Computational Problems from the 2026 3DEM GRC

Marc Aurèle Gilles* Eric Verbeke†

June 26, 2026

The purpose of this note is to record and advertise computational problems that were discussed at the 3DEM GRC of 2026. We give only some context and a rough description of the problems and don't try to be very precise or complete. Feel free to contact us for more details.

First, some notation: images in cryo-EM/cryo-ET are typically assumed to be generated by the following model in Fourier space:

$$y_i = C(c_i)\, T(t_i)\, P(r_i)\, V + \epsilon_i. \tag{1}$$

where $y_i \in \mathbb{C}^{n^2}$ is the observed image, $V \in \mathbb{C}^{n^3}$ is the volume, $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ is Gaussian noise, and $C(c), T(t), P(r)$ should be understood as linear operators or matrices acting on the volume V. $C(c) \in \mathbb{C}^{n^2 \times n^2}$ is the contrast transfer function (CTF), parametrized by c (often c is assumed to be known, but not always), $T(t) \in \mathbb{C}^{n^2 \times n^2}$ is a translation operator parametrized by $t \in \mathbb{R}^2$ and $P(r) \in \mathbb{R}^{n^2 \times n^3}$ is a projection/slicing operator parametrized by $r \in SO(3)$ (a 3D rotation). All of these matrices are very nice: $T(t)$ and $C(c)$ are diagonal and $P(r)$ is very sparse and structured.

Many reconstruction algorithms in cryo-EM try to maximize the likelihood of the data given the volume $\arg\max_V P(y_1, \ldots, y_m | V)$ by marginalizing over the latent variables t, r (and possibly others, such as c or several V's). Thus, we get a problem of the form:

$$\hat{V} = \arg\max_V P(y_1, \ldots, y_m | V) = \arg\max_V \prod_{i=1}^{m} \int_{t,r} \exp\left(-\frac{1}{2\sigma^2} \| C(c_i)\, T(t)\, P(r)\, V - y_i \|_2^2 \right) dt dr \tag{2}$$

1 Two algorithmic breakthroughs

There were two algorithmic breakthroughs that were discussed at the GRC. Their impact is undeniable: both allowed practitioners to solve problems that were either impossible or very difficult and time-consuming. However, exactly why they work so well is not entirely clear to us.

1.1 Ab-initio Refinement [2][1]

For the last decade, there has been a dominant workflow for reconstructing volumes from picked particles (individual images). It is done by maximizing the marginalized likelihood in (2) in two steps. First, we do "ab-initio reconstruction" and second we do "high-resolution refinement".

The first step is about finding an initialization for the optimization algorithm. This is typically done by limiting the reconstruction to a coarse resolution and using some version of stochastic gradient descent (SGD) by sampling random images within the dataset $\{y_i\}_{i=1}^m$. The second step

*Department of Mathematics, Princeton University, Princeton, NJ 08544, United States. gilles@princeton.edu.
†WPI-PRIMe, The University of Osaka, Suita, Osaka, Japan. ericverbeke73@gmail.com.

[1] Also see [3].

is typically done by using Expectation-Maximization (EM) (which considers the full dataset at each step). Even with a good initial model, this is challenging and the method can easily get stuck at local minima if the high frequency components are allowed to dominate. Thus, the EM iteration is run by gradually increasing the resolution of the reconstruction by frequency marching. In both cases, the necessary 5-dimensional integrals over (r, t) for each image are approximated by various clever heuristics and engineering tricks, which has made the method extremely successful in practice.

Nevertheless, this has often failed for small molecules (those with molecular weight around 40 kDa or less, which suffer from lower SNR). This has been a big thorny open problem for practitioners who were simply unable to get any reconstructions for a large class of interesting structures using the conventional pipeline. Oliver Clarke's idea was to run the ab-initio refinement without limiting the resolution, i.e. use SGD (on the marginalized likelihood). And, like magic, it just works! In fact, we have been told that it improves the resolution on all structures, not just small ones, although the effect is the most dramatic for small ones. However, it can be very slow to converge (say, 10 $\times$ slower than the old workflow).

In the past, we have rationalized the old workflow by: SGD is better at escaping local minima, but eventually the very large amount of noise in the gradient (due to the very low SNR of images at high frequencies) makes the variance just too large for them to be useful, so we need proper full EM steps which will "average out" more noise. Clearly, that was wrong! But why does this work? Is it really just about SGD?

1.2 MissAlignment [4]

Our understanding of the cryo-electron tomography (cryo-ET) pipeline is that the alignment is done in two steps[2]. There is a first "tilt-series alignment" where we try to figure out what part of one tilt series image (one very large image, with many thousands of different molecules) corresponds to the same physical location of another tilt series image. Then, there is a second, more fine grained alignment where subimages are cropped out and we have a set of small images which are aligned similarly to (2).

