

# **Reconstructing 3D Protein Structures with Simulated Cryo-electron Microscopy Images**

**Benjamin Hoshal, Chris Ki,  
Junqi Qian, Xu Rao, Yandi Wu**

December 9th, 2016

# Contents

<b>1 Biological Overview of Proteins</b>	<b>4</b>
1.1 What are Proteins? . . . . .	4
1.2 DNA . . . . .	4
1.3 Central Dogma of Biology . . . . .	5
1.4 Transcription . . . . .	5
1.5 Translation . . . . .	6
1.6 Protein Structure . . . . .	7
1.6.1 Protein folding . . . . .	7
1.6.2 Levels of Protein Structure . . . . .	7
1.6.3 How is a protein's function related to a protein's structure? . . . . .	8
<b>2 3D protein structure determination</b>	<b>8</b>
2.1 NMR Spectroscopy . . . . .	9
2.2 X-Ray Crystallography . . . . .	9
2.2.1 History of X-Ray Crystallography . . . . .	9
2.2.2 What is X-Ray Crystallography? . . . . .	10
2.2.3 Diffraction and Bragg's Law . . . . .	10
2.2.4 Protein Crystallization . . . . .	12
2.2.5 Limitations of X-Ray Crystallography . . . . .	12
2.3 Electron Microscopy . . . . .	14
2.4 Importance of Cryo-EM . . . . .	14
2.4.1 What is Cryo-EM? . . . . .	14
2.4.2 Example of Cryo-EM impact . . . . .	15
<b>3 Mathematical Model</b>	<b>15</b>
3.1 Protein Structure . . . . .	16
3.2 Electron Microscopy (EM) . . . . .	16
<b>4 Simulating EM Images</b>	<b>16</b>
4.1 The Fourier Slice Theorem . . . . .	16
4.1.1 Fourier Series . . . . .	16
4.1.2 Fourier Transform . . . . .	18
4.1.3 Fourier Slice Theorem . . . . .	18
4.1.4 Poisson Summation Formula . . . . .	18
4.1.5 Poisson Summation Formula, Modified . . . . .	20
4.1.6 The Nyquist Criterion . . . . .	21
4.2 Interpolation . . . . .	22
<b>5 Back Projections</b>	<b>22</b>
<b>6 Estimating Orientations</b>	<b>26</b>

<b>7 The Contrast Transfer Function</b>	<b>29</b>
7.1 Transmission Electron Microscope . . . . .	29
7.1.1 Electron Gun and Condenser System . . . . .	30
7.1.2 Image-Producing System . . . . .	30
7.1.3 Image-recording System . . . . .	31
7.2 Fourier Optics . . . . .	31
7.3 Numerical Aperture . . . . .	31
7.4 Spherical Aberration . . . . .	32
7.5 The Point Spread Function . . . . .	32
7.6 Phase Contrast Transfer Function . . . . .	33
7.7 Difficulties in Reconstruction . . . . .	35
7.7.1 CTF Correction . . . . .	35
7.7.2 Wiener Filtration . . . . .	35
7.7.3 Envelope Functions . . . . .	36
<b>8 Results</b>	<b>37</b>
8.1 EM Simulation . . . . .	37
8.2 3D Structure Reconstruction . . . . .	37
8.3 Orientation Estimation . . . . .	39
<b>9 Code Listing</b>	<b>39</b>
9.1 Generating random rotation matrices: <code>randrot()</code> . . . . .	39
9.2 Simulating EM images: <code>project_fst()</code> . . . . .	40
9.3 Orientation estimation: <code>common_lines()</code> , <code>orientation_est()</code> . . . . .	41
9.4 Reconstructing 3D structure: <code>reconstruction</code> . . . . .	42
9.5 Function <code>apply_CTF()</code> . . . . .	44

# 1 Biological Overview of Proteins

## 1.1 What are Proteins?

**Proteins** are one of the four essential biological macromolecules (Proteins, Lipids, Nucleic Acids, and Carbohydrates). Proteins are large molecules that play integral roles in a myriad of biological processes. Proteins are responsible for coordinating all critical biological functions in the body. Their roles span from regulating the meticulous process of DNA transcription to the binding of oxygen in red blood cells. To understand exactly how proteins work, it is critical to comprehend how proteins are formed.

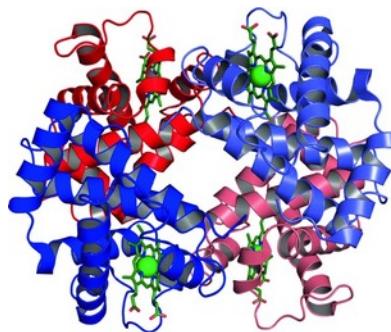


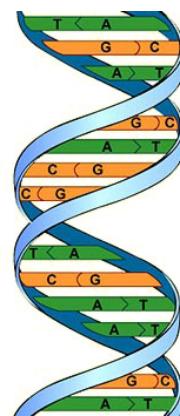
Figure 1: Hemoglobins are crucial proteins that help red blood cells bind to oxygen molecules. [[Hem](#)]

## 1.2 DNA

In living organisms, the structure and function of proteins are encoded in genetic material known as **DNA**, Deoxyribonucleic Acid.

Interestingly, DNA is also one of the four essential biomacromolecules (Nucleic Acid). DNA is a double-helix composed of different nucleotides each of which contains a nitrogenous base (Adenine, Guanine, Cytosine, and Thymine), a deoxyribose, and one or more phosphate groups. The deoxyribose of DNA molecules is critical to ensuring that DNA is synthesized in a 5' to 3' direction. These nucleotides are joined together by covalent bonds (known as phosphodiester bonds) between the 5' carbon of the sugar of one nucleotide and the 3'-hydroxyl (OH) of the adjacent nucleotide. Covalent bonds refer to chemical bonds that involve the sharing of electron pairs between atoms. These covalent bonds ultimately lead to the alternating sugar-phosphate backbone of DNA. On the other hand, the nitrogenous bases of DNA are held together by hydrogen bonds. These hydrogen bonds contribute to DNA's double-stranded structure as they help connect the individual nucleotides of each DNA strand. Hydrogen bonds are much weaker than covalent bonds as hydrogen bonding is a special case of dipole forces, the attractions between the positively and negatively charged ends of polar molecules. The sugar-phosphate

Figure 2: DNA molecules form the blueprint of life. DNA consists of 4 different types of nucleotides (Adenine, Guanine, Thymine, and Cytosine). [[dna](#)]



backbone of DNA is key to ensuring that the strands lay intact when enzymes such as helicase unwind the DNA molecule through cleavage of the hydrogen bonds. Furthermore, the two strands of DNA are *antiparallel*, as they run in opposite directions to each other. This feature of DNA is crucial to the transcription of DNA, which we will further elaborate on.

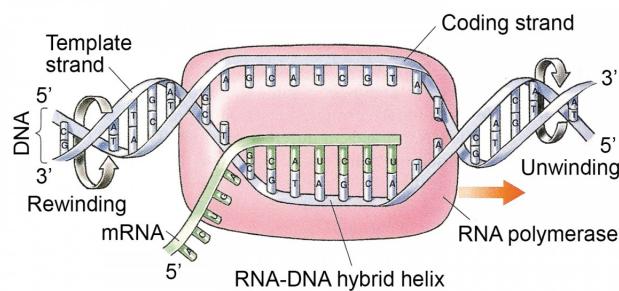
### 1.3 Central Dogma of Biology

The **central dogma of biology** explains how genetic information flows within an organism. Simply put, the central dogma states that DNA encodes RNA, which then encodes proteins. Transcription and translation are biological processes by which DNA is read to synthesize proteins. To see how the central dogma works, let us explore the first step and see exactly how DNA encodes RNA.

### 1.4 Transcription

During transcription, proteins known as **transcription factors** and **RNA Polymerase** bind DNA and reads its sequence to create a **messenger ribonucleic acid (mRNA)** molecule. There are three main steps to transcription: **initiation**, **elongation**, and **termination**. Transcription first begins with **initiation**, in which the DNA is unzipped by helicase, a special type of enzyme that separates two annealed DNA strands, to expose the now unpaired nucleotides.

RNA polymerase (with the help of other proteins, called basal transcription factors) then binds to a specific starting sequence of the desired gene. This sequence is called the promoter. Once RNA polymerase binds the promoter, initiation is complete. Subsequently, elongation takes place as the RNA polymerase uses the DNA as a template and runs down the strand to build an RNA copy of the DNA template strand. It is important to note that like DNA, RNA is composed of four nucleotides. However, RNA utilizes uracils instead of thymines. RNA also contains a ribose sugar instead of a deoxyribose sugar. This causes RNA to be far more unstable than DNA, as the 2' OH of the ribose makes it more susceptible to forming multiple phosphodiester bonds with other nucleotides. Once the mRNA is transcribed, a specific nucleotide sequence called the terminator signals RNA polymerase to release the DNA and terminate transcription. The resulting RNA molecule, mRNA, is then ready to be translated into a protein. [Suz]



*Figure 3: The elongation phase of elongation. The enzyme helicase catalyzes the unwinding of the DNA strands. [Cla]*

## 1.5 Translation

The final step in which proteins are synthesized is **translation**. Translation is facilitated by ribosomes, tiny organelles responsible for protein synthesis. The mRNA, produced by transcription, is "read" by the ribosome, which then links together nearby amino acids with the help of tRNAs to form polypeptides. Similar to transcription, translation also involves three steps: **Initiation**, **Elongation**, and **Termination**. Initiation begins as the ribosome wraps around the target mRNA at the start **codon**. **Codons** refer to a group of three bases in mRNA. Each codon refers to a specific amino acid. Interestingly, there are only 20 amino acids used in biological systems, compared to the 64 possible codons ( $4^3$ , as there are 4 possible nucleotides at 3 different positions). This abundance of codons leads to the **degeneracy** property of codons, which refers to the redundancy of the genetic code. This enables organisms to combat harmful mutations to their DNA. In initiation, the ribosome is also accompanied by tRNA, which ushers in amino acids that build upon a polypeptide chain. The ribosome can accommodate three tRNA molecules, which have a unique **anticodon** that base pairs with the relevant mRNA codon. The first amino acid that forms the polypeptide is indicated by the **start codon**. The start codon is responsible for notifying the ribosome to begin translation of the target mRNA strand. The start codon for proteins is usually AUG (Methionine). Initiation concludes with the establishment of the first amino acid corresponding to a start codon. In the elongation phase, the ribosome moves across the mRNA in a 5'-to-3' direction. tRNAs with corresponding amino acids are recruited to build upon the initial amino acid. These incoming amino acids are joined with the existing amino acids by the formation of peptide bonds. A **peptide bond** is formed by the loss of a water molecule in a reaction between the carboxyl group of one amino acid and the amino group of another amino acid. Termination occurs when a **stop codon** (UAA, UAG, or UGA) is reached. At this point, the ribosome dissociates from the target strand and the polypeptide is released. [Cla]

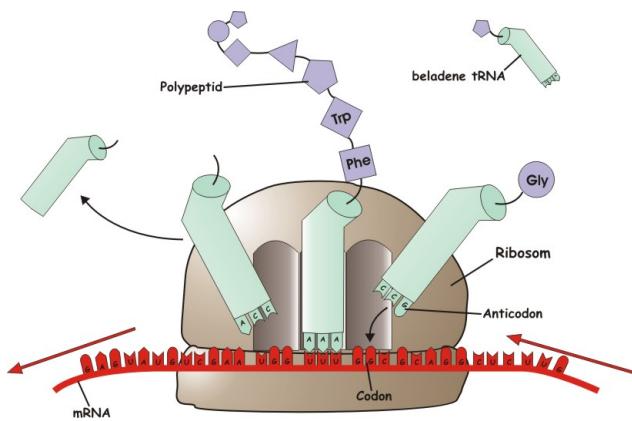


Figure 4: A ribosome reading an mRNA strand to produce a polypeptide chain. [tra]

## 1.6 Protein Structure

To understand how proteins can have such a robust variety of functions, we must analyze their structure.

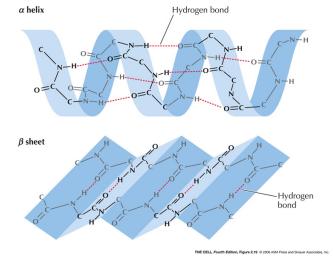
### 1.6.1 Protein folding

Once translation is completed, the polypeptide must undergo folding to obtain its native 3-dimensional structure. Folding is crucial for a protein to attain its true purpose. Failure to fold often results in the inactivity of the protein. For instance, sickle cell anemia, a type of inherited red blood cell disorder, is caused by incorrect folding of the beta-globulin of hemoglobin. To further understand protein folding, it is essential that we understand the complex levels of protein structure.

### 1.6.2 Levels of Protein Structure

Proteins have four levels of structure: **primary**, **secondary**, **tertiary**, and **quaternary**. We will employ the help of a critical oxygen-carrying protein in the blood, hemoglobin, to assist us in the explanation of these levels. The **primary** structure refers to the linear sequence of amino acids in the protein. Here, one can find the covalent bonds that hold the amino acids together, such as peptide bonds. The **primary** structure can be deduced from an analysis of the corresponding mRNA strand. One can thus understand the primary structure of hemoglobin by examining the amino acid sequence formed by translation of the mRNA encoding a globin subunit. The **secondary** structure describes the local structure of protein backbone. There are two main types of secondary structures: the  **$\alpha$ -helix** and the  **$\beta$ -sheets**.  $\alpha$ -helices occur in a right-handed spiral conformation while  $\beta$ -sheets look like pleated sheets. Hydrogen bonds, which we discussed in the DNA section, are responsible for defining the secondary structure. In our hemoglobin molecule, each globin subunit contains eight alpha helices. The **tertiary** structure refers to the 3-dimensional structure of the protein molecule that is obtained after the secondary structures of the proteins are arranged in a particular manner. Tertiary structure is greatly influenced by non-covalent interactions (hydrophobic interactions, hydrogen bonding, disulfide bonding, formation of salt bridges, etc.) By the laws of thermodynamics, proteins strive towards a stable structure. This is best achieved by minimizing the interactions between the hydrophobic (water-fearing) side chains of the amino acid backbone and the water molecules that surround the protein. The proteins thus fold in a manner such that the hydrophobic portions of the amino acids are concentrated in the center of the protein, hidden from the water molecules that interact with the external hydrophilic portions of the protein. This is seen in hemoglobin, as hydrogen bonds connect the  $\alpha$ -helices of the globin subunits. Also, covalent bonds between the globin subunits and histidine residues on the other subunits contribute to the tertiary structure. Lastly, the **quaternary** structure refers

Figure 5: There are two main types of secondary protein structures:  $\alpha$ -helices and  $\beta$ -sheets. /sec/



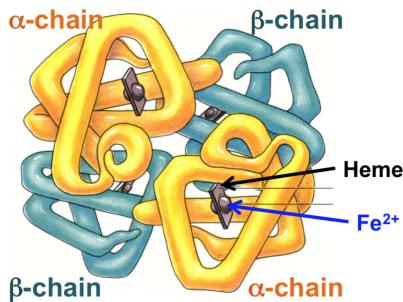


Figure 6: The quaternary structure of hemoglobin. [fou]

to the organization of folded protein subunit into complex protein structures. In our complete hemoglobin molecule, four globin proteins form hemoglobin by binding to heme and each other. These levels of organization are essential for determining the biological roles of the protein.

