

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

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 . . . . .	x
LIST OF FIGURES . . . . .	xi
TODO LIST . . . . .	1
CHAPTER 1 STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES . . . . .	4
1.1 Introduction . . . . .	4
1.2 Potential Energy Surfaces . . . . .	4
Experimentally Derived Potential Energy Surfaces . . . . .	5
Potentials . . . . .	5
Forces . . . . .	6
1.3 Ensembles . . . . .	7
Monte Carlo Modeling . . . . .	8
Hamiltonian Monte Carlo . . . . .	9
No-U-Turn Sampling . . . . .	11
Grand Canonical Ensemble . . . . .	12
Ensemble description . . . . .	12
Grand Canonical Monte Carlo . . . . .	12
GCMC biasing . . . . .	13
1.4 Conclusions . . . . .	14

<b>CHAPTER 2 ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION . . . . .</b>	<b>16</b>
2.1 Introduction . . . . .	16
2.2 Theory . . . . .	16
Derivation . . . . .	16
Analytically Gradients . . . . .	17
Without ADPs . . . . .	18
Periodic Boundary Conditions . . . . .	19
2.3 Computation . . . . .	19
HPC and GPUs . . . . .	19
GPUs and Parallelization . . . . .	20
Map from ij space to k space . . . . .	21
GPU Memory Allocation . . . . .	22
Speed and Scaling of PDF Computation . . . . .	24
2.4 Conclusions . . . . .	25
<b>CHAPTER 3 BENCHMARKING . . . . .</b>	<b>27</b>
3.1 Introduction . . . . .	27
3.2 PDF . . . . .	27
Model Parameters . . . . .	29
Au55: surface relaxed . . . . .	30
Run Parameters . . . . .	30
Au55: surface disordered . . . . .	32
Au55: amorphous . . . . .	33
Au102: triple phase . . . . .	34
Starting from fcc structure . . . . .	34
Marks decahedron . . . . .	35
Au147 . . . . .	36
3.3 PDF with ADPs . . . . .	41
ADP 50 . . . . .	41
<b>CHAPTER 4 X-RAY TOTAL SCATTERING DATA ACQUISITION AND PRO- CESSING . . . . .</b>	<b>42</b>
4.1 Introduction . . . . .	42

4.2	Detector $Q$ resolution . . . . .	42
4.3	Automated Mask Generation . . . . .	46
	Introduction . . . . .	46
	Algorithm Design . . . . .	46
	Test Cases . . . . .	47
	Results and Discussion . . . . .	47
	Conclusions . . . . .	54
4.4	Automated Image Azimuthal Integration . . . . .	54
4.5	Conclusions . . . . .	59
<b>CHAPTER 5 PHASE CHANGES AND ANNEALING DYNAMICS OF <math>\text{Pr}_2\text{NiO}_4</math> AND ITS DERIVATIVES . . . . .</b>		60
5.1	Introduction . . . . .	60
5.2	Experiments . . . . .	60
	$\text{Pr}_2\text{NiO}_4$ Synthesis . . . . .	60
	X-ray Measurements . . . . .	60
5.3	Data Processing . . . . .	60
5.4	Data Analysis . . . . .	61
	Intra Sample Comparison . . . . .	61
	PDF . . . . .	61
	$I(Q)$ . . . . .	66
	Inter Sample Comparison . . . . .	71
5.5	Simulation . . . . .	71
5.6	Conclusions . . . . .	71
<b>BIBLIOGRAPHY . . . . .</b>		74

## LIST OF TABLES

## LIST OF FIGURES

Figure 1.1	2D slice of addition biasing with Rw and Lennard Jones potentials.	14
Figure 2.1	Comparison of the CPU and GPU chip architectures	20
Figure 2.2	Speed comparison of CPU and GPU implementations	25
Figure 3.1	Au <sub>55</sub> PDF fitting of DFT-optimized <i>c</i> -Au <sub>55</sub> .	31
Figure 3.2	Au <sub>55</sub> PDF fitting of surface-disordered Au <sub>55</sub> .	37
Figure 3.3	Similar to figure 3.2 for DFT-optimized amorphous Au <sub>55</sub> .	38
Figure 3.4	Similar to figure 3.2 for Au <sub>102</sub> as in DFT-optimized Au <sub>102</sub> MBA <sub>44</sub> cluster.	39
Figure 3.5	Similar to Fig. 3.4 with Marks decahedron as the starting structure.	40
Figure 3.6	Refinement of adps	41
Figure 4.1	Scattering onto a flat detector	43
Figure 4.2	$Q$ resolution as a function of $Q$ .	44
Figure 4.3	Number of pixels as a function of $Q$ , binned at the $Q$ resolution of the detector.	45
Figure 4.4	Generated dead/hot pixel masks for a detector with 100 bad pixels.	48
Figure 4.5	Generated dead/hot pixel masks for a detector with 300 bad pixels.	48
Figure 4.6	Generated dead/hot pixel masks for a detector with 500 bad pixels.	49
Figure 4.7	Generated dead/hot pixel masks for a detector with 1000 bad pixels.	49

Figure 4.8 Generated beamstop holder masks for a beamstop holder with 10% transmittance. . . . .	50
Figure 4.9 Generated beamstop holder masks for a beamstop holder with 30% transmittance. . . . .	50
Figure 4.10 Generated beamstop holder masks for a beamstop holder with 50% transmittance. . . . .	50
Figure 4.11 Generated beamstop holder masks for a beamstop holder with 90% transmittance. . . . .	51
Figure 4.12 Generated beamstop holder masks which is rotated away from vertical . . . . .	51
Figure 4.13 Masked experimental data. . . . .	52
Figure 4.14 Masked experimental data with Pt single crystal signal. . . . .	53
Figure 4.15 Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask. . . . .	53
Figure 4.16 Masking, average, and standard deviation of an example x-ray total scattering measurement with with no mask. . . . .	56
Figure 4.17 Masking, average, and standard deviation of an example x-ray total scattering measurement with with only an edge mask. . . . .	57
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. . . . .	58
Figure 5.1 PDF as a function of temperature for as synthesized PNO showing the full PDF . . . . .	62
Figure 5.2 PDF as a function of temperature for as synthesized PNO showing a close up on the short range section . . . . .	63
Figure 5.3 PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full PDF . . . . .	64
Figure 5.4 PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the short range section . . . . .	65

Figure 5.5	$I(Q)$ as a function of temperature for as synthesized PNO showing the full XRD	67
Figure 5.6	$I(Q)$ as a function of temperature for as synthesized PNO showing a close up on the low $Q$ section	68
Figure 5.7	$I(Q)$ as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full XRD	69
Figure 5.8	$I(Q)$ as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the low $Q$ section	70

1

## TODO LIST

2	Need more references . . . . .	1
3	double check against McQuerry, cite it, also move from probabilities to par-	
4	tition functions, develop partition functions above . . . . .	8
5	Include figure which shows the configurational biasing map . . . . .	14
6	Need some introduction about why we want to use the PDF . . . . .	16
7	Also some introduction would be great . . . . .	27
8	this just needs a lot of work . . . . .	27
9	Put this somewhere . . . . .	31
10	We should discuss about why PNO is interesting, at least in brief . . . . .	60
11	need some sort of synthesis information, something along the lines of as	
12	previously reported . . . . .	60
13	Need more references	

## INTRODUCTION

15 Engineering materials and chemicals on the atomic scale has long been a goal for  
16 the chemistry, physics, materials science, and chemical engineering fields. Realizing  
17 this goal could lead to durable fuel cell catalysts, more bioavailable pharmaceuticals,  
18 and radiation damage resistant spacecraft shielding. Before we can even think of  
19 making atomistically exact structures, durable structures, or structures which change in  
20 reproducible ways, we need to know the atomic structure exactly. This work addresses  
21 these issues by developing a methodology for solving the structure of nanomaterials  
22 by matching experimental x-ray scattering data with simulated atomic structures.

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

28 Chapter 2 will discuss the mathematical and computational development of the  
29 atomic pair distribution function (PDF). §2.3 will focus on the rapid graphical pro-  
30 cessing unit based calculation of the PDF and its gradients.

31 Chapter 3 will discuss the benchmarking of the the combined statistical mechan-  
32 ical optimizer and PDF calculation systems against a series of theoretical nanoparti-  
33 cles, focusing on understanding limitations of the method and structure reproduction.

34 Chapter 4 will focus on the aquesition of experimental data, its management, and  
35 processing. §4.2, 4.3, and 4.4 will discuss the derivation of the  $Q$  resolution function,  
36 the automated masking of 2D area detectors for x-ray total scattering measurements

37 using the previously derived  $Q$  resolution, and the impact of different averaging  
38 methods and masks on azimuthal integration, respectively.

39 Chapter 5 will discuss preliminary experimental results investigating the phase  
40 changes and local structure of  $\text{Pr}_2\text{NiO}_4$

41                   CHAPTER 1

42                   STATISTICAL MECHANICAL ENSEMBLES AND  
43                   POTENTIAL ENERGY SURFACES

44           1.1 INTRODUCTION

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

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

57           1.2 POTENTIAL ENERGY SURFACES

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

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

## 69 Experimentally Derived Potential Energy Surfaces

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

## 78 Potentials

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

$$81 \chi^2 = \sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2 \quad (1.1)$$

$$82 \quad 81 \quad R_w = \sqrt{\frac{\sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}}^2}} \quad (1.2)$$

$$82 \quad \chi^2_{\text{INVERT}} = \frac{1}{N} \sum_j \sum_r [A_{\text{obs}}(r) - \alpha A_{j,\text{calc}}(r)]^2 \quad (1.3)$$

83

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

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

87 The  $Rw$  and  $\chi^2$  potentials have been reported numerous times. [31, 24, 4, 25, 33]  
 88 Essentially these potentials measure the least squares distance between the observed  
 89 scattering and the predicted scattering providing a way to quantify the agreement  
 90 between the model and experiment. While  $RW$  and  $\chi^2$  are now standard in the PDF  
 91 community, the INVERT potential is fairly new and aims to incorporate descriptions  
 92 of the structural symmetry into the PES. [7, 8] In the case of the INVERT potential  
 93 NMR or other symmetry sensitive data is used to describe the number of unique  
 94 atomic coordinations. This is then used to describe the number of unique atomwise  
 95 pair distribution functions, thus causing systems with more or less unique coordi-  
 96 nation environments to be higher in energy. This approach has been shown to be  
 97 useful for  $C_{60}$  and other systems which are highly symmetric, creating a PES with  
 98 an easier to find minima. [7, 8] However, many times this kind of data is unavailable  
 99 when refining the structure causing the potential to be less useful. Additionally, this  
 100 potential introduces an element of user bias as the refiner must decide, based on some  
 101 spectroscopic data, how many unique environments are in the material. This bias  
 102 could be removed by using one of the other potentials with a method for simulat-  
 103 ing the observed spectra, allowing the computational system decide what structures  
 104 properly reproduce all the observed data.

105 **Forces**

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

106

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

107

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

108

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

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

114 1.3 ENSEMBLES

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

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

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

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

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

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

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

double check against McQuerry, cite it, also move from probabilities to partition  
functions, develop partition functions above

132

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

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

134 where  $E$  is the energy of the system,  $W$  is the total number of states in the system,  
135 and  $\delta$  is the Dirac Delta Function.

136 However, for the canonical ensemble the probability is:

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

## 137 Monte Carlo Modeling

138 Monte Carlo can be used to simulate a statistical mechanical ensemble which can  
139 not be solved analytically. In most Monte Carlo systems the ensemble is simulated by  
140 randomly changing one of the system parameters and comparing the energy of the  
141 new system against the energy of the old system. If the energy of the new system is  
142 lower than the current energy then the new configuration is accepted. Otherwise the  
143 new system is rejected unless

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

144 where  $u$  is a random number  $[0, 1)$  and  $E_T$  is the thermal energy characteristic to the  
 145 system. The ability of Monte Carlo modeling to accept “bad” moves allows the system  
 146 to hop out of local energy minima during the search for the global minimum. Reverse  
 147 Monte Carlo (RMC) is similar to Monte Carlo except it uses  $\chi^2$  as the PES.[25]

148 Despite the utility of RMC, and its wide use in the x-ray scattering community,  
 149 as Hoffman and Gelman state “Not all MCMC [Markov Chain Monte Carlo] algo-  
 150 rithms are created equal”.[15] RMC, similar to standard Monte Carlo simulations,  
 151 samples from the PES at random, usually by translating atoms in the system ran-  
 152 domly. This creates a less efficient, random walk based, exploration of the PES.[15, 26]  
 153 Thus, methods for suppressing this random walk nature, while still searching the po-  
 154 tential energy surface fully are needed.

## 155 Hamiltonian Monte Carlo

156 Hybrid or Hamiltonian Monte Carlo (HMC) can help to address some of these issues.  
 157 HMC was developed originally in the lattice quantum chromodynamic community  
 158 and provides a more efficient, more scalable approach to PES sampling for Monte  
 159 Carlo.[10, 27] In HMC the PES is explored using Hamiltonian dynamics, essentially  
 160 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:

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

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

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

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

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

163 where  $\frac{\partial}{\partial q_{i,w}}$  is the derivative with respect to  $q$  where  $w$  denotes direction of the deriva-  
164 tive ( $x$ ,  $y$ , or  $z$ ),  $n$  is the number of atoms and  $U$  is the potential which depends on  $q$ ,  
165 and  $\vec{\mathcal{F}}_i$  is the "force" on the  $i$ th atom. Using these equations new potential configura-  
166 tions are proposed from the PES. These proposals are checked against the standard  
167 Metropolis criteria discussed above, except that the change in potential energy  $\Delta E$   
168 is replaced with the change in the Hamiltonian  $\Delta \mathcal{H}$ . Note that while this sampling  
169 closely simulates the canonical ensemble, it is not exactly the same. Usually the  
170 canonical ensemble is formulated as microcanonical ensembles in contact with an  
171 infinite heat bath at a given temperature, or a set of microcancoical ensembles which  
172 exchange thermal energy. However, the HMC ensemble presented here has a momen-  
173 tum bath instead of a temperature bath. One could imagine the atoms sitting in a  
174 simulation box which has walls which can toggle their thermal exchange. Initially the  
175 box starts in the momentum bath, allowing the atoms to come to equilibrium with  
176 the bath momentum. The box is then removed from the bath causing it to become  
177 adiabatic. Hamiltonian dynamics are then propagated inside the box, essentially run-

178 ning a microcanonical simulation. Once the dynamics are finished the energy of the  
179 system is checked with the Metropolis criteria and the box is reintroduced to the  
180 momentum bath and the process starts again.

181 **No-U-Turn Sampling**

182 Two parameters must be specified in HMC simulations, the step size  $\delta$  and the number  
183 of steps  $N$ . The step size is critical to the stability of the fitting procedure: with  
184 a too small  $\delta$  the simulation runs inefficiently producing structures too close to the  
185 previous, whereas with a too big  $\delta$  the linear approximation for the forces breaks  
186 down and often the simulated NP explodes. The number of steps to take during the  
187 dynamics is equally important and an inappropriate choice may result in backtracking  
188 or random walk characteristics in the simulations. In this work, we employ the No-  
189 U-Turn Sampling (NUTS) method recently proposed by Hoffman and Gelman to  
190 address this issue [15]. In the NUTS method  $\delta$  and  $N$  are dynamically computed  
191 by examining the ratio of accepted to rejected configurations as well as whether  
192 or not the simulation has started to take a U-turn. The U-Turn criteria makes  
193 certain that the simulation stops when it begins to backtrack, preventing excess  
194 computation on configurations that have very little new information to offer. The  
195 use of NUTS leaves us with two simulation parameters: the simulation temperature  
196 and the target acceptance. Hoffman and Gelman have empirically shown that the  
197 ideal target acceptance, which governs the dynamics time steps, is .65, which we have  
198 used for all of the simulations here. The simulation temperature sets the magnitude  
199 of the random starting momenta for the atoms at the beginning of each dynamics  
200 run [15].

201 **Grand Canonical Ensemble**

202 While NUTS-HMC simulations provide a system to find minima on PESs, the sim-  
203 ulation is fundamentally run in the Canonical Ensemble thus the variables in the  
204 simulation are limited to a fixed number of particles, simulation volume, and thermal  
205 energy. Fixing the thermal energy and simulation volume is not a problem, as they  
206 are not variables of interest in the final structure. However, specifying the number of  
207 atoms in the system can be problematic, as the exact number of atoms in a sample  
208 can be difficult to count or a sample could have a distribution of particle sizes. Thus,  
209 a new ensemble needs to be used to allow the number of atoms to vary as a function  
210 of the PES. This new ensemble is the Grand Canonical Ensemble.

211 **Ensemble description**

212 In the Grand Canonical Ensemble (GCE) two sets of variables are allowed to change,  
213 the atomic positions, and the total number of atoms and their associated identi-  
214 ties. These two variables are controlled by temperature, or average momentum, and  
215 chemical potential. The partition function is

$$\Xi = \sum_{N=0}^{\infty} Q(N, V, T) \exp \frac{N\mu\beta}{T} \quad (1.18)$$

216 where  $Q(N, V, T)$  is the Canonical partition function discussed above,  $\mu$  is the chem-  
217 ical potential. [?] This is translated into a Monte Carlo system, producing Grand  
218 Canonical Monte Carlo (GCMC).

219 **Grand Canonical Monte Carlo**

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

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

223 Similarly the removal of an atom has the probability:

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

224 However, both of these equations depend of the overall simulation volume and the  
225 thermal wavelength, which is undesirable as these are not really properties that we  
226 are of interest to these simulations. Thus, we roll them into the definition of the  
227 chemical potential, essentially setting the base chemical potential to counteract these  
228 effects. This makes certain that our simulation does not change if we change the  
229 overall cell volume. A GCMC move consists of creating a new atomic configuration,  
230 where an atom has been added or removed, and checking the above criteria. However,  
231 previous results have shown that this method is computationally expensive in dense  
232 liquids, and exceedingly expensive in solid materials. The long simulation times  
233 are due to the random nature of the atomic additions or removals which produce:  
234 over-tightly packed atoms, atoms in the middle of nowhere, or unphysical vacancies.  
235 These configurations are rejected by the GCMC criteria but their probability of being  
236 sampled is much higher than configurations which are lower in energy, since the  
237 number of incorrect ways to add/remove atoms is much larger than the correct ways.  
238 Thus, we have implemented methods for biasing the atomic addition positions and  
239 the atomic removals toward configurations which are more likely to be accepted.

240 **GCMC biasing**

241 The basic idea of GCMC biasing is mapping, in 3 dimensions, where an atomic  
242 addition or removal is most likely to be accepted. Thus, the simulation volume  
243 is broken up into voxels, 3 dimensional volumes which are contained by the total  
244 simulation volume, with a pre-set size. Each voxel is given a probability of being  
245 chosen for a trial insertion where the probability is:

$$P_{i,j,k} = \frac{e^{\beta\Delta U_{i,j,k}}}{\sum_{i,j,k} e^{\beta\Delta U_{i,j,k}}} \quad (1.21)$$

Figure 1.1: 2D slice of addition biasing with Rw and Lennard Jones potentials. Note the similarity of the minima positions

where  $\Delta U_{i,j,k}$  is the change in energy. However, calculating  $\Delta U_{i,j,k}$  can be particularly expensive, especially when calculating scattering from atomic positions. The computational expense can be mitigated by using a cheaper potential, if only for the evaluation of the voxel energy, as previously shown. Similar to previous work we can use the Lennard Jones potential to approximate the addition potential, lowering the computational burden. [36]

Atomic deletion follows a similar biasing procedure, calculating the energy of each atom and biasing the probability of each atom to be chosen for removal by its energy. This way atoms which add the most energy to the system are more likely to be removed.