MissAlignment is an algorithm (mostly) for the first step. The alignment is not rigid; between each time an image is taken, the particles move slightly. Previous algorithms seem to have taken reasonable approaches: e.g., first find features in the large images, then map those features across images and align them, or use some local correlation score across images. Unfortunately, these approaches fail sometimes, and the resulting alignment is poor. The idea of MissAlignment is to replace the loss function with a neural network that is trained to predict whether an alignment is good or not. The training is done by starting with an initial alignment (using a traditional method) and then making it "worse" by adding a perturbation to the alignment. Importantly, the training is unsupervised, and done per dataset, so it is not trained on a large dataset of tomograms. Then, that neural network is used as a loss function to gradient-descent to a better alignment. After a few iterations of gradient descent, repeat the process: retrain the neural network on the new alignment and repeat.

Why does this work? The motivation seems to be in part that even an untrained eye can "visually see" whether an alignment is good or not, as a poorly optimized alignment has particular features that are easy to identify. The neural network is trained to mimic this human intuition, and it seems to work well! We're not sure why traditional approaches fail and this one succeeds, but we suspect it is because the more traditional loss function is too non-convex, and possibly because

[2]Neither of us are very familiar with cryo-ET, and there is more than one workflow in use.

they rely on error-prone steps such as some feature detection. This is exciting not only because it works so well, but also because it appears to be a very general approach that could be applied to many problems in cryo-EM and beyond!

2 Open computational problems

2.1 Speeding up template matching

There seems to be a shift away from reconstruction and towards identification. That is, the problem is not: given a set of images y_i, reconstruct the volume V as in eq. (2), but rather: given a single, very large image (of size $\approx 10^4 \times 10^4$) with many molecules (e.g. a single tilt-angle in a tomogram) and a "template" volume V, identify where projections of the template volume are found in the large image. The volume V is obtained from an atomic model from a previously solved structure, or AlphaFold.

This is very similar to particle picking in cryo-EM with one big change: our volume V, because it is obtained from an external atomic model is poor at low resolution, and only has valuable information at high resolution[3]. Hence, the traditional approach of low-pass filtering/frequency marching to reduce the search space over poses seems doomed to fail.

The current approach seems to be to do a brute force search over a fine grid of rotations ($\approx 10^7$) of the volume, and handle the translations by a convolution/Fourier approach. On top of poses, there is a CTF defocus which depends on the z-position of the particle in the cell, so we have to search over a few defocus values (≈ 10) as well. That is, the current approach is to compute the following function for a grid of translations t_j of the size of the image:

$$\mathrm{Corr}(t_j) = \max_{r,d} \mathrm{Re}\,\langle C(d)\, T(t_j)\, P(r)\, V,\, y\rangle \tag{3}$$

and then identify its peaks[4]. This is a huge computational expense, that can only be done by throwing gigantic GPU clusters at it. To give an idea of the scale, the function $G(r,t,d) = \mathrm{Re}\,\langle C(d)\, T(t)\, P(r)\, V,\, y\rangle$ which we need to compute before taking maximums has size $\approx 10^8 \times 10^7 \times 10 = 10^{16}$!

It seems that this is destined to be an increasingly important problem, and there is a lot of room for coming up with interesting algorithms to fix this. Besides the computation, several important issues have yet to be addressed: (1) how to deal with conformational heterogeneity, (2) how to search for and identify unknown molecules, and (3) how to analyze template bias.

2.2 Heterogeneity analysis

Heterogeneity analysis is the problem of reconstructing not one volume V, but a distribution of $V(z)$ (z may be continuous or discrete) in (2). We list a few unresolved problems related to heterogeneity analysis.

- **Aligning the heterogeneity.**
  Most heterogeneity methods separate the problem into two stages. First, each image y_i is assigned a pose (r_i, t_i) and an embedding coordinate z_i. Second, the images, poses, and

[3] Finding a good model to go from atomic models to a volume which is consistent with images even at low resolution is another important open problem.

[4] There are a few complications we are ignoring here: a handful of normalization steps and padding, see [5] for details.

embedding coordinates are used to reconstruct a family of volumes $V(z)$. Methods such as cryoDRGN, RECOVAR, and 3DVA fit the map $z \mapsto V(z)$ without explicitly accounting for the possibility that nearby volumes may be related by non-rigid motions. By contrast, multibody refinement in RELION and local refinement in cryoSPARC impose a more structured model: the molecule is decomposed into a small number of approximately rigid components, and each component is assigned its own pose in each image. This can substantially improve the resolution of some states, and seems to have become a robust and widely used strategy. Hybrid methods such as 3DFlex, Dynamight, and Zernike3D attempt to go beyond rigid-body models, but this requires choosing a deformation model that is expressive enough to capture the relevant motions while remaining low-dimensional enough to be identifiable, often leading to VAE-based parametrizations.

