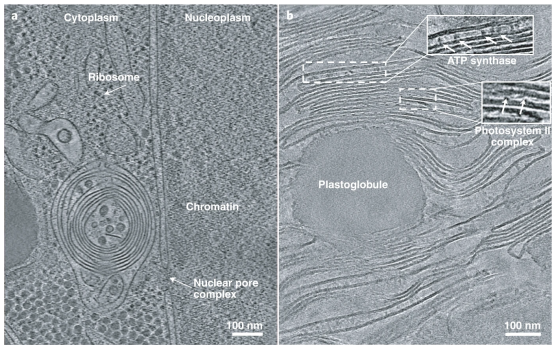


Fig. 1: High-quality cryo-ET samples enable direct observation of fine structural details of protein complexes and membranes inside cells. Figure courtesy of [1].

A significant challenge with this reconstruction is that the tilt series are collected from a limited range of angles. This missing data in the projections corresponds to a "missing wedge" in Fourier space, which results in high-frequency artifacts using (Figure 3, right) the standard filtered back-projection method [2].

The goal of my project is to reconstruct a cryo-ET volume that doesn't contain these artifacts. To do this I rely on unsupervised learning methods, i.e. training a machine learning algorithm without supervision of other image samples. I will demonstrate this on a dataset of simulated spheres, common structures used in proof-of-concept cryo-ET algorithms. I will also demonstrate preliminary results on a dataset of real vaccine particles.

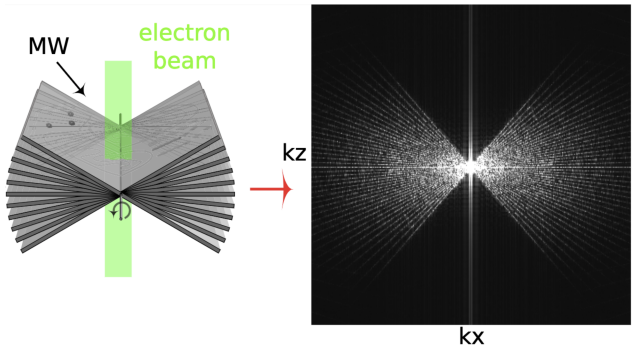


Fig. 2: Left: Cryo-ET images are acquired by transmitting an electron beam through a biological samples at different tilt angles. Right: Because the range of tilt angles is limited, this results in a missing wedge in the (x,z)-plane of frequency space. Figure courtesy of [3].

Background

Cryo-ET is a powerful tool to analyze cellular structure in its native environment. Due to recent technological developments, there is a rapidly growing availability of high-resolution data. The recent explosion of machine learning methods provides significant opportunity for data-driven algorithms, both during the image reconstruction phase and during downstream analysis of the image volume.

As discussed above, the classical method to reconstruct an image volume from its acquired tilt series is filtered back-projection [2], which yields high-frequency artifacts in the reconstruction (Figure 3, right). One recent data-driven approach [4] attempts to solve these missing wedge artifacts by artificially generating artifacts from clean images and training over a large dataset. This formulation is undesirable, however, as it assumes a reasonable simulation, and it is unable to generalize to new data. Hence my goal is to use an unsupervised learning method which is driven by only the acquired projection measurements.

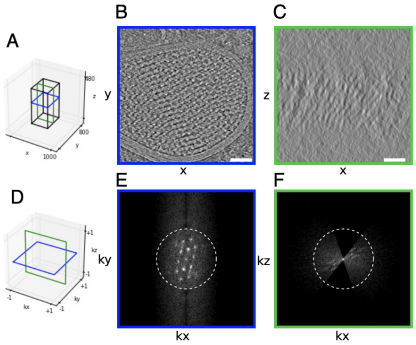


Fig. 3: Demonstration of missing wedge artifacts with filtered-back projection ("fbp"). Image reconstruction (top) and its corresponding frequency spectrum (bottom) show that the artifacts are much more prominent in the (x,z)-plane (right), as cryo-ET images are acquired in a limited-angle tomographic regime along the y-axis. Figure courtesy of [3].

Methods

Image Formation Model

Consider the original image volume $\mathbf{v} \in \mathbb{R}^{x \times y \times z}$ that we would like to reconstruct given the measurements $\mathbf{p} \in \mathbb{R}^{l \times x \times y}$, obtained via the forward model $\mathbf{p} = P\mathbf{v}$, where P is the projection operator. Hence this measurement process provides a tilt series of l two-dimensional projections of size $x \times y$, which we seek to reconstruct into a three-dimensional image. Note (x, y, z) correspond to the three image directions, where the missing wedge artifact will appear in the (x, z)-plane, e.g. see Figure 3.

