

SOLVING ATOMIC STRUCTURE USING STATISTICAL MECHANICAL SEARCHES ON  
X-RAY SCATTERING DERIVED POTENTIAL ENERGY SURFACES

by

Christopher James Wright

Bachelor of Science  
Brown University 2014

---

Submitted in Partial Fulfillment of the Requirements

for the Degree of Masters of Science in

Chemical Engineering

College of Engineering and Computing

University of South Carolina

2016

Accepted by:

Xiao-Dong Zhou, Major Professor

Thomas Vogt, Committee Member

Mark Uline, Committee Member

Jochen Lauterbach, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

© Copyright by Christopher James Wright, 2016  
All Rights Reserved.

## DEDICATION

## ACKNOWLEDGMENTS

## ABSTRACT

## TABLE OF CONTENTS

DEDICATION . . . . .	iii
ACKNOWLEDGMENTS . . . . .	iv
ABSTRACT . . . . .	v
LIST OF TABLES . . . . .	viii
LIST OF FIGURES . . . . .	ix
TODO LIST . . . . .	1
CHAPTER 1 STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES . . . . .	5
1.1 Introduction . . . . .	5
1.2 Potential Energy Surfaces . . . . .	5
1.3 Ensembles . . . . .	7
CHAPTER 2 ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION . . . . .	12
2.1 Theory . . . . .	12
2.2 Computation . . . . .	14
CHAPTER 3 BENCHMARKING . . . . .	18

3.1 PDF . . . . .	18
3.2 PDF with ADPs . . . . .	27
CHAPTER 4 X-RAY TOTAL SCATTERING DATA ACQUISITION AND PROCESSING . . . . .	33
4.1 Introduction . . . . .	33
4.2 Data Storage and Management . . . . .	33
CHAPTER 5 ANNEALING AND AGGREGATION OF 2NM AU NANOPARTICLES . . . . .	50
5.1 Experiments . . . . .	50
5.2 Data Processing . . . . .	50
5.3 Data Analysis . . . . .	50
5.4 Simulation . . . . .	50
5.5 Structural Analysis . . . . .	50
5.6 Conclusions . . . . .	50
CHAPTER 6 PHASE CHANGES AND ANNEALING DYNAMICS OF $\text{Pr}_2\text{NiO}_4$ AND ITS DERIVATIVES . . . . .	51
6.1 Experiments . . . . .	51
6.2 Data Processing . . . . .	51
6.3 Data Analysis . . . . .	51
6.4 Simulation . . . . .	70
6.5 Conclusions . . . . .	70
CHAPTER 7 CONCLUSION . . . . .	72

## LIST OF TABLES

## LIST OF FIGURES

Figure 2.1 Comparison of the CPU and GPU chip architectures . . . . .	15
Figure 3.1 Au <sub>55</sub> PDF fitting of DFT-optimized <i>c</i> -Au <sub>55</sub> . (a) the final structural solution ( $Rw=0.3\%$ ) with bond lengths color-coded by step of 0.05Å, (b) the target PDF(blue dots) overlaid with the PDF of the final structure (solid red lines) with the difference in green lines offset below, ?? the radial bond distribution, and (d) bond angle distribution. . . . .	22
Figure 3.2 Au <sub>55</sub> PDF fitting of surface-disordered Au <sub>55</sub> . a) the target structure, b) the final structural solution ( $Rw=0.6\%$ ), c) the comparison of PDFs, d) the CN distribution with the number of atoms in either the core or the surface, e) the radial bond distribution, and f) the bond angle distribution. . . . .	28
Figure 3.3 Similar to figure 3.2 for DFT-optimized amorphous Au <sub>55</sub> . . . . .	29
Figure 3.4 Similar to figure 3.2 for Au <sub>102</sub> as in DFT-optimized Au <sub>102</sub> MBA <sub>44</sub> cluster. . . . .	30
Figure 3.5 Similar to Fig. 3.6 with Marks decahedron as the starting structure. . . . .	31
Figure 3.6 . . . . .	32
Figure 4.1 Database Loading Workflow. Data is loaded from various sources, including images and text files, into the FileStore and Meta-dataStore databases. Data is then retrieved from the databases using the databroker. . . . .	34
Figure 4.2 Scattering onto a flat detector . . . . .	35
Figure 4.3 $Q$ resolution as a function of $Q$ . . . . .	36
Figure 4.4 Number of pixels as a function of $Q$ , binned at the $Q$ resolution of the detector. . . . .	37

Figure 4.5 Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the standard even bin mask and b) the $Q$ resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles. . . . .	39
Figure 4.6 Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the standard even bin mask and b) the $Q$ resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles. . . . .	39
Figure 4.7 Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the standard even bin mask and b) the $Q$ resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles. . . . .	40
Figure 4.8 Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the standard even bin mask and b) the $Q$ resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles. . . . .	40
Figure 4.9 Generated beamstop holder masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	41
Figure 4.10 Generated beamstop holder masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	41
Figure 4.11 Generated beamstop holder masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	41
Figure 4.12 Generated beamstop holder masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	42
Figure 4.13 Generated beamstop holder masks which is rotated away from verticle . . . . .	42
Figure 4.14 Masked experimental data. a) the raw image, b) the mask . . . . .	43
Figure 4.15 Masked experimental data with Pt single crystal signal. a) the raw image, b) the mask . . . . .	43

Figure 4.16 Masked experimental data with Pt single crystal signal using figure's 4.14 as a starting mask. a) the raw image, b) the mask . . . 44

Figure 4.17 Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with no mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{ \AA}^{-1}$ to  $31 \text{ \AA}^{-1}Q$  range for the mean and median, f)  $28 \text{ \AA}^{-1}$ to  $31 \text{ \AA}^{-1}Q$  range for the standard deviation . . . 47

Figure 4.18 Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with only an edge mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{ \AA}^{-1}$ to  $31 \text{ \AA}^{-1}Q$  range for the mean and median, f)  $28 \text{ \AA}^{-1}$ to  $31 \text{ \AA}^{-1}Q$  range for the standard deviation . . . . . 48

Figure 4.19 Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced combining an edge mask and the automatically generated mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{ \AA}^{-1}$ to  $31 \text{ \AA}^{-1}Q$  range for the mean and median, f)  $28 \text{ \AA}^{-1}$ to  $31 \text{ \AA}^{-1}Q$  range for the standard deviation . . . . . 49

## TODO LIST

2	Why is atomistic engineering important . . . . .	2
3	Barriers to atomistic engineering . . . . .	2
4	How are we going to attack this problem . . . . .	2
5	Fix the figures so that they don't give the full caption . . . . .	4
6	This entire section needs some rewriting to distinguish this from the paper .	18
7	Put this somewhere . . . . .	22
8	masking parameters . . . . .	51
9	integration parameters . . . . .	51
10	PDF parameters . . . . .	51

11

## INTRODUCTION

12

### Why is atomistic engineering important

13     Engineering materials and chemicals on the atomic scale has been a goal for the  
14    chemistry, physics, materials science, and chemical engineering fields long before the  
15    advent of nanomaterials. Realizing this goal could lead to durable fuel cell catalysts,  
16    more bioavailable pharmaceuticals, and radiation damage resistant spacecraft shielding.

17

### Barriers to atomistic engineering

18     Before we can even think of making atomistically exact structures, durable struc-  
19    tures, or structures which change in reproducible ways, we need to know the atomic  
20    structure exactly.

21

### How are we going to attack this problem

22     This work addresses these issues by developing a methodology for solving the  
23    structure of nanomaterials by matching experimental x-ray scattering data with sim-  
24    ulated atomic structures.

25     Chapter 1 develops the statistical mechanical system used to match the theoretical  
26    structure. §1.2 focuses on the development of potential energy surfaces, including  
27    potential energy and force equations, which have minima where experimental results  
28    and simulated structures agree the most. §1.3 will discuss statistical mechanical  
29    ensembles which are used to search for minima on the potential energy surface.

30     Chapter 2 will discuss the mathematical and computational development of the  
31    atomic pair distribution function (PDF). §2.2 will focus on the rapid graphical pro-  
32    cessing unit based calculation of the PDF and its gradients.

33     Chapter 3 will discuss the benchmarking of the the combined statistical mechan-

34 ical optimizer and PDF calculation systems against a series of theoretical nanoparti-  
35 cles, focusing on understanding limitations of the method and structure reproduction.

36 Chapter 4 will focus on the aquesition of experimental data, its management, and  
37 processing. §4.2, 4.2, and 4.2 will discuss the derivation of the  $Q$  resolution function,  
38 the automated masking of 2D area detectors for x-ray total scattering measurements  
39 using the previously derrived  $Q$  resolution, and the impact of different averaging  
40 methods and masks on azimuthal integration, respecitvly.

Fix the  
figures  
so that  
they  
don't  
give  
the full  
caption

42

# CHAPTER 1

43

## STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES

45 1.1 INTRODUCTION

46 The approach taken in this work for solving the atomic structures of materials is  
47 one of optimization. The positional variables of the system are optimized so as to  
48 minimize the value of a potential energy surface (PES). The

49 1.2 POTENTIAL ENERGY SURFACES

50 A PES simply describes the potential energy of the system as a function of all its  
51 relevant coordinates in phase space, essentially providing a mapping  $\mathbb{R}^n \rightarrow \mathbb{R}$ . Usually  
52 these coordinates are the positions of the atoms  $q$  and their conjugate momenta  $p$ .  
53 Note that there could be more variables associated with the system, for instance the  
54 magnetic moments of the atoms could play a role in describing the system. In this  
55 magnetic system there would be positional variables for the atomwise spin vectors  
56 and their "momenta". Application of the term "momenta" might seem odd here, as  
57 the magnetic spin does not have a mass or a velocity. However, since the magnetic  
58 "position" is defined on the PES we need to describe its conjugate variable to properly  
59 formulate Hamiltonian dynamics and the kinetic portion of the PES.

60 **Experimentally Derived Potential Energy Surfaces**

61 Generally PESs are obtained from purely computational experiments including: ab-  
62 initio DFT, classical approximations via the embedded atom method, or even param-  
63 eter driven models with experimentally fitted parameters. However, one can derive  
64 a PES from an experiment which describes how well the model reproduces the ex-  
65 perimental data. In this case one needs a theoretical and computational framework  
66 mapping the atomistic variables of the simulation to the same space of the data ob-  
67 tained from the experiment. This allows the experiment to be compared directly  
68 against the predicted data via an experimentally derived PES.

69 **Potentials**

70 For an experiment which produces 1D data, like powder diffraction, EXAFS or XPS,  
71 the implemented potentials are:

$$\chi^2 = \sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2 \quad (1.1)$$

$$Rw = \sqrt{\frac{\sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}}^2}} \quad (1.2)$$

$$\chi_{\text{INVERT}}^2 = \frac{1}{N} \sum_j \sum_r [A_{\text{obs}}(r) - \alpha A_{j,\text{calc}}(r)]^2 \quad (1.3)$$

$$\alpha = \frac{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}} A_{\text{calc}}}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}}^2} = \frac{\vec{A}_{\text{obs}} \cdot \vec{A}_{\text{calc}}}{|\vec{A}_{\text{calc}}|^2} \quad (1.4)$$

75 where  $A_{\text{calc}}$  and  $A_{\text{obs}}$  are the calculated and observed 1D experimental data and  $A_{j,\text{calc}}$   
76 is the calculated data for a single atom interacting with the other atoms of the system.  
77 Note that  $A_{\text{calc}}$  has a dependence on  $q$ , the positions of the system.

78 **Forces**

$$\vec{\nabla} \chi^2 = -2 \sum_{a=a_{\min}}^{a_{\max}} (\alpha \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} + A_{\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}})(A_{\text{obs}} - \alpha A_{\text{calc}}) \quad (1.5)$$

79

$$\vec{\nabla}Rw = \frac{Rw}{\chi^2} \sum_{a=a_{\min}}^{a_{\max}} (\alpha \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} + A_{\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}})(\alpha A_{\text{calc}} - (A_{\text{obs}})) \quad (1.6)$$

80

$$\frac{\partial \alpha}{\partial \gamma_{i,w}} = \frac{(\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} - 2\alpha \sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}})}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}}^2} \quad (1.7)$$

81

$$\vec{\nabla}\chi^2_{\text{INVERT}} = \frac{-2}{N} \sum_{a=a_{\min}}^{a_{\max}} \sum_j (\alpha \frac{\partial A_{j,\text{calc}}}{\partial \gamma_{i,w}} + A_{j,\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}})(A_{\text{obs}} - \alpha A_{j,\text{calc}}) \quad (1.8)$$

82 where  $\gamma_{i,w}$  is the  $i$ th arbitrary positional variable in the  $w$ th direction. The concept  
 83 of an "arbitrary positional variable" might seem a bit cumbersome but it allows us  
 84 to define the forces for any atomic parameter which can be represented as a vector  
 85 in 3-space. This comes in handy when trying to define the forces acting on variables  
 86 like anisotropic displacement parameters or atomic magnetic spins.

87 DISCUSS INVERT A BUNCH. ALSO COMPARE RW AND CHI\*\*2, POTEN-  
 88 TIALY WITH A FIGURE.

89 1.3 ENSEMBLES

90 While PESs describe which atomic configurations are the most desirable and how  
 91 the atoms would like to get there, the ensemble describes how the atoms move on  
 92 the PES. The abstraction of the PES from the ensemble is an important one, as it  
 93 allows for the reuse and exchange of both PESs and ensembles for a wide array of  
 94 problems. Statistical mechanical ensembles can be described in two ways, analytically  
 95 and stochastically. For long simulation times and fine enough numerical or analytical  
 96 integration these two descriptions should be identical. In either case one starts by  
 97 defining the Hamiltonian  $\mathcal{H}$  as the total energy of the system. Thus, the Hamiltonian  
 98 is described as the sum of the potential  $U(q)$  and kinetic  $K(p)$  energies, where  $q$  is  
 99 the positions of the atoms and  $p$  is their momenta

$$\mathcal{H}(q, p) = U(q) + K(p) \quad (1.9)$$

100 where  $K(p) = \frac{1}{2} \sum_i \frac{p_i^2}{m_i}$  and  $i$  denotes the  $i$ th particle. Analytically one generally defines  
 101 a partition function, which describes the sum of probabilities over all potential atomic  
 102 states.

$$\Xi = \sum_i P_i(q, p)$$

103 where  $P_i$  is the probability of the  $i$ th state and is a function of the total energy of  
 104 that state. This partition function can then be used to obtain the probability of any  
 105 specific state.

## 106 Hamiltonian Monte Carlo

In order to model dynamics we need to describe the motion of the particles in our system, thus:

$$\frac{dq_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} = p_i \quad (1.10)$$

$$\frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial q_i} = -\vec{\nabla}U \quad (1.11)$$

Using these equations we can derive the position and momentum vectors at any point in time using the leap-frog algorithm:

$$p_i(t + \delta t/2) = p_i(t) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t)) \quad (1.12)$$

$$q_i(t + \delta t) = q_i(t) + \delta t * p_i(t + \delta t/2) \quad (1.13)$$

$$p_i(t + \delta t) = p_i(t + \delta t/2) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t + \delta t)) \quad (1.14)$$

107 Note that  $\frac{\partial}{\partial q_i}$  is the gradient with respect to  $q$  where  $i$  denotes the  $i$ th atom being  
 108 moved. Using this notation the gradient is

$$\vec{\nabla}U = \begin{bmatrix} \frac{\partial U}{\partial q_{0,x}} & \frac{\partial U}{\partial q_{0,y}} & \frac{\partial U}{\partial q_{0,z}} \\ \vdots & \frac{\partial U}{\partial q_{i,w}} & \vdots \\ \frac{\partial U}{\partial q_{n,x}} & \frac{\partial U}{\partial q_{n,y}} & \frac{\partial U}{\partial q_{n,z}} \end{bmatrix} = \begin{bmatrix} \vec{\mathcal{F}}_0 \\ \vdots \\ \vec{\mathcal{F}}_i \\ \vdots \\ \vec{\mathcal{F}}_n \end{bmatrix} \quad (1.15)$$

109 where  $\frac{\partial}{\partial q_{i,w}}$  is the derivative with respect to  $q$  where  $w$  denotes direction of the deriva-  
 110 tive ( $x$ ,  $y$ , or  $z$ ),  $n$  is the number of atoms and  $U$  is the potential which depends on  
 111  $q$ , and  $\vec{F}_i$  is the "force" on the  $i$ th atom.

## 112 No-U-Turn-Sampling

## 113 Grand Canonical Ensemble

### 114 Ensemble description

115 In the Grand Canonical Ensemble (GCE) two sets of variables are allowed to change,  
 116 the atomic positions and the total number of atoms and their associated identities.  
 117 These two variables are controlled by temperature and chemical potential. The par-  
 118 tition function is

$$\Xi = e^{-\beta(\mathcal{H}+\mu)} \quad (1.16)$$

119 This is translated into a Monte Carlo system, producing Grand Canonical Monte  
 120 Carlo (GCMC).

### 121 Grand Canonical Monte Carlo

122 While the probabilities for atomic motion are the same as in the Canonical Ensemble,  
 123 the addition or removal of an atom have their own probabilities. For the addition of  
 124 an atom the probability is formally:

$$\min[1, \frac{V}{(N+1)\Lambda(T)^3} e^{-\beta\Delta U + \beta\mu}] \quad (1.17)$$

125 Similarly the removal of an atom has the probability:

$$\min[1, \frac{(N)\Lambda(T)^3}{V} e^{-\beta\Delta U - \beta\mu}] \quad (1.18)$$

126 However, both of these equations depend of the overall simulation volume and the  
 127 thermal wavelength, which is undesirable as these are not really properties that we

128 are of interest to these simulations. Thus, we roll them into the definition of the  
129 chemical potential, essentially setting the base chemical potential to counteract these  
130 effects. This makes certain that our simulation does not change if we change the  
131 overall cell volume. A GCMC move consists of creating a new atomic configuration,  
132 where an atom has been added or removed, and checking the above criteria. However,  
133 previous results have shown that this method is computationally expensive in dense  
134 liquids, and exceedingly expensive in solid materials. The long simulation times  
135 are due to the random nature of the atomic additions or removals which produce:  
136 over-tightly packed atoms, atoms in the middle of nowhere, or unphysical vacancies.  
137 These configurations are rejected by the GCMC criteria but their probability of being  
138 sampled is much higher than configurations which are lower in energy, since the  
139 number of incorrect ways to add/remove atoms is much larger than the correct ways.  
140 Thus we have implemented methods for biasing the atomic addition positions and  
141 the atomic removals toward configurations which are more likely to be accepted.

## 142 GCMC biasing

143 The first method is to remove some of the excess options from the probability pool.  
144 Initially the insertion positions are calculated at random using a random number gen-  
145 erator and scaled to the size of the simulation cell. This produces probabilities which  
146 have floating point level precision, which is effectively infinite. While this produces  
147 a potentially infinite number of ways to create energetically favorable configurations,  
148 the infinite ways to produce bad configurations is much larger. Thus we can limit this  
149 by moving to voxels. In this case atoms are added to the center of voxels which have  
150 a pre-set resolution, limiting our total number of valid addition points. While this  
151 could produce some problems with ergodicity, we avoid this by allowing the atoms to  
152 translate throughout the system. Each voxel has a probability of being tried:

$$P_{i,j,k} = \frac{xyz}{abc} \quad (1.19)$$

153 where  $x, y, z$  and  $a, b, c$  are the resolutions and cell side lengths in the cardinal di-  
154 rections, respectively. While this does help to limit the total probability space it  
155 does not tell us which voxels are likely to lead to better configurations, leading to  
156 many rejected atomic additions. To combat this issue we can weigh the individual  
157 voxels, giving more probability to voxels which show promise and less to those with  
158 less likelihood to be accepted.

159 The approach most likely to yield success would be to measure the change in  
160 potential energy associated with the addition of an atom at the center of the voxel  
161 where the probability of a voxel to be tried is:

$$P_{i,j,k} = \frac{e^{\beta\Delta U_{i,j,k}}}{\sum_{i,j,k} e^{\beta\Delta U_{i,j,k}}} \quad (1.20)$$

162 where  $\Delta U_{i,j,k}$  is the change in energy. However, calculating  $\Delta U_{i,j,k}$  can be particu-  
163 larly expensive, especially when calculating scattering from atomic positions. The  
164 computational expense can be mitigated by using a cheaper potential, if only for the  
165 evaluation of the voxel energy, as previously shown. Similar to previous work we can  
166 use the Lennard Jones potential to approximate the addition potential.

167

## CHAPTER 2

168

### ATOMIC PAIR DISTRIBUTION FUNCTION:

169

### THEORY AND COMPUTATION

170 2.1 THEORY

171 To properly understand the PDF and its limitations we need to derive its mathemat-  
172 ics. The following derivation has been performed numerous times but most recently  
173 and completely by Farrow and Billinge, it is reproduced here for clarity and com-  
174 pleteness.

175 **Derivation**

176 **Analytical Gradients**

177 Many optimization algorithms and simulations methodologies, including HMC, re-  
178 quire not only the potential energy of a given configuration but also the forces acting  
179 on that configuration. These forces are described by the gradient of potential energy  
180 of the system which in turn requires the gradient of the PDF. As previously shown the  
181 PDF is the Fourier Transform of the Debye equation. Since the Fourier Transform is  
182 expressed as an integral we can exchange the order of the gradient and the integral,  
183 allowing us to calculate the analytical gradient of the Debye equation and FFT the  
184 resulting function. The Debye equation, with a Debye-Waller vibrational correction  
185 is

$$F(Q) = \frac{1}{N\langle f \rangle^2} \sum_{j \neq i} f_i^*(Q) f_j(Q) \exp(-\frac{1}{2}\sigma_{ij}^2 Q^2) \frac{\sin(Qr_{ij})}{r_{ij}} \quad (2.1)$$

186 where

$$\sigma_{ij}^2 = (\vec{u}_{ij} * \hat{d}_{ij})^2 \quad (2.2)$$

$$\vec{u}_{ij} = \vec{u}_i - \vec{u}_j \quad (2.3)$$

$$\hat{d}_{ij} = \frac{\vec{d}_{ij}}{r_{ij}} \quad (2.4)$$

$$r_{ij} = \|\vec{d}_{ij}\| \quad (2.5)$$

$$\vec{d}_{ij} = \begin{bmatrix} q_{ix} - q_{jx} \\ q_{iy} - q_{jy} \\ q_{iz} - q_{jz} \end{bmatrix} \quad (2.6)$$

187 where  $Q$  is the scatter vector,  $f_i$  is atomic scattering factor of the  $i$ th atom, and  $r_{ij}$   
 188 is the distance between atoms  $i$  and  $j$  and has  $q$  dependence. For simplicities sake  
 189 we will break up  $F(Q)$  so that

$$F(Q) = \alpha \sum_{j \neq i} \beta_{ij} \tau_{ij} \Omega_{ij} \quad (2.7)$$

190 where

$$\alpha = \frac{1}{N\langle f \rangle^2} \quad (2.8)$$

$$\beta_{ij} = f_i^*(Q) f_j(Q) \quad (2.9)$$

$$\tau_{ij} = \exp(-\frac{1}{2} \sigma_{ij}^2 Q^2) \quad (2.10)$$

$$\Omega_{ij} = \frac{\sin(Qr_{ij})}{r_{ij}} \quad (2.11)$$

191 The derivatives are as follows:

$$\frac{\partial}{\partial q_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \left( \frac{\partial \tau_{ij}}{\partial q_{i,w}} \Omega_{ij} + \tau_{ij} \frac{\partial \Omega_{ij}}{\partial q_{i,w}} \right) \quad (2.12)$$

192 where

$$\frac{\partial \Omega_{ij}}{\partial q_{i,w}} = \frac{Q \cos(Qr_{ij}) - \Omega_{ij}}{r_{ij}^2} (q_{i,w} - q_{j,w}) \quad (2.13)$$

$$\frac{\partial \tau_{ij}}{\partial q_{i,w}} = \frac{\sigma_{ij} Q^2 \tau_{ij}}{r_{ij}^3} ((q_{i,w} - q_{j,w}) \sigma_{ij} - (u_{i,w} - u_{j,w}) r_{ij}^2) \quad (2.14)$$

193 Since  $\vec{u}_{ij}$  is a variable as well, we need the derivative with respect to it as well.

194 Thus

$$\frac{\partial}{\partial u_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \frac{\partial \tau_{ij}}{\partial u_{i,w}} \Omega_{ij} \quad (2.15)$$

$$\frac{\partial \tau_{ij}}{\partial u_{i,w}} = -\frac{\sigma_{ij} Q^2 \tau_{ij}}{r_{ij}} (q_{i,w} - q_{j,w}) \quad (2.16)$$

## 195 Without ADPs

196 Without ADPs the equations simplify down to

$$F(Q) = \frac{1}{N \langle f \rangle^2} \sum_{j \neq i} f_i^*(Q) f_j(Q) \frac{\sin(Qr_{ij})}{r_{ij}} \quad (2.17)$$

197 and

$$\frac{\partial}{\partial q_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \frac{\partial \Omega_{ij}}{\partial q_{i,w}} \quad (2.18)$$

198 use of these equations, when ADPs are not appropriate (like at cryogenic tempera-  
199 tures), greatly speeds up the computaiton.

