

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

This work would not have been possible without my entire ensemble of colleagues, advisors, friends, and family. To Dr. Sanjit Ghose thank you very much for getting me started in the x-ray field and sticking with me, even when I went into the mathematical deep end. To Dr. Yan Li, thank you very much for always pushing my simulations to be the best they could be, and pusing me to search every possible nook and cranny of parameter space.

I'd also like to thank the NSLS-II XPD team for their experimental, intelecual, and moral support. To Dr. Thomas Caswell, Dr. Eric Dill, and Dr. Dan Allen, I would not be  $1e^{-6}$  of the programmer I am today with out your guidence, conversation, and explinations.

To Prof. Scopatz, thank you very much for your insights, conversation, and being a spectacular developer role model.

To my many readers, I'd like to say thank you, if my thesis hasn't put you to sleep already.

To my parents, I don't know if any language could correctly render the magnitude of my gratitude for everything you have done for me.

Last, but certainly not least, Prof. Zhou. I am certain that I could not have gotten a better start in research science than working with you.

## 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.

# CONTENTS

DEDICATION . . . . .	iii
ACKNOWLEDGMENTS . . . . .	iv
ABSTRACT . . . . .	v
LIST OF TABLES . . . . .	xi
LIST OF FIGURES . . . . .	xii
TODO LIST . . . . .	1
CHAPTER 1 INTRODUCTION . . . . .	2
CHAPTER 2 STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES . . . . .	4
2.1 Introduction . . . . .	4
2.2 Potential Energy Surfaces . . . . .	4
Experimentally Derived Potential Energy Surfaces . . . . .	5
Potentials . . . . .	5
Forces . . . . .	6
2.3 Ensembles . . . . .	7
Monte Carlo Modeling . . . . .	8
Hamiltonian Monte Carlo . . . . .	9
No-U-Turn Sampling . . . . .	11
Grand Canonical Ensemble . . . . .	11
Ensemble description . . . . .	12
Grand Canonical Monte Carlo . . . . .	12
GCMC biasing . . . . .	13

2.4 Conclusions . . . . .	16
CHAPTER 3 ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION . . . . .	17
3.1 Introduction . . . . .	17
3.2 Theory . . . . .	18
Derivation . . . . .	18
Analytically Gradients . . . . .	19
Without ADPs . . . . .	20
Periodic Boundary Conditions . . . . .	21
3.3 Computation . . . . .	21
HPC and GPUs . . . . .	22
GPUs and Parallelization . . . . .	22
Map from ij space to k space . . . . .	23
GPU Memory Allocation . . . . .	25
Speed and Scaling of PDF Computation . . . . .	27
3.4 Conclusions . . . . .	28
CHAPTER 4 BENCHMARKING . . . . .	29
4.1 Introduction . . . . .	29
4.2 PDF . . . . .	29
Model Parameters . . . . .	31
Au55: surface relaxed . . . . .	32
Run Parameters . . . . .	32
Au55: surface disordered . . . . .	34
Au55: amorphous . . . . .	35
Au102: triple phase . . . . .	36
Starting from fcc structure . . . . .	36
Marks decahedron . . . . .	37
Au147 . . . . .	38
4.3 PDF with ADPs . . . . .	43
ADP 50 . . . . .	43
CHAPTER 5 X-RAY TOTAL SCATTERING DATA ACQUISITION AND PRO- CESSING . . . . .	44

5.1	Introduction . . . . .	44
5.2	Detector $Q$ resolution . . . . .	44
5.3	Automated Mask Generation . . . . .	47
	Introduction . . . . .	47
	Algorithm Design . . . . .	48
	Test Cases . . . . .	48
	Results and Discussion . . . . .	49
	Conclusions . . . . .	56
5.4	Automated Image Azimuthal Integration . . . . .	56
5.5	Conclusions . . . . .	61
<b>CHAPTER 6 PHASE CHANGES AND ANNEALING DYNAMICS OF <math>\text{Pr}_2\text{NiO}_4</math> AND ITS DERIVATIVES . . . . .</b>		<b>62</b>
6.1	Introduction . . . . .	62
6.2	Experiments . . . . .	63
	$\text{Pr}_2\text{NiO}_4$ Synthesis . . . . .	63
	X-ray Measurements . . . . .	63
6.3	Data Processing . . . . .	63
6.4	Data Analysis . . . . .	64
	Intra Sample Comparison . . . . .	64
	PDF . . . . .	64
	$I(Q)$ . . . . .	69
	Inter Sample Comparison . . . . .	74
6.5	Simulation . . . . .	74
6.6	Conclusions . . . . .	74
<b>CHAPTER 7 CONCLUSION . . . . .</b>		<b>76</b>
<b>BIBLIOGRAPHY . . . . .</b>		<b>79</b>

APPENDIX A SUPPLEMENTAL INFORMATION: PHASE CHANGES AND ANNEALING DYNAMICS OF  $\text{Pr}_2\text{NiO}_4$  AND ITS DERIVATIVES . . . . . 85

Intra Sample Comparison . . . . . 85

## LIST OF TABLES

## LIST OF FIGURES

Figure 2.1	Addition biasing with a Lennard Jones potential.	15
Figure 3.1	Comparison of the CPU and GPU chip architectures	22
Figure 3.2	Speed comparison of CPU and GPU implementations	27
Figure 4.1	Au <sub>55</sub> PDF fitting of DFT-optimized <i>c</i> -Au <sub>55</sub> .	33
Figure 4.2	Au <sub>55</sub> PDF fitting of surface-disordered Au <sub>55</sub> .	39
Figure 4.3	Similar to figure 4.2 for DFT-optimized amorphous Au <sub>55</sub> .	40
Figure 4.4	Similar to figure 4.2 for Au <sub>102</sub> as in DFT-optimized Au <sub>102</sub> MBA <sub>44</sub> cluster.	41
Figure 4.5	Similar to Fig. 4.4 with Marks decahedron as the starting structure.	42
Figure 4.6	Refinement of adps	43
Figure 5.1	Scattering onto a flat detector	45
Figure 5.2	$Q$ resolution as a function of $Q$ .	46
Figure 5.3	Number of pixels as a function of $Q$ , binned at the $Q$ resolution of the detector.	47
Figure 5.4	Generated dead/hot pixel masks for a detector with 100 bad pixels.	50
Figure 5.5	Generated dead/hot pixel masks for a detector with 300 bad pixels.	50
Figure 5.6	Generated dead/hot pixel masks for a detector with 500 bad pixels.	51
Figure 5.7	Generated dead/hot pixel masks for a detector with 1000 bad pixels.	51
Figure 5.8	Generated beamstop holder masks for a beamstop holder with 10% transmittance.	52

Figure 5.9 Generated beamstop holder masks for a beamstop holder with 30% transmittance. . . . .	52
Figure 5.10 Generated beamstop holder masks for a beamstop holder with 50% transmittance. . . . .	52
Figure 5.11 Generated beamstop holder masks for a beamstop holder with 90% transmittance. . . . .	53
Figure 5.12 Generated beamstop holder masks which is rotated away from vertical . . . . .	53
Figure 5.13 Masked experimental data. . . . .	54
Figure 5.14 Masked experimental data with Pt single crystal signal. . . . .	54
Figure 5.15 Masked experimental data with Pt single crystal signal using figure's 5.13 mask as a starting mask. . . . .	55
Figure 5.16 Masking, average, and standard deviation of an example x-ray total scattering measurement with with no mask. . . . .	58
Figure 5.17 Masking, average, and standard deviation of an example x-ray total scattering measurement with with only an edge mask. . . . .	59
Figure 5.18 Masking, average, and standard deviation of an example x-ray total scattering measurement with combining an edge mask and the automatically generated mask. . . . .	60
Figure 6.1 PDF as a function of temperature for as synthesized PNO showing the full PDF . . . . .	65
Figure 6.2 PDF as a function of temperature for as synthesized PNO showing a close up on the short range section . . . . .	66
Figure 6.3 PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full PDF . . . . .	67
Figure 6.4 PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the short range section . . . . .	68
Figure 6.5 $I(Q)$ as a function of temperature for as synthesized PNO showing the full XRD . . . . .	70

Figure 6.6	$I(Q)$ as a function of temperature for as synthesized PNO showing a close up on the low $Q$ section . . . . .	71
Figure 6.7	$I(Q)$ as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full XRD . . . . .	72
Figure 6.8	$I(Q)$ as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the low $Q$ section . . . . .	73
Figure A.1	PDF as a function of temperature for PNO annealed at 750 °C for 50 hours showing the full PDF . . . . .	86
Figure A.2	PDF as a function of temperature for PNO annealed at 750 °C for 50 hours showing a close up on the short range section . . . . .	87
Figure A.3	PDF as a function of temperature for PNO annealed at 750 °C for 100 hours showing the full PDF . . . . .	88
Figure A.4	PDF as a function of temperature for PNO annealed at 750 °C for 100 hours showing a close up on the short range section . . . . .	89
Figure A.5	PDF as a function of temperature for PNO annealed at 750 °C for 200 hours showing the full PDF . . . . .	90
Figure A.6	PDF as a function of temperature for PNO annealed at 750 °C for 200 hours showing a close up on the short range section . . . . .	91

1

## TODO LIST

2	Need more references . . . . .	1
3	Also some introduction would be great . . . . .	29
4	this just needs a lot of work . . . . .	29
5	Put this somewhere . . . . .	33
6	need to fix the problem with the $I(Q)$ figures . . . . .	92
7	Need more references	

8

# CHAPTER 1

9

## INTRODUCTION

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

18 Chapter 2 develops the statistical mechanical system used to match the theoretical  
19 structure. §2.2 focuses on the development of potential energy surfaces, including  
20 potential energy and force equations, which have minima where experimental results  
21 and simulated structures agree the most. §2.3 will discuss statistical mechanical  
22 ensembles which are used to search for minima on the potential energy surface.

23 Chapter 3 will discuss the mathematical and computational development of the  
24 atomic pair distribution function (PDF). §3.3 will focus on the rapid graphical pro-  
25 cessing unit based calculation of the PDF and its gradients.

26 Chapter 4 will discuss the benchmarking of the the combined statistical mechan-  
27 ical optimizer and PDF calculation systems against a series of theoretical nanoparti-  
28 cles, focusing on understanding limitations of the method and structure reproduction.

29 Chapter 5 will focus on the acquisition of experimental data, its management, and  
30 processing. §5.2, 5.3, and 5.4 will discuss the derivation of the  $Q$  resolution function,

31 the automated masking of 2D area detectors for x-ray total scattering measurements  
32 using the previously derived  $Q$  resolution, and the impact of different averaging meth-  
33 ods and masks on azimuthal integration, respectively.

34 Chapter 6 will discuss preliminary experimental results investigating the phase  
35 changes and local structure of  $\text{Pr}_2\text{NiO}_4$ , revealing the influence of thermal history on  
36 the structure.

37

## CHAPTER 2

38

# STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES

40 2.1 INTRODUCTION

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

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

53 2.2 POTENTIAL ENERGY SURFACES

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

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

## 65 Experimentally Derived Potential Energy Surfaces

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

## 74 Potentials

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

$$\chi^2 = \sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2 \quad (2.1)$$

$$Rw = \sqrt{\frac{\sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}}^2}} \quad (2.2)$$

$$\chi^2_{\text{INVERT}} = \frac{1}{N} \sum_j \sum_r [A_{\text{obs}}(r) - \alpha A_{j,\text{calc}}(r)]^2 \quad (2.3)$$

79

$$\alpha = \frac{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}} A_{\text{calc}}}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}}^2} = \frac{\vec{A}_{\text{obs}} \cdot \vec{A}_{\text{calc}}}{|\vec{A}_{\text{calc}}|^2} \quad (2.4)$$

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

83 The  $Rw$  and  $\chi^2$  potentials have been reported numerous times. [44, 34, 6, 35, 46]  
 84 Essentially these potentials measure the least squares distance between the observed  
 85 scattering and the predicted scattering providing a way to quantify the agreement  
 86 between the model and experiment. While  $RW$  and  $\chi^2$  are now standard in the PDF  
 87 community, the INVERT potential is fairly new and aims to incorporate descriptions  
 88 of the structural symmetry into the PES. [9, 10] In the case of the INVERT potential  
 89 NMR or other symmetry sensitive data is used to describe the number of unique  
 90 atomic coordinations. This is then used to describe the number of unique atom-wise  
 91 pair distribution functions, thus causing systems with more or less unique coordi-  
 92 nation environments to be higher in energy. This approach has been shown to be  
 93 useful for  $C_{60}$  and other systems which are highly symmetric, creating a PES with an  
 94 easier to find minima. [9, 10] However, many times this kind of data is unavailable  
 95 when refining the structure causing the potential to be less useful. Additionally, this  
 96 potential introduces an element of user bias as the refiner must decide, based on some  
 97 spectroscopic data, how many unique environments are in the material. This bias  
 98 could be removed by using one of the other potentials with a method for simulat-  
 99 ing the observed spectra, allowing the computational system decide what structures  
 100 properly reproduce all the observed data.

101 **Forces**

$$\vec{\nabla} \chi^2 = -2 \sum_{a=a_{\min}}^{a_{\max}} \left( \alpha \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} + A_{\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}} \right) (A_{\text{obs}} - \alpha A_{\text{calc}}) \quad (2.5)$$

102

$$\vec{\nabla}Rw = \frac{Rw}{\chi^2} \sum_{a=a_{\min}}^{a_{\max}} (\alpha \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} + A_{\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}})(\alpha A_{\text{calc}} - (A_{\text{obs}})) \quad (2.6)$$

103

$$\vec{\nabla}\chi^2_{\text{INVERT}} = \frac{-2}{N} \sum_{a=a_{\min}}^{a_{\max}} \sum_j (\alpha \frac{\partial A_{j,\text{calc}}}{\partial \gamma_{i,w}} + A_{j,\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}})(A_{\text{obs}} - \alpha A_{j,\text{calc}}) \quad (2.7)$$

104

$$\frac{\partial \alpha}{\partial \gamma_{i,w}} = \frac{(\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} - 2\alpha \sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}})}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}}^2} \quad (2.8)$$

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

110 2.3 ENSEMBLES

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

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

$$\mathcal{H}(q, p) = U(q) + K(p) \quad (2.9)$$

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

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

$$\Xi = \sum_i P_i(q, p) \quad (2.10)$$

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

128 For the canonical ensemble the partition function is probability is:

$$Q(N, V, T) = \exp\left(\frac{-\mathcal{H}(q, p)}{k_b T}\right) \quad (2.11)$$

129 where  $k_b$  is the Boltzmann constant and  $T$  is the temperature of the system. [36]

## 130 Monte Carlo Modeling

131 Monte Carlo can be used to simulate a statistical mechanical ensemble which can not  
132 be solved analytically. In most Monte Carlo systems the ensemble is simulated by  
133 randomly changing one of the system parameters and comparing the energy of the  
134 new system against the energy of the old system. If the energy of the new system is  
135 lower than the current energy then the new configuration is accepted. Otherwise the  
136 new system is rejected unless

$$\exp\left(\frac{-\Delta E}{E_T}\right) < u \quad (2.12)$$

137 where  $u$  is a random number  $[0, 1)$  and  $E_T$  is the thermal energy characteristic to the  
138 system. The ability of Monte Carlo modeling to accept “bad” moves allows the system  
139 to hop out of local energy minima during the search for the global minimum. Reverse  
140 Monte Carlo (RMC) is similar to Monte Carlo except it uses  $\chi^2$  as the PES.[35]

141 Despite the utility of RMC, and its wide use in the x-ray scattering community, as  
142 Hoffman and Gelman state “Not all MCMC [Markov Chain Monte Carlo] algorithms

143 are created equal".[23] RMC, similar to standard Monte Carlo simulations, samples  
 144 from the PES at random, usually by translating atoms in the system randomly. This  
 145 creates a less efficient, random walk based, exploration of the PES.[23, 37] Thus,  
 146 methods for suppressing this random walk nature, while still searching the potential  
 147 energy surface fully are needed.

## 148 Hamiltonian Monte Carlo

149 Hybrid or Hamiltonian Monte Carlo (HMC) can help to address some of these issues.  
 150 HMC was developed originally in the lattice quantum chromodynamics community  
 151 and provides a more efficient, more scalable approach to PES sampling for Monte  
 152 Carlo.[14, 38] In HMC the PES is explored using Hamiltonian dynamics, essentially  
 153 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:

$$\frac{dq_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} = p_i \quad (2.13)$$

$$\frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial q_i} = -\vec{\nabla}U \quad (2.14)$$

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

$$p_i(t + \delta t/2) = p_i(t) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t)) \quad (2.15)$$

$$q_i(t + \delta t) = q_i(t) + \delta t * p_i(t + \delta t/2) \quad (2.16)$$

$$p_i(t + \delta t) = p_i(t + \delta t/2) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t + \delta t)) \quad (2.17)$$

154 Note that  $\frac{\partial}{\partial q_i}$  is the gradient with respect to  $q$  where  $i$  denotes the  $i$ th atom being

155 moved. Using this notation the gradient is

$$\vec{\nabla}U = \begin{bmatrix} \frac{\partial U}{\partial q_{0,x}} & \frac{\partial U}{\partial q_{0,y}} & \frac{\partial U}{\partial q_{0,z}} \\ \vdots & \frac{\partial U}{\partial q_{i,w}} & \vdots \\ \frac{\partial U}{\partial q_{n,x}} & \frac{\partial U}{\partial q_{n,y}} & \frac{\partial U}{\partial q_{n,z}} \end{bmatrix} = \begin{bmatrix} \vec{\mathcal{F}}_0 \\ \vdots \\ \vec{\mathcal{F}}_i \\ \vdots \\ \vec{\mathcal{F}}_n \end{bmatrix} \quad (2.18)$$

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

174 **No-U-Turn Sampling**

175 Two parameters must be specified in HMC simulations, the step size  $\delta$  and the number  
176 of steps  $N$ . The step size is critical to the stability of the fitting procedure: with  
177 a too small  $\delta$  the simulation runs inefficiently producing structures too close to the  
178 previous, whereas with a too big  $\delta$  the linear approximation for the forces breaks  
179 down and often the simulated NP explodes. The number of steps to take during the  
180 dynamics is equally important and an inappropriate choice may result in backtracking  
181 or random walk characteristics in the simulations. In this work, we employ the No-  
182 U-Turn Sampling (NUTS) method recently proposed by Hoffman and Gelman to  
183 address this issue [23]. In the NUTS method  $\delta$  and  $N$  are dynamically computed  
184 by examining the ratio of accepted to rejected configurations as well as whether  
185 or not the simulation has started to take a U-turn. The U-Turn criteria makes  
186 certain that the simulation stops when it begins to backtrack, preventing excess  
187 computation on configurations that have very little new information to offer. The  
188 use of NUTS leaves us with two simulation parameters: the simulation temperature  
189 and the target acceptance. Hoffman and Gelman have empirically shown that the  
190 ideal target acceptance, which governs the dynamics time steps, is .65, which we have  
191 used for all of the simulations here. The simulation temperature sets the magnitude  
192 of the random starting momenta for the atoms at the beginning of each dynamics  
193 run [23].

194 **Grand Canonical Ensemble**

195 While NUTS-HMC simulations provide a system to find minima on PESs, the sim-  
196 ulation is fundamentally run in the Canonical Ensemble thus the variables in the  
197 simulation are limited to a fixed number of particles, simulation volume, and thermal  
198 energy. Fixing the thermal energy and simulation volume is not a problem, as they  
199 are not variables of interest in the final structure. However, specifying the number of

200 atoms in the system can be problematic, as the exact number of atoms in a sample  
201 can be difficult to count or a sample could have a distribution of particle sizes. Thus,  
202 a new ensemble needs to be used to allow the number of atoms to vary as a function  
203 of the PES. This new ensemble is the Grand Canonical Ensemble.

204 **Ensemble description**

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

$$\Xi = \sum_{N=0}^{\infty} Q(N, V, T) \exp \frac{N\mu\beta}{T} \quad (2.19)$$

209 where  $Q(N, V, T)$  is the Canonical partition function discussed above,  $\mu$  is the chem-  
210 ical potential. [36] This is translated into a Monte Carlo system, producing Grand  
211 Canonical Monte Carlo (GCMC).

212 **Grand Canonical Monte Carlo**

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

$$\min[1, \frac{V}{(N+1)\Lambda(T)^3} e^{-\beta\Delta U + \beta\mu}] \quad (2.20)$$

216 Similarly the removal of an atom has the probability:

$$\min[1, \frac{(N)\Lambda(T)^3}{V} e^{-\beta\Delta U - \beta\mu}] \quad (2.21)$$

217 However, both of these equations depend of the overall simulation volume and the  
218 thermal wavelength, which is undesirable as these are not really properties that we  
219 are of interest to these simulations. Thus, we roll them into the definition of the

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

### 233 GCMC biasing

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

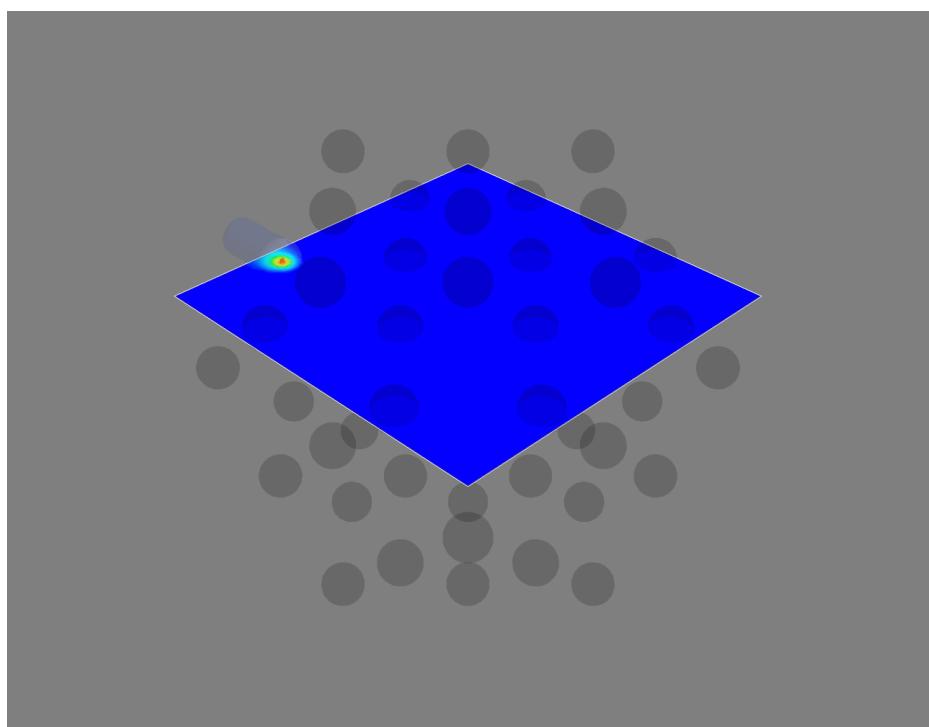
$$P_{i,j,k} = \frac{e^{\beta\Delta U_{i,j,k}}}{\sum_{i,j,k} e^{\beta\Delta U_{i,j,k}}} \quad (2.22)$$

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. [49]

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

249        Figure 2.1a shows an example map for atomic addition in a Au54 atom system,  
250    with an Au55 atom target. Figure 2.1b shows the results of a few GCMC insertions  
251    with biasing, showing the focusing of the simulation on the missing atom. The high  
252    density of insertions around the missing atom would not have been possible without  
253    the biasing.

(a)



(b)

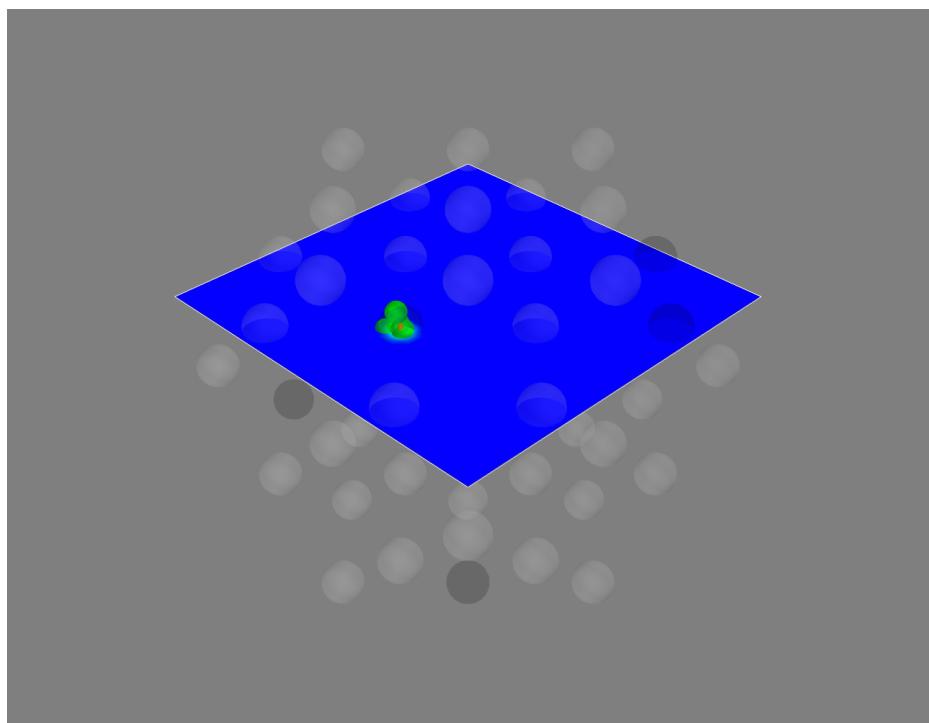


Figure 2.1: These figures show slices of the three dimensional addition potential energy surface. a) shows the addition probability, with the red area most probable. b) shows the results of trial insertions, where the green spheres denote where insertions were attempted.