### 1.6.3 How is a protein's function related to a protein's structure?

Like all other proteins, the specific structure of hemoglobin is critical to its function. As an oxygen molecule binds to one of the globin molecules, the structure of the hemoglobin shifts slightly, making its affinity for oxygen even higher. Its affinity will continue to rise with each bound oxygen (this change in affinity due to the binding of a ligand to its receptor is called **cooperativity**). This is critical for the efficiency of oxygen carrying because hemoglobin's increasing affinity for oxygen ensures that all four available binding sites for oxygen are filled. Any deficiencies in the structure of hemoglobin will ultimately affect how hemoglobin functions. For instance, if the globin subunits don't associate properly and the quaternary structure is improperly formed, hemoglobin may face impairments in its ability to capture oxygen molecules. Thus, it is absolutely critical that we understand the structure of proteins so that we may truly comprehend how proteins function. [Li]

## 2 3D protein structure determination

The intrinsic link between a protein's structure and its function confers to us a powerful tool in the analysis of proteins. However, we cannot simply look into a light microscope to determine a protein's structure. Simple light microscopes give very limited information about the 3-dimensional structure of a protein that is essential to understanding how a protein works. Fortunately, modern science has offered a solution to this conundrum as there exist three common methods to reconstruct 3D protein structure: NMR spectroscopy, X-ray crystallography and electron microscopy. This paper aims to delve into the workings of the last method, electron microscopy. To understand the relevance and usefulness of electron microscopy, it's helpful to know what its complements in 3D protein structure determination, NMR spectroscopy and X-Ray Crystallography, exactly are.

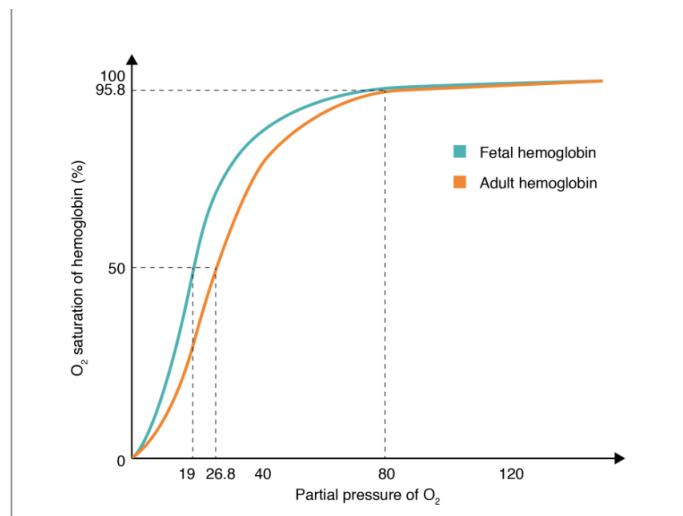


Figure 7: A graph depicting oxygen saturation of hemoglobin in response to varying partial pressures of oxygen. The sinusoidal shape indicates that hemoglobin's oxygen affinity increases as more oxygen binds. [Dis]

## 2.1 NMR Spectroscopy

**Nuclear Magnetic Resonance Spectroscopy**, otherwise known as NMR Spectroscopy, is one common technique that can be used to determine the structure of proteins. However, NMR spectroscopy is constrained to determining the structure of small proteins. It begins with the purification of the protein. The specimen is then dissolved in a buffer solution before it is placed in a very powerful magnetic field. In this magnetic field, radio waves are then transmitted through the sample. Subsequently, the resonance, the observed atom's absorption of these radio signals, of the atomic nuclei of the molecule is measured. Each atomic nuclei responds differently to a certain radio frequency signal based on the atom's environment within the protein. The absorption signals of certain nuclei can be disrupted by adjacent nuclei. Thus, it becomes possible to determine the distances between individual nuclei based on the perturbation of these absorption signals. The main measure of resonant frequency is known as the **chemical shift**. The chemical shift of a particular atom is determined by its unique electronic environment. With the observed resonances and chemical shifts, one can construct a complete structure of a protein by deducing the relative location of each atom.

## 2.2 X-Ray Crystallography

### 2.2.1 History of X-Ray Crystallography

Until the 20<sup>th</sup> century, scientists were unable to actually see the 3-dimensional structures of proteins. This changed when in 1912, Max von Laue surmised that x-rays could be used on crystalline structures to deduce the crystal's structure from the diffraction of the incident x-rays. His ideas were later confirmed by two scientists, Walter Friedrich and Paul Knipping, who successfully photographed the diffraction pattern resulting from the x-ray

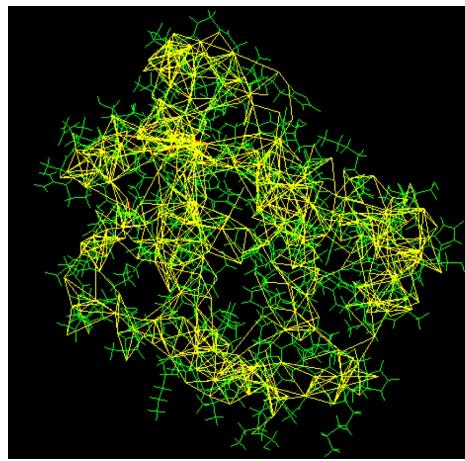


Figure 8: An elementary structure of a hemoglobin molecule formed with the results of NMR spectroscopy. [NMR]

radiation of  $CuSO_4 \cdot 5H_2O$ . However, X-Ray Crystallography wasn't applied to proteins until approximately 45 years after von Laue's discovery. Ever since then, X-ray crystallography has been used to determine countless proteins' 3D structure.

### 2.2.2 What is X-Ray Crystallography?

**X-ray Crystallography** is a very popular method used to determine structures of proteins and other biological macromolecules. This process involves focusing a beam of incident x-rays on a crystallized sample of proteins and then measuring the angles and intensities of the diffracted beams. With those recordings, one can create a 3D electron density photo of the crystal.

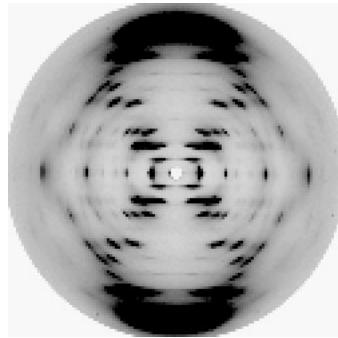


Figure 9: An X-Ray diffraction pattern of DNA obtained by Rosalind Franklin [XDi]

### 2.2.3 Diffraction and Bragg's Law

The power of X-ray Crystallography derives primarily from the phenomenon known as **diffraction**. Diffraction refers to what occurs when an electromagnetic wave encounters a slit or obstacle. Through diffraction, waves can either bounce off the obstacle or travel through to create a unique diffraction pattern. In our case with x-ray crystallography, the

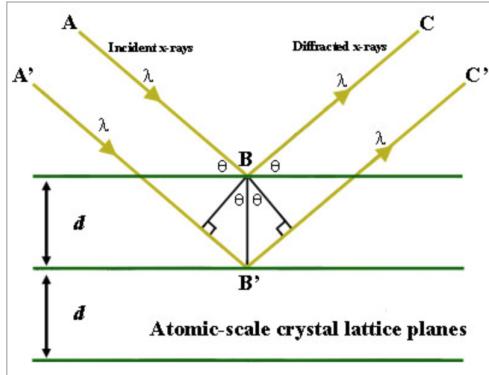
incident x-ray beam is focused on the crystalline sample of proteins. **Crystals** of proteins form regular arrays of atoms which reflect the incident x-ray beam and create a diffraction pattern. The crystal structure of our sample is absolutely necessary, as it minimizes noise and amplifies the intensities of the reflected x-rays. The scattering of the incident x-rays are caused by the electron clouds that circle each individual atom of the crystalline solid. However, knowing how diffraction functions in x-ray crystallography is not enough. We must also understand how one could measure the angles and intensities of the resulting diffraction patterns. Let's first start with understanding how to measure the angles of the scattered x-ray waves. When an X-ray beam hits the surface of a crystal, the diffracted beam can travel in 3 different paths:

1. It can diffract at the same angle away from the solid.
2. It can pass through the crystal.
3. It can diffract an angle  $\theta$  that is different from the incident angle.

This happens for many beams in many different planes of the crystal. After diffraction, the resulting beams will only constructively interfere (be in phase) if the path length difference before hitting the different planes in the crystal is equal to an integer number of wavelengths. **Bragg's Law** captures this relation:

$$\sin(\theta) = \frac{n\lambda}{2d} \quad (1)$$

where  $\lambda$  is the wavelength,  $n$  is an integer, while  $d$  is an interplanar distance between two planes. With Bragg's law, it is possible to measure the angles of the diffraction patterns. Lastly, the intensity of the reflected beams is equivalent to the vector sum of all the electron



*Figure 10: Diagram to clarify Bragg's law. Incident rays labeled A and A' are two beams with identical wavelength  $\lambda$ . The beams scatter off two different planes. The lower beam traverses an extra length of  $2ds\sin\theta$ . Constructive interference occurs when this length is equal to an integer multiple of the wavelength of the radiation.*

scattering caused by the incident beam of X-Ray. The intensity is related to the number of X-ray photons measured by constructive interference of the reflected beams in each diffraction spot. This intensity can be measured by the formula given by:

$$I = I_0 e^{-\mu t} \quad (2)$$

where  $I_0$  is the intensity of the incident beam,  $\mu$  is the linear absorption coefficient, and  $t$  is the path length through which the X-rays are moving. Note that the scattered electrons are the electrons of the protein, scattered by the incident x-ray beam. Electron scattering intensity can be characterized by the Thompson Equation:

$$I = \frac{I_0}{r^2} \left( \frac{e^2}{m_e c^2} \right)^2 \frac{1 + \cos^2(2\theta)}{2} \quad (3)$$

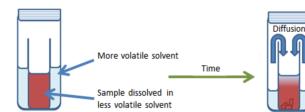
where  $r$ =distance from the electron to detector,  $2\theta$ =angle between scattering direction and incident beam,  $c$ =speed of light, and  $m_e$  = mass of an electron.

With knowledge of how to measure intensity along with Bragg's law, we can analyze the diffraction pattern produced to yield information about the protein's structure.[\[Lib\]](#)

#### 2.2.4 Protein Crystallization

Now we may know how the structure of the protein is obtained. However, it's essential that the protein of concern is crystallized. Protein crystallization is the process by which a crystal lattice of the pure protein is formed, a crucial step in X-ray crystallography. By forming a crystal lattice, the protein orients itself in the same position throughout the crystal. This is critical in deciphering the 3D structure of the protein. A **goniometer** positions the crystal at different orientations so that x-ray diffraction patterns may be analyzed from multiple angles so that the 3D structure can be resolved. There are several ways to crystallize a protein, the most common of which is **vapor diffusion**. During this process, a droplet containing a solution of purified protein and low concentrations of precipitant and buffers is placed into a larger vat of similar buffers at higher concentrations. As the droplet integrates with the larger solution, its surrounding precipitant and buffer concentrations increase via diffusion. This catalyzes the crystal formation in the drop. However, not every protein can be made into a crystal. In general, a protein must be 97 percent pure to form a crystal, and even this is not a sufficient condition. Further issues arise during attempts to form a crystal of a membrane bound protein, as a special solution called a **detergent** must be added to facilitate the crystallization process. Even then, these proteins tend to form a weakly maintained or x-ray sensitive crystal. In fact, it took over 60 years to crystallize the ribosome, which as mentioned above is critical in the synthesis of new proteins.

*Figure 11: A visualization of the vapor diffusion technique.*



#### 2.2.5 Limitations of X-Ray Crystallography

Unfortunately, as powerful as X-Ray crystallography is, there are still many challenges x-ray crystallography faces in deciphering the structures of proteins. In this section, we will discuss three major shortcomings of X-ray crystallography:

1. Difficulty of Crystallization
2. Phase Problem

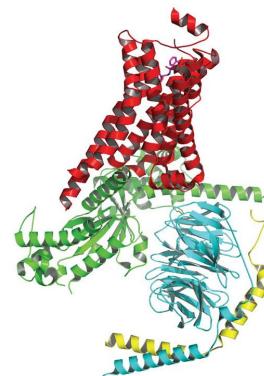
### 3. Restricted static examination of protein structure

The first challenge is found in the utmost important requirement of x-ray crystallography: crystallization of the specimen. Crystallization is an extremely difficult and painstaking process. It requires a huge period of time as the cloning and purification process of a protein with 99% purification and flawless execution requires a minimum of 3-6 months. Once the protein is purified to a high extent, one must then crystallize the protein which can take almost a whole year. Along with these time issues, there are technical difficulties that must also be considered. Proteins may be difficult to express in a huge amount and this poses a problem for x-ray crystallography as significantly large amounts of crystals are needed for accurate and clear scattering results. One may also face problems in the purification and precipitation of proteins. The detergent mentioned in the above section may even disrupt the crystal structure of the protein. X-ray radiation may even cause harm to the specimen and damage it which makes it a necessity to have a large amount of the protein. There are simply too many factors to consider in crystallization and this can ultimately render crystallization of certain proteins to be extremely costly.