## 200 2.2 COMPUTATION

201 Simply deriving the equations for the PDF is not enough. The many body nature of  
202 the PDF equation make analytical solution of the structure from the PDF impossible.  
203 Thus, the PDF must be computed from a structural candidates and compared against  
204 experimental results to evaluate the reliability of the model.

## 205 HPC and GPUs

206 To properly solve the structure of materials the PDF will need to be computed many  
207 times and checked against experimental results. This requires computation of the  
208 PDF, potentialy over many atoms. Calculating these PDFs requires a fast, highly  
209 parallized, computational framework.

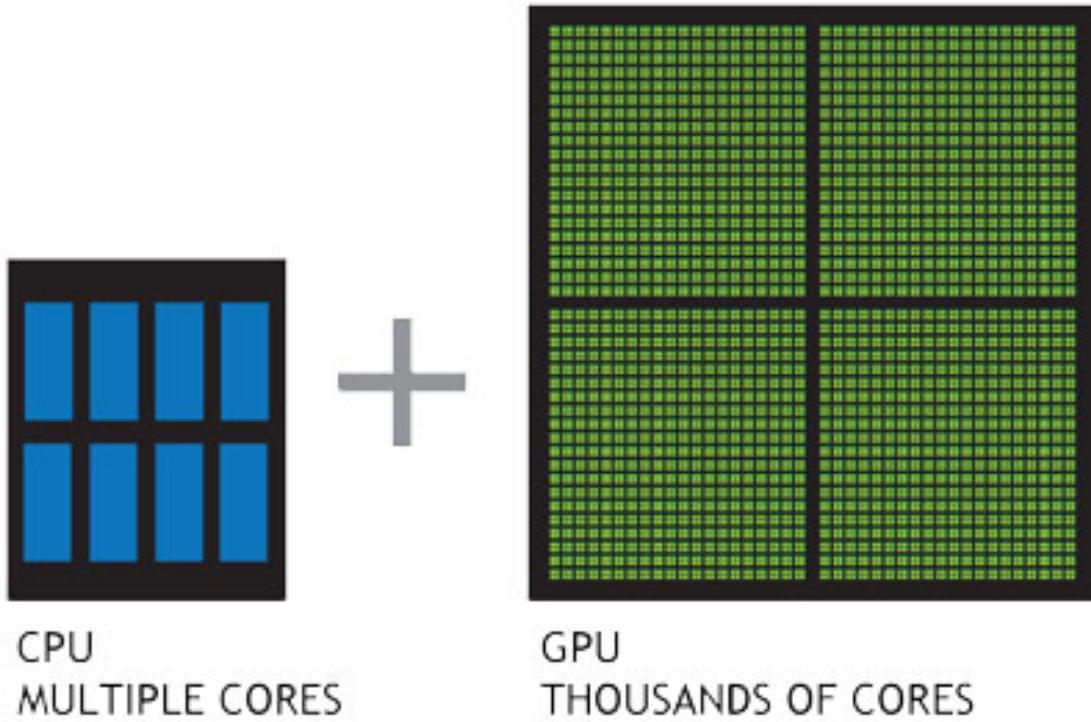


Figure 2.1: Comparison of the CPU and GPU chip architectures

## 210    GPUs and Parallelization

211    Computing the PDF is an embarrassingly parallel problem. The basic procedure is  
 212    to calculate the reduced structure factor  $F(Q)$  for each atom pair and momentum  
 213    transfer vector, sum over all the atom pairs, and Fourier transform the structure to  
 214    the PDF. The first part of this procedure is perfectly parallelizable, as each atom pair is  
 215    separate from the others. The summation over all the atomic reduced structure factors  
 216    can be parallelized via distributed summing. Lastly the FFT can be parallelized using  
 217    existing parallel FFT algorithms.

218    GPUs are particularly well suited to the task of computing PDFs. GPU chip  
 219    architecture is designed to perform many tasks simultaneously by having potentially  
 220    thousands of cores.

221 Map from  $ij$  space to  $k$  space

222 The above equations, although formally correct, are very ineffiecent.  $F(Q)$  and its  
 223 gradient are indexed over all the atoms twice, however there are symmetries that  
 224 allow us to only compute over the atom pairs esentially mapping from an  $n \times n$  space,  
 $ij$  space, to a  $\frac{n(n-1)}{2}$  space,  $k$  space. For  $F(Q)$  we apply the following mapping where

$$\begin{array}{ccccc} & & \psi & & \\ E & \xrightarrow{\quad} & E' & \xrightarrow{\Sigma} & Z \\ \phi \downarrow & & & \nearrow \Sigma' & \\ B & \xrightarrow{\quad} & B' & & \end{array}$$

225

226  $E$  denotes the atomic coordinates in  $ij$  space,  $E'$  denotes  $F(Q)$  before the summation  
 227 in  $ij$  space,  $B$  denotes the atomic pairs in  $k$  space,  $B'$  denotes  $F(Q)$  in  $k$  space, and  
 228  $Z$  denotes the final summed  $F(Q)$ . For the operators,  $\phi$  denotes the mapping from  
 229  $ij$  space to  $k$  space  $k = j + i * \frac{i-1}{2}$ ,  $\psi$  and  $\psi'$  denote the  $F(Q)$  operation in  $ij$  and  $k$   
 230 space, respectivly.  $\Sigma$  denotes the sum over all the atoms.

231 To properly define  $\Sigma'$  we must establish whether  $F(Q)$  is an even function. We  
 232 can accomplish this by examining each of the portions of  $F(Q)$ ,  $\alpha, \beta, \tau, \Omega$ .  $\Omega$  is even,  
 233 since  $r_{ij}$  is the interatomic distance, which is the same despite a flip of indicies,  $Q$   
 234 does not depend on the atomic indicies, and since  $Qr_{ij}$  is even so is  $\sin Qr_{ij}$ . Thus,  
 235  $\Omega$  is even. Providing similar analysis to  $\tau$  we can see that while  $\vec{u}_{ij}$  is odd, so is  
 236 the unit displacement vector between the two atoms, thus the two odds cancel out.  
 237 Intuitivly this makes sense, since the  $F(Q)$  equation is fundamentally interested in the  
 238 interatomic distances which is even. Thus, switching atom indicies does not change  
 239  $F(Q)$ . Due to the even nature of the  $F(Q)$  operator the  $\Sigma'$  operator sums over all the  
 240 atom pairs, and multiplies by two to reflect the double counting of the  $\Sigma$  operator.

241 For the gradient a similar mapping is used:

242 In this mapping, however, we use the  $\tilde{\phi}\Sigma$  operator. This operator simultaniously

$$\begin{array}{ccccc}
& & \psi & & \Sigma \\
E & \xrightarrow{\quad} & E' & \xrightarrow{\quad} & Z \\
\phi \downarrow & & & \nearrow \tilde{\phi}\Sigma & \\
& & B & \xrightarrow{\quad} & B' \\
& & \psi' & &
\end{array}$$

243 performs a reverse mapping from  $k$  to  $ij$  space, and a summation with the correct  
 244 symmetry. In this case the  $\psi$  and  $\psi'$  operators, which denote the  $\vec{\nabla}F(Q)$  operator  
 245 in  $ij$  and  $k$  space, are antisymmetric. Intuitivly this makes sense as an extension of  
 246 Newton's Second Law, since each particle's interation is felt oppositely by its partner.

## 247 Periodic Boundary Conditions

248 Periodic boundary conditions can be helpful when simulating extended solids or large  
 249 nanoparticles. In this case all the non-crystallinity is contained within the simulation  
 250 box and the box is repeated to create the longer distance peaks observed in the PDF.  
 251 To perform this we can break up the Debye equation into two main parts, the part  
 252 that describes the interatomic distances within the simulation box and those between  
 253 boxes. Neglecting the thermal motion portion:

$$F(Q) = \frac{1}{N\langle f \rangle^2} \left( \sum_{j \neq i} f_i^*(Q) f_j(Q) \frac{\sin(Qr_{ij})}{r_{ij}} + \sum_{i,j} f_i^*(Q) f_j(Q) \frac{\sin(QR_{ij})}{R_{ij}} \right) \quad (2.19)$$

254 where

$$R = |\vec{r} + \vec{u}| \quad (2.20)$$

$$\vec{u} = \gamma_1 * \vec{a} + \gamma_2 * \vec{b} + \gamma_3 * \vec{c} \quad (2.21)$$

255

## CHAPTER 3

256

### BENCHMARKING

257

This entire section needs some rewriting to distinguish this from the paper

258

The NUTS-HMC system was tested on a series of nanoparticle (NP) benchmarks.

259

The purpose of these benchmarks is to test the ability of the NUTS-HMC system to reproduce the target PDF and its associated structure. Systems were chosen for their size, crystallinity, and interfacial differences.

262

#### 3.1 PDF

263

The formation of NPs with both crystallographic and non-crystallographic structures [?] and with different chemical patterns [?] are well documented. For simplicity, we chose monometallic Au clusters as benchmarks and considered two groups of structures with different size and degrees of structural disorder in order to assess the reliability and efficiency of our HMC method for solving atomic structures from PDFs. The first group consists of  $\text{Au}_{55}$  clusters with different degrees of disorder, including a crystalline cluster structure in  $O_h$  (Octahedral) symmetry, a structure with a disordered surface, and an amorphous structure. The second group consists of the crystallographically solved  $\text{Au}_{102}$  structure as in the  $\text{Au}_{102}\text{MBA}_{44}$  nanocrystals [?, ?]. We used optimized structures from the Density Functional Theory (DFT) as target structures and generated the corresponding PDF,  $G_{\text{obs}}$ , according to

$$G_{\text{obs}} = \frac{2}{\pi} \int_{Q_{\min}}^{Q_{\max}} Q[S_{\text{obs}}(Q) - 1] \sin(Qr) dQ \quad (3.1)$$

274 where  $S_{\text{obs}}$  is the target structure's structure factor. Since all the target structures  
275 were optimized by DFT at zero Kelvin the target and model PDF profiles were  
276 calculated at zero temperature, with no atomic displacement parameters (ADPs).  
277 However, ADPs would have a considerable impact on the calculation of the PDF,  
278 especially for nanoparticles at non-zero temperatures.

279 Spin-polarized DFT calculations were carried out using the Vienna ab initio sim-  
280 ulation package (VASP) [?, ?] within the Perdew-Burke-Ernzerhof (PBE) exchange-  
281 correlation functional [?]. The projected augmented wave method [?] and a kinetic  
282 energy cutoff of 400 eV were used. Structural optimization was performed until the  
283 total energy and ionic forces were converged to  $10^{-6}$  eV and 10 meV/Å, respec-  
284 tively. The amorphous Au<sub>55</sub> structures were generated by simulated annealing using  
285 the classical embedded atom method potential [?]. Different annealing temperatures  
286 between 1200 K and 1670 K (bulk melting temperature of Au) were used and the  
287 thermally equilibrated structures were cooled down to 300 K before minimization at  
288 0 K. Further optimization using DFT leads to total energies that vary within 1-2  
289 eV among different amorphous structures and the lowest energy one was used as the  
290 target structure. The target structure of Au<sub>102</sub> was taken as the Au<sub>102</sub> core of the  
291 DFT-optimized Au<sub>102</sub>MBA<sub>44</sub> cluster [?].

292 All systems were refined using a PES which consists of a linear combination of  
293  $Rw$ , the repulsive and attractive thresholded spring potentials. The total potential  
294 energy in the Hamiltonian in Eq. (1.9) is expressed as:

$$U(q) = U_{Rw}(q) + U_{\text{spring}}(q, R_{\min}) + U_{\text{spring}}(q, R_{\max}) \quad (3.2)$$

295 The thresholded spring potentials are based on those previously proposed on by Pe-  
296 terson [?], i.e.  $U_{\text{spring}}(q, r_t) = \frac{\kappa}{2} \sum_{i,j} (r_{i,j} - r_t)^2$  for all atomic distance  $r_{i,j}$  outside the  
297 bounds of the spring threshold  $r_t$ . The resulting restoring forces on the out-of-bound  
298 atoms bring the system back within the bounds of the PDF,  $R_{\min}$  and  $R_{\max}$ , and  
299 therefore preventing the system from exploding or collapsing. Otherwise, incorrect

300 refinements may result by having atomic pair distances out of the PDF bounds.  $\kappa$  is  
301 the spring constant in eV/Å and the  $Rw$  potential is converted from unitless to eV  
302 via multiplication by a conversion factor  $\lambda$ .

303 Whereas the choice of the absolute values of  $\lambda$  and  $\kappa$  is somewhat arbitrary, their  
304 relative values are important in determining which term in Eq. (3.2) dominates the  
305 PES, especially when considering the effect of the simulation temperature. Generally,  
306 the ratio between the total potential energy and the temperature determines how  
307 much random motion will dominate the dynamics; a lower ratio implies that random  
308 motion will play a large role in the dynamics. The ratio between  $\lambda$  and  $\kappa$  of each  
309 spring describes how far the PDF can push the system below or above the bounds set  
310 by the spring potentials. Heuristically, too stiff a spring forbids the system to access  
311 new configurations, e.g. high energy “transition states” which may involve shorter  
312 bonds or a larger system size. Conversely, too small a spring constant makes it slower  
313 for the system to snap back within bounds and may lead to an explosion or implosion  
314 of the system, leaving the dynamics to drift aimlessly.

## 315 Model Parameters

316 Unless otherwise stated, the PDFs of the target and starting structures were generated  
317 using Eqn. (3.1) with a step of  $\delta R = .01$  Å,  $Q_{\min} = 0.1$  Å<sup>-1</sup>,  $Q_{\max} = 25.0$  Å<sup>-1</sup>.  $R_{\min}$   
318 and  $R_{\max}$  correspond to the first minimum before the first PDF peak and that after  
319 the last PDF peak, respectively, which ensure that the full meaningful region of the  
320 PDF is modeled. For each of the simulations, the Q resolution was calculated by

$$\delta Q = \frac{\pi}{R_{\max} + \frac{12\pi}{Q_{\max}}} \quad (3.3)$$

321 The HMC simulation was run with  $N = 300$  iterations, a target acceptance rate  
322 of 0.65, and an average starting momentum for each NUTS iteration of 10 eVfs/Å.  
323 Both repulsive and attractive spring potentials are used with  $\kappa = 200$  eV/Å and

324 thresholds matching  $R_{\max}$  and  $R_{\min}$  of the PDF, respectively.  $\lambda = 300$  eV was used  
325 as conversion factor for  $Rw$ . Each simulation was run with a pair of Nvidia GTX970  
326 graphics cards, with one card partially occupied with desktop visualization.

327 **Au55: surface relaxed**

328 We first test our algorithm by solving the crystalline Au<sub>55</sub> (*c*-Au<sub>55</sub>) cluster structure  
329 from its PDF. The starting structure is taken as the bulk-cut cuboctahedron of Au<sub>55</sub>  
330 with a uniform bond length of 2.89 Å. Due to finite-size and surface effects, the DFT-  
331 relaxed cluster structure shows a distinctively different bond length distribution as a  
332 function of the bond's distance to the cluster center of mass, and therefore is difficult  
333 to model with a small box approach which assumes an identical unit cell throughout  
334 the whole system.

335 **Run Parameters**

336  $R_{\min}$  and  $R_{\max}$  for this simulation were 2.45 Å and 11.4 Å, respectively, with  $\delta Q =$   
337 0.24 Å<sup>-1</sup>. The simulation ran for approximately 34 minutes, over a total of ~40  
338 thousand configurations. The results are shown in Fig. 3.1.

339 The PDF, radial bond distribution, and bond angle distribution show good agree-  
340 ment between the target and final fitted structures, with a  $Rw$  of 0.3% whereas  $Rw$   
341 of the starting structure is as high as 44.8%. DFT calculations yield a total energy of  
342 the final structure very close to that of the target structure (within a few meV). The  
343 success in the fitting is largely attributed to the factor that the target structure is  
344 only locally (and mildly) disturbed from its bulk-like counterpart and therefore there  
345 is no need to overcome any high PES barriers to reach the correct solution. As shown  
346 below, the situation is rather different for much more disordered target structures.  
347 Interestingly, the small-box solution using PDFgui[?] yields a rather large  $Rw$  of 43%,

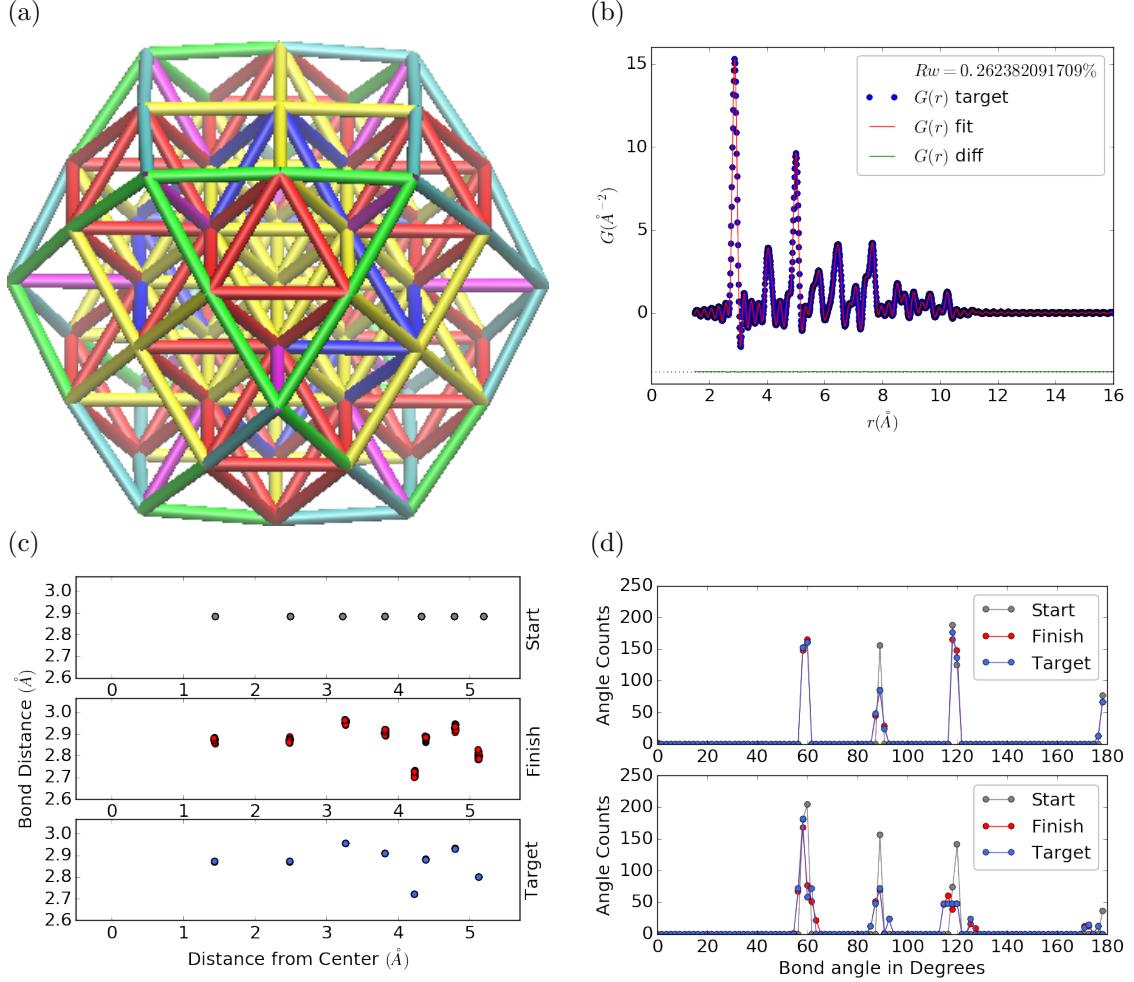


Figure 3.1:  $\text{Au}_{55}$  PDF fitting of DFT-optimized  $c\text{-}\text{Au}_{55}$ . (a) the final structural solution ( $Rw=0.3\%$ ) with bond lengths color-coded by step of  $0.05\text{\AA}$ , (b) the target PDF(blue dots) overlaid with the PDF of the final structure (solid red lines) with the difference in green lines offset below, ?? the radial bond distribution, and (d) bond angle distribution.

due to the failure to fit the surface contracted atoms with a unit cell. The PDF fits of the starting structure and small-box solution are shown

Put this somewhere

351 .

352 **Au55: surface disordered**

353 In addition to surface relaxation, the structure of a cluster or nanoparticle is often  
354 disrupted by the presence of defects and/or ligand bound to the surface. To mimic  
355 such surface disorders, we took the DFT-optimized *c*-Au<sub>55</sub> structure from case I as  
356 the starting structure and randomly displaced the surface atoms with a normal distri-  
357 bution of  $\sigma = 0.2 \text{ \AA}$ . All atoms are allowed to move in the HMC simulation, including  
358 the originally undisturbed core, which is a Au<sub>13</sub> cluster with  $O_h$  symmetry.

359  $R_{\min}$  and  $R_{\max}$  for this simulation were 1.95 Å and 12.18 Å, respectively, with  
360  $\delta Q = 0.23 \text{ \AA}^{-1}$ . The simulation ran for approximately 3.6 hours, over a total of  $\sim 270$   
361 thousand configurations. The results of the simulation are shown in Fig. ??.

362 Overall, good agreement is found between PDFs of the target structure and the  
363 final structural solution, even out to larger  $r$ , with an  $Rw = 0.6\%$  starting from an  
364  $Rw = 50.4\%$  (see Fig. S2). The radial bond distribution and angle distribution  
365 show reasonably good agreement, but with lower degree of crystallinity in the final  
366 structure compared to the target structure. The discrepancy is most obvious in  
367 the core: despite the identical core structure in the starting and target structures,  
368 the core atoms were displaced in the HMC simulations in order to achieve a “best”  
369 solution. This is because PDF measures the global average of interatomic distances  
370 between each atomic pair and does not contain direct information about the locality  
371 of such pairs, e.g. on the surface or cores. If such information is obtained a priori, for  
372 example, from theoretical prediction or other experimental measurements, the core  
373 structure can then be fixed and excluded from HMC dynamics.

