

SOLVING ATOMIC STRUCTURE 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

Thomas Vogt, Committee Member

Mark Uline, Committee Member

Jochen Lauterbach, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

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

## DEDICATION

## ACKNOWLEDGMENTS

# ABSTRACT

# TABLE OF CONTENTS

DEDICATION . . . . .	iii
ACKNOWLEDGMENTS . . . . .	iv
ABSTRACT . . . . .	v
LIST OF TABLES . . . . .	ix
LIST OF FIGURES . . . . .	x
CHAPTER 1 ATOMIC STRUCTURE: EXTRACTION AND APPLICATION . . .	2
1.1 Atomistic Goals . . . . .	2
1.2 Atomistic Experiments . . . . .	3
1.3 Atomistic Simulations . . . . .	3
CHAPTER 2 STATISTICAL MECHANICAL ENSEMBLES AND POTENTIAL ENERGY SURFACES . . . . .	4
2.1 Introduction . . . . .	4
2.2 Potential Energy Surfaces . . . . .	4
2.3 Ensembles . . . . .	6
CHAPTER 3 ATOMIC PAIR DISTRIBUTION FUNCTION: THEORY AND COMPUTATION . . . . .	11
3.1 Theory . . . . .	11

3.2	Computation . . . . .	11
3.3	Experiment . . . . .	14
3.4	Data Processing Workflow . . . . .	14
CHAPTER 4 BENCHMARKING . . . . .		25
4.1	PDF . . . . .	25
4.2	PDF with ADPs . . . . .	25
CHAPTER 5 ANNEALING AND AGGREGATION OF 2NM AU NANOPARTICLES . . . . .		26
5.1	Experiments . . . . .	26
5.2	Data Processing . . . . .	26
5.3	Data Analysis . . . . .	26
5.4	Simulation . . . . .	26
5.5	Structural Analysis . . . . .	26
5.6	Conclusions . . . . .	26
CHAPTER 6 PHASE CHANGES AND ANNEALING DYNAMICS OF $\text{Pr}_2\text{NiO}_4$ AND ITS DERIVATIVES . . . . .		27
6.1	Experiments . . . . .	27
6.2	Data Processing . . . . .	27
6.3	Data Analysis . . . . .	27
6.4	Simulation . . . . .	27
6.5	Structural Analysis . . . . .	27
6.6	Conclusions . . . . .	27

CHAPTER 7	CONCLUSION . . . . .	28
-----------	----------------------	----



## LIST OF TABLES

## LIST OF FIGURES

Figure 3.1	Database Loading Workflow. Data is loaded from various sources, including images and text files, into the FileStore and Meta-dataStore databases. Data is then retrieved from the databases using the databroker. . . . .	15
Figure 3.2	$Q$ resolution as a function of $Q$ . . . . .	16
Figure 3.3	Generated beamstop masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	19
Figure 3.4	Generated beamstop masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	20
Figure 3.5	Generated beamstop masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	21
Figure 3.6	Generated beamstop masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels . . . . .	22
Figure 3.7	Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the poorly binned mask and b) the properly binned mask	23
Figure 3.8	Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the poorly binned mask and b) the properly binned mask	23
Figure 3.9	Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the poorly binned mask and b) the properly binned mask	24
Figure 3.10	Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the poorly binned mask and b) the properly binned mask	24

1

## INTRODUCTION

2 This is the introduction to the thesis.

3

## CHAPTER 1

### 4 ATOMIC STRUCTURE: EXTRACTION AND APPLICATION

#### 5 1.1 ATOMISTIC GOALS

6 The only way to truly understand the fundamental source of material and chemical  
7 properties is through atomic structure. The goal of atomistic engineering is to pro-  
8 duce novel structures and combinations of structures to engender new properties and  
9 functions. This includes producing stronger materials, more durable catalysts, more  
10 energy dense batteries, and many more engineering applications. The true power of  
11 atomistic engineering has been shown in biochemistry and pharmaceutical design. Al-  
12 though the production of drugs and biomedical treatments is usually considered to be  
13 rather far from the field of catalyst design and materials science, the atomistic nature  
14 of these fields can not be denied. The field of protein structural analysis stands  
15 as an example of structural science, elucidating the three dimensional coordinates of  
16 thousands of atoms. These structures are then used to describe how the molecular  
17 machinery of the biological world works, enabling the development of new drugs and  
18 treatments for diseases and a deeper understanding of how we evolved. The develop-  
19 ment of protein inhibitor drugs, which are important to so many treatments, would  
20 have not been possible without very detailed atomic structures. The aspiration of  
21 this work is to create this level of accuracy and utility, generating structures which  
22 allow for the understanding of how materials work on a fundamental level.

## 23 1.2 ATOMISTIC EXPERIMENTS

### 24 **Single Crystal Diffraction**

### 25 **Electron Microscopy**

### 26 **X-ray Total Scattering**

## 27 1.3 ATOMISTIC SIMULATIONS

28 The goals of atomistic simulations are usually to produce atomic structures from  
29 quantum mechanical first principles, as in the case of Density Functional Theory  
30 (DFT), or classical approximations to quantum mechanics.

### 31 **Density Functional Theory**

### 32 **Classical Force Field**

### 33 **Monte Carlo and Statistical Mechanics**

34 Maybe put the ensemble and PES work here, since it is more general than the PDF  
35 per say. Also the rational for the gradients and fast computation make much more  
36 sense knowing we are going to be very sample happy and follow the gradient of the  
37 PES.

38

## CHAPTER 2

39

## STATISTICAL MECHANICAL ENSEMBLES AND

40

## POTENTIAL ENERGY SURFACES

41

### 2.1 INTRODUCTION

42

The approach taken in this work for solving the atomic structures of materials is

43

one of optimization. The positional variables of the system are optimized so as to

44

minimize the value of a potential energy surface (PES). The

45

### 2.2 POTENTIAL ENERGY SURFACES

46

A PES simply describes the potential energy of the system as a function of all its

47

relevant coordinates in phase space, essentially providing a mapping  $\mathbb{R}^n \rightarrow \mathbb{R}$ . Usually

48

these coordinates are the positions of the atoms  $q$  and their conjugate momenta  $p$ .