For instance, researchers have struggled to procure a high-resolution image of the  $\beta_2$ -adrenergic receptor for over 50 years. Kobilka, the 2012 Nobel Prize winner in Chemistry, was responsible for discovering the structure of the  $\beta_2$ -adrenergic receptor, a neurotransmitter receptor involved in the fight or flight response, through x-ray crystallography. Understanding the structure of  $\beta_2$ -adrenergic receptors has major implications in the pharmaceutical and medical realm as these receptors are crucial in regulating stress and metabolic levels.  $\beta_2$ -adrenergic receptors are GPCRs (G-protein coupled receptors), which play major roles in signal transduction pathways. For years, scientists failed to crystallize the  $\beta_2$ -adrenergic receptor due to its low natural abundance, its structural flexibility, and instability in detergent solutions. Like most GPCRs,  $\beta_2$ -adrenergic receptors are expressed in low levels in native tissues and easily interchange among multiple conformational states. However, Kobilka and his team were only able to succeed after crystallizing the receptor in a lipid environment or modifying the protein's structures. For example, in one experiment, crystallization of the  $\beta_2$ -adrenergic receptor was facilitated by creating a fusion protein which a T4 lysozyme (a cell organelle that digests intracellular components) replaced a significant portion of the GPCR. Thus, this difficulty in crystallization of the  $\beta_2$ -adrenergic receptor truly shows how time-consuming and arduous the process of x-ray crystallography is.

Another issue of X-ray crystallography is known as the **phase problem**. In x-ray crystallography, the picture of the electron density can be obtained by taking the inverse Fourier transform of the diffraction pattern. This notion is reasonable as each individual diffraction spot of the pattern corresponds to a certain point on the crystal lattice of the protein. However, in order to take the inverse Fourier transform, we must have the amplitude and

Figure 12: The crystal structure of  $\beta_2$ -adrenergic receptor Kobilka discovered.



the relative phase of each diffraction spot or structure factor. The **phase problem** arises from the fact that we can only recover the amplitude. We can recover the amplitude as our measurement in x-ray crystallography gives us the count of X-ray photons in each diffraction spot. The number of the X-ray photons is related to the intensity which is proportional to the square of the amplitude of the diffracted wave. However, it is impossible to find the phase shift of each diffracted wave. This phase problem poses a major setback for x-ray crystallography as the exact structure can't ever be recorded.

The last limitation of X-ray crystallography is the rigidity of the crystalline protein structure. Since the proteins in a crystal form a repeating regular pattern, the proteins are locked in a single static position. This makes it very difficult to deduce the structure of a protein in different conformations. [Che] [Ras]

## 2.3 Electron Microscopy

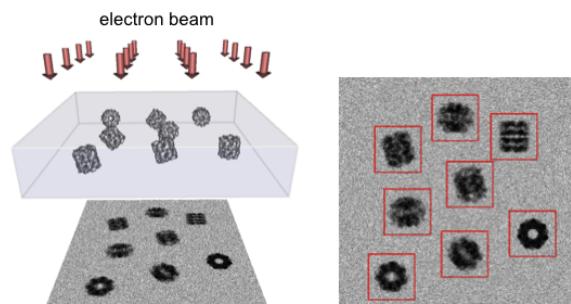
**Electron microscopy (EM)** is another extremely useful method for determining the 3D structure of a protein. EM also happens to be the major focus of our paper. EM differs from X-ray Crystallography in many ways. Most importantly, it employs a beam of electrons rather than X-rays to create an image of a macromolecule. Contrary to light microscopes, electron microscopes can attain much higher resolutions. Unlike X-ray crystallography, it's not necessary for the specimen to be crystallized. This allows electron microscopy to determine the structures of very small proteins as very small proteins are extremely challenging to crystallize. There is a multitude of preparation methods for a sample. Recently, the most popular preparation method has been cryofixation which has led to the development of an entirely new branch of EM called Cryo-Electron Microscopy. [Jac]

## 2.4 Importance of Cryo-EM

### 2.4.1 What is Cryo-EM?

**Cryo-Electron Microscopy** is a special type of Transmission Emission Microscopy (TEM) that involves the visualization of minuscule biological samples preserved in ice. **TEM** is a microscopy technique that is capable of producing very high resolution images through the transmission of electrons through very thin slices of specimen. Cryo-EM is a unique form of electron microscopy as it requires the the flash-freezing of proteins in solution. When cryo-EM first arrived in the 3D protein structure determination scene, it was far outmatched by x-ray crystallography and NMR as it only produced blurry pictures. Initially, cryoEM could only resolve images to a measly 10 Angstroms compared to the 4 Angstrom images of x-ray crystallography. However, recently, cryoEM has been rev-

Figure 13: Cryo-EM involves the freezing of proteins in random conformations. The image on the right shows the same proteins in different orientations.

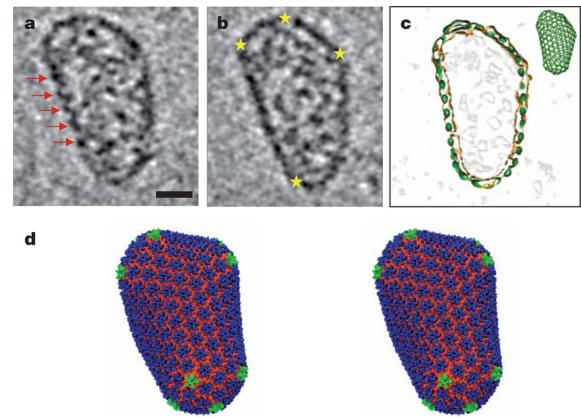


solutionizing the world of structural biology as it can now achieve high resolution images without experiencing the setbacks of its rivals. For instance, cryoEM can produce atomic-level images without facing the issue of crystallization. Plus, the lack of crystallization in cryoEM allows proteins to be examined in their native state in multiple conformations. Unlike x-ray crystallography, CryoEM is not restricted to the same static pose of each protein molecule. This allows scientists to learn more about the function of proteins. Lastly, cryoEM does not suffer from the phase problem as they produce real-space images containing both amplitude and phase information. Cryo-EM is able to do so as it preserves the phase shifts through the focusing of the diffracted electron beam into an image. This focusing is accomplished by addressing the lack of lens in x-ray crystallography. The presence of lens in cryo-EM is critical in producing the electron wavefunction immediately after transmission. We can then obtain the phase shifts and amplitudes to apply the inverse Fourier transform to procure a real-space image. [Calb]

#### 2.4.2 Example of Cryo-EM impact

Cryo-EM's emergence into the 3D protein structure determination scene is captured in the structural determination of a tubular HIV-1 capsid-protein. Although X-ray crystallography identified the structures of the capsid-protein assembly units, X-ray crystallography failed to capture structural knowledge related to the assembled capsid and contacts between the assembly units. X-ray crystallography could not resolve these structures *in vitro* as HIV-1 capsid assembly required high salt which would have interfered with the crystallization process. Furthermore, it's very difficult to replicate the contacts between the assembly units in a crystal lattice. Cryo-EM however did not face these problems as cryo-EM can accommodate for any conformation of the protein which allows for the direct examination of the CA assembly. It is crucial to understand the complete atomic model of the HIV-1 capsid model as it expands the studies of capsid function and broadens pharmacological approaches towards stopping the potency of HIV. Understanding the structure of HIV is crucial to knowing how to stop it from proliferating as researchers can work to prevent capsid assembly to defuse the danger of HIV. [Cala] [Zha]

*Figure 14: Cryo-EM played a vital role in identifying the atomic model and structure of the HIV-1 capsid model*



## 3 Mathematical Model

Given the problem, determination of 3D structure, in a biological context, we now develop a mathematical model for simulating EM images.

### 3.1 Protein Structure

We represent the protein by its electron density function which represents electronic potential to  $e^-$  in the protein.

$$\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$$

### 3.2 Electron Microscopy (EM)

#### Assumptions.

1. We assume that each copy of the same protein is identical.
2. Since each copy of the proteins is randomly embedded in ice, we further assume that by Electron Microscopy, the image of each single copy of the protein is equivalent to a projection of randomly rotated protein.

In sum, we assume that each image is a projection of a single protein from randomly different viewing directions.

**Definition 3.1. Orthogonal Matrix.** An orthogonal matrix is a square matrix with real entries whose columns and rows are orthogonal unit vectors, i.e.  $QQ^T = Q^TQ = I$ .

We use a 3-dimensional orthogonal matrix with determinant 1 to represent the rotation matrix. For a rotation matrix  $F$ , let  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  denote the 1st, 2nd and 3rd column of  $F$  (i.e.  $F = [\mathbf{a} \ \mathbf{b} \ \mathbf{c}]$ ). Then we have the following properties: (1)  $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} = 0$ ; (2)  $|\mathbf{a}| = |\mathbf{b}| = |\mathbf{c}| = 1$ .

From the assumption, we know that each image is corresponding to some rotation matrix  $F$ , such that  $\mathbf{a}$  and  $\mathbf{b}$  is the axis of the image and  $\mathbf{c}$  is the viewing direction. The image we obtained from microscopy in direction  $F$  is a projection of  $\rho$  on to the image plane  $\text{span}\{\vec{a}, \vec{b}\}$ . Denote the EM image of  $\rho$  from direction  $F$  as a function  $I_F : \mathbb{R}^2 \rightarrow \mathbb{R}$ :

$$I_F(x, y) = \int_{-\infty}^{\infty} \rho(\mathbf{a}x + \mathbf{b}y + \mathbf{c}z) dz$$

where  $x, y, z$  are the indices of the input  $\rho$  and  $I_F$  is the line integral of electron density along the line perpendicular to the image plane.

## 4 Simulating EM Images

### 4.1 The Fourier Slice Theorem

#### 4.1.1 Fourier Series

Before introducing the Fourier Transform, we look at Fourier series, which are decompositions of any periodic functions into a series of sine and cosine functions. More precisely, given  $f$ , any sufficiently "well-behaved" periodic function with period  $T$ , we can write  $f$  as:

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos\left(nx \frac{2\pi}{T}\right) + b_n \sin\left(nx \frac{2\pi}{T}\right)$$

where  $a_0$ ,  $b_n$ , and  $a_n \in \mathbb{R}$  denote the Fourier coefficients and are given by the following formulae:

$$\begin{aligned} a_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) \cos\left(nx \frac{2\pi}{T}\right) dx \\ b_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) \sin\left(nx \frac{2\pi}{T}\right) dx \\ a_0 &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) dx \end{aligned}$$

For example, we can find the Fourier series of  $f(x) = x^2$ . First, notice that  $x^2$  is an even function, as is sine, so  $b_n = 0$  since  $x^2$  cannot have any odd components. Then it suffices to find  $a_0$  and  $a_n$ :

$$\begin{aligned} a_n &= \frac{2}{2\pi} \int_{-\pi}^{\pi} x^2 \cos\left(nx \frac{2\pi}{2\pi}\right) dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} x^2 \cos(nx) dx \\ &= - \int_{-\pi}^{\pi} \frac{\sin(nx)}{n} 2x dx + \frac{x^2 \sin(nx)}{n} \quad (\text{Integration by Parts}) \end{aligned}$$

$$\begin{aligned} &= \frac{x^2 \sin(nx)}{n} + \frac{2x \cos(nx)}{n^2} - \frac{2 \sin(nx)}{n^3} \Big|_{-\pi}^{\pi} \\ &= \frac{4\pi \cos(n\pi)}{n^2} \\ &= \frac{2(-1)^n}{n^2} \end{aligned}$$

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} x^2 dx \\ &= \frac{1}{2\pi} \left( \frac{x^3}{3} \right) \Big|_{-\pi}^{\pi} \\ &= \frac{\pi^2}{3} \end{aligned}$$

Then our Fourier series representation becomes:

$$x^2 = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos(nx)$$

We can use this to show that:

$$\begin{aligned} 0 &= \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \\ -\frac{\pi^2}{3} &= 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \\ -\frac{\pi^2}{12} &= \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \end{aligned}$$

#### 4.1.2 Fourier Transform

For a function  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ , the Fourier Transform of  $f$ , denote  $\mathcal{F}\{f\}$ , or  $\hat{f}$ , is defined:

$$\hat{f}(\vec{\omega}) := \iiint_{-\infty}^{\infty} f(\vec{x}) \exp(-2\pi i \vec{x} \cdot \vec{\omega}) d\vec{x}, \quad (4)$$

where  $\vec{x}, \vec{\omega} \in \mathbb{R}^3$ . Assuming that  $f$  is well-behaved,  $f$  also has an inverse Fourier Transform,  $\mathcal{F}^{-1} : \mathbb{C}^3 \rightarrow \mathbb{R}$ :

$$\mathcal{F}^{-1}(\hat{f}(\omega)) := \int_{\mathbb{C}^3} f(\vec{\omega}) \exp(-2\pi i \vec{\omega} \cdot \vec{\omega}) d\vec{\omega}. \quad (5)$$

#### 4.1.3 Fourier Slice Theorem

Recall that  $I_F$  is the projection of  $\rho$ , an electric potential, onto a plane spanned by orthonormal vectors  $\vec{a}$  and  $\vec{b}$  in the viewing direction  $\vec{c}$  perpendicular to the image plane,  $\vec{a}, \vec{b}$ , and  $\vec{c}$  the column space of a rotation matrix.

**Theorem.** Suppose  $\rho$  is a function whose Fourier Transform is  $\mathcal{F}$ . Then:

$$\mathcal{F}^{-1}\{\mathcal{F}\{\rho_{\uparrow \text{span}\{\vec{a}, \vec{b}\}}\}\} = I_F \quad (6)$$

(7) is known as the Fourier Slice Theorem, and is a powerful tool that allows us to reconstruct images from a set of projections in lower dimensions.

In particular, the Fourier Transform of  $\rho$  restricted onto the span of  $\vec{a}$  and  $\vec{b}$ , is equivalent to its 2-D projection,  $I_F$ :

$$\hat{\rho}_{\uparrow \text{span}\{\vec{a}, \vec{b}\}}(\vec{\omega}) = \hat{I}_F(\omega_x, \omega_y) \quad (7)$$

#### 4.1.4 Poisson Summation Formula

The Poisson Summation Formula serves as a bridge between the Fourier Transform and Fourier Series. More specifically, the PST connects the abstract world of infinite, continuous integrals calculated over the entire real line to the applicable world of finite integrals, which can be approximated through finite sums over finite intervals. In this section, we derive the three-dimensional version of PST.

Recall the electron density function  $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$  where:

$$\rho(\vec{\omega}) = \int_{\mathbb{R}^3} \rho(\vec{x}) \exp(2\pi \vec{x} \cdot i\vec{\omega}) d\vec{\omega},$$

denote  $\vec{\omega} := (\omega_1, \omega_2, \omega_3)$  and  $\vec{x} := (x_1, x_2, x_3)$ .

Suppose  $\rho$  has period  $N$ , of in other words,

$$\hat{\rho}_N(\omega_i) = \sum_{m_i=-\infty}^{\infty} \hat{\rho}(\omega_i + m_i N)$$

for  $i \in \{1, 2, 3\}$ .