System Overview

The backbone of this reconstruction relies on a coordinate network [5, 6], a deterministic function $G_\theta : \mathbb{R}^c \to \mathbb{R}$ with trainable parameters θ which maps a set of coordinates $\mathbf{c}$ to a pixel value at that location. Note in our case $c = 3$, as our image volume contains three dimensions. This network is precisely defined by:

$$G_\theta := \mathbf{W}_m(\alpha_{m-1} \circ \alpha_{m-2} \circ \cdots \circ \alpha_0)(\mathbf{c}) + \mathbf{b}_m,$$

$$\mathbf{c}_j \to \alpha_j(\mathbf{c}_j) := \phi(\mathbf{W}_j\mathbf{c}_j + \mathbf{b}_j),$$

where α_j corresponds to the j^{th} network layer composed of weight matrix $\mathbf{W}_j$ and biases $\mathbf{b}_j$ operating on the input $\mathbf{c}_j$, followed by application of a nonlinear function ϕ.

Our goal then is to find a set of parameters θ such that the forward projection operator applied to the network output, i.e. $PG_\theta(\mathbf{C})$ matches the measurements $\mathbf{y}$ we are given, where $\mathbf{C} := \{\mathbf{c}_i\}_{i=1}^{x*y*z}$ is the set of all individual coordinates for our image volume. Hence we train this network in an unsupervised fashion, i.e. only relying on the measurements from one image volume, using the loss

$$\theta^* = \mathbf{argmin}_\theta \|\mathbf{p} - PG_\theta(\mathbf{C})\|.$$

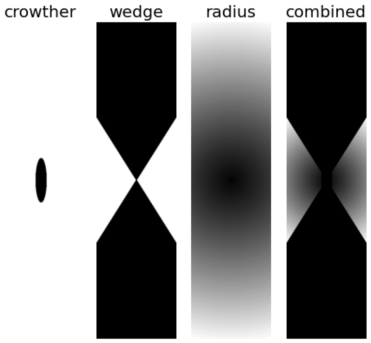


Fig. 4: Optionally we include a regularization term which penalizes the absolute energy of Fourier space which is (1) above the Crowther criterion frequency [10] (far left) and (2) within the missing wedge (middle left), such that higher frequencies are penalized higher at a linear rate (middle right). Cumulatively, this corresponds to a mask applied in Fourier space (far right).

Implementation

I use a five-layer multi-layer perceptron (MLP) for G_θ and train the network to convergence. I demonstrate results on two datasets: (1) a simulated collection of hollow spheres such that $\{l = 60, x = 1024, y = 1024, z = 256\}$, and (2) a real acquisition of vaccine particles such that $\{l = 51, x = 1023, y = 1440, z = 187\}$. Note that ground-truth is not accessible for a real-world biological sample, so the simulated dataset is very useful to evaluate our method using quantitative metrics.

I implemented the projection operator using the odl library [7] and the overall framework in PyTorch due to its data loading and machine learning capabilities. Building this took a significant amount of time, and one major challenge was handling the large size of the image volume (discussed in Section 4.2). This large data size limited the speed of my iteration process in tuning parameters, for which I relied on TensorBoard. However, in Tensorboard I was limited to two-dimensions for assessing reconstruction quality; ideally I would have a seamlessly integrated tool to monitor the reconstruction in 3D, e.g. a ChimeraX plug-in.

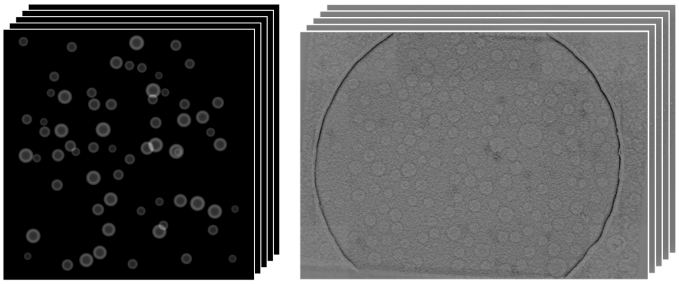


Fig. 5: Tilt series of 2D projection images generated in the cryo-ET measurement process for the simulated spheres dataset (left) and the real vaccine dataset (right).