49

Note that there could be more variables associated with the system, for instance the

50

magnetic moments of the atoms could play a role in describing the system. In this

51

magnetic system there would be positional variables for the atomwise spin vectors

52

and their "momenta". Application of the term "momenta" might seem odd here, as

53

the magnetic spin does not have a mass or a velocity. However, since the magnetic

54

"position" is defined on the PES we need to describe its conjugate variable to properly

55

formulate Hamiltonian dynamics and the kinetic portion of the PES.

## 56 Experimentally Derived Potential Energy Surfaces

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

## 65 Potentials

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

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

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

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

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

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

## 74 Forces

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

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

$$\frac{\partial \alpha}{\partial \gamma_{i,w}} = \frac{(\sum_{a=a_{\min}}^{a_{\max}} A_{\text{obs}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}} - 2\alpha \sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}} \frac{\partial A_{\text{calc}}}{\partial \gamma_{i,w}})}{\sum_{a=a_{\min}}^{a_{\max}} A_{\text{calc}}^2} \quad (2.7)$$

$$\vec{\nabla} \chi_{\text{INVERT}}^2 = \frac{-2}{N} \sum_{a=a_{\min}}^{a_{\max}} \sum_j \left( \alpha \frac{\partial A_{j\text{calc}}}{\partial \gamma_{i,w}} + A_{j\text{calc}} \frac{\partial \alpha}{\partial \gamma_{i,w}} \right) (A_{\text{obs}} - \alpha A_{j\text{calc}}) \quad (2.8)$$

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

DISCUSS INVERT A BUNCH. ALSO COMPARE RW AND CHI\*\*2, POTENTIALY WITH A FIGURE.

### 2.3 ENSEMBLES

While PESs describe which atomic configurations are the most desirable and how the atoms would like to get there, the ensemble describes how the atoms move on the PES. The abstraction of the PES from the ensemble is an important one, as it allows for the reuse and exchange of both PESs and ensembles for a wide array of problems. Statistical mechanical ensembles can be described in two ways, analytically and stochastically. For long simulation times and fine enough numerical or analytical integration these two descriptions should be identical. In either case one starts by defining the Hamiltonian  $\mathcal{H}$  as the total energy of the system. Thus, the Hamiltonian is described as the sum of the potential  $U(q)$  and kinetic  $K(p)$  energies, where  $q$  is the positions of the atoms and  $p$  is their momenta

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



96 where  $K(p) = \frac{1}{2} \sum_i \frac{p_i^2}{m_i}$  and  $i$  denotes the  $i$ th particle. Analytically one generally defines  
 97 a partition function, which describes the sum of probabilities over all potential atomic  
 98 states.

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

99 where  $P_i$  is the probability of the  $i$ th state and is a function of the total energy of  
 100 that state. This partition function can then be used to obtain the probability of any  
 101 specific state.

## 102 Hamiltonian Monte Carlo

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

$$\frac{dq_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} = p_i \quad (2.10)$$

$$\frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial q_i} = -\vec{\nabla} U \quad (2.11)$$

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

$$p_i(t + \delta t/2) = p_i(t) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t)) \quad (2.12)$$

$$q_i(t + \delta t) = q_i(t) + \delta t * p_i(t + \delta t/2) \quad (2.13)$$

$$p_i(t + \delta t) = p_i(t + \delta t/2) - \frac{\delta t}{2} \frac{\partial}{\partial q_i} U(q(t + \delta t)) \quad (2.14)$$

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

$$\vec{\nabla} U = \begin{bmatrix} \frac{\partial U}{\partial q_{0,x}} & \frac{\partial U}{\partial q_{0,y}} & \frac{\partial U}{\partial q_{0,z}} \\ \vdots & \frac{\partial U}{\partial q_{i,w}} & \vdots \\ \frac{\partial U}{\partial q_{n,x}} & \frac{\partial U}{\partial q_{n,y}} & \frac{\partial U}{\partial q_{n,z}} \end{bmatrix} = \begin{bmatrix} \vec{\mathcal{F}}_0 \\ \vdots \\ \vec{\mathcal{F}}_i \\ \vdots \\ \vec{\mathcal{F}}_n \end{bmatrix} \quad (2.15)$$

105 where  $\frac{\partial}{\partial q_{i,w}}$  is the derivative with respect to  $q$  where  $w$  denotes direction of the deriva-  
 106 tive ( $x$ ,  $y$ , or  $z$ ),  $n$  is the number of atoms and  $U$  is the potential which depends on  
 107  $q$ , and  $\vec{\mathcal{F}}_i$  is the "force" on the  $i$ th atom.

## 108 **No-U-Turn-Sampling**

## 109 **Grand Canonical Ensemble**

### 110 **Ensemble description**

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

$$\Xi = e^{-\beta(\mathcal{H}+\mu)} \quad (2.16)$$

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

### 117 **Grand Canonical Monte Carlo**

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

$$\min[1, \frac{V}{(N+1)\Lambda(T)^3} e^{-\beta\Delta U + \beta\mu}] \quad (2.17)$$

121 Similarly the removal of an atom has the probability:

$$\min[1, \frac{(N)\Lambda(T)^3}{V} e^{-\beta\Delta U - \beta\mu}] \quad (2.18)$$

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

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

### **GCMC biasing**

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

$$P_{i,j,k} = \frac{xyz}{abc} \tag{2.19}$$

149 where  $x, y, z$  and  $a, b, c$  are the resolutions and cell side lengths in the cardinal di-  
 150 rections, respectively. While this does help to limit the total probability space it  
 151 does not tell us which voxels are likely to lead to better configurations, leading to  
 152 many rejected atomic additions. To combat this issue we can weigh the individual  
 153 voxels, giving more probability to voxels which show promise and less to those with  
 154 less likelihood to be accepted.

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

$$P_{i,j,k} = \frac{e^{\beta \Delta U_{i,j,k}}}{\sum_{i,j,k} e^{\beta \Delta U_{i,j,k}}} \quad (2.20)$$

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

163

## CHAPTER 3

164

### ATOMIC PAIR DISTRIBUTION FUNCTION:

165

### THEORY AND COMPUTATION

#### 166 3.1 THEORY

167 To properly understand the PDF and its limitations we need to derive its mathemat-  
168 ics. The following derivation has been performed numerous times but most recently  
169 and completely by Farrow and Billinge, it is reproduced here for clarity and com-  
170 pleteness.

#### 171 **Derivation**

172 Consider a wave incident on a volume of variable density...

#### 173 **Analytical Gradients**

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

#### 178 3.2 COMPUTATION

179 Simply deriving the equations for the PDF is not enough. The many body nature of  
180 the PDF equation make analytical solution of the structure from the PDF impossible.

