

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

by

Christopher James Wright

Bachelor of Science
Brown University 2014

Submitted in Partial Fulfillment of the Requirements

for the Degree of Masters of Science in

Chemical Engineering

College of Engineering and Computing

University of South Carolina

2016

Accepted by:

Xiao-Dong Zhou, Major Professor

Mark Uline, Committee Member

Jochen Lauterbach, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

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

DEDICATION

For Diane & Donald Wright

My first scientific advisers

ACKNOWLEDGMENTS

ABSTRACT

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

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

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

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGMENTS	iv
ABSTRACT	v
LIST OF TABLES	x
LIST OF FIGURES	xi
TODO LIST	1
CHAPTER 1 STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES	4
1.1 Introduction	4
1.2 Potential Energy Surfaces	4
Experimentally Derived Potential Energy Surfaces	5
Potentials	5
Forces	6
1.3 Ensembles	7
Monte Carlo Modeling	8
Hamiltonian Monte Carlo	9
No-U-Turn Sampling	11
Grand Canonical Ensemble	11
Ensemble description	12
Grand Canonical Monte Carlo	12
GCMC biasing	13
1.4 Conclusions	16

CHAPTER 2 ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION	17
2.1 Introduction	17
2.2 Theory	18
Derivation	18
Analytically Gradients	19
Without ADPs	20
Periodic Boundary Conditions	21
2.3 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
2.4 Conclusions	28
CHAPTER 3 BENCHMARKING	29
3.1 Introduction	29
3.2 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
3.3 PDF with ADPs	43
ADP 50	43
CHAPTER 4 X-RAY TOTAL SCATTERING DATA ACQUISITION AND PRO- CESSING	44
4.1 Introduction	44

4.2	Detector Q resolution	44
4.3	Automated Mask Generation	48
	Introduction	48
	Algorithm Design	48
	Test Cases	49
	Results and Discussion	49
	Conclusions	56
4.4	Automated Image Azimuthal Integration	56
4.5	Conclusions	61
CHAPTER 5 PHASE CHANGES AND ANNEALING DYNAMICS OF Pr_2NiO_4 AND ITS DERIVATIVES		62
5.1	Introduction	62
5.2	Experiments	62
	Pr_2NiO_4 Synthesis	62
	X-ray Measurements	62
5.3	Data Processing	63
5.4	Data Analysis	63
	Intra Sample Comparison	63
	PDF	63
	$I(Q)$	69
	Inter Sample Comparison	74
5.5	Simulation	74
5.6	Conclusions	74
BIBLIOGRAPHY		77
APPENDIX A SUPPLEMENTAL INFORMATION: PHASE CHANGES AND ANNEALING DYNAMICS OF Pr_2NiO_4 AND ITS DERIVATIVES . . .		83
	Intra Sample Comparison	83

LIST OF TABLES

LIST OF FIGURES

Figure 1.1	Addition biasing with a Lennard Jones potential.	15
Figure 2.1	Comparison of the CPU and GPU chip architectures	22
Figure 2.2	Speed comparison of CPU and GPU implementations	27
Figure 3.1	Au ₅₅ PDF fitting of DFT-optimized <i>c</i> -Au ₅₅ .	33
Figure 3.2	Au ₅₅ PDF fitting of surface-disordered Au ₅₅ .	39
Figure 3.3	Similar to figure 3.2 for DFT-optimized amorphous Au ₅₅ .	40
Figure 3.4	Similar to figure 3.2 for Au ₁₀₂ as in DFT-optimized Au ₁₀₂ MBA ₄₄ cluster.	41
Figure 3.5	Similar to Fig. 3.4 with Marks decahedron as the starting structure.	42
Figure 3.6	Refinement of adps	43
Figure 4.1	Scattering onto a flat detector	45
Figure 4.2	Q resolution as a function of Q .	46
Figure 4.3	Number of pixels as a function of Q , binned at the Q resolution of the detector.	47
Figure 4.4	Generated dead/hot pixel masks for a detector with 100 bad pixels.	50
Figure 4.5	Generated dead/hot pixel masks for a detector with 300 bad pixels.	50
Figure 4.6	Generated dead/hot pixel masks for a detector with 500 bad pixels.	51
Figure 4.7	Generated dead/hot pixel masks for a detector with 1000 bad pixels.	51

Figure 4.8 Generated beamstop holder masks for a beamstop holder with 10% transmittance.	52
Figure 4.9 Generated beamstop holder masks for a beamstop holder with 30% transmittance.	52
Figure 4.10 Generated beamstop holder masks for a beamstop holder with 50% transmittance.	52
Figure 4.11 Generated beamstop holder masks for a beamstop holder with 90% transmittance.	53
Figure 4.12 Generated beamstop holder masks which is rotated away from vertical	53
Figure 4.13 Masked experimental data.	54
Figure 4.14 Masked experimental data with Pt single crystal signal.	55
Figure 4.15 Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask.	55
Figure 4.16 Masking, average, and standard deviation of an example x-ray total scattering measurement with with no mask.	58
Figure 4.17 Masking, average, and standard deviation of an example x-ray total scattering measurement with with only an edge mask.	59
Figure 4.18 Masking, average, and standard deviation of an example x-ray total scattering measurement with combining an edge mask and the automatically generated mask.	60
Figure 5.1 PDF as a function of temperature for as synthesized PNO showing the full PDF	65
Figure 5.2 PDF as a function of temperature for as synthesized PNO showing a close up on the short range section	66
Figure 5.3 PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full PDF	67
Figure 5.4 PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the short range section	68

Figure 5.5	$I(Q)$ as a function of temperature for as synthesized PNO showing the full XRD	70
Figure 5.6	$I(Q)$ as a function of temperature for as synthesized PNO showing a close up on the low Q section	71
Figure 5.7	$I(Q)$ as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full XRD	72
Figure 5.8	$I(Q)$ as a function of temperature for PNO annealed at 750 °C for 25 hours showing a close up on the low Q section	73
Figure A.1	PDF as a function of temperature for PNO annealed at 750 °C for 50 hours showing the full PDF	84
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	85
Figure A.3	PDF as a function of temperature for PNO annealed at 750 °C for 100 hours showing the full PDF	86
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	87
Figure A.5	PDF as a function of temperature for PNO annealed at 750 °C for 200 hours showing the full PDF	88
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	89

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	We should discuss about why PNO is interesting, at least in brief	62
7	need some sort of synthesis information, something along the lines of as 8 previously reported	62
9	need to fix the problem with the $I(Q)$ figures	90
10	Need more references	

11

INTRODUCTION

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

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

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

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

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

34 using the previously derived Q resolution, and the impact of different averaging meth-
35 ods and masks on azimuthal integration, respectively.

36 Chapter 5 will discuss preliminary experimental results investigating the phase
37 changes and local structure of Pr_2NiO_4

38

CHAPTER 1

39

STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES

40 41 1.1 INTRODUCTION

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

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

54 1.2 POTENTIAL ENERGY SURFACES

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

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

66 Experimentally Derived Potential Energy Surfaces

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

75 Potentials

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

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

$$78 \quad R_w = \sqrt{\frac{\sum_{a=a_{\min}}^{a_{\max}} (A_{\text{obs}} - \alpha A_{\text{calc}})^2}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}}^2}} \quad (1.2)$$

$$79 \quad \chi^2_{\text{INVERT}} = \frac{1}{N} \sum_j \sum_r [A_{\text{obs}}(r) - \alpha A_{j,\text{calc}}(r)]^2 \quad (1.3)$$

80

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

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

84 The Rw and χ^2 potentials have been reported numerous times. [43, 33, 6, 34, 45]
 85 Essentially these potentials measure the least squares distance between the observed
 86 scattering and the predicted scattering providing a way to quantify the agreement
 87 between the model and experiment. While RW and χ^2 are now standard in the PDF
 88 community, the INVERT potential is fairly new and aims to incorporate descriptions
 89 of the structural symmetry into the PES. [9, 10] In the case of the INVERT potential
 90 NMR or other symmetry sensitive data is used to describe the number of unique
 91 atomic coordinations. This is then used to describe the number of unique atom-wise
 92 pair distribution functions, thus causing systems with more or less unique coordi-
 93 nation environments to be higher in energy. This approach has been shown to be
 94 useful for C_{60} and other systems which are highly symmetric, creating a PES with an
 95 easier to find minima. [9, 10] However, many times this kind of data is unavailable
 96 when refining the structure causing the potential to be less useful. Additionally, this
 97 potential introduces an element of user bias as the refiner must decide, based on some
 98 spectroscopic data, how many unique environments are in the material. This bias
 99 could be removed by using one of the other potentials with a method for simulat-
 100 ing the observed spectra, allowing the computational system decide what structures
 101 properly reproduce all the observed data.

102 **Forces**

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

103

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

104

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

105

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

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

111 1.3 ENSEMBLES

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

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

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

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

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

$$\Xi = \sum_i P_i(q, p) \quad (1.10)$$

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

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

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

130 where k_b is the Boltzmann constant and T is the temperature of the system. [35]

131 Monte Carlo Modeling

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

$$\exp\left(\frac{-\Delta E}{E_T}\right) < u \quad (1.12)$$

138 where u is a random number $[0, 1)$ and E_T is the thermal energy characteristic to the
139 system. The ability of Monte Carlo modeling to accept “bad” moves allows the system
140 to hop out of local energy minima during the search for the global minimum. Reverse
141 Monte Carlo (RMC) is similar to Monte Carlo except it uses χ^2 as the PES.[34]

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

144 are created equal".[22] RMC, similar to standard Monte Carlo simulations, samples
 145 from the PES at random, usually by translating atoms in the system randomly. This
 146 creates a less efficient, random walk based, exploration of the PES.[22, 36] Thus,
 147 methods for suppressing this random walk nature, while still searching the potential
 148 energy surface fully are needed.

149 Hamiltonian Monte Carlo

150 Hybrid or Hamiltonian Monte Carlo (HMC) can help to address some of these issues.
 151 HMC was developed originally in the lattice quantum chromodynamics community
 152 and provides a more efficient, more scalable approach to PES sampling for Monte
 153 Carlo.[13, 37] In HMC the PES is explored using Hamiltonian dynamics, essentially
 154 following the gradient of the PES to find more acceptable configurations.

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

$$\frac{dq_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} = p_i \quad (1.13)$$

$$\frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial q_i} = -\vec{\nabla}U \quad (1.14)$$

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

$$p_i(t + \delta t/2) = p_i(t) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t)) \quad (1.15)$$

$$q_i(t + \delta t) = q_i(t) + \delta t * p_i(t + \delta t/2) \quad (1.16)$$

$$p_i(t + \delta t) = p_i(t + \delta t/2) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t + \delta t)) \quad (1.17)$$

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

156 moved. Using this notation the gradient is

$$\vec{\nabla}U = \begin{bmatrix} \frac{\partial U}{\partial q_{0,x}} & \frac{\partial U}{\partial q_{0,y}} & \frac{\partial U}{\partial q_{0,z}} \\ \vdots & \frac{\partial U}{\partial q_{i,w}} & \vdots \\ \frac{\partial U}{\partial q_{n,x}} & \frac{\partial U}{\partial q_{n,y}} & \frac{\partial U}{\partial q_{n,z}} \end{bmatrix} = \begin{bmatrix} \vec{\mathcal{F}}_0 \\ \vdots \\ \vec{\mathcal{F}}_i \\ \vdots \\ \vec{\mathcal{F}}_n \end{bmatrix} \quad (1.18)$$

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

175 **No-U-Turn Sampling**

176 Two parameters must be specified in HMC simulations, the step size δ and the number
177 of steps N . The step size is critical to the stability of the fitting procedure: with
178 a too small δ the simulation runs inefficiently producing structures too close to the
179 previous, whereas with a too big δ the linear approximation for the forces breaks
180 down and often the simulated NP explodes. The number of steps to take during the
181 dynamics is equally important and an inappropriate choice may result in backtracking
182 or random walk characteristics in the simulations. In this work, we employ the No-
183 U-Turn Sampling (NUTS) method recently proposed by Hoffman and Gelman to
184 address this issue [22]. In the NUTS method δ and N are dynamically computed
185 by examining the ratio of accepted to rejected configurations as well as whether
186 or not the simulation has started to take a U-turn. The U-Turn criteria makes
187 certain that the simulation stops when it begins to backtrack, preventing excess
188 computation on configurations that have very little new information to offer. The
189 use of NUTS leaves us with two simulation parameters: the simulation temperature
190 and the target acceptance. Hoffman and Gelman have empirically shown that the
191 ideal target acceptance, which governs the dynamics time steps, is .65, which we have
192 used for all of the simulations here. The simulation temperature sets the magnitude
193 of the random starting momenta for the atoms at the beginning of each dynamics
194 run [22].

195 **Grand Canonical Ensemble**

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

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

205 **Ensemble description**

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

$$\Xi = \sum_{N=0}^{\infty} Q(N, V, T) \exp \frac{N\mu\beta}{T} \quad (1.19)$$

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

213 **Grand Canonical Monte Carlo**

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

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

217 Similarly the removal of an atom has the probability:

$$\min[1, \frac{(N)\Lambda(T)^3}{V} e^{-\beta\Delta U - \beta\mu}] \quad (1.21)$$

218 However, both of these equations depend of the overall simulation volume and the
219 thermal wavelength, which is undesirable as these are not really properties that we
220 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.

234 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:

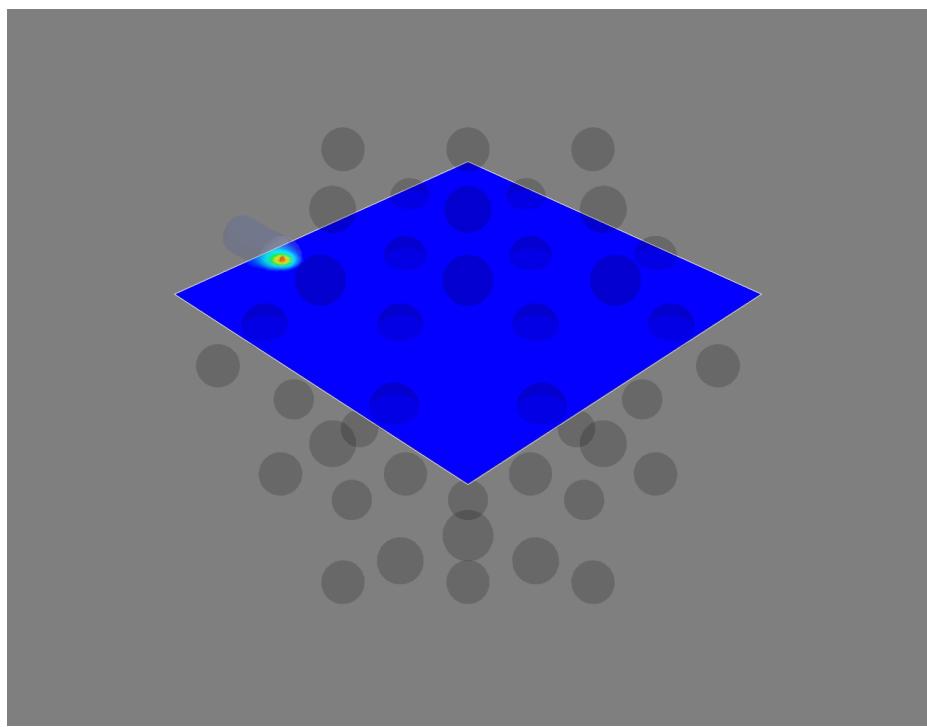
$$P_{i,j,k} = \frac{e^{\beta\Delta U_{i,j,k}}}{\sum_{i,j,k} e^{\beta\Delta U_{i,j,k}}} \quad (1.22)$$

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. [48]

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

250 Figure 1.1a shows an example map for atomic addition in a Au54 atom system,
251 with an Au55 atom target. Figure 1.1b shows the results of a few GCMC insertions
252 with biasing, showing the focusing of the simulation on the missing atom. The high
253 density of insertions around the missing atom would not have been possible without
254 the biasing.

(a)



(b)

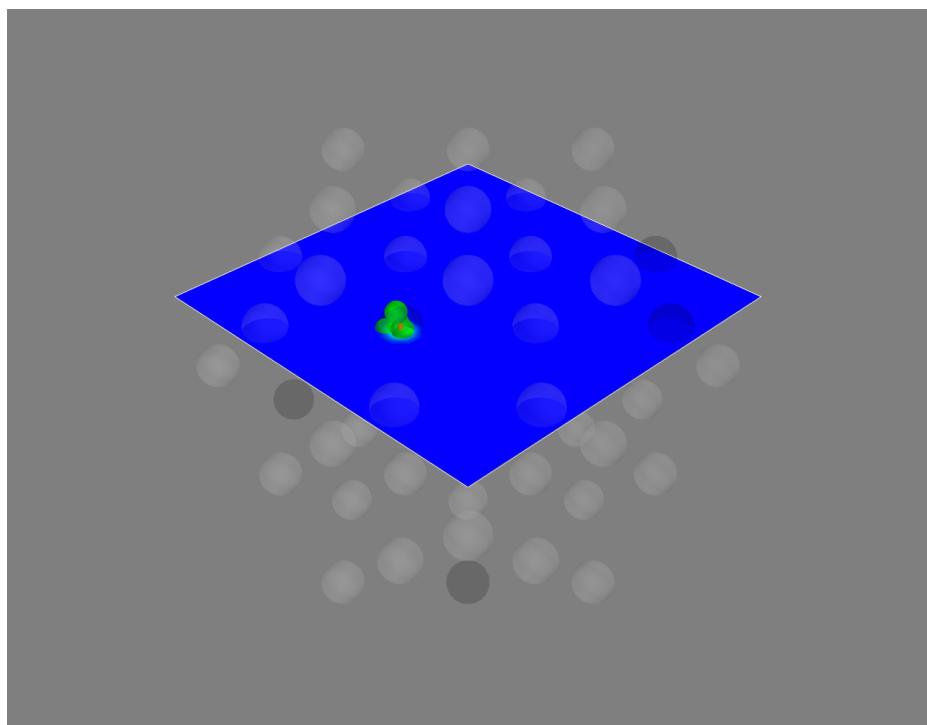


Figure 1.1: These figures show slices of the three dimensional addition potential energy surface. a) shows the addition probability, with the red area most probable. b) shows the results of trial insertions, where the green spheres denote where insertions were attempted.

255 1.4 CONCLUSIONS

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

269

CHAPTER 2

270

ATOMIC PAIR DISTRIBUTION FUNCTION:

271

THEORY AND COMPUTATION

272 2.1 INTRODUCTION

273 Atomistic structural insight is essential for understanding and controlling a mate-
274 rial's properties and functions, which has led to some of the most exciting advances
275 in modern materials science and engineering. X-ray diffraction techniques are one of
276 the most powerful tools for probing atomic structures with ultimate precision. Tradi-
277 tionally, thousands of diffraction peaks are analyzed using refinements of a structural
278 model with few parameters to determine the 3D structure of bulk single crystals with
279 high precision [19]. However, real engineered materials differ from ideal single crystals
280 by showing a complexity in morphology, crystalite size, and atomic structure. The
281 X-ray Powder Diffraction method (XPD) is among the most widely used methods
282 for solving the structure of micro-crystals. The XPD technique utilizes hundreds of
283 diffraction peaks and constrains the refinement of the structural model to few pa-
284 rameters in order to resolve the structure [40]. Solving the atomic structure becomes
285 difficult using traditional x-ray diffraction techniques when the size of the material or
286 its important features is reduced to the nanometer scale with non-periodic or short-
287 periodic atomic arrangements. Materials consisting of particles with sizes less than
288 a few tens of nanometers, often called nanoparticles (NPs), are structurally more
289 complex than their bulk cousins. This structural complexity is often attributed to
290 the large number of surface atoms which have incomplete coordination spheres [49],

291 surface relaxation [23], and surface environment effects [44, 21, 32]. These effects
292 make the precise determination of 3D atomic structure of NPs far more complicated
293 and problematic [3].

294 Over the years many advances have been made to address the famous “nanostruc-
295 ture problem”[3], for example, by using the atomic Pair Distribution Function (PDF)
296 analysis of x-ray and neutron total scattering data [15, 38, 45], PDF analysis combined
297 with molecular dynamics simulations [51, 20], bulk crystallography approaches [24],
298 and others. Attempts with non-diffraction based approaches have also been made,
299 including Transmission Electron Microscopy (TEM) [11], Raman spectroscopy[27],
300 Extended X-ray Absorption Fine-Structure Spectroscopy (EXAFS)[18] and Nuclear
301 Magnetic Resonance (NMR) [1].

302 In this chapter the PDF and its gradients will be derived. These expressions,
303 when combined with the PES and statistical mechanical treatment from chapter 1,
304 will allow for the solution of atomic structures. This chapter will also develop a
305 computational framework for evaluating the PDF and its gradients using Graphical
306 Processing Units (GPUs) to enable fast structural solution.

