Hyper-Molecules: on the Representation and Recovery of Dynamical Structures, with Application to Flexible Macro-Molecular Structures in Cryo-EM

Roy R. Lederman, [1]‡, Joakim Andén [2] and Amit Singer[3]

[1] The Department of Statistics and Data Science, Yale University, New Haven, CT

E-mail: roy.lederman@yale.edu

[2] Center for Computational Biology, Flatiron Institute, New York, NY

E-mail: janden@flatironinstitute.org

[3] Department of Mathematics and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ

E-mail: amits@math.princeton.edu

Abstract.

Cryo-electron microscopy (cryo-EM), the subject of the 2017 Nobel Prize in Chemistry, is a technology for determining the 3-D structure of macromolecules from many noisy 2-D projections of instances of these macromolecules, whose orientations and positions are unknown. The molecular structures are not rigid objects, but flexible objects involved in dynamical processes; these different conformations are manifested in different instances of the macromolecule observed in a cryo-EM experiment, each instance is recorded as a particle image. One of the great promises of cryo-EM is to map this conformation space. Remarkable progress has been made in determining rigid structures from homogeneous samples of molecules in spite of the unknown orientations, and significant progress has been made in recovering a few distinct states from mixtures of rather distinct conformations, but more complex heterogeneous samples remain a major challenge.

We introduce the "hyper-molecule" framework for modeling structures across different states of heterogeneous molecules, including continuums of states. The key idea behind this framework is representing heterogeneous macromolecules as higher-dimensional objects, with the additional dimensions representing the conformation space. This idea is then refined to model properties such as localized heterogeneity. In addition, we introduce an algorithmic framework for recovering such maps of heterogeneous objects from experimental data using a Bayesian formulation of the problem and Markov chain Monte Carlo (MCMC) algorithms to address the computational challenges in recovering these high dimensional hyper-molecules. We demonstrate these ideas in prototypes applied to synthetic datasets.

Keywords: cryo-EM, continuous heterogeneity, hyper-molecules, hyper-objects, dynamical systems, non-rigid deformations, MCMC

‡ Part of the work was done while at the Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ

Submitted to: *Inverse Problems*

1. Introduction

Cryo-electron microscopy (cryo-EM) is joining X-ray crystallography and nuclear magnetic resonance (NMR) as a technology for recovering high-resolution structures of biological molecules. A typical study produces hundreds of thousands of extremely noisy images of individual particles, where the orientation of each individual particle is unknown, giving rise to a massive computational and statistical challenge. Current algorithms have been successful in recovering remarkably high-resolution structures of static macromolecules in homogeneous samples with little variability, and have also been rather successful in recovering structures from heterogeneous samples consisting of a small number of distinct different structures (referred to as discrete heterogeneity). Even in homogeneous cases, there is ongoing work on improving resolution, and there are several open questions about validating the results and estimating the uncertainty in the solutions.

Structural variations are intrinsic to the function of many macro-molecules. Molecular motors, ion pumps, receptors, ion channels, polymerases, ribosomes and spliceosomes are some of the molecular machines for which conformational fluctuations are essential to function. As just one example, the reaction cycle of the molecular motor kinesin is seen to involve a combination of discrete states (i.e., bound kinesin monomers in different stages of ATP hydrolysis) and also a continuous motion in which one monomer "strides" ahead while it is tethered by a linker to its microtubule-bound companion [1]. As another example, fluctuations in the conformation of ligand-binding domains drive the response of neuronal glutamate receptors [2]. While technologies like X-ray crystallography and NMR measure ensembles of particles, cryo-EM produces images of individual particles, and one of the great promises of cryo-EM is that these noisy images, of individual particles at unknown states viewed from unknown directions, could potentially be compiled into maps of the dynamical processes in which these macromolecules participate. This, in turn, would help uncover the functionality of these molecular machines.

Due to the difficulties in the analysis of heterogeneous samples, researchers attempt to purify homogeneous samples; in doing so they lose information about other states/ conformations. Alternatively, they model the macromolecules observed in heterogeneous samples as a small number of distinct macromolecules (e.g., [3]); this approach overlooks relationships between states (e.g., similarity between different conformations of the molecule), and leads to an impractical number of distinct objects when the variability is complex. Currently, the analysis of heterogeneous macromolecules often misses states, achieves limited resolution, or yields remarkably high-resolution static structures, from which hang "blurry" heterogeneous pieces that cannot be accurately recovered.

In some ways, the heterogeneity problem in cryo-EM is an extreme case of related problems that appear in the analysis of other systems that exhibit some intrinsic variability; for example, the imaging of the body of a patient in computed tomography (CT) while the patient breathes [4] (in this case, the viewing directions are known, and

there are some indications for the state in the breathing cycle).

We introduce a new mathematical framework with a Bayesian formulation for describing and mapping continuous heterogeneity in macromolecules, and an algorithmic approach for computing these heterogeneous structures which addresses some of the computational and statistical challenges. We present preliminary implementations of these frameworks and results. Ultimately, the goal of this line of work is to produce scalable computational tools for mapping complex heterogeneity in macromolecules. One of the goals in this design is to allow the use of a wide range of models and solvers, which would enable the user to encode prior knowledge about the specific macromolecule being studied. For the implementation of these ideas, we envision software for modeling of complex heterogeneous molecules in computer code (or simpler interfaces for common templates) as differentiable components, analogous to deep neural network models defined in popular software for deep learning, such as TensorFlow [5] and PyTorch [6], but adapted to imaging problems, as in Operator Discretization Library (ODL) [7], and adapted to the scale and more general properties of the cryo-EM problem, especially in the case of continuous heterogeneity. These models are ideally analyzed using existing generic tools such as the optimization algorithms in TensorFlow or PyTorch, or the Bayesian algorithms in the versatile Edward [8, 9], to the extent possible. In practice, more specialized components will be required due to the scale and special technical properties of the cryo-EM problem. The prototypes presented in this paper to demonstrate these ideas are more limited in their capabilities and scalability.

The starting point of our discussion is the question what does it mean to recover a heterogeneous macromolecule compared to a homogeneous/rigid macromolecule? We propose that this boils down to the question of representing a heterogeneous macromolecule in all its states; in other words, a "solution" would allow us to view the macromolecule at any state in a user interface that would provide us with "knobs" that we could turn to observe the molecule transition between states through a continuum of states. We recall the representation of molecules as 3-D functions using a linear combination of 3-D basis functions:

$$\mathcal{V}(\boldsymbol{r}) = \sum_k a_k \psi_k(\boldsymbol{r}), \tag{1}$$

with spatial coordinates $\boldsymbol{r}$. The generalization of this representation to describe a heterogeneous macromolecule in all its states, to which we refer as a "hyper-molecule" is as follows. In Section 3.2, we propose a generic generalization of (1). We represent hyper-molecules as a linear combination of higher-dimensional basis functions $\tilde{\psi}_q$:

$$\mathcal{V}(\boldsymbol{r}, \boldsymbol{\tau}) = \sum_q a_q \tilde{\psi}_q(\boldsymbol{r}, \boldsymbol{\tau}), \tag{2}$$

where the new dimensions capture heterogeneity, so that $\boldsymbol{\tau}$ identifies a conformation, or a location in the map of states, and the macromolecule at state/conformation $\boldsymbol{\tau}$ is the 3-D density function obtained by fixing $\boldsymbol{\tau}$ in $\mathcal{V}(\cdot, \boldsymbol{\tau})$. In other words, we generalize the classic problem of "estimating a homogeneous macromolecule" to the problem of "estimating

a heterogeneous hyper-molecule," a single high-dimensional object that encodes all the conformations of the macromolecule together. The (possibly high-dimensional) variable $\boldsymbol{\tau}$ represents the map of states, or the "knobs" which a user would turn in order to transition between states. Furthermore, we argue that hyper-molecules are not merely a way to express the solution of some computation: the representation through a finite set of basis functions serves as a regularizer in the computational problem, much like band-limit assumptions in many inverse problems, including the homogeneous case of cryo-EM. In particular, the high-dimensional basis functions, each supported on multiple states, impose relations between states and define a continuum of states. This property distinguishes between hyper-molecules and a small set of independent macromolecules. This mathematical model of heterogeneous macromolecular structures is accompanied by a Bayesian formulation of recovering hyper-molecules from data, which is a generalization of the Bayesian formulation of cryo-EM, which allows a continuum of states and addresses the relationships between states.

Increasingly complex heterogeneity is formulated using increasingly high-dimensional hyper-molecules. However, in Section 3.4 we find that these hyper-molecules can be "too generic": the natural generalization of traditional algorithms to recover very high-dimensional hyper-objects requires impractically large datasets and computational resources. We address these problems in the remaining subsections of Section 3 and in Section 4.

First, in Section 3.5, we introduce "*composite hyper-molecules*," a generalization of hyper-molecules that capture additional properties of macromolecules often known to scientists or readily identifiable. Specifically, a macromolecule can often be modeled as a sum of M rigid and heterogeneous components $\mathcal{V}^{(m)}$, each with its own state $\boldsymbol{\tau}^m$. The state determines not only the shape of the component, but also its position with respect to the other components through a function denoted by f^m:

$$\mathcal{V}(\boldsymbol{r}, \boldsymbol{\tau}^1, \boldsymbol{\tau}^2, ..., \boldsymbol{\tau}^M) = \sum_{m=1}^{M} \mathcal{V}^{(m)}(f^m(\boldsymbol{r}, \boldsymbol{\tau}^m), \boldsymbol{\tau}^m). \tag{3}$$

In this case, "recovering the heterogeneous macromolecule" means recovering the coefficients that describe each individual component $\mathcal{V}^{(m)}$ of $\mathcal{V}$ and recovering the coefficients that describe the trajectory f^m of each component.

