

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

by

Christopher James Wright

Bachelor of Science
Brown University 2014

Submitted in Partial Fulfillment of the Requirements

for the Degree of Masters of Science in

Chemical Engineering

College of Engineering and Computing

University of South Carolina

2016

Accepted by:

Xiao-Dong Zhou, Major Professor

Mark Uline, Committee Member

Jochen Lauterbach, Committee Member

Thomas Vogt, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

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

DEDICATION

For Diane & Donald Wright

My first scientific advisers

ACKNOWLEDGMENTS

ABSTRACT

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

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

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

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGMENTS	iv
ABSTRACT	v
LIST OF TABLES	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
Grand Canonical Ensemble	11
Ensemble description	11
Grand Canonical Monte Carlo	12
GCMC biasing	12

CHAPTER 2 ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION	14
2.1 Theory	14
Derivation	14
Analytically Gradients	14
Without ADPs	16
2.2 Computation	16
HPC and GPUs	17
GPUs and Parallelization	17
Map from ij space to k space	18
GPU Memory Allocation	20
Speed and Scaling of PDF Computation	22
CHAPTER 3 BENCHMARKING	24
3.1 Introduction	24
3.2 PDF	24
Model Parameters	26
Au55: surface relaxed	27
Run Parameters	27
Au55: surface disordered	29
Au55: amorphous	30
Au102: triple phase	31
Starting from fcc structure	31
Marks decahedron	32
Au147	33
3.3 PDF with ADPs	33
ADP 50	33
CHAPTER 4 X-RAY TOTAL SCATTERING DATA ACQUISITION AND PRO- CESSING	39
4.1 Introduction	39
4.2 Detector Q resolution	39
4.3 Automated Mask Generation	41
Introduction	41

Algorithm Design	43
Test Cases	43
Results and Discussion	44
Conclusions	50
4.4 Automated Image Azimuthal Integration	51
4.5 Conclusions	53
CHAPTER 5 PHASE CHANGES AND ANNEALING DYNAMICS OF Pr_2NiO_4 AND ITS DERIVATIVES	56
5.1 Experiments	56
Pr_2NiO_4 Synthesis	56
X-ray Measurements	56
5.2 Data Processing	56
5.3 Data Analysis	57
Intra Sample Comparison	57
Inter Sample Comparison	67
5.4 Simulation	80
5.5 Conclusions	80
CHAPTER 6 CONCLUSION	82
BIBLIOGRAPHY	83

LIST OF TABLES

LIST OF FIGURES

Figure 1.1	2D slice of addition biasing with Rw and Lennard Jones potentials.	13
Figure 2.1	Comparison of the CPU and GPU chip architectures	17
Figure 2.2	Speed comparison of CPU and GPU implementations	22
Figure 3.1	Au ₅₅ PDF fitting of DFT-optimized <i>c</i> -Au ₅₅ .	28
Figure 3.2	Au ₅₅ PDF fitting of surface-disordered Au ₅₅ .	34
Figure 3.3	Similar to figure 3.2 for DFT-optimized amorphous Au ₅₅ .	35
Figure 3.4	Similar to figure 3.2 for Au ₁₀₂ as in DFT-optimized Au ₁₀₂ MBA ₄₄ cluster.	36
Figure 3.5	Similar to Fig. 3.4 with Marks decahedron as the starting structure.	37
Figure 3.6	Refinement of adps	38
Figure 4.1	Scattering onto a flat detector	40
Figure 4.2	Q resolution as a function of Q .	41
Figure 4.3	Number of pixels as a function of Q , binned at the Q resolution of the detector.	42
Figure 4.4	Generated dead/hot pixel masks for a detector with 100 bad pixels.	44
Figure 4.5	Generated dead/hot pixel masks for a detector with 300 bad pixels.	45
Figure 4.6	Generated dead/hot pixel masks for a detector with 500 bad pixels.	45
Figure 4.7	Generated dead/hot pixel masks for a detector with 1000 bad pixels.	46

Figure 4.8 Generated beamstop holder masks for a beamstop holder with 10% transmittance.	46
Figure 4.9 Generated beamstop holder masks for a beamstop holder with 30% transmittance.	47
Figure 4.10 Generated beamstop holder masks for a beamstop holder with 50% transmittance.	47
Figure 4.11 Generated beamstop holder masks for a beamstop holder with 90% transmittance.	47
Figure 4.12 Generated beamstop holder masks which is rotated away from vertical	48
Figure 4.13 Masked experimental data.	48
Figure 4.14 Masked experimental data with Pt single crystal signal.	49
Figure 4.15 Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask.	49
Figure 4.16 Masking, average, and standard deviation of an example x-ray total scattering measurement with with no mask.	52
Figure 4.17 Masking, average, and standard deviation of an example x-ray total scattering measurement with with only an edge mask.	53
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.	54
Figure 5.1 PDF as a function of temperature for as synthesized PNO showing the full PDF	58
Figure 5.2 PDF as a function of temperature for as synthesized PNO showing a close up on the short range section	59
Figure 5.3 PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full PDF	60
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	61

Figure 5.5 XRD as a function of temperature for S1	62
Figure 5.6 XRD as a function of temperature for S2	63
Figure 5.7 XRD as a function of temperature for S3	64
Figure 5.8 XRD as a function of temperature for S4	65
Figure 5.9 XRD as a function of temperature for S5	66

TODO LIST

2	Need more references	1
3	Why is atomistic engineering important	2
4	Barriers to atomistic engineering	2
5	How are we going to attack this problem	2
6	double check against McQuerry, cite it, also move from probabilities to par-	
7	tition functions, develop partition functions above	8
8	Talk about the advances NUTS gives us	11
9	This is wrong and needs a redo	12
10	Include figure which shows the configurational biasing map	13
11	This entire section needs some rewriting to distinguish this from the paper .	24
12	Also some introduction would be great	24
13	this just needs a lot of work	24
14	Put this somewhere	28
15	need some sort of synthesis information	56
16	masking parameters	56
17	integration parameters	56
18	PDF parameters	56
19	Need to redo all the figures:	
20	need to use the new stylesheet	
21	need to be more readable	
22	need to cull through them to pull out the interesting features	57
23	Need more references	

24

INTRODUCTION

25

Why is atomistic engineering important

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

30

Barriers to atomistic engineering

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

34

How are we going to attack this problem

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

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

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

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

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

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

54

CHAPTER 1

55

STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES

57 1.1 INTRODUCTION

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

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

70 1.2 POTENTIAL ENERGY SURFACES

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

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

82 Experimentally Derived Potential Energy Surfaces

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

91 Potentials

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

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

94

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

95

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

96

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

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

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

118 **Forces**

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

119

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

120

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

121

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

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

127 **1.3 ENSEMBLES**

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

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

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

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

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

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

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

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

145

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

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

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

149 However, for the canonical ensemble the probability is:

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

150 Monte Carlo Modeling

151 Monte Carlo can be used to simulate a statistical mechanical ensemble which can
152 not be solved analytically. In most Monte Carlo systems the ensemble is simulated by
153 randomly changing one of the system parameters and comparing the energy of the
154 new system against the energy of the old system. If the energy of the new system is
155 lower than the current energy then the new configuration is accepted. Otherwise the
156 new system is rejected unless

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

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

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

168 Hamiltonian Monte Carlo

169 Hybrid or Hamiltonian Monte Carlo (HMC) can help to address some of these issues.
170 HMC was developed originally in the lattice quantum chromodynamic community
171 and provides a more efficient, more scalable approach to PES sampling for Monte
172 Carlo.[10, 27] In HMC the PES is explored using Hamiltonian dynamics, essentially
173 following the gradient of the PES to find more acceptable configurations.

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

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

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

Using these equations we can derive the position and momentum vectors at any point

in time using the leap-frog algorithm:

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

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

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

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

175 moved. Using this notation the gradient is

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

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

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

194 Talk about the advances NUTS gives us

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

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

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:

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

217 Similarly the removal of an atom has the probability:

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

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

234 **GCMC biasing**

235 This is wrong and needs a redo

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

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

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

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

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

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

253 Include figure which shows the configurational biasing map

254

CHAPTER 2

255

ATOMIC PAIR DISTRIBUTION FUNCTION:

256

THEORY AND COMPUTATION

257 2.1 THEORY

258 To properly understand the PDF and its limitations we need to derive its mathemat-
259 ics. The PDF has been previously derived many times so it is not re-derived here.
260 This discussion of the PDF and its gradients use the notation of Farrow and Billinge.
261 [13]

262 **Derivation**

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

269 **Analytically Gradients**

270 Many optimization algorithms and simulations methodologies, including HMC, re-
271 quire not only the potential energy of a given configuration but also the forces acting
272 on that configuration. These forces are described by the gradient of potential energy

273 of the system which in turn requires the gradient of the PDF. As previously shown the
 274 PDF is the Fourier Transform of the Debye equation. Since the Fourier Transform is
 275 expressed as an integral we can exchange the order of the gradient and the integral,
 276 allowing us to calculate the analytical gradient of the Debye equation and FFT the
 277 resulting function. The Debye equation, with a Debye-Waller vibrational correction
 278 is

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

279 where

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

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

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

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

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

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

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

283 where

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

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

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

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

284 The derivatives are as follows:

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

285 where

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

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

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

287 Thus

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

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

288 Without ADPs

289 Without ADPs the equations simplify down to

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

290 and

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

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

293 2.2 COMPUTATION

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

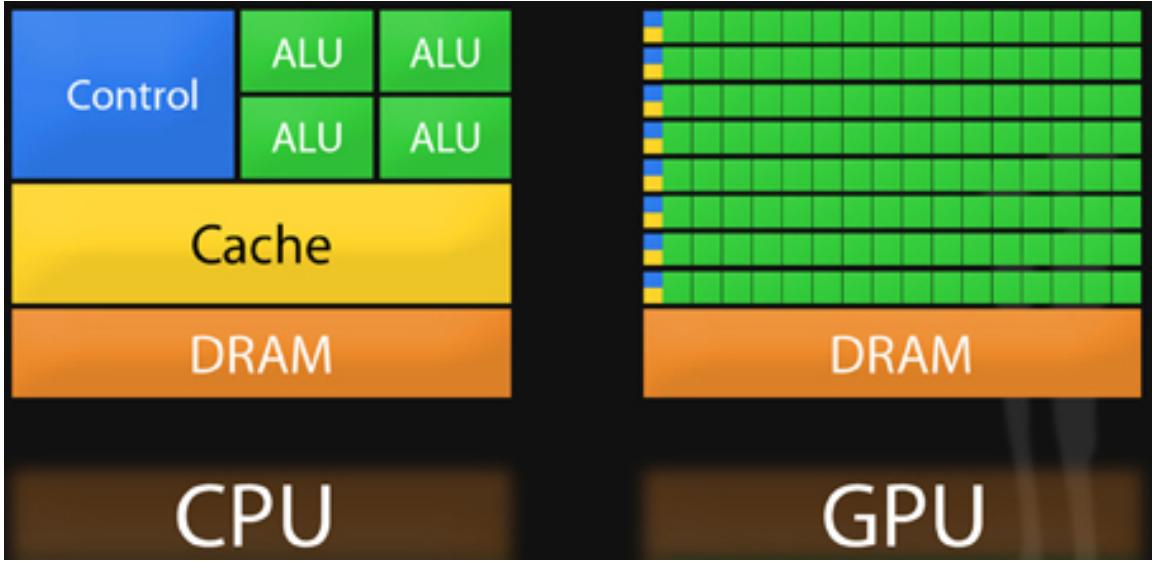


Figure 2.1: Comparison of the CPU and GPU chip architectures from [3]. The ALU are the arithmetic logic units which perform the mathematical operations, the DRAM holds most of the data, although it is slower to access, the Cache holds rapidly accessed data, and the Control controls the execution of software.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

298 HPC and GPUs

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

303 GPUs and Parallelization

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

310 existing parallel FFT algorithms.

311 GPUs are particularly well suited to the task of computing PDFs. GPU chip
312 architecture is designed to perform many tasks simultaneously by having potentially
313 thousands of cores. Figure 2.1 shows the comparison of CPU and GPU architectures.
314 As the figure shows the GPUs have a very different layout of computational proces-
315 sors (ALUs) and memory. While each ALU is simpler on the GPU, requiring the
316 instructions to be less demanding in terms of memory, there are many more of them.
317 The greater number of processors allows each atomic pairing to be placed on its own
318 processor, so long as the math can be broken into simpler operations. The equations
319 are broken up on the GPUs into various pieces which correspond to the α , β , τ and Ω
320 as shown in equations 2.8-2.11 and subequations as needed. For example, while β is
321 computed in one step, Ω requires the computation of the displacement array, then
322 the distance array and finally the Ω array. The exact breakdown of processes, how
323 the problems are broken down and spread across the processor has been optimized
324 for speed and reliability.

325 Map from ij space to k space

326 The above equations, although formally correct, are very inefficient. $F(Q)$ and its
327 gradient are indexed over all the atoms twice, however there are symmetries that
328 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

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

329

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

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

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

For the gradient a similar mapping is used:

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

345

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

351 GPU Memory Allocation

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

$$q[=]3n \quad (2.19)$$

$$d[=]3k \quad (2.20)$$

$$r[=]k \quad (2.21)$$

$$scatter[=]nQ \quad (2.22)$$

$$normalization[=]kQ \quad (2.23)$$

$$\Omega[=]kQ \quad (2.24)$$

$$F_k(Q)[=]kQ \quad (2.25)$$

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

$$Sum2[=]kQ \quad (2.27)$$

$$F(Q)[=]Q \quad (2.28)$$

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

$$k_{perGPU} = \frac{1}{16Q + 16} (-4Qn - 4Q + am - 12n) \quad (2.29)$$

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

$$adps = 3n \quad (2.30)$$

$$\sigma = k \quad (2.31)$$

$$\tau = kQ \quad (2.32)$$

366 Thus the pair allotment is:

$$k_{perGPU} = \frac{-4Qn - 4Q + am - 24n}{20Q + 20} \quad (2.33)$$

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

$$\vec{\nabla}\Omega = 3kQ \quad (2.34)$$

$$\vec{\nabla}F_k(Q) = 3kQ \quad (2.35)$$

$$\vec{\nabla}F_n(Q) = 3nQ \quad (2.36)$$

369 Thus the gradient allotment is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 12n}{32Q + 16} \quad (2.37)$$

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

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

371 Thus the allocation is:

$$\vec{\nabla}k_{perGPU} = \frac{-16Qn + am - 24n}{48Q + 20} \quad (2.39)$$

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

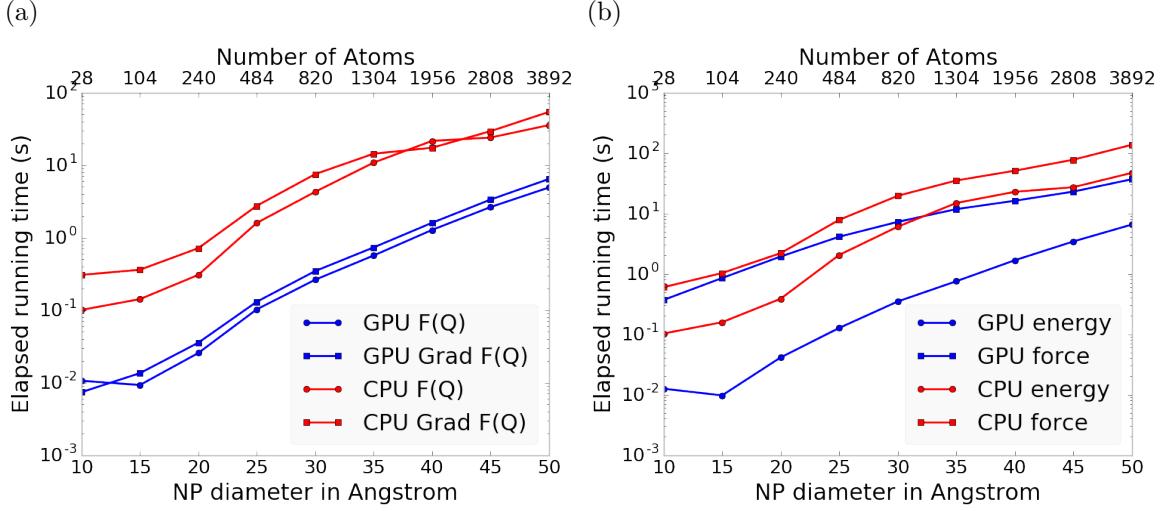


Figure 2.2: Speed comparison of CPU and GPU implementations. a) shows the time to compute the $F(Q)$ by itself. b) shows the time to compute the Rw based energy for Au NPs of various sizes, which includes computing $F(Q)$, its FFT, and the Rw .

375 Speed and Scaling of PDF Computation

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

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

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

395

CHAPTER 3

396

BENCHMARKING

397

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

398

Also some introduction would be great

399

this just needs a lot of work

400 3.1 INTRODUCTION

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

405 3.2 PDF

406 The formation of NPs with both crystallographic and non-crystallographic structures
407 [23] and with different chemical patterns [14] are well documented. For simplicity,
408 we chose monometallic Au clusters as benchmarks and considered two groups of
409 structures with different size and degrees of structural disorder in order to assess
410 the reliability and efficiency of our HMC method for solving atomic structures from
411 PDFs. The first group consists of Au_{55} clusters with different degrees of disorder,
412 including a crystalline cluster structure in O_h (Octahedral) symmetry, a structure
413 with a disordered surface, and an amorphous structure. The second group consists
414 of the crystallographically solved Au_{102} structure as in the $\text{Au}_{102}\text{MBA}_{44}$ nanocrystals
415 [16, 22]. We used optimized structures from the Density Functional Theory (DFT)

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

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

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

422 Spin-polarized DFT calculations were carried out using the Vienna ab initio simu-
423 lation package (VASP) [20, 21] within the Perdew-Burke-Ernzerhof (PBE) exchange-
424 correlation functional [29]. The projected augmented wave method [2] and a kinetic
425 energy cutoff of 400 eV were used. Structural optimization was performed until the
426 total energy and ionic forces were converged to 10^{-6} eV and 10 meV/Å, respectively.
427 The amorphous Au_{55} structures were generated by simulated annealing using the
428 classical embedded atom method potential [35]. Different annealing temperatures
429 between 1200 K and 1670 K (bulk melting temperature of Au) were used and the
430 thermally equilibrated structures were cooled down to 300 K before minimization at
431 0 K. Further optimization using DFT leads to total energies that vary within 1-2
432 eV among different amorphous structures and the lowest energy one was used as the
433 target structure. The target structure of Au_{102} was taken as the Au_{102} core of the
434 DFT-optimized $\text{Au}_{102}\text{MBA}_{44}$ cluster [22].

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

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

438 The thresholded spring potentials are based on those previously proposed on by Pe-
439 terson [30], i.e. $U_{\text{spring}}(q, r_t) = \frac{\kappa}{2} \sum_{i,j} (r_{i,j} - r_t)^2$ for all atomic distance $r_{i,j}$ outside the

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

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

458 Model Parameters

459 Unless otherwise stated, the PDFs of the target and starting structures were generated
460 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}
461 and R_{\max} correspond to the first minimum before the first PDF peak and that after
462 the last PDF peak, respectively, which ensure that the full meaningful region of the
463 PDF is modeled. For each of the simulations, the Q resolution was calculated by

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

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

470 **Au55: surface relaxed**

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

478 **Run Parameters**

479 R_{\min} and R_{\max} for this simulation were 2.45 Å and 11.4 Å, respectively, with $\delta Q =$
480 0.24 Å⁻¹. The simulation ran for approximately 34 minutes, over a total of ~40
481 thousand configurations. The results are shown in Fig. 3.1.