307 2.2 THEORY

308 To properly understand the PDF and its limitations we need to derive its mathemat-
309 ics. The PDF has been previously derived many times so it is not re-derived here.
310 This discussion of the PDF and its gradients use the notation of Farrow and Billinge.
311 [16]

312 Derivation

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

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

319 Analytically Gradients

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

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

329 where

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

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

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

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

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

330 where Q is the scatter vector, f_i is atomic scattering factor of the i th atom, and r_{ij} is
 331 the distance between atoms i and j and has q dependence. [25] For simplicities sake

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

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

333 where

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

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

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

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

334 The derivatives are as follows:

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

335 where

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

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

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

337 Thus

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

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

338 **Without ADPs**

339 Without ADPs the equations simplify down to

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

340 and

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

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

343 Periodic Boundary Conditions

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

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

350 where

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

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

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

353 2.3 COMPUTATION

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

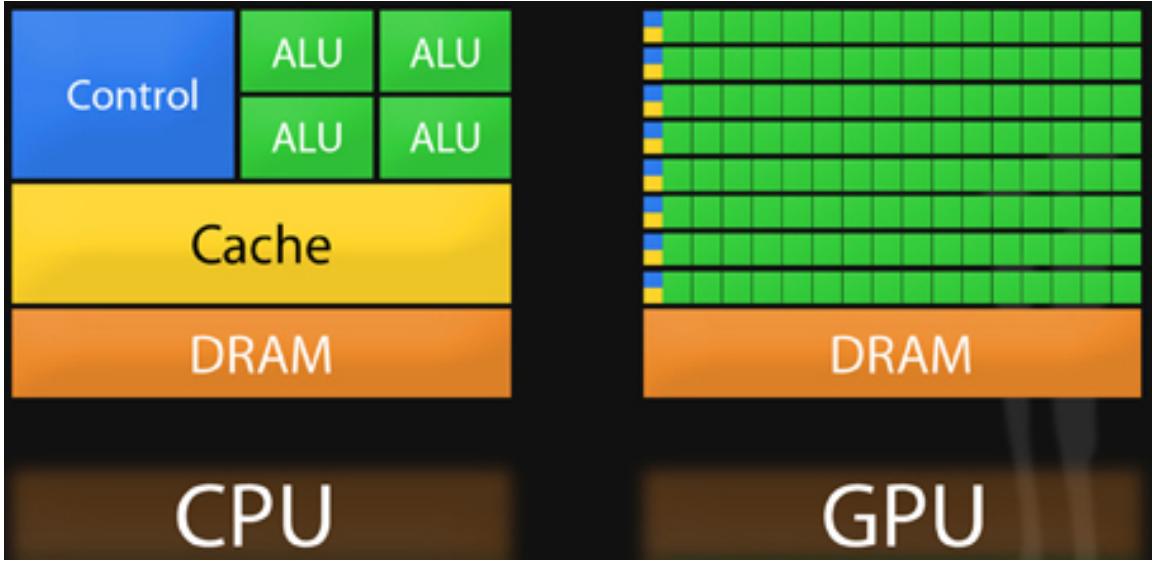


Figure 2.1: Comparison of the CPU and GPU chip architectures from [5]. The ALU are the arithmetic logic units which perform the mathematical operations, the DRAM holds most of the data, although it is slower to access, the Cache holds rapidly accessed data, and the Control controls the execution of software. 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.

358 **HPC and GPUs**

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

363 **GPUs and Parallelization**

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

370 existing parallel FFT algorithms.

371 Graphical Processing Units (GPUs) are particularly well suited to the task of com-
372 puting PDFs. GPU chip architecture is designed to perform many tasks simultaneously
373 by having potentially thousands of cores. Figure 2.1 shows the comparison of CPU
374 and GPU architectures. As the figure shows the GPUs have a very different layout
375 of computational processors (ALUs) and memory. While each ALU is simpler on the
376 GPU, requiring the instructions to be less demanding in terms of memory, there are
377 many more of them. The greater number of processors allows each atomic pairing
378 to be placed on its own processor, so long as the math can be broken into simpler
379 operations. The equations are broken up on the GPUs into various pieces which cor-
380 respond to the α, β, τ and Ω as shown in equations 2.8-2.11 and subequations as
381 needed. For example, while β is computed in one step, Ω requires the computation
382 of the displacement array, then the distance array and finally the Ω array. The exact
383 breakdown of processes, how the problems are broken down and spread across the
384 processor has been optimized for speed and reliability.

385 **Map from ij space to k space**

386 The above equations, although formally correct, are very inefficient. $F(Q)$ and its
387 gradient are indexed over all the atoms twice, however there are symmetries that
388 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

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

389

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

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

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

For the gradient a similar mapping is used:

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

405

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

411 GPU Memory Allocation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

426 Thus the pair allotment is:

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

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

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

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

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

429 Thus the gradient allotment is:

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

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

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

431 Thus the allocation is:

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

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

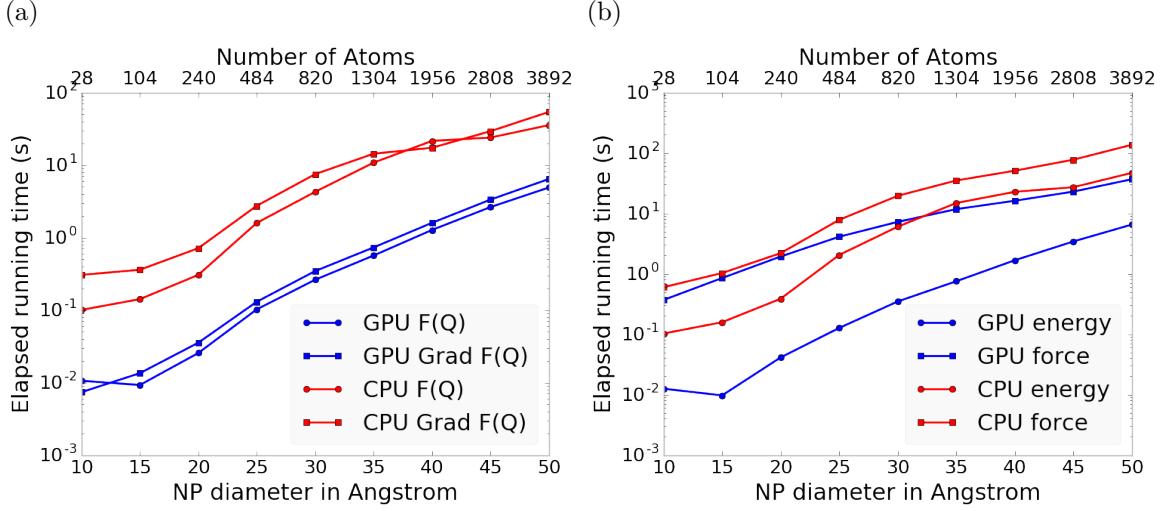


Figure 2.2: Speed comparison of CPU and GPU implementations. a) shows the time to compute the $F(Q)$ by itself. b) shows the time to compute the Rw based energy for Au NPs of various sizes, which includes computing $F(Q)$, its FFT, and the Rw .

435 Speed and Scaling of PDF Computation

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

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

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

455 2.4 CONCLUSIONS

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

463

CHAPTER 3

464

BENCHMARKING

465

Also some introduction would be great

466

this just needs a lot of work

467 3.1 INTRODUCTION

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

472 3.2 PDF

473 The formation of NPs with both crystallographic and non-crystallographic structures
474 [32] and with different chemical patterns [17] are well documented. For simplicity,
475 we chose monometallic Au clusters as benchmarks and considered two groups of
476 structures with different size and degrees of structural disorder in order to assess
477 the reliability and efficiency of our HMC method for solving atomic structures from
478 PDFs. The first group consists of Au_{55} clusters with different degrees of disorder,
479 including a crystalline cluster structure in O_h (Octahedral) symmetry, a structure
480 with a disordered surface, and an amorphous structure. The second group consists
481 of the crystallographically solved Au_{102} structure as in the $\text{Au}_{102}\text{MBA}_{44}$ nanocrystals
482 [24, 31]. We used optimized structures from the Density Functional Theory (DFT)

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

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

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

489 Spin-polarized DFT calculations were carried out using the Vienna ab initio simu-
490 lation package (VASP) [29, 30] within the Perdew-Burke-Ernzerhof (PBE) exchange-
491 correlation functional [41]. The projected augmented wave method [4] and a kinetic
492 energy cutoff of 400 eV were used. Structural optimization was performed until the
493 total energy and ionic forces were converged to 10^{-6} eV and 10 meV/Å, respectively.
494 The amorphous Au_{55} structures were generated by simulated annealing using the
495 classical embedded atom method potential [47]. Different annealing temperatures
496 between 1200 K and 1670 K (bulk melting temperature of Au) were used and the
497 thermally equilibrated structures were cooled down to 300 K before minimization at
498 0 K. Further optimization using DFT leads to total energies that vary within 1-2
499 eV among different amorphous structures and the lowest energy one was used as the
500 target structure. The target structure of Au_{102} was taken as the Au_{102} core of the
501 DFT-optimized $\text{Au}_{102}\text{MBA}_{44}$ cluster [31].

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

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

505 The thresholded spring potentials are based on those previously proposed on by Pe-
506 terson [42], 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

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

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

525 Model Parameters

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

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

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

537 **Au55: surface relaxed**

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

545 **Run Parameters**

546 R_{\min} and R_{\max} for this simulation were 2.45 Å and 11.4 Å, respectively, with $\delta Q =$
547 0.24 Å⁻¹. The simulation ran for approximately 34 minutes, over a total of ~40
548 thousand configurations. The results are shown in Fig. 3.1.

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

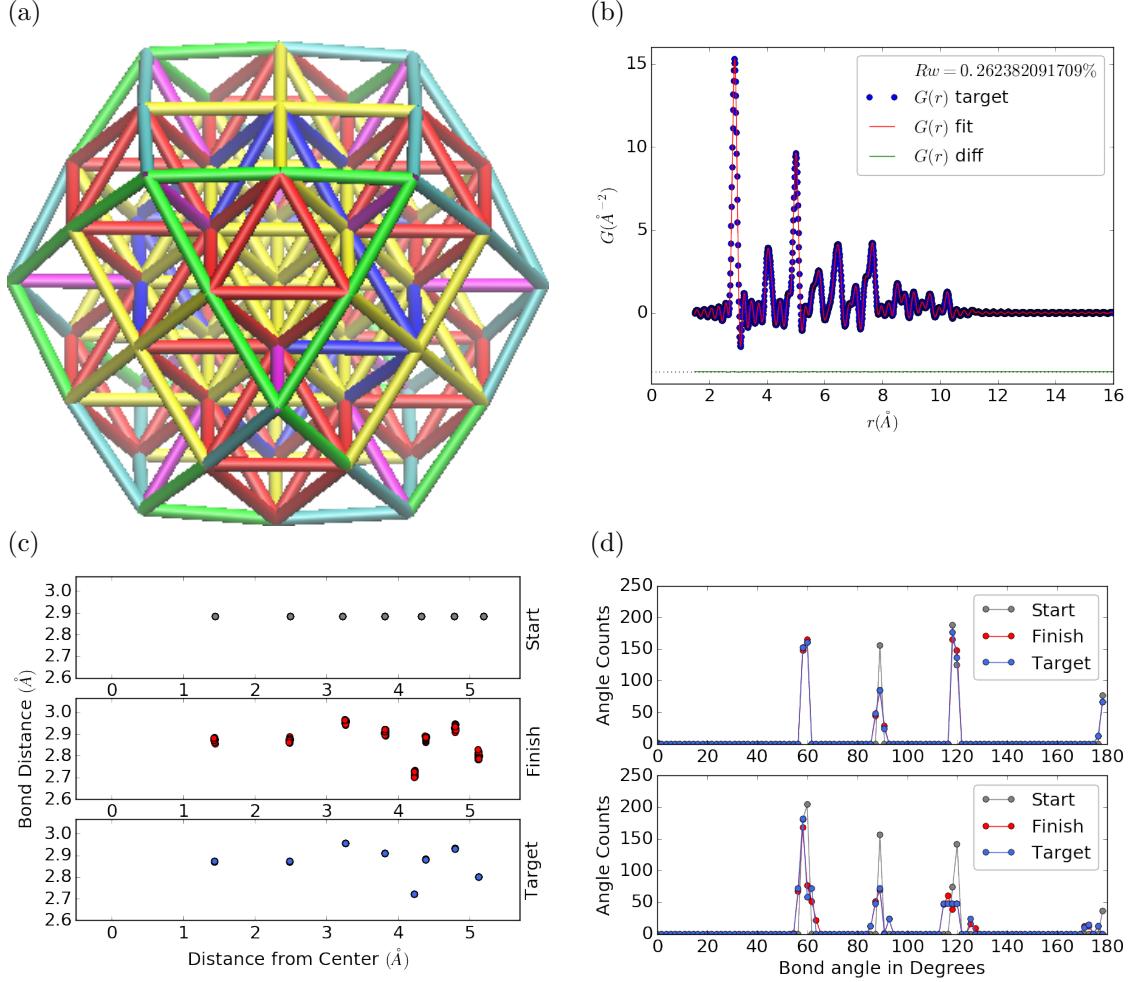


Figure 3.1: Au_{55} PDF fitting of DFT-optimized $c\text{-Au}_{55}$. (a) the final structural solution ($Rw=0.3\%$) with bond lengths color-coded by step of 0.05\AA , (b) the target PDF(blue dots) overlaid with the PDF of the final structure (solid red lines) with the difference in green lines offset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

556 below, the situation is rather different for much more disordered target structures.
 557 Interestingly, the small-box solution using PDFgui[15] yields a rather large Rw of
 558 43%, due to the failure to fit the surface contracted atoms with a unit cell. The PDF
 559 fits of the starting structure and small-box solution are shown

560 Put this somewhere

561 .

562 **Au55: surface disordered**

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

569 R_{\min} and R_{\max} for this simulation were 1.95 Å and 12.18 Å, respectively, with
570 $\delta Q = 0.23 \text{ \AA}^{-1}$. The simulation ran for approximately 3.6 hours, over a total of ~ 270
571 thousand configurations. The results of the simulation are shown in Fig. ??.

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

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

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

594 Au55: amorphous

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

600 R_{\min} and R_{\max} for this simulation were 2.6 Å and 11.26 Å, respectively, with
601 $\delta Q = 0.25 \text{ \AA}^{-1}$. The simulation ran for approximately an hour, over a total of ~ 87
602 thousand configurations. The results of the simulation are shown in Fig. ??.

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

615 **Au102: triple phase**

616 Our final benchmark is Au₁₀₂, whose structure was initially solved by Jadzinsky and
617 co workers using x-ray crystallography [24] and further confirmed by DFT studies
618 [31]. The Au₁₀₂ structure consists of three main parts, a 49-atom Marks decahedron
619 core, two C_5 caps consisting of 20 atoms each, and 13 equatorial atoms. Unlike
620 previous cases, the multi-symmetry nature of the structure, i.e. each part has its own
621 distinct symmetry, poses a challenge for PDF-based solution of the structure. This is
622 because of the atomically centralized nature of the PDF, in which each atom “sees”
623 a density of other atoms surrounding it and has a strong tendency towards becoming
624 the center of the main symmetry group. Such tendency may lead to a solution where
625 some of the correct atomic symmetries are discarded in favor of the core symmetry.

626 **Starting from fcc structure**

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

629 R_{\min} and R_{\max} for this simulation were 2.7 Å and 16. Å, respectively, with $\delta Q =$
630 0.18 Å⁻¹. The simulation ran for approximately two hours, over a total of ~82
631 thousand configurations. The results of the simulation are shown in Fig. 3.4.

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

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

651 **Marks decahedron**

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

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

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

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

670 **Au147**

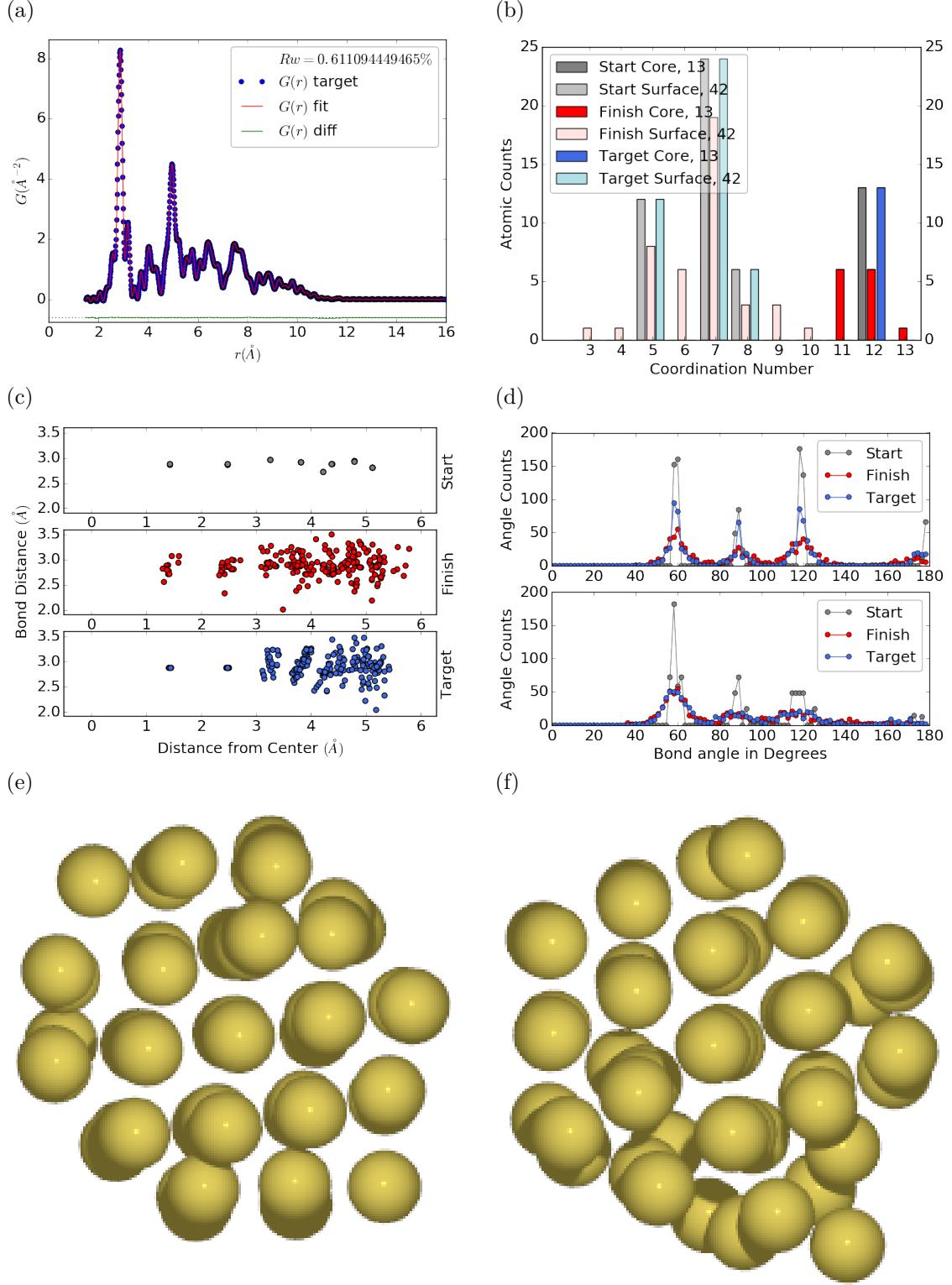


Figure 3.2: Au_{55} PDF fitting of surface-disordered Au_{55} . a) the target structure, b) the final structural solution ($R_w=0.6\%$), c) the comparison of PDFs, d) the CN distribution with the number of atoms in either the core or the surface, e) the radial bond distribution, and f) the bond angle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

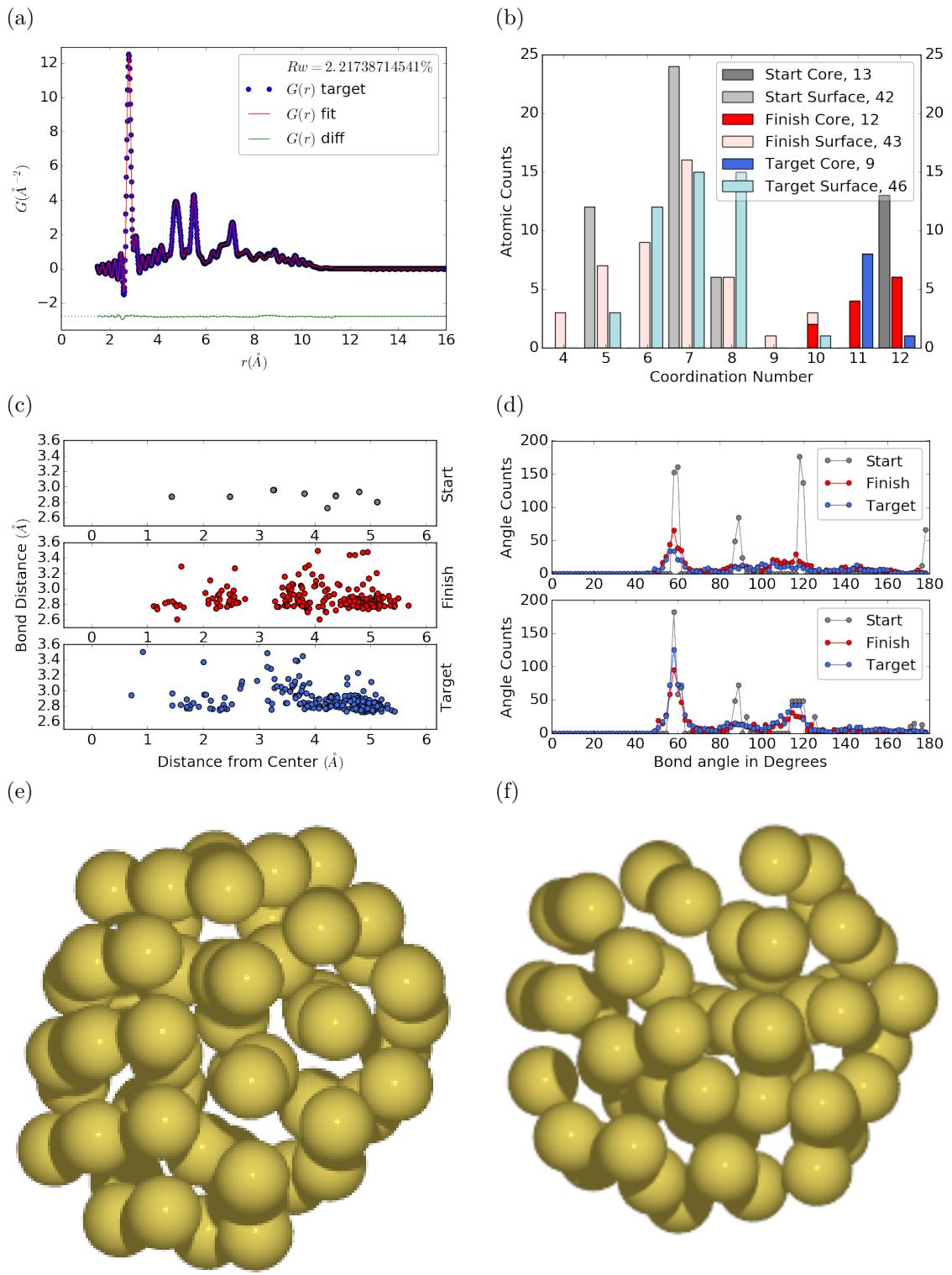


Figure 3.3: Similar to figure 3.2 for DFT-optimized amorphous Au_{55} .