Figure 1.1 shows an example map for atomic addition in a Au54 atom system, with an Au55 atom target.

Include figure which shows the configurational biasing map

#### 1.4 CONCLUSIONS

In this chapter we have presented the development of both PES and the statistical mechanical ensembles used to search them. We expanded the classical concept of a PES to a more general mapping from positional variable space to energy space. This expansion allowed for the implementation of experimentally derived PES, where the disagreement between experimental and computed results can be included in the PES. Common experimental PESs were discussed, and their forces derived. The implementation of various statistical mechanical ensembles, used for searching the PES for minima, was also discussed with a special focus on No-U-Turn-Sampling Hamiltonian Monte Carlo. Grand Canonical Monte Carlo was also discussed, with

269 an emphasis on the us of biasing to increase the overall acceptance rate. Future  
270 work in this area may include the development of PESs which leverage 2 dimensional  
271 data, like STEM images, or ensembles which help to eliminate tuned parameters like  
272 parallel tempering.

273

## CHAPTER 2

274

### ATOMIC PAIR DISTRIBUTION FUNCTION:

275

### THEORY AND COMPUTATION

276 2.1 INTRODUCTION

277

Need some introduction about why we want to use the PDF

278 2.2 THEORY

279 To properly understand the PDF and its limitations we need to derive its mathemat-  
280 ics. The PDF has been previously derived many times so it is not re-derived here.  
281 This discussion of the PDF and its gradients use the notation of Farrow and Billinge.  
282 [13]

283 **Derivation**

284 Many of the above techniques require the gradient of the PES. This in turn requires  
285 the gradient of the PDF to be derived. Mathematically treating thermal vibrations will  
286 also be discussed in this section. Systems which are truly extended materials, like  
287 powders with particle sizes larger than 10nm, are best formulated as systems with  
288 periodic boundaries. Thus, the equations for a periodically bound PDF need to be  
289 developed as well, with their gradients.

290 **Analytically Gradients**

291 Many optimization algorithms and simulations methodologies, including HMC, re-  
 292 quire not only the potential energy of a given configuration but also the forces acting  
 293 on that configuration. These forces are described by the gradient of potential energy  
 294 of the system which in turn requires the gradient of the PDF. As previously shown the  
 295 PDF is the Fourier Transform of the Debye equation. Since the Fourier Transform is  
 296 expressed as an integral we can exchange the order of the gradient and the integral,  
 297 allowing us to calculate the analytical gradient of the Debye equation and FFT the  
 298 resulting function. The Debye equation, with a Debye-Waller vibrational correction  
 299 is

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

300 where

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

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

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

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

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

301 where  $Q$  is the scatter vector,  $f_i$  is atomic scattering factor of the  $i$ th atom, and  $r_{ij}$  is  
 302 the distance between atoms  $i$  and  $j$  and has  $q$  dependence. [17] For simplicities sake  
 303 we will break up  $F(Q)$  so that

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

304 where

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

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

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

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

305 The derivatives are as follows:

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

306 where

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

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

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

308 Thus

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

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

309 **Without ADPs**

310 Without ADPs the equations simplify down to

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

311 and

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

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

314 **Periodic Boundary Conditions**

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

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

321 where

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

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

322 where  $\gamma_i$  is the number of copies of the simulation box in the  $i$ th direction, and  $\vec{a}, \vec{b}, \vec{c}$   
323 are the lattice or superlattice directions.

324 **2.3 COMPUTATION**

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

329 **HPC and GPUs**

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

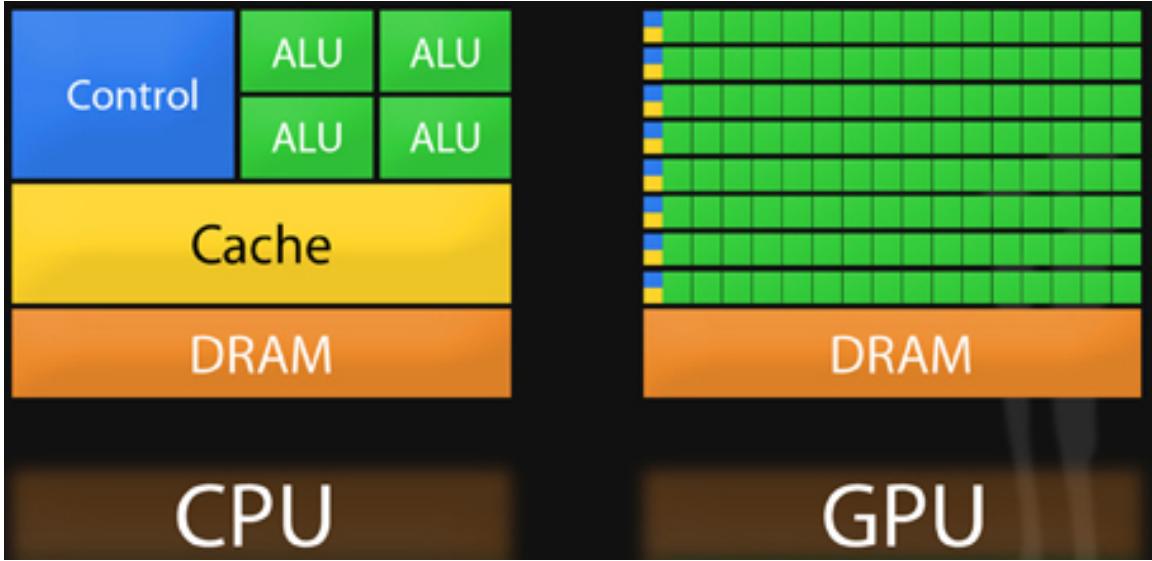


Figure 2.1: Comparison of the CPU and GPU chip architectures from [3]. The ALU are the arithmetic logic units which perform the mathematical operations, the DRAM holds most of the data, although it is slower to access, the Cache holds rapidly accessed data, and the Control controls the execution of software. Note the greater number of ALUs on the GPU, the comparatively smaller cache, and the allocation of caches and controls to entire rows of processors.

### 334 GPUs and Parallelization

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

342 Graphical Processing Units (GPUs) are particularly well suited to the task of com-  
 343 puting PDFs. GPU chip architecture is designed to perform many tasks simultaneously  
 344 by having potentially thousands of cores. Figure 2.1 shows the comparison of CPU  
 345 and GPU architectures. As the figure shows the GPUs have a very different layout  
 346 of computational processors (ALUs) and memory. While each ALU is simpler on the

347 GPU, requiring the instructions to be less demanding in terms of memory, there are  
 348 many more of them. The greater number of processors allows each atomic pairing  
 349 to be placed on its own processor, so long as the math can be broken into simpler  
 350 operations. The equations are broken up on the gpus into various pieces which cor-  
 351 respond to the  $\alpha, \beta, \tau$  and  $\Omega$  as shown in equations 2.8-2.11 and subequations as  
 352 needed. For example, while  $\beta$  is computed in one step,  $\Omega$  requires the computation  
 353 of the displacement array, then the distance array and finally the  $\Omega$  array. The exact  
 354 breakdown of processes, how the problems are broken down and spread across the  
 355 processor has been optimized for speed and reliability.

### 356 Map from ij space to k space

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

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

360  
 361  $E$  denotes the atomic coordinates in  $ij$  space,  $E'$  denotes  $F(Q)$  before the summation  
 362 in  $ij$  space,  $B$  denotes the atomic pairs in  $k$  space,  $B'$  denotes  $F(Q)$  in  $k$  space, and  
 363  $Z$  denotes the final summed  $F(Q)$ . For the operators,  $\phi$  denotes the mapping from  
 364  $ij$  space to  $k$  space  $k = j + i * \frac{i-1}{2}$ ,  $\psi$  and  $\psi'$  denote the  $F(Q)$  operation in  $ij$  and  $k$   
 365 space, respectively.  $\Sigma$  denotes the sum over all the atoms.

366 To properly define  $\Sigma'$  we must establish whether  $F(Q)$  is an even function. We  
 367 can accomplish this by examining each of the portions of  $F(Q)$ ,  $\alpha, \beta, \tau, \Omega$ .  $\Omega$  is even,

368 since  $r_{ij}$  is the interatomic distance, which is the same despite a flip of indicies,  $Q$   
 369 does not depend on the atomic indicies, and since  $Qr_{ij}$  is even so is  $\sin Qr_{ij}$ . Thus,  
 370  $\Omega$  is even. Providing similar analysis to  $\tau$  we can see that while  $\vec{u}_{ij}$  is odd, so is  
 371 the unit displacement vector between the two atoms, thus the two odds cancel out.  
 372 Intuitivly this makes sense, since the  $F(Q)$  equation is fundamentally interested in the  
 373 interatomic distances which is even. Thus, switching atom indicies does not change  
 374  $F(Q)$ . Due to the even nature of the  $F(Q)$  operator the  $\Sigma'$  operator sums over all the  
 375 atom pairs, and multiplies by two to reflect the double counting of the  $\Sigma$  operator.

For the gradient a similar mapping is used:

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

376

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

## 382 GPU Memory Allocation

383 While GPUs are very fast computational engines they tend to be memory bound.  
 384 While a gradient array for a 10nm Au nanoparticle, consisting of 31,000 atoms and  
 385 half a billion unique distances, occupies 1.5 TB of memory a single GPU's RAM  
 386 allotment varies from 4GB on a NVIDIA GTX970 to 24 GB on a NVIDIA Tesla K80.  
 387 Thus, it is important to determine exactly how many atoms can fit on a GPU of  
 388 arbitrary size as a funciton of the number of atoms and the  $Q$  range. The memory

389 required per array is:

$$q[=]3n \quad (2.22)$$

$$d[=]3k \quad (2.23)$$

$$r[=]k \quad (2.24)$$

$$scatter[=]nQ \quad (2.25)$$

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

$$\Omega[=]kQ \quad (2.27)$$

$$F_k(Q)[=]kQ \quad (2.28)$$

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

$$Sum2[=]kQ \quad (2.30)$$

$$F(Q)[=]Q \quad (2.31)$$

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

$$k_{perGPU} = \frac{1}{16Q + 16} (-4Qn - 4Q + am - 12n) \quad (2.32)$$

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

$$adps = 3n \quad (2.33)$$

$$\sigma = k \quad (2.34)$$

$$\tau = kQ \quad (2.35)$$

397 Thus the pair allotment is:

$$k_{perGPU} = \frac{-4Qn - 4Q + am - 24n}{20Q + 20} \quad (2.36)$$

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

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

$$\vec{\nabla}F_k(Q) = 3kQ \quad (2.38)$$

$$\vec{\nabla}F_n(Q) = 3nQ \quad (2.39)$$

400   Thus the gradient allotment is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 12n}{32Q + 16} \quad (2.40)$$

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

$$\vec{\nabla}\tau = 3kQ \quad (2.41)$$

402   Thus the allocation is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 24n}{48Q + 20} \quad (2.42)$$

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

## 406   Speed and Scaling of PDF Computation

407   To understand exactly how much the GPUs speed up the computation of  $F(Q)$  and  
 408   the PDF a series of time studies were run Au nanoparticles of varying size. Figure 2.2  
 409   shows the results of these time studies. CPU and GPU calculations were carried out  
 410   on an Intel i7-4820K @3.70GHz Quad-Core and one Nvidia GTX970s, respectively.  
 411   The  $F(Q)$  computations show a 100x to 10x speedup using the GPUs over the CPUs.  
 412   Additionally, the  $\vec{\nabla}F(Q)$  and  $F(Q)$  computations seem to have similar computation  
 413   time and scaling relationships on the GPU. This implies that the two processes

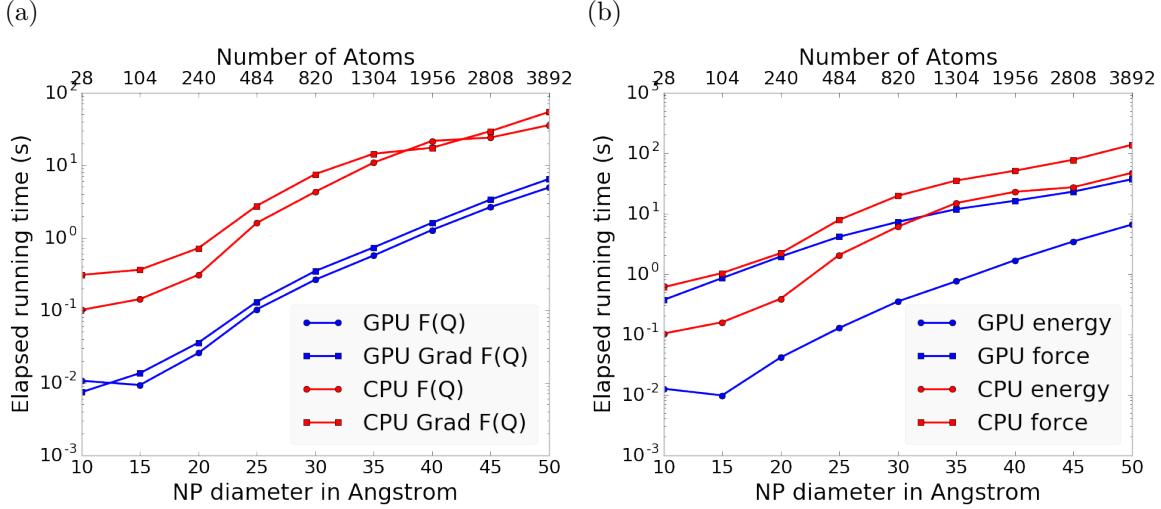


Figure 2.2: Speed comparison of CPU and GPU implementations. a) shows the time to compute the  $F(Q)$  by itself. b) shows the time to compute the  $Rw$  based energy for Au NPs of various sizes, which includes computing  $F(Q)$ , its FFT, and the  $Rw$ .

414 may have similar bottlenecks, most likely in the  $F(Q)$  computation workflow. This  
 415 relationship is similarly preserved, although to a lesser extent, in the CPU scaling.

416 Interestingly, the tight runtime relationship between  $F(Q)$  and its gradient are  
 417 not preserved in the  $Rw$  based force calculations. While the energy calculations are  
 418 very similar to the  $F(Q)$  calculations in terms of runtime, the GPU and CPU force  
 419 calculations are much closer, with the GPU calculations being much slower. This is  
 420 due to the force bottleneck being the  $3n$  FFT operations which must be performed  
 421 on the  $\vec{\nabla}F(Q)$  array to produce the  $\vec{\nabla}PDF$  array. While the GPU is leveraged  
 422 to perform the FFT, the data must be loaded off the GPU and back on, causing a  
 423 potential slowdown. Larger systems of atoms were not tried as the CPU computation  
 424 quickly becomes very slow. Even higher GPU speedup is expected on more advanced  
 425 GPUs like the Nvidia Tesla series.

426 2.4 CONCLUSIONS

427 In this chapter we developed the gradients of the PDF in the discrete and periodic  
 428 boundary condition case. We also developed the computational implementation of

429 the PDF equations. This implementaiton emphasized use of GPUs to compute the  
430 PDF and its gradient. The GPU software was futhur sped up by mapping the com-  
431 putation to atom pairs rather than atom by atom. Finally, the speed of the GPU  
432 implementation was checked against the CPU implementation via speed benchmark-  
433 ing.

434

## CHAPTER 3

435

### BENCHMARKING

436

Also some introduction would be great

437

this just needs a lot of work

438 3.1 INTRODUCTION

439 The NUTS-HMC system was tested on a series of nanoparticle (NP) benchmarks.  
440 The purpose of these benchmarks is to test the ability of the NUTS-HMC system to  
441 reproduce the target PDF and its associated structure. Systems were chosen for their  
442 size, crystallinity, and interfacial differences.

443 3.2 PDF

444 The formation of NPs with both crystallographic and non-crystallographic structures  
445 [23] and with different chemical patterns [14] are well documented. For simplicity,  
446 we chose monometallic Au clusters as benchmarks and considered two groups of  
447 structures with different size and degrees of structural disorder in order to assess  
448 the reliability and efficiency of our HMC method for solving atomic structures from  
449 PDFs. The first group consists of  $\text{Au}_{55}$  clusters with different degrees of disorder,  
450 including a crystalline cluster structure in  $O_h$  (Octahedral) symmetry, a structure  
451 with a disordered surface, and an amorphous structure. The second group consists  
452 of the crystallographically solved  $\text{Au}_{102}$  structure as in the  $\text{Au}_{102}\text{MBA}_{44}$  nanocrystals  
453 [16, 22]. We used optimized structures from the Density Functional Theory (DFT)

454 as target structures and generated the corresponding PDF,  $G_{\text{obs}}$ , according to

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

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

460 Spin-polarized DFT calculations were carried out using the Vienna ab initio simu-  
461 lation package (VASP) [20, 21] within the Perdew-Burke-Ernzerhof (PBE) exchange-  
462 correlation functional [29]. The projected augmented wave method [2] and a kinetic  
463 energy cutoff of 400 eV were used. Structural optimization was performed until the  
464 total energy and ionic forces were converged to  $10^{-6}$  eV and 10 meV/Å, respectively.  
465 The amorphous  $\text{Au}_{55}$  structures were generated by simulated annealing using the  
466 classical embedded atom method potential [35]. Different annealing temperatures  
467 between 1200 K and 1670 K (bulk melting temperature of Au) were used and the  
468 thermally equilibrated structures were cooled down to 300 K before minimization at  
469 0 K. Further optimization using DFT leads to total energies that vary within 1-2  
470 eV among different amorphous structures and the lowest energy one was used as the  
471 target structure. The target structure of  $\text{Au}_{102}$  was taken as the  $\text{Au}_{102}$  core of the  
472 DFT-optimized  $\text{Au}_{102}\text{MBA}_{44}$  cluster [22].

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

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

476 The thresholded spring potentials are based on those previously proposed on by Pe-  
477 terson [30], 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

478 bounds of the spring threshold  $r_t$ . The resulting restoring forces on the out-of-bound  
479 atoms bring the system back within the bounds of the PDF,  $R_{\min}$  and  $R_{\max}$ , and  
480 therefore preventing the system from exploding or collapsing. Otherwise, incorrect  
481 refinements may result by having atomic pair distances out of the PDF bounds.  $\kappa$  is  
482 the spring constant in eV/Å and the  $Rw$  potential is converted from unitless to eV  
483 via multiplication by a conversion factor  $\lambda$ .

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

## 496 Model Parameters

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

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

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

## 508      **Au55: surface relaxed**

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

## 516      **Run Parameters**

517       $R_{\min}$  and  $R_{\max}$  for this simulation were 2.45 Å and 11.4 Å, respectively, with  $\delta Q =$   
518      0.24 Å<sup>-1</sup>. The simulation ran for approximately 34 minutes, over a total of ~40  
519      thousand configurations. The results are shown in Fig. 3.1.