374 Similar discrepancies are found in the CN distribution. Since the initial displace-  
375 ments of the surface atoms are relatively mild, the interatomic connectivities remain  
376 more or less the same and therefore the target structure has an identical CN distri-  
377 bution to the starting (unperturbed) structure. This is, however, not the case for  
378 the final fitted structure, which shows discernible differences, especially at the low

379 and high CN numbers. This is partly caused by the displacement of the core atoms  
380 mentioned above, and partly by the lack of CN constraints for the PDF fitting, which  
381 has been previously demonstrated in the case of  $\alpha$ -Si [?]. Additional experimental  
382 data, e.g. from EXAFS or NMR, may help to steer the simulations towards better  
383 agreement in both PDF and CN distribution.

384 **Au55: amorphous**

385 Next, we turn to the case in which the entire cluster structure is disordered. We used  
386 a DFT-optimized amorphous  $\text{Au}_{55}$  ( $a\text{-Au}_{55}$ ) as the target structure, and the DFT-  
387 relaxed  $c\text{-Au}_{55}$  cluster from Case I as the starting structure. The total energy of  
388  $a\text{-Au}_{55}$  was computed to be *lower* than that of  $c\text{-Au}_{55}$  by as large as 2.9 eV, consistent  
389 with the 3.0 eV found in previous DFT work [?].

390  $R_{\min}$  and  $R_{\max}$  for this simulation were 2.6 Å and 11.26 Å, respectively, with  
391  $\delta Q = 0.25 \text{ \AA}^{-1}$ . The simulation ran for approximately an hour, over a total of  $\sim 87$   
392 thousand configurations. The results of the simulation are shown in Fig. ??.

393 Our PDF fitting yielded a final structure of  $Rw$  of 1.7%, whereas that of the  
394 initial structure is as high as 76.1% (see Fig. S3 ), due to the drastically different  
395 atomic structure of the crystalline and amorphous  $\text{Au}_{55}$  clusters. Overall reasonable  
396 agreement in PDF, bond angle distribution, and radial bond distance distribution  
397 was found, and the wide spread of the bond lengths was qualitatively reproduced.  
398 However, the mismatch in CNs is problematic, partly due to the lack of information  
399 and/or constraints on the CNs. The total energy of the final structure is computed to  
400 be  $\sim 6$  eV higher than that of the target structure and the difference is substantially  
401 larger than the variation among different amorphous structures computed by DFT  
402 ( $\Delta E_{\text{tot}} \sim \pm 1\text{-}2$  eV). Such a fitting result, despite the rather small  $Rw$ , clearly  
403 indicates the importance of complementary informations and/or constraints necessary  
404 for reliably solving disordered NP structures from PDF.

405 **Au102: triple phase**

406 Our final benchmark is Au<sub>102</sub>, whose structure was initially solved by Jadzinsky and  
407 co workers using x-ray crystallography [?] and further confirmed by DFT studies [?].  
408 The Au<sub>102</sub> structure consists of three main parts, a 49-atom Marks decahedron core,  
409 two  $C_5$  caps consisting of 20 atoms each, and 13 equatorial atoms. Unlike previous  
410 cases, the multi-symmetry nature of the structure, i.e. each part has its own distinct  
411 symmetry, poses a challenge for PDF-based solution of the structure. This is because  
412 of the atomically centralized nature of the PDF, in which each atom “sees” a density  
413 of other atoms surrounding it and has a strong tendency towards becoming the center  
414 of the main symmetry group. Such tendency may lead to a solution where some of  
415 the correct atomic symmetries are discarded in favor of the core symmetry.

416 **Starting from fcc structure**

417 The starting structure was generated by a spherical cut of the fcc bulk lattice, with  
418 two surface atoms removed to conserve the total number of Au atoms.

419  $R_{\min}$  and  $R_{\max}$  for this simulation were 2.7 Å and 16. Å, respectively, with  $\delta Q =$   
420 0.18 Å<sup>-1</sup>. The simulation ran for approximately two hours, over a total of ~82  
421 thousand configurations. The results of the simulation are shown in Fig. 3.6.

422 The initial structure of an fcc bulk-cut cluster, had a starting  $Rw$  of 77.6% (see Fig.  
423 S4), whereas the final structure has a  $Rw$  as low as 8.1%. The disagreement between  
424 the final and target PDFs shows that the majority of the error is in the high  $R$  region,  
425 which is related to the long range distances between the core, caps, and equatorial  
426 atoms. The agreement for other structural metrics is less satisfactory. The bond  
427 angle distribution for core atoms in the final structure has a poor correlation with  
428 those in the target structure, with much broader peak widths. This is likely caused  
429 by the high kinetic barrier to change from one high-symmetry core structure (fcc)  
430 to another (Marks Decahedron). In contrast, the bond angle distribution for surface

431 atoms, which are of lower symmetry than the core, show a much better agreement.  
432 This is due to the preference of Monte Carlo techniques for higher entropy, and thus  
433 lower symmetry, structures. Similarly, the radial bond distance does not show the  
434 correct clustering of bond lengths as expected from an ordered structure, indicating  
435 the amorphous nature of our fit. Finally, the CN distribution shows the largest  
436 discrepancy at CN=12, again due to the amorphous nature of the fit. Overall, the  
437 structural metrics beyond the PDF indicate the poor agreement between the final  
438 and target structures. A higher simulation temperature, potentially combined with  
439 CN or bond length aware potentials (such as DFT or EXAFS derived PESs) may  
440 help to resolve this discrepancy.

441 **Marks decahedron**

442 The starting structure, a Marks Decahedron, was generated by the ASE Cluster tool  
443 with 2 atoms on the [100] face normal to the 5-fold axis, 3 atoms on the [100] plane  
444 parallel to the 5-fold axis and, 1 atom deep Marks reentrance. This produced a  
445 structure with 101 atoms which was extended by one more Au atom to fill out the  
446  $\text{Au}_{102}$  structure.

447  $R$  bounds and Q resolution were the same as the previous case. The simulation  
448 ran for approximately 2.5 hours over a total of  $\sim$ 90 thousand configurations. The  
449 results of the simulation are shown in Fig. ??.

450 The starting structure of Marks decahedron ( $Rw=56.6\%$ , see Fig. S5) yielded  
451 a better structural solution, with a final  $Rw$  of 3.3%. However, the discrepancies at  
452 high  $R$  remains as in the previous case. By examining the final structure, we can see  
453 that these high  $R$  errors are due to a lack of the two 20-atom caps and 13 equatorial  
454 atoms. Similarly, the radial bond distance distribution displays a diffusive behavior  
455 unlike the bond length clustering in the target structure. Compared to the previous  
456 case, the agreement in the CN and bond angle distributions are improved, with the

457 latter capturing nearly all peaks in the target structure with the exception of the 110  
458 bond angle. Relatively large discrepancies are found in the CN distribution at the  
459 low and high ends.

460 **Au147**

461 **3.2 PDF WITH ADPs**

462 **ADP 50**

463 1. Basic 50% larger magnitude

464 2. Random addition to APDs

465 3. Janus ADPs

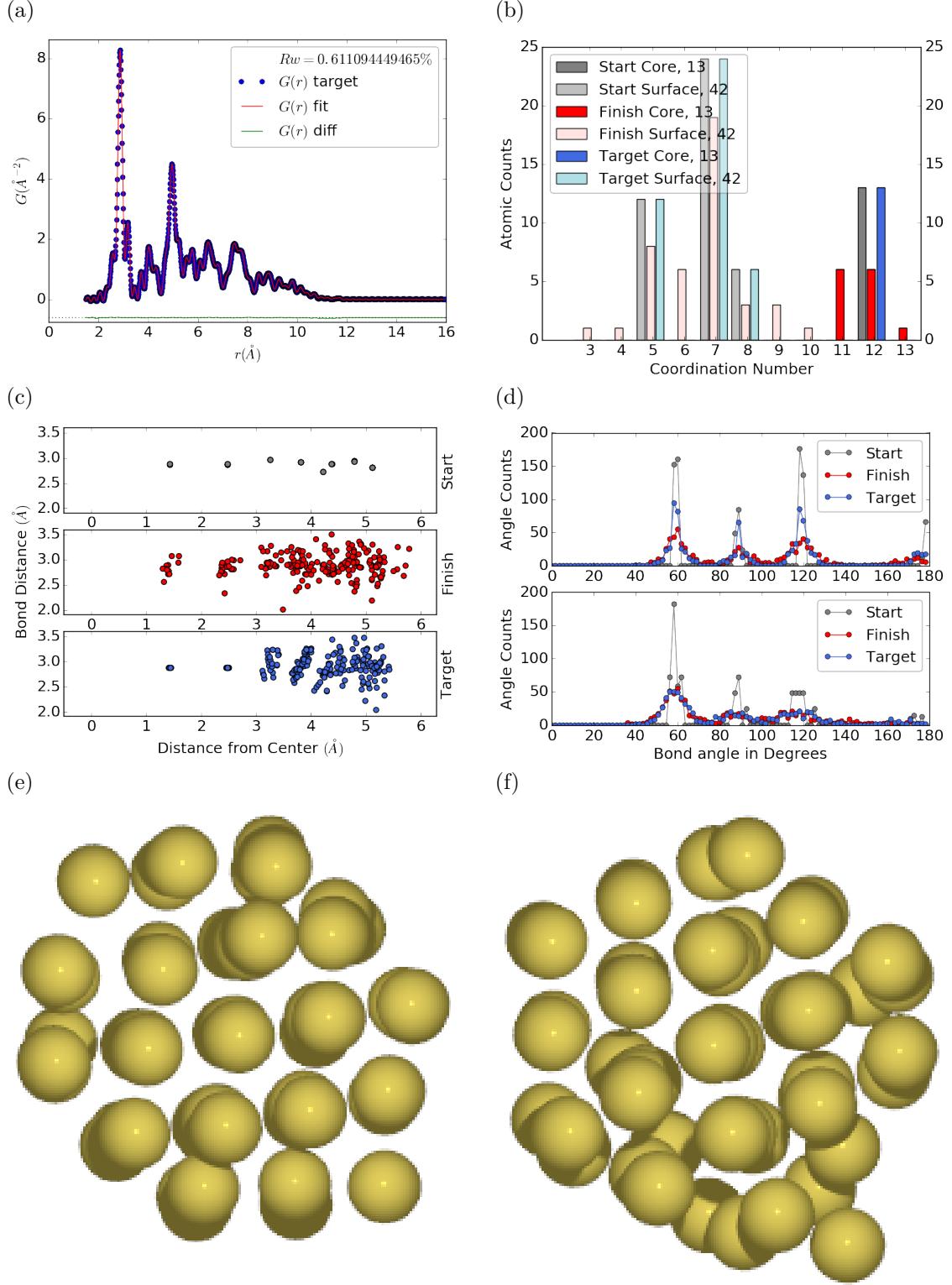


Figure 3.2:  $\text{Au}_{55}$  PDF fitting of surface-disordered  $\text{Au}_{55}$ . a) the target structure, b) the final structural solution ( $R_w=0.6\%$ ), c) the comparison of PDFs, d) the CN distribution with the number of atoms in either the core or the surface, e) the radial bond distribution, and f) the bond angle distribution.

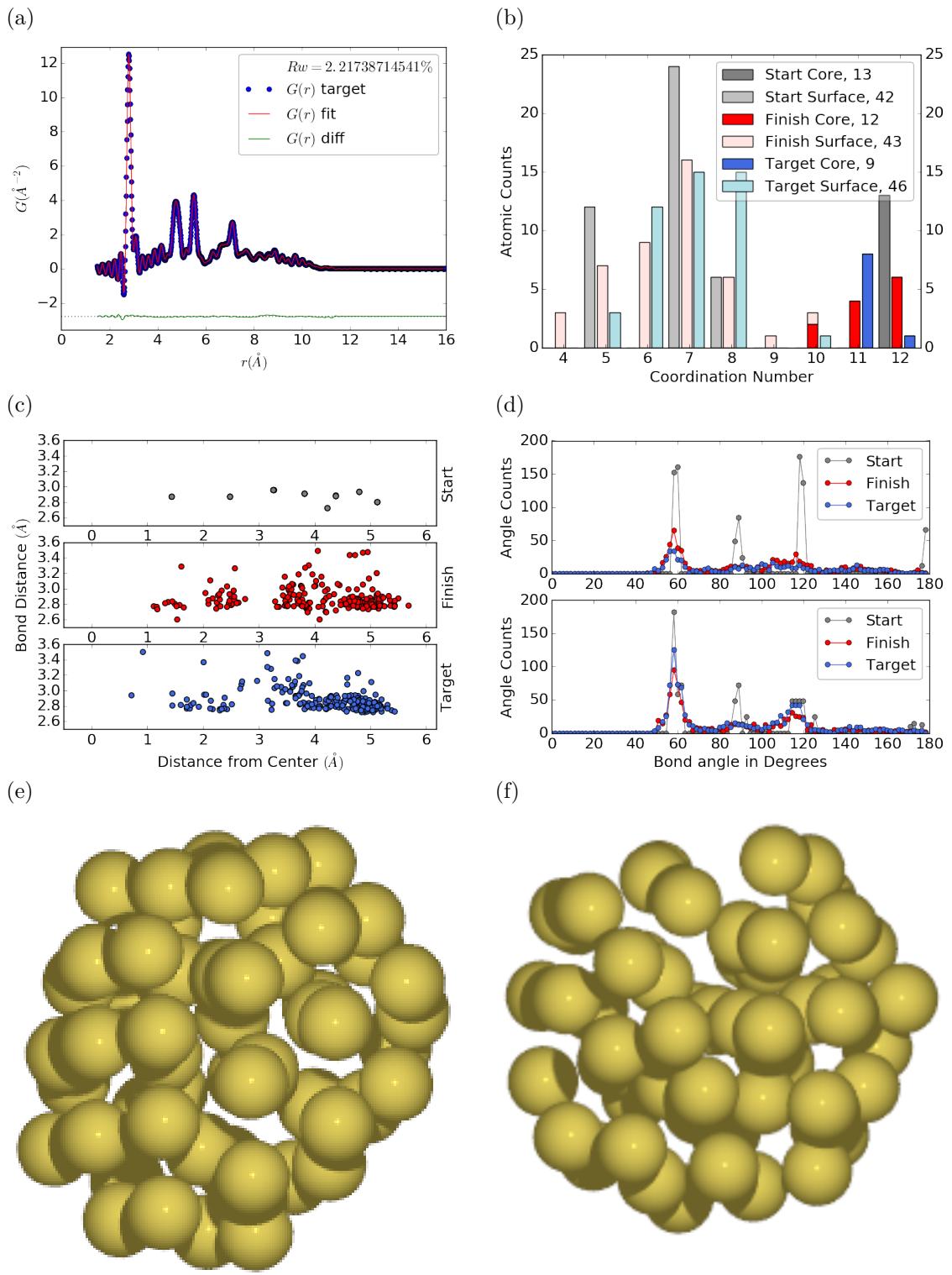


Figure 3.3: Similar to figure 3.2 for DFT-optimized amorphous  $\text{Au}_{55}$ .

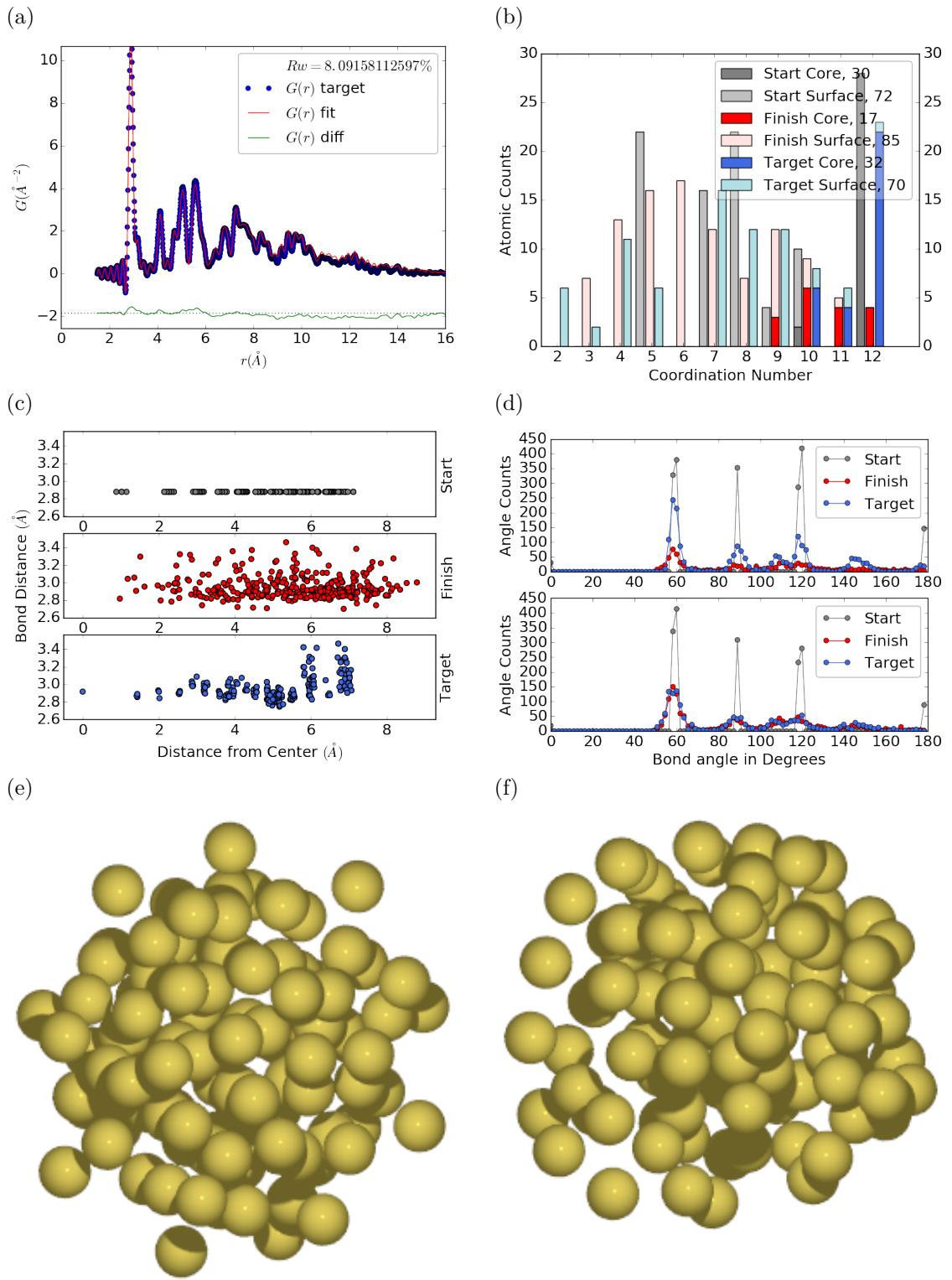


Figure 3.4: Similar to figure 3.2 for  $\text{Au}_{102}$  as in DFT-optimized  $\text{Au}_{102}\text{MBA}_{44}$  cluster.

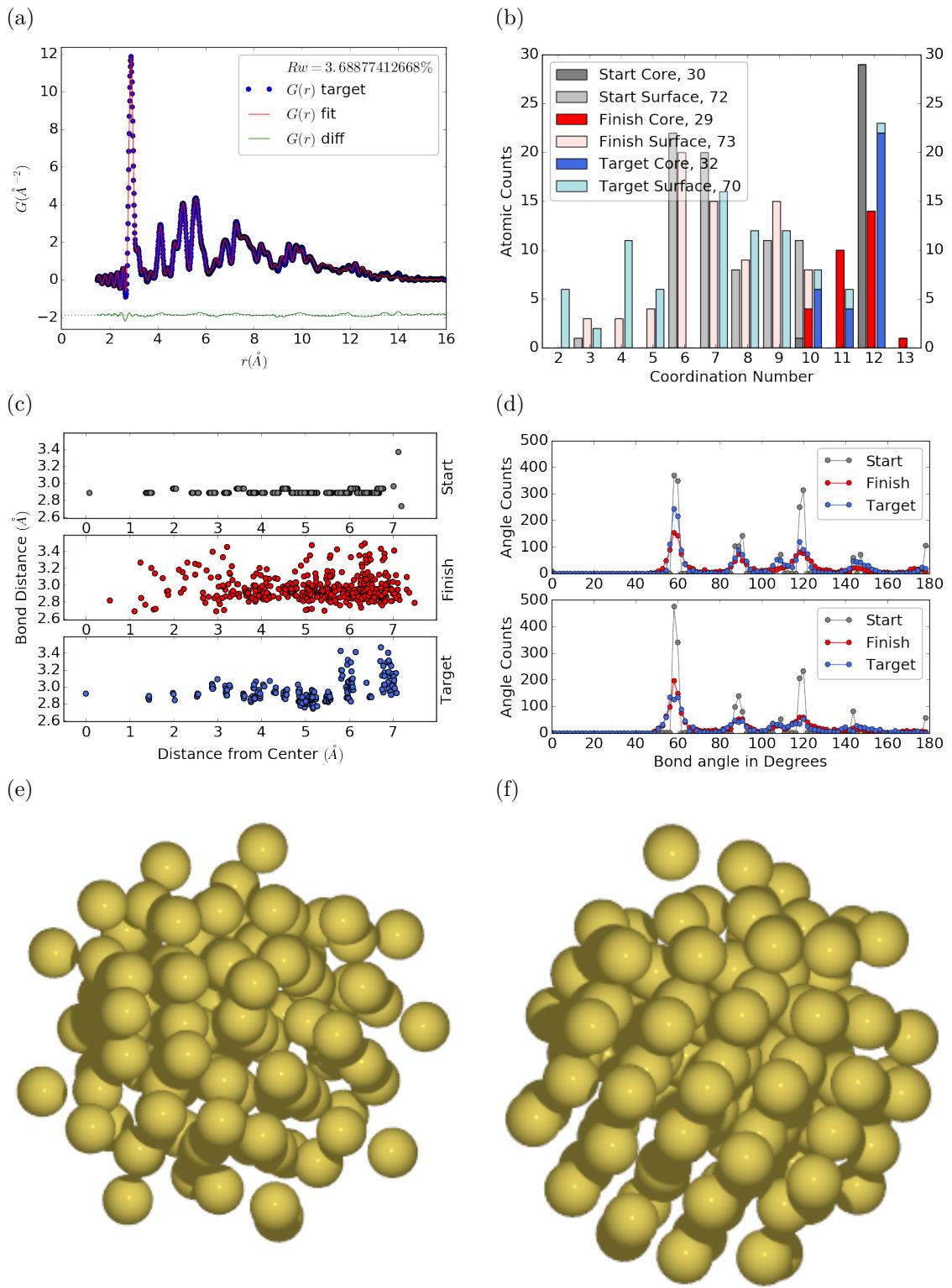


Figure 3.5: Similar to Fig. 3.6 with Marks decahedron as the starting structure.