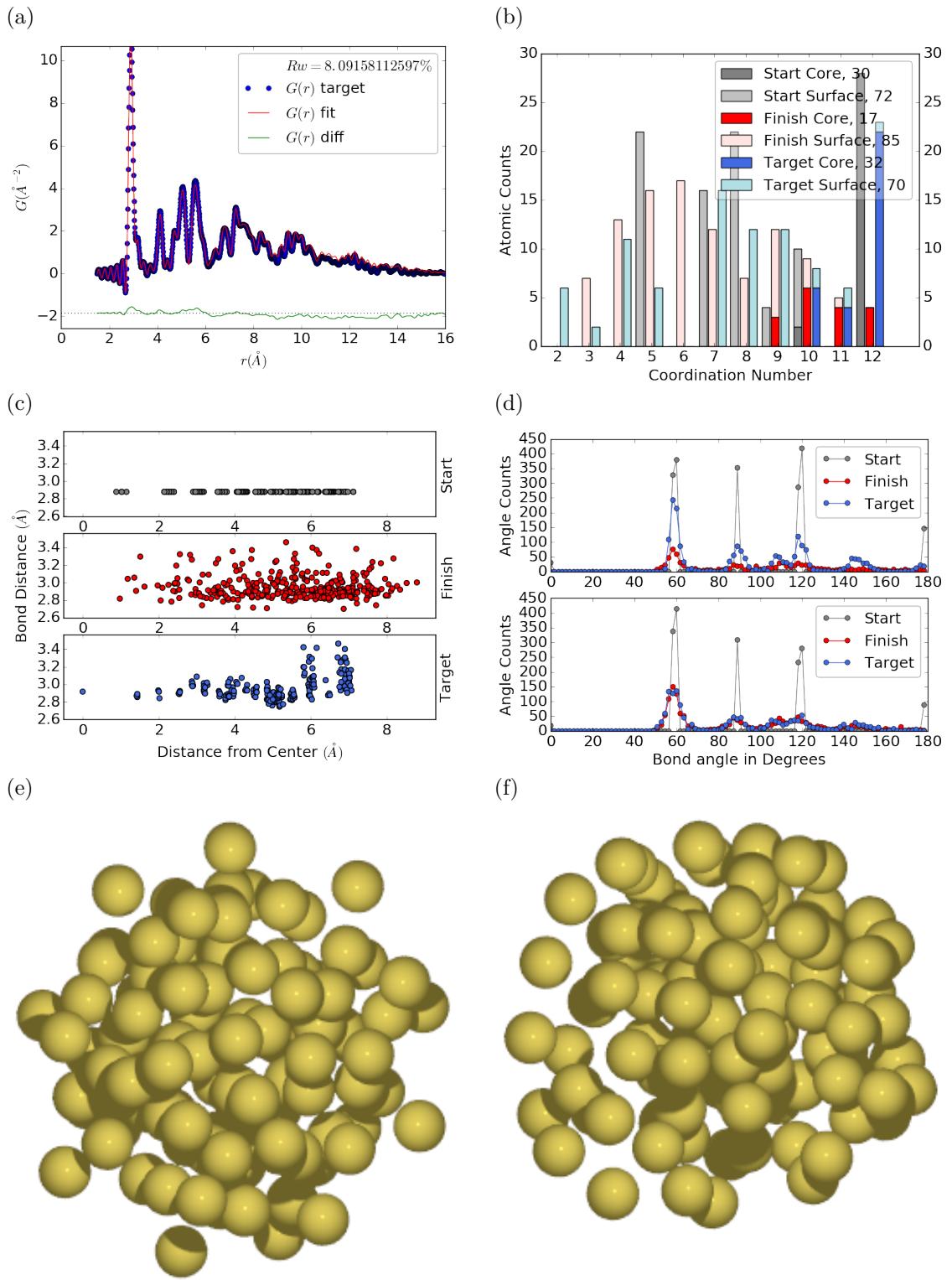


Figure 3.4: Similar to figure 3.2 for Au_{102} as in DFT-optimized $\text{Au}_{102}\text{MBA}_{44}$ cluster.

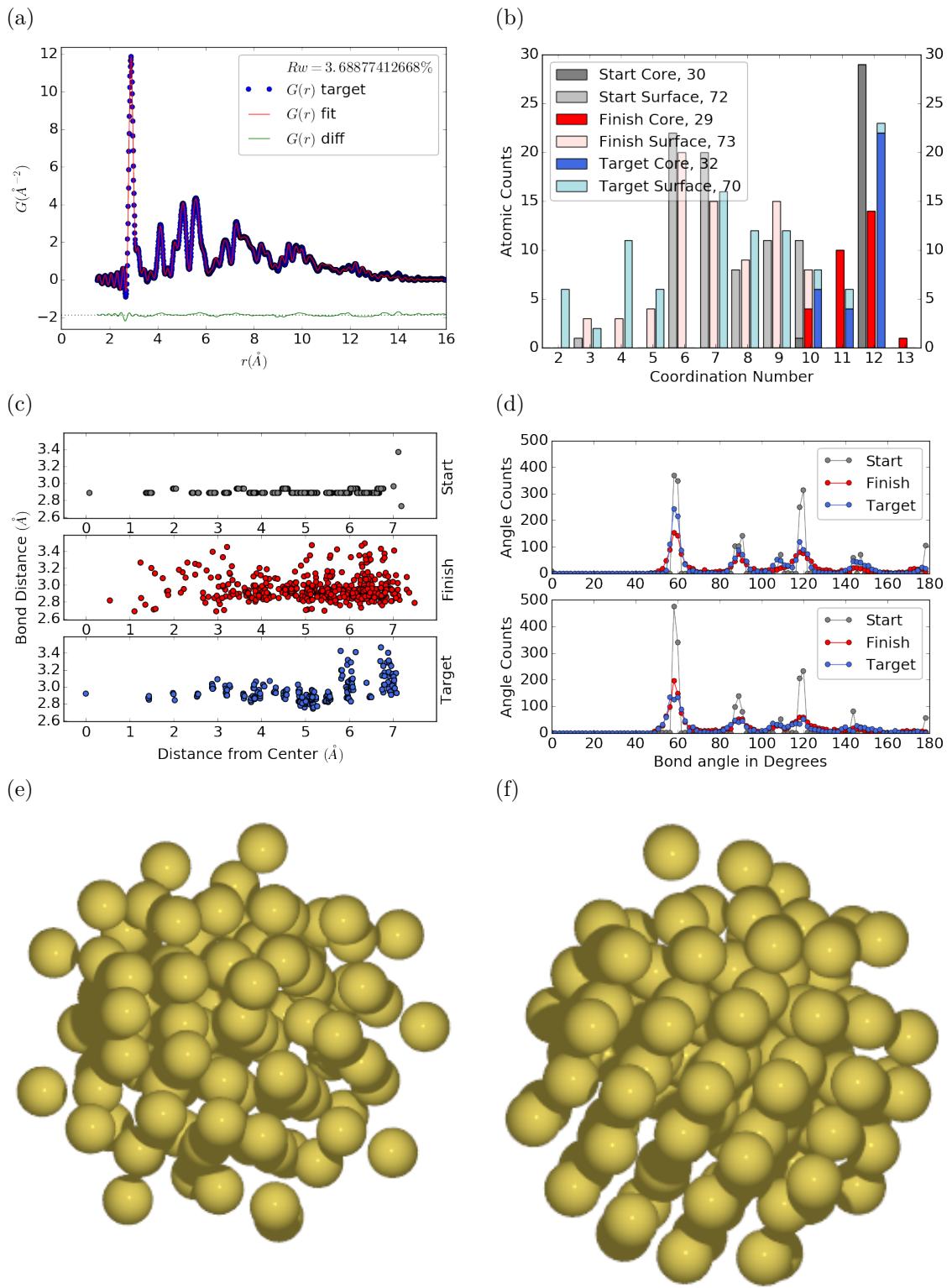


Figure 3.5: Similar to Fig. 3.4 with Marks decahedron as the starting structure.

671 3.3 PDF WITH ADPs

672 **ADP 50**

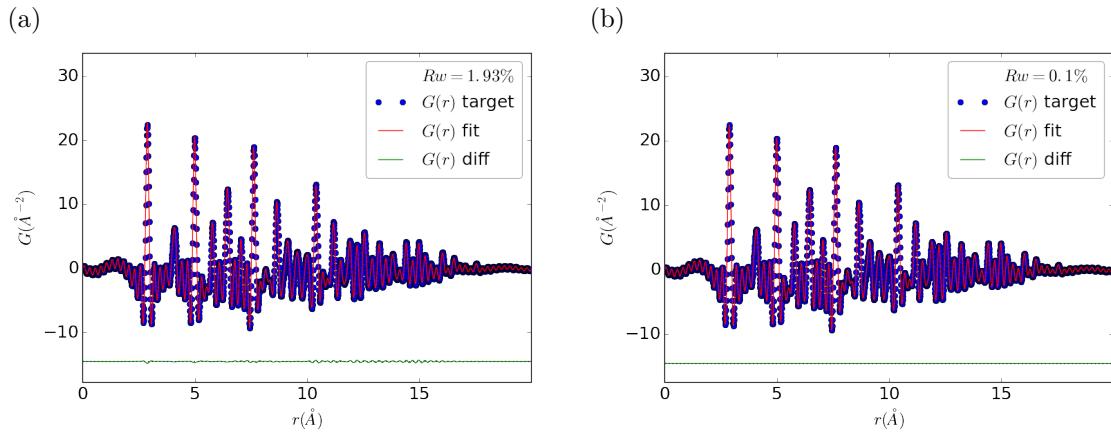


Figure 3.6: Refinement of adps

673 1. Basic 50% larger magnitude

674 2. Random addition to APDs

675 3. Janus ADPs

676

CHAPTER 4

677

X-RAY TOTAL SCATTERING DATA ACQUISITION AND

678

PROCESSING

4.1 INTRODUCTION

X-ray total scattering experiments are generally performed at synchrotron light sources, as only these sources can provide the needed flux, energy, and high momentum transfer vectors needed to obtain reliable PDFs. [8, 14] Despite the need for a dedicated facility to perform the total scattering experiments, the experiments themselves are fairly forgiving, allowing for reactive gaseous environments, experiment temperatures ranging from 2 K to 1200 K, and even electrochemical cycling. [7, 44, 46] The rapid PDF data acquisition associated with 2D area detectors creates a data management problem, as 96 hours of beamtime could result in almost 10,000 images which need to be associated with the experimental conditions and detector metadata. [8] Finally, all this data needs to be processed by masking bad pixels and regions, integrating azimuthally, and converting the scattering data to the PDF. [28, 26, 50, 39, 2]

4.2 DETECTOR Q RESOLUTION

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

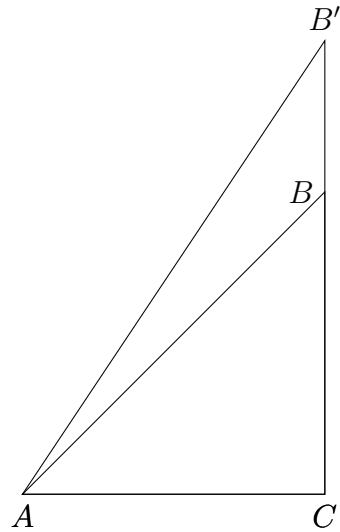


Figure 4.1: Scattering onto a flat detector

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

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

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

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

703 where λ is the x-ray wavelength.

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

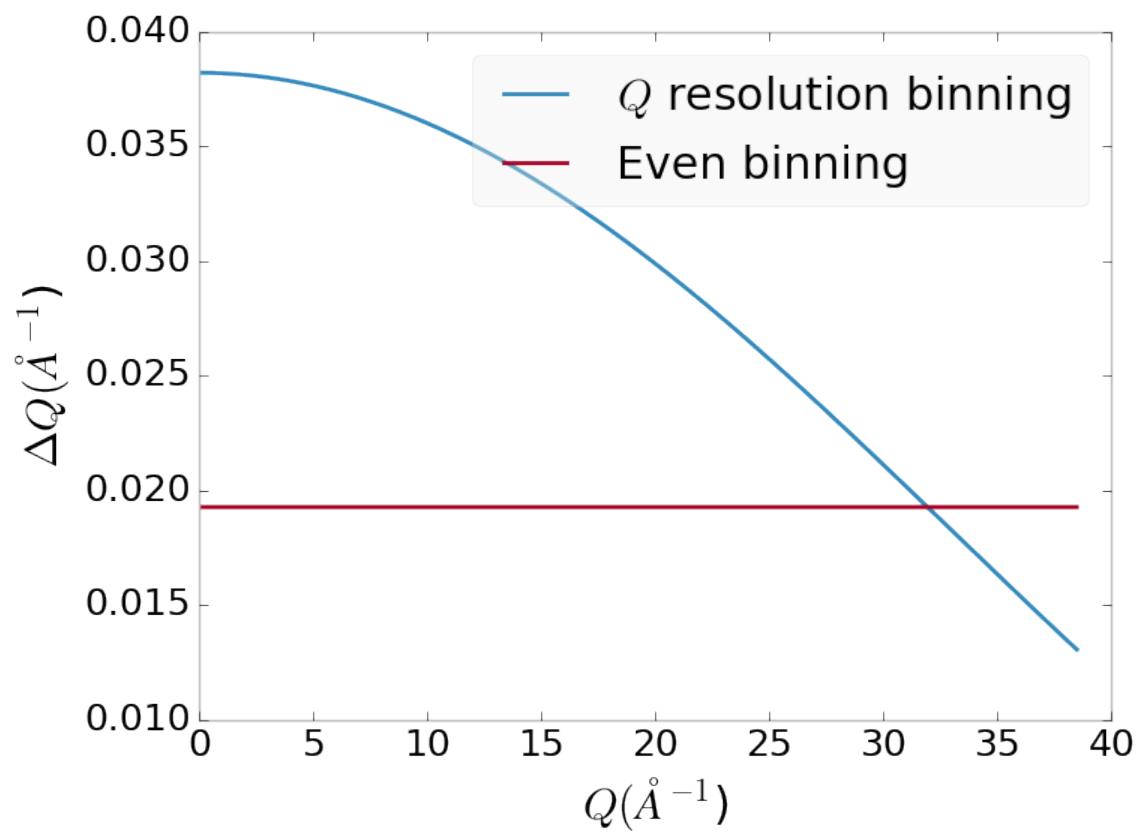


Figure 4.2: Q resolution as a function of Q .

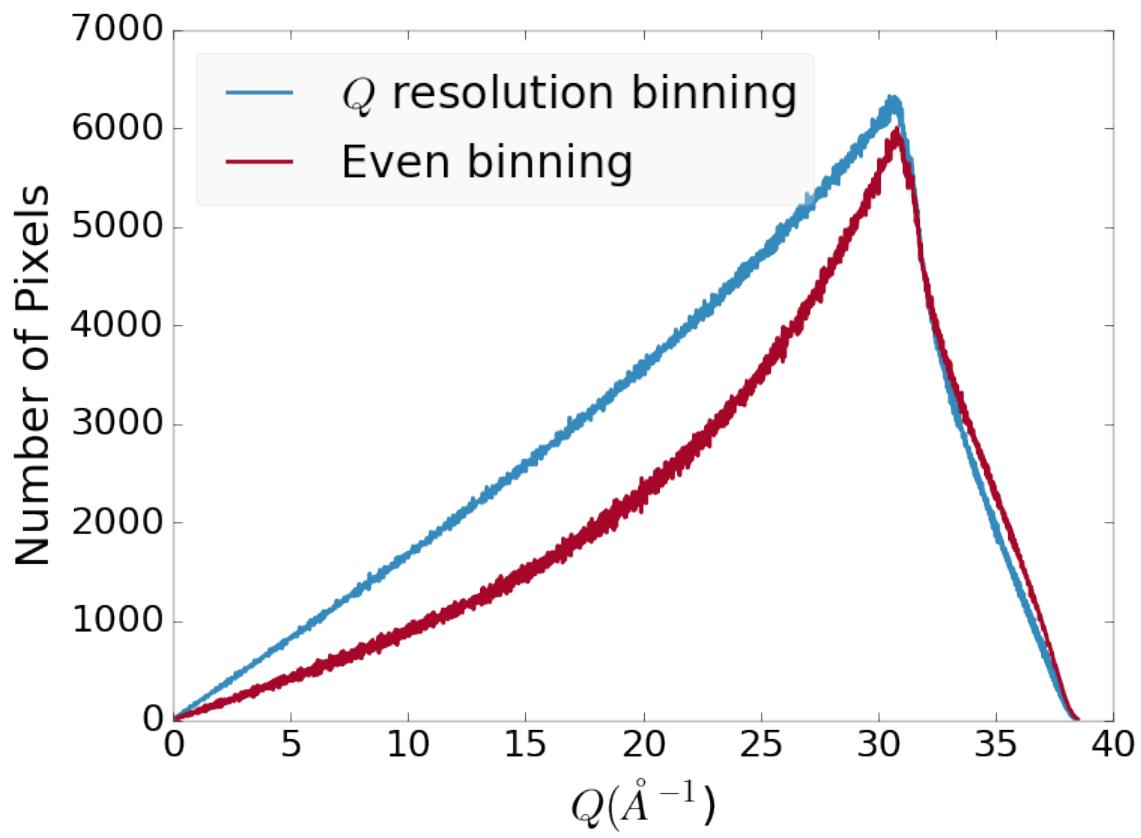


Figure 4.3: Number of pixels as a function of Q , binned at the Q resolution of the detector.

707 4.3 AUTOMATED MASK GENERATION

708 **Introduction**

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

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

727 **Algorithm Design**

728 The masking algorithm procedure takes in the image and a description of the pixel
729 positions in either distance from the point of incidence or in Q . The image is then
730 integrated twice, producing both the mean $I(Q)$ and the standard deviation of each
731 $I(Q)$ ring. The mask is created by comparing the pixel values against each ring's

732 standard deviation and threshold α . Note that the threshold can be a function of
733 distance from the point of incidence or Q .

