
# Statistics in Diffractive Imaging

by Ne-Te Loh

---

# STATISTICS IN DIFFRACTIVE IMAGING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ne-Te Loh

August 2010

STATISTICS IN DIFFRACTIVE IMAGING

Ne-Te Loh, Ph.D.

Cornell University 2010

This dissertation describes reconstruction techniques in diffractive imaging when the data is exceptionally noisy and when crucial experimental parameters are unmeasured. In particular, this work focuses on two applications of diffractive imaging, single particle imaging with unoriented data and ultrafast magnetic imaging with unmeasured charge distribution, both of which are exciting experiments planned for free electron laser facilities.

Concerning single particle imaging, in chapter 2 we introduce the EMC algorithm for reconstructing a particle's 3D diffraction intensity from very many photon-shot-noise limited 2D measurements, when the particle orientation in each measurement is unknown. We coin such an imaging technique cryptotomography. In this chapter, we also study the noise limits beyond which cryptotomography is impossible. This is followed by an experimental demonstration of EMC in chapter 3, where we reconstruct the 3D Fourier intensity distribution of mono-disperse prolate nano-particles using single-shot 2D diffraction patterns collected at DESYs FLASH facility when a bright, coherent, ultrafast X-ray pulse intercepted individual particles of random, unmeasured orientations. This experimental demonstration of cryptotomography extended the Expansion-Maximization-Compression (EMC) framework to accommodate unmeasured fluctuations in photon fluence and loss of data due to saturation or background scatter.

In chapter 4 we discuss magnetic imaging. We study, using simulated

experiments, the feasibility of phase retrieval in X-ray diffractive imaging of thin-film magnetic domains in the presence of intrinsic charge scattering given only photon-shot-noise limited diffraction data. We also chart out the limits of diffractive imaging when we vary both photon-shot-noise and the intensity of charge-scattering noise. This work is directly relevant to the time-resolved imaging of magnetic dynamics using coherent and ultrafast radiation from X-ray free electron lasers.

## BIOGRAPHICAL SKETCH

Ne-Te was born on the 9th of December 1979, in the tiny island of Singapore. He attended St. Michael's Primary School, went on to The Chinese High School, followed by Hwachong Junior College, then onwards to obtain a Bachelor's degree at Harvey Mudd College before arriving at Cornell University to begin a Ph.D. in Physics.

Unbeknownst to many, Ne-Te's first name was a convention adopted across his paternal cousins, who were all named "Ne 'something else' ", regardless of their gender. Ne-Te's uncles and aunts have also adopted a similar convention but with different prefixes. It has never been explicitly stated, at least to Ne-Te, why such a convention was required or why how the prefix "Ne" was chosen. Perhaps it was because Ne-Te had eight uncles and five aunts and it would save his paternal grandmother much grief by only having to associate variations of the last syllable with her numerous grandchildren[1]. Fortunately, this naming convention has ceased for Ne-Te's nieces and nephews.

As a child, Ne-Te went to schools which introduced him to new teachers every year, to whom he had to teach the pronunciation of his name[2]. The intended pronunciation of his name was "knee-tea" [3], but with his parents's limited English training they had to rely on the officials at the Singaporean Registry of Births and Deaths to produce an approximate phonetic spelling of Ne-Te's Mandarin name in English [4]. The officials assumed that Singaporeans would readily

[1] It is common amongst Singaporeans of Chinese descent to have two-syllable first names.

[2] The hyphen and uppercase "T" were mean to encourage the pronunciation of the first name as two syllables. Removing the hyphen would wrongly suggest that "Ne" and "Te" are actually first and middle names. In Mandarin, widely spoken by peoples of Chinese descent, "Ne" would indeed be a very odd first name especially since it is homonymous to the word "ear".

[3] Think He-Ne laser.

[4] Strangely, "Ne-Te" sounds nothing like his name in Mandarin, but resembles what it would be if it were poorly pronounced in the dialect of Hokkien.

say “Ne-Te” correctly [5]. This assumption was proved largely untrue.

In the winter of 1994, Ne-Te decided to solve this problem. He adopted a middle name — Duane [6]— partly to reinvent himself, and partly to distract mispronunciations of his first name. In fact, Duane has flatly refused to pronounce his first name when introducing himself.

Unfortunately, this choice for a middle name turned out to be even worse for Singaporeans. At least with “Ne-Te”, they would still attempt some statistically recognizable utterance. With “Duane”, many Singaporeans would pause, smile to themselves, and usually mumble something inappropriate [7]. As if matters could not become worse, there are at least two other spellings of “Duane” which are more common and frequently assumed for Duane. The pronunciation confusion has doubled now and grown an ugly spelling counterpart.

Duane has gradually developed a sense of humor about his names.

[5] An alternative would be “Nee-tee”.

[6] The reassignment of “Ne” to the end of his middle name was particularly endearing to Ne-Te. The high vowel-to-consonant ratio in “Duane” was also appealing.

[7] There is an unfortunate pun in the name, if forcibly pronounced in the dialect of Hokkien. Since its discovery by Duane’s sister, Nekee, this pun has never left Duane’s family.

This document is dedicated to my parents, who fought to provide for me.

## ACKNOWLEDGEMENTS

I would like to give the biggest and sincerest thanks to Veit Elser for nurturing me through my time graduate school. He has been an intellectual father to me and often a source of scientific rectitude. Since I began working with Veit in 2006, he has showed me a fertile scientific playground filled with various curious puzzles. Beyond cultivating my curiosity, he has also inspired me to tackle problems which I never thought possible, and in a manner more intuitive and elegant than if I had tried by myself. Veit is also remarkably generous with his time and energy — he is willing to spend many hours (daily!) considering my half-formed ideas, showing me computational tricks and correcting my incoherent and ungrammatical writing. He is truly a first-class academic role model.

I would also like to thank my colleagues: Pierre Thibault, Ivan Rakenburg, Simon Gravel and Yoav Kallus. Each of them have endured my numerous trivial questions, suffered my ideas and, most importantly, supplied me with many opportunities to grow as a scientist. My understanding of the world was greatly enriched by discussing it with them.

I am thankful for the support and encouragement from Carl Franck. His perseverance and earnestness left an indelible impression on me.

I am deeply appreciative of the patience and guidance of my many mentors at Cornell University: Jim Sethna, Erich Mueller, Christopher Henley, Sol Gruner, Paul McEuen, Jeevak Parpia and Jim Alexander.

Special thanks to Chris Jacobsen and Yoshinori Nishino for their valuable guidance and help. Also to Michael Bogan, Anton Barty, Henry Chapman, Janos Hajdu and Stefano Marchesini for their help and encouragment concerning the work in Chapter 3.

I am very grateful for my friendship with Albert Bae, Benjamin Smith, Sharon Gerbode, John Gregoire, Nathaniel Gabor, Angela Andiorio, Ryan Shannon, Matthew Warkentin, Xin Shi, Huimin Chen, David Quint, Frank Petruzielo, Arend van der Zande, Joern Kupferschmidt, Johannes Lischner, Milan Allan and many others. They have enriched and sustained me through graduate school and given me wonderful memories to mark many valuable moments during these years.

I would like to thank my family for supporting me through my education away from home. It has been a long ten years and I am truly grateful to them for being my emotional ballast during this time.

Finally, I would like to thank my wonderful wife, Stephanie, for being with me during my most trying and humbling times in graduate school. Her love and thoughtfulness were reliable defenses against punishing days.

## TABLE OF CONTENTS

## LIST OF FIGURES

CHAPTER 1

# DISSERTATION OVERVIEW

The idea of diffractive imaging would certainly appeal to a minimalist, especially when compared to conventional phase-contrast microscopy. Whereas a conventional microscope requires a series of physical optics to manipulate the scattered exit wave from an object of interest to explicate its phase contrast, in diffractive imaging one accomplishes this by simply measuring and interpreting the scattered exit wave directly, absent the optics[1]. This interpretation of the scattered radiation in diffractive imaging is usually done on a computer which, acknowledging it as an alternative to conventional microscopy, is named the "computational lens".

Of course, the simplicity of the apparatus in diffractive imaging is afforded only through the careful construction of a predictable incident light source. This predictability is captured by two characteristics of the light source: its coherence and brightness.

The importance of coherence to diffractive imaging is easily understood — the interaction of a predictable series of radiation wave-fronts with an unknown sample is easier to decode than that with unpredictable wave-fronts. Within reasonable limits, the scattered radiation using coherent light is mathematically simple and easily modeled with a computer [46]. Quite naturally, recovering the object's phase contrast should be accessible when such modeling is possible.

The importance of beam brightness to diffractive imaging is equally straightforward. The quantum description of radiation informs us that radiation com-

[1]Each optical element introduces measurement noise / aberration which is minimized in diffractive imaging.

prises discrete packets of energy, or photons. The brighter the incident beam of a fixed radiation frequency, the greater the number of photons it contains. The task of collecting scattered radiation from an object is equivalent to measuring the spatial distribution of arriving photons on a measurement surface (such as a CCD or a pixel array detector). As such, the statistics of photon arrivals is no different from those of raindrops impinging a rooftop: it obeys Poisson's probability distribution. Naturally, the spatial variation of the scattered radiation is better determined with more photons (or higher photon fluence), and noisier with fewer photons. Since the scattered radiation encodes much usable information about the object, it is clear that a high photon fluence is desirable.

Despite the separate discussions on coherence and brightness in the previous two paragraphs, these two qualities are closely related. In particular, increasing the beam brightness can improve its spatial coherence. Given a quasi-monochromatic beam (contains only frequencies within a narrow band), the spatial variations in its electromagnetic wave-fronts will have a minimum length scale $d_{\text{min}}$. If one were to place a radiation-opaque material in the path of this quasi-monochromatic beam, allowing radiation to pass through only a pinhole (of diameter $d_{\text{min}}$), light emerging from this pinhole will have high spatial coherence (see Figure 1.1). The brightness of the emergent radiation, however, will be drastically reduced by this measly pinhole (especially true for short wavelength X-ray radiation). If one could now increase the brightness of the incident radiation, the rate of photons emerging from this pinhole will also increase perhaps to levels practical for imaging[2]. Nevertheless, this method of increasing beam coherence by increasing the its brightness is wasteful and does

[2]One could increase the fluence by increasing the exposure time. However, exposure times of days or even hours may be experimentally forbidding. Furthermore, fast dynamics are beyond of the reach of such exposure times.

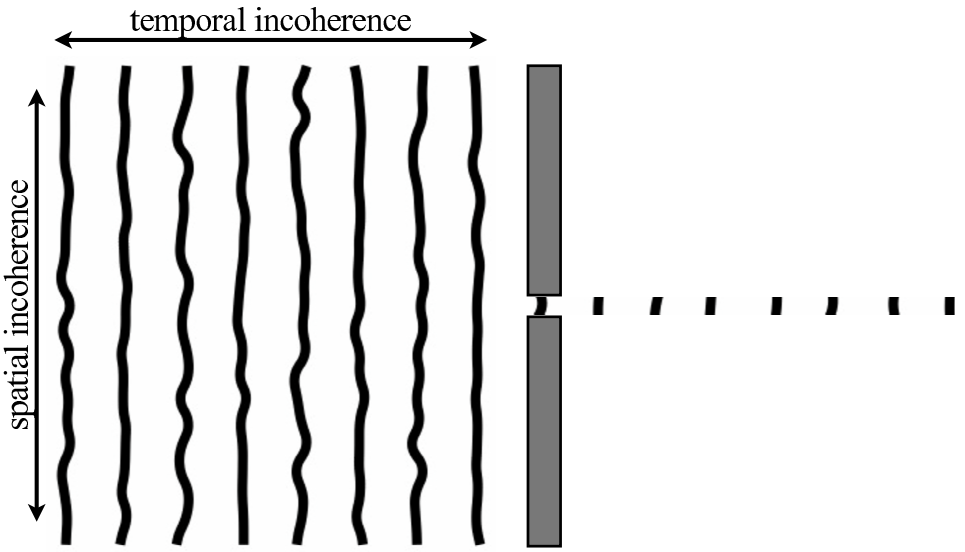



Figure 1.1: A pinhole can be used to isolate spatially coherent regions of a quasi-monchromatic plane wave. This is commonly referred to as a spatial coherence filter.

not improve the beam's temporal coherence.

Alternative strategies have emerged to improve beam's coherence without sacrificing brightness. These strategies primarily involve orchestrating the dynamics of radiating electrons (or positrons) through a feedback mechanism between themselves and the light they generate. Roughly speaking, this was first achieved by sending electrons between two opposing arrays of strong magnets (wigglers and undulators), then more recently preparing the source electrons in a controllable fashion (the subject of seeding in free electron lasers).

## 1.1 Shannon and the history of diffractive imaging

X-ray diffractive imaging began in the early 20th century, first in the study of crystal structure, then to crystals of proteins to determine the structure of these

proteins. In X-ray protein crystallography, the crystal lattice is crucial in fixing the positions and orientations of nearly identical proteins. While the diffraction signal in X-ray protein crystallography is enhanced with multiple copies of the scattering protein, the radiation dose to each protein in the crystal remains small.

In 1952, David Sayre reflected upon the work of Claude Shannon regarding sampling conditions necessary to image objects of limited Fourier complexity [39]. Then in 1980, he postulated the possibility of using soft X-rays to image single biological cells [40]. Many years had passed before diffractive imaging was experimentally demonstrated on such non-periodic samples [31]. This was partly because of improvements in beam qualities of dedicated synchrotron facilities[3]; partly because of the advent of better X-ray detectors; and partly because of algorithmic advances in the computational lenses [20] and the invention of computers which can swiftly execute these algorithms. Since the invention of wigglers and undulators, our ability to produce bright and coherent X-ray radiation has vastly improved. More recently, the technological triumphs which culminated in the construction of free electron lasers (FELs) where the engineered coherent dynamics of electrons produce X-ray (and EUV) lasers of unparalleled coherence and brightness. In step with these advancements is a diligent effort to develop diffractive imaging into a serious high-resolution imaging technique [48, 8].

[3]The coherence and brightness of X-ray sources in the mid 20th century were modest, which were more suitable for crystalline samples (where the signal is amplified by many identical objects) than single, non-periodic, microscopic objects.

## 1.2 Shannon and the future of diffractive imaging

With FELs producing coherent radiation of unprecedented brilliance in very short pulses (SLAC's Linac Coherent Light Source can produce 100 fs pulses with $\sim 10^{12}$ hard X-ray photons each), the limits of diffractive imaging were greatly advanced. People began to seriously consider imaging single particles at high resolution and short time scales. However, such endeavors approached the achievable limits of diffractive imaging since their diffraction signal will be very noisy. The determination of these limits are becoming increasingly relevant, not only to the planning of current experiments but may also to the design of future light sources. Fortunately, the tools needed to define these limits were already provided by Claude Shannon.

Claude Shannon was famous for his work in information theory, especially in establishing the bounds on transmitting information through noisy communication channels [42]. Our connection to his work begins when one considers diffractive imaging abstractly as such a communication channel [16]. In this case, the message would be the object's phase contrast encoded in its true diffraction intensities, which is transmitted using photons through the Poisson noise channel. Considering diffractive imaging in this language, one could apply the principles of information theory to diffractive imaging to determine bounds on attainable imaging complexities.

Whereas these Shannon bounds are useful in defining the limits of noisy imaging, they are silent about the algorithms for decoding the noisy diffraction images when such is possible with arbitrarily small errors. Developing and implementing such algorithms are the central goals of this dissertation.

### 1.2.1 Single particle imaging

As diffractive imaging matures, the scientific community revisited Sayre's original idea of imaging single nanometer-sized objects rather than a crystal of such objects. This revisitation is beyond a mere academic curiosity — many biomolecules of interest resist crystallization, which makes it impossible to determine their structure using X-ray crystallography. Solem first proposed using extremely short pulses of X-ray to image single particles, where the photons interact with the particle before it catastrophically explodes from ionization-induced Coulombic forces [45]. Numerical simulations by Neutze et. al showed that Angstrom-resolution imaging may be possible on nanometer-sized biomolecules with intense X-ray pulses of shorter than tens of femtoseconds [32].

Not withstanding the challenge of designing and building light sources capable of producing such intense femtosecond X-ray laser pulses, or delivery of single particles into the path of X-ray pulses focused to a few micrometers, there remained a fundamental uncertainty about the effectiveness of computational lenses in such diffractive imaging experiments: even with the expected fluence of FEL radiation, only mere hundreds of scattered photons are expected from each target particle [32]. This severe signal-to-noise ratio would not threaten diffractive imaging if many such noisy diffraction patterns were averaged, presumably from exposing many identical particles to intense X-ray pulses. However, this direct averaging demands that the 3D orientation of each target particle to be precisely known or controlled. In Chapter 2, we describe an expectation maximization algorithm which succeeds at recovering the particle's phase contrast *even in the absence of orientation information*. In the same chapter, we also

study the limits of reconstruction using Shannon's noisy coding channel theorem.

Our algorithm was first tested using numerical simulations in chapter 2, which also showed that single particle diffractive imaging should be possible with merely hundreds of scattered photons. We coin this mode of diffractive imaging "cryptotomography". In chapter 3, our reconstruction algorithm was applied to experimental cryptotomography data which included fluctuations in pulse-to-pulse (or shot-to-shot) fluctuations in photon-fluence and severe background noise.

### 1.2.2 Imaging fast magnetic dynamics

Very short FEL pulses also enable the imaging of ultrafast dynamics: a series of short pulses would interrogate an object many times to produce a sequence of diffraction intensities which could be decoded into a movie of dynamics in the object. One obvious application of this is to image fast switching dynamics in thin-film magnetic nanostructures. The magnetization distribution on these magnetic samples are apparent via resonance scattering, but the combination of short pulse duration and intensity reduction to minimize sample heating results again in few scattered photons and hence noisy diffraction images. Shannon's noisy channel theorem was applied in a lower-dimensional diffractive imaging problem [17] to determine the limits of diffractive imaging on binary-contrast images. The reconstruction (decoding) algorithm in [17] formed the foundation for work in Chapter 4, where we describe an algorithm which a computational lens could implement to recover the magnetic contrast given extremely noisy

diffraction data even in the presence of unmeasured, intrinsic charge scattering.

## 1.3 Acknowledgements

The following sources were useful in the preparation of this introductory chapter: Keith Nugent's review article [33]; Pierre Thibault's summary of diffractive imaging [47]; material from PHYS698 at Cornell University — a course on synchrotron Physics taught by Sol Gruner, Don Bilderback and Hugh Philipp.

CHAPTER 2

# SIMULATED CRYPTOTOMOGRAPHY

This chapter is an adaption of [28] (copyright 2009 by American Physical Society), which was a collaboration with Veit Elser.

## 2.1 Introduction

If the goal of single-particle imaging by free electron x-ray lasers [32] is realized in the next few years, the disciplines of imaging and microscopy will have partly merged with elementary particle physics. Even with the enormous flux of the new light sources, the scattered radiation will be detected as individual photons and hardly resemble diffraction "images" (Figure 2.1). The data in these experiments will instead resemble the particle debris produced in elementary particle collisions.

The particle physics analogy is imperfect, however. Data analysis in elementary particle experiments is complicated more as the result of complex interactions than complexity of the structures — consider the pions produced when a proton is probed with a photon. By contrast, the fundamental interactions between x-ray photons and electrons in a molecule are very simple and the complexity in the analysis of the data is entirely the result of structure.

There are two different data analysis challenges that x-ray laser studies of single-particles will have to face. Consider the two simulated detector outputs shown in Figure 2.1. Are the photon counts different because the molecule presented a different orientation to the x-ray beam; is the difference attributable to the statistics of a shot-noise limited signal; or does some combination of the

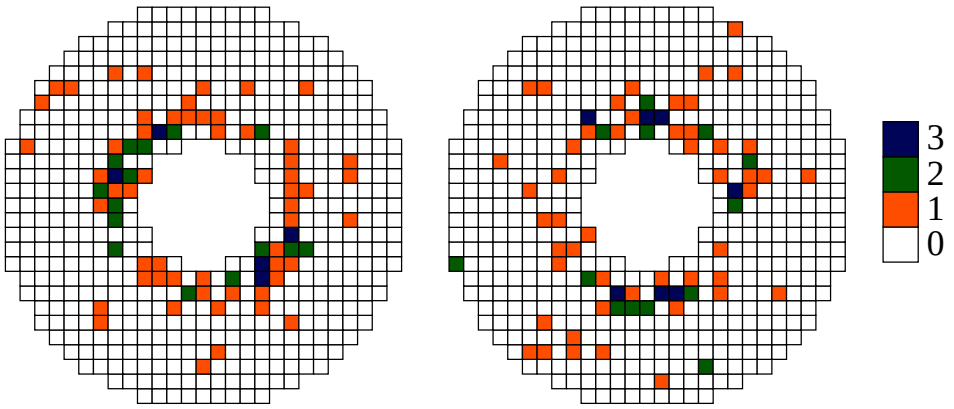



Figure 2.1: The same or different? Two simulated measurements (noisy diffraction patterns) in a single particle imaging experiment, where color (white, orange, green, blue) represents recorded photon counts (0, 1, 2, 3). Are the differences in the measurements purely statistical, or do they reflect a different view (orientation) of the particle?

two apply? It is reasonable to conjecture that by collecting sufficiently many data, the orientational and statistical uncertainties can be disentangled to produce molecular reconstructions with acceptable noise and resolution. In this chapter we present strong evidence in support of this conjecture by means of an algorithm that succeeds with simulated data.

Due to the length of this chapter, a survey of its contents may be useful to the reader. Section 2.2 explains the theoretical basis of our algorithm whose success is contingent upon an information theoretic noise criterion. Sections 2.3, 2.4 and 2.5 respectively describe the test target particles, experimental diffraction conditions and algorithmic parameters used in these single-particle imaging simulations. The limits and the encouraging results of these simulations, whose code implementation we elaborate in section 2.6, are presented in section 2.7. Finally, section 2.8 discusses the scaling of our algorithm's computational requirements

with reconstructed resolution.

## 2.2 Theory

The statistical noise and missing orientational information can be addressed by imposing internal consistency of two kinds. First consider the shot noise of the detected signal. Suppose a collection of data sets, such as the pair in Figure 2.1, have been identified as candidates for data taken with the molecule in nearly the same orientation. Whereas simply averaging the photon counts yields the continuous signal we are after, we have available the stronger test that the distribution of counts for the measurement ensemble, at each pixel, has the correct Poissonian form. If the test fails, then a different subset of the data must be identified which has this property. Statistical consistency, by itself, is thus a means for classifying like-oriented data sets.

The purely statistical analysis makes no reference to the structure implicit in the missing orientational information. This structure begins with the basic fact that the missing information comprises just three continuous variables (*e.g.* Euler angles), and extends to more detailed constraints, such as the fact that the data samples a signal on a spherical (Ewald) surface in three dimensions and different spherical samples have common values along their intersection, *etc.* A successful data analysis scheme for the single particle imaging experiments will not just have to signal-average shot noise, but must also reconstruct the missing orientational information by relying on internal consistency associated with the rotation group. The two forms of uncertainty, statistical and orientational, are not independent. In particular, when the statistical noise is large (few de-

tected photons), we expect the reconstruction of the orientational information to be probabilistic in character (*i.e.* distributions of angles as opposed to definite values).

### 2.2.1 Noise criterion

A natural question to ask is whether there exists an information theoretic criterion that would apply to any reconstruction algorithm and that can be used to evaluate the feasibility of reconstructions for particular experimental parameters. Elser [16] studied this question for a minimal model with a single rotation angle and obtained an explicit noise criterion formula. Although such a detailed analysis is difficult to extend to the three dimensional geometry of the single-particle imaging experiments, the mathematical statement of the criterion is completely general and, when evaluated numerically by the reconstruction algorithm, serves as a useful diagnostic. We include a brief discussion of the criterion here and refer the reader to the original article for details.

We recall that the mutual information $I(X,Y)$ associated with a pair of random variables $X$ and $Y$ is an information theoretic measure of their degree of correlation: $I(X,Y)$ is the average information in bits that a measurement of $X$ reveals about $Y$ (or conversely). Keeping with the notation of references [16, 25], we denote the three dimensional intensity distribution by $W$, the photon counts recorded by the detector on a two dimensional spherical surface by $K$, and the three unknown parameters that specify the orientation of the surface within the intensity space by $\Omega$ [1]. The intensities $W$ have the interpretation of random variables (just as $K$ and $\Omega$), since the particle being reconstructed (and so the

[1] $W$, $K$ and $\Omega$ are representation-independent variables.

associated $W$) belongs to a statistical ensemble with known characteristics (size, intrinsic resolution, *etc.*).

There are three forms of mutual information that arise in the framework where information about a model $W$ is obtained through measurement of data $K$ that is both statistically uncertain and incomplete (because $\Omega$ is not measured). The first is $I(K,W)$ and measures the information obtained about the model intensities $W$ from a typical unoriented measurement $K$. A second mutual information is $I(K,\Omega)|_W$, the correlation between the measurement $K$ and the orientation $\Omega$ conditional on a typical model $W$. We may also think of $I(K,\Omega)|_W$ as the entropy of $\Omega$ reduced by the finite entropy in its distribution when given typical measurements $K$ and models $W$. Finally, a simple identity [2] yields the third form

$$I(K,W)|_\Omega = I(K,W) + I(K,\Omega)|_W \tag{2.6}$$

as the sum of the other two. The mutual information $I(K,W)|_\Omega$ is the simplest

[2]Given a trio of random variables $K$, $\Omega$ and $W$, we can evaluate the mutual information between one of them, say $K$, and the other pair, $(\Omega,W)$, treated as a single random variable. Writing the mutual information in terms of the entropy function $H$, we have

$$\begin{aligned} I(K,(\Omega,W)) &= H(K) - H(K)|_{(\Omega,W)} && (2.1)\\ &= H(K) - H(K)|_\Omega + H(K)|_\Omega - H(K)|_{(\Omega,W)} \\ &= I(K,\Omega) + I(K,W)|_\Omega && (2.2)\end{aligned}$$

Interchanging $\Omega$ and $W$ in this derivation gives the identity

$$I(K,(\Omega,W)) = I(K,W) + I(K,\Omega)|_W. \tag{2.3}$$

Combining the two identities above we obtain the general result

$$I(K,\Omega) + I(K,W)|_\Omega = I(K,W) + I(K,\Omega)|_W. \tag{2.4}$$

For our specific choice of random variables the mutual information $I(K,\Omega)$ vanishes identically because a measurement $K$ confers no information about the orientation $\Omega$ since the ensemble of models $W$ itself has an orientational degree of freedom that is uniformly distributed. Our identity thus involves only three terms:

$$I(K,W)|_\Omega = I(K,W) + I(K,\Omega)|_W. \tag{2.5}$$

of the three, as it measures the direct correlation between the continuous signal $W$ and its Poisson samples $K$ because it is conditional on a known orientation $\Omega$. In the limit where the mean photon count per detector pixel is much less than 1, this mutual information is given simply in terms of the total number of photons $N$ detected in an average measurement [16],

$$I(K,W)|_\Omega = (1-\gamma)N \ , \tag{2.7}$$

where $\gamma \approx 0.577$ is Euler's constant.

In order to sufficiently sample the particle orientations and improve the signal-to-noise, information is accumulated over the course of many measurements. The information delivered in a stream of measurements will initially grow in proportion to the number of measurements, since typically each 2D measurement $K$ samples a different part of the 3D signal $W$. Two of the mutual information quantities introduced above may therefore be interpreted as *information rates*:

$$I(K,W)|_\Omega \ = \ \text{data rate in a hypothetical experiment with known particle orientations.}$$

$$I(K,W) \ = \ \text{data rate in the actual experiment with unknown orientations.}$$

The time unit in these rates is the time for one measurement. The larger of these rates, $I(K,W)|_\Omega$, applies to the situation where the noisy data $K$ can simply be signal-averaged to obtain $W$. From the ratio

$$r = \frac{I(K,W)}{I(K,W)|_\Omega} \tag{2.8}$$

we can assess the reduction in the data rate relative to the signal-averaging scenario. Because this reduction can be severe when shot noise is large, we are primarily interested in the dependence of $r$ on the mean photon number per measurement, $N$. Not only does an experiment with small $r(N)$ require a correspondingly larger number of measurements to obtain the same signal-to-noise in the reconstructed particle, our reconstruction algorithm (Section 2.2.3) requires many more iterations in this case.

Upon using equation (2.6), the case $r(N) = 1/2$ corresponds to the situation $I(K, W) = I(K, \Omega)|_W$, that is, the information in one unoriented measurement exactly matches the information acquired about its orientation. This interpretation does not imply that reconstruction is impossible for smaller $r(N)$, since the criterion refers to the properties of a single measurement while the reconstruction algorithm may, in principle, process many measurements in aggregate. Nevertheless, the criterion $r(N) > 1/2$ correctly identifies the cross-over region separating easy and hard reconstructions. Using (2.6) we can rewrite the feasibility criterion in the form

$$r(N) = 1 - \frac{I(K, \Omega)|_W}{I(K, W)|_\Omega} > \frac{1}{2} \, . \tag{2.9}$$

Our algorithm, based on the expectation maximization principle [11], evaluates $r(N)$ with no overhead since the probability distributions in $\Omega$ of the measurements $K$, from which $I(K, \Omega)|_W$ is derived, are computed in the course of updating the model. When the inequality above is strongly violated we should expect a much lower signal-to-noise in the finished reconstruction than a naive signal-averaging estimate would predict.