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

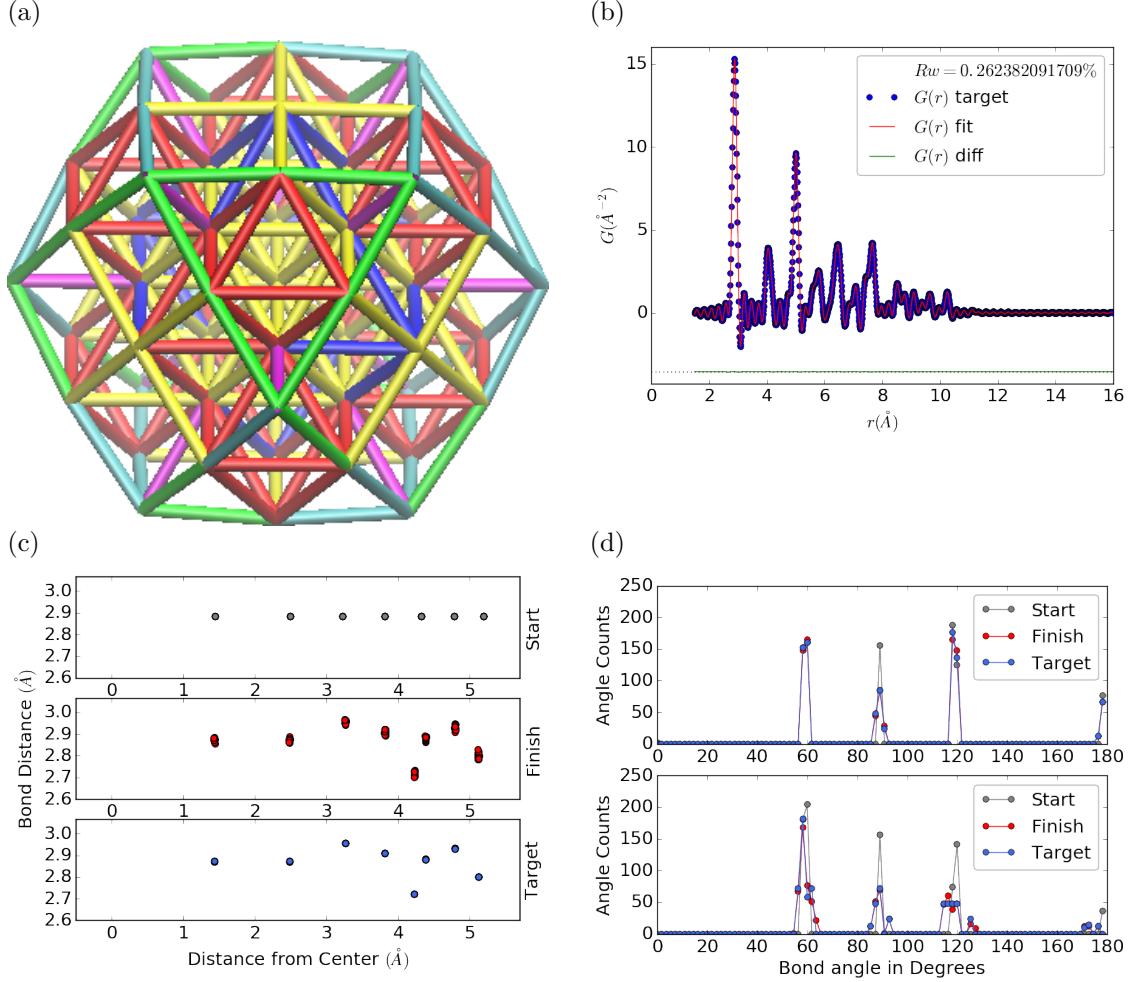


Figure 3.1:  $\text{Au}_{55}$  PDF fitting of DFT-optimized  $c\text{-Au}_{55}$ . (a) the final structural solution ( $Rw=0.3\%$ ) with bond lengths color-coded by step of  $0.05\text{\AA}$ , (b) the target PDF(blue dots) overlaid with the PDF of the final structure (solid red lines) with the difference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

527 below, the situation is rather different for much more disordered target structures.  
 528 Interestingly, the small-box solution using PDFgui[12] yields a rather large  $Rw$  of  
 529 43%, due to the failure to fit the surface contracted atoms with a unit cell. The PDF  
 530 fits of the starting structure and small-box solution are shown

531 Put this somewhere

532 .

533 **Au55: surface disordered**

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

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

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

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

560 and high CN numbers. This is partly caused by the displacement of the core atoms  
561 mentioned above, and partly by the lack of CN constraints for the PDF fitting, which  
562 has been previously demonstrated in the case of  $\alpha$ -Si [7]. Additional experimental  
563 data, e.g. from EXAFS or NMR, may help to steer the simulations towards better  
564 agreement in both PDF and CN distribution.

## 565 Au55: amorphous

566 Next, we turn to the case in which the entire cluster structure is disordered. We used  
567 a DFT-optimized amorphous  $\text{Au}_{55}$  ( $a\text{-Au}_{55}$ ) as the target structure, and the DFT-  
568 relaxed  $c\text{-Au}_{55}$  cluster from Case I as the starting structure. The total energy of  
569  $a\text{-Au}_{55}$  was computed to be *lower* than that of  $c\text{-Au}_{55}$  by as large as 2.9 eV, consistent  
570 with the 3.0 eV found in previous DFT work [9].

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

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

586 **Au102: triple phase**

587 Our final benchmark is Au<sub>102</sub>, whose structure was initially solved by Jadzinsky and  
588 co workers using x-ray crystallography [16] and further confirmed by DFT studies  
589 [22]. The Au<sub>102</sub> structure consists of three main parts, a 49-atom Marks decahedron  
590 core, two  $C_5$  caps consisting of 20 atoms each, and 13 equatorial atoms. Unlike  
591 previous cases, the multi-symmetry nature of the structure, i.e. each part has its own  
592 distinct symmetry, poses a challenge for PDF-based solution of the structure. This is  
593 because of the atomically centralized nature of the PDF, in which each atom “sees”  
594 a density of other atoms surrounding it and has a strong tendency towards becoming  
595 the center of the main symmetry group. Such tendency may lead to a solution where  
596 some of the correct atomic symmetries are discarded in favor of the core symmetry.

597 **Starting from fcc structure**

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

600  $R_{\min}$  and  $R_{\max}$  for this simulation were 2.7 Å and 16. Å, respectively, with  $\delta Q =$   
601 0.18 Å<sup>-1</sup>. The simulation ran for approximately two hours, over a total of ~82  
602 thousand configurations. The results of the simulation are shown in Fig. 3.4.

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

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

622 **Marks decahedron**

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

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

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

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

641 **Au147**

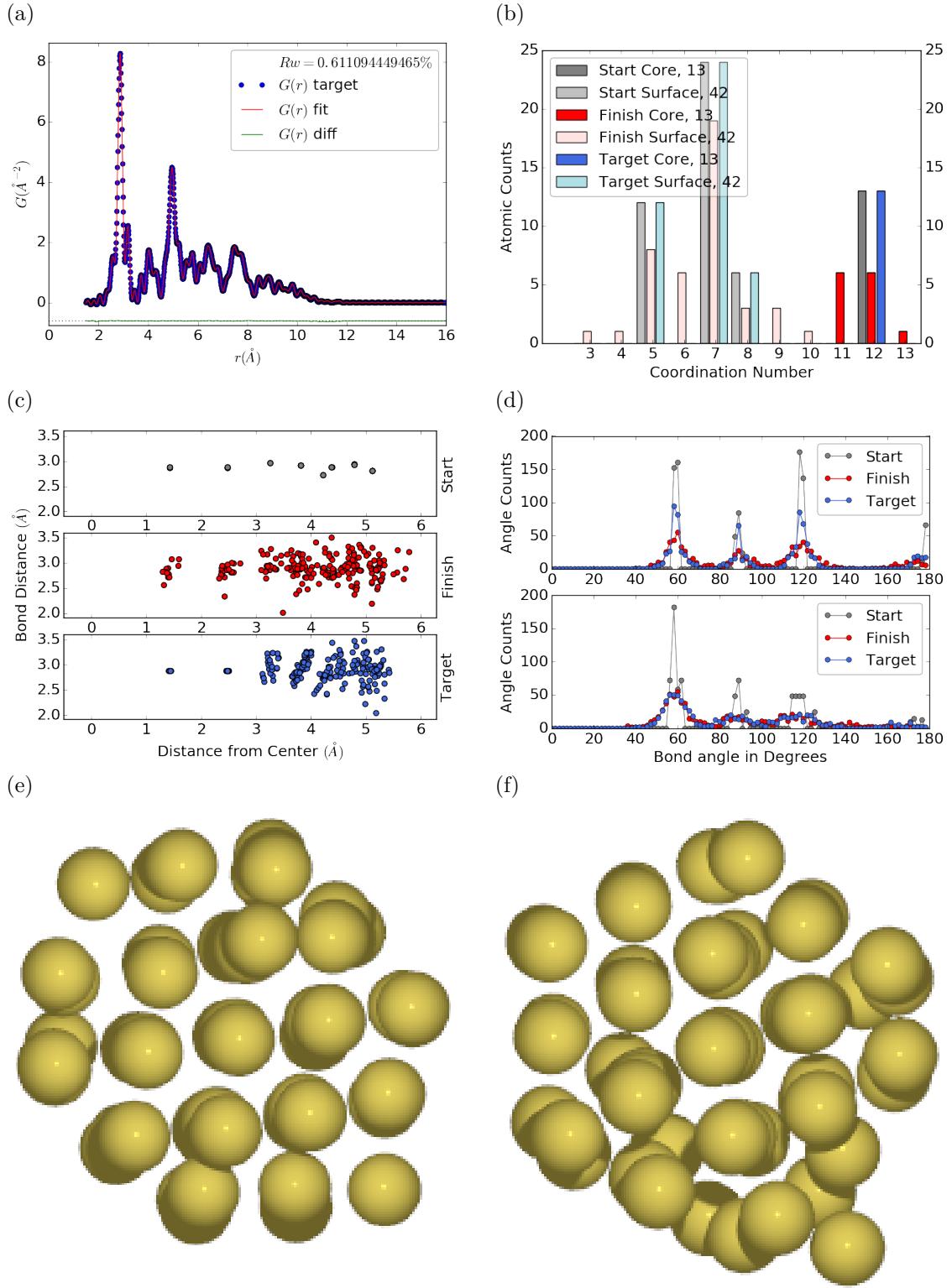


Figure 3.2:  $\text{Au}_{55}$  PDF fitting of surface-disordered  $\text{Au}_{55}$ . a) the target structure, b) the final structural solution ( $R_w=0.6\%$ ), c) the comparison of PDFs, d) the CN distribution with the number of atoms in either the core or the surface, e) the radial bond distribution,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

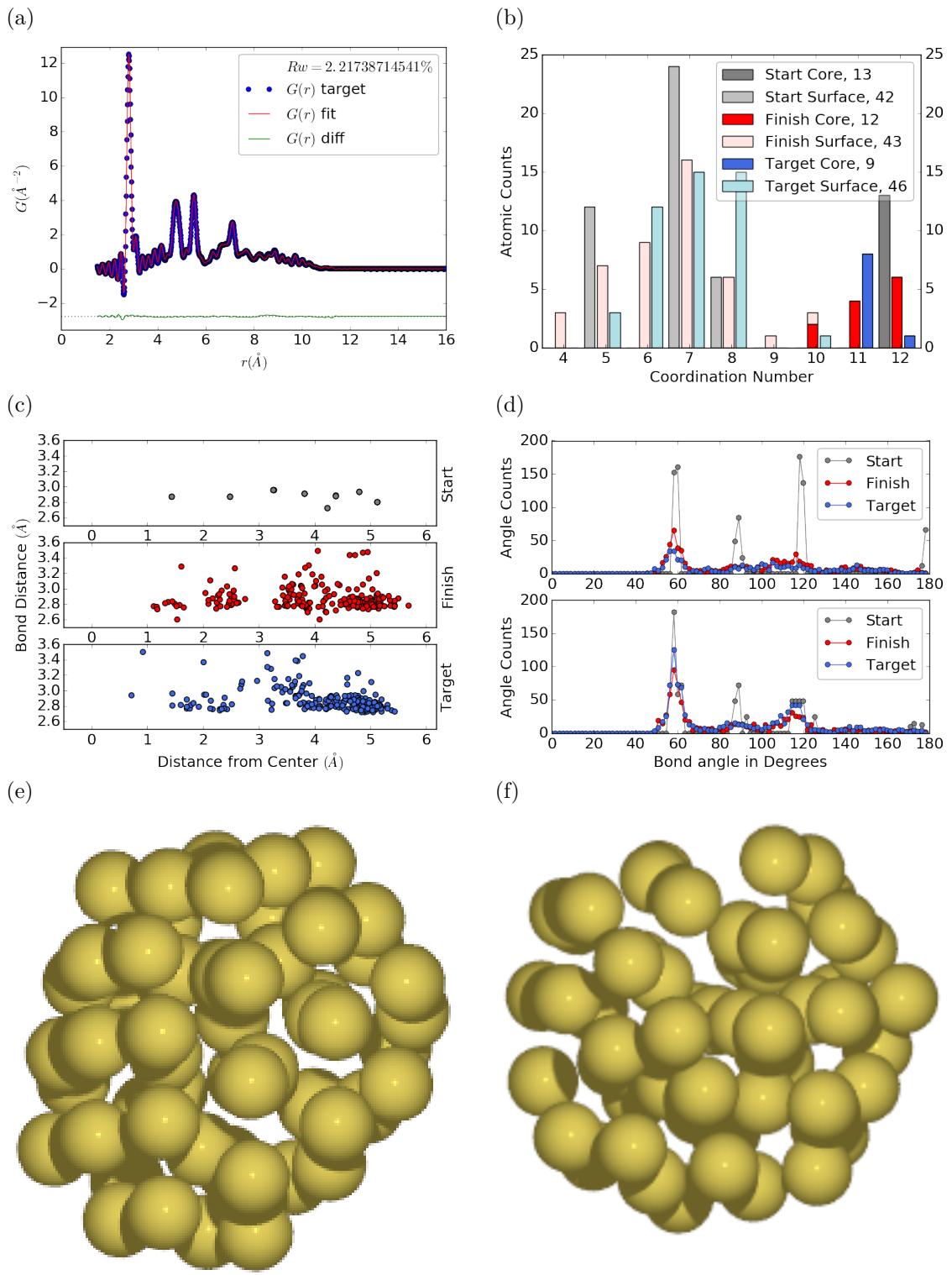


Figure 3.3: Similar to figure 3.2 for DFT-optimized amorphous Au<sub>55</sub>.

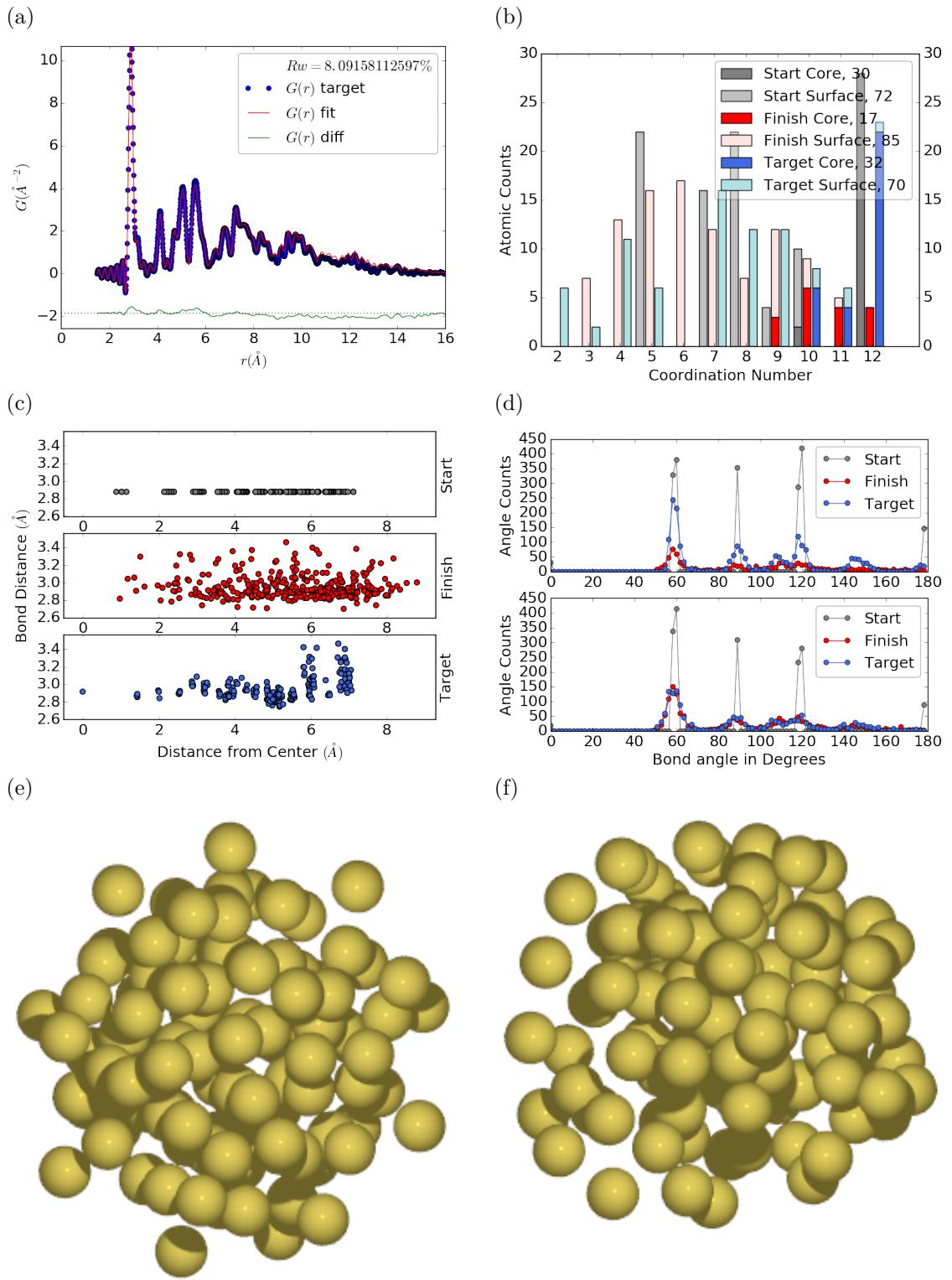


Figure 3.4: Similar to figure 3.2 for  $\text{Au}_{102}$  as in DFT-optimized  $\text{Au}_{102}\text{MBA}_{44}$  cluster.

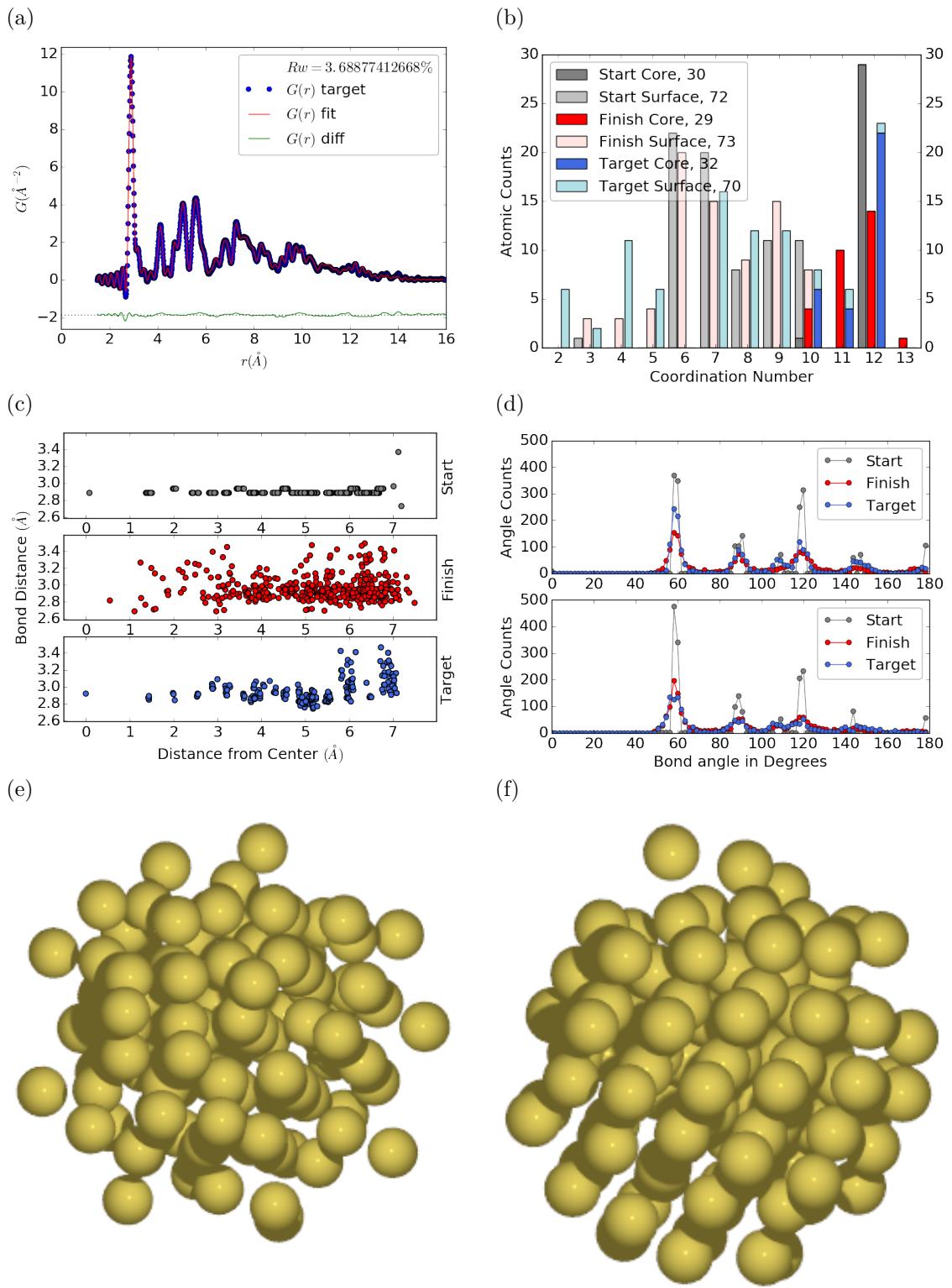


Figure 3.5: Similar to Fig. 3.4 with Marks decahedron as the starting structure.