734 **Test Cases**

735 To study the effectiveness of the masking we ran the algorithm against both simulated
736 experimental data. In the case of the simulated data four systems were created: 1)
737 dead/hot pixels with varying numbers of defective pixels, 2) beamstop holder with
738 varying beamstop holder transmittance, 3) rotated beamstop holder with varying
739 beamstop holder transmittance, and 4) beamstop holder with dead/hot pixels. The
740 base scattering was produced by

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

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

751 **Results and Discussion**

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

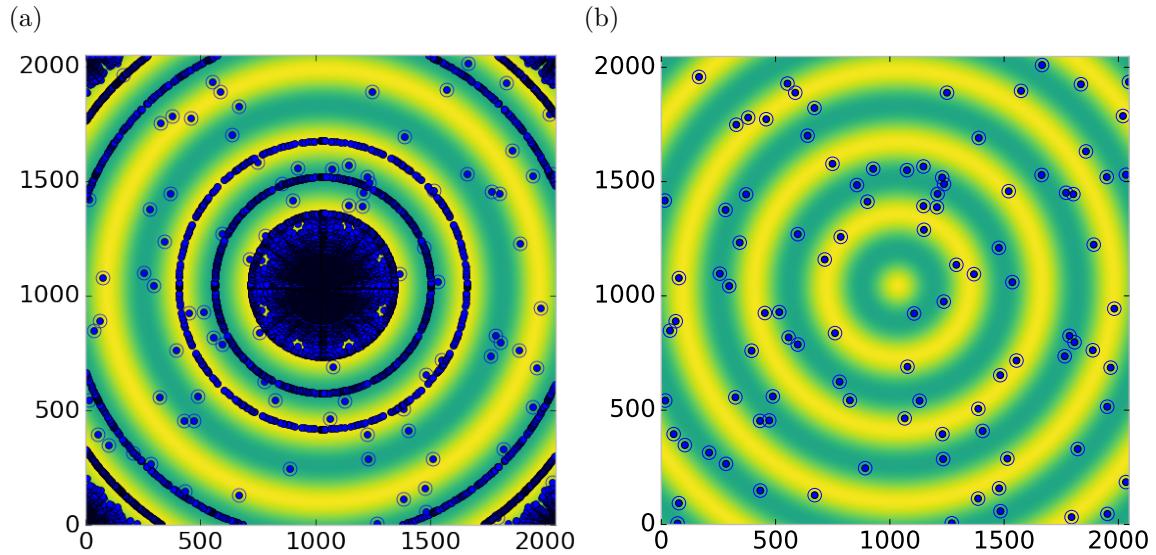


Figure 4.4: Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

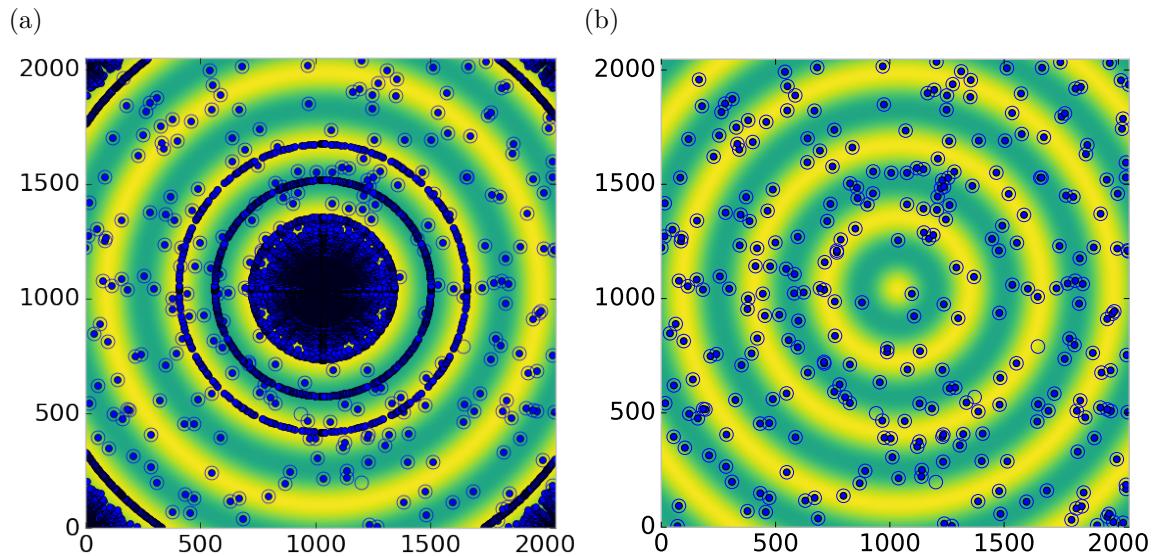


Figure 4.5: Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

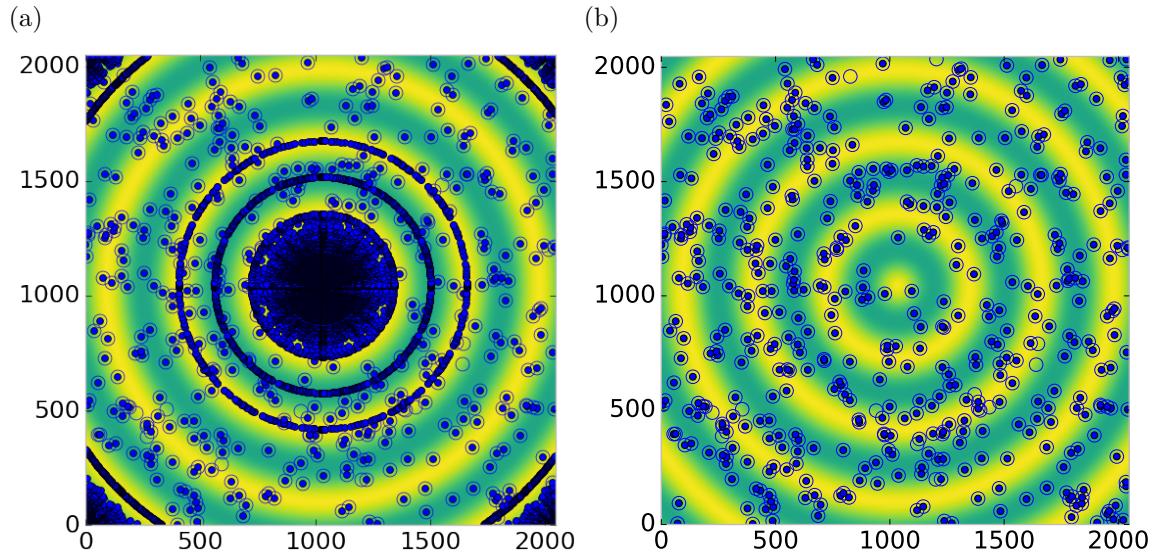


Figure 4.6: Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

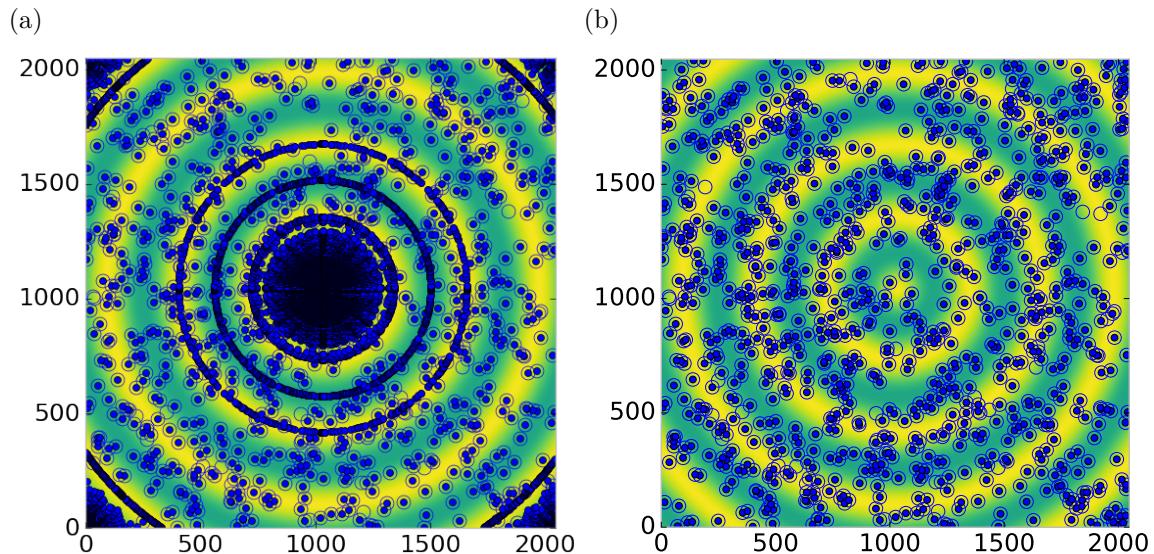


Figure 4.7: Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

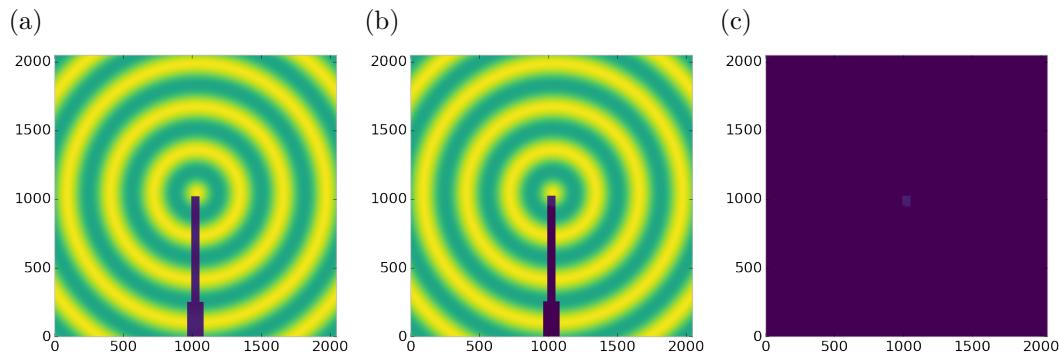


Figure 4.8: Generated beamstop holder masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

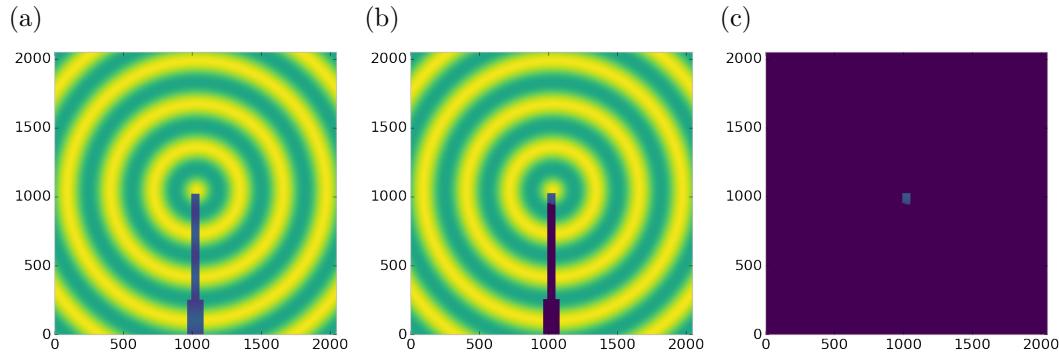


Figure 4.9: Generated beamstop holder masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

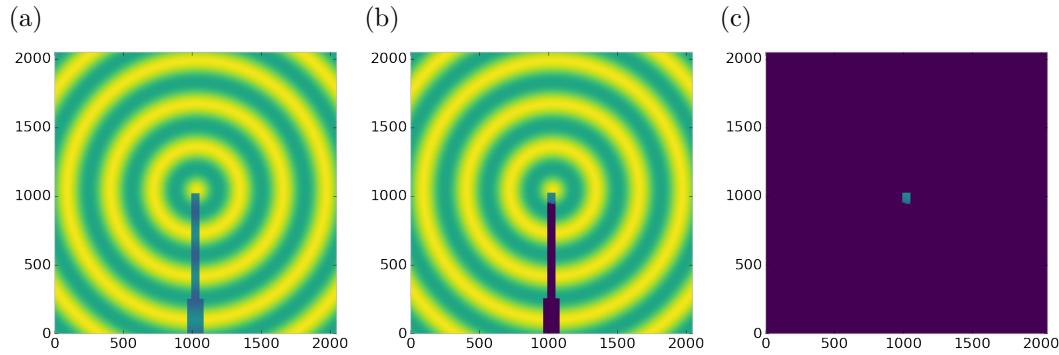


Figure 4.10: Generated beamstop holder masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

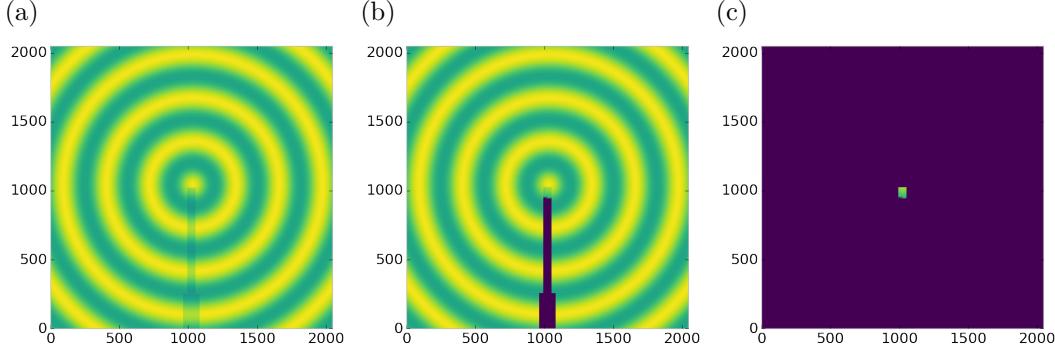


Figure 4.11: Generated beamstop holder masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

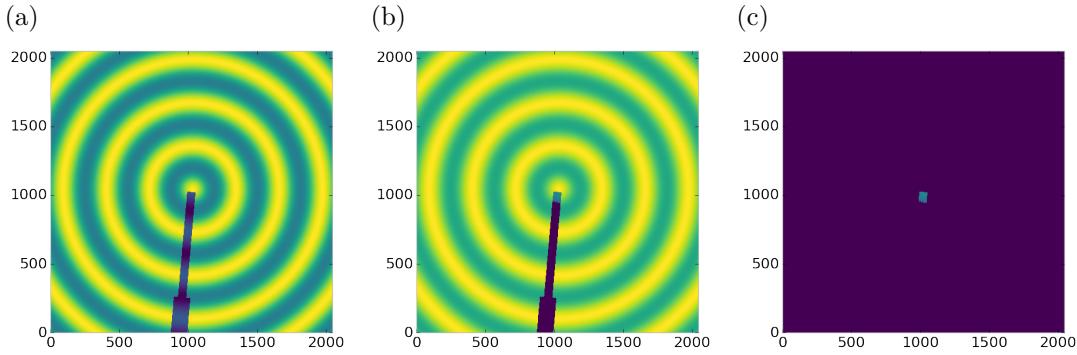


Figure 4.12: Generated beamstop holder masks which is rotated away from vertical

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

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

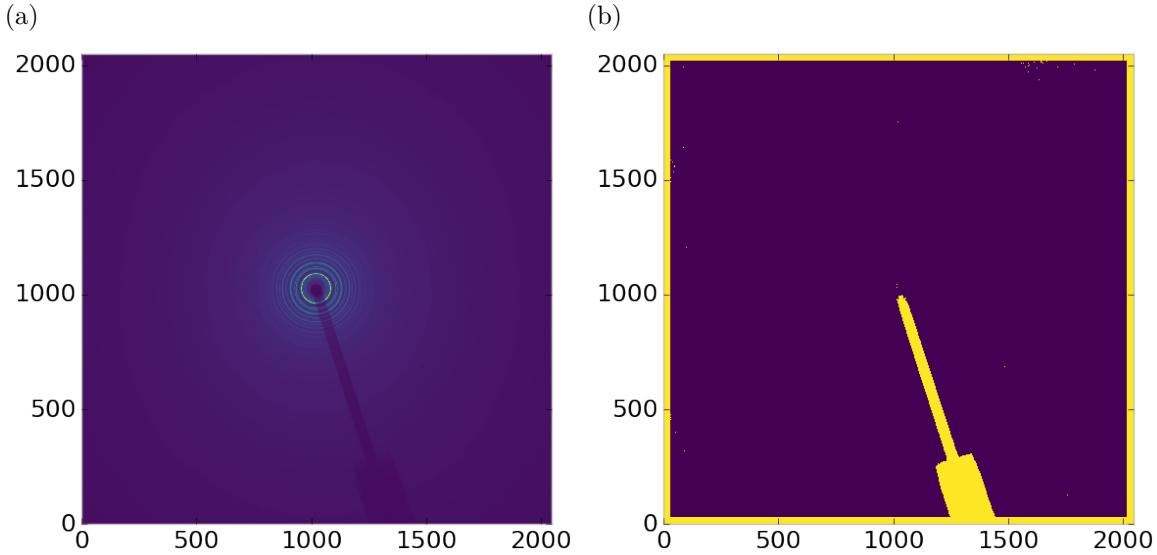


Figure 4.13: Masked experimental data. a) the raw image, b) the mask

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

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

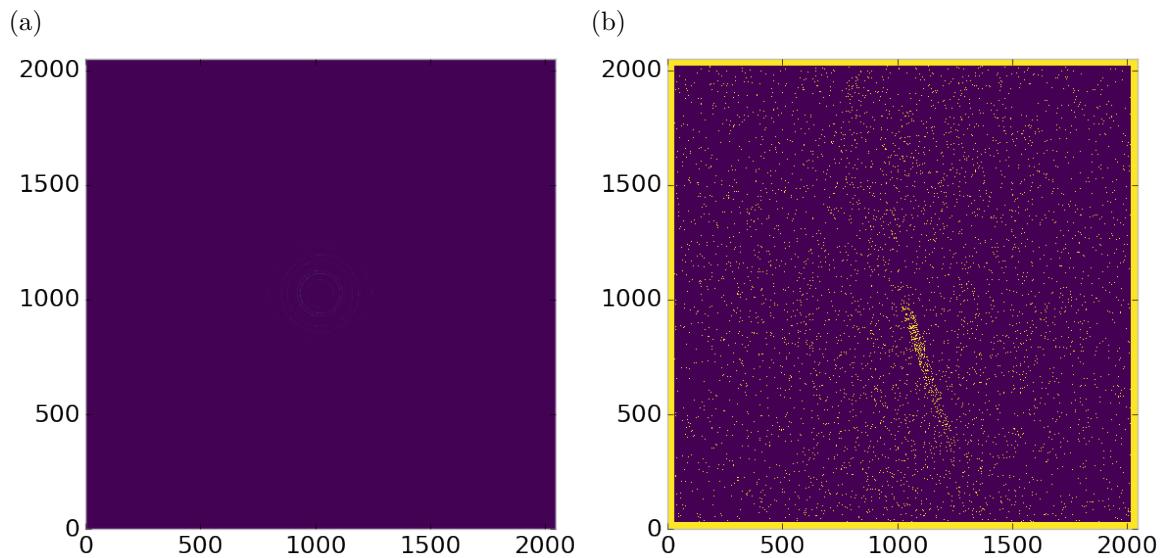


Figure 4.14: Masked experimental data with Pt single crystal signal. a) the raw image, b) the mask

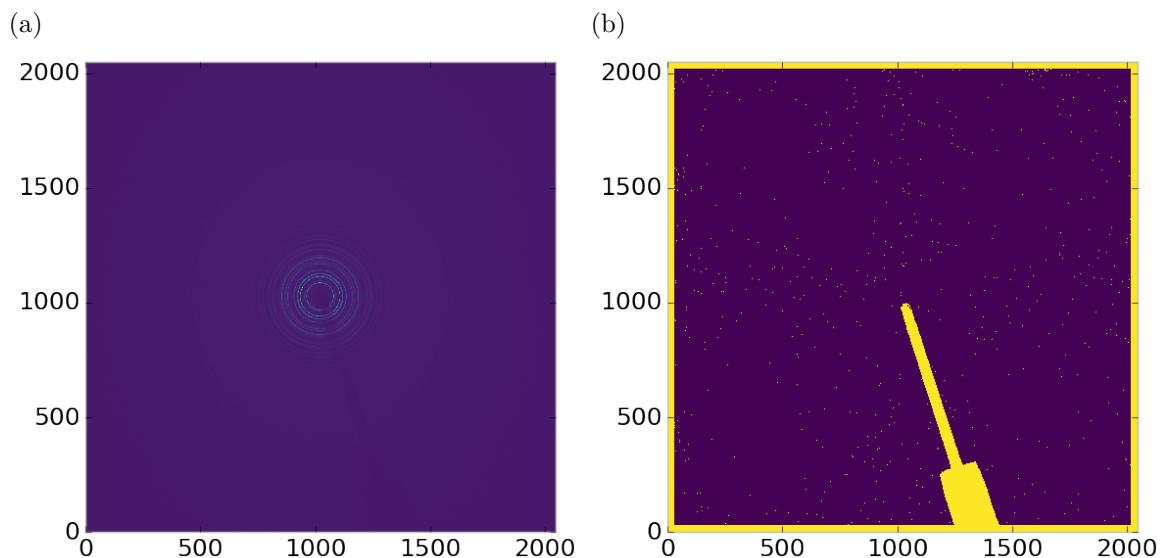


Figure 4.15: Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask. a) the raw image, b) the mask

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

789 **Conclusions**

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

802 **4.4 AUTOMATED IMAGE AZIMUTHAL INTEGRATION**

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

808 $I(Q)$.[50]

809 Figures 4.16-4.18 show the importance of masking and the choice of average func-
810 tion. All the figures were produced using the same dataset, 50 °C Pr_2NiO_4 taken at
811 the APS's 11-ID-B on a Perkin Elmer area detector. The automatic masking alpha
812 was 3 standard deviations from the mean. While it is difficult to observe the changes
813 the mask causes in the full $I(Q)$ plot (subfigures a) and b)), the standard deviation
814 plots show the effect of bad pixels on the data (subfigure c)). Subfigure c) for figures
815 4.16-4.18 shows that removal of the beamstop holder lowers the low Q standard de-
816 viation from around .1 to almost .01 out to 15 \AA^{-1} . The high Q subfigures d) and f)
817 in figures 4.16-4.18 show the “kink” effect of the detector edge and beamstop holder,
818 where there is a dip in the $I(Q)$ scattering when the rings include the edge of the
819 detector. This effect seems to be due to both errors in the edge pixel intensity and the
820 beamstop holder as masking of the edges only seems to provide only partial removal
821 of the issue. It is important to note that while integration using the mean of the
822 ring has issues with only the edge mask, as evidenced by the change in slope in 4.17
823 d) around 29.5 \AA^{-1} , the median integration does not include this error. Ideally the
824 detector would have a normal distribution of pixel intensity for a given ring, which
825 would imply an equivalency between the mean and median $I(Q)$ values. Despite the
826 closeness of the mean and median once the final mask has been created, it seems that
827 the median is more reliable, as it was less effected by the beamstop holder in figure
828 4.17. Thus, for subsequent integrations discussed in this work the median is used to
829 avoid any defective features that the masking algorithm may have missed.

