

Aspects of Fourier imaging

Wen-Hsin Hsiao, B.E.(Hons.), B.Sci.

Department of Electrical and Computer Engineering

A thesis presented for the degree of
Doctor of Philosophy

University of Canterbury
Christchurch, New Zealand
May 2008

Abstract

A number of topics related to Fourier imaging are investigated. Relationships between the magnitude of errors in the amplitude and phase of the Fourier transform of images and the mean square error in reconstructed images are derived. The differing effects of amplitude and phase errors are evaluated, and “equivalent” amplitude and phase errors are derived. A model of the probability density function of the Fourier amplitudes of images is derived. The fundamental basis of phase dominance is studied and quantitated. Inconsistencies in published counter-examples of phase dominance are highlighted. The key characteristics of natural images that lead to their observed power spectral behaviour with spatial frequency are determined.

Acknowledgements

With the deepest gratitude I wish to thank my supervisor, Professor Rick P. Millane who has supported me throughout my research. I have been extremely fortunate to have Rick as my supervisor as he has generously shared his ideas and spent his valuable time to ensure I completed my work successfully. I would also like to thank Professor Philip Bones for his help and support.

I am grateful to the Department of Electrical and Computer Engineering at the University of Canterbury for financial support.

I would like to extend my thanks to the following people for their friendship and support; Nicholas Blakeley, who showed me the world of LaTex, Kristin McLoughlin, who allowed me to observe during her Ph.D oral examination, Wei-Lun Chiu, Yong Chew, Christopher Forne, Marilyn Lim and Chun Hoon Yoon for their friendship, help and many useful suggestions.

Lastly, I am forever thankful to my parents, sister and wife Diana, for their love and encouragement. I could not have done it without their support, and I love them dearly.

Contents

Abstract	iii
Acknowledgements	v
Contents	vii
Preface	xi
1 Introduction	1
1.1 Imaging	1
1.2 Electromagnetic radiation and the scalar approximation	1
1.3 Scalar diffraction theory	4
1.3.1 Fresnel diffraction	6
1.3.2 Fraunhofer diffraction	7
1.4 Geometric optics	8
1.5 Fourier optics and imaging	8
1.5.1 Fourier transforming property of a lens	8
1.5.2 Image forming property of a lens	10
1.5.3 Generalised imaging system model	12
1.5.4 Coherent and incoherent imaging	14
1.6 Diffraction imaging	16
1.6.1 Phase problems	17

1.6.2	Phase dominance	20
1.6.3	Optical astronomical imaging	21
1.6.4	Electron microscopy	25
1.6.5	X-ray crystallography	29
1.6.6	X-ray microscopy	32
1.7	The human visual system	33
1.7.1	Structure of the eye and early stage visual system	33
1.7.2	Organisation of the visual system	35
1.7.3	Efficiency, redundant information, and natural images	36
2	Effects of Fourier Amplitude and Phase Errors on Image Reconstruction. I: Small Amplitude Errors	39
2.1	Introduction	39
2.2	Theory	40
2.3	Effects of large amplitude errors	47
2.4	Results	48
2.5	Discussion	56
3	Effects of Fourier Amplitude and Phase Errors on Image Reconstruction. II: Large Amplitude Errors	57
3.1	Introduction	57
3.2	A general error metric	58
3.3	Saturation	61
3.4	Results	71
3.5	Discussion	80
4	Image Fourier amplitude distributions	81
4.1	Introduction	81
4.2	Spectral amplitude distributions	82

4.3	Results	85
4.4	Discussion	90
5	Phase dominance	93
5.1	Introduction	93
5.2	Examples of phase dominance	93
5.3	Theory	99
5.4	Phase dominance counter-examples	101
5.4.1	Counter-example of Juvells <i>et al.</i>	101
5.4.2	Counter-examples of Lohmann <i>et al.</i>	107
5.5	Discussion	115
6	Power spectra of natural images	117
6.1	Introduction	117
6.2	Modelling natural images with synthetic images	121
6.3	Simulation results	124
6.4	Circularly averaged power spectra of nonoccluding models	129
6.4.1	Power law size distribution	135
6.4.2	Exponential size distribution	137
6.5	Comparison with Ruderman's results	139
6.6	Discussion	141
7	Conclusions and suggestions for future research	143
A	Relevant mathematics and probability	145
A.1	Fourier transform	145
A.1.1	Properties of Fourier transform	147
A.2	Probability	148
A.2.1	Probability theory	148

A.2.2	Conditional probability and Bayes' theorem	148
A.2.3	Random Variables	149
A.2.4	Expected value and Variance	150
A.2.5	PDF of a function of a random variable	150
A.2.6	Joint probability density function	150
A.2.7	Density function for the sum of 2 random variables	151
A.2.8	Covariance and Correlation	151
A.2.9	Gaussian distribution	152
A.2.10	Rayleigh distribution	152
A.2.11	Central limit theorem	152
A.2.12	Maximum likelihood and Maximum <i>a posteriori</i> estimation	153
A.2.13	Minimum mean squared error estimate	154
A.3	Special Functions	154
A.3.1	Bessel functions of the first kind	154
A.3.2	Complete elliptic integrals	154
A.3.3	Gamma function	155
A.3.4	Hypergeometric function	155
A.3.5	Lommel's function	155
A.3.6	Probability integral	156
A.3.7	Sinc function	156
A.3.8	Jinc function	156

Preface

This thesis is concerned with relationships between an image and its spatial Fourier transform. The Fourier transform of a (real) image is complex, and the focus is on properties of the amplitude (magnitude) and phase of the transform. A majority of the thesis is concerned with the effects that errors in the amplitude and phase of the transform have on images that are reconstructed by inverse Fourier transformation. The remainder concerns related properties of the amplitude of the transform.

The primary practical relevance of this work is to Fourier imaging; i.e., the use of Fourier transforms in image processing and image reconstruction. This arises, in particular, in diffraction imaging where one needs to reconstruct an image from partial and noisy Fourier data. Diffraction imaging covers a wide range of imaging systems from optical astronomy to electron microscopy. It also arises in optical processing systems.

The thesis covers four main topics: (1) The effects of Fourier amplitude and phase errors on image reconstruction (Chapters 2 and 3); (2) Image Fourier amplitude distributions (Chapter 4); (3) Phase dominance (Chapter 5); and (4) Power spectra of natural images (Chapter 6). Chapter 1 contains background information, and original work is reported in Chapters 2 to 6 (aside from the introductory material at the beginning of each chapter). The content of each chapter is briefly outlined below.

In the first chapter, relevant background information on imaging, diffraction, Fourier optics, and diffraction imaging is reviewed. Chapter 6 of the thesis is concerned with Fourier amplitude properties of natural images and implications for visual processing. Therefore, a brief review of the human visual system is included in the final section of Chapter 1. Background information specific to the topic of each chapter is also included at the beginning of the chapters.

In Chapter 2, the expected mean square error (mse) in an image that is reconstructed from its Fourier transform for which the amplitude and phase are individually subjected to errors is studied. Both theoretical analysis and simulations are conducted. Expressions are derived for the relationship between the mean square error in the image domain and am-

plitude and phase errors in the Fourier domain. Simulations are used to illustrate these relationships, and the differing effects of amplitude and phase errors on image characteristics are examined. It is explained why the analysis in this chapter is restricted to small amplitude errors.

In Chapter 3, the work in Chapter 2 is extended to the case of large amplitude errors. The mse metric is modified to account for the additional energy introduced by the noise, and the effects of the saturation are incorporated. The effects of large amplitude errors are evaluated as is the dependance on the Fourier amplitude and error distributions.

The distribution of image Fourier amplitudes is shown to be of particular significance in Chapters 3 and 5, and in Chapter 4 a density function for image Fourier amplitudes is derived based on a simple model of images. The density function is shown to well-represent that of a variety of images.

The phenomenon of phase dominance is studied in Chapter 5. The expected image mse is derived for the case of exchanging the Fourier amplitude or phase between two images. This provides a rigorous and quantitative explanation for phase dominance. Reported counter-examples of phase dominance are studied and are shown to be erroneous.

The basis of the observed power law behaviour of the power spectrum (squared Fourier amplitude) of natural images with spatial frequency is studied in Chapter 6. Using an image model consisting of independent objects of different sizes shows that the presence of edges and a self-similar object size distribution are the essential ingredients that lead to the observed power spectral behaviour.

Brief conclusions and suggestions for future research are given in Chapter 7.

This research has resulted several journal and conference papers, and they are listed below in chronological order:

W. H. Hsiao and R. P. Millane, "Effects of Fourier amplitude and phase errors on image reconstruction," in Proceedings of the Image & Vision Computing New Zealand 2002 (IVCNZ'02), D. N. Kenwright, ed., Auckland, New Zealand, November 2002, pp. 221-226.

R. P. Millane and W. H. Hsiao, "On apparent counterexamples to phase dominance," J. Opt. Soc. Am. A, Vol. 20, no. 4, pp. 753-756 (2003).

R. P. Millane and W. H. Hsiao, "Effects of amplitude and phase errors on perceived images," International Workshop on Noncrystallographic Phase Retrieval, June 2003, Cairns, Australia.

R. P. Millane, S. Alzaidi and W. H. Hsiao, "Scaling and power spectra of natural images," in Proceedings of the Image & Vision Computing New Zealand 2003 (IVCNZ'03), D. Bailey, ed., Palmerston North, New Zealand, November 2003, pp. 148-153.

W. H. Hsiao, E. Ip and R. P. Millane, "Effects of Fourier amplitude and phase errors on interpretability of images," in Proceedings of the Image & Vision Computing New Zealand 2003 (IVCNZ'03), D. Bailey, ed., Palmerston North, New Zealand, November 2003, pp. 175-180.

W. H. Hsiao and R. P. Millane, "Effects of Fourier amplitude and phase errors on image reconstruction," in Proceedings of SPIE, Volume 5562: Image Reconstruction from incomplete data III, P. J. Bones, M. A. Fiddy and R. P. Millane, eds., Denver, Colorado, USA, August 2004, pp. 27-37.

W. H. Hsiao and R. P. Millane, "Effects of occlusion, edges and scaling on the power spectra of natural images," *J. Opt. Soc. Am. A*, Vol. 22, no. 9, pp. 1789-1797 (2005).

R. P. Millane and W. H. Hsiao, "Natural images and visual processing," *Optics & Photonics News*, vol. 16, no. 12, pp. 24 (2005).

W. H. Hsiao and R. P. Millane, "Image spectral amplitude distributions," *J. Opt. Soc. Am. A*, Vol. 23, no. 8, pp. 1823-1826 (2006).

W. H. Hsiao and R. P. Millane, "Effects of Fourier-plane amplitude and phase errors on image reconstruction. I. Small amplitude errors," *J. Opt. Soc. Am. A*, Vol. 24, no. 10, pp. 3180-3188 (2007).

W. H. Hsiao and R. P. Millane, "Effects of Fourier-plane amplitude and phase errors on image reconstruction. II. Large amplitude errors," *J. Opt. Soc. Am. A*, submitted.

Chapter 1

Introduction

1.1 Imaging

An image, in the broad sense, can be considered as a (usually spatial) distribution of a physical parameter that provides information about a physical object. *Imaging* refers to the physical process involved in acquiring images, which may be one-, two- or three dimensional. Images are often produced using visible light and optical devices such as lenses, cameras and mirrors. Examples are images formed by our eyes, optical microscopes, telescopes, etc.. However, many imaging processes do not use visible light. Most imaging processes involve “remote sensing,” i.e., data is collected at some distance from the object. In almost all cases, information is transmitted from the object to the measurement location by a propagating field. Often this is an electromagnetic field, but it might also be an acoustic field, elastodynamic field, etc.. Microscopy, optical imaging, magnetic resonance imaging, electron microscopy, X-ray crystallography, neutron diffraction, synthetic aperture radar, etc., all make use of a part of the electromagnetic spectrum. Medical ultrasound, seismology, non destructive testing of materials, etc., all make use of propagation of an elastodynamic field.

In general, therefore, imaging involves the measurement of a field diffracted or scattered from an object, and processing of this data to estimate the distribution of a physical parameter that characterises the object.

1.2 Electromagnetic radiation and the scalar approximation

Electromagnetic radiation is a propagating wave with both electric and magnetic components and is generally classified by wavelength into radio waves, microwaves, infrared

(heat), light, ultraviolet, X-rays and gamma rays. Light generally refers to electromagnetic radiation with a wavelength ranging from 380 nm to 770 nm and which is visible to the human eye [1, 2]. The speed of propagation of electromagnetic radiation in free space is a constant at $\approx 3.0 \times 10^8 \text{ m/s}$ regardless of the wavelength or frequency, which means that the higher frequencies have shorter wavelengths [3]. The best theory to explain electromagnetic radiation accurately is the theory of quantum electrodynamics (QED). However, for the purposes of the applications considered in this thesis, quantum effects can be ignored and the electromagnetic field is described using classical electromagnetics.

In a linear, isotropic, dielectric, homogeneous, non-dispersive and time-invariant medium, Maxwell's equations are given by [4]

$$\begin{aligned} \nabla \cdot \epsilon \mathcal{E} &= 0 \\ \nabla \cdot \mu \mathcal{H} &= 0 \\ \nabla \times \mathcal{E} &= -\mu \frac{\partial \mathcal{H}}{\partial t} \\ \nabla \times \mathcal{H} &= \epsilon \frac{\partial \mathcal{E}}{\partial t}, \end{aligned} \quad (1.1)$$

where $\mathcal{E} = \mathcal{E}(\mathbf{x})$ and $\mathcal{H} = \mathcal{H}(\mathbf{x})$ are the electric field intensity and magnetic field intensity, respectively, and $\mathbf{x} = (x, y, z)$ is position. In a three-dimensional Cartesian coordinate system, both \mathcal{E} and \mathcal{H} have three rectilinear components, $(\mathcal{E}_x, \mathcal{E}_y, \mathcal{E}_z)$ and $(\mathcal{H}_x, \mathcal{H}_y, \mathcal{H}_z)$. μ and ϵ are the permeability and permittivity, respectively, of the propagation medium, and ∇ is the vector differential operator

$$\nabla = \frac{\partial}{\partial x} \hat{i} + \frac{\partial}{\partial y} \hat{j} + \frac{\partial}{\partial z} \hat{k}, \quad (1.2)$$

where \hat{i} , \hat{j} and \hat{k} are unit vectors in the x , y and z directions, respectively. Applying the curl $\nabla \times$ operation to both sides of the last two equations in Eq. (1.1) gives,

$$\begin{aligned} \nabla \times (\nabla \times \mathcal{E}) &= -\mu \frac{\partial}{\partial t} (\nabla \times \mathcal{H}) \\ \nabla \times (\nabla \times \mathcal{H}) &= \epsilon \frac{\partial}{\partial t} (\nabla \times \mathcal{E}). \end{aligned} \quad (1.3)$$

Substituting the last two equations of Eq. (1.1) into the right hand-side of Eq. (1.3) and applying the vector identity

$$A \times (B \times C) = B(A \cdot C) - C(A \cdot B) \quad (1.4)$$

to the left hand-side of Eq. (1.3) gives

$$\begin{aligned}\nabla(\nabla \cdot \mathcal{E}) - \nabla^2 \mathcal{E} &= -\mu\epsilon \frac{\partial^2 \mathcal{E}}{\partial t^2} \\ \nabla(\nabla \cdot \mathcal{H}) - \nabla^2 \mathcal{H} &= -\mu\epsilon \frac{\partial^2 \mathcal{H}}{\partial t^2},\end{aligned}\quad (1.5)$$

where ∇^2 is the Laplacian operator

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (1.6)$$

For a homogeneous medium Eq. (1.1) shows that $\nabla \cdot \mathcal{E} = 0$ and $\nabla \cdot \mathcal{H} = 0$, which yields

$$\begin{aligned}\nabla^2 \mathcal{E} - \frac{n^2}{c^2} \frac{\partial^2 \mathcal{E}}{\partial t^2} &= 0 \\ \nabla^2 \mathcal{H} - \frac{n^2}{c^2} \frac{\partial^2 \mathcal{H}}{\partial t^2} &= 0,\end{aligned}\quad (1.7)$$

where c is the speed of propagation in vacuum defined as

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}, \quad (1.8)$$

ϵ_0 and μ_0 are the vacuum permittivity and permeability, respectively, and n is the refractive index of the medium

$$n = \sqrt{\epsilon_r \mu_r}, \quad (1.9)$$

where $\epsilon_r = \epsilon/\epsilon_0$ and $\mu_r = \mu/\mu_0$ are the relative permittivity and permeability, respectively. The speed of propagation is $v = c/n$. Equations (1.7) are known as the electromagnetic vector wave equations.

The vector components of \mathcal{E} and \mathcal{H} in Eq. (1.7) are uncoupled so that each component satisfies a scalar wave equation. Representing one of the components by a scalar variable $u(\mathbf{x}, t)$, the scalar wave equation is then

$$\nabla^2 u(\mathbf{x}, t) - \frac{n^2}{c^2} \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} = 0. \quad (1.10)$$

The wave equation Eq. (1.7) is for a homogeneous medium. In all imaging applications however, the object represents a region of space where the medium is spatially variable. However, it can be shown that if the spatial variation of the medium properties is slower than that of \mathcal{E} and \mathcal{H} , i.e., if the medium properties varying on a scale $\Delta \gg \lambda$, where λ is the wavelength, then each component of \mathcal{E} and \mathcal{H} satisfies a scalar wave equation of the

form

$$\nabla^2 u(\mathbf{x}, t) - \frac{n^2(\mathbf{x})}{c^2} \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} = 0. \quad (1.11)$$

If the medium contains sharp boundaries (where $n(\mathbf{x})$ changes rapidly) then additional techniques are required.

1.3 Scalar diffraction theory

The electromagnetic field $u(\mathbf{x}, t)$ of a monochromatic wave has the form [4]

$$u(\mathbf{x}, t) = \text{Re}\{U(\mathbf{x}) \exp(-i2\pi ft)\}, \quad (1.12)$$

where f is frequency, $\text{Re}\{\cdot\}$ denotes the real part, and $U(\mathbf{x})$ is the complex wave function. $U(\mathbf{x})$ can be written as

$$U(\mathbf{x}) = A(\mathbf{x}) \exp(i\phi(\mathbf{x})), \quad (1.13)$$

where $A(\mathbf{x})$ and $\phi(\mathbf{x})$ are the magnitude and phase, respectively, of the wave function. The wave number k is defined by $k = 2\pi f/v = 2\pi/\lambda$. For an inhomogeneous medium $k = k(\mathbf{x})$ is spatially varying as

$$k(\mathbf{x}) = k_0 n(\mathbf{x}), \quad (1.14)$$

where $k_0 = 2\pi f/c$ is the free space wave number. The frequency dependence of $U(\mathbf{x})$ is often denoted by $U(\mathbf{x}, k)$. $U(\mathbf{x}, k)$ is the temporal Fourier transform of $u(\mathbf{x}, t)$, i.e.,

$$U(\mathbf{x}, k) = \int_{-\infty}^{\infty} u(\mathbf{x}, t) \exp(i2\pi kt) dt. \quad (1.15)$$

Substituting Eq. (1.12) into Eq. (1.10) gives the time-independent wave equation

$$(\nabla^2 + k^2(\mathbf{x}))U(\mathbf{x}, k) = (\nabla^2 + k_0^2 n^2(\mathbf{x}))U(\mathbf{x}, k) = 0, \quad (1.16)$$

which is known as the Helmholtz equation.

