

SOLVING ATOMIC STRUCTURES 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

Mark Uline, Committee Member

Jochen Lauterbach, Committee Member

Thomas Vogt, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

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

DEDICATION

For Diane & Donald Wright

My first scientific advisers

ACKNOWLEDGMENTS

ABSTRACT

Engineering the next generation of materials, especially nanomaterials, requires a detailed understanding of the material's underlying atomic structure. Understanding these structures we can obtain a better understanding of the structure-property relationship, allowing for a property driven rational design of the material structure on the atomic level. Even more importantly, understanding the structure in-situ will allow the reversal of the flow of information, translating stimuli and responses on the macroscopic system level to changes in the atomic structure. Despite the importance of atomic structures for designing materials, solving the atomic structure of materials is difficult both experimentally and computationally. Atomic pair distribution functions (PDFs) have been shown to provide information on atomic structure, although extracting the PDF from x-ray total scattering measurements can be difficult. Computationally, translating the PDF to an atomic structure requires the search of a very high dimensional space and can be computationally expensive.

This work aims to address these issues by developing 1) novel statistical mechanical approaches to solving material structures, 2) fast simulation of x-ray total scattering and atomic pair distribution functions (PDFs), and 3) data processing procedures for experimental x-ray total scattering measurements. First, experimentally derived potential energy surfaces (PES) and the statistical mechanical ensembles used to search them are developed. Then the mathematical and computational framework for the PDF and its gradients will be discussed. The combined PDF-PES-ensemble system will be bench-marked against a series of nanoparticle structures to ascertain the efficiency and effectiveness of the system. Experimental data processing proce-

dures, which maximize the usable data, will be presented. Finally, preliminary results from experimental x-ray total scattering measurements will be discussed. This work presents one of the most complete end-to-end systems for processing and modeling x-ray total scattering PDF data, potentially allowing for high-throughput structural solution.

TABLE OF CONTENTS

| | |
|--|-----|
| DEDICATION | iii |
| ACKNOWLEDGMENTS | iv |
| ABSTRACT | v |
| LIST OF TABLES | xi |
| LIST OF FIGURES | xii |
| TODO LIST | 1 |
| CHAPTER 1 STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES | 5 |
| 1.1 Introduction | 5 |
| 1.2 Potential Energy Surfaces | 5 |
| Experimentally Derived Potential Energy Surfaces | 6 |
| Potentials | 6 |
| Forces | 7 |
| 1.3 Ensembles | 8 |
| Monte Carlo Modeling | 9 |
| Hamiltonian Monte Carlo | 10 |
| Grand Canonical Ensemble | 12 |
| Ensemble description | 12 |
| Grand Canonical Monte Carlo | 12 |
| GCMC biasing | 13 |

| | |
|--|----|
| CHAPTER 2 ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION | 15 |
| 2.1 Theory | 15 |
| Derivation | 15 |
| Analytically Gradients | 15 |
| Without ADPs | 17 |
| Periodic Boundary Conditions | 17 |
| 2.2 Computation | 18 |
| HPC and GPUs | 18 |
| GPUs and Parallelization | 18 |
| Map from ij space to k space | 19 |
| GPU Memory Allocation | 21 |
| Speed and Scaling of PDF Computation | 23 |
| CHAPTER 3 BENCHMARKING | 25 |
| 3.1 Introduction | 25 |
| 3.2 PDF | 25 |
| Model Parameters | 27 |
| Au55: surface relaxed | 28 |
| Run Parameters | 28 |
| Au55: surface disordered | 30 |
| Au55: amorphous | 31 |
| Au102: triple phase | 32 |
| Starting from fcc structure | 32 |
| Marks decahedron | 33 |
| Au147 | 34 |
| 3.3 PDF with ADPs | 34 |
| ADP 50 | 34 |
| CHAPTER 4 X-RAY TOTAL SCATTERING DATA ACQUISITION AND PRO- CESSING | 40 |
| 4.1 Introduction | 40 |
| 4.2 Detector Q resolution | 40 |
| 4.3 Automated Mask Generation | 42 |

| | |
|--|----|
| Introduction | 42 |
| Algorithm Design | 44 |
| Test Cases | 44 |
| Results and Discussion | 45 |
| Conclusions | 51 |
| 4.4 Automated Image Azimuthal Integration | 52 |
| 4.5 Conclusions | 54 |
| CHAPTER 5 ANNEALING AND AGGREGATION OF 2NM AU NANOPARTICLES | 57 |
| 5.1 Experiments | 57 |
| NP Synthesis | 57 |
| X-ray Total Scattering Measurements | 57 |
| 5.2 Data Processing | 57 |
| 5.3 Data Analysis | 57 |
| 5.4 Simulation | 57 |
| 5.5 Structural Analysis | 57 |
| 5.6 Conclusions | 57 |
| CHAPTER 6 PHASE CHANGES AND ANNEALING DYNAMICS OF Pr_2NiO_4 AND ITS DERIVATIVES | 58 |
| 6.1 Experiments | 58 |
| Pr_2NiO_4 Synthesis | 58 |
| X-ray Measurements | 58 |
| 6.2 Data Processing | 58 |
| 6.3 Data Analysis | 59 |
| Intra Sample Comparison | 59 |
| Inter Sample Comparison | 65 |
| 6.4 Simulation | 78 |
| 6.5 Conclusions | 78 |

| | |
|--------------------------------|----|
| CHAPTER 7 CONCLUSION | 80 |
| BIBLIOGRAPHY | 81 |

LIST OF TABLES

LIST OF FIGURES

| | | |
|------------|--|----|
| Figure 2.1 | Comparison of the CPU and GPU chip architectures | 19 |
| Figure 2.2 | Speed comparison of CPU and GPU implementations | 23 |
| Figure 3.1 | Au ₅₅ PDF fitting of DFT-optimized <i>c</i> -Au ₅₅ | 29 |
| Figure 3.2 | Au ₅₅ PDF fitting of surface-disordered Au ₅₅ | 35 |
| Figure 3.3 | Similar to figure 3.2 for DFT-optimized amorphous Au ₅₅ | 36 |
| Figure 3.4 | Similar to figure 3.2 for Au ₁₀₂ as in DFT-optimized Au ₁₀₂ MBA ₄₄ cluster. | 37 |
| Figure 3.5 | Similar to Fig. 3.4 with Marks decahedron as the starting structure. | 38 |
| Figure 3.6 | Refinement of adps | 39 |
| Figure 4.1 | Scattering onto a flat detector | 41 |
| Figure 4.2 | Q resolution as a function of Q | 42 |
| Figure 4.3 | Number of pixels as a function of Q , binned at the Q resolution of the detector. | 43 |
| Figure 4.4 | Generated dead/hot pixel masks for a detector with 100 bad pixels. | 45 |
| Figure 4.5 | Generated dead/hot pixel masks for a detector with 300 bad pixels. | 46 |
| Figure 4.6 | Generated dead/hot pixel masks for a detector with 500 bad pixels. | 46 |
| Figure 4.7 | Generated dead/hot pixel masks for a detector with 1000 bad pixels. | 47 |
| Figure 4.8 | Generated beamstop holder masks for a beamstop holder with 10% transmittance. | 47 |

| | |
|---|----|
| Figure 4.9 Generated beamstop holder masks for a beamstop holder with 30% transmittance. | 48 |
| Figure 4.10 Generated beamstop holder masks for a beamstop holder with 50% transmittance. | 48 |
| Figure 4.11 Generated beamstop holder masks for a beamstop holder with 90% transmittance. | 48 |
| Figure 4.12 Generated beamstop holder masks which is rotated away from vertical | 49 |
| Figure 4.13 Masked experimental data. | 49 |
| Figure 4.14 Masked experimental data with Pt single crystal signal. | 50 |
| Figure 4.15 Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask. | 50 |
| Figure 4.16 Masking, average, and standard deviation of an example x-ray total scattering measurement with with no mask. | 53 |
| Figure 4.17 Masking, average, and standard deviation of an example x-ray total scattering measurement with with only an edge mask. | 54 |
| Figure 4.18 Masking, average, and standard deviation of an example x-ray total scattering measurement with combining an edge mask and the automatically generated mask. | 55 |
| Figure 6.1 XRD as a function of temperature for S1 | 60 |
| Figure 6.2 XRD as a function of temperature for S2 | 61 |
| Figure 6.3 XRD as a function of temperature for S3 | 62 |
| Figure 6.4 XRD as a function of temperature for S4 | 63 |
| Figure 6.5 XRD as a function of temperature for S5 | 64 |

TODO LIST

| | | |
|----|--|----|
| 2 | Need more references | 2 |
| 3 | Why is atomistic engineering important | 3 |
| 4 | Barriers to atomistic engineering | 3 |
| 5 | How are we going to attack this problem | 3 |
| 6 | double check against McQuerry, cite it | 9 |
| 7 | Talk about the advances NUTS gives us | 11 |
| 8 | What limitations in RMC/HMC/NUTS/Cannonical ensembles in general force us to move to GCMC | 12 |
| 10 | Include figure which shows the configurational biasing map | 14 |
| 11 | cite something with debye waller most likely simon's paper | 15 |
| 12 | Need to include PBC ADP math | 17 |
| 13 | Also should include PBC gradients, although they are trivial, maybe? | 18 |
| 14 | How does this compare against the Ewald simulation technique for ionic solutions | 18 |
| 15 | use the figure from the presentation, it is better | 18 |
| 16 | I wish we had a picture of the GPU profiling | 18 |
| 17 | This entire section needs some rewritting to distinguish this from the paper . | 25 |
| 18 | Also some introduction would be great | 25 |
| 19 | this just needs a lot of work | 25 |
| 20 | Put this somewhere | 29 |
| 21 | If we are going to keep this we need to get a lot done in terms of modeling . | 57 |
| 22 | need some sort of synthesis information | 58 |
| 23 | masking parameters | 58 |

| | | |
|----|--|----|
| 24 | integration parameters | 58 |
| 25 | PDF parameters | 58 |
| 26 | Need to redo all the figures: | |
| 27 | need to use the new stylesheet | |
| 28 | need to be more readable | |
| 29 | need to cull through them to pull out the interesting features | 59 |
| 30 | Need more references | |

31

INTRODUCTION

32

Why is atomistic engineering important

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

37

Barriers to atomistic engineering

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

41

How are we going to attack this problem

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

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

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

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

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

56 Chapter 4 will focus on the aquesition of experimental data, its management, and
57 processing. §4.2, 4.3, and 4.4 will discuss the derivation of the Q resolution function,
58 the automated masking of 2D area detectors for x-ray total scattering measurements
59 using the previously derrived Q resolution, and the impact of different averaging
60 methods and masks on azimuthal integration, respecitvly.

61 CHAPTER 1

62 STATISTICAL MECHANICAL ENSEMBLES AND
63 POTENTIAL ENERGY SURFACES

64 1.1 INTRODUCTION

65 The approach taken in this work for solving the atomic structures of materials is one
66 of optimization. The plan is to develop a potential energy surface (PES) which has
67 minima associated with atomic structures who's properties match the experimentally
68 observed properties. Thus, the various positional variables of the structure can be
69 solved by optimizing the structure against the PES. This approach is popular in the
70 PDF community for solving the structure of materials using both extensive large box
71 models and simpler small box models.

72 In this chapter we discuss the development of the various PESs used in the PDF
73 community for comparing theoretical and experimental PDFs. Special attention will
74 be paid to the gradients of the potential energy functions, as these are important
75 to some optimization techniques. Additionally, we also discuss the use of statistical
76 mechanical ensembles for finding minima on the PES.

77 1.2 POTENTIAL ENERGY SURFACES

78 A PES simply describes the potential energy of the system as a function of all its
79 relevant coordinates in phase space, essentially providing a mapping $\mathbb{R}^n \rightarrow \mathbb{R}$, where \mathbb{R}
80 is the set of real numbers and n is the number of positional parameters in the system.
81 Usually these coordinates are the positions of the atoms q and their conjugate the

82 momenta p . Note that there could be more variables associated with the system,
 83 for instance the magnetic moments of the atoms could play a role in describing the
 84 system. In this magnetic system there would be positional variables for the atomwise
 85 spin vectors and their "momenta". Application of the term "momenta" might seem
 86 odd here, as the magnetic spin does not have a mass or a velocity. However, since the
 87 magnetic "position" is defined on the PES we need to describe its conjugate variable
 88 to properly formulate Hamiltonian dynamics and the kinetic portion of the PES.

89 Experimentally Derived Potential Energy Surfaces

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

98 Potentials

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

$$101 \quad \chi^2 = \sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2 \quad (1.1)$$

$$102 \quad R_w = \sqrt{\frac{\sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}}^2}} \quad (1.2)$$

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

103

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

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

107 The Rw and χ^2 potentials have been reported numerous times. [20, 15, 2, 16, 22]
 108 Essentially these potentials measure the least squares distance between the observed
 109 scattering and the predicted scattering providing a way to quantify the agreement
 110 between the model and experiment. While RW and χ^2 are now standard in the PDF
 111 community, the INVERT potential is fairly new and aims to incorporate descriptions
 112 of the structural symmetry into the PES. [3, 4] In the case of the INVERT potential
 113 NMR or other symmetry sensitive data is used to describe the number of unique
 114 atomic coordinations. This is then used to describe the number of unique atomwise
 115 pair distribution functions, thus causing systems with more or less unique coordi-
 116 nation environments to be higher in energy. This approach has been shown to be
 117 useful for C_{60} and other systems which are highly symmetric, creating a PES with
 118 an easier to find minima. [3, 4] However, many times this kind of data is unavailable
 119 when refining the structure causing the potential to be less useful. Additionally, this
 120 potential introduces an element of user bias as the refiner must decide, based on some
 121 spectroscopic data, how many unique environments are in the material. This bias
 122 could be removed by using one of the other potentials with a method for simulat-
 123 ing the observed spectra, allowing the computational system decide what structures
 124 properly reproduce all the observed data.

125 **Forces**

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

126

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

127

$$\vec{\nabla}\chi_{\text{INVERT}}^2 = \frac{-2}{N} \sum_{a=a_{\min}}^{a_{\max}} \sum_j (\alpha \frac{\partial A_{j,\text{calc}}}{\partial \gamma_{i,w}} + A_{j,\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}})(A_{\text{obs}} - \alpha A_{j,\text{calc}}) \quad (1.7)$$

128

$$\frac{\partial \alpha}{\partial \gamma_{i,w}} = \frac{(\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} - 2\alpha \sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}})}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}}^2} \quad (1.8)$$

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

134 1.3 ENSEMBLES

135 While PESs describe which atomic configurations are the most desirable and how
 136 the atoms would like to get there, the ensemble describes how the atoms move on
 137 the PES. The abstraction of the PES from the ensemble is an important one, as it
 138 allows for the reuse and exchange of both PESs and ensembles for a wide array of
 139 problems. Statistical mechanical ensembles can be described in two ways, analytically
 140 and stochastically. For long simulation times and fine enough numerical or analytical
 141 integration these two descriptions should be identical.

142 In either case one starts by defining the Hamiltonian, \mathcal{H} , as the total energy of
 143 the system. Thus, the Hamiltonian is described as the sum of the potential $U(q)$ and
 144 kinetic $K(p)$ energies, where q is the positions of the atoms and p is their momenta

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

145 where $K(p) = \frac{1}{2} \sum_i \frac{p_i^2}{m_i}$ and i denotes the i th particle.

146 Analytically one generally defines a partition function, which describes the sum of
147 probabilities over all potential atomic states.

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

148 where P_i is the probability of the i th state and is a function of the total energy of
149 that state. This partition function can then be used to obtain the probability of any
150 specific state. The relationship of the probability of a state to the state's energy and
151 other properties depends on the ensemble being used.

152 double check against McQuerry, cite it

153 For the microcanonical ensemble the probability of a state is:

$$P(q, p) = \frac{\delta(\mathcal{H}(q, p) - E)}{W} \quad (1.10)$$

154 where E is the energy of the system, W is the total number of states in the system,
155 and δ is the Dirac Delta Function.

156 However, for the canonical ensemble the probability is:

$$P(q, p) = \exp\left(\frac{E - \mathcal{H}(q, p)}{k_b T}\right) \quad (1.11)$$

157 Monte Carlo Modeling

158 Monte Carlo can be used to simulate a statistical mechanical ensemble which can
159 not be solved analytically. In most Monte Carlo systems the ensemble is simulated by
160 randomly changing one of the system parameters and comparing the energy of the
161 new system against the energy of the old system. If the energy of the new system is
162 lower than the current energy then the new configuration is accepted. Otherwise the
163 new system is rejected unless

$$\exp\left(\frac{-\Delta E}{E_T}\right) < u$$

164 where u is a random number $[0, 1]$ and E_T is the thermal energy characteristic to the
165 system. The ability of Monte Carlo modeling to accept "bad" moves allows the system

166 to hop out of local energy minima during the search for the global minimum. Reverse
167 Monte Carlo (RMC) is similar to Monte Carlo except it uses χ^2 as the PES.[16]

168 Despite the utility of RMC, and its wide use in the x-ray scattering community,
169 as Hoffman and Gelman state “Not all MCMC [Markov Chain Monte Carlo]
170 algorithms are created equal”.[9] RMC, similar to standard Monte Carlo simulations,
171 samples from the PES at random, usually by translating atoms in the system ran-
172 domly. This creates a less efficient, random walk based, exploration of the PES.[9, ?]
173 Thus, methods for suppressing this random walk nature, while still searching the
174 potential energy surface fully are needed.

175 Hamiltonian Monte Carlo

176 Hybrid or Hamiltonian Monte Carlo (HMC) can help to address some of these issues.
177 HMC was developed originally in the lattice quantum chromodynamic community
178 and provides a more efficient, more scalable approach to PES sampling for Monte
179 Carlo.[?, 17] In HMC the PES is explored using Hamiltonian dynamics, essentially
180 following the gradient of the PES to find more acceptable configurations.

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

$$\frac{dq_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} = p_i \quad (1.12)$$

$$\frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial q_i} = -\vec{\nabla}U \quad (1.13)$$

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

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

$$q_i(t + \delta t) = q_i(t) + \delta t * p_i(t + \delta t/2) \quad (1.15)$$

$$p_i(t + \delta t) = p_i(t + \delta t/2) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t + \delta t)) \quad (1.16)$$

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

$$\vec{\nabla}U = \begin{bmatrix} \frac{\partial U}{\partial q_{0,x}} & \frac{\partial U}{\partial q_{0,y}} & \frac{\partial U}{\partial q_{0,z}} \\ \vdots & \frac{\partial U}{\partial q_{i,w}} & \vdots \\ \frac{\partial U}{\partial q_{n,x}} & \frac{\partial U}{\partial q_{n,y}} & \frac{\partial U}{\partial q_{n,z}} \end{bmatrix} = \begin{bmatrix} \vec{\mathcal{F}}_0 \\ \vdots \\ \vec{\mathcal{F}}_i \\ \vdots \\ \vec{\mathcal{F}}_n \end{bmatrix} \quad (1.17)$$

183 where $\frac{\partial}{\partial q_{i,w}}$ is the derivative with respect to q where w denotes direction of the deriva-
 184 tive (x , y , or z), n is the number of atoms and U is the potential which depends on q ,
 185 and $\vec{\mathcal{F}}_i$ is the "force" on the i th atom. Using these equations new potential configura-
 186 tions are proposed from the PES. These proposals are checked against the standard
 187 Metropolis criteria discussed above, except that the change in potential energy ΔE
 188 is replaced with the change in the Hamiltonian $\Delta \mathcal{H}$. Note that while this sampling
 189 closely simulates the canonical ensemble, it is not exactly the same. Usually the
 190 canonical ensemble is formulated as microcanonical ensembles in contact with an
 191 infinite heat bath at a given temperature, or a set of microcancoical ensembles which
 192 exchange thermal energy. However, the HMC ensemble presented here has a momen-
 193 tum bath instead of a temperature bath. One could imagine the atoms sitting in a
 194 simulation box which has walls which can toggle their thermal exchange. Initially the
 195 box starts in the momentum bath, allowing the atoms to come to equilibrium with
 196 the bath momentum. The box is then removed from the bath causing it to become
 197 adiabatic. Hamiltonian dynamics are then propagated inside the box, essentially run-
 198 ning a microcanonical simulation. Once the dynamics are finished the energy of the
 199 system is checked with the Metropolis criteria and the box is reintroduced to the
 200 momentum bath and the process starts again.

201 Talk about the advances NUTS gives us

202 **Grand Canonical Ensemble**

203 What limitations in RMC/HMC/NUTS/Cannonical ensembles in general force us
to move to GCMC

204 **Ensemble description**

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

$$\Xi = e^{-\beta(\mathcal{H}+\mu)} \quad (1.18)$$

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

211 **Grand Canonical Monte Carlo**

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

$$\min[1, \frac{V}{(N+1)\Lambda(T)^3} e^{-\beta\Delta U + \beta\mu}] \quad (1.19)$$

215 Similarly the removal of an atom has the probability:

$$\min[1, \frac{(N)\Lambda(T)^3}{V} e^{-\beta\Delta U - \beta\mu}] \quad (1.20)$$

216 However, both of these equations depend of the overall simulation volume and the
217 thermal wavelength, which is undesirable as these are not really properties that we
218 are of interest to these simulations. Thus, we roll them into the definition of the
219 chemical potential, essentially setting the base chemical potential to counteract these
220 effects. This makes certain that our simulation does not change if we change the

221 overall cell volume. A GCMC move consists of creating a new atomic configuration,
222 where an atom has been added or removed, and checking the above criteria. However,
223 previous results have shown that this method is computationally expensive in dense
224 liquids, and exceedingly expensive in solid materials. The long simulation times
225 are due to the random nature of the atomic additions or removals which produce:
226 over-tightly packed atoms, atoms in the middle of nowhere, or unphysical vacancies.
227 These configurations are rejected by the GCMC criteria but their probability of being
228 sampled is much higher than configurations which are lower in energy, since the
229 number of incorrect ways to add/remove atoms is much larger than the correct ways.
230 Thus we have implemented methods for biasing the atomic addition positions and
231 the atomic removals toward configurations which are more likely to be accepted.

232 GCMC biasing

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

$$P_{i,j,k} = \frac{xyz}{abc} \quad (1.21)$$

243 where x, y, z and a, b, c are the resolutions and cell side lengths in the cardinal di-
244 rections, respectively. While this does help to limit the total probability space it
245 does not tell us which voxels are likely to lead to better configurations, leading to

246 many rejected atomic additions. To combat this issue we can weigh the individual
247 voxels, giving more probability to voxels which show promise and less to those with
248 less likelihood to be accepted.

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

$$P_{i,j,k} = \frac{e^{\beta\Delta U_{i,j,k}}}{\sum_{i,j,k} e^{\beta\Delta U_{i,j,k}}} \quad (1.22)$$

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

257 Include figure which shows the configurational biasing map

258

CHAPTER 2

259

ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION

261 2.1 THEORY

262 To properly understand the PDF and its limitations we need to derive its mathemat-
263 ics. The PDF has been previously derived many times so it is not re-derived here.
264 This discussion of the PDF and its gradients use the notation of Farrow and Billinge.
265 [7]

266 **Derivation**

267 **Analytically Gradients**

268 Many optimization algorithms and simulations methodologies, including HMC, re-
269 quire not only the potential energy of a given configuration but also the forces acting
270 on that configuration. These forces are described by the gradient of potential energy
271 of the system which in turn requires the gradient of the PDF. As previously shown the
272 PDF is the Fourier Transform of the Debye equation. Since the Fourier Transform is
273 expressed as an integral we can exchange the order of the gradient and the integral,
274 allowing us to calculate the analytical gradient of the Debye equation and FFT the
275 resulting function. The Debye equation, with a Debye-Waller vibrational correction
276 is

277 cite something with debye waller most likely simon's paper

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

278 where

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

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

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

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

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

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

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

282 where

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

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

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

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

283 The derivatives are as follows:

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

284 where

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

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

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

286 Thus

$$\frac{\partial}{\partial u_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \frac{\partial \tau_{ij}}{\partial u_{i,w}} \Omega_{ij} \frac{\partial \tau_{ij}}{\partial u_{i,w}} = -\frac{\sigma_{ij} Q^2 \tau_{ij}}{r_{ij}} (q_{i,w} - q_{j,w}) \quad (2.15)$$

287 **Without ADPs**

288 Without ADPs the equations simplify down to

$$F(Q) = \frac{1}{N\langle f \rangle^2} \sum_{j \neq i} f_i^*(Q) f_j(Q) \frac{\sin(Qr_{ij})}{r_{ij}} \quad (2.16)$$

289 and

$$\frac{\partial}{\partial q_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \frac{\partial \Omega_{ij}}{\partial q_{i,w}} \quad (2.17)$$

290 use of these equations, when ADPs are not appropriate (like at cyrogenic tempera-
291 tures), greatly speeds up the computaiton.