254 2.4 CONCLUSIONS

255 In this chapter we have presented the development of both PES and the statistical  
256 mechanical ensembles used to search them. We expanded the classical concept of  
257 a PES to a more general mapping from positional variable space to energy space.  
258 This expansion allowed for the implementation of experimentally derived PES, where  
259 the disagreement between experimental and computed results can be included in the  
260 PES. Common experimental PESs were discussed, and their forces derived. The  
261 implementation of various statistical mechanical ensembles, used for searching the  
262 PES for minima, was also discussed with a special focus on No-U-Turn-Sampling  
263 Hamiltonian Monte Carlo. Grand Canonical Monte Carlo was also discussed, with  
264 an emphasis on the us of biasing to increase the overall acceptance rate. Future  
265 work in this area may include the development of PESs which leverage 2 dimensional  
266 data, like STEM images, or ensembles which help to eliminate tuned parameters like  
267 parallel tempering.

268

## CHAPTER 3

269

### ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION

271 3.1 INTRODUCTION

272 Atomistic structural insight is essential for understanding and controlling a mate-  
273 rial's properties and functions, which has led to some of the most exciting advances  
274 in modern materials science and engineering. X-ray diffraction techniques are one of  
275 the most powerful tools for probing atomic structures with ultimate precision. Tradi-  
276 tionally, thousands of diffraction peaks are analyzed using refinements of a structural  
277 model with few parameters to determine the 3D structure of bulk single crystals with  
278 high precision [20]. However, real engineered materials differ from ideal single crystals  
279 by showing a complexity in morphology, crystallite size, and atomic structure. The  
280 X-ray Powder Diffraction method (XPD) is among the most widely used methods  
281 for solving the structure of micro-crystals. The XPD technique utilizes hundreds of  
282 diffraction peaks and constrains the refinement of the structural model to few pa-  
283 rameters in order to resolve the structure [41]. Solving the atomic structure becomes  
284 difficult using traditional x-ray diffraction techniques when the size of the material or  
285 its important features is reduced to the nanometer scale with non-periodic or short-  
286 periodic atomic arrangements. Materials consisting of particles with sizes less than  
287 a few tens of nanometers, often called nanoparticles (NPs), are structurally more  
288 complex than their bulk cousins. This structural complexity is often attributed to  
289 the large number of surface atoms which have incomplete coordination spheres [50],

290 surface relaxation [24], and surface environment effects [45, 22, 33]. These effects  
291 make the precise determination of 3D atomic structure of NPs far more complicated  
292 and problematic [3].

293 Over the years many advances have been made to address the famous “nanostruc-  
294 ture problem”[3], for example, by using the atomic Pair Distribution Function (PDF)  
295 analysis of x-ray and neutron total scattering data [16, 39, 46], PDF analysis combined  
296 with molecular dynamics simulations [53, 21], bulk crystallography approaches [25],  
297 and others. Attempts with non-diffraction based approaches have also been made,  
298 including Transmission Electron Microscopy (TEM) [11], Raman spectroscopy[28],  
299 Extended X-ray Absorption Fine-Structure Spectroscopy (EXAFS)[19] and Nuclear  
300 Magnetic Resonance (NMR) [1].

301 In this chapter the PDF and its gradients will be derived. These expressions,  
302 when combined with the PES and statistical mechanical treatment from chapter 2,  
303 will allow for the solution of atomic structures. This chapter will also develop a  
304 computational framework for evaluating the PDF and its gradients using Graphical  
305 Processing Units (GPUs) to enable fast structural solution.

### 306 3.2 THEORY

307 To properly understand the PDF and its limitations we need to derive its mathemat-  
308 ics. The PDF has been previously derived many times so it is not re-derived here.  
309 This discussion of the PDF and its gradients use the notation of Farrow and Billinge.  
310 [17]

### 311 Derivation

312 Many of the above techniques require the gradient of the PES. This in turn requires  
313 the gradient of the PDF to be derived. Mathematically treating thermal vibrations  
314 will also be discussed in this section. Systems which are truly extended materials, like

315 powders with particle sizes larger than 10nm, are best formulated as systems with  
 316 periodic boundaries. Thus, the equations for a periodically bound PDF need to be  
 317 developed as well, with their gradients.

## 318 Analytically Gradients

319 Many optimization algorithms and simulations methodologies, including HMC, re-  
 320 quire not only the potential energy of a given configuration but also the forces acting  
 321 on that configuration. These forces are described by the gradient of potential energy  
 322 of the system which in turn requires the gradient of the PDF. As previously shown the  
 323 PDF is the Fourier Transform of the Debye equation. Since the Fourier Transform is  
 324 expressed as an integral we can exchange the order of the gradient and the integral,  
 325 allowing us to calculate the analytical gradient of the Debye equation and FFT the  
 326 resulting function. The Debye equation, with a Debye-Waller vibrational correction  
 327 is

$$F(Q) = \frac{1}{N\langle f \rangle^2} \sum_{j \neq i} f_i^*(Q) f_j(Q) \exp(-\frac{1}{2}\sigma_{ij}^2 Q^2) \frac{\sin(Qr_{ij})}{r_{ij}} \quad (3.1)$$

328 where

$$\sigma_{ij}^2 = (\vec{u}_{ij} * \hat{d}_{ij})^2 \quad (3.2)$$

$$\vec{u}_{ij} = \vec{u}_i - \vec{u}_j \quad (3.3)$$

$$\hat{d}_{ij} = \frac{\vec{d}_{ij}}{r_{ij}} \quad (3.4)$$

$$r_{ij} = \|\vec{d}_{ij}\| \quad (3.5)$$

$$\vec{d}_{ij} = \begin{bmatrix} q_{ix} - q_{jx} \\ q_{iy} - q_{jy} \\ q_{iz} - q_{jz} \end{bmatrix} \quad (3.6)$$

329 where  $Q$  is the scatter vector,  $f_i$  is atomic scattering factor of the  $i$ th atom, and  $r_{ij}$  is  
 330 the distance between atoms  $i$  and  $j$  and has  $q$  dependence. [26] For simplicity's sake

331 we will break up  $F(Q)$  so that

$$F(Q) = \alpha \sum_{j \neq i} \beta_{ij} \tau_{ij} \Omega_{ij} \quad (3.7)$$

332 where

$$\alpha = \frac{1}{N \langle f \rangle^2} \quad (3.8)$$

$$\beta_{ij} = f_i^*(Q) f_j(Q) \quad (3.9)$$

$$\tau_{ij} = \exp\left(-\frac{1}{2} \sigma_{ij}^2 Q^2\right) \quad (3.10)$$

$$\Omega_{ij} = \frac{\sin(Qr_{ij})}{r_{ij}} \quad (3.11)$$

333 The derivatives are as follows:

$$\frac{\partial}{\partial q_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \left( \frac{\partial \tau_{ij}}{\partial q_{i,w}} \Omega_{ij} + \tau_{ij} \frac{\partial \Omega_{ij}}{\partial q_{i,w}} \right) \quad (3.12)$$

334 where

$$\frac{\partial \Omega_{ij}}{\partial q_{i,w}} = \frac{Q \cos(Qr_{ij}) - \Omega_{ij}}{r_{ij}^2} (q_{i,w} - q_{j,w}) \quad (3.13)$$

$$\frac{\partial \tau_{ij}}{\partial q_{i,w}} = \frac{\sigma_{ij} Q^2 \tau_{ij}}{r_{ij}^3} ((q_{i,w} - q_{j,w}) \sigma_{ij} - (u_{i,w} - u_{j,w}) r_{ij}^2) \quad (3.14)$$

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

336 Thus

$$\frac{\partial}{\partial u_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \frac{\partial \tau_{ij}}{\partial u_{i,w}} \Omega_{ij} \quad (3.15)$$

$$\frac{\partial \tau_{ij}}{\partial u_{i,w}} = -\frac{\sigma_{ij} Q^2 \tau_{ij}}{r_{ij}} (q_{i,w} - q_{j,w}) \quad (3.16)$$

337 **Without ADPs**

338 Without ADPs the equations simplify down to

$$F(Q) = \frac{1}{N \langle f \rangle^2} \sum_{j \neq i} f_i^*(Q) f_j(Q) \frac{\sin(Qr_{ij})}{r_{ij}} \quad (3.17)$$

339 and

$$\frac{\partial}{\partial q_{i,w}} F(Q) = \alpha \sum_j \beta_{ij} \frac{\partial \Omega_{ij}}{\partial q_{i,w}} \quad (3.18)$$

340 use of these equations, when ADPs are not appropriate (like at cryogenic tempera-  
341 tures), greatly speeds up the computation.

## 342 Periodic Boundary Conditions

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

$$F(Q) = \frac{1}{N\langle f \rangle^2} \left( \sum_{j \neq i} f_i^*(Q) f_j(Q) \frac{\sin(Qr_{ij})}{r_{ij}} + \sum_{i,j} f_i^*(Q) f_j(Q) \frac{\sin(QR_{ij})}{R_{ij}} \right) \quad (3.19)$$

349 where

$$R = |\vec{r} + \vec{\nu}| \quad (3.20)$$

$$\vec{\nu} = \gamma_1 * \vec{a} + \gamma_2 * \vec{b} + \gamma_3 * \vec{c} \quad (3.21)$$

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

### 352 3.3 COMPUTATION

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

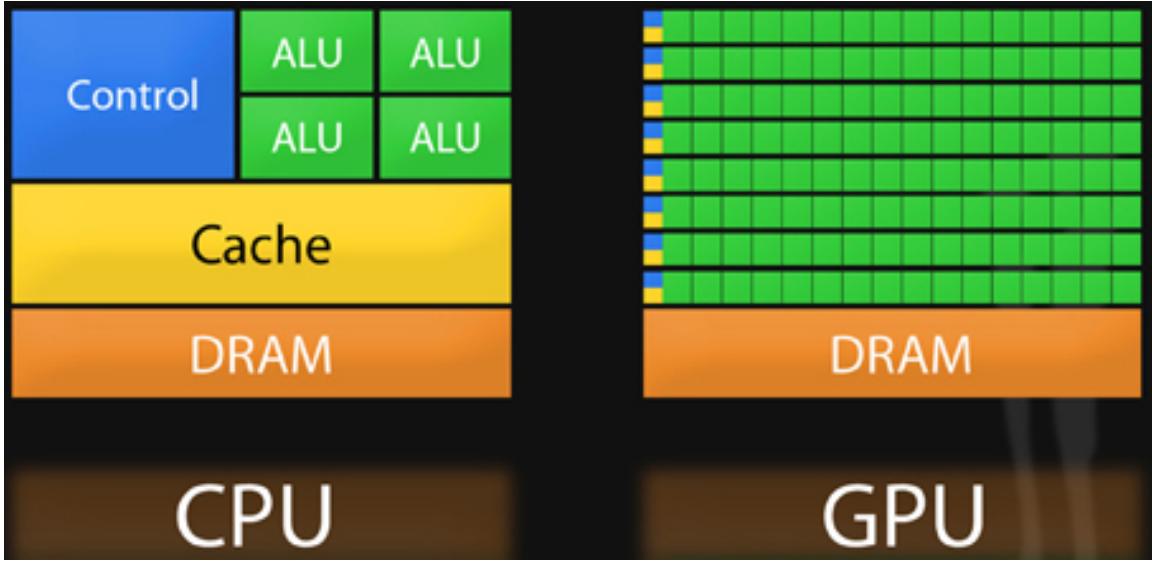


Figure 3.1: Comparison of the CPU and GPU chip architectures from [5]. The ALU are the arithmetic logic units which perform the mathematical operations, the DRAM holds most of the data, although it is slower to access, the Cache holds rapidly accessed data, and the Control controls the execution of software. 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.

### 357 **HPC and GPUs**

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

### 362 **GPUs and Parallelization**

363 Computing the PDF is an embarrassingly parallel problem. The basic procedure is  
 364 to calculate the reduced structure factor  $F(Q)$  for each atom pair and momentum  
 365 transfer vector, sum over all the atom pairs, and Fourier transform the structure to  
 366 the PDF. The first part of this procedure is perfectly parallelizable, as each atom pair  
 367 is separate from the others. The summation over all the atomic reduced structure fac-  
 368 tors can be parallelized via distributed summing. Lastly the FFT can be parallelized

369 using existing parallel FFT algorithms.

370 Graphical Processing Units (GPUs) are particularly well suited to the task of  
371 computing PDFs. GPU chip architecture is designed to perform many task simulta-  
372 neously by having potentially thousands of cores. Figure 3.1 show the comparison  
373 of CPU and GPU architectures. As the figure shows the GPUs have a very different  
374 layout of computational processors (ALUs) and memory. While each ALU is simpler  
375 on the GPU, requiring the instructions to be less demanding in terms of memory,  
376 there are many more of them. The greater number of processors allows each atomic  
377 pairing to be placed on its own processor, so long as the math can be broken into sim-  
378 pler operations. The equations are broken up on the GPUs into various pieces which  
379 correspond to the  $\alpha, \beta, \tau$  and  $\Omega$  as shown in equations 3.8-3.11 and sub-equations as  
380 needed. For example, while  $\beta$  is computed in one step,  $\Omega$  requires the computation  
381 of the displacement array, then the distance array and finally the  $\Omega$  array. The exact  
382 breakdown of processes, how the problems are broken down and spread across the  
383 processor has been optimized for speed and reliability.

384 **Map from ij space to k space**

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

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

388

389  $E$  denotes the atomic coordinates in  $ij$  space,  $E'$  denotes  $F(Q)$  before the summation

390 in  $ij$  space,  $B$  denotes the atomic pairs in  $k$  space,  $B'$  denotes  $F(Q)$  in  $k$  space, and  
 391  $Z$  denotes the final summed  $F(Q)$ . For the operators,  $\phi$  denotes the mapping from  
 392  $ij$  space to  $k$  space  $k = j + i * \frac{i-1}{2}$ ,  $\psi$  and  $\psi'$  denote the  $F(Q)$  operation in  $ij$  and  $k$   
 393 space, respectively.  $\Sigma$  denotes the sum over all the atoms.

394 To properly define  $\Sigma'$  we must establish whether  $F(Q)$  is an even function. We  
 395 can accomplish this by examining each of the portions of  $F(Q)$ ,  $\alpha, \beta, \tau, \Omega$ .  $\Omega$  is even,  
 396 since  $r_{ij}$  is the interatomic distance, which is the same despite a flip of indices,  $Q$   
 397 does not depend on the atomic indices, and since  $Qr_{ij}$  is even so is  $\sin Qr_{ij}$ . Thus,  
 398  $\Omega$  is even. Providing similar analysis to  $\tau$  we can see that while  $\vec{u}_{ij}$  is odd, so is  
 399 the unit displacement vector between the two atoms, thus the two odds cancel out.  
 400 Intuitively this makes sense, since the  $F(Q)$  equation is fundamentally interested  
 401 in the interatomic distances which is even. Thus, switching atom indices does not  
 402 change  $F(Q)$ . Due to the even nature of the  $F(Q)$  operator the  $\Sigma'$  operator sums  
 403 over all the atom pairs, and multiplies by two to reflect the double counting of the  $\Sigma$   
 404 operator.

For the gradient a similar mapping is used:

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

405

406 In this mapping, however, we use the  $\tilde{\phi}\Sigma$  operator. This operator simultaneously  
 407 performs a reverse mapping from  $k$  to  $ij$  space, and a summation with the correct  
 408 symmetry. In this case the  $\psi$  and  $\psi'$  operators, which denote the  $\vec{\nabla}F(Q)$  operator  
 409 in  $ij$  and  $k$  space, are antisymmetric. Intuitively this makes sense as an extension  
 410 of Newton's Second Law, since each particle's interaction is felt oppositely by its  
 411 partner.

412 GPU Memory Allocation

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

$$q[=]3n \quad (3.22)$$

$$d[=]3k \quad (3.23)$$

$$r[=]k \quad (3.24)$$

$$scatter[=]nQ \quad (3.25)$$

$$normalization[=]kQ \quad (3.26)$$

$$\Omega[=]kQ \quad (3.27)$$

$$F_k(Q)[=]kQ \quad (3.28)$$

$$Sum[=]kQ \quad (3.29)$$

$$Sum2[=]kQ \quad (3.30)$$

$$F(Q)[=]Q \quad (3.31)$$

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

$$k_{perGPU} = \frac{1}{16Q + 16} (-4Qn - 4Q + am - 12n) \quad (3.32)$$

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

$$adps = 3n \quad (3.33)$$

$$\sigma = k \quad (3.34)$$

$$\tau = kQ \quad (3.35)$$

427 Thus the pair allotment is:

$$k_{perGPU} = \frac{-4Qn - 4Q + am - 24n}{20Q + 20} \quad (3.36)$$

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

$$\vec{\nabla}\Omega = 3kQ \quad (3.37)$$

$$\vec{\nabla}F_k(Q) = 3kQ \quad (3.38)$$

$$\vec{\nabla}F_n(Q) = 3nQ \quad (3.39)$$

430 Thus the gradient allotment is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 12n}{32Q + 16} \quad (3.40)$$

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

$$\vec{\nabla}\tau = 3kQ \quad (3.41)$$

432 Thus the allocation is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 24n}{48Q + 20} \quad (3.42)$$

433 These equations were solved by sympy as their validity is very important to the overall  
 434 reliability of the software. If the GPU is over allocated then the system may crash  
 435 or return meaningless results.

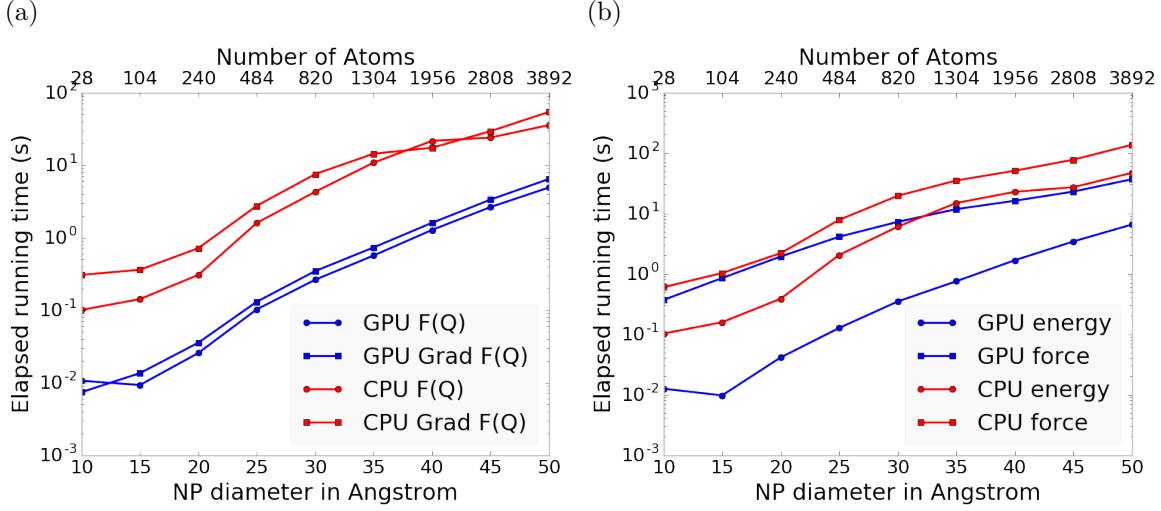


Figure 3.2: Speed comparison of CPU and GPU implementations. a) shows the time to compute the  $F(Q)$  by itself. b) shows the time to compute the  $Rw$  based energy for Au NPs of various sizes, which includes computing  $F(Q)$ , its FFT, and the  $Rw$ .

## 436 Speed and Scaling of PDF Computation

437 To understand exactly how much the GPUs speed up the computation of  $F(Q)$  and  
 438 the PDF a series of time studies were run Au nanoparticles of varying size. Figure  
 439 3.2 shows the results of these time studies. CPU and GPU calculations were carried  
 440 out on an Intel i7-4820K @3.70GHz Quad-Core and one NVIDIA GTX970s, respec-  
 441 tively. The  $F(Q)$  computations show a 100x to 10x speedup using the GPUs over  
 442 the CPUs. Additionally, the  $\vec{\nabla}F(Q)$  and  $F(Q)$  computations seem to have similar  
 443 computation time and scaling relationships on the GPU. This implies that the two  
 444 processes may have similar bottlenecks, most likely in the  $F(Q)$  computation work-  
 445 flow. This relationship is similarly preserved, although to a lesser extent, in the CPU  
 446 scaling.

447 Interestingly, the tight run time relationship between  $F(Q)$  and its gradient are  
 448 not preserved in the  $Rw$  based force calculations. While the energy calculations are  
 449 very similar to the  $F(Q)$  calculations in terms of run time, the GPU and CPU force  
 450 calculations are much closer, with the GPU calculations being much slower. This is

451 due to the force bottleneck being the  $3n$  FFT operations which must be performed  
452 on the  $\vec{\nabla}F(Q)$  array to produce the  $\vec{\nabla}\text{PDF}$  array. While the GPU is leveraged  
453 to perform the FFT, the data must be loaded off the GPU and back on, causing a  
454 potential slowdown. Larger systems of atoms were not tried as the CPU computation  
455 quickly becomes very slow. Even higher GPU speedup is expected on more advanced  
456 GPUs like the NVIDIA Tesla series.

457 3.4 CONCLUSIONS

458 In this chapter we developed the gradients of the PDF in the discrete and periodic  
459 boundary condition case. We also developed the computational implementation of  
460 the PDF equations. This implementation emphasized use of GPUs to compute the  
461 PDF and its gradient. The GPU software was further sped up by mapping the com-  
462 putation to atom pairs rather than atom by atom. Finally, the speed of the GPU  
463 implementation was checked against the CPU implementation via speed benchmark-  
464 ing.

465

## CHAPTER 4

466

### BENCHMARKING

467

Also some introduction would be great

468

this just needs a lot of work

469 4.1 INTRODUCTION

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

474 4.2 PDF

475 The formation of NPs with both crystallographic and non-crystallographic structures  
476 [33] and with different chemical patterns [18] are well documented. For simplicity,  
477 we chose monometallic Au clusters as benchmarks and considered two groups of  
478 structures with different size and degrees of structural disorder in order to assess  
479 the reliability and efficiency of our HMC method for solving atomic structures from  
480 PDFs. The first group consists of  $\text{Au}_{55}$  clusters with different degrees of disorder,  
481 including a crystalline cluster structure in  $O_h$  (Octahedral) symmetry, a structure  
482 with a disordered surface, and an amorphous structure. The second group consists  
483 of the crystallographically solved  $\text{Au}_{102}$  structure as in the  $\text{Au}_{102}\text{MBA}_{44}$  nanocrystals  
484 [25, 32]. We used optimized structures from the Density Functional Theory (DFT)

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

$$G_{\text{obs}} = \frac{2}{\pi} \int_{Q_{\min}}^{Q_{\max}} Q[S_{\text{obs}}(Q) - 1] \sin(Qr) dQ \quad (4.1)$$

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

491 Spin-polarized DFT calculations were carried out using the Vienna ab initio simu-  
492 lation package (VASP) [30, 31] within the Perdew-Burke-Ernzerhof (PBE) exchange-  
493 correlation functional [42]. The projected augmented wave method [4] and a kinetic  
494 energy cutoff of 400 eV were used. Structural optimization was performed until the  
495 total energy and ionic forces were converged to  $10^{-6}$  eV and 10 meV/Å, respectively.  
496 The amorphous  $\text{Au}_{55}$  structures were generated by simulated annealing using the  
497 classical embedded atom method potential [48]. Different annealing temperatures  
498 between 1200 K and 1670 K (bulk melting temperature of Au) were used and the  
499 thermally equilibrated structures were cooled down to 300 K before minimization at  
500 0 K. Further optimization using DFT leads to total energies that vary within 1-2  
501 eV among different amorphous structures and the lowest energy one was used as the  
502 target structure. The target structure of  $\text{Au}_{102}$  was taken as the  $\text{Au}_{102}$  core of the  
503 DFT-optimized  $\text{Au}_{102}\text{MBA}_{44}$  cluster [32].

504 All systems were refined using a PES which consists of a linear combination of  
505  $Rw$ , the repulsive and attractive thresholded spring potentials. The total potential  
506 energy in the Hamiltonian in Eq. (2.9) is expressed as:

$$U(q) = U_{Rw}(q) + U_{\text{spring}}(q, R_{\min}) + U_{\text{spring}}(q, R_{\max}) \quad (4.2)$$

507 The thresholded spring potentials are based on those previously proposed on by Pe-  
508 terson [43], 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

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

515 Whereas the choice of the absolute values of  $\lambda$  and  $\kappa$  is somewhat arbitrary, their  
 516 relative values are important in determining which term in Eq. (4.2) dominates the  
 517 PES, especially when considering the effect of the simulation temperature. Generally,  
 518 the ratio between the total potential energy and the temperature determines how  
 519 much random motion will dominate the dynamics; a lower ratio implies that random  
 520 motion will play a large role in the dynamics. The ratio between  $\lambda$  and  $\kappa$  of each  
 521 spring describes how far the PDF can push the system below or above the bounds set  
 522 by the spring potentials. Heuristically, too stiff a spring forbids the system to access  
 523 new configurations, e.g. high energy “transition states” which may involve shorter  
 524 bonds or a larger system size. Conversely, too small a spring constant makes it slower  
 525 for the system to snap back within bounds and may lead to an explosion or implosion  
 526 of the system, leaving the dynamics to drift aimlessly.

