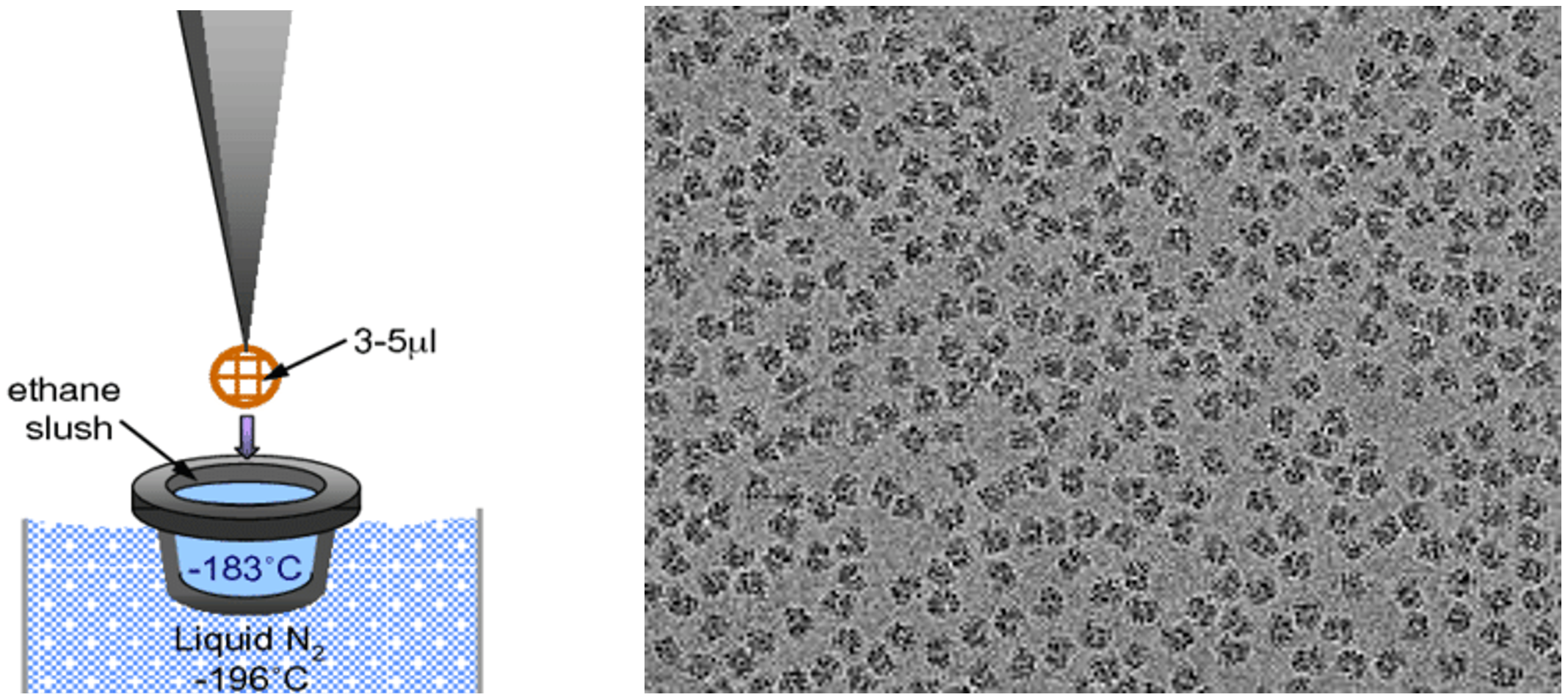
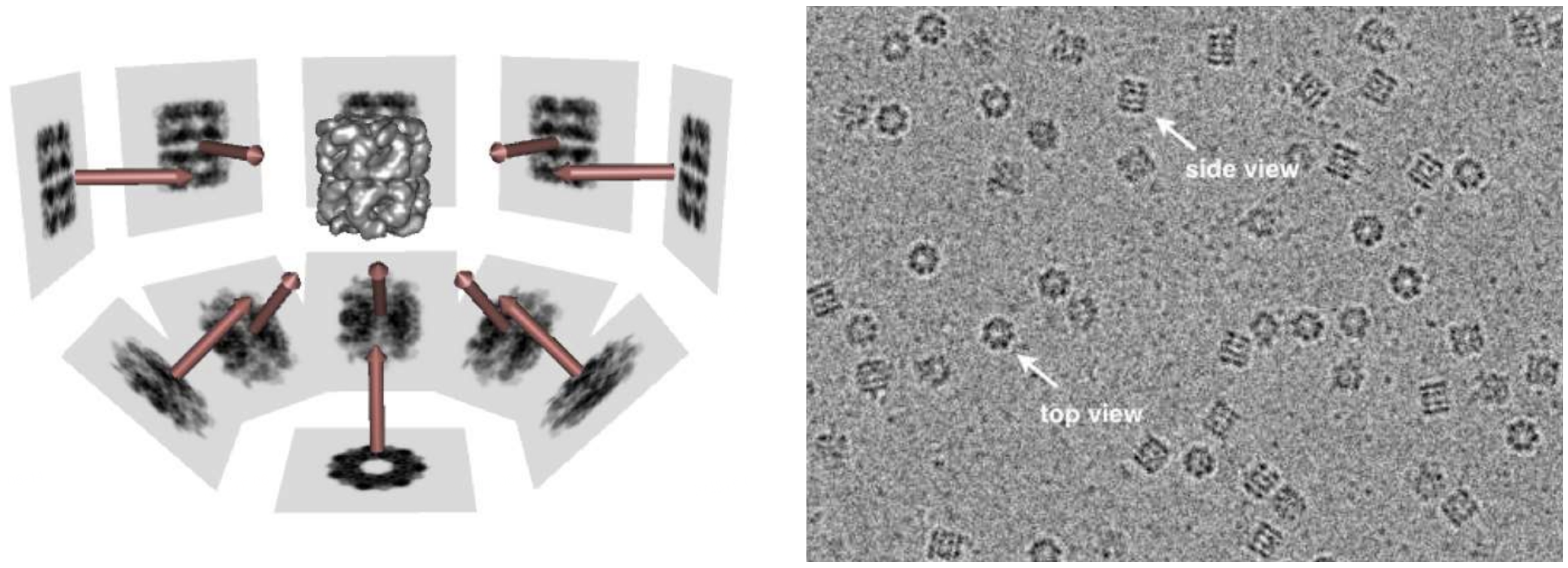
**冷冻电镜三维重构（生成模型）**