Calculation of the complex wavefield U at a point \mathbf{x} in space, given an incident field and the material properties of the medium is referred to as a diffraction problem. A simple diffraction problem is the propagation of a plane wave through a hole (an aperture) in an opaque screen as illustrated in Fig. 1.1. This setup is common to many diffraction problems and image formation. A plane wave $U_i(\xi, \eta)$ incident on an aperture A in the (ξ, η) -plane propagates to the measurement plane (x, y) , which is parallel to the aperture plane, and

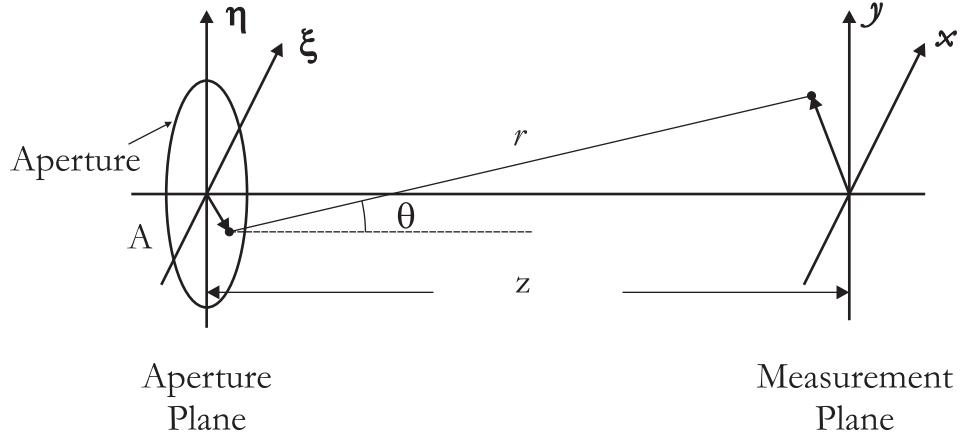


Figure 1.1: Diffraction by an aperture: the optical field $U(x, y)$ in the measurement plane due to the incident field $U_i(\xi, \eta)$ in the aperture plane.

the distance between the two planes is z . The wave function on the measurement plane is given by the Rayleigh-Sommerfeld diffraction integral [4],

$$U(x, y) = \frac{1}{i\lambda} \iint_A U_i(\xi, \eta) \frac{\exp(ikr)}{r} \cos \theta d\xi d\eta, \quad (1.17)$$

where r and θ are as shown in Fig. 1.1. Equation (1.17) can also be expressed as

$$U(x, y) = \iint_A U_i(\xi, \eta) h(x, y; \xi, \eta) d\xi d\eta, \quad (1.18)$$

where the point spread function

$$h(\xi, \eta; x, y) = \frac{1}{i\lambda} \frac{\exp(ikr)}{r} \cos \theta. \quad (1.19)$$

Equation (1.17) is a mathematical expression for the Huygens-Fresnel principle which states that every point on a propagating wavefront serves as a point source of spherical secondary wavelets, multiplied by $\cos \theta / i\lambda$ [4], i.e., the field on the measurement plane is the superposition of the spherical waves, $\exp(ikr)/r$ [5], produced by an array of point sources in the aperture (Fig. 1.2). Furthermore, Eq. (1.18) can be expressed as a convolution integral, writing Eq. (1.19) as (noting that $\cos \theta = z/r$)

$$h(x, y; \xi, \eta) = \frac{z}{i\lambda((x - \xi)^2 + (y - \eta)^2 + z^2)} \exp\left(ik\sqrt{(x - \xi)^2 + (y - \eta)^2 + z^2}\right). \quad (1.20)$$

Equation (1.17) can then be written as

$$U(x, y) = \iint_{-\infty}^{\infty} U_i(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta = U_i(x, y) \odot h(x, y), \quad (1.21)$$

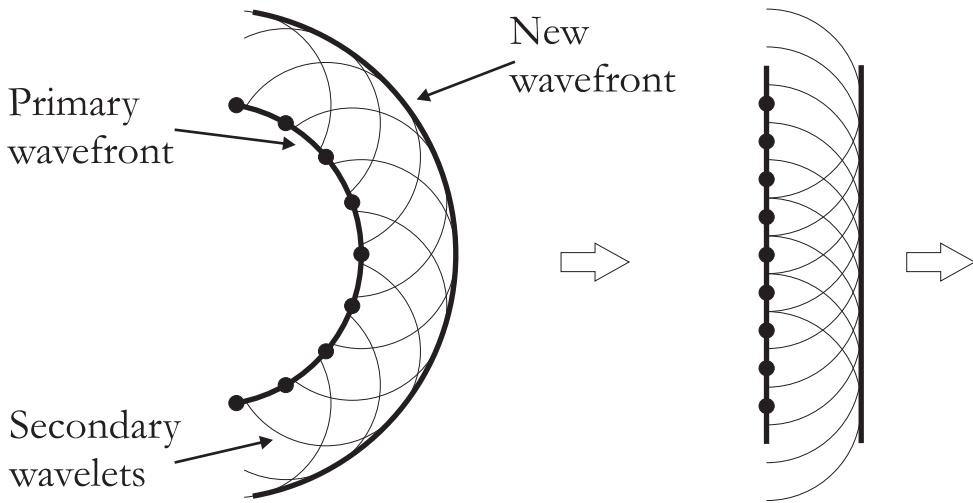


Figure 1.2: Graphical illustration of Huygens-Fresnel principle (left) a spherical wavefront and (right) a planar wavefront.

where \odot denotes convolution and

$$h(x, y) = \frac{z}{i\lambda(x^2 + y^2 + z^2)} \exp(ik\sqrt{x^2 + y^2 + z^2}), \quad (1.22)$$

which is known as the Huygens point spread function (psf) [6].

The Rayleigh Sommerfeld diffraction formula can be simplified if the measurement plane is far from the aperture. Two useful approximations are the Fresnel and Fraunhofer diffraction formulae.

1.3.1 Fresnel diffraction

In Fresnel diffraction, the measurement plane is far enough from the aperture plane so that $\cos \theta$ and r can be replaced with $\cos \theta \approx 1$ and $r \approx z$, and Eq. (1.19) becomes

$$h(x, y; \xi, \eta) = \frac{1}{i\lambda z} \exp(ikr). \quad (1.23)$$

Note that r in the exponential is still in place as a small approximation error in r can lead to a large phase error after multiplying a large k [4]. The r in the exponential is approximated with first two terms of the binomial expansion

$$r = z \left[1 + \frac{1}{2} \left(\frac{x - \xi}{z} \right)^2 + \frac{1}{2} \left(\frac{y - \eta}{z} \right)^2 \right]. \quad (1.24)$$

The psf $h(x, y; \xi, \eta)$ for Fresnel diffraction is then given by

$$h_{\text{Fres}}(x, y; \xi, \eta) = \frac{\exp(ikz)}{i\lambda z} \exp\left(i\frac{k}{2z} [(x - \xi)^2 + (y - \eta)^2]\right) \quad (1.25)$$

and the optical field for Fresnel diffraction is

$$\begin{aligned} U(x, y) &= \frac{\exp(ikz)}{i\lambda z} \exp\left(i\frac{k}{2z}(x^2 + y^2)\right) \\ &\quad \times \iint_{-\infty}^{\infty} U_i(\xi, \eta) \exp\left(i\frac{k}{2z}(\xi^2 + \eta^2)\right) \exp\left(-i\frac{2\pi}{\lambda z}(\xi x + \eta y)\right) d\xi d\eta \\ &= \frac{\exp(ikz)}{i\lambda z} \exp\left(i\frac{k}{2z}(x^2 + y^2)\right) \mathcal{F}\left\{U_i(\xi, \eta) \exp\left(i\frac{k}{2z}(\xi^2 + \eta^2)\right)\right\} \Big|_{(x/\lambda z, y/\lambda z)}. \end{aligned} \quad (1.26)$$

The Fresnel pointspread function is

$$h_{\text{Fres}}(x, y) = \frac{\exp(ikz)}{i\lambda z} \exp\left(i\frac{k}{2z}[x^2 + y^2]\right) \quad (1.27)$$

and

$$U(x, y) = U_i(x, y) \odot h_{\text{Fres}}(x, y). \quad (1.28)$$

In the Fresnel approximation, the spherical phase term is replaced by the quadratic phase $\exp((ik/2z)(\xi^2 + \eta^2))$. Fresnel diffraction can be described in terms of a Fourier transform (Eq. (1.26)) or a convolution (Eq. (1.28)).

1.3.2 Fraunhofer diffraction

If the measurement plane is further from the aperture plane, i.e., in the *far field*, then

$$z \gg \frac{k(\xi^2 + \eta^2)_{\max}}{2}, \quad (1.29)$$

and the spherical wavefront of the Huygens-Fresnel integral is replaced by a planar wavefront. The quadratic phase term in Eq. (1.26) is replaced by unity, giving

$$U(x, y) = \frac{\exp(ikz)}{i\lambda z} \exp\left(i\frac{k}{2z}(x^2 + y^2)\right) \iint_{-\infty}^{\infty} U_i(\xi, \eta) \exp\left(-i\frac{2\pi}{\lambda z}(\xi x + \eta y)\right) d\xi d\eta. \quad (1.30)$$

Replacing the coordinate system (x, y) with (u, v) where $u = x/\lambda z$ and $v = y/\lambda z$ gives

$$U(u, v) = \frac{\exp(ikz)}{i\lambda z} \exp(i\pi\lambda z(u^2 + v^2)) \mathcal{F}\{U_i(\xi, \eta)\}(u, v), \quad (1.31)$$

which shows that $U(u, v)$ on the image plane is the Fourier transform of $U_i(\xi, \eta)$ on the aperture plane and multiplied by a quadratic phase factor.

1.4 Geometric optics

Geometric optics further simplifies the propagation behaviour of electromagnetic radiation by assuming that the wavelength is close to zero. In this approximation, light no longer behaves like a wave and in a homogeneous medium propagates in a rectilinear fashion as rays. The deflection of rays at a plane boundary between two media of different refractive index are described as the familiar laws of reflection and refraction.

A lens is a refracting device that modifies the direction of propagation of rays. A bispherical lens is a region of constant refractive index n bounded by two spherical surfaces of radius R_1 and R_2 . Parallel rays are focused at a distance f from the lens, referred to as the focal length of the lens, which is given by

$$\frac{1}{f} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right). \quad (1.32)$$

1.5 Fourier optics and imaging

In certain approximations, optical systems can be treated as linear, time invariant systems, and can be analysed using Fourier techniques. This is referred to as “Fourier optics.” Two simple setups with a single lens are described below, followed by the generalised imaging system model.

1.5.1 Fourier transforming property of a lens

Consider an ideal thin lens in the x - y plane. A plane wave propagating along the z direction suffers an additional phase shift

$$\phi_l = -\frac{k}{2f}(x^2 + y^2) \quad (1.33)$$

on propagating through the lens. The modification of the field is described by the “transmission factor” $t_l(x, y)$ given by

$$t_l(x, y) = \exp\left(-i\frac{k}{2f}(x^2 + y^2)\right). \quad (1.34)$$

Consider now a lens in an aperture with a plane wave incident on the aperture, as shown in Fig. 1.3. The input (illuminating source) $U_i(\xi, \eta)$ is placed at a distance f in front of the lens and the measurement plane is located at the same distance behind the lens. The optical field $U(x, y)$ on the measurement plane can be expressed in terms of $U'_l(\xi, \eta)$, the field on the right hand side of the lens, using the Fresnel approximation Eq. (1.26) (assuming Fresnel conditions are satisfied), noting that $U'_l(\xi, \eta) = t_l(\xi, \eta)U_l(\xi, \eta)$, and using Eq. (1.34) to give

$$U(u, v) = \frac{\exp(ikf)}{i\lambda f} \exp(i\pi\lambda f(u^2 + v^2)) \mathcal{F}\{U_l(\xi, \eta)\}(u, v), \quad (1.35)$$

i.e., the phase shift due to the lens cancels out that in the Fresnel diffraction formula. Equation (1.35) shows that the field at the focal distance is the Fraunhofer diffraction of the field incident on the lens.

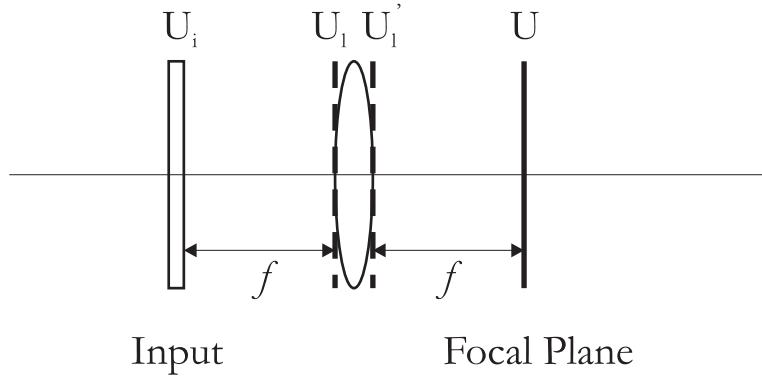


Figure 1.3: Demonstrating lens’ ability to perform the two dimensional Fourier transform.

The field incident on the lens can be expressed using the convolution (Eq. (1.28)), i.e.,

$$U_l(x, y) = U_i(x, y) \odot h_{\text{Fres}}(x, y)|_{z=f}. \quad (1.36)$$

The Fourier transform of $U_l(x, y)$ can be expressed as

$$\begin{aligned} \mathcal{F}\{U_l(x, y)\} &= \mathcal{F}\{U_i(x, y) \odot h_{\text{Fres}}(x, y)|_{z=f}\} \\ &= \mathcal{F}\{U_i(x, y)\} \times \mathcal{F}\{h_{\text{Fres}}(x, y)|_{z=f}\}, \end{aligned} \quad (1.37)$$

where

$$\begin{aligned}\mathcal{F}\{h|_{\text{Fres}}(x, y)|_{z=f}\} &= \frac{\exp(ikf)}{i\lambda f} \mathcal{F}\left\{\exp\left(i\frac{\pi}{\lambda f}[x^2 + y^2]\right)\right\} \\ &= \exp(ikf) \exp(-i\pi\lambda f(u^2 + v^2)).\end{aligned}\quad (1.38)$$

Substituting Eqs. (1.37) and (1.38) into Eq. (1.35) gives

$$U(u, v) = \frac{\exp(i2kf)}{i\lambda f} \mathcal{F}\{U_i(x, y)\}, \quad (1.39)$$

i.e., the field in the focal plane is the exact Fourier transform of $U_i(x, y)$ with a neglectable constant term.

1.5.2 Image forming property of a lens

The second setup shows image formation by a single lens. Consider an input field a distance d_i in front of the lens and a measurement (or output) plane a distance d_o behind the lens as shown in Fig. 1.4. The optical field $U_o(x, y)$ in the measurement plane can be calculated by applying the Fresnel diffraction (Eq. (1.26)) twice for light propagating from $U_i(x, y)$ to $U_l(x, y)$ and from $U_{l'}(x, y)$ to $U(x, y)$, and lens transfer function (from $U_l(x, y)$ to $U_{l'}(x, y)$) (Eq. (1.34)) with appropriate parameters as

$$U_o(x, y) = \left[\left(U_i(x, y) \odot h_{\text{Fres}}(x, y)|_{z=d_i} \right) \times \exp\left(-i\frac{k}{2f}(x^2 + y^2)\right) \right] \odot h_{\text{Fres}}(x, y)|_{z=d_o}, \quad (1.40)$$

or in terms of the Fourier transform, i.e., using Eq. (1.26), as

$$\begin{aligned}U_o(x, y) &= \frac{\exp(ikd_i)}{i\lambda d_i} \frac{\exp(ikd_o)}{i\lambda d_o} \exp\left(i\frac{k(x^2 + y^2)}{2d_o}\right) \\ &\quad \times \mathcal{F}\left\{\exp\left(\frac{ik}{2}(\xi'^2 + \eta'^2)\left(\frac{1}{d_i} + \frac{1}{d_o} - \frac{1}{f}\right)\right)\right\} \\ &\quad \times \mathcal{F}\left\{U_i(\xi, \eta) \exp\left(i\frac{k}{2d_i}(\xi^2 + \eta^2)\right)\right\} \Big|_{\left(\frac{\xi'}{\lambda d_i}, \frac{\eta'}{\lambda d_i}\right)} \Big|_{\left(\frac{x}{\lambda d_o}, \frac{y}{\lambda d_o}\right)}. \quad (1.41)\end{aligned}$$

Assume now that d_i and d_o satisfy the lens equation

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}. \quad (1.42)$$

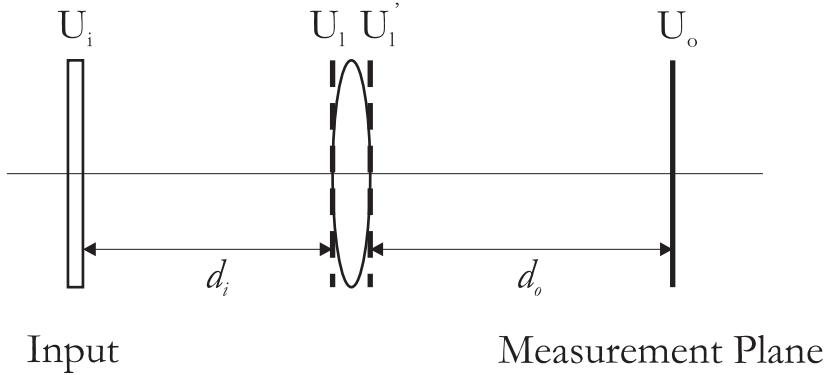


Figure 1.4: Image formation geometry of a lens.

Equation (1.41) becomes

$$U_o(x, y) = \frac{\exp(ikd_i)}{i\lambda d_i} \frac{\exp(ikd_o)}{i\lambda d_o} \exp\left(i\frac{k(x^2 + y^2)}{2d_o}\right) \times \mathcal{F} \left\{ \mathcal{F} \left\{ U_i(\xi, \eta) \exp\left(i\frac{k}{2d_i}(\xi^2 + \eta^2)\right) \right\} \Big|_{\left(\frac{\xi'}{\lambda d_i}, \frac{\eta'}{\lambda d_i}\right)} \right\} \Big|_{\left(\frac{x}{\lambda d_o}, \frac{y}{\lambda d_o}\right)}. \quad (1.43)$$

Using the identity $\mathcal{F}\{\mathcal{F}\{f(x)\}\} = f(-x)$ and the similarity theorem of the Fourier transform (Eq. (A.14) in Appendix A.1.1), the double Fourier transform in Eq. (1.43) reduce to

$$\begin{aligned} & \mathcal{F} \left\{ \mathcal{F} \left\{ U_i(\xi, \eta) \exp\left(i\frac{k}{2d_i}(\xi^2 + \eta^2)\right) \right\} \Big|_{\left(\frac{\xi'}{\lambda d_i}, \frac{\eta'}{\lambda d_i}\right)} \right\} \\ &= \lambda^2 d_i^2 U_i(-\lambda d_i \xi, -\lambda d_i \eta) \exp(i\pi \lambda d_i (\xi^2 + \eta^2)). \end{aligned} \quad (1.44)$$

Substituting Eq. (1.44) into Eq. (1.43) gives

$$U_o(x, y) = -\frac{d_i}{d_o} \exp(ik(d_i + d_o)) \exp\left(i\frac{k(1+d_i)(x^2 + y^2)}{2d_o}\right) U_i\left(-\frac{d_i}{d_o}x, -\frac{d_i}{d_o}y\right). \quad (1.45)$$

Equation (1.45) shows that $U_o(x, y)$ is a phase shifted, scaled and attenuated version of $U_i(x, y)$. Defining the magnification factor $M = -d_o/d_i$, aside from a phase factor,

$$U_o(x, y) = \frac{1}{M} U_i\left(\frac{x}{M}, \frac{y}{M}\right). \quad (1.46)$$

The image intensity $I(x, y)$ is given by

$$I(x, y) = \frac{1}{M^2} |U_i\left(\frac{x}{M}, \frac{y}{M}\right)|^2. \quad (1.47)$$

Equation (1.46) corresponds to the image predicted by geometric optics, which is denoted

by $U_g(x, y)$, i.e.,

$$U_g(x, y) = U_i \left(\frac{x}{M}, \frac{y}{M} \right). \quad (1.48)$$

1.5.3 Generalised imaging system model

The analysis in the previous two sections shows that the propagation of the field and the effect of optical elements (such as lenses) can be described in terms of Fourier transforms, or linear system theory. The resulting image obtained, for example, is that predicted by geometrical optics, i.e., the effects of diffraction are not incorporated. This is because it was assumed, for example, that all of the propagated field is captured by the lens. In practice, only the low-angle field is captured by the lens and the high-angle information is lost. This effect is included by incorporating “apertures” or “pupils” to model the elimination of the high-angle diffracted fields. Such a model that includes the diffraction effects introduced by apertures is called a *generalised imaging system model* and is often a satisfactory description of such systems. It is common to group the optical components and propagation into one imaging system as shown in Fig. 1.5. A general imaging system contains entrance and exit pupils and it is assumed that geometric optics is sufficient to estimate the field propagation within the system, and all the diffraction effects are associated with either the path from the input to the entrance pupil or from the exit pupil to the output.

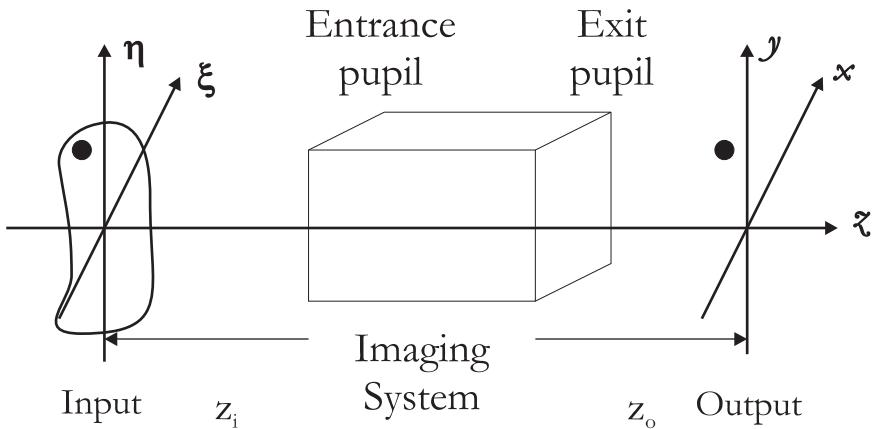


Figure 1.5: Generalised imaging system model.

Consider a generalised imaging system as shown Fig. 1.5, and define the (exit) pupil function $P(x, y)$ as

$$\begin{aligned} P(x, y) &= 1, && \text{inside the lens aperture} \\ &= 0, && \text{otherwise,} \end{aligned} \quad (1.49)$$

then the relationship between the input and output can be expressed as

$$U_o(x, y) = \iint_{-\infty}^{\infty} h(x, y; \xi, \eta) U_i(\xi, \eta) d\xi d\eta, \quad (1.50)$$

where $h(x, y; \xi, \eta)$ is the imaging system response, including the pupils. For the simplest imaging system containing a single lens and assuming that there are no aberrations, the imaging system response is

$$h(x, y; \xi, \eta) = \frac{A}{\lambda z_o} \iint_{-\infty}^{\infty} P(u, v) \exp \left(-i \frac{2\pi}{\lambda z_o} [(x - M\xi)u + (y - M\eta)v] \right) du dv, \quad (1.51)$$

where A is a constant amplitude, and z_o is the distance from the exit pupil to the image plane (Fig. 1.5). Replacing $M\xi$ and $M\eta$ with ξ' and η' , respectively, Eq. (1.51) becomes

$$h(x - \xi', y - \eta') = \frac{A}{\lambda z_o} \iint_{-\infty}^{\infty} P(u, v) \exp \left(-i \frac{2\pi}{\lambda z_o} [(x - \xi')u + (y - \eta')v] \right) du dv. \quad (1.52)$$

Using Eq. (1.46), Eq. (1.50) can be expressed as

$$\begin{aligned} U_o(x, y) &= \iint_{-\infty}^{\infty} h(x - \xi', y - \eta') U_g(\xi', \eta') d\xi' d\eta' \\ &= h(x, y) \odot U_g(x, y), \end{aligned} \quad (1.53)$$

where

$$\begin{aligned} h(x, y) &= \frac{A}{\lambda z_o} \iint_{-\infty}^{\infty} P(u, v) \exp \left(-i \frac{2\pi}{\lambda z_o} (ux + vy) \right) du dv \\ &= \frac{A}{\lambda z_o} \mathcal{F}\{P(u, v)\}(\lambda z_o x, \lambda z_o y), \end{aligned} \quad (1.54)$$

i.e., the output image $U_o(x, y)$ is the convolution of the geometric optics image with the impulse response of the imaging system. This imaging model is often referred to as diffraction limited, i.e., the imaging system is perfect and the resolution is limited only by diffraction. In practice, higher order aberrations (e.g., due to imperfect lenses) are present and are normally included by replacing the pupil function with a generalised pupil function that includes these effects. The coherent transfer function (CTF) is related to the (generalised) pupil function which is described in detail in the next section.