## 527 Model Parameters

528 Unless otherwise stated, the PDFs of the target and starting structures were generated  
 529 using Eqn. (4.1) with a step of  $\delta R = .01 \text{ \AA}$ ,  $Q_{\min} = 0.1 \text{ \AA}^{-1}$ ,  $Q_{\max} = 25.0 \text{ \AA}^{-1}$ .  $R_{\min}$   
 530 and  $R_{\max}$  correspond to the first minimum before the first PDF peak and that after  
 531 the last PDF peak, respectively, which ensure that the full meaningful region of the  
 532 PDF is modeled. For each of the simulations, the Q resolution was calculated by

$$\delta Q = \frac{\pi}{R_{\max} + \frac{12\pi}{Q_{\max}}} \quad (4.3)$$

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

539      **Au55: surface relaxed**

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

547      **Run Parameters**

548       $R_{\min}$  and  $R_{\max}$  for this simulation were 2.45 Å and 11.4 Å, respectively, with  $\delta Q =$   
549      0.24 Å<sup>-1</sup>. The simulation ran for approximately 34 minutes, over a total of ~40  
550      thousand configurations. The results are shown in Fig. 4.1.

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

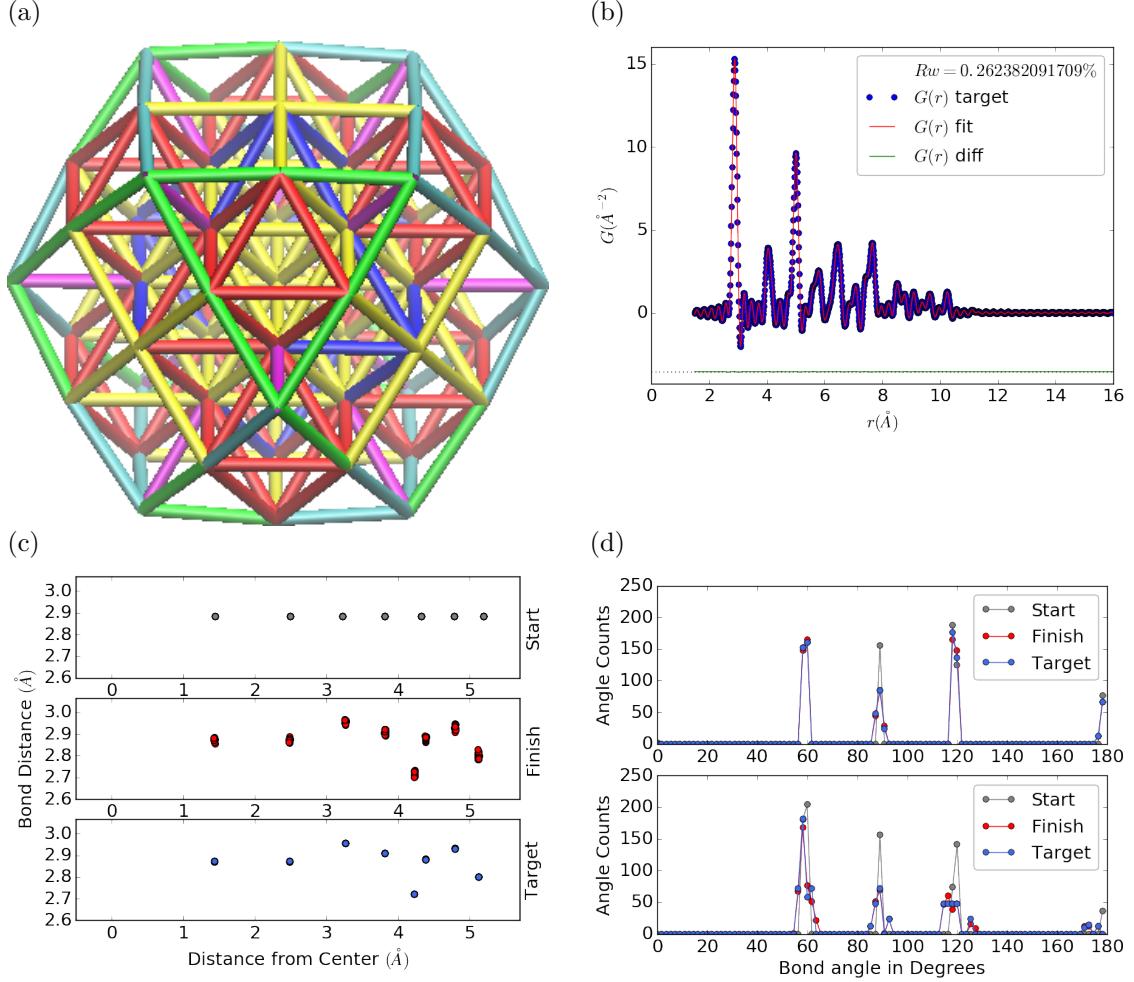


Figure 4.1:  $\text{Au}_{55}$  PDF fitting of DFT-optimized  $c\text{-Au}_{55}$ . (a) the final structural solution ( $Rw=0.3\%$ ) with bond lengths color-coded by step of  $0.05\text{\AA}$ , (b) the target PDF(blue dots) overlaid with the PDF of the final structure (solid red lines) with the difference in green lines offset below, ?? the radial bond distribution, and (d) bond angle distribution.

558 below, the situation is rather different for much more disordered target structures.  
 559 Interestingly, the small-box solution using PDFgui[16] yields a rather large  $Rw$  of  
 560 43%, due to the failure to fit the surface contracted atoms with a unit cell. The PDF  
 561 fits of the starting structure and small-box solution are shown

562 Put this somewhere

563 .

564 **Au55: surface disordered**

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

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

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

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

591 and high CN numbers. This is partly caused by the displacement of the core atoms  
592 mentioned above, and partly by the lack of CN constraints for the PDF fitting, which  
593 has been previously demonstrated in the case of  $\alpha$ -Si [9]. Additional experimental  
594 data, e.g. from EXAFS or NMR, may help to steer the simulations towards better  
595 agreement in both PDF and CN distribution.

## 596 Au55: amorphous

597 Next, we turn to the case in which the entire cluster structure is disordered. We used  
598 a DFT-optimized amorphous  $\text{Au}_{55}$  ( $a\text{-Au}_{55}$ ) as the target structure, and the DFT-  
599 relaxed  $c\text{-Au}_{55}$  cluster from Case I as the starting structure. The total energy of  
600  $a\text{-Au}_{55}$  was computed to be *lower* than that of  $c\text{-Au}_{55}$  by as large as 2.9 eV, consistent  
601 with the 3.0 eV found in previous DFT work [12].

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

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

617 **Au102: triple phase**

618 Our final benchmark is Au<sub>102</sub>, whose structure was initially solved by Jadzinsky and  
619 co workers using x-ray crystallography [25] and further confirmed by DFT studies  
620 [32]. The Au<sub>102</sub> structure consists of three main parts, a 49-atom Marks decahedron  
621 core, two  $C_5$  caps consisting of 20 atoms each, and 13 equatorial atoms. Unlike  
622 previous cases, the multi-symmetry nature of the structure, i.e. each part has its own  
623 distinct symmetry, poses a challenge for PDF-based solution of the structure. This is  
624 because of the atomically centralized nature of the PDF, in which each atom “sees”  
625 a density of other atoms surrounding it and has a strong tendency towards becoming  
626 the center of the main symmetry group. Such tendency may lead to a solution where  
627 some of the correct atomic symmetries are discarded in favor of the core symmetry.

628 **Starting from fcc structure**

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

631  $R_{\min}$  and  $R_{\max}$  for this simulation were 2.7 Å and 16. Å, respectively, with  $\delta Q =$   
632 0.18 Å<sup>-1</sup>. The simulation ran for approximately two hours, over a total of ~82  
633 thousand configurations. The results of the simulation are shown in Fig. 4.4.

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

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

653 **Marks decahedron**

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

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

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

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

672 **Au147**

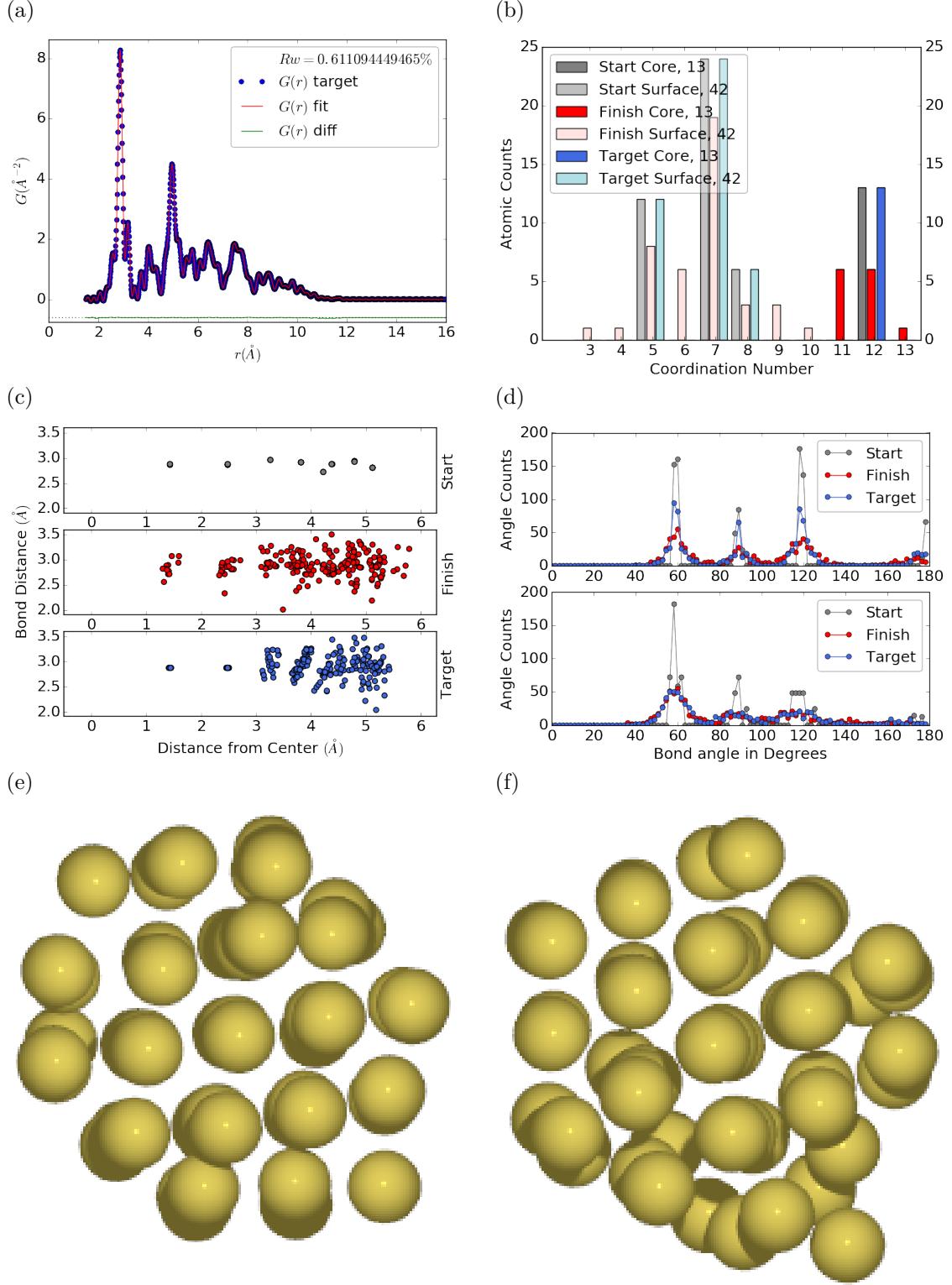


Figure 4.2:  $\text{Au}_{55}$  PDF fitting of surface-disordered  $\text{Au}_{55}$ . a) the target structure, b) the final structural solution ( $Rw=0.6\%$ ), c) the comparison of PDFs, d) the CN distribution with the number of atoms in either the core or the surface, e) the radial bond distribution, and f) the bond angle distribution.

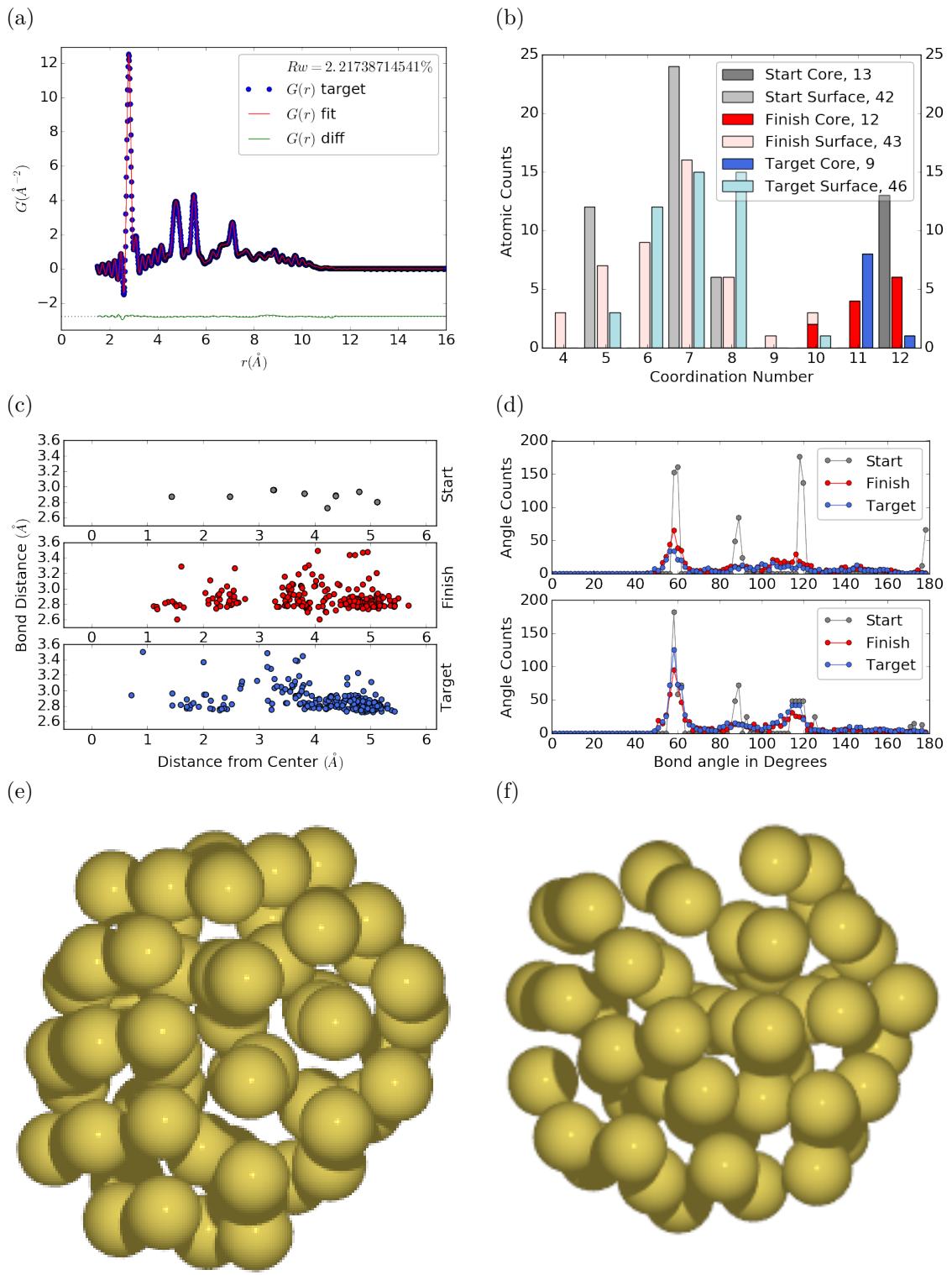


Figure 4.3: Similar to figure 4.2 for DFT-optimized amorphous Au<sub>55</sub>.

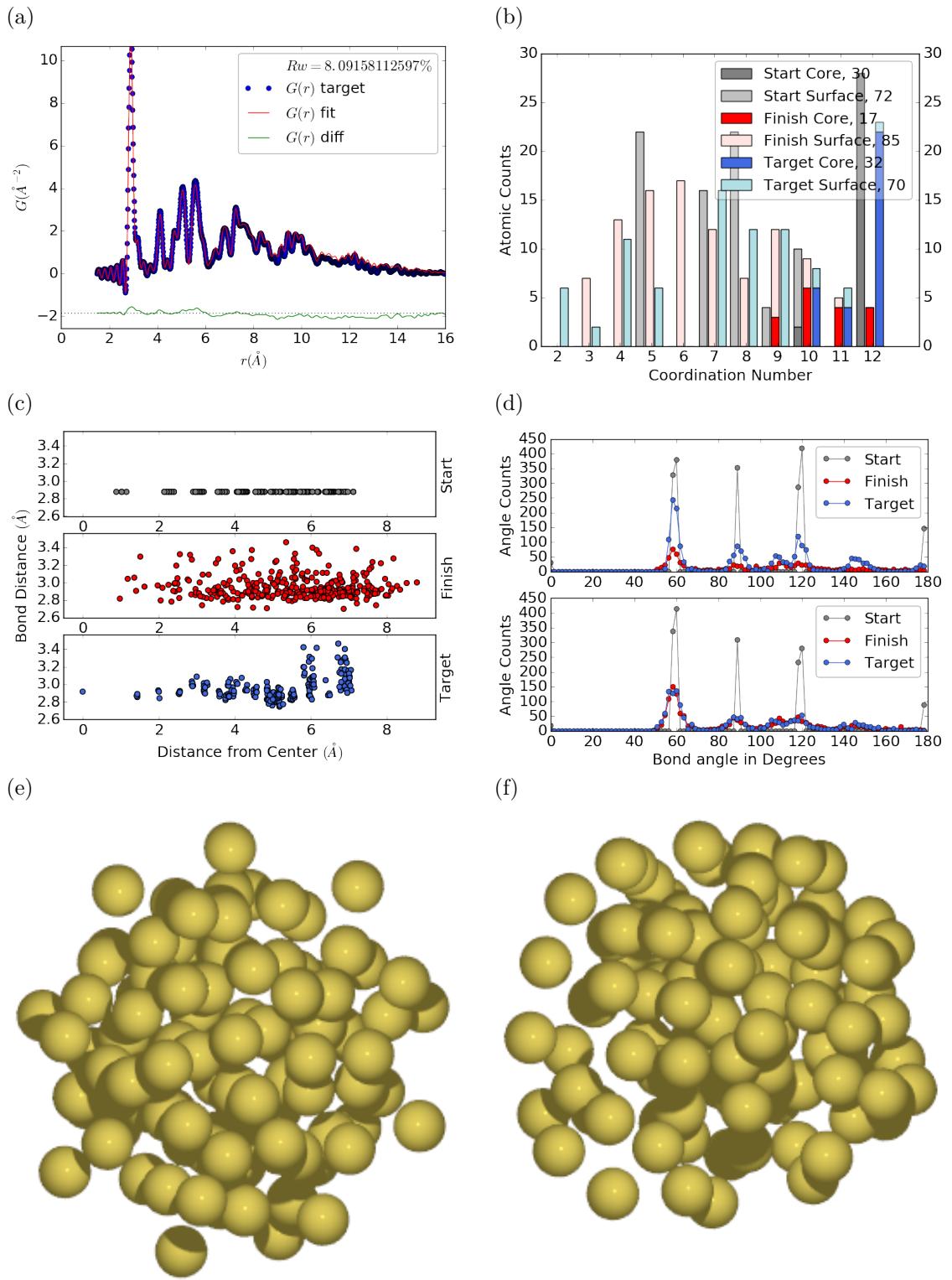


Figure 4.4: Similar to figure 4.2 for  $\text{Au}_{102}$  as in DFT-optimized  $\text{Au}_{102}\text{MBA}_{44}$  cluster.

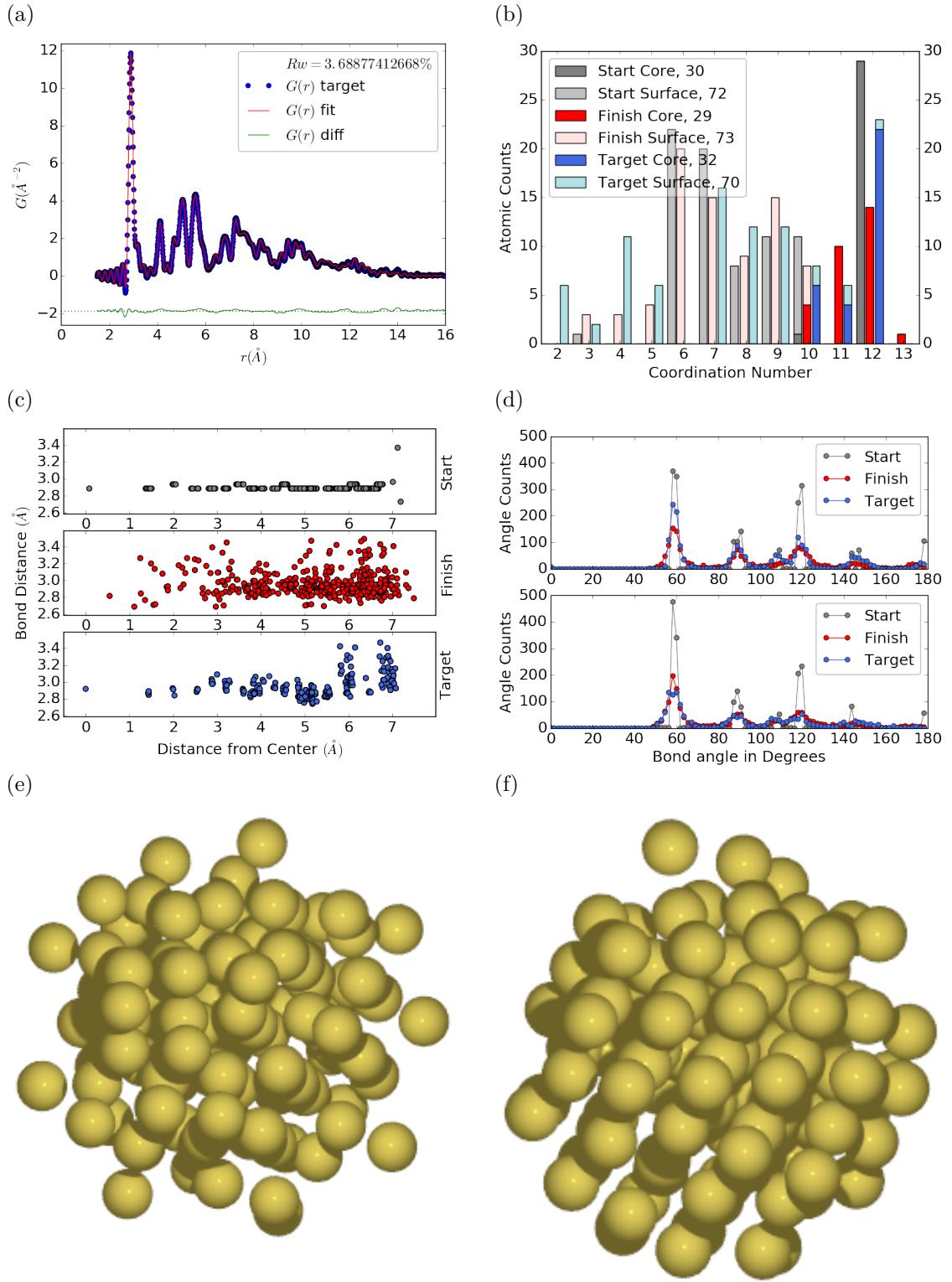


Figure 4.5: Similar to Fig. 4.4 with Marks decahedron as the starting structure.