292 **Periodic Boundary Conditions**

293 Periodic boundary conditions can be helpful when simulating extended solids or large

294 nanoparticles. In this case all the non-crystallinity is contained within the simulation

295 box and the box is repeated to create the longer distance peaks observed in the PDF.

296 To perform this we can break up the Debye equation into two main parts, the part

297 that describes the interatomic distances within the simulation box and those between

298 boxes. Neglecting the thermal motion portion:

$$F(Q) = \frac{1}{N\langle f \rangle^2} \left(\sum_{j \neq i} f_i^*(Q) f_j(Q) \frac{\sin(Qr_{ij})}{r_{ij}} + \sum_{i,j} f_i^*(Q) f_j(Q) \frac{\sin(QR_{ij})}{R_{ij}} \right) \quad (2.18)$$

299 where

$$R = |\vec{r} + \vec{u}| \quad (2.19)$$

$$\vec{u} = \gamma_1 * \vec{a} + \gamma_2 * \vec{b} + \gamma_3 * \vec{c} \quad (2.20)$$

300

301 Need to include PBC ADP math
302 Also should include PBC gradients, although they are trivial, maybe?
303 How does this compare against the Ewald simulation technique for ionic solutions

304 **2.2 COMPUTATION**

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

309 **HPC and GPUs**

310 To properly solve the structure of materials the PDF will need to be computed many
311 times and checked against experimental results. This requires computation of the
312 PDF, potentially over many atoms. Calculating these PDFs requires a fast, highly
313 parallelized, computational framework.

314 **GPUs and Parallelization**

315 use the figure from the presentation, it is better
316 I wish we had a picture of the GPU profiling

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

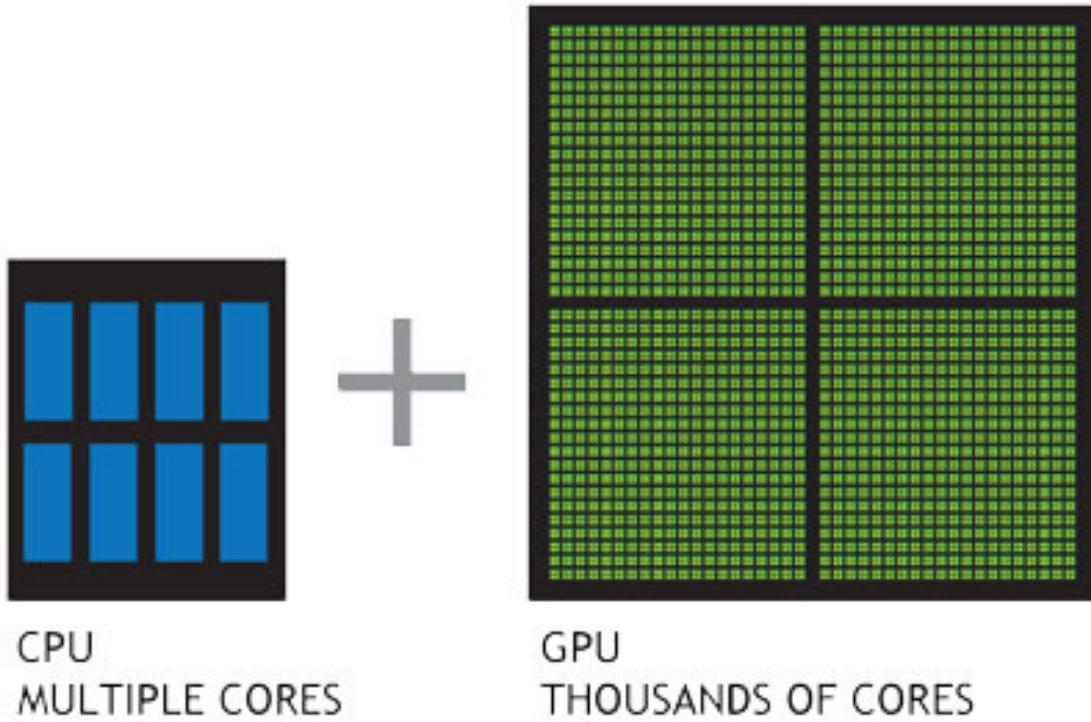


Figure 2.1: Comparison of the CPU and GPU chip architectures

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

327 **Map from ij space to k space**

328 The above equations, although formally correct, are very inefficient. $F(Q)$ and its
 329 gradient are indexed over all the atoms twice, however there are symmetries that
 330 allow us to only compute over the atom pairs essentially mapping from an $n \times n$ space,
 331 ij space, to a $\frac{n(n-1)}{2}$ space, k space. For $F(Q)$ we apply the following mapping where
 332 E denotes the atomic coordinates in ij space, E' denotes $F(Q)$ before the summation
 333 in ij space, B denotes the atomic pairs in k space, B' denotes $F(Q)$ in k space, and
 334 Z denotes the final summed $F(Q)$. For the operators, ϕ denotes the mapping from
 335 ij space to k space $k = j + i * \frac{i-1}{2}$, ψ and ψ' denote the $F(Q)$ operation in ij and k

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

336 space, respectivly. Σ denotes the sum over all the atoms.

337 To properly define Σ' we must establish whether $F(Q)$ is an even function. We
 338 can accomplish this by examining each of the portions of $F(Q)$, $\alpha, \beta, \tau, \Omega$. Ω is even,
 339 since r_{ij} is the interatomic distance, which is the same despite a flip of indicies, Q
 340 does not depend on the atomic indicies, and since Qr_{ij} is even so is $\sin Qr_{ij}$. Thus,
 341 Ω is even. Providing similar analysis to τ we can see that while \vec{u}_{ij} is odd, so is
 342 the unit displacement vector between the two atoms, thus the two odds cancel out.
 343 Intuitivly this makes sense, since the $F(Q)$ equation is fundamentally interested in the
 344 interatomic distances which is even. Thus, switching atom indicies does not change
 345 $F(Q)$. Due to the even nature of the $F(Q)$ operator the Σ' operator sums over all the
 346 atom pairs, and multiplies by two to reflect the double counting of the Σ operator.

For the gradient a similar mapping is used:

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

347

348 In this mapping, however, we use the $\tilde{\phi}\Sigma$ operator. This operator simultaniously
 349 performs a reverse mapping from k to ij space, and a summation with the correct
 350 symmetry. In this case the ψ and ψ' operators, which denote the $\vec{\nabla}F(Q)$ operator
 351 in ij and k space, are antisymmetric. Intuitivly this makes sense as an extension of
 352 Newton's Second Law, since each particle's interation is felt oppositely by its partner.

353 GPU Memory Allocation

354 While GPUs are very fast computational engines they tend to be memory bound.
355 While a gradient array for a 10nm Au nanoparticle, consisting of 31,000 atoms and
356 half a billion unique distances, occupies 1.5 TB of memory a single GPU's RAM
357 allotment varies from 4GB on a NVIDIA GTX970 to 24 GB on a NVIDIA Tesla K80.
358 Thus, it is important to determine exactly how many atoms can fit on a GPU of
359 arbitrary size as a function of the number of atoms and the Q range. The memory
360 required per array is:

$$q[=]3n \quad (2.21)$$

$$d[=]3k \quad (2.22)$$

$$r[=]k \quad (2.23)$$

$$scatter[=]nQ \quad (2.24)$$

$$normalization[=]kQ \quad (2.25)$$

$$\Omega[=]kQ \quad (2.26)$$

$$F_k(Q)[=]kQ \quad (2.27)$$

$$Sum[=]kQ \quad (2.28)$$

$$Sum2[=]kQ \quad (2.29)$$

$$F(Q)[=]Q \quad (2.30)$$

361 where n is the number of atoms, k is the number of unique distances, Q is the scatter
362 vector, and the $[=]$ operator denote the number of single precision floating point
363 values in memory. Each of the above arrays are used in the computation and thus
364 must be able to be held in memory. Thus the number of atom pairs that can fit on
365 a GPU with am bytes of available memory is:

$$k_{perGPU} = \frac{1}{16Q + 16} (-4Qn - 4Q + am - 12n) \quad (2.31)$$

366 If ADPs are included in the calculation, then the following arrays are also added to
 367 the memory allocation:

$$adps = 3n \quad (2.32)$$

$$\sigma = k \quad (2.33)$$

$$\tau = kQ \quad (2.34)$$

368 Thus the pair allotment is:

$$k_{perGPU} = \frac{-4Qn - 4Q + am - 24n}{20Q + 20} \quad (2.35)$$

369 For the Gradient we need to calculate $F(Q)$ and its gradient, so the total memory
 370 overhead is equal to the previously mentioned arrays plus:

$$\vec{\nabla}\Omega = 3kQ \quad (2.36)$$

$$\vec{\nabla}F_k(Q) = 3kQ \quad (2.37)$$

$$\vec{\nabla}F_n(Q) = 3nQ \quad (2.38)$$

371 Thus the gradient allotment is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 12n}{32Q + 16} \quad (2.39)$$

372 For the gradient with ADPs the ADP gradient array is:

$$\vec{\nabla}\tau = 3kQ \quad (2.40)$$

373 Thus the allocation is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 24n}{48Q + 20} \quad (2.41)$$

374 These equations were solved by sympy as their validity is very important to the overall
 375 reliability of the software. If the GPU is overallocated then the system may crash or
 376 return meaningless results.

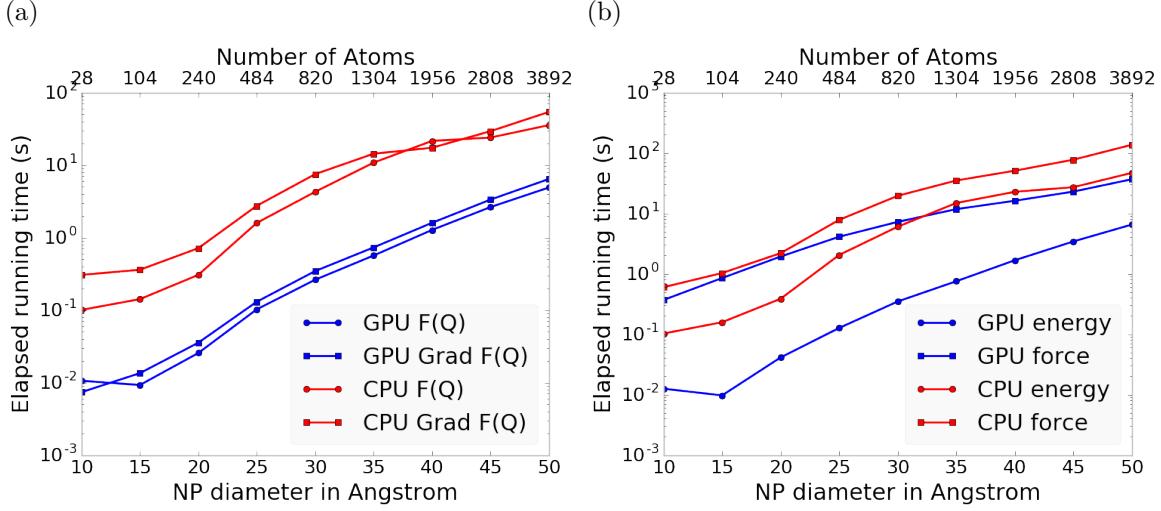


Figure 2.2: Speed comparison of CPU and GPU implementations. a) shows the time to compute the $F(Q)$ by itself. b) shows the time to compute the Rw based energy for Au NPs of various sizes, which includes computing $F(Q)$, its FFT, and the Rw .

377 Speed and Scaling of PDF Computation

378 To understand exactly how much the GPUs speed up the computation of $F(Q)$ and
 379 the PDF a series of time studies were run Au nanoparticles of varying size. Figure 2.2
 380 shows the results of these time studies. CPU and GPU calculations were carried out
 381 on an Intel i7-4820K @3.70GHz Quad-Core and one Nvidia GTX970s, respectively.
 382 The $F(Q)$ computations show a 100x to 10x speedup using the GPUs over the CPUs.
 383 Additionally, the $\vec{\nabla}F(Q)$ and $F(Q)$ computations seem to have similar computation
 384 time and scaling relationships on the GPU. This implies that the two processes
 385 may have similar bottlenecks, most likely in the $F(Q)$ computation workflow. This
 386 relationship is similarly preserved, although to a lesser extent, in the CPU scaling.

387 Interestingly, the tight runtime relationship between $F(Q)$ and its gradient are
 388 not preserved in the Rw based force calculations. While the energy calculations are
 389 very similar to the $F(Q)$ calculations in terms of runtime, the GPU and CPU force
 390 calculations are much closer, with the GPU calculations being much slower. This is
 391 due to the force bottleneck being the $3n$ FFT operations which must be performed

392 on the $\vec{\nabla}F(Q)$ array to produce the $\vec{\nabla}PDF$ array. While the GPU is leveraged
393 to perform the FFT, the data must be loaded off the GPU and back on, causing a
394 potential slowdown. Larger systems of atoms were not tried as the CPU computation
395 quickly becomes very slow. Even higher GPU speedup is expected on more advanced
396 GPUs like the Nvidia Tesla series.

397

CHAPTER 3

398

BENCHMARKING

399

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

400

Also some introduction would be great

401

this just needs a lot of work

402

3.1 INTRODUCTION

403

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

404

The purpose of these benchmarks is to test the ability of the NUTS-HMC system to

405

reproduce the target PDF and its associated structure. Systems were chosen for their

406

size, crystallinity, and interfacial differences.

407

3.2 PDF

408

The formation of NPs with both crystallographic and non-crystallographic structures

409

[14] and with different chemical patterns [8] are well documented. For simplicity,

410

we chose monometallic Au clusters as benchmarks and considered two groups of

411

structures with different size and degrees of structural disorder in order to assess

412

the reliability and efficiency of our HMC method for solving atomic structures from

413

PDFs. The first group consists of Au_{55} clusters with different degrees of disorder,

414

including a crystalline cluster structure in O_h (Octahedral) symmetry, a structure

415

with a disordered surface, and an amorphous structure. The second group consists

416

of the crystallographically solved Au_{102} structure as in the $\text{Au}_{102}\text{MBA}_{44}$ nanocrystals

417

[10, 13]. We used optimized structures from the Density Functional Theory (DFT)

418 as target structures and generated the corresponding PDF, G_{obs} , according to

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

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

424 Spin-polarized DFT calculations were carried out using the Vienna ab initio simu-
425 lation package (VASP) [11, 12] within the Perdew-Burke-Ernzerhof (PBE) exchange-
426 correlation functional [18]. The projected augmented wave method [1] and a kinetic
427 energy cutoff of 400 eV were used. Structural optimization was performed until the
428 total energy and ionic forces were converged to 10^{-6} eV and 10 meV/Å, respectively.
429 The amorphous Au_{55} structures were generated by simulated annealing using the
430 classical embedded atom method potential [23]. Different annealing temperatures
431 between 1200 K and 1670 K (bulk melting temperature of Au) were used and the
432 thermally equilibrated structures were cooled down to 300 K before minimization at
433 0 K. Further optimization using DFT leads to total energies that vary within 1-2
434 eV among different amorphous structures and the lowest energy one was used as the
435 target structure. The target structure of Au_{102} was taken as the Au_{102} core of the
436 DFT-optimized $\text{Au}_{102}\text{MBA}_{44}$ cluster [13].

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

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

440 The thresholded spring potentials are based on those previously proposed on by Pe-
441 terson [19], 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

442 bounds of the spring threshold r_t . The resulting restoring forces on the out-of-bound
443 atoms bring the system back within the bounds of the PDF, R_{\min} and R_{\max} , and
444 therefore preventing the system from exploding or collapsing. Otherwise, incorrect
445 refinements may result by having atomic pair distances out of the PDF bounds. κ is
446 the spring constant in eV/Å and the Rw potential is converted from unitless to eV
447 via multiplication by a conversion factor λ .

448 Whereas the choice of the absolute values of λ and κ is somewhat arbitrary, their
449 relative values are important in determining which term in Eq. (3.2) dominates the
450 PES, especially when considering the effect of the simulation temperature. Generally,
451 the ratio between the total potential energy and the temperature determines how
452 much random motion will dominate the dynamics; a lower ratio implies that random
453 motion will play a large role in the dynamics. The ratio between λ and κ of each
454 spring describes how far the PDF can push the system below or above the bounds set
455 by the spring potentials. Heuristically, too stiff a spring forbids the system to access
456 new configurations, e.g. high energy “transition states” which may involve shorter
457 bonds or a larger system size. Conversely, too small a spring constant makes it slower
458 for the system to snap back within bounds and may lead to an explosion or implosion
459 of the system, leaving the dynamics to drift aimlessly.

460 Model Parameters

461 Unless otherwise stated, the PDFs of the target and starting structures were generated
462 using Eqn. (3.1) with a step of $\delta R = .01 \text{ \AA}$, $Q_{\min} = 0.1 \text{ \AA}^{-1}$, $Q_{\max} = 25.0 \text{ \AA}^{-1}$. R_{\min}
463 and R_{\max} correspond to the first minimum before the first PDF peak and that after
464 the last PDF peak, respectively, which ensure that the full meaningful region of the
465 PDF is modeled. For each of the simulations, the Q resolution was calculated by

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

466 The HMC simulation was run with $N = 300$ iterations, a target acceptance rate
467 of 0.65, and an average starting momentum for each NUTS iteration of 10 eVfs/Å.
468 Both repulsive and attractive spring potentials are used with $\kappa = 200$ eV/Å and
469 thresholds matching R_{\max} and R_{\min} of the PDF, respectively. $\lambda = 300$ eV was used
470 as conversion factor for Rw . Each simulation was run with a pair of Nvidia GTX970
471 graphics cards, with one card partially occupied with desktop visualization.

472 **Au55: surface relaxed**

473 We first test our algorithm by solving the crystalline Au₅₅ (*c*-Au₅₅) cluster structure
474 from its PDF. The starting structure is taken as the bulk-cut cuboctahedron of Au₅₅
475 with a uniform bond length of 2.89 Å. Due to finite-size and surface effects, the DFT-
476 relaxed cluster structure shows a distinctively different bond length distribution as a
477 function of the bond's distance to the cluster center of mass, and therefore is difficult
478 to model with a small box approach which assumes an identical unit cell throughout
479 the whole system.

480 **Run Parameters**

481 R_{\min} and R_{\max} for this simulation were 2.45 Å and 11.4 Å, respectively, with $\delta Q =$
482 0.24 Å⁻¹. The simulation ran for approximately 34 minutes, over a total of ~40
483 thousand configurations. The results are shown in Fig. 3.1.

484 The PDF, radial bond distribution, and bond angle distribution show good agree-
485 ment between the target and final fitted structures, with a Rw of 0.3% whereas Rw
486 of the starting structure is as high as 44.8%. DFT calculations yield a total energy of
487 the final structure very close to that of the target structure (within a few meV). The
488 success in the fitting is largely attributed to the factor that the target structure is
489 only locally (and mildly) disturbed from its bulk-like counterpart and therefore there
490 is no need to overcome any high PES barriers to reach the correct solution. As shown

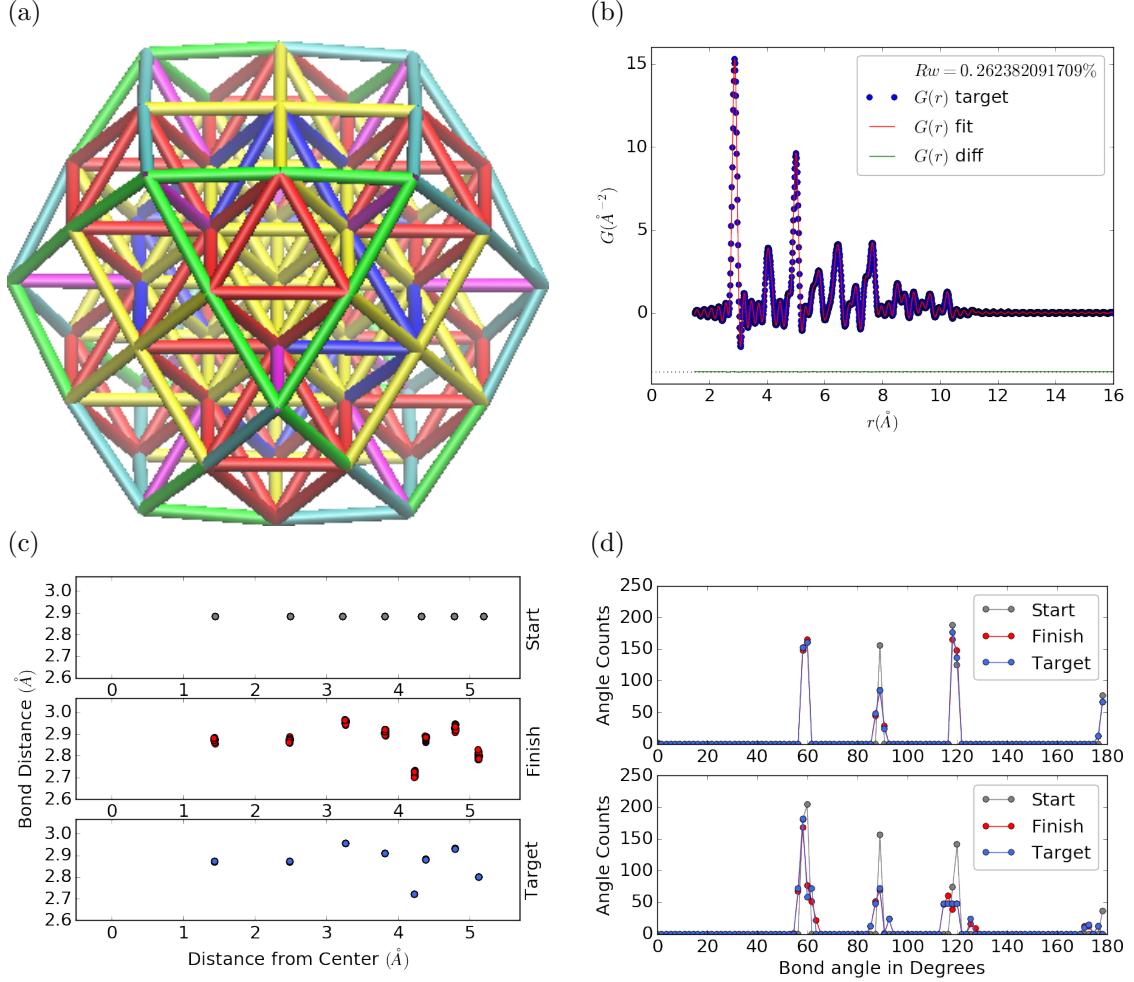


Figure 3.1: Au_{55} PDF fitting of DFT-optimized $c\text{-Au}_{55}$. (a) the final structural solution ($Rw=0.3\%$) with bond lengths color-coded by step of 0.05\AA , (b) the target PDF(blue dots) overlaid with the PDF of the final structure (solid red lines) with the difference in green lines offset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

491 below, the situation is rather different for much more disordered target structures.
 492 Interestingly, the small-box solution using PDFgui[6] yields a rather large Rw of 43%,
 493 due to the failure to fit the surface contracted atoms with a unit cell. The PDF fits
 494 of the starting structure and small-box solution are shown

495 Put this somewhere

496 .