**问题背景：**冷冻电子显微镜技术是一种利用冷冻透射电子显微镜（cryo-EM）进行生物结构三维重建的技术，包括单颗粒分析和冷冻断层扫描。冷冻电镜因其提供的高分辨率、天然状态成像和分析，成为结构细胞生物学领域的一颗冉冉升起的新星。下图给出冷冻电镜的拍摄成像过程，通过拍摄被冷冻的生物结构样本（左图），获得溶液中众多三维生物结构的二维的投影（右图）：



**角度优化问题：**

有了三维生物结构的众多二维投影，我们需要协同地实现角度估计和三维重构。**Relion**基于最大似然估计的方法，结合贝叶斯推理来估计每个颗粒的角度。它通过优化似然函数，逐步提高对角度的估计。**CryoSPARC**则采用了基于快速非均匀傅里叶变换（NUFFT）的角度估计方法。这种方法引入了一种“分级、逐层精细的优化”策略，即通过一次性估计出近似的角度，随后在每个层次上对角度估计进行细化。在本次大作业中，我们尝试复现第一种方法。



**实验数据：**

给定1000张结构不变的蛋白质结构以及对应的参考poses

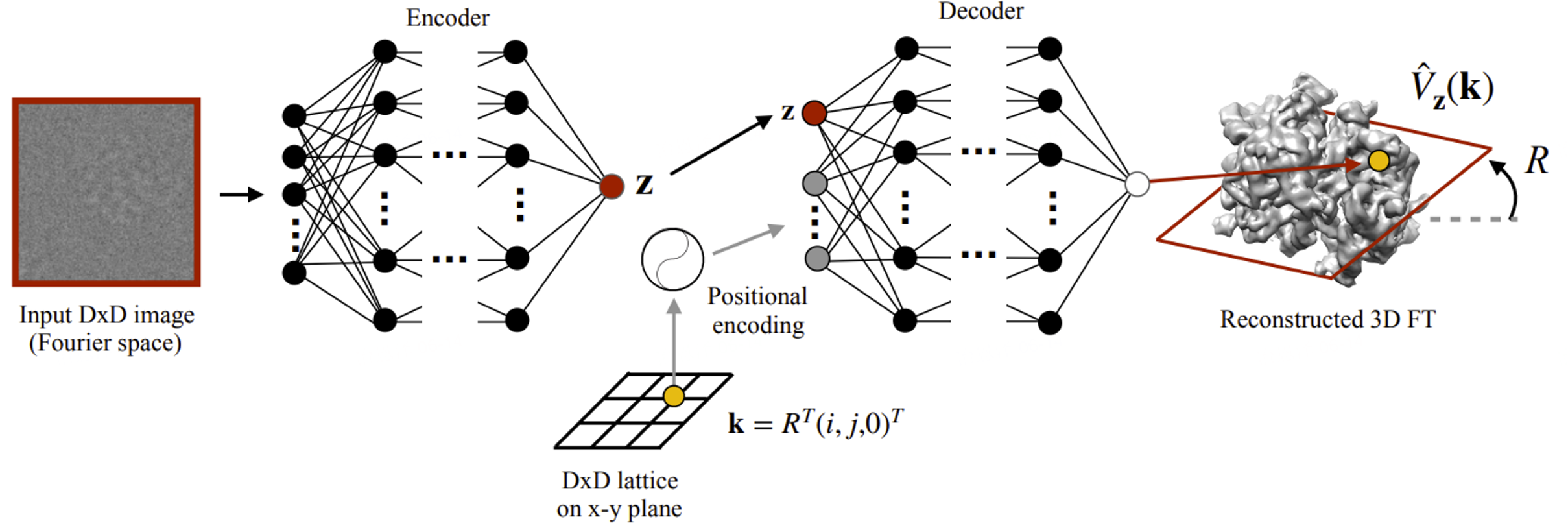
**程序1 (6’)：**编写程序实现基于傅里叶切片定理方法的三维重构算法，实现在给定角度下的三维重构。