An important general observation about the noise criterion (2.9) is that it is remarkably optimistic. As an information measure, $I(K, \Omega)|_W$ grows only

logarithmically with the complexity of the particle. Recall that $I(K,\Omega)|_W$ is the entropy reduction in $\Omega$ of a typical measurement $K$. Suppose a particle of radius $R$ has its density resolved to contrast elements of size $\delta R$. Its rotational structure (in its own space or the Fourier transform space of $W$) will then only extend to an angular resolution $\delta R/R$. A sampling of the rotation group comprising $(R/\delta R)^3$ elements thus provides a fair estimate of the entropy and $I(K,\Omega)|_W$ evaluates to a number of order $3\log(R/\delta R)$ [3]. This estimate and equations (2.7) and (2.9) imply that values of $N$ of only a few hundred should be sufficient to reconstruct even the most complex particles encountered in biology.

### 2.2.2 Classification by cross correlating data

The theoretical noise criterion discussed above is beyond the reach of the kind of algorithm that would seem to offer the most direct solution [25]. In this scenario the task is divided into two steps. The first is concerned with classifying the diffraction data into sets that, with some level of confidence, describe the particle in a small range of orientations. After averaging photon counts for the like-sets to improve the signal quality, the diffraction pattern averages would then be assembled into a consistent three dimensional intensity distribution in the second step.

The most direct method of assessing the similarity of two diffraction data is to compute the cross correlation. A pair of like-views of the particle would be identified by a large cross correlation. Because this measure also fluctuates as the result of shot noise, its statistical significance must be estimated as well. The result of such an analysis [25] is that the cross correlation based classification

[3] $I(K,\Omega)|_W$ also has a weak logarithmic variation with $N$ [16].

can succeed only if the average number of photons per diffraction pattern, $N$, and the number of detector pixels, $M_{\mathrm{pix}}$, satisfy

$$N \gg \sqrt{M_{\mathrm{pix}}} \,. \tag{2.10}$$

This criterion imposes a higher threshold on $N$ than the information theoretic criterion (2.9) because $M_{\mathrm{pix}}$ grows algebraically, and not logarithmically, with the complexity of the reconstructed particle.

Since the number of measured diffraction patterns will be very large, and the number of pairs to be cross correlated grows as its square, the execution of this approach also seems prohibitive. As Bortel *et al.* [5] have shown, however, this estimate is overly pessimistic since by selecting suitable representatives of the orientational classes the number of cross correlation computations can be drastically reduced. The expectation maximization (EM) algorithm described below is an alternative classification method where the most time consuming step is again the computation of very many cross correlations. But unlike methods where both vectors of the cross correlation are data and criterion (5) applies, in the EM algorithm only one of the vectors is data while the other is derived from a model. This has the added bonus that the time of the EM calculation is linear, rather than quadratic, in the number of measurements.

### 2.2.3 Classification by expectation maximization

The algorithm we have developed for the single particle imaging experiments and studied previously in the context of noise limits [16] is based on the idea of expectation maximization (EM) [11] . In general, EM seeks to reconstruct a model from statistical data that is incomplete. The model in the present setting

is the intensity signal $W$, the data are the sets of photon counts $K$ recorded by the detector, and the latter are incomplete because the orientation $\Omega$, of the particle relative to the detector, is not measured.

The EM algorithm is an update rule on the model, $W \rightarrow W'$, based on maximizing a log-likelihood function $Q(W')$. The algorithm derives its name from the fact that $Q(W')$ is actually an expectation value of log-likelihood functions, where a probability distribution based on the current model parameters $W$ is applied to the missing data $\Omega$. We will derive $Q(W')$ for the single particle imaging problem in stages, beginning with the log-likelihood function for the photon counts at a single detector pixel.

Let $W(\mathbf{q})$ be the time-integrated scattered intensity at spatial frequency $\mathbf{q}$ when the particle is in some reference orientation. The detector pixels, labeled by the index $i$, approximately measure $M_{\text{pix}}$ point samples $W(\mathbf{q}_i)$. When multiplied by the pixel area and divided by the photon energy, $W(\mathbf{q}_i)$ corresponds to the average photon number recorded at pixel $i$. Since these normalization factors are constants, we will refer to $W$ interchangeably as "intensity" or "average photon number." If we now give the particle some arbitrary orientation $\Omega$, the average photon number at pixel $i$ is $W(\mathbf{R}_\Omega \cdot \mathbf{q}_i)$, where $\mathbf{R}_\Omega$ is the orthogonal matrix corresponding to the rotation between the reference orientation and $\Omega$. Because the implementation of the algorithm approximates the continuous $\Omega$ with a discrete sampling of $M_{\text{rot}}$ points labeled by the index $j$, we define $W_{ij} = W(\mathbf{R}_j \cdot \mathbf{q}_i)$ as the average photon number at detector pixel $i$ when the particle has orientation $j$.

The log-likelihood function for the mean photon number $W'_{ij}$, given a photon count $K_{ik}$ at pixel $i$ in measurement $k$, is the logarithm of the Poisson distribu-

tion (apart from an irrelevant constant):

$$Q_{ijk}(W') = K_{ik} \log W'_{ij} - W'_{ij} \ . \tag{2.11}$$

Summing this function over the detector pixels gives the log-likelihood function associated with the joint and independent Poisson distributions on the photons detected in a single measurement (labeled by $k$):

$$Q_{jk}(W') = \sum_{i=1}^{M_{\rm pix}} Q_{ijk}(W') \ . \tag{2.12}$$

If we knew the orientation $j$ that applied to the counts $K_{ik}$ of measurement $k$, we would try to maximize the corresponding $Q_{jk}$ with respect to the model values $W'_{ij}$. The EM algorithm deals with the missing information by making an educated estimate of $j$, for each measurement $k$, based on the current model values. However, before we enter into these details we should point out that the EM algorithm in our formulation works with many more model parameters than there are in the physical model. That is because $W_{ij}$ and $W_{i'j'}$ are treated as independent parameters even in the event that the corresponding spatial frequencies $\mathbf{R}_j \cdot \mathbf{q}_i$ and $\mathbf{R}_{j'} \cdot \mathbf{q}_{i'}$ are nearly the same. This overspecification of parameters will be rectified by the "compression step" described below.

The EM algorithm defines the log-likelihood function $Q(W')$ on the updated model parameters $W'$ by assigning a provisional distribution of orientations $j$ to each measurement $k$ based on the current model parameters $W$. The $j$-distribution is given as the normalized likelihood function for the measurements $K_{ik}$ conditional on $j$ and the model parameters $W$. Up to an irrelevant $j$-independent factor, the conditional probability in question is the product of Poisson probabilities at each detector pixel:

$$R_{jk}(W) = \prod_{i=1}^{M_{\rm pix}} W_{ij}^{K_{ik}} \exp\left(-W_{ij}\right) \ . \tag{2.13}$$

The normalized likelihood function allows for an arbitrary prior distribution of the orientations $j$ which we denote by the normalized weights $w_j$:

$$P_{jk}(W) = \frac{w_j R_{jk}(W)}{\sum_j w_j R_{jk}(W)} \ . \tag{2.14}$$

This form is necessary even when the prior distribution on orientations $\Omega$ is uniform (the usual assumption for the single particle experiments) because in general the discrete samples $j$ cannot be chosen in such a way that the weights $w_j$ are uniform. The EM log-likelihood function may now be written explicitly:

$$Q(W') = \sum_{k=1}^{M_{\rm data}} \sum_{j=1}^{M_{\rm rot}} P_{jk}(W) Q_{jk}(W') \ . \tag{2.15}$$

Maximizing $Q(W')$ [4] results in the (M) update rule, which is simple and intuitive:

$$\mathbf{M}: \quad W_{ij} \to W'_{ij} = \frac{\sum_{k=1}^{M_{\rm data}} P_{jk}(W) K_{ik}}{\sum_{k=1}^{M_{\rm data}} P_{jk}(W)} \ . \tag{2.19}$$

We see that the data $K_{ik}$ are averaged over all the data sets ($k$ index) with the unknown orientation index $j$ distributed according to probabilities $P_{jk}(W)$ defined by the current model. It is instructive to check that the update rule applied to an arbitrary rotation of the true signal leaves the signal unchanged (see Appendix A for details).

[4] Rearranging the order of the sums in the definition of the log-likelihood function we obtain

$$Q(W') = \sum_{i=1}^{M_{\rm pix}} \sum_{j=1}^{M_{\rm rot}} (A_{ij} \log W'_{ij} - B_j W'_{ij}), \tag{2.16}$$

where

$$A_{ij} = \sum_{k=1}^{M_{\rm data}} P_{jk}(W) K_{ik} \tag{2.17}$$

$$B_j = \sum_{k=1}^{M_{\rm data}} P_{jk}(W). \tag{2.18}$$

Each term of the sum (2.16) is of the form $a \log W - b W$ where $a$ and $b$ are positive constants. Since the terms are independent, the global maximum is achieved when each term is maximized with the value $W = a/b$.

We now return to the point that the parameters $W_{ij}$ overspecify the true model parameters. For fixed $j$, the $M_{\rm pix}$ numbers $W_{ij}$ correspond to a tomographic sampling of the 3D space of intensities on a spherical surface with orientation in the 3D space specified by $j$. To recover a signal in the 3D space we define a "condensation / compression" ($\mathbf{C}$) mapping

$$\mathbf{C}: \quad W_{ij} \rightarrow W(\mathbf{p}) \; , \tag{2.20}$$

where $\mathbf{p}$ denotes a spatial frequency sampling point in the 3D intensity space. Since the samples $\mathbf{p}$ will be arranged on a regular 3D grid, we define interpolation weights $f(\mathbf{q})$ for a general point $\mathbf{q}$ in the 3D space which vanish for large $|\mathbf{q}|$ and have the property

$$\sum_{\mathbf{p}} f(\mathbf{p}-\mathbf{q}) = 1 \tag{2.21}$$

for arbitrary $\mathbf{q}$. Recalling that the value $W_{ij}$ corresponds to the 3D sampling point $\mathbf{R}_j \cdot \mathbf{q}_i$, the signal values after the compression mapping are given by

$$W(\mathbf{p}) = \frac{\sum_{i=1}^{M_{\rm pix}} \sum_{j=1}^{M_{\rm rot}} f(\mathbf{p}-\mathbf{R}_j \cdot \mathbf{q}_i) W_{ij}}{\sum_{i=1}^{M_{\rm pix}} \sum_{j=1}^{M_{\rm rot}} f(\mathbf{p}-\mathbf{R}_j \cdot \mathbf{q}_i)} \; . \tag{2.22}$$

To begin another round of the EM algorithm, after the condensation step, the signal values on the 3D grid have to be "exported / expanded" ($\mathbf{E}$) to the tomographic representation:

$$\mathbf{E}: \quad W(\mathbf{p}) \rightarrow W'_{ij} \; . \tag{2.23}$$

Using the same interpolation weights and rotation samples $j$ as before, we have

$$W'_{ij} = \sum_{\mathbf{p}} f(\mathbf{p}-\mathbf{R}_j \cdot \mathbf{q}_i) W(\mathbf{p}) \; . \tag{2.24}$$

The combined mappings $\mathbf{E} \cdot \mathbf{C} : W_{ij} \rightarrow W'_{ij}$ then have the effect of imposing on the redundant tomographic representation of the signal the property that it is

derived from values on a 3D intensity grid. A slightly different way of grouping the three mappings defines one iteration of what we will call the EMC algorithm (**E**xpansion followed by expectation **M**aximization followed by **C**ompression):

$$\mathbf{C}\cdot\mathbf{M}\cdot\mathbf{E}: \quad W(\mathbf{p}) \rightarrow W'(\mathbf{p})\ . \tag{2.25}$$

The most time consuming step of the EMC algorithm is the computation of the probabilities $P_{jk}(W)$. Prior to normalization these are the likelihood functions $R_{jk}(W)$ whose logarithms are given by

$$\log\left(R_{jk}(W)\right) = \sum_{i=1}^{M_{\mathrm{pix}}} K_{ik}\log W_{ij} - W_{ij}. \tag{2.26}$$

At the heart of the algorithm we have to compute the cross correlation (sum on $i$) between the photon counts in each measurement $k$ and the logarithm of the signal at each tomographic sampling (particle orientation) $j$. Since data are not cross correlated with data, as in some classification methods, the time scaling is linear in the number of measurements. After normalizing to get $P_{jk}(W)$, the mutual information needed for the noise criterion (2.9) is obtained without significant additional effort (see Appendix B):

$$I(K,\Omega)|_W = \frac{1}{M_{\mathrm{data}}}\sum_{k=1}^{M_{\mathrm{data}}}\sum_{j=1}^{M_{\mathrm{rot}}} P_{jk}(W)\log\left(\frac{P_{jk}(W)}{w_j}\right)\ . \tag{2.27}$$

The expectation maximization technique described above is very similar to that used by Scheres *et al.* [41] for cryo-EM reconstructions. Cryo-EM and single-particle x-ray imaging differ in two important physical respects. The first is that the diffraction data in the x-ray experiments has a known origin (zero frequency), thereby reducing the missing information. This is not completely an advantage because the diffraction data, after a successful reconstruction, must undergo an additional stage of phase retrieval before the results can be compared with cryo-EM. The second difference is the noise model that applies to the

two techniques. In the absence of background, the shot noise in the x-ray experiments is a fundamental and parameter-free process, whereas the background ice scattering in cryo-EM requires phenomenological models. The expectation maximization algorithm is general enough that these differences do not change the overall structure of the reconstruction process. In fact, the work of Scheres *et al.* [41] points out that the algorithm is readily adapted to include additional missing data, such as conformational variants of the molecule.

Redundant representations of the model parameters, and operations that impose consistency with a physical (3D) model, are also shared features. Scheres *et al.* [41] obtain the 3D model using ART [12], a least-squares projection technique. The corresponding operation in our reconstruction algorithm is the linear compression-expansion mapping $\mathbf{E} \cdot \mathbf{C}$. The redundancy question did not arise in the same way for the minimal model studied previously [16], with only a single rotation axis. There the intensity tomographs did not intersect and the speckle structure had to be imposed by an additional support constraint on the Fourier transform of the intensity distribution.

Finally, we note that Fung *et al.* [21] have developed a technique that, like our expectation maximization approach, uses the entire body of data in a single update of the model parameters.

## 2.3 Test particles

When simulating the single-particle experiments it is important to distinguish between the resolution limit imposed by the maximum measured scattering angle and the resolution intrinsic to the scattering particle as a result of its dy-

namics. The scattering cross section for a complex molecule is generally significantly smaller, at large momentum transfer, than what is predicted by atomic form factors and a static molecular structure. This phenomenon is well known in crystallography, where the coherent illumination of numerous molecules in various states of perturbation is equivalent — when considering the information recorded in Bragg peaks — to a single molecule with blurred contrast. The same effect, but in the temporal domain, will diminish the scattering at large angles in the single-particle experiments.

The dynamics of molecules subject to intense x-ray pulses is complicated by the presence of several physical processes [24]. In the case where the degree of ionization of the atoms is relatively low during the passage of the pulse, the x-ray scattering is dominated by the atomic cores and can be analyzed by modeling the atomic motions. Let $A_{\mathbf{k}}(t)$ be the amplitude of the incident radiation in a particular momentum mode $\mathbf{k}$. The cross section for scattering a photon into mode $\mathbf{k}+\mathbf{q}$ with frequency $\omega$ contains as a factor the expression

$$\Gamma(\mathbf{q}) = \left| \int dt\, A_{\mathbf{k}}(t) e^{i\omega t} \sum_p f_p(\mathbf{q}) \exp\left[i\mathbf{q}\cdot\mathbf{r}_p(t)\right] \right|^2 , \tag{2.28}$$

where $f_p(\mathbf{q})$ is the atomic form factor of atom $p$ whose position $\mathbf{r}_p(t)$ changes with time as a result of large scale ionization, *etc.* We are interested in modeling the $\mathbf{q}$-dependence of the molecular form factor (2.28). Without access to detailed simulations of the Coulomb explosion, we have adopted for our data modeling the simple one-parameter form,

$$\Gamma(\mathbf{q})/\Gamma(0) = \exp\left(-B|\mathbf{q}|^2\right) S(\mathbf{q}) , \tag{2.29}$$

where $S(\mathbf{q})$ is the normalized static ($t = 0$) structure factor of the molecule. This Gaussian form results if the dynamics of the positions $\mathbf{r}_p(t)$ can be approximated by independent Gaussian fluctuations over the coherent time scale $T$

of the pulse. The period $T$, or the time during which the function $A_{\mathbf{k}}(t)e^{i\omega t}$ is approximately constant, is significantly shorter than the pulse duration in a non-seeded free electron laser [44].

We expect more detailed dynamical simulations of the scattering cross section to show significant departures from the form above, of a simple Gaussian factor modulating a static structure factor. Rather, the effective structure factor of an exploding molecule should resemble that of an atomistic density that is primarily blurred radially, with contrast at the surface of the molecule experiencing the greatest degradation [24]. Given such complications, our test particle modeling ignores atomicity and treats the particle more simply as a distribution of positive contrast on a specified support with a phenomenologically defined intrinsic resolution given by the form (2.29).

### 2.3.1 Binary contrast particles

It is a great convenience, when developing algorithms, to have a simply defined ensemble of problem instances that offers direct control over the key parameters. We have chosen to work with an ensemble having a single parameter that controls the complexity of the particle, where our measure of complexity is the dimensionless radius $R$ which specifies the physical particle radius in units of the intrinsic resolution. Our particles have the following properties:

(1) Spherical shape,

(2) binary contrast, and

(3) Gaussian form factor.

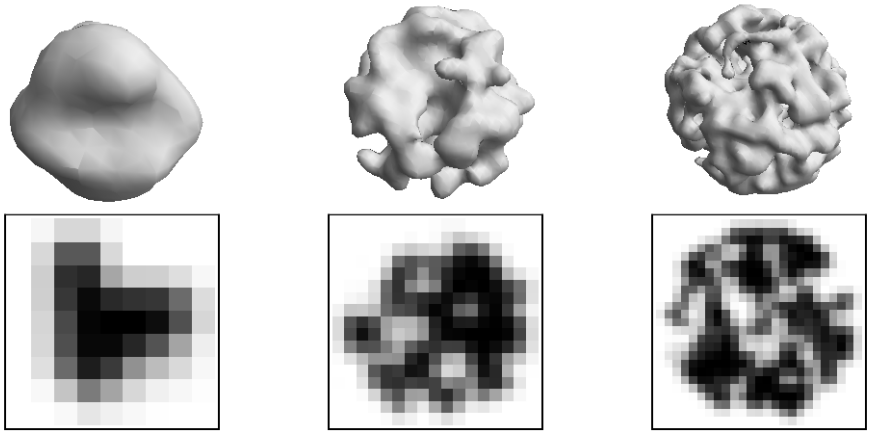


Figure 2.2: Examples of the random binary contrast particles used in our simulations. The contrast is nearly uniform inside a labyrinthine region that fills half the volume of a sphere. Shown are particles of radius $R = 4, 8, 12$ (left to right), where the dimensionless $R$ is in units of the intrinsic resolution, *i.e.* the scale of the voxels shown in cross-section for each particle in the lower panel. The particle renderings in this chapter, such as those above, show the iso-contrast surface appropriate for the labyrinth walls (half the maximum contrast).

Property 1 is chosen to make the reconstructions as hard as possible, both for the assembly of the 3D intensity and later in the phase retrieval stage (the spherical support offering the fewest constraints [18]). We chose property 2 to mimic a large biomolecule that because of damage can only be resolved into solvent (empty) and non-solvent regions. This property also has the convenience that most of the information about the structure is conveyed by rendering a single 3D contour at an intermediate contrast. Property 3 defines the intrinsic resolution.

The construction of a test particle begins by choosing a value for the dimensionless radius $R$; the final contrast values will be defined on a cubic grid of size $2R + 1$. Our test particle construction algorithm is diagramed in pseudocode (algorithms 1 and 2) . It uses the particle support $S$ (voxels inside the sphere of

radius $R$) and a Gaussian low-pass filter $F(\mathbf{q})$ to impose the form factor (2.29) on the Fourier transform of the contrast. To keep the dynamic range of the intensity measurements in our simulations constant for different particle sizes $R$, we parameterize the filter as

$$F(\mathbf{q}) = \exp\left[-1.5(|\mathbf{q}|/q_{\max})^2\right], \tag{2.30}$$

where $q_{\max} = \pi/R$. Since only scattering with $|\mathbf{q}| < q_{\max}$ is measured, the discarded power is always

$$\frac{\int_1^\infty \exp\left(-3q^2\right) q^2 dq}{\int_0^\infty \exp\left(-3q^2\right) q^2 dq} = 11\%. \tag{2.31}$$

**Output**: particle contrast on cubic grid $C$

$C \leftarrow$ `RandomContrast`

**for** $i \leftarrow 1$ **to** $4$ **do**
  $C \leftarrow$ `BinaryContrast(`$C$`)`
  $C \leftarrow$ `LowPassFilter(`$C$`)`
**end**

**return** $C$

**Algorithm 1**: test particle construction.

In Figure 2.2 we show examples of binary contrast test particles constructed for three values of the dimensionless radius $R$. The largest particles considered in this study had $R = 8$ because the reconstruction computations grow, in both time and memory, very rapidly with $R$. In Section 2.8 we discuss the scaling of the computations with $R$.

**Input**: arbitrary contrast on $C$, support $S$

**Output**: `BinaryContrast`$(C)$

**foreach** $\mathbf{r} \notin S$ **do** $C[\mathbf{r}] \leftarrow 0$

$v \leftarrow$ `MedianValue`$(C[\mathbf{r} \in S])$

**foreach** $\mathbf{r} \in S$ **do**

  **if** $C[\mathbf{r}] < v$ **then**
    $C[\mathbf{r}] \leftarrow 0$
  **else**
    $C[\mathbf{r}] \leftarrow 1$
  **end**

**end**

**return** $C$

**Algorithm 2**: binary contrast projection.

### 2.3.2 Degraded resolution biomolecules

To put our dimensionless radius $R$ in perspective, the same low-pass filter used for test particles was applied to the roughly $0.8$ MDa biomolecule GroEL [36]. After binning the coordinates of the non-hydrogen atoms of the PDB structure on a cubic grid of resolution 2Å, the discrete Fourier transform was computed and truncated at the size $2R + 1$. The result was then multiplied by the filter (2.30) and inverse transformed to give the contrast used in the simulation.

Figure 2.3 shows GroEL processed in this way for three values of $R$. Handedness in the protein secondary structure begins to appear at about $R = 6$. These degraded resolution models of GroEL, that mimic the effects of dynamics and a finite duration pulse, are of course completely phenomenological. It may not

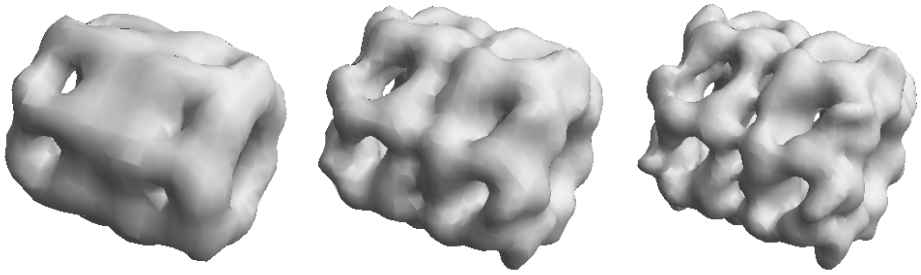


Figure 2.3: Contrast of the protein complex GroEL [36] degraded by the same type of low-pass filter used in the construction of random test particles (Figure 2.2). The axial length of GroEL is approximately 15 nm. At high damage, $R = 4$ (left), the contrast reveals only the gross particle shape ("cornEL"). Handedness of the protein secondary structure appears at about $R = 6$ (middle) and is fully evident by $R = 8$ (right).

even be true that the diffraction signal can be modeled by an appropriately blurred contrast function. This will be the case, for example, if the damage dynamics strongly varies with the orientation of the particle. Finally, we cannot ignore the fact that, as a result of thermal motion and solvent, at some level of resolution even the model of a unique ($t = 0$) diffraction signal breaks down.

## 2.4 Experimental parameters

### 2.4.1 Detector parameters

The detector geometry, pixel dimensions, and position relative to the scattering particle determine the spatial frequency samples $\mathbf{q}_i$ of the experimental data.

Our simulations use a detector model with three parameters: oversampling factor $\sigma$, maximum scattering angle $\theta$, and a dimensionless central data cutoff $\alpha$.

The oversampling factor has the most direct interpretation in real space. Oversampling $\sigma$ corresponds to embedding the particle, with contrast defined on a grid of size $2R + 1$, on a grid magnified in size by the factor $\sigma$. This also defines the dimensions of the 3D intensity grid, on which most of the computations of the EMC algorithm take place. Since Fourier transforms play no role in these computations there is no incentive to make the dimensions of the intensity grid a product of small primes. It is more natural in the EMC calculations, which have rotational symmetry about zero frequency, to have intensity grids of odd dimension with indices that run between $-q_{\max}$ and $+q_{\max}$. Here $q_{\max}$ is given by $\sigma\, R$ rounded up to the nearest integer. Speckles in the intensity distribution will have a linear size $\sigma$ in grid units.

A real detector does not measures point samples with respect to spatial frequency but convolves the true intensity signal with a point spread function defined by the pixel response [26]. To minimize this effect the oversampling in experiments should be kept large. Another reason for keeping $\sigma$ large is algorithmic: the expansion and compression steps of the EMC algorithm, which interpolate between tomographic and grid samples, introduce errors that are also minimized when the oversampling is large. In this study we used $\sigma = 6$.

The maximum scattering angle is determined by the radius of the detector, $L$, and the distance $D$ between the particle and the detector, by $\tan\theta = L/D$. We define $L$ to be the radius of the largest disk that fits inside the actual detector. This corresponds to discarding data recorded in pixels outside the disk, in the corners of the detector. With this minor truncation of the data, all of it can be

embedded in the 3D intensity grid for any particle orientation (relative to the reference orientation).

The actual choice of spatial frequencies $\mathbf{q}_i$ used by the EMC algorithm is largely arbitrary, and so we start by considering detector pixels at arbitrary positions $[x_i, y_i]$ in the detector plane. Up to a constant factor, the photon momentum detected at pixel $i$ is

$$\mathbf{p}_i = \frac{[x_i,\, y_i,\, D]}{\sqrt{x_i^2 + y_i^2 + D^2}} \,, \tag{2.32}$$

and the corresponding spatial frequency, or momentum transfer from the incident beam, is $\mathbf{q}_i = \mathbf{p}_i - \mathbf{p}_0$, where $\mathbf{p}_0$ is given by (2.32) with $x_i = y_i = 0$. Intensities at these spatial frequencies will be represented as interpolated values with respect to the 3D intensity grid. Since the latter has unit grid spacing, we choose an appropriate rescaling of the $\mathbf{q}_i$ that is well matched with this. Because most detectors will have pixel positions on a square lattice with some pixel spacing $d$, our simulations are based on this model. We note, however, that the EMC algorithm operates with general tables of frequencies $\mathbf{q}_i$, and whether these are derived from a standard detector or a more complex tiled design is invisible to the workings of the algorithm. For the square-array detector $x_i = m_i\, d$, $y_i = n_i\, d$, where $m_i$ and $n_i$ are integers, and the rescaled spatial frequencies are given by

$$\mathbf{q}_i = \frac{[m_i,\, n_i,\, D/d]}{\sqrt{(m_i^2 + n_i^2)(d/D)^2 + 1}} - [0, 0,\, D/d] \,. \tag{2.33}$$

These samples lie on the surface of a sphere that passes through the origin and the scaling is such that samples near the origin match the 3D integer grid:

$$\mathbf{q}_i \approx [m_i,\, n_i,\, 0] \,. \tag{2.34}$$

The pixels at the edge of the detector have

$$\sqrt{m_i^2 + n_i^2} = L/d \,, \tag{2.35}$$

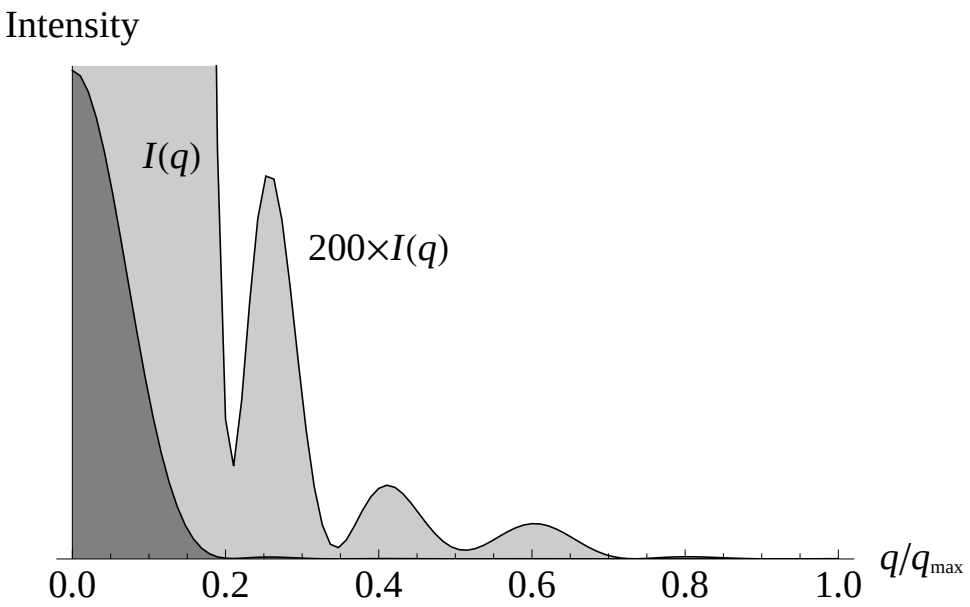



Figure 2.4: Radial intensity scan for an $R = 8$ test particle on a linear scale. Our simulations only use data collected outside the central speckle, $q/q_{\max} > \alpha/R = 0.18$.

and should correspond to frequencies at the highest resolution shell, or $|\mathbf{q}_i| = q_{\max}$. This condition, evaluated for (2.33) with $D = L\cot\theta$, reduces to

$$q_{\max} = (L/d)\cos\theta\sec(\theta/2)\ . \tag{2.36}$$

For a small maximum scattering angle this reduces to the equality between the pixel size of the detector, $2(L/d)$, and the number of samples in one dimension of the intensity grid, $2q_{\max}$. The $\theta$-dependence of expression (2.36) is a result of the spherical shape of the Ewald sphere.