181 Thus, the PDF must be computed from a structural candidates and compared against  
182 experimental results to evaluate the reliability of the model.

## 183 **HPC and GPUs**

184 To properly solve the structure of materials the PDF will need to be computed many  
185 times and checked against experimental results. This requires computation of the  
186 PDF, potentially over many atoms. Calculating these PDFs requires a fast, highly  
187 parallized, computational framework.

## 188 **GPUs and Parallelization**

189 Computing the PDF is an embarrassingly parallel problem. The basic procedure is  
190 to calculate the reduced structure factor  $F(Q)$  for each atom pair and momentum  
191 transfer vector, sum over all the atom pairs, and Fourier transform the structure to  
192 the PDF. The first part of this procedure is perfectly parallizable, as each atom pair is  
193 seperate from the others. The summation over all the atomic reduced structure factors  
194 can be parallelized via distributed summing. Lastly the FFT can be parallelized using  
195 existing parellel FFT algorithms.

196 GPUs are particularly well suted to the task of computing PDFs. GPU chip  
197 architecture is designed to perform many task simultaneously by having potentially  
198 thousands of cores.

## 199 **Map from $ij$ space to $k$ space**

200 The above equations, although formally correct, are very ineffiecent.  $F(Q)$  and its  
201 gradient are indexed over all the atoms twice, however there are symmetries that  
202 allow us to only compute over the atom pairs esentially mapping from an  $n \times n$  space,  
203  $ij$  space, to a  $\frac{n(n-1)}{2}$  space,  $k$  space. For  $F(Q)$  we apply the following mapping where  
204  $E$  denotes the atomic coordinates in  $ij$  space,  $E'$  denotes  $F(Q)$  before the summation

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

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

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

For the gradient a similar mapping is used:

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

219

220 In this mapping, however, we use the  $\tilde{\phi}\Sigma$  operator. This operator simultaneously  
 221 performs a reverse mapping from  $k$  to  $ij$  space, and a summation with the correct

222 symmetry. In this case the  $\psi$  and  $\psi'$  operators, which denote the  $\vec{\nabla}F(Q)$  operator  
 223 in  $ij$  and  $k$  space, are antisymmetric. Intuitively this makes sense as an extension of  
 224 Newton's Second Law, since each particle's interaction is felt oppositely by its partner.

## 225 Periodic Boundary Conditions

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

$$F(Q) = \frac{1}{N\langle f \rangle^2} \left( \sum_{j \neq i} f_i^*(Q) f_j(Q) \frac{\sin(Qr_{ij})}{r_{ij}} + \sum_{i,j} f_i^*(Q) f_j(Q) \frac{\sin(QR_{ij})}{R_{ij}} \right) \quad (3.1)$$

232 where

$$R = |\vec{r} + \vec{u}| \quad (3.2)$$

$$\vec{u} = \gamma_1 * \vec{a} + \gamma_2 * \vec{b} + \gamma_3 * \vec{c} \quad (3.3)$$

## 233 3.3 EXPERIMENT

234 PDF experiments are generally performed at synchrotron light sources, as only these  
 235 sources can provide the need flux, energy, and high momentum transfer vectors needed  
 236 to obtain reliable PDFs.

## 237 3.4 DATA PROCESSING WORKFLOW

238 Processing the raw pixel intensities to the PDF is very important as we are extracting  
 239 most of our interesting information out of very high  $Q$  data. This data relies on good  
 240 statistics and sound background subtraction. Talk about papers from Billinge Group



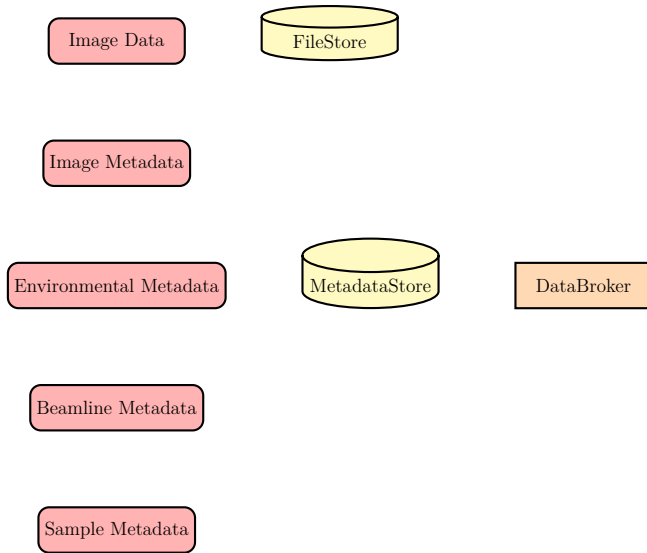


Figure 3.1: Database Loading Workflow. Data is loaded from various sources, including images and text files, into the FileStore and MetadataStore databases. Data is then retrieved from the databases using the databroker.

241 with thin film PDF and dilute NP solutions. Diagram of the overall data processing  
 242 workflow. Discuss the NSLS-II data stack.

## 243 MetadataStore Side Loading

244 Design of sidewinder-spec for loading the data into metadatastore. Most of the design  
 245 considerations went into the loaders, which are different for each experiment.

## 246 Automated Image Azimuthal Integration

247 Mux data as needed. Use pyFAI to get the radial distance array. Note that to  
 248 properly mask and integrate the system we need to compute the bin edges for the  
 249 pixels. The bin edges change as a function of  $Q$ , as the angle subtended by a pixel  
 250 shrinks essentially giving high  $Q$  pixels more resolution.