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

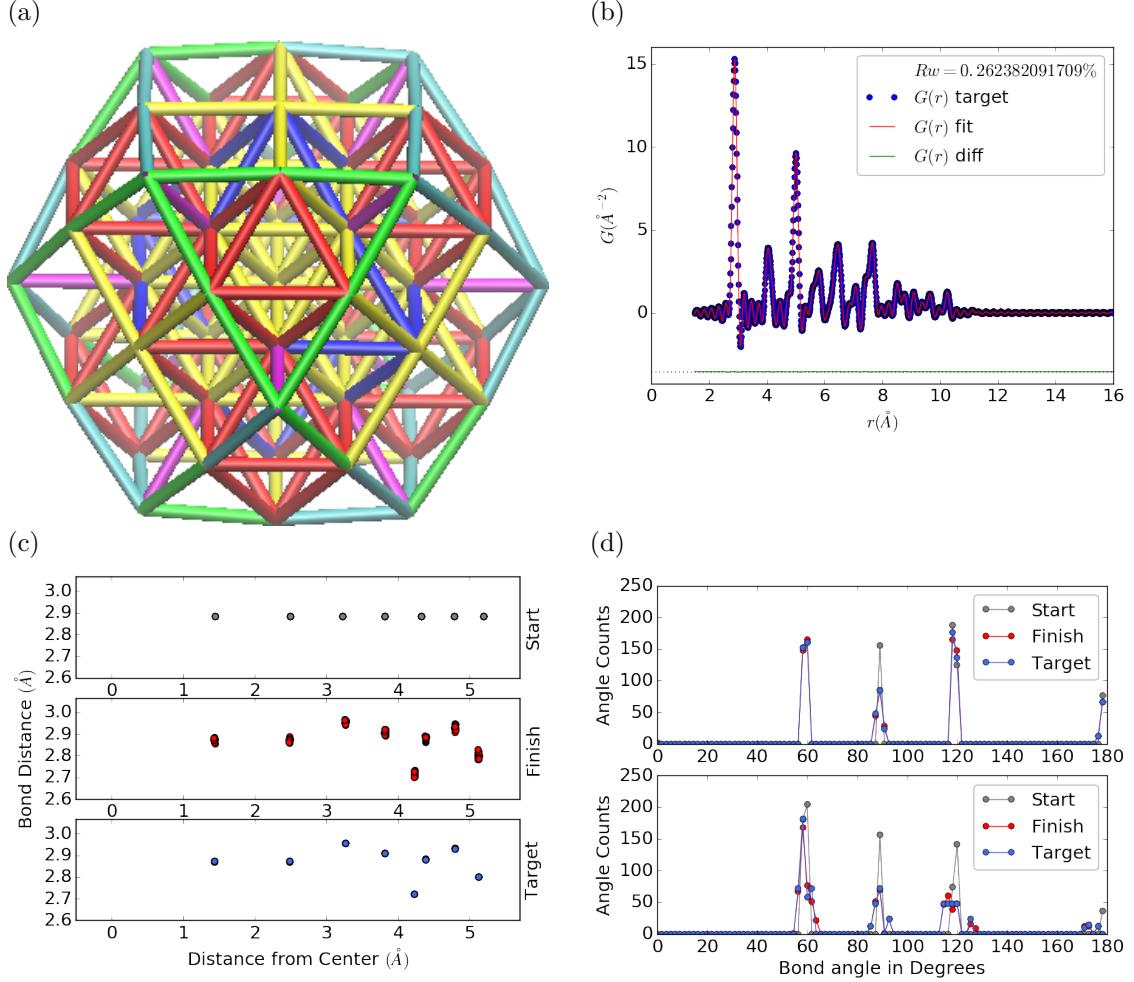


Figure 3.1: Au_{55} PDF fitting of DFT-optimized $c\text{-Au}_{55}$. (a) the final structural solution ($Rw=0.3\%$) with bond lengths color-coded by step of 0.05\AA , (b) the target PDF(blue dots) overlaid with the PDF of the final structure (solid red lines) with the difference in green lines offset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

489 below, the situation is rather different for much more disordered target structures.
 490 Interestingly, the small-box solution using PDFgui[12] yields a rather large Rw of
 491 43%, due to the failure to fit the surface contracted atoms with a unit cell. The PDF
 492 fits of the starting structure and small-box solution are shown

493 Put this somewhere

494 .

495 **Au55: surface disordered**

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

502 R_{\min} and R_{\max} for this simulation were 1.95 Å and 12.18 Å, respectively, with
503 $\delta Q = 0.23 \text{ \AA}^{-1}$. The simulation ran for approximately 3.6 hours, over a total of ~ 270
504 thousand configurations. The results of the simulation are shown in Fig. ??.

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

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

522 and high CN numbers. This is partly caused by the displacement of the core atoms
523 mentioned above, and partly by the lack of CN constraints for the PDF fitting, which
524 has been previously demonstrated in the case of α -Si [7]. Additional experimental
525 data, e.g. from EXAFS or NMR, may help to steer the simulations towards better
526 agreement in both PDF and CN distribution.

527 Au55: amorphous

528 Next, we turn to the case in which the entire cluster structure is disordered. We used
529 a DFT-optimized amorphous Au_{55} ($a\text{-Au}_{55}$) as the target structure, and the DFT-
530 relaxed $c\text{-Au}_{55}$ cluster from Case I as the starting structure. The total energy of
531 $a\text{-Au}_{55}$ was computed to be *lower* than that of $c\text{-Au}_{55}$ by as large as 2.9 eV, consistent
532 with the 3.0 eV found in previous DFT work [9].

533 R_{\min} and R_{\max} for this simulation were 2.6 Å and 11.26 Å, respectively, with
534 $\delta Q = 0.25 \text{ \AA}^{-1}$. The simulation ran for approximately an hour, over a total of ~ 87
535 thousand configurations. The results of the simulation are shown in Fig. ??.

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

548 **Au102: triple phase**

549 Our final benchmark is Au₁₀₂, whose structure was initially solved by Jadzinsky and
550 co workers using x-ray crystallography [16] and further confirmed by DFT studies
551 [22]. The Au₁₀₂ structure consists of three main parts, a 49-atom Marks decahedron
552 core, two C_5 caps consisting of 20 atoms each, and 13 equatorial atoms. Unlike
553 previous cases, the multi-symmetry nature of the structure, i.e. each part has its own
554 distinct symmetry, poses a challenge for PDF-based solution of the structure. This is
555 because of the atomically centralized nature of the PDF, in which each atom “sees”
556 a density of other atoms surrounding it and has a strong tendency towards becoming
557 the center of the main symmetry group. Such tendency may lead to a solution where
558 some of the correct atomic symmetries are discarded in favor of the core symmetry.

559 **Starting from fcc structure**

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

562 R_{\min} and R_{\max} for this simulation were 2.7 Å and 16. Å, respectively, with $\delta Q =$
563 0.18 Å⁻¹. The simulation ran for approximately two hours, over a total of ~82
564 thousand configurations. The results of the simulation are shown in Fig. 3.4.

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

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

584 **Marks decahedron**

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

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

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

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

603 **Au147**

604 **3.3 PDF WITH ADPs**

605 **ADP 50**

606 1. Basic 50% larger magnitude

607 2. Random addition to APDs

608 3. Janus ADPs

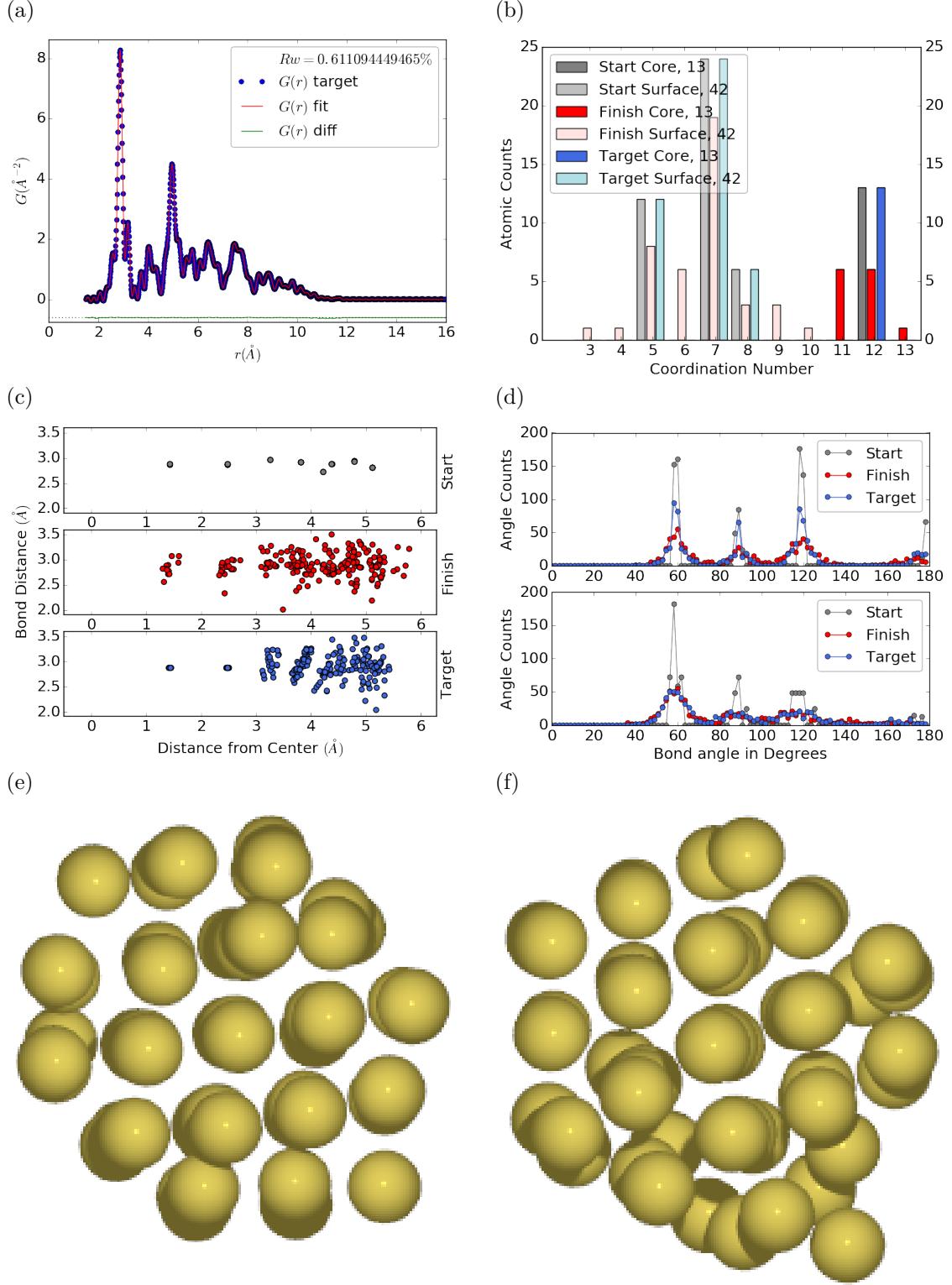


Figure 3.2: Au_{55} PDF fitting of surface-disordered Au_{55} . a) the target structure, b) the final structural solution ($R_w=0.6\%$), c) the comparison of PDFs, d) the CN distribution with the number of atoms in either the core or the surface, e) the radial bond distribution, and f) the bond angle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

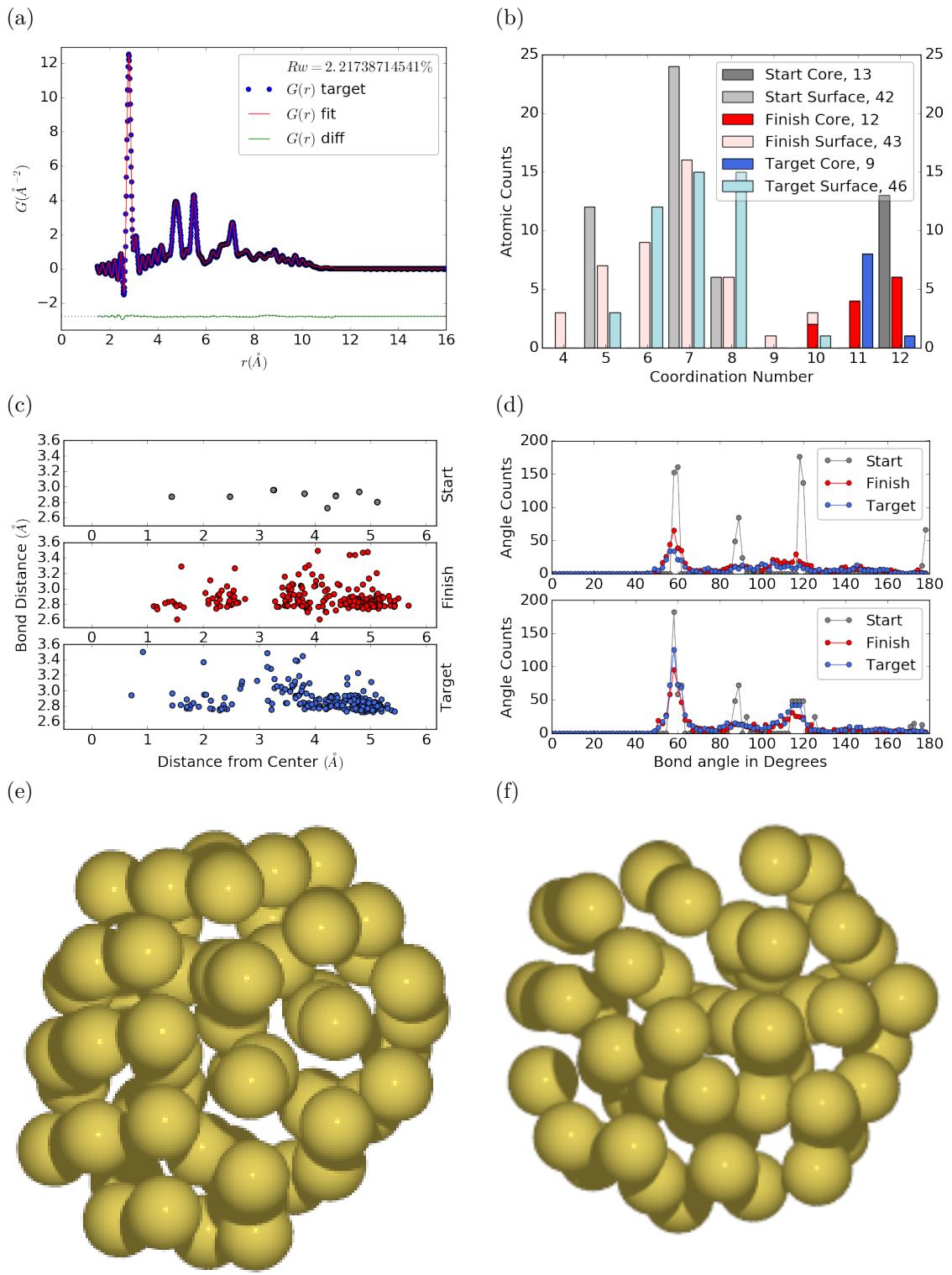


Figure 3.3: Similar to figure 3.2 for DFT-optimized amorphous Au_{55} .

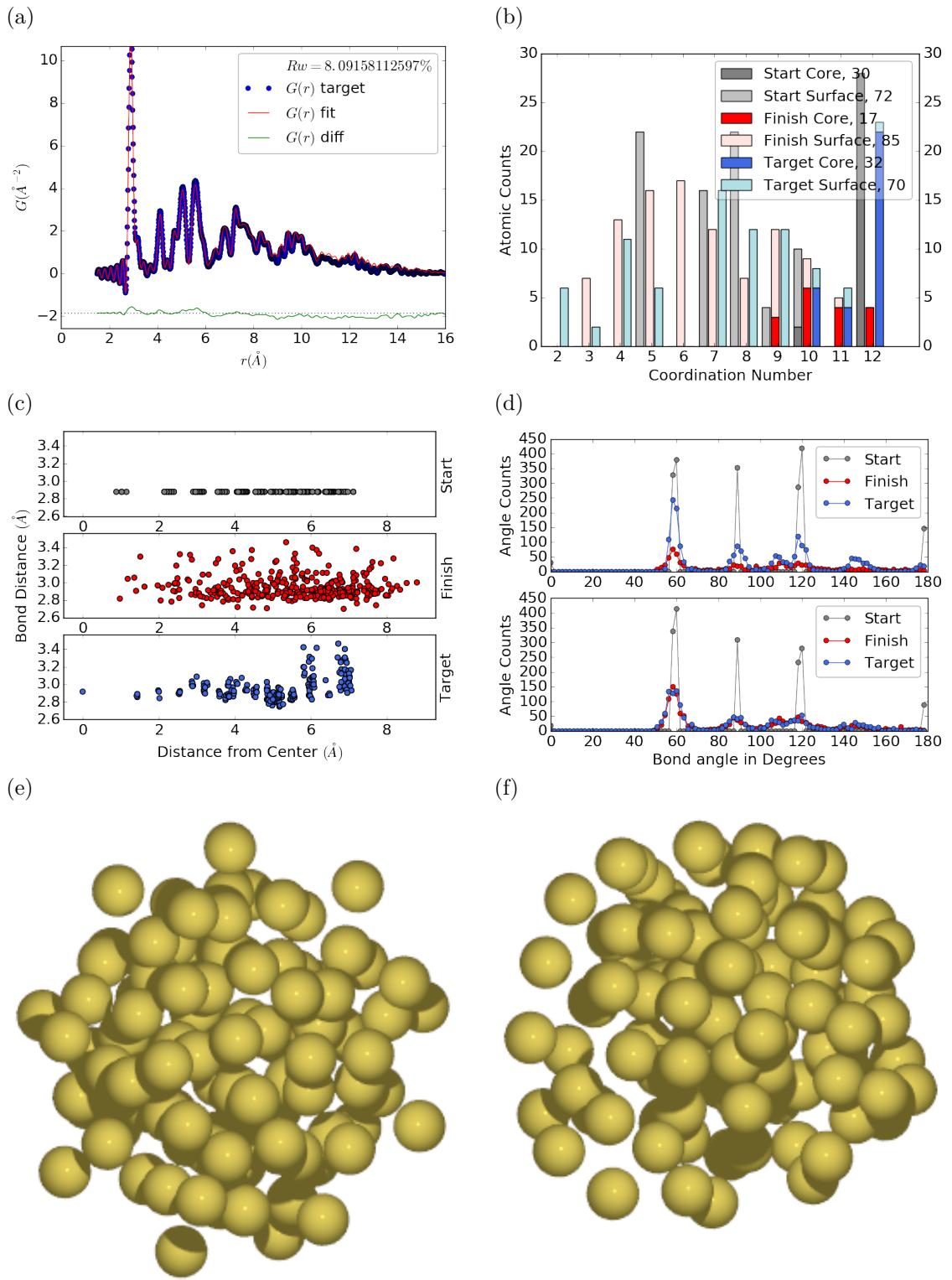


Figure 3.4: Similar to figure 3.2 for Au_{102} as in DFT-optimized $\text{Au}_{102}\text{MBA}_{44}$ cluster.

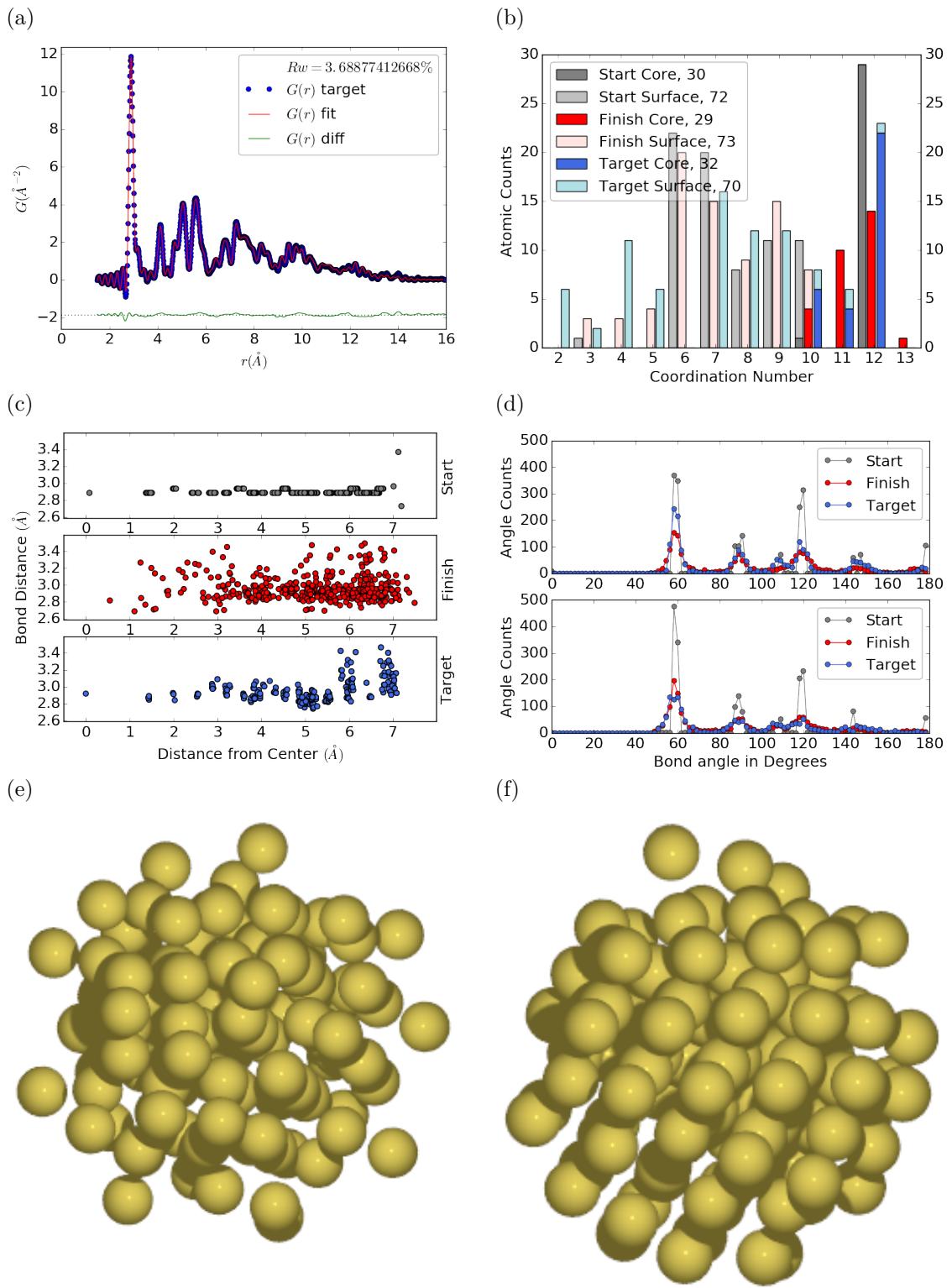


Figure 3.5: Similar to Fig. 3.4 with Marks decahedron as the starting structure.