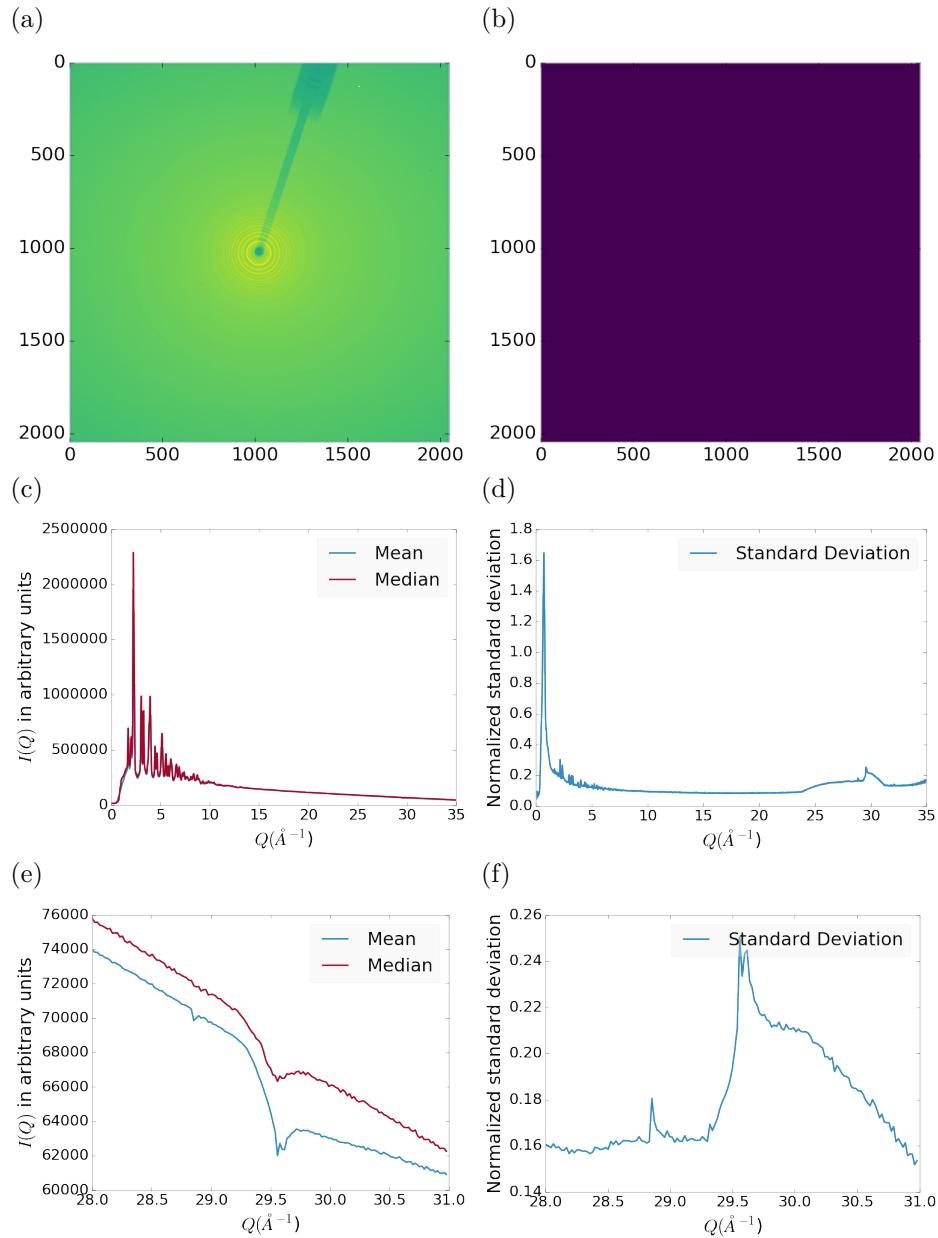


Figure 4.16: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with no mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the 28 \AA^{-1} to 31 \AA^{-1}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

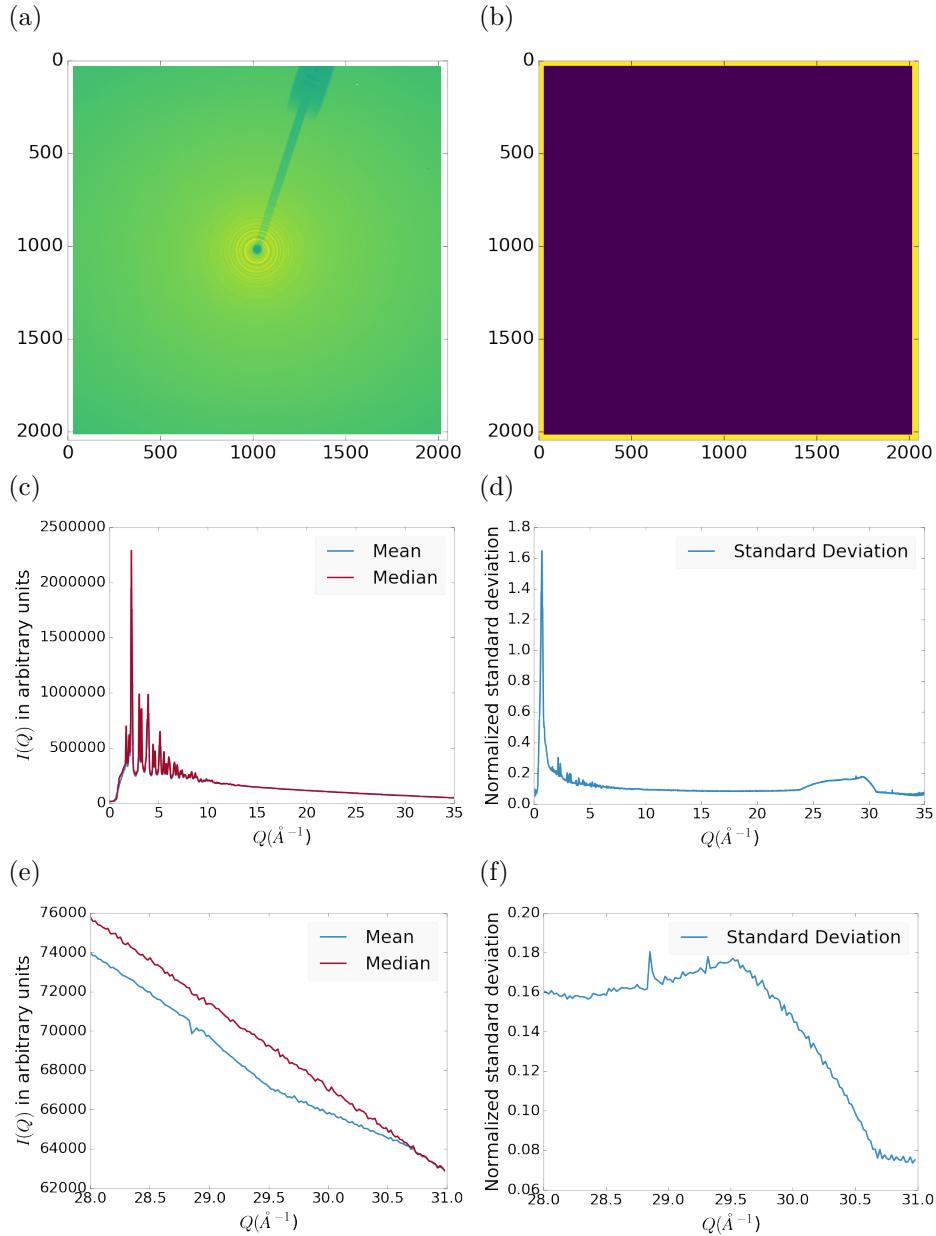


Figure 4.17: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with only an edge mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the 28 \AA^{-1} to 31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

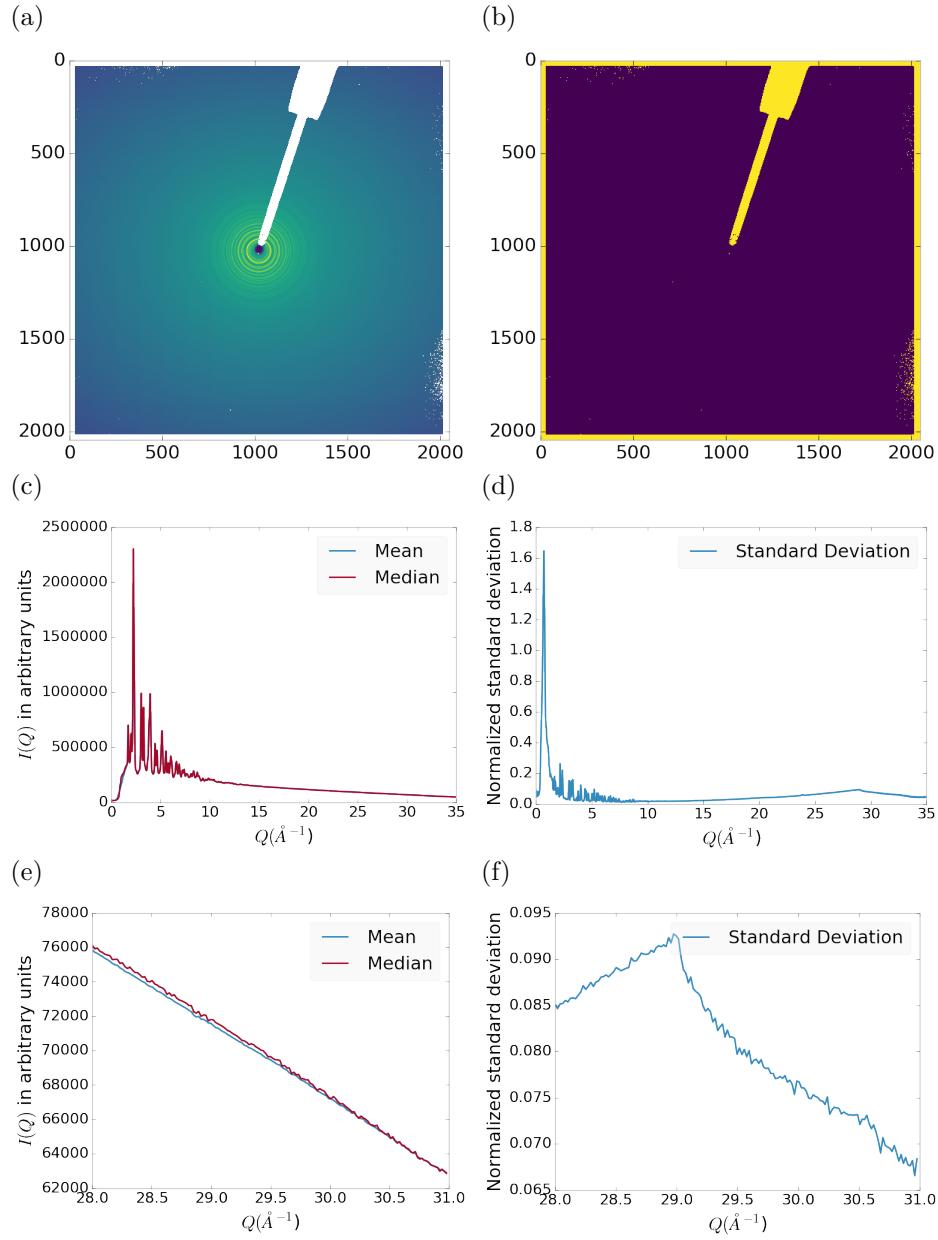


Figure 4.18: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced combining an edge mask and the automatically generated mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median),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

830 4.5 CONCLUSIONS

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

848

CHAPTER 5

849

PHASE CHANGES AND ANNEALING DYNAMICS OF

850

Pr_2NiO_4 AND ITS DERIVATIVES

851 5.1 INTRODUCTION

852

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

853 5.2 EXPERIMENTS

854 **Pr_2NiO_4 Synthesis**

855

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

856 **X-ray Measurements**

857 X-ray total scattering and x-ray powder diffraction experiments were performed at
858 the APS's 11-ID-B beamline. An x-ray energy of 86.7 keV, .145 Å was provided
859 by the beamline monochromator. The detector was moved between a 20cm and a
860 95 cm sample to detector distance to measure the x-ray total scattering and x-ray
861 diffraction patterns. Various PNO samples were annealed on the beamline during
862 x-ray measurement.

863 5.3 DATA PROCESSING

864 The data was calibrated at each of the detector positions using a CeO₂ standard
865 via pyFAI. [28] The images were corrected for a .95 x-ray polarization. Masks were
866 produced for both the foreground and background images. The foreground masks were
867 produced using both a 30 pixel edge mask and a 2.5σ automatic mask as discussed
868 in chapter 4. The background masks were produced by using the foreground mask as
869 a starting mask with a 2.5σ automatic mask.

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

874 Each PDF was generated with a Q_{min} of 1.5, Q_{max} of 29., R_{poly} of .9, R_{max} of 40.
875 descriptions of these parameters can be found in the work by Juhas et. al. [26]

876 5.4 DATA ANALYSIS

877 **Intra Sample Comparison**

878 **PDF**

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

884 The annealed samples figures, 5.3 and 5.4, tell a rather different story. In this case
885 the PDF shows significant peak shifts and broadening, especially at higher interatomic
886 distances. Some peaks completely disappear, like the peak at 12 Å. Similar results were

887 also observed for samples with longer annealing times, as shown in the appendix.

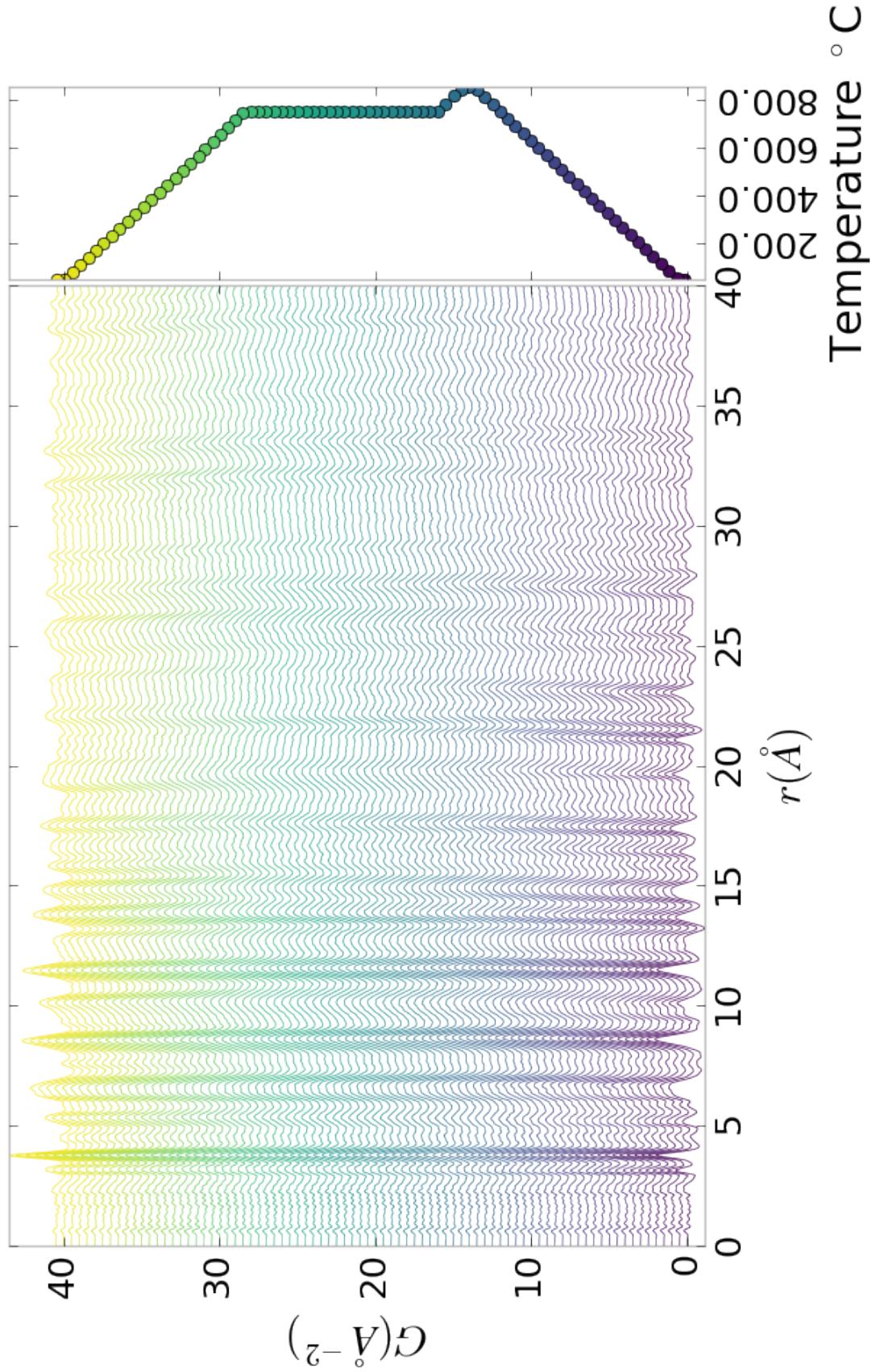


Figure 5.1: PDF as a function of temperature for as synthesized PNO showing the full PDF

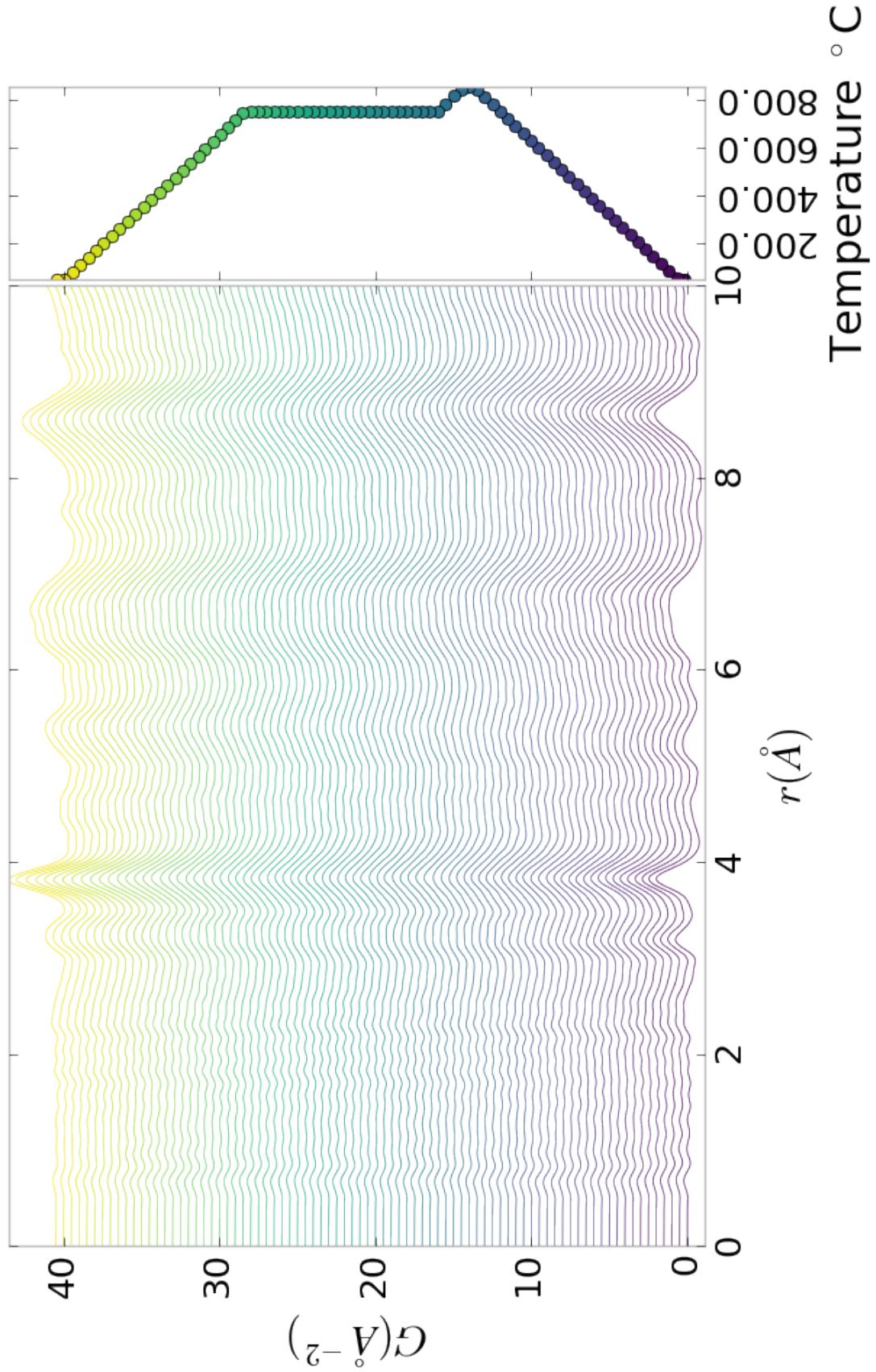


Figure 5.2: PDF as a function of temperature for as synthesized PNO showing a close up on the short range section