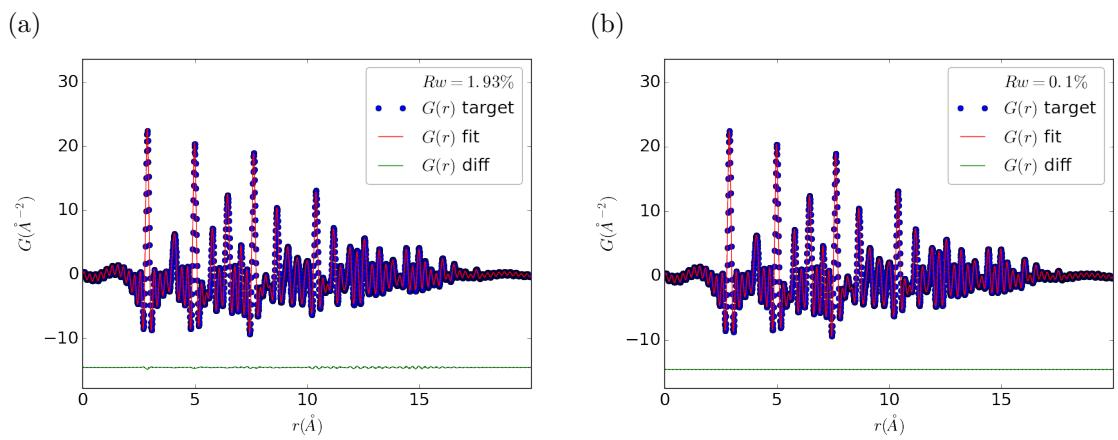


Figure 3.6

466

## CHAPTER 4

467

### X-RAY TOTAL SCATTERING DATA ACQUISITION AND PROCESSING

469

#### 4.1 INTRODUCTION

470 X-ray total scattering experiments are generally performed at synchrotron light sources,  
471 as only these sources can provide the needed flux, energy, and high momentum trans-  
472 fer vectors needed to obtain reliable PDFs. [?, ?] Despite the need for a dedicated  
473 facility to perform the total scattering experiments, the experiments themselves are  
474 fairly forgiving, allowing for reactive gaseous environments, experiment temperatures  
475 ranging from 2 K to 1200 K, and even electrochemical cycling. [?, ?, ?] The rapid  
476 PDF data acquisition associated with 2D area detectors creates a data management  
477 problem, as 96 hours of beamtime could result in almost 10,000 images which need  
478 to be associated with the experimental conditions and detector metadata. [?] Finally,  
479 all this data needs to be processed by masking bad pixels and regions, integrating  
480 azimuthally, and converting the scattering data to the PDF. [?, ?, ?, ?, ?]

481

#### 4.2 DATA STORAGE AND MANAGEMENT

482 Processing the raw pixel intensities to the PDF is very important as we are extracting  
483 most of our interesting information out of very high  $Q$  data. This data relies on good  
484 statistics and sound background subtraction. Talk about papers from Billinge Group  
485 with thin film PDF and dilute NP solutions. Diagram of the overall data processing  
486 workflow. Discuss the NSLS-II data stack.

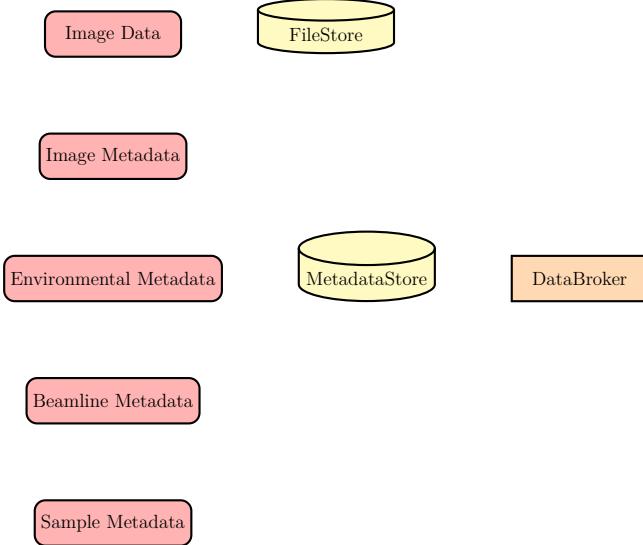


Figure 4.1: Database Loading Workflow. Data is loaded from various sources, including images and text files, into the FileStore and MetadataStore databases. Data is then retrieved from the databases using the databroker.

## 487    **MetadataStore Side Loading**

488    Design of sidewinder-spec for loading the data into metadatastore. Most of the design  
 489    considerations went into the loaders, which are different for each experiment.

## 490    **Detector $Q$ resolution**

491    To properly azimuthally integrate the images taken from the detector the  $Q$  resolution  
 492    of the pixels must be calculated. Integrating using even bins will cause pixels which  
 493    are not on the same ring to be binned together, causing the incorrect value of  $I(Q)$   
 494    to be obtained and a larger standard deviation in the integrated data. To properly  
 495    calculate the  $Q$  resolution the resolution of each of the pixels in  $2\theta$  must be calculated.  
 496    Figure 4.2 shows the scattering of x-rays onto a flat image plate detector. In this  
 497    diagram the bottom of the  $n$ th pixel is  $B$  while the top is  $B'$ . The resolution of this  
 498    pixel in  $2\theta$  is  $\angle BAC - \angle B'AC$ . Thus the resolution, calculated from the distances is

$$\Delta 2\theta = \arctan \frac{b}{d} - \arctan \frac{t}{d} \quad (4.1)$$

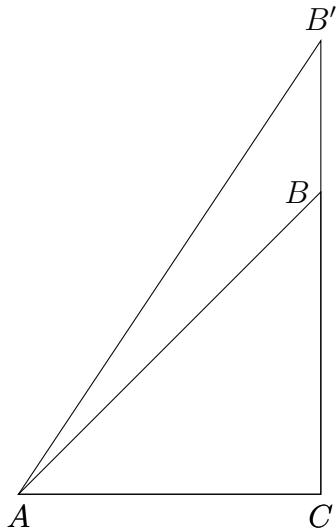


Figure 4.2: Scattering onto a flat detector

499 where  $d$  is the sample to detector distance,  $b$  is the distance to the bottom of a pixel,  
 500 and  $t$  is the distance to the top of that pixel. Note that these distances need to have  
 501 been corrected for detector tilt and rotation. Thus the resolution of a pixel in  $Q$  is

$$\Delta Q = \frac{4\pi(\sin \arctan \frac{b}{d} - \sin \arctan \frac{t}{d})}{\lambda} \quad (4.2)$$

502 where  $\lambda$  is the x-ray wavelength.

503 For a Perkin Elmer image plate, like the one used at the NSLS-II's XPD and the  
 504 APS's 11-ID-B, the resolution function is shown in 4.3. For the same detector the  
 505 number of pixels per  $Q$  is shown in 4.4

## 506 Automated Mask Generation

### 507 Introduction

508 Detector masking is an important part of any x-ray scattering workflow as dead/hot  
 509 pixels, streak errors, and beamstop associated features can be averaged into the data  
 510 changing the signal and its statistical significance. While some features, like the  
 511 beamstop holder, can be easily observed and masked by hand other are much more

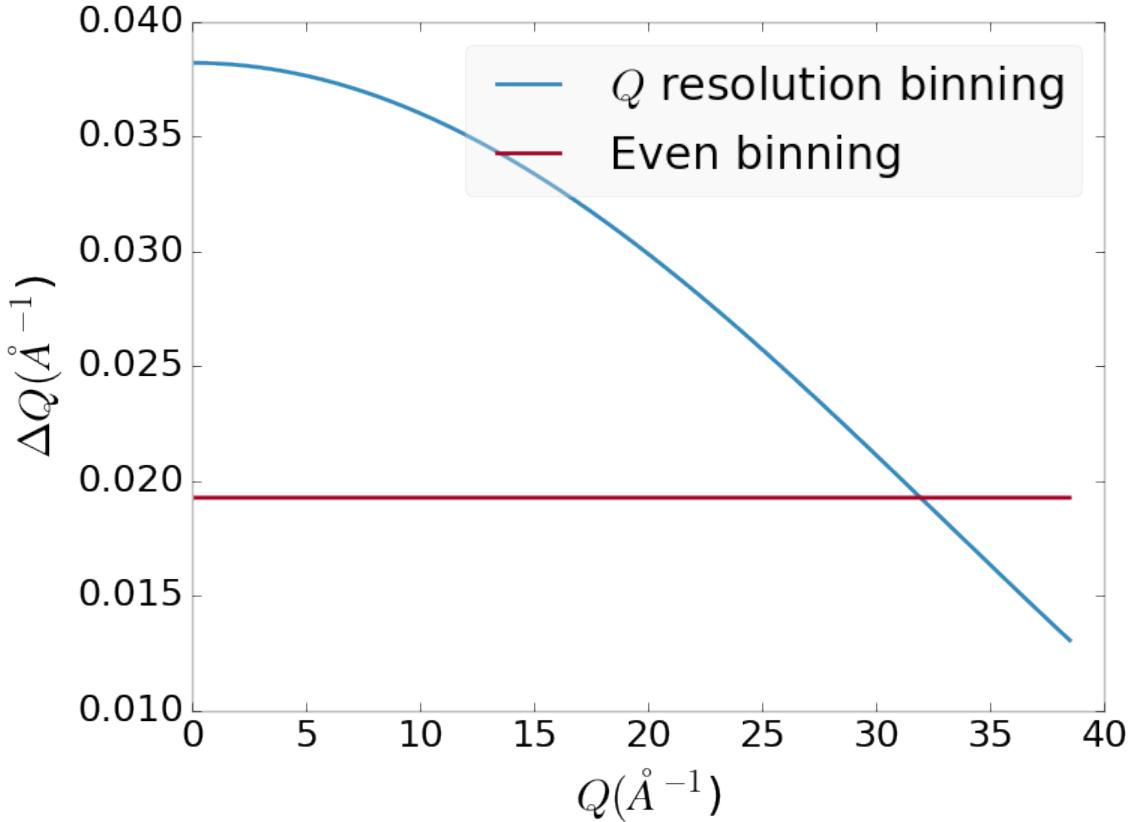


Figure 4.3:  $Q$  resolution as a function of  $Q$ .

512 difficult to observe even on large computer monitors. Additionally, while dead/hot  
 513 pixels and streaks are usually static the hot pixels associated with textured or sin-  
 514 gle crystal scattering or cosmic rays are not. Thus, coming up with an automated  
 515 method for finding such erroneous pixels is important, especially as high flux diffrac-  
 516 tion beamlines can generate data very quickly.

517 While this problem can be quite complex in the most general case, we can use the  
 518 annular symmetry of the powder scattering pattern to our advantage, by comparing  
 519 a pixel against pixels in the same ring. Since non-textured powder scattering should  
 520 produce the same pixel intensity for a given ring we can mask any pixels which are  $\alpha$   
 521 standard deviations away from the mean. This method relies on the aforementioned  
 522 pixel binning algorithm, as using miss sized bins will cause some pixels which should  
 523 be in separate rings to be put together, and others which should be in the same ring

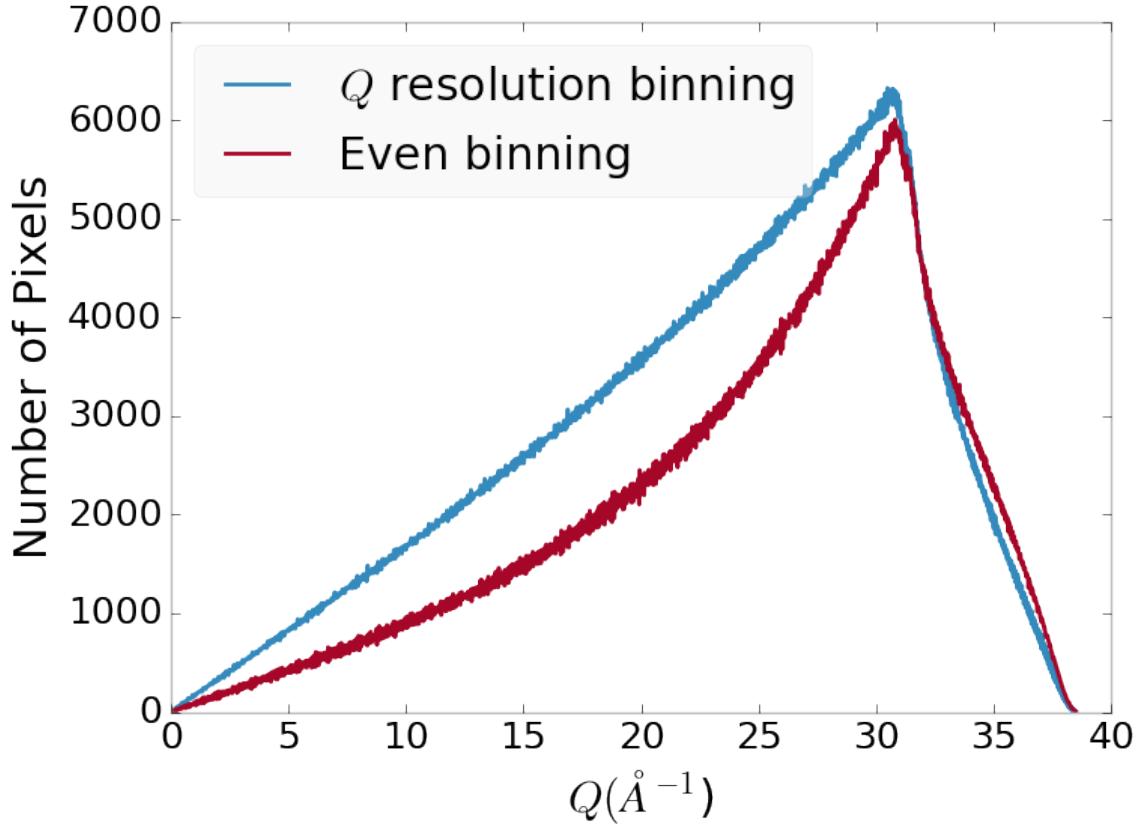


Figure 4.4: Number of pixels as a function of  $Q$ , binned at the  $Q$  resolution of the detector.

524 to be separated. In that case the masking algorithm will overestimate the number of  
 525 pixels to be masked due to the additional statistical variation in the sample.

## 526 Algorithm Design

527 The masking algorithm procedure takes in the image and a description of the pixel  
 528 positions in either distance from the point of incidence or in  $Q$ . The image is then  
 529 integrated twice, producing both the mean  $I(Q)$  and the standard deviation of each  
 530  $I(Q)$  ring. The mask is created by comparing the pixel values against each ring's  
 531 standard deviation and threshold  $\alpha$ . Note that the threshold can be a function of  
 532 distance from the point of incidence or  $Q$ .

533 **Test Cases**

534 To study the effectiveness of the masking we ran the algorithm against both simulated  
535 experimental data. In the case of the simulated data four systems were created: 1)  
536 dead/hot pixels with varying numbers of defective pixels, 2) beamstop holder with  
537 varying beamstop holder transmittance, 3) rotated beamstop holder with varying  
538 beamstop holder transmittance, and 4) beamstop holder with dead/hot pixels. The  
539 base scattering was produced by

$$I = 100 \cos(50r)^2 + 150 \quad (4.3)$$

540 where  $r$  is a pixel's distance from the beam point of incidence. The positions of  
541 the dead/hot pixels were chosen at random as was the dead or hot nature of the  
542 defect. Dead pixels had values from 0 to 10, while hot pixels had values from 200  
543 to 255. The beamstop was positioned at the vertical center of the detector with an  
544 initial width of 60 pixels and final width of 120 pixels. The hight of the beamstop  
545 was 1024 pixels. The beamstop was calculated to attenuate the x-ray scattering  
546 signal at various transmittance, as various beamstop holder materials have different  
547 transmittance. Two version of the masking algorithm were run for each test case, one  
548 using the standard even bin sizes for the integration step, and one where the bin sizes  
549 are tuned to the pixel  $Q$  resolution as discussed in 4.2.

550 **Results and Discussion**

551 Figures 4.5-4.12 show the results of the masking algorithm on simulated images. The  
552 dead/hot pixel masking shows the importance of using the  $Q$  resolution based bin sizes  
553 as the even bin based mask have a tendency to over mask the image, removing pixels  
554 which contain valuable signal. This overmasking is caused by pixels being improperly  
555 associated with one another by the even bins. Figure 4.5 indicates that the masking  
556 algorithm, with the proper binning, masks the image perfectly, with no missed bad

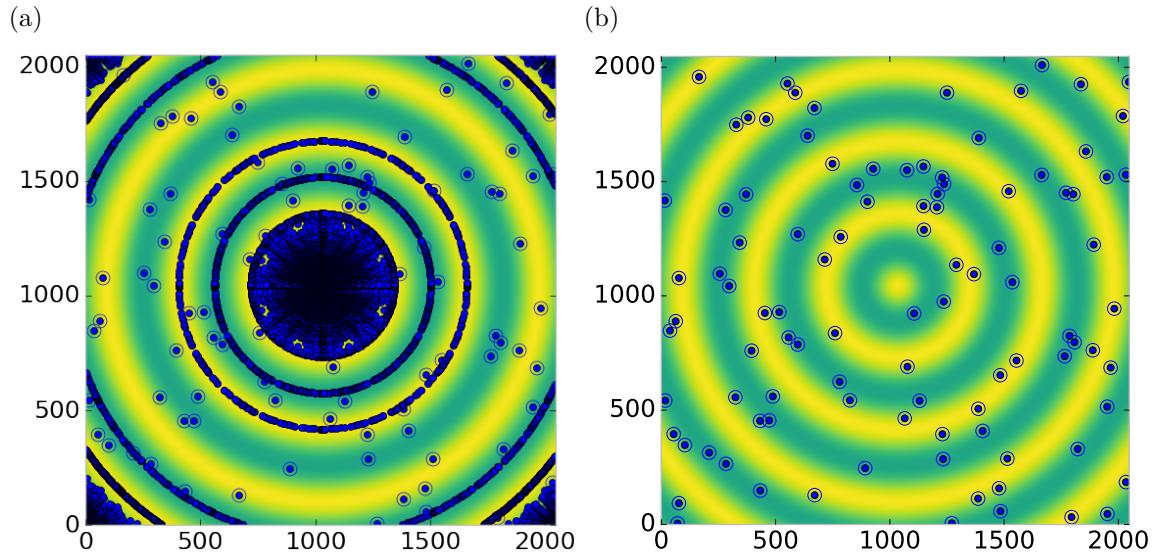


Figure 4.5: Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

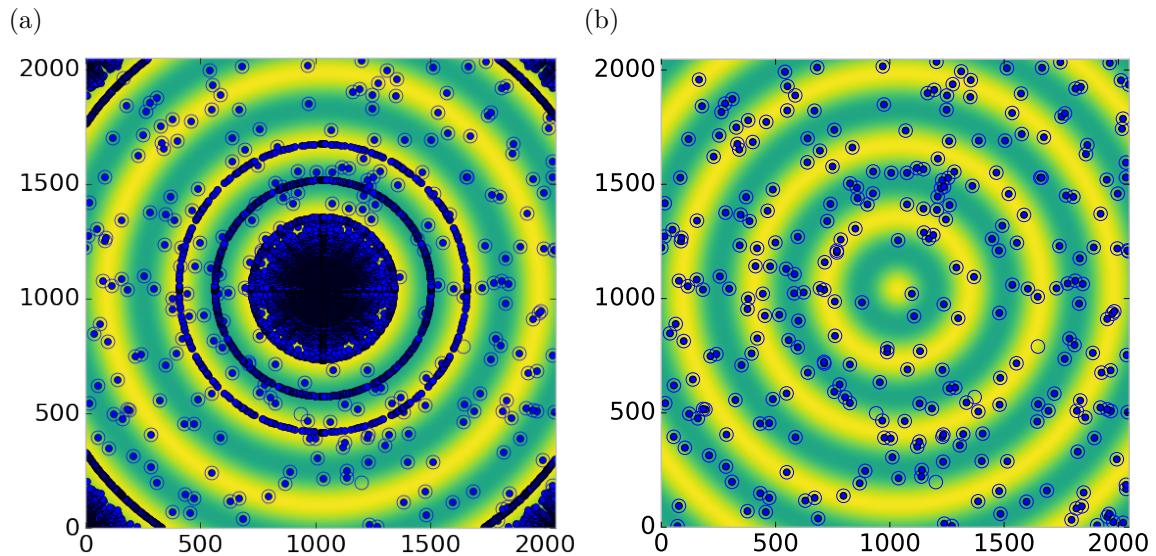


Figure 4.6: Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

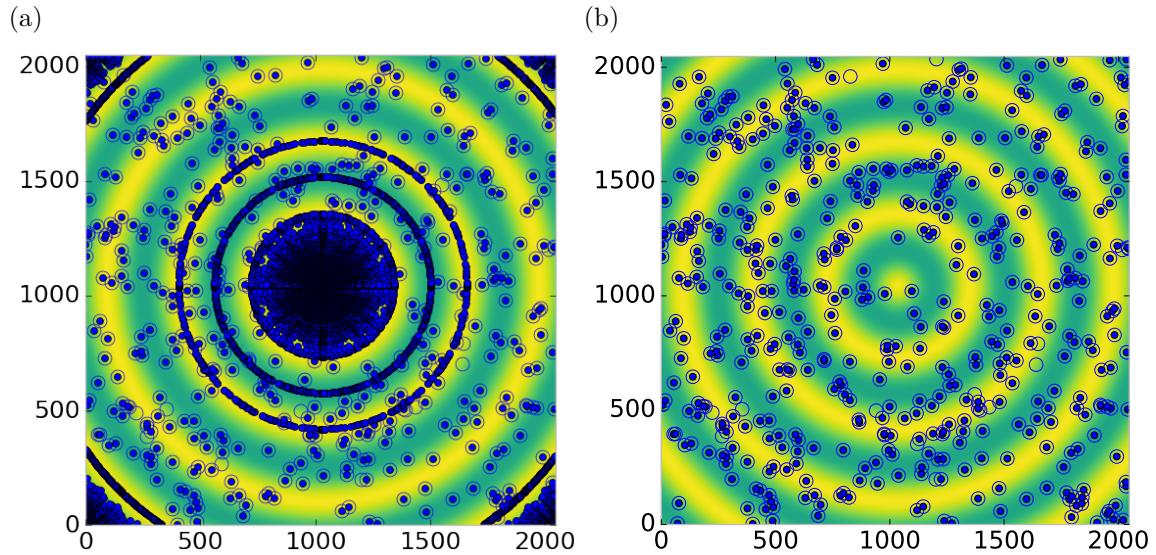


Figure 4.7: Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

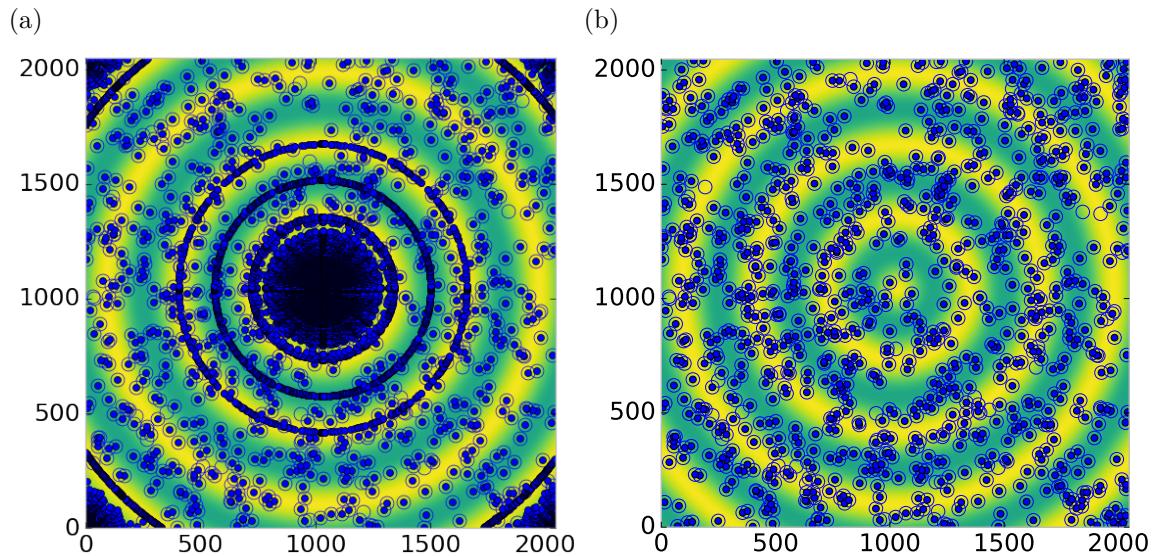


Figure 4.8: Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

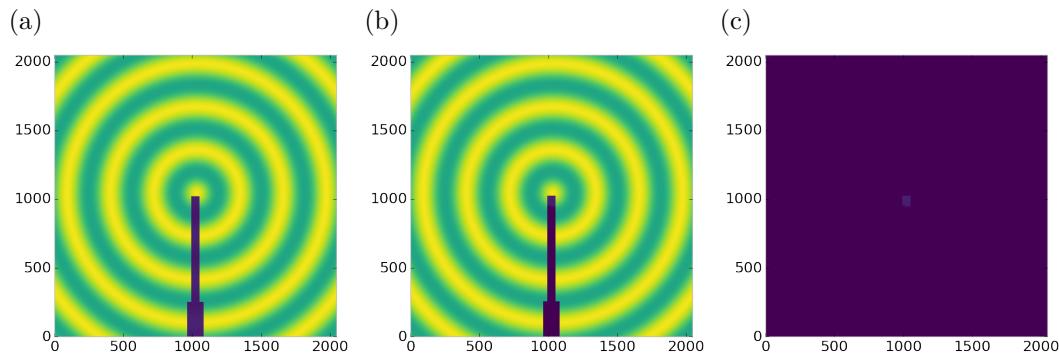


Figure 4.9: Generated beamstop holder masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