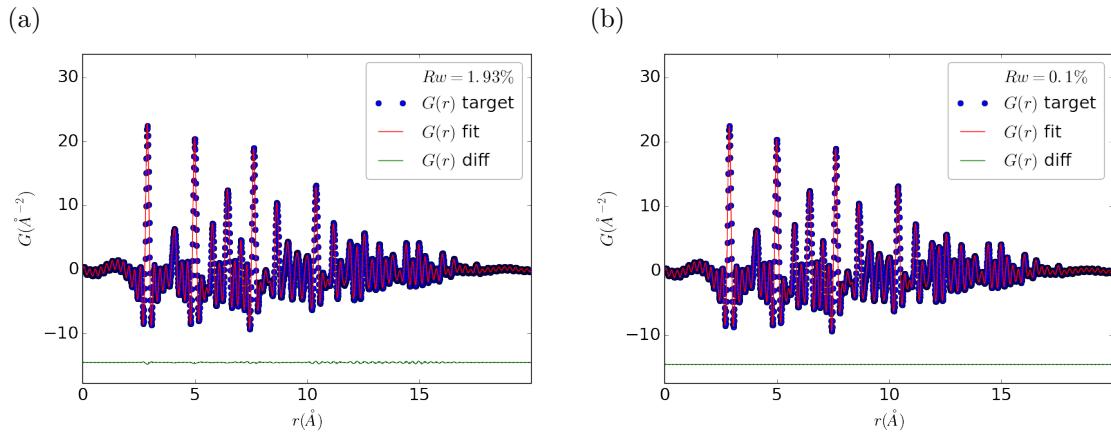


Figure 3.6: Refinement of adps

609

CHAPTER 4

610 X-RAY TOTAL SCATTERING DATA ACQUISITION AND

611 PROCESSING

612 4.1 INTRODUCTION

613 X-ray total scattering experiments are generally performed at synchrotron light sources,
614 as only these sources can provide the needed flux, energy, and high momentum trans-
615 fer vectors needed to obtain reliable PDFs. [6, 11] Despite the need for a dedicated
616 facility to perform the total scattering experiments, the experiments themselves are
617 fairly forgiving, allowing for reactive gaseous environments, experiment temperatures
618 ranging from 2 K to 1200 K, and even electrochemical cycling. [5, 32, 34] The rapid
619 PDF data acquisition associated with 2D area detectors creates a data management
620 problem, as 96 hours of beamtime could result in almost 10,000 images which need
621 to be associated with the experimental conditions and detector metadata. [6] Finally,
622 all this data needs to be processed by masking bad pixels and regions, integrating
623 azimuthally, and converting the scattering data to the PDF. [19, 18, 36, 28, 1]

624 4.2 DETECTOR Q RESOLUTION

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

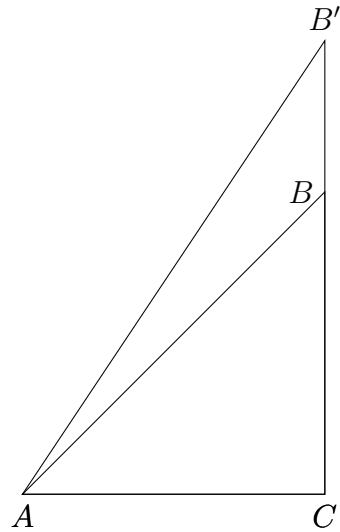


Figure 4.1: Scattering onto a flat detector

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

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

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

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

636 where λ is the x-ray wavelength.

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

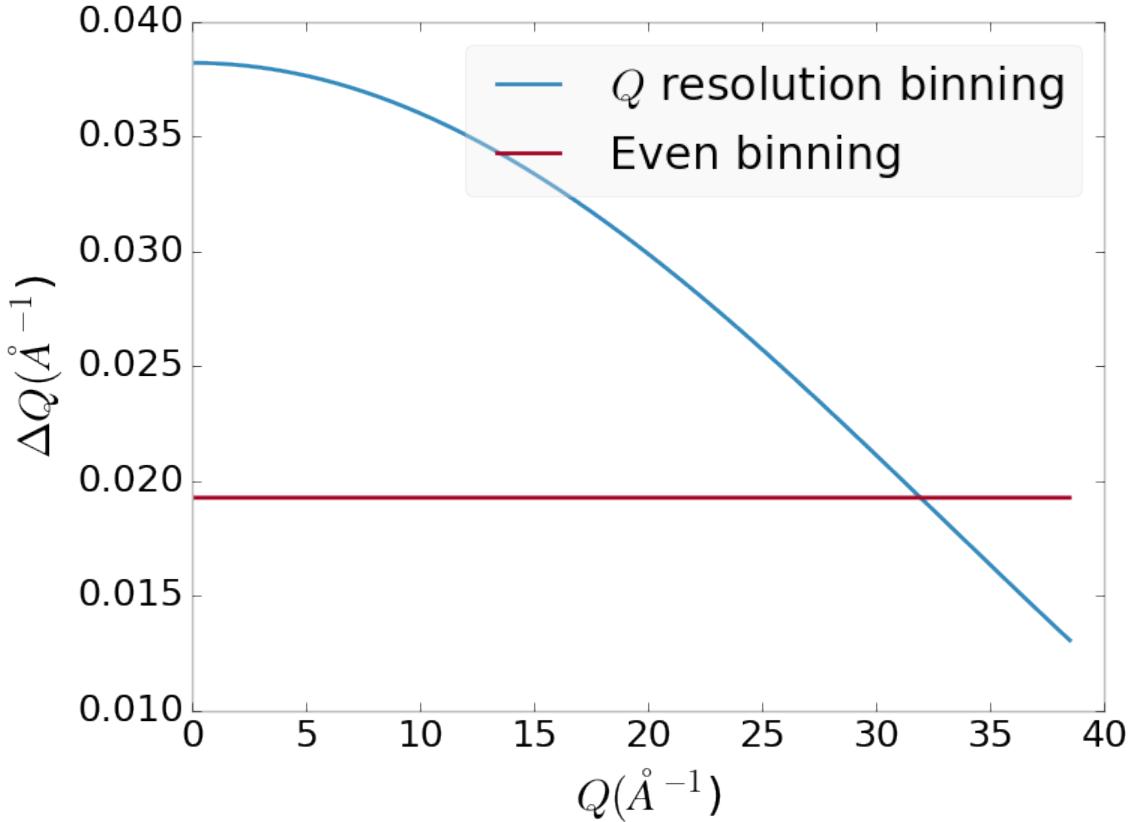


Figure 4.2: Q resolution as a function of Q .

640 4.3 AUTOMATED MASK GENERATION

641 **Introduction**

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

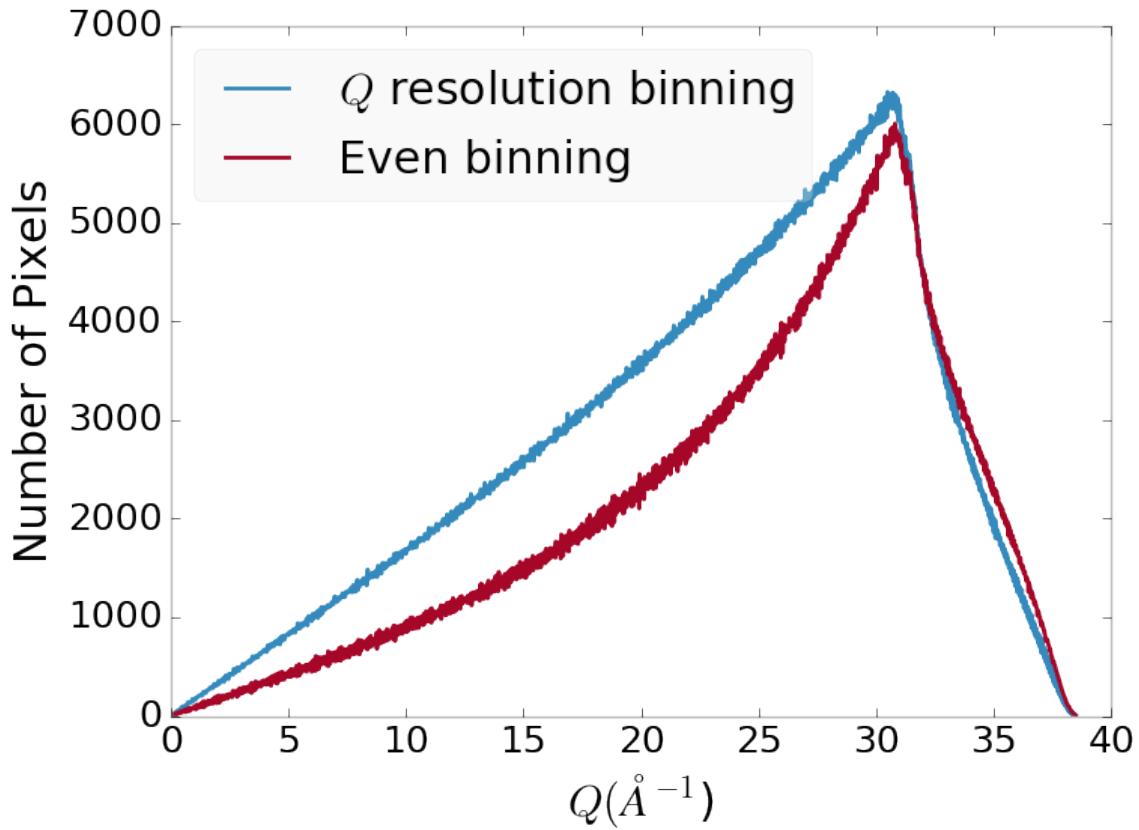


Figure 4.3: Number of pixels as a function of Q , binned at the Q resolution of the detector.

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

660 **Algorithm Design**

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

667 **Test Cases**

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

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

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

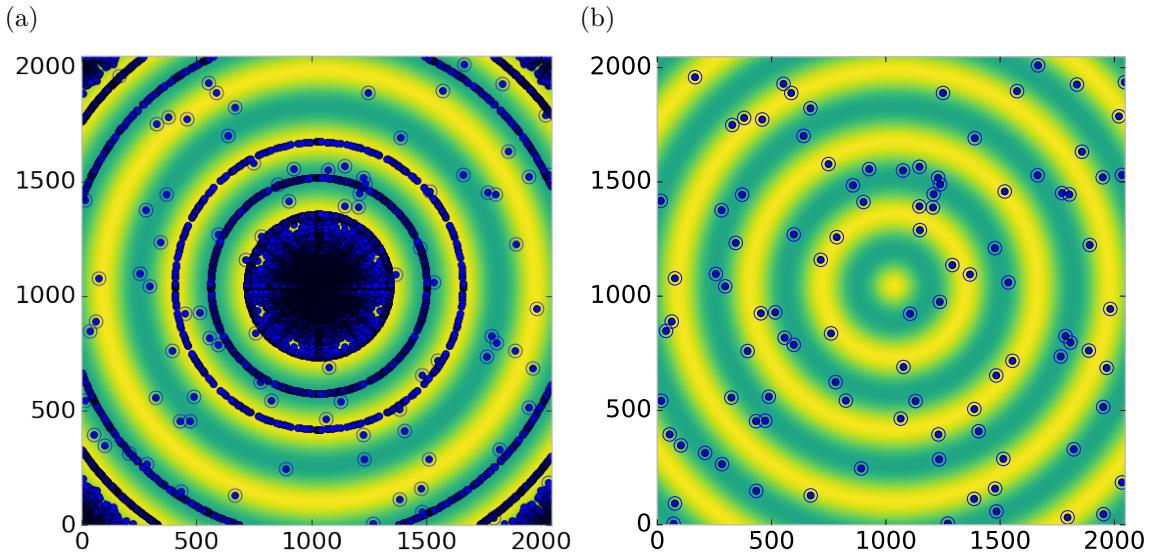


Figure 4.4: Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

684 Results and Discussion

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

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

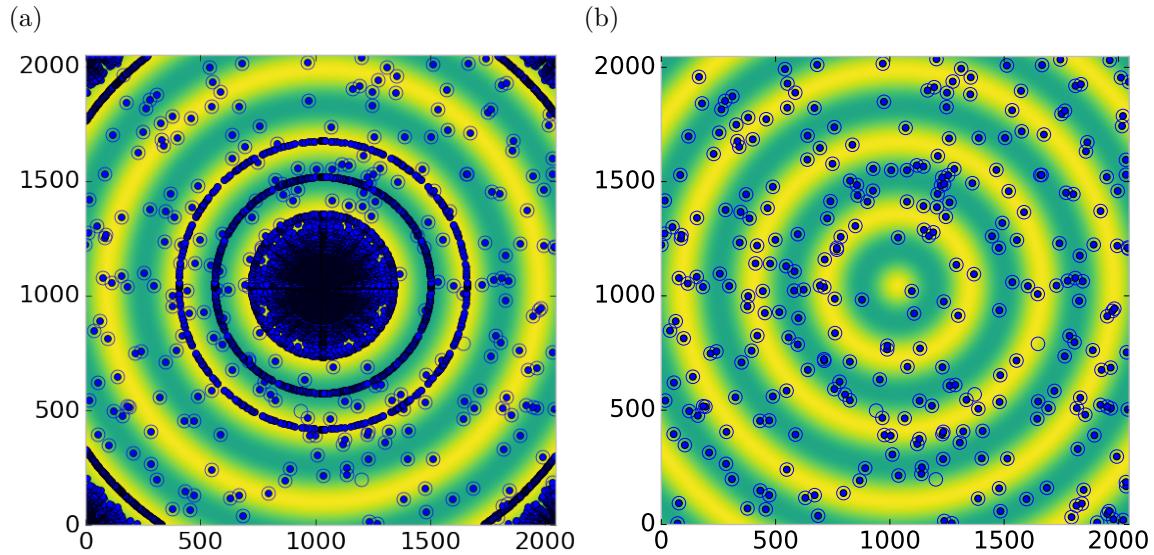


Figure 4.5: Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

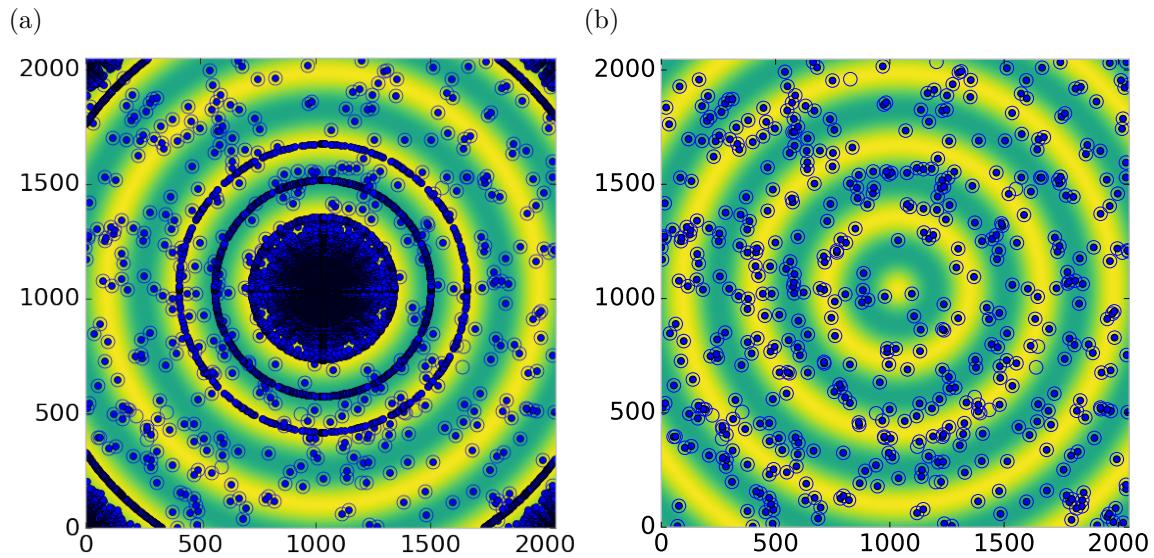


Figure 4.6: Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

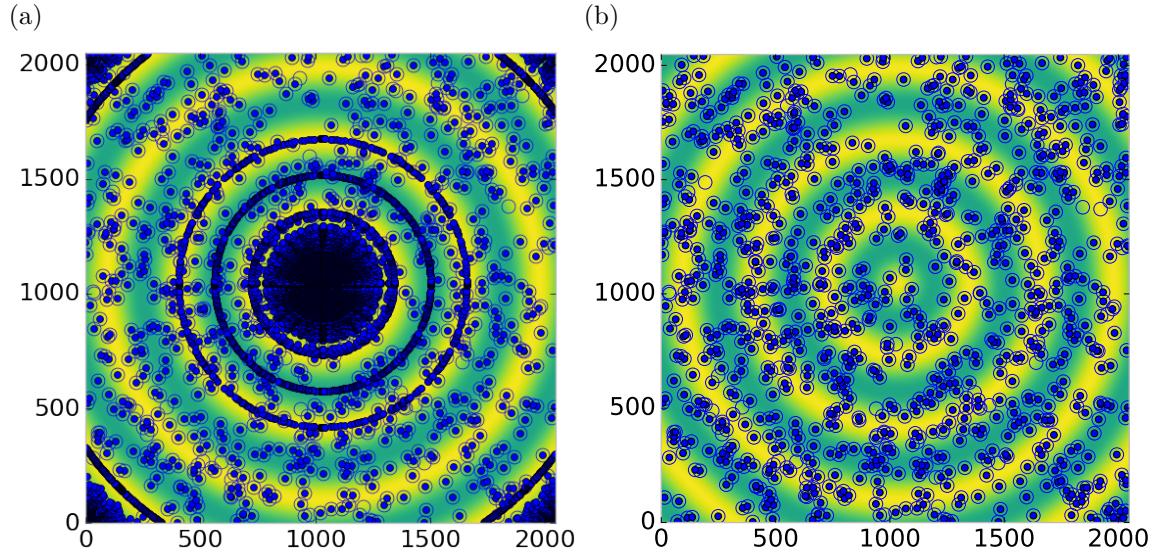


Figure 4.7: Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the standard even bin mask and b) the Q resolution binned mask. The bad pixels are noted with open circles, masked pixels are noted with closed circles.

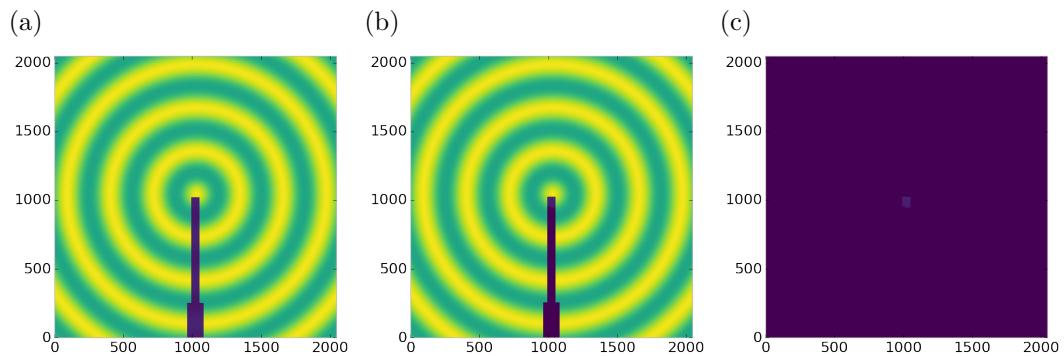


Figure 4.8: Generated beamstop holder masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

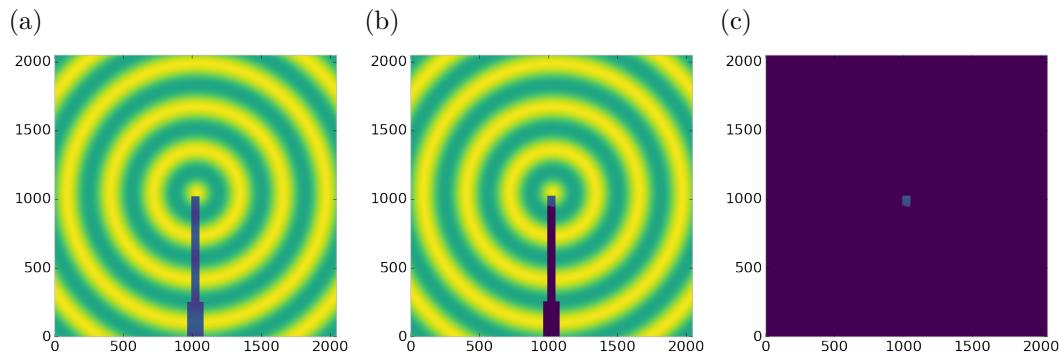


Figure 4.9: Generated beamstop holder masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

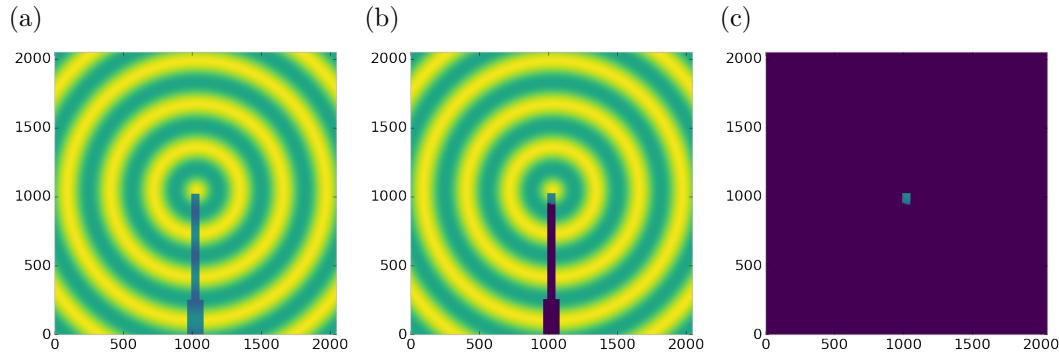


Figure 4.10: Generated beamstop holder masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