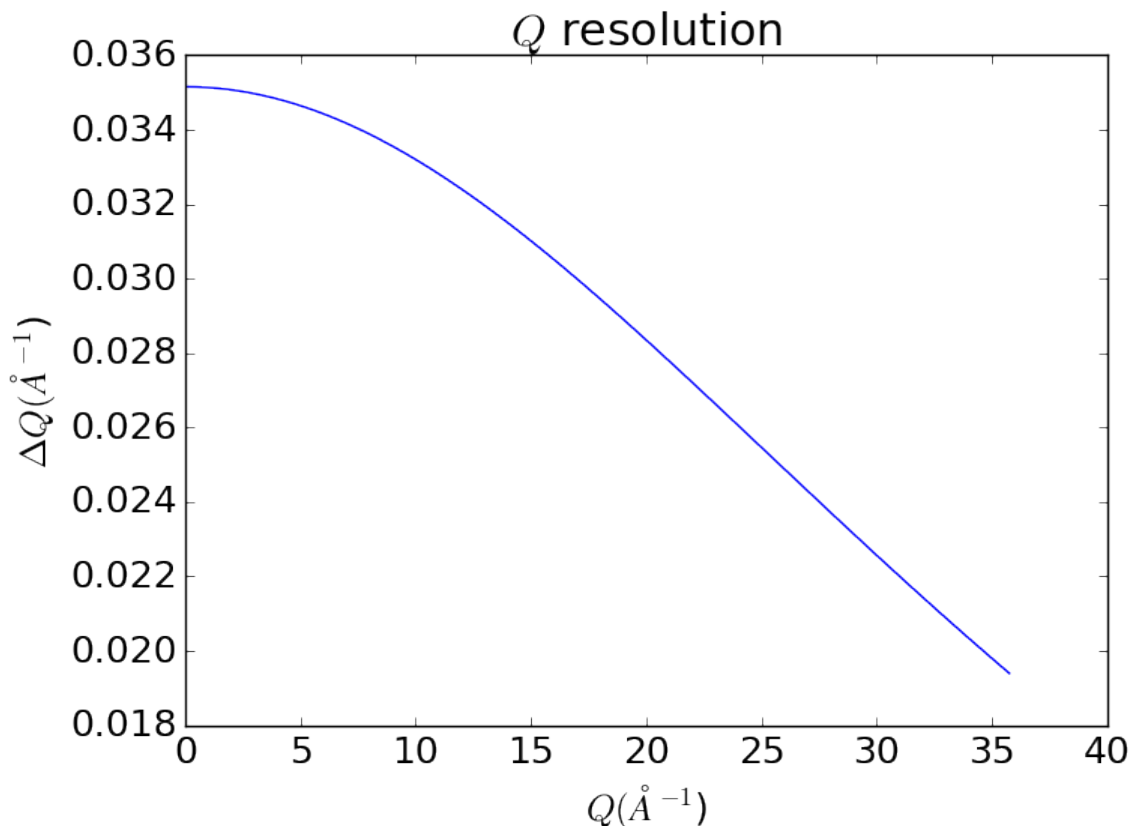


Figure 3.2:  $Q$  resolution as a function of  $Q$ .

## 251 **Detector $Q$ resolution**

## 252 **Automated Mask Generation**

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

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

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

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

$$I = 100 \cos(50r)^2 + 150 \quad (3.4)$$

where  $r$  is a pixel's distance from the beam point of incidence. The positions of the dead/hot pixels were chosen at random as was the dead or hot nature of the defect. The beamstop was positioned at the vertical center of the detector with an initial width of 60 pixels and final width of 120 pixels. The height of the beamstop was 1024 pixels. The beamstop was calculated to attenuate the x-ray scattering

288 signal at various transmittance, as various beamstop holder materials have different  
289 transmittance.

290 Figures 3.3-3.10 show the results of the masking algorithm on simulated images.

291 Note that we do miss some pixels when the number of dead pixels grows very  
292 large. However, most detectors do not have that many dead pixels. We can run into  
293 these kinds of situations when running samples with some single crystal or textured  
294 components. However, when this is the case the contrast between the affected pixels  
295 and the desired signal is very large enabling easier masking.

296 Additionally the standard deviation threshold can be a function of the pixel dis-  
297 tance from the center, allowing the mask generator to be more forgiving at certain  
298 points and stricter at others. This is particularly helpful as the small number of  
299 pixels near the point of incidence combined with the very sharp peaks causes some  
300 pixels to be improperly masked. Similarly it is important to remove dead pixels at  
301 the edge of the detector as these have an outsized effect on the integration as the pixel  
302 intensity is low to begin with. In practice this results in the removal of almost all  
303 dead pixels and potentially the beamstop holder. Removal of the holder depends on  
304 its individual properties, since a holder which is more x-ray opaque will cause a larger  
305 shift in the pixel intensity distribution. The method was benchmarked on synthetic  
306 data, with both hot and cold pixels added. Additional benchmarking was performed  
307 with synthetic beamstop holders of various x-ray transmittance. Anomalous corner  
308 masking most likely due to the small number of pixels out at the corners.