The forward scattering by the uniform or uninteresting part of the charge density of a compact particle is so much more intense than the scattering at larger angles from the non-uniform, interesting part, that most detectors need to have the central pixels blocked (beyond what is needed to avoid the incident beam). Figure 2.4 shows a simulated intensity scan passing through the origin,

for one of the test particles described in Section 2.3. The huge central speckle contains essentially only information about the total charge and almost no information about the structure of the particle. A natural size for the detector block is such that scattering at frequencies inside the main speckle is discarded. We obtain the cutoff frequency $q_{\min}$ by evaluating the scattering amplitude of a uniform ball having the same radius $R$ as the test particle. The first vanishing of this amplitude determines $q_{\min}$:

$$(R/q_{\max})q_{\min} = \alpha \tag{2.37}$$

where $\alpha \approx 1.43$ is the first non-zero root of $\pi x = \tan \pi x$. Since $q_{\max} = \sigma\, R$, more generally we define

$$q_{\min} = \sigma\, \alpha\ , \tag{2.38}$$

which shows that the low frequency cutoff is $\alpha$ times the speckle size in grid units ($\sigma$).

We conclude this section by reviewing the procedure for generating the spatial frequencies $\mathbf{q}_i$ used by the EMC algorithm. Prior to this a dimensionless test particle radius $R$ has been selected. The half-size of the intensity grid is then given by $q_{\max} = \sigma\, R$, and our simulations used $\sigma = 6$. Given the maximum scattering angle $\theta$ we then determine the detector radius in pixel units, $L/d$, from (2.36), as well as the detector-particle distance $D/d$ using $D = L \cot \theta$. All our simulations used $\theta = 45°$. Having determined $L/d$, we determine (for our choice of square array detector) the indices $m_i$, $n_i$ satisfying $m_i^2 + n_i^2 < (L/d)^2$. These are used in formula (2.33) to give the table of frequency samples in the reference orientation of the particle. Finally, to model the discarded central data we remove from the table all samples with $|\mathbf{q}_i| < q_{\min}$, where $q_{\min} = 1.43\sigma$.

### 2.4.2 Diffracted signal strength

A key experimental parameter is the flux of photons incident on the particle. For the purpose of simulating the reconstruction process, however, a more convenient form of this parameter is the average number of photons scattered to the detector in one measurement, $N$. This normalization of the diffraction signal can be carried out once the detector's spatial frequency samples $\mathbf{q}_i$ are determined.

In order to generate diffraction data with the property that the mean photon number is $N$, we first compute the squared magnitude of the Fourier transform of the particle contrast embedded on the intensity grid (having size $2q_{\max}$). We interpret the numbers on this grid as the photon flux scattered into the respective spatial frequencies, at this point with arbitrary global normalization. A detector pixel at one such frequency sample will record an integer photon count drawn from the Poisson distribution having the (time and pixel area-integrated) flux as mean. The quantity we wish to normalize is the net flux at all the detector pixels. When this quantity is $N$, then the mean photon number per measurement will also be $N$.

Because the particle contrast is not spherically symmetric, the net diffracted flux to the detector pixels will fluctuate with the particle orientation. We avoid bias arising from this effect by sampling a few hundred orientations of the particle and applying the associated rotations to the frequency samples in order to estimate the orientation-averaged flux. This number is then used to rescale the flux values on our 3D intensity grid. With this normalization in place, we generate data by repeatedly sampling random orientations, rotating the frequency samples, and then drawing Poisson samples at each of the rotated frequencies for the means given on the normalized intensity grid. Linear interpolation of the

grid intensities is used to obtain the diffracted fluxes at the rotated frequencies.

## 2.5 Reconstruction parameters

In addition to the diffraction data, prior information about the particle provides additional parameters to the reconstruction algorithm. The only such information we consider in our simulations is the dimensionless particle radius $R$. This parameter is used in both the intensity reconstruction by the EMC algorithm, as well as the phase retrieval stage that reconstructs the particle contrast from the intensity.

### 2.5.1 Rotation group sampling

The EMC algorithm orients 2D data tomographs within the 3D intensity distribution using a discrete sampling of the rotation group. An optimal sampling is one where the samples are uniformly distributed and at a sufficient density to resolve the smallest angular features. Because speckles in the intensity distribution have linear dimension $\sigma$, features of this size (in voxel units) at the highest resolution shell, $q_{\max}$, determine the angular scale:

$$\delta\theta = \sigma/q_{\max} = 1/R \; . \tag{2.39}$$

The rotation group parameterization that is best suited for generating uniform samplings is based on quaternions. Unit quaternions are points on the unit sphere in four dimensions and encode 2-to-1 the elements of the continuous rotation group in three dimensions. Their key property is the fact that the distance

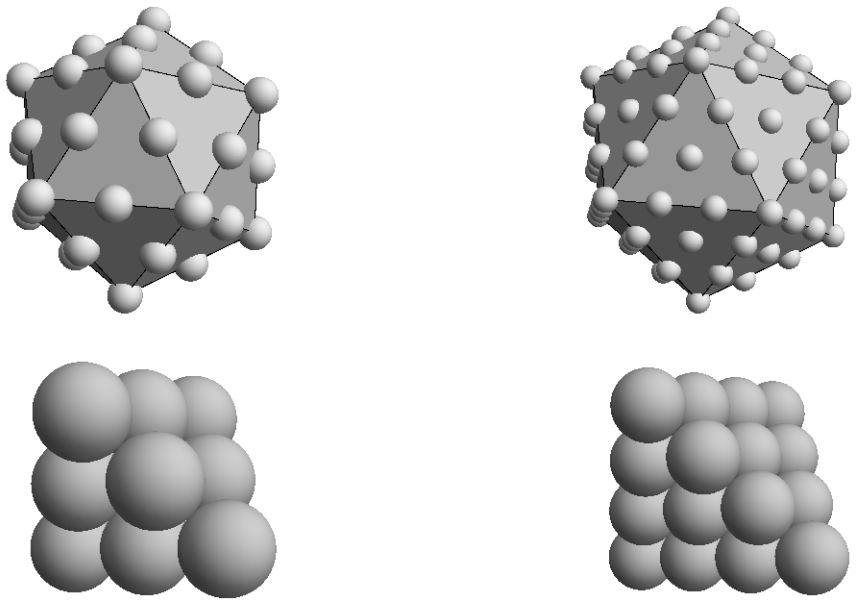


Figure 2.5: Sampling the 3D rotation group is equivalent to sampling the surface of a sphere in 4D. Shown in the top row is a scheme for sampling the surface of a sphere in one lower dimension based on subdivisions of the 20 faces of the icosahedron. The analogous construction in one higher dimension subdivides the 600 tetrahedral faces of the 600-cell, two examples of which are shown in the bottom row. The resolution of the sampling is specified by the number of subdivisions of each edge; shown are $n = 2$ (left) and $n = 3$ (right).

between quaternions $q$ and $q'$, in the usual sense, is simply related to the angle of the relative rotation between the group elements associated with $q$ and $q'$. For small relative rotations $\delta\theta$ this relationship is:

$$\|q - q'\| \approx \delta\theta/2 \; . \tag{2.40}$$

Given a $\delta\theta$, the problem of selecting rotation group samples, such that any rotation is within a relative rotation $\delta\theta$ of some sample, is thus equivalent to the standard problem of constructing efficient coverings [9] of the 3-sphere. We

solve this covering problem by using a design based on a highly symmetric polytope, the 600-cell [10]. This polytope is the four dimensional analog of the icosahedron in that it approximates the curved surface of the sphere by a union of regular simplices — 3D tetrahedra rather than triangles in four dimensions. The regular tetrahedron is efficiently covered by points in the fcc arrangement; coverings with increasing resolution are shown in Figure 2.5. The resolution of the covering is parametrized by an integer $n$ that gives the number of subdivisions of each edge of the tetrahedron. Our 3-sphere coverings are obtained by rescaling the points that cover the tetrahedral faces of the 600-cell to unit length. Details of the construction, including the computation of the sample weights, are given in Appendix C.

There are only two properties of the rotation group sampling that have direct relevance to the EMC algorithm for intensity reconstruction: the angular resolution $\delta\theta$ and the number of rotation samples, $M_{\rm rot}$. Defining $\delta\theta$ as the covering radius of the sampling, the $n$-dependence is given by (see Appendix C):

$$\delta\theta(n) \approx 0.944/n \ . \tag{2.41}$$

When combined with the estimate (2.39), this implies that $n$ should roughly coincide with the dimensionless particle radius $R$. Moreover, since $n \approx R$, the number of samples (see Appendix C),

$$M_{\rm rot}(n) = 10(5n^3 + n) \ , \tag{2.42}$$

grows in proportion with the volume of the particle.

### 2.5.2 Particle support

Our phase reconstruction of the complex diffraction amplitude is carried out with the diffraction magnitude on the same grid as used by the EMC algorithm for the intensity reconstruction. The support constraint is therefore that the particle contrast can be non-zero only within a sphere of radius $R$ grid units. In our simulations, which were limited to $R \leq 8$, we increased the support radius by one or two units because precise knowledge of the support is usually not available in real experiments.

## 2.6 Details on implementation of algorithm

Our algorithm for reconstructing the scattering contrast of a particle begins by reconstructing the 3D intensity with the EMC algorithm for classifying diffraction data. This section describes in concise algorithmic language the EMC process already sketched in Section 2.2.3. For the relatively much easier final step, of reconstructing the particle contrast from the intensity, we use the difference-map (DM) phase reconstruction algorithm. A short description of the DM algorithm, described in greater detail elsewhere [19], is included for completeness.

### 2.6.1 EMC intensity reconstruction

The EMC algorithm builds a model of the 3D intensity from a large collection of non-oriented, shot-noise limited diffraction data. The orientational classification of the data is probabilistic, where the data are assigned probability dis-

tributions in the rotation group and these are systematically refined so as to maximize the likelihood of the intensity model. The EMC algorithm comprises three steps:

E-step: Expand the grid intensities into the tomographic representation: $W[\mathbf{q}] \rightarrow W_{ij}$.

M-step: Update the tomographic intensities by expectation maximization: $W_{ij} \rightarrow W'_{ij}$.

C-step: Compress the tomographic model back into a grid model: $W'_{ij} \rightarrow W'[\mathbf{q}]$.

We use pseudocode to describe these steps in the next sections. The notation matches the theoretical discussion in Section 2.2.3. Spatial frequencies are denoted by $\mathbf{q}$ and $\mathbf{p}$, detector pixel indices by $i$, rotation group samples by $j$, and $k$ is always a data index.

**E-step: model expansion**

In the E-step the grid model of the intensities is expanded into a redundant tomographic representation (model) to make the expectation maximization step (M) easier. Intensities $W_{ij}$ in the tomographic model are treated as independent variables by the M-step.

Each element of the tomographic model is associated with a particular detector spatial frequency $\mathbf{q}_i$ and rotation matrix $\mathbf{R}_j$. The $M_{\mathrm{pix}}$ frequencies $\mathbf{q}_i$ refer to the detector (or particle) in an arbitrary reference orientation; the $M_{\mathrm{rot}}$ matrices $\mathbf{R}_j$ are rotations relative to the reference orientation. The construction of the $\mathbf{q}_i$

for a simple square array detector is given in Section 2.4. Our rotation matrices are generated from pre-computed lists of quaternions that sample the rotation group at the desired resolution (see Appendix C). We use linear interpolation to extract intensity values at the rotated spatial frequencies $\mathbf{q}' = \mathbf{R}_j \cdot \mathbf{q}_i$.

**Input**: grid model $W[\mathbf{q}]$, reference tomograph spatial frequencies $\mathbf{q}_i$, rotation matrices $\mathbf{R}_j$

**Output**: tomographic model $W_{ij}$

**for** $j \leftarrow 1$ **to** $M_{\text{rot}}$ **do**

  **for** $i \leftarrow 1$ **to** $M_{\text{pix}}$ **do**

    $\mathbf{q}' \leftarrow \mathbf{R}_j \cdot \mathbf{q}_i$

    $W_{ij} \leftarrow \texttt{Interpolate}(W[\mathbf{q}], \mathbf{q}')$

  **end**

**end**

**return** $W_{ij}$

**Algorithm 3**: E-step, model expansion.

**M-step, expectation maximization**

The probabilistic classification of the data, and then their aggregation into an improved tomographic model, is performed in the M-step. Central in this process is the computation of the conditional probabilities $P_{jk}$. These are based on the current intensity model, $W_{ij}$. When the diffraction data (photon counts) $K_{ik}$ are averaged with respect to these probabilities (equation (2.19)), the result is a tomographic model $W'_{ij}$ with increased likelihood.

The most time-intensive parts of the computation, and indeed of the whole reconstruction algorithm, are the nested loops over $k$, $j$, and $i$ that would imply an operation count that scales as $M_{\mathrm{data}} \times M_{\mathrm{rot}} \times M_{\mathrm{pix}}$. However, the innermost loop, on the pixel index $i$, can be greatly streamlined in both places where it occurs by skipping all the pixels that have zero photons. We use a sparse representation of the photon counts that reduces the time scaling to $M_{\mathrm{data}} \times M_{\mathrm{rot}} \times N$, since most non-zero counts will be a single photon and the average total photon number is $N$.

Two copies of the intensity model are held in memory at any time: the current model for conditional probabilities, and the updated model obtained by photon averaging. In the innermost loop computations of the conditional probabilities only the logarithm of the current model is used. The actual computation in this most time-intensive step involves incrementing the conditional probabilities $P_{jk}$ by $\log W_{ij}$ for the pixels $i$ that recorded photons in diffraction pattern $k$ (or a multiple of this if multiple photons were recorded). After the conditional probability for a particular $k$ is computed, the second time-intensive step is executed. In this the updated model $W'_{ij}$ is incremented by $P_{jk}$, again, for only the pixels $i$ where photons were recorded (or a multiple of this).

**Input**: tomographic model $W_{ij}$, data $K_{ik}$, rotation group weights $w_j$

**Output**: updated model $W'_{ij}$, mutual information $I$

$I \leftarrow 0$

**for** $j \leftarrow 1$ **to** $M_{\text{rot}}$ **do**
  $S_j \leftarrow 0$
  **for** $i \leftarrow 1$ **to** $M_{\text{pix}}$ **do**
    $W'_{ij} \leftarrow 0$
  **end**
**end**

**for** $k \leftarrow 1$ **to** $M_{\text{data}}$ **do**
  **for** $j \leftarrow 1$ **to** $M_{\text{rot}}$ **do**
    $P_{jk} \leftarrow \texttt{CondProb}(W_{ij}, K_{ik})$
    **for** $i \leftarrow 1$ **to** $M_{\text{pix}}$ **do**
      $W'_{ij} \leftarrow W'_{ij} + P_{jk} K_{ik}$
    **end**
    $S_j \leftarrow S_j + P_{jk}$
    $I \leftarrow I + P_{jk} \cdot \text{Log}(P_{jk}/w_j)$
  **end**
**end**

**for** $j \leftarrow 1$ **to** $M_{\text{rot}}$ **do**
  **for** $i \leftarrow 1$ **to** $M_{\text{pix}}$ **do**
    $W'_{ij} \leftarrow W'_{ij}/S_j$
  **end**
**end**

$I \leftarrow I/M_{\text{data}}$

**return** $W'_{ij}$, $I$

**Algorithm 4**: M-step, expectation maximization.

The pseudocode shows how directly the mutual information $I(K,\Omega)|_W$ is computed from the conditional probabilities $P_{jk}$. In Section 2.7.2 we show how this quantity provides a useful diagnostic for reconstructions in addition to having intrinsic value as a measure of information. There are a few places not shown in the pseudocode that require special attention in the implementation. As an example, it is important to check for over / underflow in the computation of the conditional probabilities when the logarithms of the not yet normalized probabilities are exponentiated.

**Input**: data index $k$, tomographic model $W_{ij}$, data $K_{ik}$, rotation group weights $w_j$

**Output**: $P_{jk} = \texttt{CondProb}(W_{ij}, K_{ik})$

$S \leftarrow 0$

**for** $j \leftarrow 1$ **to** $M_{\mathrm{rot}}$ **do**
  $P_{jk} \leftarrow \mathrm{Log}(w_j)$
  **for** $i \leftarrow 1$ **to** $M_{\mathrm{pix}}$ **do**
    $P_{jk} \leftarrow P_{jk} + K_{ik}\,\mathrm{Log}(W_{ij}) - W_{ij}$
  **end**
  $P_{jk} \leftarrow \mathrm{Exp}(P_{jk})$
  $S \leftarrow S + P_{jk}$
**end**

**for** $j \leftarrow 1$ **to** $M_{\mathrm{rot}}$ **do**
  $P_{jk} \leftarrow P_{jk}/S$
**end**

**return** $P_{jk}$

**Algorithm 5**: conditional probability.

**C-step: model compression**

The C-step (2.22) is the reverse of the model expansion, or E-step, and both of these operations use far less time than the M-step that comes in between. Over the course of multiple EMC iterations, the combination of C-step followed by E-step has the effect of making the tomographic model of the intensity, $W'_{ij}$, consistent with a 3D model $W'[\mathbf{q}]$ defined on a grid. We use linear interpolation (as in the E-step) when collapsing the tomographically sampled intensities onto the grid.

Because of noise, the averaging of the data $K_{ik}$ in the M-step produces a tomographic model that does not respect the Friedel symmetry $W'[\mathbf{q}] = W'[-\mathbf{q}]$ when compressed to the grid model. This symmetry is restored at the conclusion of the C-step by replacing $W'[\mathbf{q}]$ and $W'[-\mathbf{q}]$ with their average.

### 2.6.2 Phase retrieval

We use the difference-map algorithm [19] to reconstruct the phases associated with the Fourier magnitudes obtained by the EMC algorithm, as well as the Fourier magnitudes in the central missing data region ($q < q_{\min}$).

Our pseudocode for the difference-map emphasizes its generic character as a method for reconstructing models subject to two constraints. One constraint, provided by the EMC intensities, is the Fourier magnitude constraint. The difference map implements this constraint by the operation `FourierProj`. `FourierProj` takes an arbitrary, real-valued input contrast $C$ and projects to another contrast $F$ called the Fourier estimate. The action of `FourierProj` is

**Input**: tomographic model $W'_{ij}$, reference tomograph spatial frequencies $\mathbf{q}_i$, rotation matrices $\mathbf{R}_j$

**Output**: grid model $W'[\mathbf{q}]$

**foreach** $\mathbf{q}$ **do**
  $W'[\mathbf{q}] \leftarrow 0$
  $S[\mathbf{q}] \leftarrow 0$
**end**

**for** $j \leftarrow 1$ **to** $M_{\mathrm{rot}}$ **do**
  **for** $i \leftarrow 1$ **to** $M_{\mathrm{pix}}$ **do**
    $\mathbf{q}' \leftarrow \mathbf{R}_j \cdot \mathbf{q}_i$
    $G \leftarrow \texttt{GridNeighbors}(\mathbf{q}')$
    **foreach** $\mathbf{p} \in G$ **do**
      $f \leftarrow \texttt{InterpolationWeight}(\mathbf{q}' - \mathbf{p})$
      $W'[\mathbf{p}] \leftarrow W'[\mathbf{p}] + f\, W'_{ij}$
      $S[\mathbf{p}] \leftarrow S[\mathbf{p}] + f$
    **end**
  **end**
**end**

**foreach** $\mathbf{q}$ **do**
  $W'[\mathbf{q}] \leftarrow W'[\mathbf{q}]/S[\mathbf{q}]$
**end**

$W'[\mathbf{q}] \leftarrow \texttt{FriedelSym}(W'[\mathbf{q}])$

**return** $W'[\mathbf{q}]$

**Algorithm 6**: C-step, model compression.

most transparent on the Fourier transforms of the input and output contrasts. In the data region $q_{\min} < q < q_{\max}$, the output $F$ inherits the Fourier phases of the input $C$ with the Fourier magnitudes provided by the square roots of the

EMC intensities. In the central missing data region, $q < q_{\min}$, both the Fourier phase and magnitude of the input are preserved in the output. Finally, for spatial frequencies above the data cutoff, $q > q_{\max}$, the Fourier amplitudes of $F$ are identically set to zero.

The other difference map constraint is implemented by the operation `SupportProj`. When acting on an arbitrary real-valued input contrast $C$, `SupportProj` outputs the support estimate $S$. The output $S$ is obtained by zeroing the contrast in $C$ outside the support of the particle and all the negative contrast within the support. Since the phase reconstructions are performed on exactly the same size grids as the EMC intensity reconstruction, the radius of the spherical support region implemented by `SupportProj` is the same dimensionless radius $R$ (increased by a few grid units) that defines our binary contrast test particles and degraded resolution biomolecules (Section 2.5.2).

**Difference map phase reconstruction**

The difference-map reconstruction begins with a randomly generated initial real-valued contrast $X$ and is otherwise completely deterministic. As $X$ is updated by the operations `FourierProj` and `SupportProj`, the corresponding Fourier and support estimates, $F$ and $S$, are generated. In reconstruction problems that reach fixed points $X^*$, where the magnitude $\epsilon$ of the update $\Delta X$ vanishes, either $F$ or $S$ can be output as the solution since they are the same when $\epsilon$ vanishes. This is not the case even for our phase reconstructions with simulated data. There are multiple sources of error that make it impossible for both constraints to be satisfied simultaneously. The intensity truncation for $q > q_{\max}$, for example, introduces a small inconsistency even when the diffraction data are

oriented perfectly by the EMC algorithm. Of the two alternatives to choose for the reconstructed contrast, we use the Fourier estimate $F$ for reasons that will be clear below, when we discuss the modulation transfer function (MTF) [43].

**Input**: constraint projections `FourierProj` and `SupportProj`

**Output**: real-valued particle contrast $C$, error series $E$

$C \leftarrow 0$

$M \leftarrow 0$

$X \leftarrow$ `RandomRealContrast`

**for** $i \leftarrow 1$ **to** `IterationCount` **do**

  $S \leftarrow$ `SupportProj` $(X)$

  $F \leftarrow$ `FourierProj` $(2\,S - X)$

  $\Delta X \leftarrow F - S$

  $X \leftarrow X + \Delta X$

  $\epsilon \leftarrow \|\Delta X\|$

  $E \leftarrow$ `Append`$(E, \epsilon)$

  **if** $\epsilon <$ `ErrorMax` **then**

    $C \leftarrow C + F$

    $M \leftarrow M + 1$

  **end**

**end**

$C \leftarrow C/M$

**return** $C$, $E$

**Algorithm 7**: difference map algorithm for phase reconstruction.

The behavior of the difference-map error metric $\epsilon = \|F - S\|$ typically has two regimes in phase reconstructions. In the early iterations $\epsilon$ decreases nearly monotonically, thereby improving the consistency between the Fourier and support estimates. In the second regime $\epsilon$ is relatively constant, with small ampli-

tude fluctuations suggestive of a steady-state. Because $\epsilon = \|\Delta X\|$ also measures the magnitude of the update, the iterated contrast $X$ and the estimates $F$ and $S$ are also fluctuating in this regime. To produce reproducible results, we average the Fourier estimate $F$ in the steady-state and call this the result of the phase reconstruction.

Since our spherical support is consistent with either enantiomer of the particle (and the intensity data does not distinguish these either), a successful reconstruction will be the inversion of the true particle in about half of all attempts that start with a different initial $X$. In those cases we invert the reconstruction before making comparisons.

**Modulation transfer function**

The Fourier estimates $F$ have, by construction, always the same Fourier magnitudes in the data region $q_{\min} < q < q_{\max}$ (provided by the EMC intensities). This means that the fluctuations of $F$ at these spatial frequencies are purely the result of phase fluctuations. By averaging the difference-map estimates $F$ (after the steady-state is established), we are performing the average

$$\mathrm{MTF}(\mathbf{q}) = \langle \exp\left(i\phi_{\mathbf{q}}\right) \rangle, \tag{2.43}$$

where $\phi_{\mathbf{q}}$ is the Fourier phase at spatial frequency $\mathbf{q}$. Phases that are reconstructed well and fluctuate weakly give $\mathrm{MTF} \approx 1$, while strongly fluctuating phases lead to a small $\mathrm{MTF}$. Since the degree of fluctuation is correlated with the magnitude of $\mathbf{q}$, we additionally perform a spherical average of $\mathrm{MTF}(\mathbf{q})$ to define a modulation transfer function that concisely conveys the quality of the phase reconstruction as a function of the resolution $q = |\mathbf{q}|$.

## 2.7 Simulations to establish our algorithm's efficacy

This section explores the conditions necessary to reconstruct particles in numerical simulations. We are primarily interested in understanding the behavior of the reconstruction algorithm as a function of the dimensionless particle radius $R$. For any $R$, the feasibility and quality of reconstructions depends critically on three additional parameters:

1. Does the average number of photons per measurement, $N$, satisfy the criterion (2.9) on the reduced information rate?
2. Do the $M_{\mathrm{rot}}$ discrete samples of the rotation group provide a sufficient approximation of the continuous group for particles of the given complexity?
3. Are the total number of measurements $M_{\mathrm{data}}$ sufficient to reconstruct the particle with acceptable signal-to-noise?

Although the parameters $N$ and $M_{\mathrm{data}}$ are determined by the experiment while $M_{\mathrm{rot}}$ is algorithmic, this distinction is artificial when we recognize that both the physical and computational components of the imaging process are subject to limited resources. We have studied the effects of these parameters systematically by reconstructing the binary contrast test particles described in Section 2.3. These particles resemble biomolecules at a resolution above the atomic scale and can be generated for any $R$. Our simulations culminate in a desktop computer reconstruction of the GroEL protein complex (Section 2.3) at a resolution corresponding to $R = 8$.

### 2.7.1 Data generation

All our simulations begin with the construction of the contrast of a particle at a specified dimensionless radius $R$ (Section 2.3). After embedding the contrast on a grid with the chosen oversampling (usually $\sigma = 6$), the squared Fourier magnitudes are computed as a model of the intensities. All particles in our simulations thus have only a single discernable structure/conformation at the measured resolution.

Simulated data — tables of photon counts — are generated by repeatedly Poisson-sampling the intensity model at a set of spatial frequency samples specified by our detector parameters. In each simulated measurement all the spatial frequencies are rotated by a random element of the 3D rotation group. We generate uniform rotation group samples by uniformly sampling points on the surface of the unit sphere in four dimensions (quaternions) and mapping these to orthogonal matrices (equation (C.1)). The rotation element used to produce each measurement is not recorded. Because the data is simulated with uniform rotation group samples, the group weights used by the reconstruction algorithm are the uniform sampling weights (C.13). Since the rotated spatial frequency samples will fall between the grid points of the intensity model, interpolation is used to define the mean of each Poisson-sampled photon count. Finally, by normalizing the intensity model as described in Section 2.4.2, our data have the property that the mean photon number per measurement is $N$.

The number of data used in the reconstructions ($M_{\text{data}}$) can be very large, sometimes exceeding $10^6$ when $R$ is large and $N$ is small. By using sparse records of the photon counts (Section 2.8.1), however, the total storage required for the data is not much greater than the storage needed for a single 3D intensity

model having the corresponding signal-to-noise. The reconstruction algorithm has, of course, no access to the 3D intensity model that was used to generate the data.

### 2.7.2 Convergence with rotation group sampling

We argued in Section 2.5.1 that an adequate sampling of the rotation group, for reconstructing particles of dimensionless radius $R$, is obtained when the rotation group sampling parameter $n$ (edge-subdivisions of the 600-cell) matches this value. Whereas the proportionality $n \propto R$ is clear, we present here some additional assessments that support the simple rule $n \approx R$. While larger $n$ are even better, the $n^3$ growth in the memory used by the algorithm motivates us to identify the smallest $n$ that achieves good results.