**程序2 (10’)：**基于Expectation-Maximization(EM) 方法实现角度-3D重构的迭代优化过程，从随机初始化的角度的开始迭代优化角度，实现冷冻电镜图像在无已知角度情况下的角度估计和三维重构。

**实验1 (2’)：**将迭代分析的角度估计结果和角度参考值进行对比分析。

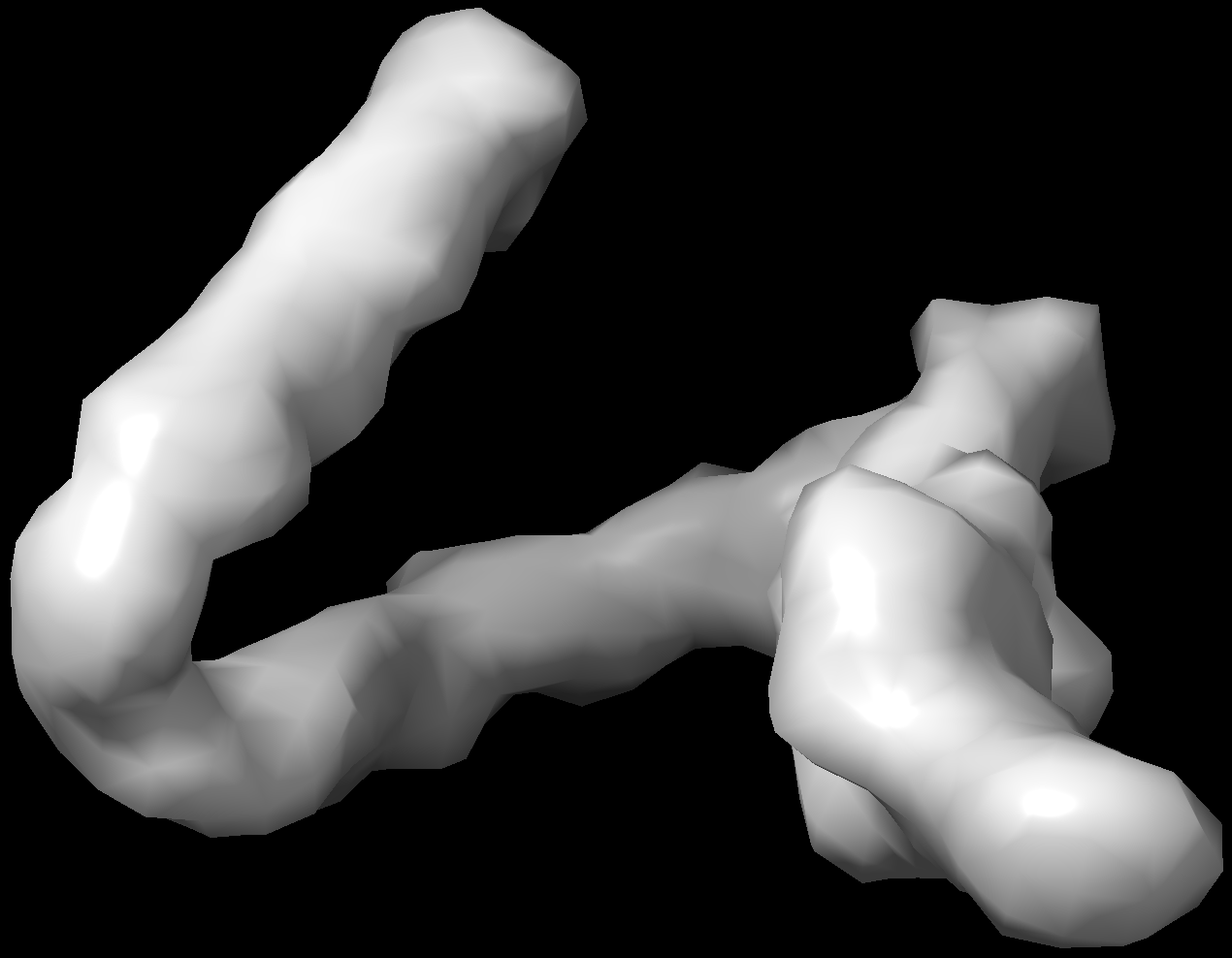
**生成模型问题：**

生成模型赋予强大的能力，使得单颗粒冷冻电镜三维重构方法可以被用来分析蛋白质连续的构象变化过程。下面我们考虑这一问题：给定已知角度的众多不同构象下的二维投影，设计基于深度神经网络的生成模型，实现从二维投影到三维生物结构的重构。这里我们根据VAE的框架进行实现[2]：将二维投影（Particle image）通过encoder得到一个隐变量（latent variable），然后与给定的pose组合，输入decoder得到最终三维结构的一个切片。



**实验数据：**

给定1000张蛋白质结构动态变化的二维投影以及对应的Poses，我们可以尝试重建蛋白质3D结构的变化过程（大致如下图所示）。



**程序1 (6’)：**编写Encoder与对应的Encode过程（注意参照VAE的范式）。将二维图像输入给Encoder，最终得到提取的特征隐变量z。

**程序2 (6’)：**编写decoder与对应的decode过程（注意参照Implicit Neural Representation的范式）。将pose信息与坐标结合起来，再拼接encoder的输出z，输入给decoder。

**程序3 (5’)：**设计合理的loss function，跑通二维投影-Encoder-Decoder-三维结构的模型，实现最终的三维重构。