Then we have:

$$\hat{\rho}(\vec{\omega}) = \sum_{\vec{k}=-\infty}^{\infty} A_{\vec{k}} \exp(-2\pi i \frac{\vec{k} \cdot \vec{\omega}}{N}) \quad (8)$$

for  $\vec{k} \in \mathbb{R}^3$  and  $A_{\vec{k}}$  the Fourier coefficients.

Recall the equation for Fourier coefficients:

$$\begin{aligned} A_{\vec{k}} &= \frac{1}{N} \iiint_{-\frac{N}{2}}^{\frac{N}{2}} \hat{\rho}_N(\vec{x}) \exp(2\pi i \frac{\vec{k} \cdot \vec{x}}{N}) d\vec{x} \\ &= \frac{1}{N^3} \int_{[-\frac{N}{2}, \frac{N}{2}]^3} \sum_{\vec{m}=-\infty}^{\infty} \hat{\rho}(\vec{\omega} + \vec{m}N) \exp(2\pi i \frac{\vec{x} \cdot \vec{k}}{N}) d\vec{x} \\ &= \frac{1}{N^3} \sum_{\vec{m}=-\infty}^{\infty} \int_{[-\frac{N}{2}, \frac{N}{2}]^3} \hat{\rho}(\vec{\omega} + \vec{m}N) \exp(2\pi i \frac{\vec{x} \cdot \vec{k}}{N}) d\vec{x} \\ &= \frac{1}{N^3} \iiint_{-\infty}^{\infty} \hat{\rho}(\vec{\omega}) \exp(2\pi i \frac{\vec{x} \cdot \vec{k}}{N}) d\vec{x} \\ &= \frac{1}{N^3} \rho\left(\frac{\vec{k}}{N}\right) \end{aligned} \quad (9)$$

for  $i \in \{1, 2, 3\}$ .

Suppose  $\vec{k} = (s, r, q)$ . Then by (1), we have:

$$A_{\vec{k}} = \frac{1}{N^3} \rho\left(\frac{s}{N}, \frac{r}{N}, \frac{q}{N}\right).$$

We combine (1) and (2) to obtain the following derivation:

$$\begin{aligned} \hat{\rho}_N(\omega_1, \omega_2, \omega_3) &= \sum_{s=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} (A_s, A_r, A_q) \exp(-2\pi i \frac{\omega_1 s + \omega_2 r + \omega_3 q}{N}) \\ &= \sum_{s,r,q=-\infty}^{\infty} \frac{1}{N^3} \rho\left(\frac{s}{N}, \frac{q}{N}, \frac{r}{N}\right) \exp(-2\pi i \frac{\omega_1 s + \omega_2 r + \omega_3 q}{N}) \end{aligned} \quad (10)$$

(3) implies that:

$$\hat{\rho}(\vec{\omega} + \vec{m}N) = \frac{1}{N^3} \sum_{s,r,q=-\infty}^{\infty} \rho\left(\frac{s}{N}, \frac{q}{N}, \frac{r}{N}\right) \exp\left(-2\pi i \frac{\omega_1 s + \omega_2 r + \omega_3 q}{N}\right)$$

Finally, we obtain the 3-D Poisson Summation Formula:

$$\hat{\rho}_N(\vec{\omega}) = \frac{1}{N^3} \sum_{s,r,q=-\infty}^{\infty} \rho\left(\frac{s}{N}, \frac{q}{N}, \frac{r}{N}\right) \exp\left(-2\pi i \frac{\omega_1 s + \omega_2 r + \omega_3 q}{N}\right) \quad (11)$$

#### 4.1.5 Poisson Summation Formula, Modified

Suppose we want to find the relationship between fourier transforms of two functions:

$$r(x, y, z) \text{ and } \rho\left(\frac{-L}{2} + xL, \frac{L}{2} - yL, \frac{z}{L} + zL\right)$$

(Note this is a very practical application since sample sets of dimension  $L^3$  are often centered at  $(0, 0, 0)$ .) We apply the change of variables technique to find the relationship between  $\hat{f}$  and  $\hat{\rho}$ , where the conversion formulas are as follows:

$$\begin{cases} x' = \frac{-L}{2} + xL \\ y' = \frac{L}{2} - yL \\ z' = \frac{L}{2} + zL \end{cases} \quad (12)$$

To find the Jacobian, we take:

$$\frac{\partial(x, y, z)}{\partial(x', y', z')} = \begin{vmatrix} \frac{1}{L} & 0 & 0 \\ 0 & \frac{1}{-L} & 0 \\ 0 & 0 & \frac{1}{L} \end{vmatrix} = \frac{1}{-L^3}$$

All that's left to do is manipulate the change of variables equation:

$$\begin{aligned} \hat{r}(\omega_x, \omega_y, \omega_z) &= \int_{\mathbb{R}^3} r(x, y, z) \exp(-2\pi i(x, y, z) \cdot (\omega_x, \omega_y, \omega_z)) dx dy dz \\ &= \hat{\rho}(\omega_x, \omega_y, \omega_z) \\ &= \frac{1}{-L^3} \int_{\mathbb{R}^3} \rho(x', y', z') \exp(-2\pi i(x', y', z') \cdot (\omega_x, \omega_y, \omega_z)) dx' dy' dz' \\ &= \frac{1}{-L^3} \int_{\mathbb{R}^3} \rho(x', y', z') \exp\left(-2\pi i\left(\frac{x' + \frac{L}{2}}{L}, \frac{-y' + \frac{L}{2}}{L}, \frac{z' - \frac{L}{2}}{L}\right) \cdot (\omega_x, \omega_y, \omega_z)\right) dx' dy' dz' \\ &= \frac{1}{-L^3} \int_{\mathbb{R}^3} \rho(x', y', z') \exp\left(-2\pi i\left(\frac{x}{L}, \frac{y}{L}, \frac{z}{L}\right) \cdot (\omega_x, \omega_y, \omega_z)\right) dx' dy' dz' \end{aligned} \quad (13)$$

Observe From (10), we obtain a formula relating  $\hat{r}(\omega_x, \omega_y, \omega_z)$  and  $\hat{\rho}(\omega_{x'}, \omega_{y'}, \omega_{z'})$ :

$$\hat{r}(\omega_x, \omega_y, \omega_z) = \frac{-\exp(-\pi i(\omega_x, \omega_y, \omega_z))N^3}{-L^3} \sum_{-\infty}^{\infty} \hat{\rho}\left(\frac{\omega_{x'} + uN}{L}, \frac{\omega_{y'} + vN}{L}, \frac{\omega_{z'} + wN}{L}\right) \quad (14)$$

where  $u, v, w \in \mathbb{Z}$  and represent shifting in frequency space. This modified and shifted version of the PSF was ultimately the one used in our code.

#### 4.1.6 The Nyquist Criterion

Recall a function  $g(\vec{x})$  has bounded support if  $g(\vec{x}) = 0$  if  $|x_i| > B_i$  for  $0 \leq i \leq n$ ,  $\vec{x} \in \mathbb{R}^n$ .

We revist the Poisson Summation Theorem and note that in the special case  $f$  is supported on  $[0,1)$ , we obtain:

$$\begin{aligned} \sum_{s,r,q=-\infty}^{\infty} \hat{f}(\vec{\omega} \cdot (sN, rN, qN)) &= \frac{1}{N^3} \sum_{s,r,q=0}^{N-1} f\left(\frac{s}{N}, \frac{r}{N}, \frac{q}{N}\right) \exp\left(-2\pi i \frac{(s, r, q) \cdot \vec{\omega}}{N}\right) \\ &= \hat{f}(\vec{\omega}) + \sum_{s,r,q=1}^{N-1} f\left(\frac{s}{N}, \frac{r}{N}, \frac{q}{N}\right) \exp\left(-2\pi i \frac{(s, r, q) \cdot \vec{\omega}}{N}\right) \end{aligned}$$

Since  $\hat{f}$  is the true Fourier Transform, the term  $\sum_{s,r,q=1}^{N-1} f\left(\frac{s}{N}, \frac{r}{N}, \frac{q}{N}\right) \exp\left(-2\pi i \frac{(s, r, q) \cdot \vec{\omega}}{N}\right)$  denotes the aliasing error.

Naturally, the question of how to minimize, or better yet eliminate, aliasing error arises. We thus change  $N$ , the sample rate (in samples/sec), accordingly.

**Theorem.** Suppose  $f$  is a square integrable, or  $\int_{-\infty}^{\infty} |f(x)|dx$  is finite. Let its Fourier Transform, denote  $\hat{f}$ , have bounded support  $[-\frac{L}{2}, \frac{L}{2}]^3$ . Then  $f$  can be reconstructed from samples  $\{f(\frac{2\pi}{L}s), f(\frac{2\pi}{L}r), f(\frac{2\pi}{L}q) : (s, r, q) \in \mathbb{Z}^3\}$ . Furthermore, to prevent aliasing, the sampling rate must be twice the maximum frequency of  $f$  (the Nyquist Criterion).

If the sampling rate in samples per second,  $N < 2(\frac{L}{2\pi}) = \frac{L}{\pi}$ , there will be a discrepancy, aliasing, between the reconstructed function and the original function.

To summarize, for the purposes of electron microscopy experiments, we only examine the discrete case of the PST. Suppose  $\rho$  is an  $N \times N \times N$  array of  $N^3$  elements. Let DFT denote the Discrete Fourier Transform of  $\rho$ . Then we have:

$$DFT(\rho, \vec{\omega}) = \sum_{s,r,q=0}^{N-1} \rho[s, r, q] \exp\left(-2\pi i \frac{(q, r, s) \cdot (\omega_1, \omega_2, \omega_3)}{N}\right)$$

For the purposes of electron microscopy experiments, we only examine the discrete case of the PST (see Eq 6, 3.1.3). Suppose  $\rho$  is an  $N \times N \times N$  array of  $N^3$  elements. Let DFT

denote the Discrete Fourier Transform of  $\rho$ . Then we have:

$$DFT(\rho, \vec{\omega}) = \sum_{s,r,q=0}^{N-1} \rho[s, r, q] \exp\left(-2\pi i \frac{(q, r, s) \cdot (\omega_1, \omega_2, \omega_3)}{N}\right)$$

## 4.2 Interpolation

Roughly speaking, interpolation involves calculating values between known datasets surrounding the values. In particular, datasets are often collected as discrete points for the Discrete Fourier Transform, so we need to find  $f(\omega)$ , where  $\omega \notin \left[\frac{k}{N}\right]$ , where  $k = 1, 2, \dots, N - 1$ . One widely-used method is linear interpolation, where the formula is given as follows:

$$f(\omega) \approx \alpha f\left(\frac{k-1}{N}\right) + \beta f\left(\frac{k}{N}\right)$$

where  $k - 1 < \omega < k$ :

$$\begin{aligned} \alpha &= \left(1 - \frac{(\omega - \frac{k-1}{N})}{\frac{1}{N}}\right) \\ \beta &= \left(1 - \frac{(\omega - \frac{k}{N})}{\frac{1}{N}}\right) \end{aligned}$$

## 5 Back Projections

Electron microscopes record projection images of a specimen tilted at various angles. Back projection is a method of recovering a 3D volume,  $\rho$ , from its 2D projections,  $I_F$ . The final volume is a summation of back projection profiles, or a series of stacked thin slices. Before we talk about back projections, we need to state a few definitions.

**Definition.** Let  $S$  represent a linear operator. Suppose  $f, g$  are two functions. Then:

$$S(af + bh) = aS(f) + bS(g).$$

**Definition.** Let  $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$  where  $|\int_{\mathbb{R}^n} f(x)dx| < \infty$  and  $|\int_{\mathbb{R}^n} g(x)dx| < \infty$ . Then the convolution of  $f$  and  $g$ , denote  $f * g$ , is defined as:

$$(f * g)(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx. \quad (15)$$

**Theorem (Convolution Theorem)** The Fourier transform of a convolution of  $f$  and  $g$  is the product of the Fourier transforms of  $f$  and  $g$ , or:

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\}\mathcal{F}\{g\}$$

*Proof.* Without loss of generality, it suffices to prove that the convolution theorem holds for one dimension.

$$\begin{aligned}
 \mathcal{F}\{f * g\} &= \mathcal{F}\left\{\int_{-\infty}^{\infty} f(x)g(z-x)dx\right\} \\
 &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} f(x)g(z-x)dx \right] \exp(-2\pi i \omega \cdot z) dz \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)g(z-x) \exp(-2\pi i \omega \cdot z) dx dz \quad (\text{Fubini's Theorem}) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)g(y) \exp(-2\pi i \omega \cdot (x+y)) dx dy \quad (y = z-x, \text{ linear change of variables}) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \exp(-2\pi i \omega x) g(y) \exp(-2\pi i \omega y) dx dy \\
 &= \int_{-\infty}^{\infty} f(x) \exp(-2\pi i \omega x) dx \int_{-\infty}^{\infty} g(y) \exp(-2\pi i \omega y) dy \\
 &= \mathcal{F}\{f\} \mathcal{F}\{g\}
 \end{aligned}$$

□

The Convolution Theorem is used to cut down on runtime in calculating Fourier transforms; the runtime for calculating the convolution using the convolution theorem and FFT is  $\mathcal{O}(N \log N)$ . This is a great improvement from the runtime for computing the convolution directly,  $\mathcal{O}(N^2)$ . For example, take  $N=200$ . Then  $N \log(N) \approx 460$ , but  $N^2 = 40000$ . This is difference of two orders of magnitude.

**Definition.**  $\delta(x, y, z)$  is the dirac delta function that satisfies:

1.  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y, z) dx dy dz = 1$  and
2.  $\delta(x, y, z) = 0 \forall x, y, z \neq 0$ .

Specifically, the dirac delta has the property that:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z) \delta(x-a, y-b, z-c) dx dy dz = f(a, b, c) \tag{16}$$

Knowing (10), consider a decomposition of a function  $\rho$  into delta functions:

$$\rho(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x', y', z') \delta(x-x', y-y', z-z') dx' dy' dz'$$

Consider the linear operator  $S$ , and observe:

$$S(\rho)(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x', y', z') S(\delta(x-x', y-y', z-z')) dx' dy' dz'$$

We call  $h(x, y, z, x', y', z') = S\delta(x-x', y-y', z-z')$  the transfer function. Suppose  $h$  only depends on  $x-x'$ ,  $y-y'$ , and  $z-z'$  (or,  $h$  is shift invariant). Note this is a reasonable

assumption since transfer properties of a good electron transmission line should not depend on time, but rather only on the difference between times. Then we have:

$$S(\rho)(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x', y', z') h(x - x', y - y', z - z') dx' dy' dz' \quad (17)$$

Then  $S(\rho)(x, y, z) = \rho * h$ .