The most direct test is to perform a single iteration of the EMC algorithm, beginning with the true intensity model. For this we generated data with sufficient total recorded photons ($N \times M_{\mathrm{data}}$) that signal-to-noise is not a factor. Since the data are generated by the same intensity model that begins the EMC update, the only thing that can spoil the preservation of the intensity by the update is the insufficient sampling of the rotation group. In Figure 2.6 we show planar slices of the intensity model after one EMC update for a particle with $R = 8$. The extreme case $n = 1$, with only $M_{\mathrm{rot}} = 60$ samples, is clearly inadequate because large regions of the intensity grid are never visited by a rotation of one of the detector's spatial frequency samples. This shortcoming is eliminated at about $n = 4$ ($M_{\mathrm{rot}} = 3240$), however, the intensity in the highest resolution shell lacks the expected speckle structure. These features first become established at

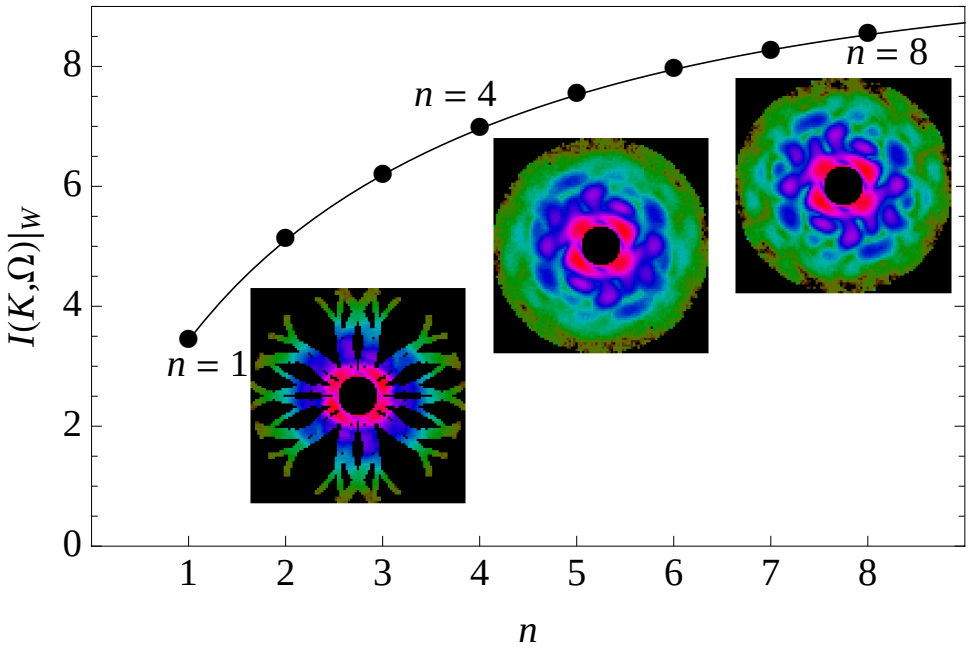



Figure 2.6: Convergence with respect to rotation group sampling. The plot shows the increase in the mutual information $I(K,\Omega)|_W$ as the discrete sampling of the rotation group is increased; the integer $n$ is inversely proportional to the angular resolution. Saturation of the mutual information with $n$ indicates the data $K$ have exhausted the available orientational information in the intensity $W$, here for the case of an $R = 8$ particle. The insets show corresponding cross sections of the intensity $W'$ generated by a single EMC update starting from the true intensity $W$. Speckles in the highest resolution shell appear at about $n = 8$. The intensity scale is logarithmic and missing data regions are rendered black.

level $n = 8$ ($M_{\text{rot}} = 25680$).

There is another assessment of the rotation group sampling that does not require the true intensity model (or converged reconstruction). In this test we ask to what extent the data can detect additional rotational structure just by increasing the sampling parameter $n$. The measure of rotational structure most relevant

to the available data is the mutual information $I(K,\Omega)|_W$. For a given intensity model $W$, this gives the information a typical measurement $K$ provides about its location in the rotation group. Clearly this depends on $W$ — possibly a poor approximation of the true intensity — as well as the noise in the data (mean photon number $N$). In any case, if the value of $I(K,\Omega)|_W$ is significantly increased upon increasing $n$, then there is information in the data that can detect additional rotational structure that should be exploited. In our implementation of the EMC algorithm $I(K,\Omega)|_W$ is calculated with negligible overhead in every iteration (see algorithm 4). To test convergence with respect to $n$ we simply increase $n$ and observe how much $I(K,\Omega)|_W$ increases. In Figure 2.6 we also show the behavior of $I(K,\Omega)|_W$ as a function of $n$ for the same particle and data used to create the intensity cross sections. The leveling off at $n \approx 8$ is consistent with our earlier observations.

### 2.7.3 Feasibility with respect to mean photon number

The total number of photons recorded in an imaging experiment is $N \times M_{\mathrm{data}}$. If the particle orientations were known for each of the diffraction data, then the quality of reconstructions would be independent of how the photon budget is allocated: simple signal averaging will give the same result if half the number of photons ($N/2$) are recorded on twice the number of data sets ($2M_{\mathrm{data}}$). This changes when the orientations are unknown, and we rely on the reduced information rate $r(N)$ for guidance (Section 2.2.1).

We computed $r(N)$ using equation (2.9) by numerically evaluating $I(K,\Omega)|_W$ for binary contrast test particles. The strict definition of $r(N)$ calls for an average

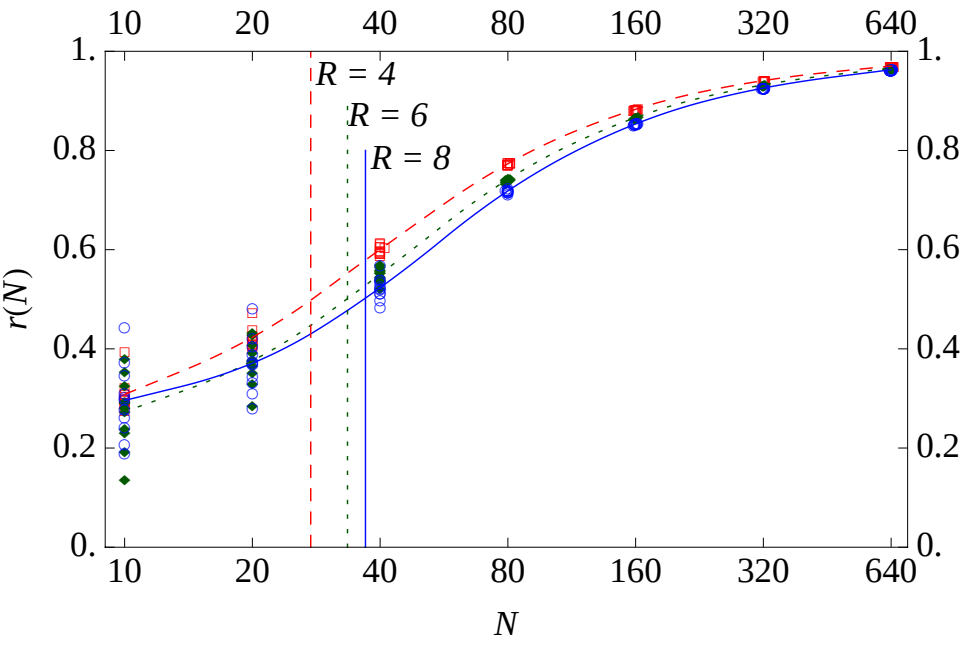



Figure 2.7: Numerically computed reduced information rate $r(N)$ as a function of mean photon number $N$, for particle sizes $R = 4, 6, 8$ — red (dashed), green (dotted) and blue (solid) respectively. The interpolating curve shows reduced information rate averaged over 11 particles — represented by the scattered points — at each $R$ for various $N$. The vertical lines intersect the curves at their respective $r(N) = 1/2$ points.

over an intensity ensemble $W$; in our case this corresponds to particles of a particular radius $R$. Figure 2.7 shows plots of $r(N)$ as a function of the mean photon number $N$ for $R = 4, 6$, and $8$. As shown by the small scatter in the results for the larger $N$, fluctuations of $I(K, \Omega)|_W$ within each radius-ensemble are very small, thus establishing $r(N)$ as a useful statistic when given only the particle radius. The single most important conclusion to draw from the $r(N)$ plots is that the reduced information rate is negligibly reduced (from unity) for even relatively small $N$. Taking $r(N) > 1/2$ as the feasibility criterion, we obtain mean photon number thresholds of $N = 27.5, 33.5$, and $36.9$ for the three sizes

of particles. From studies of the 1D minimal model [16], where this behavior can be analyzed in greater detail, we expect the threshold $N$ values to grow logarithmically with $R$.

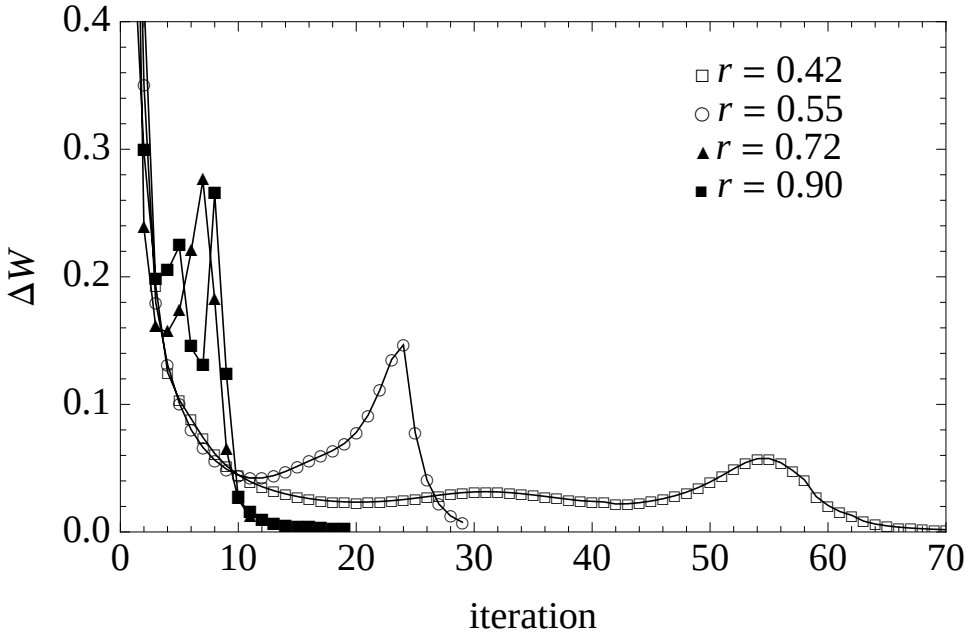



Figure 2.8: Update magnitudes $\Delta W$ for the intensity reconstruction of an $R = 8$ particle at four values of the reduced information rate: $r(25) = 0.42$, $r(45) = 0.55$, $r(80) = 0.72$ and $r(225) = 0.90$. The normally rapid convergence ($\Delta W \to 0$) of the EMC algorithm becomes protracted as $r(N)$ approaches the value $1/2$. The corresponding particle reconstructions are shown in Figure 2.9.

The consequences of $r(N)$ being below the feasibility threshold are noticed in two ways. First, there is a marked change in the behavior of the EMC intensity reconstruction algorithm, the progress of which we monitor by the time series of the update magnitudes:

$$\Delta W^2 = \langle |W'(\mathbf{q}) - W(\mathbf{q})|^2 \rangle_{\mathbf{q}}. \tag{2.44}$$

Here $W'$ is the update of $W$ and the average is uniform over all spatial frequencies between $q_{\min}$ and $q_{\max}$. Figure 2.8 contrasts $\Delta W$ time series for four EMC in-

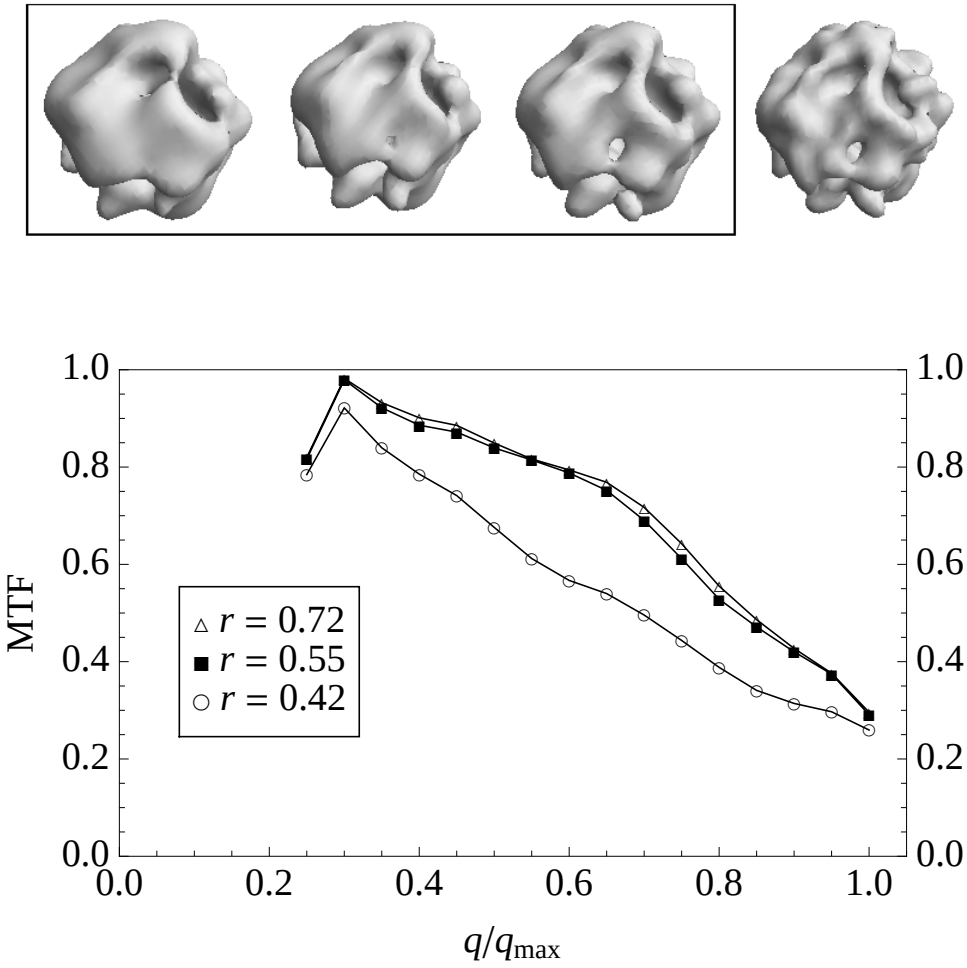



Figure 2.9: Reconstructions of an $R = 8$ test particle at three values of the reduced information rate $r(N)$, all other parameters unchanged. Shown in the top panel, from left to right, are $r(25) = 0.42$, $r(45) = 0.55$ and $r(80) = 0.72$; the true particle is reproduced on the right. The bottom panel shows the corresponding modulation transfer function (MTF) computed by the phasing algorithm. Behavior of the EMC algorithm for these reconstructions is shown in Figure 2.8.

tensity reconstructions of the same $R = 8$ particle, the data differing only in the value of $N$ with all other parameters, including the total photon number, identical. The rapid decrease to zero, seen in the reconstructions with $r(80) = 0.72$

and $r(225) = 0.90$, is typical when $r(N) > 1/2$. Likewise, the strongly non-monotonic behavior that stretches out over many iterations is normal for data that according to our $r(N)$ criterion is below the feasibility threshold. At the first broad minimum of $\Delta W$ in the plot for the case $r(45) = 0.55$ the reconstructed intensity is nearly spherically symmetrical (a powder pattern). It takes the EMC algorithm many iterations to develop speckle structure by the gradual amplification of small features.

A second manifestation of being below our feasibility criterion is a loss of resolution in the final particle reconstruction. This is demonstrated in Figure 2.9, which shows the results of three of the reconstructions described above. The degradation of high spatial frequency detail in the reconstructions at small $r(N)$ is a direct consequence of the reduced information rate in these data. With less total information available to reconstruct an accurate intensity, the resulting phase reconstruction of the particle is compromised.

### 2.7.4 Reconstruction noise and number of measurements

Even when the orientations of the diffraction patterns are known, shot noise in the intensity measurements will limit the signal-to-noise of the reconstructed particle. The resolution of the reconstruction will be compromised if the signal-to-noise at the highest spatial frequencies is poor. Because the intensities within a speckle are correlated, a natural quantity to consider is the total number of photons recorded in a typical, high spatial frequency speckle. If we denote this number by $\mu$, the associated shot noise magnitude is $\sqrt{\mu}$, and the signal-to-noise in the highest frequency features of the reconstruction is $\sqrt{\mu}$. The scaling of $\mu$

with the experimental parameters is obtained by dividing the total number of recorded photons, $N \times M_{\text{data}}$, by the number of speckles. Since the latter scales as $R^3$, the bound on the signal-to-noise given by oriented diffraction patterns and perfect phase reconstruction scales as

$$\text{SNR} \propto \sqrt{\frac{NM_{\text{data}}}{R^3}} \propto \sqrt{\frac{NM_{\text{data}}}{M_{\text{rot}}}} = S, \tag{2.45}$$

where the second proportionality follows from (2.42) and defines a convenient dimensionless measure of the signal. We see that $S^2$ is simply the average number of photons assigned, in a classification scheme, to each tomograph in a tomographic representation of the 3D intensity when the rotation group is sampled at the appropriate resolution.

When the diffraction data are not oriented, the reduction in the information rate, as measured by $r(N)$, leads to a loss in signal-to-noise. For small data sets (which may not be sufficient to reconstruct the particle) this effect is modeled by replacing $M_{\text{data}}$ with $r(N)M_{\text{data}}$ in (2.45). However, as argued in Section 2.7.3, for even modest $N$, $r(N)$ is close to unity and loss of signal-to-noise by this mechanism is minor.

Provided the photon numbers are reasonable, say $r(N) > 1/2$, reconstructions from non-oriented diffraction data will fail for the same reason they fail for oriented data: the signal-to-noise is simply too small. This is shown in Figure 2.10, where three $R = 8$ test particles are reconstructed at the same $r(N) = 0.75$ and decreasing values of $S$ (decreasing $M_{\text{data}}$). We see that below $S \approx 30$ the resolution of the reconstructed particle is far less than the intrinsic resolution of the $R = 8$ particle used to generate the data. The EMC algorithm succeeds at reconstructing a low resolution model of the particle at the smaller $S$ because the speckles at small spatial frequency may have sufficient numbers of photons

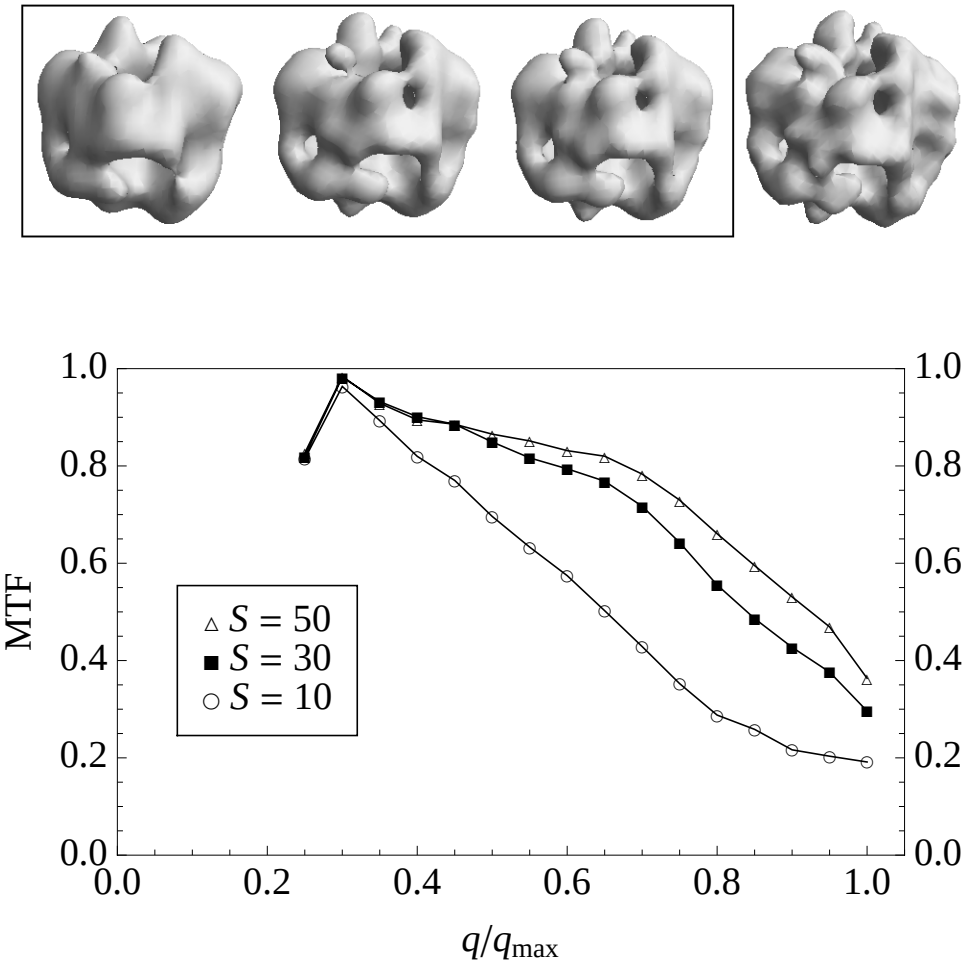



Figure 2.10: The same particle reconstructed with the same average number of photons per diffraction pattern, $N = 100$, and increasing data. From left to right in the top panel are shown reconstructions with signal-to-noise parameter $S = 10, 30, 50$; the true particle is on the right. The panel below shows the corresponding modulation transfer function (MTF) computed by the phasing algorithm.

while the speckles at high frequencies do not. We have defined $S$ so that the same standard, say $S \approx 30$, is meaningful for particles of arbitrary size.

### 2.7.5 A biomolecule at 2 nm resolution

We prepared an $R = 8$ GroEL particle (1 nm half-period resolution) and simulated diffraction data from its Fourier intensities as described in Sections 2.3 and 2.7.1, respectively. The mean photon number was set at $N = 100$. For particles of size $R = 8$ this implies a reduced information rate $r(N) = 0.75$ (Figure 2.7). The EMC intensity reconstruction algorithm was run with rotation group sampling up to $n = 8$ ($M_{\mathrm{rot}} = 25680$) and up to one million diffraction patterns ($M_{\mathrm{data}} = 10^6$). All other parameters used in these simulations are listed in Sections 2.3, 2.4 and 2.5.

As discussed above (Section 2.7.4), the EMC algorithm automatically reconstructs a lower resolution model when the number of data is such that only the speckles at low spatial frequency have adequate signal-to-noise. Because such lower resolution models have coarser angular features, a lower resolution sampling of the rotation group can be used to reconstruct them. These observations suggest a simple protocol for accelerating the solution process: first obtain a low resolution model and then refine this by increasing $M_{\mathrm{rot}}$ and $M_{\mathrm{data}}$ until the conditions for reconstructing up to the intrinsic resolution are reached. This strategy takes advantage of the fact that in practice very few EMC iterations are needed for the final, time-intensive refinements. Time and memory scaling of the algorithm with $M_{\mathrm{rot}}$ and $M_{\mathrm{data}}$ are discussed in Section 2.8.

We began the intensity reconstruction of the GroEL particle using $n = 5$ and $M_{\mathrm{data}} = 2 \times 10^5$ ($S = 54$). Most of the time saving is the result of the reduced rotational sampling, which scales as $n^3$. A single EMC iteration on a 3 GHz machine at these parameter values takes 40 minutes. The reconstruction of low frequency speckles can be monitored by visually inspecting planar cross sections

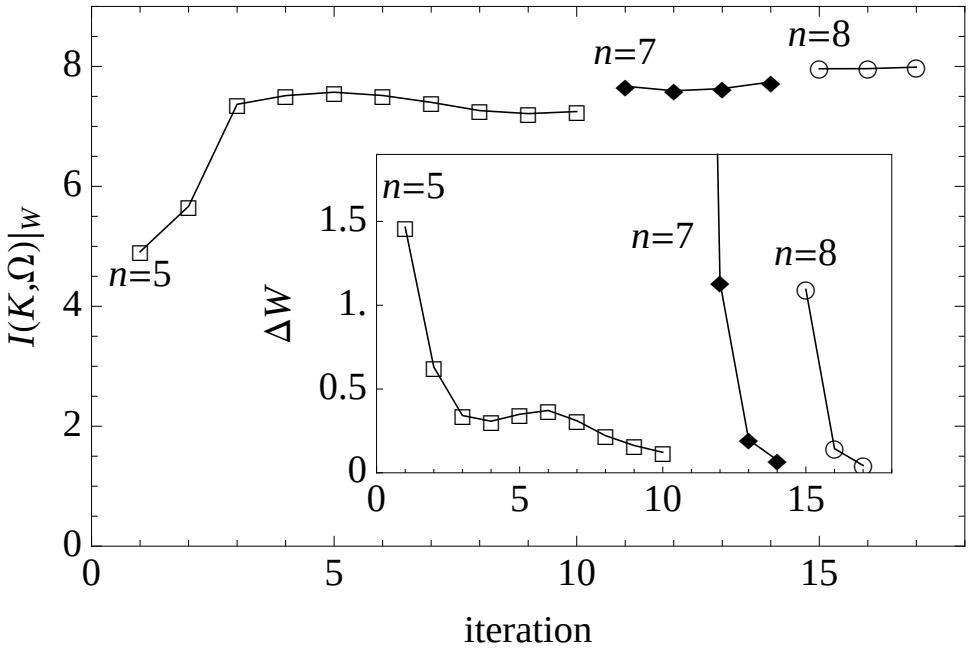



Figure 2.11: Mutual information $I(K,\Omega)|_W$ and EMC update magnitude $\Delta W$ as a function of iteration and increasing rotational sampling $n$ in the intensity reconstruction of the GroEL particle. The initial $n$ is small to save time and increased when the mutual information saturates and $\Delta W$ vanishes. With each increase of $n$ the number of data were also increased: $2 \times 10^5$, $5 \times 10^5$, and $10^6$. The intensity model after the 17th iteration was used in the phase reconstruction of the particle shown in Figure 2.13.

of the intensity model. A more quantitative approach makes use of the mutual information $I(K,\Omega)|_W$, where $W$ denotes the current intensity model. When this quantity saturates, the EMC algorithm is unable to improve the orientational accommodation of the data when evaluated at the limited rotational sampling. Stagnation of the EMC algorithm ($\Delta W = 0$) also implies the likelihood function, of $W$ given the data, cannot be improved with the current settings. To improve both the orientational assignments of the data and the likelihood function, the rotational sampling and the number of data must be increased.

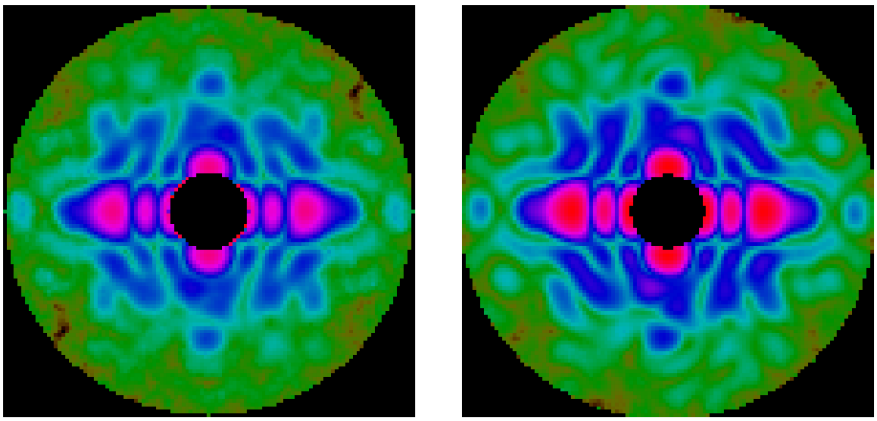


Figure 2.12: Log-intensity cross sections after the $n = 5$ (left) and $n = 8$ (right) stages of the intensity reconstruction of the GroEL particle. The colors red (high) to yellow (low) span 4 orders of magnitude. The low frequency speckles are already fully reconstructed at rotational sampling $n = 5$.

Figure 2.11 shows the effects of gradually increasing $n$ and $M_{\text{data}}$ on the value of $I(K,\Omega)|_W$. The refinements with $n = 7$ and $n = 8$ take very few iterations to reach the next (higher) plateau. We used the vanishing of $\Delta W$ as the stopping criterion for all the EMC stages, including the refinements. One EMC iteration in the final refinement stage ($n = 8$, $S = 62$) took about 24 hours. Figure 2.12 compares intensity cross sections after the early ($n = 5$) and late ($n = 8$) stages of rotational refinement.

The GroEL particle was reconstructed from the final EMC intensity model using the phase retrieval algorithm described in Section 2.6.2. The only other input to this algorithm is the support of the particle contrast and the constraint that the contrast is non-negative. Altogether there are four sources of error that can degrade the quality of the reconstructed particle. Three of these are responsible for errors in the intensity model: finite sampling of the rotation group,

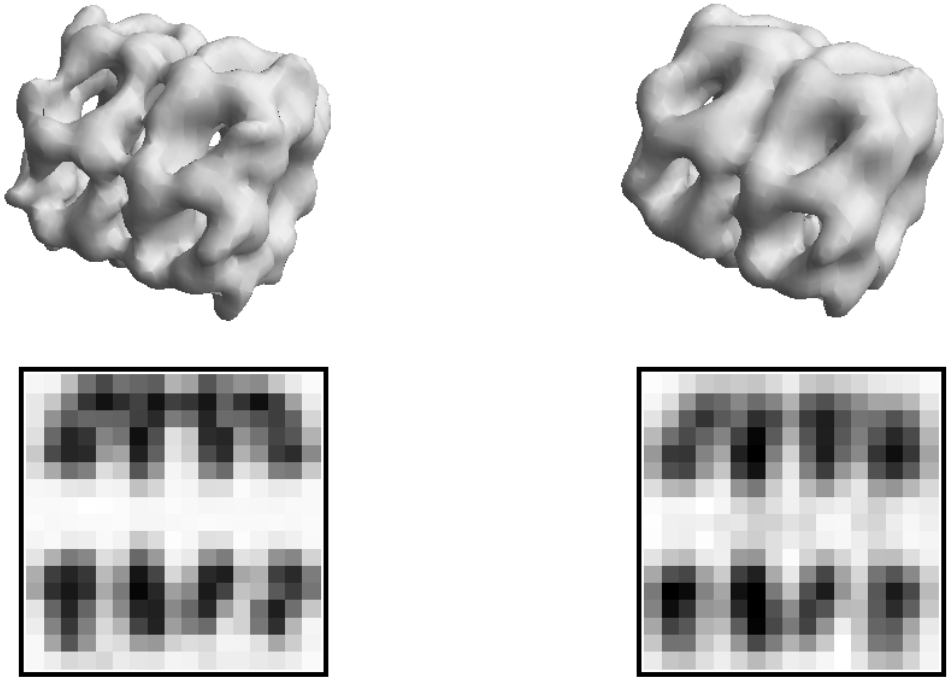


Figure 2.13: The $R = 8$ GroEL particle (top left) compared with the results of our reconstruction (top right). The reconstruction was rotated to bring the two particles into alignment. The resolution of the reconstruction is degraded by about a factor of two relative to the model used to generate the data. Cross sections of the contrast are compared in the bottom row (left: model, right: reconstruction).

finite data sets, and grid interpolation errors (finite oversampling). The fourth error source is the truncation of the data at spatial frequencies outside the range $q_{\min}$ to $q_{\max}$. The missing information for $q > q_{\max}$ has the greater effect, since the non-negativity constraint is very effective at reconstructing the missing beamstop ($q < q_{\min}$) intensities when this region includes few speckles. Because the weak intensities beyond $q_{\max}$ are not reconstructed — the algorithm treats them as zero — a small incompatibility is introduced in the constraints used by the difference-map phasing algorithm. This together with the other sources of error

lead to the nonvanishing of the difference-map error metric $\epsilon$ shown in Figure 2.14, and averaging with respect to residual phase fluctuations is required to arrive at a reproducible result. The phase retrieval MTF function (Section 2.6.2) computed during the averaging period provides a comprehensive assessment of the internal self-consistency of the entire reconstruction process.