## 642 3.3 PDF WITH ADPs

## 643 ADP 50

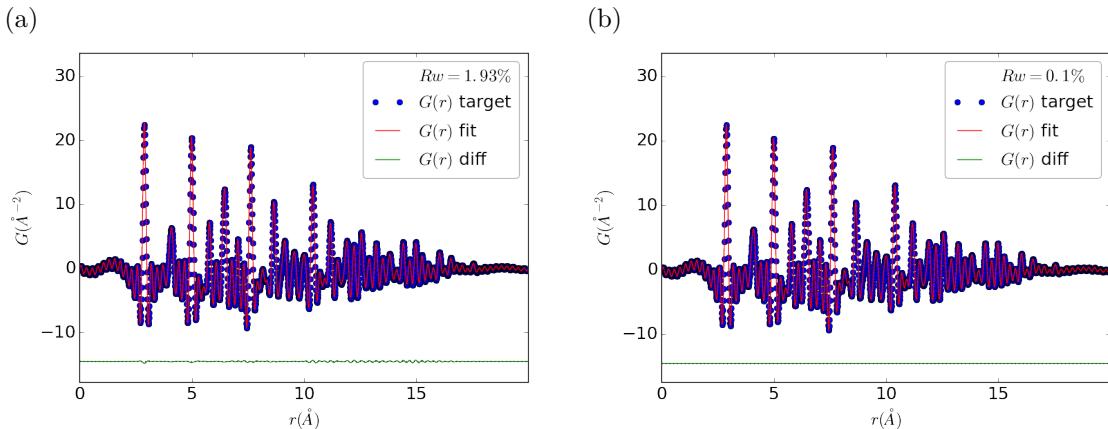


Figure 3.6: Refinement of adps

644 1. Basic 50% larger magnitude

645 2. Random addition to APDs

646 3. Janus ADPs

647

## CHAPTER 4

648

# X-RAY TOTAL SCATTERING DATA ACQUISITION AND PROCESSING

650

4.1 INTRODUCTION

X-ray total scattering experiments are generally performed at synchrotron light sources, as only these sources can provide the needed flux, energy, and high momentum transfer vectors needed to obtain reliable PDFs. [6, 11] Despite the need for a dedicated facility to perform the total scattering experiments, the experiments themselves are fairly forgiving, allowing for reactive gaseous environments, experiment temperatures ranging from 2 K to 1200 K, and even electrochemical cycling. [5, 32, 34] The rapid PDF data acquisition associated with 2D area detectors creates a data management problem, as 96 hours of beamtime could result in almost 10,000 images which need to be associated with the experimental conditions and detector metadata. [6] Finally, all this data needs to be processed by masking bad pixels and regions, integrating azimuthally, and converting the scattering data to the PDF. [19, 18, 37, 28, 1]

662

4.2 DETECTOR  $Q$  RESOLUTION

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

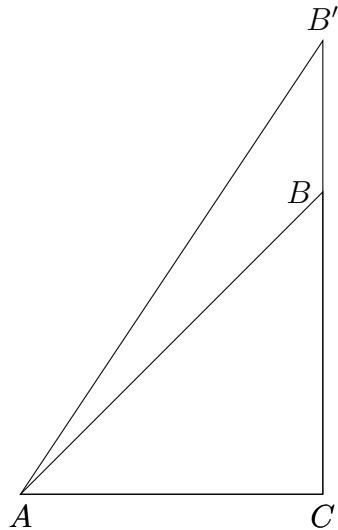


Figure 4.1: Scattering onto a flat detector

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

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

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

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

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

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

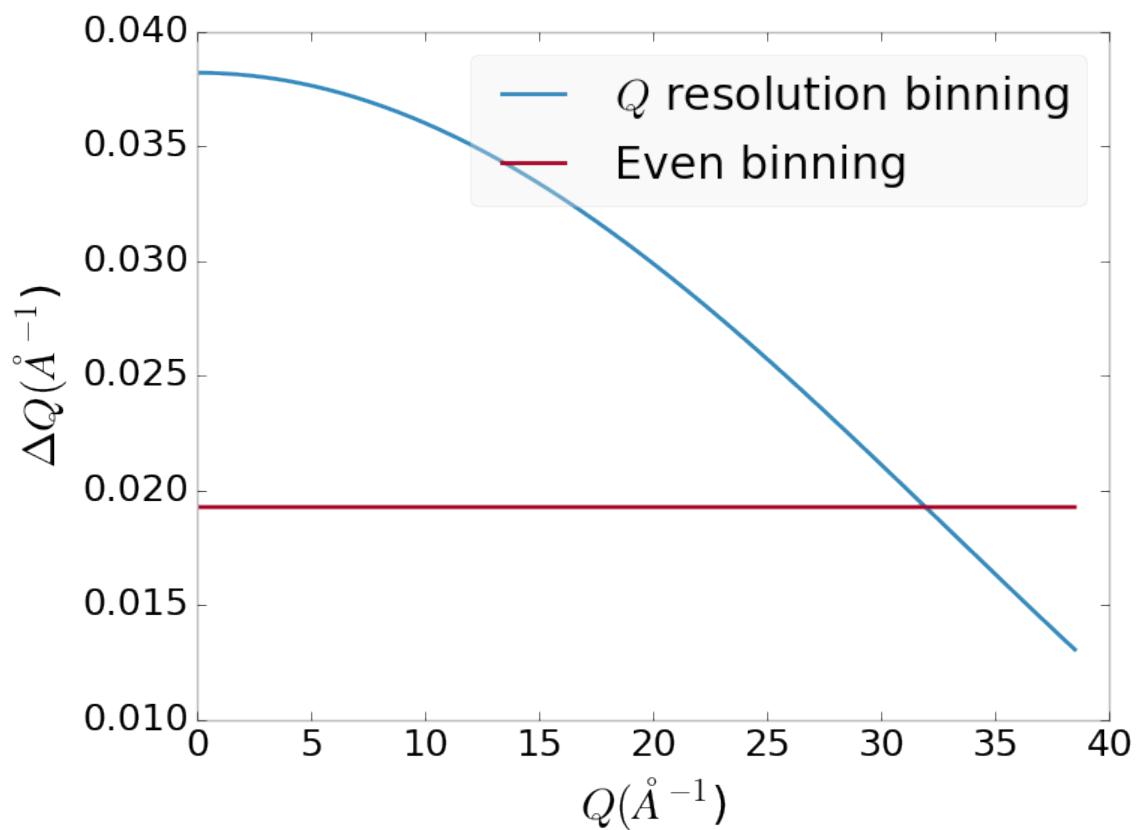


Figure 4.2:  $Q$  resolution as a function of  $Q$ .

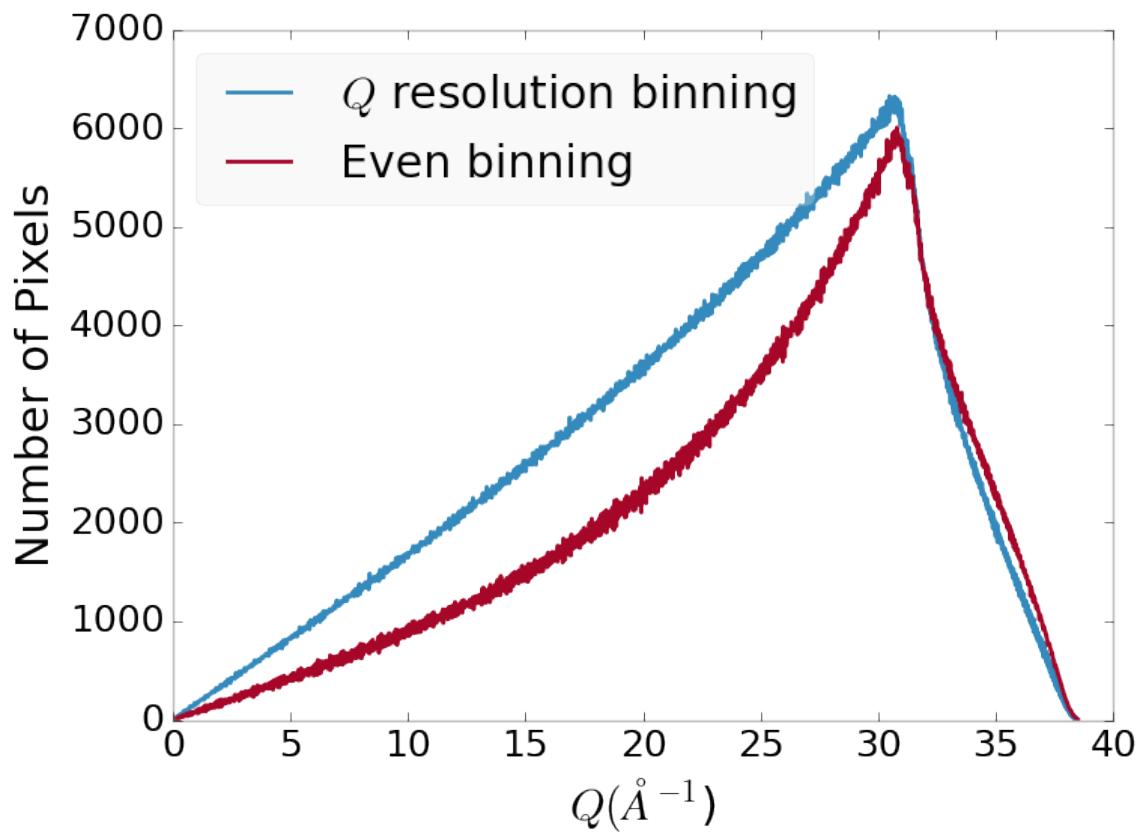


Figure 4.3: Number of pixels as a function of  $Q$ , binned at the  $Q$  resolution of the detector.

678 4.3 AUTOMATED MASK GENERATION

679 **Introduction**

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

689 While this problem can be quite complex in the most general case, we can use the  
690 annular symmetry of the powder scattering pattern to our advantage, by comparing  
691 a pixel against pixels in the same ring. Since non-textured powder scattering should  
692 produce the same pixel intensity for a given ring we can mask any pixels which are  $\alpha$   
693 standard deviations away from the mean. This method relies on the aforementioned  
694 pixel binning algorithm, as using miss sized bins will cause some pixels which should  
695 be in separate rings to be put together, and others which should be in the same ring  
696 to be separated. In that case the masking algorithm will overestimate the number of  
697 pixels to be masked due to the additional statistical variation in the sample.

698 **Algorithm Design**

699 The masking algorithm procedure takes in the image and a description of the pixel  
700 positions in either distance from the point of incidence or in  $Q$ . The image is then  
701 integrated twice, producing both the mean  $I(Q)$  and the standard deviation of each  
702  $I(Q)$  ring. The mask is created by comparing the pixel values against each ring's

703 standard deviation and threshold  $\alpha$ . Note that the threshold can be a function of  
704 distance from the point of incidence or  $Q$ .

705 **Test Cases**

706 To study the effectiveness of the masking we ran the algorithm against both simulated  
707 experimental data. In the case of the simulated data four systems were created: 1)  
708 dead/hot pixels with varying numbers of defective pixels, 2) beamstop holder with  
709 varying beamstop holder transmittance, 3) rotated beamstop holder with varying  
710 beamstop holder transmittance, and 4) beamstop holder with dead/hot pixels. The  
711 base scattering was produced by

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

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

722 **Results and Discussion**

723 Figures 4.4-4.11 show the results of the masking algorithm on simulated images. The  
724 dead/hot pixel masking shows the importance of using the  $Q$  resolution based bin  
725 sizes as the even bin based mask have a tendency to over mask the image, removing