Next, in Section 3.6, we note that the Bayesian formulation of hyper-molecule does not rely on a specific representation of the hyper-molecule and it interacts with the model of the hyper-molecule mainly through the comparison of particle images with the hyper-molecule at certain viewing directions and states, and through priors on the hyper-molecule structure. Therefore, we may replace our proposed hyper-molecules and composite hyper-molecules with an other models, with coefficients $\boldsymbol{\theta}$, and discuss an algorithm which accesses a black-box function $\mathcal{V}[\boldsymbol{\theta}](\boldsymbol{r}, \boldsymbol{\tau})$ and a prior $P(\boldsymbol{\theta})$, and update the coefficients $\boldsymbol{\theta}$, the viewing directions, state variables etc. without explicit knowledge of the detailed in the model of $\mathcal{V}$. This formulation, which separates the model and prior of the hyper-molecule from the algorithm allows users to define elaborate models,

based on prior knowledge, available tools and technical expertise.

The high-dimensional nature of hyper-molecules leads to a computational challenge. Specifically, the main computational challenge in current popular algorithms is that each iteration of the algorithms involves a comparison of each particle image to the current estimate of the molecule as viewed from any possible direction. In hyper-molecules, we add the high-dimensional state variable $\boldsymbol{\tau}$, so that the natural generalization of current algorithms would require comparison of each particle image to each possible molecule (i.e., the hyper-molecule at any possible state) at each possible viewing direction, increasing the computational complexity exponentially with the increase in dimensionality. Furthermore, current algorithms are generally designed to model single molecules or a few distinct molecules/conformation rather than more generic black-box models with custom subtle relations between variables. In Section 4 we propose a framework based on Markov chain Monte Carlo (MCMC) algorithms to address some of the computational complexity. This framework, allows complex, flexible, programmable black-box models and bypasses the need for exhaustive searches in each iteration.

In Section 5, we present two prototypes which implement hyper-molecules and MCMC, and present results from experiments with synthetic data. These prototypes demonstrate the applicability of hyper-molecules, composite hyper-molecules and MCMC to the mapping of continuous heterogeneity. We are currently developing the next versions of these prototypes, which will be more scalable and allow more general models of hyper-molecules.

2. Preliminaries

The purpose of this section is to briefly review some of the technical tools used in this paper. In addition, we present the cryo-EM problem and related work on the problem, and we formulate the mathematical and statistical models which we will generalize in the remainder of the paper.

Table 1. Table of Notation

$\mathcal{V}$	three or higher dimensional function
$\hat{\mathcal{V}}$	the Fourier transform of $\mathcal{V}$ in spacial coordinates
$R\boldsymbol{x}$	the vector $\boldsymbol{x}$ rotated by R
$R\mathcal{V}$	the function $\mathcal{V}$ rotated by R, so that $(R\mathcal{V})(\boldsymbol{x}) = \mathcal{V}(R^{-1}\boldsymbol{x})$
$\boldsymbol{x}$	bold fonts are used to emphasize that a certain variable may be a vector, not just a scalar, when this is not obvious from the context.

2.1. Representation of Functions

A function such as $f : \mathcal{X} \to \mathbb{R}$ can be represented in many ways. In this discussion, we assume a default representation which is a linear combination of a finite set of basis functions ψ_k:

$$\mathcal{V}(\boldsymbol{x}) = \sum_k a_k \psi_k(\boldsymbol{x}). \tag{4}$$

The linear combinations of these functions yield an objects which are regular; the specific type of regularity is determined by the choice of basis functions. Typical examples of such functions would be low-frequency (band-limited) sine and cosine functions, and low-order polynomials. The key properties of these representations that we use are that once the model is formulated (i.e., once the basis functions are chosen), the function $\mathcal{V}$ is completely determined by the coefficients a_k, and that the choice of basis functions imposes constraints or regularizes the function (a sum of low frequency sines cannot yield a higher-frequency sine).

In cryo-EM, the functions are sometimes described, loosely speaking, as "band-limited" and "compactly supported." Often, these functions are defined through samples on 3-D grid, with different interpolations in different implementations. We represent functions with these properties in this work using generalized prolate spheroidal functions (see [10] and Section 5.2), however the particular choice of basis functions is not the main topic of this paper, which applies to various forms of representation.

A linear combination of basis functions is not the only way to represent functions. In particular, a Gaussian mixture model (GMM) has been proposed in [11] for low-resolution representation of molecules in cryo-EM; in this representation, the function is a sum of Gaussian masses; In this case, the coefficients determine the amplitude, centers and covariances of the masses. The discussion in this paper also applies to representations like these, with some modifications.

In Sections 3.5 and 3.6 we extend the discussion to more general forms more explicitly.

Remark 1 (Terminology: "representation") *Our use of the term "representation" in the context of this paper is different from the context in which we use the term in [12]. However, we have not found a better term that would avoid this confusion. In this paper "representation" is a way of expressing a function or a problem, typically an expansion of a function in some basis, whereas in [12] it is a technical representation theory term. These two works are independent; the conceptual relation between the two is the motivation to treat heterogeneity as "just another variable," analogous to the viewing direction variable.*

2.2. Cryo-EM and the Forward Model

The purpose of this section is formulate the standard cryo-EM problem in the homogeneous case. We review the main characteristics of the cryo-EM imaging process

and the forward model briefly, and discuss the Bayesian formulation of the problem of recovering the structure of a macromolecule. One of the ideas in this paper is to introduce a flexible framework where components can be exchanged for others to reflect slightly different models, therefore, we restrict the discussion in this section to the general structure of the formulation and highlight the key difficulties; while it is certainly tempting to delve into the mathematical and numerical properties of the forward operator and the different parameters associated with it, the finer details are beyond the scope of this section. A broader discussion of the imaging model and challenges can be found in many surveys such as [13, 14, 15, 16, 17], and further discussions of a Bayesian framework for cryo-EM (in the context of a maximum a posteriori (MAP) formulation) can be found in [3, 18]. We diverge slightly from the standard numerical representation of the homogeneous case in our use of generalized prolate spheroidal functions as natural basis functions for the problem (see Section 5), but otherwise make use of a rather standard imaging model.

Electron microscopy is an important tool for recovering the 3-D structure of molecules. Of particular interest in the context of this paper is single particle reconstruction (SPR), and, more specifically, cryo-EM, where multiple noisy 2-D projections, ideally of identical particles viewed from different directions, are used in order to recover the 3-D structure.

The following formula is a simplified noiseless imaging model of SPR for obtaining the noiseless particle image $I^{(i)}$ from a function $\mathcal{V}$ (representing the molecule's density or a potential):

$$I^{(i)} = a_i H_i * \int_{\mathbb{R}} \mathcal{V}(R_i^{-1}\boldsymbol{x} + \boldsymbol{s}_i) dx_3. \tag{5}$$

where $\boldsymbol{x} = (x_1, x_2, x_3)^\intercal$, R_i is the rotation that determines the direction from which the molecule is viewed, $\boldsymbol{s}_i$ is the in-plane shift, and H_i is the contrast transfer function (CTF) applied to each particle image, and a_i is a positive real valued contrast (amplitude). The viewing direction R_i and the in-plane shift s_i are typically unknown. The parameters of the CTF are not all known; for simplicity, we will assume in this simplified model that they are known or estimated by other means.

A Fourier transform of both sides of Equation (5) reveals that in the Fourier domain, the Fourier transform of the image $\hat{I}^{(i)}$ is related to the 3-D Fourier transform $\hat{\mathcal{V}}$ of the density by the formula

$$\hat{I}^{(i)}(\omega_1, \omega_2) = a_i \hat{H}_i(\omega_1, \omega_2) S[\mathbf{s}_i](\omega_1, \omega_2) \hat{\mathcal{V}}(R_i^{-1}\boldsymbol{\omega}), \tag{6}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, 0)^\intercal$, S is the shift operator in the Fourier domain (which is a point-wise multiplication in the Fourier domain), and $\hat{H}_i$ is the Fourier transform of the CTF. In other words, in the Fourier domain, this imaging model reduces to an evaluation of the Fourier $\hat{\mathcal{V}}$ in the plane perpendicular to the viewing direction, and to point-wise multiplications to compute the effects of CTF, shift and contrast.

In practice, the particle image $Y^{(i)}$ obtained in experiments is discrete (composed of pixels) and noisy. We will study $Y^{(i)}$ through its *discrete* Fourier transform (as

implemented in FFT) $\hat{Y}^{(i)}$ of $Y^{(i)}$, evaluated at regular grid points $\{(\omega_1(k), \omega_2(k))\}$ in the Fourier domain. First, with a minor abuse of notation, we define the discrete noiseless particle image $\hat{I}^{(i)}[\cdot]$ by sampling $\hat{I}^{(i)}(\cdot)$ at the points $\{(\omega_1(k), \omega_2(k))\}$ in the Fourier domain:

$$\hat{I}^{(i)}[k] = \hat{I}^{(i)}(\omega_1(k), \omega_2(k)). \tag{7}$$

We note that $\hat{I}^{(i)}(\omega_1(k), \omega_2(k)) = \overline{\hat{I}^{(i)}(-\omega_1(k), -\omega_2(k))}$ and $\hat{I}^{(i)}(0,0)$ is real valued, because $I^{(i)}$ is real valued by definition.

For brevity and generality, we absorb the various imaging parameters such as the in-plane shift s_i and contrast a_i (as well as noise and CTF variables where applicable) of each particle image into an imaging variable which we denote by $\boldsymbol{q}_i$. For the purpose of this discussion, we denote the forward model operator by $A(R_i, \boldsymbol{q}_i)$. The noiseless imaging model is then summarized by the formula

$$I^{(i)} = A(R_i, \boldsymbol{q}_i)\mathcal{V}. \tag{8}$$

The map $A(R_i, \boldsymbol{q}_i)$ is typically linear.

Next, we model the noise in a simplified imaging model for $\hat{Y}^{(i)}$:

$$\hat{Y}^{(i)}[k] = \hat{I}^{(i)}[k] + \sigma_k \eta_{i,k} = \left(A(R_i, \boldsymbol{q}_i)\mathcal{V}\right)[k] + \sigma_k \eta_{i,k}, \tag{9}$$

where $Re(\eta_{i,k}) \sim N(0, 1/2)$ and $Im(\eta_{i,k}) \sim N(0, 1/2)$ are i.i.d, except for $\eta_{i,k} = \overline{\eta_{i,k'}}$ if $(\omega_1(k), \omega_2(k)) = (-\omega_1(k'), -\omega_2(k'))$ since the noisy image is real valued in the spatial domain. The sample at $\omega = 0$ has no imaginary component for the same reason. The noise variance σ_k depends on the frequency; in this simplified model, we assume that the noise variance is similar for all images and is known; in practice it can be one of the model variables.