The reconstructed $0.8$ MDa GroEL particle is compared in Figure 2.13 with the $R = 8$ resolution model used to generate the diffraction data. There is clearly a loss in resolution as a result of all the factors described above. From the phase retrieval MTF function, shown in Figure 2.15, we see that contrast begins to deteriorate beginning with spatial frequencies about half the maximum of those measured (1 nm). Since the MTF begins to decline at about $0.5\, q_{max}$, the reconstructed resolution is conservatively half the half-period resolution, or 2 nm.

## 2.8 Computational requirements

Our simulations show that particles can be reconstructed at low resolution, $R < 8$, and modest computational resources even when as few as $N = 100$ photons are recorded on the average diffraction pattern. Because the parameters $R$ and $N$ are dictated by the physical properties — including damage mechanisms — of the sample and available light source and are therefore least under the control of the imaging experiment, it makes sense to assess the feasibility of real reconstructions as a function of these parameters. We will see that the computational resources are essentially independent of $N$ and scale as simple powers of $R$. This analysis assumes a fixed oversampling $\sigma$ and fixed signal-to-noise in the reconstructed contrast.

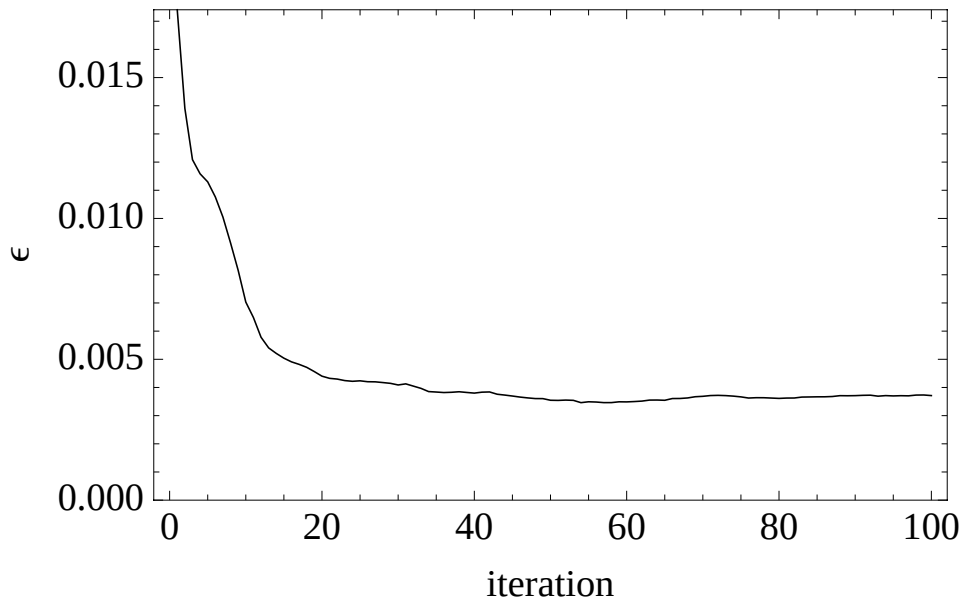



Figure 2.14: Difference map error metric $\epsilon$ in the phase retrieval of the GroEL intensity model obtained by the EMC algorithm. A steady state of residual phase fluctuations is reached after about 50 iterations. Averaging the phase reconstruction over the subsequent 200 iterations produced the particle contrast shown in Figure 2.13 and the MTF function in Figure 2.15.

### 2.8.1 Memory scaling

The data storage demands are modest and minor relative to the memory used by the algorithm when photon counts are recorded in a sparse format. A sparse-encoded measurement comprises on the order of $N$ integers identifying the detector pixels that have non-zero counts. For the pixels with single counts the pixel index provides a complete record; the small minority of pixels with multiple counts require an additional integer. Sparse data storage therefore scales in proportion to the total number of measured photons, $M_{\mathrm{data}} \times N$. At fixed signal-to-noise the number of photons aggregated per grid point in the intensity reconstruction is fixed. Since the number of points in the intensity grid scales

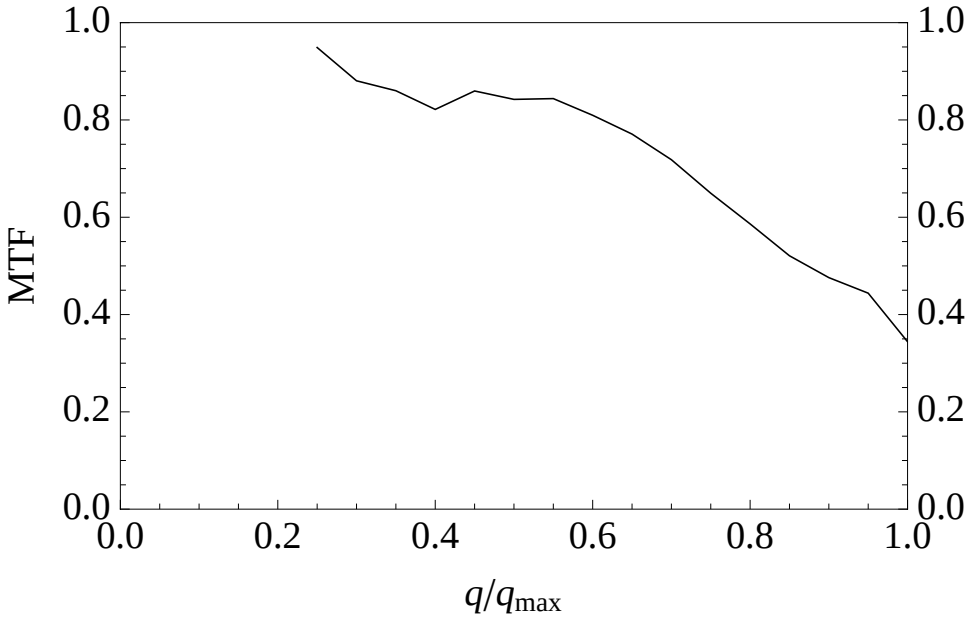



Figure 2.15: Phase retrieval MTF function for the GroEL particle reconstruction.

as $R^3$, the total number of detected photons — and the sparse storage space — also scales as $R^3$. In a parallel implementation of the reconstruction algorithm it makes sense for all the data to be resident in each processor.

The largest set of variables used by the algorithm are the current and updated tomographic representations of the intensity models, $W_{ij}$ and $W'_{ij}$. These have size $M_{\rm pix} \times M_{\rm rot}$ and scale as $R^2 \times R^3 = R^5$. The actual memory used depends strongly also on $\sigma$, which we assume is kept fixed as $R$ is varied. Our largest simulations ($R = 8, \sigma = 6$) used about 1 Gb with the memory dominated by these arrays.

### 2.8.2 Time scaling

The least certain part of our analysis is estimating the number of iterations needed by the EMC algorithm. The simulations described in Section 2.7.3 are consistent with the hypothesis that the number of iterations is fixed and small provided $N$ exceeds a modest information theoretic threshold (see Figure 2.8). Because this minimum $N$ (see Figure 2.7) is believed to grow only logarithmically with $R$, we assume the criterion can always be satisfied and thus the number of EMC iterations is practically independent of $R$.

The most time-intensive operation in each iteration of the EMC algorithm is the expectation maximization step (M) where the photon counts in each measurement, $K_{ik}$, are cross-correlated with the model log-intensities in each rotational sample, $\log W_{ij}$. The number of operations scales as $N \times M_{\mathrm{data}} \times M_{\mathrm{rot}}$ after the sum over the $i$-index, using the sparse data representation, is reduced to a sum of $N$ terms. As argued above, $N \times M_{\mathrm{data}}$ scales as $R^3$ and $M_{\mathrm{rot}}$ as $R^3$, giving a time per iteration that scales as $R^6$. This represents the time scaling of the reconstruction, since the number of iterations is independent of $R$ and the other two steps of the EMC update (E and C) are much faster because they do not involve the data.

### 2.8.3 Parallel implementation

Since both the memory and time scaling is dominated by operations on the tomographic intensity models $W_{ij}$ and $W'_{ij}$, in a parallel implementation these and the cross-correlations on them would be distributed among the processors so as to minimize message passing. A natural approach is to have a separate master

node perform the **E** step in such a way that blocks of the tomographic models (ranges of the $j$-index) are sent to different processors. After cross-correlating with all the data, which requires no message passing, each processor will send its results back to the master node for aggregation in the **C** step. In this scheme both memory per processor and time of the reconstruction are reduced in proportion to the number of processors.

## 2.9 Conclusions

The aim of this study was to provide a detailed assessment of the feasibility and quality of reconstructions for the proposed single-particle imaging experiments by testing the performance of a particular algorithm developed for this purpose. The many dozens of reconstructions required to map out the parameters space could not have been carried out had the operation of the algorithm not become a fairly routine process. We never encountered a situation where the intensity reconstruction (EMC algorithm) had to be abandoned or restarted, or where the subsequent phase reconstruction did not reproduce the true particle to the expected resolution. Our attention to experimental details (*e.g.* missing central data) in the simulations gives us confidence that the algorithm developed here will also succeed with real data.

A variant of our algorithm was previously studied in the context of a minimal model having a single rotation angle as missing data [16]. In the introduction we argued that in some respects the minimal model reconstructions might be harder than the reconstructions in a realistic model, where the diffraction data span the entire 3D rotation group. This scenario has been confirmed by

our simulations. Recall that in the minimal model a 2D intensity distribution is sampled by non-intersecting 1D diffraction patterns, while in single-particle imaging the intensity is 3D and the 2D diffraction patterns intersect pairwise along arcs. In the minimal model the tomographic representation of the intensity is non-redundant, while in the 3D problem the tomographs are highly redundant and mutual consistency has to be imposed with the E and C steps of the EMC algorithm. The structure of the intersecting diffraction patterns and redundant variables in the 3D problem provide a mechanism, absent in the minimal model, that accelerates the reconstruction. The redundancy is greatest near the origin, where many diffraction patterns pass through the same speckles. Because the signal-to-noise is also greatest in this region, the reconstruction can begin there in a consistent fashion and then progress to higher resolution shells. By contrast, the rotational classification of diffraction patterns in the minimal model is not incremental and requires many more model-update iterations.

Reconstruction algorithms for the first round of experiments will have to address two additional complications not considered in our simulations. The first is the large shot-to-shot fluctuation in the incident photon flux of the source when the FEL process is unseeded [44]. This adds another missing datum to the three orientational parameters, per diffraction pattern, that the expectation maximization step of the algorithm will have to reconstruct. The second complication is the background of photon counts arising from non-target particles, upstream beam optics and inelastic scattering. To deal with this the reconstruction algorithm should make use of the averaged dark count (no target particle). This represents additional data and requires a modification of the conditional probability computations.

## 2.10 Acknowledgements

The work in this chapter was supported by Department of Energy grant DE-FG02-05ER46198 and was completed while Veit Elser was a guest of the Centre for Free-Electron Laser Science at DESY, Hamburg. I would also like to thank the referees of [28] for their valuable patience and advice.

CHAPTER 3

# EXPERIMENTAL DEMONSTRATION OF CRYPTOTOMOGRAPHY

This chapter is an adaptation of [29] (copyright 2010 by American Physical Society), which was a collaboration between Michael Bogan, Veit Elser, Anton Barty, Sebastien Boutet, Sasa Bajt, Janos Hajdu, Tomas Ekeberg, Filipe Maia, Joachim Schulz, Marvin Seibert, Bianca Iwan, Nicusor Timneanu, Stefano Marchesini, Ilme Schlichting, Robert Shoeman, Lukas Lomb, Matthis Frank, Meng Liang, Henry Chapman and I.

## 3.1 Introduction

Single-shot diffractive imaging is a developing technique aimed at determining the structure of very small biomolecules, such as proteins and viruses, not easily imaged through more established methods. In one scheme of single-shot diffractive imaging a serial stream of particles is injected into a pulse train of highly coherent and energetic X-ray free electron laser (FEL) radiation [4]. Photons from a single FEL pulse are diffracted when they encounter a particle. Although the FEL pulse destroys this particle, its structure is recorded in the diffraction data if the pulse is sufficiently short [32, 7].

As it is difficult to manipulate or determine the orientations of very small particles, they are currently injected into the FEL radiation at *random*, unmeasured orientations. Nevertheless, sufficiently many 2D diffraction patterns from an ensemble of identical albeit randomly oriented particles can in principle overdetermine the particle's band-limited 3D Fourier intensities. Earlier numerical simulations also show that the particle's 3D intensities can be recovered

even if the random 2D diffraction patterns are remarkably *noisy* (see [16, 21] and Chapter 2).

## 3.2 Experimental data

This chapter demonstrates the experimental feasibility of single-shot diffractive imaging with noisy 2D diffraction patterns from randomly-oriented identical particles, which we coin *cryptotomography*. In this first exercise in cryptotomography, we reconstruct the 3D intensities of a relatively large and simple iron oxide nano-particle. These nearly mono-disperse particles are approximately solids of revolution with principal radii 25 nm and 100 nm (SEM measurements shown in Figure 3.1).

Our primary objective is not to study these simple nano-particles in any greater detail than what is already available via SEM, but to show that cryptotomography, despite its random and noisy data, is experimentally viable through the union of experimental and theoretical innovations. We do so in the style of previous papers in diffraction imaging, which demonstrated novel techniques with simple test subjects [7, 8].

We performed our experiment at DESY's FLASH facility with each FEL pulse train comprising 100 pulses (7 nm radiation; $30\pm10\mu$m beam focus, extrapolated from [6]) separated by 10 $\mu$s, repeated at 5 pulse trains per second. Each detector exposure was 1 second long. Additional details on experimental parameters and data collection were similar to [4] and [3, 1] respectively, except we used a nebulizer instead to aerosolize the nano-particles, which were then directed into the FEL radiation using an aerodynamic lens stack [2].

Our experiment imposed two considerable challenges. First, the FEL pulse fluence fluctuated due to electron bunch dynamics in the lasing process [38]. Second, there was considerable diffuse background scattering (Figure 3.1) from a silicon aperture used to shield downstream instruments from beamline scatter. All $5\times100$ pulses in a single detector exposure contributed to this scatter whereas only 1 incident pulse creates a useful diffraction pattern if it illuminates a particle. This diffuse background changed gradually as the FEL pulses eroded the edges of the aperture, exacerbating this issue.

To overcome the second challenge, we assumed that the background scatter added incoherently to the diffraction pattern from a pulse illuminating a particle. From 2000 diffraction patterns (data) we identified those that contained coherent scattering (hits), by checking for intensity lobes expected of single prolate nano-particles (Figure 3.1) while excluding anomalous data with scattering from injected debris or those with diffraction patterns from more than one particle. Only non-hits without anomalies were considered background data.

Since the character of the background changed slowly over many data-acquisition cycles, for each hit we compared the results after separately subtracting five background data acquired nearest in time. From these five subtractions we selected the one that gave roughly equal numbers of positive and negative photon counts at higher spatial frequencies, where the particle's signal is presumably negligible [1].

Row defects, scattering from the aperture in the multi-layer planar mirror [1] and missing photon counts at the lowest spatial frequencies were always limited to certain pixels of the X-ray sensitive detector. These pixels were excised from

[1]Some potential hits still had strong pure background, due to the mismatched total fluence between potential hits and background data, and had to be discarded.

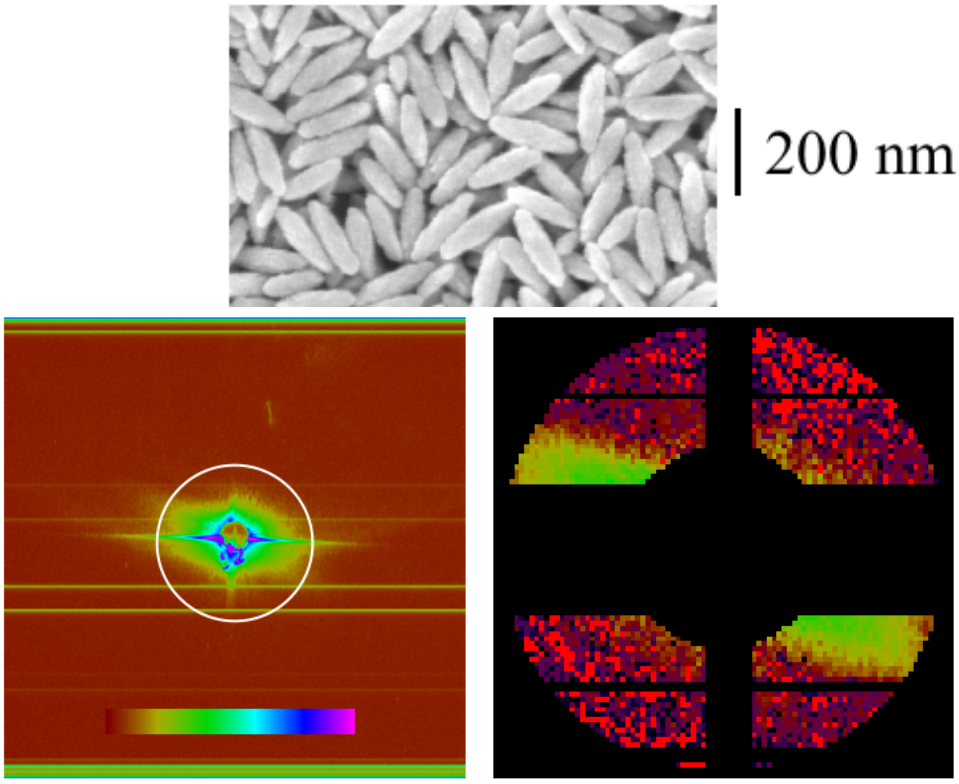



Figure 3.1: Top panel shows an SEM image of the prolate iron oxide nanoparticles (oblate 3D Fourier intensities) used in our experiment. Lower panels show a typical noisy and random diffraction pattern from such a nano-particle before ($512 \times 512$ array, left) and after processing ($91 \times 91$ array, right): truncated high spatial frequencies (discarded photons counts outside left white circle); binned photon counts; subtracted background; excised non-signal pixels (redacted). The logarithm of photon counts are colored according to the inset color bar (max. at pink); negative counts in red. The mean dynamic range of photon counts in these post-processed data spans 2 orders of magnitude, subject to fluctuations in fluence.

all data — their measurements did not constrain intensity reconstruction. 2880 ($M_{\mathrm{pix}}$) usable pixels remained, identical in all hits (non-redacted pixels in Figure 3.1, lower right).

## 3.3 Elements of reconstruction algorithm

Each of the resultant 54 ($M_{\mathrm{data}}$) background-subtracted hits, despite missing information about their orientation ($\Omega$) and fluence ($\phi$), are noisy Ewald sphere sections of the particle's true 3D intensities. Our goal: to recover the set of 3D intensities and fluences most statistically compatible with these hits.

Our algorithm for recovering the particle's 3D intensities ([16] and Chapter 2) is based on expectation maximization [11], where we iteratively apply a simple rule to increase the compatibility of any model intensities ($W$, even random ones) with all hits. Consider the simplest case where we are given only one hit. We use a statistical test to determine which Ewald sphere sections in $W$, here on known as *tomograms*, are most compatible with this hit, then replace those tomograms with said hit, weighted by probabilities taken from the statistical test. This prescription on $W$, which is iterated to a fixed point, determines the most likely orientations of this hit with respect to an increasingly compatible $W$. We can generalize this prescription to many hits by updating the tomograms in consensus with all hits. This chapter extends our algorithm to also determine the fluence distribution of the hits.

The model 3D intensity $W$ has various representations. It can be compactly written as $W(\mathbf{q})$: the time-integrated scattered intensity at spatial frequency $\mathbf{q}$ when the particle is in some reference orientation. We represented $W(\mathbf{q})$ in our reconstructions as a cubic array of floating-point numbers, indexed by equally spaced samples of $\mathbf{q}$. Detector pixels, labeled by index $i$, measure $M_{\mathrm{pix}}$ point samples $W(\mathbf{q}_i)$.

If we gave the particle some arbitrary orientation $\Omega$ from the reference po-

sition, the intensity recorded at detector pixel $i$ is $W(\mathbf{R}_\Omega \cdot \mathbf{q}_i)$, where $\mathbf{R}_\Omega$ is the orthogonal matrix of this rotation. We approximated the continuous $\Omega$ with a discrete sampling of $M_{\rm rot}$ points labeled by the index $j$ (sampling discussed in Appendix C). As a shorthand, we define the intensity at detector pixel $i$, after dividing away the incident pulse fluence, from diffraction off a particle at approximate orientation $j$ as tomogram samples $W_{ij} = W(\mathbf{R}_j \cdot \mathbf{q}_i)$.

The statistical test central to our expectation-maximization algorithm assess the likelihood that measurements on the $i$th detector pixel of the $k$th hit ($K_{ik}$), with photon fluence $\phi_k$, corresponds to tomogram samples of our intensity model $W_{ij}$. Intensity fluctuations at each detector pixel of a particular hit due to background-subtraction are assumed to be mutually independent. Assuming the noise from background-subtraction dominates over Poisson statistics, the likelihood that the $k$th hit comes from the $j$th tomogram of $W$ is

$$R_{jk}(W, \Omega, \phi) \propto \exp\left(-\frac{\sum_i^{M_{\rm pix}} (K_{ik}/\phi_k - W_{ij})^2}{2\,\sigma^2}\right). \tag{3.1}$$

The global noise parameter $\sigma$ in equation (3.1) is the only unconstrained parameter in our algorithm. We seek an ideal $\sigma$ that quantifies the true noise in the diffraction data.

Determining the most likely model parameters $(W, \Omega, \phi)$ given all hits is unattainable in a single step, hence we adopt an iterative procedure that we call Expansion-Maximization-Compression (EMC) to implement the model updates (Section 2.2.3).

In the EMC prescription we first expand (E-step) $W(\mathbf{q})$ into the set of $M_{\rm rot}$ tomograms $W_{ij}$ (see Section 2.6.1). Although this expansion allows efficient comparisons between hits and model tomograms, it creates redundancy since each intensity sample of the 3D model $W(\mathbf{q})$ is represented in multiple tomograms

$W_{ij}$. Such redundancy ameliorates the effects of pixel excisions when populating $W(\mathbf{q})$ with data measurements.

After expanding our model $W$ into tomograms, we maximize (M-step) the data likelihood of each model tomogram $W_{ij} \rightarrow W'_{ij}$ independently (see Section 2.6.1). Specifically, we determine the new model $(W'_{ij}, \Omega'_k, \phi'_k)$ which maximizes Eq. (3.2), conditional on probabilities of the current model $(W_{ij}, \Omega_k, \phi_k)$:

$$\underset{W',\Omega',\phi'}{\operatorname{argmax}} \sum_{k}^{M_{\text{data}}} \sum_{j}^{M_{\text{rot}}} R_{jk}(W, \Omega, \phi)\, w_j \ \log\left(R_{jk}(W', \Omega', \phi')\right) \ . \tag{3.2}$$

The numbers $w_j$ are weights applied to our rotation group samples that approximate a uniform prior distribution (see Appendix C). Despite this uniform prior distribution over the rotation group, we can still detect orientational biases by evaluating equation (3.1) for all hits over the converged $W$ reconstructions.

The requirement that $W'_{ij}$ from different orientations are consistent with a single intensity model $W'(\mathbf{q})$ is enforced in the final compression (C-step) by averaging interpolated intensity samples in all tomograms (details in Section 2.6.1), which represent a particular intensity sample in $W'(\mathbf{q})$. We also impose Friedel symmetry on $W'(\mathbf{q})$ since we are operating in the limit of weak elastic scattering. This compressed and symmetrized model $W'(\mathbf{q})$ is now ready for another round of EMC.

We exploited a side-effect of EMC's redundant intensity representation to find the ideal noise parameter $\sigma_{\text{min}}$. If $\sigma$ is too small, even though the data likelihood of each updated tomogram is provisionally increased, such tomograms are mutually incompatible, thus diminishing the data likelihood of the compressed updated intensities. We determined $\sigma_{\text{min}}$ knowingly: if $\sigma < \sigma_{\text{min}}$, the log-likelihood of reconstructions, equation (3.2), decline and reconstructions

vary dramatically.

EMC updates of the fluence and intensity model have a regrettable degeneracy: if scaling $W'(\mathbf{q})$ by a multiplicative constant increases the model's log-likelihood in equation (3.2) so will a commensurate scaling in $\phi'$. As a consequence, simultaneous EMC updates for $\phi_k$ and $W_{ij}$ cannot be decoupled easily. However, if we updated only $\phi_k$ or $W_{ij}$, while keeping the other fixed, the net log-likelihood still increases:

$$W'_{ij} = \frac{\sum_k R_{jk} w_j \; K_{ik}/\phi_k}{\sum_k R_{jk} w_j}, \tag{3.3}$$

$$\phi'_k = \frac{\sum_j R_{jk} w_j \; \sum_i K_{ik}^2}{\sum_j R_{jk} w_j \; \sum_i K_{ik} W_{ij}}. \tag{3.4}$$

The likelihood $R_{jk}$ in the last two equations is evaluated at the current model parameters $(W, \Omega, \phi)$. We imposed $\sum_j R_{jk} w_j = 1$ during each EMC iteration to assert that every hit (index $k$) must be found at some orientation (index $j$).

## 3.4 Reconstruction

We reconstructed 3D Fourier intensities from random starts using EMC with only diffraction data while imposing Friedel symmetry since these are the minimal constraints expected in future cryptotomography experiments. We later evaluated each converged reconstruction (Figures 3.2) by fitting them to intensity distributions of ideal ellipsoidal particles $I_{\text{ellip}}$.

We began each reconstruction with random intensities $W(\mathbf{q})$, represented by random numbers on a cubic array. Intensities were reconstructed only in the

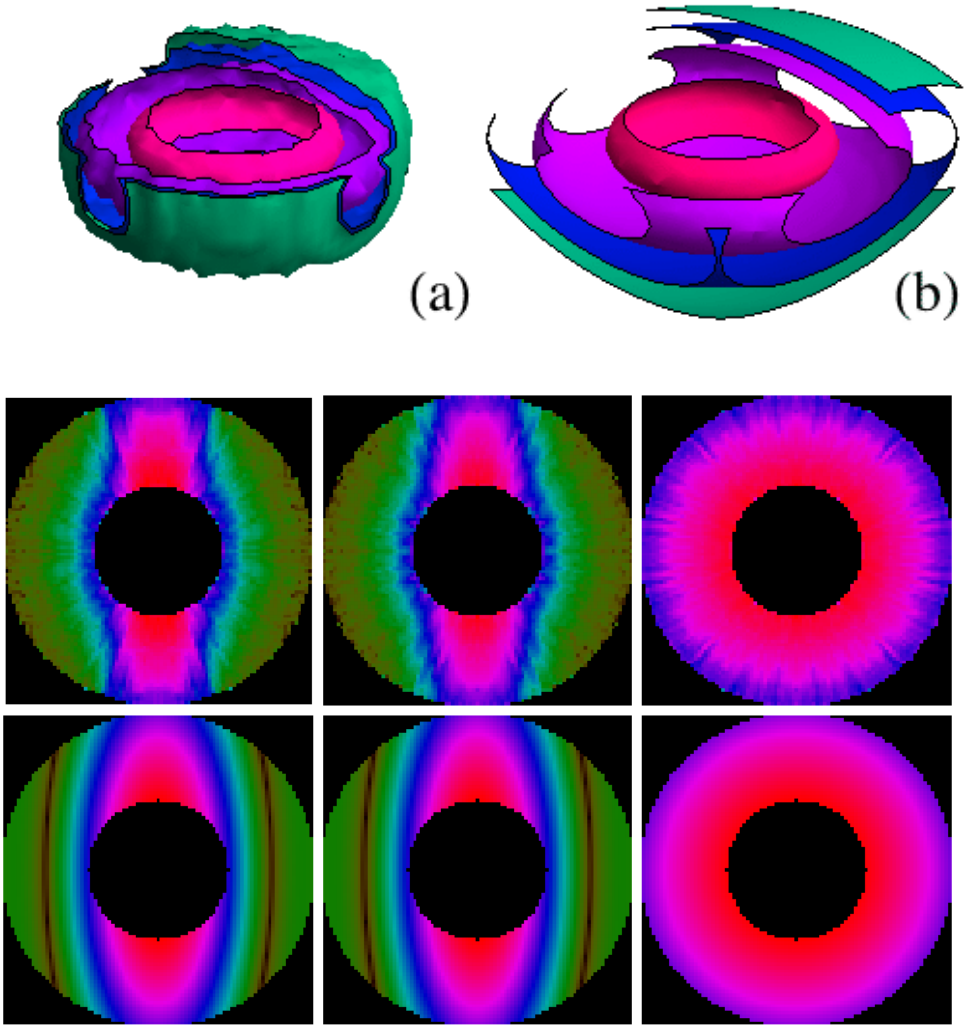



Figure 3.2: Comparing reconstructed intensities to those of an ideal ellipsoidal particle $I_{\text{ellip}}$. (a) Cutaway view of choice 3D isointensity surfaces of a reconstruction which show an oblate intensity distribution and (b) those of $I_{\text{ellip}}$ with the best R-factor fit to this reconstruction; (middle row) mutually perpendicular cross-sections of this reconstruction; (bottom row) same cross-sections of $I_{\text{ellip}}$ in (b). Logarithm of intensities are shown as hues (color bar in Figure 3.1). Intensities in reconstructions span 3 orders of magnitude.