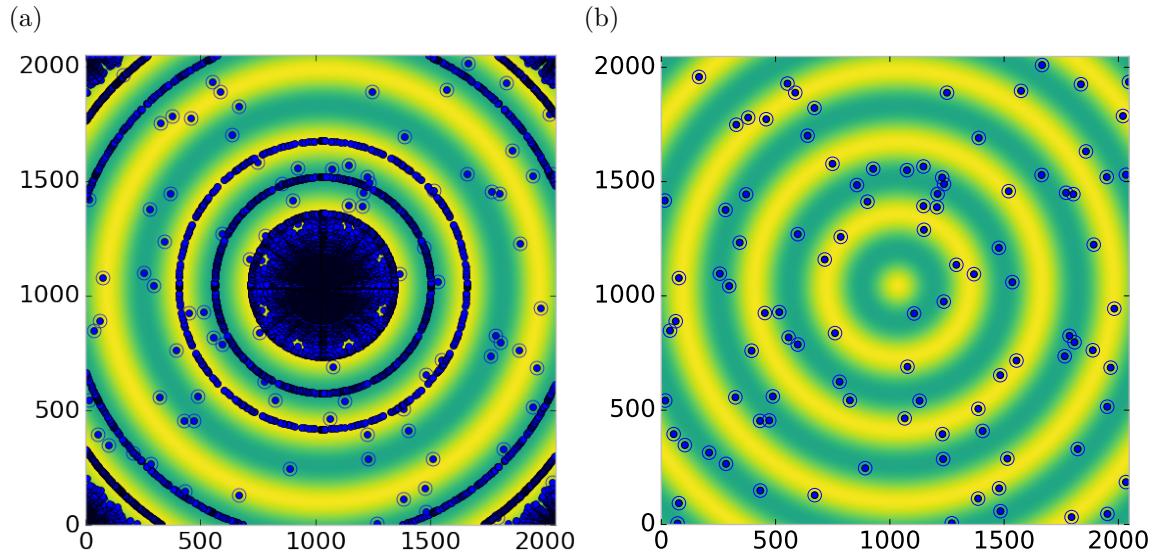


Figure 4.4: Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

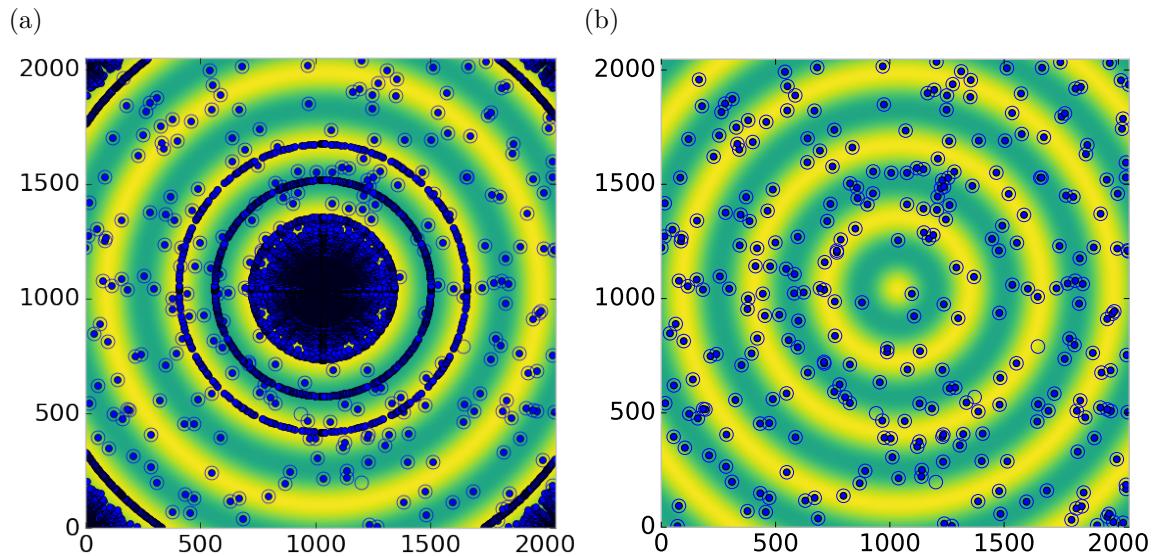


Figure 4.5: Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

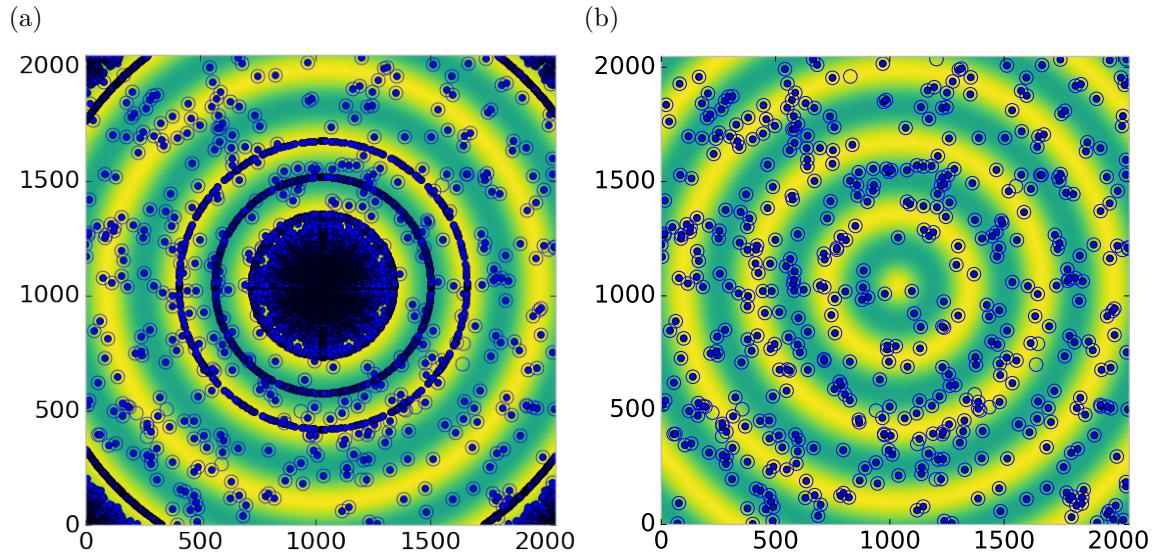


Figure 4.6: Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

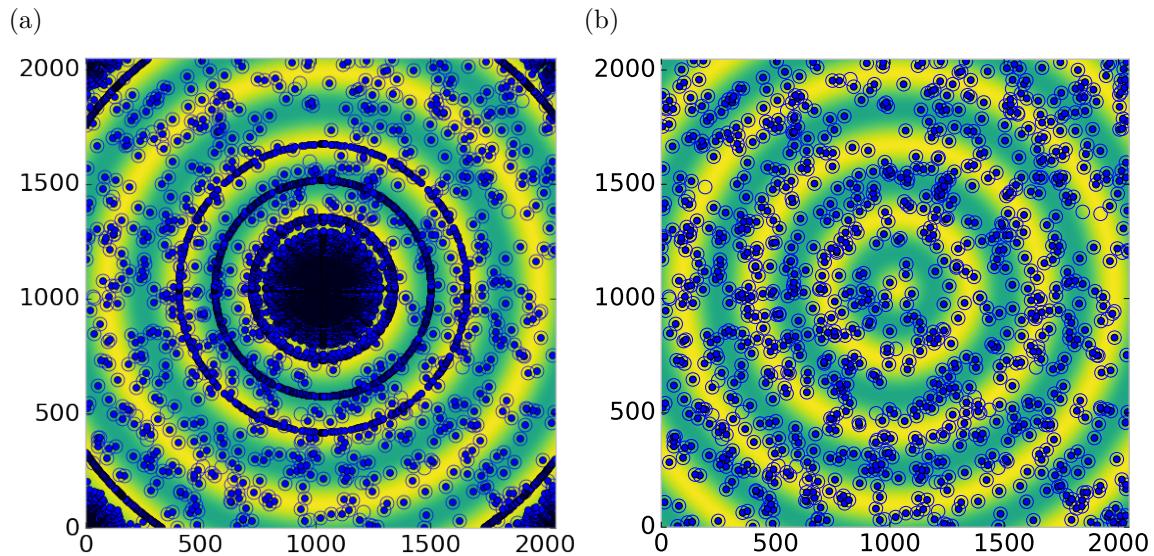


Figure 4.7: Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

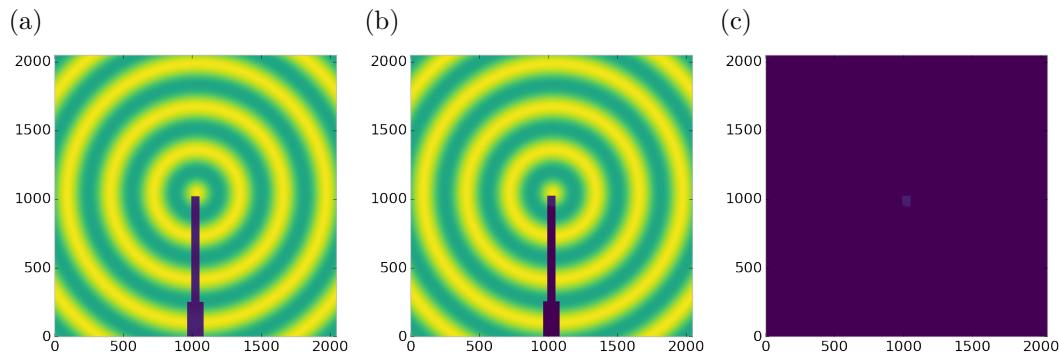


Figure 4.8: Generated beamstop holder masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

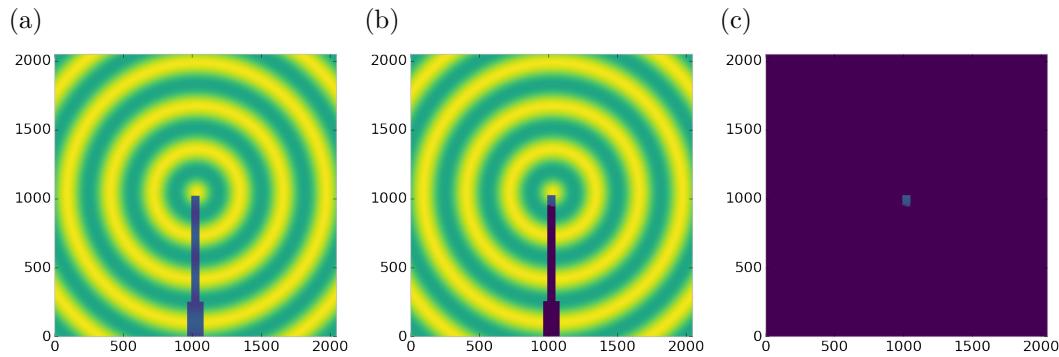


Figure 4.9: Generated beamstop holder masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

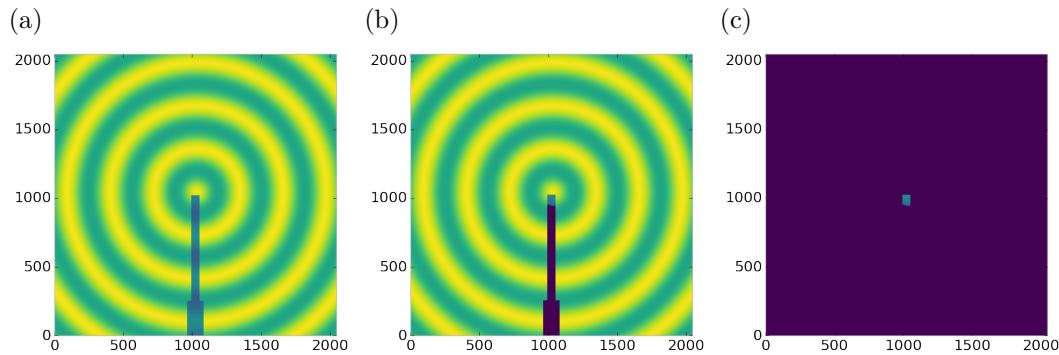


Figure 4.10: Generated beamstop holder masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

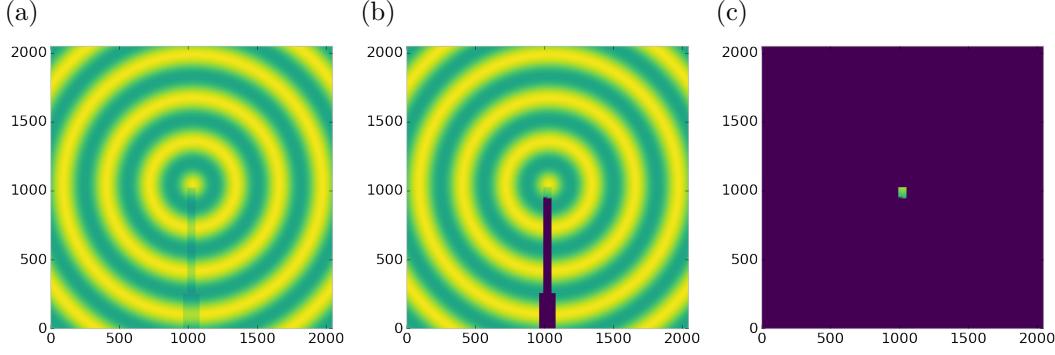


Figure 4.11: Generated beamstop holder masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

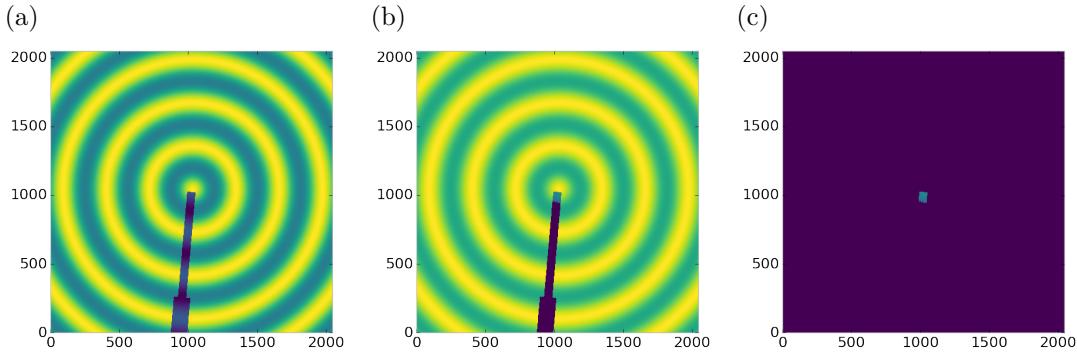


Figure 4.12: Generated beamstop holder masks which is rotated away from vertical

726 pixels which contain valuable signal. This over-masking is caused by pixels being  
 727 improperly associated with one another by the even bins. Figure 4.4 indicates that  
 728 the masking algorithm, with the proper binning, masks the image perfectly, with no  
 729 missed bad pixels or good pixels masked. This is not the case in figures 4.5 - 4.7 as  
 730 we can see pixels which should have been masked but were not. Despite these missed  
 731 pixels no pixels were improperly masked in any of the well binned images. These  
 732 test cases are actually more difficult than experimental data, as the dynamic range  
 733 of most detector causes the dead/hot pixels and single crystal/textured peaks to be  
 734 orders of magnitude away from the desired signal.

735 The beamstop holder masks shown in figures 4.8 - 4.11, which were all run with  
 736 the  $Q$  resolution binning show similar results across the transmittance range, missing  
 737 only a small part of the beamstop holder near the point of incidence. Near this point

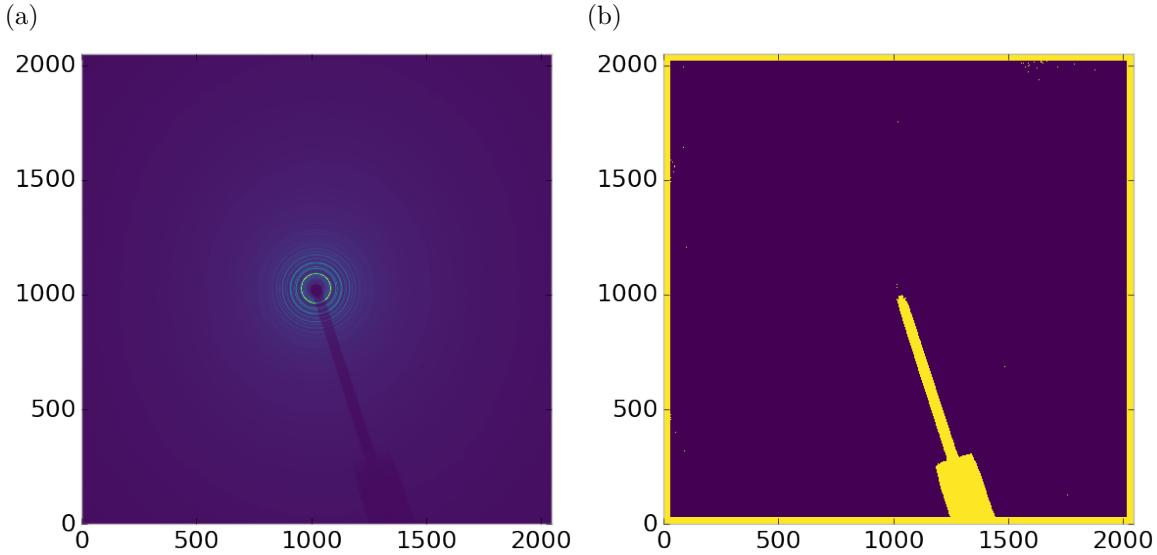


Figure 4.13: Masked experimental data. a) the raw image, b) the mask

738 the beamstop holder becomes a statistically significant part of the total number of  
 739 pixels in a given ring, thus it can not be masked out using a statistical search of the  
 740 rings. For most PDF and XRD studies this small area can be masked automatically  
 741 by masking all the pixels who's distance from the point of incidence is smaller than a  
 742 given radius  $r$ , or can be neglected outright as the area is not used in the analysis or  
 743 refinement. Similar results were produced for beamstop holders which were rotated  
 744 away from the vertical position, as shown in figure 4.12

745 Working with actual experimental data, obtained at the Advanced Photon Source  
 746 beamline 11-ID-B, shows the difficulty of masking images which have low photon  
 747 counts. While the masking of experimental data taken with longer exposures, con-  
 748 sisting of 250 .2 second shots, shown in figure 4.13 provides very sharp edges to the  
 749 beamstop holder, and very little extra masking beyond the occasional dead pixel, this  
 750 is not the case for the single crystal data. The single crystal data is more problem-  
 751 atic because of its short exposure time and low flux, with 500 frame at a .1 second  
 752 exposure and having shrunk the beam size. The low flux is to prevent the very strong  
 753 single crystal peaks from damaging the detector. However, this causes the image

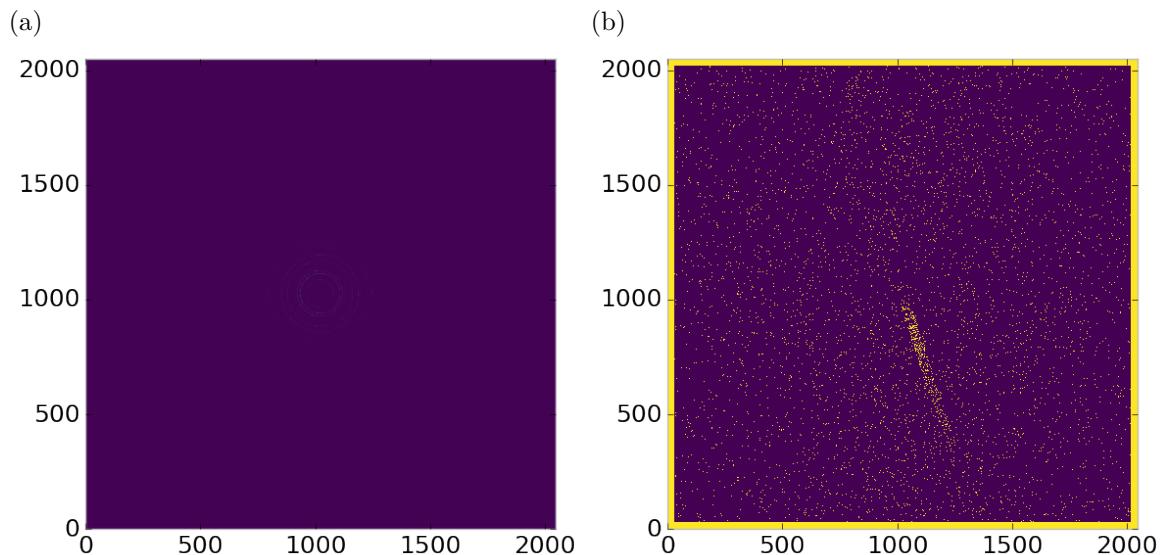


Figure 4.14: Masked experimental data with Pt single crystal signal. a) the raw image, b) the mask

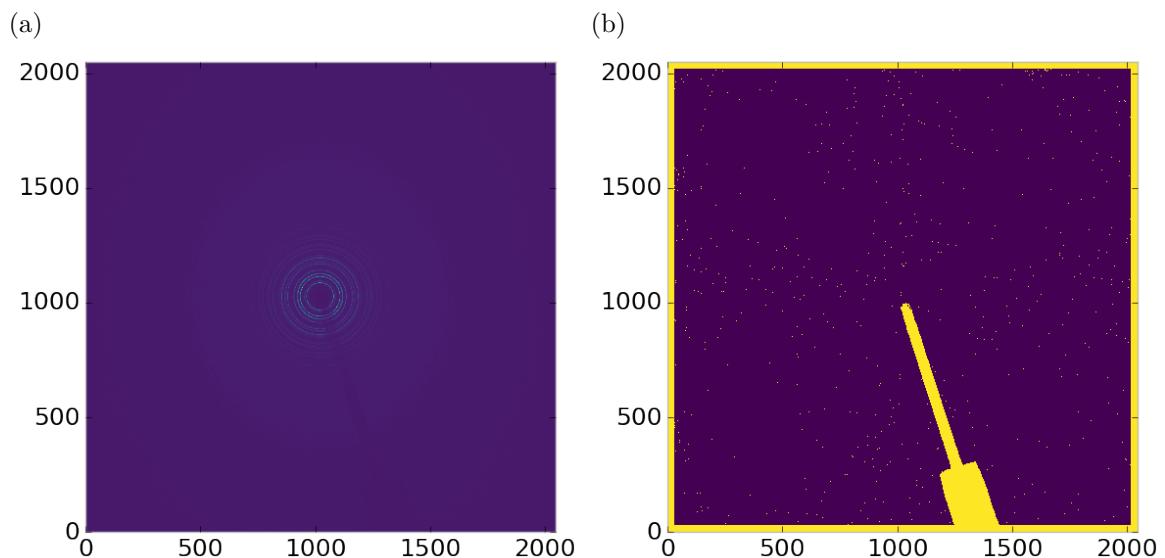


Figure 4.15: Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask. a) the raw image, b) the mask

754 to be less statistically viable then ideal, causing problems with the mask as seen in  
755 figure 4.14. This can be alleviated to some degree by using the previously generated  
756 mask as a starting mask for the single crystal image, as shown in 4.15. While the  
757 masking algorithm still produces many diffuse masked pixels, they are far fewer, this  
758 may be due to the removal of the beamstop which could have contributed to the large  
759 standard deviation in figure 4.14.

760 **Conclusions**

761 In this section the masking algorithm, which relies on both  $Q$  resolution based bin-  
762 ning and a statistical approach to azimuthal symmetry, was developed. The focus of  
763 this algorithm was to remove many unwanted detector features associated with pixel  
764 defect, beamstop holder associated scattering attenuation, and single crystal/texture  
765 peaks. Simulated data was used to evaluate the beamstop holder and dead/hot pixel  
766 masking capacity, while experimental data was used to check for single crystal and  
767 texture based masking.  $Q$  resolution based binning was shown to be very important  
768 to avoid over-masking. The ability of the mask writer to mask images is somewhat  
769 limited by the overall statistical image quality, although some deficiencies can be  
770 obtained by using previously generated masks as starting points. This masking algo-  
771 rithm is now in use in the data processing workflow and will be available in scikit-beam  
772 soon.

773 **4.4 AUTOMATED IMAGE AZIMUTHAL INTEGRATION**

774 Using the  $Q$  resolution binning and masking developed in sections 4.2 and 4.3 the  
775 images can be properly integrated. Generally, images are integrated by taking the  
776 mean value of the pixels in a ring. However, other statistical measures of the average  
777 value can be used, like the median. Note that all the integrations done here use the  
778 pixels as they are, without pixel splitting, minimizing the covariance of the resulting

779  $I(Q)$ .[37]

780 Figures 4.16-4.18 show the importance of masking and the choice of average func-  
781 tion. All the figures were produced using the same dataset, 50 °C  $\text{Pr}_2\text{NiO}_4$  taken at  
782 the APS’s 11-ID-B on a Perkin Elmer area detector. The automatic masking alpha  
783 was 3 standard deviations from the mean. While it is difficult to observe the changes  
784 the mask causes in the full  $I(Q)$  plot (subfigures a) and b)), the standard deviation  
785 plots show the effect of bad pixels on the data (subfigure c)). Subfigure c) for figures  
786 4.16-4.18 shows that removal of the beamstop holder lowers the low  $Q$  standard de-  
787 viation from around .1 to almost .01 out to  $15 \text{ \AA}^{-1}$ . The high  $Q$  subfigures d) and f)  
788 in figures 4.16-4.18 show the “kink” effect of the detector edge and beamstop holder,  
789 where there is a dip in the  $I(Q)$  scattering when the rings include the edge of the  
790 detector. This effect seems to be due to both errors in the edge pixel intensity and the  
791 beamstop holder as masking of the edges only seems to provide only partial removal  
792 of the issue. It is important to note that while integration using the mean of the  
793 ring has issues with only the edge mask, as evidenced by the change in slope in 4.17  
794 d) around  $29.5 \text{ \AA}^{-1}$ , the median integration does not include this error. Ideally the  
795 detector would have a normal distribution of pixel intensity for a given ring, which  
796 would imply an equivalency between the mean and median  $I(Q)$  values. Despite the  
797 closeness of the mean and median once the final mask has been created, it seems that  
798 the median is more reliable, as it was less effected by the beamstop holder in figure  
799 4.17. Thus, for subsequent integrations discussed in this work the median is used to  
800 avoid any defective features that the masking algorithm may have missed.