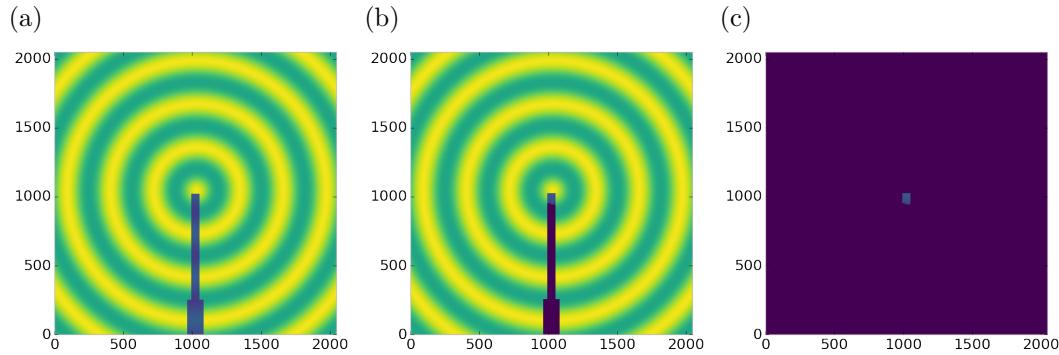


Figure 4.10: Generated beamstop holder masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

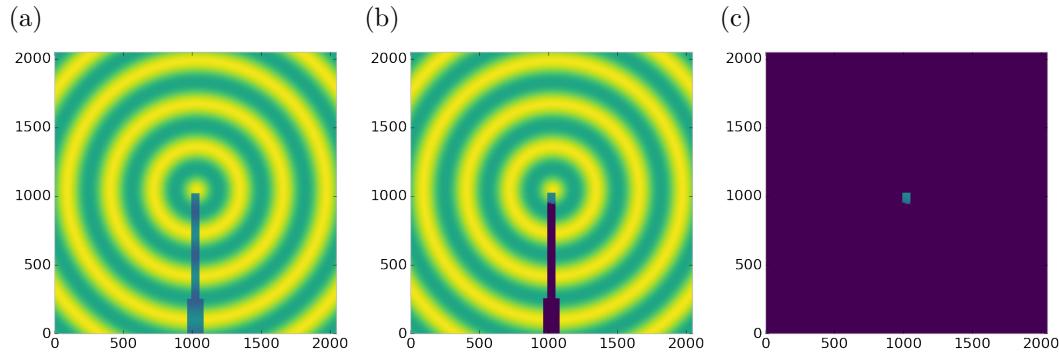


Figure 4.11: Generated beamstop holder masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

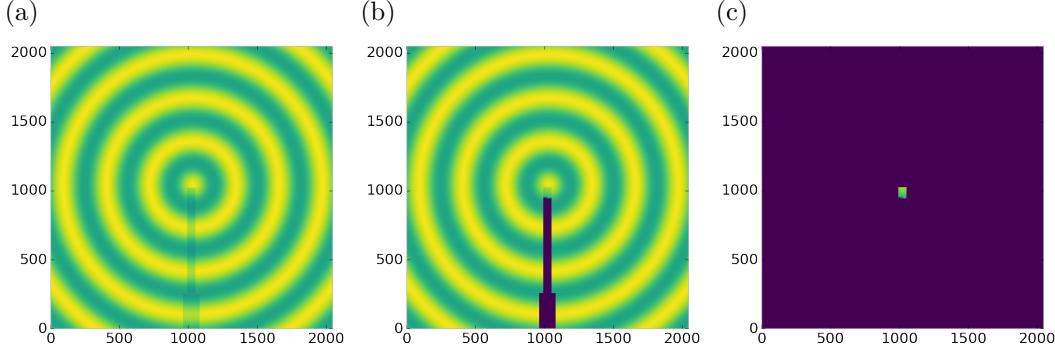


Figure 4.12: Generated beamstop holder masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

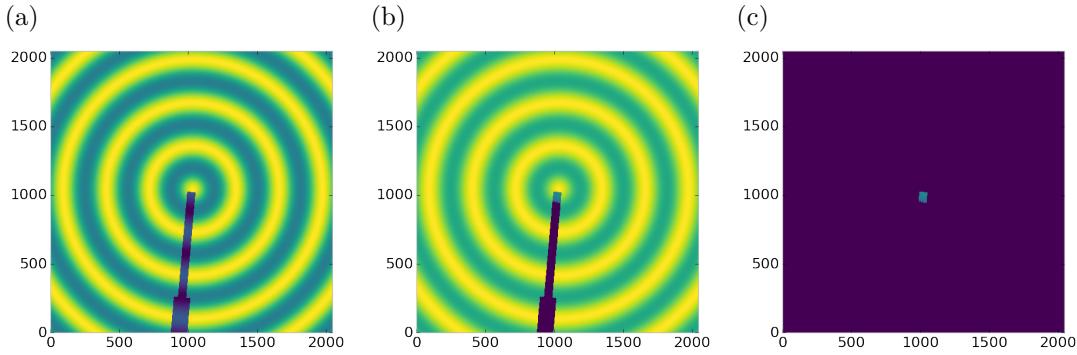


Figure 4.13: Generated beamstop holder masks which is rotated away from verticle

557 pixels or good pixels masked. This is not the case in figures 4.6 - 4.8 as we can see  
 558 pixels which should have been masked but were not. Despite these missed pixels no  
 559 pixels were improperly masked in any of the well binned images. These test cases  
 560 are actually more difficult than experimental data, as the dynamic range of most  
 561 detector causes the dead/hot pixels and single crystal/textue peaks to be orders of  
 562 magnitude away from the desired signal.

563 The beamstop holder masks shown in figures 4.9 - 4.12, which were all run with  
 564 the  $Q$  resolution binning show similar results across the transmittance range, missing  
 565 only a small part of the beamstop holder near the point of incidence. Near this point  
 566 the beamstop holder becomes a statistically significant part of the total number of  
 567 pixels in a given ring, thus it can not be masked out using a statistical search of the  
 568 rings. For most PDF and XRD studies this small area can be masked automatically

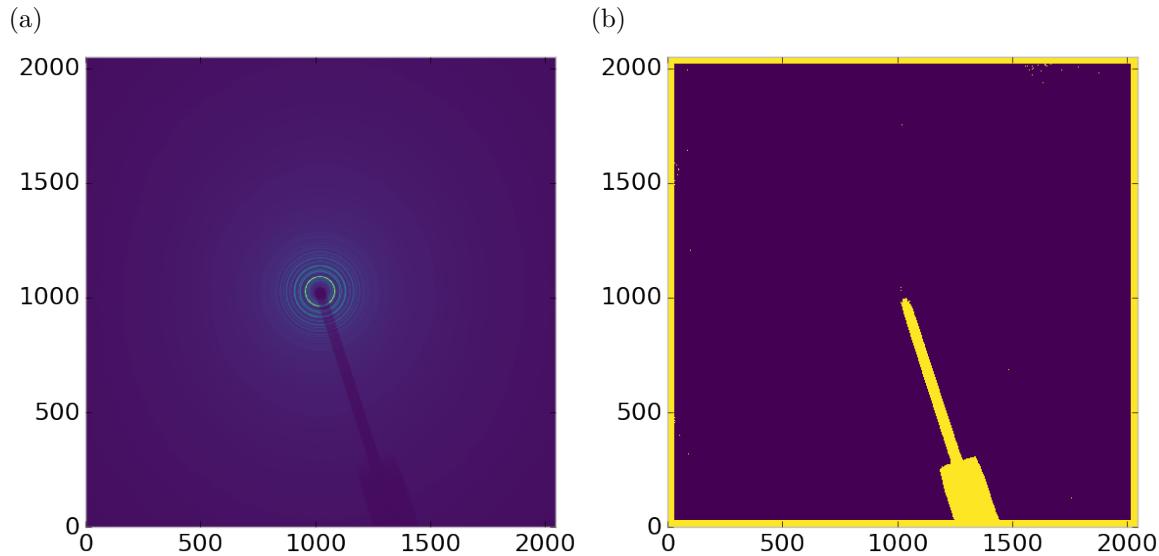


Figure 4.14: Masked experimental data. a) the raw image, b) the mask

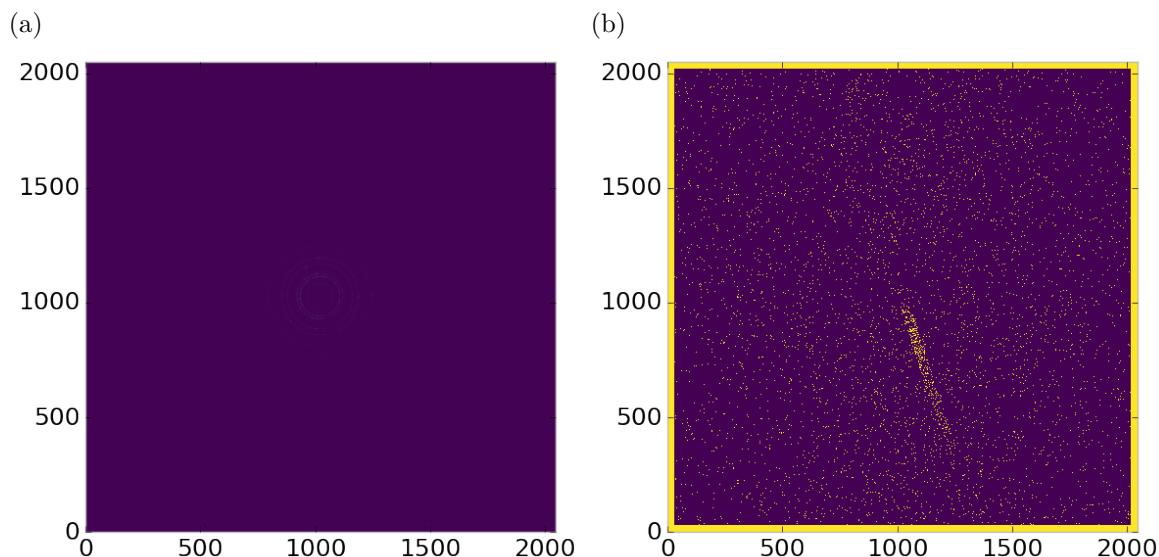


Figure 4.15: Masked experimental data with Pt single crystal signal. a) the raw image, b) the mask

569 by masking all the pixels who's distance from the point of incidence is smaller than a  
 570 given radius  $r$ , or can be negelected outright as the area is not used in the analysis or  
 571 refinement. Similar results were produced for beamstop holders which were rotated  
 572 away from the vericle position, as shown in figure 4.13

573 Working with actual experimental data, obtained at the Advanced Photon Source

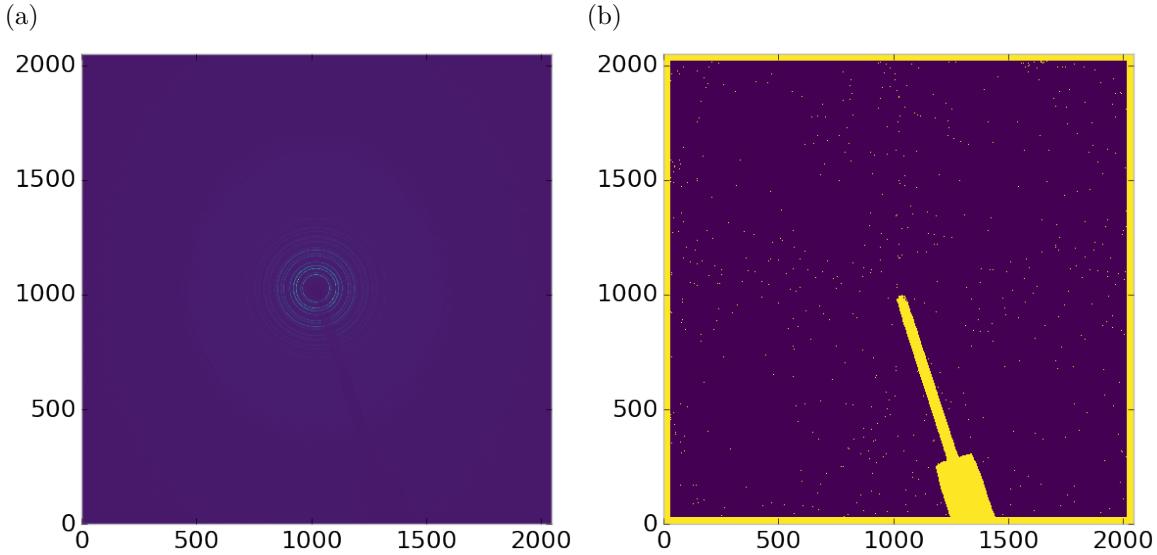


Figure 4.16: Masked experimental data with Pt single crystal signal using figure's 4.14 as a starting mask. a) the raw image, b) the mask

beamline 11-ID-B, shows the difficulty of masking images which have low phonon counts. While the masking of experimental data taken with longer exposures, consisting of 250 .2 second shots, shown in figure 4.14 provides very sharp edges to the beamstop holder, and very little extra masking beyond the occasional dead pixel, this is not the case for the single crystal data. The single crystal data is more problematic because of its short exposure time and low flux, with 500 frame at a .1 second exposure and having shrunk the beam size. The low flux is to prevent the very strong single crystal peaks from damaging the detector. However, this causes the image to be less statistically viable then ideal, causing problems with the mask as seen in figure 4.15. This can be alleviated to some degree by using the previously generated mask as a starting mask for the single crystal image, as shown in 4.16. While the masking algorithm still produces many diffuse masked pixels, they are far fewer, this may be due to the removal of the beamstop which could have contributed to the large standard deviation in figure 4.15.

588 **Conclusions**

589 In this section the masking algorithm, which relies on both  $Q$  resolution based binning  
590 and a statistical approach to azimuthal symmetry, was developed. The focus of  
591 this algorithm was to remove many unwanted detector features associated with pixel  
592 defect, beamstop holder associated scattering attenuation, and single crystal/texture  
593 peaks. Simulated data was used to evaluate the beamstop holder and dead/hot pixel  
594 masking capacity, while experimental data was used to check for single crystal and  
595 texture based masking.  $Q$  resolution based binning was shown to be very important to  
596 avoid overmasking. The ability of the mask writer to mask images is somewhat limited  
597 by the overall statistical image quality, although some deficiencies can be obtained by  
598 using previously generated masks as starting points. This masking algorithm is now  
599 in use in the data processing workflow and will be available in scikit-beam soon.

600 **Automated Image Azimuthal Integration**

601 Using the  $Q$  resolution binning and masking developed in sections 4.2 and 4.2 the  
602 images can be properly integrated. Generally, images are integrated by taking the  
603 mean value of the pixels in a ring. However, other statistical measures of the average  
604 value can be used, like the median.

605 Figures 4.17-4.19 show the importance of masking and the choice of average func-  
606 tion. All the figures were produced using the same dataset, 50 °C  $\text{Pr}_2\text{NiO}_4$  taken at  
607 the APS's 11-ID-B on a Perkin Elmer area detector. The automatic masking alpha  
608 was 3 standard deviations from the mean. While it is difficult to observe the changes  
609 the mask causes in the full  $I(Q)$  plot (subfigures a) and b)), the standard deviation  
610 plots show the effect of bad pixels on the data (subfigure c)). Subfigure c) for figures  
611 4.17-4.19 shows that removal of the beamstop holder lowers the low  $Q$  standard de-  
612 viation from around .1 to almost .01 out to 15  $\text{\AA}^{-1}$ . The high  $Q$  subfigures d) and f)  
613 in figures 4.17-4.19 show the “kink” effect of the detector edge and beamstop holder,

614 where there is a dip in the  $I(Q)$  scattering when the rings include the edge of the  
615 detector. This effect seems to be due to both errors in the edge pixel intensity and the  
616 beamstop holder as masking of the edges only seems to provide only partial removal  
617 of the issue. It is important to note that while integration using the mean of the  
618 ring has issues with only the edge mask, as evidenced by the change in slope in 4.18  
619 d) around  $29.5 \text{ \AA}^{-1}$ , the median integration does not include this error. Ideally the  
620 detector would have a normal distribution of pixel intensity for a given ring, which  
621 would imply an equivalency between the mean and median  $I(Q)$  values. Despite the  
622 closeness of the mean and median once the final mask has been created, it seems that  
623 the median is more reliable, as it was less effected by the beamstop holder in figure  
624 4.18. Thus, for subsequent integrations discussed in this work the median is used to  
625 avoid any defective features that the masking algorithm may have missed.

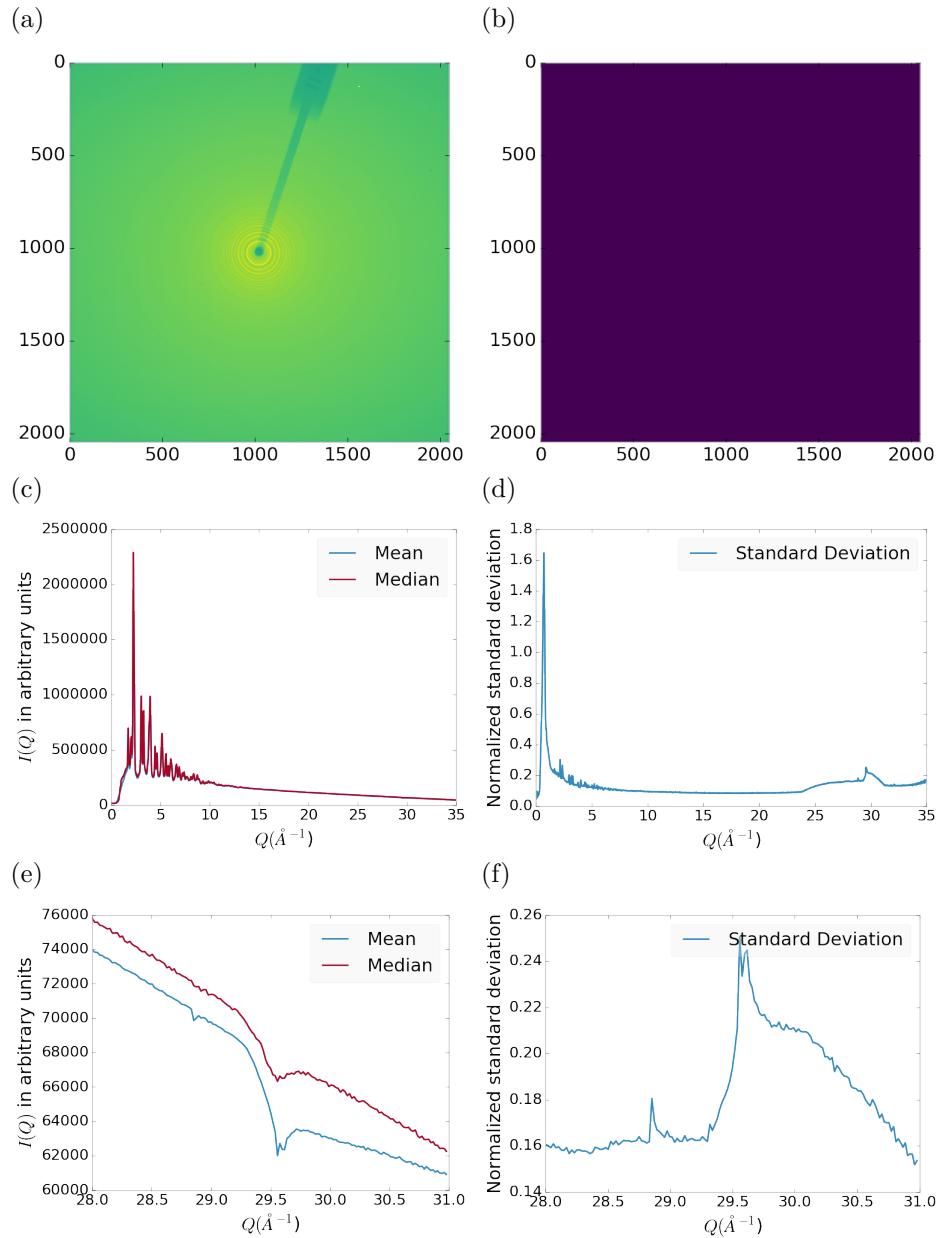


Figure 4.17: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with no mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{ \AA}^{-1}$  to  $31 \text{ \AA}^{-1}$   $Q$  range for the mean and median, f)  $28 \text{ \AA}^{-1}$  to  $31 \text{ \AA}^{-1}$   $Q$  range for the standard deviation