Applying the Convolution Theorem, we have:

$$\begin{aligned} \mathcal{F}\{S(\rho)\} &= \mathcal{F}\{\rho * h\} \\ &= \mathcal{F}\{\rho\} \mathcal{F}\{h\} \end{aligned}$$

Then  $\mathcal{F}\{\rho\} = \frac{\mathcal{F}\{S(\rho)\}}{\mathcal{F}\{h\}}$ , and finally:

$$\rho = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{S(\rho)\}}{\mathcal{F}\{h\}} \right\} \quad (18)$$

**Definition.** We denote the rectangular function:

$$\text{rect}(z) = \begin{cases} 1, & \text{if } -D \leq z \leq D \\ 0, & \text{otherwise} \end{cases}$$

where  $D$  is the diameter of  $\rho$ .

We now have the background to begin our discussion on back projections.

**Definition.** If  $I_F$  is the image obtained from the rotation matrix:

$$F = \begin{bmatrix} | & | & | \\ \vec{a} & \vec{b} & \vec{c} \\ | & | & | \end{bmatrix}$$

then the back projection of  $I_F$  is the function:

$$b(x\vec{a} + y\vec{b} + z\vec{c}) = I_F(x, y) * (\text{rect}(z)\delta(x, y)), \quad (19)$$

Suppose we have a collection of images,  $I_1, I_2, I_3, \dots, I_N$  corresponding to the rotation matrices  $F_1, F_2, F_3, \dots, F_N$ , and:

$$F_j = \begin{bmatrix} | & | & | \\ \vec{a}_j & \vec{b}_j & \vec{c}_j \\ | & | & | \end{bmatrix}$$

for  $1 \leq j \leq N$ . Note that  $a_j, b_j, c_j$  is a basis for  $\rho$ , so we can write:

$$\rho = x^j a_j + y^j b_j + z^j c_j$$

We recall (13) and obtain that the back projection of  $I_j$ , denote  $b_{\rho_j}$ , is:

$$b_{\rho_j} = I_j(x^j, y^j) * l_j$$

Recall that a reconstruction is the summation of a collection of back projections at various rotations, or:

$$\begin{aligned}
 b_\rho &= \sum_{j=1}^N b_{\rho_j}(x^j, y^j, z^j) \\
 &= I_j(x^j, y^j) * l_j \\
 &= \sum_{j=1}^N \mathcal{F}^{-1}\{\mathcal{F}\{I_j(x^j, y^j) * l_j\}\} \\
 &= \sum_{j=1}^N \mathcal{F}^{-1}\{\mathcal{F}\{I_j(x^j, y^j)\}\mathcal{F}\{l_j\}\}
 \end{aligned} \tag{20}$$

We calculate, in particular,  $\mathcal{F}\{l_j\}$ :

$$\begin{aligned}
 \mathcal{F}\{l_j\} &= \sum_{j=1}^N \int_{-D}^D \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x^j, y^j) rect(z^j) \exp(-2\pi i)(\omega_{x^j} x^j + \omega_{y^j} y^j + \omega_{z^j} z^j) dx^j dy^j dz^j \\
 &= \sum_{j=1}^N \int_{-D}^D rect(z^j) dz^j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x^j, y^j) \exp(-2\pi i(\omega_{x^j}, \omega_{y^j}, \omega_{z^j}) \cdot (x^j, y^j, z^j)) dx^j dy^j dz^j \\
 &= \sum_{j=1}^N \int_{-D}^D 1 \cdot \exp(-2\pi i(\omega_{z^j} z^j)) dz^j \\
 &= \sum_{j=1}^N \int_{-D}^D \cos(-2\pi(\omega_{z^j} z^j)) + i \sin(-2\pi(\omega_{z^j} z^j)) dz^j \\
 &= \sum_{j=1}^N \frac{\sin(2\pi(D\omega_{z^j}))}{\pi\omega_{z^j}}
 \end{aligned} \tag{21}$$

Recall the sinc function:

$$sinc(x) = \frac{\sin(x)}{x}$$

Then (14) can be rewritten as:

$$\mathcal{F}\{b_\rho\} = \sum_{j=1}^N I_j(x^j, y^j) \times 2D \text{sinc}(2D\pi\omega_{z^j}) \tag{22}$$

In particular, (16) tells us that a back projection is just the inverse Fourier Transform of the summation of the images, collected from various orientations, multiplied by a transformed sinc function:

$$b_\rho = \mathcal{F}^{-1} \left\{ \sum_{j=1}^N I_j(x^j, y^j) \times 2D \text{sinc}(2D\pi\omega_{z^j}) \right\} \tag{23}$$

## 6 Estimating Orientations

Suppose we have images  $I_1, I_2, \dots, I_n$  and a set of orientations  $F_1, F_2, \dots, F_n$ .

Our goal is to properly align the back projection of  $I_j, b_{\rho_j}$ , so that it is along the projection direction. These alignments are unknown, but we can resolve them from our images  $I_j$ .

Let us consider two images,  $I_1$  with orientation  $F_1$  and  $I_2$  with orientation  $F_2$ . Both of these images are of the function  $\rho$ , but they are in different orientations.

Recall the Fourier Slice Theorem:

$$\mathcal{F}\{I_i\} = \mathcal{F}\{\rho\}|_{\text{span}\{a_i, b_i\}} \quad (24)$$

Applying this to  $I_1$  and  $I_2$ , we see that we have two planes:  $\text{span}\{a_1, b_1\}$  and  $\text{span}\{a_2, b_2\}$ . Since these two planes are in 3D space, they must intersect at some line. Note this assumes the planes are not parallel.

We define this line as follows:

$$L_{12} = \text{span}\{a_1, b_1\} \cap \text{span}\{a_2, b_2\} \quad (25)$$

This line is the common line between  $\mathcal{F}\{I_1\}$  and  $\mathcal{F}\{I_2\}$ . So we have the following:

$$\mathcal{F}\{I_f\}|_{L_{12}} = \mathcal{F}\{I_1\}|_{l_{21}} = \mathcal{F}\{I_2\}|_{l_{21}} \quad (26)$$

Where  $l_{12}$  is the projection of  $L_{12}$  onto  $\text{span}\{a_1, b_1\}$  and  $l_{21}$  is the projection of  $L_{12}$  onto  $\text{span}\{a_2, b_2\}$ :

$$l_{12} = \begin{bmatrix} -a_1 - \\ -b_1 - \end{bmatrix} * L_{12} \quad (27)$$

$$l_{21} = \begin{bmatrix} -a_2 - \\ -b_2 - \end{bmatrix} * L_{12} \quad (28)$$

Note that  $l_{12}$  is the line in  $I_1$  that agrees with  $I_2$  and  $l_{21}$  is the line in  $I_2$  that agrees with  $I_1$ .

Now that we see how to generate these common lines, let us use them to orient our images. Suppose we have 3 images:  $I_1, I_2, I_3$  from unknown orientations  $F_1, F_2, F_3$ . Assume we know all common line pairs. We can see that there are 3, since we have 3 images and we choose 2 of them to form a line. We can always find the number of common line pairs from  $nC_2$ , n choose 2, where n is the number of images.

We cannot use these lines to find the exact  $F_1, F_2, F_3$ , but we can globally orient them by arbitrarily setting the first image and aligning the others accordingly.

Let

$$F_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (29)$$

Then  $\text{span}\{a_1, b_1\}$  is the x-y plane. Now,  $F_2$  is guaranteed to intersect  $F_1$  at line  $L_{12}$ . However, this leaves a rotational ambiguity because  $F_2$  can still revolve around  $l_{12}$  in 3D space without becoming misaligned. This is because we are missing a constraint, so  $F_2$  has an extra degree of freedom to rotate.

We can resolve this by using a third image,  $F_3$ . First, append  $F_2$  onto  $F_1$  so that  $l_{12}$  and  $l_{21}$  align. Then, we glue  $F_3$  to  $F_1$  along  $L_{13}$ . We now roate  $F_2$  around  $L_{12}$  and  $F_3$  around  $L_{13}$  until  $l_{23} = l_{32}$ . Now, all three images are fixed in space by the constraints:

$$l_{12} = l_{21} : l_{13} = l_{31} : l_{23} = l_{32} \quad (30)$$

However, we need to know how much we must rotate our images so that they meet the constraints. We can solve exactly for these angles A, B, C.

First, overlay a unit sphere onto our 3D coordinate space. Then we intersect this sphere 6 times, twice for each line. Now, we have

$$\alpha \in \text{span}\{a_1, b_1\} : \beta \in \text{span}\{a_2, b_2\} : \gamma \in \text{span}\{a_3, b_3\} \quad (31)$$

lines in each plane forming a spherical triangle on the surface of the unit sphere.

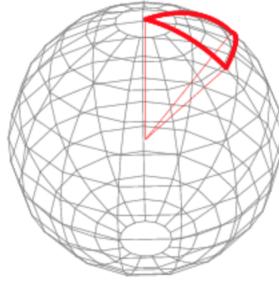


Figure 15: Three planes form a spherical triangle on the unit sphere.

We claim that A,B,C are the angles between the spans. We will use the spherical law of cosines to solve for these angles:

$$\cos(C) = \frac{\cos(\gamma) - \cos(\alpha) * \cos(\beta)}{\sin(\alpha) * \sin(\beta)} \quad (32)$$

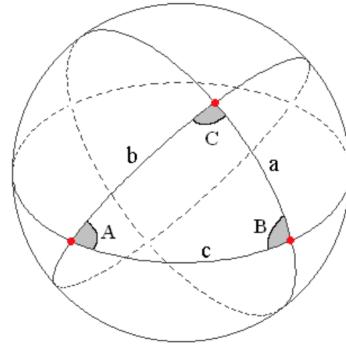


Figure 16: A labeled spherical triangle. In the derivation below, the arc lengths  $a, b, c$  are written  $\alpha, \beta, \gamma$ .

*Proof.* Let  $u, v, w$  be unit vectors from the center of the unit sphere to the corners of the spherical triangle. Then the side lengths are as follows:

$$\cos(\alpha) = u \cdot v \quad : \quad \cos(\beta) = u \cdot w \quad : \quad \cos(\gamma) = v \cdot w \quad (33)$$

To find angle C, we need tangent vectors  $a_t, b_t$  at  $u$  along sides a and b. Specifically,  $a_t$  is the unit vector perpendicular with  $u$  in the  $u-v$  plane with direction given by the component of  $v$  perpendicular to  $u$ . So we have:

$$a_t := \frac{v - u(u \cdot v)}{\|v - u(u \cdot v)\|} \quad (34)$$

We now use the identity  $\sin^2(a) = 1 - \cos^2(a)$  to substitute:

$$a_t := \frac{v - u\cos(\alpha)}{\sin(\alpha)} \quad (35)$$

Similarly for  $b_t$ :

$$b_t := \frac{w - u\cos(\beta)}{\sin(\beta)} \quad (36)$$

So then we have:

$$\cos(C) = a_t * b_t = \frac{\cos(\gamma) - \cos(\alpha)\cos(\beta)}{\sin(\alpha)\sin(\beta)} \quad (37)$$

□

Using this, we see that

$$A = \cos^{-1}\left(\frac{\cos(\alpha) - \cos(\beta)\cos(\gamma)}{\sin(\beta)\sin(\gamma)}\right) \quad (38)$$

$$B = \cos^{-1}\left(\frac{\cos(\beta) - \cos(\alpha)\cos(\gamma)}{\sin(\alpha)\sin(\gamma)}\right) \quad (39)$$

$$C = \cos^{-1}\left(\frac{\cos(\gamma) - \cos(\alpha)\cos(\beta)}{\sin(\alpha)\sin(\beta)}\right) \quad (40)$$

Now, we may use the angles A, B, C to orient  $F_1, F_2, F_3$  as follows:

- (I): set  $F_1$  to be the x-y plane.
- (II): attach  $F_2$  to  $L_{12}$  with angle C.
- (III): attach  $F_3$  to  $L_{12}$  with angle B.

We may rotate  $F_2$  and  $F_3$  by their corresponding angles using the rotation matrix:

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (41)$$

For the remaining planes,  $F_i$ , we need to determine a linear transformation  $T: l_{i,j} \mapsto L_{j,i}, 1 \leq j \leq 3$ . Notice the  $3 \times 2$  matrix representation of  $T$  is the change-of-basis from the standard

basis in  $\mathbb{R}^2$  to  $\text{span}\{a_i, b_i\}$ , simply the plane onto which  $L_{i,j}$  is projected to create  $l_{j,i}$  or  $l_{i,j}$ . (Note  $a_i$  and  $b_i$  are also the columns of  $F_i$ ). Or, more explicitly:

$$l_{i,j} = \begin{bmatrix} a_i \\ b_i \end{bmatrix} L_{i,j}$$

$$a_i = T(e_1)$$

$$b_i = T(e_2)$$

$T$  is an isometry (congruent transformation), is a distance-preserving injective transformation between  $\mathbb{R}^2$  and  $\mathbb{R}^3$ .

For example, consider  $F_{17}$ , the orientation of  $I_{17}$ . To resolve its orientation, we need only two pieces of information:  $l_{1,17}$  and  $l_{17,1}$ . Since we have three planes "locked" into place appropriately in 3D space, it is sufficient to align  $F_{17}$  based only on these two lines to properly orient it.

If we use all the maps from  $l_{i,j} \mapsto L_{j,i}$ , we now have a system of linear equations that will allow us to determine  $T$  uniquely. Apply  $T$  to  $l_{i,j}$ , and we are done.

## 7 The Contrast Transfer Function

To understand how Contrast Transfer Function(CTF) will affect the cryogenic electron microscopy(Cryo-EM), we need to know first that Cryo-EM is a type of TEM where the sample is studied at cryogenic temperatures (generally liquid-nitrogen temperatures).

**Transmission electron microscopy(TEM)** Transmission electron microscopy (TEM) is a microscopy technique where a high energy beam of electrons is transmitted through a very thin sample, and the interactions between the electrons and the atoms can be used to observe features of the atoms. The interaction when the electrons transmitted through the specimen forms the image.

The Phase Contrast Transfer Function (CTF) is a mathematical description of the imaging process expressed in Fourier Space in TEM. Ideally, TEM images would represent exact projections of the electron density of the specimen on the image plane. However, in the real world, images are distorted by the microscope optics, primarily phase contrast, and contain high levels of noise. Additionally, an electron beam bends when it encounters small objects and spreads when it goes through an opening, a phenomenon called diffraction.