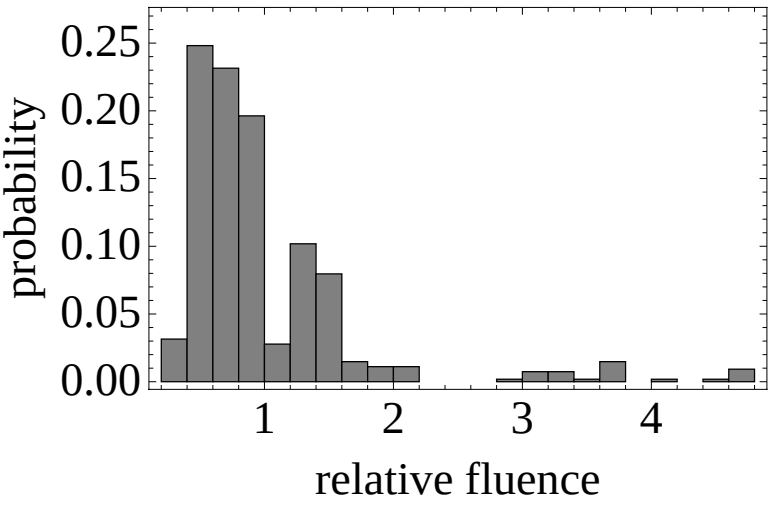



Figure 3.3: Probability distribution of reconstructed relative fluence of the hits (expected distributions studied in [38]), a result of fluctuations in pulse fluence and the positions of particles when illuminated.

range $q_{\min} < |\mathbf{q}| \leq q_{\max}$ (non-redacted regions in 2D sections of Figure 3.2) with $q_{\min}/q_{\max} = 20/44$ — determined from low spatial frequencies missing-signal region and maximum scattering angle of $5.23°$ in processed hits (white circle in Figure 3.1, lower left).

We normalized each random initial $W(\mathbf{q})$ to have tomograms that matched the mean power received per hit; initial fluences were set to $\sum_i K_{ik}$; we used a $M_{\text{rot}} = 3240$ tomographic representation of our model ($M_{\text{rot}}$ sufficiency discussed in Section 2.5.1). The noise parameter $\sigma$ was measured in units of $\Delta_K$: the square root of the sum of variances in measured intensities of each pixel [2]. We determined $\sigma_{\min} = 0.07\,\Delta_K$, where reconstructions below this showed diminished likelihood and significant diversity [3].

[2]Undesirable rotational averaging occurs at $\sigma = \Delta_K$. Such reconstructions resemble 3D powder diffraction pattern.

[3]Reconstructions using $\sigma_{\min}$ routinely converged to one of two quantitatively distinguishable varieties: log-likelihood of one was about $1\%$ higher than the other. We rejected the lower-likelihood variety as candidate solutions. Such multiplicity is expected since reconstructions

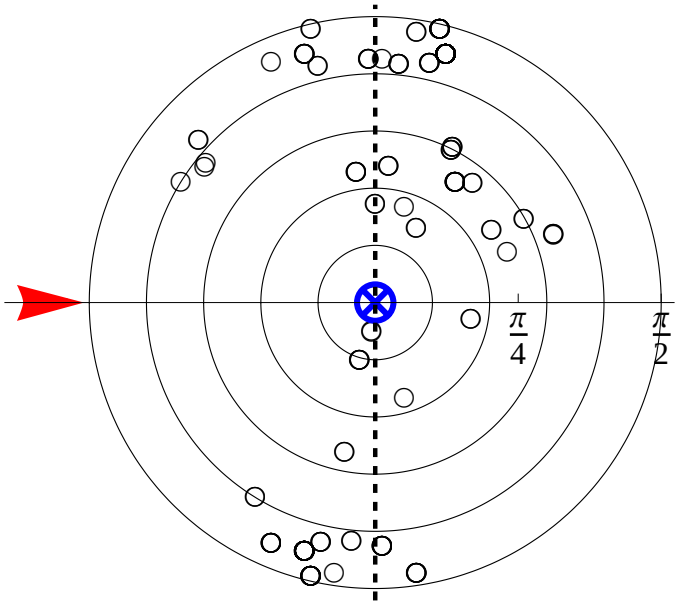



Figure 3.4: We superimpose the most likely orientations of the nano-particle symmetry axis in the 54 hits found in each of 10 reconstructions in this azimuthal projection (marked by equidistant rings of constant latitude in this top hemisphere). Orientations are inversion-symmetric to those in the complementary bottom hemisphere. Many orientations coincide. The vertical dashed line shows the detector plane; FEL pulses traveled with the red arrow; nano-particles were injected into the page (blue crosshair).

We could in principle recover the nano-particle's real-space contrast from the reconstructed 3D intensities $W$ via phase retrieval [8], but there will be pixellation effects (38.3 nm half-resolution in processed hits). Instead we determined the principal radii of the nano-particle from our reconstructed intensities using R-factor comparisons [4] with those of an ideal ellipsoidal particle $I_{\text{ellip}}(\mathbf{q})$. The

near $\sigma_{\text{min}}$ were only marginally constrained by so few hits (Figure 3.4).

[4] R-factor between $I_{\text{ellip}}(\mathbf{q})$ and reconstructed $W$ were computed as

$$\left(\sum_{\mathbf{q}} \left| I_{\text{ellip}}(\mathbf{q})^{1/2} - W(\mathbf{q})^{1/2} \right| \right) / \left(\sum_{\mathbf{q}} W(\mathbf{q})^{1/2}\right) , \tag{3.5}$$

particle's principal radii were found to be $30.5 \pm 0.8 : 30.6 \pm 0.7 : 76.1 \pm 1.1$ nm with R-factor $0.093 \pm 0.006$ ($0.162 \pm 0.003$ when reconstructions were compared to the intensity function expected of the SEM particle measurements) [5]. We expect a larger error in the longer direction of the prolate nano-particle since it corresponds to the compact direction of its oblate intensities, which was more susceptible to background noise. Figure 3.3 shows the concomitant fluence distribution we reconstructed.

Figure 3.4 shows the most likely orientations of the particles corresponding to the 54 hits used in intensity reconstruction [6]. Reconstructed orientational bias in the data could arise from either systematic effects in particle delivery or biases during hit-selection. If a nano-particle's axis of symmetry is colinear with the incident direction of FEL pulses its diffraction pattern will not have identifiable lobes (unlike Figure 3.1). Alternatively, data may contain intensity lobes but are obscured by the redacted pixels or background noise. Such data might have been missed during hit-selection.

---

where

$$I_{\text{ellip}}(\mathbf{q}) \propto \left| \left(\sin(\pi|\tilde{\mathbf{q}}|) - \pi|\tilde{\mathbf{q}}|\cos(\pi|\tilde{\mathbf{q}}|)\right) \,/\, |\tilde{\mathbf{q}}|^3 \right|^2 \;; \tag{3.6}$$

and

$$|\tilde{\mathbf{q}}| = \sqrt{q_x^2 x_0^2 + q_y^2 y_0^2 + q_z^2 z_0^2}\,/q_{\max} \,. \tag{3.7}$$

In these equations, $q_{\min} < |\mathbf{q}| \leq q_{\max}$ and $x_0, y_0, z_0$ are the principal radii of the nano-particle. $I_{\text{ellip}}(\mathbf{q})$ and $W$ were normalized to the same total power for R-factor comparisons.

[5] We reconstructed ten intensity distributions (e.g. Figure 3.2) from random initial intensities. For each reconstruction we found the principal radii that gave the best R-factor fit to an ellipsoidal intensity $I_{\text{ellip}}(\mathbf{q})$; the quoted radii are their averages.

[6] EMC found the particle's axial symmetry without imposition. For each hit, EMC automatically assigned nearly equal probability to orientations related by this symmetry, hence mitigating data-scarcity effects expected from limited hits, pixel excision and orientation biases.

## 3.5 Conclusions

Despite the simplicity of our nano-particle, we emphasize that reconstructing its 3D intensities, using only diffraction data, is non-trivial primarily because of ambiguities from unmeasured data orientation and fluence. These ambiguities make direct interpretation of the data very challenging. EMC circumvents these ambiguities without prior assumptions about the intensity distribution beyond Friedel symmetry, enforcing only simple statistics (equation (3.1)) to determine that the particle was prolate instead of oblate. Subsequent R-factor fits of converged EMC reconstructions also gave reasonable particle dimensions, corroborating the effectiveness of EMC on experimental cryptotomography data.

## 3.6 Acknowledgements

Our work was supported by: Helmholtz Association; the PULSE Institute at the SLAC National Accelerator Laboratory by the U.S. Department of Energy, Office of Basic Energy Sciences (MJB); Lawrence Livermore National Laboratory (LLNL) under contracts W-7405-Eng-48 and DE-AC52-07NA27344 ; Laboratory of Directed Research and Development Program of LLNL pertaining to project 05-SI-003; CHESS through NSF and NIH / NIGMS via NSF award DMR-0225180; Deutsches Elektronen-Synchrotron, DFG Cluster of Excellence at Munich Centre for Advanced Photonics, Virtual Institute Program of Helmholtz Society, Joachim Herz Stiftung, Max Planck Society and Swedish Research Council. I would also like to thank the reviewers of [29] for their insightful suggestions, which improved the paper this chapter was based on, and Yoav Kallus for his invaluable discussions.

CHAPTER 4

# ULTRAFAST IMAGING OF MAGNETIZATION DISTRIBUTIONS

This chapter was produced in collaboration with Stefan Eisebitt and Samuel Flewett, in preparation for publication at Physical Review E.

## 4.1 Introduction

There has been a growing interest in studying and manipulating thin-film magnetic nanostructures [17, 37, 13, 14, 35, 34, 22]. Besides the immediate commercial applicability of such studies, experimental data on the formation, dynamics and stability of magnetic nanostructures will provide clues for constructing predictive models of magnetic materials [34], which may in turn drive the invention of novel devices.

A comprehensive understanding of these magnetic nanostructures involves studying extremely fast magnetic dynamics at high resolution. Ideally, this can be achieved by sequentially illuminating an evolving magnetic specimen using very short, intense pulses of coherent X-ray radiation (cartoon of such in Figure 4.1). Such radiation has become available at X-ray free electron laser (XFEL) facilities, which can produce femtosecond pulses with upwards of $10^{12}$ X-ray photons each. However, the intensity of the XFEL pulses must be reduced to prevent sample heating by energetic X-ray photons. As a result of reducing pulse intensity, the diffraction signal from the weakly scattering magnetic contrast in specimens is expected to be *photon-shot-noise* limited [22]. To make matters worse, the magnetic signal will also be contaminated by strong scattering from the non-uniform charge density intrinsic to magnetic specimens.

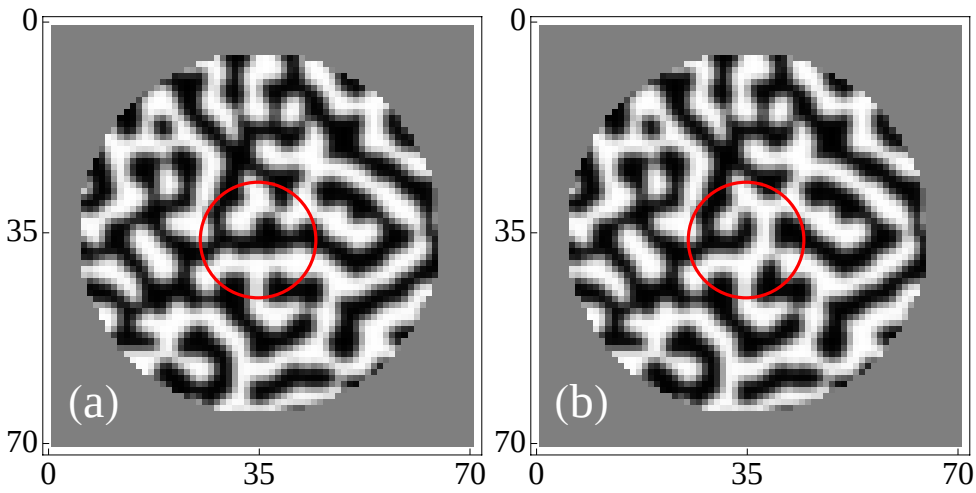



Figure 4.1: Simulated example of the type of magnetic domain motion that could be investigated by an XFEL source. The magnetization distributions above differ only within the central circle. The average domain is 5 pixels wide, or approximately 170 nm when related to the experiment data in [14] (each pixel hence measures 34 nm). We explain how to generate an ensemble of such similar distributions in the text.

Currently, X-ray Fourier transform holography [14] and X-ray speckle metrology [35] are two leading techniques already used to study magnetic nanostructures. Despite their effectiveness, they still have limitations: speckle metrology is a statistical technique restricted to the reciprocal space (unable to resolve local dynamics in Figure 4.1); Fourier transform holography affords direct-space imaging but it requires the crafting of a reference illumination.

This chapter describes a reconstruction procedure that directly images extended magnetization distributions when given only transmission diffraction data, without the need for a reference illumination. To make our procedure applicable to magnetic imaging with XFEL radiation, we use diffraction data with the severest levels of noise, in addition to missing data in the beamstop. We

compare our diffractive imaging reconstruction with the performance of Fourier transform holography at comparable noise levels. In our simulations of such experiments, we also had to generate an ensemble of credible magnetization distributions to be used as scattering sources (we explain this in the earlier parts of this chapter). Finally, we include a feasibility study of noisy magnetic imaging when subjected to our described methods, which may be useful in the design of future experiments.

## 4.2 Resonance scattering

Multilayer magnetic thin films with perpendicular magnetic anisotropy [13, 35, 14, 34, 22, 23, 27] exhibit a notable phase comprising magnetic nanostructures that can be described by a 2D coarse-grained magnetization distribution $\mathbf{m}(\mathbf{r})$. The magnetization in this phase is effectively parallel or antiparallel to the sample's layer normal. In this section, we briefly discuss how such magnetization distributions are encoded in the diffraction data.

Diffractive imaging measures the sample's elastic photon scattering cross section. The scattering mechanism includes virtual transitions between core electron states and unoccupied electron states above the Fermi level [23]. Since these unoccupied states are spin-polarized by the sample's local magnetization, photons scattered from the sample are sensitive to the sample's magnetization distribution $\mathbf{m}(\mathbf{r})$.

There are, naturally, other components of the sample's elastic scattering amplitude that are insensitive to the magnetization: $f^0(\mathbf{r})$, the Thomson contribution; $f^c(\mathbf{r})$, the anomalous charge scattering. Both of these contributions are

integrated along the incident beam direction. Like $\mathbf{m}(\mathbf{r})$, $f^0(\mathbf{r})$ and $f^c(\mathbf{r})$ are also treated as 2D distributions.

A magnetic specimen's total elastic scattering amplitude is given by [23] as

$$\begin{aligned} f_{\text{tot}}(\mathbf{r}) \;\approx\; & f^0(\mathbf{r}) + (\mathbf{e}_f{}^* \cdot \mathbf{e}_i)\, f^c(\mathbf{r}) - \mathrm{i}(\mathbf{e}_f{}^* \times \mathbf{e}_i) \cdot \mathbf{m}(\mathbf{r})\, f^M \\ & + (\mathbf{e}_f{}^* \cdot \mathbf{m}(\mathbf{r}))\, (\mathbf{e}_i{}^* \cdot \mathbf{m}(\mathbf{r}))\; f^m \;, \end{aligned} \tag{4.1}$$

where $\mathbf{e}_i$ and $\mathbf{e}_f$ are the polarization vectors of the incident and scattered radiation. The magnetization-sensitive scattering amplitudes $f^M$ and $f^m$ are scaled to allow the magnetization to be normalized as $\max(|\mathbf{m}(\mathbf{r})|) = 1$.

The total elastic scattering amplitude $f_{\text{tot}}(\mathbf{r})$ of multilayer magnetic films can be simplified with a few experimental constraints. First, since the magnetization is parallel or antiparallel to the sample's layer normal, we can replace $\mathbf{m}(\mathbf{r})$ with the longitudinal scalar $m(\mathbf{r})$. More importantly, light transmission along this longitudinal direction suppresses the contribution of the last term in equation (4.1). Second, we restrict ourselves to circularly polarized incident radiation, which is a scattering eigenstate of the 3rd term in equation (4.1). This choice causes the diffraction patterns from magnetic and charge distributions to interfere [1]. Third, in the small-angle scattering limit, we can combine the non-magnetic scattering contributions as $f^C(\mathbf{r})$. These conditions produce a simpli-

[1] There will be no interference between charge and magnetic scattering terms if the incident radiation were linearly polarized. In this case, diffraction intensities from charge and magnetic distributions are separately added, as demonstrated in reference [13], and the former, ideally, can be subtracted away. Determining the static, random charge scattering for subtraction is possible when the photon energy is detuned away from the core-level resonance, hence suppressing magnetic scattering. But this subtraction may be unreliable at noisy, high-$q$ signal regions where the magnetization distribution is primarily encoded. Subtraction might also be problematic in single-shot imaging, when the incident photon fluence may fluctuate between shots — guesswork is needed to match the intensities of the charge-plus-magnetic data to those of charge-only data for reliable subtraction.

fied total scattering amplitude:

$$f_{\text{tot}}(\mathbf{r}) \approx f^C(\mathbf{r}) + f^M m(\mathbf{r}) \quad . \tag{4.2}$$

Experimently, the magnetic scattering amplitude $f^M$ can be increased through resonance scattering: by tuning the energy of the incident photons to match those of core-level electron transitions in the sample (L or M edges) [13, 35, 14, 34, 22, 23, 27]. This enhances the scattering signature of the magnetization with respect to the charge distribution, which is useful since we are interested only in the former.

A difference in the correlation length of the charge distribution and that of the magnetization distribution is common, which causes a separation in the peaks of their respective diffracted power [27]. Potentially, one could then ignore the charge distribution when imaging the magnetization at a lower resolution. However, later sections of this chapter show that magnetic imaging may still be difficult despite such a separation.

The Fraunhofer diffraction intensity from samples obeying equation (4.2) is

$$I(\mathbf{q}) \propto \phi \left| \int \mathrm{d}^2\mathbf{r}\ \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\, f_{\text{tot}}(\mathbf{r}) \right|^2 \quad , \tag{4.3}$$

where $\phi$, the photon fluence, crucially determines the number of diffracted photons and hence the severity of photon-shot-noise. Since we are only interested in the total number of diffracted photons in our simulations, the absolute scale of the magnetic and charge scattering amplitudes, $f^M$ and $f^C$ respectively, is immaterial since we can vary the number of diffracted photons by varying $\phi$. From here on, magnetic scattering amplitudes and magnetization become interchangeable because they differ only by this unimportant absolute scale. The same is true between charge scattering amplitudes and charge. To be consistent,

we normalize the magnetic scattering amplitude $f^M = 1$ in equation (4.2). The ratio $f^C/f^M$, however, depends on the experimental specimen. This, of course, means that $\phi$ is no longer strictly the photon fluence, but a variable to control the number of scattered photons.

## 4.3 Generating magnetic domain patterns

To simulate realistic magnetic imaging, we first need to generate magnetization distributions, or *domain patterns*, that resemble a wide and interesting variety of actual specimens, even if they were produced in an unphysical fashion. At the minimum, the ensemble of such domain patterns should conform to these experimental observations:

1. in Fourier-space, an azimuthally symmetric diffracted power which peaks at a particular spatial frequency (compare simulated example in Figure 4.3 to those from experiments in [27, 13, 14, 34, 22]);
2. in direct-space, a statistical distribution on the magnetization (Figure 4.5) of ferromagnetic domains with finite-width domain walls (Figure 4.2).

The clues to generating realistic domain patterns lie in the careful examination of the diffraction envelope shown in Figure 4.3. The spatial frequency dependence of this envelope reveals two competing effects: short-range exchange interaction that produces ferromagnetic domains and long-range demagnetizing fields which in turn destabilize these domains.

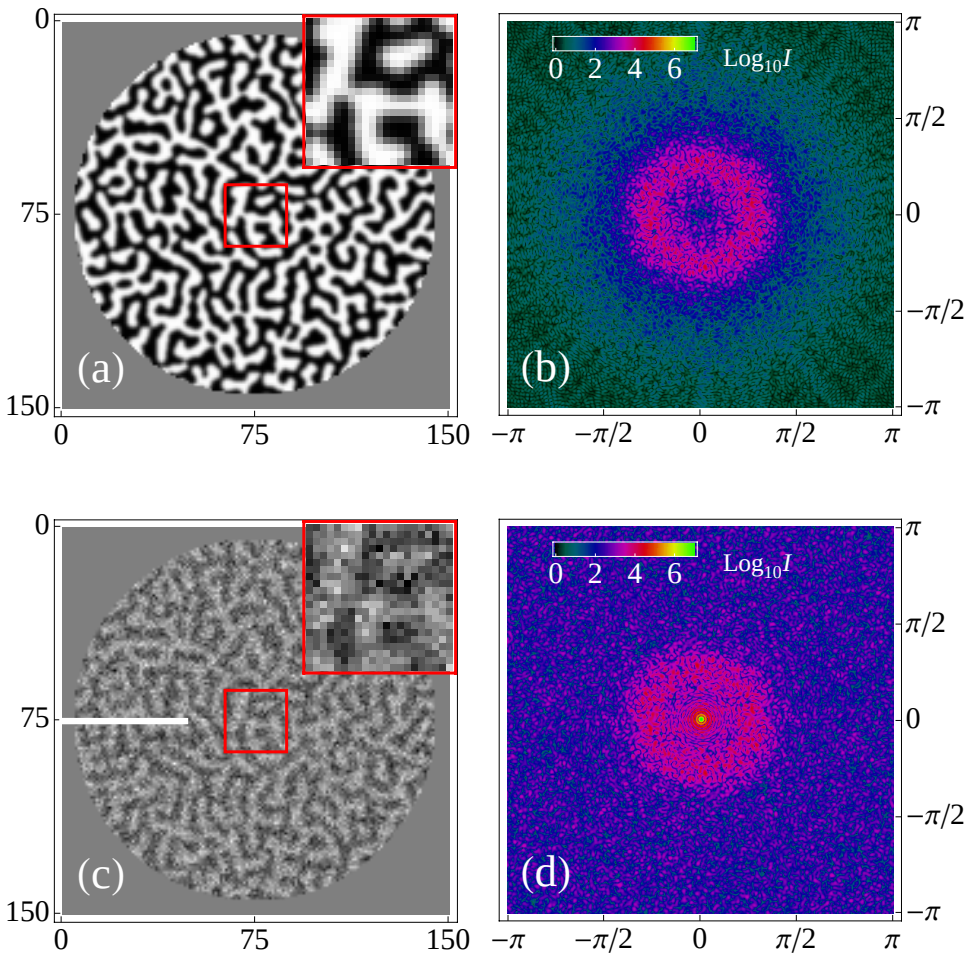



Figure 4.2: Simulated magnetic domains and their associated diffraction intensities, which together illustrate the effects of charge scattering. Domains without charge scattering in (a) and the logarithm of their diffraction intensities in (b). To the same domains we add random charge distribution ($\Delta_m/\Delta_c = 1$) to produce (c) and the logarithm of its resultant diffraction intensities in (d).

These effects are modeled by the 2D Landau-Ginzburg free energy density

$$\mathcal{F}(\mathbf{r}) = A\left(m(\mathbf{r})^2 - 1\right)^2 + B\left|\nabla m(\mathbf{r})\right|^2 + C\int_{|\mathbf{r}'-\mathbf{r}|>l} \frac{m(\mathbf{r})m(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\, d^2\mathbf{r}' \,, \qquad (4.4)$$

where $A$, $B$ and $C$ are temperature dependent positive quantities, and $l = \pi/q_{\max}$ is a cutoff that defines the maximum spatial frequency. Rewriting (4.4)

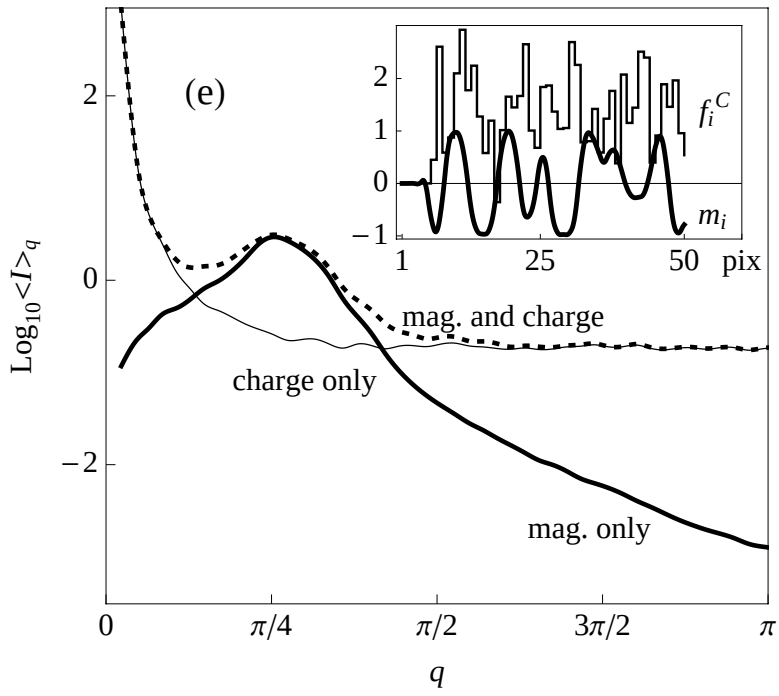



Figure 4.3: Power envelope of magnetic and charge scattering. In panel (e), we plot the azimuthally averaged diffraction intensities from the charge only (thin, solid line), magnetic only (thick, solid line) and charge-plus-magnetization distributions (dashed line) belonging to the domain pattern in Figure 4.2c. The direct-space scattering amplitudes ($f_i^C$ for charge; $m_i$ for magnetic) along the horizontal white line of the simulated domain pattern in Figure 4.2c are inset in panel (e).

in terms of the Fourier modes of the magnetization, $m(\mathbf{q})$, we obtain in the limit $|\mathbf{q}| = q \ll q_{\max}$ the following expression:

$$\mathcal{F}(\mathbf{q}) = |m(\mathbf{q})|^2 \left[-2A + Bq^2 + 2\pi C \left(q_{\max}/\pi - q + (\pi/4)\, q^2/q_{\max}\right)\right] + O\left(|m(\mathbf{q})|^4\right). \tag{4.5}$$

Defining new constants $a > 0$ and $b$, and rescaling $q_{\max}$ by a constant, this can be rewritten in the much simplified form

$$\mathcal{F}(\mathbf{q}) = \left[(q/q_{\max} - a)^2 + b\right] |m(\mathbf{q})|^2 + O\left(|m(\mathbf{q})|^4\right). \tag{4.6}$$

The ferromagnetic instability corresponds to $b \to 0$ and the $q$-dependence of the fluctuations as this limit is approached is controlled by the coefficient of the term quadratic in the magnetization. As a simple model for the formation of magnetic domains in real materials we will assume the distribution of fluctuations in the paramagnetic phase ($b > 0$), given by the equipartition theorem, is preserved when the system freezes into a particular domain pattern. The intensity in this model is given by

$$I(q) \propto \frac{1}{(q/q_{\max} - a)^2 + b} \; . \tag{4.7}$$

Our simulations will use this form for the power spectrum with $a$ and $b$ fitted to agree with experimental data [14]. We use dimensionless units where the maximum spatial frequency $q_{\max}$ is scaled to the value $\pi$.

The generation of each domain pattern begins with an array of random, uniformly distributed real numbers between -1 and +1, mimicking the high-temperature magnetization distribution in the absence of external fields. On this random state $m(\mathbf{r})$, we apply two nonphysical operations in turn:

1. band-pass Fourier filter using equation (4.7) —

$$m(\mathbf{q}) \to \frac{m(\mathbf{q})}{\sqrt{(q/q_{\max} - 0.27)^2 - 0.015}} \; , \tag{4.8}$$

   where $m(\mathbf{q})$ is the discrete Fourier transform of $m(\mathbf{r})$;

2. binary projection on $m(\mathbf{r})$ —

$$m(\mathbf{r}) \to \begin{cases} +1, & \text{if } m(\mathbf{r}) \geq 0 \\ -1, & \text{if } m(\mathbf{r}) < 0 \end{cases} \; . \tag{4.9}$$

The composition of these two operations is iterated on $m(\mathbf{r})$ until it *converges* (array values unchanged upon iteration). Thereafter, we multiplied the converged distribution $m(\mathbf{r})$ with a final low-pass Fourier filter, $\exp\left(-2.5\,(q/q_{\max})^2\right)$, to simulate the finite width of the magnetic domain walls [2]. This domain pattern is then normalized to $\max(|m(\mathbf{r})|) = 1$.

Different, random initial arrays result in different domain patterns $m(\mathbf{r})$, defining an ensemble of simulated patterns. Whereas we generated domains with zero net magnetization, this recipe can be easily modified to change this net magnetization.

This recipe for generating domain patterns is easily extended to create perturbed versions of any domain pattern: we replace randomly selected circular areas in a previously converged source domain pattern with random numbers, then reapply the domain generation recipe until this perturbed pattern converges. This perturbation occurs before the low-pass Fourier filter is applied to the source pattern. As an example, the pattern in Figure 4.1b is a converged perturbation of the pattern in Figure 4.1a.