309 ALSO NEED ROTATED BEAMSTOP MASKS

310 ALSO ALSO NEED TO SHOW THE EFFECT OF CHANGING ALPHA FOR  
311 THE MASKING

312 ALSO ALSO ALSO SHOW SOME MASKS OF REAL DATA, INCLUDING  
313 DATA WITH SINGLE CRYSTAL/TEXTURE

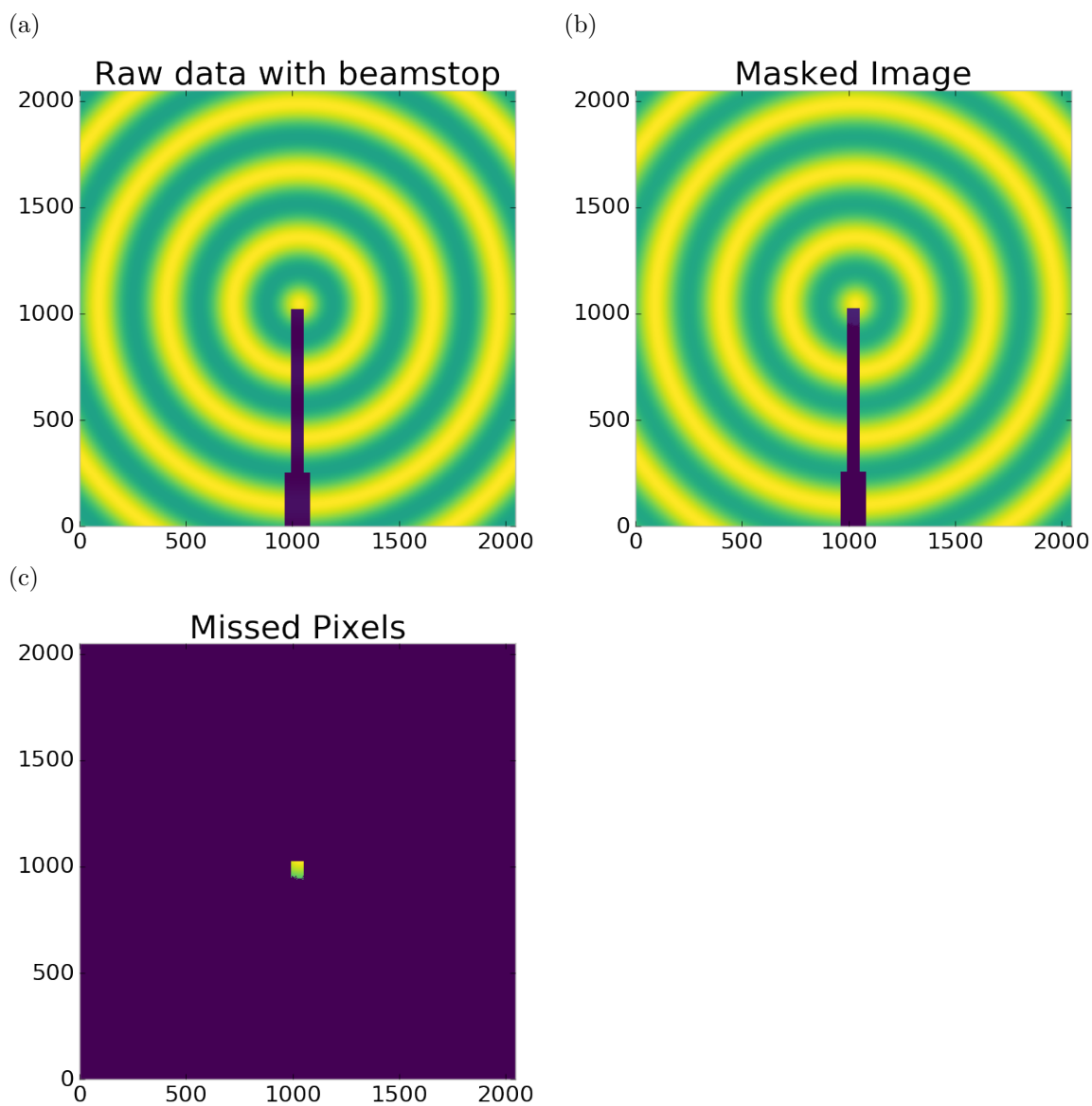


Figure 3.3: Generated beamstop masks for a beamstop holder with 10% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

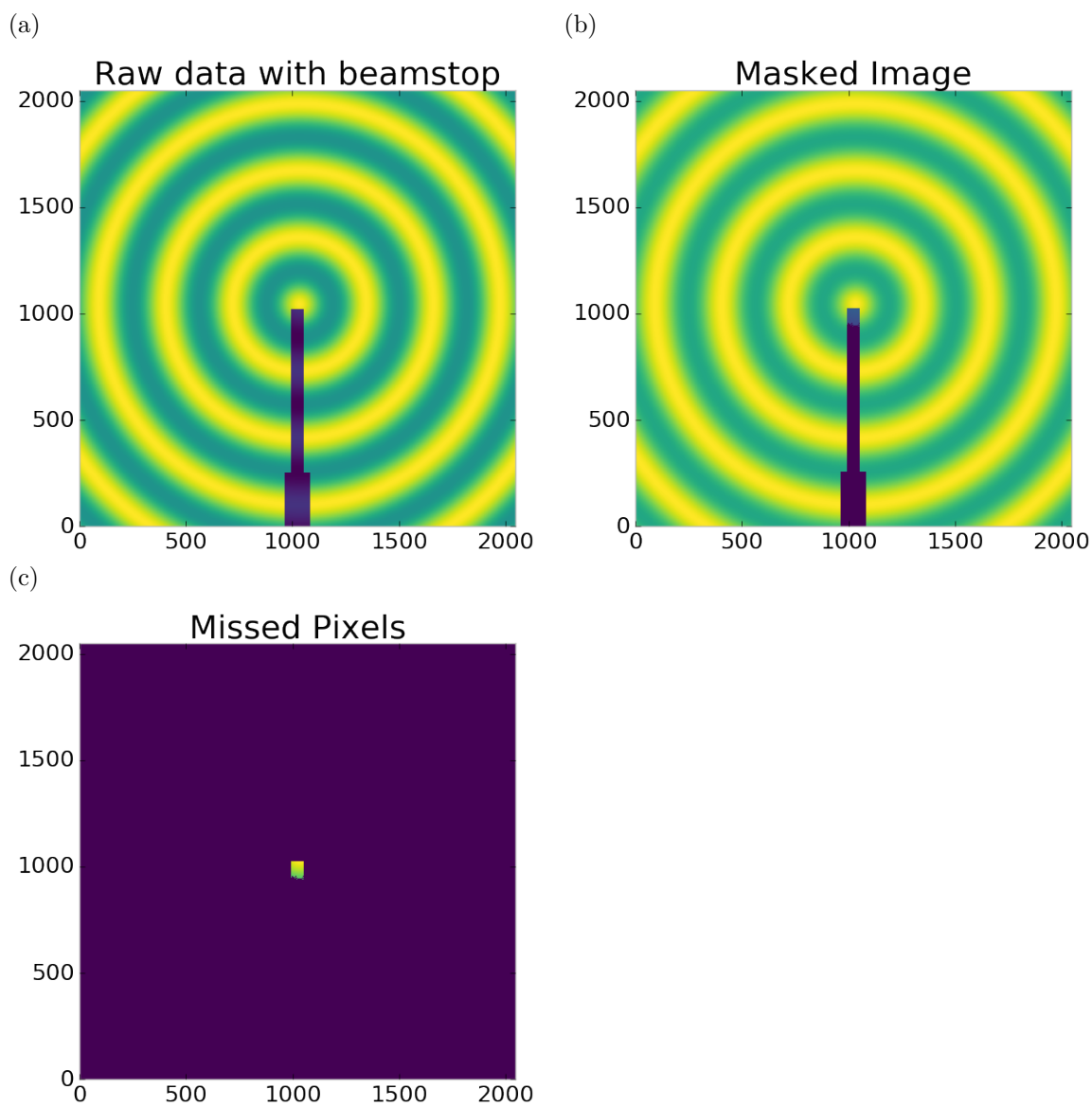


Figure 3.4: Generated beamstop masks for a beamstop holder with 30% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