497 **Au55: surface disordered**

498 In addition to surface relaxation, the structure of a cluster or nanoparticle is often
499 disrupted by the presence of defects and/or ligand bound to the surface. To mimic
500 such surface disorders, we took the DFT-optimized *c*-Au₅₅ structure from case I as
501 the starting structure and randomly displaced the surface atoms with a normal distri-
502 bution of $\sigma = 0.2 \text{ \AA}$. All atoms are allowed to move in the HMC simulation, including
503 the originally undisturbed core, which is a Au₁₃ cluster with O_h symmetry.

504 R_{\min} and R_{\max} for this simulation were 1.95 Å and 12.18 Å, respectively, with
505 $\delta Q = 0.23 \text{ \AA}^{-1}$. The simulation ran for approximately 3.6 hours, over a total of ~ 270
506 thousand configurations. The results of the simulation are shown in Fig. ??.

507 Overall, good agreement is found between PDFs of the target structure and the
508 final structural solution, even out to larger r , with an $Rw = 0.6\%$ starting from an
509 $Rw = 50.4\%$ (see Fig. S2). The radial bond distribution and angle distribution
510 show reasonably good agreement, but with lower degree of crystallinity in the final
511 structure compared to the target structure. The discrepancy is most obvious in
512 the core: despite the identical core structure in the starting and target structures,
513 the core atoms were displaced in the HMC simulations in order to achieve a “best”
514 solution. This is because PDF measures the global average of interatomic distances
515 between each atomic pair and does not contain direct information about the locality
516 of such pairs, e.g. on the surface or cores. If such information is obtained a priori, for
517 example, from theoretical prediction or other experimental measurements, the core
518 structure can then be fixed and excluded from HMC dynamics.

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

524 and high CN numbers. This is partly caused by the displacement of the core atoms
525 mentioned above, and partly by the lack of CN constraints for the PDF fitting, which
526 has been previously demonstrated in the case of α -Si [3]. Additional experimental
527 data, e.g. from EXAFS or NMR, may help to steer the simulations towards better
528 agreement in both PDF and CN distribution.

529 Au55: amorphous

530 Next, we turn to the case in which the entire cluster structure is disordered. We used
531 a DFT-optimized amorphous Au_{55} ($a\text{-Au}_{55}$) as the target structure, and the DFT-
532 relaxed $c\text{-Au}_{55}$ cluster from Case I as the starting structure. The total energy of
533 $a\text{-Au}_{55}$ was computed to be *lower* than that of $c\text{-Au}_{55}$ by as large as 2.9 eV, consistent
534 with the 3.0 eV found in previous DFT work [5].

535 R_{\min} and R_{\max} for this simulation were 2.6 Å and 11.26 Å, respectively, with
536 $\delta Q = 0.25 \text{ \AA}^{-1}$. The simulation ran for approximately an hour, over a total of ~ 87
537 thousand configurations. The results of the simulation are shown in Fig. ??.

538 Our PDF fitting yielded a final structure of Rw of 1.7%, whereas that of the
539 initial structure is as high as 76.1% (see Fig. S3), due to the drastically different
540 atomic structure of the crystalline and amorphous Au_{55} clusters. Overall reasonable
541 agreement in PDF, bond angle distribution, and radial bond distance distribution
542 was found, and the wide spread of the bond lengths was qualitatively reproduced.
543 However, the mismatch in CNs is problematic, partly due to the lack of information
544 and/or constraints on the CNs. The total energy of the final structure is computed to
545 be ~ 6 eV higher than that of the target structure and the difference is substantially
546 larger than the variation among different amorphous structures computed by DFT
547 ($\Delta E_{\text{tot}} \sim \pm 1\text{-}2$ eV). Such a fitting result, despite the rather small Rw , clearly
548 indicates the importance of complementary informations and/or constraints necessary
549 for reliably solving disordered NP structures from PDF.

550 **Au102: triple phase**

551 Our final benchmark is Au₁₀₂, whose structure was initially solved by Jadzinsky and
552 co workers using x-ray crystallography [10] and further confirmed by DFT studies
553 [13]. The Au₁₀₂ structure consists of three main parts, a 49-atom Marks decahedron
554 core, two C_5 caps consisting of 20 atoms each, and 13 equatorial atoms. Unlike
555 previous cases, the multi-symmetry nature of the structure, i.e. each part has its own
556 distinct symmetry, poses a challenge for PDF-based solution of the structure. This is
557 because of the atomically centralized nature of the PDF, in which each atom “sees”
558 a density of other atoms surrounding it and has a strong tendency towards becoming
559 the center of the main symmetry group. Such tendency may lead to a solution where
560 some of the correct atomic symmetries are discarded in favor of the core symmetry.

561 **Starting from fcc structure**

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

564 R_{\min} and R_{\max} for this simulation were 2.7 Å and 16. Å, respectively, with $\delta Q =$
565 0.18 Å⁻¹. The simulation ran for approximately two hours, over a total of ~82
566 thousand configurations. The results of the simulation are shown in Fig. 3.4.

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

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

586 **Marks decahedron**

587 The starting structure, a Marks Decahedron, was generated by the ASE Cluster tool
588 with 2 atoms on the [100] face normal to the 5-fold axis, 3 atoms on the [100] plane
589 parallel to the 5-fold axis and, 1 atom deep Marks reentrance. This produced a
590 structure with 101 atoms which was extended by one more Au atom to fill out the
591 Au_{102} structure.

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

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

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

605 **Au147**

606 **3.3 PDF WITH ADPs**

607 **ADP 50**

608 1. Basic 50% larger magnitude

609 2. Random addition to APDs

610 3. Janus ADPs

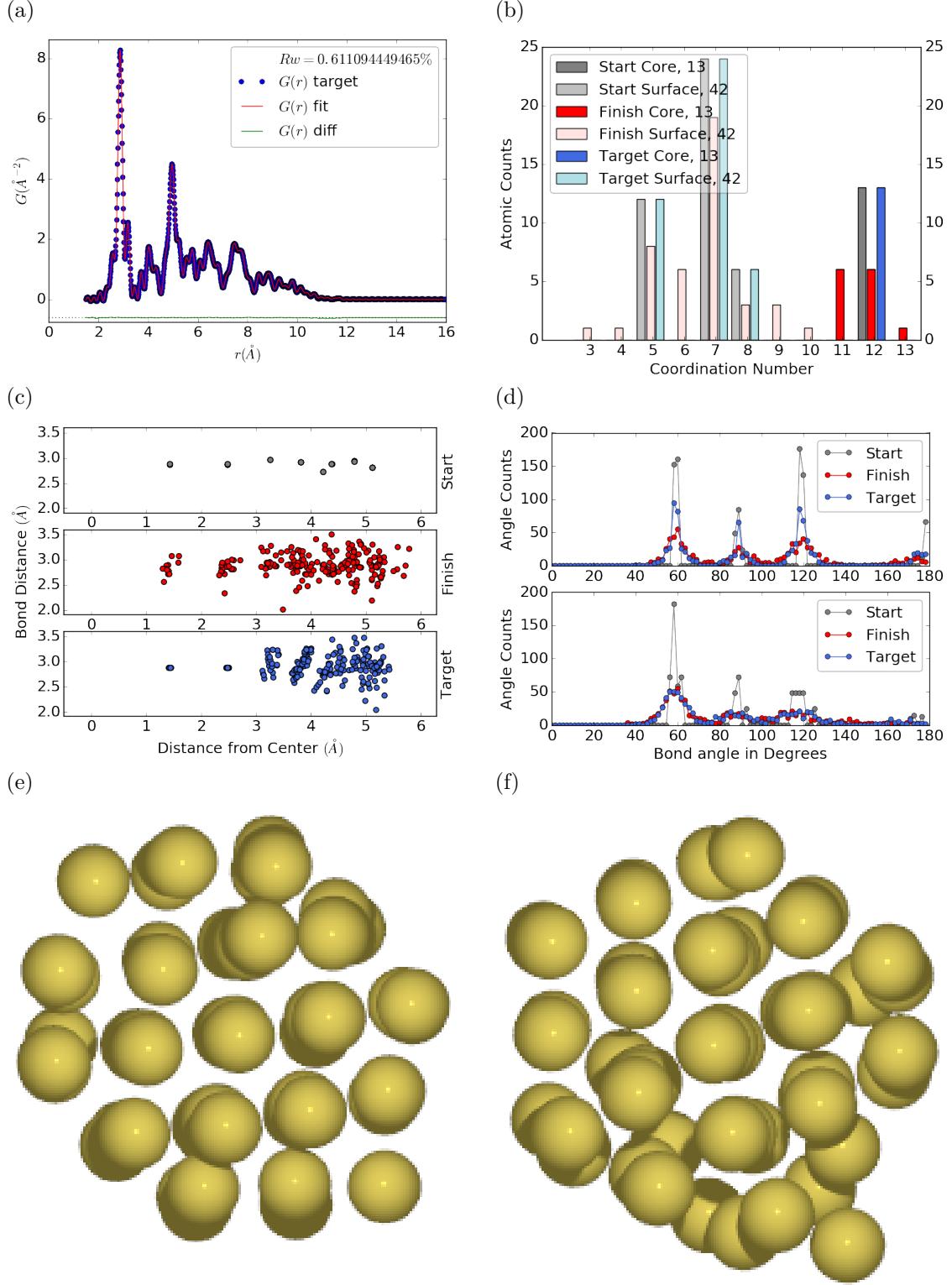


Figure 3.2: Au_{55} PDF fitting of surface-disordered Au_{55} . a) the target structure, b) the final structural solution ($R_w=0.6\%$), c) the comparison of PDFs, d) the CN distribution with the number of atoms in either the core or the surface, e) the radial bond distribution, and f) the bond angle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

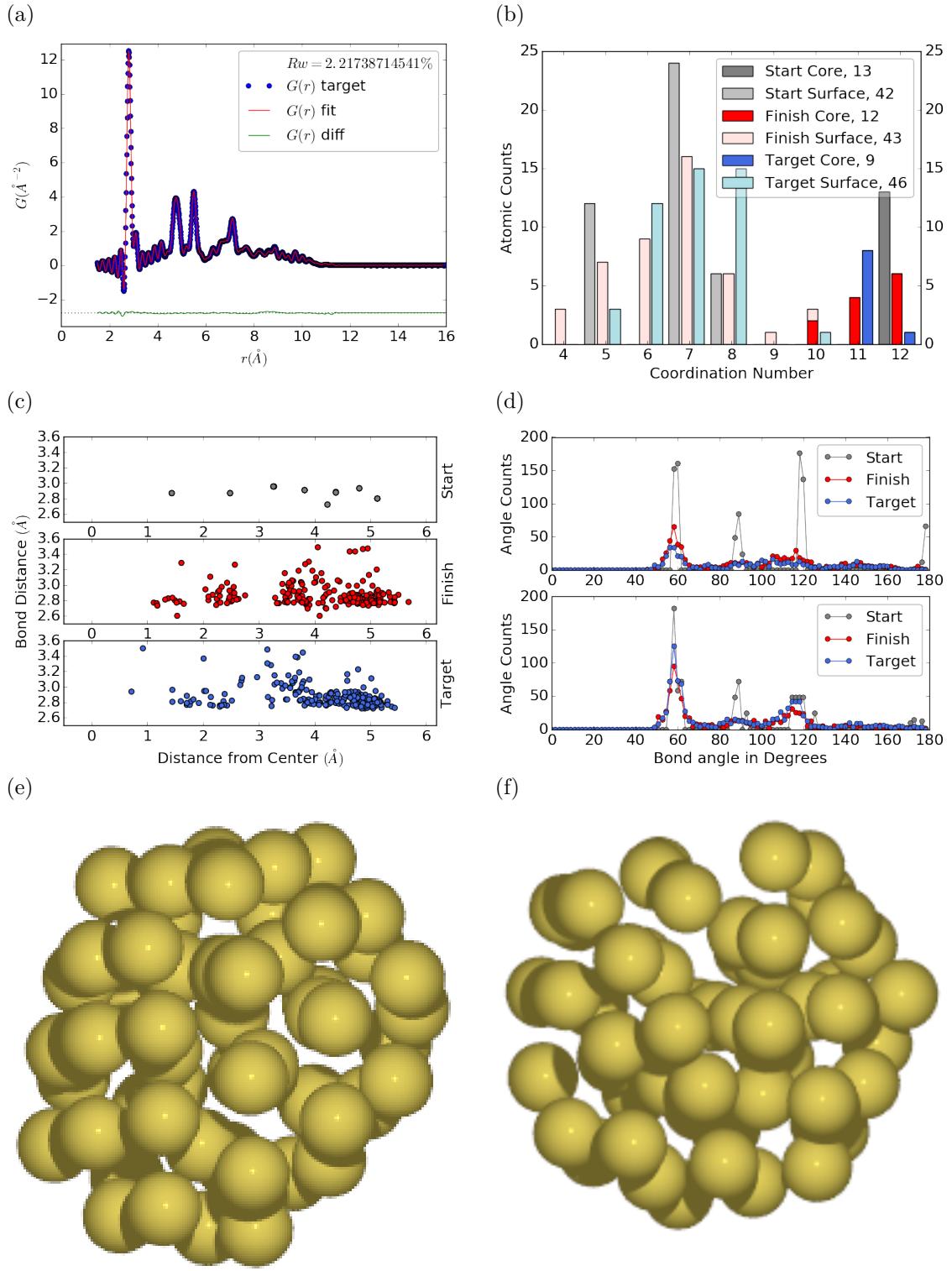


Figure 3.3: Similar to figure 3.2 for DFT-optimized amorphous Au_{55} .

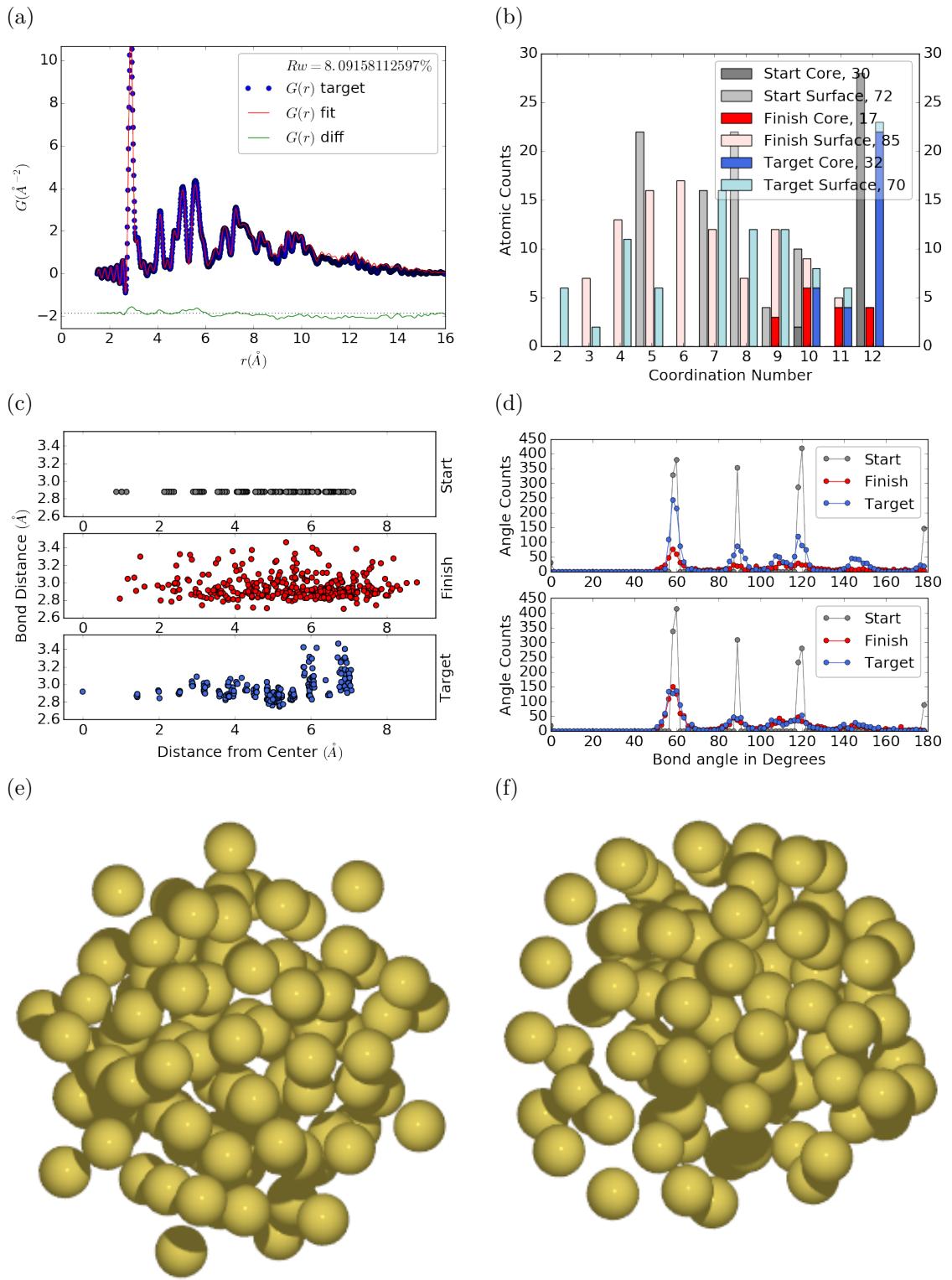


Figure 3.4: Similar to figure 3.2 for Au_{102} as in DFT-optimized $\text{Au}_{102}\text{MBA}_{44}$ cluster.

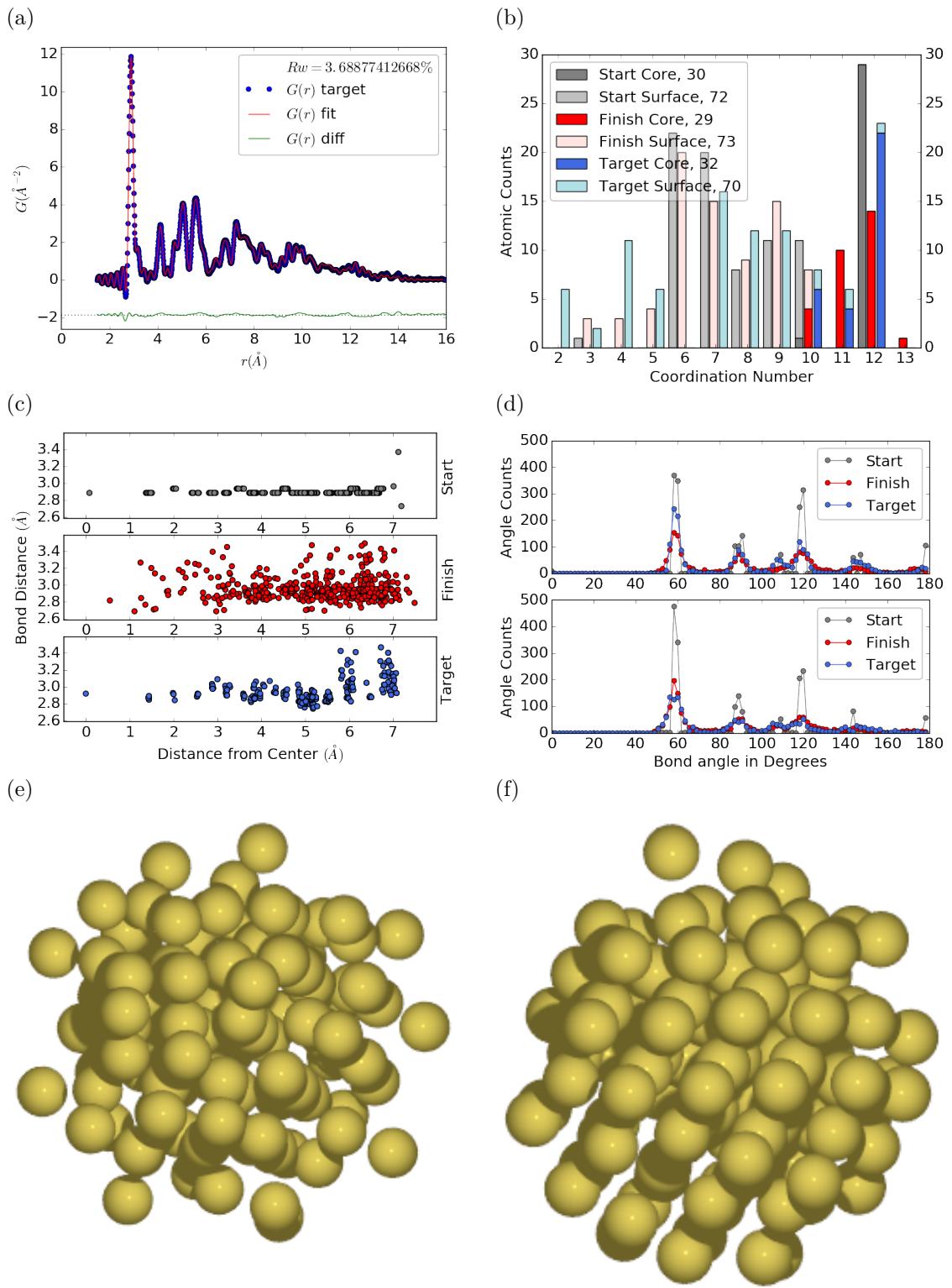


Figure 3.5: Similar to Fig. 3.4 with Marks decahedron as the starting structure.

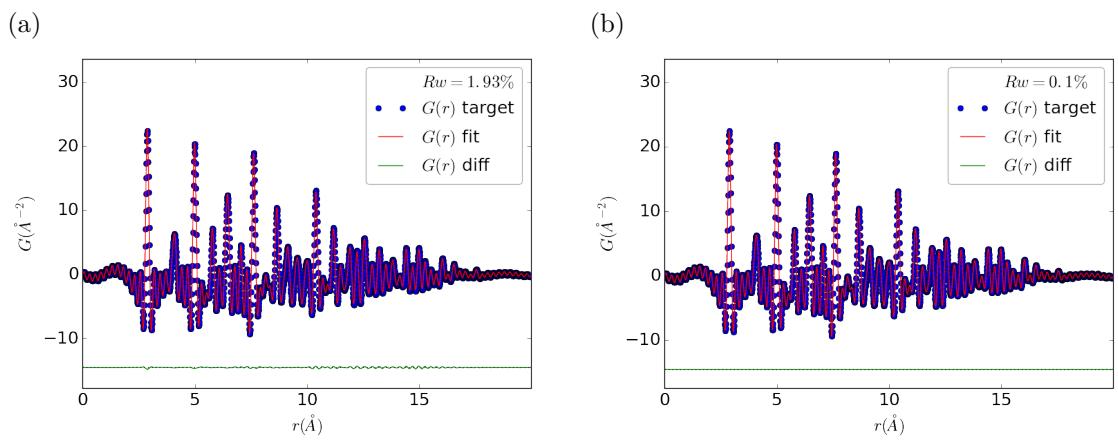


Figure 3.6: Refinement of adps

611

CHAPTER 4

612

X-RAY TOTAL SCATTERING DATA ACQUISITION AND

613

PROCESSING

614

4.1 INTRODUCTION

615 X-ray total scattering experiments are generally performed at synchrotron light sources,
616 as only these sources can provide the needed flux, energy, and high momentum trans-
617 fer vectors needed to obtain reliable PDFs. [?, ?] Despite the need for a dedicated
618 facility to perform the total scattering experiments, the experiments themselves are
619 fairly forgiving, allowing for reactive gaseous environments, experiment temperatures
620 ranging from 2 K to 1200 K, and even electrochemical cycling. [?, 21, ?] The rapid
621 PDF data acquisition associated with 2D area detectors creates a data management
622 problem, as 96 hours of beamtime could result in almost 10,000 images which need
623 to be associated with the experimental conditions and detector metadata. [?] Finally,
624 all this data needs to be processed by masking bad pixels and regions, integrating
625 azimuthally, and converting the scattering data to the PDF. [?, ?, ?, ?, ?]