## 4.4 Model of charge scattering

Since it is reasonable to expect the charge distribution $f^C(\mathbf{r})$ to be spatially uncorrelated at the resolution of the resonant scattering experiments [27, 22], we model it as a 2D array of random, real numbers $f^C(\mathbf{r})$. Each array element of

[2] We assumed that the coarse-graining length is considerably smaller than the width represented by one array pixel. To minimize finite-size effects from sampling the distribution on a numerical array, we generated domains with twice the resolution ($|q_x| \leq 2q_{\max}$ and $|q_y| \leq 2q_{\max}$ without changing $q_{\max}$ in equation (4.8)) then truncated the Fourier-space of the converged domain pattern back to the lower resolution $q_{\max}$.

$f^C(\mathbf{r})$ represents the charge scattering amplitude averaged over a resolution element, or a *pixel*.

The statistics of the spatially uncorrelated charge distributions is characterized by its mean $\langle f^C(\mathbf{r})\rangle$ and standard deviation,

$$\Delta_c = \sqrt{\langle\, (f^C(\mathbf{r}) - \langle f^C(\mathbf{r})\rangle)^2 \,\rangle}\,, \tag{4.10}$$

which we coin *charge contrast*. The angle brackets denote the average over each distribution. The charge contrast should be compared to the *magnetic contrast*,

$$\Delta_m = \sqrt{\langle\, (m(\mathbf{r}) - \langle m(\mathbf{r})\rangle)^2 \,\rangle}\,. \tag{4.11}$$

The diffraction intensity in equation (4.3) does not distinguish between charge and magnetic scattering, so any reconstruction can only determine their sum (see equation (4.2)). Since we are interested only in recovering the magnetization distribution, the intrinsic charge distribution will contribute an inextricable scattering noise, characterized only by the signal-to-noise ratio $\Delta_m/\Delta_c$. When $\Delta_m \approx \Delta_c$, it becomes visually impossible to differentiate between these distributions even if their sum were correctly reconstructed (compare Figure 4.2a and 4.2c).

In contrast, the mean charge scattering amplitude $\langle f^C(\mathbf{r})\rangle$, as later sections will show, is an immaterial constant to the reconstruction of $m(\mathbf{r})$. Nevertheless, to be consistent, we fixed $\langle f^C(\mathbf{r})\rangle = 1.33$ using experimental data from [14].

## 4.5 Diffractive imaging as constraint satisfaction

We can interpret the diffractive imaging experiments in the language of constraint-satisfaction problems. Essentially, the goal is to recover the true magnetization distribution subject to two constraints: its measured noisy diffraction data (Fourier constraint) and the assumed statistics on its expected magnetization (direct-space constraint). This section discusses how we generated and characterized these two constraints.

### 4.5.1 Fourier constraint

The Fourier constraint requires that the diffraction intensities of the true magnetization distribution, which we wish to recover, be statistically compatible with the measured photon data, mindful that the data includes intrinsic charge scattering.

To simulate the diffraction data, we first added each pair of randomly-generated magnetization and charge distributions, $m(\mathbf{r})$ and $f^C(\mathbf{r})$ respectively. We confined this total scattering amplitude to a circular support $S$ (Figure 4.2a, for example). Its continuous intensity distribution was scaled by $\phi$ to give the desired average number of scattered photons, then Poisson sampled to simulate photon-shot-noise. We averaged each data with its Friedel-symmetry counterpart to make it consistent with the real-valued direct-space contrast. Finally, we applied a beamstop to this symmetrized data, thus removing photon counts that in experiments would be contaminated by intense, unscattered radiation. The size of the beamstop was adjusted such that the remaining photon counts

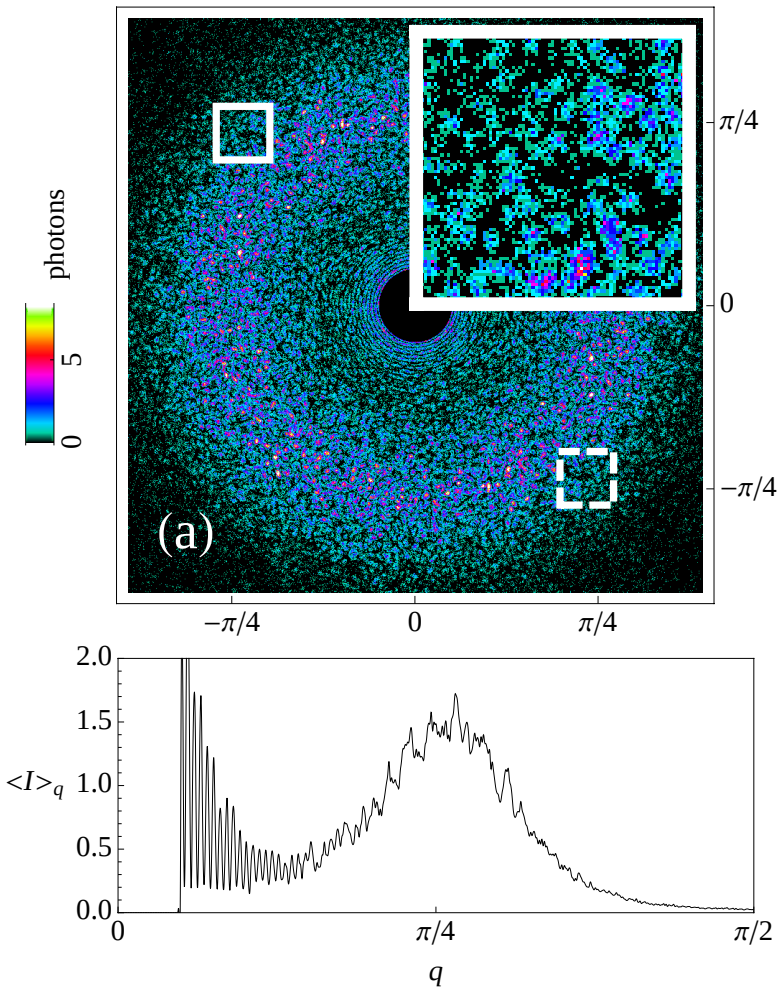



Figure 4.4: Noisy photon data from simulated diffraction experiments. Panel (a) shows the diffraction data from the domain pattern in Figure 4.10a, with signal-to-noise of point B in Figure 4.9. The intensities in the dashed-line box (lower right) of panel (a) are inversion symmetric to those in the solid-line box (upper left); larger inset is a magnified view of photon data in the solid-line box; the central black disk is the beamstop. Panel (b) shows the azimuthally averaged photon counts in panel (a).

span two orders of magnitude (example photon data in Figure 4.4). The Fourier amplitudes at spatial frequencies within the beamstop are unconstrained in our reconstruction algorithm.

### 4.5.2 Direct-space constraint

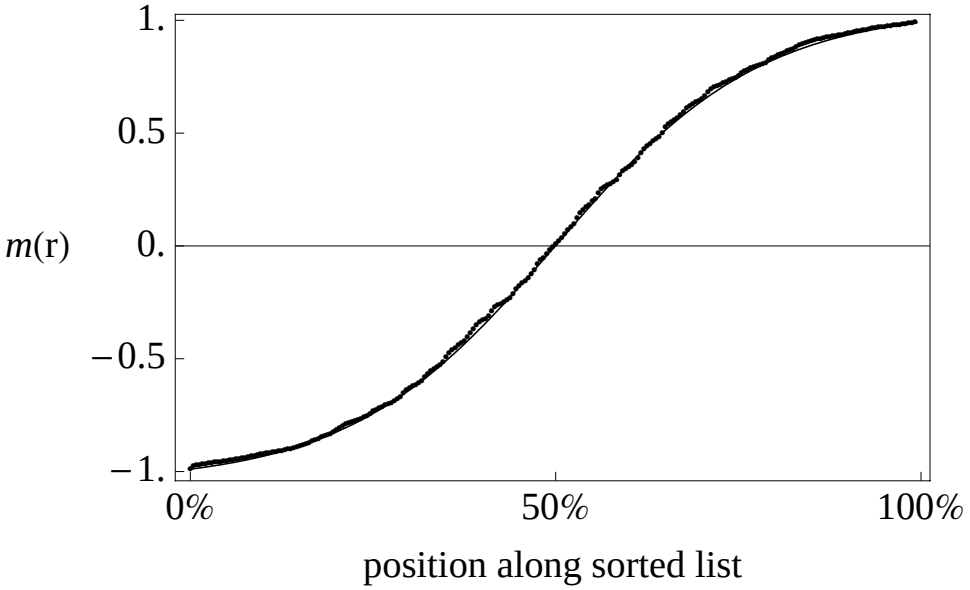



Figure 4.5: Model magnetizations of direct-space constraint. Sorted list of normalized magnetization $\widetilde{m}$ when averaged over many simulated domain patterns (curve) and those from one pattern (dots).

The statistics on the magnetization should be invariant within our ensemble of randomly generated domains patterns. This assumption improves with increasing support size, even on patterns with dynamically varying magnetization. The sorted magnetizations within a domain pattern, when averaged over many patterns, is shown in Figure 4.5. We denote this as the list of *model magnetizations* $\widetilde{m}$. Although this list of expected magnetizations is a result of our domain-generation recipe, it should still be qualitatively similar to magnetic domain patterns with finite domain walls. We will describe, in the later section on projecting to the direct-space constraint, a strategy for determining the histogram constraint when it is not initially unknown. These expected magnetizations constitute a strong *histogram constraint* on the ensemble of true magnetic

distributions we wish to reconstruct.

The histogram constraint and the knowledge of the size and shape of the support $S$ constitute the direct-space constraint in our reconstructions.

### 4.5.3 Noisiness of constraints

To prepare for systematic studies of reconstruction feasibility, we classify our diffractive imaging simulations using convenient signal-to-noise parameters. One such consideration is the photon-shot-noise in diffraction data.

Photon-shot-noise is conventionally related to the average number of scattered photons per pixel $\mu_{\text{tot}}$, regardless of whether it came from the charge or magnetization distribution (refer to equation (4.3)). Increasing $\mu_{\text{tot}}$ ought to improve the chances of reconstructing the total scattering distributions. However, using $\mu_{\text{tot}}$ as a signal-to-noise parameter is too optimistic since we are only interested in recovering the magnetization distribution [3]. One must still identify the magnetization distribution from the total scattering distributions, even if the latter is correctly determined (i.e. to extract the magnetization Figure 4.2a from only Figure 4.2c).

To characterize the noisiness of the photon data, instead of the total scattered power $\mu_{\text{tot}}$, we use $\mu_m$: the average number of photons scattered due to the magnetization in each pixel within the support $S$. In experiments, $\mu_m$ can be estimated directly from magnetic elastic scattering amplitude $f^M$, photon fluence and exposure time of diffraction measurement. In our simulations, $\mu_m$ is

[3] The size and shape of the beamstop also affect $\mu_{\text{tot}}$ without practical significance to reconstruction success.

computed as

$$\mu_m = \frac{\phi}{N_S} \sum_{\mathbf{q}} |m(\mathbf{q})|^2 \ , \tag{4.12}$$

where $m(\mathbf{q})$ is the discrete Fourier transform of the magnetization distribution $m(\mathbf{r})$; $N_S$ is the number of support pixels; $\phi$ is the same scalar in equation (4.3) which we vary to give the desired total number of scattered photons; $f^M$ is again set to unity inconsequentially. The product $\mu_m N_S$ corresponds to the total number of photons scattered per pulse in the absence of charge scattering.

The other noise consideration comes from charge scattering. We assume that the specimen's random charge distribution is unknown, which results in a harder reconstruction problem. As a consequence, the model magnetizations in Figure 4.5 will not agree with those in the total scattering amplitudes of equation (4.2), which includes the charge distribution. Essentially, this makes our direct-space constraint noisy [4]. Experimental measurement of the charge distribution would certainly reduce this noise and simplify the reconstruction.

## 4.6 Reconstruction algorithm

### 4.6.1 Modifying the difference map

Seeking the true magnetization distribution is equivalent to finding the intersection of the Fourier and direct-space constraint sets. Such intersections, or

[4] One could include the expected statistics on the charge distribution in Figure 4.5. This will certainly make the direct-space and Fourier constraints more compatible, potentially improving the reconstruction success rate. Even having included the charge statistics it may still be fairly challenging afterwards to isolate the magnetization distribution from these reconstructions chiefly because the exact charge distribution is unknown. Smoothing operations can remove charge contrast only if it is small compared to the magnetic contrast.

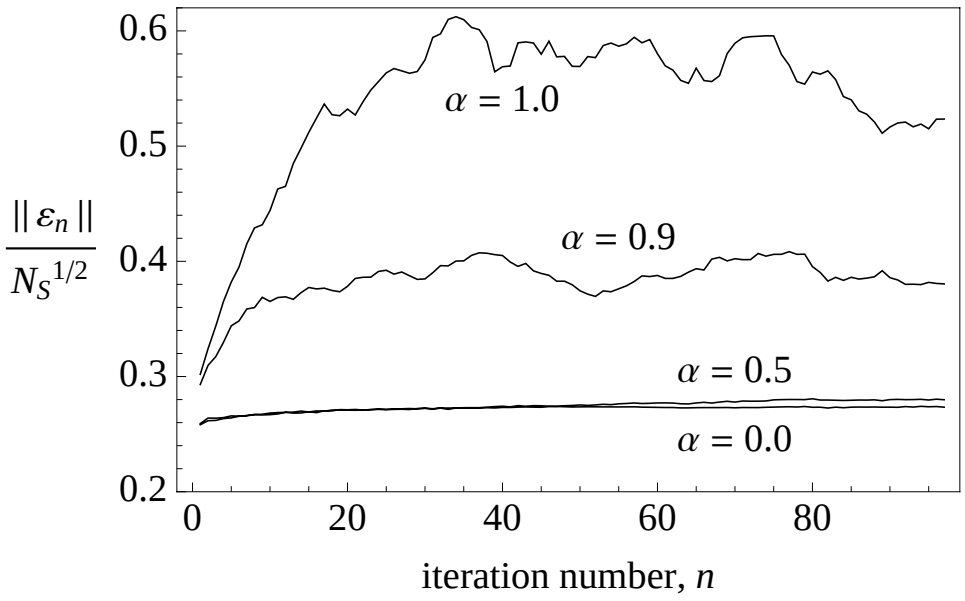



Figure 4.6: Stability of modified difference map, equation (4.13), around a solution. We estimate this stability by tracking how quickly the iterate leaves the solution because of the noisy constraints. Starting from the solution magnetization distribution, the normalized magnitude of the iterate's updates ($||\varepsilon_n||/N_S^{1/2}$) remains low for $\alpha < 0.5$, indicating iterate stability around a noisy solution.

*solutions*, can be discovered using an iterative constraint-satisfaction algorithm: the difference map [19], which uses simple projections to these two constraints ($P_D$, projection to direct-space constraint; $P_F$, projection to Fourier constraint).

The difference map algorithm was optimized for noiseless constraints sets with true intersections [15]. This algorithm accelerates the discovery of a solution, primarily by reducing the dimension of the search space [19]. It is also particularly efficient in extricating the iterate from near intersections (false solutions) to prevent the search from stalling.

Unfortunately, photon and charge scattering noise distorts our measurement

of the true Fourier constraint, demoting its intersections with the direct-space constraint to near intersections, from which iterates are jettisoned. This prohibits the search from reaching the true magnetization distributions encoded in these near intersections.

To increase its reconstruction success rate, the difference map was modified to improve the stability of the iterate around a near intersection. This is accomplished by an intermediate step to the iteration $m_n \rightarrow m'_n \rightarrow m_{n+1}$ (where $n = 0, 1, 2, \ldots$), which keeps the iterate close to the Fourier constraint [5]:

$$\begin{aligned} m'_n &= \alpha\, m_n + (1-\alpha) P_F(m_n)\ , \\ \varepsilon_n &= P_F\left(\, 2P_D(m'_n) - m'_n \,\right) - P_D(m'_n)\ , \\ m_{n+1} &= m'_n + \varepsilon_n\ , \end{aligned} \tag{4.13}$$

with $P_D$ and $P_F$ as the direct-space and Fourier constraint projections respectively and $\alpha$ as the map's modification parameter[6]. The update on the iterate is denoted $\varepsilon_n$, so that it may be referenced concisely in later paragraphs.

[5] We prefer the iterate to orbit near the Fourier constraint since it is a direct experimental measurement of a particular magnetization distribution, as opposed to the direct-space constraint which is a broader description of the ensemble of distributions.

[6] The modified difference map in equation (4.13) resembles the relaxed averaged alternating reflections algorithm (RAAR) used in iterative phase retrieval [30]. Like the modified difference map, RAAR was designed to stabilize iterates in the domain of attraction of a solution given noisy diffraction data. To see their resemblance, we combine the last 2 lines of equation (4.13) as a single operation $D$:

$$m'_n = \alpha\, m_n + (1-\alpha) P_F(m_n)\ , \tag{4.14}$$

$$m_{n+1} = D(m'_n)\ . \tag{4.15}$$

The first step in the next iteration would be

$$m'_{n+1} = \alpha\, m_{n+1} + (1-\alpha) P_F(m_{n+1}) \tag{4.16}$$

$$= \alpha\, D(m'_n) + (1-\alpha) P_F(D(m'_n))\ , \tag{4.17}$$

which is similar in structure to the RAAR update:

$$m_{n+1} = \alpha\, D(m_n) + (1-\alpha) P_F(m_n)\ . \tag{4.18}$$

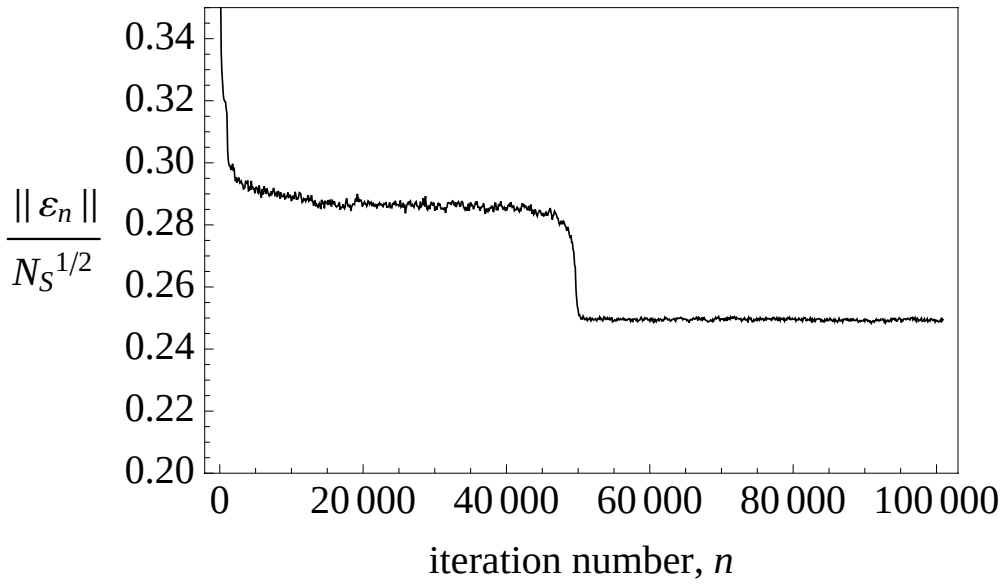



Figure 4.7: Signature of a successful reconstruction. The normalized error metric ($||\varepsilon_n||/N_S^{1/2}$) suffers a noticeable transition at iteration number $n \approx 50000$ during the successful reconstruction of the distribution in Figure 4.10b.

In our reconstructions we chose $\alpha = 0$, which substantially improves the iterate's stability (Figure 4.6) while reducing the number of computations in the first step of the algorithm. With $\alpha = 1$, equation (4.13) reduces to an instance of the original difference map.

The modified difference map is iteratively applied to a random, initial magnetization distribution $m_0$. The norm of the map's update $||\varepsilon_n||$, which we term the *error metric*, measures the average change of the iterate during the search. When the error metric drastically declines, it indicates that the difference map updates have experienced a dynamic transition and the search has likely converged (Figure 4.7). Because of the inherent noise in the constraints, the error metric will never vanish as it would, had an intersection of the two constraints been found in the noiseless case. When a noticeable transition in $||\varepsilon_n||$ occurs

and is stable, we harvest the *candidate solution* of the magnetization distribution, $P_D\,(P_F\,(\,2P_D(m'_n) - m'_n\,))$. The correctness of this candidate solution is tested when compared against other candidate solutions from different, random, initial iterates $m_0$. Consistent recovery of nearly identical candidate solutions, up to an overall multiplicative sign or inversion, from random restarts asserts their credibility as the true magnetization distribution. One can smooth out the fluctuations between the candidate solutions by averaging them.

In searches using the noisiest photon data, the error metric $||\varepsilon_n||$ will never show a clear transition. In such cases, recovering the true magnetization distribution is plainly impossible. Nonetheless, we can still evaluate the search results, however wrong they may be. From equation (4.13), notice that $||\varepsilon_n||$ also measures the distance between two points on the two constraints: $P_F\,(\,2P_D(m'_n) - m'_n\,)$ and $P_D(m'_n)$. Hence the minimum $||\varepsilon_n||$ during a search signals the nearest distance between the two constraints — the best alternative to discovering an intersection. Unlike more robust candidate solutions with less noisy data, these *faux solutions* are never repeated with random restarts.

### 4.6.2 Projection to direct-space constraint

The projection to the direct-space constraint, $P_R(m)$, comprises the following operations on $m$:

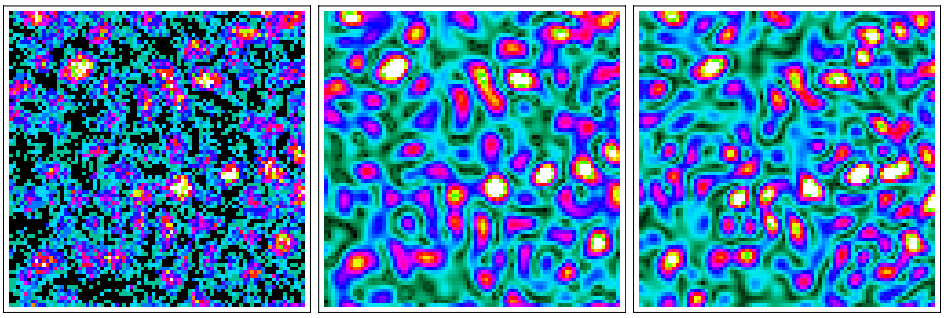
1. set all values of $m$ outside the support $S$ to zero;
2. modify the model magnetizations $\widetilde{m}$ to have the same mean and variance as those of $m$ within the support $S$;

3. replace the values of $m$ inside $S$ with the modified $\widetilde{m}$, such that the ordering of the former is preserved.

It is clear from step 2 that the mean scattering amplitude $\langle f^C(\mathbf{r})\rangle$ and the magnetic contrast $\Delta_m$ are not directly constrained during phase retrieval but are indirectly constrained by the diffraction data [7].

In actual experiments where the magnetization histogram $\widetilde{m}$ is not readily available or simulated, one could instead fit the iterate's sorted values to a function qualitatively similar to Figure 4.5, then replace the iterate's values with those sampled from its fit while preserving the ordering of values in the original iterate.

### 4.6.3 Projection to Fourier constraint



photon data | healed photon data | true intensities

Figure 4.8: Speckle-healing by applying an autocorrelation support constraint to the photon data during reconstruction. We show this healing of a magnified section of the photon data in Figure 4.4.

[7] The mean charge scattering amplitude $\langle f^C(\mathbf{r})\rangle$ is non-critical to the reconstruction since it constitutes mainly the missing intensities in the data where the diffraction intensities from the sample's magnetization is low (see Figure 4.4).

Before discussing the projection to the Fourier constraint, $P_F(m)$, we describe a modification to the diffraction data which lowers the photon-shot-noise using the direct-space constraint. If the scattering distribution is contained within a direct-space support $S$, the Fourier transform of the diffraction intensities — or the autocorrelation of the direct-space scattering distribution — should be contained within the autocorrelation support $S_A$.

We can lower the noise in the diffraction data using our knowledge of the support, hence constraining the photon data to have the expected speckles. We did so by applying an autocorrelation support constraint to the Fourier transform of the photon data — setting all values outside $S_A$ in the data's Fourier transform to zero. Empirically, this *speckle-healing operation* increases the R-factor between the processed photon data and the true intensities (see Figure 4.8).

However, the missing data within the beamstop may confuse speckle-healing. These central Fourier amplitude are indirectly constrained by the diffraction data and after numerous iterations the difference map iterate proposes preliminary intensities for them. We replace the missing photon data with these preliminary intensities before applying the speckle-healing operation. In our reconstructions, the photon data was healed this way every 1000 iterations, which were then used to constrain iterations until the next healing. When the reconstruction converges under this adiabatic healing process, we are assured that it is still compatible with the photon data.

With this adiabatic speckle-healing procedure in mind, the projection to the Fourier constraint, $P_F(m)$, comprises the following operations on the iterate's Fourier transform $m(\mathbf{q})$:

1. set the amplitudes of $m(\mathbf{q})$ outside the beamstop to the square root of the speckle-healed photon data, while retaining the phases of $m(\mathbf{q})$;
2. the $m(\mathbf{q})$ values within the beamstop are unchanged.

## 4.7 Feasibility

### 4.7.1 Difference map reconstructions

Unlike an actual experiment, the true magnetization distributions are known in our simulated experiments. This allows us to directly compare the reconstructions $m^{\text{rec}}$ with the true distribution $m^{\text{true}}$ within the support $S$ via the following deviation measure:

$$\delta = \frac{1}{2} \sum_{\mathbf{r} \in S} \left( u^{\text{rec}}(\mathbf{r}) - u^{\text{true}}(\mathbf{r}) \right)^2 \tag{4.19}$$

$$u(\mathbf{r}) = \frac{m(\mathbf{r})}{\left( \sum_{\mathbf{r} \in S} m(\mathbf{r})^2 \right)^{1/2}} \, . \tag{4.20}$$

The deviation $\delta$ is proportional to the square of the distance between $u^{\text{rec}}$ and $u^{\text{true}}$, which are the respective distributions normalized as unit vectors. Allowing for an overall sign in the reconstructed magnetization, deviations lie within the range $0 \leq \delta \leq 1$.

In our simulations, reconstructions with $\delta < 0.2$ were consistently recovered from random restarts. In actual experiments, only the consistency test is available to evaluate the reliability of the reconstructions.

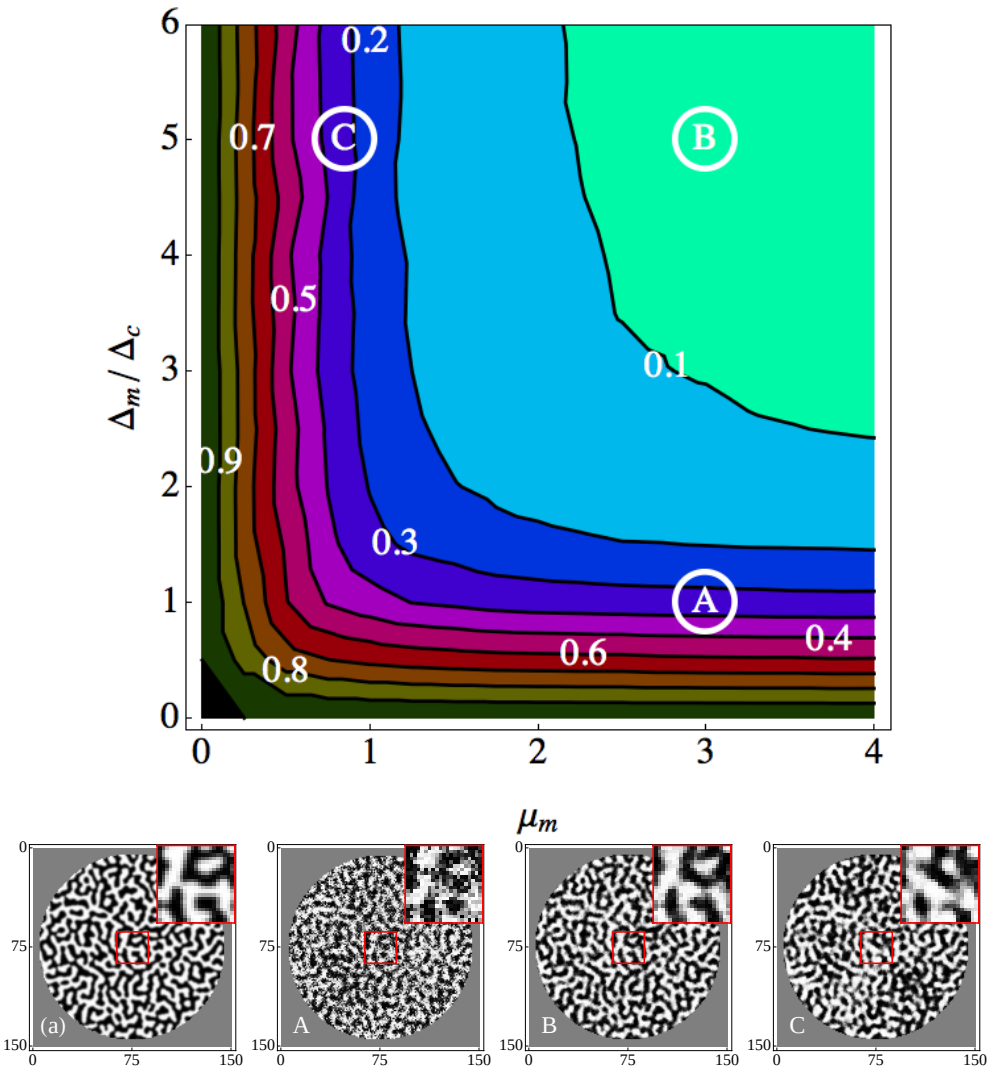



Figure 4.9: Feasibility of diffractive imaging at various noise levels. The top panel shows a contour plot of the reconstruction deviation $\delta$ (in equation (4.20)) generated from many simulated reconstructions of the pure magnetization distribution similar to that in (a) as the signal-to-noise parameters ($\Delta_m/\Delta_c$ and $\mu_m$) were independently varied. Panels A, B and C show reconstructions of (a) subject to corresponding signal-to-noise parameters marked in the top panel.