1.5.4 Coherent and incoherent imaging

Coherence is an important concept in wave propagation, optics, imaging, and is concerned with the correlation between field measurements at different locations and times. The correlation between measurements of a field $f(\mathbf{x}, t)$ at position \mathbf{x}_1 and \mathbf{x}_2 and at times separated by τ is called the mutual coherence function and is given by

$$\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \langle f(\mathbf{x}_1, t) f^*(\mathbf{x}_2, t + \tau) \rangle, \quad (1.55)$$

where $\langle \cdot \rangle$ denotes the mean value (Section A.2.4). The intensity, averaged over time, of the field at \mathbf{x} is

$$\begin{aligned} I(\mathbf{x}) &= \langle f(\mathbf{x}, t) f^*(\mathbf{x}, t) \rangle \\ &= \Gamma(\mathbf{x}, \mathbf{x}, 0). \end{aligned} \quad (1.56)$$

The normalised mutual coherence function

$$\gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \frac{\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)}{\sqrt{I(\mathbf{x}_1)I(\mathbf{x}_2)}}. \quad (1.57)$$

is called the complex degree of coherence. Coherence can be separated into its spatial and temporal components. Two commonly used coefficients are the complex degree of temporal and spatial coherence, respectively, given by

$$\gamma_{11}(\mathbf{x}, \tau) = \gamma(\mathbf{x}, \mathbf{x}, \tau) \quad (1.58)$$

and

$$\gamma_{12}(\mathbf{x}_1, \mathbf{x}_2) = \gamma(\mathbf{x}_1, \mathbf{x}_2, 0). \quad (1.59)$$

The degree of coherence lies between -1 and 1 . $\gamma = 1$ implies perfect coherence and $\gamma = 0$ implies complete incoherence. $0 < \gamma < 1$ implies partial coherence. $\gamma < 0$ implies that the field measurements are anti-correlated. An example of a highly coherent source (both spatial and temporal) is a laser. Thermal lights and self-luminant objects are incoherent (both spatial and temporal) sources. A point source is spatially coherent.

For a coherent optical system (i.e., the fields are perfectly correlated), the response at each point is the superposition of the complex fields, and the system output is the convolution of the system impulse response with the input, i.e.,

$$U_o(x, y) = U_i(x, y) \odot h(x, y). \quad (1.60)$$

The Fourier transform of Eq. (1.60) is

$$\tilde{U}_o(u, v) = \mathcal{F}\{U_o(x, y)\} = \tilde{U}_i(u, v)H(u, v), \quad (1.61)$$

where the \sim is used here to denote the Fourier transform, and

$$H(u, v) = \mathcal{F}\{h(x, y)\} \quad (1.62)$$

is called the coherent transfer function (CTF). Using Eq. (1.54) and the identity $\mathcal{F}\{\mathcal{F}\{f(x)\}\} = f(-x)$, the CTF is given as

$$\begin{aligned} H(u, v) &= \mathcal{F}\{\mathcal{F}\{P(\lambda z_o u, \lambda z_o v)\}\} \\ &= P(-\lambda z_o u, -\lambda z_o v), \end{aligned} \quad (1.63)$$

which shows that the CTF is the pupil function scaled and inverted in the origin.

Now consider two stationary complex fields $f(\mathbf{x}_1, t)$ and $g(\mathbf{x}_2, t + \tau)$ separated by time τ and distance $\mathbf{x}_1 - \mathbf{x}_2$. The intensity of the sum of the two fields, averaged over time, is given by

$$\begin{aligned} \langle |f(\mathbf{x}_1, t) + f(\mathbf{x}_2, t + \tau)|^2 \rangle &= \langle |f(\mathbf{x}_1, t)|^2 \rangle + \langle |f(\mathbf{x}_2, t + \tau)|^2 \rangle \\ &\quad + \langle f(\mathbf{x}_1, t)f^*(\mathbf{x}_2, t + \tau) \rangle + \langle f^*(\mathbf{x}_1, t)f(\mathbf{x}_2, t + \tau) \rangle \\ &= I(\mathbf{x}_1) + I(\mathbf{x}_2) + 2 \operatorname{Real}\{\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)\}. \end{aligned} \quad (1.64)$$

For an incoherent system (i.e., spatially and temporally fields propagating in the system) the fields are uncorrelated, i.e., $\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = 0$ and Eq. (1.64) shows the system is linear in intensity. Therefore, the output intensity is related to the input intensity by a convolution and, referring to Eq. (1.60), this relationship is given by

$$I_o(x, y) = I_i(x, y) \odot |h(x, y)|^2. \quad (1.65)$$

The Fourier transform of Eq. (1.65) is

$$\tilde{I}_o(u, v) = \mathcal{F}\{I_o(x, y)\} = \tilde{I}_i(u, v)\mathcal{H}(u, v), \quad (1.66)$$

where $\mathcal{H}(u, v) = \mathcal{F}\{|h(x, y)|^2\}$ is called the incoherent transfer function, and is often referred to as the optical transfer function (OTF). Referring to Eq. (1.54) and the autocorrelation theorem for Fourier transforms (Eq. (A.21) in Appendix A.1.1), the OTF is given as

$$\mathcal{H}(u, v) = P(\lambda z_o u, \lambda z_o v) \otimes P(-\lambda z_o u, -\lambda z_o v), \quad (1.67)$$

i.e., the OTF is the autocorrelation of the pupil function. The modulation transfer function (MTF) is defined as the absolute value of the OTF.

1.6 Diffraction imaging

Imaging of an object as described in Section 1.5 relies upon the construction of a physical system (e.g., a lens) to convert the diffracted field into an image of the object. However, there are numerous situations where this is not possible. For example, lenses cannot be constructed for very long (e.g., radio) or very short (e.g., X-ray) wavelengths. This is not problematic in principle, since if the diffracted field can be measured over a sufficient range of spatial frequencies then the image can be calculated numerically by Fourier or Fresnel transformation. This is referred to as *diffraction imaging* and takes advantage of the Fourier relationship between an object and its diffraction pattern. Diffraction imaging includes a wide range of imaging techniques such as astronomy, interferometry, radio astronomy, microwave imaging, synthetic aperture radar, ultrasonic and acoustic imaging, diffraction tomography, and electron and x-ray microscopy and crystallography [7, 8, 9, 10, 11, 12]. Another important imaging modality for which data are obtained in the Fourier domain is magnetic resonance imaging [13]. Since the spectrum is complex, both the amplitude and the phase are needed in order to calculate the image by inverse Fourier transformation.

Difficulties arise if not enough information on the diffracted field can be measured. Two common situations occur. Firstly the diffracted field can generally be measured only over a limited range of propagation directions. The earlier sections of this chapter show that this corresponds to a limited range of spatial frequencies of the object. Therefore, such data limit the resolution of the reconstructed image. Methods are available to extend the data to higher resolution although these are generally limited in scope. This is referred to super resolution and is not discussed further here. The second situation arises when the amplitude of the scattered field can be measured but the phase either cannot be measured or can be measured only inaccurately. Clearly, both the amplitude and the phase are needed in order to calculate the image by inverse Fourier transformation. This situation can arise if a wavefield is measured after propagation through a random medium that introduces phase errors, or if the wavelength of the radiation is too small for coherent detection. Since the phase needs to be determined in order to reconstruct the image, this is referred as a *phase problem*. As the phase of the diffracted field cannot be measured in many diffraction imaging problems, the phase problem is of prominent importance in this area. The characteristics of phase problems and methods for phase retrieval are described in the next subsection. In the following subsections, the characteristics of some specific diffraction imaging methods are described.

1.6.1 Phase problems

Phase problems occur in many areas of image recovery, for example, in optical astronomy, X-ray crystallography and electron microscopy. In optical astronomical imaging, the phase of the detected wavefield is distorted when propagating through regions of variable refractive index in the turbulent atmosphere. In X-ray crystallography, only the intensity of the diffraction pattern can be measured due to the short wavelength of X-rays. In electron microscopy, although a low resolution image can be measured, high resolution information can be obtained only in the diffraction plane and the phase of the diffracted electrons can not be measured. Various experimental and computational techniques have been developed to solve phase problems. Computational techniques, which are of most relevance here, are generally referred to as “phase retrieval algorithms.”

If only the amplitude of the Fourier transform of an image is measured, an immediate question is the degree to which the data uniquely define the image. In principle, there are infinite number of possible phase functions that can be used, together with the amplitude, to reconstruct an image. However, since an image represents a physical object, there are usually restrictions on what represents a feasible image. A particularly important constraint that applies to almost any physical object is *compact support* (nonzero within a finite region). This simply represents the property of any imaged object being of finite size. Such constraints limit the number of possible solutions, and ideally lead to a single unique solution. Therefore, the practical uniqueness question is: “Given the Fourier amplitude data and appropriate constraints (or *a priori* information), is it possible to retrieve a unique phase solution?”

Some characteristics of an image are irretrievably lost when the phase is not available. This can be seen as follows: The Fourier amplitude of the images

$$\begin{aligned} f(\mathbf{x}) \\ f(\mathbf{x} + \mathbf{x}_1) \exp(i\phi_1) \\ f^*(-\mathbf{x} + \mathbf{x}_2) \exp(i\phi_2) \end{aligned}$$

are the same, where $\mathbf{x}_1, \mathbf{x}_2, \phi_1, \phi_2$ are real constants and $*$ denotes complex conjugation. This indicates that the information regarding the absolute position, inversion in the origin and conjugation, and a phase constant of the image are lost. However, these ambiguities are generally trivial in most applications and so of little significance. A more serious ambiguity occurs when the image $f(\mathbf{x})$ can be written as a convolution, i.e.,

$$f(\mathbf{x}) = g(\mathbf{x}) \odot h(\mathbf{x}). \quad (1.68)$$

Consider the case where f , g and h are real. The two images

$$\begin{aligned} f(\mathbf{x}) &= g(\mathbf{x}) \odot h(\mathbf{x}) \Rightarrow |F(\mathbf{u})| = |G(\mathbf{u})||H(\mathbf{u})| \\ f_1(\mathbf{x}) &= g(\mathbf{x}) \odot h(-\mathbf{x}) \Rightarrow |F_1(\mathbf{u})| = |G(\mathbf{u})||H(\mathbf{u})|, \end{aligned} \quad (1.69)$$

therefore have the same Fourier amplitude and cannot be distinguished from the amplitude information alone. If an image is a convolution of N components, then there are 2^N images with the same Fourier amplitude. In general however, it is unlikely that an image will have the form Eq. (1.68). The uniqueness properties of the phase problem are different in the one-dimensional images and two (and more)-dimensional cases.

The uniqueness of phase problem in one dimension was first studied by O'Neill and Walther [14, 15], which is briefly explained here. For a finite size image ($f(\mathbf{x})$ with compact support), the Fourier transform $F(u)$ can be analytically continued into the complex z -plane where $u = \text{Real}\{z\}$. The function $F(z)$ is called an *analytic* function (i.e., is infinitely differentiable in the finite complex plane) and can be factorised as

$$F(z) = k \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right), \quad (1.70)$$

where k is a complex constant, i.e., is completely characterised by its zeros z_n . The intensity $I = |F|^2$ is then given as

$$I(z) = F(z)F^*(z^*) = |k|^2 \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) \left(1 - \frac{z}{z_n^*}\right). \quad (1.71)$$

Therefore, exchanging (flipping) one of the zeros z_i for z_i^* does not change the intensity $I(u)$. There are therefore an infinite number of images $f(x)$ with the same Fourier amplitude $|F(\mathbf{u})|^2$ and the problem is highly non-unique. Each of these images has the same support. A similar result exists for a discrete image of N pixels [16]. In this case the Fourier transform is an N -th order polynomial which, by the fundamental theorem of algebra, can be factorised into N linear factors, and characterised by N zeros. Flipping each zero leads to 2^N possible solutions that give the same Fourier amplitude. In general then, the phase problem is highly non-unique in the one-dimensional case.

In the case for two (or more) dimensions, however, the situation is quite different, which was noted first by Bruck and Sodin [16]. In the continuous case, a two-dimensional analytic function cannot generally be factorised, and in the discrete case there is no fundamental theorem of algebra in more than one dimension, i.e., multi-dimensional polynomials are almost always irreducible (cannot be factorised) except rare cases. In general therefore, the two (or more) dimensional phase problem is unique. Uniqueness, for two (or more)

dimensions, can be illustrated with the concept of zero sheets [17, 18]. They showed that the Fourier amplitude of a K -dimensional image is zero on a single continuous surface in a $2K$ -dimensional space, which is referred to as a zero sheet. The zero sheet has a dimension of $2K - 2$. The single zero sheet in the multi-dimensional case replaces the point zeros in one-dimensional case, so that only the whole sheet can be flipped.

Given that, in the multi-dimensional case (which are usually concerned), the phase problem has a unique solution, the problem is to find this unique image from measurements of the Fourier amplitude. This is referred to as *phase retrieval* as it is equivalent to reconstructing the Fourier phase [19, 20, 21].

A multi-dimensional image can be reconstructed directly from its Fourier-amplitude using the concept of zero sheets [17, 18]. However, this approach is computationally too expensive and too noise sensitive for practical applications. The most effective and commonly used phase retrieval algorithms are known as *iterative transform algorithms* (ITAs). The approach is to find an image that satisfies the Fourier amplitude data as well as the image-plane constraints (compact support, etc.). This is achieved by iterating between image space and Fourier space, alternately satisfying constraints in each. ITAs for phase retrieval began with the ideas of Gerchberg and Saxton [22] and later developed by Fienup [23]. The basic procedure of iterative transform algorithms is shown in Fig. 1.6. The algorithm generally starts with a random image $f'_0(x, y)$ which is Fourier transformed, and the transform modified to satisfy the Fourier constraints, i.e., the updated Fourier transform $F_n(u, v)$ at the n -th iteration is given by

$$F_n(u, v) = |F(u, v)| \exp(i\phi\{F'_n(u, v)\}), \quad (1.72)$$

where $\phi\{\}$ denotes the phase. For the case of a support constraint, the image is set to zero outside the support region, i.e.,

$$\begin{aligned} f'_{n+1}(x, y) &= f_n(x, y) && \text{where the constraints are satisfied} \\ &= 0 && \text{otherwise.} \end{aligned} \quad (1.73)$$

This is referred to as the *error reduction* (ER) *algorithm* and can also be used to incorporate a positivity constraint [23]. Although this is the most obvious way of applying image space constraints, the ER algorithm often exhibits *stagnation* in which progress towards the solution stalls [21]. The probability of stagnation is greatly reduced by using the hybrid-input-output algorithm [23] in which the image is updated as

$$\begin{aligned} f'_{n+1}(x, y) &= f_n(x, y) && \text{where the constraints are satisfied} \\ &= f'_n(x, y) - \beta f_n(x, y) && \text{otherwise.} \end{aligned} \quad (1.74)$$

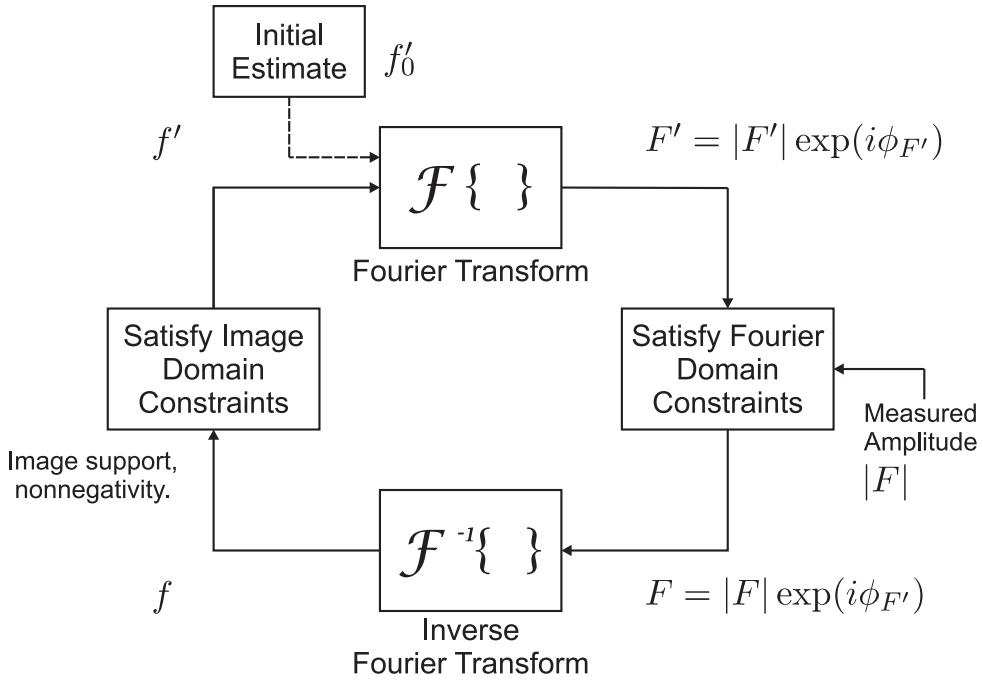


Figure 1.6: The iterative transform algorithm.

where β is a constant, usually between 0.5 and 1, called feedback parameter. The HIO algorithm is generally quite effective. Iterative transform algorithms have recently been formulated in the more general framework of *iterative projection algorithms* [24, 25]. This has allowed the development of more general algorithms and a wider variety of constraints.

1.6.2 Phase dominance

A characteristic of Fourier imaging that is related to the phase problem is the phenomenon of *phase dominance* [26, 27]. Phase dominance refers to the general observation that loss of the spectral phase information tends to lead to a less recognisable image than does loss of the spectral amplitude information. This implies that the phase contains more information than the amplitude. This characteristic of image spectra has been known for some time in a variety of fields including crystallography, image processing, visual perception and holography, as well as in signal coding and speech [28, 29, 19, 30, 26, 31, 32]. It is also of relevance in coding, compression and phase-only holograms (kinoforms) [32].

Phase dominance for images is commonly demonstrated as follows [26, 28, 27, 33, 34]. Consider two independent images $f(x, y)$ and $g(x, y)$ that are used to reconstruct two synthetic images $f_a g_\phi(x, y)$ and $f_\phi g_a(x, y)$ as

$$\begin{aligned} f_a g_\phi(x, y) &= \mathcal{F}^{-1} \{ |F(u, v)| \exp(i\phi \{ G(u, v) \}) \} \\ f_\phi g_a(x, y) &= \mathcal{F}^{-1} \{ |G(u, v)| \exp(i\phi \{ F(u, v) \}) \}, \end{aligned} \quad (1.75)$$

and the reconstructed images are compared with the original images. It is generally observed that $f_a g_\phi(x, y)$ resembles $g(x, y)$ more than it does $f(x, y)$, and $f_\phi g_a(x, y)$ resembles $f(x, y)$ more than it does $g(x, y)$. An example is shown in Fig. 5.1. This is what is usually meant by phase dominance, in the sense that the Fourier phase carries more information than does the Fourier amplitude.

Some qualitative explanations have been given for phase dominance. As a result of the Fourier shift theorem, the spectral phase is related to the relative position of objects in an image. Loss of information on the relative positions of objects is expected to render an image unrecognisable. The spectral amplitude, however, represents the magnitude of the spatial frequency components, which may be considered to affect the contrast in the image and is expected to have a lesser effect on the features in the image. Most of these statements are rather vague and qualitative, however. Phase dominance has to be carefully defined, as does its relationship to errors in the Fourier amplitude and phase. Phase dominance is studied in detail in Chapter 5.

1.6.3 Optical astronomical imaging

Optical astronomical imaging is concerned with imaging astronomical objects by detecting the light that they emit using (usually ground-based) telescopes. The objects of interest such as stars, planets, and galaxies are located at a large distance from the observer and hence the field at the telescope is described by Fraunhofer diffraction. The field at the aperture (pupil) plane is the Fourier transform of the astronomical source brightness distribution which is to be imaged (Fig. 1.7). The telescope optical system forms an image in the image plane. The resolution of the image is determined by the dimensions of the aperture in the pupil plane. The resolution is quoted as *angular resolution*, the angle subtended at the telescope between two objects that can just be resolved.