In the previous sections, we discuss how the true projections of the electron density of the specimen look like and also how they look with filters. However, we did not consider distortions of the image by phase contrast and did not apply the Contrast Transfer Function. In this section, we first discuss Transmission Electron Microscopes and their image-producing systems to provide some real-life context. Then we will examine the "distortions" caused by the properties of electron beams and instrument-influenced errors.

### 7.1 Transmission Electron Microscope

Usually, Transmission Electron Microscope is constructed by three main systems.

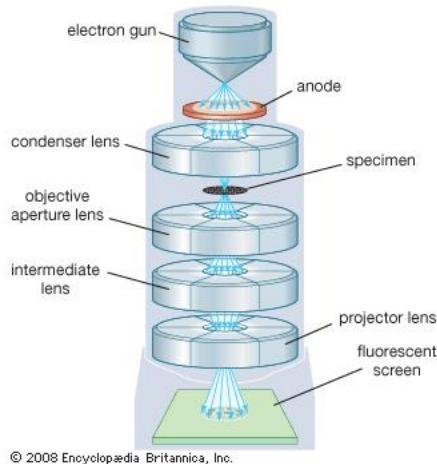


Figure 17: TEM Model [Bri].

### 7.1.1 Electron Gun and Condenser System

This system produces focused beam onto the object. Electron gun produces a electron beam and condenser lens help to focus the beam.

- 1. Condenser Lens** Usually, there are two condenser lenses in a TEM, which purpose is to control how strongly the beam is focuses onto the specimen. This lens is used to form the beam, limits the amount of the current in the beam and also control the diameter of the beam. If the condenser current increases, the beam focuses above the aperture and the aperture will intercept a large part of electron beam which reduces the current. To eliminate the high-angle electrons from the beam, it always works in conjunction with the condenser aperture.
- 2. Coherence of Electron Source** Coherence is a definition in physics which describes that two wave sources are perfectly coherent if they have a constant phase difference and the same frequency. It is an ideal property of waves that enables stationary interference. It is a concept describes all properties of the correlation between physical quantities of a single wave, or between several waves or wave packets. In the image formation process, coherence is the degree to which separate parts of the electron wave have defined phase relationships and interfere with each other. When all parts of the wave have defined phase relationships, it is called full coherence. When there is no defined phase relationships, it is called incoherence. The intermediate situation is partial coherence. In bright field electron microscopy, the contrast transfer function is normally given for the fully coherent situation, and the influence of partial coherence is expressed by an envelope function. [Bri]

### 7.1.2 Image-Producing System

This system consists the objective lens, movable specimen stage, and intermediate and projector lenses, which focus the electrons passing through the specimen to form a real, highly magnified image. [Bri]

1. **Objective Lens** Objective lens is the part of microscope responsible for magnifying the image of specimen. Usually there are three to four objective lenses in a standard microscope of 4X, 10X, 40X and 100X. This is the first lens of a microscope which is closest to the specimen being observed where the electron beams coming from object fall and make an image somewhere between it and eye piece which then is observed by observer via an eye piece. [Bri]
2. **Apertures** The intensity and angular aperture of the beam are controlled by the condenser lens system between the gun and the specimen. The function of the objective aperture is to select those electrons which will contribute to the image and affect the appearance of the image. It also improves the contrast of the final image. By inserting and moving the position of objective aperture, different types of images could be formed. [Bri]

### 7.1.3 Image-recording System

This system consists a fluorescent screen for viewing and focusing the image and a digital camera for permanent records, which converts the electron image into some form perceptible to the human eye. [Bri]

1. **Defocus** Defocus is the aberration in which an image is simply out of focus. It refers to a translation along the optical axis away from the plane or surface of best focus. Ideally, the image should focus on the image plane to get clear image. However, in the real world, the image usually focuses slightly up or down of the image plane which will cause fine details of the image are blurred and even invisible. In general, defocus reduces the sharpness and contrast of the image.[Bri]

## 7.2 Fourier Optics

Fourier optics is the study of classical optics using Fourier transforms, which is particularly useful for studying cascades of lenses in microscopes and other linear systems (such as telescopes). Microscopes output data in real space, in the image plane, but lenses perform exact Fourier Transforms on waves exiting an object, so microscopes are formally be described as performing a "double Fourier Transform"- the first from real space to reciprocal space, and the second back to real space. A lens focuses beams diffracted from an object to form a diffraction pattern- the Fourier Transform- in reciprocal space on the back focal plane of a lens. A second Fourier Transform produces an inverted image of the object magnified by a ratio of the lens-image distance to the lens-object distance. However, the image does not completely faithfully represent the object due to various aberrations that alter the Fourier Transform.[Fre]

## 7.3 Numerical Aperture

In addition to various aberrations caused by the lenses, lens aperture can limit the resolution of an image. The numerical aperture depends on the  $n$ , the refractive index of the

medium between the objective front lens and the specimen, and  $\alpha$ , the aperture half-angle of the objective lens: [Fre]

$$N.A. = n \sin(\alpha)$$

Recall resolution is defined as the shortest distance between two points on a sample that can still be distinguished by the imaging system as separate points. Thus, the numerical aperture greatly limits the resolution of the object. Given two image points separated by a distance  $d$  (the resolution), the German physicist Ernst Abbe developed the following relationship between resolution and numerical aperture in an idealized optical system: [Rus]

$$d = \frac{0.612 \times \lambda}{n \sin(\alpha)}$$

In other words, the higher the numerical aperture, the lower the resolution. Note, however, that this equation is not complete; the total resolving power of the microscope depends not only on the numerical aperture of the objective lens, but also on that of the condenser.

## 7.4 Spherical Aberration

In this section, we discuss one type of aberration, errors introduced by an optical system that distorts perfectly spherical wavefronts to cause defocuses in an image. The surface of lenses is spherical. When the electrons beam pass marginal zones of the lenses, they are refracted more than those who pass near the center. The beams that pass though the axis are refracted the least. Because of this, the electron beams cannot be focused in one single point. Ideally, all the beams should be imaged as a single point in the imaged plane. However, aberration causes beams to be imaged as the spread of a single point because of the aberration, they are imaged as a spread of a point over a finite disc on the image plane. As a result, the spherical aberration is a blurring effect because the lenses are not able to converge all incoming beams at higher angles to the focus point but rather focus them in the close positions.

The spherical aberration value  $C_s$  is a constant unique to a lens. It is quite complicated and depends on the refractive index, lens shape factor, and lens position factor. [MF] However, the radius of spherical aberration (the radius of the finite disc that represents a blurred point on the image plane) can be calculated as follows: [MF]

$$r_{sphere} = C_s \beta^3$$

where  $\beta$  is the maximum half-angle of the direction a beam travels of the objective lens aperture (see figure below).

## 7.5 The Point Spread Function

We can use transfer functions in reciprocal space to describe the effects of aberrations characterizing an instrument, a common theme with all linear systems. As mentioned before, aberrations can "smear" an image. A point spread function models this "smearing" effect seen on an image, quite literally describing how a single point "spreads" across an image. Note that although there are many different kinds of aberrations, for the sake of simplicity,

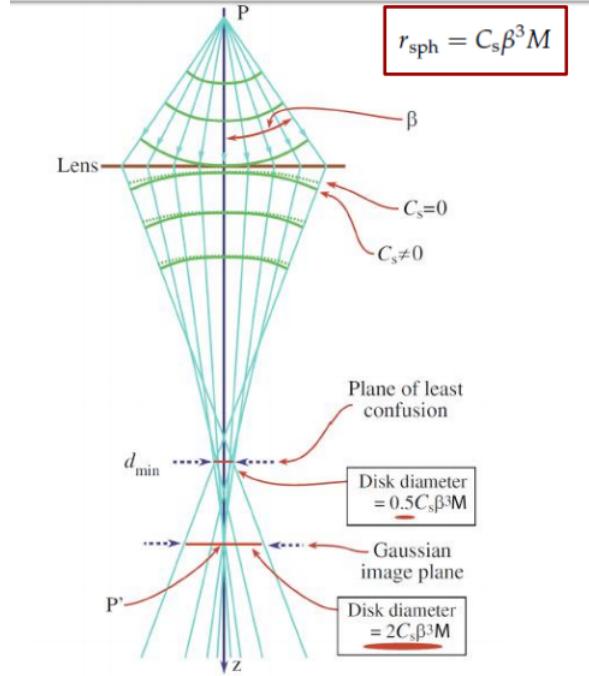


Figure 18: Various calculations of the radius of spherical aberration.

we only address spherical aberration in this paper. In linear systems, an output image ( $g$ ) is the point spread function ( $h$ ), convolved with an input image ( $f$ ):

$$g(x, y) = h(x, y) * f(x, y) \quad (42)$$

The convolution theorem allows for convenient transitions between real and reciprocal space; equation (39) can be Fourier Transformed to yield:

$$\hat{g}(k_x, k_y) = \hat{h}(k_x, k_y) \hat{f}(k_x, k_y) \quad (43)$$

The Fourier Transform of the Point Spread Function (the  $\hat{h}(k_x, k_y)$  term) of an electron microscope is known as the Contrast Transfer Function.

## 7.6 Phase Contrast Transfer Function

We now discuss the Fourier Transform of the point spread function mentioned in the previous section. Recall that when an electron hits the sample, it behaves in a wave-like nature and scatters in all different directions of the full circle. As a result, many of the scattered beams are out of phase with the unscattered beam and therefore cannot be detected on the image plane. Since a scattered electron travels a longer path between the sample and the image plane, the scattered beam will be periodically out of phase with the unscattered beam.

A higher scattering angle ( $\theta$ ) means a higher spatial frequency ( $k$ ), given by the relationship:

$$\theta = \lambda k$$

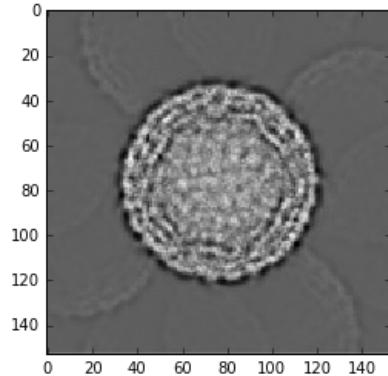


Figure 19: Original image of the Zika virus before corruption by the CTF.

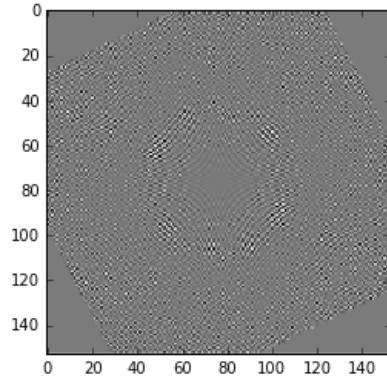


Figure 20: The Fourier Transform form of the same image corrupted by the CTF.

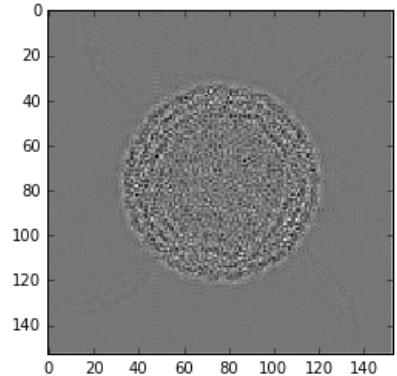


Figure 21: The final CTF-corrupted image, the inverse FT of the center figure.

Thus, many of the higher spatial frequencies are also out of phase with the unscattered beam and thus cannot be detected on the image plane. The CTF , which gives information on how much contrast is transferred into the image as a function of spatial frequency, is formulated to address this concern.[[Jena](#)]

Mathematically speaking, the Phase Contrast Transfer Function (or CTF for short), in terms of spatial frequency, quantifies image distortions due to defocus and spherical aberration. The Contrast Transfer Function( $K(k)$ ) is:

$$K(k) = \sin((2\pi\lambda)W(k))$$

where

$$W(k) = C_s \frac{\lambda^3 k^4}{4} - \Delta z \frac{\lambda k^2}{2}$$

represents the wave aberration,  $\Delta z$  is the defocus of the objective lens,  $C_s$  the spherical aberration of the objective lens,  $k$  is the spatial frequency, and  $\lambda$  represents the wavelength of the electron, which can be calculated from an approximation derived from de Broglie's hypothesis ( $\lambda = \frac{h}{p}$ ,  $h$  is Planck's constant, and  $p$  the momentum of the particle):

$$\lambda \approx \frac{12}{\sqrt{V}}$$

where  $V$  is the acceleration voltage of the microscope. In 1949, German physicist Otto Scherzer published a paper relating the defocus value,  $\Delta z$ , and the spherical aberration,  $C_s$ , now an equation known as the Scherzer defocus:

$$z_s = (C_s \lambda)^{1/2}$$

where  $z_s$  indicates the Scherzer defocus, used by many microscopists to counteract the effects of spherical aberration to obtain the optimal contrast (which a defocus value of 0 does not always obtain).

## 7.7 Difficulties in Reconstruction

Recall from what we talked before that CTF is a function accounts for the imperfection of images information transmission due to the microscope lenses. It is like a filter which only pass parts of the image information to the detector. Therefore, what we observe from the microscope is an incomplete distorted image. Without any CTF correction, the model produced by a reconstruction may contain significant local density displacements. Thus, to do the reconstruction, we need to correct the image first.

### 7.7.1 CTF Correction

In an ideal world, to obtain information about the original image, we can correct the image by simply dividing the CTF function. Recall an image that is formed is the inverse transform of the object times the contrast transfer function (see section on CTF). The object is therefore the Fourier transform of the image divided by the CTF, if we think of the CTF as a scaling factor of the image. Thus, if we divide inverse Fourier transform of the image by the CTF, we obtain the object itself. But, of course, in the real world, certain problems arise, which we discuss below.

### 7.7.2 Wiener Filtration

One elementary problem is: because the CTF is a sine function, there are zeros contained in the function which will cause errors if we divide by it. There are several ways to address this problem, one of which is to simply omit the frequencies where CTF is equal to 0. [Jenb] However, the most commonly used method to recover the image is to collect images at different defocus values so the images that carry no information at certain frequencies will be complemented by those that do. At lower resolutions, the CTF does not depend on defocus, but at higher resolutions, the CTF's begin to interfere, so it becomes necessary to correct the phases and amplitudes. A common way to address this problem is *Wiener Filtration*, which has shown, through experiments, to mathematically yield the best results.