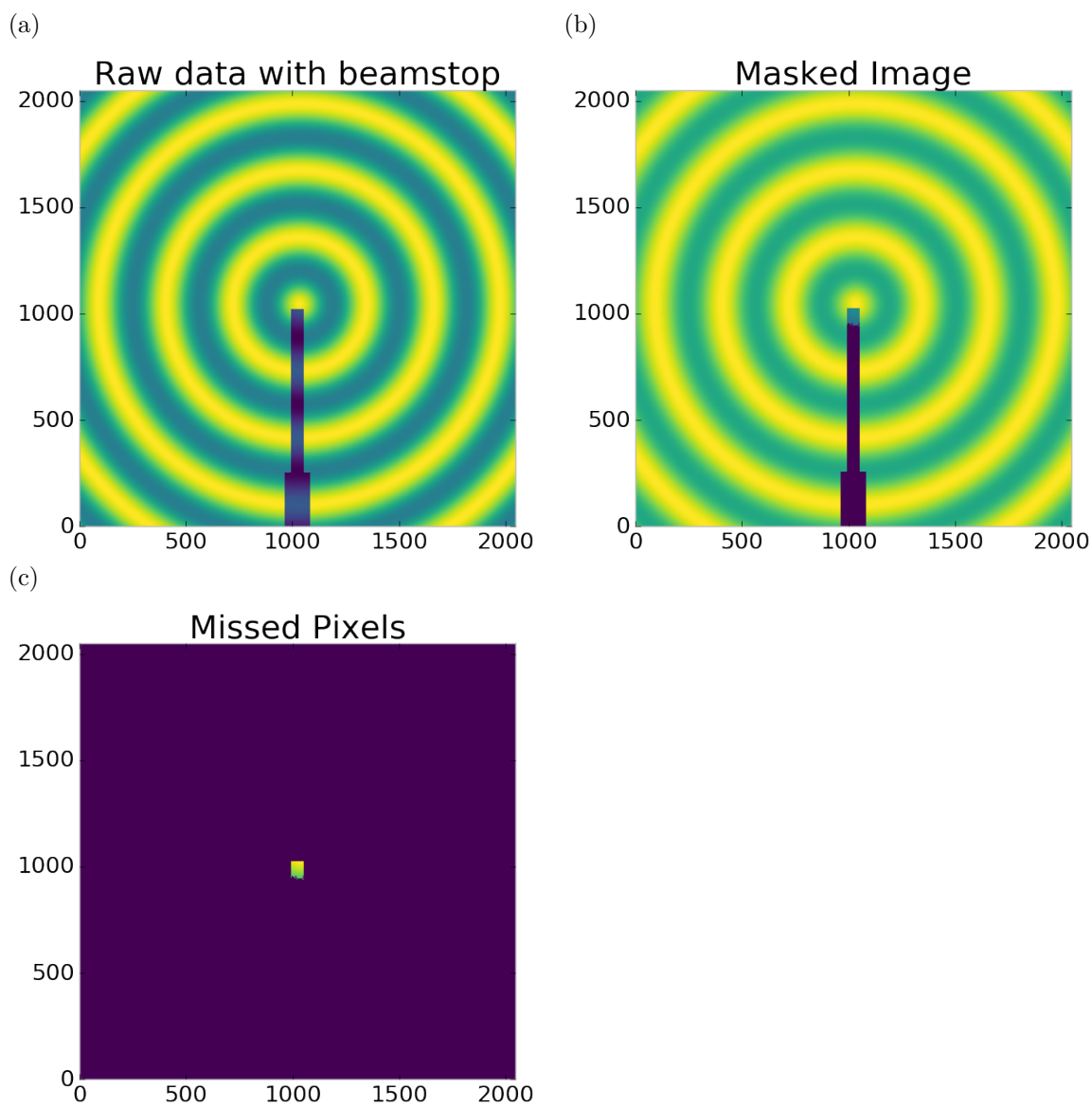


Figure 3.5: Generated beamstop masks for a beamstop holder with 50% transmittance. a) the raw image, b) the masked image, c) and the missed pixels