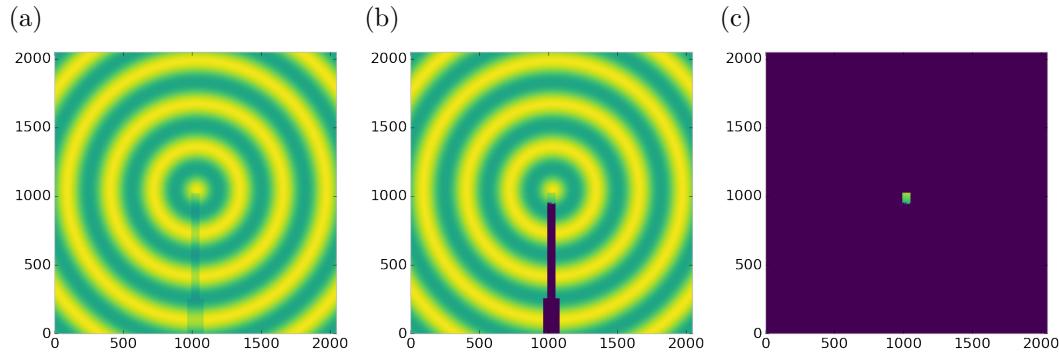


Figure 4.11: Generated beamstop holder masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

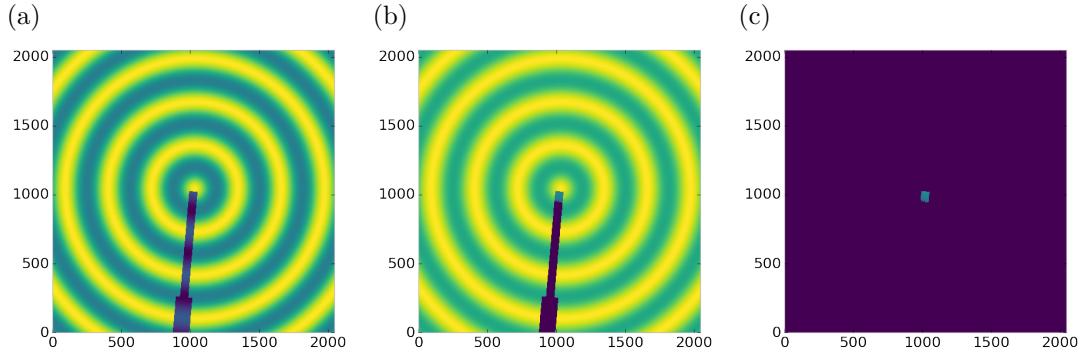


Figure 4.12: Generated beamstop holder masks which is rotated away from vertical

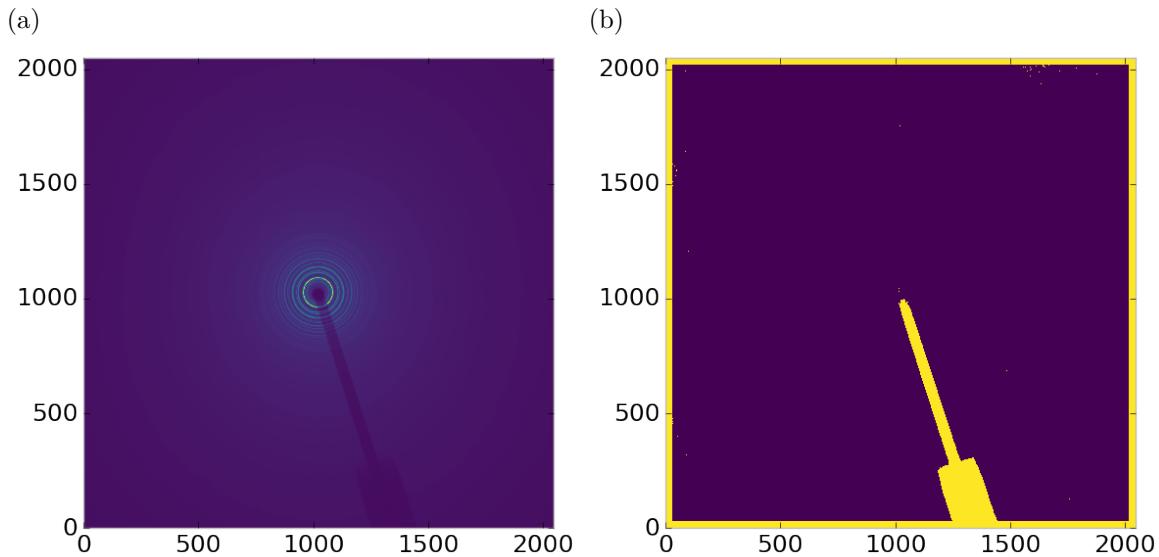


Figure 4.13: Masked experimental data. a) the raw image, b) the mask

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

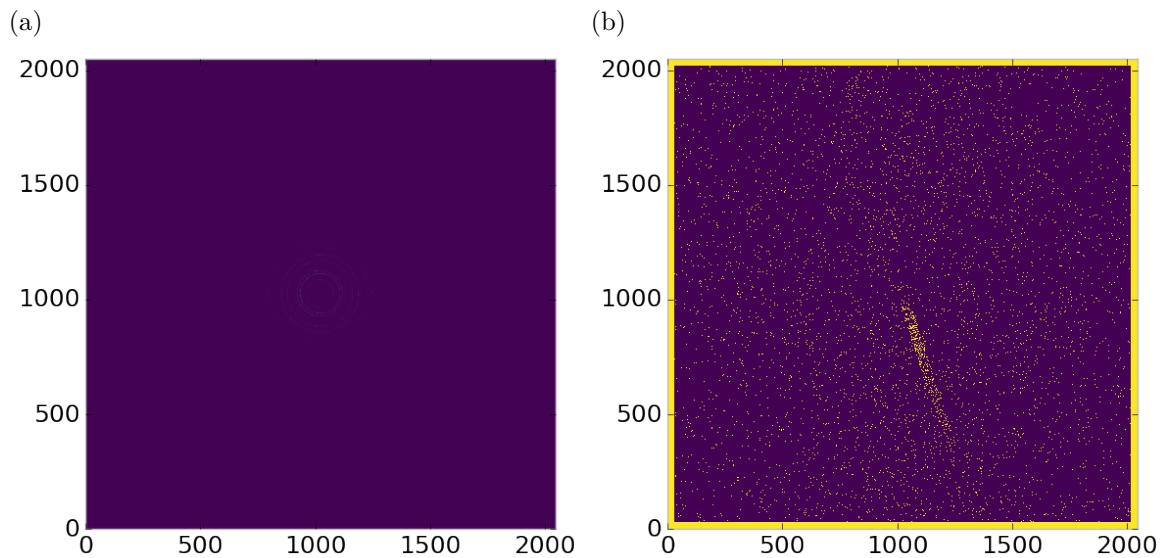


Figure 4.14: Masked experimental data with Pt single crystal signal. a) the raw image, b) the mask

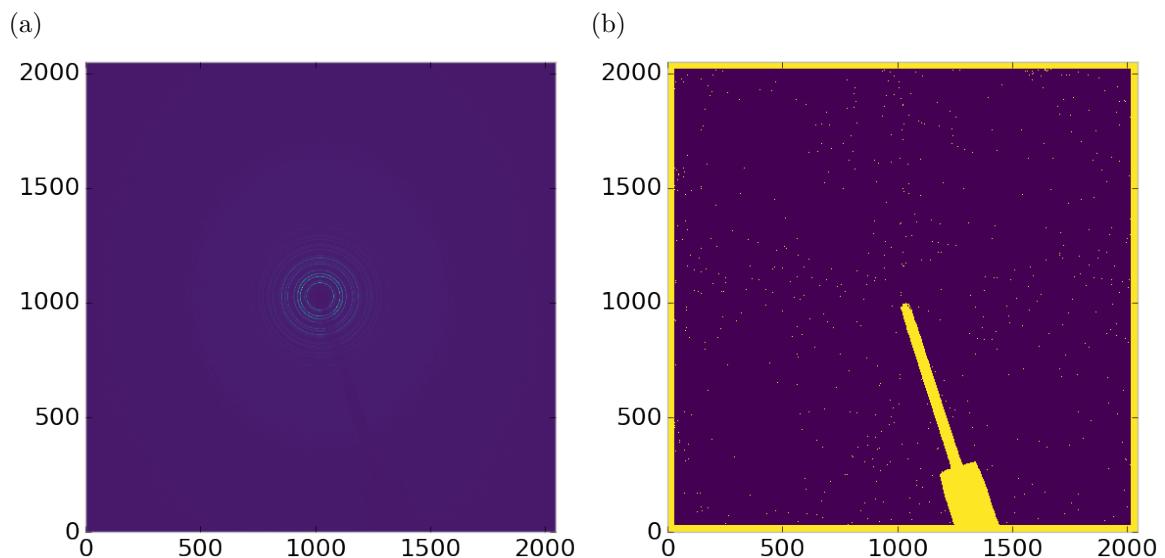


Figure 4.15: Masked experimental data with Pt single crystal signal using figure's 4.13 mask as a starting mask. a) the raw image, b) the mask

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

722 Conclusions

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

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

735 4.4 AUTOMATED IMAGE AZIMUTHAL INTEGRATION

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

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

759 4.17. Thus, for subsequent integrations discussed in this work the median is used to
 760 avoid any defective features that the masking algorithm may have missed.

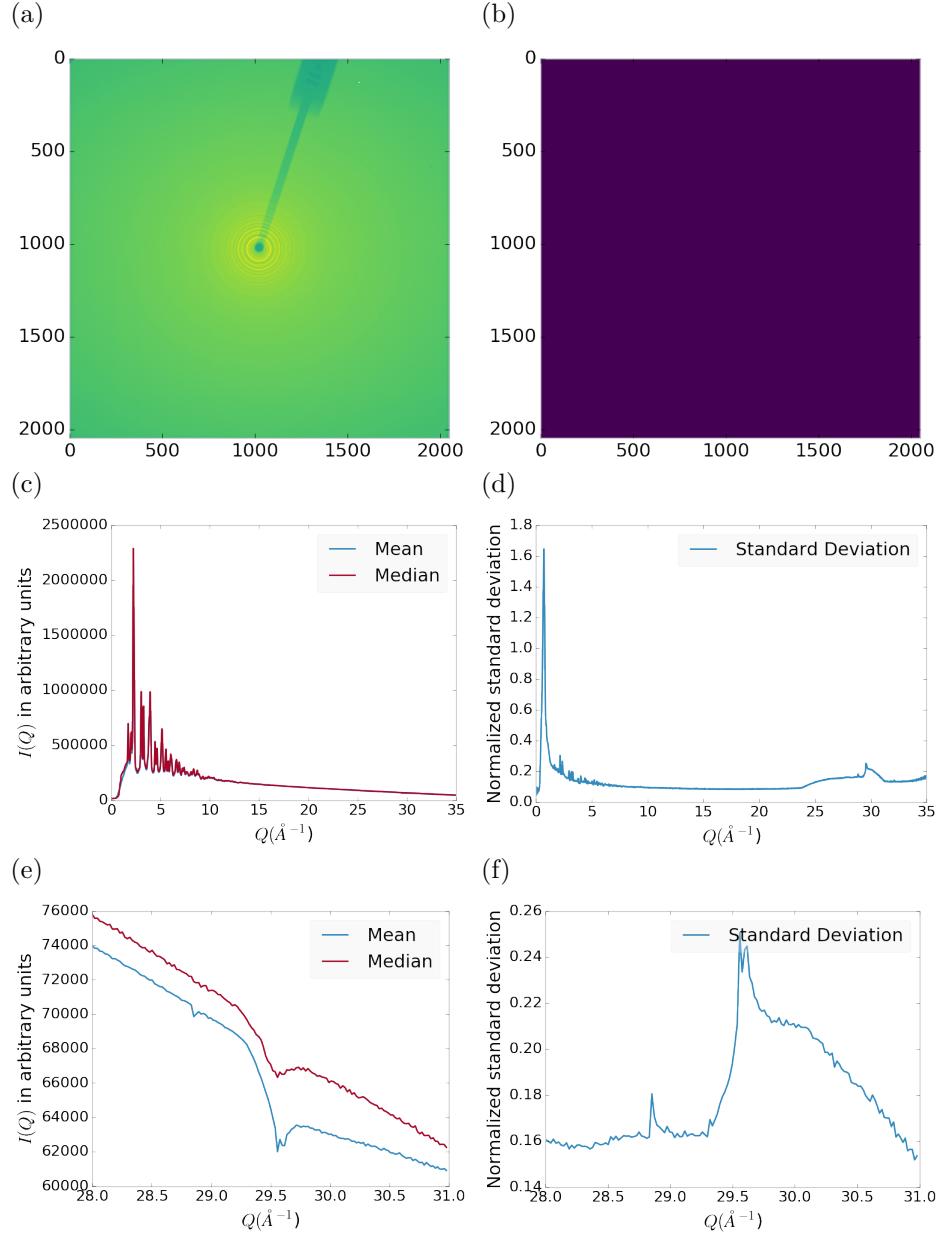


Figure 4.16: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with no mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the 28 \AA^{-1} to 31 \AA^{-1}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

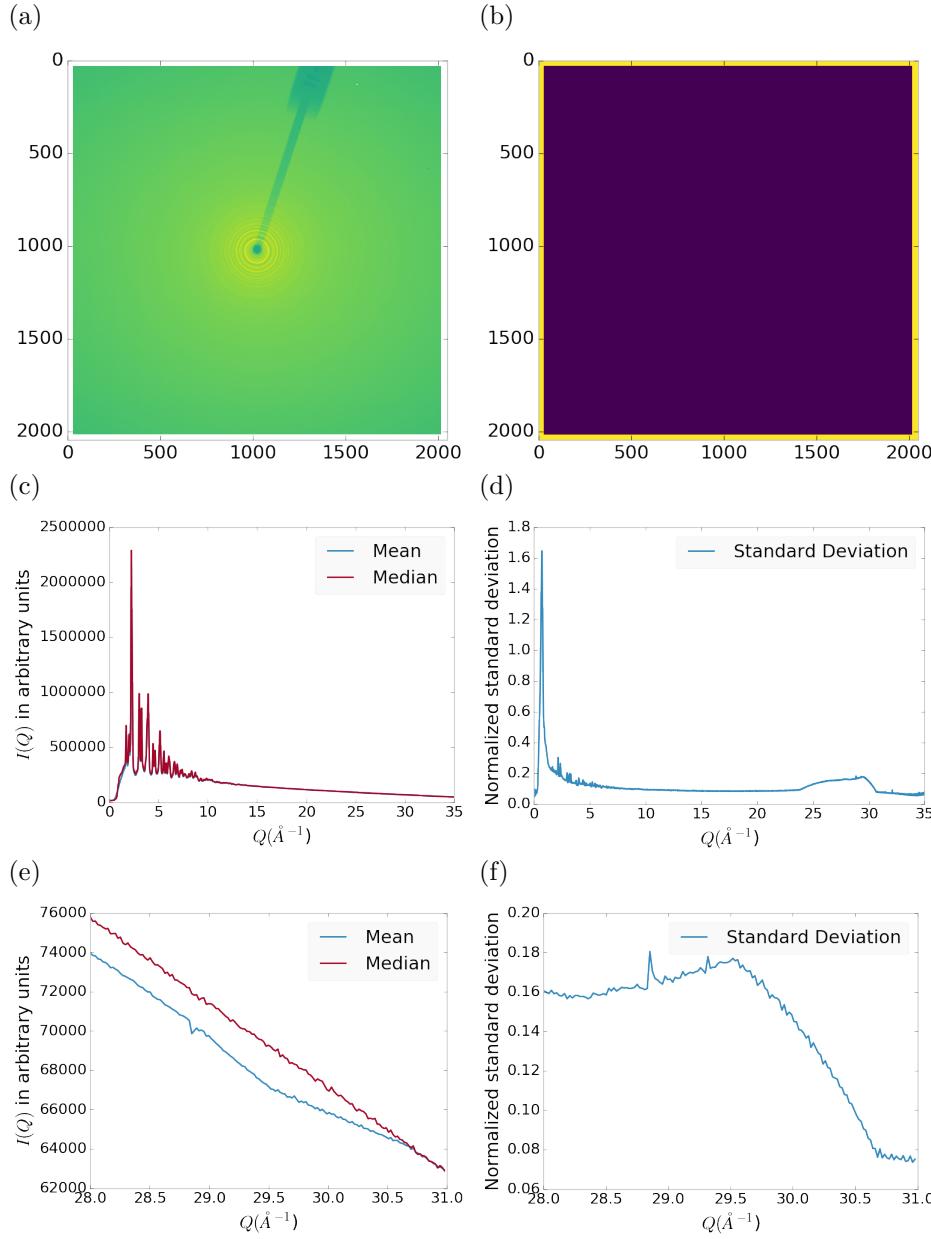


Figure 4.17: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced with only an edge mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median), e) a closeup of the 28 \AA^{-1} to 31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

761 4.5 CONCLUSIONS

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

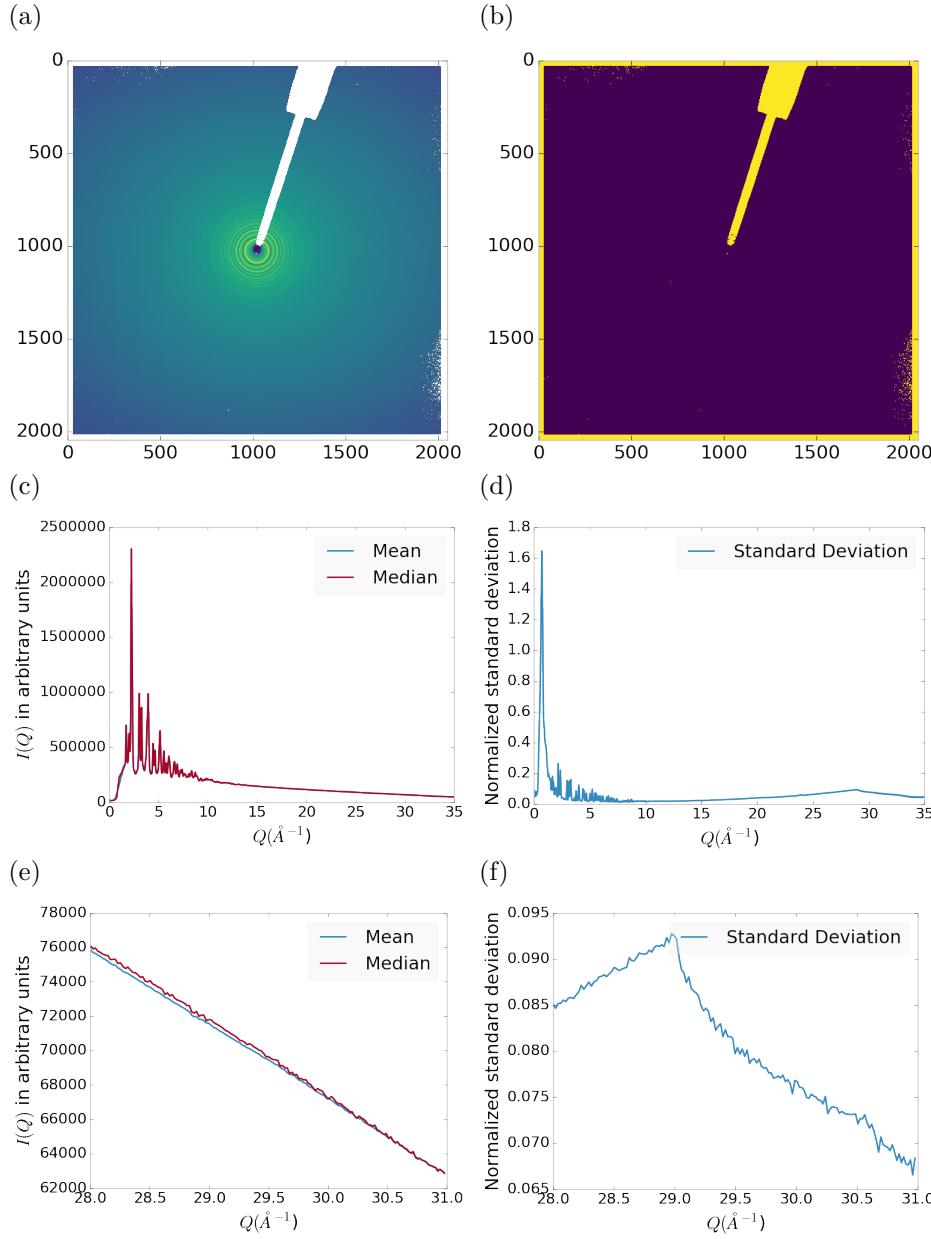


Figure 4.18: Masking, average, and standard deviation of an example x-ray total scattering measurement. This image was produced combining an edge mask and the automatically generated mask. a) the image, b) the mask, c) the mean and median values, d) the standard deviation (normalized to the median),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

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

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

779

CHAPTER 5

780

PHASE CHANGES AND ANNEALING DYNAMICS OF

781

Pr_2NiO_4 AND ITS DERIVATIVES

782 5.1 EXPERIMENTS

783 Pr_2NiO_4 Synthesis

784

need some sort of synthesis information

785 X-ray Measurements

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

791 5.2 DATA PROCESSING

792

masking parameters

793

integration parameters

794

PDF parameters

795 5.3 DATA ANALYSIS

Need to redo all the figures:
need to use the new stylesheet
need to be more readable
need to cull through them to pull out the interesting features

