Appendix

Supplementary Information

A.1 Introduction

The appendix provides further details about the results of Prec refinement cycles, the effect of additional images in conventional reconstructions on the Ewald sphere resolution limit, and a comparison of Ewald sphere resolution limit predictions with reconstructions from simulated data in the first three sections. This information could not be included with Chapter 3 due to the brevity required of academic papers. In addition, the orientation conventions of software packages used during the testing of CPV, ε15, and DLP are described. Lastly, a list of all the important programs used in Chapters 2 and 3 is provided.

A.2 Prec Refinement in Practice

As described in Chapter 3, Prec possesses an iterative capability, which allows for errors due to the Ewald sphere curvature in the 3D Fourier transform (FT) of the reconstruction to be reduced by successive applications of the Prec algorithm.

In order to determine the significance of an additional refinement cycle, the improvement in the FSC curves of reconstructions of one application of the refinement algorithm versus an additional refinement were calculated for simulated data sets of 25, 50, 100, 250, 500, 1000, 2500, and 5000 images using Prec in Bsoft on Ewald projections of
FMDV at 15 kV. The results of the tests indicated that the additional refinement produced an insignificant improvement in the FSC curves and this improvement decreased as the number of images used increased (Figure A-1).

To understand these results, we observe the form of the Fourier values after the first iteration as described in Chapter 3:

\[ F_{R_0} \approx \bar{F}_R + \varepsilon \]  \hspace{1cm} (1)

where \( \bar{F}_R \) is the average \( F_{R_k} \) and \( \varepsilon \) is the residual error which consists of the average of the \( F_{L_k} (\alpha - i \beta)^2 e^{-i \alpha x} \) terms, which is a random walk with step size of approximately \( |F_{L_k}| \).

The residual error after the first iteration falls off as \( \sim \frac{1}{\sqrt{N}} \), thus the error is small for large numbers of images and only small improvements can be expected from additional iterations.

In practice, large numbers of images, on the order of \( 10^4 \) (Jiang, Baker et al. 2008; Yu, Jin et al. 2008; Zhang, Settembre et al. 2008), are used for reconstructions that achieve high resolution, thus no additional refinement is necessary.

A.3 The Effect of the Ewald Sphere Resolution Limit on Conventional Algorithms

In Chapter 3, the Ewald sphere curvature problem was characterized by observing the resolution achieved by conventional algorithms as the number of images with significant Ewald sphere curvature increased. Reconstructions were generated using sets of 25, 50, 100, 250, 500, 1000, 2500, and 5000 multi-slice images at 15 kV. The results of these
tests (Figure A-2) indicated that, regardless of how many images were used, the Ewald sphere curvature problem could not be overcome by additional images.

In contrast, current state-of-the-art high-resolution reconstructions of large particles (~700–750 Å in diameter) (Jiang, Baker et al. 2008; Yu, Jin et al. 2008; Zhang, Settembre et al. 2008) have not reached the Ewald sphere resolution limit of ~ 2.5 Å as predicted by our simulations of a 754 Å diameter virus particle at 300 kV, despite the large number of images being used. Once these limits are approached, significant improvements in resolution should be observed without an increase in the number of images when the Prec algorithm is applied.

A.4 Comparison of Ewald Sphere Resolution Limit Predictions with Simulations

Currently there are two formulas for predicting the resolution limits imposed by the curvature of the Ewald sphere. The first is an envelope function by Jensen and Kornberg (Jensen and Kornberg 2000), which indicates the percentage of information content remaining as resolution increases. The second formula by DeRosier (DeRosier 2000) indicates a resolution limit. A third approach to predicting the resolution limit is through simulations where Ewald projections of a model generated from pdb files are used to produce a reconstruction, which is subsequently compared with a reference model using an FSC curve (Chapter 3).

According to the resolution of the reconstructions of simulated data sets (Figure A-3), the simulation method predicts the highest resolution limits due to the Ewald sphere
curvature, indicating that both formula predictions may be too strict. When taking into account only the Ewald sphere curvature effect, the simulation method is the most accurate as it produces resolution limits without making any other assumptions about the information content and also simulates entirely the reconstruction process. Its drawback is that it requires a large amount of computation time in the generation of simulated images and the reconstruction process.