Memory Management

One major challenge in this project was representing the entire image volume with limited GPU memory. A naive computation of $G_\theta(\mathbf{C})$ would require ~400GB of GPU memory for a single gradient step. While I am fortunate to have access to a NVIDIA Quadro RTX 8000 (48GB), an engineering solution was still essential. One such solution would be to downsample the images until the entire volume can be used; I demonstrate this in Figure 6, requiring 8x downsampling to fit on my GPU. However, this downsampling is very undesirable, as the entire goal of cryo-ET is to capture small detail. Hence instead I created β batches of y-slices and iterated over each batch. This allowed me to model the entire cryo-ET volume at high-resolution at the cost of more training steps. See Figure 6 for an analysis of this tradeoff.

Batch size	128	64	32	16
Memory (GB)	6.3	3.7	2.4	1.8
Time (s)	53	66	95	160
Steps	500	1000	2000	4000

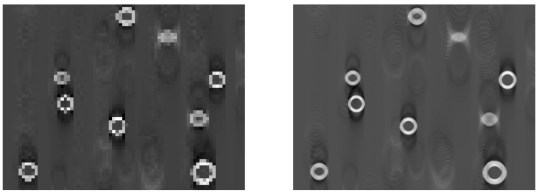


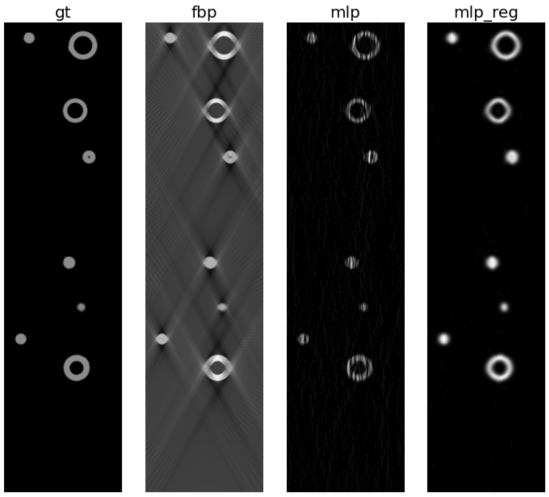
Fig. 6: Left: Tradeoff analysis between memory and runtime with different batch sizes using an 8× downsampled version of the spheres image volume, i.e. 128×128×32. While batch size can be reduced to meet system memory requirements, this results in longer runtime but has negligible effect on reconstruction quality. Right: Random 2D slice of the ground-truth image volume downsampled 8× (left) and full resolution (right). This demonstrates the value of reducing batch size; otherwise system memory limitations would require downsampling the image volume, which is very undesirable.

Results

This section includes experimental results. For simplicity, each figure caption contains detailed observations and discussion. Hence to reduce redundancy, the main text is significantly simplified.

Spheres: Results

Our method provides encouraging qualitative and quantitative results when compared against the fbp. The Crowther criterion regularization term is very effective at improving image quality. However, further work is required to improve the slice-wise consistency of the reconstructions.



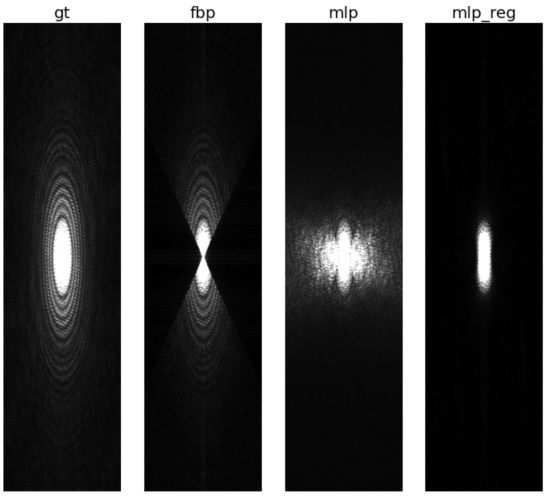


Fig. 7: Reconstruction results (left grouping contains images, right grouping contains the corresponding frequency spectra) for a 2D-slice, comparing ground-truth ("gt" - far left in each) to filtered back-projection ("fbp" - middle left), our multi-layered perceptron result without ("mlp" - middle right) and with ("mlp_reg" - far right) Crowther criterion regularization. Observations: (1) The mlp significantly reduces missing wedge artifacts visible in the fbp; however, this introduces vertical, high-frequency streaking artifacts visible in the frequency spectrum. (2) The regularization term effectively mitigates the vertical streaking artifacts, yielding encouraging results. Note this comes at the expense of some resolution, as noted by the slightly blurry appearance of the spheres output in mlp_reg.