626

4.2 DETECTOR Q RESOLUTION

627 To properly azimuthally integrate the images taken from the detector the Q resolution
628 of the pixels must be calculated. Integrating using even bins will cause pixels which
629 are not on the same ring to be binned together, causing the incorrect value of $I(Q)$
630 to be obtained and a larger standard deviation in the integrated data. To properly
631 calculate the Q resolution the resolution of each of the pixels in 2θ must be calculated.

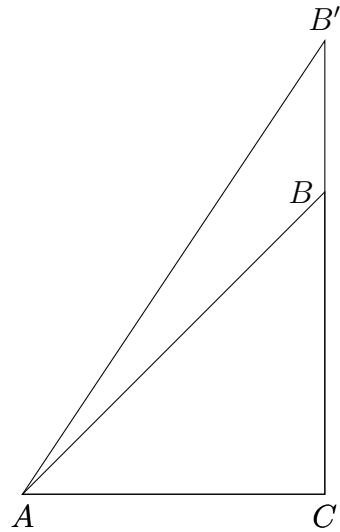


Figure 4.1: Scattering onto a flat detector

632 Figure 4.1 shows the scattering of x-rays onto a flat image plate detector. In this
 633 diagram the bottom of the n th pixel is B while the top is B' . The resolution of this
 634 pixel in 2θ is $\angle BAC - \angle B'AC$. Thus the resolution, calculated from the distances is

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

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

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

638 where λ is the x-ray wavelength.

639 For a Perkin Elmer image plate, like the one used at the NSLS-II's XPD and the
 640 APS's 11-ID-B, the resolution function is shown in 4.2. For the same detector the
 641 number of pixels per Q is shown in 4.3

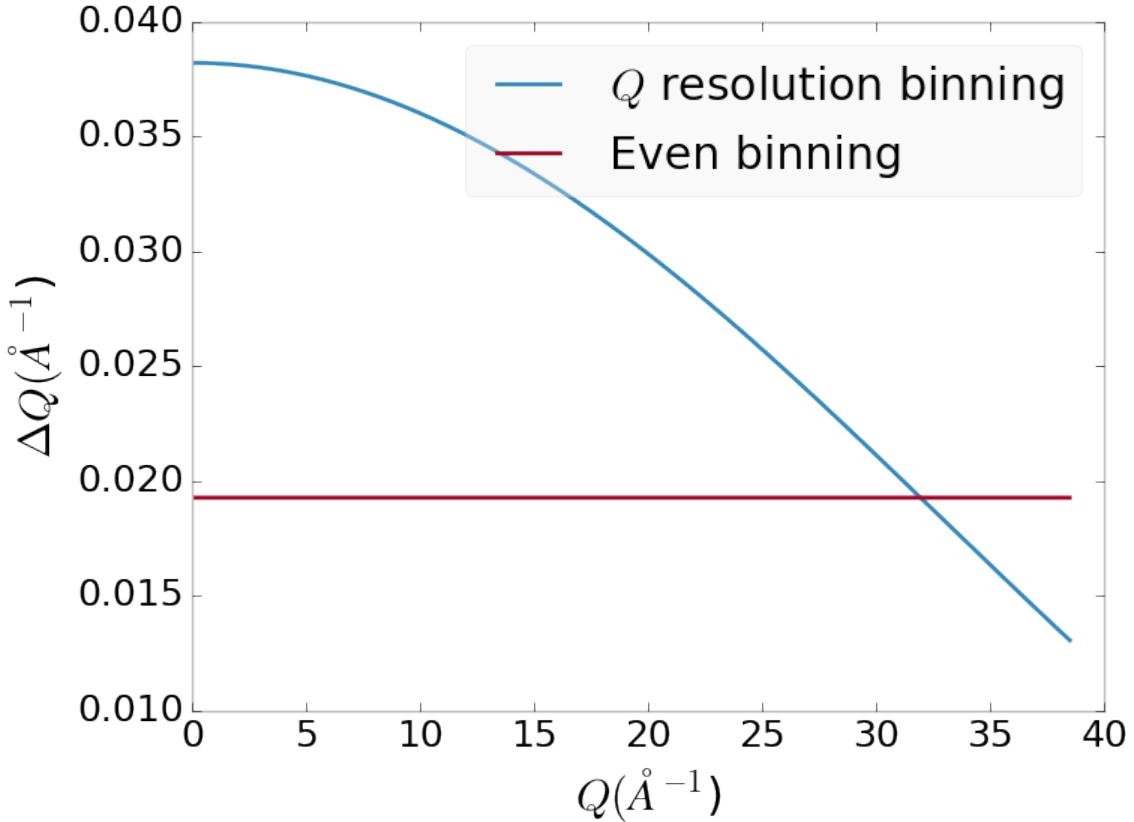


Figure 4.2: Q resolution as a function of Q .

642 4.3 AUTOMATED MASK GENERATION

643 **Introduction**

644 Detector masking is an important part of any x-ray scattering workflow as dead/hot
 645 pixels, streak errors, and beamstop associated features can be averaged into the data
 646 changing the signal and its statistical significance. While some features, like the
 647 beamstop holder, can be easily observed and masked by hand other are much more
 648 difficult to observe even on large computer monitors. Additionally, while dead/hot
 649 pixels and streaks are usually static the hot pixels associated with textured or sin-
 650 gle crystal scattering or cosmic rays are not. Thus, coming up with an automated
 651 method for finding such erroneous pixels is important, especially as high flux diffrac-
 652 tion beamlines can generate data very quickly.

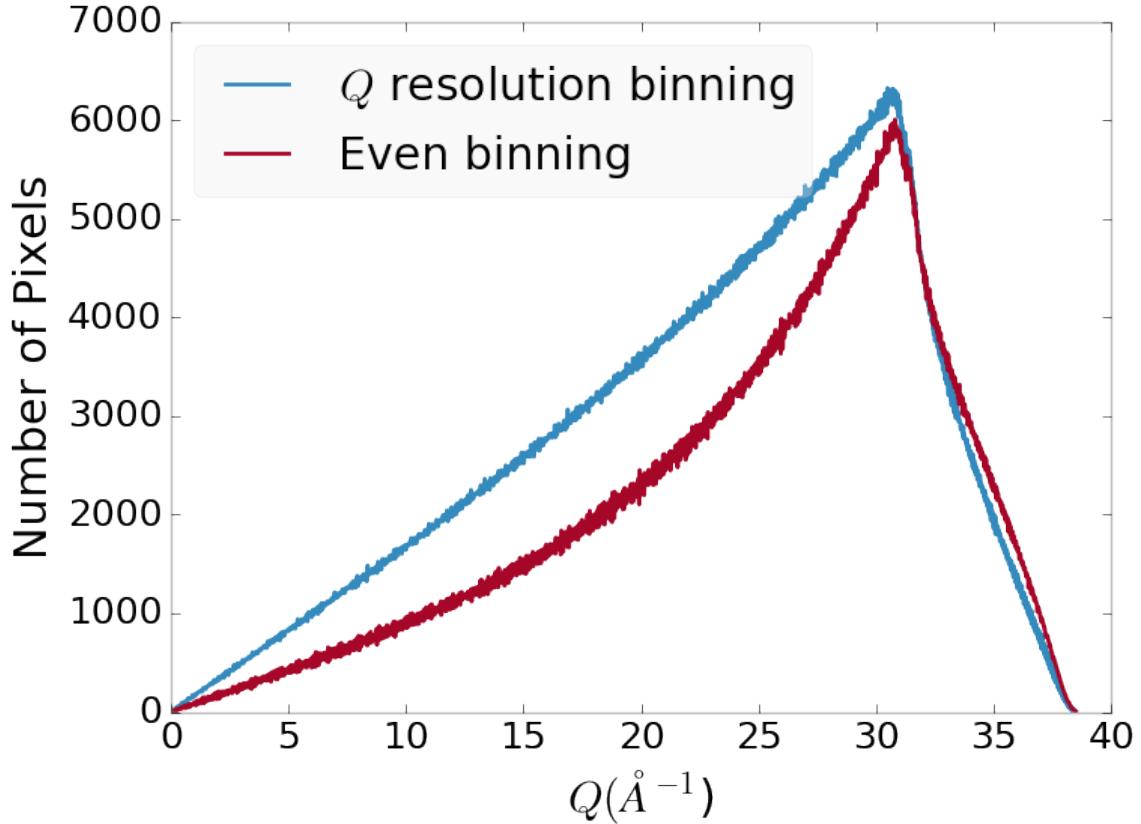


Figure 4.3: Number of pixels as a function of Q , binned at the Q resolution of the detector.

653 While this problem can be quite complex in the most general case, we can use the
 654 annular symmetry of the powder scattering pattern to our advantage, by comparing
 655 a pixel against pixels in the same ring. Since non-textured powder scattering should
 656 produce the same pixel intensity for a given ring we can mask any pixels which are α
 657 standard deviations away from the mean. This method relies on the aforementioned
 658 pixel binning algorithm, as using miss sized bins will cause some pixels which should
 659 be in separate rings to be put together, and others which should be in the same ring
 660 to be separated. In that case the masking algorithm will overestimate the number of
 661 pixels to be masked due to the additional statistical variation in the sample.

662 **Algorithm Design**

663 The masking algorithm procedure takes in the image and a description of the pixel
664 positions in either distance from the point of incidence or in Q . The image is then
665 integrated twice, producing both the mean $I(Q)$ and the standard deviation of each
666 $I(Q)$ ring. The mask is created by comparing the pixel values against each ring's
667 standard deviation and threshold α . Note that the threshold can be a function of
668 distance from the point of incidence or Q .

669 **Test Cases**

670 To study the effectiveness of the masking we ran the algorithm against both simulated
671 experimental data. In the case of the simulated data four systems were created: 1)
672 dead/hot pixels with varying numbers of defective pixels, 2) beamstop holder with
673 varying beamstop holder transmittance, 3) rotated beamstop holder with varying
674 beamstop holder transmittance, and 4) beamstop holder with dead/hot pixels. The
675 base scattering was produced by

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

676 where r is a pixel's distance from the beam point of incidence. The positions of
677 the dead/hot pixels were chosen at random as was the dead or hot nature of the
678 defect. Dead pixels had values from 0 to 10, while hot pixels had values from 200
679 to 255. The beamstop was positioned at the vertical center of the detector with an
680 initial width of 60 pixels and final width of 120 pixels. The height of the beamstop
681 was 1024 pixels. The beamstop was calculated to attenuate the x-ray scattering
682 signal at various transmittance, as various beamstop holder materials have different
683 transmittance. Two version of the masking algorithm were run for each test case, one
684 using the standard even bin sizes for the integration step, and one where the bin sizes
685 are tuned to the pixel Q resolution as discussed in 4.2.

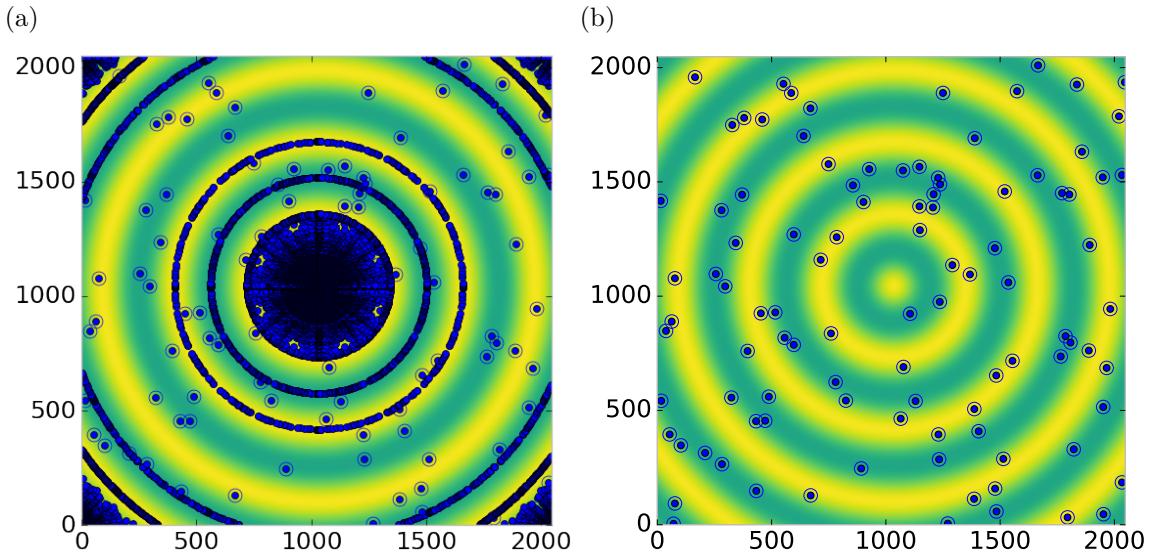


Figure 4.4: Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

686 Results and Discussion

687 Figures 4.4-4.11 show the results of the masking algorithm on simulated images. The
 688 dead/hot pixel masking shows the importance of using the Q resolution based bin
 689 sizes as the even bin based mask have a tendency to over mask the image, removing
 690 pixels which contain valuable signal. This over-masking is caused by pixels being
 691 improperly associated with one another by the even bins. Figure 4.4 indicates that
 692 the masking algorithm, with the proper binning, masks the image perfectly, with no
 693 missed bad pixels or good pixels masked. This is not the case in figures 4.5 - 4.7 as
 694 we can see pixels which should have been masked but were not. Despite these missed
 695 pixels no pixels were improperly masked in any of the well binned images. These
 696 test cases are actually more difficult than experimental data, as the dynamic range
 697 of most detector causes the dead/hot pixels and single crystal/textured peaks to be
 698 orders of magnitude away from the desired signal.

699 The beamstop holder masks shown in figures 4.8 - 4.11, which were all run with

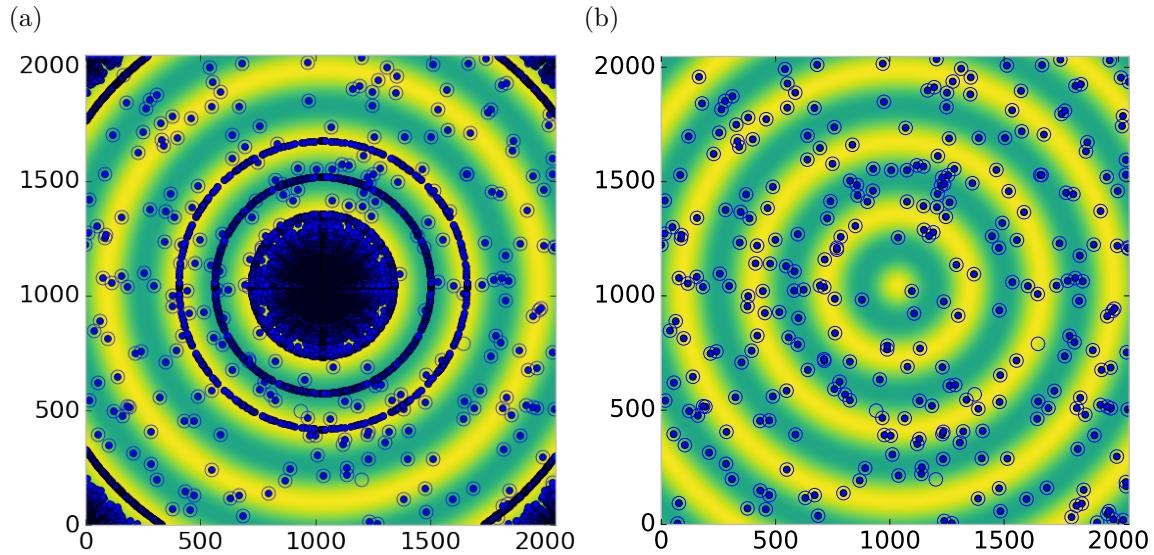


Figure 4.5: Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

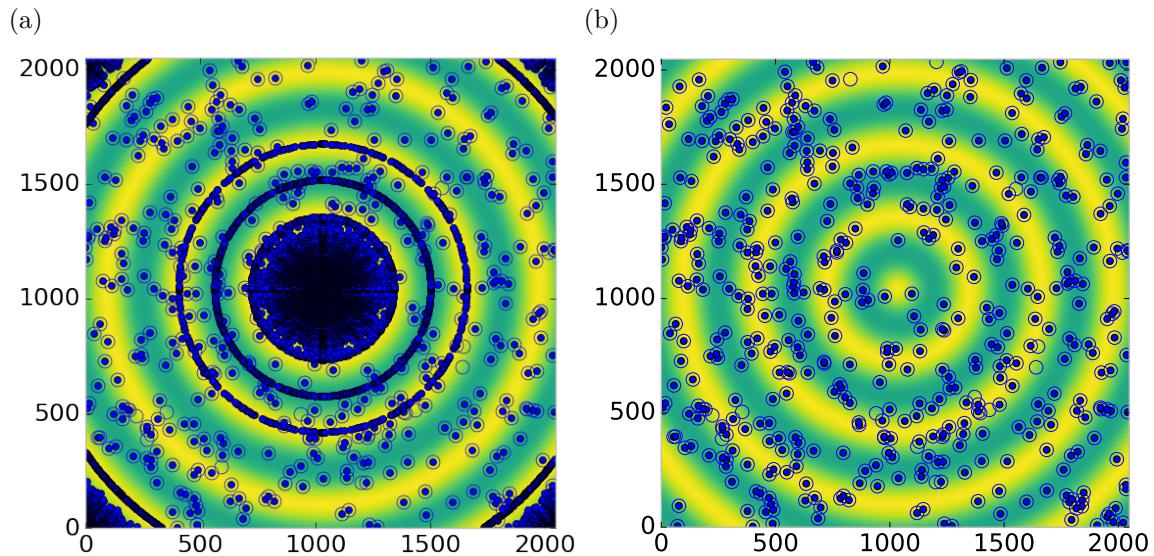


Figure 4.6: Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

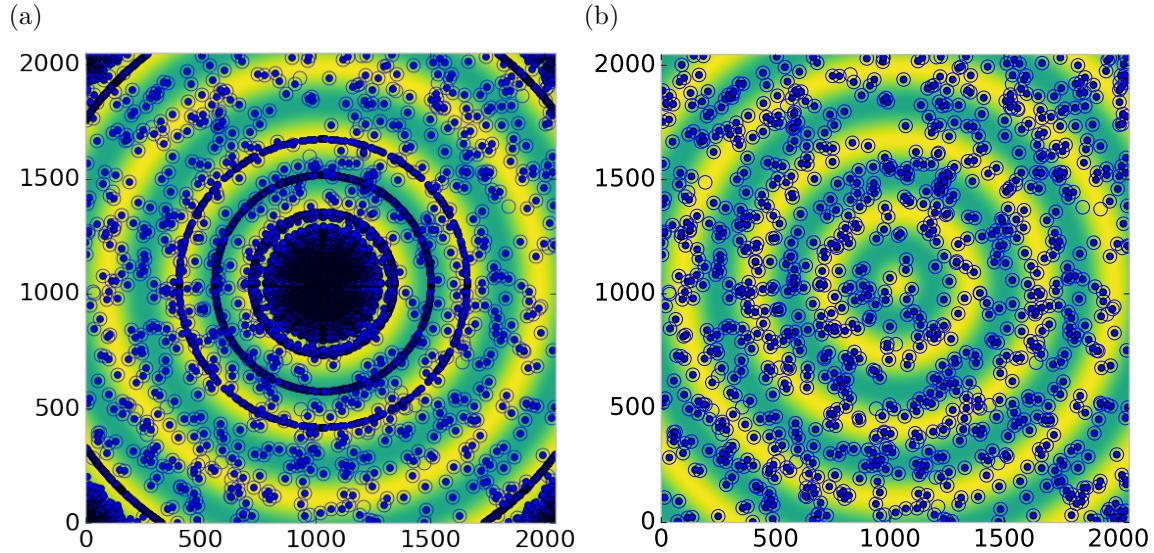


Figure 4.7: Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

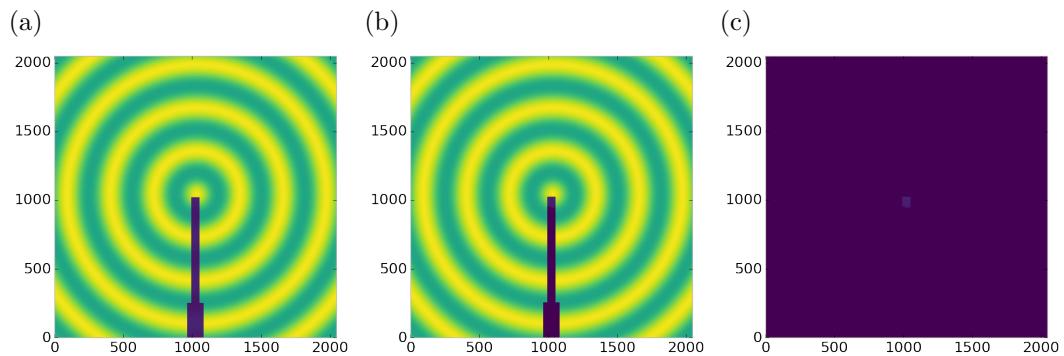


Figure 4.8: Generated beamstop holder masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

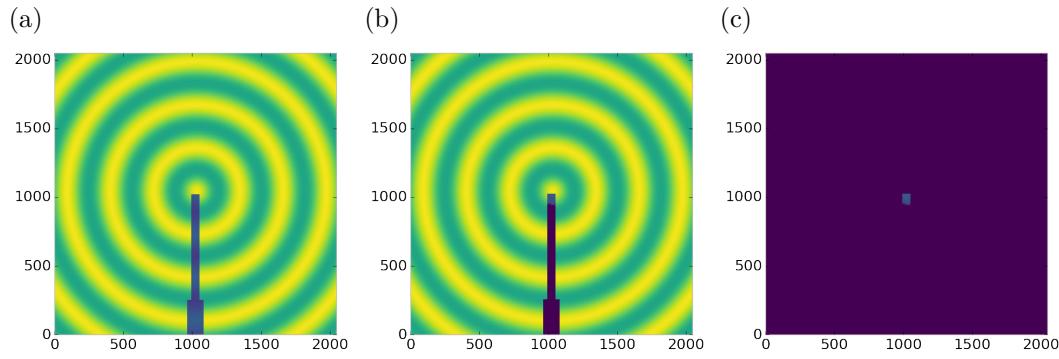


Figure 4.9: Generated beamstop holder masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

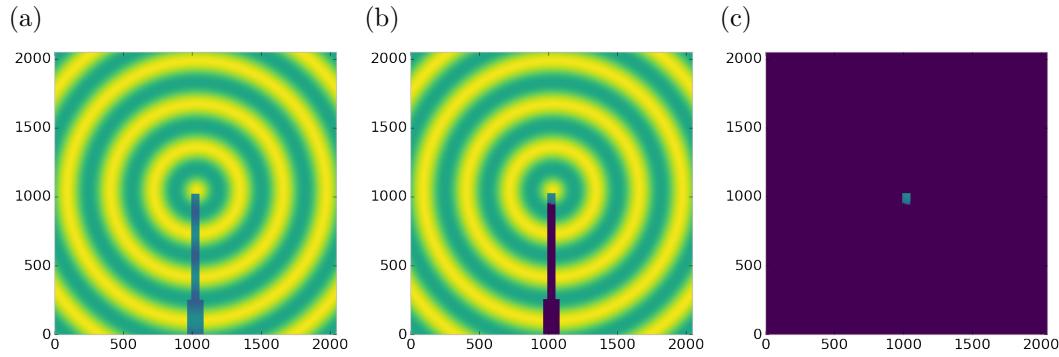


Figure 4.10: Generated beamstop holder masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

