

COMPUTATIONAL CHALLENGES IN HIGH-RESOLUTION CRYO-ELECTRON
MICROSCOPY

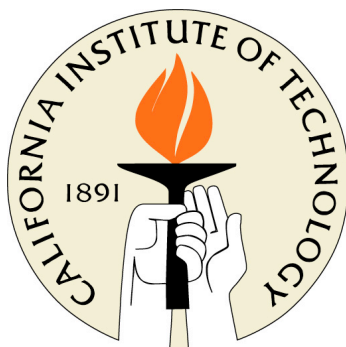
Thesis by

Peter Anthony Leong

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2009

(Defended Aug 04, 2008)

© 2009

Peter Anthony Leong

All Rights Reserved

To God, who made all this possible

Acknowledgements

I would like to thank Prof. Grant Jensen for being an excellent advisor and role model for me during my time at Caltech. He has always shown tremendous understanding, kindness, and encouragement throughout, and supported me to the fullest extent in my research, career, and personal development — making this a wonderful life experience.

I am also very indebted to Drs. Bernard Heymann and Andrew Rawlinson. Bernard, who was a mentor to me when I first joined the lab, taught me much about computer hardware and software. Andy, who worked closely with me during the middle of my thesis work, helped me greatly in our discussions about the mathematics and physics related to our research work. Their mentoring has been very significant in my development as a scientist. In addition, I would also like to thank all my other lab mates, both past and present, who have been wonderful colleagues and friends. This thesis work could not have been completed without their help.

I would also like to thank my thesis and candidacy committee members Profs. Scott Fraser, Douglas Rees, Brent Fultz, Robert Phillips, and Z. Hong Zhou. Their advice and feedback about my research projects and about academics in general have been extremely helpful.

I would also like to thank my collaborators Profs. Hong Zhou, Wen Jiang, and Nikolaus Grigorieff, and their respective lab members, especially Drs. Xuekui Yu and Weimin Wu, for supporting me and helping in the completion of my thesis work.

Lastly and most importantly, I would like to thank my family, especially my parents, for all the love, support, and encouragement they have always shown me.

Abstract

To avoid the challenges of crystallization and the size limitations of NMR, it has long been hoped that single-particle cryo-electron microscopy (cryo-EM) would eventually yield atomically interpretable reconstructions. For the most favorable class of specimens (large icosahedral viruses), two of the key obstacles are the large computational requirements of high-resolution reconstructions and the curvature of the Ewald sphere, which leads to a breakdown of the projection theorem used by conventional 3D reconstruction programs. Here, two solutions to these obstacles are presented.

First, a simple distributed processing system named Peach was developed to meet the rising computational demands of modern structural biology (and other) laboratories without additional expense by using existing hardware resources more efficiently. A central server distributes jobs to idle workstations in such a way that each computer is used maximally, but without disturbing intermittent interactive users. As compared to other distributed systems, Peach is simple, easy to install, easy to administer, easy to use, scalable, and robust. While it was designed to queue and distribute large numbers of small tasks to participating computers, it can also be used to send single jobs automatically to the fastest currently available computer and/or survey the activity of an entire laboratory's computers. Tests of robustness and scalability are reported, as are three specific cryo-EM applications where Peach enabled projects that would not otherwise have been feasible without an expensive, dedicated cluster.

Second, an iterative refinement reconstruction algorithm, *Prec*, is described that overcomes the curvature of the Ewald sphere resolution limitation by averaging information from images recorded from different points of view, as are present in typical micrographs. *Prec* was implemented in the popular software packages IMIRS, EMAN, and Bsoft. In preliminary tests with both simple and multi-slice simulated images, *Prec* overcame the curvature problem even in the presence of noise. *Prec* was then used to refine the three recently published, ~ 4 Å resolution, icosahedral virus reconstructions from experimental cryo-EM images, but unfortunately no significant improvements in resolution were realized. Further simulations showed that limitations other than the Ewald sphere curvature problem must still be dominant in these experimental studies.