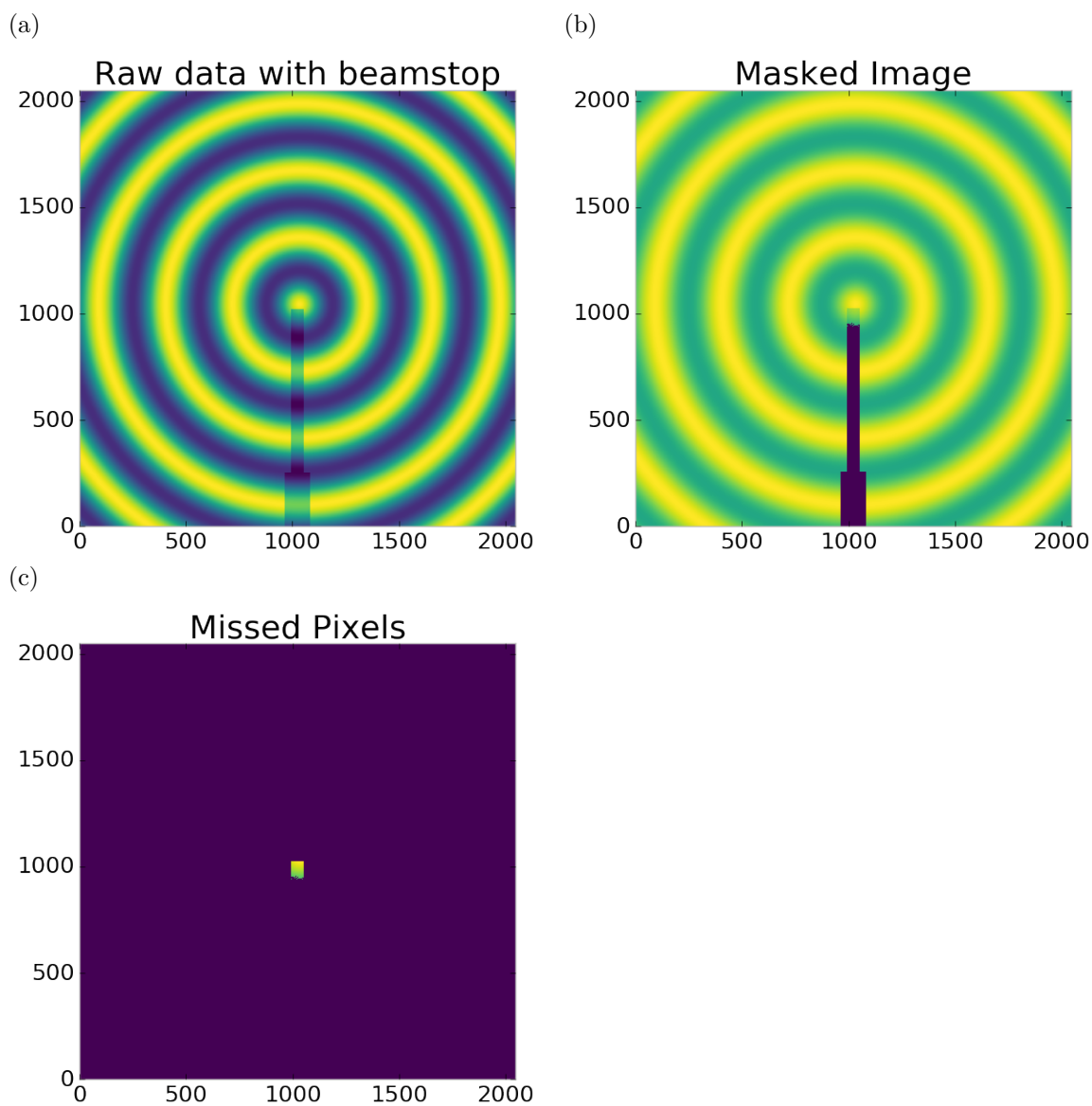


Figure 3.6: Generated beamstop masks for a beamstop holder with 90% transmittance. a) the raw image, b) the masked image, c) and the missed pixels



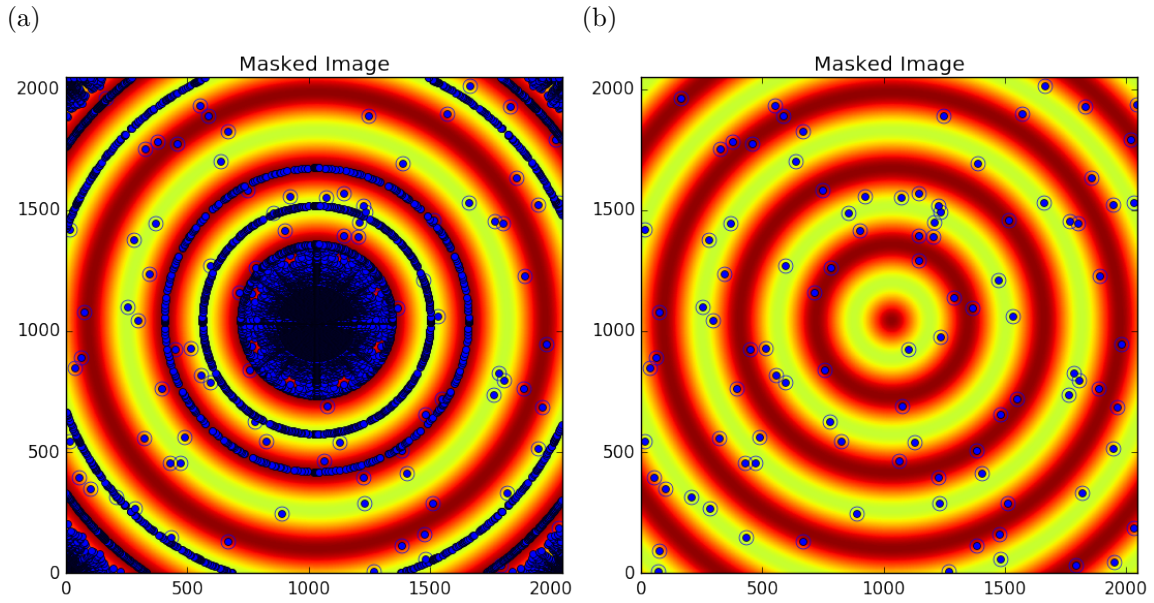


Figure 3.7: Generated dead/hot pixel masks for a detector with 100 bad pixels. a) the poorly binned mask and b) the properly binned mask

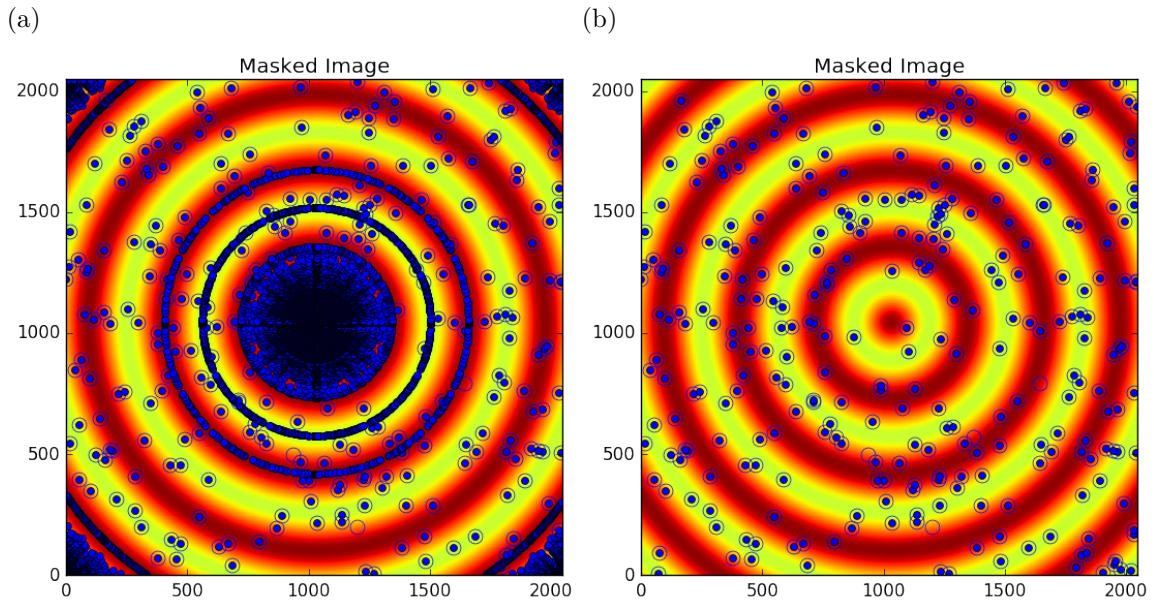


Figure 3.8: Generated dead/hot pixel masks for a detector with 300 bad pixels. a) the poorly binned mask and b) the properly binned mask