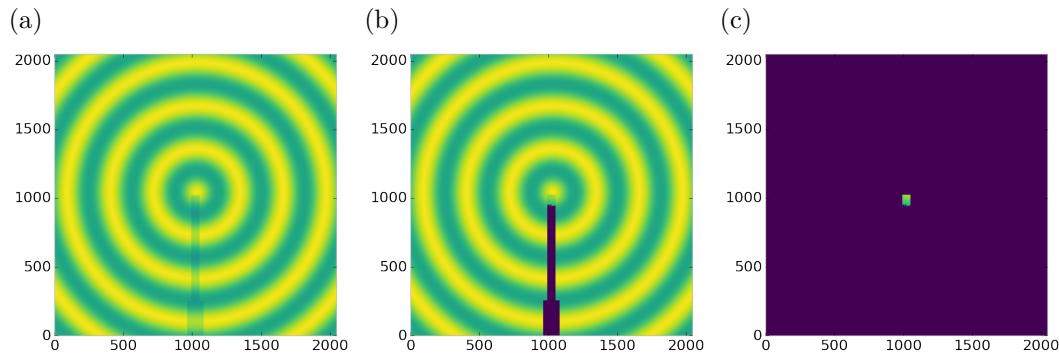


Figure 4.11: Generated beamstop holder masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

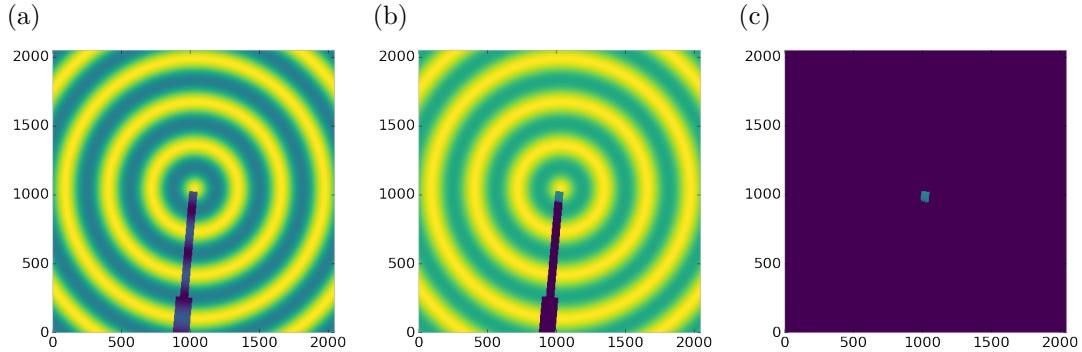


Figure 4.12: Generated beamstop holder masks which is rotated away from vertical

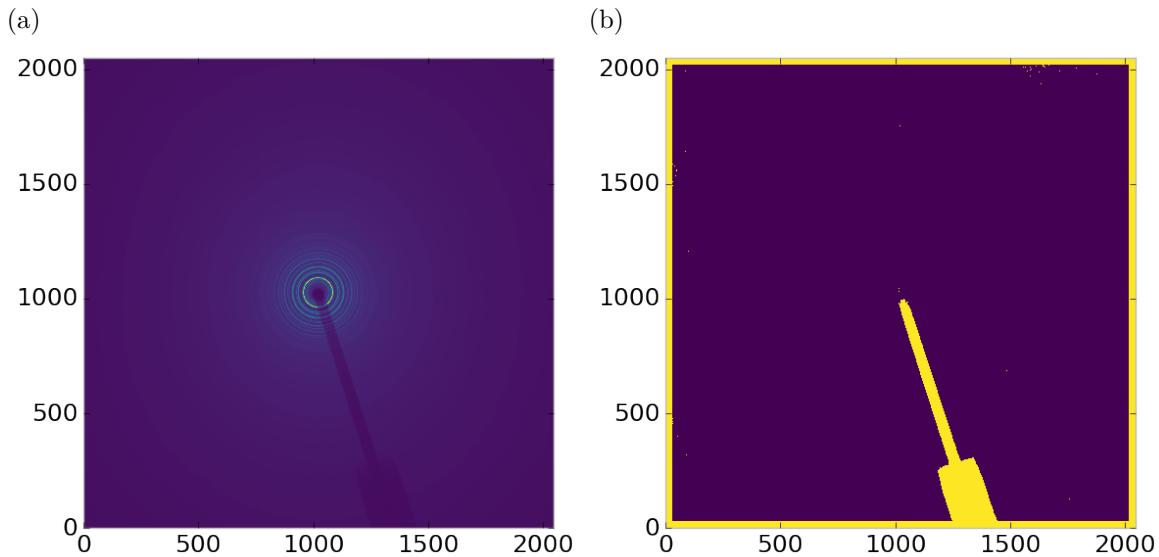


Figure 4.13: Masked experimental data. a) the raw image, b) the mask

700 the Q resolution binning show similar results across the transmittance range, missing
 701 only a small part of the beamstop holder near the point of incidence. Near this point
 702 the beamstop holder becomes a statistically significant part of the total number of
 703 pixels in a given ring, thus it can not be masked out using a statistical search of the
 704 rings. For most PDF and XRD studies this small area can be masked automatically
 705 by masking all the pixels who's distance from the point of incidence is smaller than a
 706 given radius r , or can be neglected outright as the area is not used in the analysis or
 707 refinement. Similar results were produced for beamstop holders which were rotated
 708 away from the vertical position, as shown in figure 4.12

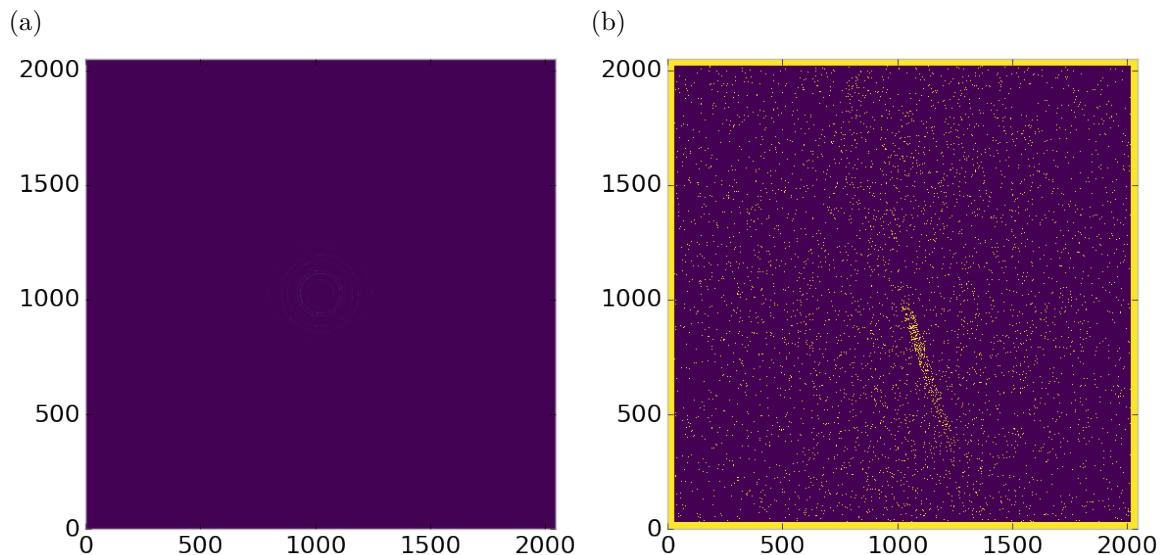


Figure 4.14: Masked experimental data with Pt single crystal signal. a) the raw image, b) the mask

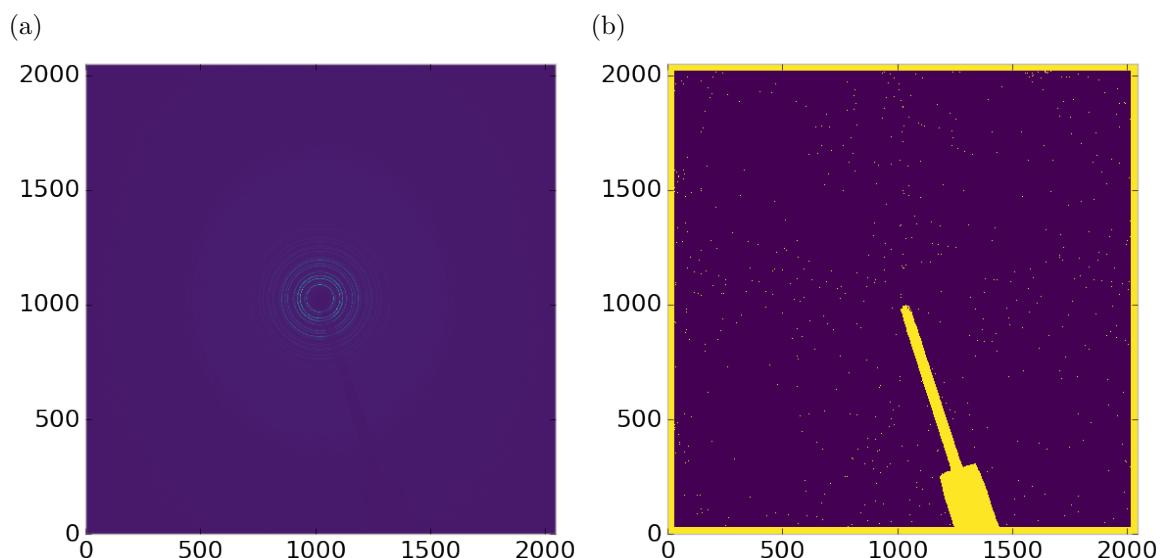


Figure 4.15: Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask. a) the raw image, b) the mask

709 Working with actual experimental data, obtained at the Advanced Photon Source
710 beamline 11-ID-B, shows the difficulty of masking images which have low photon
711 counts. While the masking of experimental data taken with longer exposures, con-
712 sisting of 250 .2 second shots, shown in figure 4.13 provides very sharp edges to the
713 beamstop holder, and very little extra masking beyond the occasional dead pixel, this
714 is not the case for the single crystal data. The single crystal data is more problem-
715 atic because of its short exposure time and low flux, with 500 frame at a .1 second
716 exposure and having shrunk the beam size. The low flux is to prevent the very strong
717 single crystal peaks from damaging the detector. However, this causes the image
718 to be less statistically viable then ideal, causing problems with the mask as seen in
719 figure 4.14. This can be alleviated to some degree by using the previously generated
720 mask as a starting mask for the single crystal image, as shown in 4.15. While the
721 masking algorithm still produces many diffuse masked pixels, they are far fewer, this
722 may be due to the removal of the beamstop which could have contributed to the large
723 standard deviation in figure 4.14.

724 Conclusions

725 In this section the masking algorithm, which relies on both Q resolution based bin-
726 ning and a statistical approach to azimuthal symmetry, was developed. The focus of
727 this algorithm was to remove many unwanted detector features associated with pixel
728 defect, beamstop holder associated scattering attenuation, and single crystal/texture
729 peaks. Simulated data was used to evaluate the beamstop holder and dead/hot pixel
730 masking capacity, while experimental data was used to check for single crystal and
731 texture based masking. Q resolution based binning was shown to be very important
732 to avoid over-masking. The ability of the mask writer to mask images is somewhat
733 limited by the overall statistical image quality, although some deficiencies can be
734 obtained by using previously generated masks as starting points. This masking algo-

735 rithm is now in use in the data processing workflow and will be available in scikit-beam
736 soon.

737 4.4 AUTOMATED IMAGE AZIMUTHAL INTEGRATION

738 Using the Q resolution binning and masking developed in sections 4.2 and 4.3 the
739 images can be properly integrated. Generally, images are integrated by taking the
740 mean value of the pixels in a ring. However, other statistical measures of the average
741 value can be used, like the median.

742 Figures 4.16-4.18 show the importance of masking and the choice of average func-
743 tion. All the figures were produced using the same dataset, 50 °C Pr₂NiO₄ taken at
744 the APS's 11-ID-B on a Perkin Elmer area detector. The automatic masking alpha
745 was 3 standard deviations from the mean. While it is difficult to observe the changes
746 the mask causes in the full $I(Q)$ plot (subfigures a) and b)), the standard deviation
747 plots show the effect of bad pixels on the data (subfigure c)). Subfigure c) for figures
748 4.16-4.18 shows that removal of the beamstop holder lowers the low Q standard de-
749 viation from around .1 to almost .01 out to 15 Å⁻¹. The high Q subfigures d) and f)
750 in figures 4.16-4.18 show the “kink” effect of the detector edge and beamstop holder,
751 where there is a dip in the $I(Q)$ scattering when the rings include the edge of the
752 detector. This effect seems to be due to both errors in the edge pixel intensity and the
753 beamstop holder as masking of the edges only seems to provide only partial removal
754 of the issue. It is important to note that while integration using the mean of the
755 ring has issues with only the edge mask, as evidenced by the change in slope in 4.17
756 d) around 29.5 Å⁻¹, the median integration does not include this error. Ideally the
757 detector would have a normal distribution of pixel intensity for a given ring, which
758 would imply an equivalency between the mean and median $I(Q)$ values. Despite the
759 closeness of the mean and median once the final mask has been created, it seems that
760 the median is more reliable, as it was less effected by the beamstop holder in figure

761 4.17. Thus, for subsequent integrations discussed in this work the median is used to
 762 avoid any defective features that the masking algorithm may have missed.

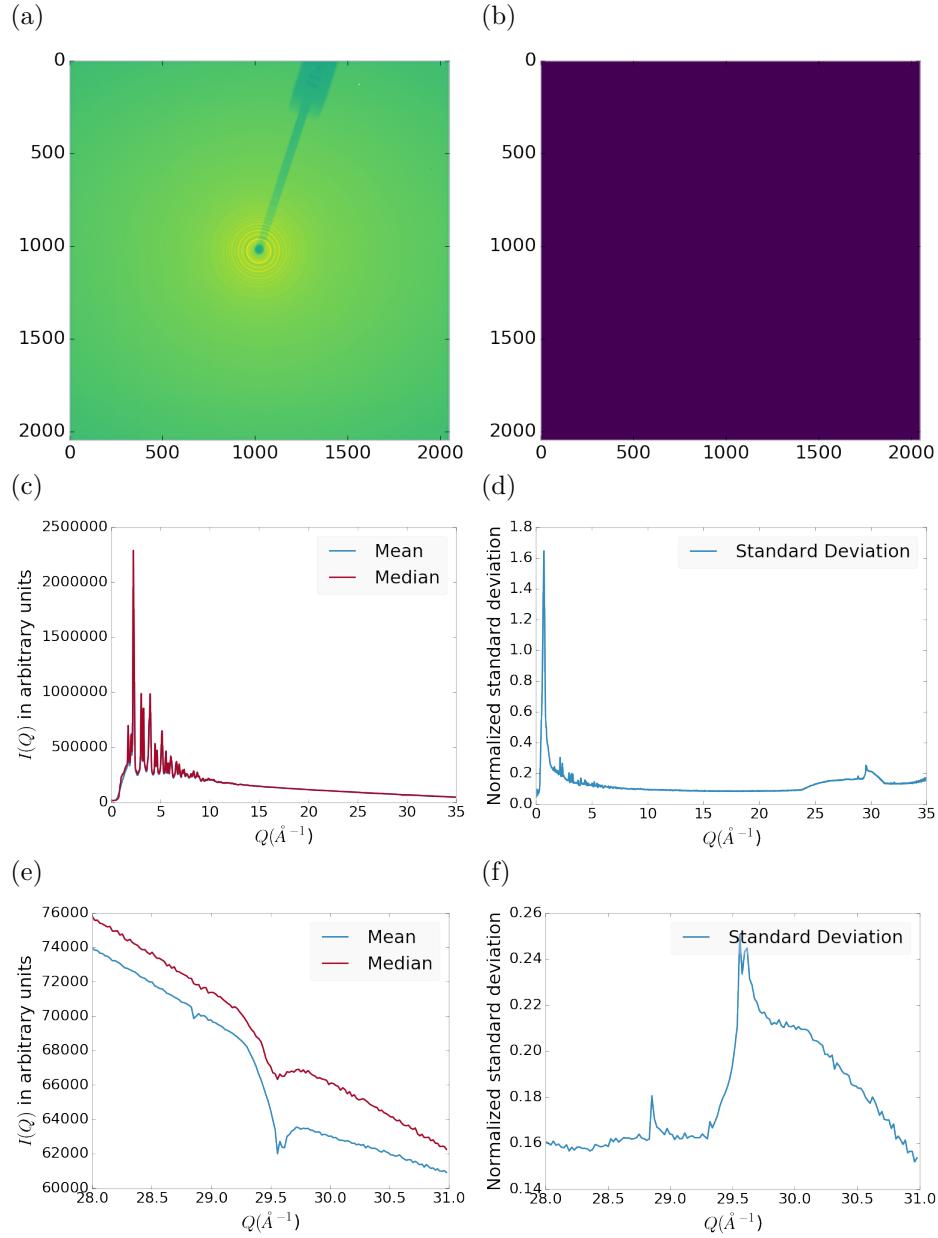


Figure 4.16: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with no mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the 28 \AA^{-1} to 31 \AA^{-1} Q range for the mean and median, f) 28 \AA^{-1} to 31 \AA^{-1} Q range for the standard deviation

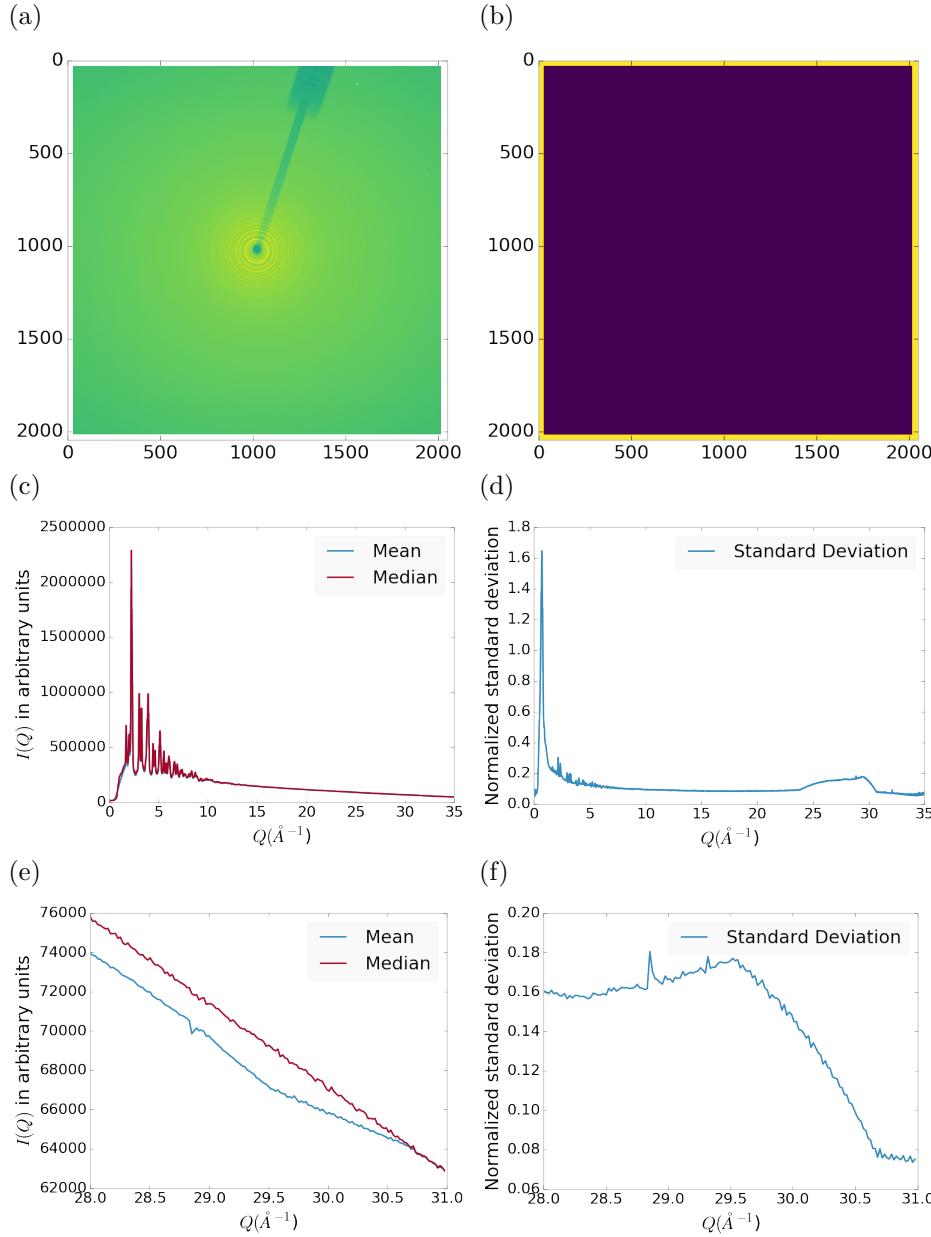


Figure 4.17: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with only an edge mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the 28 \AA^{-1} to 31 \AA^{-1} Q range for the mean and median, f) 28 \AA^{-1} to 31 \AA^{-1} Q range for the standard deviation

763 4.5 CONCLUSIONS

764 This chapter developed and analyzed the proper data processing and reduction method-
 765 ology for producing reliable $F(Q)$ data from x-ray total scattering measurements.

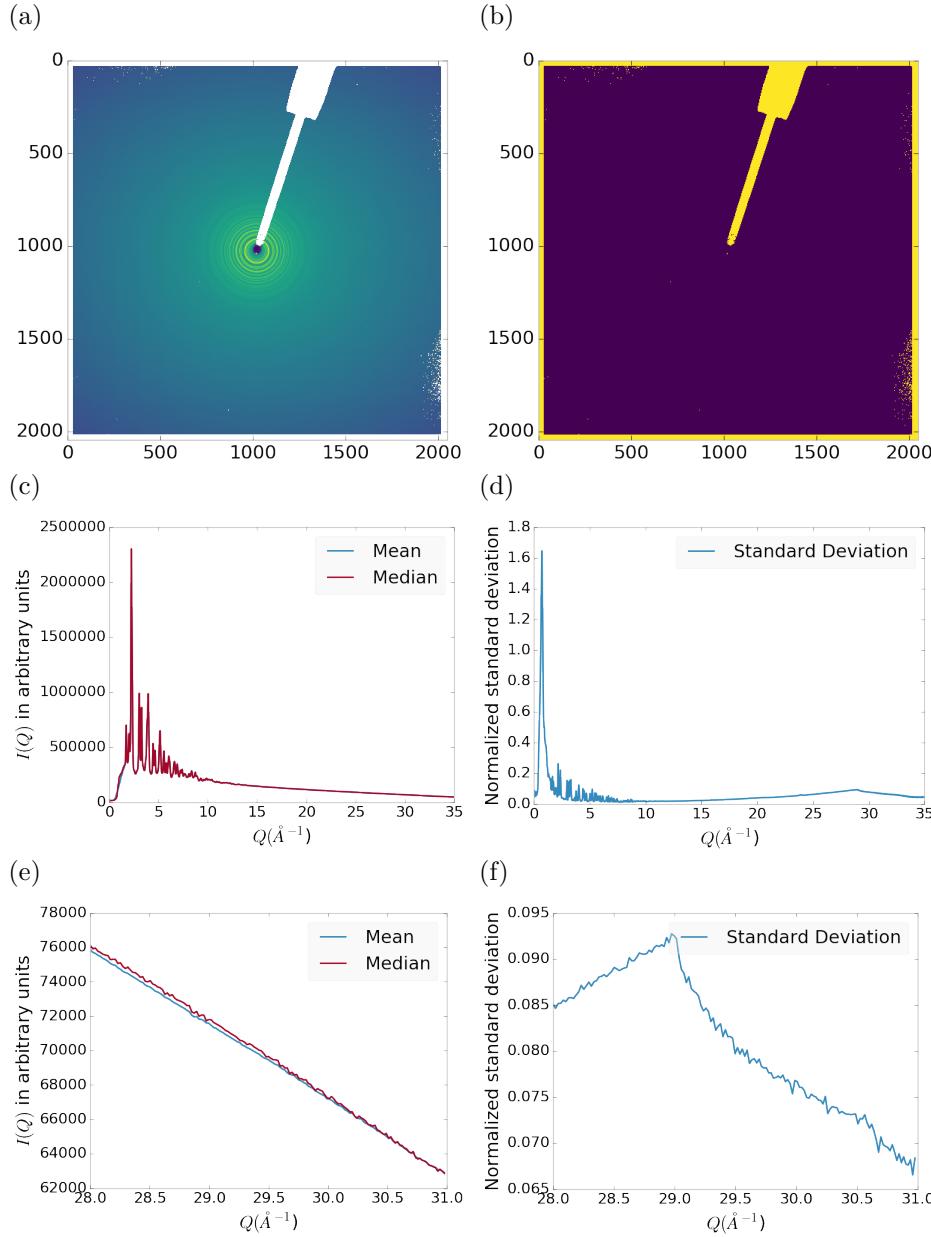


Figure 4.18: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced combining an edge mask and the automatically generated mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median),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

766 Binning at the Q resolution of the detector was found to be key to the data process-
 767 ing. The primary outcome of using the Q resolution binning was an enhancement in
 768 effectiveness for the masking algorithm, producing much fewer false positives for dead

769 pixels. This masking approach was then applied to the integration of experimental
770 data taken at the APD's 11-ID-B beamline. The automatically generated masks,
771 when combined with edge masks, were found to greatly reduce the overall standard
772 deviation of the pixel intensity and produce a smoother $F(Q)$ at high Q , enabling
773 the use of much higher Q data in the PDF. Different statistical measures used in the
774 azimuthal integration was also compared. This comparison showed that the median
775 was a more reliable statistic for integration with data which had more detector de-
776 fects. However, upon properly masking it was shown that these metrics were almost
777 identical. The masking induced similarity between the mean and median shows that
778 the rings, when integrated, may form a Gaussian distribution. The distribution of
779 the pixel intensities for strongly and weakly scattering samples may be investigated
780 in future work.