## 673 4.3 PDF WITH ADPs

## 674 ADP 50

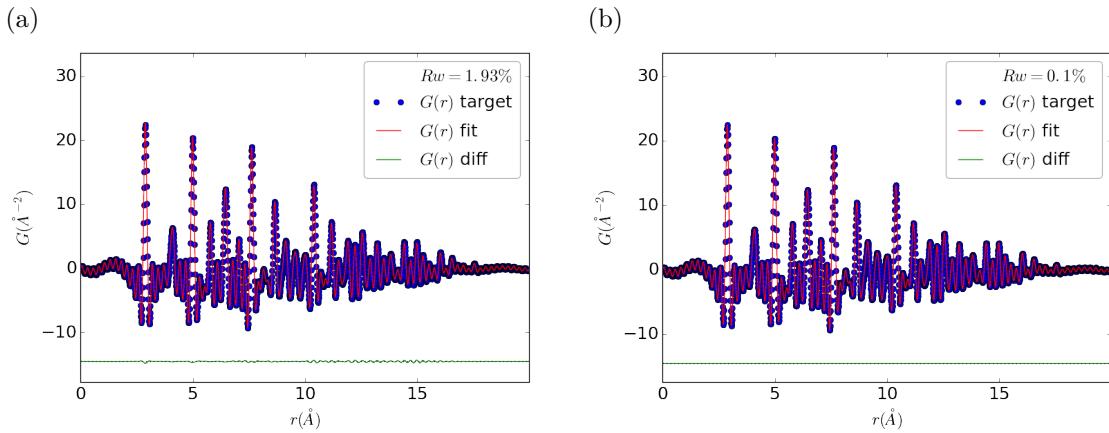


Figure 4.6: Refinement of adps

675 1. Basic 50% larger magnitude

676 2. Random addition to APDs

677 3. Janus ADPs

678

## CHAPTER 5

679

# X-RAY TOTAL SCATTERING DATA ACQUISITION AND PROCESSING

681 5.1 INTRODUCTION

682 X-ray total scattering experiments are generally performed at synchrotron light sources,  
683 as only these sources can provide the needed flux, energy, and high momentum trans-  
684 fer vectors needed to obtain reliable PDFs. [8, 15] Despite the need for a dedicated  
685 facility to perform the total scattering experiments, the experiments themselves are  
686 fairly forgiving, allowing for reactive gaseous environments, experiment temperatures  
687 ranging from 2 K to 1800 K, and even electrochemical cycling. [7, 45, 47] The rapid  
688 PDF data acquisition associated with 2D area detectors creates a data management  
689 problem, as 96 hours of beamtime could result in almost 10,000 images which need  
690 to be associated with the experimental conditions and detector metadata. [8] Finally,  
691 all this data needs to be processed by masking bad pixels and regions, integrating  
692 azimuthally, and converting the scattering data to the PDF. [29, 27, 51, 40, 2]

693 5.2 DETECTOR  $Q$  RESOLUTION

694 To properly azimuthaly integrate the images taken from the detector the  $Q$  resolution  
695 of the pixels must be calculated. Azimuthal integration is the process of deviding the  
696 pixels up into “bins”. Each bin has a set width, usually in  $Q$ , which describes which  
697 pixels can go into the bin. During the integration process each pixel is placed into  
698 its correspondng bin. Finally a statistical measurement of the average of all the

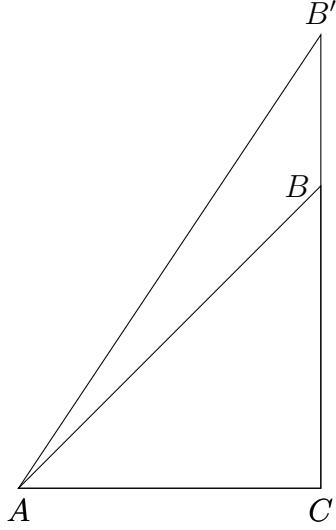


Figure 5.1: Scattering onto a flat detector

699 pixels in each bin is taken to produce the  $I(Q)$  data. Although commonly performed,  
 700 integrating using evenly spaced bins will cause pixels which are not on the same ring  
 701 to be binned together, causing the incorrect value of  $I(Q)$  to be obtained and a larger  
 702 standard deviation in the integrated data. To properly calculate the  $Q$  resolution  
 703 the resolution of each of the pixels in  $2\theta$  must be calculated. Figure 5.1 shows the  
 704 scattering of x-rays onto a flat image plate detector. In this diagram the bottom of the  
 705  $n$ th pixel is  $B$  while the top is  $B'$ . The resolution of this pixel in  $2\theta$  is  $\angle BAC - \angle B'AC$ .  
 706 Thus the resolution, calculated from the distances is

$$\Delta 2\theta = \arctan \frac{b}{d} - \arctan \frac{t}{d} \quad (5.1)$$

707 where  $d$  is the sample to detector distance ( $AC$  in figure 5.1),  $b$  is the distance to  
 708 the bottom of a pixel ( $CB$  in figure 5.1)), and  $t$  is the distance to the top of that  
 709 pixel ( $CB'$  in figure 5.1)). Note that these distances need to have been corrected for  
 710 detector tilt and rotation. Thus the resolution of a pixel in  $Q$  is

$$\Delta Q = \frac{4\pi(\sin \arctan \frac{b}{d} - \sin \arctan \frac{t}{d})}{\lambda} \quad (5.2)$$

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

712 This effect is analagous to looking at windows on a very tall building. The windows  
713 are all the same size, but due to the nature of the perspecitve, the windows seem to  
714 shrink as one looks higher up the building.

715 For a Perkin Elmer image plate, like the one used at the NSLS-II's XPD and the  
716 APS's 11-ID-B, the resolution function is shown in 5.2. For the same detector the  
number of pixels per  $Q$  is shown in 5.3

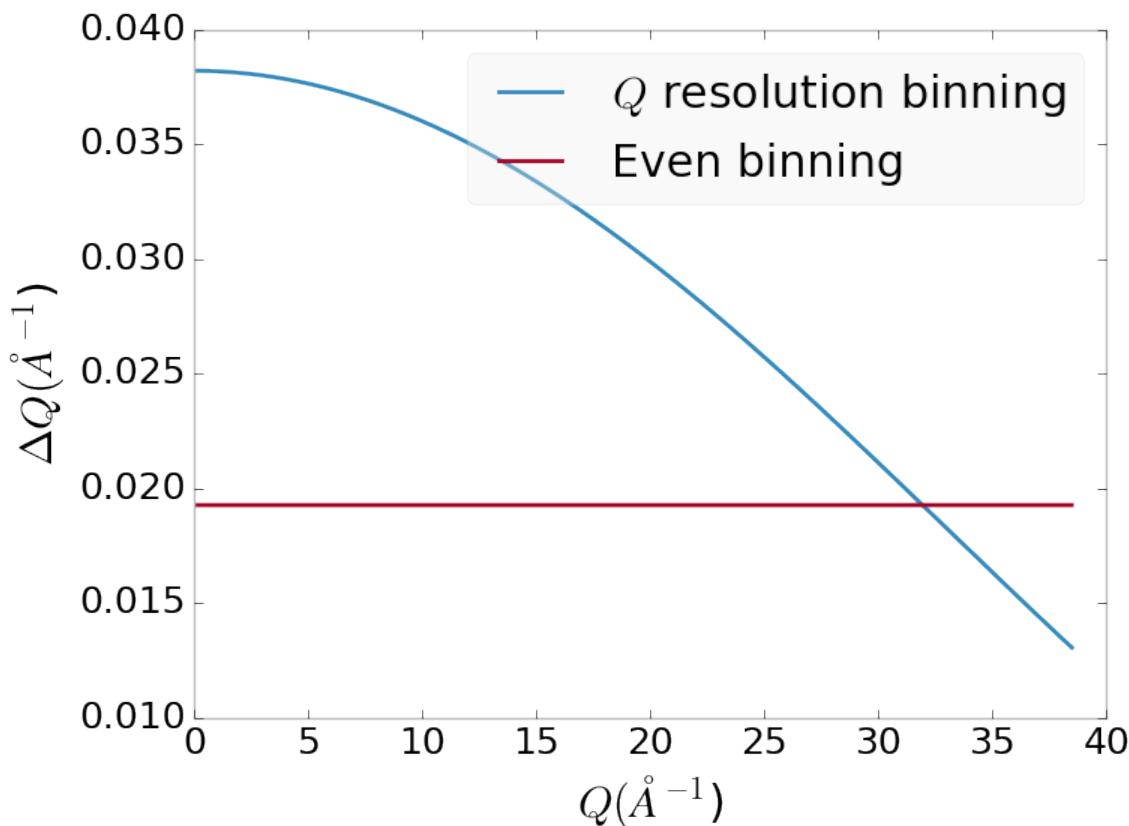


Figure 5.2:  $Q$  resolution as a function of  $Q$ .

717

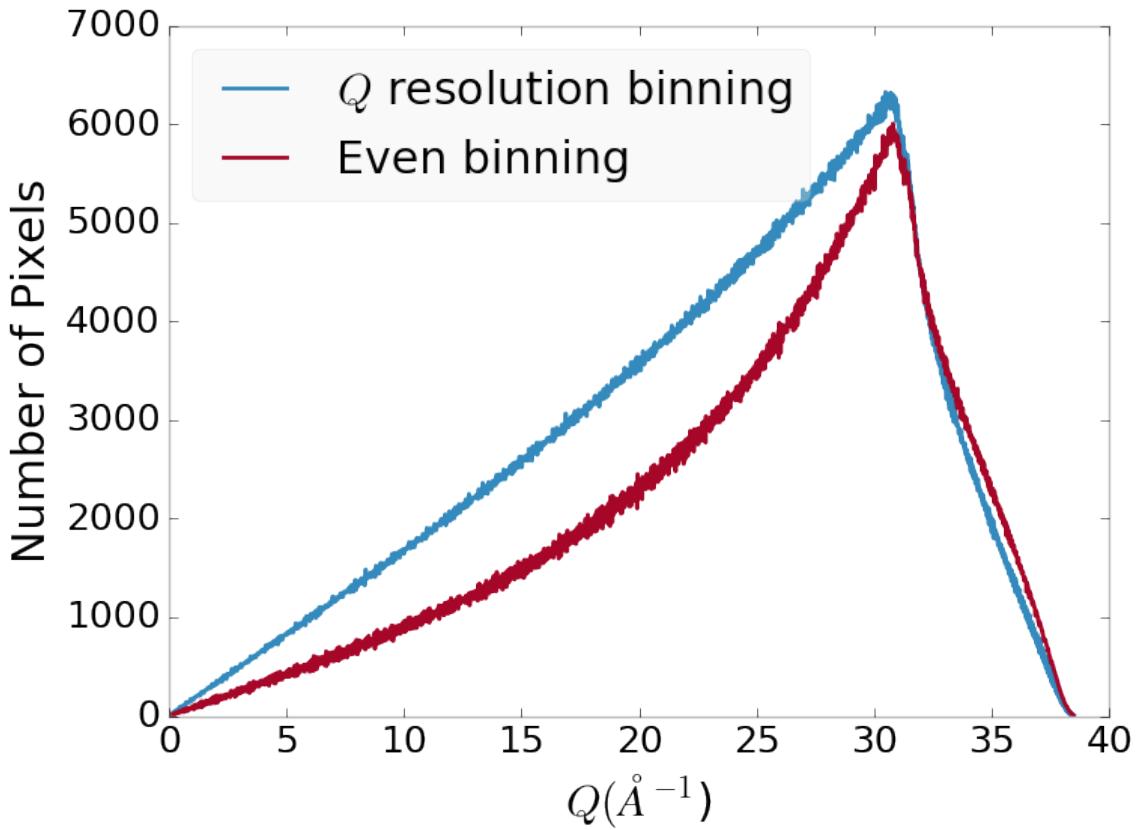


Figure 5.3: Number of pixels as a function of  $Q$ , binned at the  $Q$  resolution of the detector.

718 5.3 AUTOMATED MASK GENERATION

719 **Introduction**

720 Detector masking is an important part of any x-ray scattering workflow as dead/hot  
 721 pixels, streak errors, and beamstop associated features can be averaged into the data  
 722 changing the signal and its statistical significance. While some features, like the  
 723 beamstop holder, can be easily observed and masked by hand other are much more  
 724 difficult to observe even on large computer monitors. Additionally, while dead/hot  
 725 pixels and streaks are usually static the hot pixels associated with textured or sin-  
 726 gle crystal scattering or cosmic rays are not. Thus, coming up with an automated  
 727 method for finding such erroneous pixels is important, especially as high flux diffrac-

728 tion beamlines can generate data very quickly.

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

## 738 **Algorithm Design**

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

## 745 **Test Cases**

746 To study the effectiveness of the masking we ran the algorithm against both simulated  
747 experimental data. In the case of the simulated data four systems were created: 1)  
748 dead/hot pixels with varying numbers of defective pixels, 2) beamstop holder with  
749 varying beamstop holder transmittance, 3) rotated beamstop holder with varying  
750 beamstop holder transmittance, and 4) beamstop holder with dead/hot pixels. The

751 base scattering was produced by

$$I = 100 \cos(50r)^2 + 150 \quad (5.3)$$

752 where  $r$  is a pixel's distance from the beam point of incidence. The positions of  
753 the dead/hot pixels were chosen at random as was the dead or hot nature of the  
754 defect. Dead pixels had values from 0 to 10, while hot pixels had values from 200  
755 to 255. The beamstop was positioned at the vertical center of the detector with an  
756 initial width of 60 pixels and final width of 120 pixels. The height of the beamstop  
757 was 1024 pixels. The beamstop was calculated to attenuate the x-ray scattering  
758 signal at various transmittance, as various beamstop holder materials have different  
759 transmittance. Two version of the masking algorithm were run for each test case, one  
760 using the standard even bin sizes for the integration step, and one where the bin sizes  
761 are tuned to the pixel  $Q$  resolution as discussed in 5.2.

## 762 Results and Discussion

763 Figures 5.4-5.11 show the results of the masking algorithm on simulated images. The  
764 dead/hot pixel masking shows the importance of using the  $Q$  resolution based bin  
765 sizes as the even bin based mask have a tendency to over mask the image, removing  
766 pixels which contain valuable signal. This over-masking is caused by pixels being  
767 improperly associated with one another by the even bins. Figure 5.4 indicates that  
768 the masking algorithm, with the proper binning, masks the image perfectly, with no  
769 missed bad pixels or good pixels masked. This is not the case in figures 5.5 - 5.7 as  
770 we can see pixels which should have been masked but were not. Despite these missed  
771 pixels no pixels were improperly masked in any of the well binned images. These  
772 test cases are actually more difficult than experimental data, as the dynamic range  
773 of most detector causes the dead/hot pixels and single crystal/texture peaks to be  
774 orders of magnitude away from the desired signal.

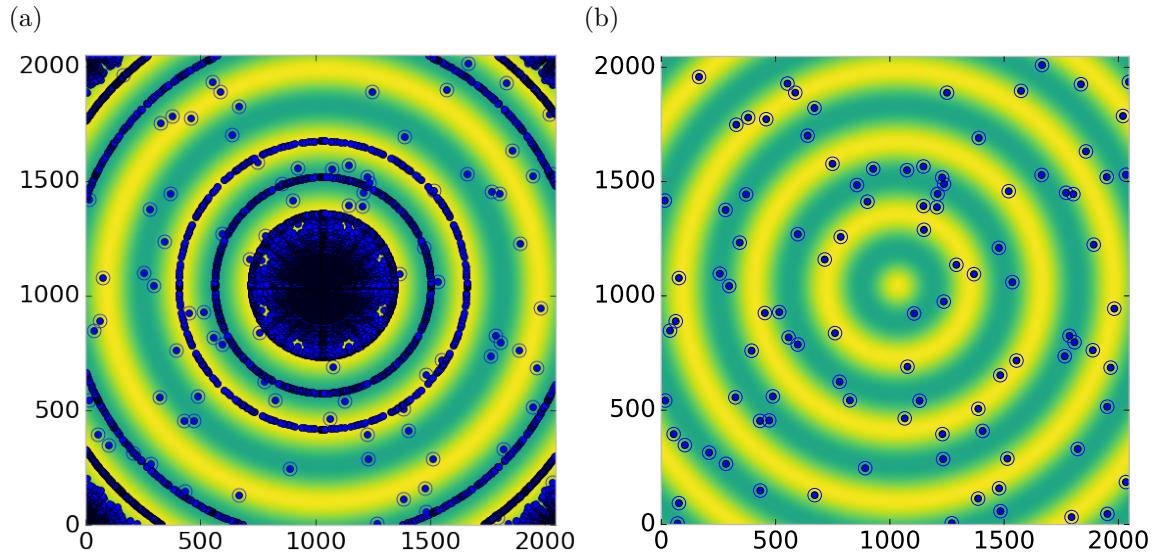


Figure 5.4: Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

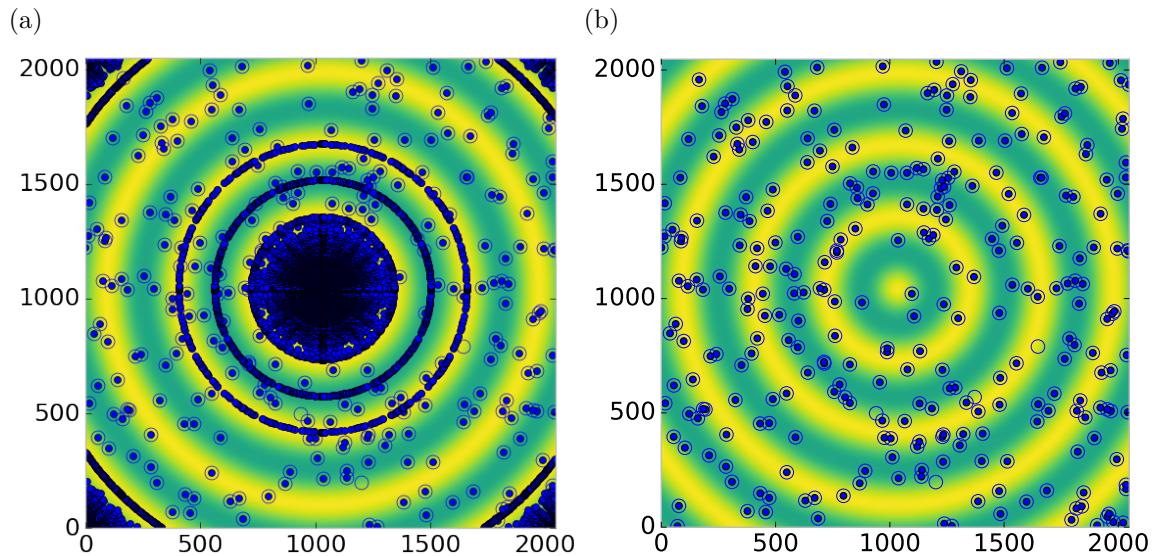


Figure 5.5: Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

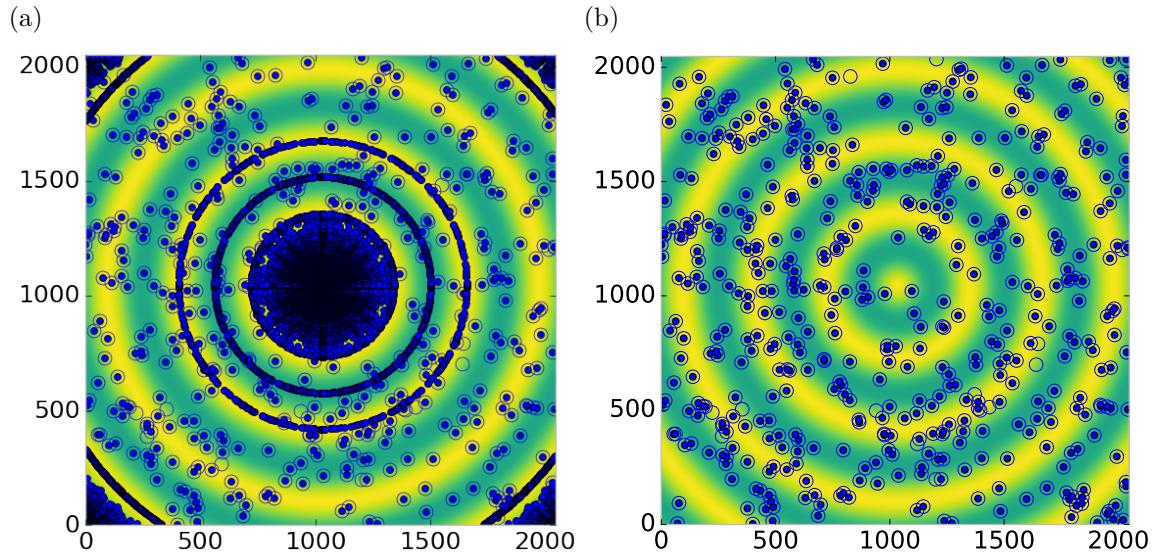


Figure 5.6: Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

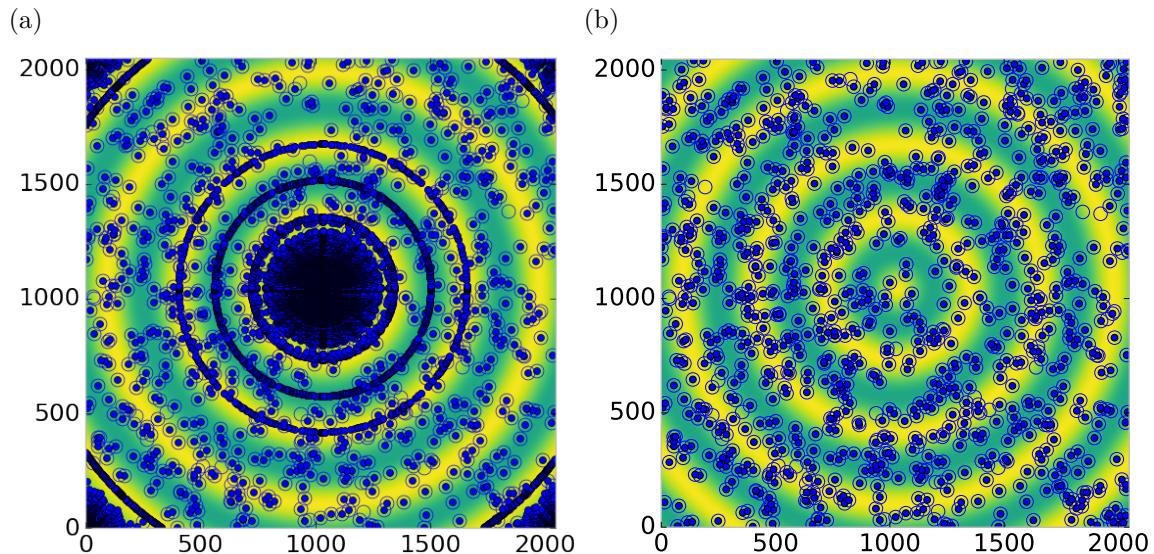


Figure 5.7: Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the standard even bin mask and b) the  $Q$  resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

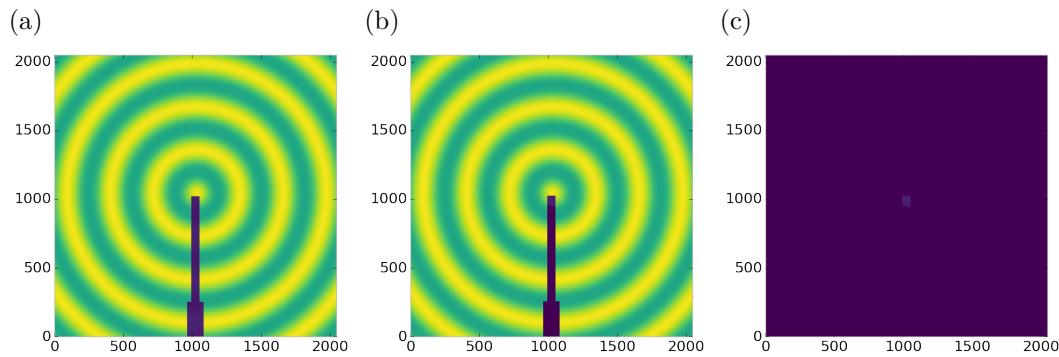


Figure 5.8: Generated beamstop holder masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

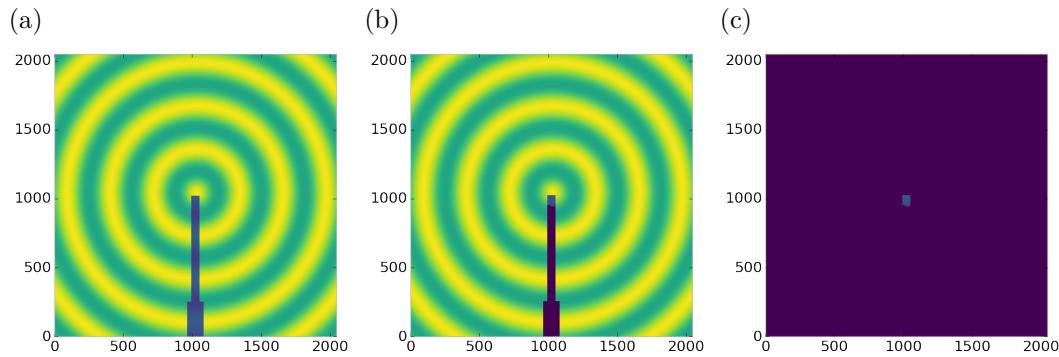


Figure 5.9: Generated beamstop holder masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

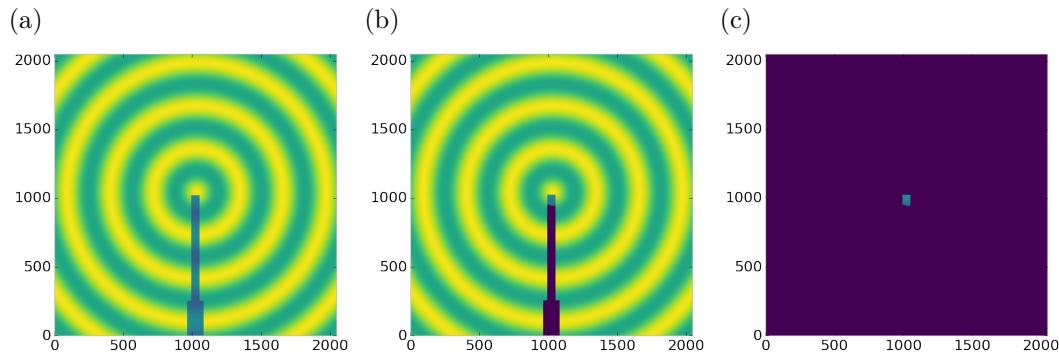


Figure 5.10: Generated beamstop holder masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

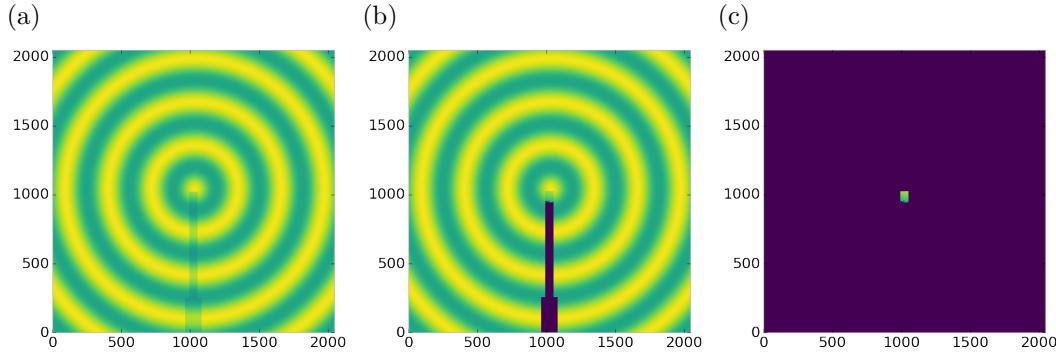


Figure 5.11: Generated beamstop holder masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