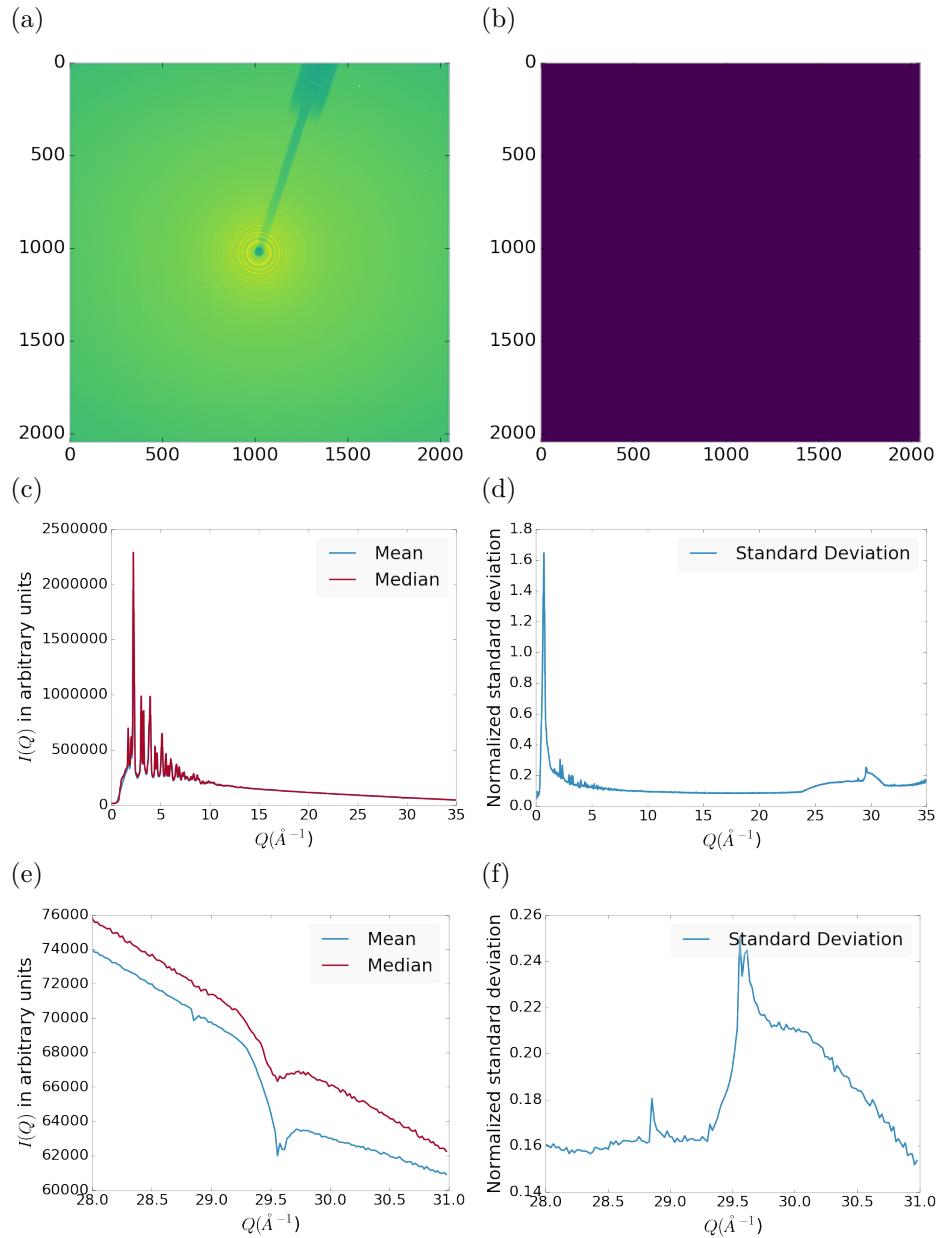


Figure 4.16: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with no mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{\AA}^{-1}$  to 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

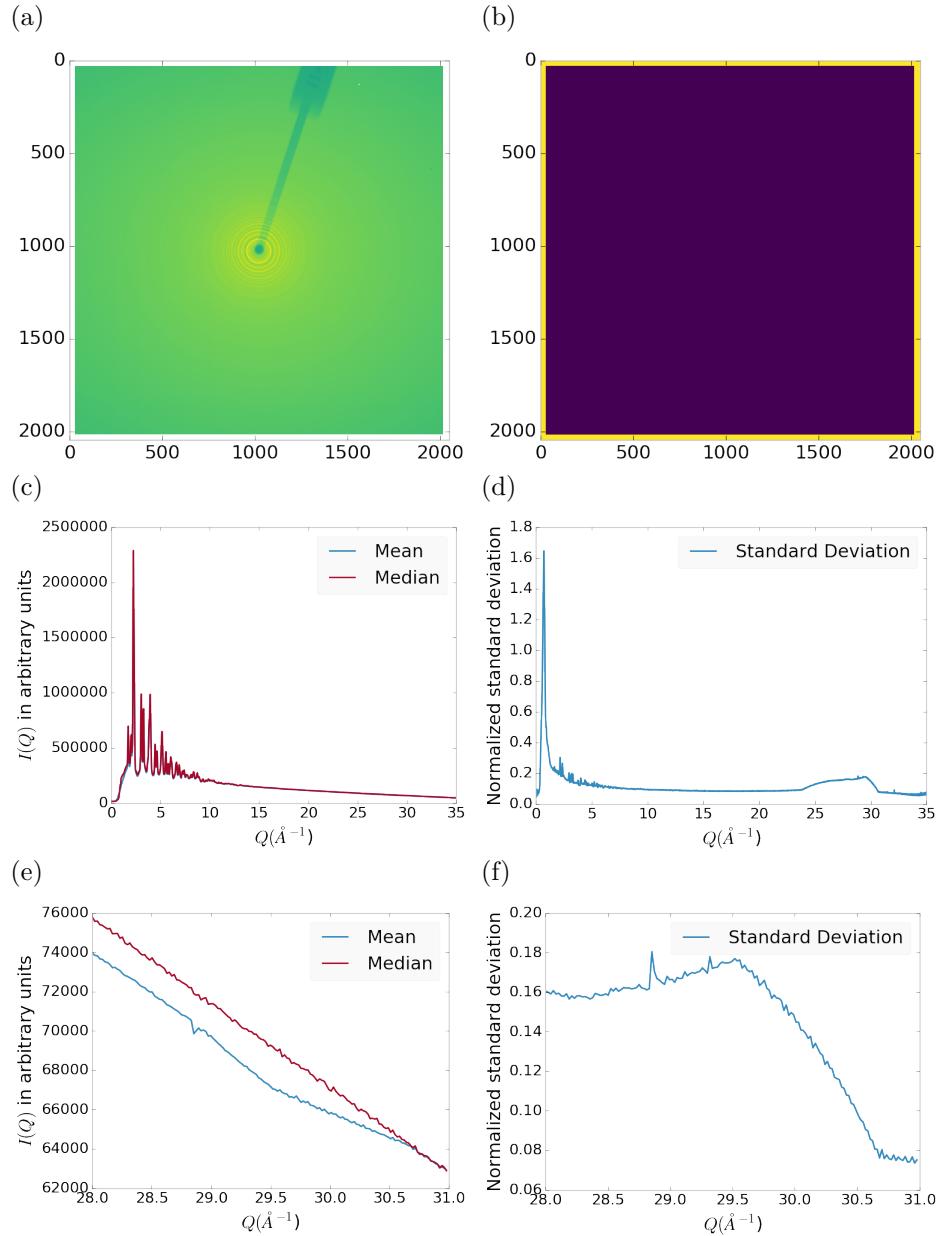


Figure 4.17: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with only an edge mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{ \AA}^{-1}$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

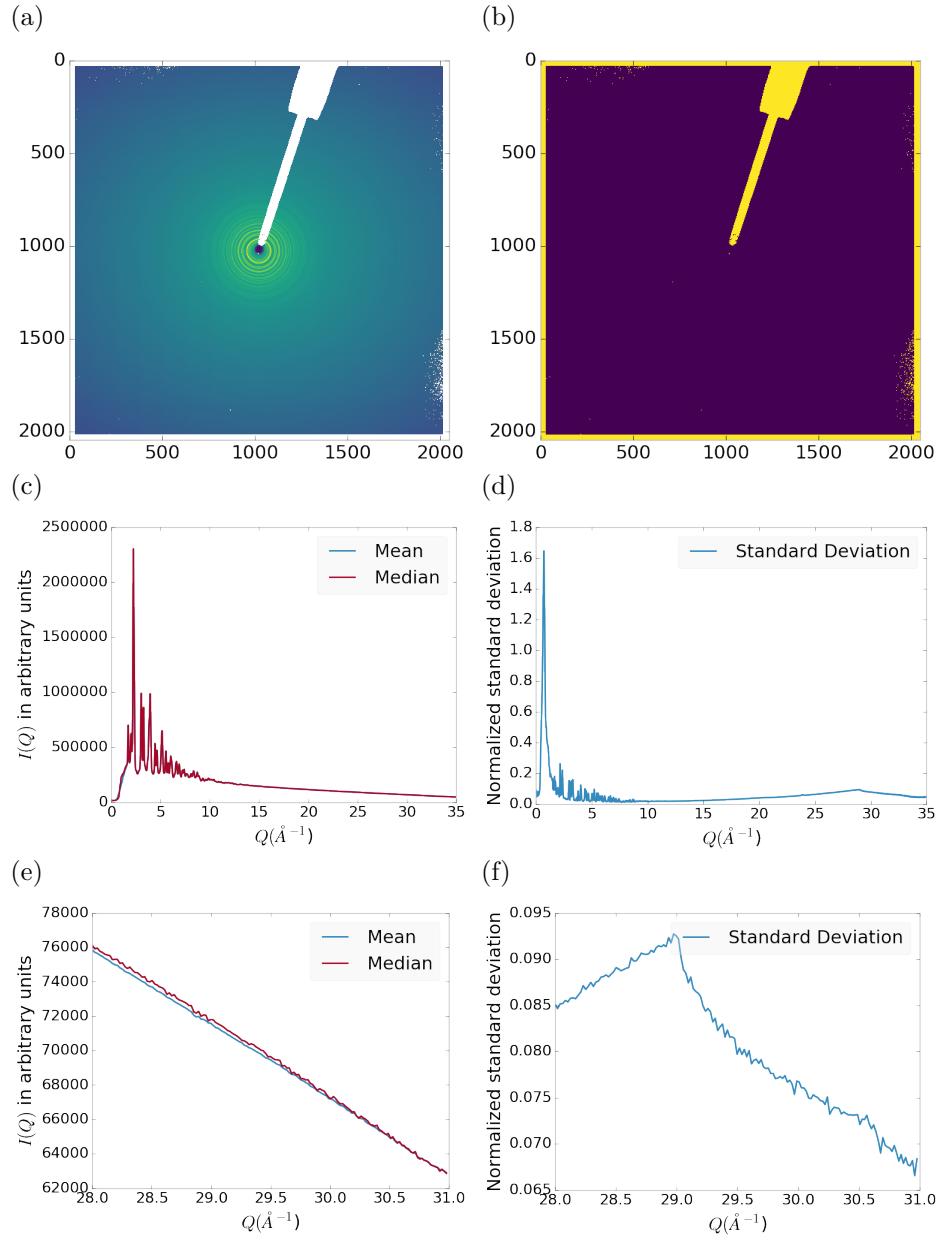


Figure 4.18: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced combining an edge mask and the automatically generated mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median),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

801 4.5 CONCLUSIONS

802 This chapter developed and analyzed the proper data processing and reduction method-  
803 ology for producing reliable  $F(Q)$  data from x-ray total scattering measurements.  
804 Binning at the  $Q$  resolution of the detector was found to be key to the data process-  
805 ing. The primary outcome of using the  $Q$  resolution binning was an enhancement in  
806 effectiveness for the masking algorithm, producing much fewer false positives for dead  
807 pixels. This masking approach was then applied to the integration of experimental  
808 data taken at the APD’s 11-ID-B beamline. The automatically generated masks,  
809 when combined with edge masks, were found to greatly reduce the overall standard  
810 deviation of the pixel intensity and produce a smoother  $F(Q)$  at high  $Q$ , enabling  
811 the use of much higher  $Q$  data in the PDF. Different statistical measures used in the  
812 azimuthal integration was also compared. This comparison showed that the median  
813 was a more reliable statistic for integration with data which had more detector de-  
814 fects. However, upon properly masking it was shown that these metrics were almost  
815 identical. The masking induced similarity between the mean and median shows that  
816 the rings, when integrated, may form a Gaussian distribution. The distribution of  
817 the pixel intensities for strongly and weakly scattering samples may be investigated  
818 in future work.

819

## CHAPTER 5

820

### PHASE CHANGES AND ANNEALING DYNAMICS OF

821

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

822

#### 5.1 INTRODUCTION

823

We should discuss about why PNO is interesting, at least in brief

824

#### 5.2 EXPERIMENTS

825

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

826

need some sort of synthesis information, something along the lines of as previously reported

827

##### X-ray Measurements

828

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

833

#### 5.3 DATA PROCESSING

834

The data was calibrated at each of the detector positions using a  $\text{CeO}_2$  standard via pyFAI. [19] The images were corrected for a .95 x-ray polarization. Masks were

836 produced for both the foreground and background images. The foreground masks were  
837 produced using both a 30 pixel edge mask and a  $2.5\sigma$  automatic mask as discussed  
838 in chapter 4. The background masks were produced by using the foreground mask as  
839 a starting mask with a  $2.5\sigma$  automatic mask.

840 The foreground and background images were then integrated using the  $Q$  resolu-  
841 tion binning discussed in chapter 4. The resulting  $I(Q)$  data were corrected for their  
842 number of frames and  $I_{00}$ . Finally the corrected background  $I(Q)$  was subtracted  
843 from the foreground  $I(Q)$ .

844 Each PDF was generated with a  $Q_{min}$  of 1.5,  $Q_{max}$  of 29.,  $R_{poly}$  of .9,  $R_{max}$  of 40.  
845 descriptions of these parameters can be found in the work by Juhas et. al. [18]

846 **5.4 DATA ANALYSIS**

847 **Intra Sample Comparison**

848 **PDF**

849 As figures 5.1 and 5.2 show the as synthesized PNO undergoes very little change in  
850 structure according to the PDF. The PDF does show some broadening at around 3.5  
851 and 8.5 Å, but the peak shifts themselves are fairly limited. This implies that the as  
852 synthesized PNO structure is stable at least for the 1 hour that the sample was held  
853 at 750 °C.

854 The annealed samples figures, 5.3 and 5.4, tell a rather different story. In this case  
855 the PDF shows significant peak shifts and broadening, especially at higher interatomic  
856 distances. Some peaks completely disappear, like the peak at 12 Å. Similar results were  
857 also observed for samples with longer annealing times, as shown in the appendix.

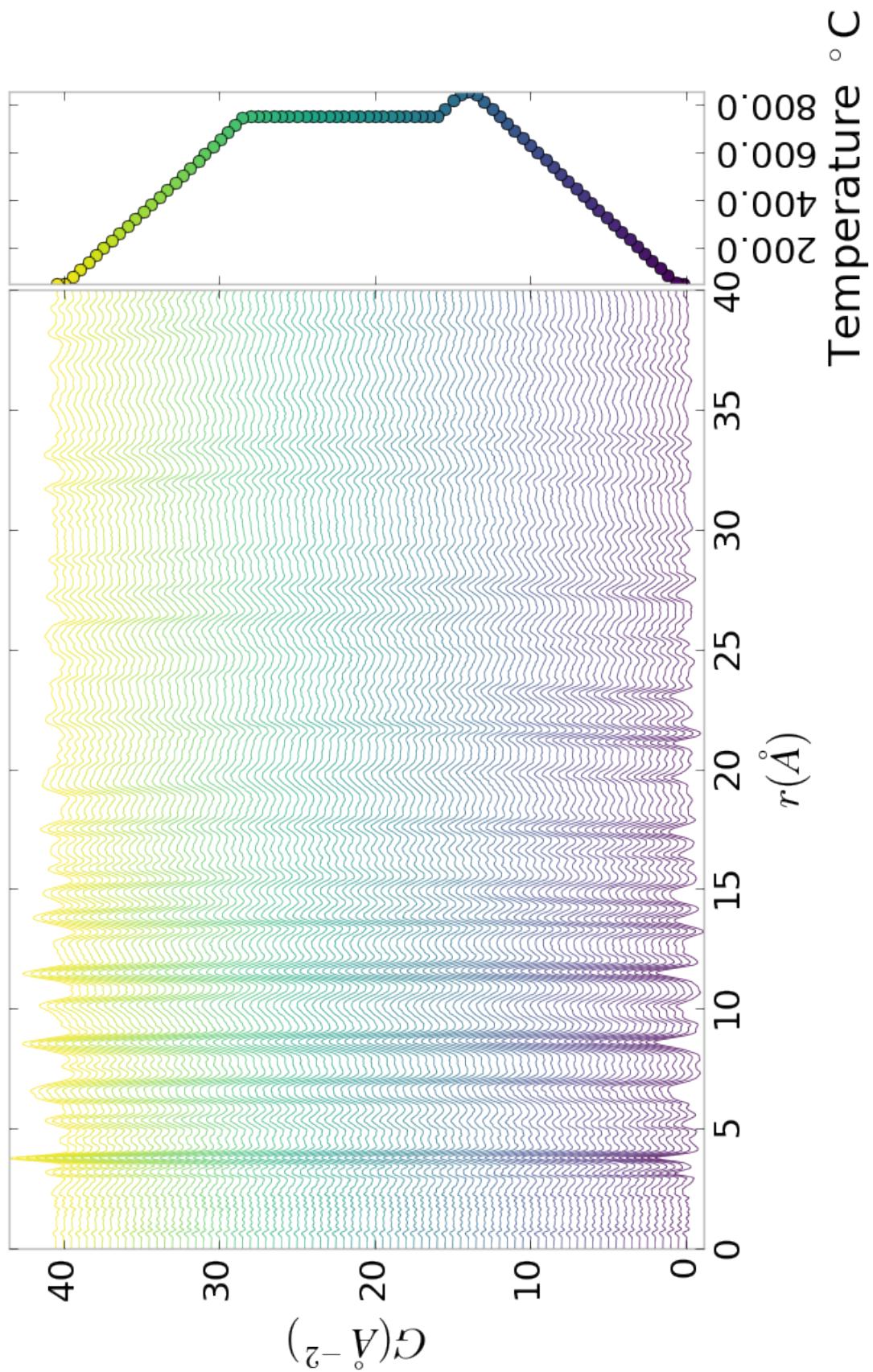


Figure 5.1: PDF as a function of temperature for as synthesized PNO showing the full PDF