A.5 Icosahedral Symmetry Conventions

During the testing of Prec, four image-processing packages were used. These were Bsoft (Heymann 2001), IMIRS (Liang, Ke et al. 2002), EMAN (Ludtke, Baldwin et al. 1999) and FREALIGN (Grigorieff 2007). Prec was implemented in all the packages except FREALIGN, as it possessed its own Ewald sphere correction functionality. The packages were chosen because the three highest resolution reconstructions by cryo-EM to date, CPV, DLP, and ε15 were achieved using IMIRS, FREALIGN, and EMAN, respectively, while Bsoft was used for testing and generating simulated images.

For images to be used in the reconstruction process, their orientations have to be specified. While a standardization of these conventions has been proposed (Heymann, Chagoyen et al. 2005), each of the packages still possessed their own orientation conventions. There are multiple ways that orientations can be specified, one way is by defining three Euler angles (used in IMIRS, EMAN, and FREALIGN) that correspond to rotation matrices such as
\[
R_x(\alpha) = \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \alpha & \sin \alpha \\
0 & -\sin \alpha & \cos \alpha
\end{pmatrix}
\]

(2)

\[
R_y(\alpha) = \begin{pmatrix}
\cos \alpha & 0 & -\sin \alpha \\
0 & 1 & 0 \\
\sin \alpha & 0 & \cos \alpha
\end{pmatrix}
\]

(3)

\[
R_z(\alpha) = \begin{pmatrix}
\cos \alpha & \sin \alpha & 0 \\
-\sin \alpha & \cos \alpha & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

(4)

where \(R_x(\alpha)\), \(R_y(\alpha)\), and \(R_z(\alpha)\) represent right-handed rotations of the axes by angle \(\alpha\) around the x, y, and z axes, respectively.

The alternative approach, which is used in Bsoft, is to define an axis of rotation by a normalized 3D vector known as a “view vector” and an angle of rotation.

The confusion which surrounds the use of Euler angles is due to the numerous possible combinations of rotation axes and angles directions that are possible. Through careful examination of the software code, the Euler angle conventions (Table A-1), as well as their order of listing in the various orientation file formats (Table A-2), were determined.

In addition to the different Euler conventions, each of the packages had different reference orientations (Table A-3), i.e., when all three Euler angles are equal to zero.

Once the correct conventions had been determined for each of the packages, the conversion of angles between packages was straightforward for Bsoft (I90), IMIRS, and FREALIGN (I2). However, conversions from EMAN to Bsoft (I90) required additional
rotations of $R_z(-90^\circ)$, $R_z(31.7175^\circ)$, and $R_z(90^\circ)$ before the application of EMAN Euler angles, due to a different reference orientation.

A.6 List of Important Programs in Peach and Prec

**Peach — Distributed computation system**

- **Pjobd**: Job daemon
- **Pserv**: Job server
- **Pview**: Interactive client
- **Psubmit**: Client for submission of jobs

**Prec in Bsoft**

- **Brec**: Multi-threaded version of Breconstruct
- **Prec**: Multi-threaded implementation of Prec
- **Pref**: Multi-threaded implementation of Prec refinement loops
- **Ewald_proj**: Generates Ewald projections

**Prec in IMIRS**

- **Prec**: Implementation of Prec
- **Pref**: Implementation of Prec refinement loops
- **Reconstruct_ast**: Modified version of reconstruct with astigmatism correction
- **Prec_ast**: Modified version of Prec with astigmatism correction
Prec in EMAN

Make3d  Contains implementation of Prec compatible with multi-threaded and (modified distributed computation capabilities of EMAN version)

Euler Conversion Programs

Eman_to_bsoft  Converts EMAN Euler angles to Bsoft view vector and angle
Bsoft_to_eman  Converts Bsoft view vector and angle to EMAN Euler angles

A.7 References


A.8 Figures and Tables

![Graph showing improvements in FSC with an additional refinement loop.](image)

Figure A-1. **Effect of additional refinement loop.** Improvements in FSC for reconstructions using 25, 50, 100, 250, 500, 1000, 2500, and 5000 Ewald projections at an acceleration voltage of 15 kV using Prec in Bsoft demonstrate the decreasing significance of the improvement between the first and second cycles of refinement. When a large number of images are used in the reconstruction, the additional refinement has no significant effect.
Figure A-2. **Effect of number of images on Ewald sphere curvature resolution limit.**