fbp	vif:0.737	ssim:0.318	psnr:24.532
mlp	vif:0.436	ssim:0.667	psnr:23.950
mlp_reg	vif:0.524	ssim:0.843	psnr:29.216

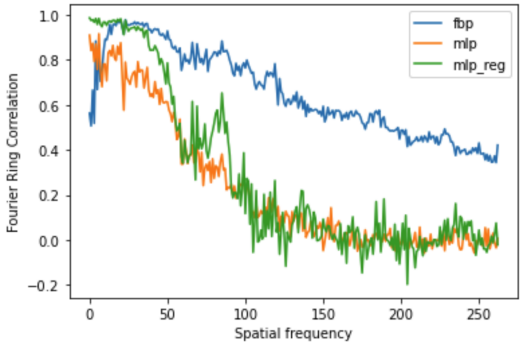


Fig. 8: Quantitative results comparing ground-truth ("gt" - far left) to filtered back-projection ("fbp" - middle left), our multi-layered perceptron result without ("mlp" - middle right) and with ("mlp_reg" - far right) Crowther criterion regularization. The table (top) includes the metrics VIF [8], SSIM, and PSNR. The graph (bottom) contains Fourier Ring Correlation [9], which computes the correlation of the reconstruction's frequency spectra with that of ground-truth. Observations: (1) The regularization term provides significant value, as mlp_reg outperforms mlp in all metrics. (2) Our reconstructions (mlp, mlp_reg) are comparable to or better than fbp in terms of PSNR and SSIM. Meanwhile fbp wins in VIF, a metric which places emphasis on higher frequency components. (3) The VIF insight is supported by the Fourier ring correlation, which shows fbp outperforming mlp_reg at higher spatial frequencies.

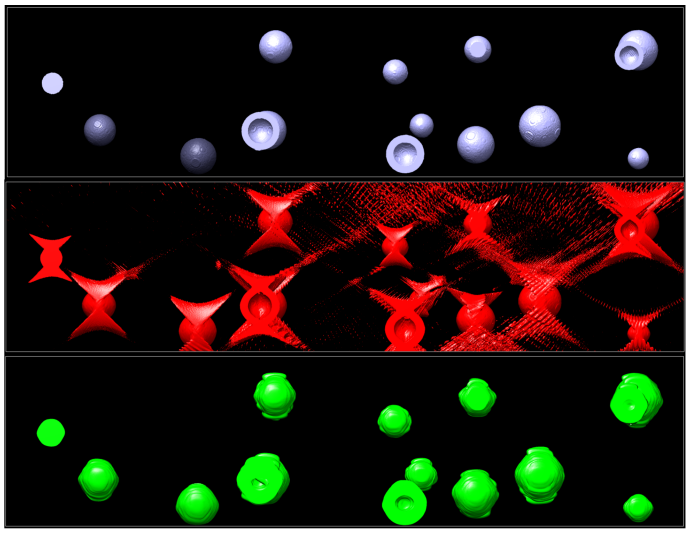
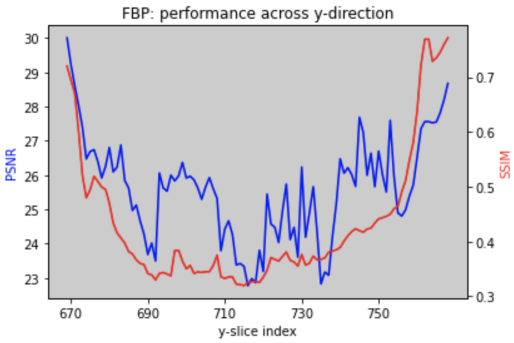


Fig. 9: Qualitative reconstructions for a portion of the 3D-volume displayed in ChimeraX, demonstrating ground-truth (top), filtered back-projection (middle), and our result with Crowther criterion regularization (bottom). Observations: These qualitative results are very promising; however, one area for improvement is to mitigate the "wavy" through-plane artifacts and encourage a smoother sphere surface.

Spheres: Ablation Studies

We analyze quantitative performance of our method on the spheres dataset in two different contexts: across different y-slices and in the presence of noise.