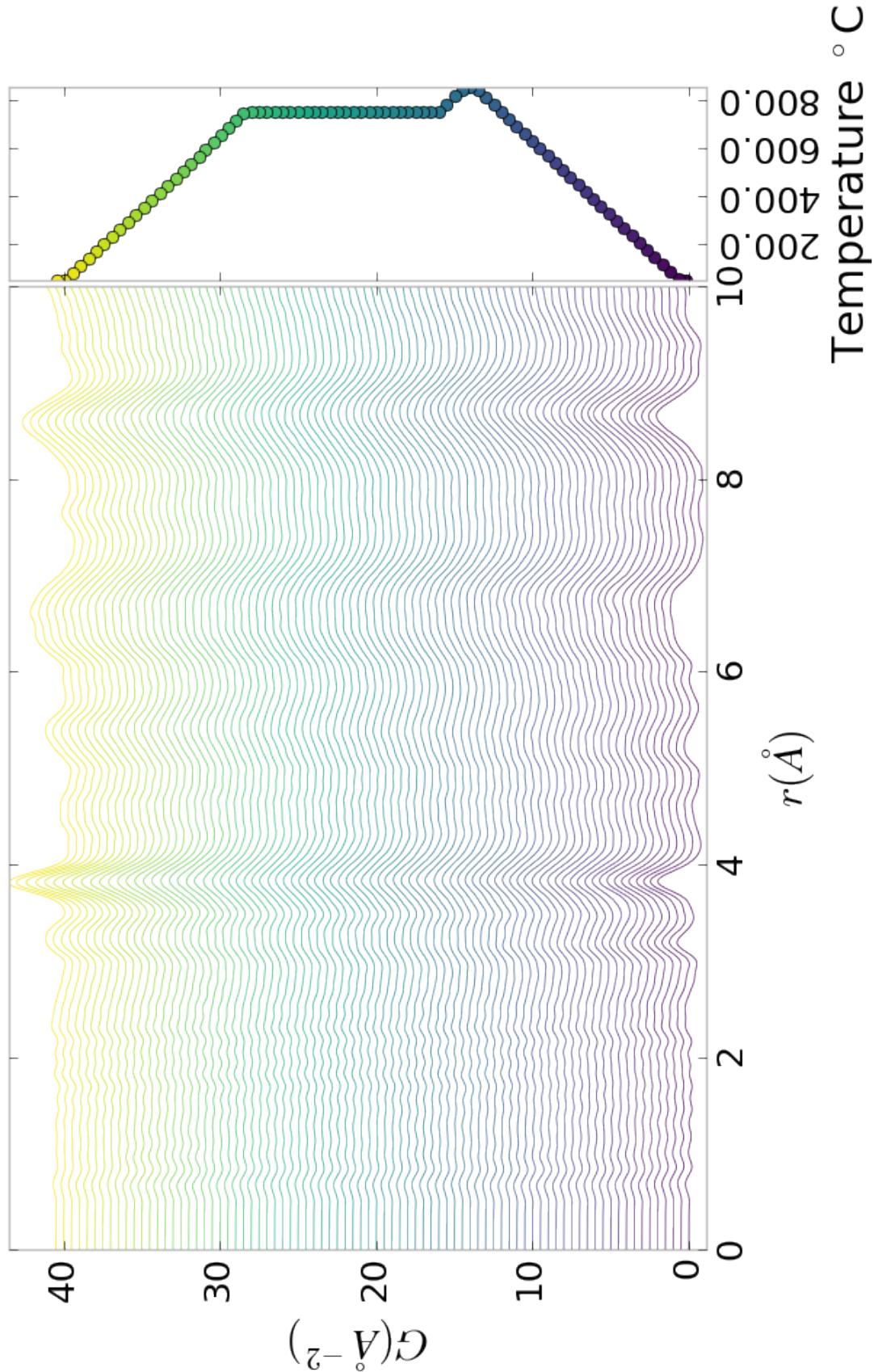


Figure 5.2: PDF as a function of temperature for as synthesized PNO showing a close up on the short range section

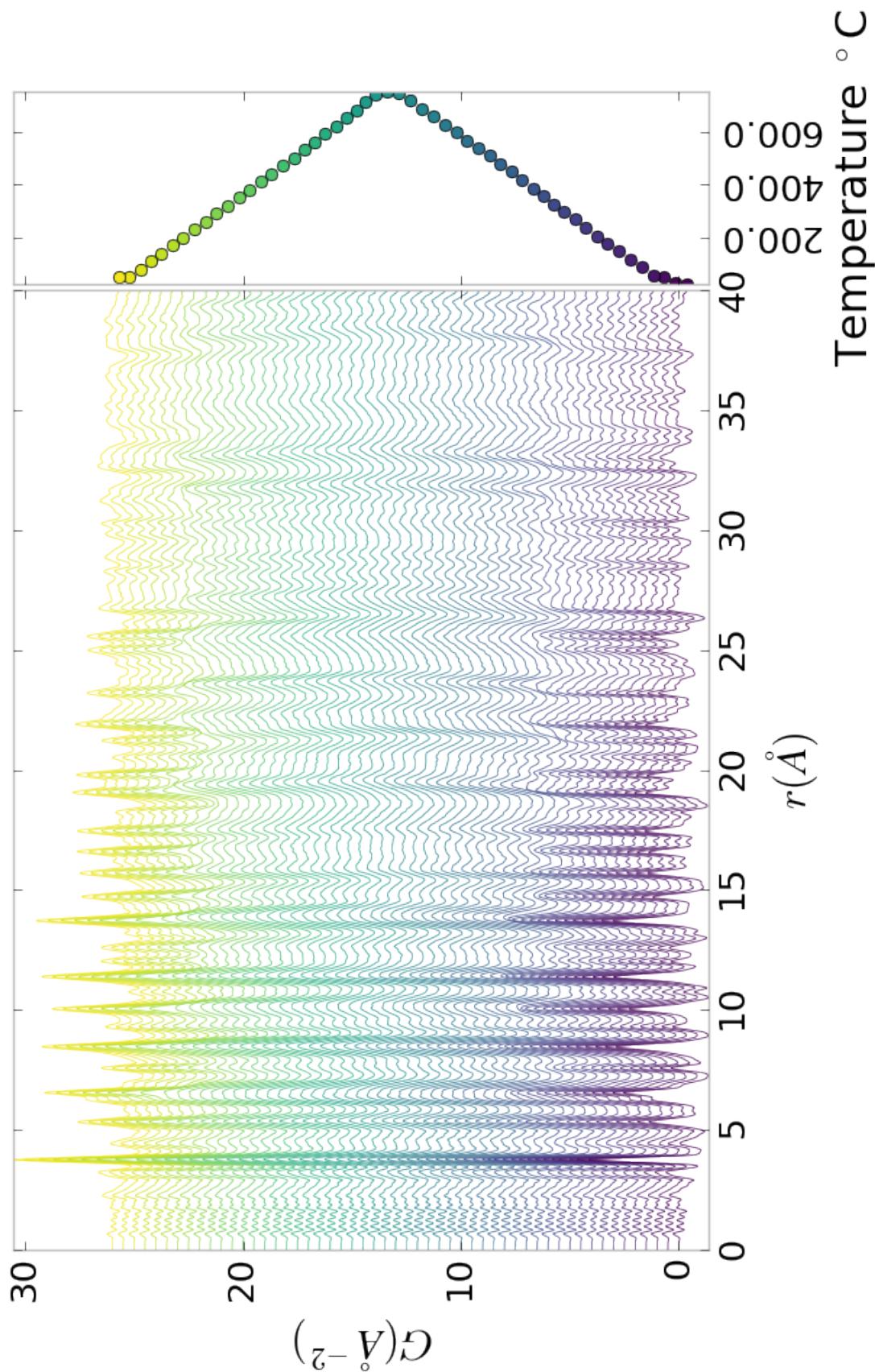


Figure 5.3: PDF as a function of temperature for PNO annealed at 750  $^{\circ}\text{C}$  for 25 hours showing the full PDF

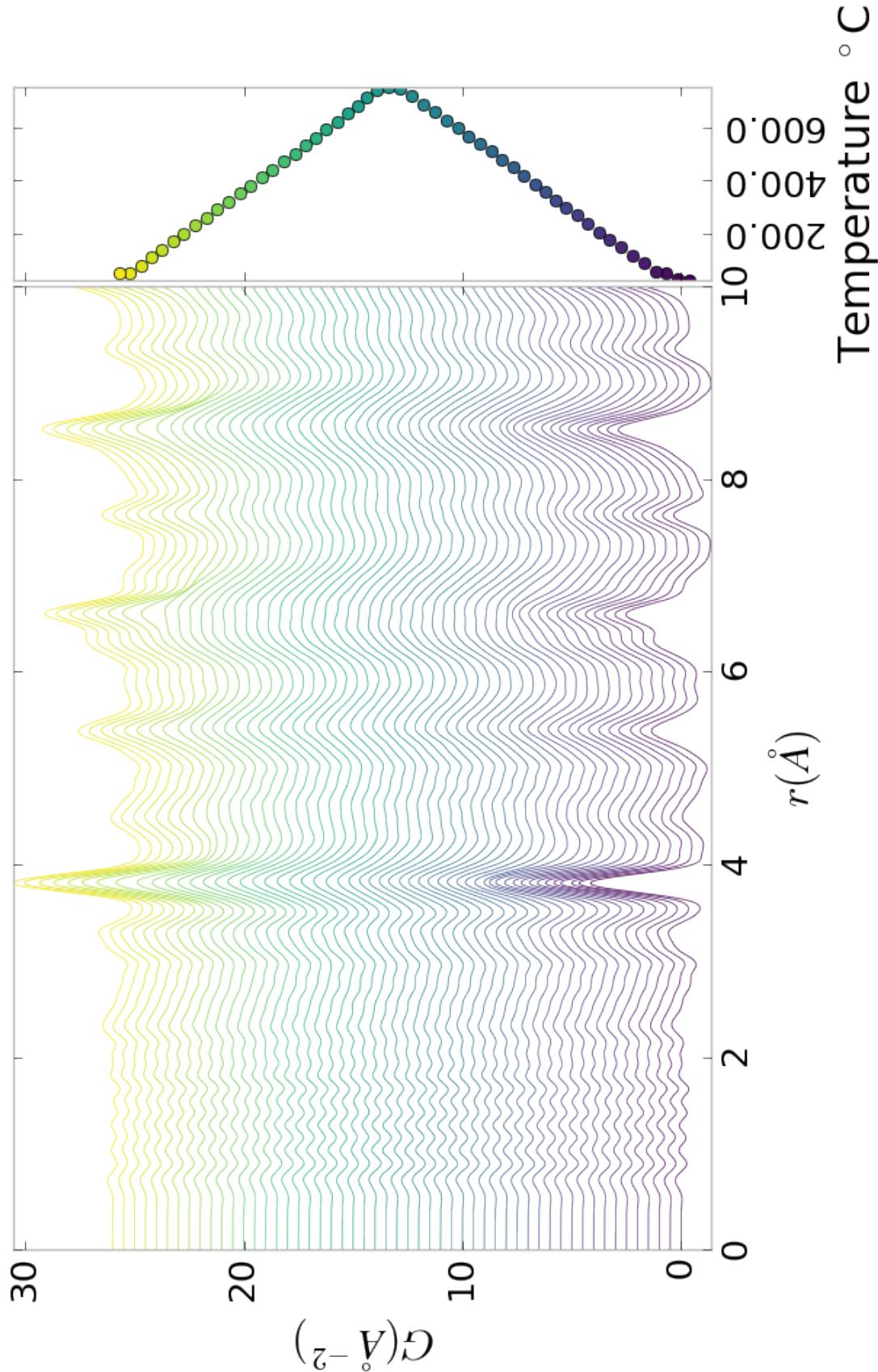


Figure 5.4: PDF as a function of temperature for PNO annealed at 750  $^{\circ}\text{C}$  for 25 hours showing a close up on the short range section

858  $I(Q)$

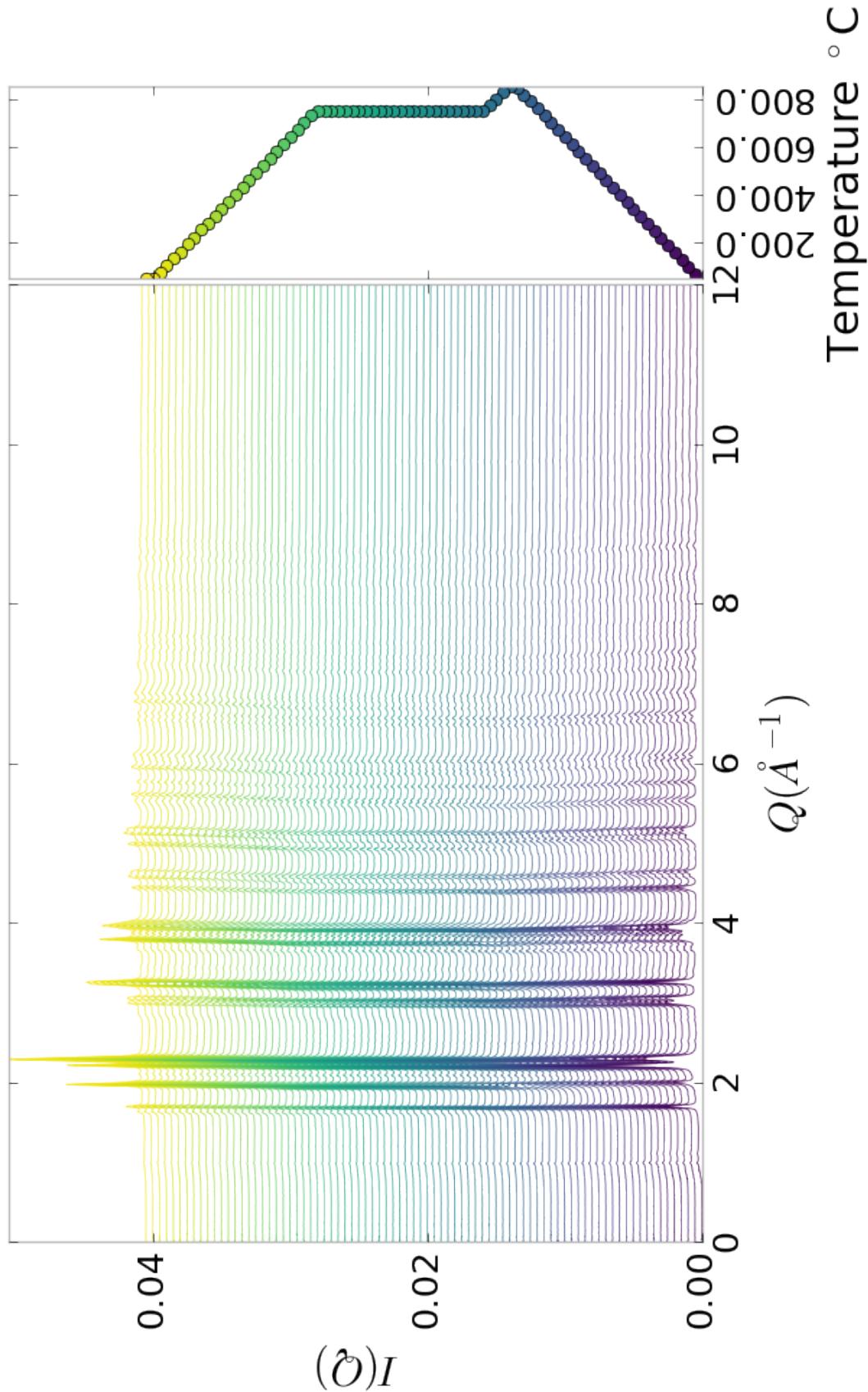


Figure 5.5:  $I(Q)$  as a function of temperature for as synthesized PNO showing the full XRD

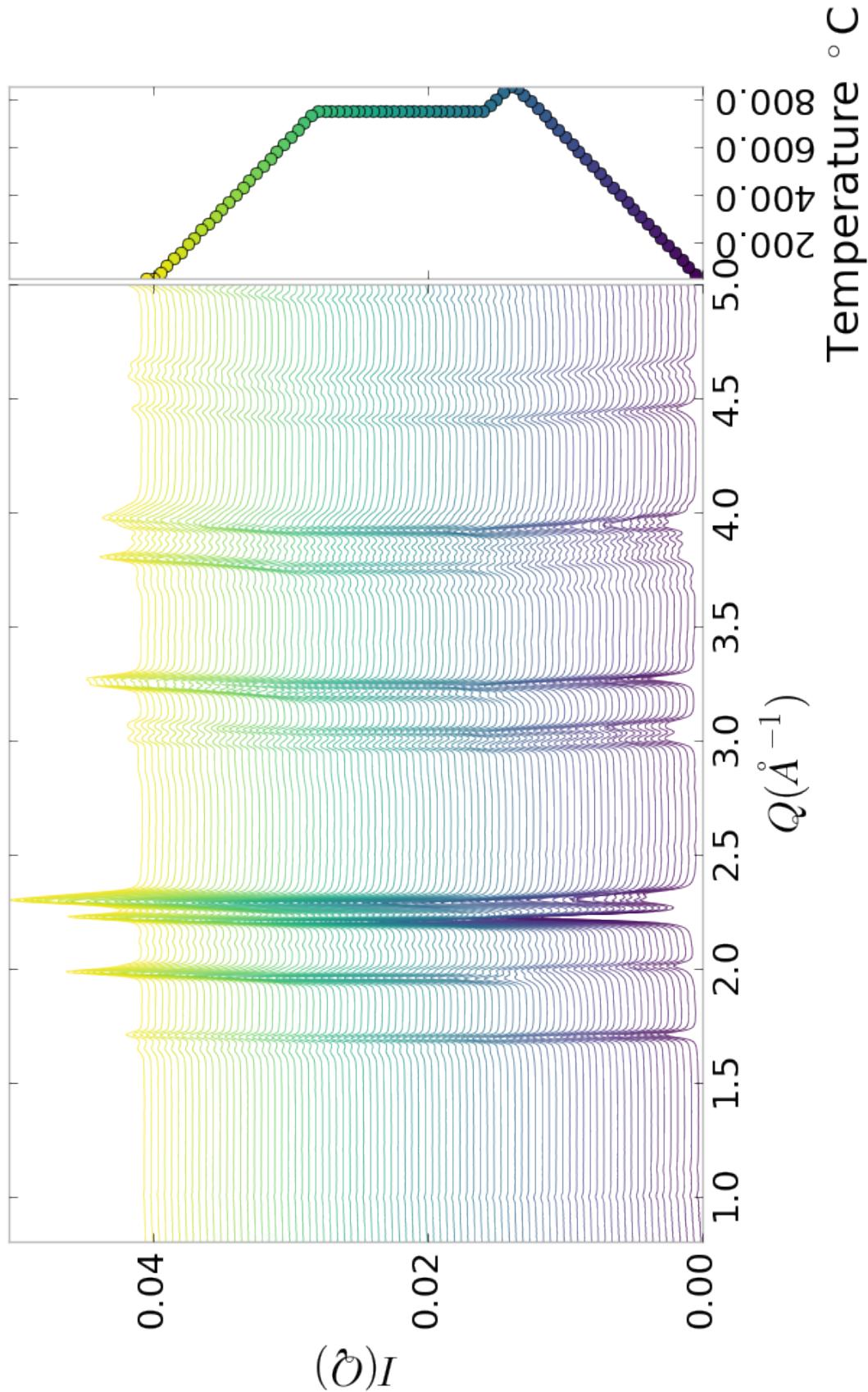


Figure 5.6:  $I(Q)$  as a function of temperature for as synthesized PNO showing a close up on the low  $Q$  section