To obtain an average image from the various defocuses, we start by summing the images collected, described by the object multiplied by the CTF of each image:

$$I_1 + I_2 + \dots + I_n = S(CTF_1 + CTF_2 + CTF_3 + \dots + CTF_n)$$

Suppose that we multiply each CTF by a Wiener factor  $\frac{\sum_{i=1}^n CTF_i}{\sum_{i=1}^n CTF_i^2 + \frac{n}{s}}$ :

$$\sum_{i=1}^n I_i = S(CTF_1 + CTF_2 + CTF_3 + \dots + CTF_n) \left( \frac{\sum_{i=1}^n CTF_i}{\sum_{i=1}^n CTF_i^2 + \frac{n}{s}} \right)$$

where  $\frac{n}{s}$  is the noise-to-signal ratio. The  $\frac{n}{s}$  term is used to prevent division by a small value, thus blowing up the Fourier data. Roughly speaking, applying the Wiener Filter multiplies the image by the CTF before dividing the product by the CTF-squared, which is equivalent to dividing the image by the CTF. The effect is to smooth the image while inverse filtering at the same time. [Bar95]

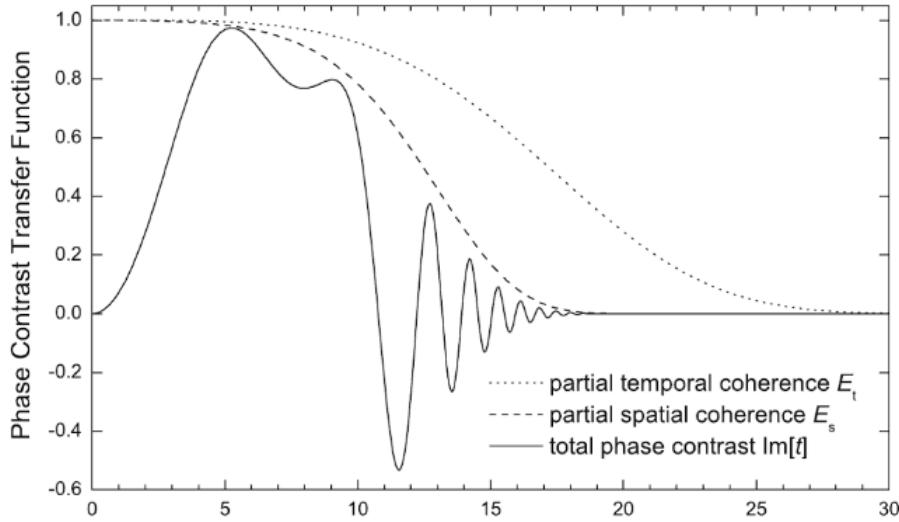
### 7.7.3 Envelope Functions

Let's revisit the idea of coherence of an electron source, mentioned before. An ideal electron beam is fully coherent, but real electron sources produce electrons of various energies at various angles and thus lead to a partially coherent imaging process. This leads problems such as chromatic aberrations, the effect of refraction of different wavelengths through slightly different angles, which prevents the ability to focus correctly. The CTF presented before models an idealized, fully coherent electron source, so it becomes necessary to address additional problems to create a new CTF model. Aberrations, such as chromatic aberration, tend to affect higher spatial frequencies more; thus, envelope functions suppress data at higher spatial frequencies, until, at a certain point, no more signal can pass through (the Information Limit).

We now present a model taking into account that true electron sources are only partially coherent, making use of damping equations called envelope functions. There are two kinds of envelope functions: temporal and spatial envelope functions. Temporal envelopes take into account that not all electrons have the same energy, while spatial frequencies address the issue that they travel towards the sample at different angles. Both envelope functions affect the effective CTF, denoted  $K_{eff}$ :

$$K_{eff}(k) = E_s E_t (\sin((2\pi\lambda)W(k)))$$

where  $E_s$  and  $E_t$  denote the temporal and spatial envelope functions, respectively. In short, envelope functions provide a more practical model of the CTF's effect on the output image in unidealized settings.[\[Jenc\]](#)

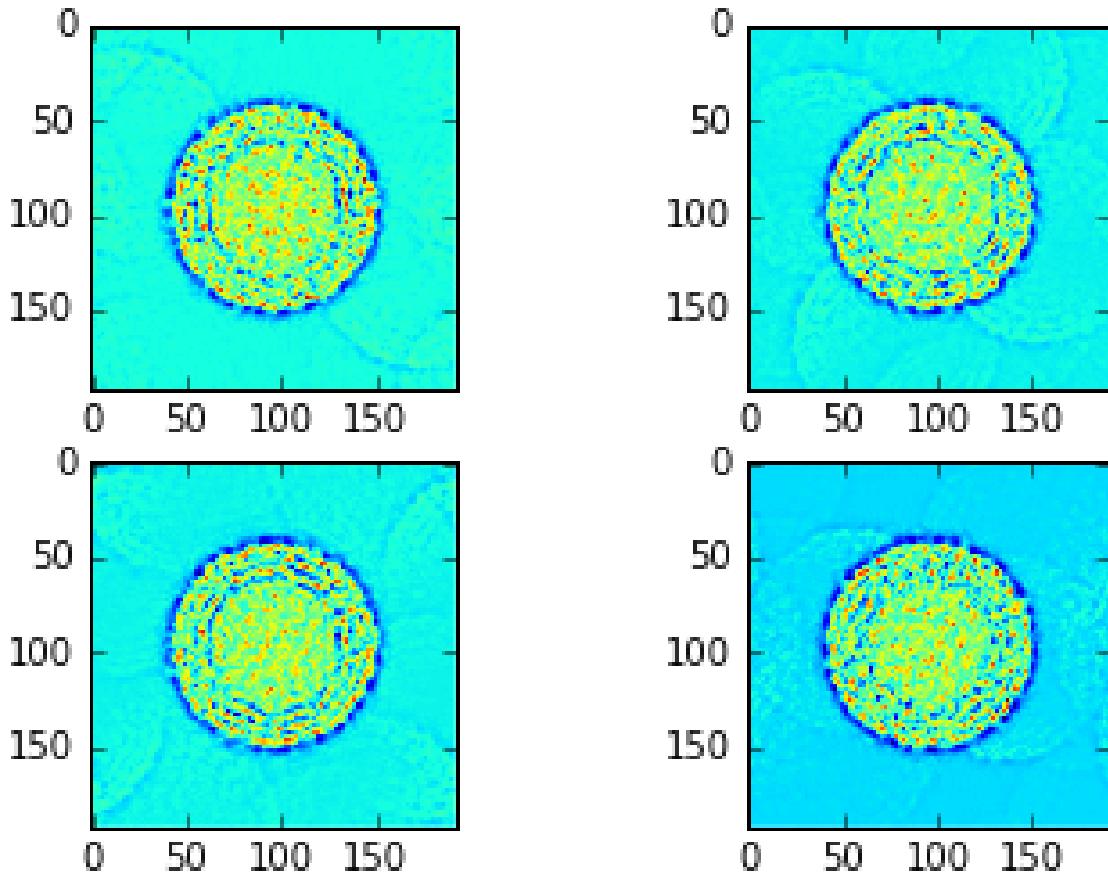


*Figure 22: Temporal and spatial envelope functions take into account the incoherence of unidealized electron sources and dampens the CTF, suppressing data at higher frequencies. [\[Ern15\]](#)*

## 8 Results

### 8.1 EM Simulation

Simulated EM projection images, from *project\_fst* of a zika virus from 4 random rotational matrices (generated by *randrot*) shown below:



### 8.2 3D Structure Reconstruction

We reconstructed the zika virus using images stimulated by *project\_fst* and ran the reconstructions with various signal-to-noise ratios ranging from little noise (1000, see first figure on the top left) to signal-to-noise level ratios as low as 0.1.

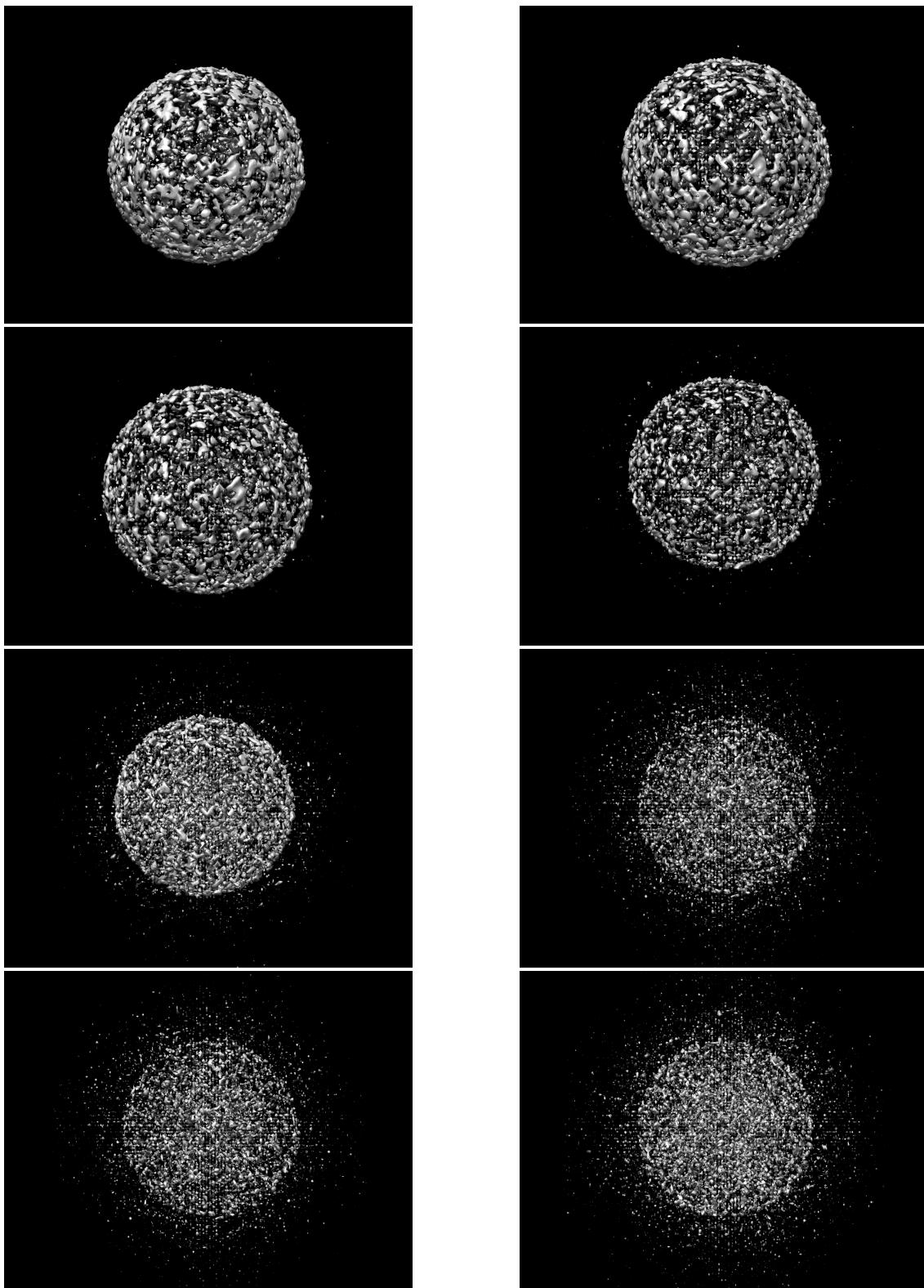


Figure 23: Our projection results with various signal-to-noise ratios: Top Left: 10000; Top right: 5; 2nd Row Left: 2; 2nd Row Right: 1; 3rd Row Left: 0.5; 3rd Row Right: 0.25; Last Left: 0.125; Last Right: 0.1.

### 8.3 Orientation Estimation

Since the 3D structure of Zika virus is highly symmetric, projecting it from several directions would yield similar results, making it hard to find the common lines. We test the orientation estimation method on a asymmetric structure, RelA bound to the 70S ribosome (shown above). Our code could find the proper rotation matrices that "glue" the common lines. But it's hard to tell whether it finds the correct common lines.

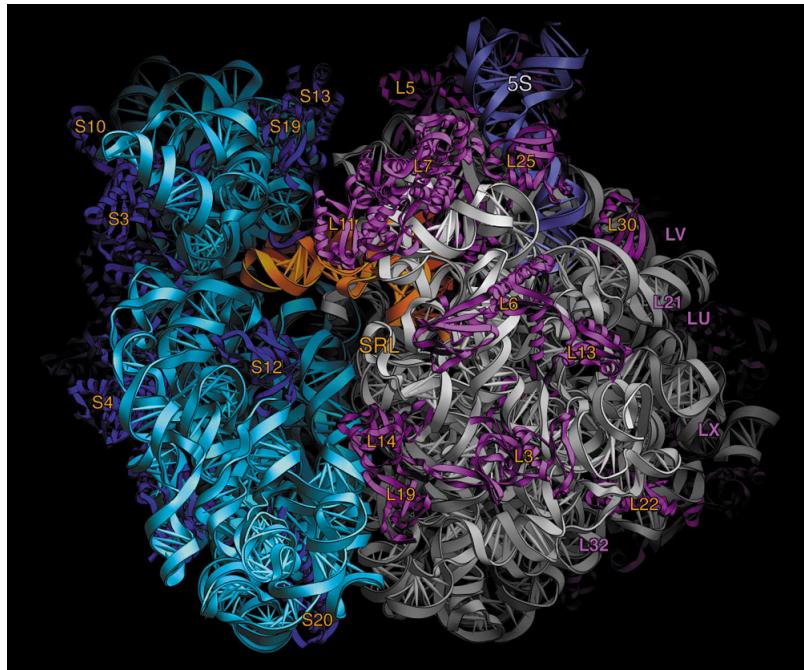


Figure 24: The 70S ribosome. [70S]

## 9 Code Listing

### 9.1 Generating random rotation matrices: randrot()

```
import numpy as np

def randrot():
    """
    This is a function to generate a random orthogonal matrix.
    :Output: an orthogonal matrix
    """
    a = np.random.rand(3)      #random vector 1
    b = np.random.rand(3)
    b /= np.linalg.norm(b)    #want b to be normalized
    a -= a.dot(b) * b         #want a to be orthogonal to b
```

```

a /= np.linalg.norm(a)    #want a to be normalized
c = np.cross(a, b)        #take cross product of a and b
c /= np.linalg.norm(c)    #normalize c

return np.array((a), (b), (c))