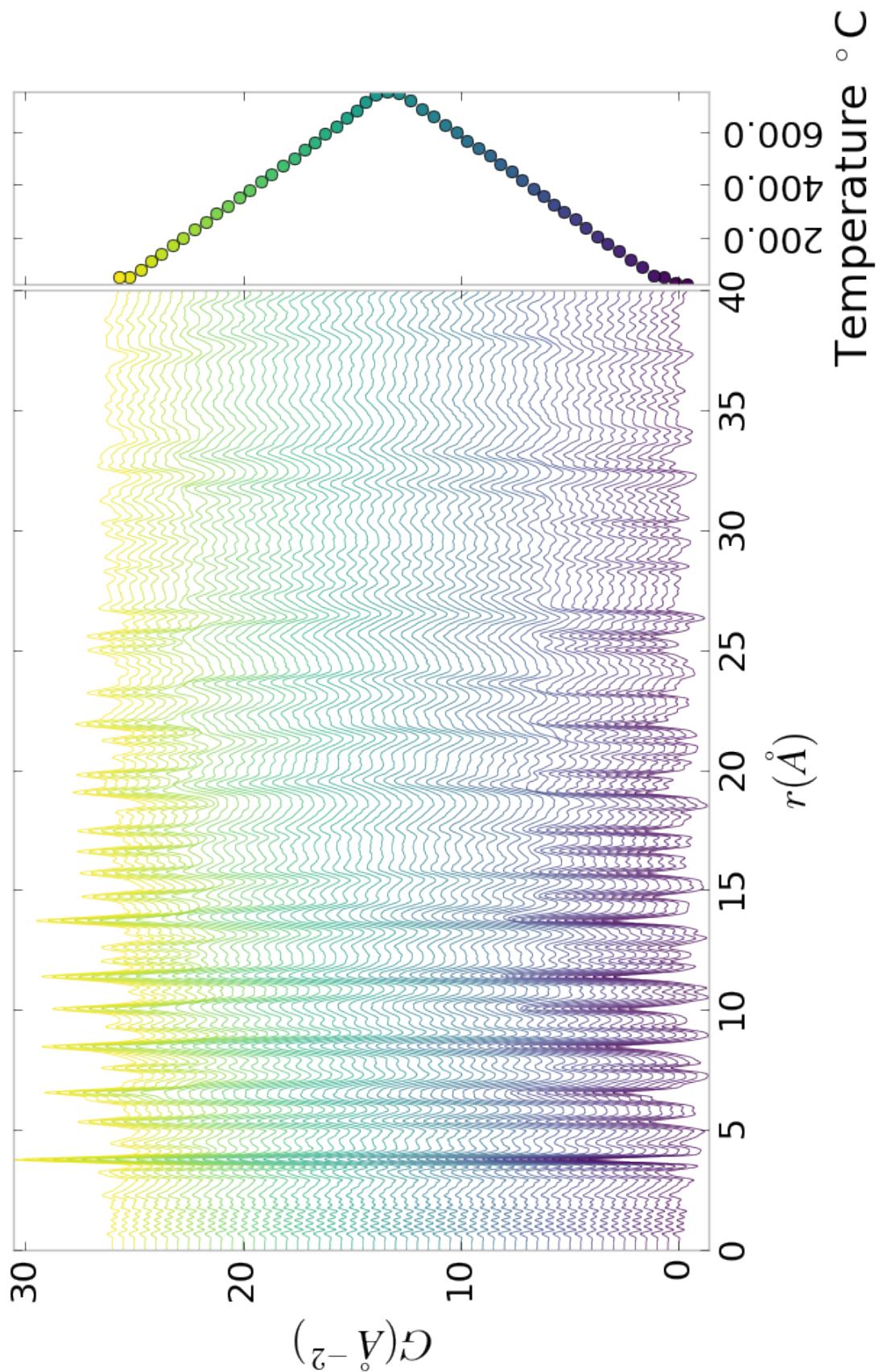


Figure 5.3: PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full PDF

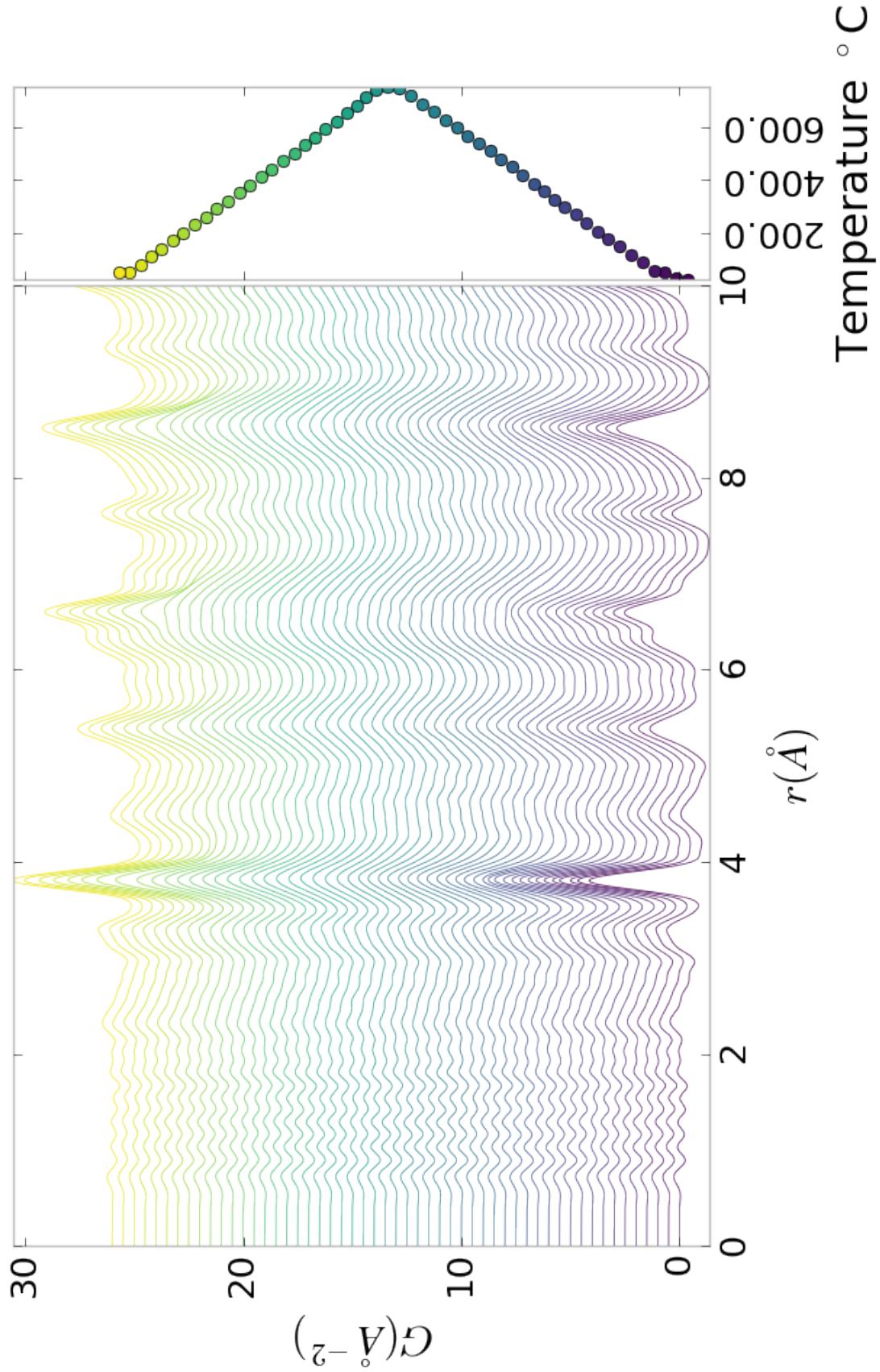


Figure 5.4: PDF as a function of temperature for PNO annealed at 750 $^{\circ}\text{C}$ for 25 hours showing a close up on the short range section

$$_{888} \quad I(Q)$$

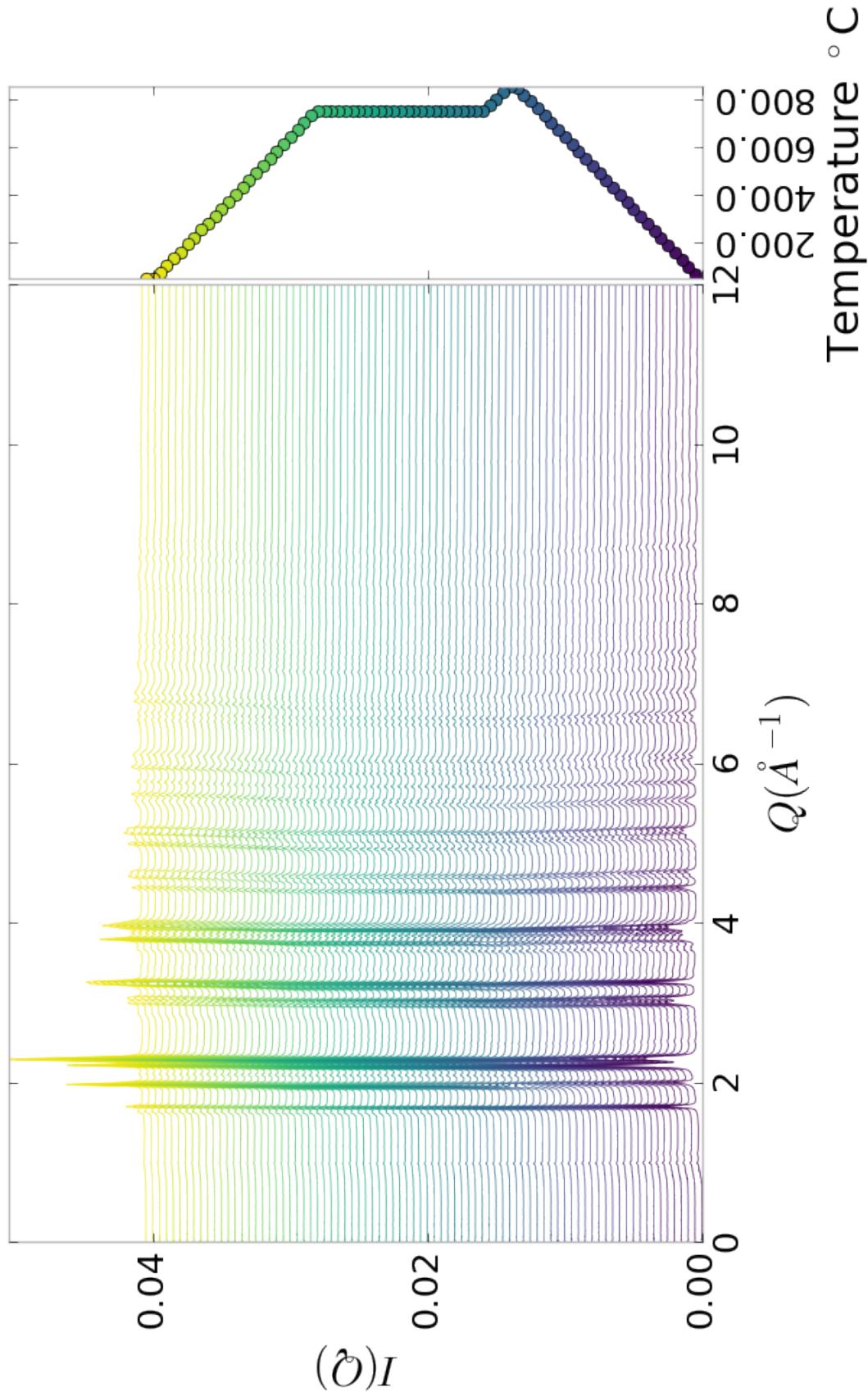


Figure 5.5: $I(Q)$ as a function of temperature for as synthesized PNO showing the full XRD

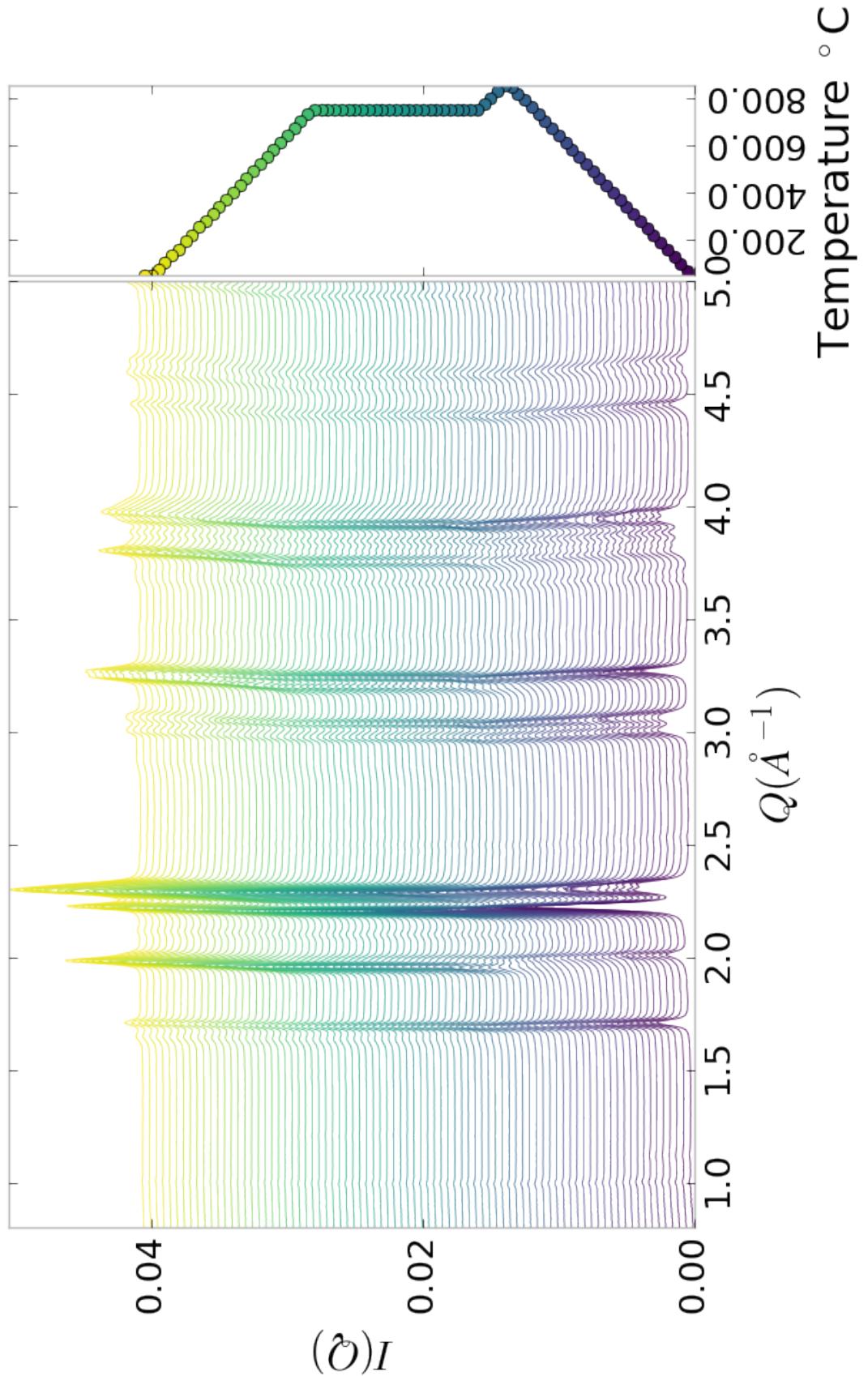
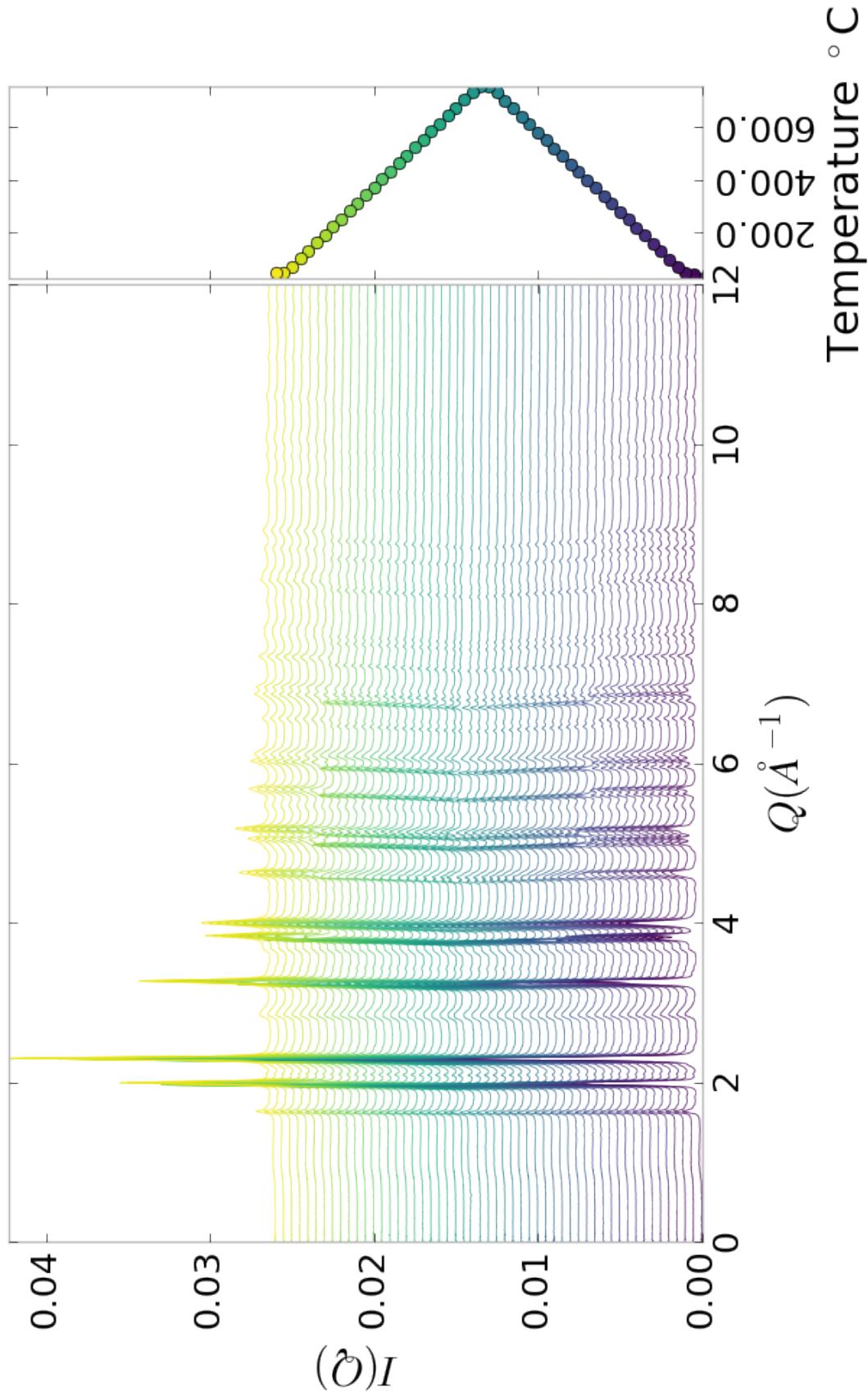


Figure 5.6: $I(Q)$ as a function of temperature for as synthesized PNO showing a close up on the low Q section