781

CHAPTER 5

782

ANNEALING AND AGGREGATION OF 2NM

783

AU NANOPARTICLES

784

If we are going to keep this we need to get a lot done in terms of modeling

785 5.1 EXPERIMENTS

786 NP Synthesis

787 X-ray Total Scattering Measurements

788 5.2 DATA PROCESSING

789 5.3 DATA ANALYSIS

790 5.4 SIMULATION

791 5.5 STRUCTURAL ANALYSIS

792 5.6 CONCLUSIONS

793

CHAPTER 6

794

PHASE CHANGES AND ANNEALING DYNAMICS OF

795

Pr_2NiO_4 AND ITS DERIVATIVES

796 6.1 EXPERIMENTS

797 Pr_2NiO_4 Synthesis

798

need some sort of synthesis information

799 X-ray Measurements

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

805 6.2 DATA PROCESSING

806

masking parameters

807

integration parameters

808

PDF parameters

809 6.3 DATA ANALYSIS

Need to redo all the figures:

need to use the new stylesheet

need to be more readable

need to cull through them to pull out the interesting features

810

811 **Intra Sample Comparison**

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

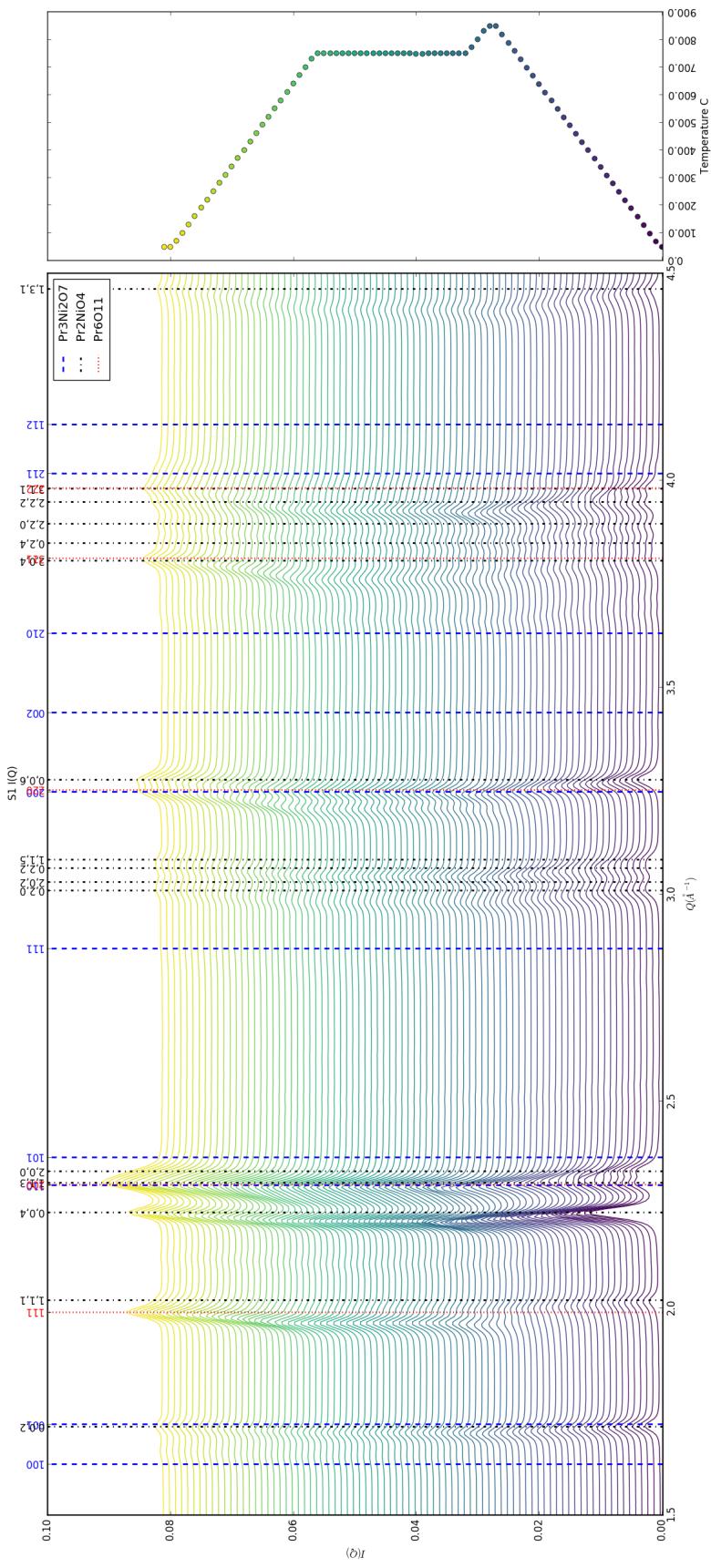


Figure 6.1: XRD as a function of temperature for S1

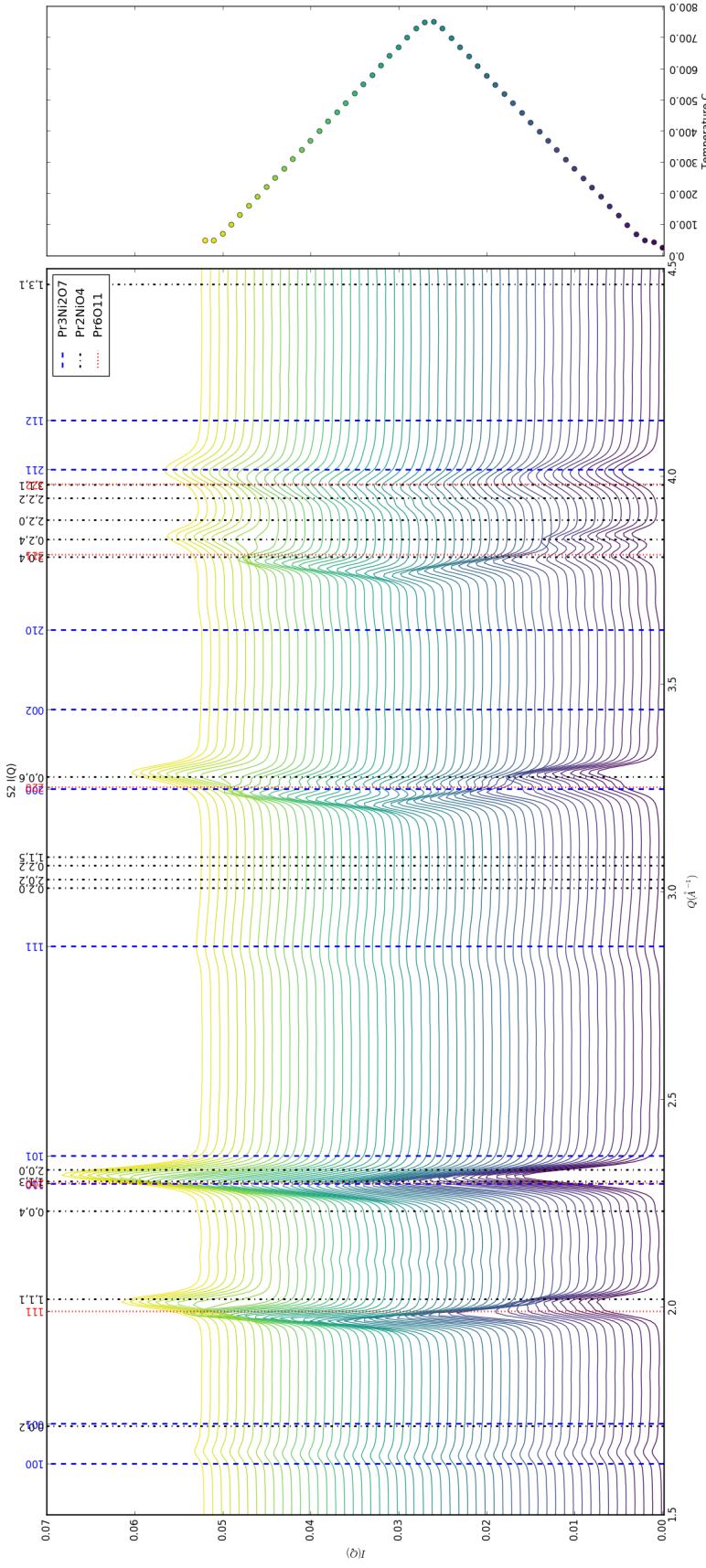


Figure 6.2: XRD as a function of temperature for S2

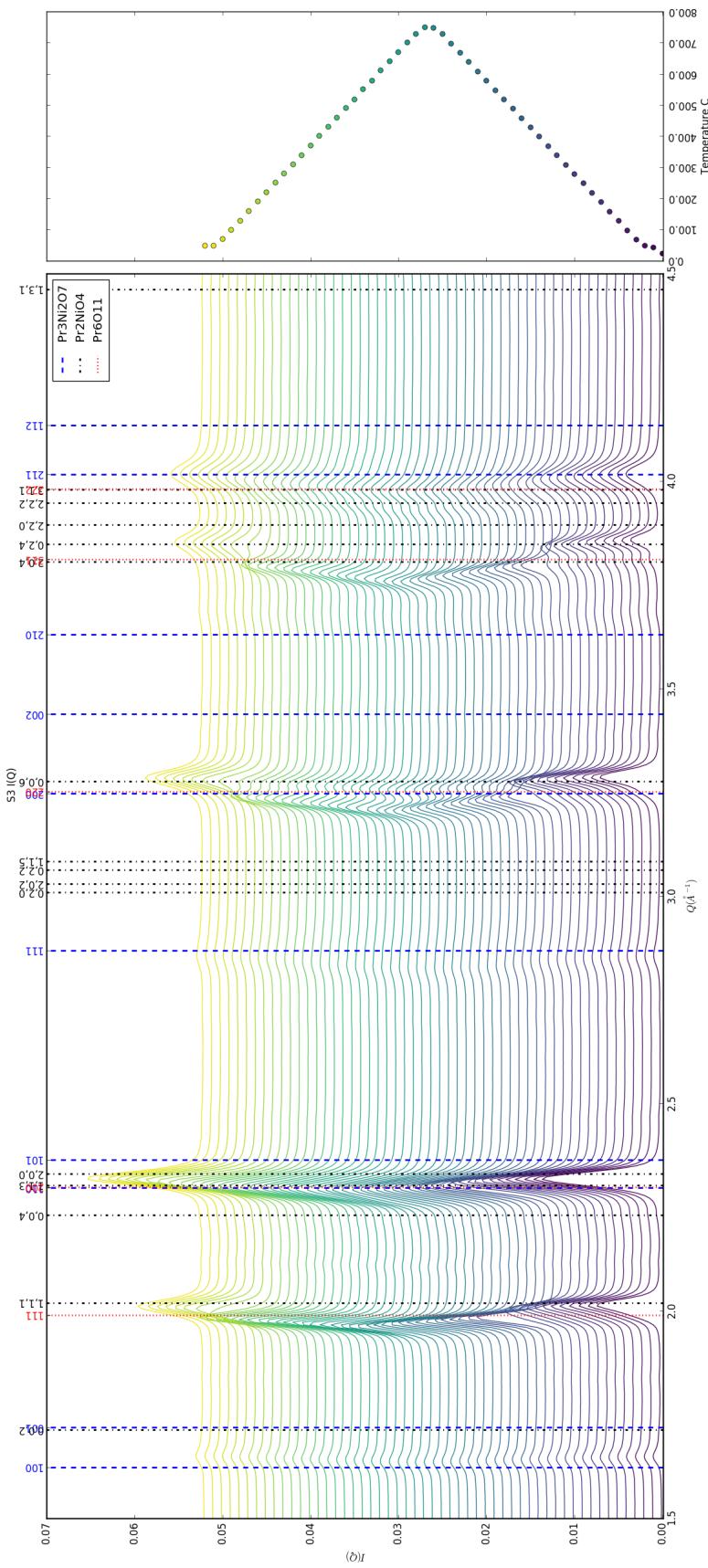


Figure 6.3: XRD as a function of temperature for S3

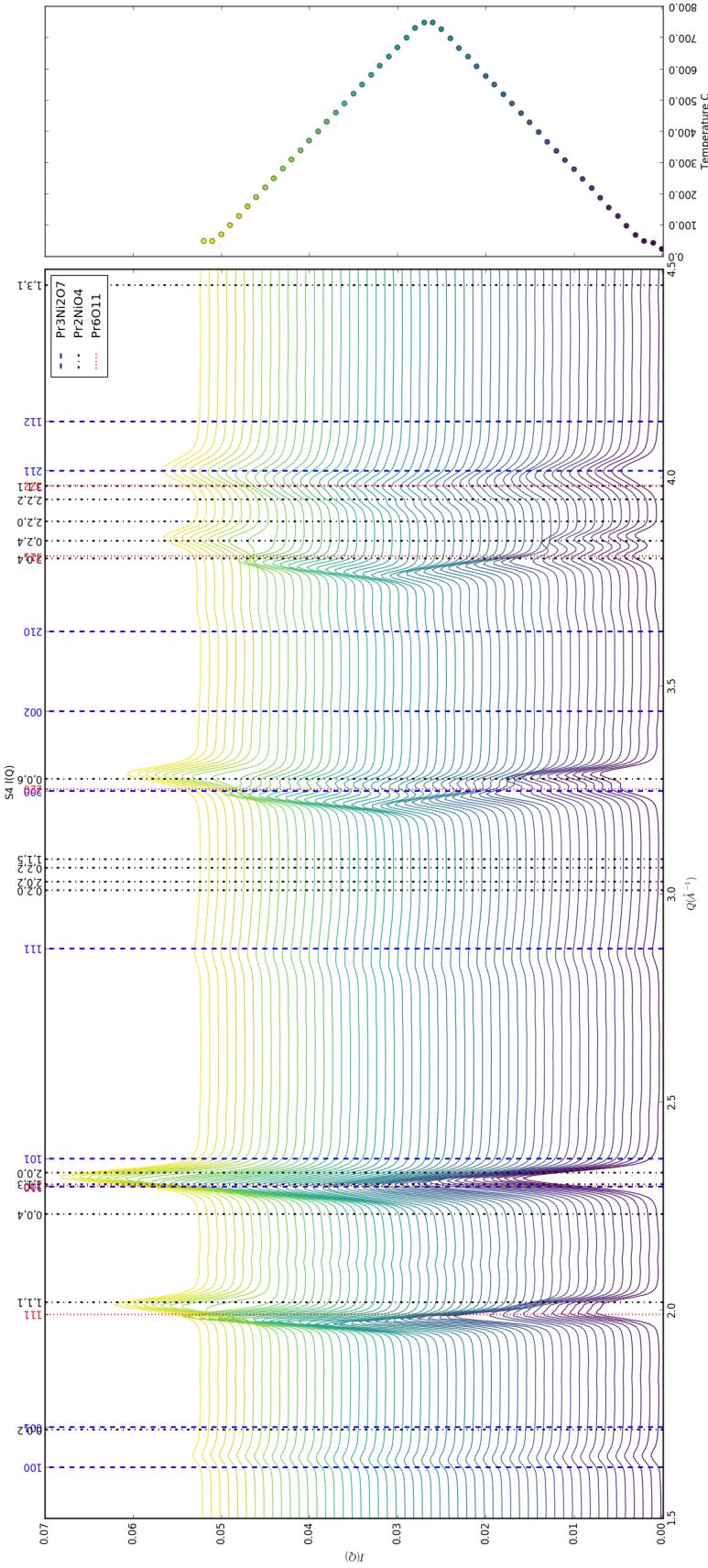


Figure 6.4: XRD as a function of temperature for S4

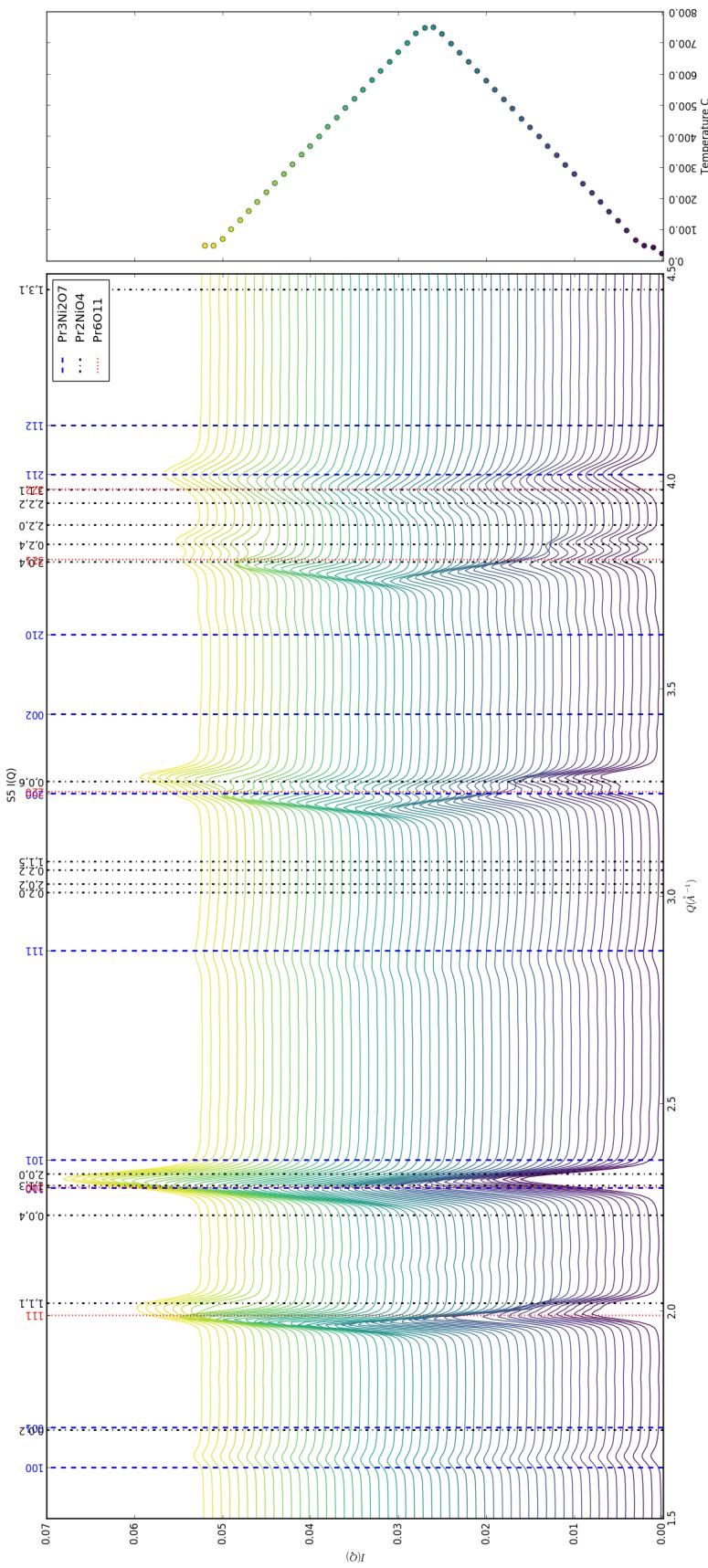
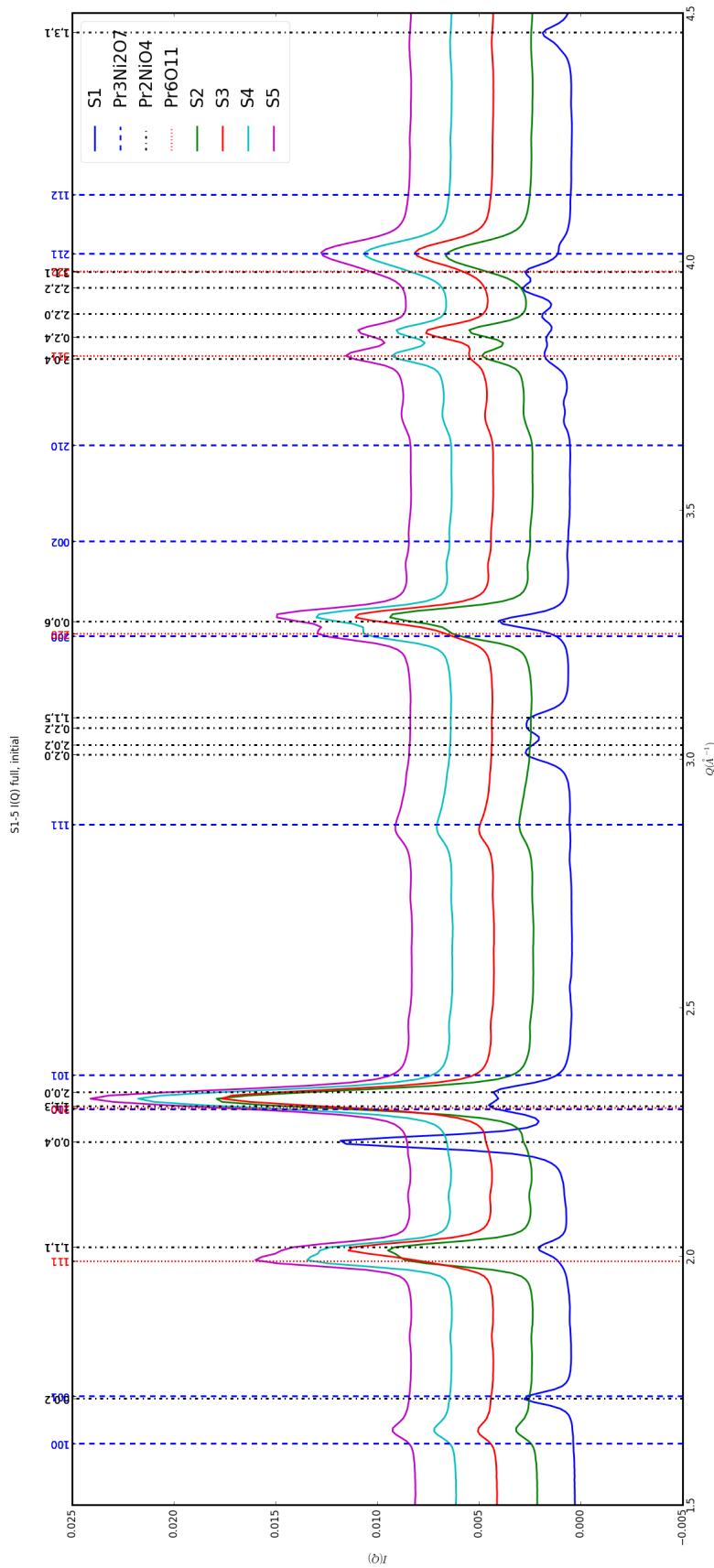
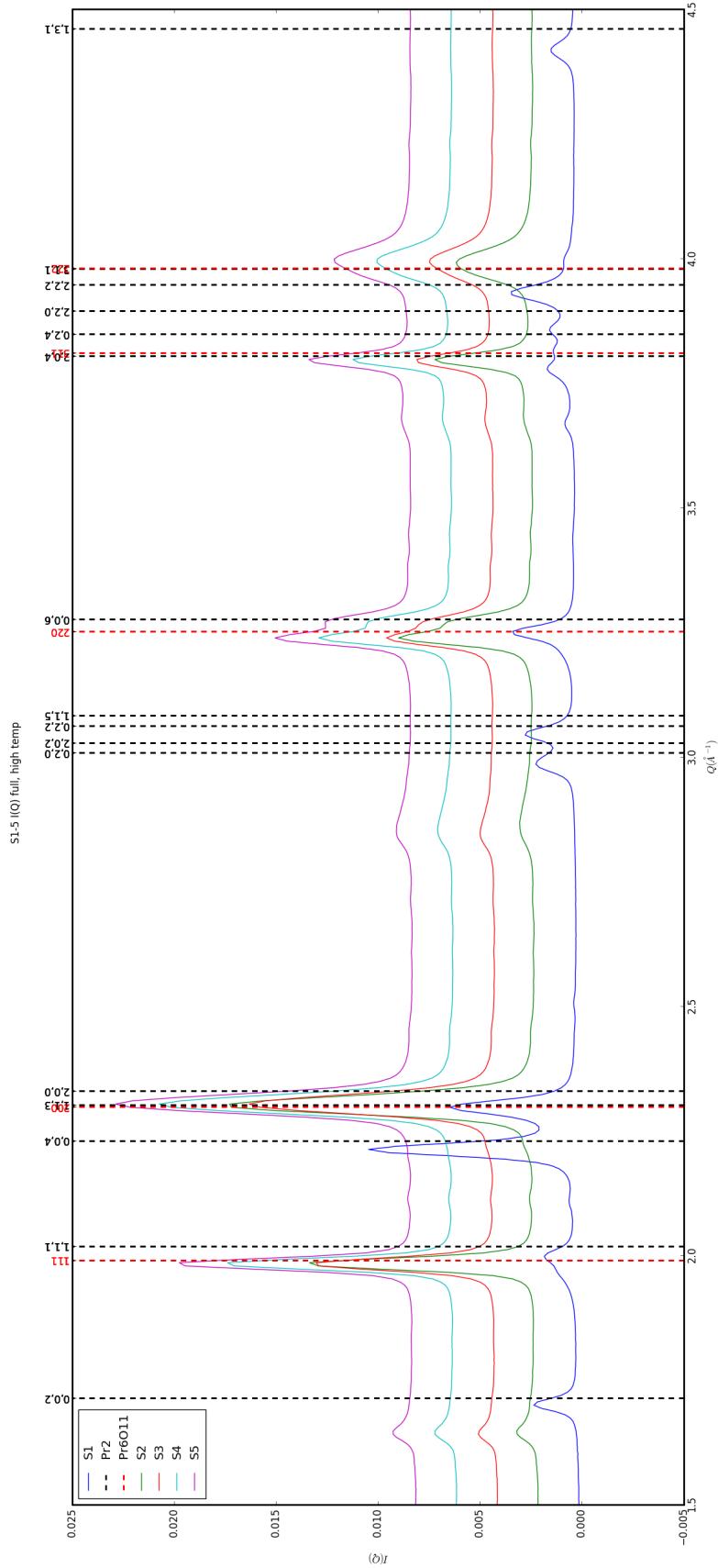
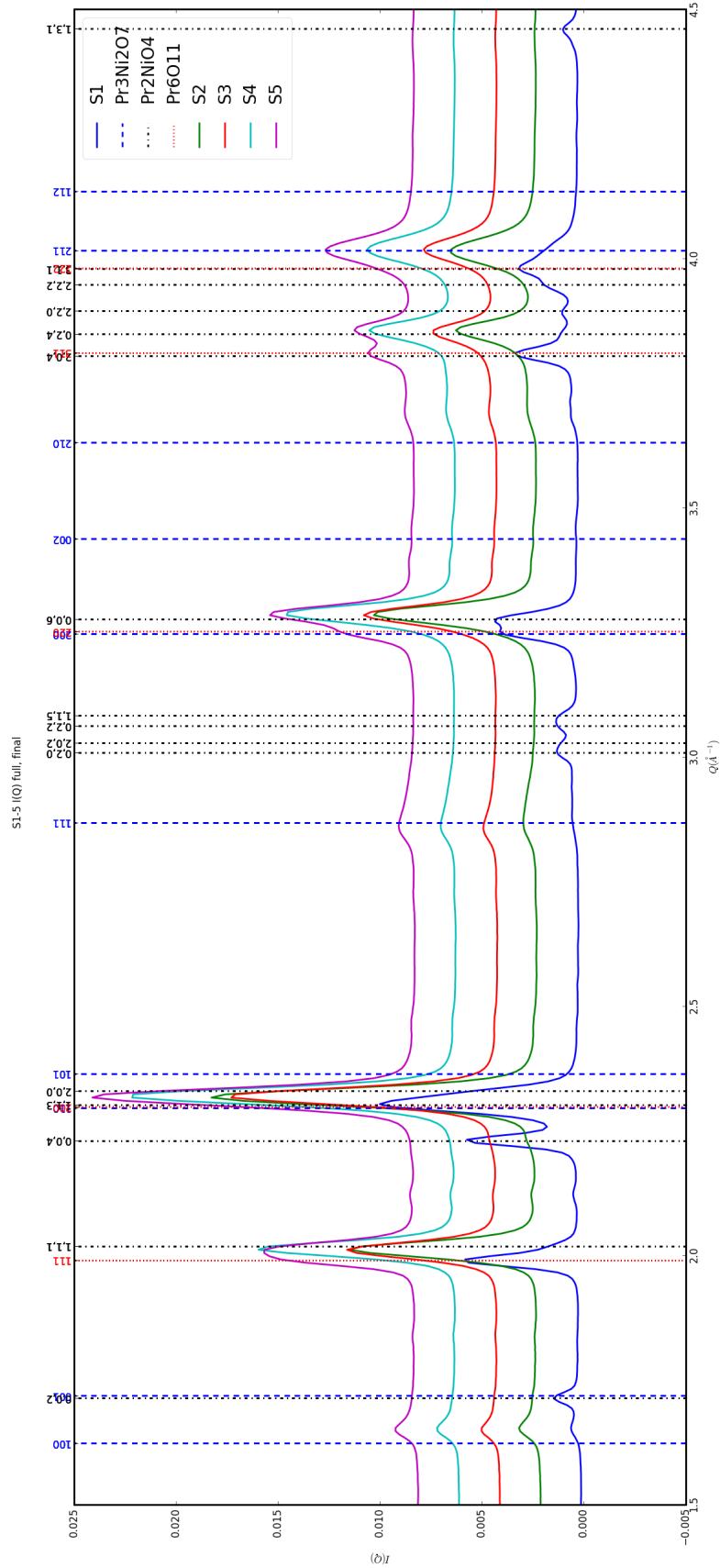


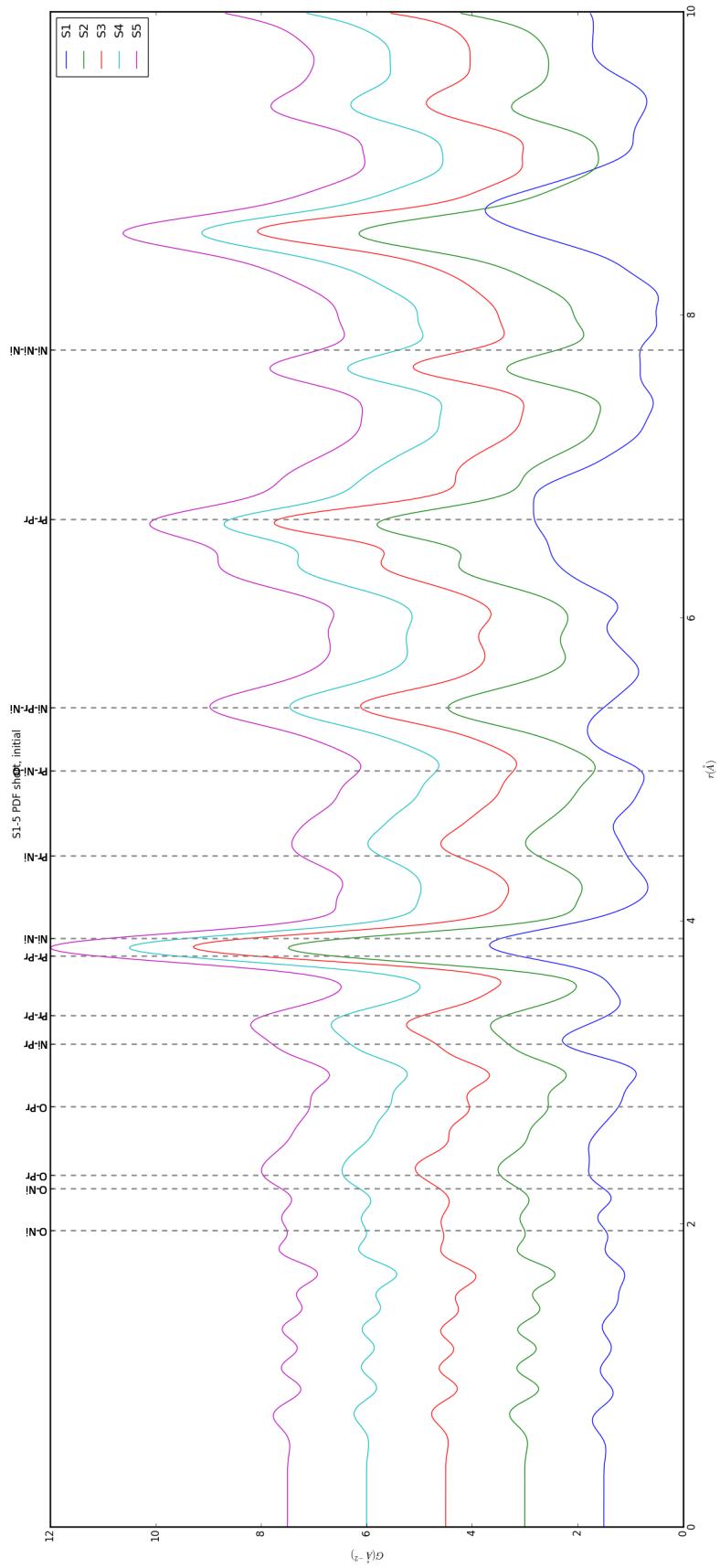
Figure 6.5: XRD as a function of temperature for S5