We systematically studied the efficacy of our reconstruction algorithm when we varied the two signal-to-noise parameters: relative magnetic contrast $\Delta_m/\Delta_c$

and the average number of photons scattered from the magnetic distribution $\mu_m$. As Figure 4.9 indicates, increasing $\Delta_m/\Delta_c$ and $\mu_m$ improves the accuracy of the reconstructions. However the same figure shows that the effects of suppressing either $\Delta_m/\Delta_c$ or $\mu_m$ are qualitatively different — lowering one variety of noise can not compensate for the reconstruction errors caused by the other.

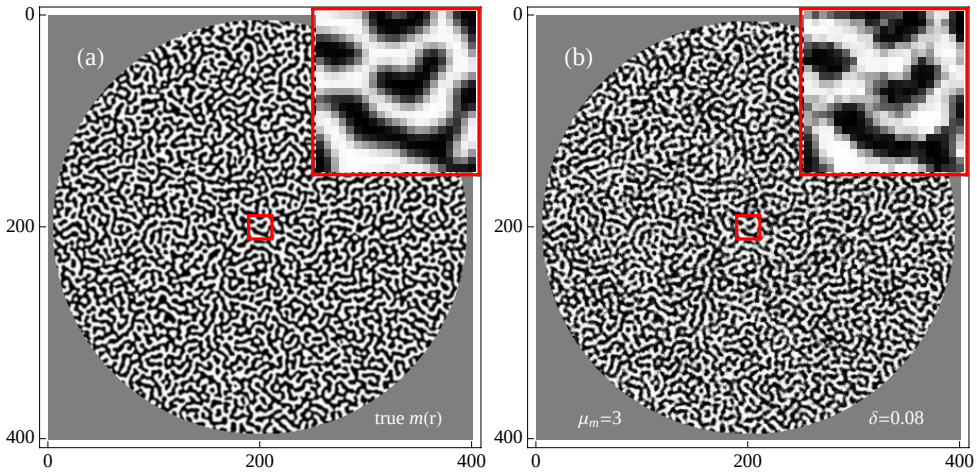



Figure 4.10: In panel (b), a modified difference map reconstruction of a simulated magnetization distribution, panel (a), with signal-to-noise parameters corresponding to point B in Figure 4.9. The diffraction data used for this reconstruction is shown in Figure 4.4. If the magnetic domains are 170 nm wide, then having only three photons scattered by the magnetic contrast within each 34 nm pixel was sufficient to reconstruct the pattern in the bottom panel.

Magnetization distributions shown in Figure 4.10 are routinely reconstructed with low deviations $\delta \leq 0.1$ given remarkably noisy data typical of Figure 4.4: relative magnetic contrast $\Delta_m/\Delta_c = 5.0$ and average scattered photons due to the magnetization in each pixel $\mu_m = 3$. The deviation of reconstructions from the true domain pattern at various noise levels is numerically computed in Figure 4.9 and appears to be independent of the support size at a

constant domain resolution (the reconstructions in Figure 4.9B and Figure 4.10b suffered comparable noise levels).

Reconstructions with the unmodified difference map, $\alpha = 1$ in equation (4.13), do not converge within the range of noise parameters in Figure 4.9: neither in the sense of achieving a dynamic transition in the error metric (Figure 4.7) nor repeatability given random restarts. We witness this lesser performance even with reconstructions using the modified difference map when we omit either the histogram constraint in the direct-space projection or the speckle-healing procedure or both.

### 4.7.2 Comparison with Fourier transform holography

To provide perspective, we compared our reconstructions with those from simulated Fourier transform holography (FTH) in Figure 4.11. In FTH, the domain pattern is obtained directly from its cross-correlation with an aptly machined reference pinhole [14]. This cross-correlation is obtained from a simple Fourier transform of the measured diffraction intensities without the need for phase retrieval.

To make the comparison more compelling, we provided our FTH simulations with the following advantages over the non-holographic method:

1. diffraction signal within the beamstop region was provided;
2. single-pixel reference pinhole for highest possible reconstruction resolution (pinhole diameter roughly 34 nm if magnetic domains are 170 nm wide) whereas the pinhole in [14] which, at its narrowest part, had a di-

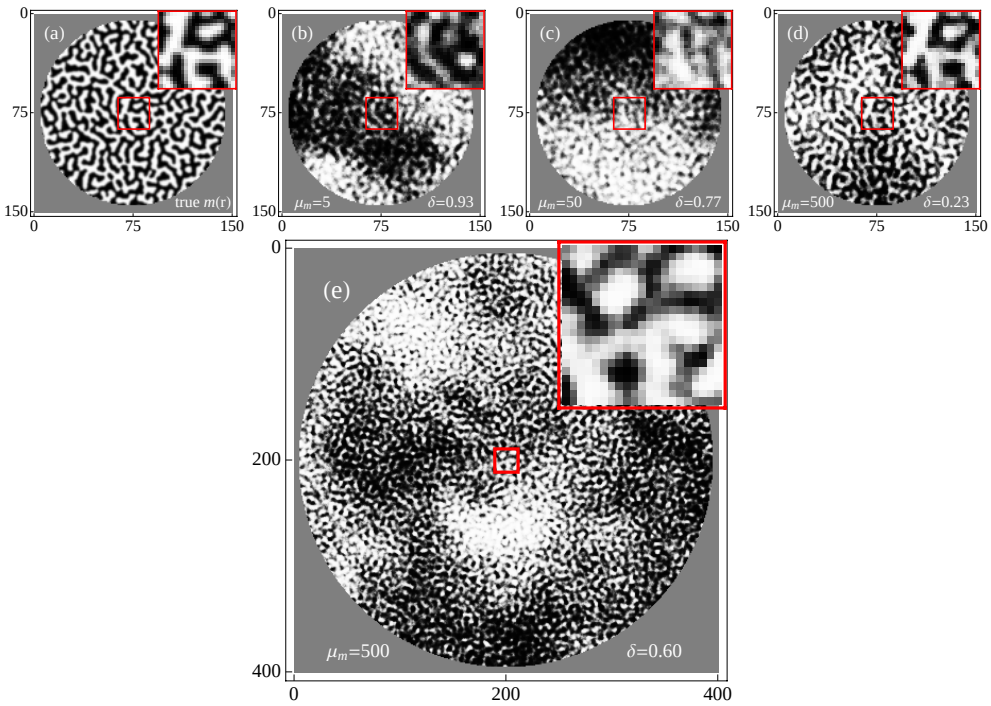



Figure 4.11: Simulated reconstructions with Fourier transform holography (FTH). The panels (b), (c), (d) show how FTH require more scattered photons for acceptably low reconstruction deviations. Panel (e) is a FTH reconstruction of Figure 4.10a. Panels (d) and (e) together show that reconstructions worsen with increasing support size. Direct-space projections were applied to these FTH reconstructions to enhance their contrast; their relative magnetic contrast were $\Delta_m/\Delta_c = 10$.

ameter of 100 nm.

In our simulated FTH reconstructions, the ratio of the number of photons scattered by the magnetic contrast in each support pixel to the number which pass through each pixel of the reference pinhole is 1:50, as estimated from [14].

At the low signal-to-noise levels of Figure 4.9, the low deviation reconstructions using our proposed non-holographic diffractive imaging technique are out of the reach of FTH (compare Figures 4.9 to 4.11). Figure 4.11 also illus-

trates that FTH reconstructions worsen with increasing support size because the photon fluence through the pinhole does not increase commensurately. Non-holographic diffractive imaging does not suffer this size dependency since only the noise per support pixel is important. In addition, non-holographic diffractive imaging also does not need the experimental fabrication of a small reference pinhole and uses the beam's spatial coherence more efficiently, although it requires accurate knowledge of the support [46].

## 4.8 Conclusions

Ultrafast imaging of magnetic nanostructures is presumably possible within the noise limits predicted by Figure 4.9. This, of course, is valid only in the absence of other varieties of noise. Our study is limited to magnetic imaging without prior measurement of the random charge distribution. We speculate that the reconstruction noise limits would improve if the specimen's charge distribution, which may fluctuate, were available.

We also showed that non-holographic diffractive imaging is more competitive than Fourier transform holography given exceptionally noisy data and when the conditions of our simulations apply. Although our feasibility demonstration relies on the knowledge of the histogram of magnetization values, minor variations of our assumed histogram do not change their validity for imaging ferromagnetic contrast. What is certain is that with a working knowledge of the histogram of magnetization values, we expect our methods to be relevant to magnetic imaging experiments with remarkably noisy data.

The modified difference map, equation (4.13), may be relevant to constraint-

satisfaction problems that suffer from imprecise or noisy constraints. Furthermore, the speckle-healing procedure in this chapter is pertinent to the recovery of missing global information common in diffractive imaging.

## 4.9 Acknowledgements

This work was supported by CHESS through NSF and NIH / NIGMS via NSF award DMR-0225180. I would like to thank Yoav Kallus, for his indispensable insights with regards to the domain-generation prescription, and Victor Lo for pointing out the similarity of our modified difference map to the RAAR algorithm [30] (see footnote 6).

APPENDIX A

## FIXED-POINT ROTATIONAL INVARIANCE

We wish to show that the intensities of the true model, $\widetilde{W}$, or more generally any rotation of this model, $\widetilde{W}^{\mathbf{R}}$, is a fixed point of the maximization update rule (2.19). Given the true model we can write down the probability distribution of the photon counts $K$:

$$P(K) = \sum_{j=1}^{M_{\rm rot}} w_j \, R_j(\widetilde{W}, K) \; , \tag{A.1}$$

where the joint Poisson distribution

$$R_j(\widetilde{W}, K) = \prod_{i=1}^{M_{\rm pix}} \frac{\widetilde{W}_{ij}^{K_i}}{K_i!} \exp\left(-\widetilde{W}_{ij}\right) \tag{A.2}$$

is the same (up to an irrelevant factor) as (2.13) but the data index $k$ has been replaced by the function argument $K$ representing an arbitrary vector of photon counts. Because the probability distribution on $K$ is unchanged if the model is rotated, we have

$$\sum_{j=1}^{M_{\rm rot}} w_j \, R_j(\widetilde{W}^{\mathbf{R}}, K) = \sum_{j=1}^{M_{\rm rot}} w_j \, R_j(\widetilde{W}, K) \; , \tag{A.3}$$

for arbitrary rotations $\mathbf{R}$. The distribution (A.1) and the invariance (A.3) are approximations that become exact in the limit $M_{\rm rot} \to \infty$.

Taking the numerator of the maximization update rule (2.19) and evaluating it for the rotated model $\widetilde{W}^{\mathbf{R}}$ with the data sum replaced by a sampling of $P(K)$, we have

$$\sum_{k=1}^{M_{\rm data}} P_{jk}(\widetilde{W}^{\mathbf{R}}) K_{ik} = M_{\rm data} \sum_K P(K) \left( \frac{w_j R_j(\widetilde{W}^{\mathbf{R}}, K)}{\sum_{j'=1}^{M_{\rm rot}} w_{j'} R_{j'}(\widetilde{W}^{\mathbf{R}}, K)} \right) K_i \; . \tag{A.4}$$

Substituting (A.1) for $P(K)$ and using (A.3), this reduces to

$$\begin{aligned} \sum_{k=1}^{M_{\rm data}} P_{jk}(\widetilde{W}^{\mathbf{R}}) K_{ik} &= M_{\rm data} \sum_K w_j \, R_j(\widetilde{W}^{\mathbf{R}}, K) \, K_i & \text{(A.5)} \\ &= M_{\rm data} \, w_j \widetilde{W}_{ij}^{\mathbf{R}} & \text{(A.6)} \end{aligned}$$

by the property that the mean of $K_i$ for the Poisson distribution is $\widetilde{W}_{ij}^{\mathbf{R}}$. By the same steps, the denominator of the update rule gives

$$\sum_{k=1}^{M_{\mathrm{data}}} P_{jk}(\widetilde{W}^{\mathbf{R}}) = M_{\mathrm{data}}\, w_j \ , \tag{A.7}$$

thus showing the desired fixed point property

$$\mathbf{M}: \quad \widetilde{W}_{ij}^{\mathbf{R}} \to \widetilde{W}_{ij}^{\mathbf{R}} \ . \tag{A.8}$$

APPENDIX B

## MUTUAL INFORMATION FORMULA

To evaluate the mutual information $I(K,\Omega)|_W$ using the quantities used by the EMC algorithm we need to approximate the integral over orientations $\Omega$ by a sum over the discrete samples $j$ and expectation values over photon counts $K_i$ by a normalized sum over the counts $K_{ik}$ in the actual data ($k$ is the data index).

Suppressing the model variables $W$, which we treat as a fixed quantity whenever $I(K,\Omega)|_W$ is calculated by the EMC algorithm, the mutual information is given by

$$I(K,\Omega) = \int_\Omega \sum_K P(K)P(\Omega|K)\log\frac{P(\Omega|K)}{P(\Omega)}\ . \tag{B.1}$$

Replacing the $\Omega$-integral by a weighted sum over samples $j$ and the $K$-expectation by a sum over the data, we obtain

$$I(K,\Omega) = \sum_{j=1}^{M_{\rm rot}} w_j \frac{1}{M_{\rm data}} \sum_{k=1}^{M_{\rm data}} P(\Omega_j|K_k)\log\frac{w_j P(\Omega_j|K_k)}{w_j}\ .$$

We recover formula (2.27) with the identification

$$w_j P(\Omega_j|K_k) = P_{jk}(W)\ . \tag{B.2}$$

APPENDIX C

## ROTATION GROUP SAMPLING BASED ON THE 600-CELL

The quaternion parameterization of $3 \times 3$ orthogonal matrices is given by the formula

$$\mathbf{R}(q) = \begin{pmatrix} 1-2q_2^2-2q_3^2 & 2q_1q_2+2q_0q_3 & 2q_1q_3-2q_0q_2 \\ 2q_2q_1-2q_0q_3 & 1-2q_1^2-2q_3^2 & 2q_2q_3+2q_0q_1 \\ 2q_3q_1+2q_0q_2 & 2q_3q_2-2q_0q_1 & 1-2q_1^2-2q_2^2 \end{pmatrix}, \qquad \text{(C.1)}$$

where the unit quaternions $q = (q_0, q_1, q_2, q_3)$ are points on the unit 3-sphere. Because of the property $\mathbf{R}(q) = \mathbf{R}(-q)$, the unit quaternions are mapped 2-to-1 to the elements of the 3D rotation group. The multiplication rule for quaternions is most transparent in the $2 \times 2$ spin-1/2 representation

$$q = q_0 + i\,\boldsymbol{\sigma} \cdot \mathbf{q}\,, \qquad \text{(C.2)}$$

($\boldsymbol{\sigma}$ are the Pauli matrices) which defines the "scalar" ($q_0$) and "vector" ($\mathbf{q}$) parts of the 4-vector. The scalar part encodes just the rotation angle, $\theta$, while the vector part also carries information about the rotation axis, $\mathbf{n}$:

$$q_0 = \cos(\theta/2) \qquad \mathbf{q} = \sin(\theta/2)\,\mathbf{n}\,. \qquad \text{(C.3)}$$

From (C.2) and properties of the Pauli matrices one can verify that the inverse of a rotation is obtained by reversing the sign of the vector part. This fact is consistent with equations (C.1) and (C.3).

Since the rotation required to move the element $q$ to the element $q'$ is the quaternion $q'q^{-1}$, a group-invariant distance between these elements should only be a function of the rotation angle (conjugacy class) of $q'q^{-1}$, that is, its scalar part: $q_0q_0' + \mathbf{q}\cdot\mathbf{q}'$. Since the latter equals

$$q \cdot q' = 1 - \|q - q'\|^2/2\,, \qquad \text{(C.4)}$$

the standard Euclidean distance on the quaternion 3-sphere is a group invariant distance between rotation group elements. This last remark is key for generating efficient samplings of the 3D rotation group: the unit quaternions should be placed to efficiently *cover* the 3-sphere [9]. A cover is optimal if it has the minimum number of points with the property that an arbitrary point of the space is always within a given covering radius $r_{\rm c}$ of some point in the cover. In our case the covering radius corresponds to a rotation angle $\delta\theta$, such that an arbitrary rotation group element is always within a rotation by $\delta\theta$ from one of our samples. If $q$ is a point of the cover and $q'$ an arbitrary point, then

$$\begin{aligned} r_{\rm c}^2 &> \|q - q'\|^2 \\ &= 2 - 2q \cdot q' \\ &= 2 - 2\left(q'q^{-1}\right)_0 \\ &= 2 - 2\cos\left(\theta/2\right) \\ &\approx (\theta/2)^2. \end{aligned}$$

The covering radius, of the 3-sphere by quaternions, and the angular resolution $\delta\theta$ of the rotation group sampling are therefore related by:

$$r_{\rm c} \approx \delta\theta/2\ . \tag{C.5}$$

Our covers of the 3-sphere are based on the highly symmetric 4-dimensional polytope that has the greatest number of regular tetrahedra — 600 — as its 3D facets. Known as the $\{3,3,5\}$ polytope, or 600-cell, it is the 4D analog of the regular icosahedron [10]. The tetrahedral facets of the 600-cell are well covered by points arranged in the fcc lattice. We will use the integer $n$ to describe the degree of the refinement of each tetrahedron by points of the fcc lattice. For example, when $n = 4$, each tetrahedron edge is divided into 4 equal segments

thus introducing $n-1=3$ edge points. This value of $n$ also introduces $(n-1)(n-2)/2=3$ points on each tetrahedron face and $(n-1)(n-2)(n-3)/6=1$ points in the interior of the tetrahedron. The $\{3,3,5\}$ polytope has 120 vertices, 720 edges, 1200 faces, and 600 cells. Combining this information with the point counts of the tetrahedron refinements, we obtain the following formula for the number of sample points on the 3-sphere:

$$\begin{aligned} & 2M_{\mathrm{rot}}(n) \\ &= 120+720(n-1)+1200\frac{(n-1)(n-2)}{2}+600\frac{(n-1)(n-2)(n-3)}{6} \\ &= 20(n+5n^3). \end{aligned}$$

The factor of 2 on the left side takes into account the overcounting by $\pm$ quaternion pairs.

To determine the appropriate level of refinement $n$, we need to compute the corresponding angular resolution $\delta\theta(n)$. Consider one tetrahedral cell of $\{3,3,5\}$; in canonical coordinates [10] the vertices are

$$\begin{aligned} v_1 &= (1,0,0,0) \\ v_2 &= \tfrac{1}{2}(\tau,1,1/\tau,0) \\ v_3 &= \tfrac{1}{2}(\tau,1/\tau,0,1) \\ v_4 &= \tfrac{1}{2}(\tau,0,1,1/\tau)\ , \end{aligned}$$

where $\tau=(1+\sqrt{5})/2$ is the golden mean. The edge length of this tetrahedron is $1/\tau$; the tetrahedra of the refinement will have edge length $1/(n\,\tau)$ and this is also the minimum distance between fcc lattice points and $\sqrt{2}$ times the covering radius $r_{\mathrm{c}}$. When the fcc lattice points are projected to the unit 3-sphere by rescaling, the points near the center of the tetrahedron are expanded by the greatest amount. The linear expansion factor is given by the reciprocal of the distance

between the origin and the tetrahedron center:

$$4/\|v_1 + v_2 + v_3 + v_4\| = \sqrt{8}/\tau^2 \approx 1.080 \; . \tag{C.6}$$

The coarsest part of the sampling thus has minimum distance $r_c = 2/(n\,\tau^3)$. Using (C.5) we obtain:

$$\delta\theta(n) = 4/(n\,\tau^3) \approx 0.944/n \; . \tag{C.7}$$

The samples of the 3D rotation group in this scheme carry non-uniform weights. Because the measure on the continuous group is just the volume element of the 3-sphere, the weight associated with a sample is proportional to the volume of its associated Voronoi cell. These weights / volumes are non-uniform as a result of two effects. First, there is a correction that affects the samples at the vertices and edges of the 600-cell, where the joining of the tetrahedral cells results in angular deficits. The second correction to the weights arises from the non-uniform distortion of the Voronoi cells, when these are projected from the 600-cell to the 3-sphere.

A regular (flat) tetrahedron has dihedral angle

$$\alpha = \cos^{-1}(1/3) \approx 70.5^\circ \; . \tag{C.8}$$

Because five tetrahedra meet at every edge of the 600-cell, the fractional volume associated with samples on edges is given by

$$f_1 = 5\alpha/(2\pi) \approx 0.979566 \; . \tag{C.9}$$

The spherical angle subtended at a vertex of the regular tetrahedron, by spherical trigonometry, is $3\alpha - \pi$. Because 20 tetrahedra meet at every vertex of the 600-cell, each vertex sample has fractional volume

$$f_0 = 20(3\alpha - \pi)/(4\pi) \approx 0.877398 \; . \tag{C.10}$$

Since there are no deficits at samples on faces or within the cells of the polytope,

$$f_2 = f_3 = 1 \ . \tag{C.11}$$

The volume change upon projecting a Voronoi cell from a $\{3,3,5\}$ facet to the unit 3-sphere is the result of two things: (1) a uniform expansion by the linear scale factor $1/\|\tilde{q}\|$, where $\tilde{q}$ is the (non-unit) quaternion of the sample on $\{3,3,5\}$, and (2) projection of the 3-space of the tetrahedral cell to the tangent space of the 3-sphere at $q$, the projection of $\tilde{q}$. The second of these produces a reduction by the factor $q \cdot c$, where $c$ is the unit outward normal vector to the facet on which $\tilde{q}$ resides. In terms of the four cell-vertices,

$$c = \frac{v_1 + v_2 + v_3 + v_4}{\|v_1 + v_2 + v_3 + v_4\|} \ . \tag{C.12}$$

The overall (unnormalized) weight of a sample $q$, originating from sample $\tilde{q}$ on the 600-cell, is given by

$$w(q) = f_k \, \frac{q \cdot c}{\|\tilde{q}\|^3} \ , \tag{C.13}$$

where $k = 0, 1, 2, 3$ is the associated dimensionality of the sample (vertex, edge, etc.). Vertex samples have the lowest weight, samples at the cell center the highest; their ratio is

$$\frac{w_{\text{vertex}}}{w_{\text{center}}} = \frac{f_0}{f_3} \left( \frac{\tau^2}{\sqrt{8}} \right)^4 \approx 0.644 \ . \tag{C.14}$$

In our pre-computed tables (for given $n$) of quaternion samples $q_j$ we include the weights $w_j \propto w(q_j)$ as the fifth component. All formulas in this chapter assume the normalization

$$\sum_j w_j = 1 \ . \tag{C.15}$$

## BIBLIOGRAPHY


[1] S. Bajt et al. Camera for coherent diffractive imaging and holography with a soft x-ray free-electron laser. *Applied Optics*, 47:1673–1683, 2008.

[2] W. H. Benner et al. Non-destructive characterization and alignment of aerodynamically focused particle beams using single particle charge detection. *Journal of Aerosol Science*, 39:917–928, 2008.

[3] M. J. Bogan et al. 2009. unpublished.

[4] M. J. Bogan et al. Aerosol imaging with a soft x-ray free electron laser. *Aerosol Science and Technology*, 44(3):i–iv, 2010.

[5] G. Bortel and G. Faigel. Classification of continuous diffraction patterns: A numerical study. *Journal of Structural Biology*, 158:10–18, 2007.

[6] J. Chalupsky et al. Characteristics of focused soft x-ray free-electron laser beam determined by ablation of organic molecular solids. *Optics Express*, 15(10), 2007.

[7] H. N. Chapman et al. Femtosecond diffractive imaging with a soft-x-ray free-electron laser. *Nature Physics*, 2:839–843, 2006.

[8] H. N. Chapman et al. High-resolution ab initio three-dimensional x-ray diffraction microscopy. *Journal of Optical Society America, A*, 23(5), May 2006.

[9] J. H. Conway and N. J. A. Sloane. *Sphere packing, lattices and groups*. Springer-Verlag, NY, 1998.

[10] H. S. M. Coxeter. *Regular Polytopes*. Dover, 1973.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39:1–38, 1977.

[12] P. P. B. Eggermont, G. T. Herman, and A. Lent. Iterative algorithms for large partitioned linear systems, with applications to image reconstruction. *Linear Algebra and Its Applications*, 40:37–67, 1981.

[13] S. Eisebitt et al. Polarization effects in coherent scattering from magnetic specimen: Implications for x-ray holography, lensless imaging, and correlation spectroscopy. *Physical Review B*, 68:104419, Jan 2003.

[14] S. Eisebitt et al. Lensless imaging of magnetic nanostructures by x-ray spectro-holography. *Nature*, 432(7019):885–8, Dec 2004.

[15] V. Elser. Phase retrieval by iterated projections. *Journal of the Optical Society of America A*, 20(1):40–55, 2003.

[16] V. Elser. Noise limits on reconstructing diffraction signals from random tomographs. *IEEE Transactions of Information Theory*, 55:4715–4722, 2009.

[17] V. Elser and S. Eisebitt. Uniqueness transition in noisy phase retrieval. *New Journal of Physics*, 2010. Submitted.

[18] V. Elser and R. P. Millane. Reconstruction of an object from its symmetry-averaged diffraction pattern. *Acta Crystallography*, A64:273–279, 2008.

[19] V. Elser, I. Rankenburg, and P. Thibault. Searching with iterated maps. *Proceedings of the National Academy of Sciences (USA)*, 104:418–423, 2007.

[20] J. R. Fienup. Phase-retrieval algorithms: a comparison. *Applied Optics*, 21(15), 1982.

[21] R. Fung et al. Structure from fleeting illumination of faint spinning objects in flight. *Nature Physics*, 5:64–67, 2009.

[22] C. Gutt et al. Single-pulse resonant magnetic scattering using a soft x-ray free-electron laser. *Physical Review B*, 81(10):100401, Mar 2010.

[23] J. P. Hannon, G. T. Trammell, M. Blume, and D. Gibbs. X-ray resonance exchange scattering. *Physical Review Letters*, 61(10):1245, Sep 1988.

[24] S. P. Hau-Riege, R. A. London, and A. Szoke. Dynamics of biological molecules irradiated by short x-ray pulses. *Physical Review E*, 69:051906, 2004.

[25] G. Huldt, A. Szoke, and J. Hajdu. Diffraction imaging of single particles and biomolecules. *Journal of Structural Biology*, 144:219–227, 2003.

[26] L. Koerner, H. Philipp, M. Hromalik, and M. Tate. X-ray tests of a pixel array detector for coherent x-ray imaging at the LINAC coherent light source. *Journal of Instrumentation*, 4(03):P03001, 2009.

[27] J. B. Kortright et al. Soft x-ray small-angle scattering as a sensitive probe of magnetic and charge heterogenity. *Physical Review B*, 64:092401, Jan 2001.

[28] N. D. Loh and V. Elser. A reconstruction algorithm for single-particle diffraction imaging experiments. *Physical Review E*, 80:026705, 2009.

[29] N. D. Loh, V. Elser, et al. Cryptotomography: Reconstructing 3d fourier intensities from randomly oriented single-shot diffraction patterns. *Physical Review Letters*, 104:225501, 2010.

[30] D. R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse problems*, 21(1):37–50, 2005.

[31] J. Miao et al. Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400:342–344, 1999.

[32] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu. Potential for biomolecular imaging with femtosecond x-ray pulses. *Nature*, 406:752–757, 2000.

[33] K. A. Nugent. Coherent methods in the x-ray sciences. *Advances in Physics*, 59(1):1–99, 2010.

[34] M. S. Pierce, C. Buechler, L. Sorensen, and S. Kevan. Disorder-induced magnetic memory: Experiments and theories. *Physical Review B*, 75:144406, Jan 2007.

[35] M. S. Pierce et al. Quasistatic x-ray speckle metrology of microscopic magnetic return-point memory. *Physical Review Letters*, 90(17):175502, May 2003.

[36] N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, and H. R. Saibil. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, 107:869–879, 2001.

[37] H. Saga et al. New recording method combining thermo-magnetic writing and flux detection. *Japanese Journal of Applied Physics*, 38:1839–1840, 1999.

[38] E. L. Saldin et al. *The Physics of free electron lasers*. Springer-Verlag, Berlin, 1999.

[39] D. Sayre. Some implications of a theorem due to shannon. *Acta Crystallography*, 5:843, 1952.

[40] D. Sayre. Imaging processes and coherence in physics. In J. Ehlers et al., editors, *Lectures in Physics*. Springer-Verlag. Berlin, Heidelberg, New York, 1980.

[41] S. Scheres et al. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods*, 4:27–29, 2007.

[42] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623656, 1948.

[43] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. Neiman, and D. Sayre. Biological imaging by soft x-ray diffraction microscopy. *Proceedings of the National Academy of Sciences (USA)*, 102(43):15343–15346, 2005.

[44] S. D. Shastri, P. Zambianchi, and D. M. Mills. Dynamical diffraction of ultrashort x-ray free-electron laser pulses. *Journal of Synchrotron Radiation*, 8:1131–1135, 2001.

[45] J. Solem. Imaging biological specimens with high-intensity soft x-rays. *Journal of the Optical Society of America B*, 3(11):1551–1565, November 1986.

[46] P. Thibault. *Algorithmic methods in diffraction microscopy*. PhD thesis, Cornell University, 2007.

[47] P. Thibault and V. Elser. X-ray diffraction microscopy. *Annual Review of Condensed Matter Physics*, 1:5.1–5.19, 2010.

[48] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, and D. Sayre. Reconstruction of a yeast cell from x-ray diffraction data. *Acta Crystallography A*, 62:248–261, 2006.