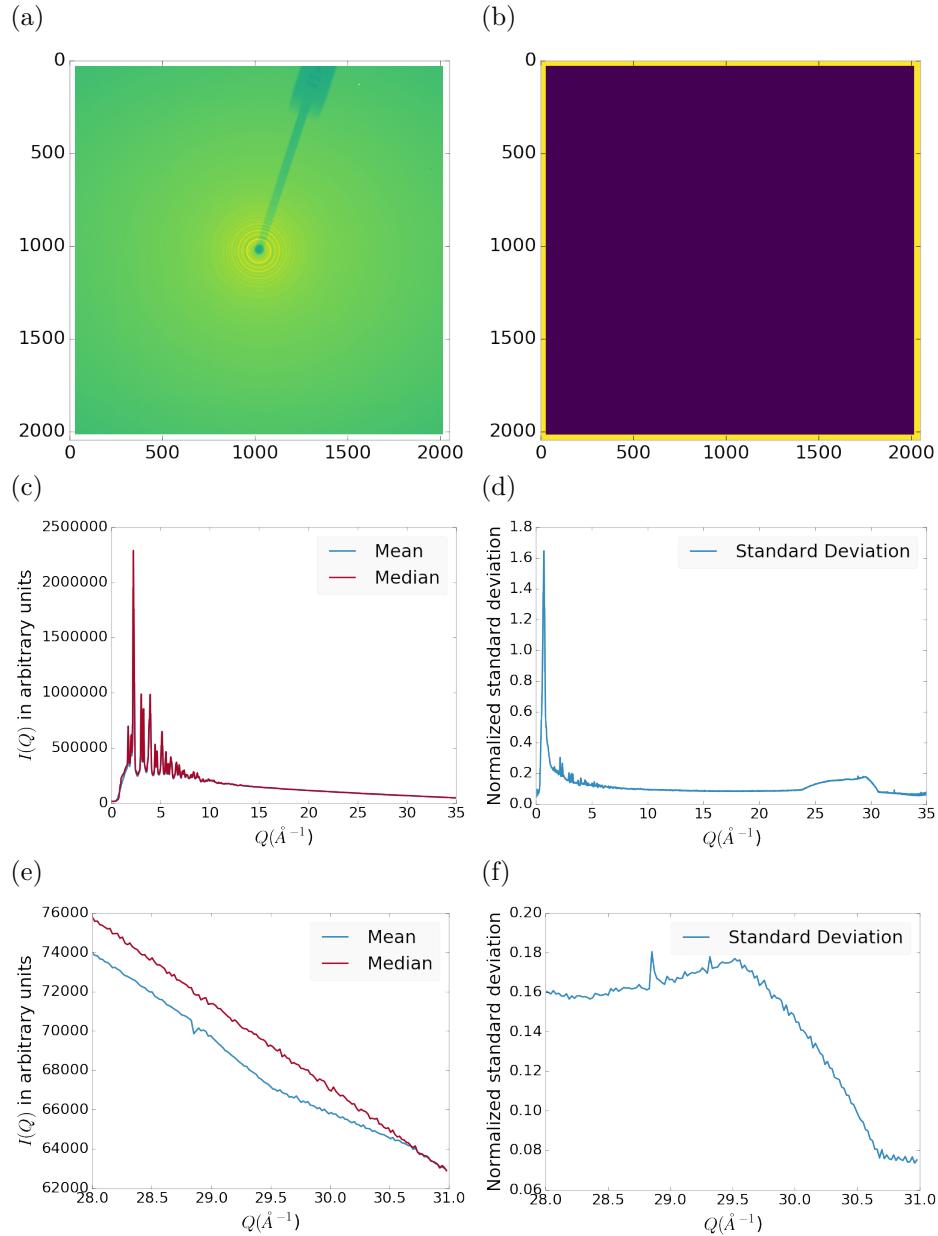


Figure 4.18: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with only an edge mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{ \AA}^{-1}$  to  $31 \text{ \AA}^{-1}$   $Q$  range for the mean and median, f)  $28 \text{ \AA}^{-1}$  to  $31 \text{ \AA}^{-1}$   $Q$  range for the standard deviation

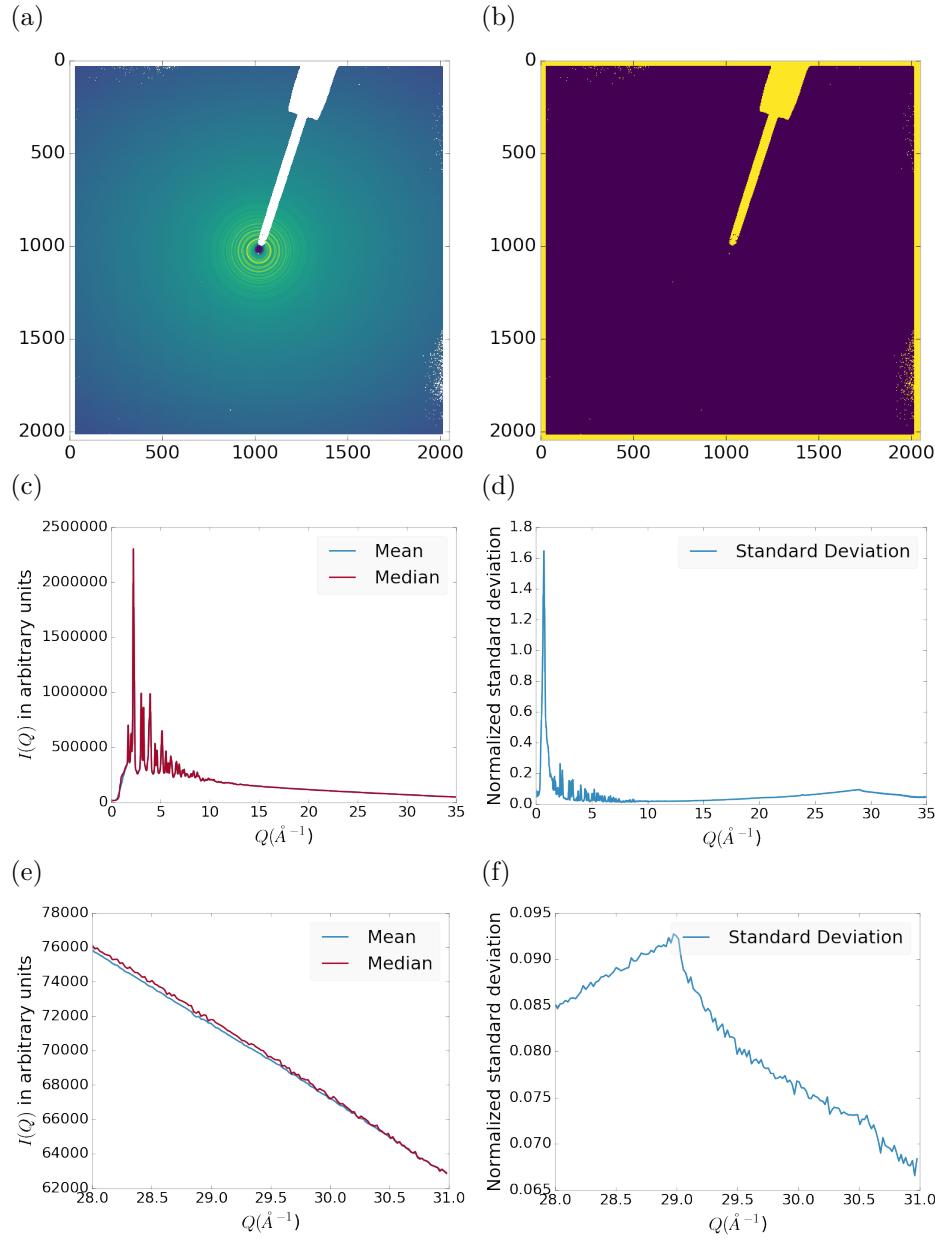


Figure 4.19: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced combining an edge mask and the automatically generated mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{ \AA}^{-1}$  to  $31 \text{ \AA}^{-1}$   $Q$  range for the mean and median, f)  $28 \text{ \AA}^{-1}$  to  $31 \text{ \AA}^{-1}$   $Q$  range for the standard deviation

626

## CHAPTER 5

627

### ANNEALING AND AGGREGATION OF 2NM

628

### AU NANOPARTICLES

629 5.1 EXPERIMENTS

630 NP Synthesis

631 X-ray Total Scattering Measurements

632 5.2 DATA PROCESSING

633 5.3 DATA ANALYSIS

634 5.4 SIMULATION

635 5.5 STRUCTURAL ANALYSIS

636 5.6 CONCLUSIONS

637

## CHAPTER 6

638

### PHASE CHANGES AND ANNEALING DYNAMICS OF

639

### $\text{Pr}_2\text{NiO}_4$ AND ITS DERIVATIVES

640 6.1 EXPERIMENTS

#### 641 $\text{Pr}_2\text{NiO}_4$ Synthesis

#### 642 X-ray Measurements

643 X-ray total scattering and x-ray powder diffraction experiments were performed at  
644 the APS's 11-ID-B beamline. An x-ray energy of XXX keV, YYY Å was used. The  
645 detector was moved between a 20cm and a 95 cm sample to detector distance to mea-  
646 sure the x-ray total scattering and x-ray diffraction patterns. Various PNO samples  
647 were annealed on the beamline during x-ray measurement.

648 6.2 DATA PROCESSING

649

masking parameters

650

integration parameters

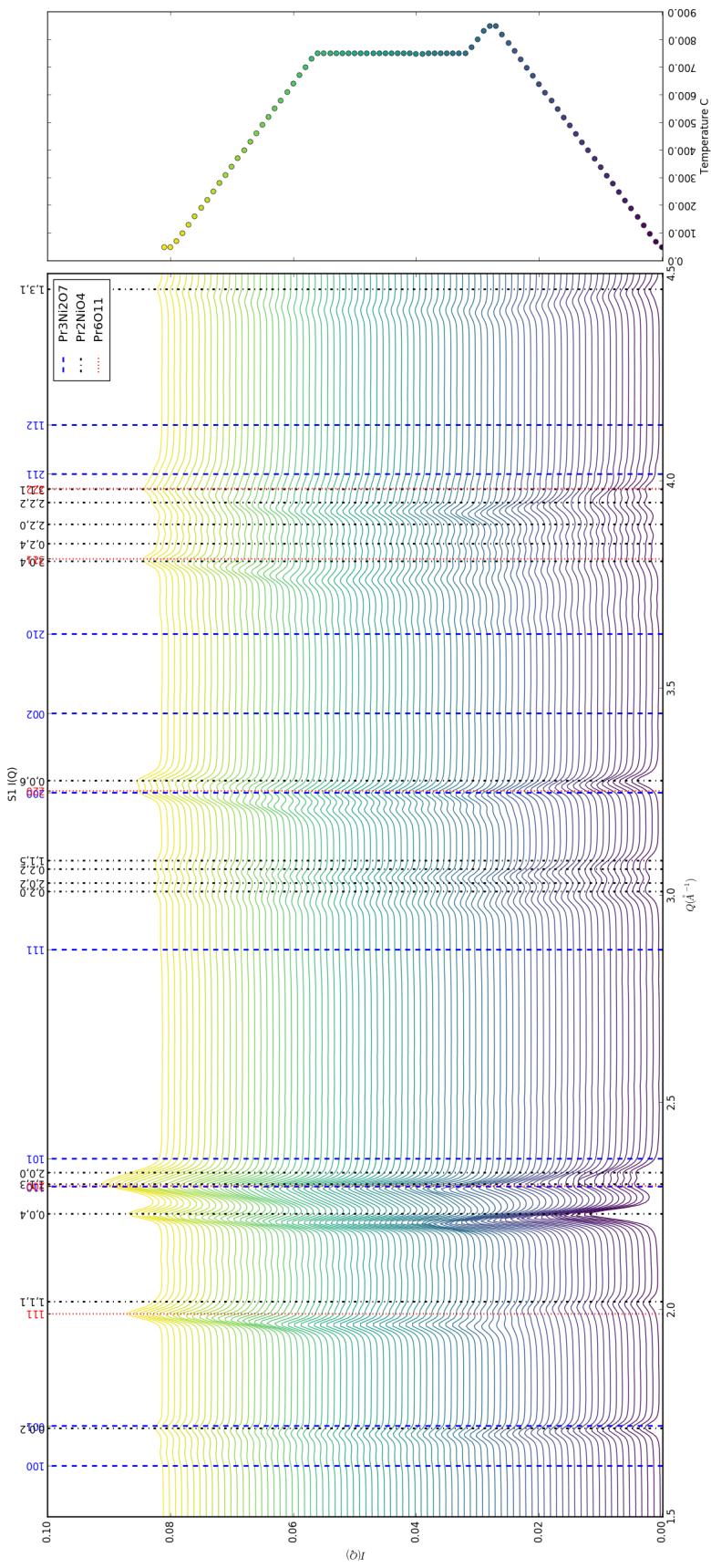
651

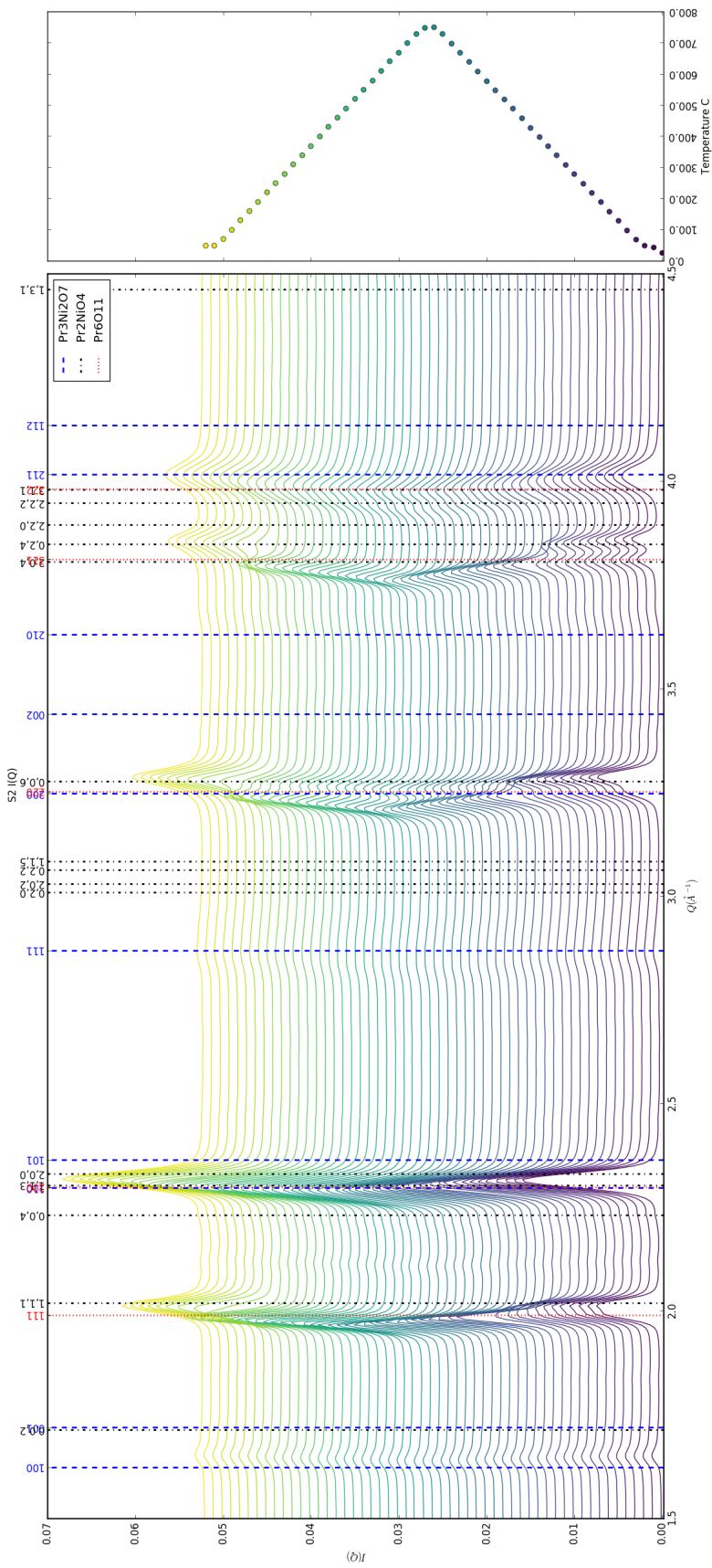
PDF parameters

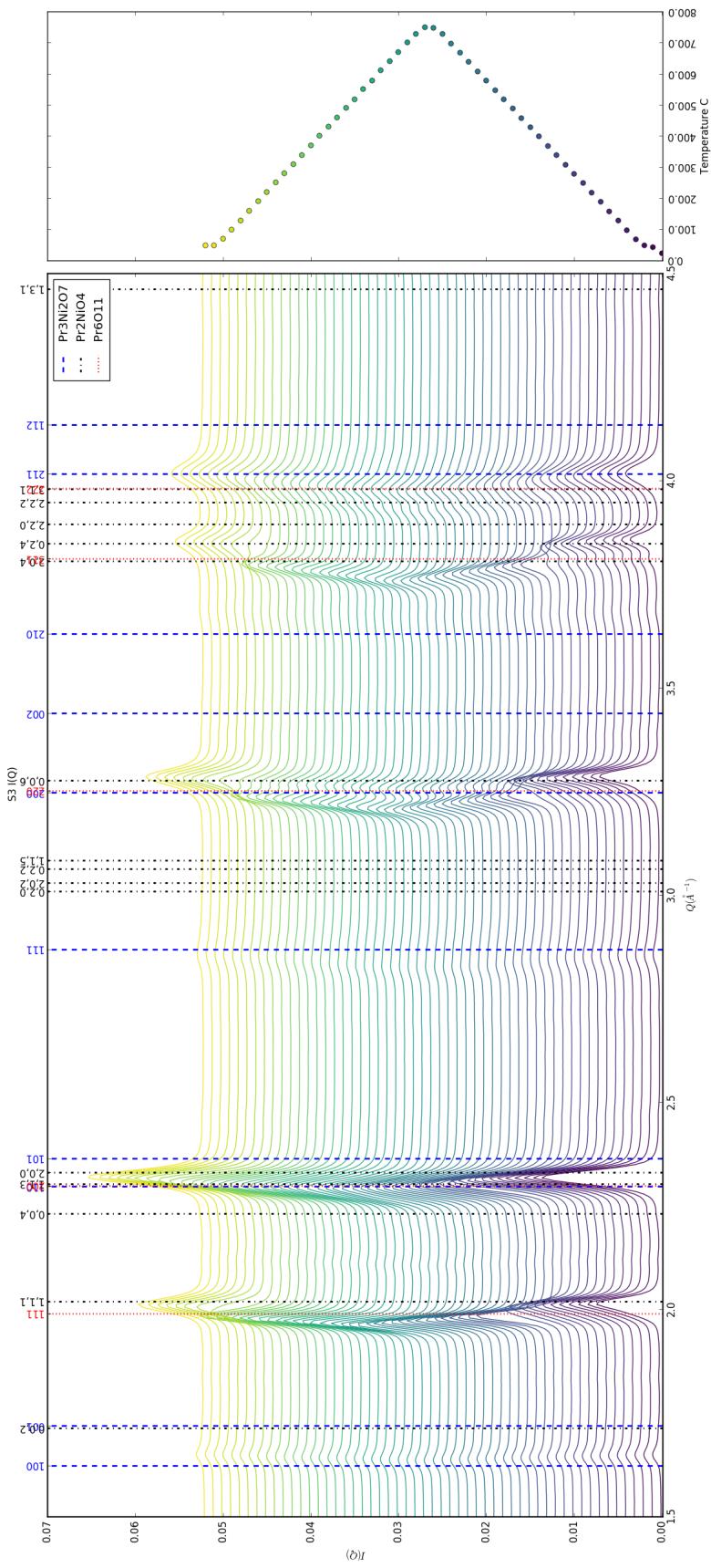
652 6.3 DATA ANALYSIS

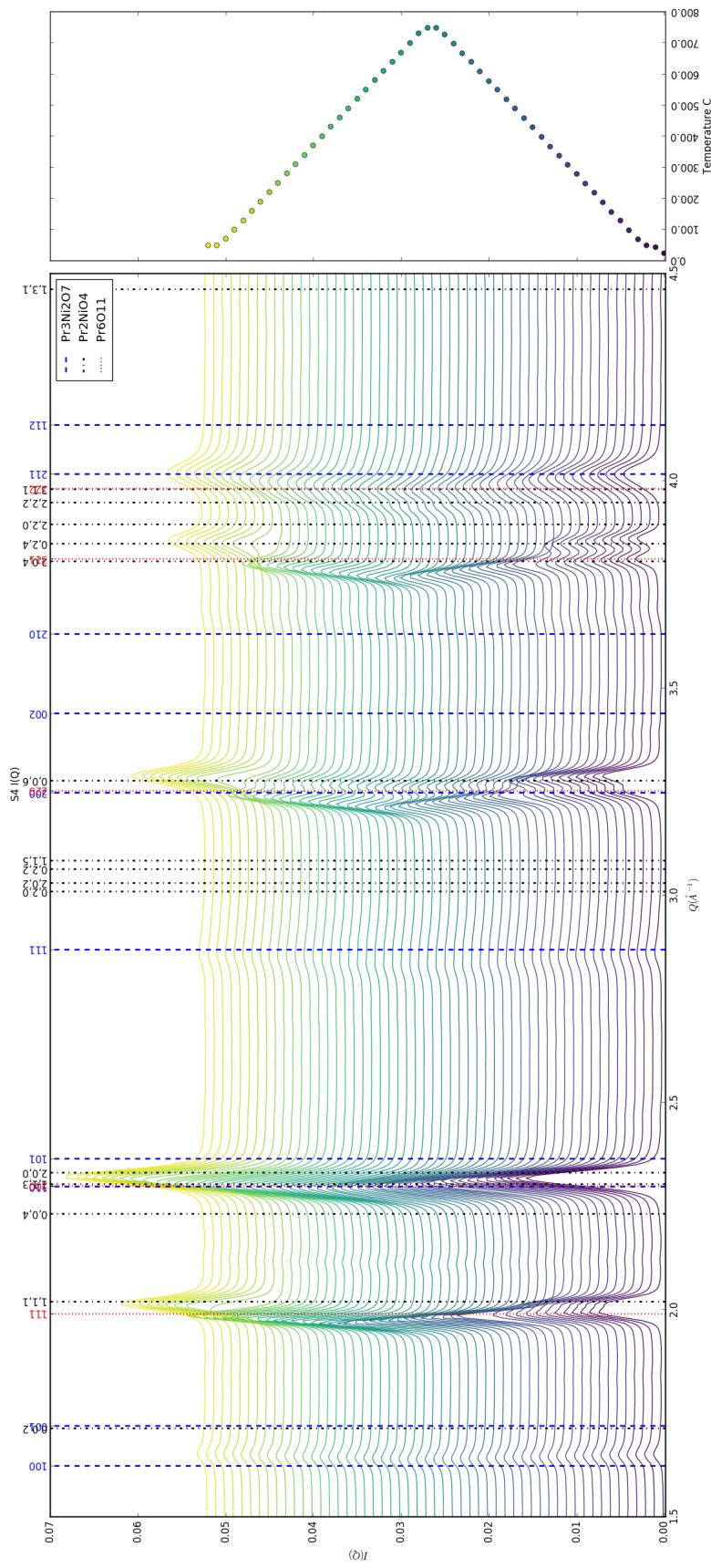
#### 653 Intra Sample Comparison

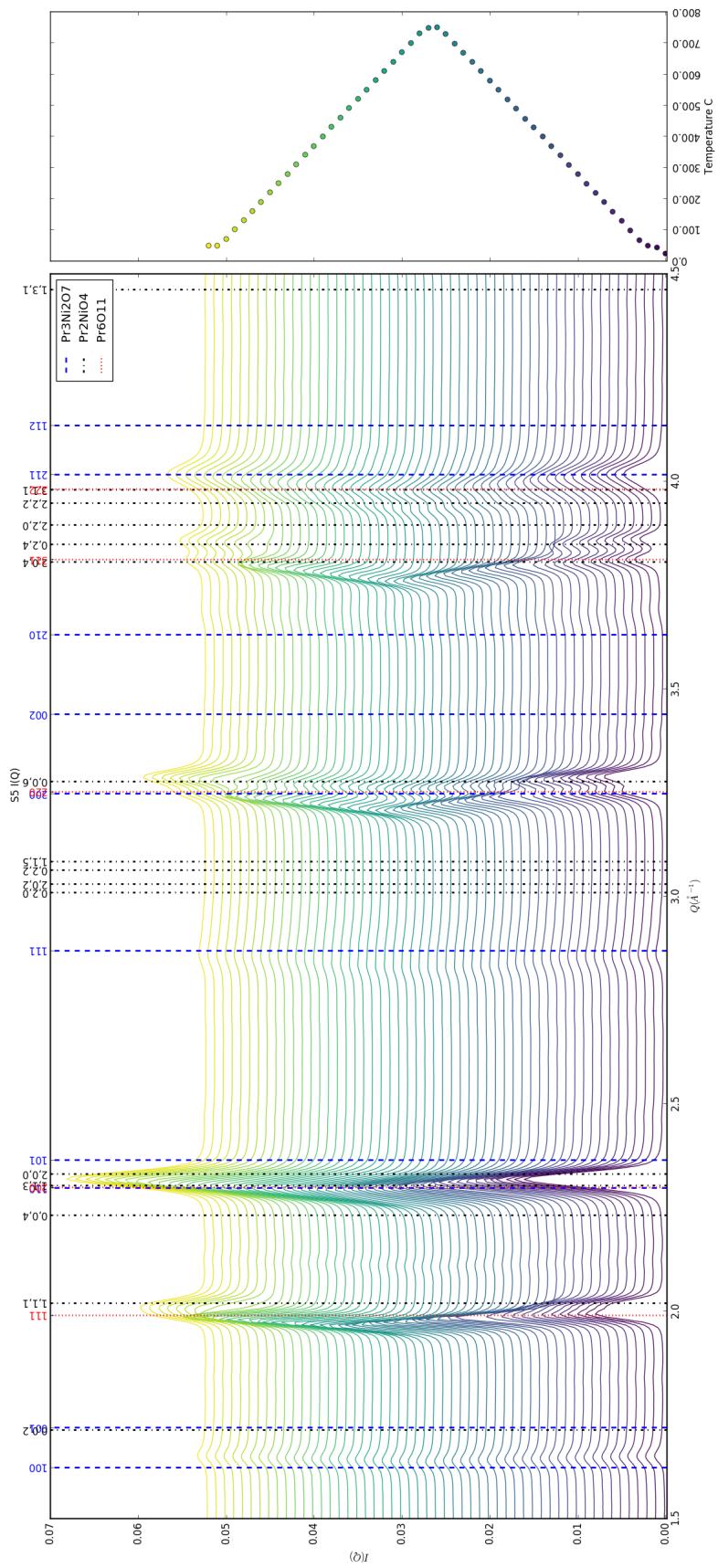
654 Changes in S1 but very little in S2-5.