FSC curves of conventional reconstructions performed using `reconstruct` of the IMIRS package, using 15 kV multi-slice images, demonstrate that the maximum resolution imposed by the curvature of the Ewald sphere cannot be overcome by increasing the number of images. Insets a–h show volume extracts of a single α-helix from the reconstruction from 25, 50, 100, 250, 500, 1000, 2500, and 5000 images, respectively.
Figure A-3. **Comparison of Ewald sphere resolution limitations.** The comparison of the maximum achievable resolutions at acceleration voltages of 15 (red), 25 (blue), 50 (green), and 100 kV (black) for the foot and mouth virus and 300 kV (pink) for the reovirus core using the FSCs of the reconstructions (solid curves) from Ewald projections, the sinc envelopes by Jensen and Kornberg (dashed curves), and the empirical threshold by DeRosier (vertical lines), where the dimensionless constant $p$ is 0.7. The envelopes and the limit formula both predict limits significantly lower than the resolutions achieved by reconstructions from simulated images.
<table>
<thead>
<tr>
<th>Software Package</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; angle</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; axis</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Angle</th>
<th>4&lt;sup&gt;th&lt;/sup&gt; axis</th>
<th>5&lt;sup&gt;th&lt;/sup&gt; Angle</th>
<th>6&lt;sup&gt;th&lt;/sup&gt; axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bsoft</td>
<td>Phi</td>
<td>Z</td>
<td>Theta</td>
<td>Y</td>
<td>Psi</td>
<td>Z</td>
</tr>
<tr>
<td>IMIRS</td>
<td>Phi</td>
<td>-Z</td>
<td>Theta</td>
<td>Y</td>
<td>Omega</td>
<td>Z&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>EMAN</td>
<td>Az</td>
<td>Z</td>
<td>Alt</td>
<td>X</td>
<td>Phi</td>
<td>Z</td>
</tr>
<tr>
<td>FREALIGN</td>
<td>Phi</td>
<td>Z</td>
<td>Theta</td>
<td>Y</td>
<td>Psi</td>
<td>Z</td>
</tr>
</tbody>
</table>

Table A-1. Table of Euler angle conventions

<sup>1</sup>IMIRS Omega angle requires an addition of 180° for it to be correct.

<table>
<thead>
<tr>
<th>Software Package</th>
<th>Orientation File Format</th>
<th>Orientation File Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bsoft</td>
<td>View vector and angle (degrees)</td>
<td>Star</td>
</tr>
<tr>
<td>IMIRS</td>
<td>Theta, Phi, Omega (degrees)</td>
<td>Dat</td>
</tr>
<tr>
<td>EMAN</td>
<td>Alt, Az, Phi (degrees)</td>
<td>Lst</td>
</tr>
<tr>
<td>FREALIGN</td>
<td>Psi, Theta, Phi (degrees)</td>
<td>Par</td>
</tr>
</tbody>
</table>

Table A-2. Table of orientation file formats
<table>
<thead>
<tr>
<th>Software Package</th>
<th>Symmetry Axis along Z-Axis</th>
<th>Additional Symmetry Axis for Orientation Clarification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bsoft (I)</td>
<td>2-fold</td>
<td>5-fold axis along (0, 1, ϕ) vector</td>
</tr>
<tr>
<td>Bsoft (I90)</td>
<td>2-fold</td>
<td>5-fold axis along (1, 0, ϕ) vector</td>
</tr>
<tr>
<td>IMIRS</td>
<td>2-fold</td>
<td>5-fold axis along (1, 0, ϕ) vector</td>
</tr>
<tr>
<td>EMAN</td>
<td>5-fold</td>
<td>2-fold axis along (0, -1, ϕ) vector</td>
</tr>
<tr>
<td>FREALIGN (I)</td>
<td>2-fold</td>
<td>5-fold axis along (0, 1, ϕ) vector</td>
</tr>
<tr>
<td>FREALIGN (I2)</td>
<td>2-fold</td>
<td>5-fold axis along (1, 0, ϕ) vector</td>
</tr>
</tbody>
</table>

Table A-3. **Table of reference orientations**

1 Bsoft (I90) was used for all simulations and for compatibility with other packages

2 IMIRS uses a 5-fold axis along z internally during the reconstruction process

3 EMAN uses a 5-fold axis along z but 2-fold axis along (1, 0, golden ratio) internally

4 FREALIGN (I2) was used during rotavirus reconstruction

ϕ is the golden ratio which is defined as $\frac{1 + \sqrt{5}}{2}$