One question that was raised is whether one can instead align the heterogeneous pieces directly. That is, can we use the heterogeneity embedding itself to determine which images, or which local regions of images, should be jointly aligned and averaged in order to improve resolution, without committing in advance to a particular parametrization? That is, one would like a refinement procedure that is local in conformational space: nearby or compatible particles would be aligned in the way that best increases resolution for the corresponding state. RECOVAR does a version of this, figuring out on the fly which images provide "useful" information for parts of the volume, but the alignment is fixed in advance and it seems unlikely that this fixed choice is optimal.

- **Metrics for heterogeneity analysis.** There are now dozens of methods that have been proposed. In practice, practitioners typically try a few on their dataset, which inevitably give different results. There is a lot of interest in metrics that will be able to quantify how much confidence should be given to a particular reconstruction of a particular method. The usual FSC-based scores catastrophically fail at this: they essentially only capture the variance of the estimates whereas the error is often dominated by the bias of the algorithms. Here are some specific open questions: (1) how do we define the resolution of a heterogeneous reconstruction, (2) how can we compare the outputs between different algorithms, and (3) how do we know if a particular feature reconstructed by a method is actually there or is a hallucination?

  A particular idea: for some methods, we can estimate the variance of the embedding estimate z_i as a function of the image SNR. Can we use this to come up with an upper bound on the achievable resolution? Can we use this to "de-bias" the FSC, or come up with an alternative metric?

- **Recovering the conformational landscape.** It is common for heterogeneity analysis method developers to say we want to recover the "conformational landscape". This is not a well-defined quantity. The only somewhat well-defined quantities are the stable states (the volumes corresponding to them, and their occupancy/density), and perhaps the transition paths between those states, although transition probabilities are not directly identifiable from standard cryo-EM data alone. Can we actually recover this information with current methods? How would we measure if we have succeeded?

3 CTF modeling problems

Several modeling problems were recently raised due to advances in hardware and methods.

1. **Tilt-dependent CTF.** The classical CTF formula used in cryo-EM depends on the defocus (essentially the depth position along the optical axis). If the sample is tilted, the defocus varies across the image as a function of position in the tilted plane. It would be valuable to derive an analytical description of this. It will be particularly useful for template matching. This probably only requires some knowledge of Fourier analysis.

2. **Chromatic aberration (Cc) correction.** The chromatic aberration corrector is a piece of hardware that may improve the SNR or dose efficiency of cryo-EM images by making some low-loss inelastically scattered electrons usable for reconstruction. Its use was recently demonstrated in cryo-EM [6]. However, there is currently no good CTF model for this setting that can be used in reconstruction, so deriving one would be very valuable. This probably requires some understanding of diffraction theory.

3. **Laser phase plates.** Laser phase plates have been promising hardware for several years, but only recently have they been successfully applied on cryo-EM and cryo-ET datasets. From a mathematical perspective, they provide a "nicer" CTF with much better low-frequency transfer, rather than simply making the CTF higher. However, there is the current CTF model for the laser phase plate may not be a perfect description of what is seen in practice, and we don't know what effect it will have on reconstructions. Preliminary simulations (made by MAG) indicate it could have a huge effect for heterogeneity analysis!

Acknowledgments

These problems were raised during presentations and discussions at the GRC, and in conversations with many participants, particularly John Briggs (about heterogeneity), Chris Russo (about CTFs), Bronwyn Lucas and Josh Dickerson (about template matching).

References

[1] Amit Singer and Fred J. Sigworth. "Computational Methods for Single-Particle Electron Cryomicroscopy." IEEE Signal Processing Magazine. https://web.math.princeton.edu/~amits/publications/IEEE_Signal_Processing_Magazine.pdf.

[2] "High-resolution ab initio reconstruction enables cryo-EM structure determination of small particles." bioRxiv preprint, 2025. https://www.biorxiv.org/content/10.1101/2025.09.08.674935v1.

[3] CryoSPARC blog. "The Resolution Revolution Is Still Ongoing." May 27, 2026. https://cryosparc.com/blog/high-resolution-ab-initio-reconstruction.

[4] "MissAlignment Teaches Itself Better Cryo-ET Tilt-Series Alignment by Making It Worse." bioRxiv preprint, 2026. https://www.biorxiv.org/content/10.64898/2026.04.29.721716v1.

[5] M. D. Giammar, J. L. Dickerson, L. N. Hall, and B. A. Lucas. "Leopard-EM: an extensible 2D template-matching package to accelerate in situ structural biology." Acta Crystallographica Section D: Structural Biology, 82, 30–42, 2026. https://doi.org/10.1107/S2059798325009982.

[6] J. Wu, C. Liu, A. J. Wang, Y. Z. Gao, L. T. Fu, Z. Liu, J. L. Dickerson, C. J. Russo, and P. Wang. "Chromatic aberration (Cc) corrected cryo-EM: The structure of pseudorabies virus

(PRV) using both zero-loss and energy loss electrons." Ultramicroscopy, 276, 114182, 2025. https://doi.org/10.1016/j.ultramic.2025.114182.