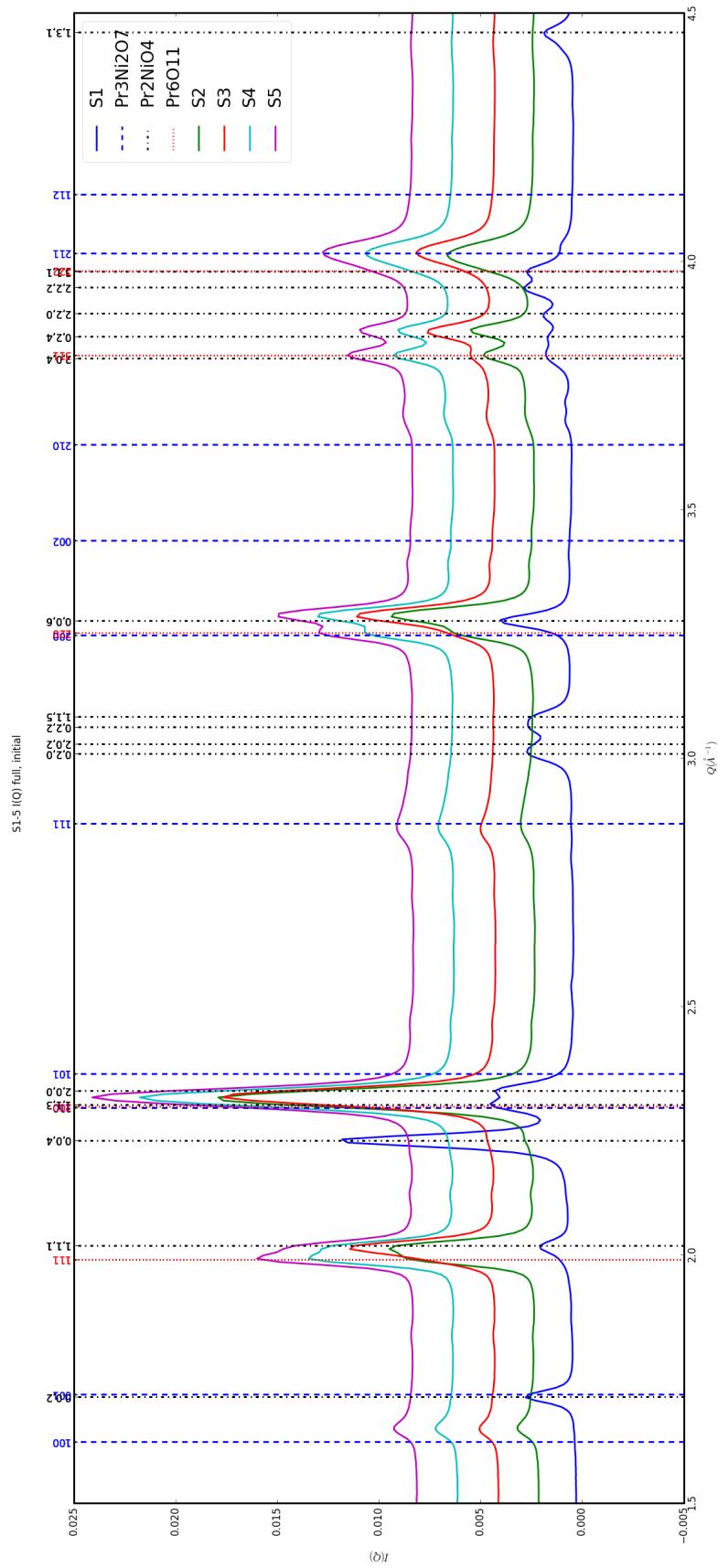


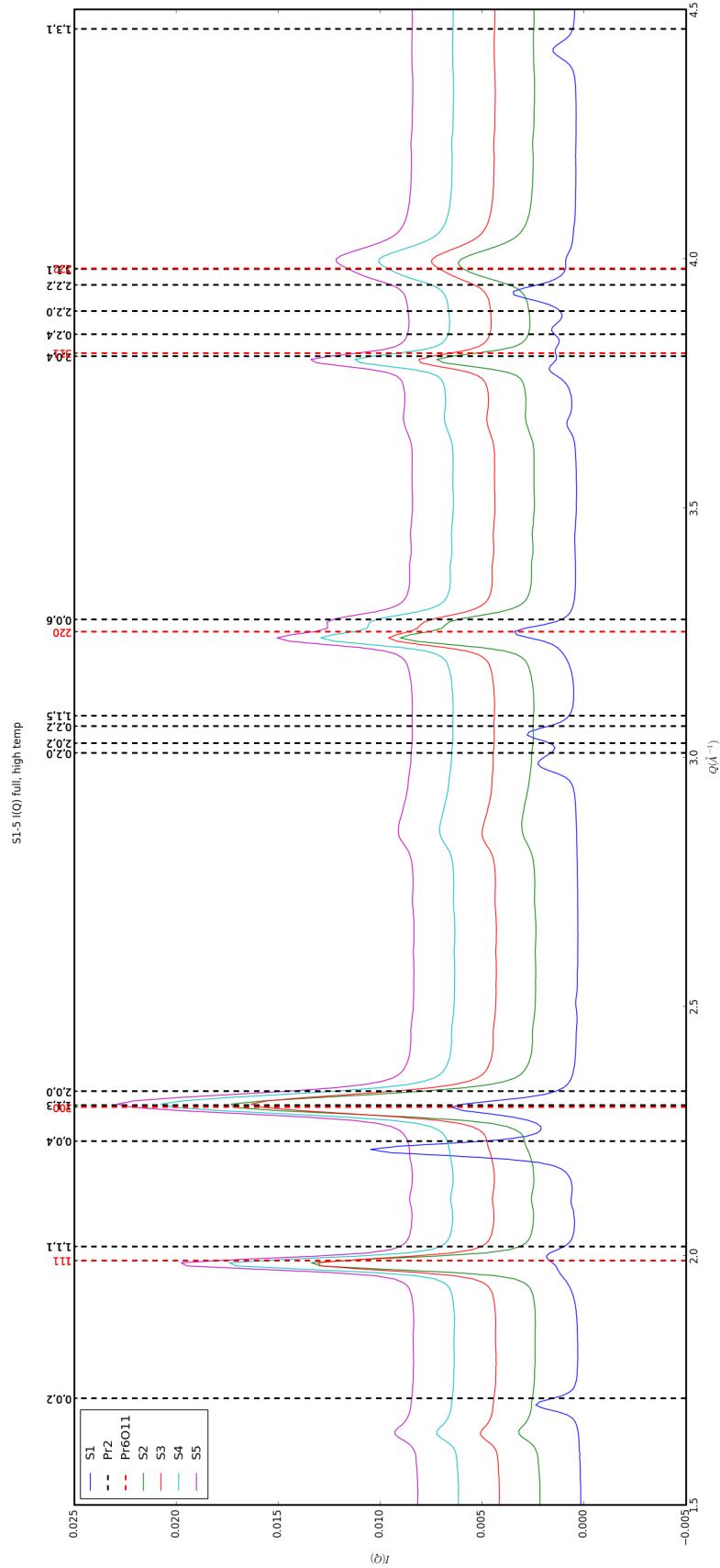


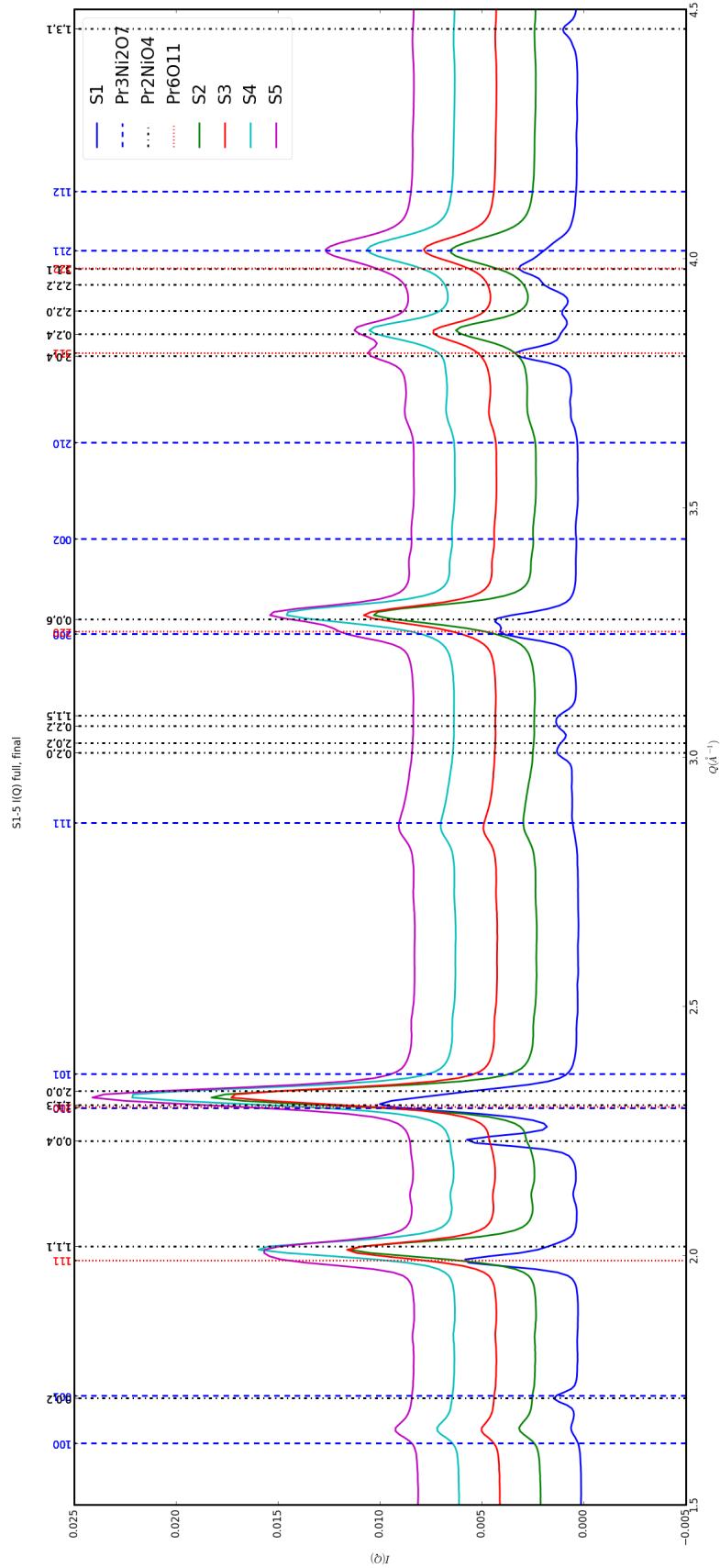


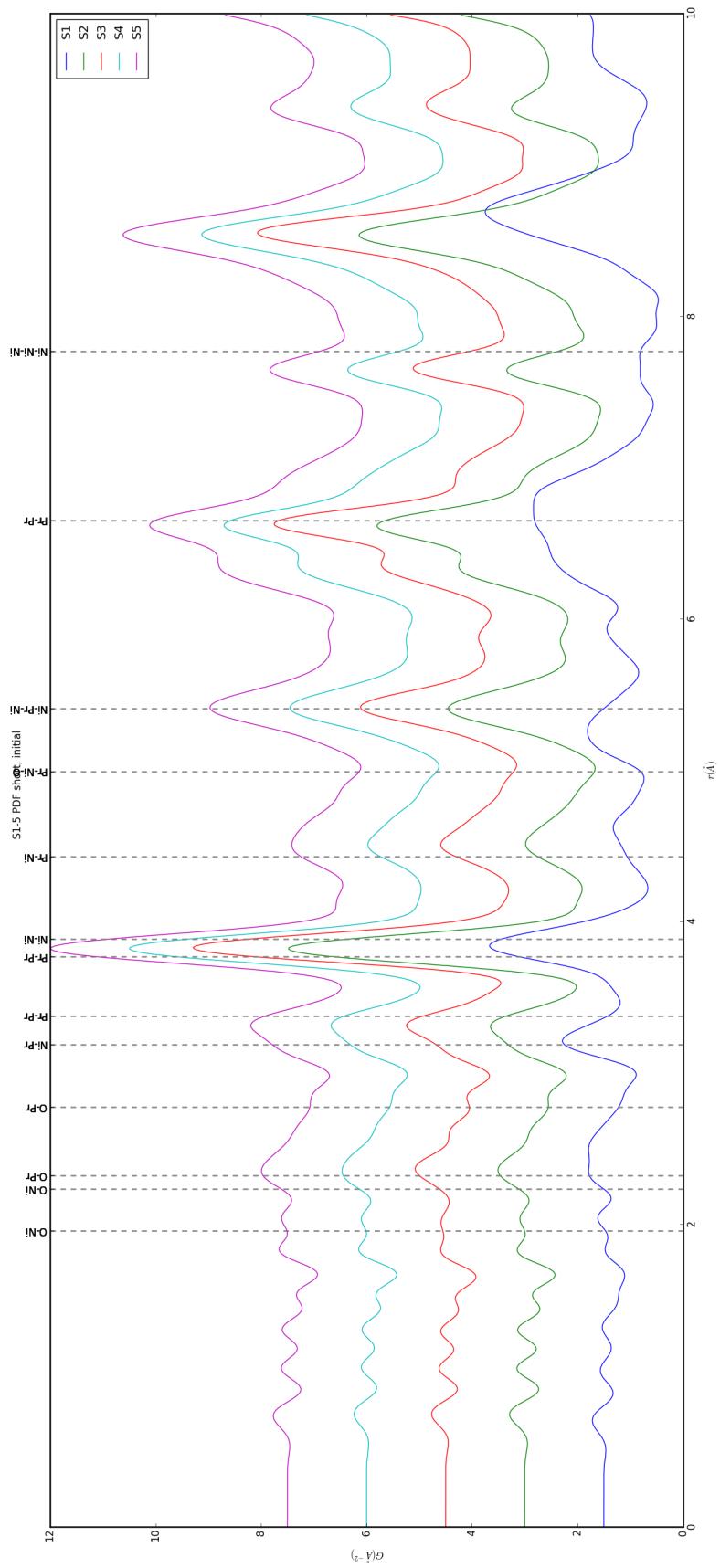


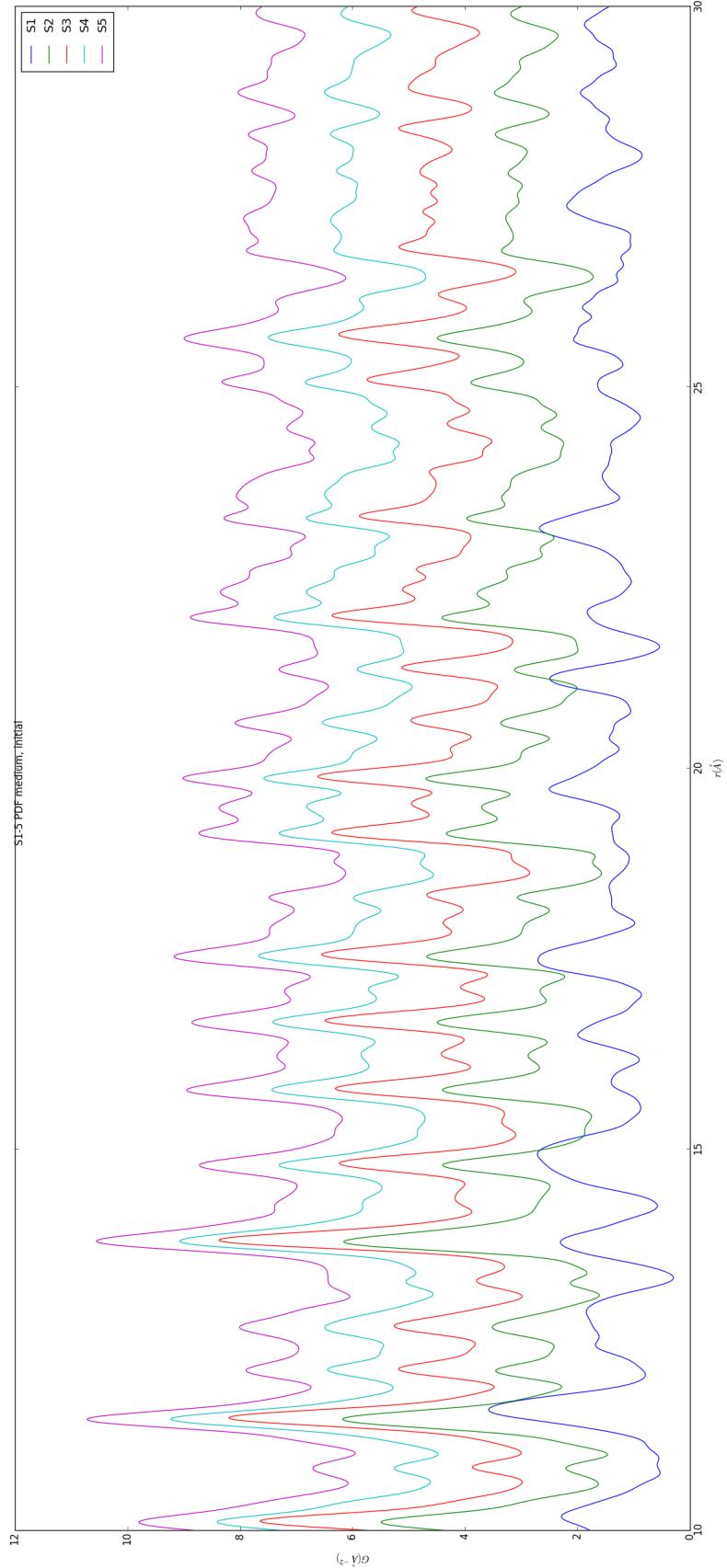
655 Inter Sample Comparison

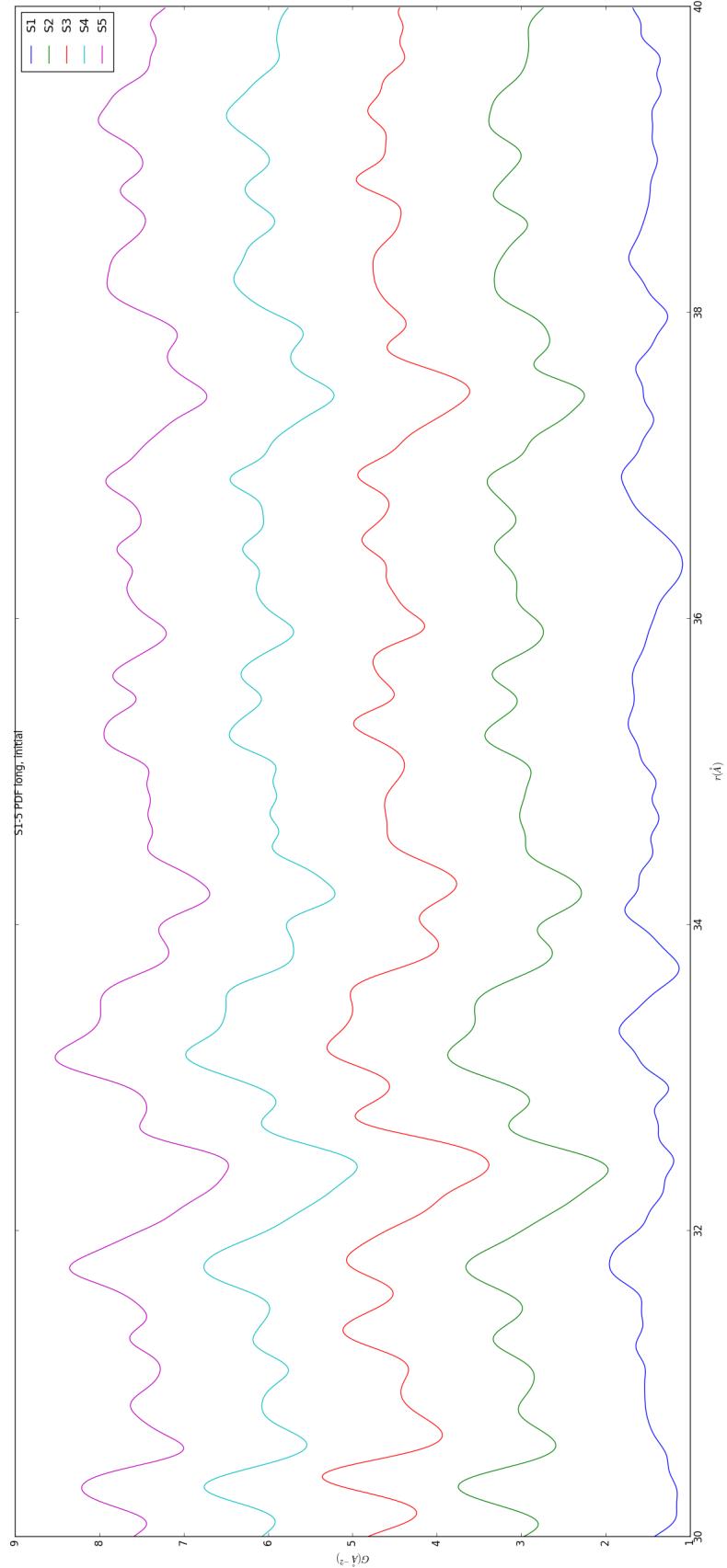


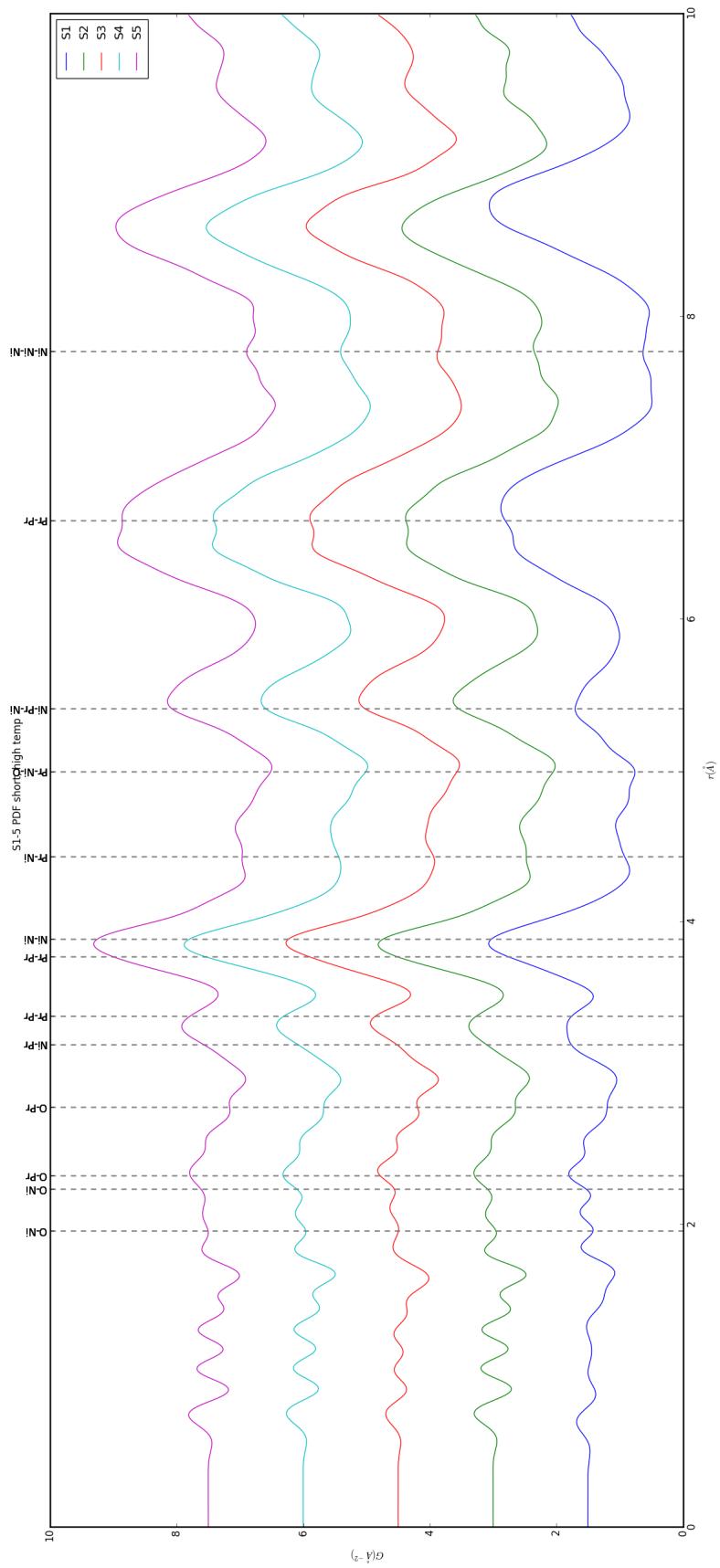


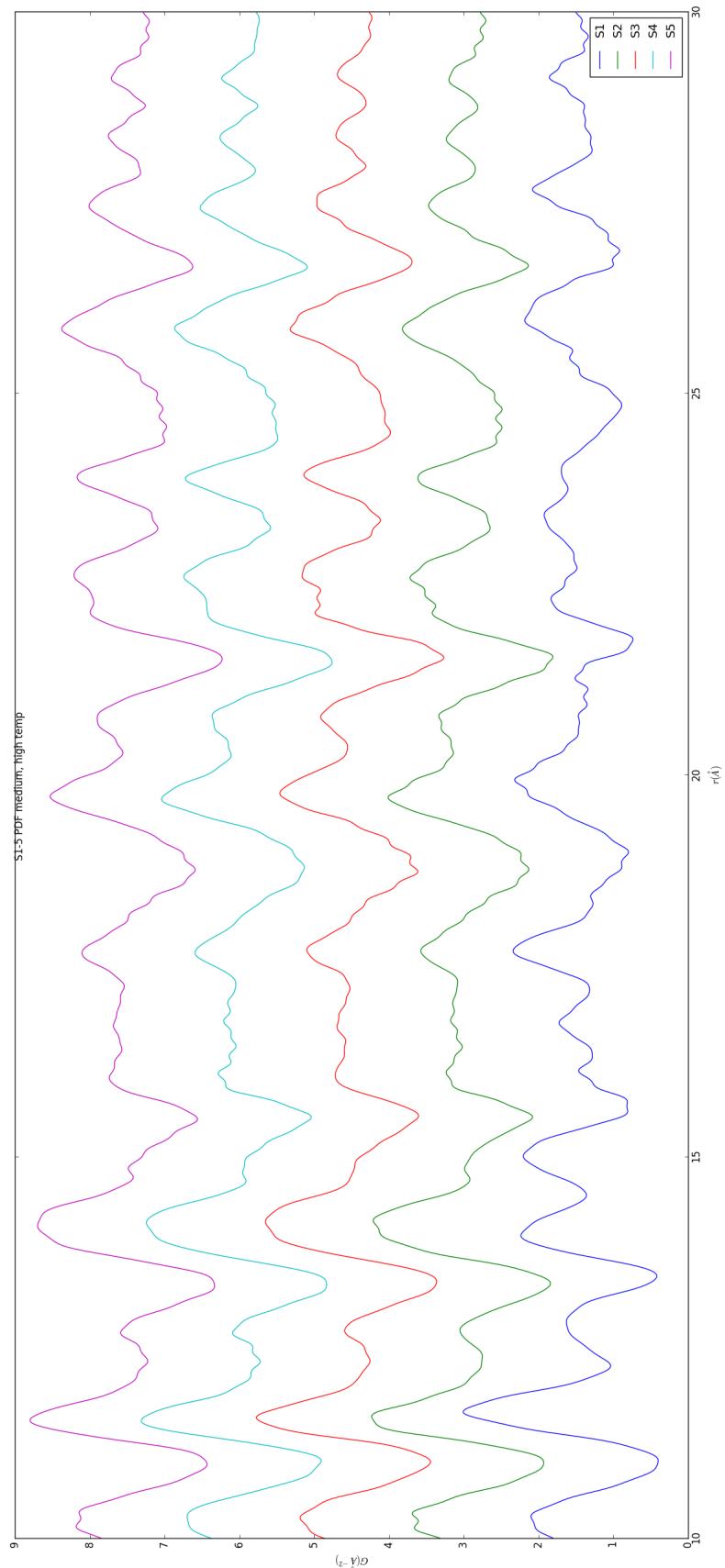


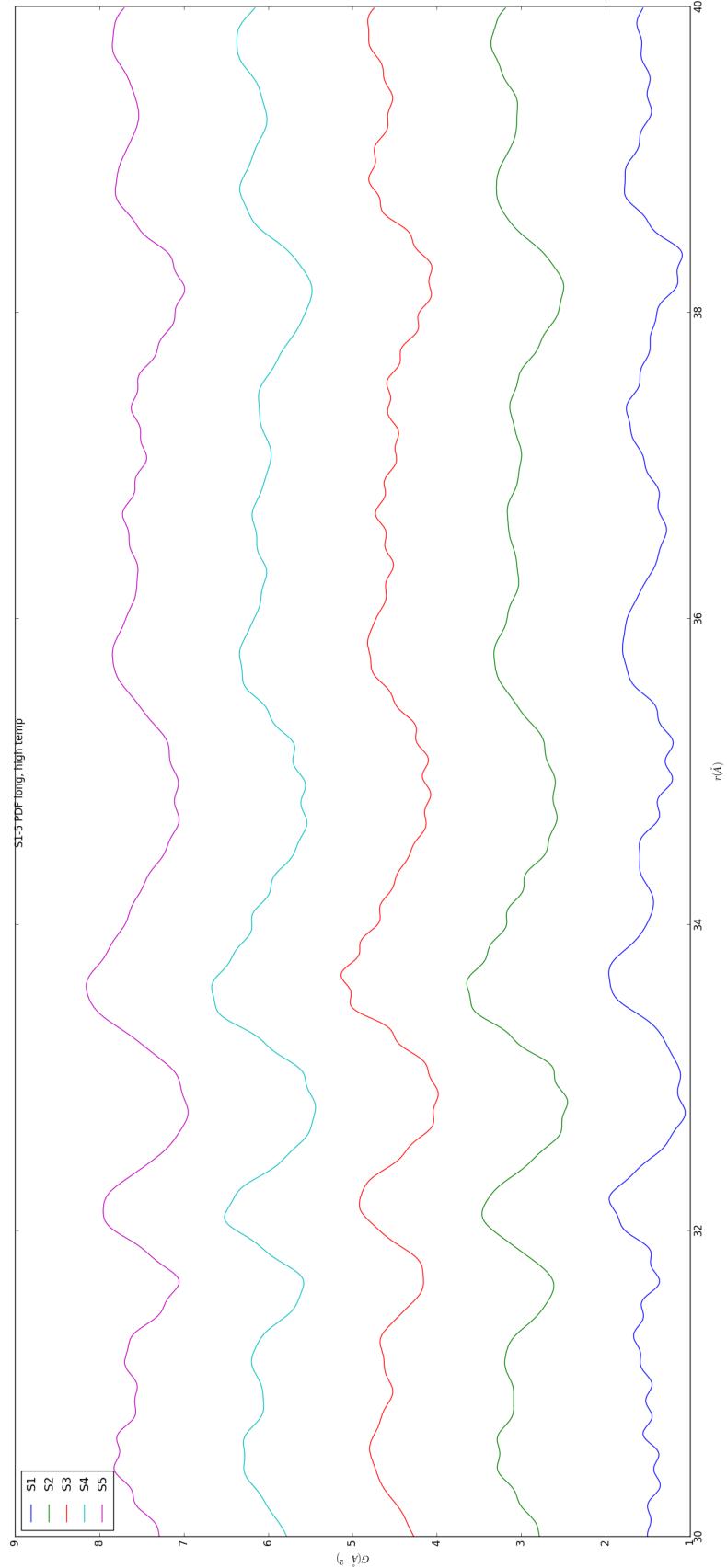


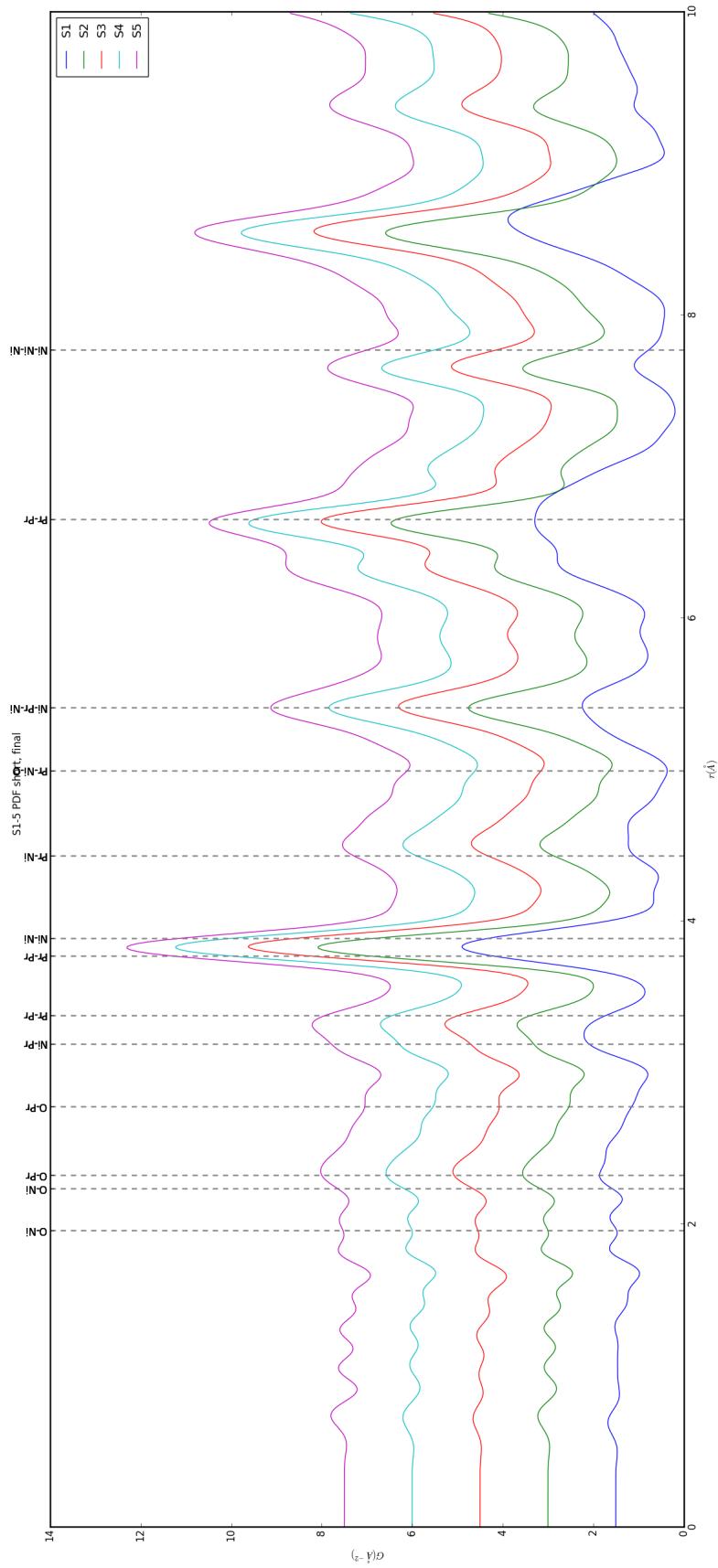


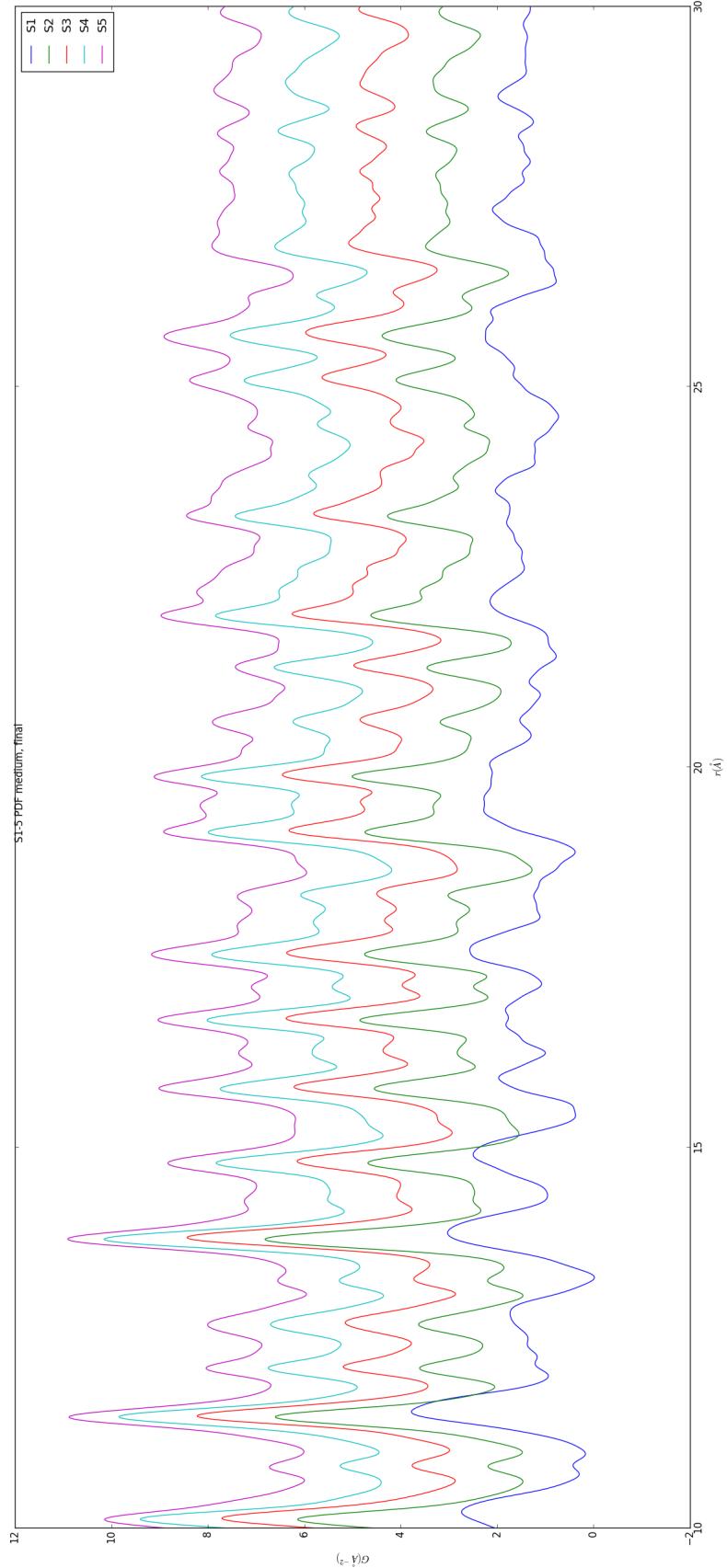


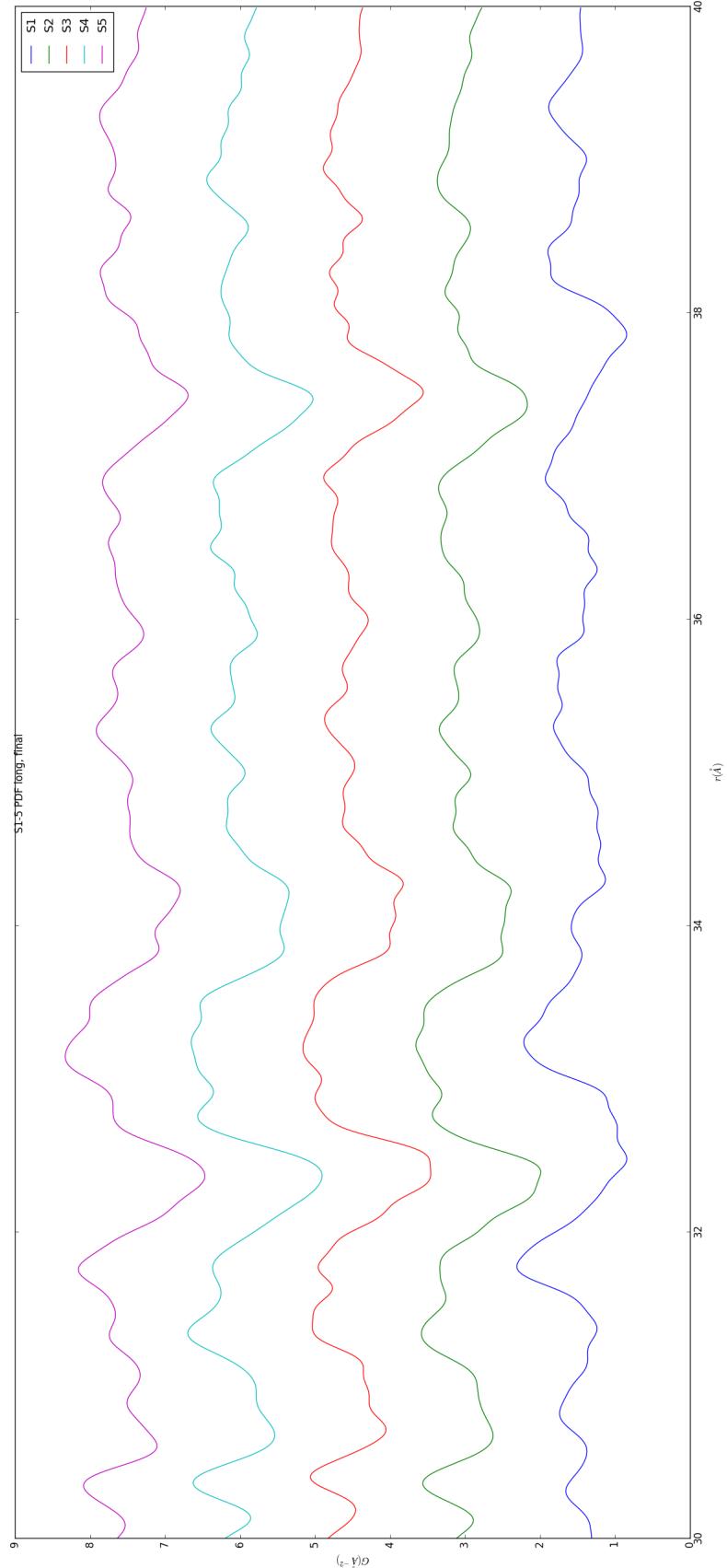












656        The PDF for S1 at operating temperature looks like S2-5 at the same temperature.

657    6.4    SIMULATION

658    Simulations have not been run yet on these PNO samples. Solving the structures of  
659    these samples is expected to be more difficult than the NP benchmarks previously  
660    solved. The difficulty of these simulations is due to:

- 661        1. The PDF's insensitivity to the oxygen positions, due to the poor x-ray scattering  
662              off the very electorn poor oxygens.
- 663        2. The large difference in mass between the oxygen and other atoms, causing the  
664              dynamics of the simulation to be governed by oxygen motion, nessecitating long  
665              simulation times to obtain movement of the other atoms.
- 666        3. The large parameter space caused by potential defects and degradation prod-  
667              ucts. Without knowing that the starting phase is pure, it is difficult to even  
668              produce starting structures, since the simulation will need to explore all the  
669              potential defect/degenerated structures.

670    6.5    CONCLUSIONS