The complication in optical astronomy arises from the passage of the radiation through the Earth's atmosphere (Fig. 1.7). The atmosphere is a layer gases and particles surrounding the earth, and temperature fluctuations, variations in wind speed and direction and the concentration of water vapour generate turbulence, which results in a position- and time-varying refractive index in the atmosphere. The effect of the refractive index variations is that the wavefront, and in particular its phase, is distorted while passing through the atmosphere. At low resolution, or at small spatial frequencies or for wavefronts close to the optical axis (i.e., small apertures), the radiation passes through small lateral regions of the atmosphere where the refractive index is almost constant and phase distortions are not significant. Good images are therefore formed at low resolution, corresponding to < 1 arc-sec at a good observing site and this is the resolution limit imposed by atmospheric turbulence [7]. This resolution limit is in practice quite small for astronomical purposes. At

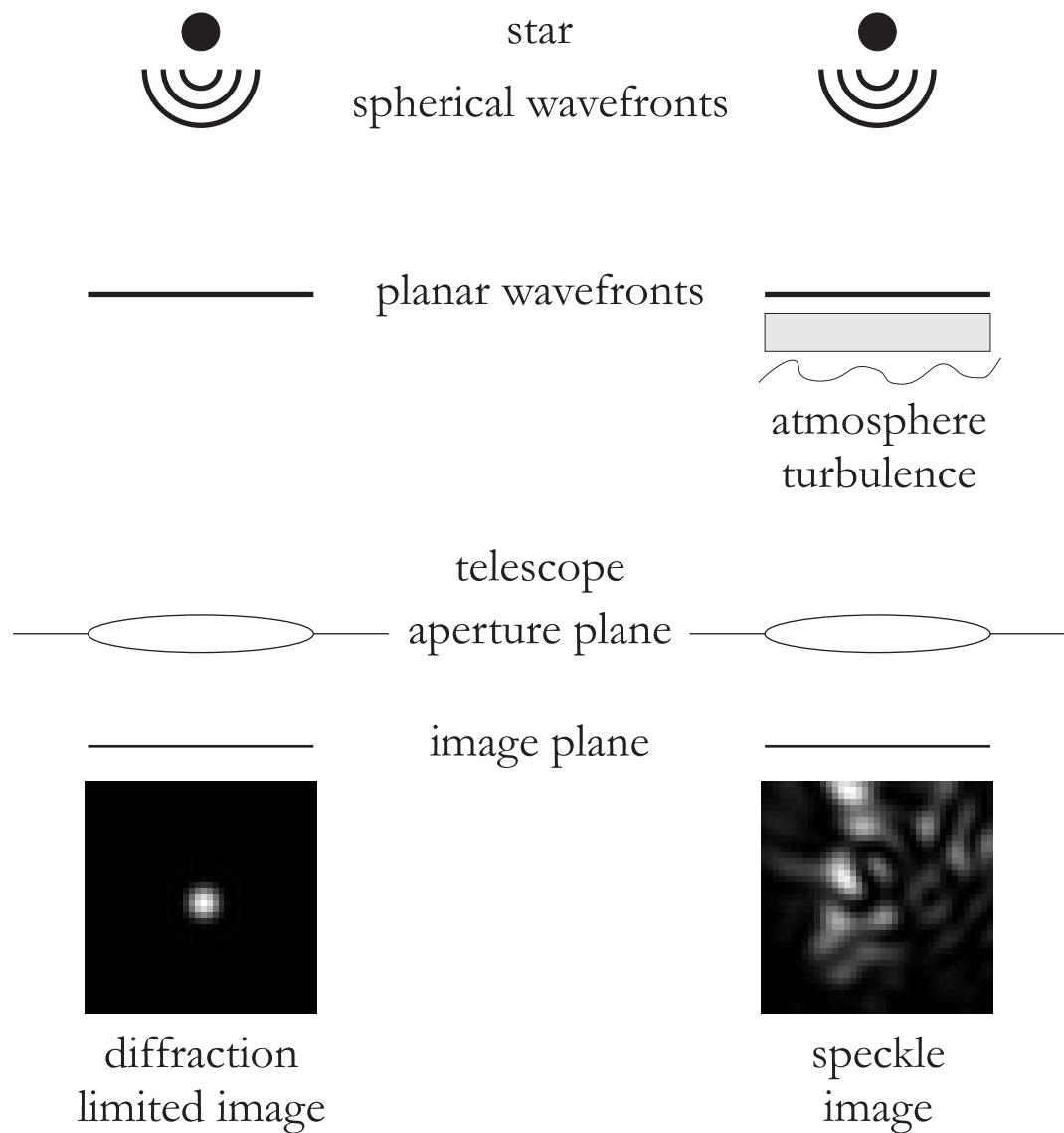


Figure 1.7: Astronomical imaging. Left: A diffraction limited image in the absence of atmosphere turbulence. Right: A speckle image due to atmospheric turbulence.

higher resolution the radiation passes through atmospheric “patches” of different refractive index which introduces phase variations at the pupil plane. This leads to highly distorted images in the image plane. To extract higher resolution images from a large aperture that are diffraction-limited, both image-plane and diffraction (aperture)-plane processing must be used. Many techniques have been developed to overcome the effects of turbulence such as speckle interferometry, blind deconvolution and adaptive optics. Adaptive optics uses a flexible mirror to detect and correct for the wavefront distortions in real time. Although effective, it requires very expensive equipment.

Astronomical speckle interferometry uses phase retrieval algorithms to solve the astronomical problem. Many short exposure images are captured within milliseconds (less than the fluctuation time of the atmosphere), which are referred to as speckle images. The atmospheric turbulence is effectively frozen (static) during each image acquisition. The images have a speckle appearance as a result of the atmosphere existing as patches with approximately constant refractive index. The m -th speckle image is given by

$$s_m(x, y) = f(x, y) \odot h_m(x, y) + n_m(x, y) \quad (1.76)$$

where $f(x, y)$ is the high-resolution astronomical image, $h_m(x, y)$ is the psf of the atmosphere above the aperture plane at the time the image was taken and $n_m(x, y)$ is the noise. The Fourier transform of each speckle image is then given by

$$S_m(u, v) = F(u, v)H_m(u, v) + N_m(u, v), \quad (1.77)$$

where $H_m(u, v)$ is the m -th OTF (Eq. (1.66)). Averaging many speckle images simply gives a long exposure (low resolution) image. However, if the spectral intensity of the speckle images is averaged, high resolution information is retained, i.e.,

$$\langle |S_m(u, v)|^2 \rangle = |F(u, v)|^2 \langle |H_m(u, v)|^2 \rangle + \langle |N_m(u, v)|^2 \rangle, \quad (1.78)$$

where $\langle |H_m(u, v)|^2 \rangle$ is referred to as the speckle transfer function. The speckle transfer function can be estimated by viewing an unresolvable star (point object) under similar conditions and using the same processing. $|F(u, v)|^2$ can be then calculated from Eq. (1.78) using deconvolution. The high resolution image $f(x, y)$ can then be recovered from $|F(u, v)|$ using phase retrieval.

Another method to reconstruct a high resolution image from the speckle images is the triple-correlation (also called bispectrum) method. The triple correlation and the bispectrum, which are a Fourier transform pair, of an image are defined as

$$\hat{f}(\mathbf{x}_1, \mathbf{x}_2) = \int_{-\infty}^{\infty} f^*(\mathbf{x})f(\mathbf{x} + \mathbf{x}_1)f(\mathbf{x} + \mathbf{x}_2) d\mathbf{x} \quad (1.79)$$

and

$$\hat{F}(\mathbf{u}_1, \mathbf{u}_2) = F(\mathbf{u}_1)F(\mathbf{u}_2)F^*(\mathbf{u}_1 + \mathbf{u}_2), \quad (1.80)$$

respectively. The speckle bispectrum can be calculated from the speckle images and (ignoring the noise) is given by

$$\langle \hat{S}_m(\mathbf{u}_1, \mathbf{u}_2) \rangle = \hat{F}(\mathbf{u}_1, \mathbf{u}_2) \langle \hat{H}(\mathbf{u}_1, \mathbf{u}_2) \rangle, \quad (1.81)$$

where $\langle \hat{H}(\mathbf{u}_1, \mathbf{u}_2) \rangle$ is called the bispectral speckle transfer function and can be determined in the similar way to that for the speckle transfer function. Therefore $\hat{F}(\mathbf{u}_1, \mathbf{u}_2)$ can be calculated. An important property of the bispectral speckle transfer function is that it is positive [35, 36], i.e., its phase is zero, so that the phase of the object bispectrum is equal to the phase of the image bispectrum, i.e., $\phi \{ \langle \hat{S}_m(\mathbf{u}_1, \mathbf{u}_2) \rangle \} = \phi \{ \hat{F}(\mathbf{u}_1, \mathbf{u}_2) \}$. Using this relationship and Eq. (1.80) gives relationships between the phases of $F(\mathbf{u})$, $\phi \{ F(\mathbf{u}) \}$, at two points \mathbf{u}_m and \mathbf{u}_n in Fourier space of the form

$$\phi \{ F(\mathbf{u}_n) \} = -\phi \{ F(\mathbf{u}_m) \} - \phi \{ F(\mathbf{u}_{-m-n}) \} + \phi \{ \hat{F}(\mathbf{u}_m, \mathbf{u}_n) \}. \quad (1.82)$$

There are many such equations for different \mathbf{u}_m and \mathbf{u}_n and these can be solved for the phases $\phi \{ F(\mathbf{u}_m) \}$, allowing the image to be reconstructed [35, 37, 38, 39].

Another approach to astronomical phase retrieval that uses multiple data is phase diversity (Gonsalves [40] and Paxman [41, 42]). This method requires two or more images of the same object. The first image is the focused image taken at the focal plane, which is a low resolution image due to the phase distortion. Additional images of the same object are captured but intentionally defocused. Possible setups for acquiring images from different planes is described by a number of authors [41]. Phase diversity can also be applied to correct other phase aberrations such as misalignment [41, 42]. The image at the focal plane can be expressed as

$$g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x}), \quad (1.83)$$

and the Fourier transform of the image is

$$G(\mathbf{u}) = F(\mathbf{u})H(\mathbf{u}), \quad (1.84)$$

where $h(\mathbf{x})$ and $H(\mathbf{u})$ are the psf and OTF of the system. The OTF of the defocused image is given by

$$H_d(\mathbf{u}) = H(\mathbf{u}) \exp(i\Delta\phi(\mathbf{u})), \quad (1.85)$$

where the subscript d denotes the defocused plane and $\Delta\phi(\mathbf{u})$ is the known phase shift due to defocusing. The Fourier transform of the defocused image is then

$$G_d(\mathbf{u}) = F(\mathbf{u})H_d(\mathbf{u}). \quad (1.86)$$

The goal is then to reconstruct $f(\mathbf{x})$ from $g(\mathbf{x})$ and $g_d(\mathbf{x})$ with $h(\mathbf{x})$ unknown. This can be approached by minimising the mean-square error

$$MSE = \int_{-\infty}^{\infty} [|G(\mathbf{u}) - \check{H}(\mathbf{u})\check{F}(\mathbf{u})|^2 + |G_d(\mathbf{u}) - \check{H}(\mathbf{u})\exp(i\Delta\phi(\mathbf{u}))\check{F}(\mathbf{u})|^2] d\mathbf{u} \quad (1.87)$$

over the estimates indicated by $\check{\cdot}$. Various methods have been used to solve the minimisation problem including maximum likelihood estimates for Gaussian and Poisson noise models [42].

1.6.4 Electron microscopy

Electron microscopes are microscopes that use electrons as the illuminating source instead of visible light, with the same basic principles as a light microscope. This is due to the fact that electrons exhibit wave-particle duality. The discussion here is limited to transmission electron microscopes (TEM) that are used to study thin semi-transparent specimens, particularly biological specimens. A diagram of a TEM is shown in Fig. 1.8. An advantage of using electrons is that their wavelength is small, of the order of 0.01 to 0.002nm, so that a very high resolution is potentially available. Another advantage of electron microscopy is that “electron lenses” can be formed by magnetic fields produced using electromagnets. However, in practice resolution is limited by aberrations due to imperfect magnetic lenses. The resolution obtainable varies, but is typically limited to about 0.1-0.5nm [43, 44, 45].

The basic principle of a transmission electron microscope is described as follows (Fig. 1.9). Electrons emitted from an electron gun first travel through a vacuum column in the microscope (as electrons are easily scattered by air), and then pass through a condenser lens before encountering the specimen. The electron wave on passing through the specimen suffers an attenuation and phase shift, and the refractive index of the specimen is therefore complex. The specimen must be thin to reduce multiple scattering and excessive attenuation. The scattered electrons, or the “exit wave”, then pass through the objective lens which forms the first intermediate image. Subsequently, the image is enlarged by the following intermediate (and projector) lenses and finally the data is recorded/displayed as electrons hitting a fluorescent screen, photographic plate, or light sensitive sensor (e.g., CCD camera).

Information in an electron microscope can be measured in the image plane or the diffraction pattern can be recorded, depending on the configuration of the lenses. This is one of

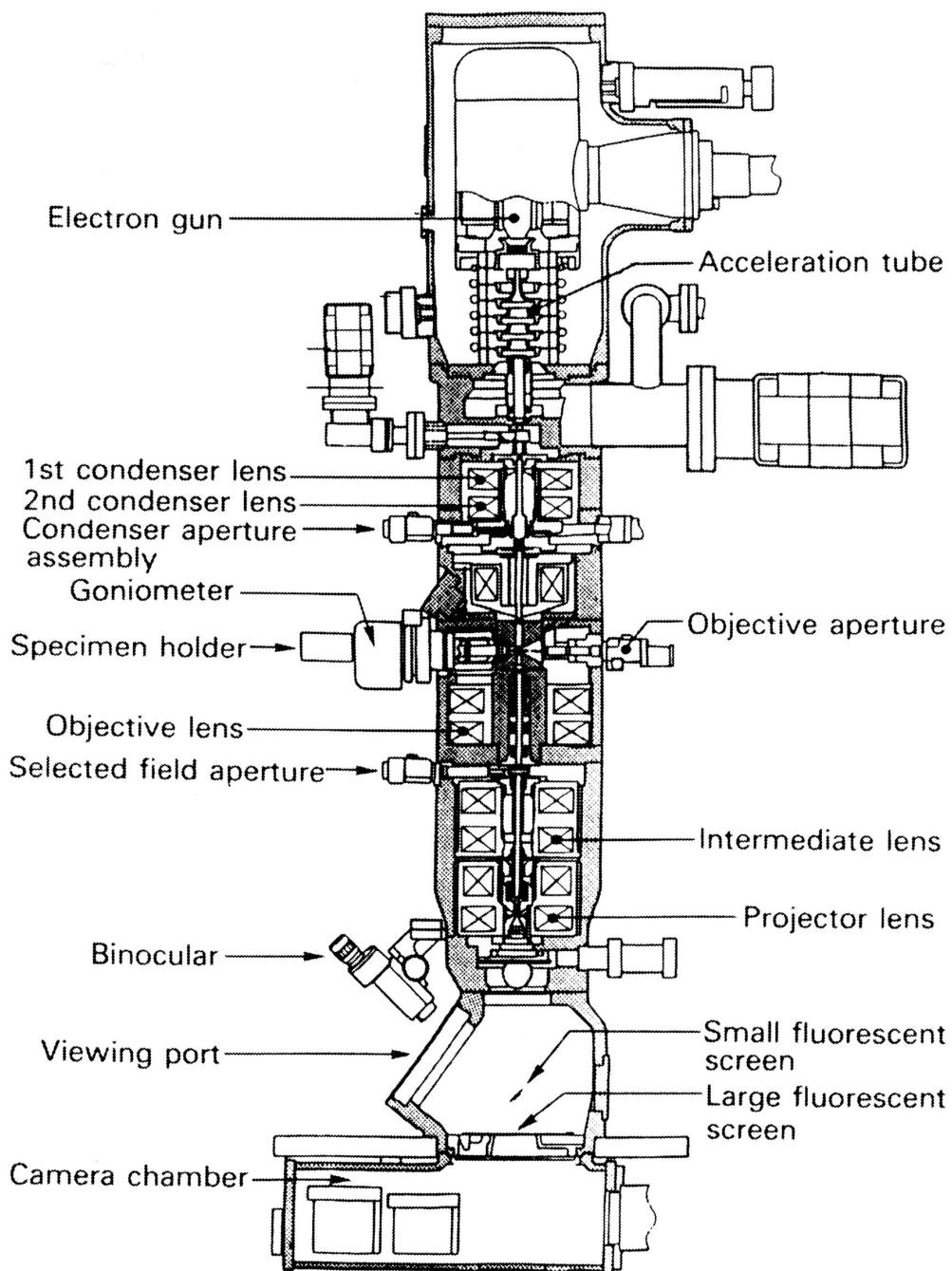


Figure 1.8: Diagram of a transmission electron microscope [46].

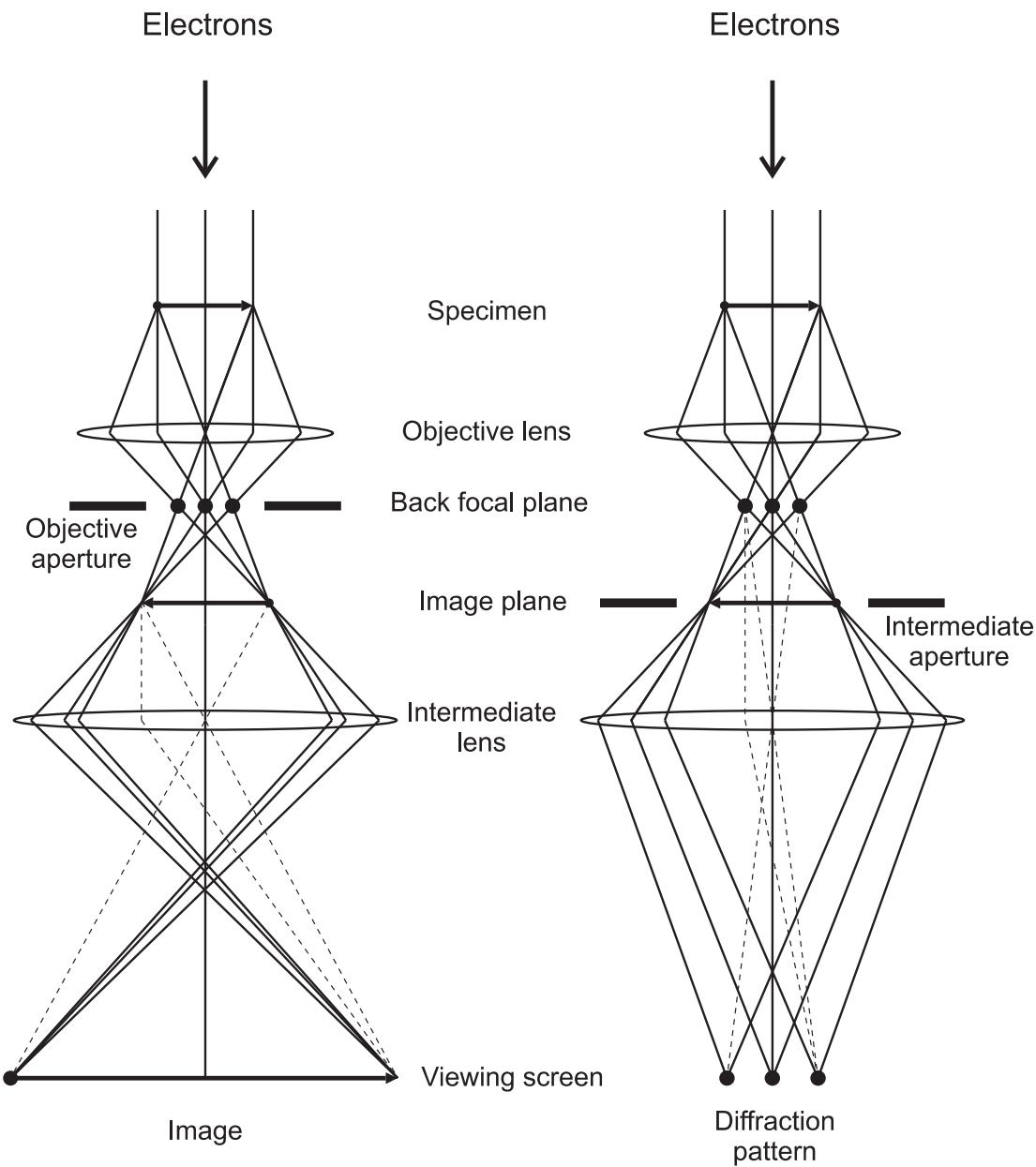


Figure 1.9: An electron microscope in image mode (left) and diffraction mode (right) [46].

the main advantages of TEM. The main principles of electron microscopy can be illustrated by geometrical optics, and are shown in Fig. 1.9. In both the image and diffraction mode, a diffraction pattern is firstly formed by the objective lens at its back focal plane. In image mode (Fig. 1.9), an objective aperture is placed at the back focal plane of the objective lens to select the electron beam for the final image and intermediate image is formed in the image plane. A bright field image is formed if the undiffracted electrons pass through, and a dark field image is formed if the undiffracted electrons are blocked. In diffraction mode (Fig. 1.9), the intermediate lens is adjusted such that it is focused on the back focal plane of the objective lens and the diffraction pattern is magnified by the following lenses. An intermediate aperture can be placed at the image plane of the objective lens to limit the diffraction pattern to a selected area of the specimen, and this is referred to as selected area electron diffraction (SAED). Furthermore, by tilting the specimen over various angles with respect to the incident electron beam, diffraction patterns of different orientations can be recorded, allowing reconstruction of a three-dimensional (thick) specimen.