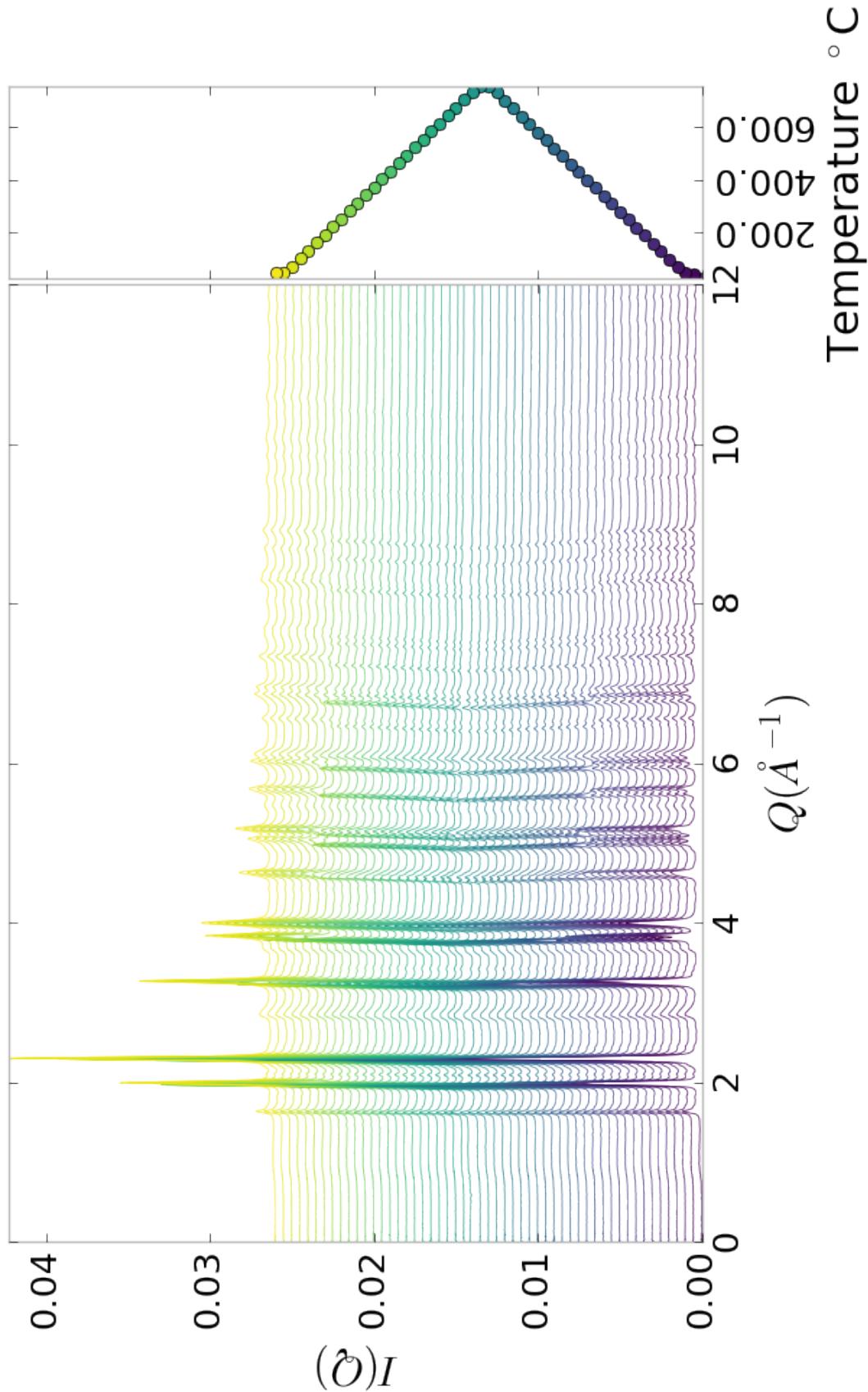


Figure 5.7:  $I(Q)$  as a function of temperature for PNO annealed at  $750^\circ\text{C}$  for 25 hours showing the full XRD

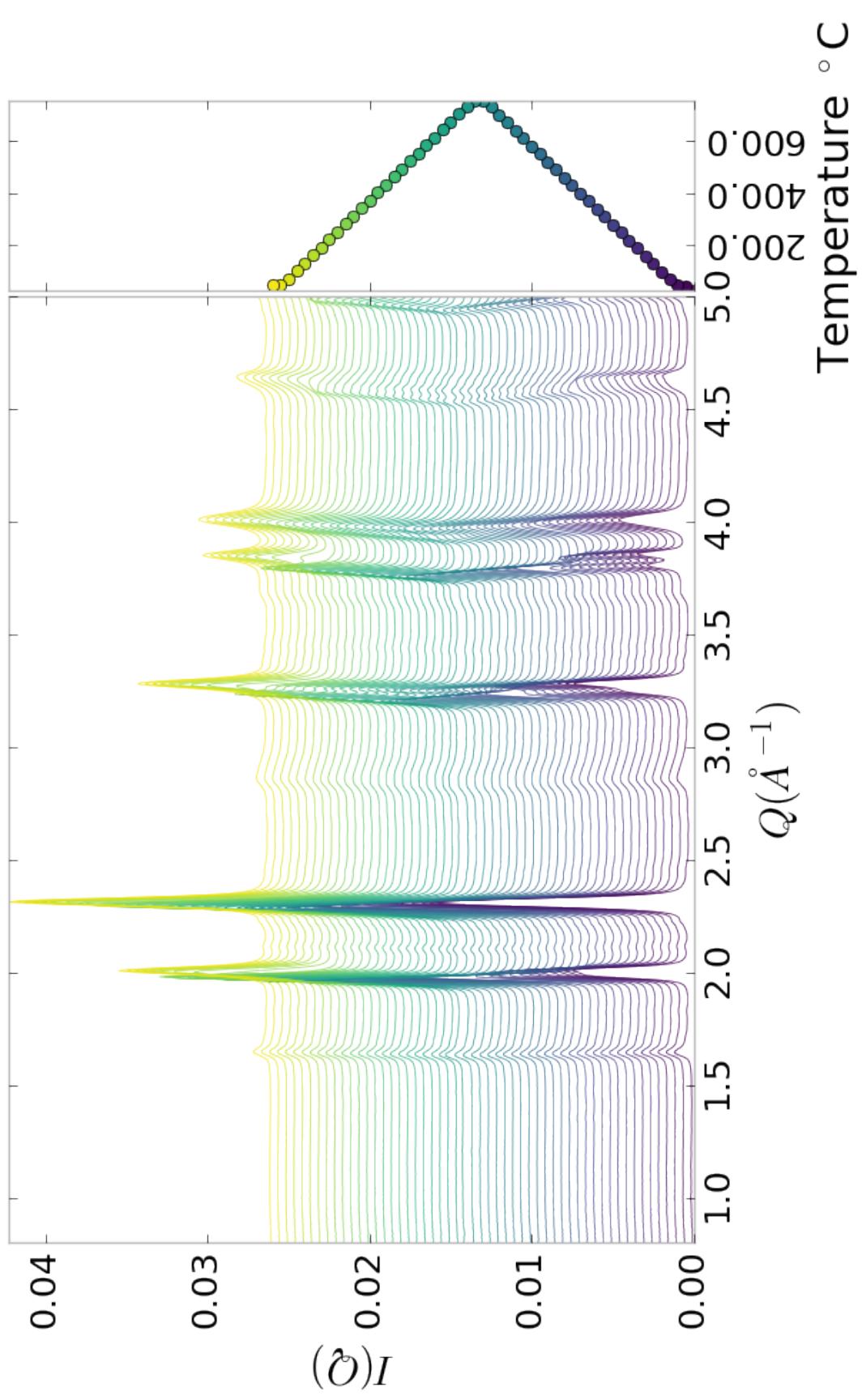


Figure 5.8:  $I(Q)$  as a function of temperature for PNO annealed at  $750\text{ }^{\circ}\text{C}$  for 25 hours showing a close up on the low  $Q$  section

859 **Inter Sample Comparison**

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

861 **5.5 SIMULATION**

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

865 1. The PDF's insensitivity to the oxygen positions, due to the poor x-ray scattering  
866 off the very electorn poor oxygens.

867 2. The large difference in mass between the oxygen and other atoms, causing the  
868 dynamics of the simulation to be governed by oxygen motion, nessecitating long  
869 simulation times to obtain movement of the other atoms.

870 3. The large parameter space caused by potential defects and degradation prod-  
871 ucts. Without knowing that the starting phase is pure, it is difficult to even  
872 produce starting structures, since the simulation will need to explore all the  
873 potential defect/degenerated structures.

874 **5.6 CONCLUSIONS**

875 X-ray total scattering and x-ray powder diffraction data was obtained on  $\text{Pr}_2\text{NiO}_4$   
876 powder samples annealed for various lengths of time. In-situ studies on the beamline  
877 were performed to understand how the structure of each of these powders changes  
878 at operating temperatures. The data was processed with the previously discussed  $Q$   
879 binning, masking, and integration methodology. The PDF results show very little  
880 change in the structure for the as synthesized sample. However, the PDFs show a  
881 large change in the previously annealed samples. These changes seem to reporduce

882 PDFs similar to the as-synthesized PNO at operating temperatures. This would seem  
883 to imply that the source of the anomalous PNO phase/power density relationship may  
884 be due to the adoption of an active structure upon heating which is universal despite  
885 the amount of thermal degradation observed at room temperature. In contrast to the  
886 PDF results, the XRD results seem to show significant changes in the PNO structure,  
887 both with ex-situ and in-situ annealing. The XRDs show the degradation of the PNO  
888 into various phases, potentially including  $\text{Pr}_2\text{O}_{11}$ , and higher ordered Pr based phases.  
889 The discrepancy between these two results is quite interesting as it seems that the  
890 XRD and PDF results are contradictory. Turbostratic displacements between the  
891 layers may be a cause of the PDF/XRD disagreement, as these changes would cause  
892 very little change in the local structure observed in the PDF, while causing large  
893 changes in the XRD.

## CONCLUSION

895

## BIBLIOGRAPHY

- 896 [1] Simon J L Billinge and Takeshi Egami, *Underneath the Bragg Peaks: Structural  
897 Analysis of complex Materials*, vol. Volume 16, Pergamon, 2012.
- 898 [2] P. E. Blöchl, *Projector augmented-wave method*, Physical Review B **50** (1994),  
899 no. 24, 17953–17979.
- 900 [3] Gary K. Chen and Yunfei Guo, *Discovering epistasis in large scale genetic as-  
901 sociation studies by exploiting graphics cards*, Frontiers in Genetics **4** (2013),  
902 no. DEC, 1–12.
- 903 [4] Joshua J Choi, Xiaohao Yang, Zachariah M Norman, Simon J L Billinge, and  
904 Jonathan S Owen, *Structure of methylammonium lead iodide within mesoporous  
905 titanium dioxide: Active material in high-performance perovskite solar cells*,  
906 Nano Letters **14** (2014), no. 1, 127–133.
- 907 [5] Peter J. Chupas, Karena W. Chapman, Charles Kurtz, Jonathan C. Hanson,  
908 Peter L. Lee, and Clare P. Grey, *A versatile sample-environment cell for non-  
909 ambient X-ray scattering experiments*, Journal of Applied Crystallography **41**  
910 (2008), no. 4, 822–824.
- 911 [6] PJ Chupas, X Qiu, JC Hanson, PL Lee, CP Grey, and SJL Billinge, *Rapid-  
912 acquisition pair distribution function (RA-PDF) analysis*, Journal of Applied  
913 Crystallography **36** (2003), 1342–1347.
- 914 [7] Matthew J. Cliffe, Martin T. Dove, D. a. Drabold, and Andrew L. Goodwin,  
915 *Structure determination of disordered materials from diffraction data*, Physical  
916 Review Letters **104** (2010), no. 12, 1–4.
- 917 [8] Matthew J Cliffe and Andrew L Goodwin, *Nanostructure determination from the  
918 pair distribution function: a parametric study of the INVERT approach.*, Journal  
919 of physics. Condensed matter : an Institute of Physics journal **25** (2013), no. 45,  
920 454218.
- 921 [9] Juarez L F Da Silva, Hyoung Gyu Kim, Maurício J. Piotrowski, Maurício J. Pri-  
922 perto, and Germano Tremiliosi-Filho, *Reconstruction of core and surface nanopar-*

- 923        *ticles: The example of Pt 55 and Au55*, Physical Review B - Condensed Matter  
924        and Materials Physics **82** (2010), no. 20, 1–6.
- 925 [10] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth, *Hybrid  
926        Monte Carlo*, Physics Letters B **195** (1987), no. 2, 216–22.
- 927 [11] Timur Dykhne, Ryan Taylor, Alastair Florence, and Simon J L Billinge, *Data  
928        requirements for the reliable use of atomic pair distribution functions in amor-  
929        phous pharmaceutical fingerprinting.*, Pharmaceutical research **28** (2011), no. 5,  
930        1041–8.
- 931 [12] C L Farrow, P Juhas, J W Liu, D Bryndin, E S Božin, J Bloch, Th Proffen, and  
932        S J L Billinge, *PDFfit2 and PDFgui: computer programs for studying nanostruc-  
933        ture in crystals.*, Journal of Physics. Condensed Matter : an Institute of Physics  
934        journal **19** (2007), no. 33, 335219.
- 935 [13] Christopher L Farrow and Simon J L Billinge, *Relationship between the atomic  
936        pair distribution function and small-angle scattering: implications for modeling  
937        of nanoparticles.*, Acta Crystallographica Section A Foundations of Crystallog-  
938        raphy **65** (2009), no. Pt 3, 232–9 (en).
- 939 [14] Riccardo Ferrando, Julius Jellinek, and Roy L Johnston, *Nanoalloys: From The-  
940        ory to Applications of Alloy Clusters and Nanoparticles*, Chemical Reviews **108**  
941        (2008), no. 3, 846–904.
- 942 [15] Md Matthew D. Hoffman and Andrew Gelman, *The No-U-Turn Sampler: Adap-  
943        tively Setting Path Lengths in Hamiltonian Monte Carlo*, The Journal of Machine  
944        Learning Research **15** (2014), no. 2008, 1593–1623.
- 945 [16] Pablo D Jadzinsky, Guillermo Calero, Christopher J Ackerson, David A Bushnell,  
946        and Roger D Kornberg, *Structure of a thiol monolayer-protected gold nanoparti-  
947        cle at 1.1 Å resolution.*, Science (New York, N.Y.) **318** (2007), no. 5849, 430–433.
- 948 [17] I. K. Jeong, R. H. Heffner, M. J. Graf, and S. J. L. Billinge, *Lattice dynamics  
949        and correlated atomic motion from the atomic pair distribution function*, (2002),  
950        9.
- 951 [18] P. Juhás, T. Davis, C. L. Farrow, and S. J. L. Billinge, *PDFgetX3 : a  
952        rapid and highly automatable program for processing powder diffraction data into  
953        total scattering pair distribution functions*, Journal of Applied Crystallography  
954        **46** (2013), no. 2, 560–566 (en).

- 955 [19] Jérôme Kieffer and Dimitrios Karkoulis, *PyFAI, a versatile library for azimuthal*  
956 *regrouping*, Journal of Physics: Conference Series **425** (2013), 202012.
- 957 [20] G. Kresse and J. Hafner, *Ab initio molecular dynamics for liquid metals*, Physical  
958 Review B **47** (1993), no. 1, 558–561.
- 959 [21] ———, *Ab initio molecular-dynamics simulation of the liquid-*  
960 *metal–amorphous-semiconductor transition in germanium*, Physical Review  
961 B **49** (1994), no. 20, 14251–14269.
- 962 [22] Yan Li, Giulia Galli, and François Gygi, *Electronic structure of thiolate-covered*  
963 *gold nanoparticles: Au<sub>102</sub>(MBA)<sub>44</sub>*, ACS Nano **2** (2008), no. 9, 1896–1902.
- 964 [23] L D Marks, *Experimental studies of small particle structures*, Reports on  
965 Progress in Physics **57** (1994), no. 6, 603–649 (en).
- 966 [24] A. S. Masadeh, E. S. Božin, C. L. Farrow, G. Paglia, P. Juhas, S. J. L. Billinge,  
967 A. Karkamkar, and M. G. Kanatzidis, *Quantitative size-dependent structure and*  
968 *strain determination of CdSe nanoparticles using atomic pair distribution func-*  
969 *tion analysis*, Physical Review B - Condensed Matter and Materials Physics **76**  
970 (2007), no. 11, 115413.
- 971 [25] R L McGreevy and L Pusztai, *Reverse Monte Carlo Simulation: A New Tech-*  
972 *nique for the Determination of Disordered Structures*, Molecular Simulation **1**  
973 (1988), no. 6, 359–367.
- 974 [26] Radford M Neal, *Probabilistic Inference Using Markov Chain Monte Carlo Meth-*  
975 *ods*, Intelligence **45** (1993), no. September, 144.
- 976 [27] Radford M. Neal, *MCMC Using Hamiltonian Dynamics*, Handbook of Markov  
977 Chain Monte Carlo (Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li  
978 Meng, eds.), Chapman and Hall/CRC, 2011, pp. 113–162.
- 979 [28] B R Pauw, *Corrigendum: Everything SAXS: small-angle scattering pattern col-*  
980 *lection and correction (2013 J. Phys.: Condens. Matter 25 383201)*, Journal of  
981 Physics: Condensed Matter **26** (2014), no. 23, 239501.
- 982 [29] John P. Perdew, Kieron Burke, and Matthias Ernzerhof, *Generalized Gradient*  
983 *Approximation Made Simple*, Physical Review Letters **77** (1996), no. 18, 3865–  
984 3868.

- 985 [30] Andrew A. Peterson, *Global optimization of adsorbate-surface structures while*  
986 *preserving molecular identity*, Topics in Catalysis **57** (2014), no. 1-4, 40–53.
- 987 [31] Valeri Petkov, Binay Prasai, Yang Ren, Shiyao Shan, Jin Luo, Pharrah Joseph,  
988 and Chuan-Jian Zhong, *Solving the nanostructure problem: exemplified on metal-*  
989 *alloy nanoparticles*, Nanoscale **6** (2014), no. 17, 1–11.
- 990 [32] Valeri Petkov, Shiyao Shan, Peter Chupas, Jun Yin, Lefu Yang, Jin Luo, and  
991 Chuan-Jian Zhong, *Noble-transition metal nanoparticle breathing in a reactive*  
992 *gas atmosphere.*, Nanoscale **5** (2013), no. 16, 7379–87.
- 993 [33] Th. Proffen and R. B. Neder, *DISCUS: a Program for Diffuse Scattering and*  
994 *Defect-Structure Simulation*, Journal of Applied Crystallography **30** (1997), 171–  
995 175.
- 996 [34] Erin L. Redmond, Brian P. Setzler, Pavol Juhas, Simon J. L. Billinge, and  
997 Thomas F. Fuller, *In-Situ Monitoring of Particle Growth at PEMFC Cathode*  
998 *under Accelerated Cycling Conditions*, Electrochemical and Solid-State Letters  
999 **15** (2012), no. 5, B72 (en).
- 1000 [35] H. W. Sheng, M. J. Kramer, A. Cadien, T. Fujita, and M. W. Chen, *Highly*  
1001 *optimized embedded-atom-method potentials for fourteen FCC metals*, Physical  
1002 Review B - Condensed Matter and Materials Physics **83** (2011), no. 13, 134118.
- 1003 [36] Randall Q. Snurr, Alexis T. Bell, and Doros N. Theodorou, *Prediction of ad-*  
1004 *sorption of aromatic hydrocarbons in silicalite from grand canonical Monte Carlo*  
1005 *simulations with biased insertions*, The Journal of Physical Chemistry **97** (1993),  
1006 no. 51, 13742–13752 (EN).
- 1007 [37] X. Yang, P. Juhás, and S. J L Billinge, *On the estimation of statistical uncertain-*  
1008 *ties on powder diffraction and small-angle scattering data from two-dimensional*  
1009 *X-ray detectors*, Journal of Applied Crystallography **47** (2014), no. 4, 1273–1283.