671    X-ray total scattering and x-ray powder diffraction data was obtained on  $\text{Pr}_2\text{NiO}_4$   
672    powder samples annealed for various lengths of time. In-situ studies on the beamline  
673    were performed to understand how the structure of each of these powders changes  
674    at operating temperatures. The data was processed with the previously discussed  $Q$   
675    binning, masking, and integration methodology. The PDF results show very little  
676    change in the structure for the as synthesized sample. However, the PDFs show a  
677    large change in the previously annealed samples. These changes seem to reporduce  
678    PDFs similar to the as-synthesized PNO at operating temperatures. This would seem  
679    to imply that the source of the anamolus PNO phase/power density relationship may

680 be due to the adoption of an active structure upon heating which is universal despite  
681 the amount of thermal degradation observed at room temperature. In contrast to the  
682 PDF results, the XRD results seem to show signifigant changes in the PNO structure,  
683 both with ex-situ and in-situ annealing. The XRDs show the degradation of the PNO  
684 into various phases, potentially including  $\text{Pr}_2\text{O}_{11}$ , and higher ordered Pr based phases.  
685 The discrepancy between these two results is quite interesting as it seems that the  
686 XRD and PDF results are contradictory. Turbostratic diplacements between the  
687 layers may be a cause of the PDF/XRD disagreement, as these changes would cause  
688 very little change in the local structure observed in the PDF, while causing large  
689 changes in the XRD.

690

## CHAPTER 7

691

## CONCLUSION