796

797 Intra Sample Comparison

798 As figures 5.4

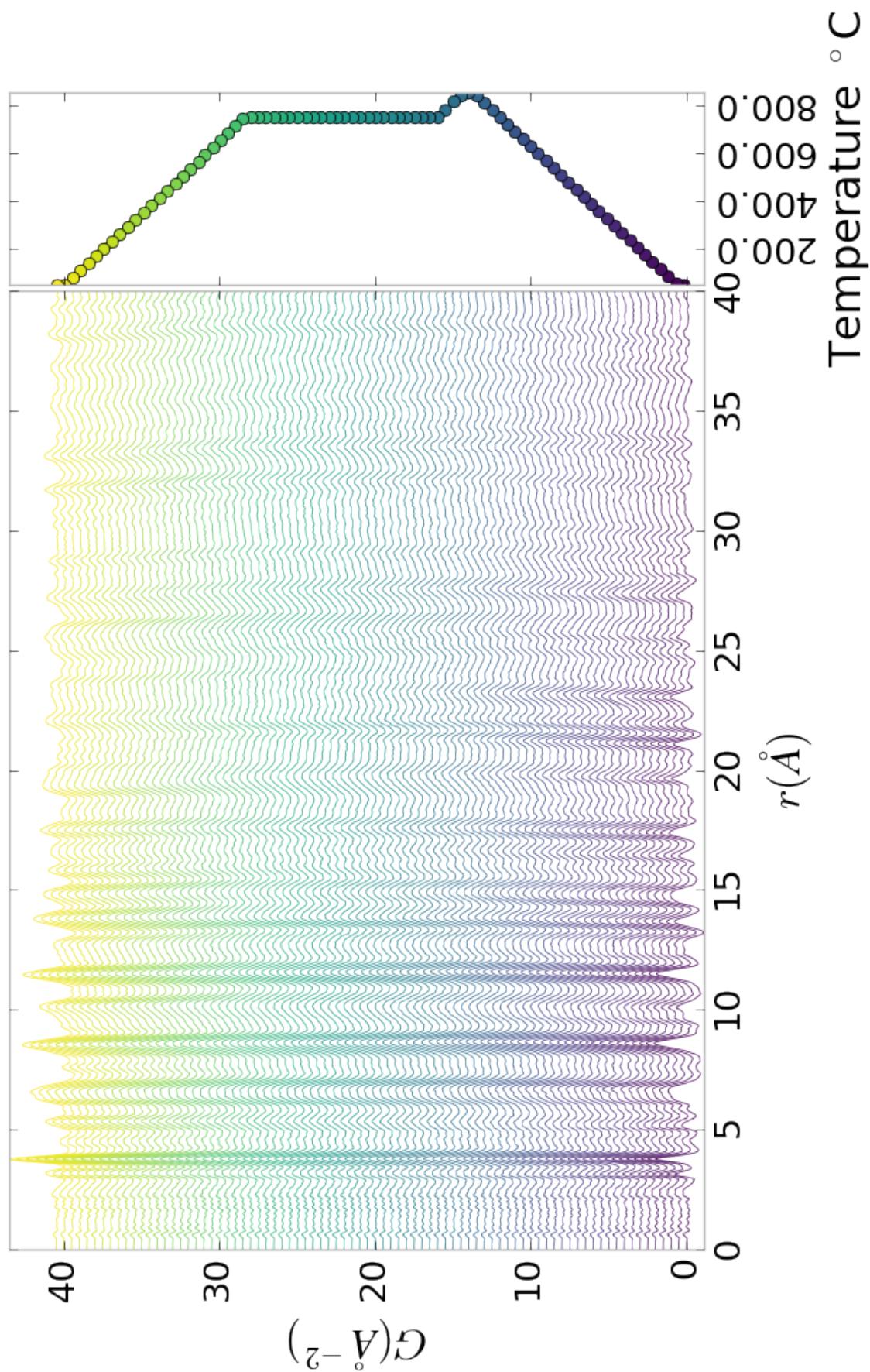


Figure 5.1: PDF as a function of temperature for as synthesized PNO showing the full PDF

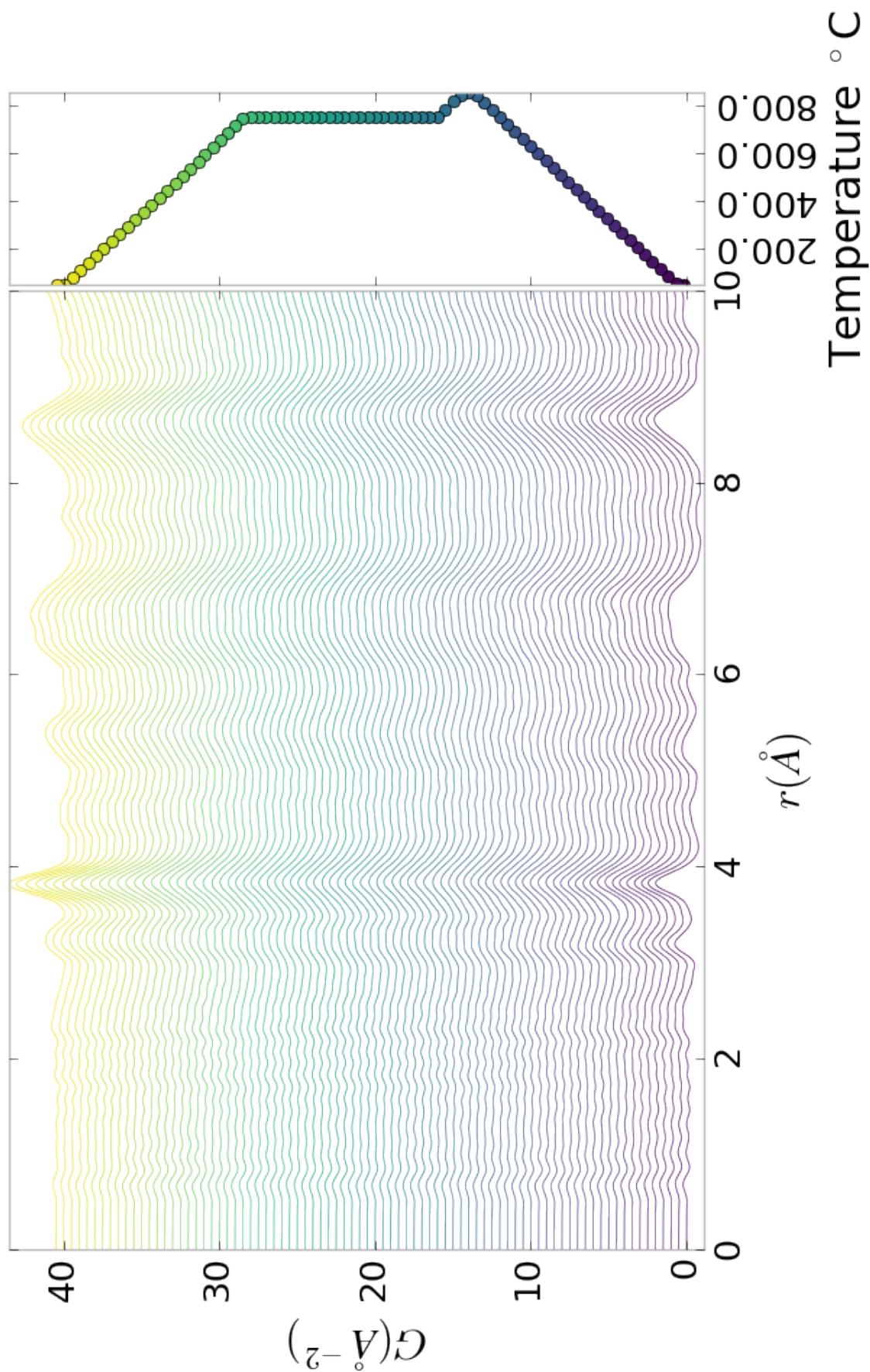


Figure 5.2: PDF as a function of temperature for as synthesized PNO showing a close up on the short range section

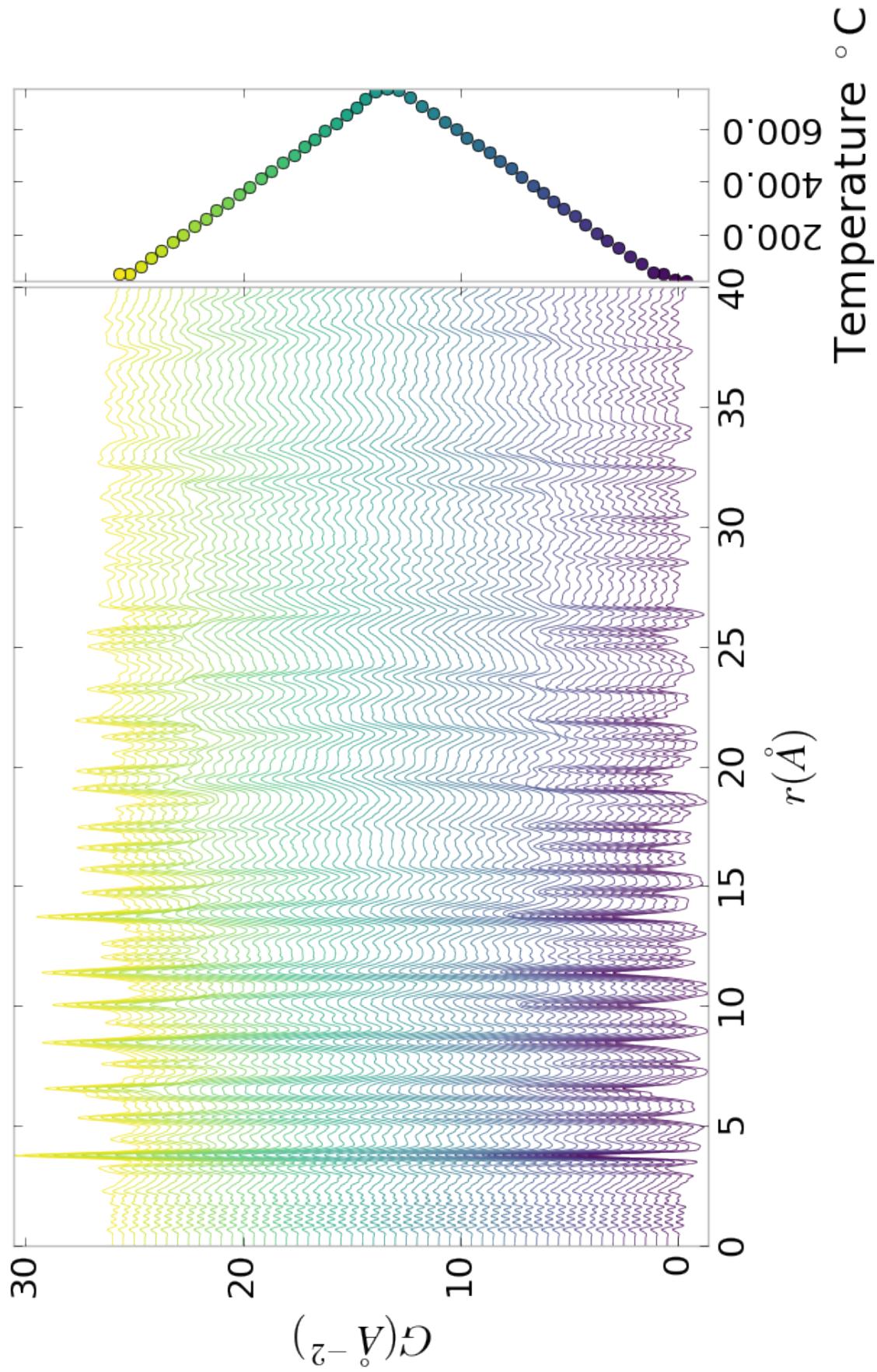


Figure 5.3: PDF as a function of temperature for PNO annealed at 750 °C for 25 hours showing the full PDF

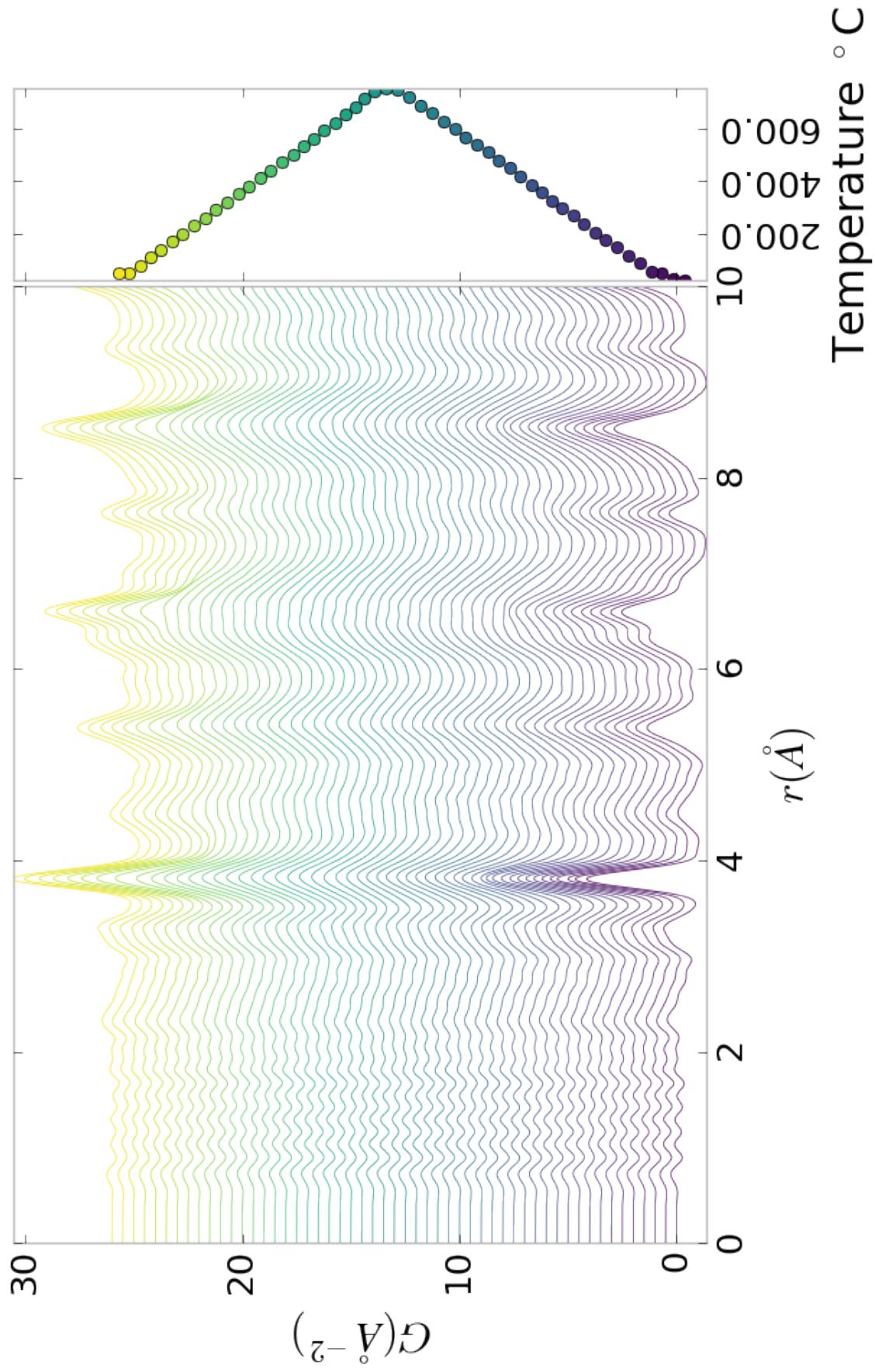


Figure 5.4: PDF as a function of temperature for PNO annealed at 750 $^{\circ}\text{C}$ for 25 hours showing a close up on the short range section