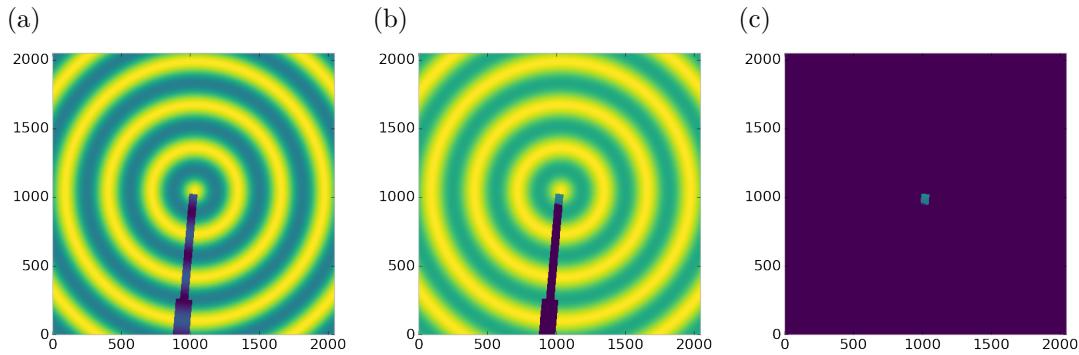


Figure 5.12: Generated beamstop holder masks which is rotated away from vertical

775 The beamstop holder masks shown in figures 5.8 - 5.11, which were all run with  
 776 the  $Q$  resolution binning show similar results across the transmittance range, missing  
 777 only a small part of the beamstop holder near the point of incidence. Near this point  
 778 the beamstop holder becomes a statistically significant part of the total number of  
 779 pixels in a given ring, thus it can not be masked out using a statistical search of the  
 780 rings. For most PDF and XRD studies this small area can be masked automatically  
 781 by masking all the pixels who's distance from the point of incidence is smaller than a  
 782 given radius  $r$ , or can be neglected outright as the area is not used in the analysis or  
 783 refinement. Similar results were produced for beamstop holders which were rotated  
 784 away from the vertical position, as shown in figure 5.12

785 Working with actual experimental data, obtained at the Advanced Photon Source  
 786 beamline 11-ID-B, shows the difficulty of masking images which have low photon

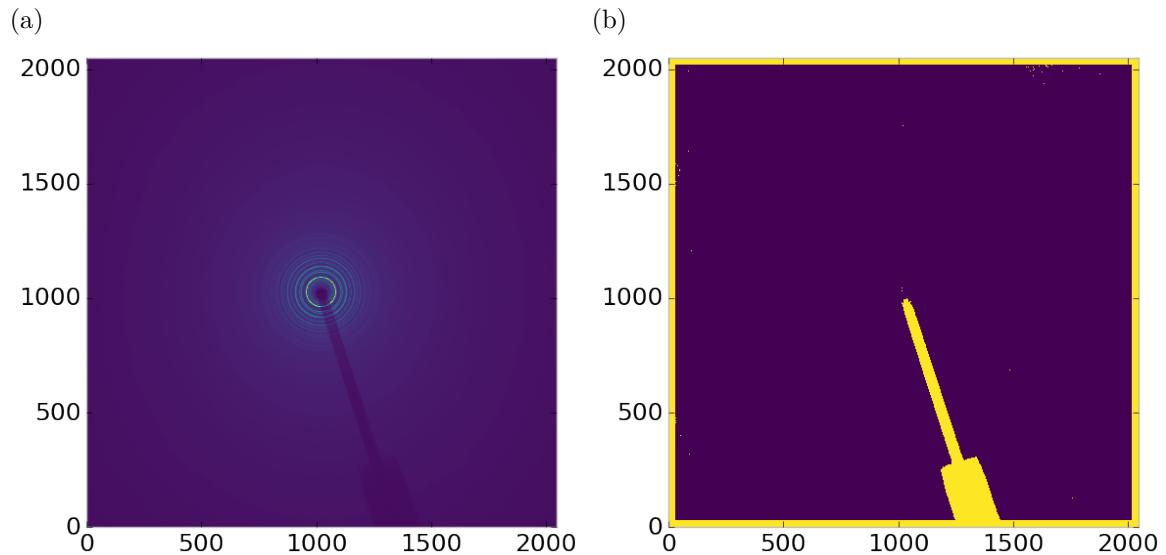


Figure 5.13: Masked experimental data. a) the raw image, b) the mask

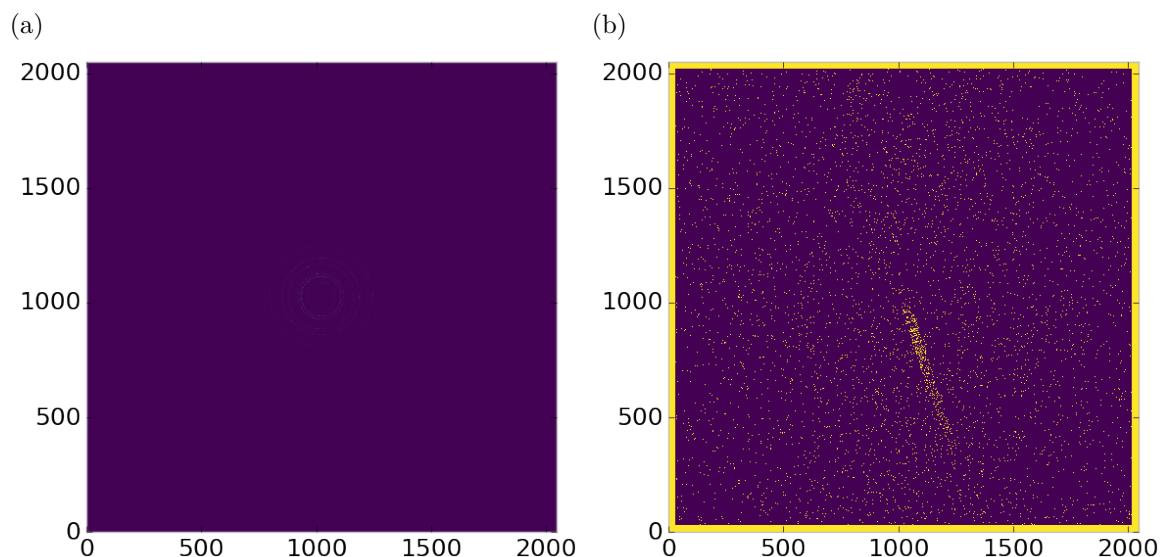


Figure 5.14: Masked experimental data with Pt single crystal signal. a) the raw image, b) the mask

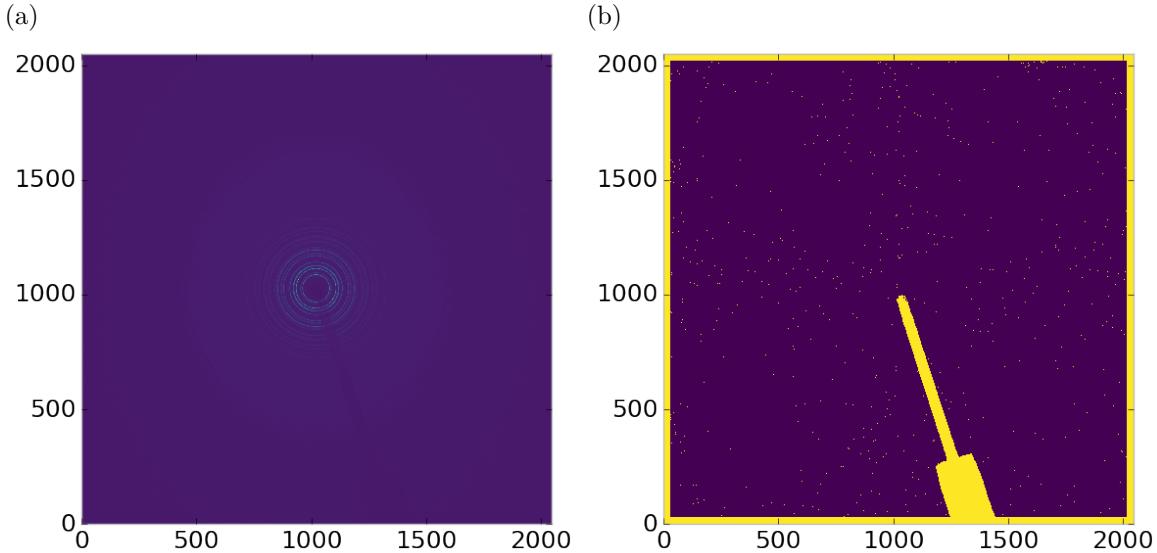


Figure 5.15: Masked experimental data with Pt single crystal signal using figure's 5.13 mask as a starting mask. a) the raw image, b) the mask

counts. While the masking of experimental data taken with longer exposures, consisting of 250 .2 second shots, shown in figure 5.13 provides very sharp edges to the beamstop holder, and very little extra masking beyond the occasional dead pixel, this is not the case for the single crystal data. The single crystal data is more problematic because of its short exposure time and low flux, with 500 frame at a .1 second exposure and having shrunk the beam size. The low flux is to prevent the very strong single crystal peaks from damaging the detector. However, this causes the image to be less statistically viable than ideal, causing problems with the mask as seen in figure 5.14. This can be alleviated to some degree by using the previously generated mask as a starting mask for the single crystal image, as shown in 5.15. While the masking algorithm still produces many diffuse masked pixels, they are far fewer, this may be due to the removal of the beamstop which could have contributed to the large standard deviation in figure 5.14.

800 **Conclusions**

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

813 **5.4 AUTOMATED IMAGE AZIMUTHAL INTEGRATION**

814 Using the  $Q$  resolution binning and masking developed in sections 5.2 and 5.3 the  
815 images can be properly integrated. Generally, images are integrated by taking the  
816 mean value of the pixels in a ring. However, other statistical measures of the average  
817 value can be used, like the median. Note that all the integrations done here use the  
818 pixels as they are, without pixel splitting, minimizing the covariance of the resulting  
819  $I(Q)$ .[51]

820 Figures 5.16-5.18 show the importance of masking and the choice of average func-  
821 tion. All the figures were produced using the same dataset, 50 °C  $\text{Pr}_2\text{NiO}_4$  taken at  
822 the APS's 11-ID-B on a Perkin Elmer area detector. The automatic masking alpha  
823 was 3 standard deviations from the mean. While it is difficult to observe the changes  
824 the mask causes in the full  $I(Q)$  plot (subfigures a) and b)), the standard deviation  
825 plots show the effect of bad pixels on the data (subfigure c)). Subfigure c) for figures

5.16-5.18 shows that removal of the beamstop holder lowers the low  $Q$  standard deviation from around .1 to almost .01 out to  $15 \text{ \AA}^{-1}$ . The high  $Q$  subfigures d) and f) in figures 5.16-5.18 show the “kink” effect of the detector edge and beamstop holder, where there is a dip in the  $I(Q)$  scattering when the rings include the edge of the detector. This effect seems to be due to both errors in the edge pixel intensity and the beamstop holder as masking of the edges only seems to provide only partial removal of the issue. It is important to note that while integration using the mean of the ring has issues with only the edge mask, as evidenced by the change in slope in 5.17 d) around  $29.5 \text{ \AA}^{-1}$ , the median integration does not include this error. Ideally the detector would have a normal distribution of pixel intensity for a given ring, which would imply an equivalency between the mean and median  $I(Q)$  values. Despite the closeness of the mean and median once the final mask has been created, it seems that the median is more reliable, as it was less effected by the beamstop holder in figure 5.17. Thus, for subsequent integrations discussed in this work the median is used to avoid any defective features that the masking algorithm may have missed.

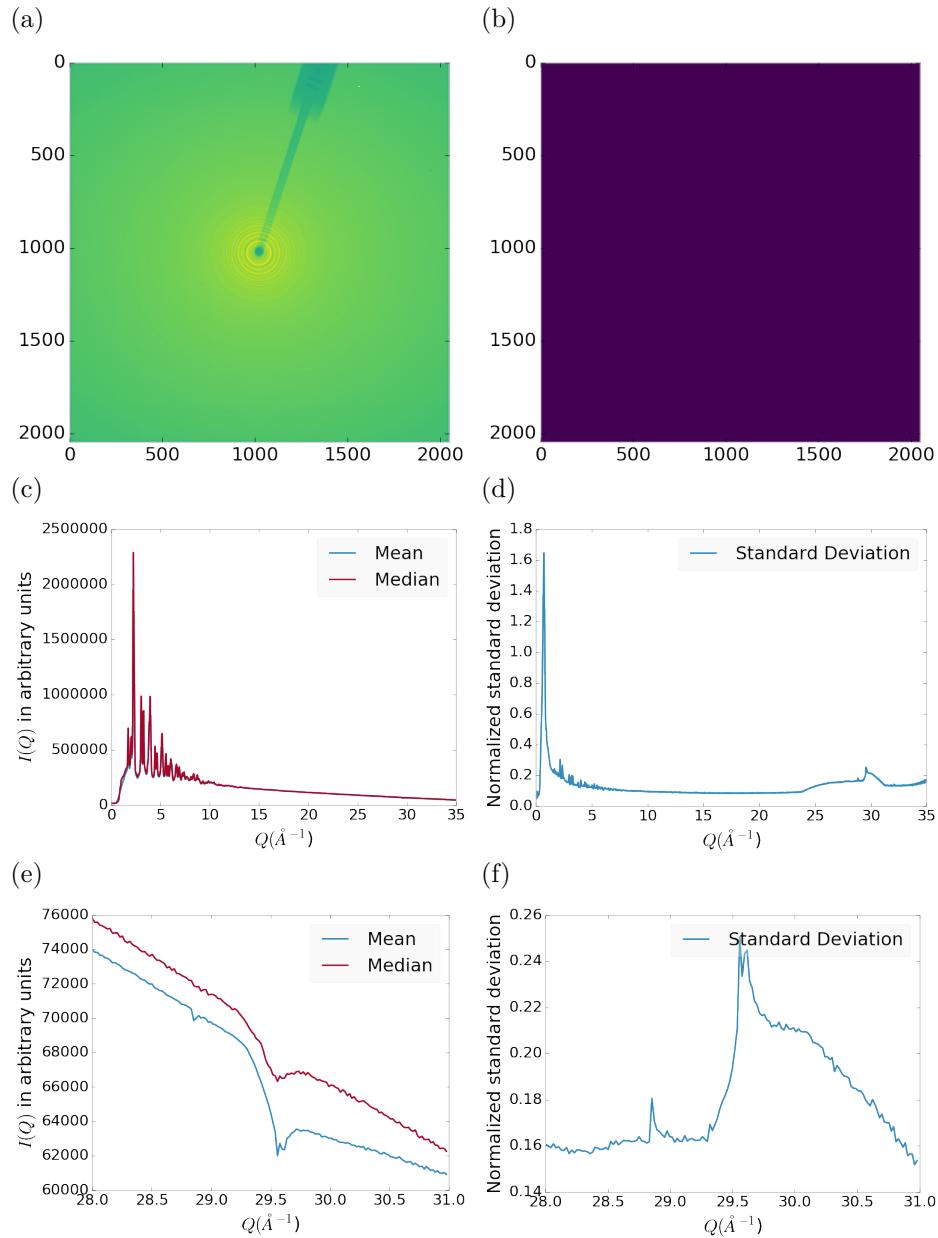


Figure 5.16: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with no mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{\AA}^{-1}$  to  $31 \text{\AA}^{-1}$   $Q$  range for the mean and median, f)  $28 \text{\AA}^{-1}$  to  $31 \text{\AA}^{-1}$   $Q$  range for the standard deviation

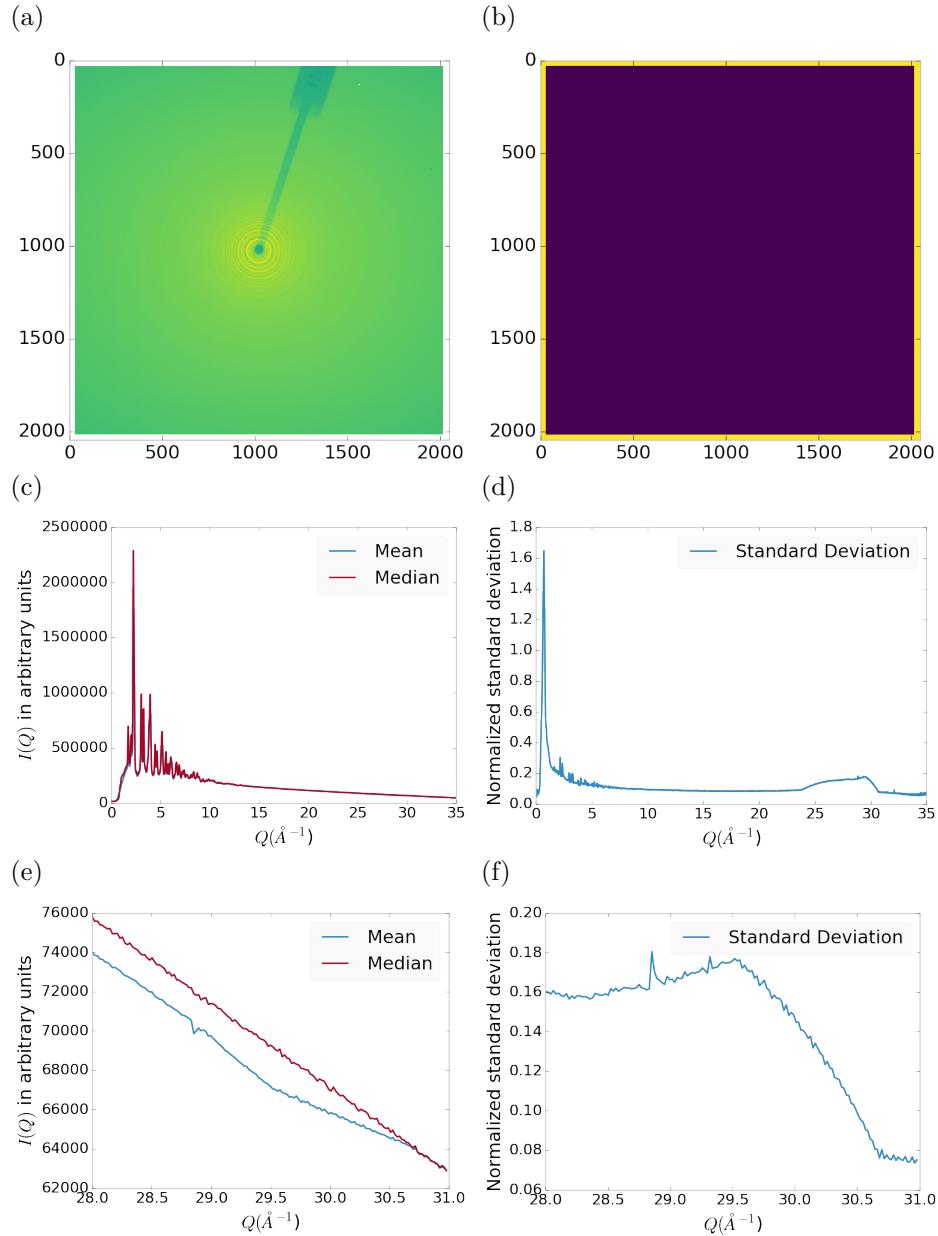


Figure 5.17: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with only an edge mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the  $28 \text{\AA}^{-1}$  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

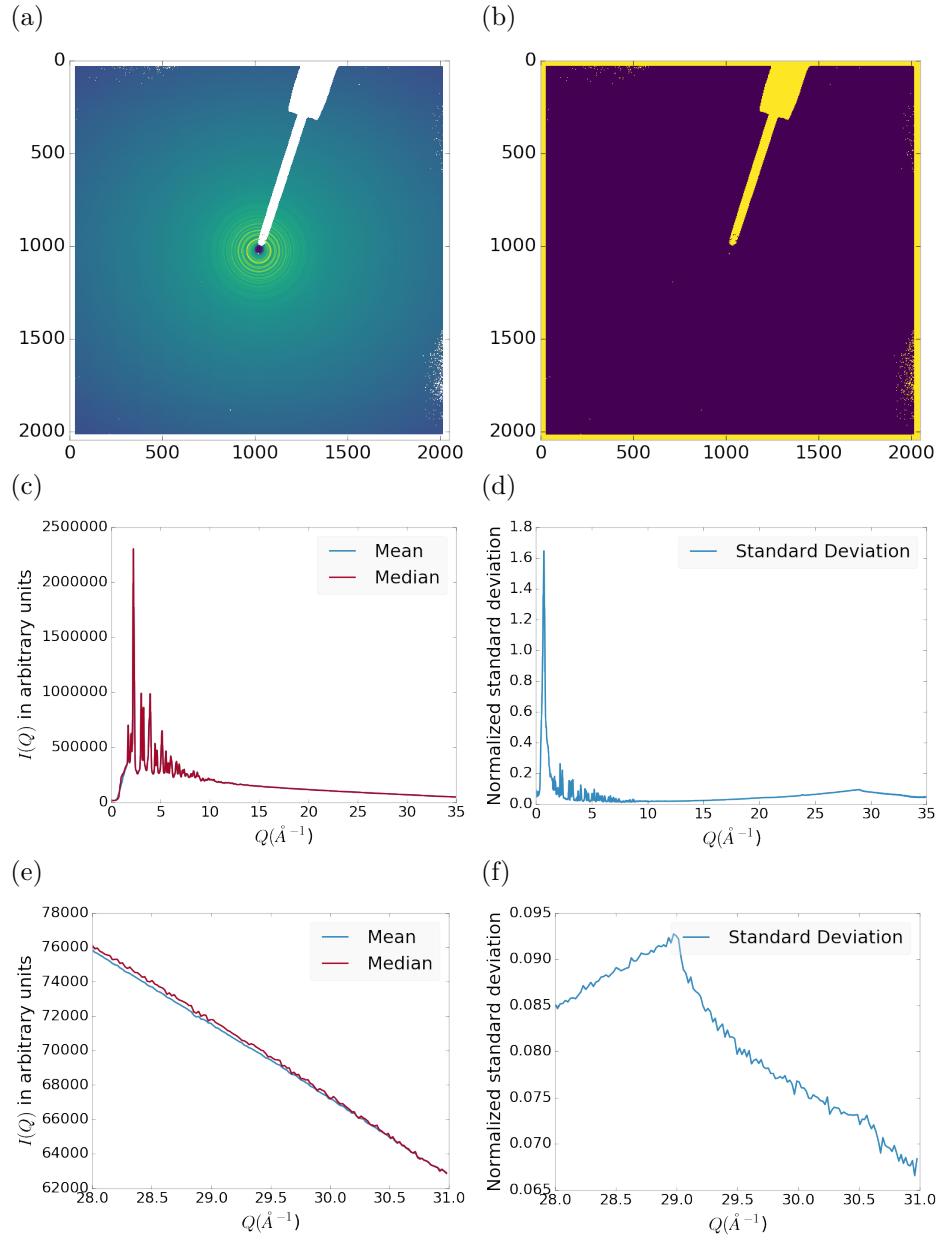


Figure 5.18: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced combining an edge mask and the automatically generated mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), 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

841 5.5 CONCLUSIONS

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

859                   CHAPTER 6

860                   PHASE CHANGES AND ANNEALING DYNAMICS OF  
861                    $\text{Pr}_2\text{NiO}_4$  AND ITS DERIVATIVES

862                   6.1 INTRODUCTION

863                    $\text{Pr}_2\text{NiO}_4$  (PNO) electrodes provide higher power density than  $\text{La}_{0.8}\text{Sr}_{0.2}\text{MnO}_3$  (LSM),  
864                   and is more stable than  $(\text{La}_{0.60}\text{Sr}_{0.40})_{0.95}(\text{Co}_{0.20}\text{Fe}_{.80})\text{O}_{3-\text{x}}$  (LSCF), which is known to  
865                   rapidly degrade in performance. [54] PNO's high performance between 600-900 °C is  
866                   associated with its high activity towards the oxygen reduction reaction (ORR), which  
867                   stems from PNO's high oxygen diffusion and surface exchange coefficients, substantial  
868                   oxygen over-stoichiometry, and large oxygen ion conduction paths through the unit  
869                   cell. [52] Despite these advantages, PNO's tendency to partially decompose into  
870                    $\text{PrOx}$  and other phases is particularly challenging. [13] Full cell operation after 500  
871                   hours at 750 °C and 0.8 V shows major decomposition of the parent PNO phase,  
872                   while the performance degrades by only 4%. Such significant changes in phase and  
873                   relatively small changes in performance further assure the necessity for understanding  
874                   the phase evolution in nickelate cathodes during operation. To address these disparity  
875                   in performance and phase degradation PDF and XRD analysis may be able to examine  
876                   these issues from both long and short range ordering perspectives.

877 6.2 EXPERIMENTS

878 **Pr<sub>2</sub>NiO<sub>4</sub> Synthesis**

879 Pr<sub>2</sub>NiO<sub>4</sub> was synthesized using the standard approach, as detailed in the work by  
880 Dogdibegovic et. al. [13] The nickelate powder was initially aprepared via the  
881 glycine-nitrate process. This was followd by thermal annealing at 1080 °C for 10  
882 hours in air.