In TEM, the diffracted field can be approximated by Fraunhofer diffraction, and the wave function in the image plane and in the diffraction pattern, denoted $f(x, y)$ and $F(u, v)$ respectively, form a Fourier transform pair. In electron microscopy, both $f(x, y)$ and $F(u, v)$ are complex, both containing the same information, and the phase information of each is lost, i.e., only $|f(x, y)|$ and $|F(u, v)|$ can be recorded. In practice, the resolution in image space is further degraded by aberrations, mainly caused by the lenses (other aberrations include changes in the energy level of the electrons and changes in the strength of magnetic fields) [45, 47]. These aberrations affect only the phase in Fourier space, i.e., an unknown phase term is introduced in the CTF, so that it is possible to recover a higher resolution image using the amplitude of the diffraction pattern if the phase information can be recovered.

Recovering the complex image $f(x, y)$ from measurement of its amplitude $|f(x, y)|$ and the amplitude of its Fourier transform $|F(u, v)|$ is a variation on the phase problem described in Section 1.6.1, and was first studied by Gerchberg and Saxton [48, 49]. Iterative transform algorithms similar to those described in Section 1.6.1 can be applied in which the known amplitudes in each domain are applied. Furthermore, the resolution of the reconstructed image can be extended beyond that recorded in the image plane by also estimating the amplitudes in the image using the high resolution diffraction amplitudes. If the specimen is thin enough that it affects only the phase of the diffracted field (known as a phase object), then the image amplitude is known to be constant, simplifying the problem.

An alternative approach is to recover the phase by analysing a set of defocused images directly using the transport of intensity equation (TIE). The TIE is a wave propagation equation derived by Teague [50] in terms of the intensity and phase under the small angle

approximation and shows that the phase can be estimated from the intensity measurements in the Fresnel region. This method has been applied successfully [51, 52].

1.6.5 X-ray crystallography

X-rays are electromagnetic radiation with a wavelength shorter than ultraviolet light but greater than gamma radiation, typically between 0.1\AA and 100\AA . X-rays give a suitable resolution for imaging molecules. They can be normally generated by accelerating electrons and colliding them with a metal target, which decelerates electrons rapidly and X-ray photons are emitted [53]. The wavelength of the X-rays depend on the energy of the electrons and therefore the accelerating voltage. High intensity X-rays are also generated by synchrotrons by accelerating electrons around a circular path [54].

X-ray crystallography is a technique used to determine the structures (the relative positions of the atoms) of molecules. Since the minimum distance between atoms is of the order of 1\AA , a resolution of $1 - 2 \text{\AA}$ is required to image the individual atoms. Radiation with a wavelength of the order of 1\AA is therefore required so that high-energy X-rays are used. The X-rays interact with the electron in the atoms and the quantity that is imaged is the electron density of the molecule. The atomic positions are easily located from a high resolution image of the electron density. Because the diffraction by one molecule would be very weak and undetectable, crystalline specimens that contain many molecules all in the same orientation and regularly spaced are used in order to amplify the diffracted signal. The smallest non repeating unit of the crystal is referred to as the unit cell, and may contain one, or a few, molecules (Fig. 1.10).

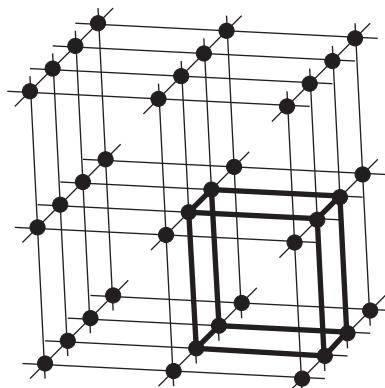


Figure 1.10: Three dimensional lattice structure in a crystal; unit cell (thick line) and the lattice (circles).

Although X-ray lenses have been developed for long wavelengths, at the wavelength used in crystallography X-ray lenses are not available, and so it is not possible to focus the diffracted waves to form an image. It is therefore necessary to record the diffraction pat-

tern and calculate the image numerically by inverse Fourier transformation. However, due to the short wavelength of X-rays, only the intensity of the diffraction pattern can be measured, but not the phase. X-ray crystallography therefore represents a classic phase problem. The phase problem in X-ray crystallography is different from the general phase problem described in Section 1.6.1 due to the periodic nature of the crystalline specimen.

The effect of periodicity is as follows. For a rectangular unit cell the electron density in a crystal can be expressed as

$$\begin{aligned} f(\mathbf{x}) &= e(\mathbf{x}) \odot l(\mathbf{x}) \\ &= e(x, y, z) \odot \sum_{h,k,l=-\infty}^{\infty} \delta(x - ha)\delta(y - kb)\delta(z - lc), \end{aligned} \quad (1.88)$$

where $e(\mathbf{x})$ is the electron density in a single unit cell, $l(\mathbf{x})$ is the lattice (Fig. 1.10), and a , b and c are lengths of the unit cell vectors. Taking the Fourier transform of Eq. (1.88) shows that the X-ray amplitudes diffracted by the crystal are

$$F(u, v, w) = E(u, v, w) \sum_{h,k,l=-\infty}^{\infty} \delta(u - h/a)\delta(u - k/b)\delta(u - l/c), \quad (1.89)$$

where $E(\mathbf{u}) = E(u, v, w)$ is the Fourier transform of $e(\mathbf{x})$, the diffraction pattern from a single unit cell. Equation (1.89) shows that the measured diffraction pattern consists of samples of $E(\mathbf{u})$ at the points $(h/a, k/b, l/c)$ where h , k and l are integers. The values are denoted

$$F_{\mathbf{h}} = F(h/a, k/b, l/c) = E(h/a, k/b, l/c), \quad (1.90)$$

where $\mathbf{h} = (h, k, l)$ and are referred to as the structure factors, and the set of points $\{(h/a, k/b, l/c)\}$ is called the reciprocal lattice. Since $|E(\mathbf{u})|^2$ is the Fourier transform of the autocorrelation of $e(\mathbf{x})$, which has extent $2a \times 2b \times 2c$, the samples $|F_{\mathbf{h}}|^2$ undersample $|E(\mathbf{u})|^2$, relative to the Nyquist frequency, by a factor 2 in each dimension. The uniqueness properties described in Section 1.6.1 therefore do not apply. The autocorrelation of $e(\mathbf{x})$ cannot be calculated, but the so-called Patterson function, which is the Fourier transform of the intensity samples $|F_{\mathbf{h}}|^2$, i.e.,

$$P(\mathbf{x}) = \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \exp(i2\pi\mathbf{h} \cdot \mathbf{x}), \quad (1.91)$$

can be calculated. This can be rewritten in terms of $A(\mathbf{x})$, the autocorrelation of $e(\mathbf{x})$, as

$$P(\mathbf{x}) = \sum_{\mathbf{h}} A(\mathbf{x} - \mathbf{h} \cdot \mathbf{a}). \quad (1.92)$$

Hence, the undersampling problem described above can be considered as a aliasing problem in autocorrelation space.

Several phase retrieval methods have been developed for X-ray crystallography to deal with the undersampling of the Fourier amplitude data.

So-called direct methods are used to solve small molecule (less than about 200 atoms) structures. These methods utilities the atomicity property of the electron density, i.e., that it is made up of sharp separated peaks that correspond to the atoms. For a molecule made up of identical, resolvable atoms, it can be shown that the structure factors are related by [55]

$$F_{\mathbf{h}} = \theta_{\mathbf{h}} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}}, \quad (1.93)$$

where $\theta_{\mathbf{h}}$ can be determined from the atomic scattering factors and this equation is referred to as Sayre's equation. Equation (1.93) gives the relationships between the phases $\phi_{\mathbf{h}}$, of the structure factors. If one of the terms on the rhs of Eq. (1.93) is dominant, the phases are related by

$$\phi_{\mathbf{h}} \approx \phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}}. \quad (1.94)$$

Many equations such as Eq. (1.94) exist for different values of \mathbf{h} and \mathbf{k} which can be solved for the phases. Although it is assumed that the atoms in the molecule are identical, Eq. (1.94) can be extended to non-identical atoms. The power of direct methods can be increased by using a probabilistic approach which treats the atomic positions as random variables. These techniques are well-developed and small molecule atomic structures can be routinely solved from high resolution diffraction amplitudes [56, 57].

For large molecules, direct methods cannot be used. Ancillary information, usually of an experimental nature, is needed to determine these structure [21]. One of the main techniques is called the isomorphous replacement method [21, 58]. An isomorphous derivative is created by attaching a heavy atom to the molecule, and its diffraction pattern measured. The structure factor for the derivative crystal, denoted by $F_{\mathbf{h}}^D$, is given by

$$F_{\mathbf{h}}^D = F_{\mathbf{h}} + F_{\mathbf{h}}^H, \quad (1.95)$$

where $F_{\mathbf{h}}^H$ is the diffraction pattern of the heavy atom alone. Rewriting Eq. (1.95) in terms of amplitude and phase gives

$$|F_{\mathbf{h}}^D|^2 = |F_{\mathbf{h}}|^2 + |F_{\mathbf{h}}^H|^2 - 2|F_{\mathbf{h}}||F_{\mathbf{h}}^H| \cos(\phi_{\mathbf{h}} - \phi_{\mathbf{h}}^H), \quad (1.96)$$

and

$$\phi_{\mathbf{h}} = \phi_{\mathbf{h}}^H + \cos^{-1} \left(\frac{|F_{\mathbf{h}}^D|^2 - |F_{\mathbf{h}}|^2 - |F_{\mathbf{h}}^H|^2}{2|F_{\mathbf{h}}||F_{\mathbf{h}}^H|} \right). \quad (1.97)$$

Using Eq. (1.97), $\phi_{\mathbf{h}}$ can be calculated from the amplitude data if the position of the heavy atom (hence $F_{\mathbf{h}}^H$) can be determined. The latter can be calculated by using an isomorphous difference Patterson map, which is calculated from the original and derivative amplitude data [21, 59]. Because of the arc-cosine function in Eq. (1.97), there are two solutions, which can be resolved by using two different isomorphous derivatives. A similar method called anomalous scattering uses multiple X-ray wavelengths as opposed to multiple heavy-atom derivatives [21, 60].

Methods commonly referred to as electron density modification implement iterative transform algorithms of the kind described in Section 1.6.1. Because the amplitude data are undersampled, these are successful only if they are started close to the correct solution. This can be achieved by starting the algorithm with approximate phases derived from isomorphous replacement, or from phases calculated from a similar molecule of known structure. Additionally, other known constraints on the molecule (such as non-crystallographic symmetry or solvent boundaries) can be incorporated into these algorithms. Such algorithms are an important component of crystallographic phase retrieval.

1.6.6 X-ray microscopy

Determination of the structures of small molecules has traditionally required the use of crystals in order to obtain diffracted X-ray amplitudes of sufficient intensity to be measured. However, recent construction of extremely high intensity synchrotron X-ray sources allows diffracted intensities to be measured from extremely small non-crystalline specimens. The X-ray source is required to maintain phase coherence across the entire width of the object in order to retrieve structural information. This is a new area of research. The advantage of using a non-crystalline sample is that the diffracted amplitudes are not sampled (are continuous), and so can be sampled sufficiently finely to represent the continuous Fourier transform amplitude. There is then sufficient data to determine the phases using the techniques described in Section 1.6.1. This is referred to as X-ray microscopy and uses numerical phase retrieval in place of a lens. These methods have been applied successfully to some nanoparticles and cells using both short and long wavelength X-rays [61, 62, 63, 10]. With more intense X-ray sources coming on line and high-speed data acquisition, it may be possible in the future to apply this technique to individual protein molecules [64, 65].

1.7 The human visual system

In this section, the human visual system and its relation to work presented in Chapter 6 are described.

1.7.1 Structure of the eye and early stage visual system

The eye is roughly spherical in shape with a diameter of approximately 20 mm. It is surrounded by three layers of membranes (Fig. 1.11). The outer layer consists of the sclera and the cornea. The sclera is an opaque membrane which connects with the cornea and encloses the whole eye ball. The cornea is the transparent sclera at the front of the eye that allows the entry of light. The choroid layer is rich in blood vessels and provides nutrients to the eye. The choroid layer also contains a light absorbing black pigment, melanin, which prevents reflection of light within the eyeball. There is a ring of muscles behind the cornea called the iris, which connects to the choroid layer. The pupil refers to the opening at the centre of iris and the amount of light entering the eye is controlled by the pupil's diameter. The lens in a human eye acts like a lens in other optical systems. The lens is made of transparent proteins called crystallins, which are water-soluble structural proteins. The anterior chamber is the space between the pupil and the lens, which is filled with a clear, watery fluid called the aqueous humour which transports nutrients to its surrounding structures and removes waste. The light entering the eye is focused by the cornea and the lens. The cornea is a non adjustable focusing device, and focus is adjusted by modifying the shape of the lens. Light focused by the lens passes through the posterior chamber, which contains a clear aqueous fluid called vitreous humour, before reaching the innermost membrane, the retina, at the back of the eyeball. The focused light pattern on the retina, called the "retinal image," is first transformed into neural responses by the photoreceptors (light sensitive receptor cells) in the retina and then transmitted to the brain through the optic nerve.

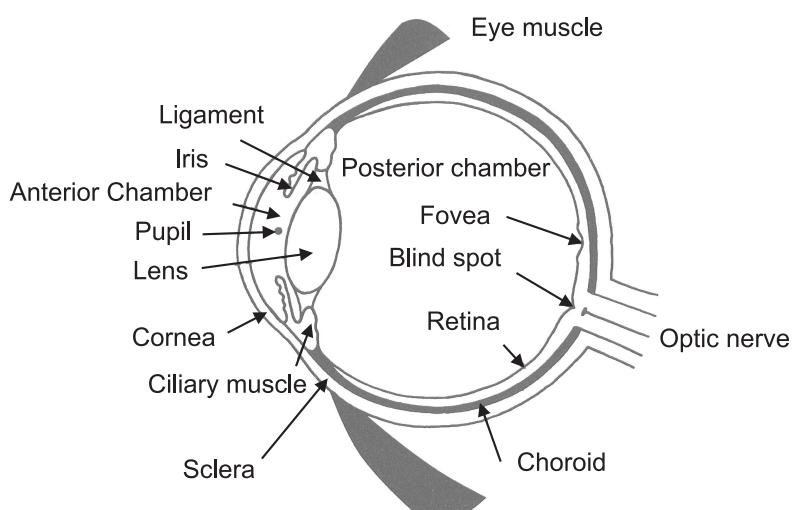


Figure 1.11: Cross sectional diagram of the human eye [66].

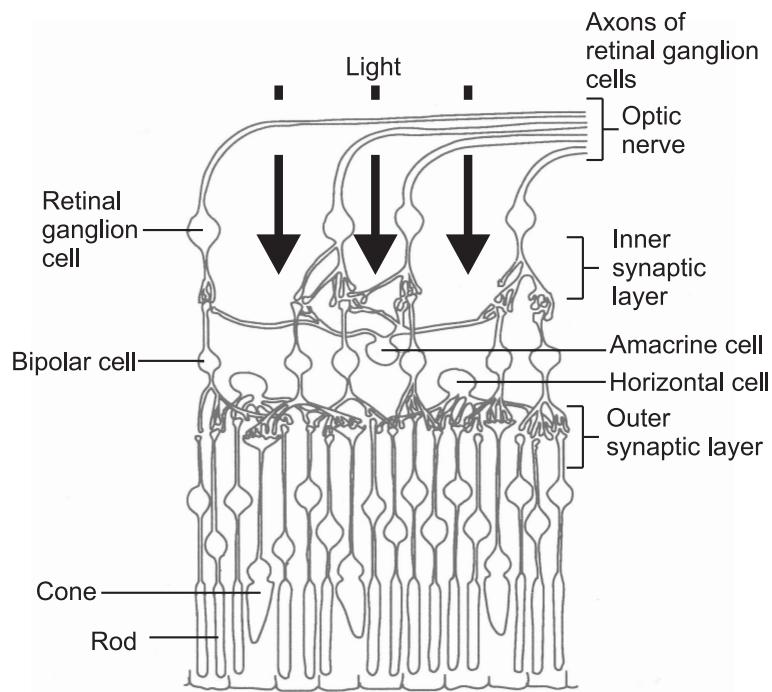


Figure 1.12: A diagram of the retinal structure [66].

There are two types of photoreceptor; rods and the cones. The human retina (Fig. 1.12) contains around 120 million rods and 6 million cones [66]. The rods are very sensitive light detectors which are active in scotopic light (dim/low light levels), are suitable for nocturnal vision, but provide only low acuity monochrome vision. Visual acuity defines the spatial resolving capacity of the visual system. The reason for low acuity under scotopic light levels, despite the large number and highly light sensitive rods, is attributed to the fact that many rods converge into one neuron in the optic nerve fibres. The high spatial density of the rods improves sensitivity, not spatial resolution. The cones are roughly four times larger than rods, are active only in photopic light levels (diurnal vision) and provide high resolution colour (trichromatic) vision [67]. The fovea (Fig. 1.11) is the center region of retina which has the highest concentration of cones (there are no rods in the center of fovea) and hence with the highest visual acuity. The high density of cones in the foveal region allows good spatial resolution. Also, the foveal cone signal is encoded by several neurons, which suggests that spatial sampling of the cones is important to the visual system.

Cones are classified into three types based on the wavelength sensitivity of the photoreceptors: L(ong)-cones, M(edium)-cones, and S(hort)-cones (also referred as red, green, and blue cones) [67, 66]. The number of S-cones is about 5 to 10% of the total number of cones, and they form a ring around the boundaries of the fovea. They are also much more sensitive than the L- and M- cones. The mosaic (spatial arrangement pattern) describes how the retinal image is sampled. The characteristics of the S-cone mosaic is probably a con-

sequence of the optical limitations of the lens (chromatic aberration). i.e., the lens blurs short-wavelength light more significantly. The M-cones and L-cones are in a 1:2 ratio and form a randomly mixed arrayed in small clusters [68, 69, 70]. Some authors suggest that this random distribution of cone types might represent a spatial and spectral sampling scheme that is a compromise between regular and clumped arrays of the cones [70].

Ganglion cells (Fig. 1.12) are neurons located in the retina that receive signals from the photoreceptors. The axons of ganglion cells merge and form the optic nerve, which connects to the brain, and leaves the eye at a point called “optic disk.” Since there are no photoreceptors at the optic disk, there is a blind spot. The response range of the optic nerve fibres is very limited compared to the range of detectable intensity ($\approx 10^{10} : 1$) of the photoreceptor [66]. This is coped with by a flexible operating range of the visual system which shifts with the average light intensity, referred to as “light adaptation.”

1.7.2 Organisation of the visual system

The visual system can be viewed as a collection of separate pathways to meet various functional requirements. The separate pathways start at the photoreceptors (rods and cones) which extend the range of viewable illuminations. There are three main kinds of ganglion cell in the retina; the magnocellular (M class), parvocellular (P class) and Koniocellular (K class) ganglion cells, which constitute about 10%, 80% and 10% of the total cells, respectively [66, 71]. The optic nerves from the two eyes merge at the optic chiasm and connect mainly to the lateral geniculate nucleus (LGN) (Fig. 1.13). All three types of cell connect to different layers in the LGN, and the subsequent visual pathways are referred as M, P and K pathways. The P class cells receive inputs from relatively few rods and cones. They have a slow conduction velocity, and are sensitive for wavelength and high spatial frequencies [66]. The M class cells receive signals from relatively many rods and cones. They respond rapidly, are sensitive to low spatial frequencies, but are not colour or wavelength sensitive. K class cells are less responsive to high spatial frequencies and they have a moderate conduction velocity [72]. Also, the K class cells are colour sensitive [73]. Neurons in the LGN receive information from ganglion cells through the same pathway, then most of the the LGN neurons project to the primary visual cortex (V1) where information is partially processed and sent to other visual cortices (V2, V3, V4 and V5) for more elaborate analysis.

Two primary pathways, called the ventral and dorsal (also called the “what and where”) pathways have been proposed [74]. The ventral stream begins with V1, goes through V2, V4 and to the inferior temporal lobe, which performs object recognition. The dorsal stream starts from V1, then goes through V2, V3, V5 and to the inferior parietal lobule which handles information associated with the spatial position of objects. Although this is an

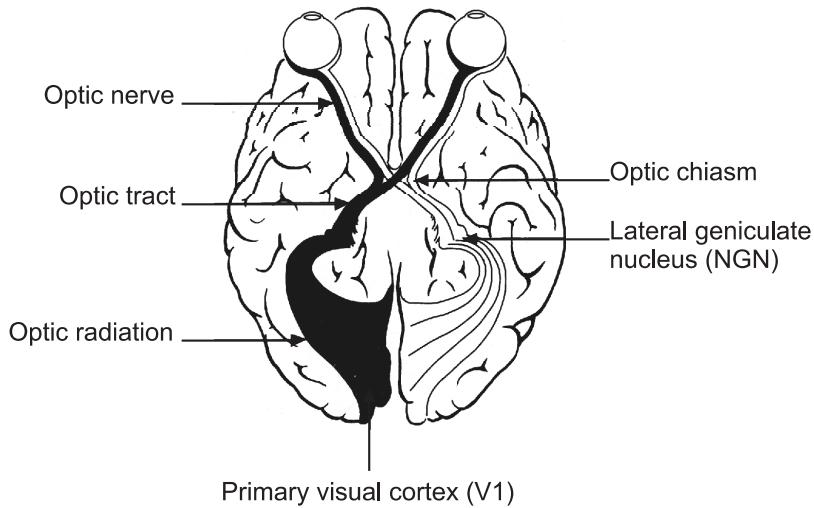


Figure 1.13: The connections from the retina to the brain [66].

active area of research, many characteristics and behaviours of the visual cortex and the visual system are still not well understood.