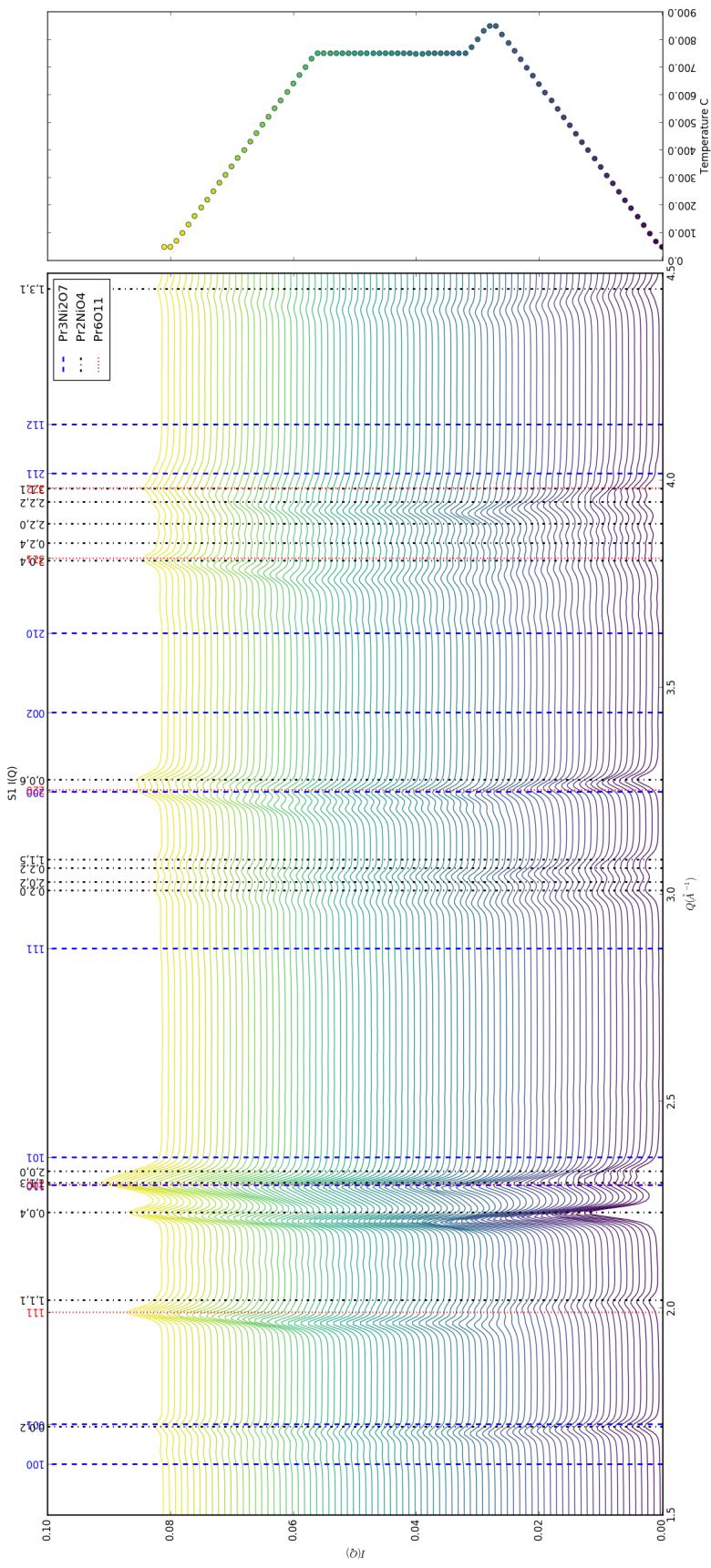


Figure 5.5: XRD as a function of temperature for S1

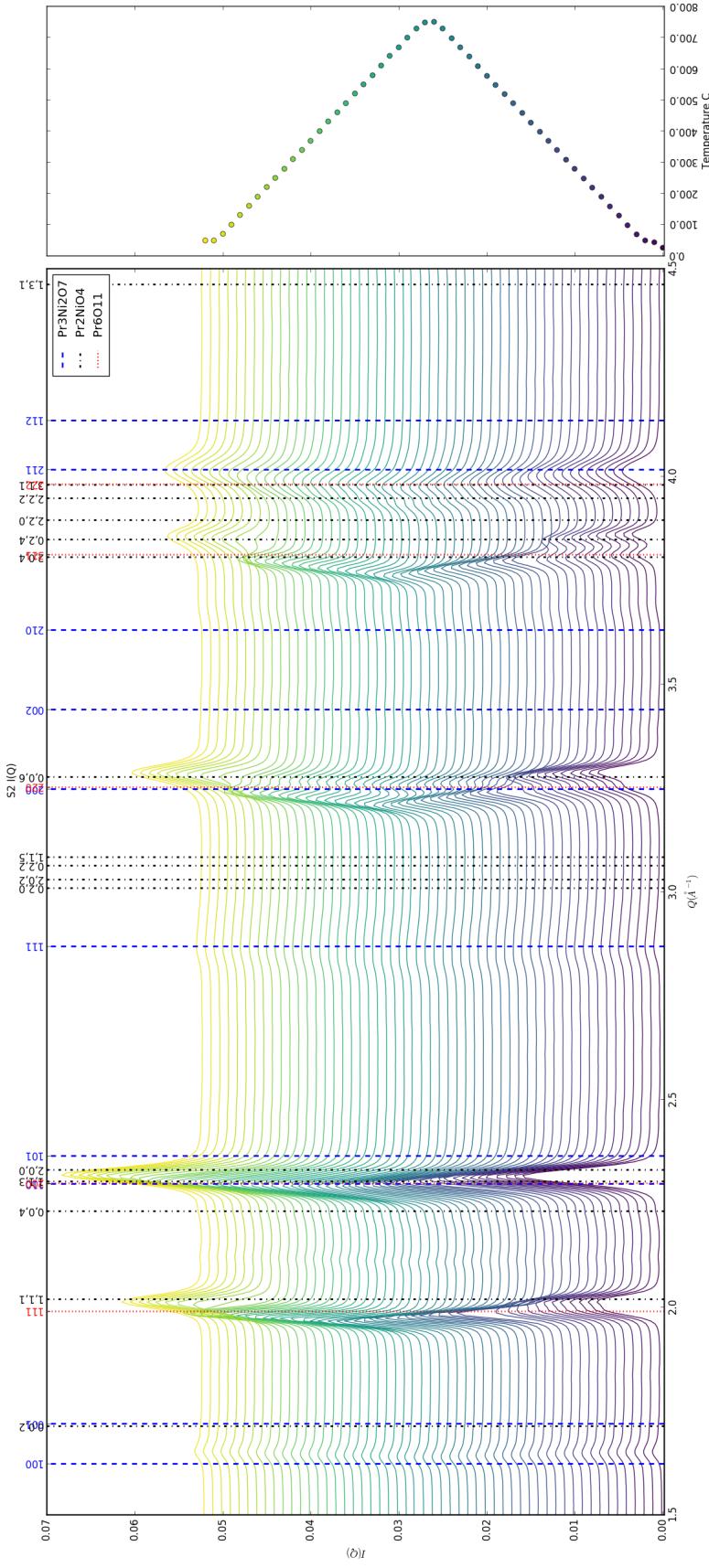


Figure 5.6: XRD as a function of temperature for S2

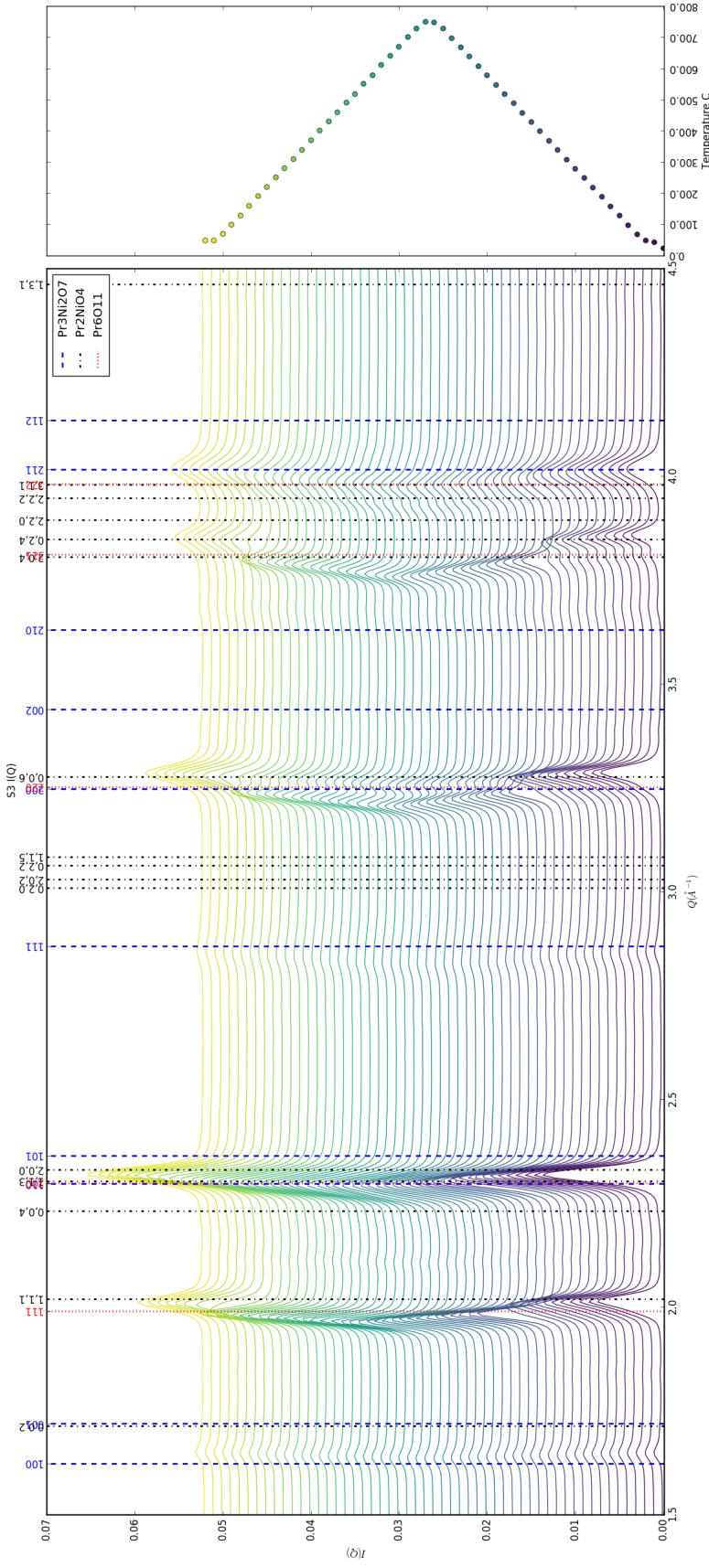


Figure 5.7: XRD as a function of temperature for S3

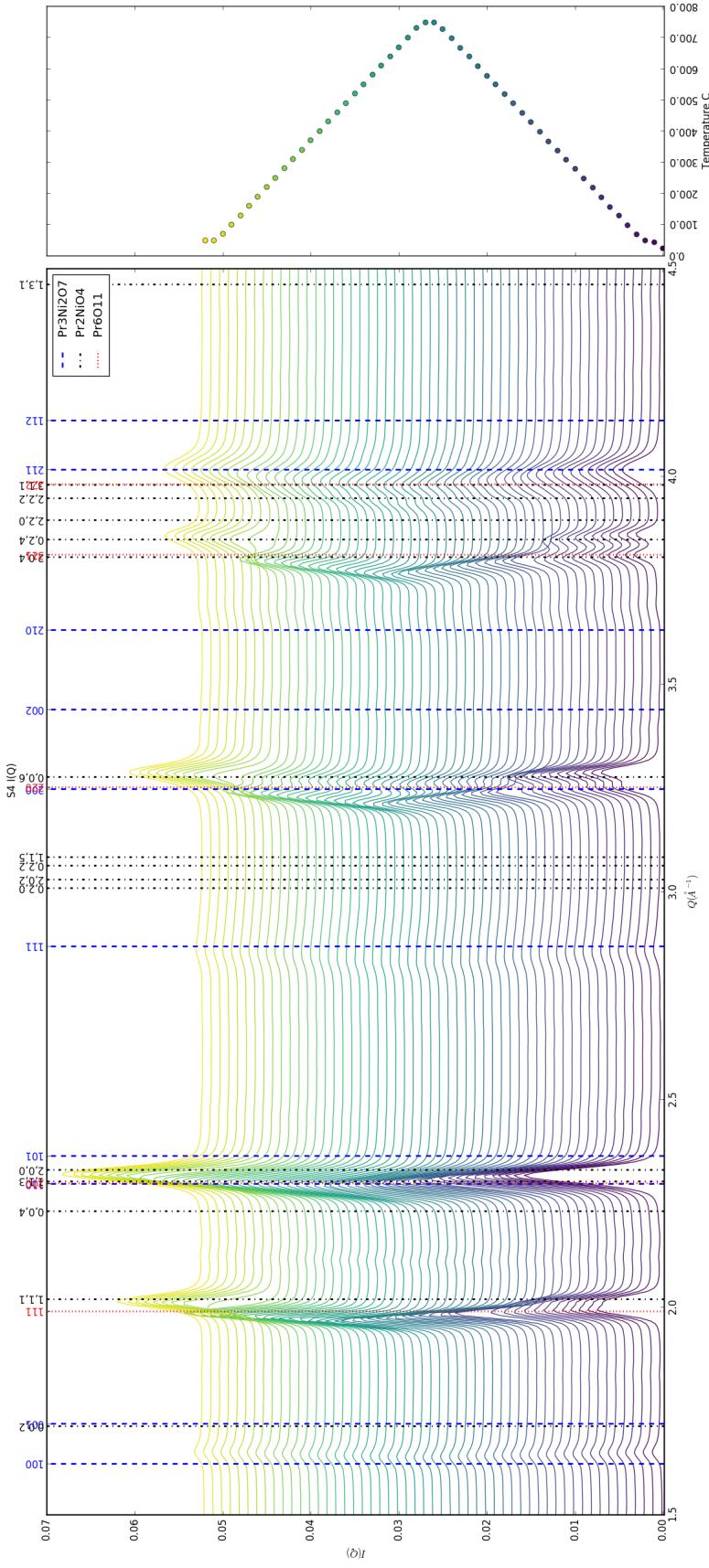


Figure 5.8: XRD as a function of temperature for S4

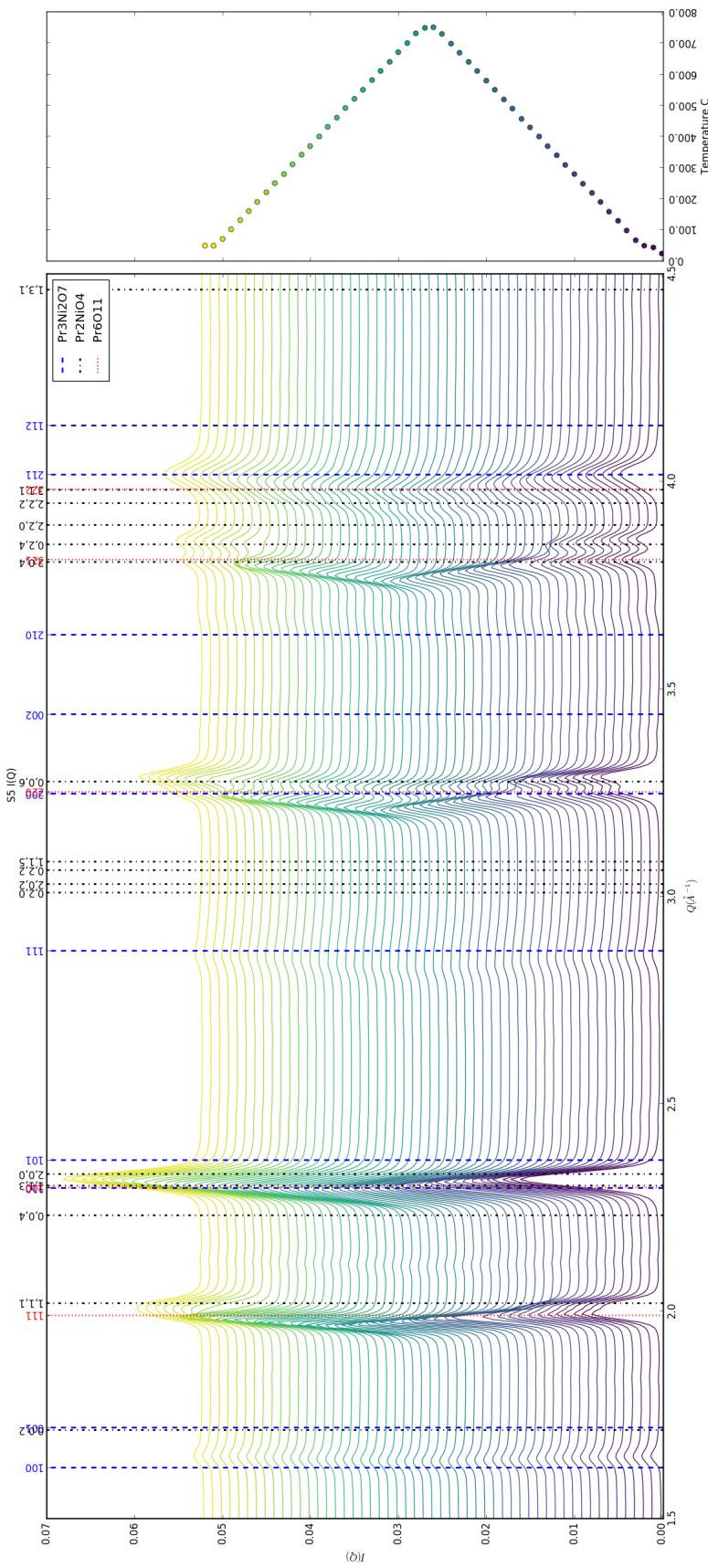
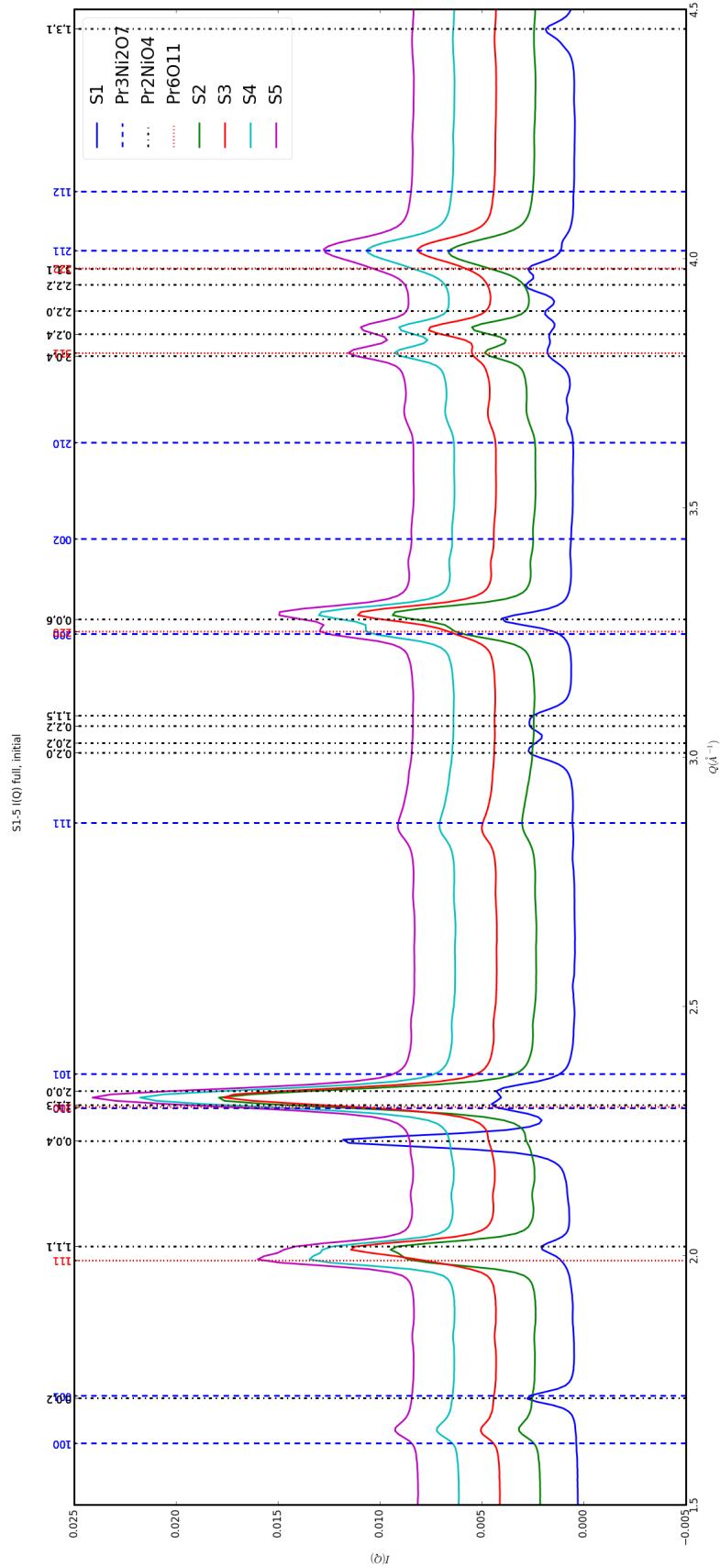
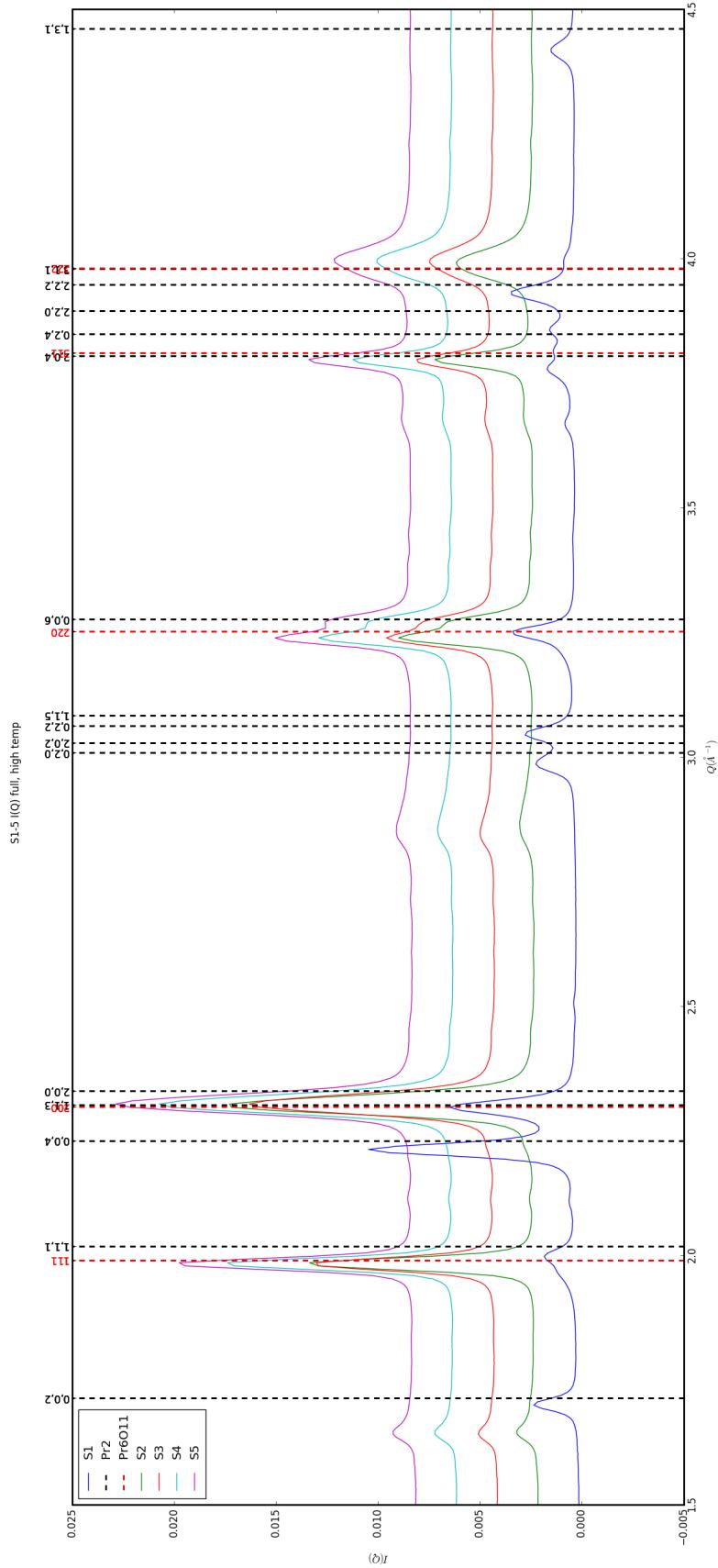
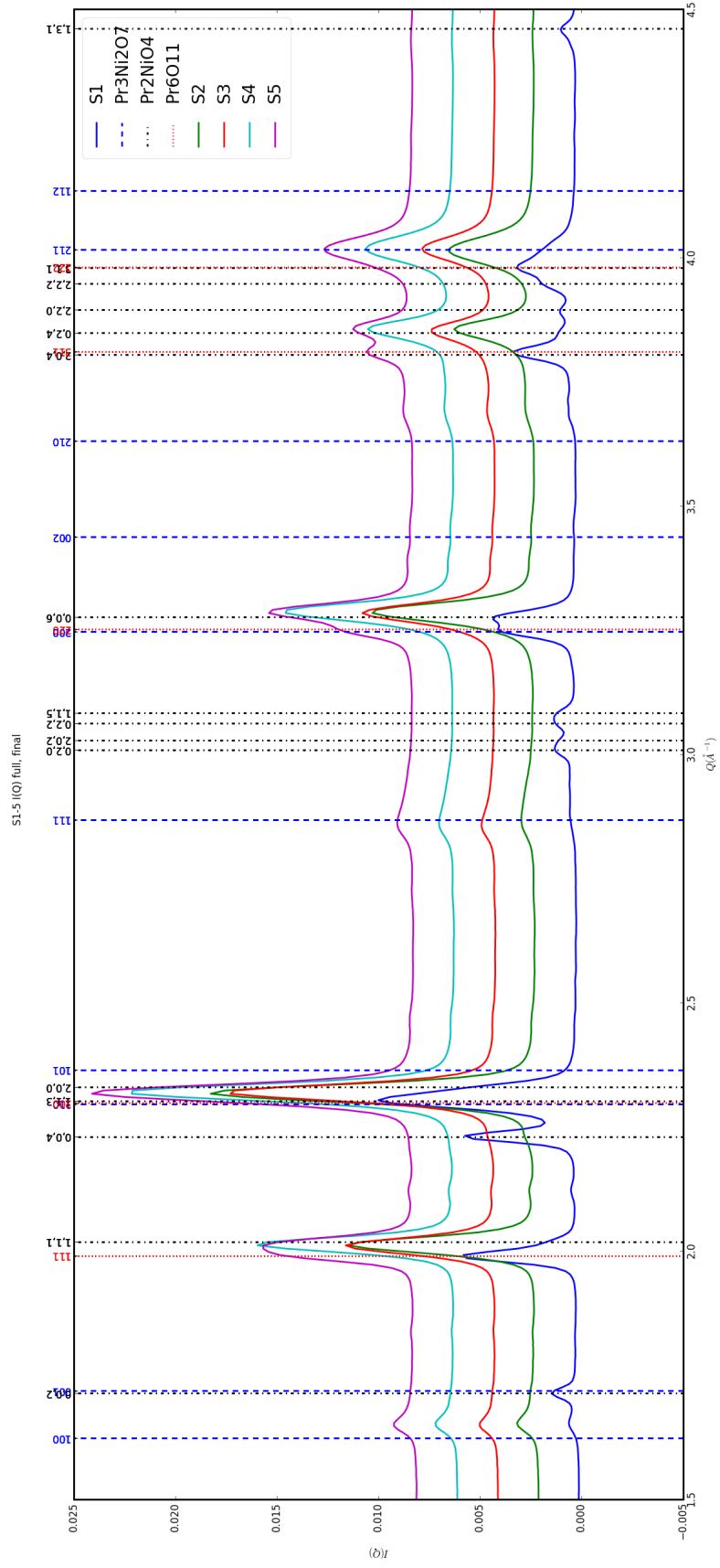