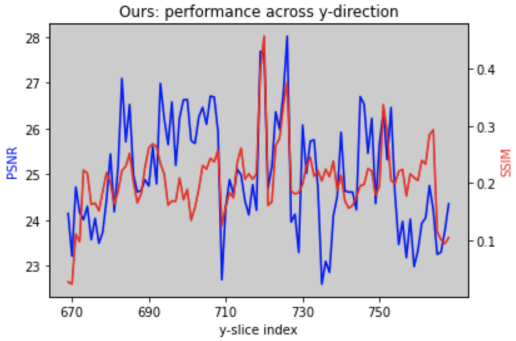


Fig. 10: Ablation study comparing quantitative metrics (PSNR in blue, SSIM in red) across different slices in the y-direction for filtered back-projection ("fbp" - left) and our result with regularization (right). Observations: Both fbp and ours have significant variation across different y-slices. This is congruent with the "wavy" thru-plane artifacts observed in the 3D reconstruction and further motivates mechanisms which encourage thru-plane consistency.

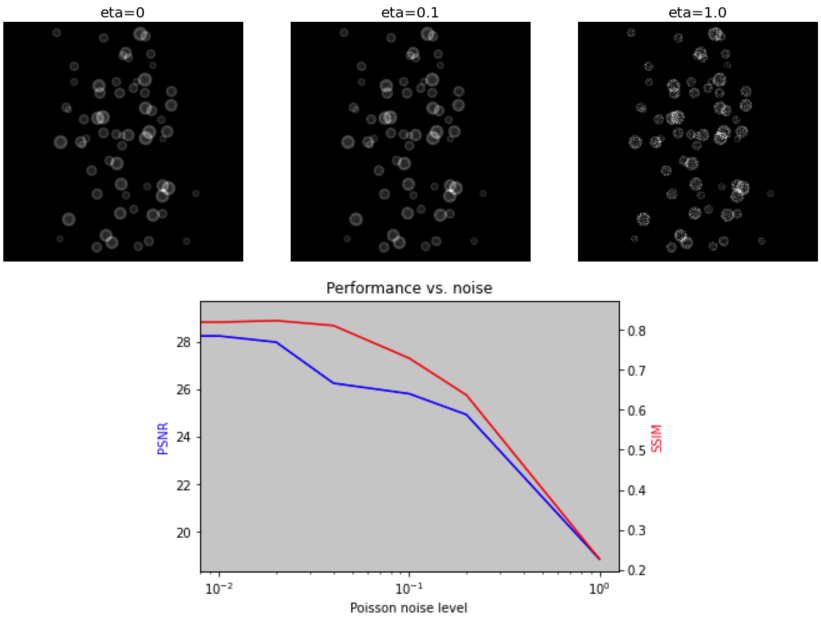


Fig. 11: Ablation study which analyzes reconstruction performance (PSNR in blue, SSIM in red) across different levels of additive Poisson noise (bottom). Note the noise is added to the projections, as demonstrated for varying noise levels (top). Observations: Our method seems fairly robust to low levels of noise but drops off significantly for the highest level, i.e. eta=1. Reconstruction in the presence of noise is crucial for real datasets, so using simulated noise in the sphere dataset will be a useful area for further study.

Vaccine: Initial Results

We provide initial results for the real vaccine dataset, which is much higher noise and lower contrast than the sphere dataset. Further it lacks ground-truth, which inhibits the use of quantitative metrics requiring a reference image.

Fig. 12: Preliminary results with the real vaccine dataset. We show results obtained from two different network initializations ("init" - far left, middle left) and their averaged result ("avg" - middle right) compared to the filtered back-projection baseline ("fbp" - far right). Different rows corresponds to different weighting of the Crowther criterion regularization term: 0 (top), 1e-7 (middle), and 1e-6 (bottom). Observations: (1) Our reconstruction does show similar structure compared to the fbp. (2) Averaging across different network initializations significantly improves image quality. (3) The regularization term likely perhaps provides some benefit at lam_c=1e-7 but is likely overdone for lam_c=1e-6. (4) It is difficult to assess these images with my limited knowledge of structural biology. Improved evaluation methods and expert advice is required.

Conclusion and Future Work

This unsupervised learning method provides encouraging results, particularly on the simulated spheres results. However there are multiple improvements which can be made.

In the spheres dataset, “wavy” through-plane artifacts need to be addressed in order to create a smooth sphere surface. The high-noise, low-contrast vaccine dataset provides additional challenges. I noticed that my current reconstruction is not aligned with the fbp, which leads me to believe the limited-angle range in my code differs from that of the acquisition (appendix, Figure A1). Further the gold particles used for projection stack alignment result in extremely high pixel values relative to the biological sample (appendix, Figure A2). To solve these issues, I must consult my collaborators. Once those are fixed, it's clear that some additional denoising techniques must be employed during the reconstruction process to obtain reasonable images.