1.7.3 Efficiency, redundant information, and natural images

The structure of the eye, e.g., imperfect lens, fixed number of photoreceptors, mosaic pattern, uneven spatial acuity across the retina, limited number of ganglion cells' axons and neurons, etc., limits the information available to the rest of the visual system. The visual system compensates for this with several visual streams to enhance its functional ability. However, it is impractical to have a visual stream for each situation. This implies that the visual system must be flexible to adapt and/or an efficient coding scheme must be in place. Light adaptation (Section 1.7.1) has been observed in the retina, lateral geniculate nucleus and primary visual cortex. Light adaptation is one of a number of examples where the visual system adapts to various situations. It has also been suggested that an efficient image-coding scheme is an important aspect of visual processing [75].

An efficient coding scheme takes into account the statistics of the input and the information required at the output. Barlow [76] suggested that reducing some of this redundancy at an early stage of visual processing can help the visual system to transfer information more efficiently to the later stages in the visual system. It has been speculated that the visual system has developed in such a way that the coding scheme is optimised for the properties of the visual world, as a way of removing image redundancy [75].

Images from the natural environment such as trees, landscapes, and rocks are often referred to as *natural images*. It is common, in a wider definition, to include all physical objects that

the human visual system encounters, such as buildings, bridges, and urban scenes. Natural images are not random patterns, but show a large degree of structure with a number of consistent statistical properties. The study of such properties is therefore of relevance to areas such as coding, early visual processing, and image compression. Since modern techniques of image display are based in part on image models as well as models of visual perception, such properties of images are of technological interest also.

A property of natural images that has received considerable attention by many authors is the characteristic that the circularly averaged spatial power spectrum (amplitude squared of the Fourier transform) of image ensembles is a power law function of spatial frequency [75, 77, 78, 79, 80, 81, 82, 83]. This property is often referred to as *scale invariance*, although this term is reserved for properties of images in the spatial, rather than spatial frequency, domain in this thesis. This characteristic behaviour of the power spectrum implies that spatial frequency channels with constant logarithmic bandwidth contain the same energy, and has been discussed in terms of coding efficiency, the response properties of visual cortical cells, and visual processing of natural scenes [75, 80, 84, 85]. The basis of this spatial power spectral property of natural images is examined in detail in Chapter 6.

Chapter 2

Effects of Fourier Amplitude and Phase Errors on Image Reconstruction. I: Small Amplitude Errors

2.1 Introduction

In many applications (as described in Chapter 1) the spectrum (Fourier transform) of an image, instead of the image itself, is measured; therefore, both the amplitude and the phase are required to reconstruct the image by inverse Fourier transformation. The amplitude and phase are often measured separately, and they may be subjected various instrumental and detector errors, and the amplitude and phase may be measured with different degrees of precision. In the case of diffraction imaging, the precision with which the amplitude and phase can be measured depends on the wavelength of the diffracted field and the characteristics of the detector. In the case where only the amplitude is measured, the phase is effectively subjected to large errors.

The effects of spectral errors on signals have been studied previously in a number contexts [26]. A mse (Section A.2.13) analysis has been used to study magnitude-phase encoding of signals for compression and communications [32]. Error analysis has also been used to evaluate the effects of amplitude and phase quantisation on the performance of computer-generated holograms [86, 87]. Similar work has been conducted for image reconstruction from spectral amplitude data in X-ray crystallography [88] and astronomy [89]. All of these studies address the particular characteristics of particular applications. The general problem of Fourier amplitude and phase errors, and their relative effects on the quality/interpretability of reconstructed images are studied in this chapter and Chapter 3.

An inevitable question in such a study is what error metric should be used to assess image quality. Many image quality measures based on human perception have been proposed [90]. However, the mean square error, and its similar forms such as SNR (Signal to Noise Ratio) and PSNR (Peak Signal to Noise Ratio), are still the most commonly used measures for signal/image quality, despite some criticisms such as that mses do not always correspond to human visual perception [91]. Therefore, the mean square error (mse) is used here as the error metric for its simplicity and its ability to give a reasonable representation of the overall quality of an image. MSE error metric is discussed further in the next section.

The relative effects of spectral amplitude and phase errors on reconstructed images are studied in terms of the expected mean square error in the image. A theoretical analysis of the relationship between the mse in the image domain and errors in the spectral domain are presented in Section 2.2. For reasons described below, “small” amplitude errors are studied in this chapter and the case of large amplitude errors is addressed in Chapter 3. Relationships between spectral amplitude and phase errors that give rise to the same image mean square error are also derived in Section 2.2. The effects of large amplitude errors are discussed in Section 2.3 and studied in detail in Chapter 3. Simulation results are included in Section 2.4, and the overall findings are discussed in Section 2.5.

2.2 Theory

Consider an image $f(x, y)$ and its Fourier transform $F(u, v)$

$$f(x, y) \iff F(u, v), \quad (2.1)$$

where (x, y) and (u, v) are the position in image and Fourier space, respectively. It is assumed here that $f(x, y)$ is square-integrable, and since, in practice, $F(u, v)$ can only be measured in a finite region of Fourier space, denoted R , $f(x, y)$ is considered to be band-limited to $(u, v) \in R$. The Fourier transform is decomposed into the amplitude $|F(u, v)|$ and the phase $\phi(u, v)$, where

$$F(u, v) = |F(u, v)| \exp(i\phi(u, v)). \quad (2.2)$$

Now consider an image denoted $\hat{f}(x, y)$ that is reconstructed after the spectral amplitude and/or phase have been subjected to errors. The effect on the image is assessed by calculating the relative mean square error e^2 between the reconstructed and original images defined as

$$e^2 = \frac{\iint_{-\infty}^{\infty} [\hat{f}(x, y) - f(x, y)]^2 dx dy}{\iint_{-\infty}^{\infty} [f(x, y)]^2 dx dy}. \quad (2.3)$$

The mse is a commonly used measure of error in a reconstructed image, but it has to be used with some caution. For example, if the spectral data are under sampled or all phase information is lost, the reconstructed image may be shifted in space and/or multiplied by an unknown phase factor, relative to the original image [92]. The reconstructed image may then need to be registered with the original and/or a phase factor applied before the error is calculated using Eq. (2.6). In addition, spectral data are usually measured on an arbitrary scale so that the amplitude of the reconstructed image may need to be scaled to the original before calculation of Eq. (2.6). It is also assumed that the spectral data of interest are sufficiently sampled, have no systematic phase errors and have Hermitian symmetry (or are made to be Hermitian) so that the reconstructed image is not shifted and is real. While the mse is an effective measure of the difference between two images, it is not necessarily a good measure of the quality of a reconstructed image as perceived by a human observer. Various image quality measures have been proposed that attempt to incorporate aspects of visual perception and the human visual system [93]. Although psychophysical tests show that these measures can be effective, they are complicated to calculate and analyse, and their effectiveness can be quite dependent on the particular kind of distortion that is present. The mse therefore remains as an error metric that is convenient to analyse, has a clear physical meaning, and does a reasonable job of quantitating perceived error over a variety of kinds of distortion.

The Fourier transform of $\hat{f}(x, y)$ is denoted by $\hat{F}(u, v)$, and the amplitude and phase of $\hat{F}(u, v)$ are denoted by $|\hat{F}(u, v)|$ and $\hat{\phi}(u, v)$, respectively. Hence $\hat{F}(u, v)$ represents $F(u, v)$ after the addition of spectral amplitude or phase errors. The error in the Fourier transform is denoted by $\Delta F(u, v) = \hat{F}(u, v) - F(u, v)$, and at a particular spatial frequency, the errors $\Delta|F(u, v)|$ and $\Delta\phi(u, v)$ in the amplitude and phase are given by

$$\Delta|F| = |\hat{F}| - |F|, \quad (2.4)$$

$$\Delta\phi = \hat{\phi} - \phi. \quad (2.5)$$

The relationship between F and \hat{F} , for a particular spatial frequency, in the complex plane is shown in Fig. 2.1.

The objective is to relate e^2 to $\Delta|F|$ and $\Delta\phi$. Using Parseval's theorem the error e^2 can be expressed in Fourier space as

$$e^2 = \frac{\iint_R |\Delta F(u, v)|^2 du dv}{\iint_R |F(u, v)|^2 du dv}. \quad (2.6)$$

Referring to Fig. 2.1,

$$|\Delta F|^2 = |\hat{F}|^2 + |F|^2 - 2|F||\hat{F}|\cos(\Delta\phi), \quad (2.7)$$

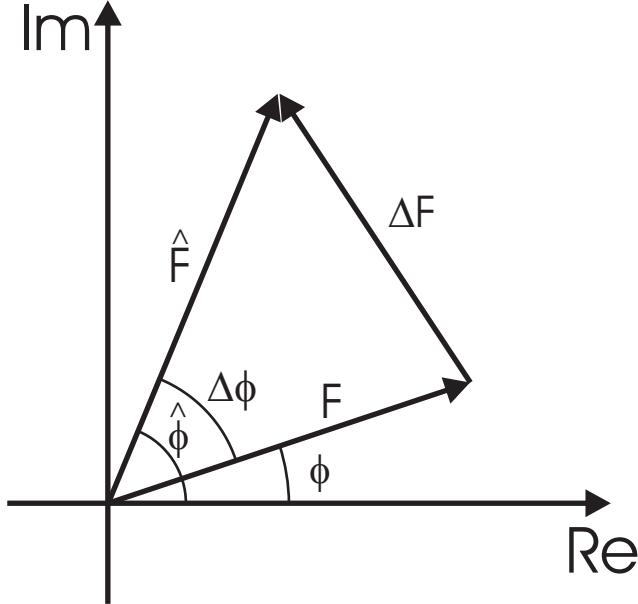


Figure 2.1: Relationship between F and \hat{F} in the complex plane.

and eliminating $|\hat{F}|$ using Eq. (2.4) allows Eq. (2.7) to be written as

$$\begin{aligned} |\Delta F|^2 &= (\Delta|F| + |F|)^2 + |F|^2 - 2|F|(\Delta|F| + |F|)\cos(\Delta\phi) \\ &= 2|F|(|F| + \Delta|F|)(1 - \cos(\Delta\phi)) + (\Delta|F|)^2. \end{aligned} \quad (2.8)$$

Using Eq. (2.8) and taking the expected value of Eq. (2.6) over the spectral errors shows that the expected mse, $\langle e^2 \rangle$, is given by

$$\langle e^2 \rangle = \frac{\left\langle \iint_R [2|F(u, v)|(|F(u, v)| + \Delta|F(u, v)|)(1 - \cos(\Delta\phi(u, v))) + \Delta|F(u, v)|^2] du dv \right\rangle}{\iint_R |F(u, v)|^2 du dv}. \quad (2.9)$$

Two assumptions are now made. First, the characteristics of the amplitude errors $\Delta|F(u, v)|$ and phase errors $\Delta\phi(u, v)$ on a particular image are described by random variables $\Delta|F|$ and $\Delta\phi$ with densities $P(\Delta|F|)$ and $P(\Delta\phi)$, respectively, and both densities are zero mean and even, i.e., $P(x) = P(-x)$. The second assumption is that the errors $\Delta|F|$ and $\Delta\phi$ are independent of each other and of the original image, and independent of (u, v) , i.e., that the statistics of $\Delta|F(u, v)|$ and $\Delta\phi(u, v)$ are independent of position in Fourier space. Then, Eq. (2.9) can be simplified as

$$\begin{aligned} \langle e^2 \rangle &= 2\langle 1 - \cos(\Delta\phi) \rangle + \frac{\langle \Delta|F|^2 \rangle}{|R^{-1}| \iint_R |F(u, v)|^2 du dv} \\ &= e_\phi^2 + e_a^2, \end{aligned} \quad (2.10)$$

where e_ϕ^2 and e_a^2 represent the mse due to the phase errors and amplitude errors, respectively, and $|R|$ is the area of the region R . The mse $\langle e^2 \rangle$ calculated from Eq. (2.10) is the sum of two terms, one which depends only on the phase errors and the other one depends only on the amplitude errors. The amplitude term in Eq. (2.10) is independent of the distribution of $\Delta|F|$ and depends only on its variance $\langle \Delta|F|^2 \rangle$, which are independent of (u, v) . The normalised variance of the amplitude errors is therefore defined as

$$\sigma_a^2 = \frac{\langle \Delta|F|^2 \rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv}, \quad (2.11)$$

so that $e_a^2 = \sigma_a^2$ and

$$\langle e^2 \rangle = 2 \langle 1 - \cos(\Delta\phi) \rangle + \sigma_a^2. \quad (2.12)$$

Since $P(\Delta\phi)$ is even, the mse due to phase errors e_ϕ^2 is given by

$$e_\phi^2 = 2 - 4 \int_0^\pi \cos(\Delta\phi) P(\Delta\phi) d\Delta\phi, \quad (2.13)$$

i.e., e_ϕ^2 does depend on the distribution function of the phase errors $\Delta\phi$. Two cases are considered; uniform and normally distributed phase errors.

Consider phase errors uniformly distributed between $-A$ and A radians. It is convenient to express this in terms of the standard deviation $\sigma_{\phi u} = A/\sqrt{3}$. Note that the uniform distribution is valid (due to wrapping on the interval $(-\pi, \pi)$) only for $A < \pi$ or $\sigma_{\phi u} < \pi/\sqrt{3} \approx 104^\circ$, and $\sigma_{\phi u}$ is limited here to this range. Using Eq. (2.13), Eq. (2.10) can be evaluated for uniformly distributed phase errors as

$$\langle e^2 \rangle_u = \sigma_a^2 + 2 \left[1 - \frac{\sin(\sqrt{3}\sigma_{\phi u})}{\sqrt{3}\sigma_{\phi u}} \right]. \quad (2.14)$$

For small phase errors, Eq. (2.14) reduces to

$$\langle e^2 \rangle_u \approx \sigma_a^2 + \sigma_{\phi u}^2, \quad \sigma_{\phi u} \ll 1 \quad (2.15)$$

(where $\sigma_{\phi u}$ is in radians).

Consider a normal distribution with variance $\sigma_{\phi n}^2$. Since the normal distribution exists on the interval $-\infty < \Delta\phi < \infty$, the phase errors outside the interval $(-\pi, \pi)$ wrap onto this

interval. The actual phase error distribution is

$$P(\Delta\phi) = \frac{1}{\sigma_{\phi n} \sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \exp\left(\frac{-(\Delta\phi - 2n\pi)^2}{2\sigma_{\phi n}^2}\right) \quad |\Delta\phi| < \pi. \quad (2.16)$$

This can be written as

$$\begin{aligned} P(\Delta\phi) &= \frac{1}{\sigma_{\phi n} \sqrt{2\pi}} \exp\left(\frac{-\Delta\phi^2}{2\sigma_{\phi n}^2}\right) \sum_{n=-\infty}^{\infty} \exp\left(\frac{2n\pi\Delta\phi}{\sigma_{\phi n}^2}\right) \exp\left(\frac{-2n^2\pi^2}{\sigma_{\phi n}^2}\right) \quad |\Delta\phi| < \pi \\ &= \frac{1}{\sigma_{\phi n} \sqrt{2\pi}} \exp\left(\frac{-\Delta\phi^2}{2\sigma_{\phi n}^2}\right) \left[1 + 2 \sum_{n=1}^{\infty} \cosh\left(\frac{2n\pi\Delta\phi}{\sigma_{\phi n}^2}\right) \exp\left(\frac{-2n^2\pi^2}{\sigma_{\phi n}^2}\right) \right] \quad |\Delta\phi| < \pi \end{aligned} \quad (2.17)$$

showing that the phase error distribution on $(-\pi, \pi)$ is the normal distribution on that interval multiplied by the expression in the square brackets. The distribution $P(\Delta\phi)$ is shown in Fig. 2.2(a). Note that $P(\Delta\phi)$ is almost uniform for $\sigma_{\phi n} > 180^\circ$. The variance of $P(\Delta\phi)$, $\tilde{\sigma}_{\phi n}^2$, is

$$\begin{aligned} \tilde{\sigma}_{\phi n}^2 &= \int_{-\pi}^{\pi} \Delta\phi^2 P(\Delta\phi) d\Delta\phi, \\ &= \frac{1}{\sigma_{\phi n} \sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} \Delta\phi^2 \exp\left(\frac{-(\Delta\phi - 2n\pi)^2}{2\sigma_{\phi n}^2}\right) d\Delta\phi, \\ &= \frac{1}{\sigma_{\phi n} \sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_{(-1-2n)\pi}^{(1-2n)\pi} (\Delta\phi + 2n\pi)^2 \exp\left(\frac{-\Delta\phi^2}{2\sigma_{\phi n}^2}\right) d\Delta\phi, \\ &= \sigma_{\phi n}^2 - 2\sqrt{2\pi}\sigma_{\phi n} \sum_{n=-\infty}^{\infty} n \left[\exp\left(\frac{-\pi^2(1-2n)^2}{2\sigma_{\phi n}^2}\right) - \exp\left(\frac{-\pi^2(-1-2n)^2}{2\sigma_{\phi n}^2}\right) \right] \\ &\quad + 2\pi^2 \sum_{n=-\infty}^{\infty} n^2 \left[\operatorname{erf}\left(\frac{\pi(1-2n)}{\sqrt{2}\sigma_{\phi n}}\right) - \operatorname{erf}\left(\frac{\pi(-1-2n)}{\sqrt{2}\sigma_{\phi n}}\right) \right] \\ &= \sigma_{\phi n}^2 - 2\sqrt{2\pi}\sigma_{\phi n} S_1 + 2\pi^2 S_2, \end{aligned} \quad (2.18)$$

where $\operatorname{erf}(\cdot)$ is the error function. The sum S_1 can be simplified as follows:

$$\begin{aligned} S_1 &= \sum_{n=-\infty}^{\infty} n \exp\left(\frac{-\pi^2(1-2n)^2}{2\sigma_{\phi n}^2}\right) - \sum_{n'=-\infty}^{\infty} (n'-1) \exp\left(\frac{-\pi^2(1-2n')^2}{2\sigma_{\phi n}^2}\right) \\ &= \sum_{n=-\infty}^{\infty} \exp\left(\frac{-\pi^2(1-2n)^2}{2\sigma_{\phi n}^2}\right) = \sum_{\substack{m=-\infty \\ m \text{ odd}}}^{\infty} \exp\left(\frac{-\pi^2 m^2}{2\sigma_{\phi n}^2}\right) \\ &= 2 \sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} \exp\left(\frac{-\pi^2 m^2}{2\sigma_{\phi n}^2}\right). \end{aligned} \quad (2.19)$$

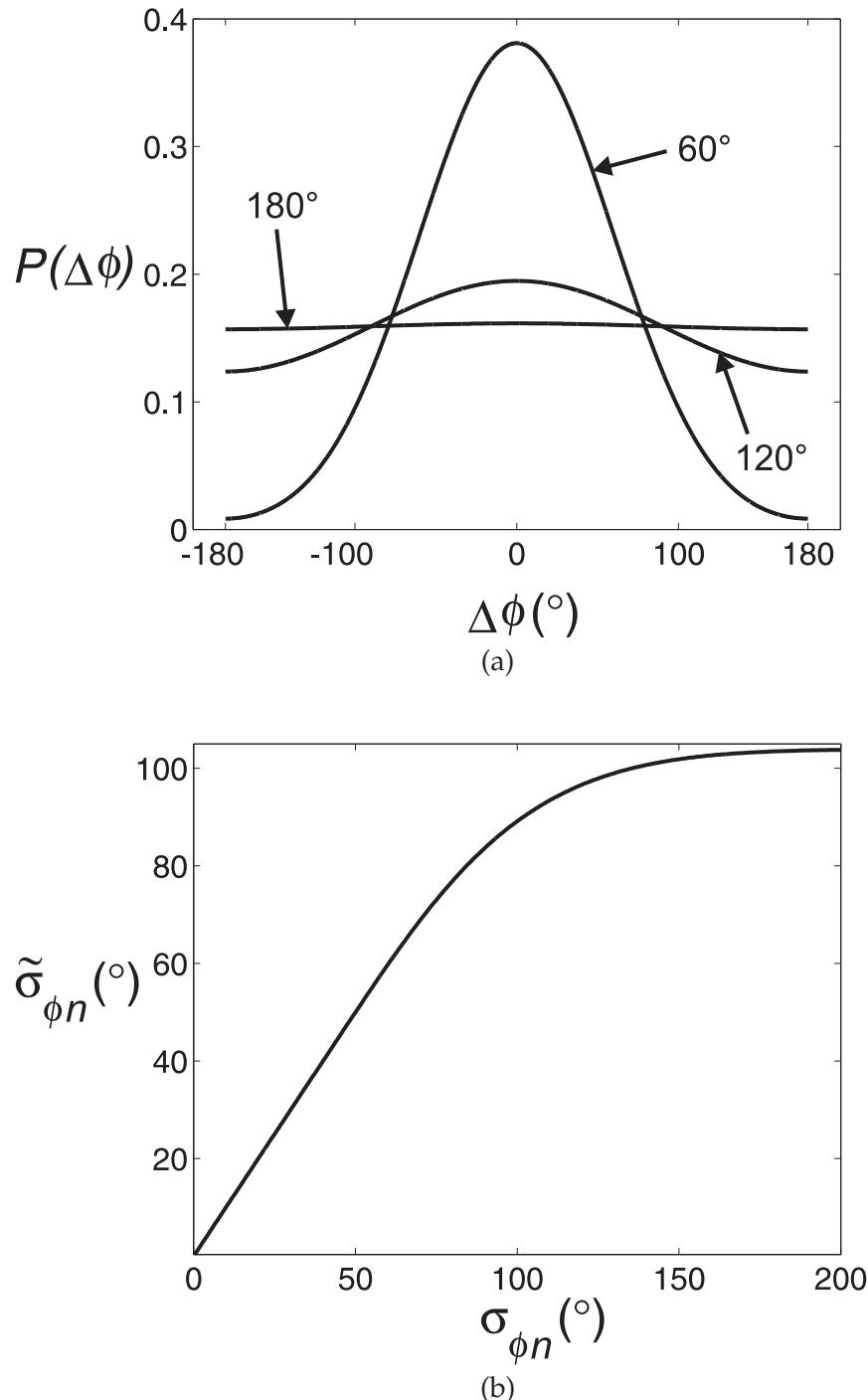


Figure 2.2: (a) The phase error distribution $P(\Delta\phi)$ for normally distributed phase errors for values of $\sigma_{\phi n}$ as shown and (b) the standard deviation $\tilde{\sigma}_{\phi n}$ of $P(\Delta\phi)$ versus the standard deviation $\sigma_{\phi n}$ of the normal distribution.