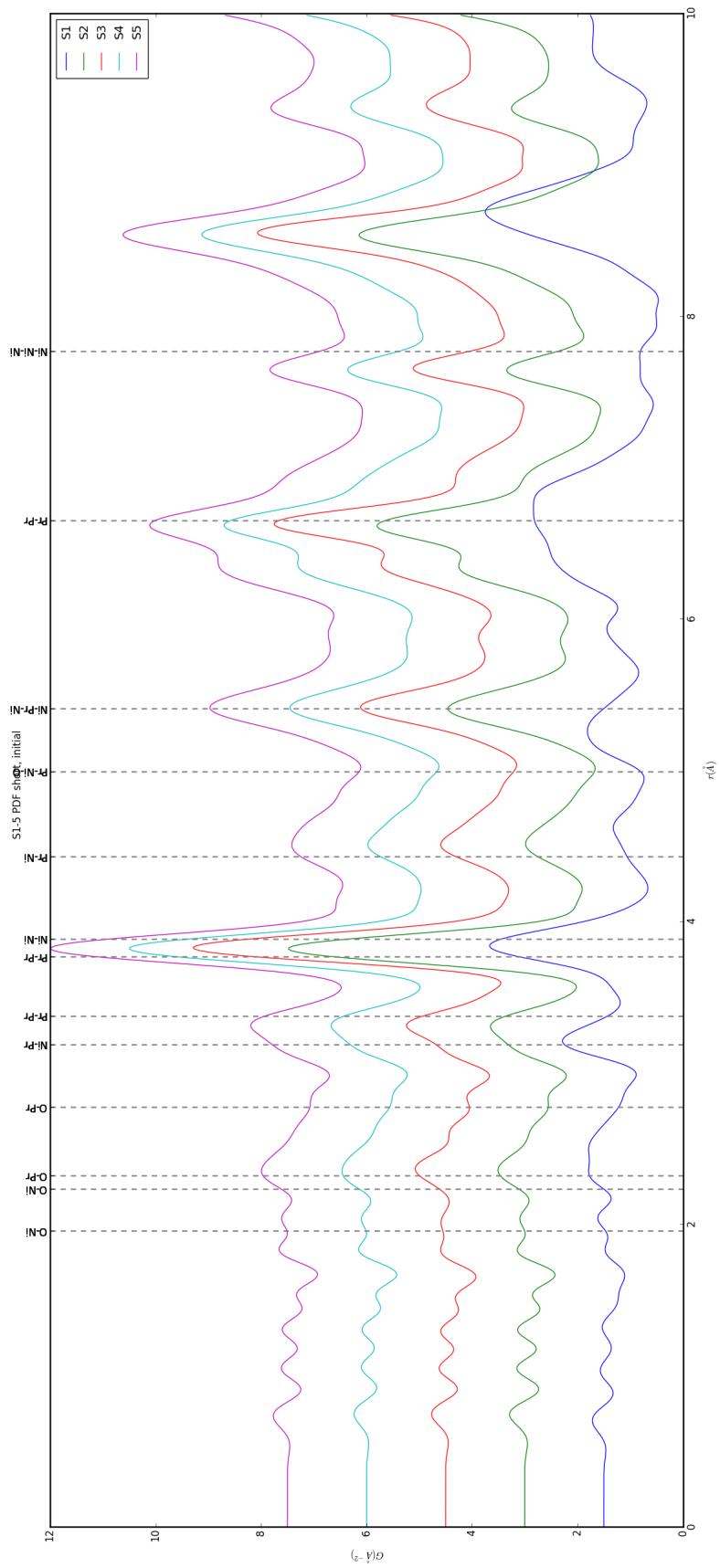
Figure 5.9: XRD as a function of temperature for S5