These simplified models neglect several aspects of the physical model, numerical computation, and experiment setup. For example, in practice, the images of individual particles must first be extracted from a larger image (micrograph). As we noted above, the parameters determining the CTF and noise profile are sometimes added to the model. To allow a more general formulation, we add the variable $\boldsymbol{\mu}$ which encodes latent variable of the experiment that are not particle specific (e.g., noise σ_k).

Given this model, the likelihood $p(Y^{(i)}|R_i, \boldsymbol{q}_i, \mathcal{V})$ of a particle image $Y^{(i)}$ given the object $\mathcal{V}$ and particle specific-variables R_i and $\boldsymbol{q}_i$:

$$P(Y^{(i)}|R_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}) \propto exp\left(\sum_k \frac{\left|\widehat{Y^{(i)}}[k] - \widehat{(A(R_i, \boldsymbol{q}_i)\mathcal{V})}[k]\right|^2}{2\sigma_{i,k}^2}\right). \tag{10}$$

This leads to a Bayesian description of the problem, with a probability density for an object, image parameters and observed images given by:

$$P\left(\{Y^{(i)}, R_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V}\right) = P\left(\{R_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V}\right) \prod_i P(Y^{(i)}|R_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V})), \tag{11}$$

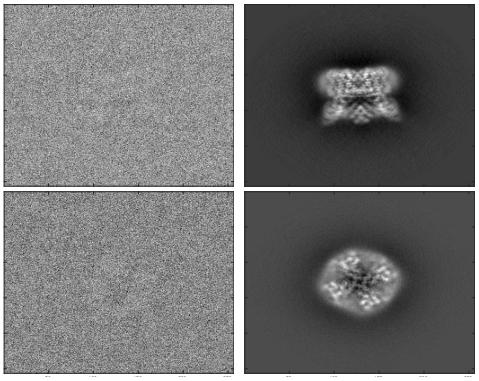

Figure 1. Left: two raw experimental images of TRPV1, available via EMDB 5778 [19]. Right: computed projections of TRPV1 which are the close to the particle images on their left.

where $P\left(\{R_i, \boldsymbol{q}_i\}_i, \mathcal{V}\right)$ is a prior for the molecule and the particle-specific variables such as the viewing direction. The posterior distribution of the variables given the data is therefore proportional to the right hand side of this equation:

$$P\left(\{R_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V} | \{Y^{(i)}\}_i\right) \propto P\left(\{R_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V}\right) \prod_i P(Y^{(i)} | R_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V})). \quad (12)$$

The variables $\{R_i, \boldsymbol{q}_i\}_i$ are particle image specific latent variables, while the object itself, represented by $\mathcal{V}$, is the variable of interest. In other words, the distribution that we are interested in is

$$P\left(\mathcal{V} | \{Y^{(i)}\}\right) = \int P\left(\{R_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V} | \{Y^{(i)}\}\right) dR_1 dR_2 ... dR_n d\boldsymbol{q}_1 d\boldsymbol{q}_2 ... d\boldsymbol{q}_n d\boldsymbol{\mu} \quad (13)$$

Often, we would use a simpler model which assumes a uniform prior for the viewing directions, and independent particles variables; we obtain the posterior

$$P\left(\{R_i, \boldsymbol{q}_i\}_i, \mathcal{V} | \{Y^{(i)}\}\right) \propto P(\mathcal{V}) P(\boldsymbol{\mu}) \prod_i P(Y^{(i)} | R_i, \boldsymbol{q}_i, \boldsymbol{\mu}\mathcal{V})) P(\boldsymbol{q}_i) \quad (14)$$

where $P(\mathcal{V})$ is a prior for molecules (e.g., weighted norms of coefficients representing the molecule), and $P(\boldsymbol{q}_i)$ is a prior for the random variables controlling each individual images, such as in-plane shifts.

While this general framework is sufficient for the purpose of this paper, we note that in the very influential work in [3, 18], a Bayesian framework was used to formulate the problem of recovering a molecule $\mathcal{V}$ as a maximum a posteriori estimation problem, implemented using an expectation maximization algorithm. We choose a slightly different formulation and different algorithms for our purpose due to several technical and computational considerations discussed below. Different algorithms use slightly different models and may absorb different components of the model into different latent variables.

2.3. Heterogeneity in Cryo-EM

The description of the cryo-EM problem in Section 2.2 assumes that all the particles in all the cropped images are identical (but viewed from different directions). However, particles in a sample are often not identical. In some cases, several different types of macromolecules or different conformations of the same macromolecule are mixed together, and sometimes there is be some flexibility in the structure of the macromolecule, which is manifested as a continuum of slightly different versions of the molecule. The first case of distinct classes of macromolecules is called *discrete heterogeneity* and the second case is called *continuous heterogeneity*. In this paper we focus on continuous heterogeneity, although much of the discussion applied to discrete heterogeneity with small modifications.

The models in Section 2.2 cover only a single object, not the heterogeneous case. Therefore, the mathematical formulation needs to be extended to the heterogeneous case, which is the primary goal of this paper.

2.4. Existing Methods in Cryo-EM and Related Work

Many of the existing algorithms for cryo-EM try to estimate the maximum-likelihood or the MAP molecule $\mathcal{V}$ from models formulated roughly like the model in Section 2.2 (see, for example, [20, 21, 3]). One of the popular methods for this is a family of expectation-maximization algorithms, implemented in software such as RELION [18, 22, 23]. Another is based on Stochastic Gradient Descent (SGD), implemented in cryoSPARC [24]. These algorithms alternate between updating an estimate (or, more precisely, estimate a distribution) of the viewing direction for each particle image given the current estimate of the macromolecule and updating the estimate of the macromolecule given the estimated viewing direction for each particle image (or its distribution). In these updates, the algorithm must compare each particle image to the estimated macromolecule as viewed from each viewing direction (discretized), at each value of the other variables (most notably, the in-plane shifts). Naturally, this comparison is expensive. In recent years, several algorithms have been very successful in solving the homogeneous case (no heterogeneity). Clever algorithms and efficient use of hardware components such as GPUs have made the recent implementation of these algorithms rather fast [18, 22, 25, 24]. Other approaches to the cryo-EM problem rely on similarity measure between images to align the images before estimating the structure of the molecule [26, 27, 28, 29]. A MCMC algorithm, using Gibbs Sampling, has been proposed for coarse modeling in the homogeneous case using a Gaussian mixture model [30].

In addition to homogeneous reconstruction, many of the methods mentioned above also accommodate discrete heterogeneity through a 3-D classification framework. Here, each projection image is assigned to a separate 3-D reconstruction by maximizing a similarity measure. While this approach has led to impressive results, it requires significant human intervention in a process of successive refinement of the datasets to

achieve a more homogeneous sample, and structures that are not well-represented in the data tend to be lost [14].

A few approaches have emerged to treat the continuous heterogeneity problem. The method proposed in [31, 32, 33] first groups images by viewing direction, then attempts to learn the manifold formed by the set of images for each of those directions. Following this, the various direction-specific manifolds are registered with one another, and a global manifold is obtained. A 3-D model may then be constructed for each point on that manifold, providing the user with a description of the continuous varying structure. This method requires a consistent assignment of viewing directions across all states, and relies on a delicate metric for comparing noisy images to which different filters have been applied. The method assumes that certain properties of the manifold are conserved across the different viewing directions and requires a successful and globally consistent registration of the manifolds observed in different directions, which is not always possible. Furthermore, complex heterogeneity with more degrees of freedom results in manifolds that are intrinsically high-dimensional; such high-dimensional manifolds are difficult to estimate without exponential increase in the number of samples, and become more difficult to align. This method has been demonstrated in the mapping of the continuous heterogeneity of the ribosome.

More recently, the RELION framework has been extended to include multi-body refinement [34] (also see [35, 36, 37, 37, 38, 39]). In this approach, the user selects different 3-D models that are to be refined separately from the main, or consensus, model. Each separate sub-model is then refined separately, with its own viewing direction and translation, allowing it to move with respect to the consensus model in a rigid-body fashion. This method is limited to rigid-body variability in a few sub-volumes, and cannot handle non-rigid deformations or other types of variability. In particular, the structure found at the interface between the sub-models is likely to vary as their relative positions vary, and it is therefore lost in this method.

Finally, the covariance estimation approach proposed in [40] does not rely on a particular model for heterogeneity, be it discrete or continuous. Indeed, the authors present a method for characterizing continuous variability in synthetic data. However, the covariance approach is adapted to a linear model of variability and is therefore not well-suited for continuous, and necessarily non-linear, variability. Furthermore, the limited resolution of the reconstruction precludes the study of heterogeneity at higher level of detail.

2.5. Markov Chain Monte Carlo (MCMC)

MCMC is a collection of methods which have been used in statistical computing for decades. The full extent of these methods is beyond the scope of this paper. The purpose of this section is to briefly mention a few properties of some MCMC methods that will be useful in our discussion, while inevitably omitting some technical details. A review of MCMC can be found in many textbooks, such as [41].

MCMC algorithms are designed to sample from a probability distribution by constructing a Markov chain (i.e., a model of transitions between states at certain probabilities), such that the desired distribution is the equilibrium distribution of the Markov chain. Often, like in this paper, the desired probability from which we wish to sample is the posterior distribution $P(\boldsymbol{X}|\boldsymbol{Y})$ of a variable $\boldsymbol{X}$, given a statistical model and data $\boldsymbol{Y}$. Very often, we have access only to an unnormalized density $h(\boldsymbol{X}|\boldsymbol{Y}) \propto P(\boldsymbol{X}|\boldsymbol{Y})$, so that we can compute the ratio $h(\boldsymbol{X}|\boldsymbol{Y})/h(\tilde{\boldsymbol{X}}|\boldsymbol{Y})$ between densities at two states $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, but not $P(\boldsymbol{X}|\boldsymbol{Y})$ and $P(\tilde{\boldsymbol{X}}|\boldsymbol{Y})$ directly.