In order to split S_2 into two convergent sums, S_2 is first rewritten as

$$S_2 = \sum_{n=-\infty}^{\infty} n^2 \left[\operatorname{erfc} \left(\frac{\pi(-1-2n)}{\sqrt{2}\sigma_{\phi n}} \right) - \operatorname{erfc} \left(\frac{\pi(1-2n)}{\sqrt{2}\sigma_{\phi n}} \right) \right], \quad (2.20)$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ is the complementary error function. The sum can then be simplified as

$$\begin{aligned} S_2 &= \sum_{n'=-\infty}^{\infty} (n'-1)^2 \operatorname{erfc} \left(\frac{\pi(1-2n')}{\sqrt{2}\sigma_{\phi n}} \right) - \sum_{n=-\infty}^{\infty} n^2 \operatorname{erfc} \left(\frac{\pi(1-2n)}{\sqrt{2}\sigma_{\phi n}} \right) \\ &= \sum_{n=-\infty}^{\infty} (1-2n) \operatorname{erfc} \left(\frac{\pi(1-2n)}{\sqrt{2}\sigma_{\phi n}} \right) = \sum_{\substack{m=-\infty \\ m \text{ odd}}}^{\infty} m \operatorname{erfc} \left(\frac{\pi m}{\sqrt{2}\sigma_{\phi n}} \right) \\ &= 2 \sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} m \operatorname{erfc} \left(\frac{\pi m}{\sqrt{2}\sigma_{\phi n}} \right). \end{aligned} \quad (2.21)$$

Using Eqs. (2.19) and (2.21), $\tilde{\sigma}_{\phi n}^2$ is then given by

$$\tilde{\sigma}_{\phi n}^2 = \sigma_{\phi n}^2 - 4\sqrt{2\pi}\sigma_{\phi n} \sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} \exp \left(\frac{-\pi^2 m^2}{2\sigma_{\phi n}^2} \right) + 4\pi^2 \sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} m \operatorname{erfc} \left(\frac{m\pi}{\sqrt{2}\sigma_{\phi n}} \right). \quad (2.22)$$

The relationship Eq. (2.22) is shown in Fig. 2.2 (b). Note that $\tilde{\sigma}_{\phi n} \approx 104^\circ$, the same as for a uniform distribution, for $\sigma_{\phi n} > 180^\circ$.

Since the phase error $\Delta\phi$ appears in the cosine function in Eq. (2.13), it can be evaluated directly using the normal distribution and replacing the upper limit of the integral by infinity, giving

$$\langle e^2 \rangle_n = \sigma_a^2 + 2 [1 - \exp(-\sigma_{\phi n}^2/2)]. \quad (2.23)$$

For small phase errors, Eq. (2.23) reduces to

$$\langle e^2 \rangle_n \approx \sigma_a^2 + \sigma_{\phi n}^2, \quad \sigma_{\phi n} \ll 1, \quad (2.24)$$

i.e., the same as for uniformly distributed phase errors.

Equivalent amplitude and phase errors can therefore be defined as the value of σ_a for amplitude only errors that gives the same mse as for phase only errors with the corresponding value of σ_ϕ . This relationship is given by

$$\sigma_a = \left(2 - \frac{2 \sin(\sqrt{3}\sigma_{\phi u})}{\sqrt{3}\sigma_{\phi u}} \right)^{1/2} \quad (2.25)$$

for uniformly distributed phase errors and by

$$\sigma_a = (2 - 2 \exp(-\sigma_{\phi n}^2/2))^{1/2} \quad (2.26)$$

for normally distributed phase errors. Note that $\sigma_{\phi n}^2$ is used instead of $\tilde{\sigma}_{\phi n}^2$ in Eq. (2.26), since the former represents the measurement errors. For small phase errors, both Eqs. (2.25) and (2.26) reduce to

$$\sigma_a \approx \sigma_\phi, \quad \sigma_\phi \ll 1, \quad (2.27)$$

where σ_ϕ is in radians. Therefore, the term “small phase errors” refers to $\sigma_\phi < 0.3$ radians = 17°

2.3 Effects of large amplitude errors

There are two effects that are not considered in the previous analysis (Section 2.2) that become important when the amplitude errors are large. These are referred to here as the scaling and saturation effects. The scaling effect arises from the calculation of $\langle e^2 \rangle$ using Eq. (2.3) misrepresenting the perceived error for large amplitude errors. This can be seen by noting an energy difference between the original and reconstructed images, i.e.,

$$\left\langle \iint_R |\hat{F}(u, v)|^2 du dv \right\rangle = \left\langle \iint_R |F(u, v)|^2 du dv \right\rangle (1 + \sigma_a^2). \quad (2.28)$$

Therefore, for large values of σ_a the energy in the reconstructed image is substantially larger than the energy in the original image. This tends to make the overall amplitude of the reconstructed image larger than that of the original image. The mse $\langle e^2 \rangle$ is therefore due in part to the overall difference in amplitude between the two images. In comparing reconstructed and original images (both in terms of visual perception and in quantitative technical applications) the overall amplitude of the whole image tends to be unimportant. The effect is that the relative mse calculated by Eq. (2.3) overestimates the actual error. This difficulty is easily addressed, for example, by scaling the reconstructed image to have the same energy as the original image before calculating the error. Calculations using such an error metric show that if $\sigma_a < 0.4$ then Eq. (2.6) overestimates the mse by less than 0.02. Such an error metric is described in Chapter 3.

The saturation effect is a shortcoming of the model used which is embodied in Eq. (2.4). A problem arises if $\Delta|F|$ is large in magnitude, negative and $\Delta|F| < -|F|$, since then Eq. (2.4) implies that $|\hat{F}| < 0$. Because the amplitude cannot be negative, the result is that $|\hat{F}|$ is replaced by $-|\hat{F}|$ and a phase error $\Delta\phi = \pi$ is introduced. In such a case then, the amplitude and phase errors are misrepresented by the model. Since the effect of saturation

is a shortcoming of the model itself, rather than the calculation of the error, a more suitable model is one in which the Fourier amplitude saturates at zero if the amplitude error is such that $\Delta|F| < -|F|$. The calculations presented in Chapter 3 show that if $\sigma_a < 0.3$ the mse is overestimated by less than 0.03. The model used here is therefore suitable for $\sigma_a < 0.3$, and it is appropriate and convenient to use the error metric $\langle e^2 \rangle$ for this range of amplitude values. Therefore, throughout this thesis, “small amplitude errors” refers to $\sigma_a < 0.3$. The error metric, appropriate for images with large amplitude errors, is further developed in Chapter 3.

Two points are worth making. First, since for most images the spectral amplitudes fall off with increasing magnitude of the spatial frequency, and the magnitude of the amplitude errors in our model is independent of frequency, the saturation effect will predominantly affect the high spatial frequencies. Second, scaling and saturation must be addressed together when considering large amplitude errors since both effects are significant. Although one can consider an error metric that compensates for scaling alone to isolate this effect (as was done in Ref. [34, 94, 95]), the results are of little significance since the model incorrectly treats the large amplitude errors (the saturation effect).

2.4 Results

Using Eqs. (2.14) and (2.23), plots of the mse versus amplitude and phase errors for the suitable ranges are shown in Fig. 2.3. The plot in Fig. 2.3(a) simply reflects the relationship $e_a^2 = \sigma_a^2$. The maximum expected mse for phase errors only is 2 as can be seen in Fig. 2.3(b). Note that for large $\sigma_{\phi n}$ the actual phase errors approach a uniform distribution. If the mse for normally distributed phase errors is plotted versus $\tilde{\sigma}_{\phi n}$ then the plots are similar for uniform and normal phase error distributions (Fig. 2.4).

Equivalent amplitude and phase errors (Eqs. (2.25) and (2.26)) are shown in Fig. 2.5. Since this analysis is restricted to $\sigma_a < 0.3$, σ_ϕ also only spans a small range. In this range, there is little distinction between uniformly and normally distributed phase errors, and between $\sigma_{\phi n}$ and $\tilde{\sigma}_{\phi n}$. The relationship in Fig. 2.5 is consistent with Eq. (2.27)

Real images were used in simulations to validate the analysis in Section 2.2 for small amplitude errors ($\sigma_a < 0.3$) and all ranges of phase errors. The pixel values of the original images were first shifted such that the average value over the image is zero since the addition of phase errors to the otherwise very large value of the Fourier transform at zero spatial frequency (generally the case for images) causes unreasonably large and erratic errors in the images. The error at zero spatial frequency does not change the image structure or contrast. The errors added to the spectral amplitude and phase were manipulated such that the Fourier transform retains Hermitian symmetry and reconstructed images are real.

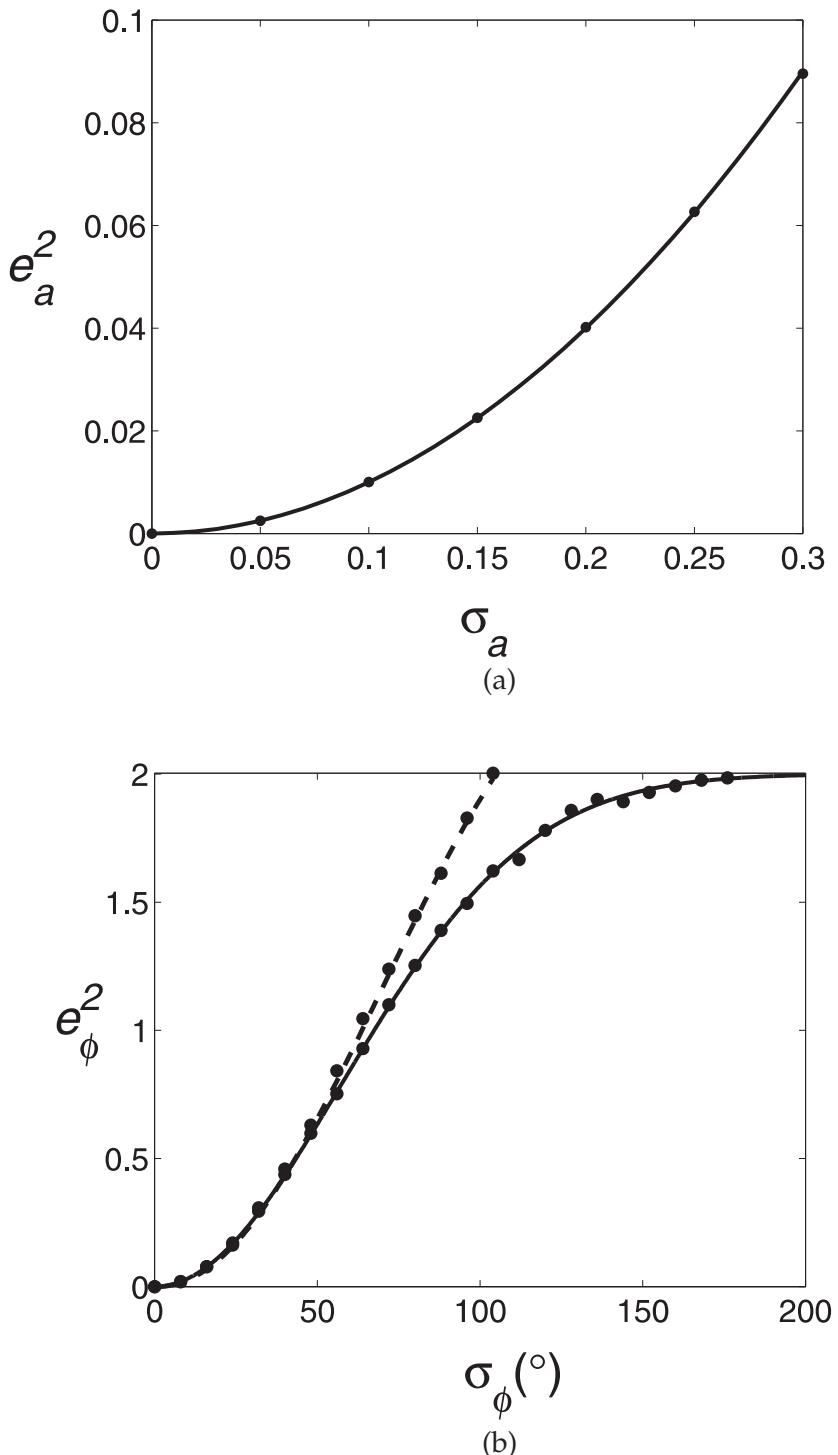


Figure 2.3: The mse $\langle e^2 \rangle$ versus (a) amplitude error σ_a (for $\sigma_\phi = 0$), and (b) phase error for uniformly σ_{ϕ_u} (---) and normally σ_{ϕ_n} (—) distributed errors (for $\sigma_a = 0$) calculated from Eqs. (2.14) and (2.23). The dots show results from the simulations.

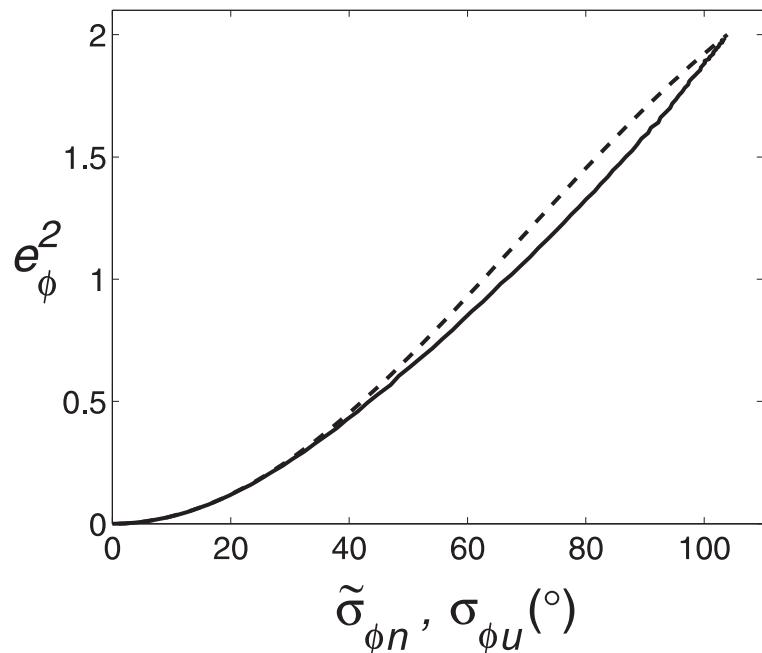


Figure 2.4: e_{ϕ}^2 versus the phase error standard deviation for uniformly (--) and normally (—) distributed phase errors.

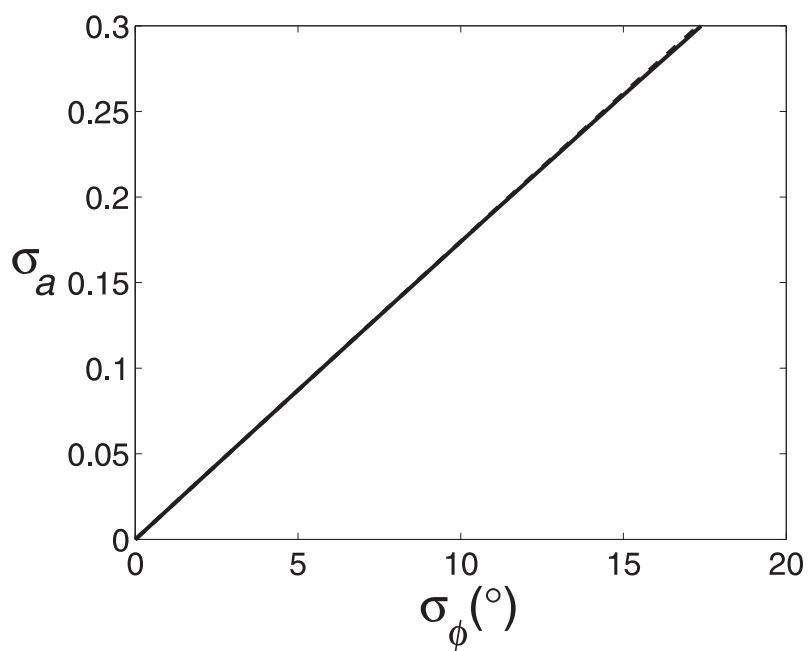


Figure 2.5: Relationship between amplitude error σ_a and phase error σ_{ϕ} for uniformly (--) and normally (—) distributed phase errors, such that the mses $\langle e^2 \rangle$ are equal.

In practice, measured Fourier data are usually averaged to enforce Hermitian symmetry to produce real image. The addition of Hermitian errors is expected to approximate this situation reasonably well. All reconstructed images are displayed such that their full dynamic range is represented as black to white.

The Fourier transform of a number of 512×512 pixel images were calculated, errors added to the Fourier amplitudes and phases and images reconstructed. The mses of the reconstructed images were calculated and averaged over 100 sets of noise signals and are plotted in Fig. 2.3. Agreement with the theoretical results is seen to be excellent. Example reconstructed images are shown in Fig. 2.6. Each column in the figure shows reconstructed images with a constant mse, the original image in the left column and the mse increasing across the page from left to the right with the values given in the caption. The images in the top row were constructed with normally distributed spectral amplitude errors only, and those in the bottom row with normally distributed phase errors only. The images in the center row were constructed using both amplitude and phase errors such that each contributes one half of the mse based on Eq. (2.10). The amplitude errors were restricted to $\sigma_a < 0.3$. Inspection of this figure shows a number of interesting features. In Fig. 2.6, the images are slightly distorted and still quite recognisable, since the errors are restricted to a small range. The quality of images in the same column (with either amplitude or phase or both types of errors) are similar in quality.

To observe the effect of small amplitude and phase errors more clearly, the image difference defined as

$$\Delta f(x, y) = \hat{f}(x, y) - f(x, y) \quad (2.29)$$

was calculated for the central 9 images in the box in Fig. 2.6, and is shown in Fig. 2.7. In order to clearly show the locations, and magnitudes of regions of positive and negative errors, the image difference is represented by a grey scale as shown by the intensity bar in Fig. 2.7. The intensity bar covers the same numerical range in each direction, which is equal to the maximum magnitude of the image difference. Pixels whose errors are less than 15% of the maximum are shown as white. Pixels with negative errors are shown by light gray and those with positive errors by dark gray as indicated. Inspection of Fig. 2.7 shows that although the images all appear similar, amplitude and phase errors introduce different kinds of error. Amplitude errors tend to affect the whole image uniformly, although the errors fluctuate rapidly from pixel to pixel. However, errors due to spectral phase errors are more correlated from pixel to pixel and vary slowly over the image, leading to patches of larger error. These different kinds of errors can be seen more clearly in the enlarged regions shown in Fig. 2.8. These characteristics are consistent with the above observations of the reconstructed images. The uniform distribution of image errors due to spatial am-



Figure 2.6: Images reconstructed with a variety of normally distributed amplitude and phase errors as described in the text. The original image is shown in the left column. The $\text{mse} \langle e^2 \rangle$ is identical in each column and increases across the page to the right with values of 0, 0.03, 0.06, 0.09 and 0.18. The top row contains amplitude errors only, the center row both amplitude and phase errors as described in the text, and the bottom row is for phase errors only. The top right corner is purposely left blank due to the restriction of $\sigma_a < 0.3$ for small amplitude errors.