883 **X-ray Measurements**

884 X-ray total scattering and x-ray powder diffraction experiments were performed at  
885 the APS's 11-ID-B beamline. An x-ray energy of 86.7 keV, .145 Åwas provided  
886 by the beamline monochromator. The detector was moved between a 20cm and a  
887 95 cm sample to detector distance to measure the x-ray total scattering and x-ray  
888 diffraction patterns. Various PNO samples were annealed on the beamline during  
889 x-ray measurement.

890 6.3 DATA PROCESSING

891 The data was calibrated at each of the detector positions using a CeO<sub>2</sub> standard  
892 via pyFAI. [29] The images were corrected for a .95 x-ray polarizetion. Masks were  
893 produced for both the foreground and background images. The forground masks were  
894 produced using both a 30 pixel edge mask and a  $2.5\sigma$  automatic mask as discussed  
895 in chapter 5. The background masks were produced by using the forground mask as  
896 a starting mask with a  $2.5\sigma$  automatic mask.

897 The foreground and background images were then integrated using the  $Q$  resolu-  
898 tion binning discussed in chapter 5. The resulting  $I(Q)$  data were corrected for their  
899 number of frames and  $I_{00}$ . Finally the corrected background  $I(Q)$  was subtracted  
900 from the foreground  $I(Q)$ .

901        Each PDF was generated with a  $Q_{min}$  of 1.5,  $Q_{max}$  of 29.,  $R_{poly}$  of .9,  $R_{max}$  of 40.  
902        descriptions of these parameters can be found in the work by Juhas et. al. [27]

903        **6.4 DATA ANALYSIS**

904        **Intra Sample Comparison**

905        **PDF**

906        As figures 6.1 and 6.2 show the as synthesized PNO undergoes very little change in  
907        structure according to the PDF. The PDF does show some broadening at around 3.5  
908        and 8.5 Å, but the peak shifts themselves are fairly limited. This implies that the as  
909        synthesized PNO structure is stable at least for the 1 hour that the sample was held  
910        at 750 °C.

911        The annealed samples figures, 6.3 and 6.4, tell a rather different story. In this case  
912        the PDF shows significant peak shifts and broadening, especially at higher interatomic  
913        distances. Some peaks completely disappear, like the peak at 12 Å. Similar results were  
914        also observed for samples with longer annealing times, as shown in the appendix.

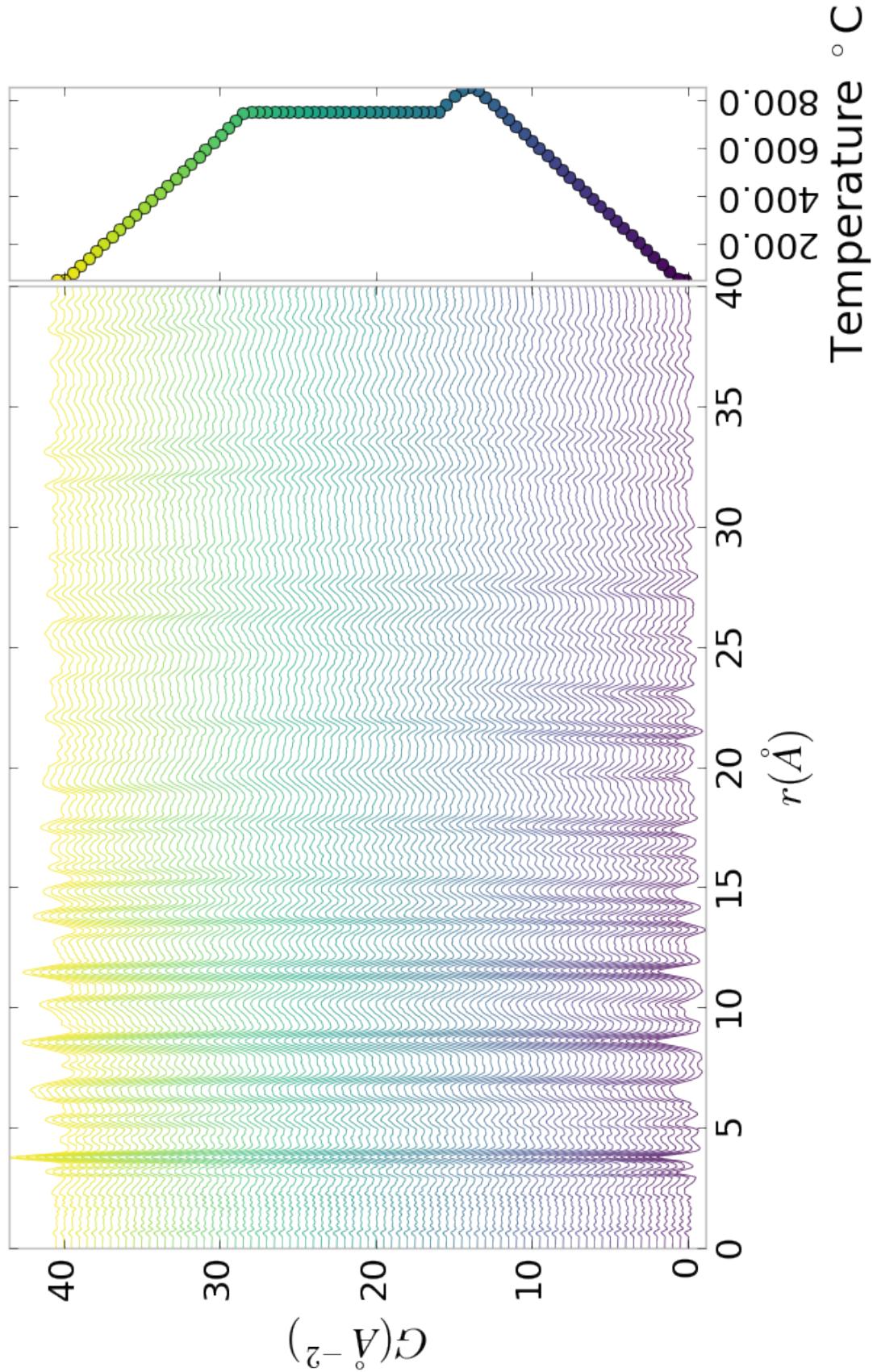


Figure 6.1: PDF as a function of temperature for as synthesized PNO showing the full PDF

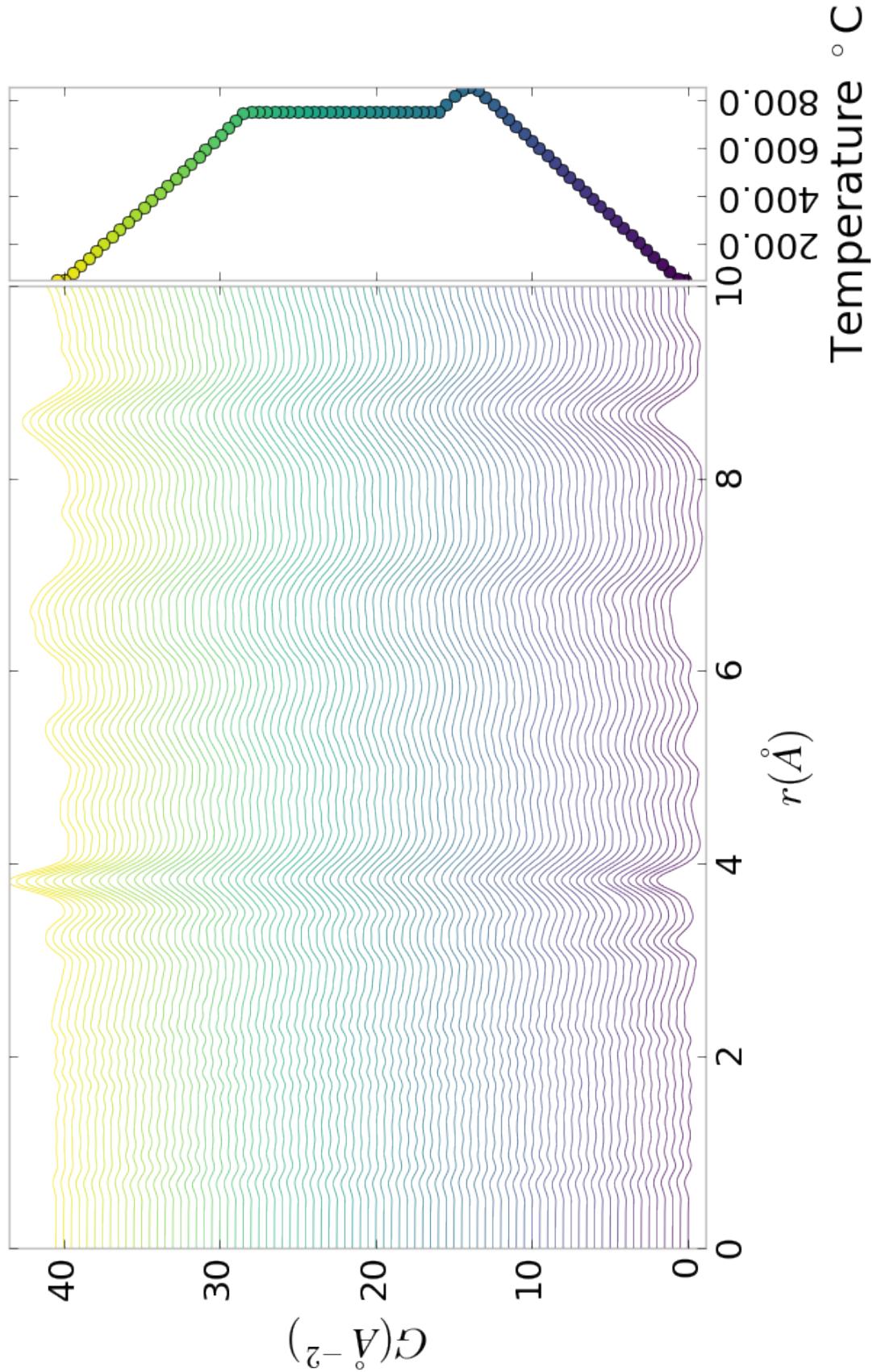


Figure 6.2: PDF as a function of temperature for as synthesized PNO showing a close up on the short range section

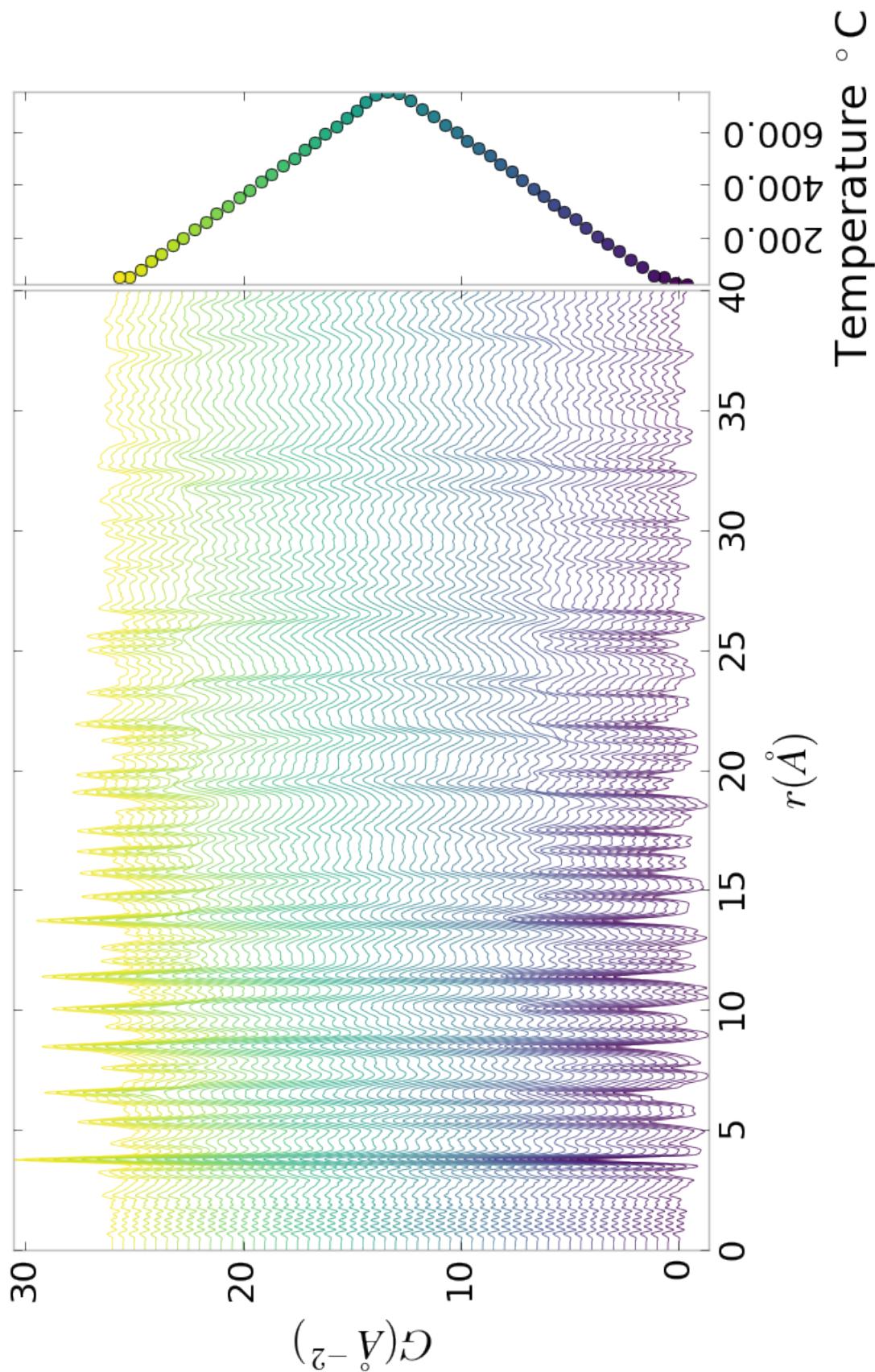


Figure 6.3: PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full PDF

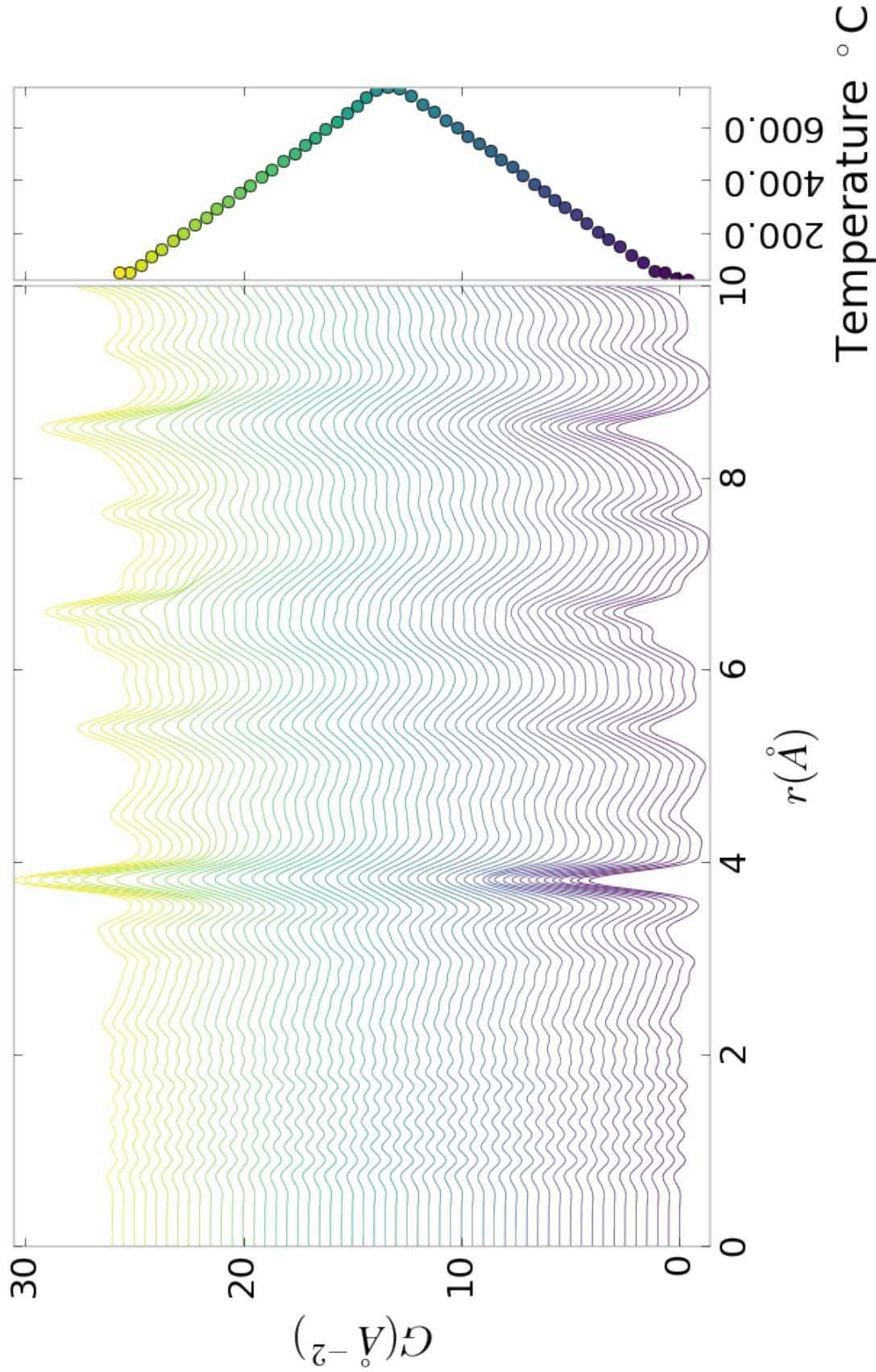


Figure 6.4: PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the short range section

$$_{915} \quad I(Q)$$

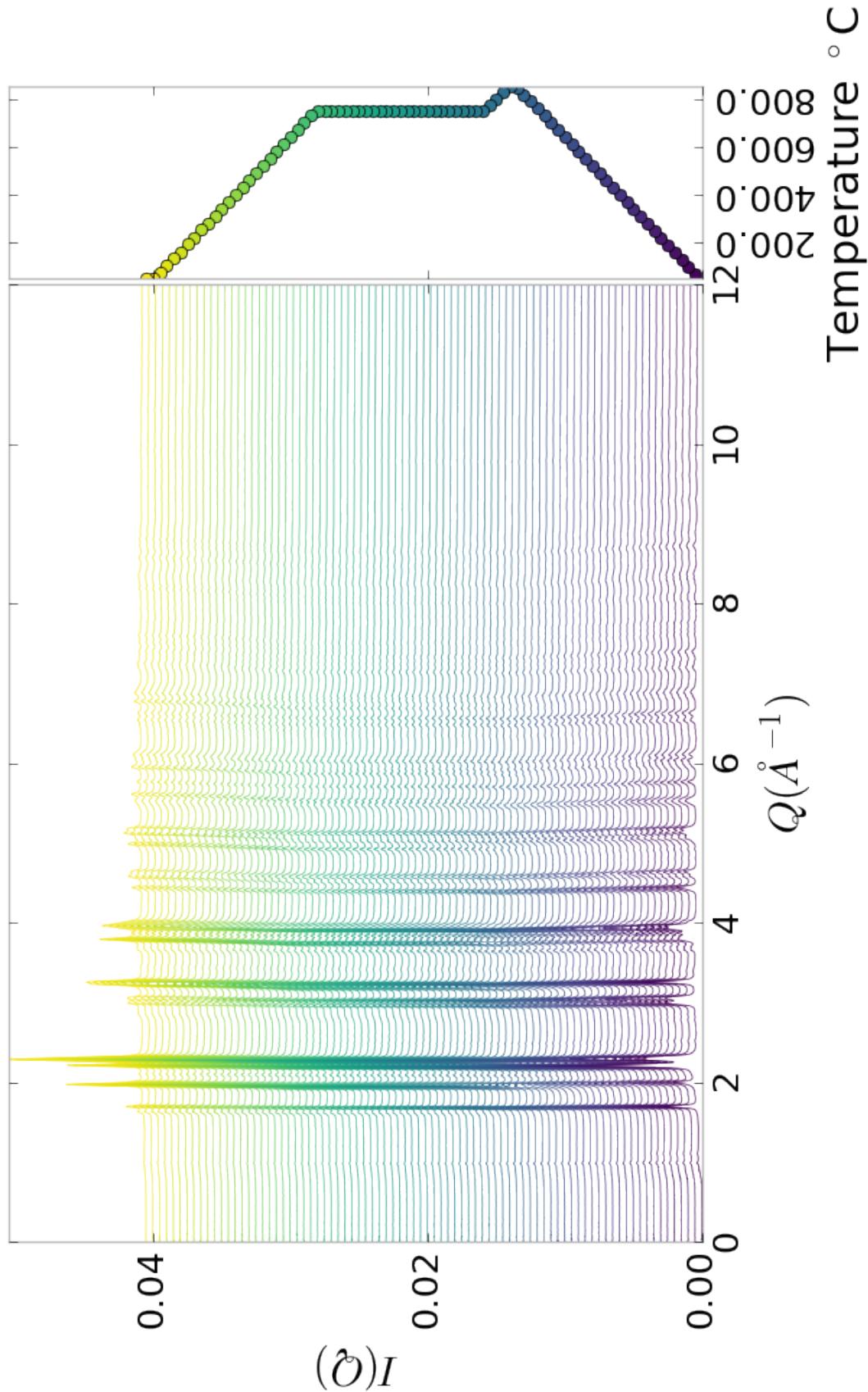


Figure 6.5:  $I(Q)$  as a function of temperature for as synthesized PNO showing the full XRD

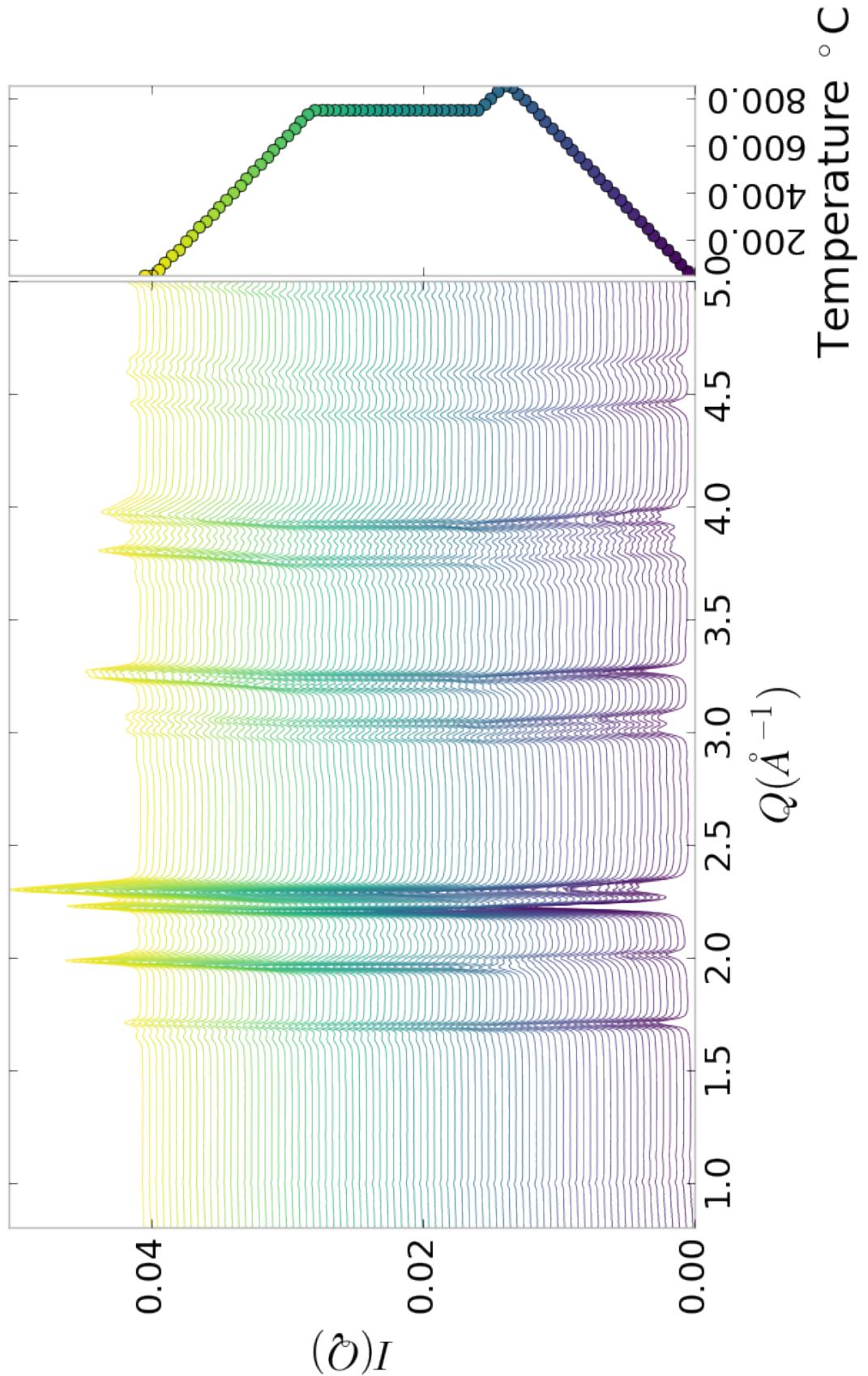


Figure 6.6:  $I(Q)$  as a function of temperature for as synthesized PNO showing a close up on the low  $Q$  section