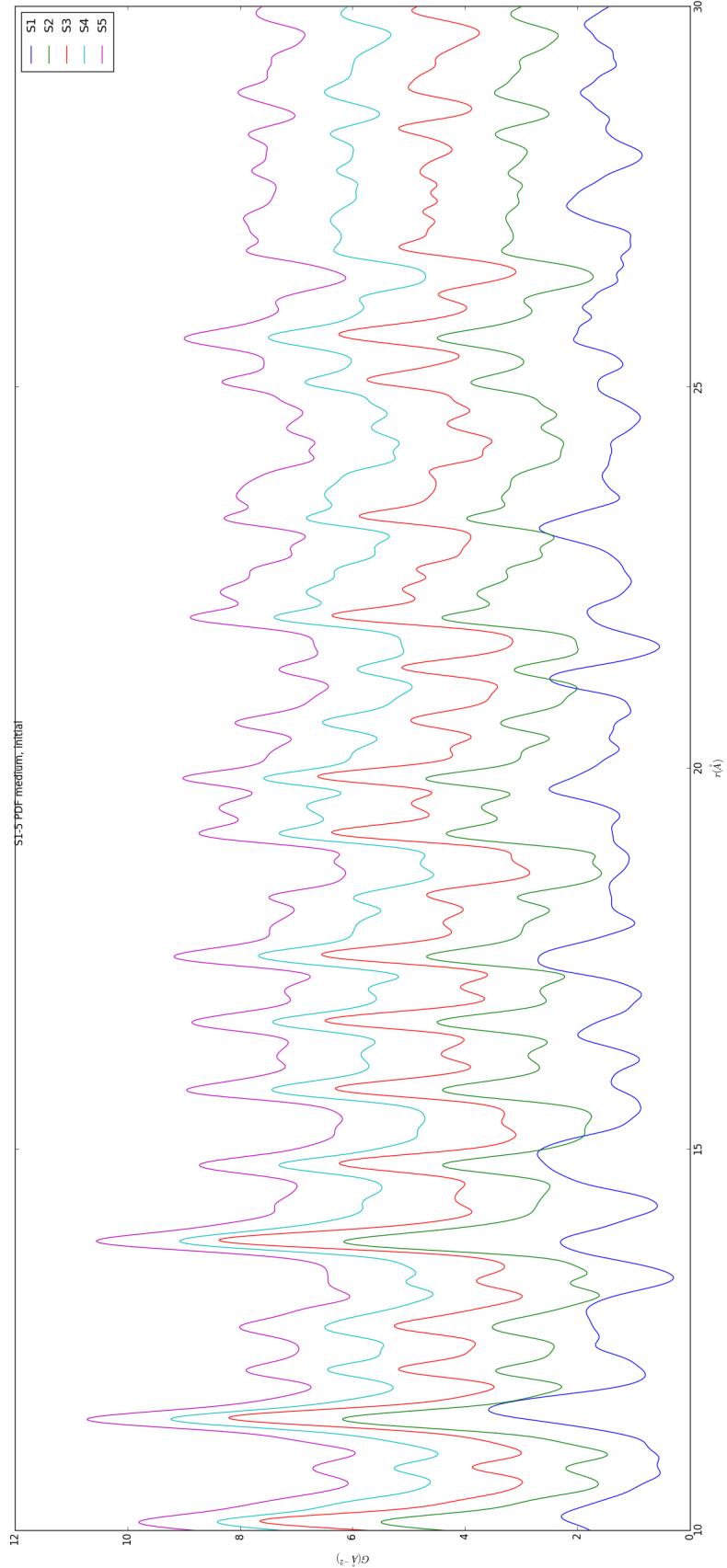
⁸¹³ Inter Sample Comparison

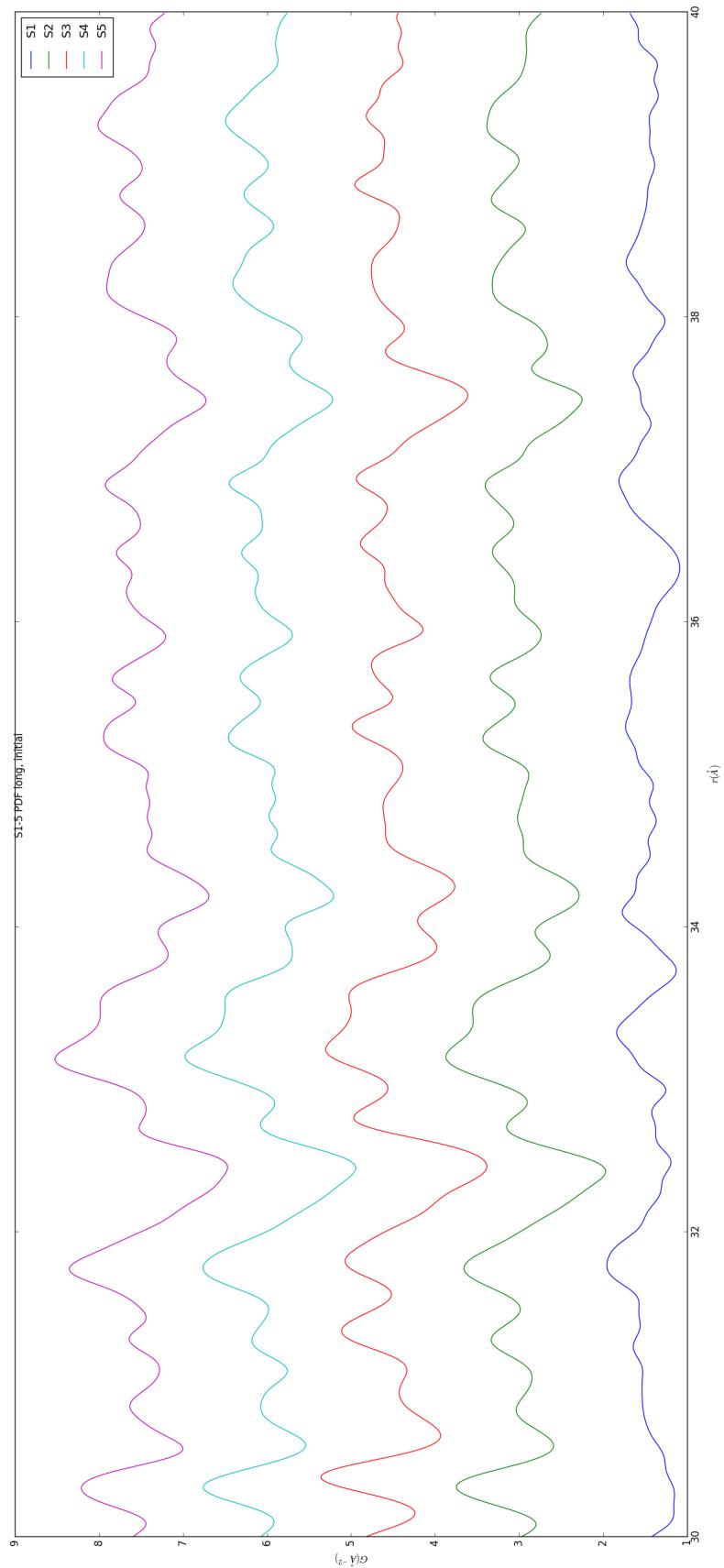


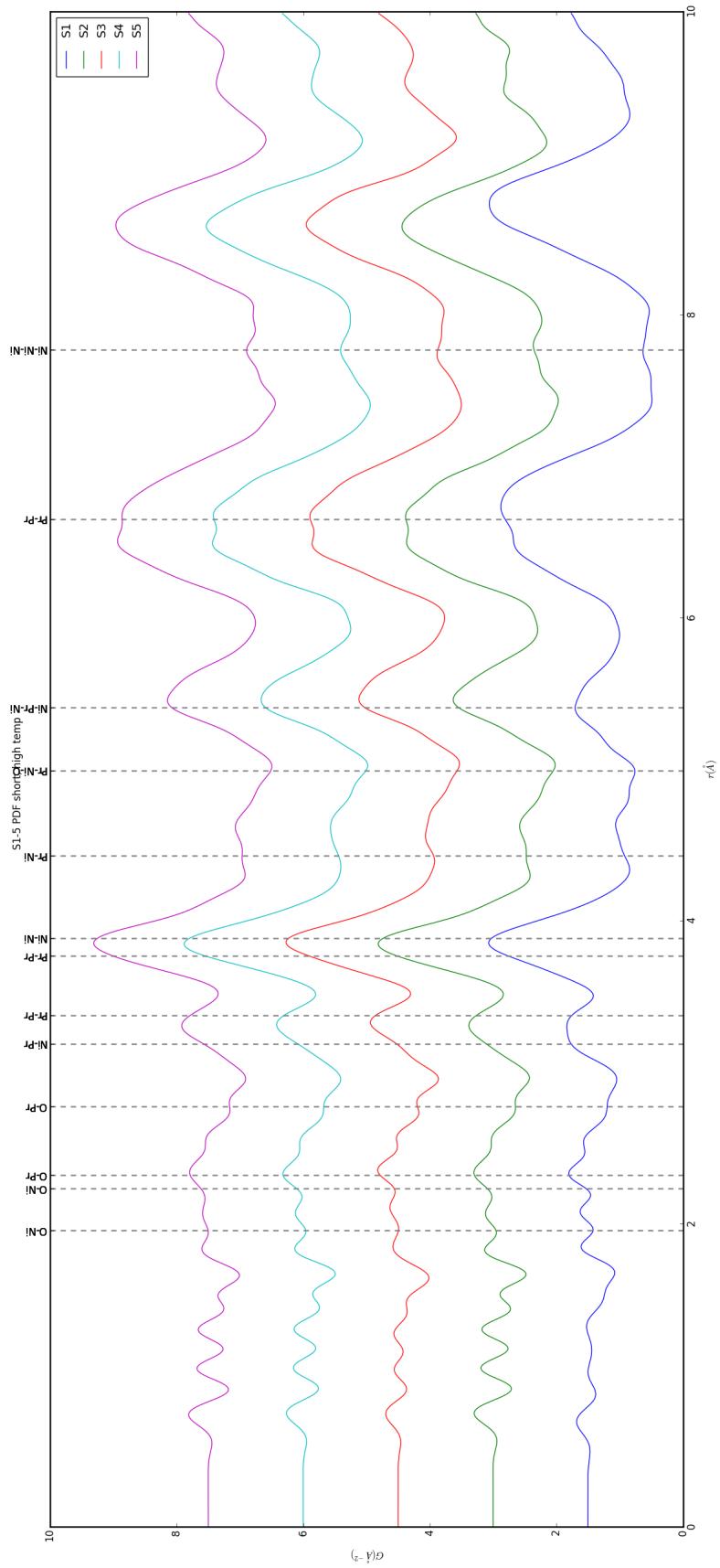


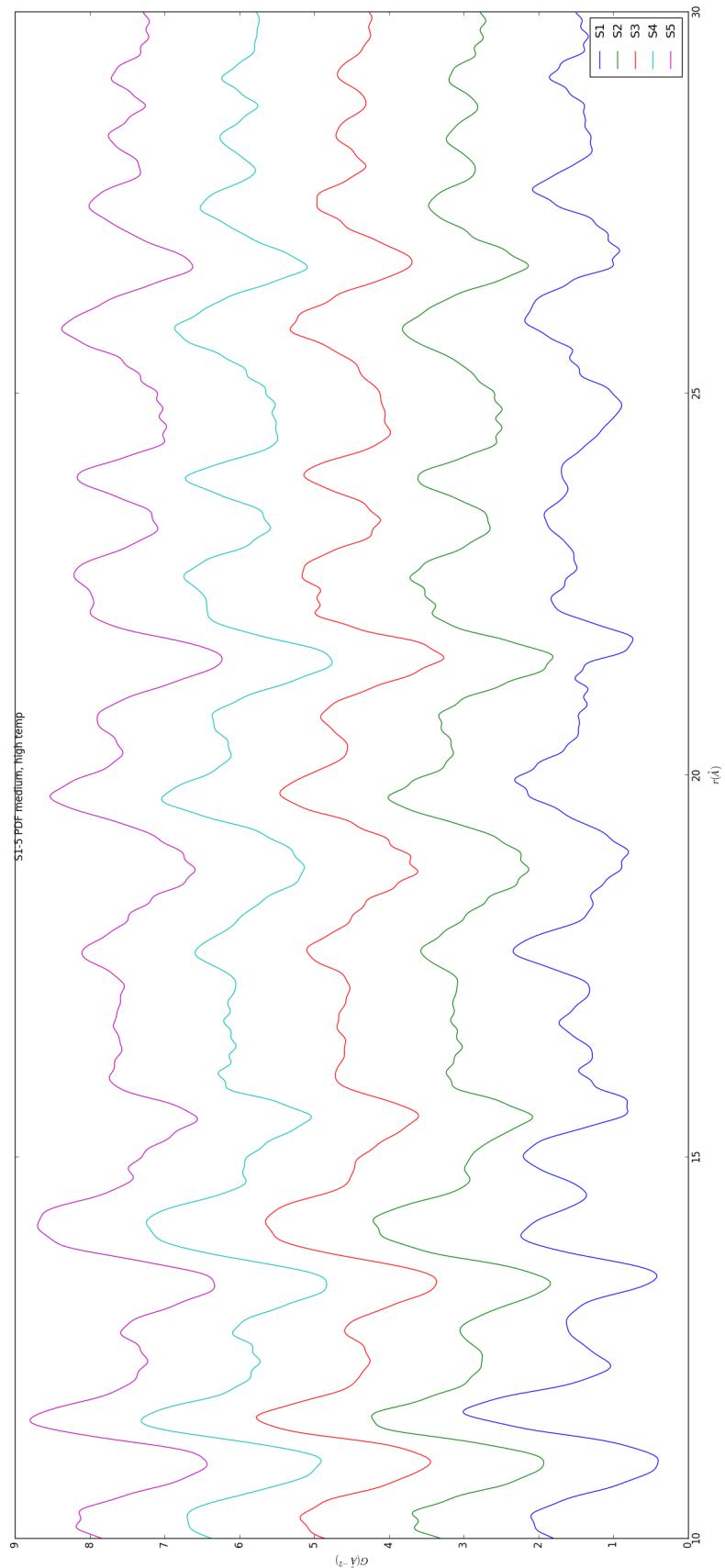


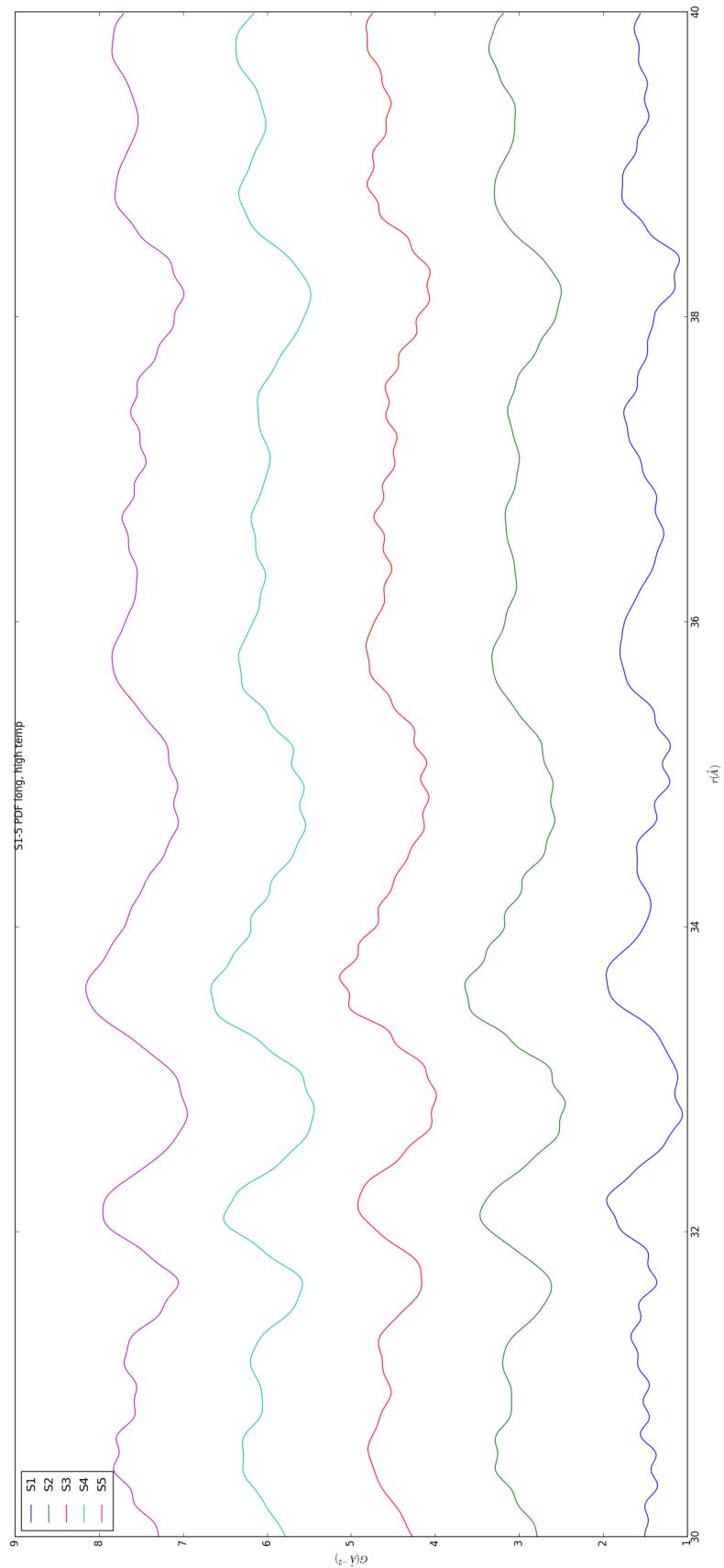


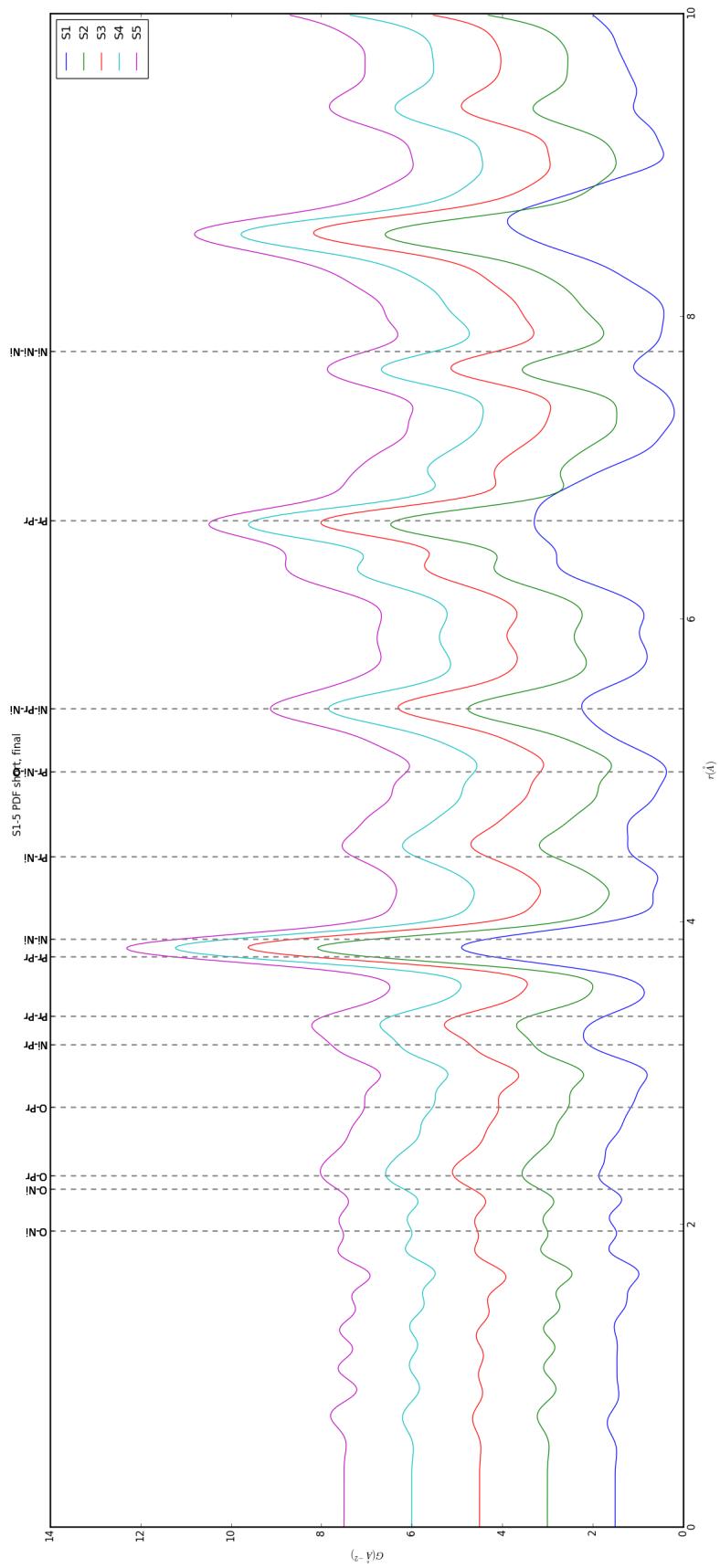


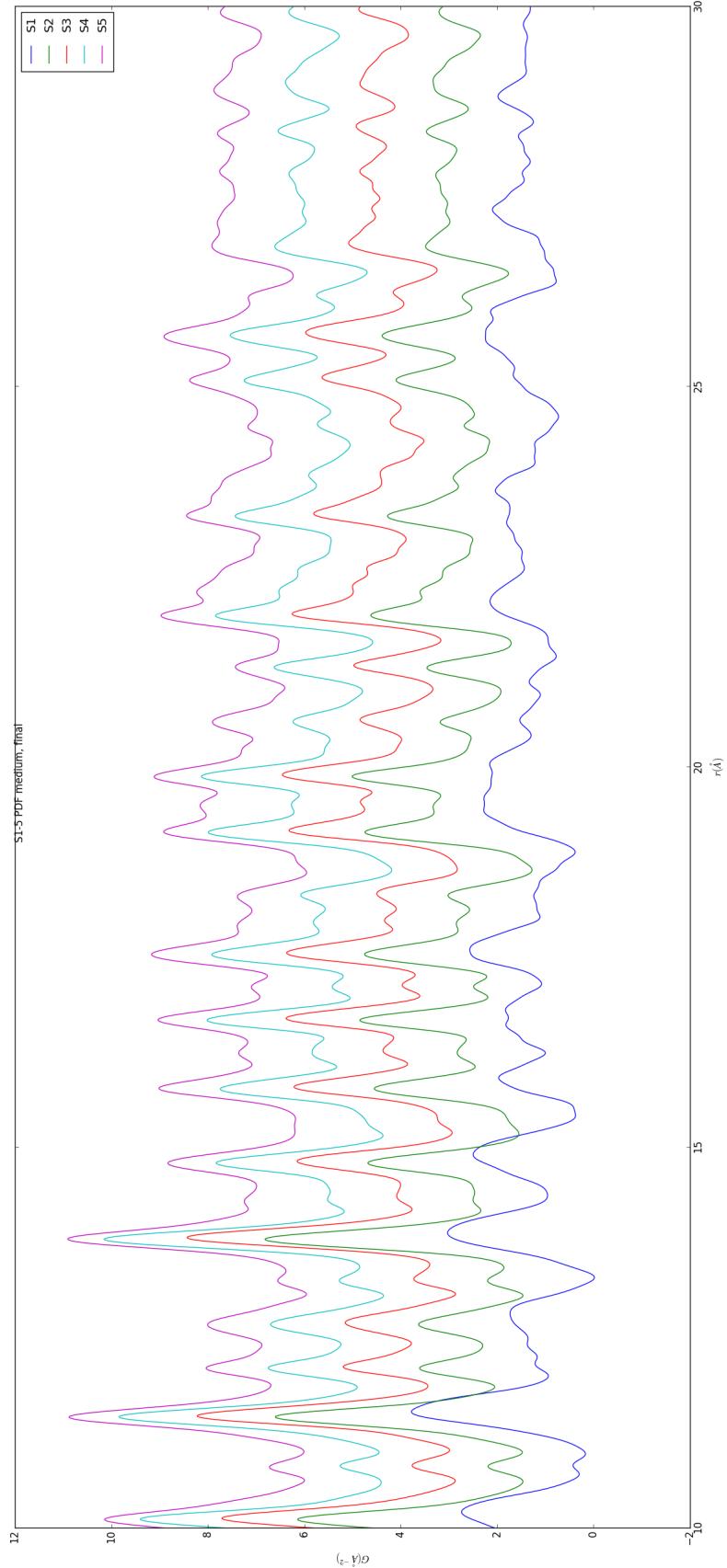


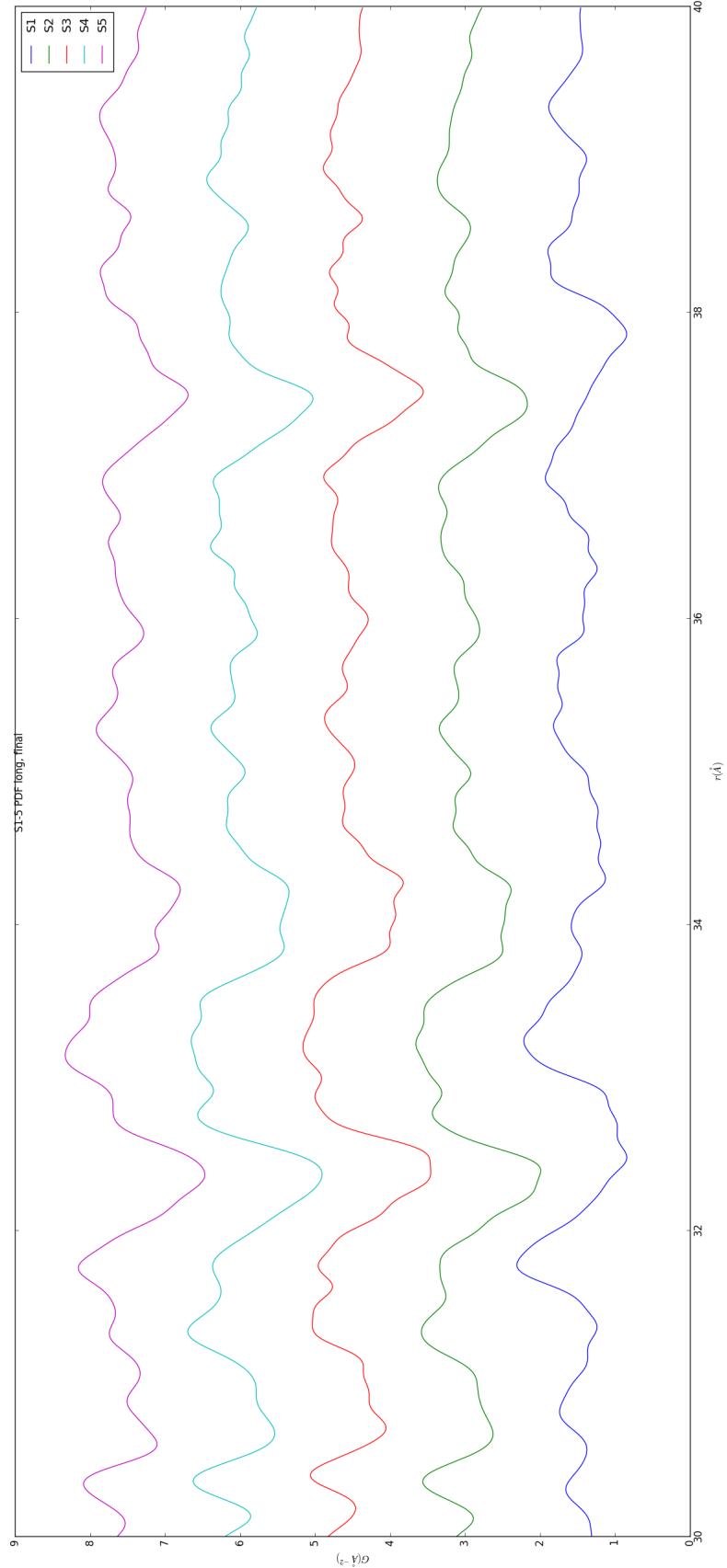












814 The PDF for S1 at operating temperature looks like S2-5 at the same temperature.

815 6.4 SIMULATION

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

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

828 6.5 CONCLUSIONS

829 X-ray total scattering and x-ray powder diffraction data was obtained on Pr_2NiO_4
830 powder samples annealed for various lengths of time. In-situ studies on the beamline
831 were performed to understand how the structure of each of these powders changes
832 at operating temperatures. The data was processed with the previously discussed Q
833 binning, masking, and integration methodology. The PDF results show very little
834 change in the structure for the as synthesized sample. However, the PDFs show a
835 large change in the previously annealed samples. These changes seem to reporduce
836 PDFs similar to the as-synthesized PNO at operating temperatures. This would seem
837 to imply that the source of the anamolus PNO phase/power density relationship may

838 be due to the adoption of an active structure upon heating which is universal despite
839 the amount of thermal degradation observed at room temperature. In contrast to the
840 PDF results, the XRD results seem to show signifigant changes in the PNO structure,
841 both with ex-situ and in-situ annealing. The XRDs show the degradation of the PNO
842 into various phases, potentially including Pr_2O_{11} , and higher ordered Pr based phases.
843 The discrepancy between these two results is quite interesting as it seems that the
844 XRD and PDF results are contradictory. Turbostratic diplacements between the
845 layers may be a cause of the PDF/XRD disagreement, as these changes would cause
846 very little change in the local structure observed in the PDF, while causing large
847 changes in the XRD.

848

CHAPTER 7

849

CONCLUSION

BIBLIOGRAPHY

- 851 [1] P. E. Blöchl, *Projector augmented-wave method*, Physical Review B **50** (1994),
852 no. 24, 17953–17979.
- 853 [2] Joshua J Choi, Xiaohao Yang, Zachariah M Norman, Simon J L Billinge, and
854 Jonathan S Owen, *Structure of methylammonium lead iodide within mesoporous*
855 *titanium dioxide: Active material in high-performance perovskite solar cells*,
856 Nano Letters **14** (2014), no. 1, 127–133.
- 857 [3] Matthew J. Cliffe, Martin T. Dove, D. a. Drabold, and Andrew L. Goodwin,
858 *Structure determination of disordered materials from diffraction data*, Physical
859 Review Letters **104** (2010), no. 12, 1–4.
- 860 [4] Matthew J Cliffe and Andrew L Goodwin, *Nanostructure determination from the*
861 *pair distribution function: a parametric study of the INVERT approach.*, Journal
862 of physics. Condensed matter : an Institute of Physics journal **25** (2013), no. 45,
863 454218.
- 864 [5] Juarez L F Da Silva, Hyoung Gyu Kim, Maurício J. Piotrowski, Maurício J. Pri-
865 eto, and Germano Tremiliosi-Filho, *Reconstruction of core and surface nanopar-*
866 *ticles: The example of Pt 55 and Au55*, Physical Review B - Condensed Matter
867 and Materials Physics **82** (2010), no. 20, 1–6.
- 868 [6] C L Farrow, P Juhas, J W Liu, D Bryndin, E S Božin, J Bloch, Th Proffen, and
869 S J L Billinge, *PDFfit2 and PDFgui: computer programs for studying nanostruc-*
870 *ture in crystals.*, Journal of Physics. Condensed Matter : an Institute of Physics
871 journal **19** (2007), no. 33, 335219.