The *Metropolis-Hastings (MH)* algorithm, which is the basis for many MCMC algorithms, is based on the following Metropolis-Hastings Update:

- Given the state $\boldsymbol{X}^{(n)}$ at step n, propose a new state $\tilde{\boldsymbol{X}}^{(n+1)}$ with conditional probability given the current state $\boldsymbol{X}^{(n)}$. The probability of proposing $\tilde{\boldsymbol{X}}^{(n+1)}$ given the current state $\boldsymbol{X}^{(n)}$ is denoted by $q(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n+1)}|\boldsymbol{Y})$. MH can be implemented in different ways, with different methods for proposing a new state, each method has a different function q associated with it.
- Compute the *Hastings ratio:*

$$r(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n+1)}) = \frac{h(\tilde{\boldsymbol{X}}^{(n+1)}|\boldsymbol{Y})q(\tilde{\boldsymbol{X}}^{(n+1)}, \boldsymbol{X}^{(n)}|\boldsymbol{Y})}{h(\boldsymbol{X}^{(n)}|\boldsymbol{Y})q(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n+1)}|\boldsymbol{Y})} \tag{15}$$

- approve the transition to the new state (i.e., $\boldsymbol{X}^{(n+1)} = \tilde{\boldsymbol{X}}^{(n+1)}$) with probability

$$a(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n+1)}) = \min(1, r(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n+1)})), \tag{16}$$

  If the proposed state is rejected, the previous state is retained with $\boldsymbol{X}^{(n+1)} = \boldsymbol{X}^{(n)}$.

Over time, under some conditions, MCMC samples states $\boldsymbol{X}^{(n)}$ from the equilibrium distribution, which is designed in MH to be $P(\boldsymbol{X}|\boldsymbol{Y})$.

Remark 2 *The Metropolis algorithm is a special case of the Metropolis-Hastings algorithm, with the transition probability chosen such that* $q(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = q(\tilde{\boldsymbol{X}}, \boldsymbol{X})$.

Remark 3 *Variable-at-a-time and Composition MCMC allows a composition of update rules in different steps. For example, at each step, a subset of variables can be updated separately, given the other variables.*

Remark 4 *Gibbs sampling is a version of MCMC where at each step the algorithm samples some of the variables conditioned on other variables. It is used when the joint distribution of all the variables is difficult to compute, but it is computationally feasible to sample some of the variables at each step, while holding other variables fixed. Formally, this is a special case of MH. We mention this important variant here for completeness, but the algorithms described in this paper do not rely on this version of MCMC, which is often not trivial to compute for all variables.*

Needless to say, this brief discussion of MCMC is not a complete overview. The purpose of this discussion is to emphasize that MCMC can, in principle, be used to sample from a complicated posterior distribution even when the normalization of this

distribution is unknown, and that various update strategies can be mixed together in MCMC algorithms. Samples from the posterior produced by MCMC can be used to approximate an expected value of a variable, but also to study the uncertainty.

2.6. Metropolis-adjusted Langevin Algorithm (MALA)

MALA is a MH algorithm where the update proposal is given by the formula

$$\tilde{\boldsymbol{X}}^{(n+1)} = \boldsymbol{X}^{(n)} + \frac{\sigma^2}{2}\nabla \log P(\boldsymbol{X}^{(n)}|\boldsymbol{Y}) + \sigma\tilde{\boldsymbol{W}}^{(n+1)}, \tag{17}$$

where

$$\tilde{\boldsymbol{W}}^{(n+1)} \sim N\left(\boldsymbol{0}, I_d\right). \tag{18}$$

$\nabla \log P(\boldsymbol{X}^{(n)}|\boldsymbol{Y})$ is the gradient of the log-likelihood with respect to the variables. Note that the unnormalized $h(\boldsymbol{X}|\boldsymbol{Y})$ is sufficient for computing the MALA steps. σ is a parameter set by the user.

A positive definite preconditioner matrix A can be added without changing the equilibrium distribution:

$$\tilde{\boldsymbol{X}}^{(n+1)} = \boldsymbol{X}^{(n)} + \frac{\sigma^2}{2}A\nabla \log P(\boldsymbol{X}^{(n)}|\boldsymbol{Y}) + \sigma\sqrt{A}\tilde{\boldsymbol{W}}^{(n+1)}. \tag{19}$$

MALA is just an update rule for which the Hastings ratio can be computed as usual, making it a standard Metropolis Hastings update. The MALA algorithm is motivated by the Langevin stochastic differential equation. Loosely speaking, the Langevin stochastic differential describes a stochastic process which is analogous to Equation (17), with infinitesimally small updates (small σ); the equilibrium distribution of this stochastic process is $P(\tilde{\boldsymbol{X}}|\boldsymbol{Y})$.

Work such as [42] find relations between the Langevin equation and stochastic gradient descent (SGD), a key algorithm in the area of deep learning, which has also been applied to cryo-EM by cryoSPARC [24].

2.7. Hamiltonian Monte Carlo (HMC)

Hamiltonian Monte Carlo (HMC) is a another MCMC algorithm, which does not use the MH propose-accept-reject algorithm, like Gibbs sampling. Unlike Gibbs sampling, HMC does not require a sampling from a conditional distribution, but rather uses a gradient of the log-likelihood (like MALA) for a combination of deterministic steps (unlike MALA) and randomized steps. Due to the limited scope of this paper, and the complexity of ideas behind HMC, we refer the reader to one of the many resources about MCMC and HMC, such as [41] for additional information. In the context of this discussion, the key property of HMC is its use of the gradient, which we discuss in the context of MALA; however, HMC often has advantageous mixing properties compared to MALA.

3. Hyper-models

3.1. Toy Examples

The purpose of this section is to introduce synthetic examples which we will use to illustrate some of the ideas and in numerical experiments.

3.1.1. The "Cat" To illustrate the problem, we constructed the "cat," an object composed of Gaussian elements in real space, where each Gaussian follows a continuous trajectory as a function of the parameter t, so that we have a continuous space of objects, corresponding to an object with extensive large-scale heterogeneity. The heterogeneity is one-dimensional, where the state corresponds to the direction in which the cat's "head" is turned. Examples of synthetic 3-D object instances and the 2-D projections are presented in Figure 2 (rows 1-3). A dataset and a simulation using this model are described in Section 5.1.

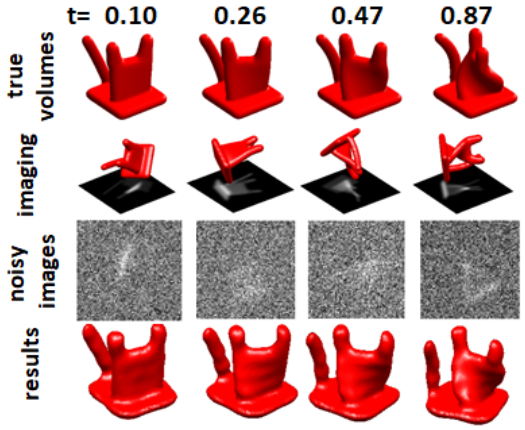

Figure 2. Sample cats: true 3-D instances (top row), rotated instance and noiseless projection images (second row), images with noise as used in the simulation (third row), and the reconstructed cat (bottom row, see Section 5.1)

3.1.2. The "Pretzel" To illustrate continuous heterogeneity with more structure, we constructed the "pretzel," which is composed of three parts: a rigid "base" and two independent "arms." The two heterogeneous regions are highlighted in the green and blue balls in Figure 3. In Figure 4(top) we present different versions of the Pretzel (for the purpose of this illustration, we hold one of the "arms" in a fixed state and sample different states of the other arm). In our simulations, each arm can take any state independently of the other. This is a simplified illustrative mock-up of a typical experiment where one part of the macromolecule is rigid and others are heterogeneous. A dataset and a simulation using this model are described in Section 5.2.

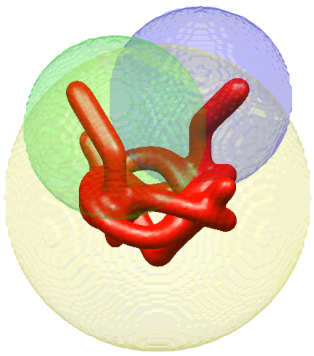

Figure 3. The anatomy of the pretzel: the green and blue regions identify the heterogeneous "arms." In the analysis in Section 5, the yellow region marked the bounds of where the rigid component, and the green and blue balls marked the bounds of the two heterogeneous components.

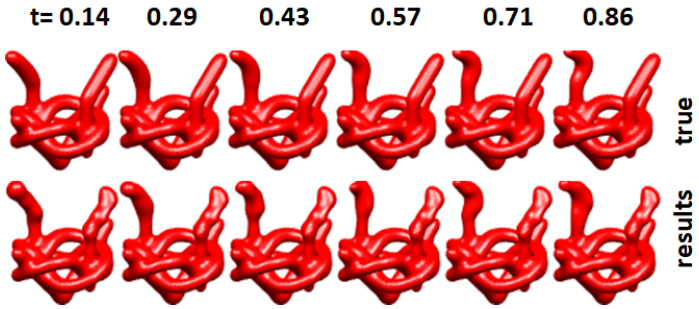

Figure 4. The pretzel: samples of true pretzels and reconstructed pretzels (see Section 5). For the purpose of this illustration only, we hold one of the arms in a fixed state. In the simulation and the recovered object, the arms move independently.

3.2. Generalizing Molecules: Hyper-Molecules

Hyper-molecules generalize 3-D density functions $\mathcal{V}(\boldsymbol{r})$ to higher dimensional functions $\mathcal{V}(\boldsymbol{r}, \boldsymbol{\tau})$, with the new state variable $\boldsymbol{\tau}$; each different conformation or state is represented by a fixed value of $\boldsymbol{\tau}$, and for a fixed $\boldsymbol{\tau}$, the 3-D density function $\mathcal{V}(\cdot, \boldsymbol{\tau})$ represents the molecule at that given conformation.