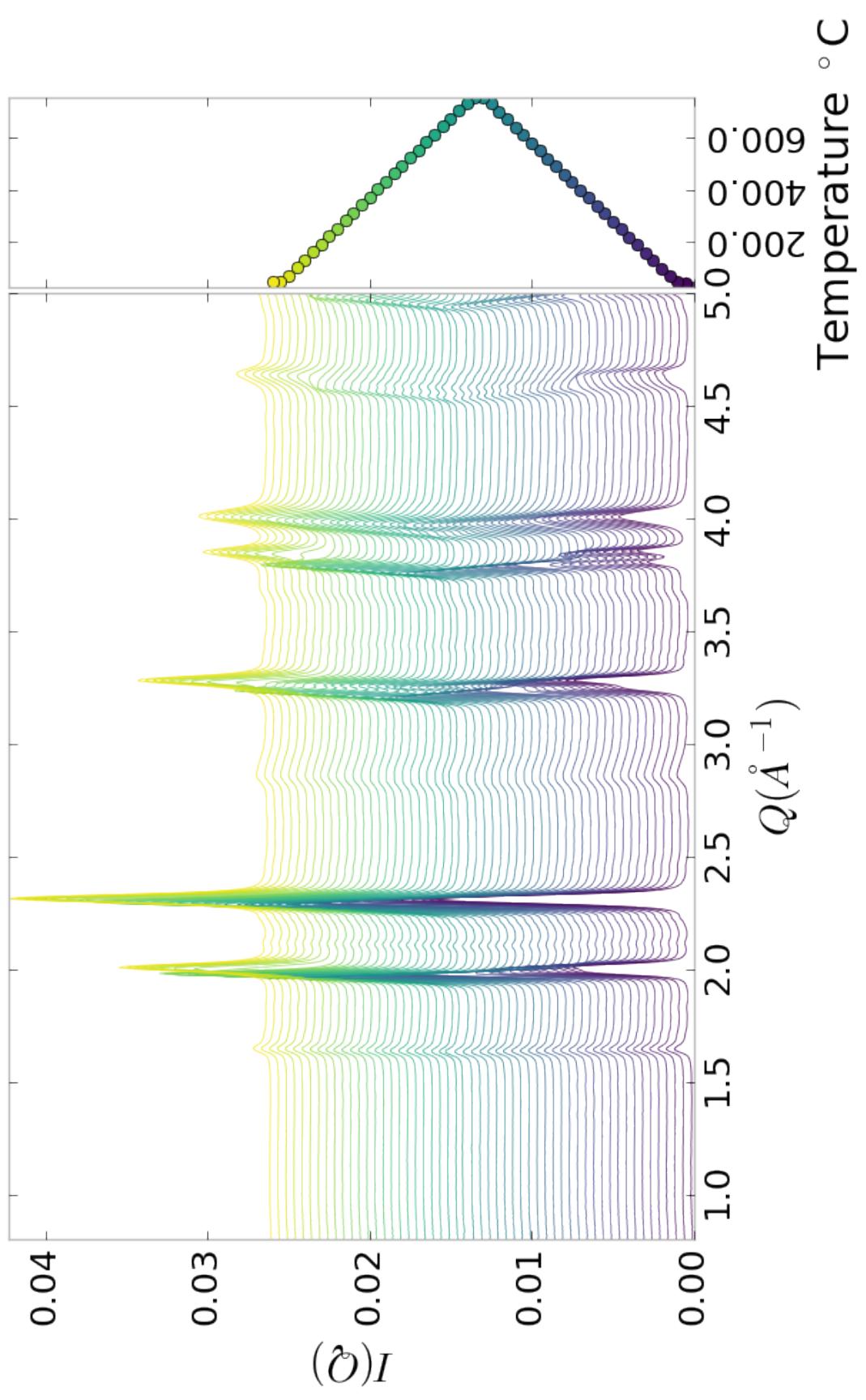


Figure 5.8: $I(Q)$ as a function of temperature for PNO annealed at $750\text{ }^{\circ}\text{C}$ for 25 hours showing a close up on the low Q section

889 **Inter Sample Comparison**

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

891 **5.5 SIMULATION**

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

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

904 **5.6 CONCLUSIONS**

905 X-ray total scattering and x-ray powder diffraction data was obtained on Pr_2NiO_4
906 powder samples annealed for various lengths of time. In-situ studies on the beamline
907 were performed to understand how the structure of each of these powders changes
908 at operating temperatures. The data was processed with the previously discussed Q
909 binning, masking, and integration methodology. The PDF results show very little
910 change in the structure for the as synthesized sample. However, the PDFs show a
911 large change in the previously annealed samples. These changes seem to reporduce

912 PDFs similar to the as-synthesized PNO at operating temperatures. This would seem
913 to imply that the source of the anomalous PNO phase/power density relationship may
914 be due to the adoption of an active structure upon heating which is universal despite
915 the amount of thermal degradation observed at room temperature. In contrast to the
916 PDF results, the XRD results seem to show significant changes in the PNO structure,
917 both with ex-situ and in-situ annealing. The XRDs show the degradation of the PNO
918 into various phases, potentially including Pr_2O_{11} , and higher ordered Pr based phases.
919 The discrepancy between these two results is quite interesting as it seems that the
920 XRD and PDF results are contradictory. Turbostratic displacements between the
921 layers may be a cause of the PDF/XRD disagreement, as these changes would cause
922 very little change in the local structure observed in the PDF, while causing large
923 changes in the XRD.

CONCLUSIONS

925

BIBLIOGRAPHY

- 926 [1] Alexandre A. Arnold, Victor Terskikh, Qian Ying Li, Rafik Naccache, Isabelle
927 Marcotte, and John A. Capobianco, *Structure of NaYF₄ upconverting nanoparticles: A multinuclear solid-state NMR and DFT computational study*, Journal
928 of Physical Chemistry C **117** (2013), no. 48, 25733–25741.
- 930 [2] Simon J L Billinge and Takeshi Egami, *Underneath the Bragg Peaks: Structural
931 Analysis of complex Materials*, vol. Volume 16, Pergamon, 2012.
- 932 [3] Simon J L Billinge and Igor Levin, *The problem with determining atomic struc-
933 ture at the nanoscale.*, Science (New York, N.Y.) **316** (2007), no. 5824, 561–565.
- 934 [4] P. E. Blöchl, *Projector augmented-wave method*, Physical Review B **50** (1994),
935 no. 24, 17953–17979.
- 936 [5] Gary K. Chen and Yunfei Guo, *Discovering epistasis in large scale genetic as-
937 sociation studies by exploiting graphics cards*, Frontiers in Genetics **4** (2013),
938 no. DEC, 1–12.
- 939 [6] Joshua J Choi, Xiaohao Yang, Zachariah M Norman, Simon J L Billinge, and
940 Jonathan S Owen, *Structure of methylammonium lead iodide within mesoporous
941 titanium dioxide: Active material in high-performance perovskite solar cells*,
942 Nano Letters **14** (2014), no. 1, 127–133.
- 943 [7] Peter J. Chupas, Karena W. Chapman, Charles Kurtz, Jonathan C. Hanson,
944 Peter L. Lee, and Clare P. Grey, *A versatile sample-environment cell for non-
945 ambient X-ray scattering experiments*, Journal of Applied Crystallography **41**
946 (2008), no. 4, 822–824.
- 947 [8] PJ Chupas, X Qiu, JC Hanson, PL Lee, CP Grey, and SJL Billinge, *Rapid-
948 acquisition pair distribution function (RA-PDF) analysis*, Journal of Applied
949 Crystallography **36** (2003), 1342–1347.
- 950 [9] Matthew J. Cliffe, Martin T. Dove, D. a. Drabold, and Andrew L. Goodwin,
951 *Structure determination of disordered materials from diffraction data*, Physical
952 Review Letters **104** (2010), no. 12, 1–4.

- 953 [10] Matthew J Cliffe and Andrew L Goodwin, *Nanostructure determination from the*
954 *pair distribution function: a parametric study of the INVERT approach.*, Journal
955 of physics. Condensed matter : an Institute of Physics journal **25** (2013), no. 45,
956 454218.
- 957 [11] Chunhua Cui, Lin Gan, Marc Heggen, Stefan Rudi, and Peter Strasser, *Composi-*
958 *tional segregation in shaped Pt alloy nanoparticles and their structural behaviour*
959 *during electrocatalysis.*, Nature materials **12** (2013), no. 12, 765–771.
- 960 [12] Juarez L F Da Silva, Hyoung Gyu Kim, Maurício J. Piotrowski, Maurício J. Pri-
961 *eto, and Germano Tremiliosi-Filho, Reconstruction of core and surface nanopar-*
962 *ticles: The example of Pt 55 and Au55,* Physical Review B - Condensed Matter
963 and Materials Physics **82** (2010), no. 20, 1–6.
- 964 [13] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth, *Hybrid*
965 *Monte Carlo*, Physics Letters B **195** (1987), no. 2, 216–22.
- 966 [14] Timur Dykhne, Ryan Taylor, Alastair Florence, and Simon J L Billinge, *Data*
967 *requirements for the reliable use of atomic pair distribution functions in amor-*
968 *phous pharmaceutical fingerprinting.*, Pharmaceutical research **28** (2011), no. 5,
969 1041–8.
- 970 [15] C L Farrow, P Juhas, J W Liu, D Bryndin, E S Božin, J Bloch, Th Proffen, and
971 S J L Billinge, *PDFfit2 and PDFgui: computer programs for studying nanostruc-*
972 *ture in crystals.*, Journal of Physics. Condensed Matter : an Institute of Physics
973 journal **19** (2007), no. 33, 335219.
- 974 [16] Christopher L Farrow and Simon J L Billinge, *Relationship between the atomic*
975 *pair distribution function and small-angle scattering: implications for modeling*
976 *of nanoparticles.*, Acta Crystallographica Section A Foundations of Crystallog-
977 *raphy* **65** (2009), no. Pt 3, 232–9 (en).
- 978 [17] Riccardo Ferrando, Julius Jellinek, and Roy L Johnston, *Nanoalloys: From The-*
979 *ory to Applications of Alloy Clusters and Nanoparticles*, Chemical Reviews **108**
980 (2008), no. 3, 846–904.
- 981 [18] Anatoly Frenkel, *Solving the 3D structure of metal nanoparticles*, Zeitschrift für
982 Kristallographie **222** (2007), no. 11, 605–611.
- 983 [19] Carmelo Giacovazzo, Hugo Luis Monaco, Gilberto Artioli, Davide Viterbo,
984 Marco Milanesio, Gastone Gilli, Paola Gilli, Giuseppe Zanotti, Giovanni Fer-

- 985 raris, and Michele Catti, *Fundamentals of Crystallography*, Oxford University
986 Press, 1992.
- 987 [20] Benjamin Gilbert, Jasmine J. Erbs, R. Lee Penn, Valeri Petkov, Dino Spagnoli,
988 and Glenn A. Waychunas, *A disordered nanoparticle model for 6-line ferrihydrite*,
989 American Mineralogist **98** (2013), no. 8-9, 1465–1476.
- 990 [21] Benjamin Gilbert, Feng Huang, Hengzhong Zhang, Glenn A. Waychunas, and
991 Jillian F Banfield, *Nanoparticles: strained and stiff.*, Science (New York, N.Y.)
992 **305** (2004), no. 5684, 651–654.
- 993 [22] Md Matthew D. Hoffman and Andrew Gelman, *The No-U-Turn Sampler: Adap-
994 tively Setting Path Lengths in Hamiltonian Monte Carlo*, The Journal of Machine
995 Learning Research **15** (2014), no. 2008, 1593–1623.
- 996 [23] W J Huang, R Sun, J Tao, L D Menard, R G Nuzzo, and J M Zuo, *Coordination-
997 dependent surface atomic contraction in nanocrystals revealed by coherent diffrac-
998 tion.*, Nature Materials **7** (2008), no. 4, 308–313.
- 999 [24] Pablo D Jadzinsky, Guillermo Calero, Christopher J Ackerson, David A Bushnell,
1000 and Roger D Kornberg, *Structure of a thiol monolayer-protected gold nanoparti-
1001 cle at 1.1 Å resolution.*, Science (New York, N.Y.) **318** (2007), no. 5849, 430–433.
- 1002 [25] I. K. Jeong, R. H. Heffner, M. J. Graf, and S. J. L. Billinge, *Lattice dynamics
1003 and correlated atomic motion from the atomic pair distribution function*, (2002),
1004 9.
- 1005 [26] P. Juhás, T. Davis, C.ÁL. Farrow, and S.ÁJ.ÁL. Billinge, *PDFgetX3 : a
1006 rapid and highly automatable program for processing powder diffraction data into
1007 total scattering pair distribution functions*, Journal of Applied Crystallography
1008 **46** (2013), no. 2, 560–566 (en).
- 1009 [27] A. Kassiba, M. Makowska-Janusik, J. Bouclé, J. Bardeau, A. Bulou, and
1010 N. Herlin-Boime, *Photoluminescence features on the Raman spectra of quasisto-
1011 chiometric SiC nanoparticles: Experimental and numerical simulations*, Physical
1012 Review B **66** (2002), no. 15, 1–7.
- 1013 [28] Jérôme Kieffer and Dimitrios Karkoulis, *PyFAI, a versatile library for azimuthal
1014 regrouping*, Journal of Physics: Conference Series **425** (2013), 202012.
- 1015 [29] G. Kresse and J. Hafner, *Ab initio molecular dynamics for liquid metals*, Physical
1016 Review B **47** (1993), no. 1, 558–561.

- 1017 [30] ———, *Ab initio molecular-dynamics simulation of the liquid-
1018 metalâ€šamorphous-semiconductor transition in germanium*, Physical Review
1019 B **49** (1994), no. 20, 14251–14269.
- 1020 [31] Yan Li, Giulia Galli, and Fran ois Gygi, *Electronic structure of thiolate-covered
1021 gold nanoparticles: Au102(MBA)44*, ACS Nano **2** (2008), no. 9, 1896–1902.
- 1022 [32] L D Marks, *Experimental studies of small particle structures*, Reports on
1023 Progress in Physics **57** (1994), no. 6, 603–649 (en).
- 1024 [33] A. S. Masadeh, E. S. Bo in, C. L. Farrow, G. Paglia, P. Juhas, S. J. L. Billinge,
1025 A. Karkamkar, and M. G. Kanatzidis, *Quantitative size-dependent structure and
1026 strain determination of CdSe nanoparticles using atomic pair distribution func-
1027 tion analysis*, Physical Review B - Condensed Matter and Materials Physics **76**
1028 (2007), no. 11, 115413.
- 1029 [34] R L McGreevy and L Pusztai, *Reverse Monte Carlo Simulation: A New Tech-
1030 nique for the Determination of Disordered Structures*, Molecular Simulation **1**
1031 (1988), no. 6, 359–367.
- 1032 [35] Donald A. McQuarrie, *Statistical Mechanics*, University Science Books, Sausal-
1033 ito, CA, 2000.
- 1034 [36] Radford M Neal, *Probabilistic Inference Using Markov Chain Monte Carlo Meth-
1035 ods*, Intelligence **45** (1993), no. September, 144.
- 1036 [37] Radford M. Neal, *MCMC Using Hamiltonian Dynamics*, Handbook of Markov
1037 Chain Monte Carlo (Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li
1038 Meng, eds.), Chapman and Hall/CRC, 2011, pp. 113–162.
- 1039 [38] Katharine Page, Taylor C. Hood, Thomas Proffen, and Reinhard B. Neder,
1040 *Building and refining complete nanoparticle structures with total scattering data*,
1041 Journal of Applied Crystallography **44** (2011), no. 2, 327–336 (en).
- 1042 [39] B R Pauw, *Corrigendum: Everything SAXS: small-angle scattering pattern col-
1043 lection and correction (2013 J. Phys.: Condens. Matter 25 383201)*, Journal of
1044 Physics: Condensed Matter **26** (2014), no. 23, 239501.
- 1045 [40] Vitalij Pecharsky and Peter Zavalij, *Fundamentals of Powder Diffraction and
1046 Structural Characterization of Materials*, Springer Science, 2009.

- 1047 [41] John P. Perdew, Kieron Burke, and Matthias Ernzerhof, *Generalized Gradient*
1048 *Approximation Made Simple*, Physical Review Letters **77** (1996), no. 18, 3865–
1049 3868.
- 1050 [42] Andrew A. Peterson, *Global optimization of adsorbate-surface structures while*
1051 *preserving molecular identity*, Topics in Catalysis **57** (2014), no. 1-4, 40–53.
- 1052 [43] Valeri Petkov, Binay Prasai, Yang Ren, Shiyao Shan, Jin Luo, Pharrah Joseph,
1053 and Chuan-Jian Zhong, *Solving the nanostructure problem: exemplified on metal-*
1054 *alloy nanoparticles*, Nanoscale **6** (2014), no. 17, 1–11.
- 1055 [44] Valeri Petkov, Shiyao Shan, Peter Chupas, Jun Yin, Lefu Yang, Jin Luo, and
1056 Chuan-Jian Zhong, *Noble-transition metal nanoparticle breathing in a reactive*
1057 *gas atmosphere.*, Nanoscale **5** (2013), no. 16, 7379–87.
- 1058 [45] Th. Proffen and R. B. Neder, *DISCUS: a Program for Diffuse Scattering and*
1059 *Defect-Structure Simulation*, Journal of Applied Crystallography **30** (1997), 171–
1060 175.
- 1061 [46] Erin L. Redmond, Brian P. Setzler, Pavol Juhas, Simon J. L. Billinge, and
1062 Thomas F. Fuller, *In-Situ Monitoring of Particle Growth at PEMFC Cathode*
1063 *under Accelerated Cycling Conditions*, Electrochemical and Solid-State Letters
1064 **15** (2012), no. 5, B72 (en).
- 1065 [47] H. W. Sheng, M. J. Kramer, A. Cadien, T. Fujita, and M. W. Chen, *Highly*
1066 *optimized embedded-atom-method potentials for fourteen FCC metals*, Physical
1067 Review B - Condensed Matter and Materials Physics **83** (2011), no. 13, 134118.
- 1068 [48] Randall Q. Snurr, Alexis T. Bell, and Doros N. Theodorou, *Prediction of ad-*
1069 *sorption of aromatic hydrocarbons in silicalite from grand canonical Monte Carlo*
1070 *simulations with biased insertions*, The Journal of Physical Chemistry **97** (1993),
1071 no. 51, 13742–13752 (EN).
- 1072 [49] Chang Q. Sun, *Size dependence of nanostructures: Impact of bond order defi-*
1073 *ciency*, Progress in Solid State Chemistry **35** (2007), no. 1, 1–159.
- 1074 [50] X. Yang, P. Juhás, and S. J L Billinge, *On the estimation of statistical uncertain-*
1075 *ties on powder diffraction and small-angle scattering data from two-dimensional*
1076 *X-ray detectors*, Journal of Applied Crystallography **47** (2014), no. 4, 1273–1283.

- 1077 [51] Hengzhong Zhang, Bin Chen, Jillian F. Banfield, and Glenn A. Waychunas,
1078 *Atomic structure of nanometer-sized amorphous TiO₂*, Physical Review B **78**
1079 (2008), no. 21, 214106.