While I was able to implement this reconstruction algorithm at full resolution, the runtime is incredibly slow, requiring 5-6 hours to reconstruct the entire volume. This could be prohibitive for labs collecting even just a few samples per day. Hence one improvement would be to model this more efficiently, perhaps leveraging the underlying physics of the measurement process.

Lastly, given the large image volumes and absence of ground-truth for real data, the experimental iterative process is very slow and cumbersome. I would like to build better tools for analyzing experimental output, e.g. a 3D visualization in TensorBoard, so that I can more efficiently evaluate and improve my methods.

Code

This link contains my code, a readme with instructions, and an option to download the raw data.

References

1: Li, Xueming. "Cryo-electron tomography: observing the cell at the atomic level." *Nature Methods* 18.5 (2021): 440-441.

2: Radermacher, Michael. "Weighted back-projection methods." *Electron tomography*. Springer, New York, NY, 2007. 245-273.

3: Croxford, Matthew, et al. "Entropy-regularized deconvolution of cellular cryotransmission electron tomograms." *Proceedings of the National Academy of Sciences* 118.50 (2021): e2108738118.

4: Liu, Yun-Tao, et al. "Isotropic reconstruction of electron tomograms with deep learning." *bioRxiv* (2021).

5: Sitzmann, Vincent, et al. "Implicit neural representations with periodic activation functions." *Advances in Neural Information Processing Systems* 33 (2020): 7462-7473.

6: Tancik, Matthew, et al. "Fourier features let networks learn high frequency functions in low dimensional domains." *Advances in Neural Information Processing Systems* 33 (2020): 7537-7547.

7: Adler, Jonas, Holger Kohr, and Ozan Öktem. "Operator discretization library (odl)." *Zenodo* (2017).

8: Sheikh, Hamid R., and Alan C. Bovik. "Image information and visual quality." *IEEE Transactions on image processing* 15.2 (2006): 430-444.

9: Van Heel, Marin, and Michael Schatz. "Fourier shell correlation threshold criteria." *Journal of structural biology* 151.3 (2005): 250-262.

10: Masenelli-Varlot, Karine, et al. "Wet-STEM tomography: principles, potentialities and limitations." *Microscopy and Microanalysis* 20.2 (2014): 366-375.

Appendix

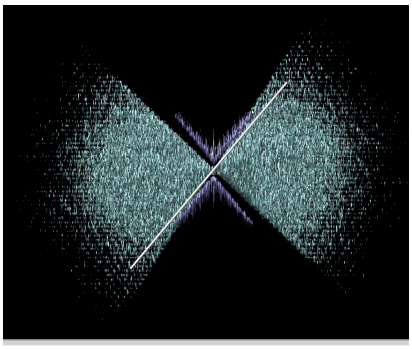


Fig. A1: My current reconstruction output for the vaccine dataset is not aligned with the fbp, which leads me to believe the limited-angle range in my code differs from that of the acquisition.

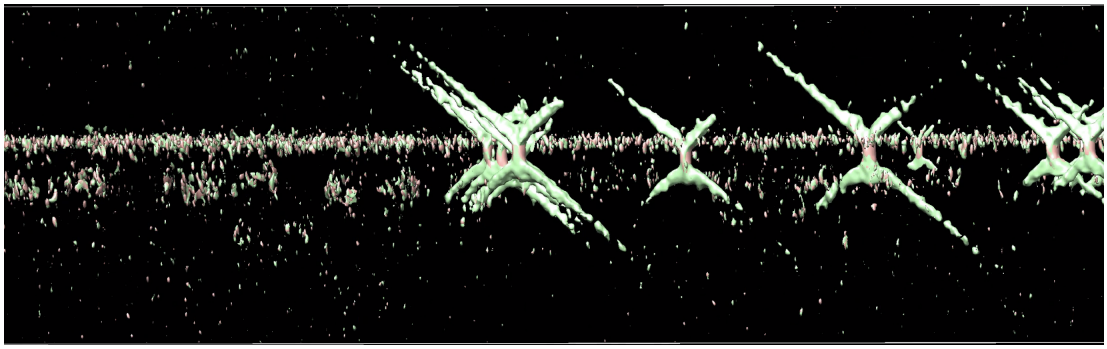


Fig. A2: Gold particles used for projection stack alignment in the vaccine dataset result in extremely high pixel values relative to the biological sample.