To illustrate the idea, we consider the cat example in Section 3.1.1. A natural way to view this cat is to produce a 3-D movie of the cat, where we would see a different conformation of the cat in each frame of the movie. In other words, each frame would present $\mathcal{V}(\cdot, \boldsymbol{\tau})$ for a different value of $\boldsymbol{\tau}$. Since the deformation of the cat is continuous, we could sample it at any arbitrary value of $\boldsymbol{\tau}$; a viewer may expect the movie to show a continuous transformation, with the cat not changing considerably as we move from one frame to the next. In other words, the movie would be expected to be relatively smooth (with several possible definitions of smoothness). This property of the movie reflects relations between different conformations. Hyper-molecules enforce such relations in the

modeling of $\mathcal{V}(\cdot, \boldsymbol{\tau})$.

We recall that density functions in cryo-EM are often assumed to be band-limited, effectively making them smooth in the spatial domain. This is enforced implicitly by the representation defined in (1), when the basis functions ψ_k are band-limited. Hyper-molecules enforce smoothness in the state-space through the definition in Equation (2), by choosing $\tilde{\psi}_k$ that have a similar property in the state variable in addition to the spatial variables. For example, in the case of 1-D state space in the cat example, with the only state variable representing the direction in which the cat is looking, a natural generalization of the representation in (1), generates 4-D basis functions $\tilde{\psi}_{k,q}(\boldsymbol{r}, t)$ from products $\tilde{\psi}_{k,q}(\boldsymbol{r}, t) = \psi_k(\boldsymbol{r})P_q(t)$ of 3-D functions ψ_k and low degree orthogonal polynomials P_q (e.g, Chebyshev polynomials), such that:

$$\mathcal{V}(\boldsymbol{r}, \tau) = \sum_{k,q} a_{k,q}\psi_k(\boldsymbol{r})P_q(t). \tag{20}$$

More generally, when there are d degrees of freedom of flexible motion, the manifold of conformations is of dimension d and the time variable t in Equation (20) is replaced by manifold coordinates $\boldsymbol{\tau} \in T$. The polynomials P_q are replaced by a truncated set of basis functions over the manifold, denoted $P_q(\boldsymbol{\tau})$, with a minor abuse of notation:

$$\mathcal{V}(\boldsymbol{r}, \boldsymbol{\tau}) = \sum_{k,q} a_{k,q}\psi_k(\boldsymbol{r})P_q(\boldsymbol{\tau}). \tag{21}$$

For example, the basis function $P_q(\boldsymbol{\tau})$ can be the product of polynomials in multiple variables.

The model in Section 2.2 then generalizes naturally, such that Equation (5) is generalized to

$$I^{(i)}(x_1, x_2) = a_i H_i * \int_{\mathbb{R}} \mathcal{V}(R_i^{-1}\mathbf{x} + \mathbf{s}_i, \boldsymbol{\tau}_i)dx_3, \tag{22}$$

the corresponding operator $A(R_i, \boldsymbol{q}_i)$ to $A(R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i)$, and the posterior (12) to

$$P\left(\{R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V} | \{Y^{(i)}\}_i\right) \propto P\left(\{R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V}\right) \prod_i P(Y^{(i)} | R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V})). \tag{23}$$

In other words, we use the formulation of the continuously heterogeneous molecules as hyper-molecules to turn generalize the Bayesian formulation of the cryo-EM problem from a problem of recovering a 3-D molecule from 2-D projections in unknown viewing directions, to a problem of recovering a higher dimensional hyper-molecule from 2-D projections. The key to this formulation, compared to a formulation as a collection of independent molecules, is that hyper-molecules encode relations between states/conformations, with the related property that they encode a continuum of states varying smoothly.

3.3. Enforcing Structure

We note that there exists an equivalent scheme using appropriate samples in the state space, which would be numerically equivalent to our use of polynomials in the state

variable; we can also construct a crude approximations using discretized states and linear interpolations. However, hyper-molecules are different from independent molecules because they provide relations between states/conformations. This regularity in the relation between states can be further reinforced by generalizing other ideas implemented for 3-D molecules, such as priors that favor lower coefficients for basis functions with high-frequency components in the state variable.

The basis functions presented above are not the only way to define such relations between states; for example, one can use a discretized state space, and penalize for large changes between adjacent states using a term of a form such as:

$$L(\mathcal{V}) = \sum_{t=1}^{T-1} \int |\mathcal{V}(\boldsymbol{r}, t) - \mathcal{V}(\boldsymbol{r}, t+1)|^2 d\boldsymbol{r}. \tag{24}$$

In fact, smoothness and continuity are crude proxies for properties that we would expect to find in the state space of molecules. For example, often, we would expect to to observe a flow of mass as we move between states. This would be captured better through a Wasserstein distance between states; additional physical properties are discussed in the remainder of the paper and in [43]. In the Bayesian formulation it is natural to add more explicit priors for hyper-molecules.

3.4. A Curse of Dimensionality

Building upon the success of the maximum likelihood and MAP frameworks in cryo-EM (see discussion in Section 2.4), it is natural to consider their application to the hyper-volume reconstruction problem. The expectation maximization algorithms are iterative refinement algorithms which attempt to recover the maximum-likelihood or MAP solution, alternating between updating the distributions of variables such as the viewing direction R_i, and updating the estimate of the molecule $\mathcal{V}$ (i.e., coefficients in the representation of the object as defined in Equation (1)). Generating the projections for all viewing directions and comparing them to all particle images are a computationally intensive operations in the implementation of expectation maximization algorithms.

In the case of hyper-molecules, expectation maximization would be generalized to alternating between updating the joint distribution over viewing directions R_i and (possibly high dimensional) state variables $\boldsymbol{\tau}_i$ (compared to a small number of discrete conformations in current algorithms), and updating the hyper-molecule (21). In other words, one would have to project the hyper-molecule in every possible state in every possible viewing direction, and compare each particle image with each to these projections, rapidly increasing the number of comparisons in this already expensive procedure. More complex models of hyper-molecules, introduced later in this paper, would make it more difficult to design specialized algorithms and heuristics to optimize this procedure.

In addition, we note that the number of coefficients required to represent a molecule as a linear combination of basis functions in Equation (1), at a resolution corresponding to about $N \times N \times N$ voxels, is $O(N^3)$. Similarly, adding d-dimensional heterogeneity

at "state resolution" Q implies $O(N^3Q^d)$ coefficients. High-dimensional heterogeneity, arising, for example, in molecules that have several independent heterogeneous regions, results in a very large number coefficients which could exceed the overall number of pixels in all the particle image in an experiment despite the high throughput of cryo-EM. Indeed, since hyper-molecules have the capacity to represent very generic molecules, it is natural to expect that a lot of data would be required to estimate them; in particular, if the number of possible states (in some discretization) grows exponentially fast with the dimension, it is natural to expect the required number of particle images to grow as fast, if not faster. Given infinite data and infinite computational resources, it is tempting to model very little and allow the data and algorithm to reveal the structure. Unfortunately, despite the rapid growth in cryo-EM throughput and computational resources, they are far from "infinite." The natural question to ask is if we can use prior knowledge and assumptions to reduce the amount of data that we need, even in the case of high-dimensional heterogeneity.

In the remainder of this paper, we address some of these challenges.

3.5. Finer Structures I: Composite Hyper-Molecules

In the previous section, we found that recovering a hyper-molecule which describes very generic complicated dynamics of a macromolecule requires massive amounts of data. Often, researchers have prior-knowledge about the structure and dynamics of a macromolecule that they study. For example, many macromolecules are composed of a relatively static component to which smaller flexible heterogeneous components are attached (for an illustrative toy example, see the pretzel example in Section 3.1.2). Often, practitioners are able to use traditional cryo-EM algorithms to recover the static component at a high resolution, but the regions of the flexible components are much more blurry. In these cases, researchers are often able to hypothesize where each component is located, which components are static, and which components are heterogeneous. Tools for estimation of local variance and resolution help researchers in identifying these regions [44, 45, 46, 47, 48, 49, 50].

We introduce *composite hyper-molecules*, a model which is the sum of M components $\mathcal{V}^{(m)}$, each of which is a hyper-molecule. The following formula describes a simple version of a composite hyper-molecule:

$$\mathcal{V}(\boldsymbol{r}, \boldsymbol{\tau}^1, \boldsymbol{\tau}^2, ..., \boldsymbol{\tau}^M) = \sum_{m=1}^{M} \mathcal{V}^{(m)}(\boldsymbol{r}, \boldsymbol{\tau}^m). \tag{25}$$

Each component is constrained to a certain region of space where it is assumed to be supported (the regions may overlap). Each component has its own set of state variables and coefficients that describe it. In our pretzel example, the yellow region in Figure 3 is modelled as a rigid static "body," and the green and blue regions represent regions of space where two one-dimensional heterogeneous components are supported. As can be seen in this example, the regions may overlap and do not have to be tight around the actual true object.

In some cases, the different components could be roughly described as moving one with respect to the other, with more subtle deformations (for example, at the interface between the components). Indeed, heterogeneous macromolecule have been modelled as a superposition of a number of rigid objects in somewhat arbitrary relative positions in work such as [34, 35, 36, 37, 37, 38, 39]. We recall that hyper-molecules and the composite hyper-molecules in Equation (25) are generic enough to describe the relative motion of these components, but if such dynamics can be assumed, capturing them in the model is advantageous for computational and statistical reasons. Therefore, a more complete version of composite hyper-objects allows both motion and heterogeneity in each component

$$\begin{aligned}&\mathcal{V}(\boldsymbol{r}, \boldsymbol{\tau}^{1,state}, \boldsymbol{\tau}^{2,state}, ..., \boldsymbol{\tau}^{M,state}, \boldsymbol{\tau}^{1,position}, \boldsymbol{\tau}^{2,position}, ..., \boldsymbol{\tau}^{M,position}) = \\ &\sum_{m=1}^{M} \mathcal{V}^{(m)}(f^m(\boldsymbol{r}, \boldsymbol{\tau}^{m,position}), \boldsymbol{\tau}^{m,state})\end{aligned} \tag{26}$$

where $f^m(\boldsymbol{r}, \boldsymbol{\tau}^{m,position})$ is a function that describes the trajectory of the m-th component, so that the component is in heterogeneity state $\boldsymbol{\tau}^{m,state}$, and its location along the "trajectory" is determined by the position variable $\boldsymbol{\tau}^{m,position}$. For example, a simple affine f^m can take the form

$$f^m(\boldsymbol{r}, \boldsymbol{\tau}^{m,position}) = \begin{pmatrix} \tau^{m,state}\theta_{x,1}^{m,position} + \theta_{x,2}^{m,position} + r_x \\ \tau^{m,state}\theta_{y,1}^{m,position} + \theta_{y,2}^{m,position} + r_y \\ \tau_i^{m,state}\theta_{z,1}^{m,position} + \theta_{z,2}^{m,position} + r_z \end{pmatrix}, \tag{27}$$

where $\boldsymbol{r} = (r_x, r_y, r_z)^\intercal$. The variables $\boldsymbol{\theta}^{m,position}$, which determine the trajectory, are part of the variables describing the hyper-molecule, much like the coefficients in Equation (25). Actual trajectory functions would presumably be more complex and could involve a rotation and deformations.