1080

APPENDIX A

1081

SUPPLEMENTAL INFORMATION: PHASE CHANGES AND

1082

ANNEALING DYNAMICS OF Pr_2NiO_4 AND ITS

1083

DERIVATIVES

1084

Intra Sample Comparison

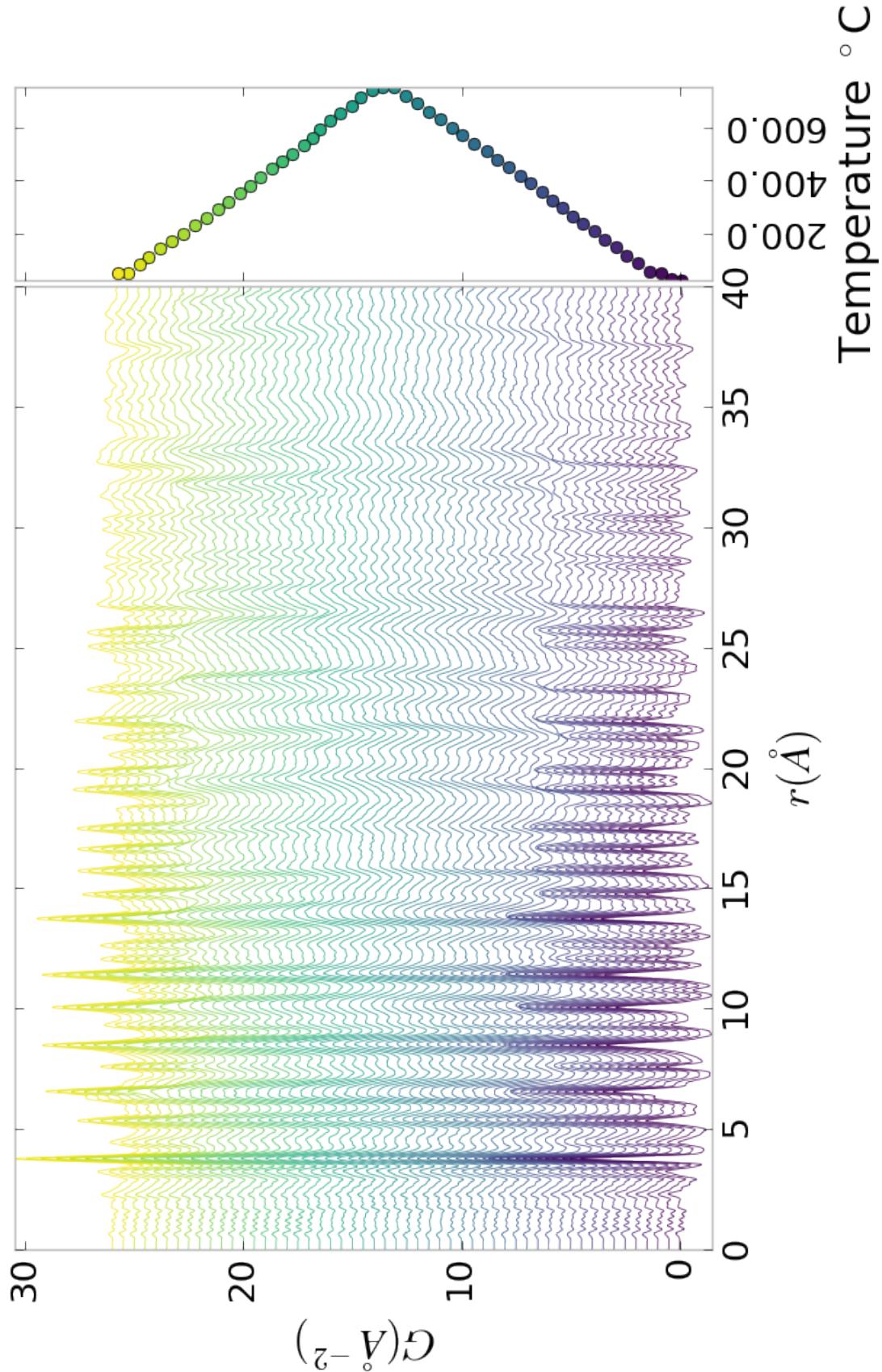


Figure A.1: PDF as a function of temperature for PNO annealed at 750 °C for 50 hours showing the full PDF

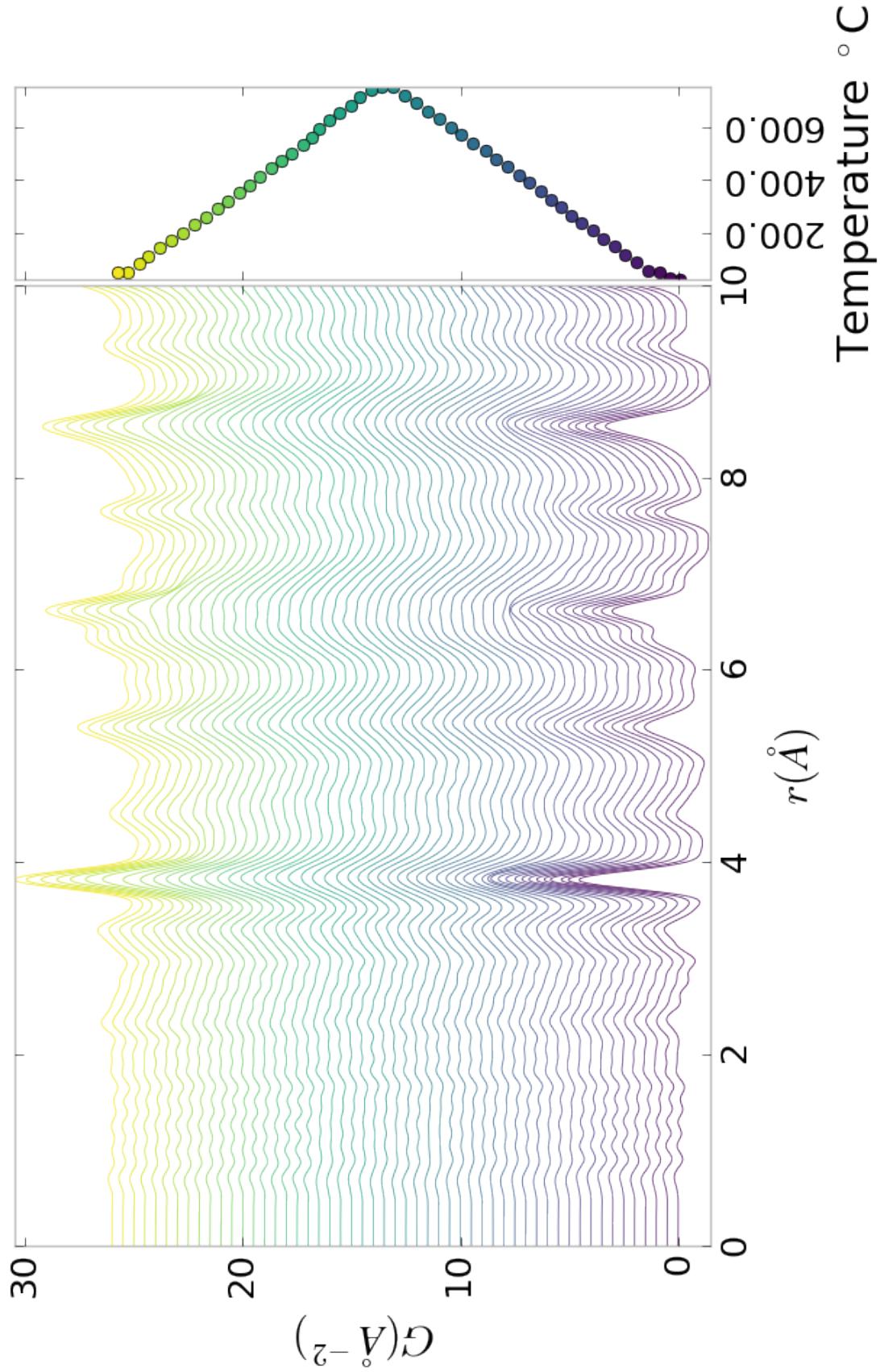


Figure A.2: PDF as a function of temperature for PNO annealed at 750 $^{\circ}\text{C}$ for 50 hours showing a close up on the short range section

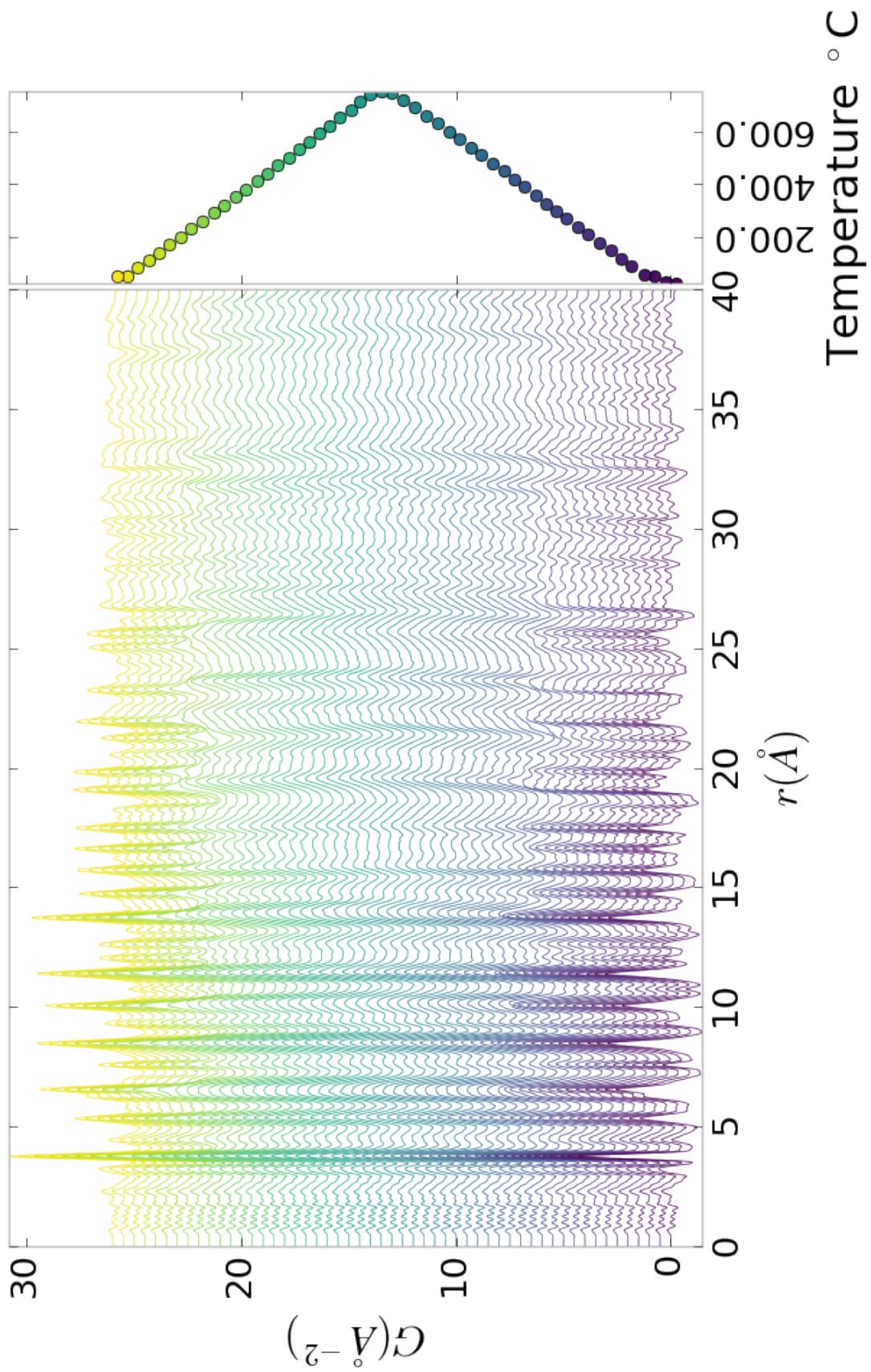


Figure A.3: PDF as a function of temperature for PNO annealed at 750 $^{\circ}\text{C}$ for 100 hours showing the full PDF

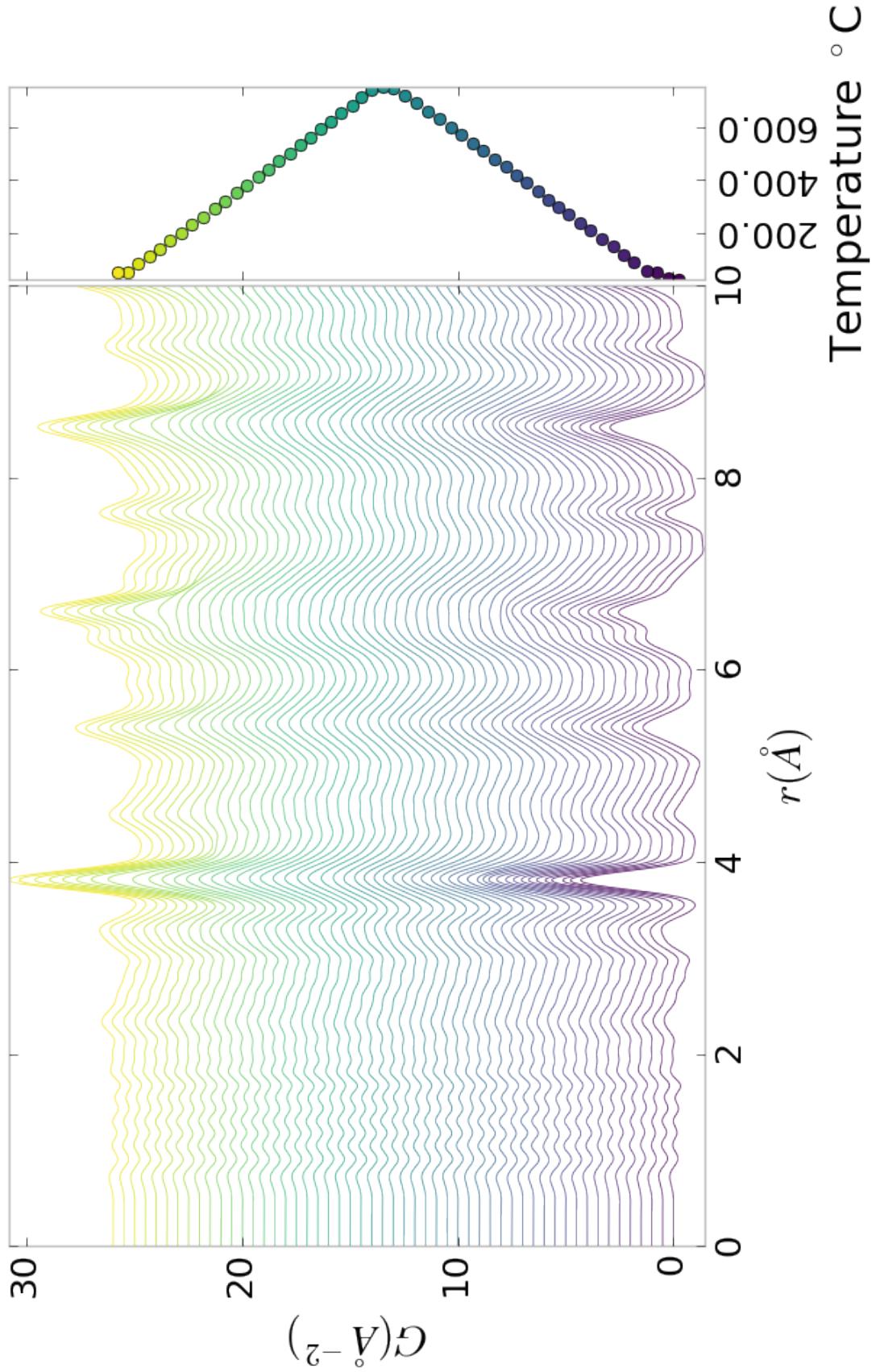


Figure A.4: PDF as a function of temperature for PNO annealed at 750 $^\circ\text{C}$ for 100 hours showing a close up on the short range section

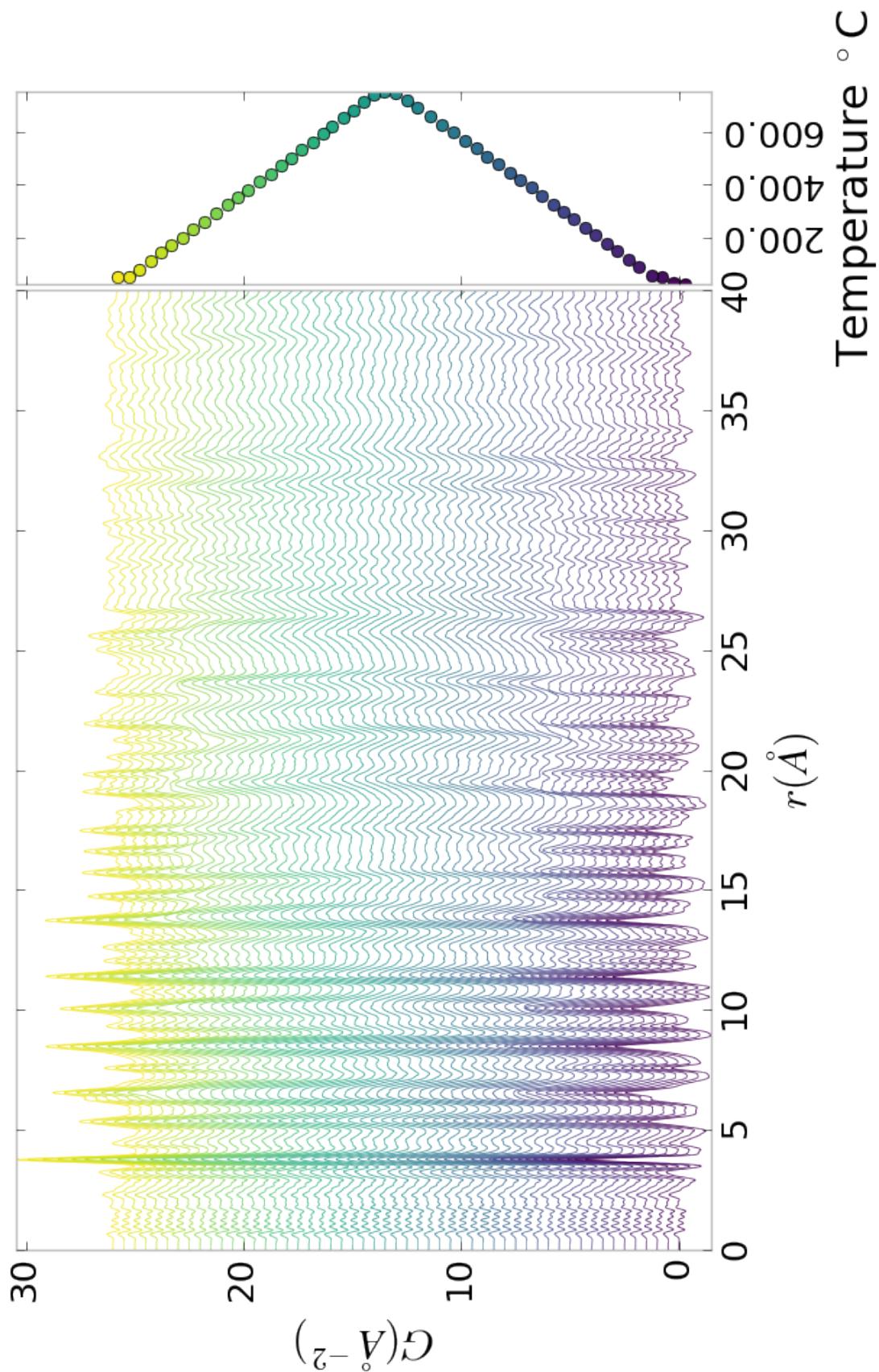


Figure A.5: PDF as a function of temperature for PNO annealed at 750 $^{\circ}\text{C}$ for 200 hours showing the full PDF

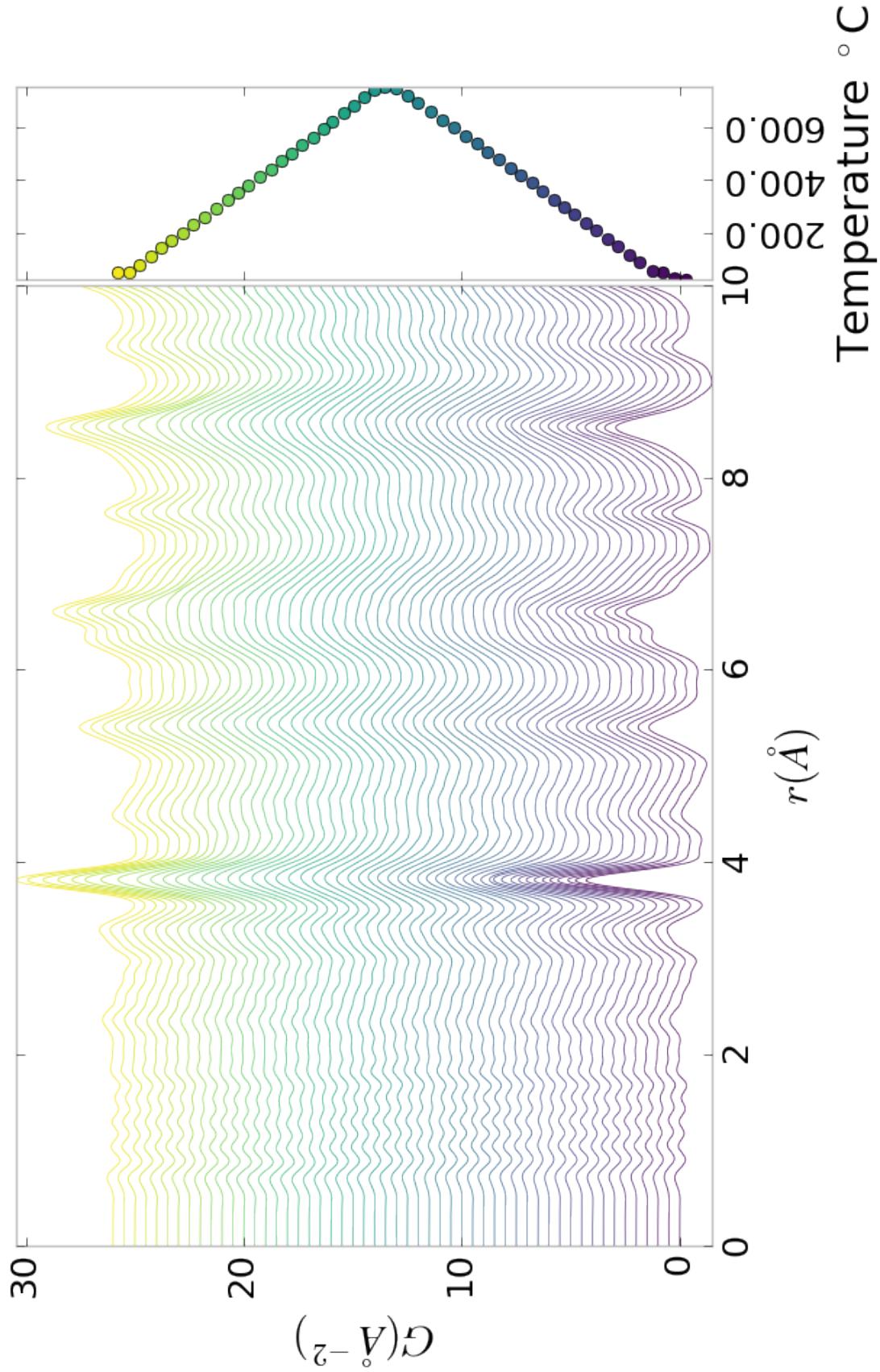


Figure A.6: PDF as a function of temperature for PNO annealed at $750\text{ }^{\circ}\text{C}$ for 200 hours showing a close up on the short range section

1085

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