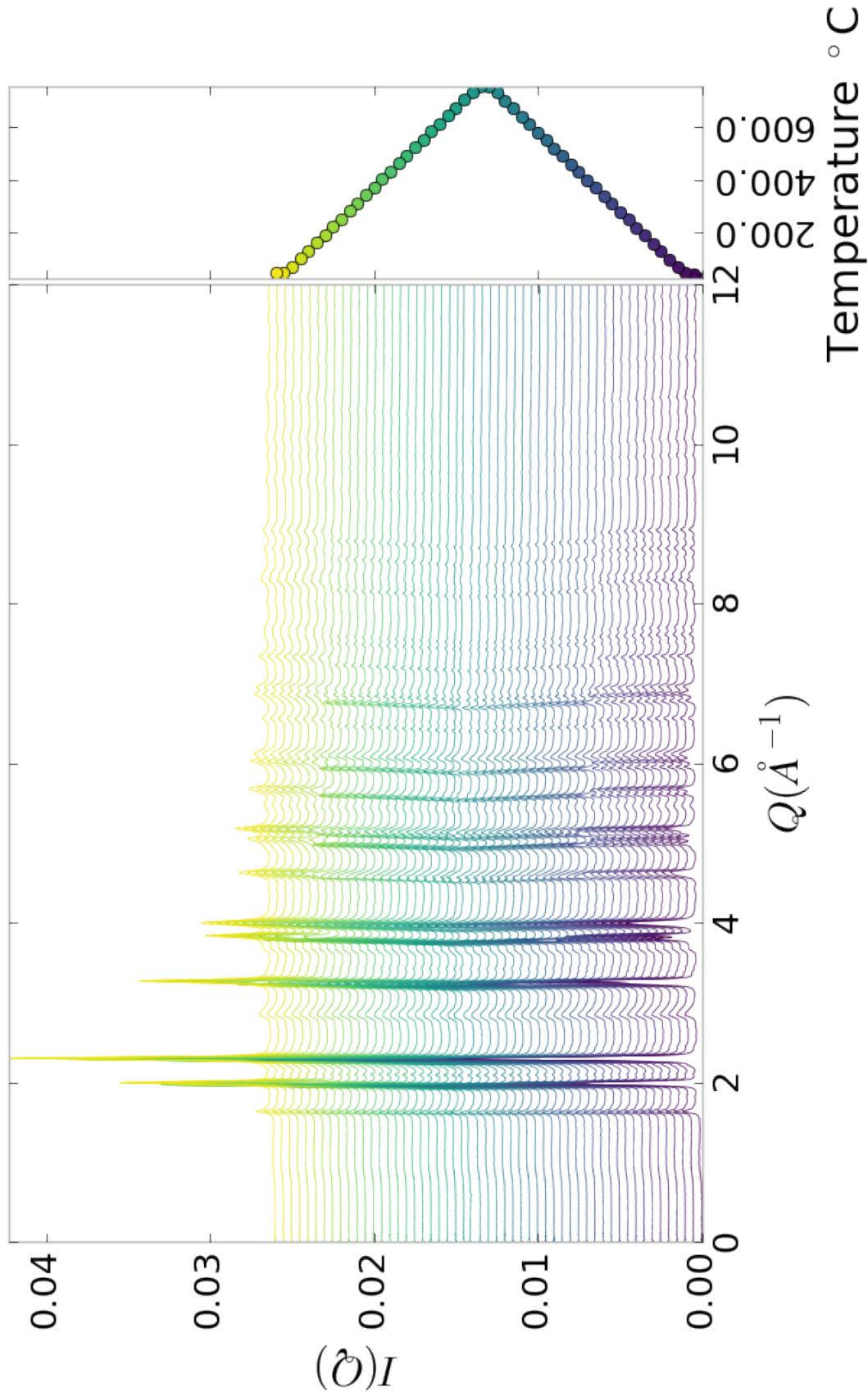


Figure 6.7:  $I(Q)$  as a function of temperature for PNO annealed at  $750^{\circ}\text{C}$  for 25 hours showing the full XRD

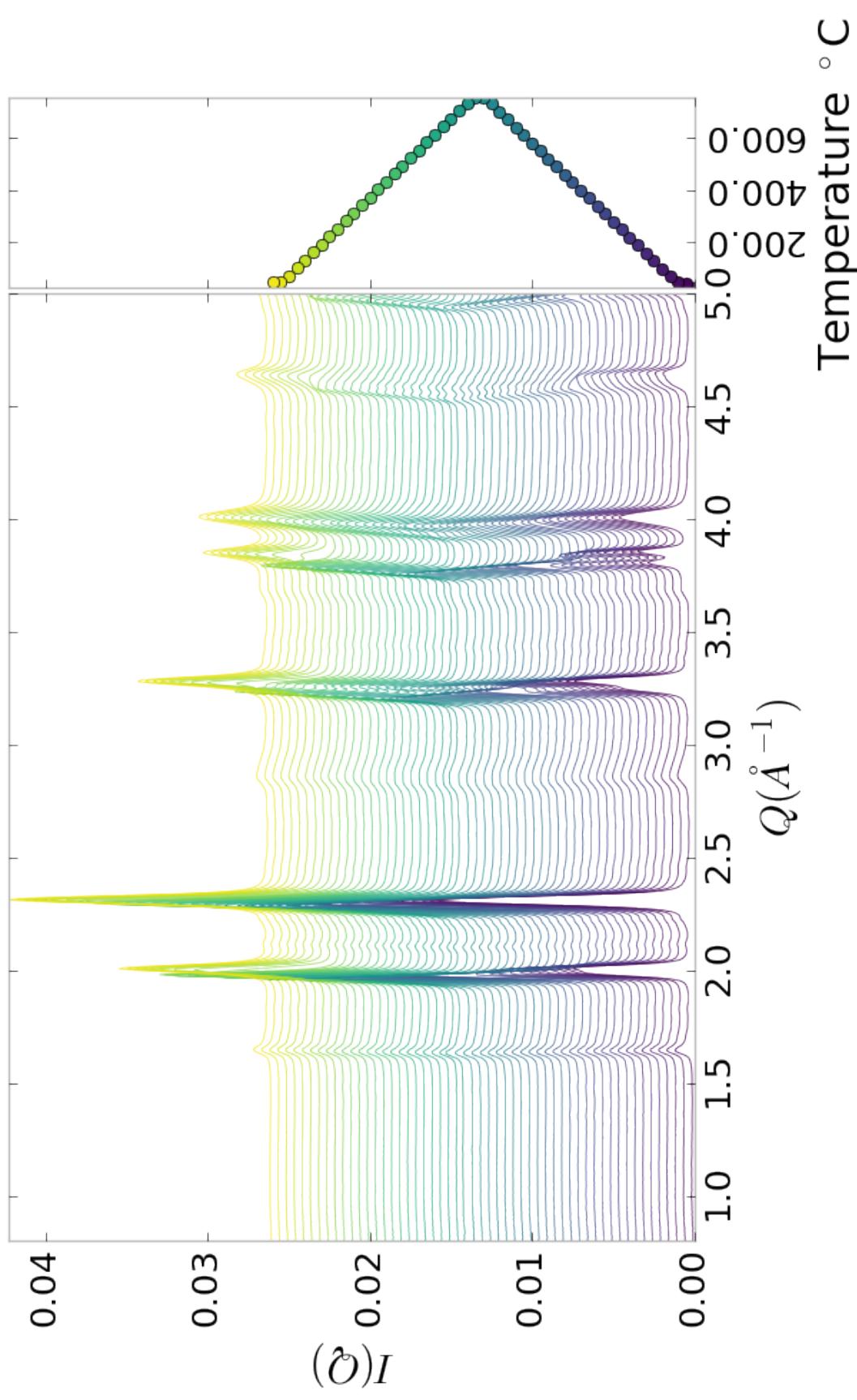


Figure 6.8:  $I(Q)$  as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the low  $Q$  section

916 **Inter Sample Comparison**

917 **6.5 SIMULATION**

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

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

930 **6.6 CONCLUSIONS**

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

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

950

## CHAPTER 7

951

### CONCLUSION

952 The work here presents one of the most complete end to end approaches to process-  
953 ing, analyzing, and simulating atomic pair distribution function data. The goals of  
954 this work were to build a modular, quick, and robust method for handeling both  
955 experimental PDF data and solving atomic structures from said data.

956 The statistical mechanical PES solvers were designed to robustly find atomic  
957 solutions which are global minima of the PES. This was accomplished by using some  
958 of the most advanced Monte Carlo algorithms and samplers. The analytical equations  
959 for the PES and its gradients were derived to provide the quickest searches.

960 The PDF gradients were derived and implemented as GPU kernels to furthur  
961 speed up the PES search. The inclusion of the GPUs, combined with the atom pair  
962 mapping, were found to provide a 10x to 100x speedup over a multiprocessed CPU  
963 methodology.

964 The extensive benchmarking of the NUTS-HMC system presented in chapters 2  
965 and 3 showcased the system's robustness, speed, and effectiveness. Interestingly it  
966 seems the the simulations also helped to elucidate the relationship between  $Rw$  and  
967 the resulting fit of the secondary metrics, including radial bond distribution. This is  
968 particularly important as it begins to establish  $Rw$  goals and a relationship between  
969  $Rw$  and the confidence that features from the underlying structure that the PDF  
970 represents are reproduced by the structural model which is produced my Monte Carlo  
971 modeling. It seems that the theshold for acceptable  $Rw$  in Monte Carlo modeling  
972 needs to be quite lower than the current literature standards to properly reproduce

973 the structure.

974 A novel data processing workflow was also developed which focused on using  $Q$   
975 resolution binning to create masks automatically and azimuthally integrate. The  $Q$   
976 resolution binning provided a significant improvement in the automated masking  
977 robustness, leading to much fewer false positives, as shown by a series of masks  
978 generated on simulated and experimental data. The effect of these masks on the  
979 median and mean azimuthal integration was also discussed, establishing masks as  
980 very important to the removal of the high  $Q$  “kink” seen in 2D area detector data.  
981 Furthermore, a comparison was drawn between the median and mean integration,  
982 showing the median to be more reliable than the mean when working with data that  
983 could have residual detector defects. Overall the masking scheme was shown to reduce  
984 the standard deviation of the data significantly.

985 Finally, preliminary results of x-ray total scattering measurements on  $\text{Pr}_2\text{NiO}_4$   
986 were presented. Interestingly, these results show a strong discrepancy between the  
987 PDF and  $I(Q)$  data. Where the PDF shows a very static as synthesized structure,  
988 despite annealing, the associated  $I(Q)$  data shows peak movement and formation. For  
989 the pre-annealed samples both the  $I(Q)$  and PDF data show peak changes. Inter-  
990 estingly, the PDFs of the as synthesized and pre-annealed samples show very similar  
991 local structure at operating temperatures.

992 Despite all the work presented here, there is of course much more to be done.  
993 Implementing new ensembles, like Parallel Tempering, and faster Grand Canonical  
994 Monte Carlo, may help to find solutions faster and with less user based parameter tun-  
995 ing. Building the mathematics and software to quickly compute the data from other  
996 atomistic experiments, including EXAFS, STEM, and neutron scattering, could help  
997 to produce structures which more fully describe all the available experimental data.  
998 Implementing the existing codebase in a more general High Performance Computing  
999 context would allow for the solution of much larger particles, and extended solids.

1000 Furthur benchmarking will help to probe the robustness of the algorithm with other  
1001 systems, including systems with periodic boundary conditions. Faster scattering data  
1002 processing will enable a quicker total turn around time from taking experimental im-  
1003 ages to producing atomic structures. However, even without these enhancements it is  
1004 expected that this work will become a standard method for solving atomic structures  
1005 from x-ray total scattering experiments.

1006

## BIBLIOGRAPHY

- 1007 [1] Alexandre A. Arnold, Victor Terskikh, Qian Ying Li, Rafik Naccache, Isabelle  
1008 Marcotte, and John A. Capobianco, *Structure of NaYF<sub>4</sub> upconverting nanoparticles: A multinuclear solid-state NMR and DFT computational study*, Journal  
1009 of Physical Chemistry C **117** (2013), no. 48, 25733–25741.
- 1011 [2] Simon J L Billinge and Takeshi Egami, *Underneath the Bragg Peaks: Structural  
1012 Analysis of complex Materials*, vol. Volume 16, Pergamon, 2012.
- 1013 [3] Simon J L Billinge and Igor Levin, *The problem with determining atomic struc-  
1014 ture at the nanoscale.*, Science (New York, N.Y.) **316** (2007), no. 5824, 561–565.
- 1015 [4] P. E. Blöchl, *Projector augmented-wave method*, Physical Review B **50** (1994),  
1016 no. 24, 17953–17979.
- 1017 [5] Gary K. Chen and Yunfei Guo, *Discovering epistasis in large scale genetic as-  
1018 sociation studies by exploiting graphics cards*, Frontiers in Genetics **4** (2013),  
1019 no. DEC, 1–12.
- 1020 [6] Joshua J Choi, Xiaohao Yang, Zachariah M Norman, Simon J L Billinge, and  
1021 Jonathan S Owen, *Structure of methylammonium lead iodide within mesoporous  
1022 titanium dioxide: Active material in high-performance perovskite solar cells*,  
1023 Nano Letters **14** (2014), no. 1, 127–133.
- 1024 [7] Peter J. Chupas, Karena W. Chapman, Charles Kurtz, Jonathan C. Hanson,  
1025 Peter L. Lee, and Clare P. Grey, *A versatile sample-environment cell for non-  
1026 ambient X-ray scattering experiments*, Journal of Applied Crystallography **41**  
1027 (2008), no. 4, 822–824.
- 1028 [8] PJ Chupas, X Qiu, JC Hanson, PL Lee, CP Grey, and SJL Billinge, *Rapid-  
1029 acquisition pair distribution function (RA-PDF) analysis*, Journal of Applied  
1030 Crystallography **36** (2003), 1342–1347.
- 1031 [9] Matthew J. Cliffe, Martin T. Dove, D. a. Drabold, and Andrew L. Goodwin,  
1032 *Structure determination of disordered materials from diffraction data*, Physical  
1033 Review Letters **104** (2010), no. 12, 1–4.

- 1034 [10] Matthew J Cliffe and Andrew L Goodwin, *Nanostructure determination from the*  
1035 *pair distribution function: a parametric study of the INVERT approach.*, Journal  
1036 *of physics. Condensed matter : an Institute of Physics journal* **25** (2013), no. 45,  
1037 454218.
- 1038 [11] Chunhua Cui, Lin Gan, Marc Heggen, Stefan Rudi, and Peter Strasser, *Composi-*  
1039 *tional segregation in shaped Pt alloy nanoparticles and their structural behaviour*  
1040 *during electrocatalysis.*, *Nature materials* **12** (2013), no. 12, 765–771.
- 1041 [12] Juarez L F Da Silva, Hyoung Gyu Kim, Maurício J. Piotrowski, Maurício J. Pri-  
1042 *eto, and Germano Tremiliosi-Filho, Reconstruction of core and surface nanopar-*  
1043 *ticles: The example of Pt 55 and Au55*, *Physical Review B - Condensed Matter*  
1044 *and Materials Physics* **82** (2010), no. 20, 1–6.
- 1045 [13] Emir Dogdibegovic, Christopher J. Wright, and Xiao-Dong Zhou, *Stability and*  
1046 *Activity of (Pr <sub>1-x</sub> Nd <sub>x</sub>)<sub>2</sub>NiO<sub>4</sub> as Cathodes for Solid Oxide Fuel Cells: I. Quantification of*  
1047 *Phase Evolution in Pr<sub>2</sub>NiO<sub>4</sub>*, *Journal of the American Ceramic Society* **5** (2016), 1–5.
- 1050 [14] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth, *Hybrid*  
1051 *Monte Carlo*, *Physics Letters B* **195** (1987), no. 2, 216–22.
- 1052 [15] Timur Dykhne, Ryan Taylor, Alastair Florence, and Simon J L Billinge, *Data*  
1053 *requirements for the reliable use of atomic pair distribution functions in amor-*  
1054 *phous pharmaceutical fingerprinting.*, *Pharmaceutical research* **28** (2011), no. 5,  
1055 1041–8.
- 1056 [16] C L Farrow, P Juhas, J W Liu, D Bryndin, E S Božin, J Bloch, Th Proffen, and  
1057 S J L Billinge, *PDFfit2 and PDFgui: computer programs for studying nanostruc-*  
1058 *ture in crystals.*, *Journal of Physics. Condensed Matter : an Institute of Physics*  
1059 *journal* **19** (2007), no. 33, 335219.
- 1060 [17] Christopher L Farrow and Simon J L Billinge, *Relationship between the atomic*  
1061 *pair distribution function and small-angle scattering: implications for modeling*  
1062 *of nanoparticles.*, *Acta Crystallographica Section A Foundations of Crystallog-*  
1063 *raphy* **65** (2009), no. Pt 3, 232–9 (en).
- 1064 [18] Riccardo Ferrando, Julius Jellinek, and Roy L Johnston, *Nanoalloys: From The-*  
1065 *ory to Applications of Alloy Clusters and Nanoparticles*, *Chemical Reviews* **108**  
1066 (2008), no. 3, 846–904.

- 1067 [19] Anatoly Frenkel, *Solving the 3D structure of metal nanoparticles*, Zeitschrift für  
1068 Kristallographie **222** (2007), no. 11, 605–611.
- 1069 [20] Carmelo Giacovazzo, Hugo Luis Monaco, Gilberto Artioli, Davide Viterbo,  
1070 Marco Milanesio, Gastone Gilli, Paola Gilli, Giuseppe Zanotti, Giovanni Fer-  
1071 raris, and Michele Catti, *Fundamentals of Crystallography*, Oxford University  
1072 Press, 1992.
- 1073 [21] Benjamin Gilbert, Jasmine J. Erbs, R. Lee Penn, Valeri Petkov, Dino Spagnoli,  
1074 and Glenn A. Waychunas, *A disordered nanoparticle model for 6-line ferrihydrite*,  
1075 American Mineralogist **98** (2013), no. 8-9, 1465–1476.
- 1076 [22] Benjamin Gilbert, Feng Huang, Hengzhong Zhang, Glenn A. Waychunas, and  
1077 Jillian F Banfield, *Nanoparticles: strained and stiff.*, Science (New York, N.Y.)  
1078 **305** (2004), no. 5684, 651–654.
- 1079 [23] Md Matthew D. Hoffman and Andrew Gelman, *The No-U-Turn Sampler: Adap-  
1080 tively Setting Path Lengths in Hamiltonian Monte Carlo*, The Journal of Machine  
1081 Learning Research **15** (2014), no. 2008, 1593–1623.
- 1082 [24] W J Huang, R Sun, J Tao, L D Menard, R G Nuzzo, and J M Zuo, *Coordination-  
1083 dependent surface atomic contraction in nanocrystals revealed by coherent diffrac-  
1084 tion.*, Nature Materials **7** (2008), no. 4, 308–313.
- 1085 [25] Pablo D Jadzinsky, Guillermo Calero, Christopher J Ackerson, David A Bushnell,  
1086 and Roger D Kornberg, *Structure of a thiol monolayer-protected gold nanoparti-  
1087 cle at 1.1 Å resolution.*, Science (New York, N.Y.) **318** (2007), no. 5849, 430–433.
- 1088 [26] I. K. Jeong, R. H. Heffner, M. J. Graf, and S. J. L. Billinge, *Lattice dynamics  
1089 and correlated atomic motion from the atomic pair distribution function*, (2002),  
1090 9.
- 1091 [27] P. Juhás, T. Davis, C. Farrow, and S. Billinge, *PDFgetX3 : a  
1092 rapid and highly automatable program for processing powder diffraction data into  
1093 total scattering pair distribution functions*, Journal of Applied Crystallography  
1094 **46** (2013), no. 2, 560–566 (en).
- 1095 [28] A. Kassiba, M. Makowska-Janusik, J. Bouclé, J. Bardeau, A. Bulou, and  
1096 N. Herlin-Boime, *Photoluminescence features on the Raman spectra of quasisto-  
1097 chiometric SiC nanoparticles: Experimental and numerical simulations*, Physical  
1098 Review B **66** (2002), no. 15, 1–7.

- 1099 [29] Jérôme Kieffer and Dimitrios Karkoulis, *PyFAI, a versatile library for azimuthal*  
1100 *regrouping*, Journal of Physics: Conference Series **425** (2013), 202012.
- 1101 [30] G. Kresse and J. Hafner, *Ab initio molecular dynamics for liquid metals*, Physical  
1102 Review B **47** (1993), no. 1, 558–561.
- 1103 [31] ———, *Ab initio molecular-dynamics simulation of the liquid-*  
1104 *metal–amorphous-semiconductor transition in germanium*, Physical Review  
1105 B **49** (1994), no. 20, 14251–14269.
- 1106 [32] Yan Li, Giulia Galli, and François Gygi, *Electronic structure of thiolate-covered*  
1107 *gold nanoparticles: Au<sub>102</sub>(MBA)<sub>44</sub>*, ACS Nano **2** (2008), no. 9, 1896–1902.
- 1108 [33] L D Marks, *Experimental studies of small particle structures*, Reports on  
1109 Progress in Physics **57** (1994), no. 6, 603–649 (en).
- 1110 [34] A. S. Masadeh, E. S. Božin, C. L. Farrow, G. Paglia, P. Juhas, S. J. L. Billinge,  
1111 A. Karkamkar, and M. G. Kanatzidis, *Quantitative size-dependent structure and*  
1112 *strain determination of CdSe nanoparticles using atomic pair distribution func-*  
1113 *tion analysis*, Physical Review B - Condensed Matter and Materials Physics **76**  
1114 (2007), no. 11, 115413.
- 1115 [35] R L McGreevy and L Pusztai, *Reverse Monte Carlo Simulation: A New Tech-*  
1116 *nique for the Determination of Disordered Structures*, Molecular Simulation **1**  
1117 (1988), no. 6, 359–367.
- 1118 [36] Donald A. McQuarrie, *Statistical Mechanics*, University Science Books, Sausal-  
1119 ito, CA, 2000.
- 1120 [37] Radford M Neal, *Probabilistic Inference Using Markov Chain Monte Carlo Meth-*  
1121 *ods*, Intelligence **45** (1993), no. September, 144.
- 1122 [38] Radford M. Neal, *MCMC Using Hamiltonian Dynamics*, Handbook of Markov  
1123 Chain Monte Carlo (Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li  
1124 Meng, eds.), Chapman and Hall/CRC, 2011, pp. 113–162.
- 1125 [39] Katharine Page, Taylor C. Hood, Thomas Proffen, and Reinhard B. Neder,  
1126 *Building and refining complete nanoparticle structures with total scattering data*,  
1127 Journal of Applied Crystallography **44** (2011), no. 2, 327–336 (en).

- 1128 [40] B R Pauw, *Corrigendum: Everything SAXS: small-angle scattering pattern col-*  
1129 *lection and correction* (2013 *J. Phys.: Condens. Matter* **25** 383201), *Journal of*  
1130 *Physics: Condensed Matter* **26** (2014), no. 23, 239501.
- 1131 [41] Vitalij Pecharsky and Peter Zavalij, *Fundamentals of Powder Diffraction and*  
1132 *Structural Characterization of Materials*, Springer Science, 2009.
- 1133 [42] John P. Perdew, Kieron Burke, and Matthias Ernzerhof, *Generalized Gradient*  
1134 *Approximation Made Simple*, *Physical Review Letters* **77** (1996), no. 18, 3865–  
1135 3868.
- 1136 [43] Andrew A. Peterson, *Global optimization of adsorbate-surface structures while*  
1137 *preserving molecular identity*, *Topics in Catalysis* **57** (2014), no. 1-4, 40–53.
- 1138 [44] Valeri Petkov, Binay Prasai, Yang Ren, Shiyao Shan, Jin Luo, Pharrah Joseph,  
1139 and Chuan-Jian Zhong, *Solving the nanostructure problem: exemplified on metal-*  
1140 *alloy nanoparticles*, *Nanoscale* **6** (2014), no. 17, 1–11.
- 1141 [45] Valeri Petkov, Shiyao Shan, Peter Chupas, Jun Yin, Lefu Yang, Jin Luo, and  
1142 Chuan-Jian Zhong, *Noble-transition metal nanoparticle breathing in a reactive*  
1143 *gas atmosphere.*, *Nanoscale* **5** (2013), no. 16, 7379–87.
- 1144 [46] Th. Proffen and R. B. Neder, *DISCUS: a Program for Diffuse Scattering and*  
1145 *Defect-Structure Simulation*, *Journal of Applied Crystallography* **30** (1997), 171–  
1146 175.
- 1147 [47] Erin L. Redmond, Brian P. Setzler, Pavol Juhas, Simon J. L. Billinge, and  
1148 Thomas F. Fuller, *In-Situ Monitoring of Particle Growth at PEMFC Cathode*  
1149 *under Accelerated Cycling Conditions*, *Electrochemical and Solid-State Letters*  
1150 **15** (2012), no. 5, B72 (en).
- 1151 [48] H. W. Sheng, M. J. Kramer, A. Cadien, T. Fujita, and M. W. Chen, *Highly*  
1152 *optimized embedded-atom-method potentials for fourteen FCC metals*, *Physical*  
1153 *Review B - Condensed Matter and Materials Physics* **83** (2011), no. 13, 134118.
- 1154 [49] Randall Q. Snurr, Alexis T. Bell, and Doros N. Theodorou, *Prediction of ad-*  
1155 *sorption of aromatic hydrocarbons in silicalite from grand canonical Monte Carlo*  
1156 *simulations with biased insertions*, *The Journal of Physical Chemistry* **97** (1993),  
1157 no. 51, 13742–13752 (EN).
- 1158 [50] Chang Q. Sun, *Size dependence of nanostructures: Impact of bond order defi-*  
1159 *ciency*, *Progress in Solid State Chemistry* **35** (2007), no. 1, 1–159.