```

## 9.2 Simulating EM images: project\_fst()

```

from MRCFile import MRCFile
import numpy as np
from matplotlib import pyplot as plt
import scipy.interpolate as si
from rotation import randrot

def project_fst( mol, R ):
    '''Simulating the projection along some random direction
    # Input: mol-3D structure of molecule; R-rotation matrix
    # Output: projected 2D image'''

    L = mol.header.CELLA[0]
    N = mol.data.shape[0]

    # Perform Fourier Transform
    r_hat = (L**3/N**3) * np.fft.fftshift(np.fft.fftn(mol.data))

    freq = np.arange(-(N-1)//2, (N+1)//2)
    wx, wy, wz = np.meshgrid(freq, freq, freq)

    r_hat = (-1)*((-1)**((wx + wy + wz))) * r_hat

    #Projecting image_hat
    wx2d, wy2d = np.meshgrid(freq, freq)
    a, b = R[:,0], R[:,1]
    #Interpolation
    coords = wx2d[..., np.newaxis]*a/L + wy2d[..., np.newaxis]*b/L
    image_hat = si.interpn((freq/L, freq/L, freq/L), r_hat, coords, method="linear",
                           bounds_error=False, fill_value=0)
    image_hat = (N**2/L**2) * (-1)*(-1)**((wx2d + wy2d)) * image_hat

    #return image
    return np.fft.ifftn(np.fft.ifftshift(image_hat))

```

### 9.3 Orientation estimation: common\_lines(), orientation\_est()

```

from MRCFile import MRCFile
import numpy as np
from matplotlib import pyplot as plt
import scipy.interpolate as si

def common_lines(img1,img2,unit_vec):
    '''Find the common line of two given 2D images from provided directions unit_vec'''
    #Fourier Transform on images
    img1_hat = (L**2/N**2) * np.fft.fftshift(np.fft.fftn(img1))
    img2_hat = (L**2/N**2) * np.fft.fftshift(np.fft.fftn(img2))
    freq = np.arange(-(N-1)//2, (N+1)//2)
    wx, wy= np.meshgrid(freq, freq)
    img1_hat = np.real((-1)*((-1)**((wx + wy))) * img1_hat)
    img2_hat = np.real((-1)*((-1)**((wx + wy))) * img2_hat)
    cor=0.0
    #Choose 2 random unit vectors
    opt_vec1=unit_vec[0]
    opt_vec2=unit_vec[0]
    for vec1 in unit_vec:
        coords1 = freq[...,np.newaxis] * vec1 * 0.99
        line1= si.interpn((freq, freq), img1_hat, coords1, method='linear')
        for vec2 in unit_vec:
            coords2 = freq[...,np.newaxis] * vec2 *0.99
            line2= si.interpn((freq, freq), img2_hat, coords2, method='linear')
            new_cor=line1.dot(line2)
            #Find the largest cross-correlation, set as the new cross-correlation
            if new_cor>cor:
                cor=new_cor
                opt_vec1=vec1
                opt_vec2=vec2
    #Return two vectors representing the projections of the common line onto the image plane
    return opt_vec1,opt_vec2

def orientation_est(I1,I2,I3):
    '''Set F1 to identity matrix and estimate other two roatation matrices
    # Input: 3 2D images'''

    # Generate several lines as unit vectors
    angel=np.arange(0,2*np.pi,0.1)
    x=np.cos(angel)
    y=np.sin(angel)

```

```

unit_vec=np.array([x,y]).T

#Find pairwise common lines of the three images
ell_12,ell_21,_=common_lines(I1,I2,unit_vec)
ell_13,ell_31,_=common_lines(I1,I3,unit_vec)
ell_23,ell_32,_=common_lines(I2,I3,unit_vec)

cos_alpha=ell_12.dot(ell_13)
cos_beta=ell_21.dot(ell_23)
cos_gamma=ell_31.dot(ell_32)
sin_alpha=np.sqrt((1-cos_alpha**2))
sin_beta=np.sqrt((1-cos_beta**2))
sin_gamma=np.sqrt((1-cos_gamma**2))

#Applying Spherical Law of Cosines
cos_A=(cos_alpha-cos_beta*cos_gamma)/sin_beta/sin_gamma
cos_B=(cos_beta-cos_alpha*cos_gamma)/sin_alpha/sin_gamma
cos_C=(cos_gamma-cos_beta*cos_alpha)/sin_beta/sin_alpha
sin_A=np.sqrt((1-cos_A**2))
sin_B=np.sqrt((1-cos_B**2))
sin_C=np.sqrt((1-cos_C**2))

#Glue l12 with angle C
R_21=np.array([[ell_21[0],ell_21[1],0],[-ell_21[1],ell_21[0],0],[0,0,1]])
R_22=np.array([[1,0,0],[0,cos_C,-sin_C],[0,sin_C,cos_C]])
R_23=np.array([[ell_12[0],ell_12[1],0],[-ell_12[1],ell_12[0],0],[0,0,1]]).T
F2=R_23.dot(R_22.dot(R_21))

#Glue l13 with angle B
R_31=np.array([[ell_31[0],ell_31[1],0],[-ell_31[1],ell_31[0],0],[0,0,1]])
R_32=np.array([[1,0,0],[0,cos_B,sin_B],[0,-sin_B,cos_B]])
R_33=np.array([[ell_13[0],ell_13[1],0],[-ell_13[1],ell_13[0],0],[0,0,1]]).T
F3=R_33.dot(R_32.dot(R_31))

return(np.eye(3),F2,F3)

```

## 9.4 Reconstructing 3D structure: reconstruction

```

from MRCPfile import MRCPfile
import numpy as np
from matplotlib import pyplot as plt
import scipy.interpolate as si

```

```

def back_project(images, Fs):
    '''Reconstructing 3D structure with 2D projected images
    # Input: a set of 2D images and corresponding directions'''

    NUM=Fs.shape[0] #number of images
    bp=np.zeros((N,N,N),dtype='complex')
    h=np.zeros((N,N,N))

    for i in range(NUM):
        image=images[i]
        F=Fs[i]
        #Fourier transform of the images with PSF applied.
        image_hat = (L**2/N**2) * np.fft.fftshift(np.fft.fftn(image))
        freq = np.arange(-(N-1)//2, (N+1)//2)
        wx, wy= np.meshgrid(freq, freq)
        image_hat = ((-1)**((wx + wy))) * image_hat
        #Tile image in the z-direction to obtain 3D volume.
        image_tile=np.tile(image_hat[...],np.newaxis), (1,1,N))
        #Fourier transform of the phase contrast function.
        l_hat = np.sinc(2*freq/L*D)*2*D
        #Tile l_hat in x and y directions to create 3D volume.
        l_tile = np.tile(l_hat[np.newaxis,np.newaxis,...],(N,N,1))
        #Convolve tiled image with the phase contrast function.
        bi_hat = image_tile * l_tile

        #Create new coordinate system based on the rotation matrix.
        ai, bi, ci = F[0], F[1], F[2]
        wx3d, wy3d, wz3d = np.meshgrid(freq, freq, freq)
        coords = wx3d[...], np.newaxis]*ai +
                  wy3d[...], np.newaxis]*bi +
                  wz3d[...], np.newaxis]*ci

        #Interpolation.
        bp_i = si.interpn((freq/L, freq/L, freq/L), bi_hat, coords/L, method='linear',
                           bounds_error=False, fill_value=0)
        bp += bp_i

        h_i = si.interpn((freq/L, freq/L, freq/L), l_tile, coords/L, method='linear',
                           bounds_error=False, fill_value=0)
        h += h_i

    bp = (N**3/L**3) *(-1)**((wx3d + wy3d + wz3d)) * bp
    h[np.logical_and(h>=0,h<0.6)]=0.6
    h[np.logical_and(h<0,h>-0.6)]=-0.6
    est_hat = (bp / h)

```

```
#Apply inverse Fourier Transform.
return np.fft.ifftn(np.fft.ifftshift(est_hat))
```

## 9.5 Function apply\_CTF()

```
from MRCFile import MRCFile
import numpy as np
from matplotlib import pyplot as plt
import scipy.interpolate as si

def apply_CTF(voltage, defocus, Cs, image):
    #Take Voltage, Defocus, Spherical Aberration Value, and Image as inputs

    #Image_hat
    image_hat = np.fft.fftshift(np.fft.fftn(image))
    freq = np.arange(-(N-1)//2, (N+1)//2)
    wx, wy = np.meshgrid(freq, freq)

    #the CTF function
    k = np.sqrt((wx/L)**2 + (wy/L)**2)
    lamb = 12/np.sqrt(voltage)
    W = Cs * (lamb**3 * k**4)/4 - defocus*(lamb*k**2)/2
    K = np.sin(2*np.pi*lamb*W)

    #Convolve the CTF function with fourier transform of the image
    image_ctf_hat = image_hat*K

    #Return the image corrupted by CTF
    return np.fft.ifftn(np.fft.ifftshift(image_ctf_hat))
```

## References

- [70S] 70s\_atrna\_labels.jpg (1006×832). [http://rna.ucsc.edu/rnacenter/images/figs/70s\\_atrna\\_labels.jpg](http://rna.ucsc.edu/rnacenter/images/figs/70s_atrna_labels.jpg). University of California, San Francisco. (Accessed on 12/09/2016).
- [Bar95] Richard G. Baraniuk. Wiener filtering. <http://www.owlnet.rice.edu/~elec539/Projects99/BACH/proj2/wiener.html>. Rice University, November 1995. (Accessed on 12/08/2016).
- [Bri] Encyclopedia Britannica. Transmission electron microscope (tem). <https://www.britannica.com/technology/transmission-electron-microscope>. (Accessed on 12/09/2016).
- [Cala] Ewen Callaway. Mature hiv-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3729984/#SD1>. (Accessed on 12/08/2016).
- [Calb] Ewen Callaway. The revolution will not be crystallized: a new method sweeps through structural biology. <http://www.nature.com/news/the-revolution-will-not-be-crystallized-a-new-method-sweeps-through-structural-biology-18335>. (Accessed on 12/08/2016).
- [Che] Vadim Cherezov. High resolution crystal structure of an engineered human beta2-adrenergic g protein-coupled receptor. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2583103/>. (Accessed on 12/08/2016).
- [Cla] Brown Clancy. Translation: Dna to mrna to protein. <http://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393>. (Accessed on 12/08/2016).
- [Dis] Hemoglobin oxygen dissociation curve. <https://lh3.googleusercontent.com/YU5tHRgL-bHg8GVaiguiPsdaTyxJ4ZwCT1dfhMdhrxhNOUIcYXw4wmbjRhdbNPiuhC98Xw=s114>. (Accessed on 12/09/2016).
- [dna] Dna. <http://www.bbc.co.uk/staticarchive/4717def808b61d7d2c90c11cea34c807928e101c.jpg>. (Accessed on 12/09/2016).
- [Ern15] Rolf Erni. *Aberration-Corrected Imaging in Transmission Electron Microscopy*. IMPERIAL COLLEGE PRESS, May 2015.
- [fou] Quaternary hemoglobin structure. [http://classconnection.s3.amazonaws.com/535/flashcards/3547535/png/quaternary\\_hemoglobin-141DBD963BC0249761D.png](http://classconnection.s3.amazonaws.com/535/flashcards/3547535/png/quaternary_hemoglobin-141DBD963BC0249761D.png). (Accessed on 12/09/2016).
- [Fre] Terry Frey. Biology 750: Electron microscopy. <http://www.sci.sdsu.edu/TFrey/Bio750/ElecMicroscopy.html>. San Diego State University. (Accessed on 12/08/2016).

- [Hem] Heme crystal. <http://www.proteopedia.org/wiki/index.php/Hemoglobin>. <http://www.proteopedia.org/wiki/index.php/Hemoglobin>, note = (Accessed on 12/09/2016).
- [Jac] Milne Jacqueline. Cryo-electron microscopy: A primer for the non-microscopist. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3537914/>. (Accessed on 12/08/2016).
- [Jena] Grant Jensen. Getting started in cryo-em: Contrast transfer function. <https://www.coursera.org/learn/cryo-em/lecture/XqrX4/the-contrast-transfer-function>. California Institute of Technology. (Accessed on 12/08/2016).
- [Jenb] Grant Jensen. Getting started in cryo-em: Ctf correction. <https://www.coursera.org/learn/cryo-em/lecture/hzV2T/ctf-correction>. California Institute of Technology. (Accessed on 12/08/2016).
- [Jenc] Grant Jensen. Getting started in cryo-em: Envelopes. <https://www.coursera.org/learn/cryo-em/lecture/VnFM9/envelopes>. California Institute of Technology. (Accessed on 12/08/2016).
- [Li] National Heart Lung and Blood institute. What is sickle cell disease? <https://www.nhlbi.nih.gov/health/health-topics/topics/sca>. (Accessed on 12/08/2016).
- [Lib] Chemistry LibreTexts. X-ray crystallography. [http://chem.libretexts.org/Core/Analytical\\_Chemistry/Instrumental\\_Analysis/Diffraction/X-ray\\_Crystallography](http://chem.libretexts.org/Core/Analytical_Chemistry/Instrumental_Analysis/Diffraction/X-ray_Crystallography). (Accessed on 12/08/2016).
- [MF] Australian Microscopy and Microanalysis Research Facility. Spherical aberration. <http://www.ammrf.org.au/myscope/tem/background/concepts/problems/spherical.php>. (Accessed on 12/08/2016).
- [NMR] Nrm visualization of hemoglobin. <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure>. (Accessed on 12/09/2016).
- [Ras] Soren Rasmussen. Crystal structure of the human beta-2 adrenergic g-protein-coupled receptor. <http://www.nature.com/nature/journal/v450/n7168/full/nature06325.html>. (Accessed on 12/08/2016).
- [Rus] Scott D. Russell. Limits to resolution in the electron microscope. <http://www.ou.edu/research/electron/bmz5364/resolutn.html>. Oklahoma University. (Accessed on 12/08/2016).
- [sec] Secondary protein structure. <http://oregonstate.edu/instruction/bi314/summer09/Fig-02-19-0.jpg>. (Accessed on 12/09/2016).
- [Suz] Clancy Suzanne. Dna transcription. <http://www.nature.com/scitable/topicpage/dna-transcription-426>. (Accessed on 12/08/2016).

- [tra] Translation. [http://images.slideplayer.com/37/10737408/slides/slide\\_30.jpg](http://images.slideplayer.com/37/10737408/slides/slide_30.jpg). (Accessed on 12/09/2016).
- [XDi] X-ray diffraction of dna. <https://writelatex.s3.amazonaws.com/zfknstjncjpq/uploads/8873/8917088/1.jpg>. (Accessed on 12/09/2016).
- [Zha] Gongpu Zhao. Mature hiv-1 capsid structure by cryo-electron microscopy. <http://www.ks.uiuc.edu/Publications/Papers/paper.cgi?tbcodes=ZHA02013>. (Accessed on 12/08/2016).