The variables for the position $\boldsymbol{\tau}^{m,position}$ and state $\boldsymbol{\tau}^{m,state}$ can be closely related (the position can be related or unrelated to the heterogeneity state variable for that component); for brevity, we use $\boldsymbol{\tau}^m$ as a state variable that encapsulates both $\boldsymbol{\tau}^{m,position}$ and $\boldsymbol{\tau}^{m,state}$.

Compared to previous work like [34, 35, 36, 37, 37, 38, 39], the composite hyper-molecule formulation models components that are inherently non-rigid, and, in particular, models the flexible interface between components. Furthermore, composite hyper-molecules model the possible relative positions of the different components with respect to each other ("trajectories") and deformations, bother are parametrized so that they can be fitted using data.

Remark 5 *In some cases, there are relations between the different regions that can be captured in the description of the composite hyper-molecule. For example, our pretzel has two identical arms (shifted and rotated with respect to each other). While each arm can appear in a different state independently from the other arm, they have the same fundamental structure (i.e., they are the same hyperobject, at a different state and*

position). A similar phenomenon is observed in some macromolecules that have certain symmetries. We capture this fact in our model by having the hyper-objects of the two arms share the coefficients in their representation. This is analogous to "weight sharing" in deep neural networks.

3.6. Finer Structures II: Priors and "Black-Box Hyper-Molecules"

The purpose of this section is to add a layer of abstraction to the modeling of hyper-molecules, where the model can be implemented as a "black-box" provided to an algorithm designed to recover hyper-molecules; the algorithms themselves are discussed in later sections, while this section focuses on the formal modeling of these components. These black-box models will allow users with different levels of technical expertise to define more elaborate models and priors which reflect assumptions and prior knowledge about the experiment, to the extent that such assumptions are necessary given the amount of data, model complexity and computational resources that are available.

We recall that the formulation of the hyper-molecule $\mathcal{V}$ as a sum of basis functions in Equation (21). We denote the coefficients of these basis functions by $\boldsymbol{\theta}$. Similarly, in the formulation in Equation (26), the coefficients of the basis functions in every component and the coefficients of the trajectories are denoted collectively by $\boldsymbol{\theta}$. We write this fact explicitly using the notation $\mathcal{V}[\boldsymbol{\theta}](r, \boldsymbol{\tau})$. We revisit Equation (23), and add this explicit notation:

$$P\left(\{R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}] | \{Y^{(i)}\}_i\right) \propto P\left(\{R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \boldsymbol{\theta}\right) \prod_i P(Y^{(i)} | R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}])). \quad (28)$$

In particular, it is compelling to decompose (28) into simpler components and formulate a slightly more specific structure such as

$$\begin{aligned} &P\left(\{R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i\}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}] | \{Y^{(i)}\}_i\right) \propto \\ &P\left(\boldsymbol{\theta}\right) P\left(\boldsymbol{\mu}\right) \prod_i P(Y^{(i)} | R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}])) P\left(R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i | \boldsymbol{\mu}\right). \end{aligned} \quad (29)$$

where $P\left(\boldsymbol{\theta}\right)$ is a black-box prior for the hyper-molecule, $P\left(\boldsymbol{\mu}\right)$ is a black-box prior for imaging variables and latent variables (e.g., noise parameters and CTF parameters for micrographs), $P\left(R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i | \boldsymbol{\mu}\right)$ is a prior for the variables of each particle image (e.g., shift from center, contrast parameters), and $P(Y^{(i)} | R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$ is the relation to the measurements.

In this formulation, $\mathcal{V}$ can be replaced by an arbitrary "black-box" function that produces a consistent notion of a hyper-molecule; this black-box formulation decouples the specifics of the model from the algorithm, giving the scientist more flexibility in defining their model. The key components Indeed in this formulation are the model $\mathcal{V}[\boldsymbol{\theta}]$ which defines the density at any spatial position and state as a function of the coefficients $\boldsymbol{\theta}$, and a prior $P\left(\boldsymbol{\theta}\right)$. These two components encode the scientist's assumptions, prior knowledge and physical constraints. Another key component is $P(Y^{(i)} | R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$, which contains the imaging model. The components $P\left(R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i | \boldsymbol{\mu}\right)$ and $P\left(R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i | \boldsymbol{\mu}\right)$ give some additional flexibility.

4. Algorithms

In this section we discuss the role of MCMC algorithms in the framework for recovering hyper-molecules.

4.1. MCMC, MALA and HMC

We consider the Bayesian formulation of recovering a hyper-molecule in Equation (29). The difficulty with expectation-maximization algorithms is that they must compute $P(R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i | Y^{(i)}, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$ as a function of all possible combinations of viewing directions R_i, states $\boldsymbol{\tau}_i$, and some of the other particle-image specific variable $\boldsymbol{q}_i$ (e.g., in-plane shift) at every iteration (the update of $\boldsymbol{\theta}$ involves another computationally expensive operation for similar reasons). This involves some crude discretization of these variables and very large number of comparisons which are computationally expensive, at every iteration. This computational challenge presents a difficulty in the homogeneous case and in the case of discrete heterogeneity when there is a small number of conformations; the natural generalization to high-dimensional continuous heterogeneity increases the computational complexity exponentially fast with the dimensionality of the heterogeneity.

We propose to an MCMC framework for sampling from the posterior in Equation (29); some of the main features of MCMC are reviewed briefly in Section 2.5. We note that MCMC is not a single algorithm, but a collection of algorithms that can be used together.

Equation (29) and the analogy to expectation maximization, suggests that different variables in the MCMC formulation can be treated separately, mixing strategies for updating a subset of variables while holding the others constant. In particular, the particle-image variables $R_i, \boldsymbol{\tau}_i$ and $\boldsymbol{q}_i$ can be evaluated separately and in parallel, because they are independent conditioned on $\boldsymbol{\mu}$ and $\mathcal{V}[\boldsymbol{\theta}$. MCMC algorithms such as a simple MH (with a simple proposition strategy), do not require the computation of the distribution $P(Y^{(i)} | R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$ for every value of $R_i, \boldsymbol{\tau}_i$ and $\boldsymbol{q}_i$, but rather require only the ratio $P(R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i | Y^{(i)}, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))/P(\tilde{R}_i, \tilde{\boldsymbol{\tau}}_i, \tilde{\boldsymbol{q}}_i | Y^{(i)}, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$ between the likelihood of different values of the variables; in other words, at every iteration, this version of MCMC requires only the evaluation at two points in the update of particle-image specific variables, and it is sufficient to have $P(Y^{(i)} | R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$ up to a multiplicative constant (so that the probability does not need to be normalized to integrate to 1). Other strategies, such as MALA and HMC, require the gradient of the log-likelihood with respect to the different variables (again, implying that the probability does not need to be normalized to integrate to 1). Similar considerations apply to the update of other variables. We note that MCMC is not a "magic solution" to the computational challenge, because it may require more steps than expectation maximization, but each step is computationally feasible, and different strategies and existing tools can easily be combined to improve performance.

MCMC yields a sample of the variables and latent variable; we can restrict our attention to variables such as $\boldsymbol{\theta}$ which are sampled hyper-molecules, and we can consider

the statistics of $\boldsymbol{\tau}$ if we wish to study the statistics of states. Most often, in practice, $\boldsymbol{\theta}$ or $\mathcal{V}$ can be averaged over all the samples to produce an "expected" hyper-molecules, although this averaging can introduce some technical difficulties do to ambiguities which we will discuss briefly latter; this issue is not uncommon in this type of problems, and most common does not present a real practical difficulty. A similar problem happens the the maximum-likelihood and MAP approaches, since there are several equivalent solutions. There too, this is not a problem in practice. The advantage of having multiple samples from the posterior however, is that they allow us to study the uncertainty in the solution by studying the variability of $\mathcal{V}$.

4.2. A Remark about Black-Box Hyper-Molecules

In this section, we revisit the Bayesian formulation in Equation (29) and discuss some aspects of the formulation of generalized hyper-molecules that are related to the algorithms and implementation. In principle, it is sufficient to define black-box functions which would evaluate the prior $P(\boldsymbol{\theta})$ and the density $\mathcal{V}[\boldsymbol{\theta}](r, \boldsymbol{\tau})$ at any spatial (or frequency) location r, and any state $\boldsymbol{\tau}$ (and possibly provide the interface for computing gradients over the difference variables); the algorithm would use these functions to compute $P(Y^{(i)}|R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$ using its imaging model. In practice, a generic implementation of function for computing $\mathcal{V}[\boldsymbol{\theta}](r, \boldsymbol{\tau})$ may present certain numerical and computational difficulties, because the implementation of the imaging model if often closely related to the representation of $\mathcal{V}[\boldsymbol{\theta}]$; for example, it would be inefficient and numerically accurate to compute projection for each particle-image (as a step in the imaging procedure) from explicit sampling of $\mathcal{V}[\boldsymbol{\theta}]$ at many sample points.