Table of Contents

Title Page	(not numbered)
Copyright Page	ii
Acknowledgements	iv
Abstract	vi
Table of Contents	viii
List of Figures and Tables	xii
1. Introduction	1
1.1. Structural Biology	1
1.2. Structure Determination Techniques	1
1.3. Cryo-Electron Microscopy	3
1.4. Reconstruction Theory	5
1.5. Resolution Measures	6
1.6. Resolution Limitations	9
1.7. Instrumentation Progress	9
1.8. Progress in Processing Techniques	10
1.9. Approaching Atomic Resolution by Cryo-EM	11
1.10. Icosahedral Virus Structures	11
1.11. Viruses	12
1.12. Approaching Atomic Resolution by Single Particle Analysis	14
1.13. Computational Complexity of 3D Reconstruction Algorithm	15
1.14. Parallel Computation	16

1.15. Distributed Computation	18
1.16. Hybrid Approach	20
1.17. Depth of Field and Ewald Sphere Curvature	20
1.18. Viruses Structures Limited by Ewald Sphere Curvature	24
1.19. References	25
1.20. Figures	32
2. Peach: A Simple Perl-Based System For Distributed Computation And Its Application To Cryo-EM Data Processing	34
2.1. Summary	35
2.2. Introduction	36
2.3. Design	38
2.3.1. Design Philosophy	38
2.3.2. Implementation	39
2.3.3. Information Flow	39
2.3.4. The Job Server	40
2.3.5. The Job Clients	41
2.3.6. Use of Existing Capabilities	41
2.3.7. Security	41
2.3.8. Peach Administration	42
2.4. Tests and Results	43
2.4.1. Installation and Test Environments	43
2.4.2. Cryo-EM Applications	44
2.4.3. Robustness	46

2.4.4. Scalability	47
2.5. Discussion	48
2.6. Acknowledgements	52
2.7. References	53
2.8. Figures	56
3. Chapter 3: Prec: An Iterative Reconstruction Method For Correction Of The Ewald Sphere	60
3.1. Abstract	61
3.2. Introduction	62
3.3. Results	64
3.3.1. The Ewald Curvature Problem and Symbols Used	64
3.3.2. The Paraboloid Method in the Context of 3-D Reconstruction	67
3.3.3. The Prec Algorithm	69
3.3.4. Implementation of the Prec Algorithm	71
3.3.5. Tests on Simulated Images	73
3.3.6. Application to the CPV, $\epsilon 15$, and DLP reconstructions	77
3.4. Discussion	79
3.5. Acknowledgements	81
3.6. References	82
3.7. Figures	86
4. Conclusion	91
4.1. Progression of Single Particle Analysis	91

4.2. Hybrid Approach to Address Lack of Computational Power	92
4.3. Paraboloid Reconstruction Algorithm to Address Ewald Sphere Curvature	92
4.4. References	93
A. Appendix	95
A.1. Introduction	95
A.2. Prec Refinement in Practice	95
A.3. Number of Images and Effect on Ewald Sphere	96
A.4. Comparison of Ewald Sphere Resolution Limit Predictions	97
A.5. Icosahedral Symmetry Conventions	98
A.6. List of Important Programs	100
A.7. References	101
A.8. Figures and Tables	103

List of Figures and Tables

Figure 1-1 Flow chart of simplified reconstruction process	32
Table 1-1 Table of biological structural features observable at different resolutions	32
Table 1-2 Table of viruses known to infect humans	33
Figure 2-1 Schematic drawing of the setup and information flow in the testing of Peach	56
Figure 2-2 An example cryo-EM image processing project made feasible by Peach	57
Figure 2-3 An example image simulation project managed by Peach	58
Figure 2-4 Scalability	59
Figure 3-1 The Ewald sphere and Prec algorithm	86
Figure 3-2 Prec overcomes the curvature problem in Ewald projections	87
Figure 3-3 Prec overcomes the curvature problem in multi-slice images and in the presence of noise	88
Figure 3-4 Application of Prec to experimental images: 3D reconstruction of CPV	89
Figure 3-5 Reconstructions of the 754 Å diameter Reovirus from 300 kV simulated images	90
Figure A-1 Effect of addition refinement loop	103
Figure A-2 Comparison of Ewald sphere resolution limitations	104
Figure A-3 Effect of number of images on Ewald sphere curvature resolution limit	105

Table A-1 Table of Euler angle conventions.	106
Table A-2 Table of orientation file formats	106
Table A-3 Table of reference orientations	107