- 1160 [51] X. Yang, P. Juhás, and S. J L Billinge, *On the estimation of statistical uncertainties*  
1161 *on powder diffraction and small-angle scattering data from two-dimensional*  
1162 *X-ray detectors*, Journal of Applied Crystallography **47** (2014), no. 4, 1273–1283.
- 1163 [52] Masatomo Yashima, Makiko Enoki, Takahiro Wakita, Roushown Ali, Yoshitaka  
1164 Matsushita, Fujio Izumi, and Tatsumi Ishihara, *Structural Disorder and Diffusional Pathway of Oxide Ions in a Doped Pr<sub>2</sub>NiO<sub>4</sub>-Based Mixed Conductor*,  
1165 Journal of the American Chemical Society **130** (2008), no. 9, 2762–2763.
- 1167 [53] Hengzhong Zhang, Bin Chen, Jillian F. Banfield, and Glenn A. Waychunas,  
1168 *Atomic structure of nanometer-sized amorphous TiO<sub>2</sub>*, Physical Review B **78**  
1169 (2008), no. 21, 214106.
- 1170 [54] X.-D. Zhou, J.W. Templeton, Z. Nie, H. Chen, J.W. Stevenson, and L.R. Pederson,  
1171 *Electrochemical performance and stability of the cathode for solid oxide fuel*  
1172 *cells: V. high performance and stable Pr<sub>2</sub>NiO<sub>4</sub> as the cathode for solid oxide fuel*  
1173 *cells*, Electrochimica Acta **71** (2012), 44–49.

1174

## APPENDIX A

1175 SUPPLEMENTAL INFORMATION: PHASE CHANGES AND

1176 ANNEALING DYNAMICS OF  $\text{Pr}_2\text{NiO}_4$  AND ITS

1177 DERIVATIVES

1178 **Intra Sample Comparison**

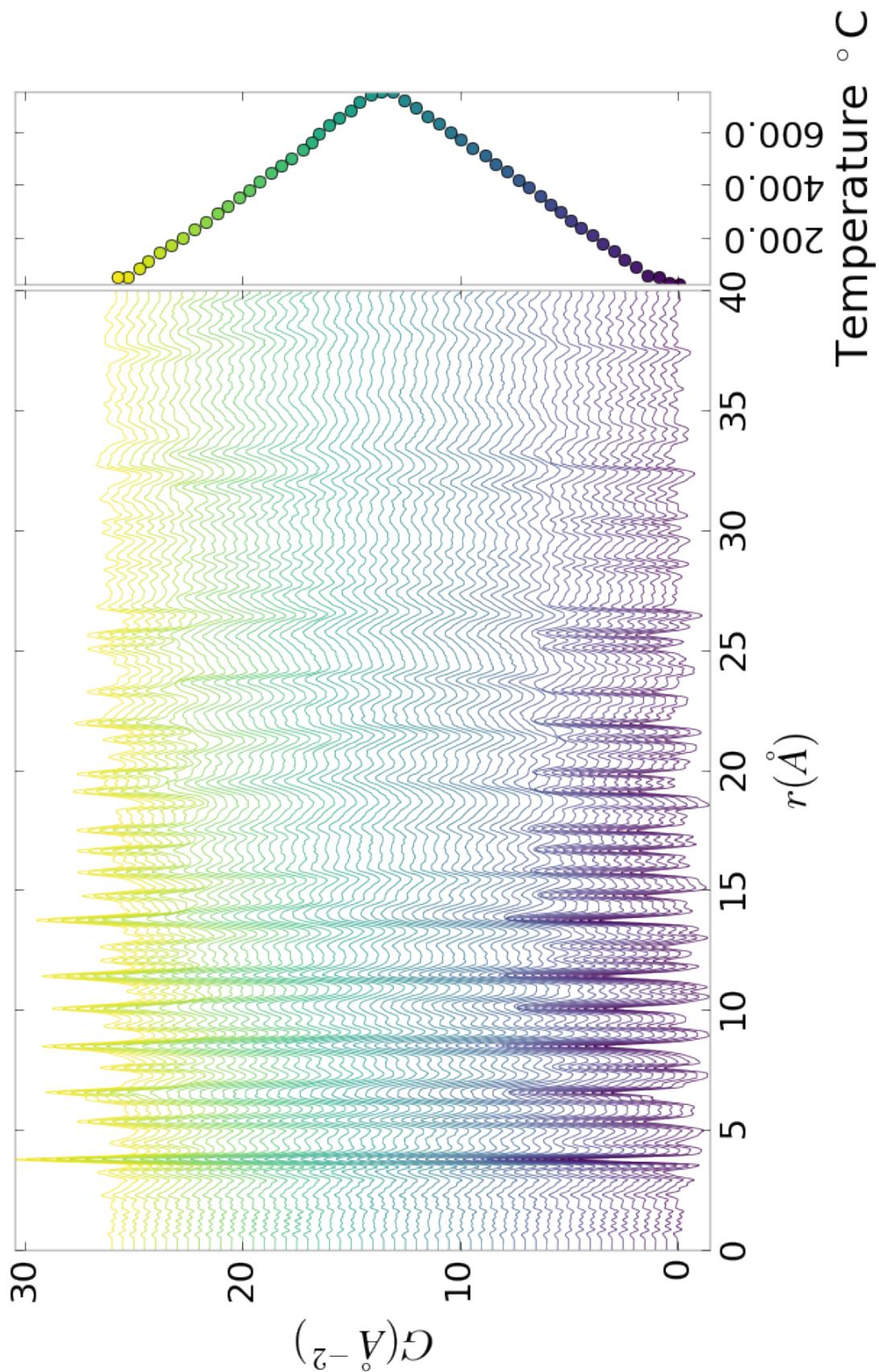


Figure A.1: PDF as a function of temperature for PNO annealed at 750 °C for 50 hours showing the full PDF

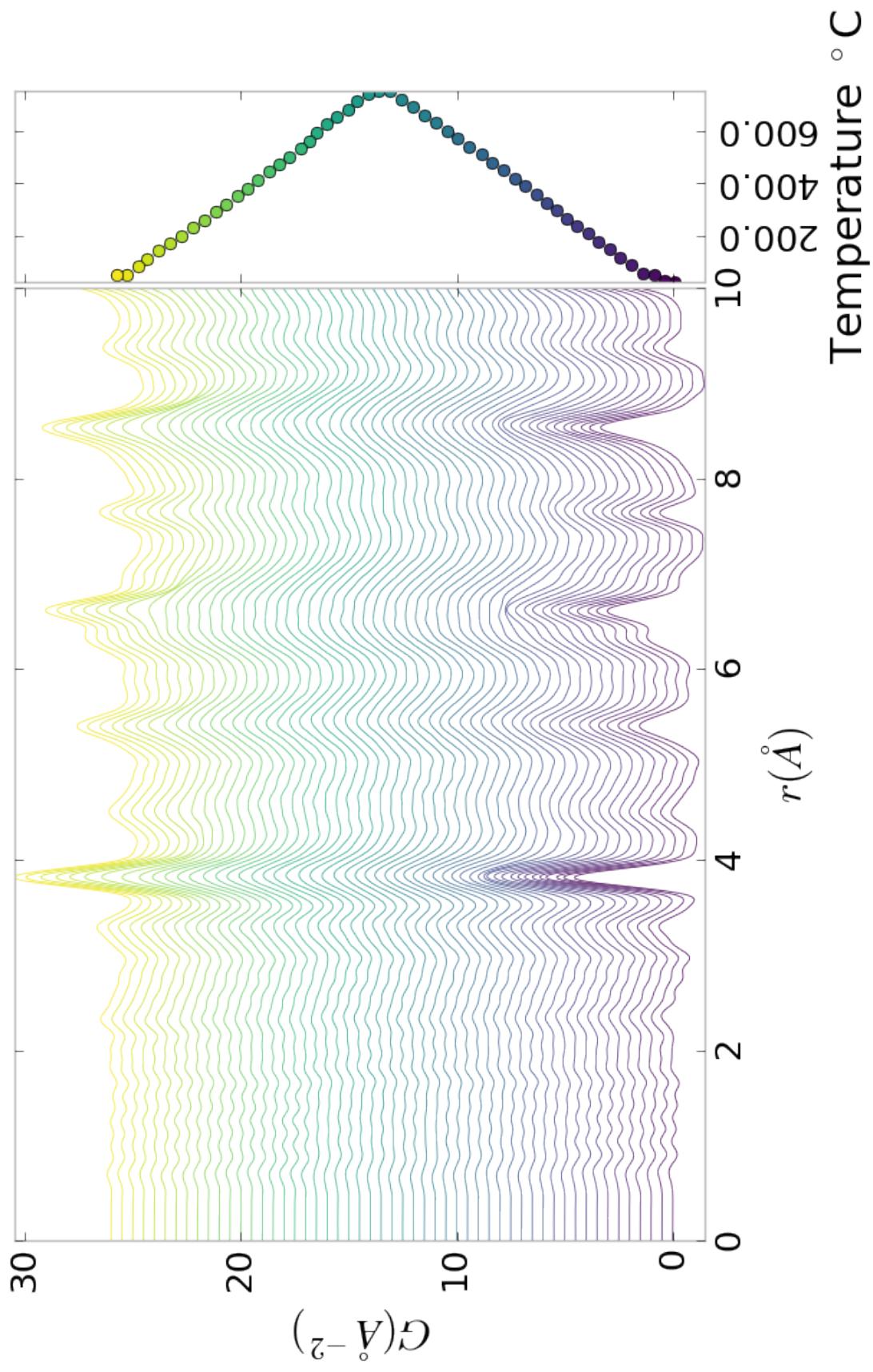


Figure A.2: PDF as a function of temperature for PNO annealed at 750 °C for 50 hours showing a close up on the short range section

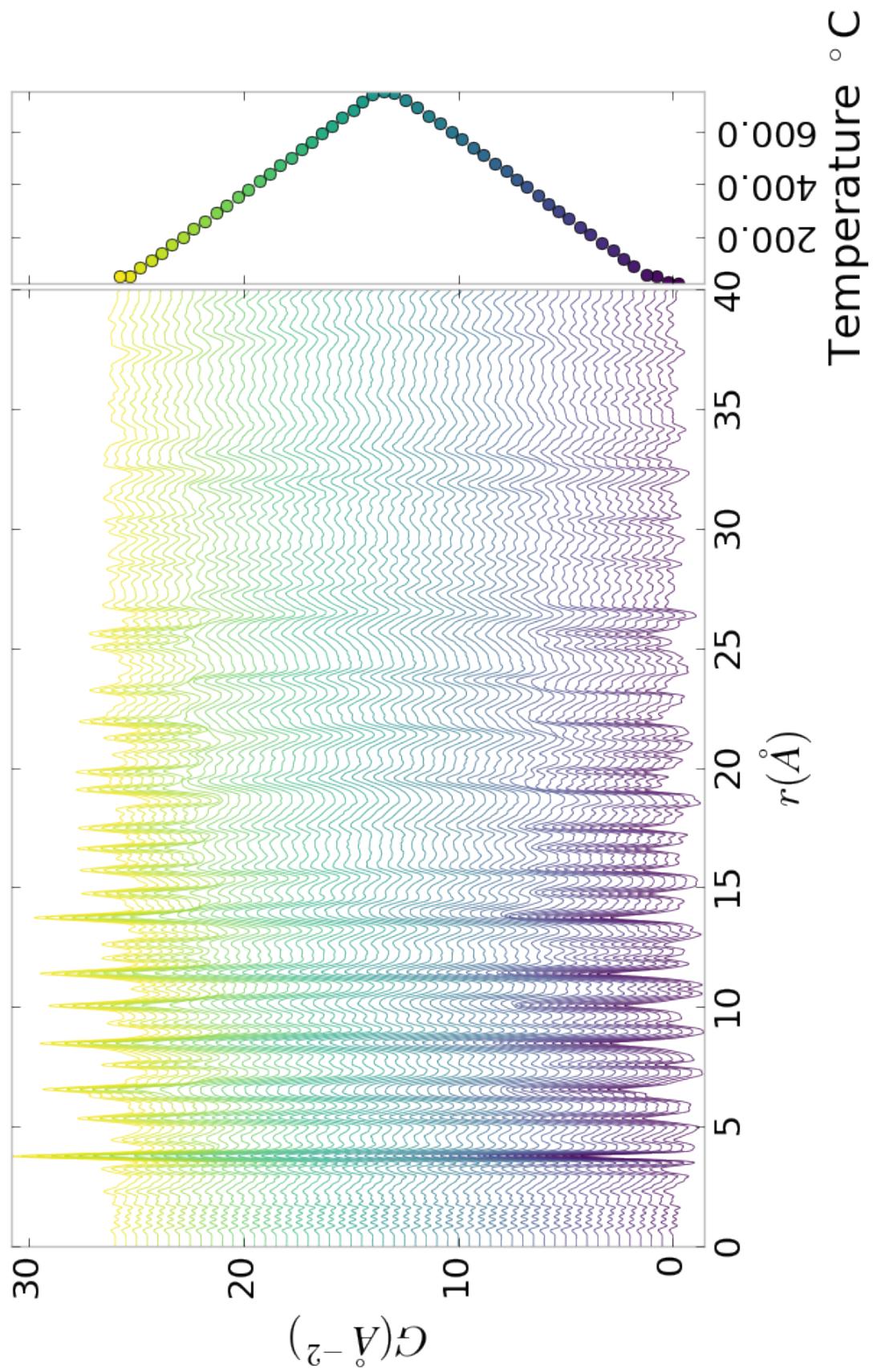


Figure A.3: PDF as a function of temperature for PNO annealed at 750 °C for 100 hours showing the full PDF

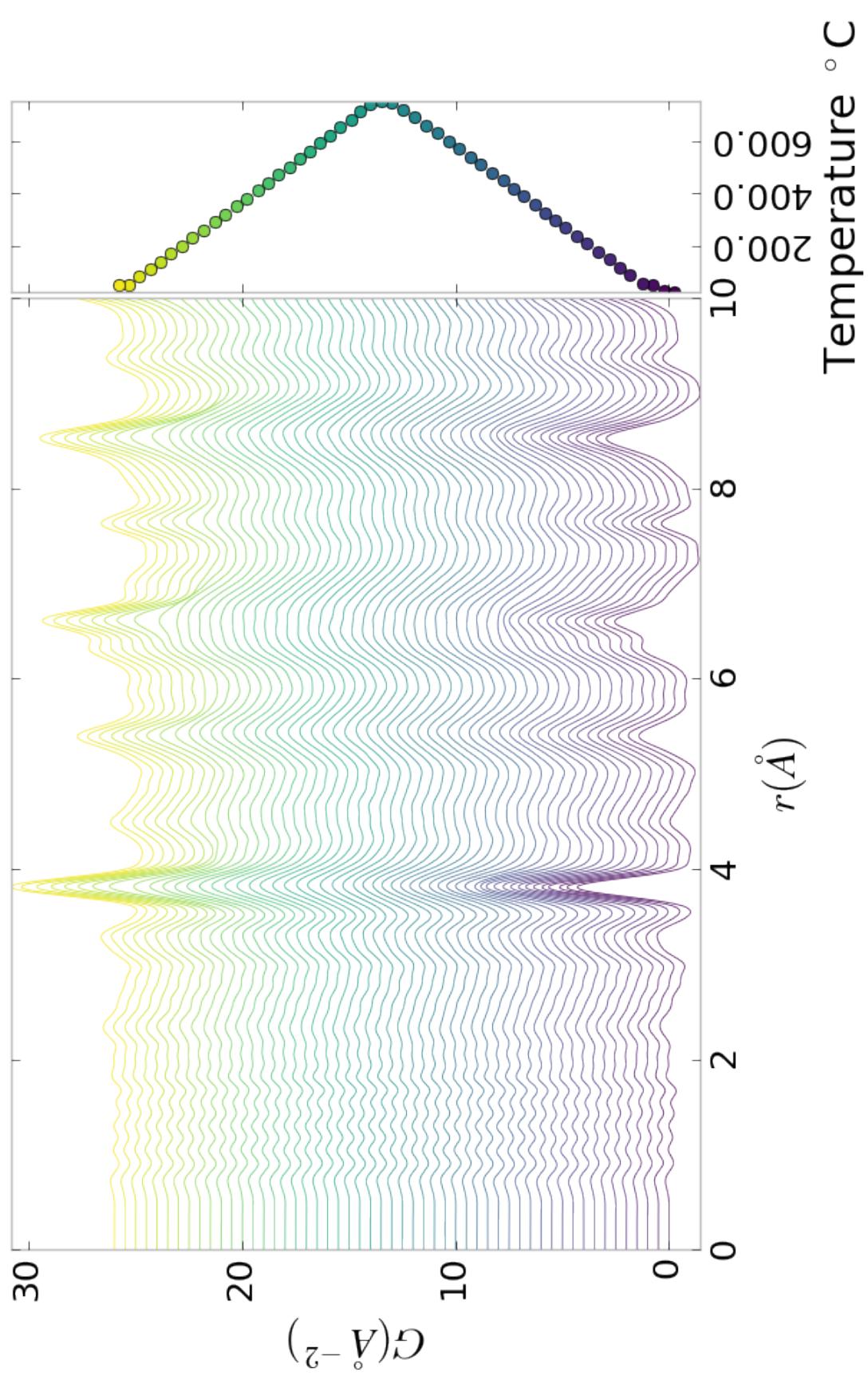


Figure A.4: PDF as a function of temperature for PNO annealed at 750 °C for 100 hours showing a close up on the short range section

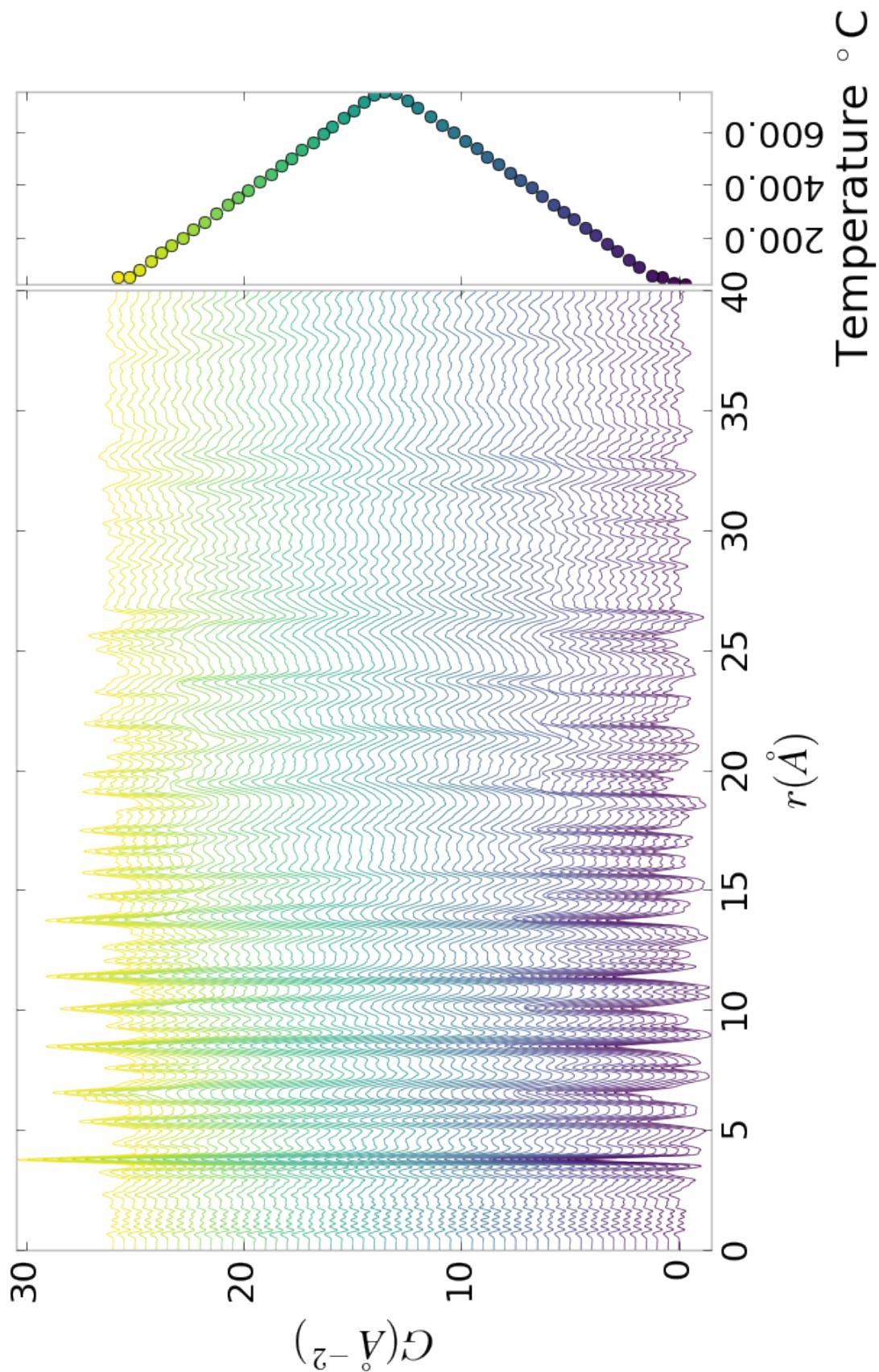


Figure A.5: PDF as a function of temperature for PNO annealed at 750  $^{\circ}\text{C}$  for 200 hours showing the full PDF

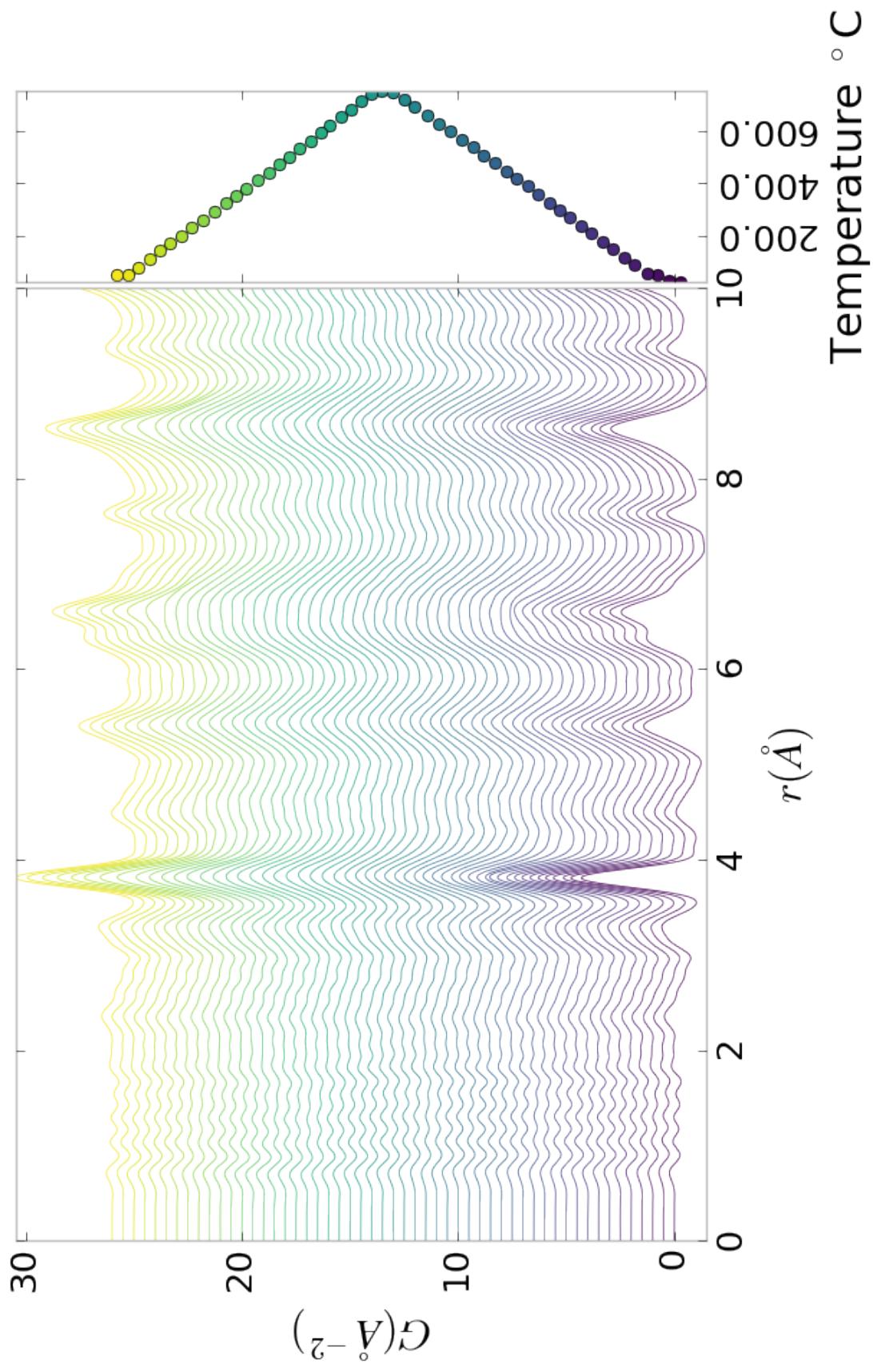


Figure A.6: PDF as a function of temperature for PNO annealed at 750  $^{\circ}\text{C}$  for 200 hours showing a close up on the short range section

1179

need to fix the problem with the  $I(Q)$  figures