- 872 [7] Christopher L Farrow and Simon J L Billinge, *Relationship between the atomic*
873 *pair distribution function and small-angle scattering: implications for modeling*
874 *of nanoparticles.*, Acta Crystallographica Section A Foundations of Crystallog-
875 *raphy* **65** (2009), no. Pt 3, 232–9 (en).
- 876 [8] Riccardo Ferrando, Julius Jellinek, and Roy L Johnston, *Nanoalloys: From The-*
877 *ory to Applications of Alloy Clusters and Nanoparticles*, Chemical Reviews **108**
878 (2008), no. 3, 846–904.

- 879 [9] Md Matthew D. Hoffman and Andrew Gelman, *The No-U-Turn Sampler: Adap-*
880 *tively Setting Path Lengths in Hamiltonian Monte Carlo*, The Journal of Machine
881 Learning Research **15** (2014), no. 2008, 1593–1623.
- 882 [10] Pablo D Jazdzinsky, Guillermo Calero, Christopher J Ackerson, David A Bushnell,
883 and Roger D Kornberg, *Structure of a thiol monolayer-protected gold nanoparti-*
884 *cle at 1.1 Å resolution.*, Science (New York, N.Y.) **318** (2007), no. 5849, 430–433.
- 885 [11] G. Kresse and J. Hafner, *Ab initio molecular dynamics for liquid metals*, Physical
886 Review B **47** (1993), no. 1, 558–561.
- 887 [12] ———, *Ab initio molecular-dynamics simulation of the liquid-*
888 *metal–amorphous-semiconductor transition in germanium*, Physical Review
889 B **49** (1994), no. 20, 14251–14269.
- 890 [13] Yan Li, Giulia Galli, and François Gygi, *Electronic structure of thiolate-covered*
891 *gold nanoparticles: Au102(MBA)44*, ACS Nano **2** (2008), no. 9, 1896–1902.
- 892 [14] L D Marks, *Experimental studies of small particle structures*, Reports on
893 Progress in Physics **57** (1994), no. 6, 603–649 (en).
- 894 [15] A. S. Masadeh, E. S. Božin, C. L. Farrow, G. Paglia, P. Juhas, S. J. L. Billinge,
895 A. Karkamkar, and M. G. Kanatzidis, *Quantitative size-dependent structure and*
896 *strain determination of CdSe nanoparticles using atomic pair distribution func-*
897 *tion analysis*, Physical Review B - Condensed Matter and Materials Physics **76**
898 (2007), no. 11, 115413.
- 899 [16] R L McGreevy and L Pusztai, *Reverse Monte Carlo Simulation: A New Tech-*
900 *nique for the Determination of Disordered Structures*, Molecular Simulation **1**
901 (1988), no. 6, 359–367.
- 902 [17] Radford M. Neal, *MCMC Using Hamiltonian Dynamics*, Handbook of Markov
903 Chain Monte Carlo (Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li
904 Meng, eds.), Chapman and Hall/CRC, 2011, pp. 113–162.
- 905 [18] John P. Perdew, Kieron Burke, and Matthias Ernzerhof, *Generalized Gradient*
906 *Approximation Made Simple*, Physical Review Letters **77** (1996), no. 18, 3865–
907 3868.
- 908 [19] Andrew A. Peterson, *Global optimization of adsorbate-surface structures while*
909 *preserving molecular identity*, Topics in Catalysis **57** (2014), no. 1-4, 40–53.

- 910 [20] Valeri Petkov, Binay Prasai, Yang Ren, Shiyao Shan, Jin Luo, Pharrah Joseph,
911 and Chuan-Jian Zhong, *Solving the nanostructure problem: exemplified on metal-*
912 *lic alloy nanoparticles*, *Nanoscale* **6** (2014), no. 17, 1–11.
- 913 [21] Valeri Petkov, Shiyao Shan, Peter Chupas, Jun Yin, Lefu Yang, Jin Luo, and
914 Chuan-Jian Zhong, *Noble-transition metal nanoparticle breathing in a reactive*
915 *gas atmosphere.*, *Nanoscale* **5** (2013), no. 16, 7379–87.
- 916 [22] Th. Proffen and R. B. Neder, *DISCUS: a Program for Diffuse Scattering and*
917 *Defect-Structure Simulation*, *Journal of Applied Crystallography* **30** (1997), 171–
918 175.
- 919 [23] H. W. Sheng, M. J. Kramer, A. Cadien, T. Fujita, and M. W. Chen, *Highly*
920 *optimized embedded-atom-method potentials for fourteen FCC metals*, *Physical*
921 *Review B - Condensed Matter and Materials Physics* **83** (2011), no. 13, 134118.