We note that the explicit evaluation of $\mathcal{V}[\boldsymbol{\theta}]$ isn't required in Equation (29). Instead, $\mathcal{V}$ is considered implicitly in prior $P(\boldsymbol{\theta})$ and explicitly in the comparison to images in $P(Y^{(i)}|R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$. The explicit use in $P(Y^{(i)}|R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}]))$ implies that the algorithm would use the black-box $\mathcal{V}$ to evaluate the the hyper-molecule at some points in order to produce an image using the algorithm's own imaging models. In fact, this can be numerically inaccurate and computationally expensive without certain assumptions on the structure of $\mathcal{V}$. It is therefore useful to include such assumptions, or to implement efficient functions that produce projections projections of the hyper-molecule that are consistent with the model implemented for the black-box $\mathcal{V}$. In our implementation, such a module computes $log(P(Y^{(i)}|R_i, \boldsymbol{\tau}_i, \boldsymbol{q}_i, \boldsymbol{\mu}, \mathcal{V}[\boldsymbol{\theta}])))$ (i.e., the projections and the statistical model for comparing the projections to the particle image in the dataset).

These considerations highlight the fact that complete decoupling of the hyper-molecule model from other components may be a challenge, but a balance can be found between the need for abstraction and implementation considerations.

5. Implementation and Numerical Results

In this section we discuss two prototypes constructed for the recovy of hyper-molecules based on the ideas presented in this paper, and present the results of experiments with synthetic data.

5.1. A First Prototype

The first prototype using hyper-objects was implemented in Matlab. This simplified proof of concept does not consider shifts in the images, CTF, etc., but it does included noise.

We applied this prototype to the Cat discussed in Section 3.1.1. We used the simplified imaging model, without in-plane shifts or a CTF (see discussion of the Cryo-EM imaging model in Section 2.2) to produce 10,000 particle images, 65×65 pixels each, in various viewing directions and states, and added Gaussian white noise to the images. The SNR in this experiment was 1/16. Examples of the simulated images are presented in Figure 2 (second and third row).

The algorithm was run on computer equipped with Intel(R) Core(TM) i7-4770 CPU, 32GB RAM server and a single NVIDIA GeForce GTX 980 Ti GPU with 6 GB GPU RAM for about 5 hours. We note that very similar results have been obtained even in the 1-2 hours of this experiment and of similar shorter experiments. This version and the experiments are discussed in more detail in our report in [43].

For the purpose of visualization of the computed hyper-molecule, we choose sample values of the parameter τ, and approximated the molecule at regular grid points in real space. We use Matlab's "isosurface" and "patch" to visualize level sets of the objects. Examples of reconstructed objects are presented in Figure 2 (bottom).

5.2. A Second Prototype

The second prototype implements simple composite hyper-objects (see Section 3.5). In each experiment, the user can define the number and position of heterogeneous components of the hyper-object. Each component can be defined to be rigid, or heterogeneous with a 1-D or 2-D state space. Finally, the user can define components that share the same parameters, but not the same state; in the Pretzel example, each of the two arms is identical, but in each image, each arm can be in a different state. Each object is represented using 3-D Generalized Prolate Spheroidal Functions, which are the optimal basis for representing objects that are as concentrated as possible in the spatial domain ("compactly supported") and in the frequency domain ("band-limited"), for more details see [10]). These 3-D basis functions are multiplied by 1-D or 2-D cosines and sines to produce higher dimensional components.

The MCMC algorithm implements MALA steps for updating the coefficients $\boldsymbol{\theta}$ of the hyper-molecule, and simpler MH steps (random perturbation of the variables to propose new values) for updating the viewing direction, state, in-plane shift and contrast

of each particle image (MALA has not been implemented for these yet). We are working on implementing MALA and HMC for additional variables. The algorithm has a second mode, provided as a crude approximation of MCMC, where in each iteration, only a subset of the particle image variables (viewing direction, state, etc.) are updated (using a MH step for each particle image); the hyper-molecule is updated using a gradient step, based only on the subset of particle-images used in that iteration. This new prototype was also written in Matlab.

We generated a dataset of 20,000 synthetic pretzel images (synthetic model described in Section 3.1.2), 151×151 pixels each, at about SNR of about 1/30, including the effects of in-plane shifts and CTF. First, we set up a homogeneous model, and ran the algorithm with random initial viewing directions and in-plane shifts, and with the initial model set to 0 everywhere. This run produces an initial alignment of the viewing directions. Next, we set up a the model depicted in Figure 3, with a rigid object support in the yellow sphere, and two heterogeneous regions, each supported in one of the other spheres. The two heterogeneous regions are identical components (share coefficients, but shifted and rotated with respect to one another), but each of them can appear in a different state in each particle image. The state variable is initialized uniformly random. The algorithm starts with a low frequency representation of hyper-molecule, then gradually increases the frequencies allowed in the representation; the gradual increase in frequency of the representation of 3-D density functions is common practice in cryo-EM, which is generalized here to gradual increase in the frequencies allowed in the state variable. The processing requires about a 5 days, using a server equipped with a E5-2680 CPU and 720GB RAM and one NVIDIA Tesla P100 GPU with 16GB of RAM. The results are presented in Figure 4(bottom).

6. Discussion and Future Work

The main goal of this paper is to introduce the idea of hyper-molecules as high-dimensional representations of 3-D molecules at all their conformations; this idea is applicable to other inverse problem such as CT. In addition to the generalization of 3-D molecules to hyper-molecules, we generalize the Bayesian formulation of cryo-EM to a Bayesian formulation of continuous heterogeneity in cryo-EM. Compared to existing work on representing molecules in a small number of discrete conformations, hyper-molecules provide a way of describing a continuum of structure and the relations between states.

These higher-dimensional objects can be represented as generic high-dimensional functions, but we discuss statistical and computational motivations to introduce more general models of hyper-molecules, that describe more specific objects, when prior knowledge is available. These statistical and computational challenges, along with the customised hyper-molecule representations that we propose, pose a challenge to the scalability of the algorithms that are currently popular in cryo-EM to these settings. Therefore, we also discuss a MCMC framework which overcomes some of the technical

difficulties in phrasing each iteration of current algorithms in the more general settings that we propose, and we discuss some of the additional benefits of this framework. Furthermore, we note that the MCMC framework provides a natural connection to atomic structures and other experiment modalities, demonstrated for example in [51], which uses a density map produced from a cryo-EM experiment together with physical models and other modalities.

Ultimately, the goal of this like of work is to provide a highly customizable framework for encoding prior knowledge about complex molecules, and to find the realistic trade-off between the bias that can be introduced by assumptions and the realistic constraints on the amount of data that can be collected. We envision this framework as a combination of imaging modules for modeling hyper-molecules, adapted to fast computation of projection images and to computing gradients with respect to variables such as the viewing direction and model coefficients. Such modules will be used in a framework inspired by TensorFlow [5], PyTorch [6] (both designed primarily for deep learning) and Edward [8, 9], which allow to construct modules analogous to the black-box modules discussed in this paper, with more focus on imaging as in ODL [7]. Ideally, a wide array of general purpose tools and algorithms constructed for optimization, Bayesian inference, deep learning and imaging could be used together with this framework. However, the large scale of the cryo-EM problem and various properties of the problem require a more specialized framework and flexibility in solver strategies; for example, the memory management in software designed for deep learning is often optimized to work with small batches, whereas in some implementations of imaging algorithms there are computational advantages in working with very large batches. Another example is the update of in-plane shift variables, which can be performed without recomputing the entire image. We demonstrated the ideas in this paper in two prototypes; we are currently building the next prototype, which would be more customizable and scalable.

Our reference to tools such as TensorFlow, PyTorch and Edward demonstrates that the lines between optimization, stochastic optimization, MCMC and other algorithms are not entirely rigid, in the sense that modules used in one framework can be used in some other frameworks. We expect to experiment with other algorithms for initialization of MCMC and approximation of steps, and to examine additional Bayesian inference algorithms. Indeed, we have already experimented with using expectation maximization algorithms to initialize crude viewing directions in cryo-EM data, and with SGD hybrids for approximating MCMC steps.

In the following sections we briefly comment on some additional aspects of the problem.

6.1. The Homogeneous case, Discrete Heterogeneity and Continuous Heterogeneity

In many cases, molecules appear mainly in a discrete set of conformations that are very similar to one another. While we mainly discuss continuous heterogeneity in the paper,

the framework proposed here applies to the discrete case (or mixtures of discrete and continuous heterogeneity in different regions) with few changes (for example, the basis functions used to capture the variability as a function of the heterogeneity parameter t can be replaced by the Haar basis). Hyper-molecules, composite hyper-molecules and the algorithms discussed here are advantageous in the discrete case as well: they allow to use the similarity between different conformations, and they allow to decompose the heterogeneity to local heterogeneity in different regions.

More generally, we hope that a generic Bayesian framework could also be used to study more elaborate models for imaging and experiment latent variables even in the homogeneous case.

6.2. Ambiguity

We note that even in the classic cryo-EM problem, certain ambiguities emerge in the macro-molecules that are recovered: any result has "equivalent" results that are identical up to global rotation, shifts and reflection. Naturally, hyper-molecules have similar ambiguities. Since hyper-objects generalize the spatial coordinate and in many ways treat the state parameters in the same way as they treat the spatial coordinate, one may expect a generalized form of ambiguity to appear. Indeed, there is ambiguity in how the molecules in different states are aligned with respect to each other, and ambiguity in parameterization of the state space. These ambiguities are reduced by regularization or priors, or when when the model contains rigid components that align other components.

One such effect can be observed in the cat example in Figure 2, where the recovered cats are aligned slightly differently with respect to each other compared to the original cats (the change in alignment is continuous, so the "movie of cat" is still continuous). Of course, our original alignment was arbitrary, so the algorithm's choice is no better or worse than ours, but it is better suited to the limited degree polynomials we allowed the algorithm to use to represent these recovered Cats.

7. Conclusions

A mathematical formulation and a Bayesian formulation has been presented for the modeling of continuously heterogeneous molecular structures. This formulation "hyper-molecules," and its generalizations, allow to model generic heterogeneous molecules, or to encode structural constraints and priors, where these are available or required for practical reasons.