**实验1 (5’):** 参考cryodrgn/commands/analyze.py部分的代码，直接使用analyze的脚本对重构结果进行分析，要求可视化动态的重建结果和隐空间变量z的分布图（利用PCA降维至二维）。

**理论作业**

Bayesian Estimator根据cost function的不同选择可以分为如下三种类型：

Minimum-Mean-Squared-Error (MMSE) Estimator

Minimum-Mean-Absolute-Error (MMAE) Estimator

Maximum A Posteriori Probability (MAP) Estimator

**作业1 (5’)：已知指数分布**

推导其MMSE, MMAE, MAP 三个Estimator的形式。

**作业2 (10’)：已知如下两种多元高斯分布（Multivariate Gaussian Distribution）**

分别推导其MMSE, MMAE, MAP 三个Estimator的形式。

**数据：**实验1给定1000张结构不变的蛋白质结构以及对应的参考poses，实验2给定1000张动态蛋白结构的二维投影数据和对应的poses。

**代码：**提供一份冷冻电镜三维重构代码框架作为参考，也可以直接填写空缺的部分。

**作业提交：**提交电子版报告，通过[北京大学教学网](course.pku.edu.cn)提交，或提交助教邮箱yez23@stu.pku.edu.cn。

**作业提交截止时间：**期末（2025.01.12晚上24:00）前。

**注：**报告不要求字数，鼓励精简，讲清楚所做内容即可。代码需要基于Python/Matlab等主流语言开发。**要求代码可复现报告内容，否则不得分。**