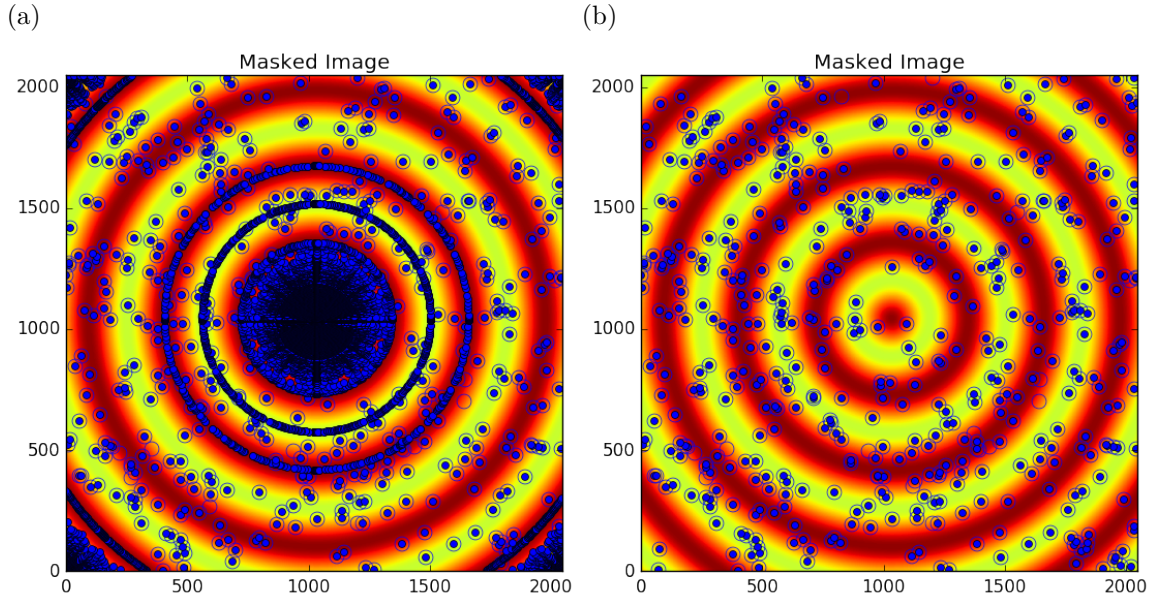


Figure 3.9: Generated dead/hot pixel masks for a detector with 500 bad pixels. a) the poorly binned mask and b) the properly binned mask

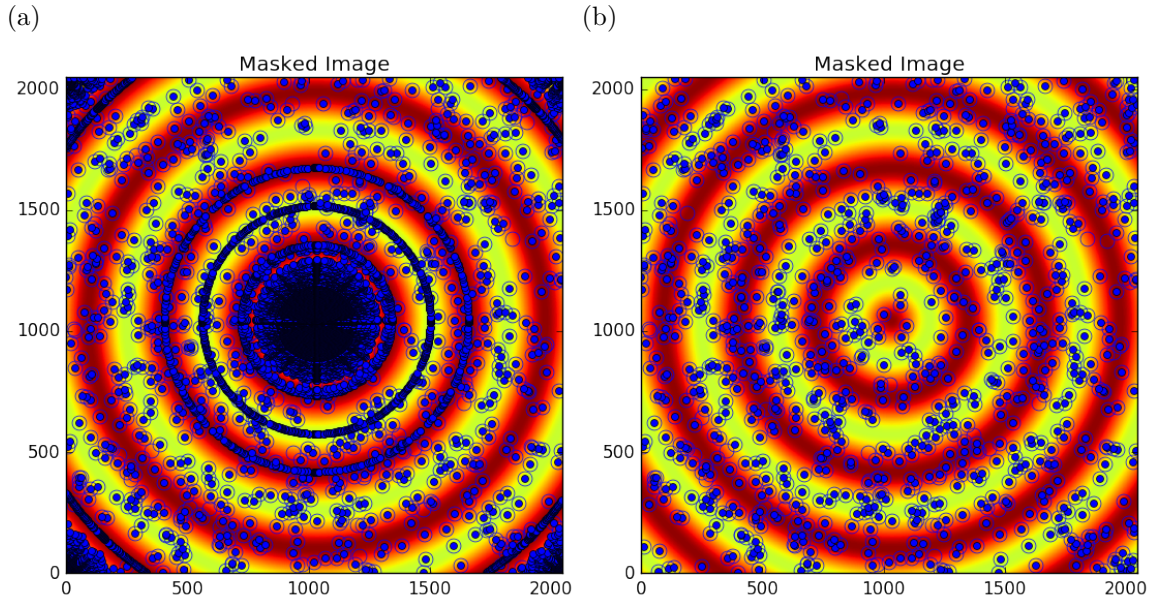


Figure 3.10: Generated dead/hot pixel masks for a detector with 1000 bad pixels. a) the poorly binned mask and b) the properly binned mask

314

## CHAPTER 4

315

## BENCHMARKING

316 4.1 PDF

317 **Au55: surface relaxed**

318 **Au55: surface disordered**

319 **Au55: amorphous**

320 **Au102: triple phase**

321 **Au147**

322 **C60**

323 4.2 PDF WITH ADPS

324

## CHAPTER 5

325

## ANNEALING AND AGGREGATION OF 2NM

326

## AU NANOPARTICLES

327 5.1 EXPERIMENTS

328 **NP Synthesis**

329 **X-ray Total Scattering Measurements**

330 5.2 DATA PROCESSING

331 5.3 DATA ANALYSIS

332 5.4 SIMULATION

333 5.5 STRUCTURAL ANALYSIS

334 5.6 CONCLUSIONS

335

## CHAPTER 6

336

### PHASE CHANGES AND ANNEALING DYNAMICS OF

337

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

338

#### 6.1 EXPERIMENTS

339

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

340

#### **X-ray Total Scattering Measurements**

341

#### 6.2 DATA PROCESSING

342

#### 6.3 DATA ANALYSIS

343

#### **Intra Sample Comparison**

344

#### **Inter Sample Comparison**

345

#### 6.4 SIMULATION

346

#### **Small Box**

347

#### **Large Box**

348

#### 6.5 STRUCTURAL ANALYSIS

349

#### 6.6 CONCLUSIONS

350

## CHAPTER 7

351

## CONCLUSION