In addition, a computational framework based on MCMC has been presented for the recovery of hyper-molecules from cryo-EM data. This framework addresses some of the computational challenges associated with generalizing existing popular algorithms to the cryo-EM problem. In particular, it bypasses the computationally-intensive estimation of the conditional distribution of variables such as the viewing direction of

each particle image at each iteration of expectation maximization, which would become more computationally demanding if additional state variables are introduced in the case of continuous heterogeneity. This framework also offers a natural way to incorporate elaborate black-box models that researchers can customize for their needs, and a tool for studying the uncertainty in solutions.

The ideas presented in this paper have been demonstrated in prototype implementations, applied to synthetic data. Work on real datasets will be discussed seperately. More scalable implementations are being constructed for more genericc models, larger datasets, and more efficient computation.

Acknowledgments

The authors would like to thank Fred Sigworth and Tejal Bhamre for their help.

A. Singer was partially supported by NIGMS Award Number R01GM090200, AFOSR FA955017-1-0291, Simons Investigator Award, the Moore Foundation Data-Driven Discovery Investigator Award, and NSF BIGDATA Award IIS-1837992. These awards also partially supported R. R. Lederman at Princeton University.

References

[1] Daifei Liu, Xueqi Liu, Zhiguo Shang, and Charles V Sindelar. Structural basis of cooperativity in kinesin revealed by 3d reconstruction of a two-head-bound state on microtubules. *Elife*, 6:e24490, 2017.

[2] Drew M Dolino, Soheila Rezaei Adariani, Sana A Shaikh, Vasanthi Jayaraman, and Hugo Sanabria. Conformational selection and submillisecond dynamics of the ligand-binding domain of the n-methyl-d-aspartate receptor. *Journal of Biological Chemistry*, 291(31):16175–16185, 2016.

[3] Sjors HW Scheres. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.*, 415(2):406–418, 2012.

[4] Daniel A Low, Michelle Nystrom, Eugene Kalinin, Parag Parikh, James F Dempsey, Jeffrey D Bradley, Sasa Mutic, Sasha H Wahab, Tareque Islam, Gary Christensen, et al. A method for the reconstruction of four-dimensional synchronized ct scans acquired during free breathing. *Medical physics*, 30(6):1254–1263, 2003.

[5] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[6] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[7] Jonas Adler, Holger Kohr, and Ozan Öktem. Odl-a python framework for rapid prototyping in inverse problems. *Royal Institute of Technology*, 2017.

[8] Dustin Tran, Alp Kucukelbir, Adji B. Dieng, Maja Rudolph, Dawen Liang, and David M. Blei. Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint arXiv:1610.09787*, 2016.

[9] Dustin Tran, Matthew D. Hoffman, Rif A. Saurous, Eugene Brevdo, Kevin Murphy, and David M. Blei. Deep probabilistic programming. In *International Conference on Learning Representations*, 2017.

[10] Roy R Lederman. Numerical algorithms for the computation of generalized prolate spheroidal functions. *arXiv preprint arXiv:1710.02874*, 2017.

[11] Takeshi Kawabata. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophysical journal*, 95(10):4643–4658, 2008.

[12] Roy R Lederman and Amit Singer. A representation theory perspective on simultaneous alignment and classification. Submitted, 2016.

[13] J. Frank. *Three-dimensional electron microscopy of macromolecular assemblies*. Academic Press, 2006.

[14] Fred J. Sigworth. Principles of cryo-EM single-particle image processing. *Microscopy*, 65(1):57–67, 12 2015.

[15] Yifan Cheng, Nikolaus Grigorieff, PawelA. Penczek, and Thomas Walz. A primer to single-particle cryo-electron microscopy. *Cell*, 161(3):438–449, 2015.

[16] Jacqueline LS Milne, Mario J Borgnia, Alberto Bartesaghi, Erin EH Tran, Lesley A Earl, David M Schauder, Jeffrey Lengyel, Jason Pierson, Ardan Patwardhan, and Sriram Subramaniam. Cryo-electron microscopy–A primer for the non-microscopist. *FEBS Journal*, 280(1):28–45, 2013.

[17] Kutti R Vinothkumar and Richard Henderson. Single particle electron cryomicroscopy: Trends, issues and future perspective. *Q. Rev. Biophys.*, 49, 2016.

[18] S. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.*, 180(3):519–530, 2012.

[19] Maofu Liao, Erhu Cao, David Julius, and Yifan Cheng. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107–112, 2013.

[20] Fred J. Sigworth. A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.*, 122(3):328–339, 1998.

[21] Fred J Sigworth, Peter C Doerschuk, Jose-Maria Carazo, and Sjors HW Scheres. Chapter ten-an introduction to maximum-likelihood methods in cryo-em. *Methods in enzymology*, 482:263–294, 2010.

[22] Dari Kimanius, Björn O Forsberg, Sjors HW Scheres, and Erik Lindahl. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife*, 5, nov 2016.

[23] Jasenko Zivanov, Takanori Nakane, Björn O Forsberg, Dari Kimanius, Wim JH Hagen, Erik Lindahl, and Sjors HW Scheres. New tools for automated high-resolution cryo-em structure determination in relion-3. *Elife*, 7:e42166, 2018.

[24] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, 14(3):290–296, 2017.

[25] Ali Punjani, Marcus Brubaker, and David Fleet. Building proteins in a day: Efficient 3d molecular structure estimation with electron cryomicroscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[26] M. Shatsky, R. Hall, E. Nogales, J. Malik, and S. Brenner. Automated multi-model reconstruction from single-particle electron microscopy data. *J. Struct. Biol.*, 170(1):98–108, 2010.

[27] Amit Singer, Ronald R Coifman, Fred J Sigworth, David W Chester, and Yoel Shkolnisky. Detecting consistent common lines in cryo-EM by voting. *J. Struct. Biol.*, 169(3):312–322, 2010.

[28] Yoel Shkolnisky and Amit Singer. Viewing direction estimation in cryo-EM using synchronization. *SIAM J. Imaging Sci.*, 5(3):1088–1110, 2012.

[29] Afonso S Bandeira, Yutong Chen, and Amit Singer. Non-unique games over compact groups and orientation estimation in cryo-em. *arXiv preprint arXiv:1505.03840*, 2015.

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

[31] Ali Dashti, Peter Schwander, Robert Langlois, Russell Fung, Wen Li, Ahmad Hosseinizadeh, Hstau Y. Liao, Jesper Pallesen, Gyanesh Sharma, Vera A. Stupina, Anne E. Simon, Jonathan D. Dinman, Joachim Frank, and Abbas Ourmazd. Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl. Acad. Sci. U.S.A.*, 111(49):17492–17497, 2014.

[32] P Schwander, R Fung, and A Ourmazd. Conformations of macromolecules and their complexes from heterogeneous datasets. *Phil. Trans. R. Soc. B*, 369(1647):20130567, 2014.

[33] Joachim Frank and Abbas Ourmazd. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-em. *Methods*, 100:61–67, 2016.

[34] Takanori Nakane, Dari Kimanius, Erik Lindahl, and Sjors HW Scheres. Characterisation of molecular motions in cryo-em single-particle data by multi-body refinement in relion. *Elife*, 7:e36861, 2018.

[35] Alexey Amunts, Alan Brown, Xiao-chen Bai, Jose L. Llácer, Tanweer Hussain, Paul Emsley, Fei Long, Garib Murshudov, Sjors H. W. Scheres, and V. Ramakrishnan. Structure of the yeast mitochondrial large ribosomal subunit. *Science*, 343(6178):1485–1489, 2014.

[36] 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.

[37] Qiang Zhou, Xuan Huang, Shan Sun, Xueming Li, Hong-Wei Wang, and Sen-Fang Sui. Cryo-em structure of snap-snare assembly in 20s particle. *Cell research*, 25(5):551, 2015.

[38] Xiao-chen Bai, Eeson Rajendra, Guanghui Yang, Yigong Shi, and Sjors HW Scheres. Sampling the conformational space of the catalytic subunit of human γ-secretase. *Elife*, 4:e11182, 2015.

[39] Serban L Ilca, Abhay Kotecha, Xiaoyu Sun, Minna M Poranen, David I Stuart, and Juha T Huiskonen. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nature communications*, 6:8843, 2015.

[40] Joakim Andén and Amit Singer. Structural Variability from Noisy Tomographic Projections. *SIAM Journal on Imaging Sciences*, 11(2):1441–1492, jan 2018.

[41] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[42] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

[43] Roy R Lederman and Amit Singer. Continuously heterogeneous hyper-objects in cryo-em and 3-d movies of many temporal dimensions. *arXiv preprint arXiv:1704.02899*, 2017.

[44] Weiping Liu and Joachim Frank. Estimation of variance distribution in three-dimensional reconstruction. I. Theory. *J. Opt. Soc. Am. A*, 12(12):2615–2627, Dec 1995.

[45] P. A. Penczek. Variance in three-dimensional reconstructions from projections. In *Proc. ISBI*, pages 749–752, 2002.

[46] Pawel A. Penczek, Chao Yang, Joachim Frank, and Christian M.T. Spahn. Estimation of variance in single-particle reconstruction using the bootstrap technique. *J. Struct. Biol.*, 154(2):168–183, 2006.

[47] H. Liao and J. Frank. Classification by bootstrapping in single particle methods. In *Proc. ISBI*, pages 169–172. IEEE, April 2010.

[48] P. Penczek, M. Kimmel, and C. Spahn. Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *Structure*, 19(11):1582–1590, 2011.

[49] J. Andén, E. Katsevich, and A. Singer. Covariance estimation using conjugate gradient for 3D classification in CRYO-EM. In *Proc. ISBI*, pages 200–204, April 2015.

[50] Joakim Andén and Amit Singer. Structural Variability from Noisy Tomographic Projections. *SIAM Journal on Imaging Sciences*, 11(2):1441–1492, jan 2018.

[51] Michael Habeck. Bayesian modeling of biomolecular assemblies with cryo-em maps. *Frontiers in*

molecular biosciences, 4:15, 2017.