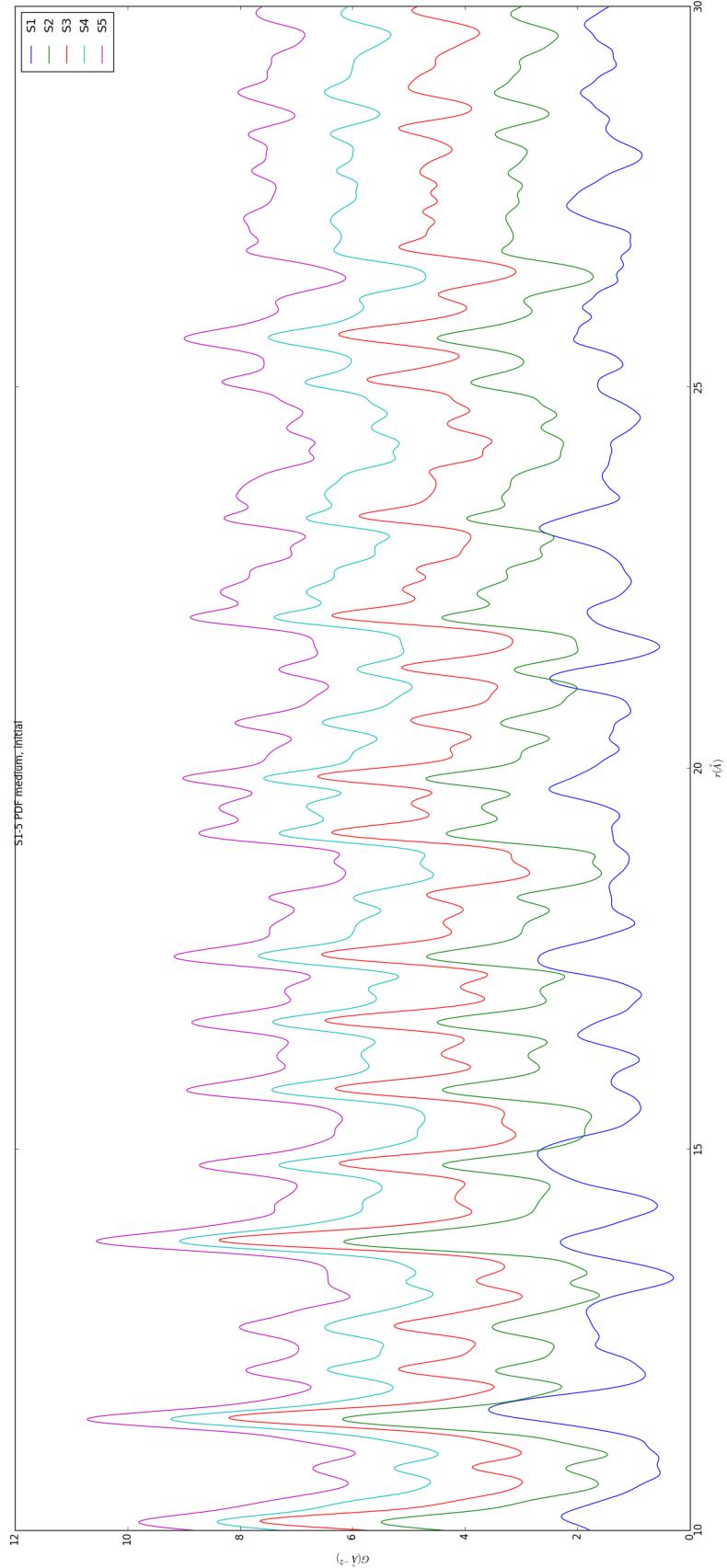
⁷⁹⁹ Inter Sample Comparison

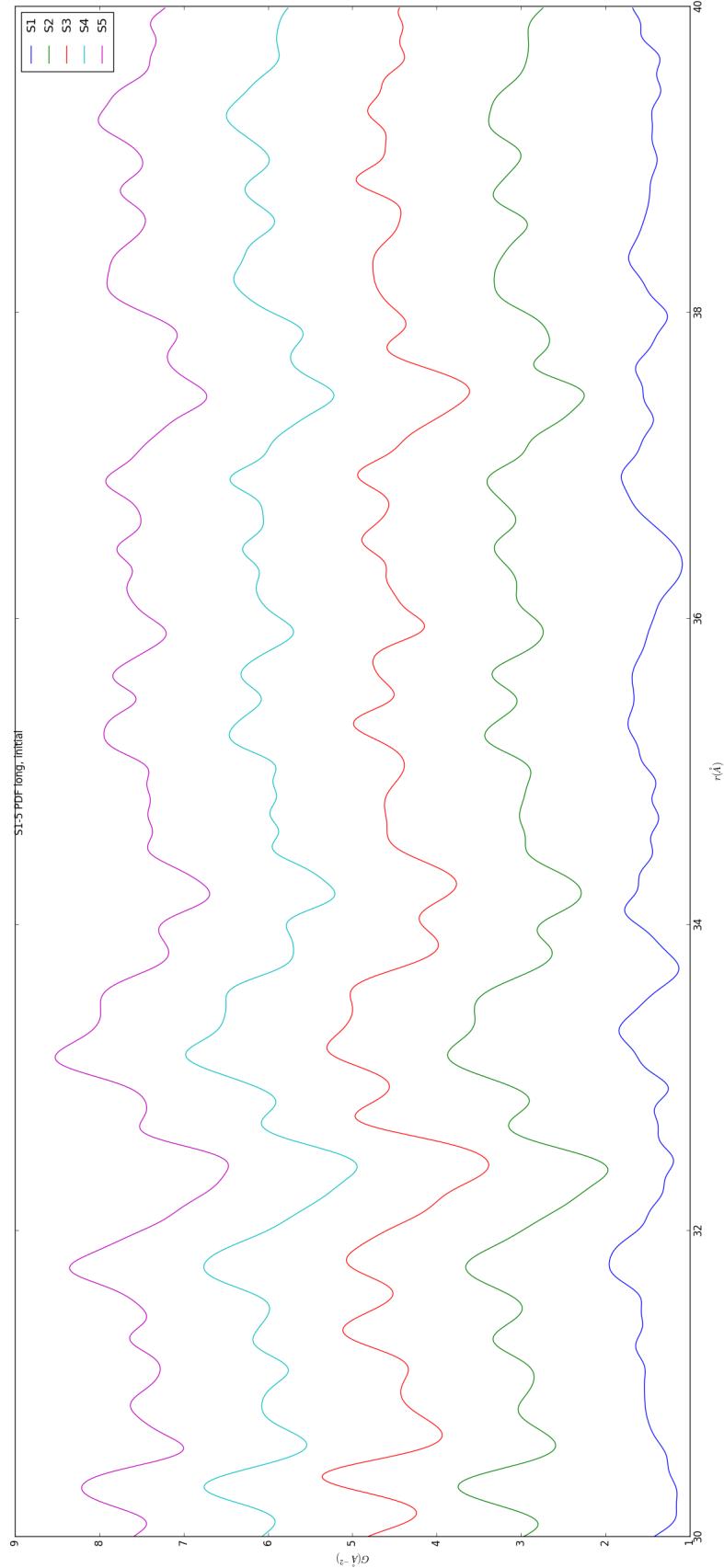


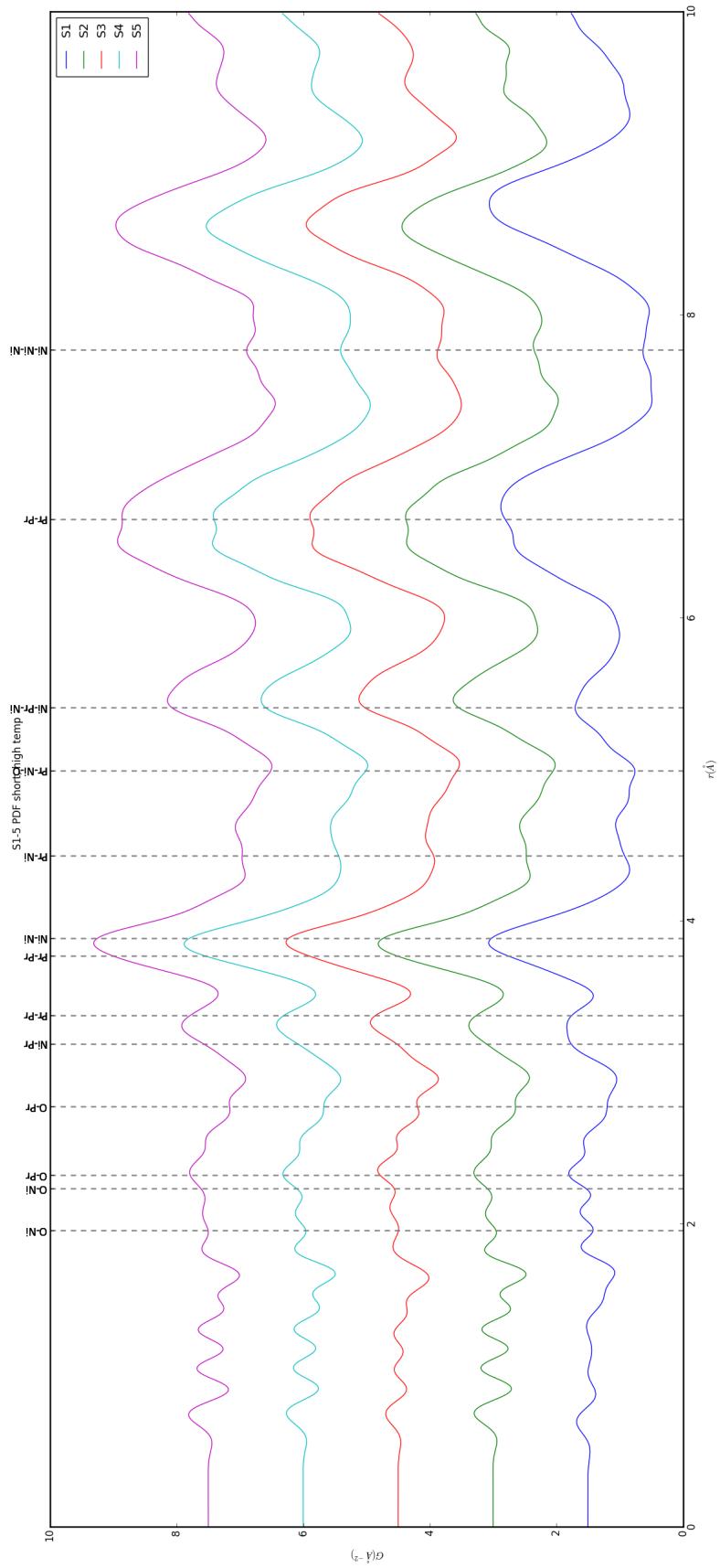


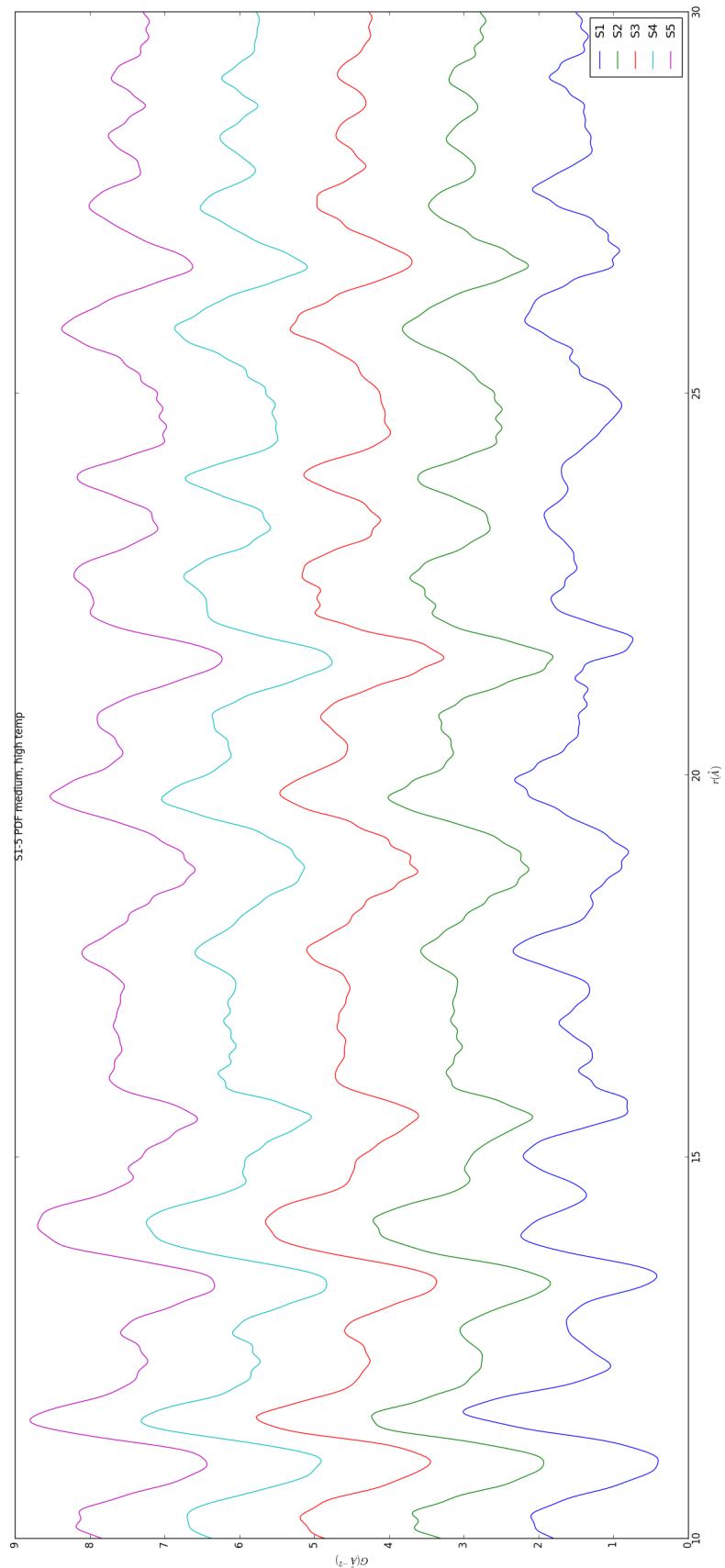


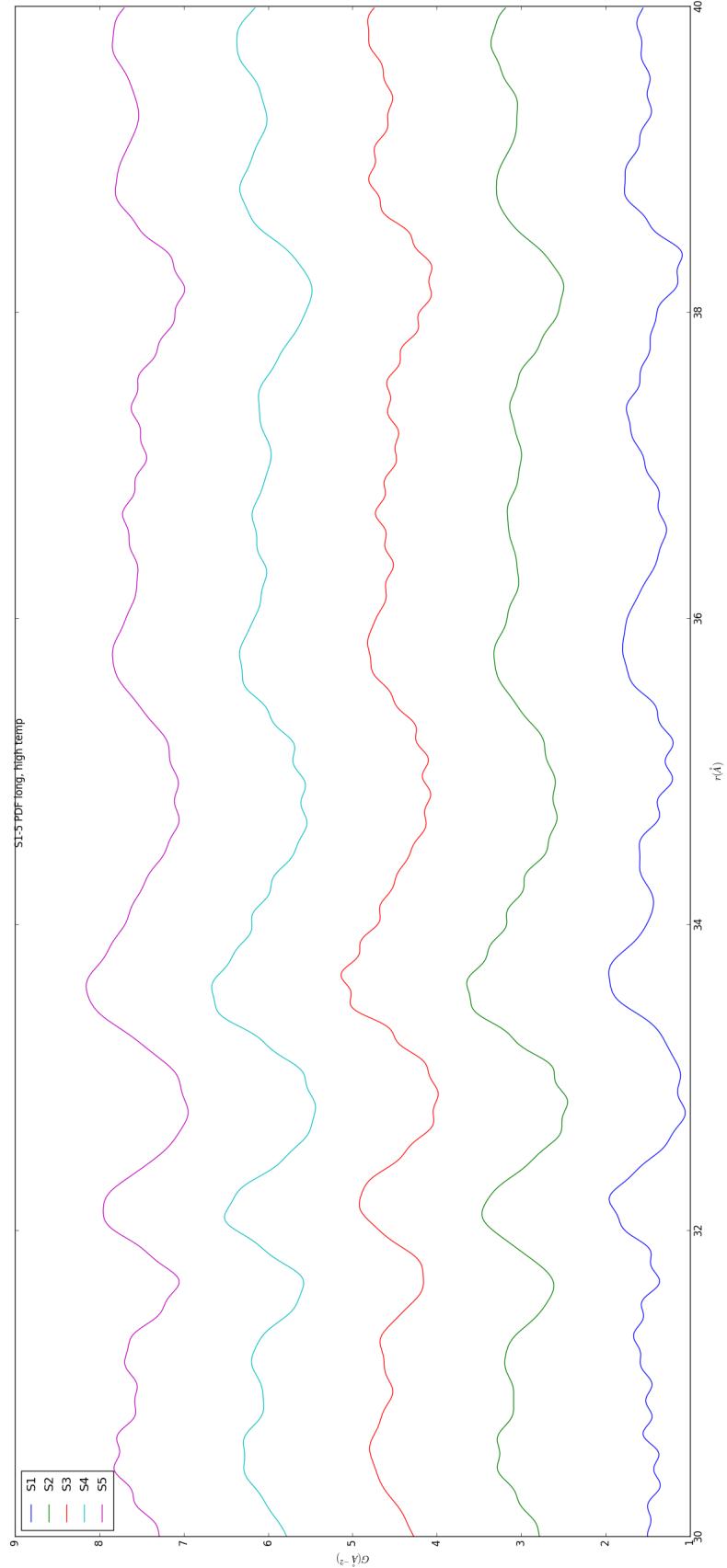


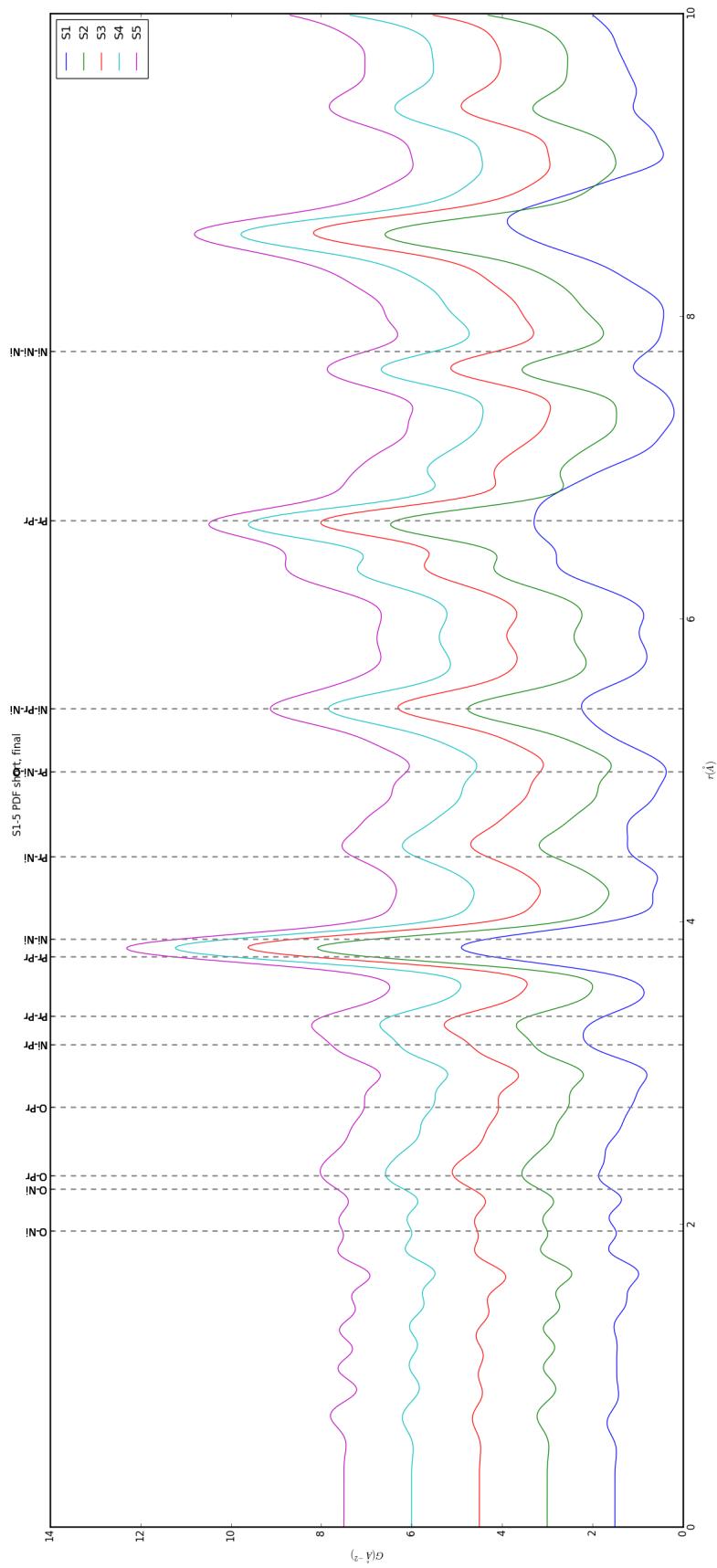


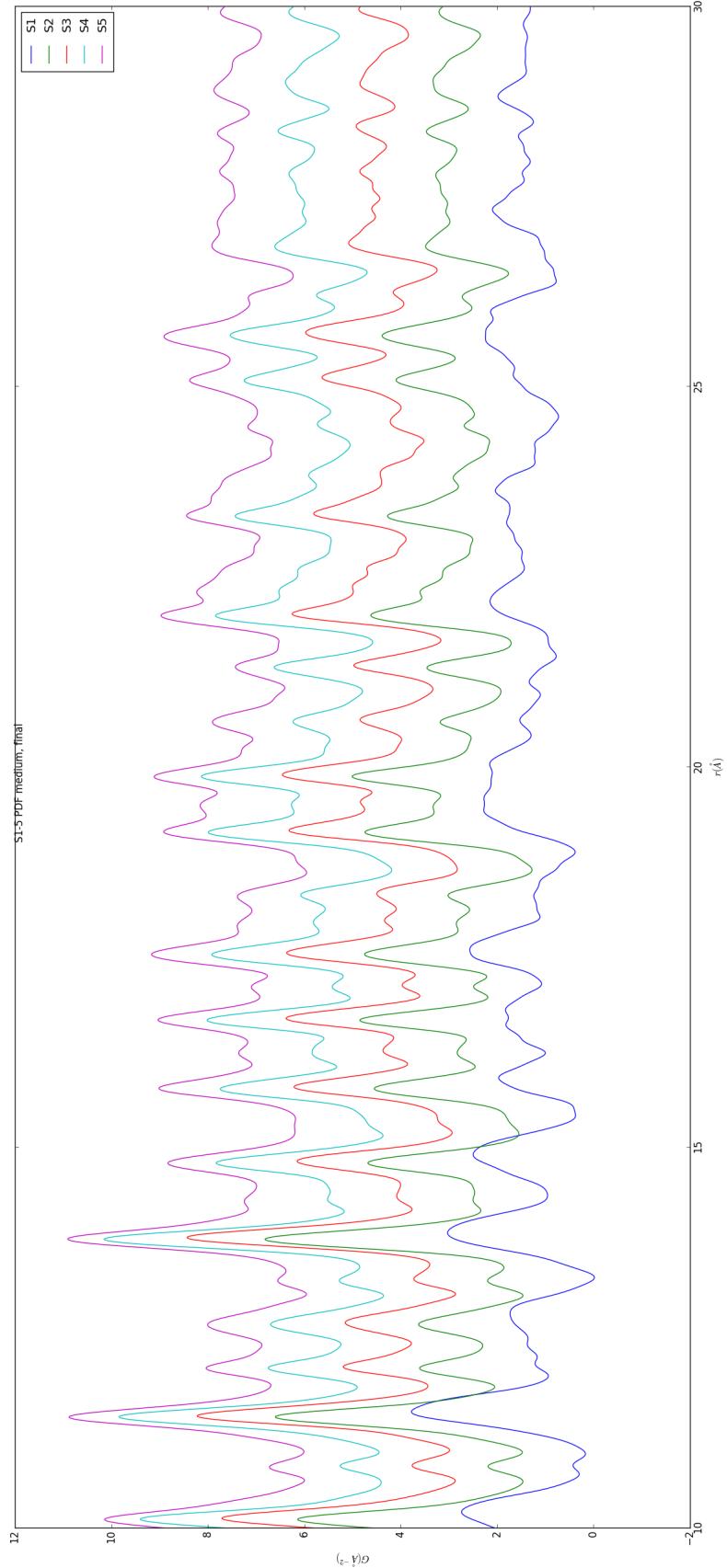


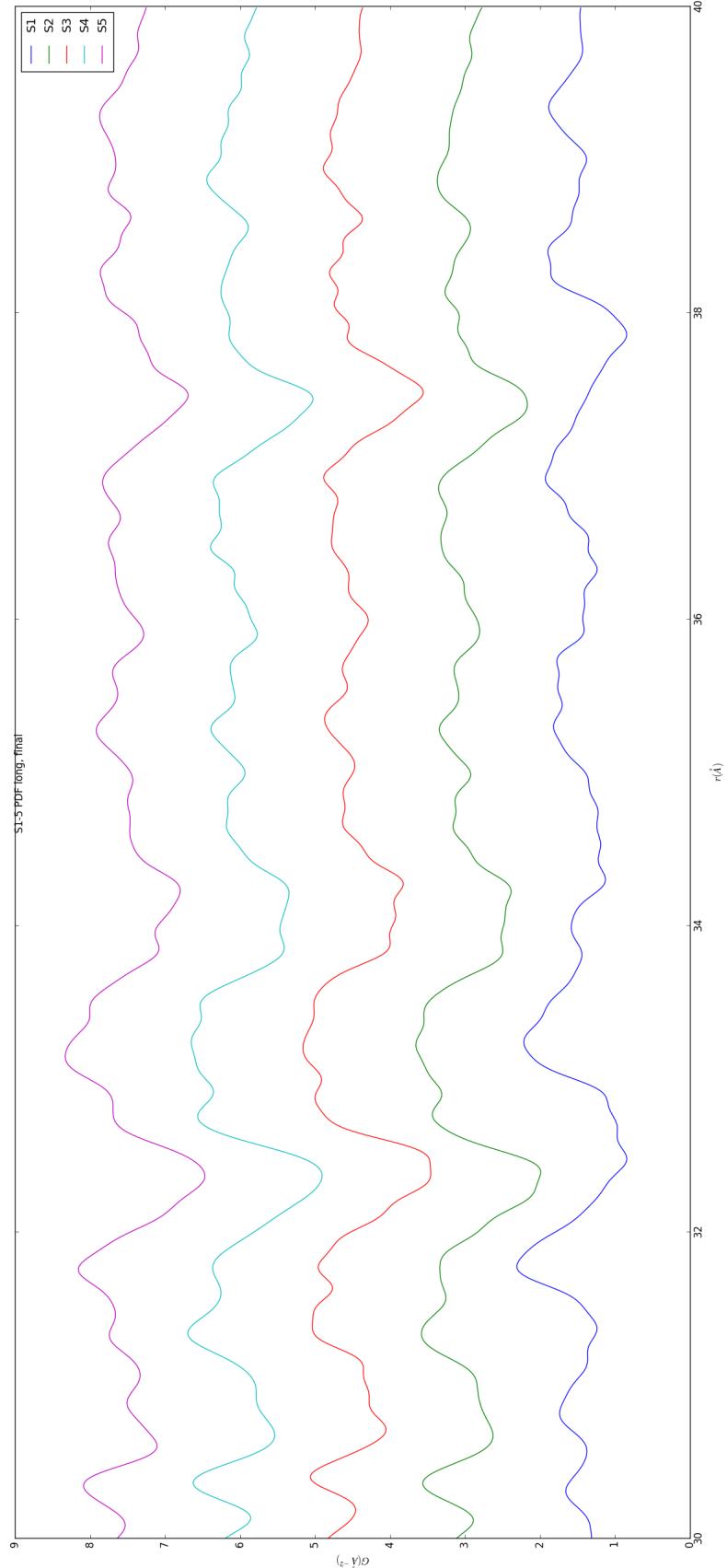












800 The PDF for S1 at operating temperature looks like S2-5 at the same temperature.

801 5.4 SIMULATION

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

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

814 5.5 CONCLUSIONS

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

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

834

CHAPTER 6

835

CONCLUSION

836

BIBLIOGRAPHY

- 837 [1] Simon J L Billinge and Takeshi Egami, *Underneath the Bragg Peaks: Structural
838 Analysis of complex Materials*, vol. Volume 16, Pergamon, 2012.
- 839 [2] P. E. Blöchl, *Projector augmented-wave method*, Physical Review B **50** (1994),
840 no. 24, 17953–17979.
- 841 [3] Gary K. Chen and Yunfei Guo, *Discovering epistasis in large scale genetic as-
842 sociation studies by exploiting graphics cards*, Frontiers in Genetics **4** (2013),
843 no. DEC, 1–12.
- 844 [4] Joshua J Choi, Xiaohao Yang, Zachariah M Norman, Simon J L Billinge, and
845 Jonathan S Owen, *Structure of methylammonium lead iodide within mesoporous
846 titanium dioxide: Active material in high-performance perovskite solar cells*,
847 Nano Letters **14** (2014), no. 1, 127–133.
- 848 [5] Peter J. Chupas, Karena W. Chapman, Charles Kurtz, Jonathan C. Hanson,
849 Peter L. Lee, and Clare P. Grey, *A versatile sample-environment cell for non-
850 ambient X-ray scattering experiments*, Journal of Applied Crystallography **41**
851 (2008), no. 4, 822–824.
- 852 [6] PJ Chupas, X Qiu, JC Hanson, PL Lee, CP Grey, and SJL Billinge, *Rapid-
853 acquisition pair distribution function (RA-PDF) analysis*, Journal of Applied
854 Crystallography **36** (2003), 1342–1347.
- 855 [7] Matthew J. Cliffe, Martin T. Dove, D. a. Drabold, and Andrew L. Goodwin,
856 *Structure determination of disordered materials from diffraction data*, Physical
857 Review Letters **104** (2010), no. 12, 1–4.
- 858 [8] Matthew J Cliffe and Andrew L Goodwin, *Nanostructure determination from the
859 pair distribution function: a parametric study of the INVERT approach.*, Journal
860 of physics. Condensed matter : an Institute of Physics journal **25** (2013), no. 45,
861 454218.
- 862 [9] Juarez L F Da Silva, Hyoung Gyu Kim, Maurício J. Piotrowski, Maurício J. Pri-
863 perto, and Germano Tremiliosi-Filho, *Reconstruction of core and surface nanopar-*

- 864 *ticles: The example of Pt 55 and Au55*, Physical Review B - Condensed Matter
865 and Materials Physics **82** (2010), no. 20, 1–6.
- 866 [10] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth, *Hybrid
867 Monte Carlo*, Physics Letters B **195** (1987), no. 2, 216–22.
- 868 [11] Timur Dykhne, Ryan Taylor, Alastair Florence, and Simon J L Billinge, *Data
869 requirements for the reliable use of atomic pair distribution functions in amor-
870 phous pharmaceutical fingerprinting.*, Pharmaceutical research **28** (2011), no. 5,
871 1041–8.
- 872 [12] C L Farrow, P Juhas, J W Liu, D Bryndin, E S Božin, J Bloch, Th Proffen, and
873 S J L Billinge, *PDFfit2 and PDFgui: computer programs for studying nanostruc-
874 ture in crystals.*, Journal of Physics. Condensed Matter : an Institute of Physics
875 journal **19** (2007), no. 33, 335219.
- 876 [13] Christopher L Farrow and Simon J L Billinge, *Relationship between the atomic
877 pair distribution function and small-angle scattering: implications for modeling
878 of nanoparticles.*, Acta Crystallographica Section A Foundations of Crystallog-
879 raphy **65** (2009), no. Pt 3, 232–9 (en).
- 880 [14] Riccardo Ferrando, Julius Jellinek, and Roy L Johnston, *Nanoalloys: From The-
881 ory to Applications of Alloy Clusters and Nanoparticles*, Chemical Reviews **108**
882 (2008), no. 3, 846–904.
- 883 [15] Md Matthew D. Hoffman and Andrew Gelman, *The No-U-Turn Sampler: Adap-
884 tively Setting Path Lengths in Hamiltonian Monte Carlo*, The Journal of Machine
885 Learning Research **15** (2014), no. 2008, 1593–1623.
- 886 [16] Pablo D Jadzinsky, Guillermo Calero, Christopher J Ackerson, David A Bushnell,
887 and Roger D Kornberg, *Structure of a thiol monolayer-protected gold nanoparti-
888 cle at 1.1 Å resolution.*, Science (New York, N.Y.) **318** (2007), no. 5849, 430–433.
- 889 [17] I. K. Jeong, R. H. Heffner, M. J. Graf, and S. J. L. Billinge, *Lattice dynamics
890 and correlated atomic motion from the atomic pair distribution function*, (2002),
891 9.
- 892 [18] P. Juhás, T. Davis, C. L. Farrow, and S. J. L. Billinge, *PDFgetX3 : a
893 rapid and highly automatable program for processing powder diffraction data into
894 total scattering pair distribution functions*, Journal of Applied Crystallography
895 **46** (2013), no. 2, 560–566 (en).

- 896 [19] Jérôme Kieffer and Dimitrios Karkoulis, *PyFAI, a versatile library for azimuthal*
897 *regrouping*, Journal of Physics: Conference Series **425** (2013), 202012.
- 898 [20] G. Kresse and J. Hafner, *Ab initio molecular dynamics for liquid metals*, Physical
899 Review B **47** (1993), no. 1, 558–561.
- 900 [21] ———, *Ab initio molecular-dynamics simulation of the liquid-*
901 *metal–amorphous-semiconductor transition in germanium*, Physical Review
902 B **49** (1994), no. 20, 14251–14269.
- 903 [22] Yan Li, Giulia Galli, and François Gygi, *Electronic structure of thiolate-covered*
904 *gold nanoparticles: Au₁₀₂(MBA)₄₄*, ACS Nano **2** (2008), no. 9, 1896–1902.
- 905 [23] L D Marks, *Experimental studies of small particle structures*, Reports on
906 Progress in Physics **57** (1994), no. 6, 603–649 (en).
- 907 [24] A. S. Masadeh, E. S. Božin, C. L. Farrow, G. Paglia, P. Juhas, S. J. L. Billinge,
908 A. Karkamkar, and M. G. Kanatzidis, *Quantitative size-dependent structure and*
909 *strain determination of CdSe nanoparticles using atomic pair distribution func-*
910 *tion analysis*, Physical Review B - Condensed Matter and Materials Physics **76**
911 (2007), no. 11, 115413.
- 912 [25] R L McGreevy and L Pusztai, *Reverse Monte Carlo Simulation: A New Tech-*
913 *nique for the Determination of Disordered Structures*, Molecular Simulation **1**
914 (1988), no. 6, 359–367.
- 915 [26] Radford M Neal, *Probabilistic Inference Using Markov Chain Monte Carlo Meth-*
916 *ods*, Intelligence **45** (1993), no. September, 144.
- 917 [27] Radford M. Neal, *MCMC Using Hamiltonian Dynamics*, Handbook of Markov
918 Chain Monte Carlo (Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li
919 Meng, eds.), Chapman and Hall/CRC, 2011, pp. 113–162.
- 920 [28] B R Pauw, *Corrigendum: Everything SAXS: small-angle scattering pattern col-*
921 *lection and correction (2013 J. Phys.: Condens. Matter 25 383201)*, Journal of
922 Physics: Condensed Matter **26** (2014), no. 23, 239501.
- 923 [29] John P. Perdew, Kieron Burke, and Matthias Ernzerhof, *Generalized Gradient*
924 *Approximation Made Simple*, Physical Review Letters **77** (1996), no. 18, 3865–
925 3868.

- 926 [30] Andrew A. Peterson, *Global optimization of adsorbate-surface structures while*
927 *preserving molecular identity*, Topics in Catalysis **57** (2014), no. 1-4, 40–53.
- 928 [31] Valeri Petkov, Binay Prasai, Yang Ren, Shiyao Shan, Jin Luo, Pharrah Joseph,
929 and Chuan-Jian Zhong, *Solving the nanostructure problem: exemplified on metal-*
930 *alloy nanoparticles*, Nanoscale **6** (2014), no. 17, 1–11.
- 931 [32] Valeri Petkov, Shiyao Shan, Peter Chupas, Jun Yin, Lefu Yang, Jin Luo, and
932 Chuan-Jian Zhong, *Noble-transition metal nanoparticle breathing in a reactive*
933 *gas atmosphere.*, Nanoscale **5** (2013), no. 16, 7379–87.
- 934 [33] Th. Proffen and R. B. Neder, *DISCUS: a Program for Diffuse Scattering and*
935 *Defect-Structure Simulation*, Journal of Applied Crystallography **30** (1997), 171–
936 175.
- 937 [34] Erin L. Redmond, Brian P. Setzler, Pavol Juhas, Simon J. L. Billinge, and
938 Thomas F. Fuller, *In-Situ Monitoring of Particle Growth at PEMFC Cathode*
939 *under Accelerated Cycling Conditions*, Electrochemical and Solid-State Letters
940 **15** (2012), no. 5, B72 (en).
- 941 [35] H. W. Sheng, M. J. Kramer, A. Cadien, T. Fujita, and M. W. Chen, *Highly*
942 *optimized embedded-atom-method potentials for fourteen FCC metals*, Physical
943 Review B - Condensed Matter and Materials Physics **83** (2011), no. 13, 134118.
- 944 [36] X. Yang, P. Juhás, and S. J L Billinge, *On the estimation of statistical uncertain-*
945 *ties on powder diffraction and small-angle scattering data from two-dimensional*
946 *X-ray detectors*, Journal of Applied Crystallography **47** (2014), no. 4, 1273–1283.