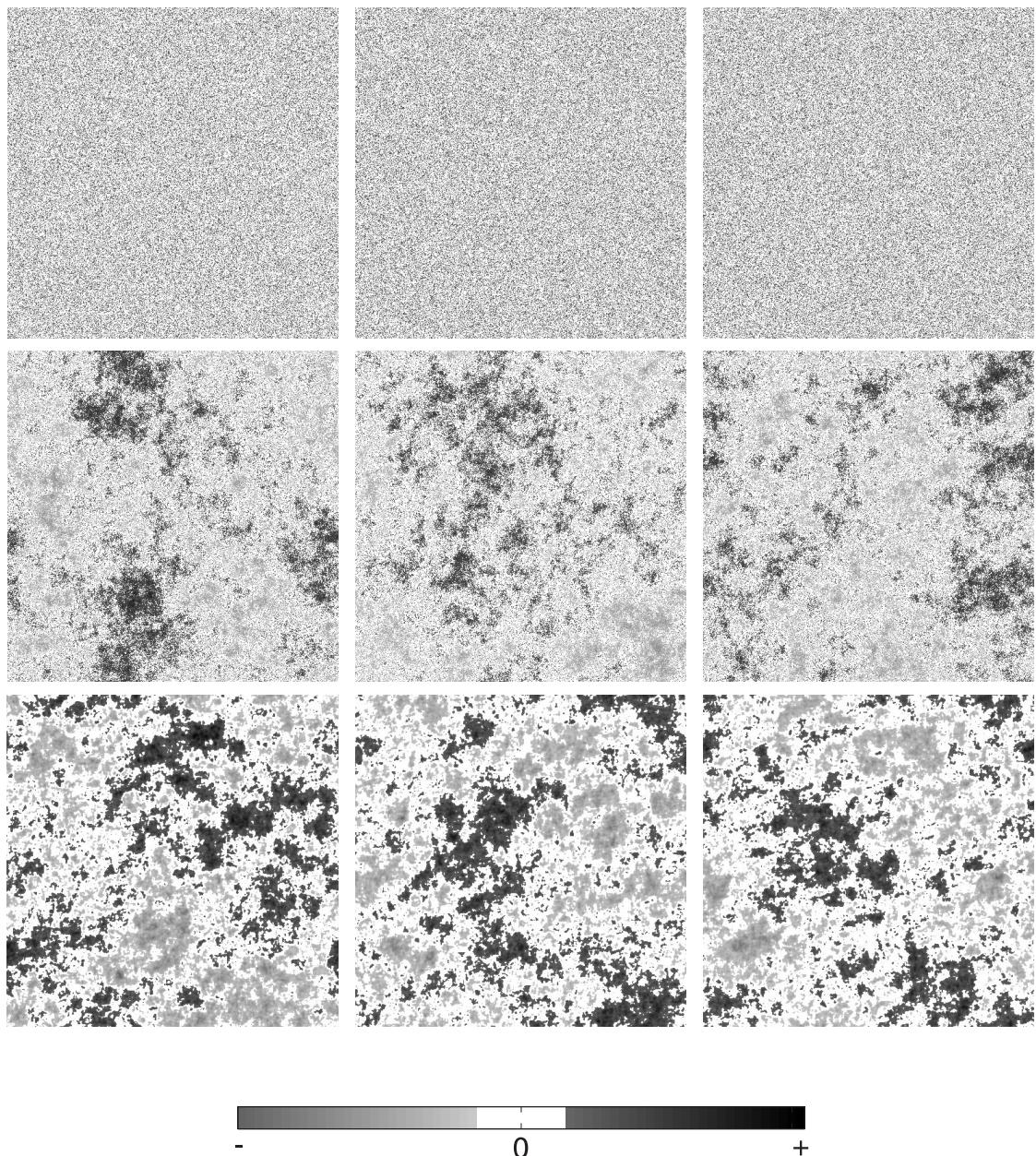


Figure 2.7: The image difference $\Delta f(x, y)$ for the images shown in the box in Fig. 2.6.

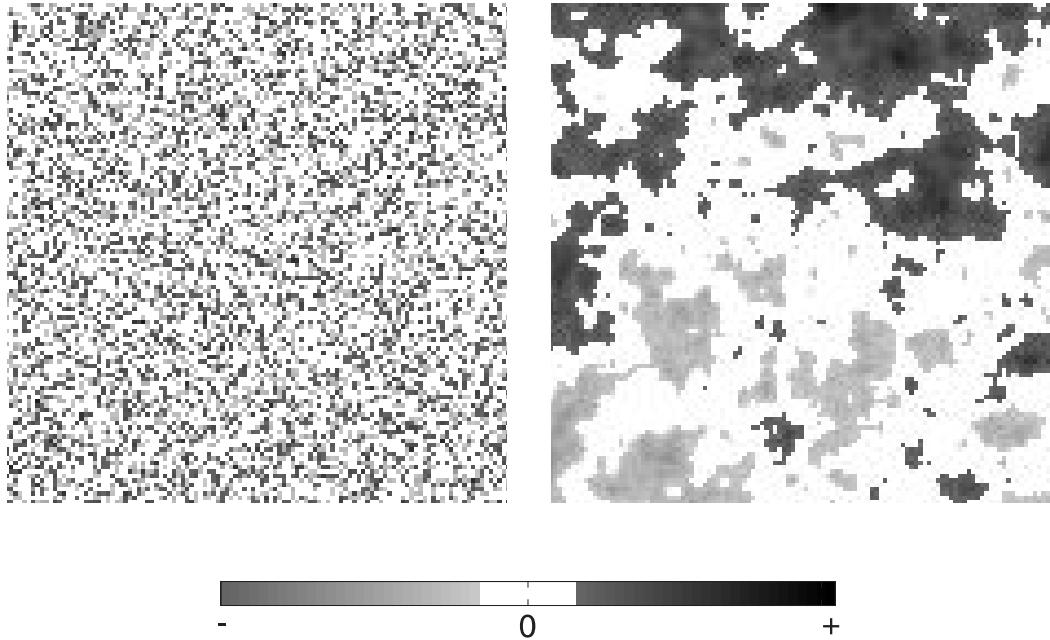


Figure 2.8: Enlarged portions of two image differences shown in Fig. 2.7. Left: top right of Fig. 2.7. Right: bottom right of Fig. 2.7.

plitude errors leads primarily to a loss of contrast in the reconstructed images, whereas the correlated patches of image errors due to phase errors lead to erroneous features in the reconstructed images. These characteristics are due in part to the model of the spectral amplitude errors. Since the magnitude of the amplitude errors is constant throughout Fourier space (i.e., the power spectrum of the noise is flat), the associated errors in the image domain are uncorrelated, leading to the observed rapid fluctuations from pixel to pixel.

Since the phase errors are not restricted to a small value in this analysis, simulations were also conducted for large normally distributed phase errors and the results are shown in Fig. 2.9. The image differences are shown in the right hand column of Fig. 2.9. Inspection of Fig. 2.9 shows that for larger phase errors the reconstructed images become unrecognisable. As in the case for small phase errors, errors in the reconstructed images are highly correlated between neighbouring pixels. There is some evidence that phase errors degrade the edges more than do amplitude errors. In general, large phase errors tend to destroy the structure/objects in an image by introducing diffuse structures, with no obvious edges, that are not present in the original image.

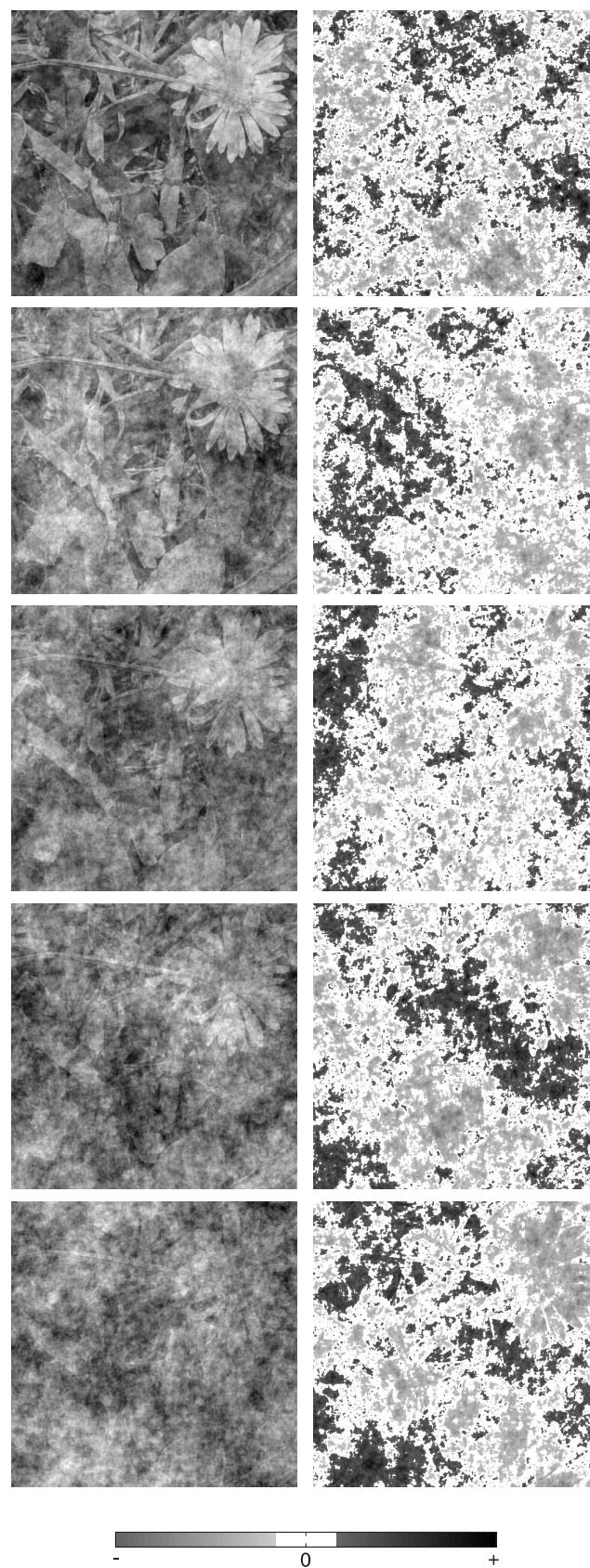


Figure 2.9: Reconstructed images with large normally distributed phase errors (left) and the corresponding image differences (right). The phase error and mse increase down the figure with the values $\sigma_{\phi n} = 30^\circ, 43^\circ, 55^\circ, 67^\circ, 80^\circ$, and $\langle e^2 \rangle = 0.25, 0.5, 0.75, 1, 1.25$.

2.5 Discussion

Expressions for the expected mse in images reconstructed from spectral amplitudes and phases that are subject to different independent errors have been derived. This mse is suitable for any phase error, but only for small amplitude errors with a normalised standard deviation less than about 0.3. Equivalent amplitude and phase errors are defined as those that give an identical mse in a reconstructed image and an expression for equivalent amplitude and phase error is derived. Simulation results support the theoretical expressions. In this range, the mse appears to correspond well to visual image quality. For small errors, the effects of amplitude and phase errors are small, and the images are still quite recognisable. Detailed inspection of errors in the images shows that while amplitude errors introduce uncorrelated errors in the reconstructed image, phase errors introduce spatially correlated errors in the reconstructed images. However, the effects on the overall quality/interpretability of images are quite similar for small errors. Large phase errors cause large correlated errors that render the image unrecognisable.

For the noise model used here the magnitude of the amplitude errors is independent of spatial frequency. This model may be appropriate when measurement errors are dominated by those that occur in the diffraction plane. The result is that, as explained in the previous section, the amplitude errors lead to spatially uncorrelated errors in the reconstructed image. In other cases, for example if the dominant noise first propagates through a system transfer function before detection in the diffraction plane, the magnitude of the amplitude errors may reduce with increasing spatial frequency. In such cases a more appropriate noise model may be one in which the noise level tracks the overall spatial frequency content of the image, i.e., $\langle \Delta|F(u, v)|^2 \rangle \propto \langle |F(u, v)|^2 \rangle$. In this case the amplitude errors will produce errors in the image domain that are spatially correlated, introducing features in the reconstructed image that are more similar to those produced by phase errors described here.

Both the scaling and saturation effects becomes prominent when amplitude errors are large ($\sigma > 0.3$). The scaling effect leads to misleading mse values caused by a larger energy in the reconstructed image. The saturation effect is a shortcoming of the model which introduce unwanted phase errors. Large amplitude errors and both of these effects are discussed in detail in Chapter 3.

Chapter 3

Effects of Fourier Amplitude and Phase Errors on Image Reconstruction. II: Large Amplitude Errors

3.1 Introduction

The effects of spectral amplitude and phase errors on image reconstruction were studied in Chapter 2. The mean square error was used there to quantitate the error between the reconstructed and the original image. It was noted in Chapter 2 that for large spectral amplitude errors, such that the normalised standard deviation of the amplitude errors is greater than about 0.3, (1) the mse as defined there does not provide a useful measure of the error, and (2) the model used there does not accommodate the effects of large amplitude errors. These are referred to here as the scaling and saturation effects. In this chapter, the theory and simulations presented in Chapter 2 are extended to the case of large amplitude errors.

The theoretical development used in Chapter 2, as noted in Section 2.3, is not appropriate for large amplitude errors. The difficulty arises when the amplitude error $\Delta|F|$ is negative and its magnitude exceeds the original amplitude $|F|$. For the additive noise model used this would lead to a negative amplitude $|\hat{F}|$. Clearly, as noted in Section 2.3, this is more likely to occur for large amplitude errors, or for large σ_a . The model therefore produces an erroneous amplitude and a π phase error in such cases. The additional phase error is expected to introduce more error into the reconstructed image so that the error calculated using this model is expected to overestimate the actual error. In most practical situations it is expected that large negative amplitude errors will saturate the detector, i.e., the am-

plitude, at zero. This is the model used here. The problem then is to incorporate this (nonlinear) saturation phenomenon into the error analysis. This is addressed in this chapter. The scaling effect, as described in Section 2.3, refers to an overestimated mse due to the additional energy contributed by the amplitude errors. This difficulty is addressed here by scaling the energy of the reconstructed image to that of the original image before calculating the mse. The scaling effect is also addressed in this chapter.

A general mse metric is developed in Section 3.2 that scales the energy of the reconstructed image and also accommodates a model that incorporates saturation. In Section 3.3 a model that incorporates saturation is developed and expressions for the mse derived. Simulations results are presented in Section 3.4 and compared with the theoretical results, and the overall findings are discussed in Section 3.5.

3.2 A general error metric

A mse is developed here that accommodates saturation and scaling. This done by first defining an error, denoted by e'^2 , that uses the reconstructed image after its energy has been scaled to that of the the original image. This is then developed so that it accommodates saturation by removing some of the assumptions used in Chapter 2. The squared error e'^2 is defined by

$$\begin{aligned} e'^2 &= \frac{\iint_{-\infty}^{\infty} [\alpha \hat{f}(x, y) - f(x, y)]^2 dx dy}{\iint_{-\infty}^{\infty} [f(x, y)]^2 dx dy} = \frac{\iint_R |\alpha \hat{F}(u, v) - F(u, v)|^2 du dv}{\iint_R |F(u, v)|^2 du dv} \\ &= \frac{\iint_R |\Delta F'(u, v)|^2 du dv}{\iint_R |F(u, v)|^2 du dv}, \end{aligned} \quad (3.1)$$

where (Fig. 3.1)

$$\Delta F' = \alpha \hat{F} - F, \quad (3.2)$$

and the scaling factor α scales the energy of the reconstructed image to that of the original image, i.e.,

$$\alpha^2 = \frac{\iint_{-\infty}^{\infty} |f(x, y)|^2 dx dy}{A \left\langle \iint_{-\infty}^{\infty} |\hat{f}(x, y)|^2 dx dy \right\rangle} = \frac{\iint_R |F(u, v)|^2 du dv}{\left\langle \iint_R |\hat{F}(u, v)|^2 du dv \right\rangle}, \quad (3.3)$$

where A denotes the area of the image, and therefore α^2 is effectively constant for a particular image and noise statistics.

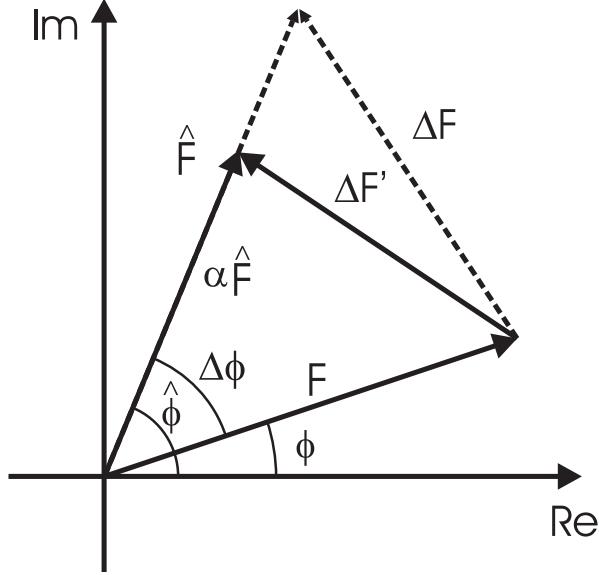


Figure 3.1: Relationship between F and \hat{F}' in the complex plane.

Referring to Fig. 3.1 and using Eq. (2.4) gives

$$\begin{aligned} |\Delta F'|^2 &= \left(\alpha|\hat{F}|\right)^2 + |F|^2 - 2\alpha|F||\hat{F}|\cos(\Delta\phi) \\ &= \alpha^2(|F| + \Delta|F|)^2 + |F|^2 - 2\alpha|F|(|F| + \Delta|F|)\cos(\Delta\phi) \\ &= |F|^2(\alpha^2 + 1 - 2\alpha\cos(\Delta\phi)) + 2|F|\Delta|F|(\alpha^2 - \alpha\cos(\Delta\phi)) + \alpha^2(\Delta|F|)^2. \end{aligned} \quad (3.4)$$

Using Eq. (3.4) and taking the expected value of Eq. (3.1) over the noise signals gives the expected mse, $\langle e'^2 \rangle$, for large errors as

$$\begin{aligned} \langle e'^2 \rangle &= \frac{\left\langle \iint_R |F(u, v)|^2 [\alpha^2 + 1 - 2\alpha\cos(\Delta\phi(u, v))] du dv \right\rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv} \\ &\quad + \frac{\left\langle \iint_R 2|F(u, v)|\Delta|F(u, v)| [\alpha^2 - \alpha\cos(\Delta\phi(u, v))] du dv \right\rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv} \\ &\quad + \frac{\left\langle \iint_R \alpha^2\Delta|F(u, v)|^2 du dv \right\rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv}. \end{aligned} \quad (3.5)$$

Equation (3.5) can be compared with Eq. (2.9), and the scaling factor α is included to remove the scaling effect. For the model (described below) that incorporates saturation, all of the assumptions following Eq. (2.9) are valid except that (1) $|F|$ and $\Delta|F|$ are not indepen-

dent of each and of (u, v) , and (2) $\langle \Delta|F| \rangle \neq 0$. Therefore, Eq. (3.5) reduces to

$$\begin{aligned} \langle e'^2 \rangle &= \alpha^2 + 1 - 2\alpha \langle \cos(\Delta\phi) \rangle \\ &+ \frac{2(\alpha^2 - \alpha \langle \cos(\Delta\phi) \rangle) \langle \iint_R |F(u, v)| \Delta|F(u, v)| du dv \rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv} \\ &+ \frac{\alpha^2 \langle \iint_R \Delta|F(u, v)|^2 du dv \rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv}, \end{aligned} \quad (3.6)$$

which should be compared with Eq. (2.9).

The quantities β and γ are introduced and defined by

$$\beta = \frac{\langle \iint_R |F(u, v)| \Delta|F(u, v)| du dv \rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv} \quad (3.7)$$

and

$$\gamma^2 = \frac{\langle \iint_R \Delta|F(u, v)|^2 du dv \rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv}. \quad (3.8)$$

Note that since it is not assumed that $\langle \Delta|F| \rangle = 0$, γ^2 is not in general the normalised variance of $\Delta|F|$. Equation (3.6) can then be written as

$$\langle e'^2 \rangle = 1 + \alpha^2 (1 + 2\beta + \gamma^2) - 2\alpha(1 + \beta) \langle \cos(\Delta\phi) \rangle. \quad (3.9)$$

Using Eqs. (2.3) and (2.7), the energy of the reconstructed image is given by

$$\iint_R |\hat{F}(u, v)|^2 du dv = \iint_R [|F(u, v)|^2 + 2|F(u, v)|\Delta|F(u, v)| + (\Delta|F(u, v)|)^2] du dv, \quad (3.10)$$

and substituting into Eq. (3.3) gives

$$\begin{aligned} \alpha^2 &= \frac{\iint_R |F(u, v)|^2 du dv}{\langle \iint_R [|F(u, v)|^2 + 2|F(u, v)|\Delta|F(u, v)| + (\Delta|F(u, v)|)^2] du dv \rangle} \\ &= \frac{1}{1 + 2\beta + \gamma^2}. \end{aligned} \quad (3.11)$$

Hence, the three parameters α , β and γ in Eq. (3.9) are not all independent, and $\langle e'^2 \rangle$ can be written in terms of only two of them. It is convenient to eliminate γ and write Eq. (3.9) in the form

$$\begin{aligned} \langle e'^2 \rangle &= 2\alpha(1 + \beta)(1 - \langle \cos(\Delta\phi) \rangle) + 2[1 - \alpha(1 + \beta)] \\ &= e_\phi'^2 + e_a'^2. \end{aligned} \quad (3.12)$$

Referring to Eq. (3.12), $\langle e'^2 \rangle$ is separated into two terms denoted e'_ϕ^2 and e'_a^2 , in a similar way that $\langle e^2 \rangle$ is separated into e_ϕ^2 and e_a^2 in Eq. (2.10). The significance is slightly different however. In the case of $\langle e^2 \rangle$, e_ϕ^2 and e_a^2 can be interpreted as being due to the phase and amplitude errors, respectively. In the case of $\langle e'^2 \rangle$, for phase errors only $\beta = 0$, $\gamma = 0$ and $\alpha = 1$, so that $\langle e'^2 \rangle = e'_\phi^2$. For amplitude errors only, $(1 - \langle \cos(\Delta\phi) \rangle) = 0$, so that $\langle e'^2 \rangle = e'_a^2$. Therefore, for amplitude *or* phase errors *only*, the interpretation is the same as for e_ϕ^2 and e_a^2 . However, if *both* phase and amplitude errors are present, then *both* contribute to e'_ϕ^2 , and the effects cannot be separated into the sum of two terms. This is in distinction to e_ϕ^2 and e_a^2 .

As described in Chapter 2, one can define equivalent amplitude and phase errors as those which, alone, give the same mse. For the case of $\langle e'^2 \rangle$ this corresponds to $e'_\phi^2 = e'_a^2$, and using Eq. (3.12) give the corresponding relationship

$$\alpha(1 + \beta) = \langle \cos(\Delta\phi) \rangle. \quad (3.13)$$

In the case of $\langle e^2 \rangle$, a relationship between σ_a and σ_ϕ can be determined for a particular phase error distribution. However, for $\langle e'^2 \rangle$, as a result of the dependence on α and β , the relationship also depends on the amplitude error distribution and the image spectral amplitude distribution.

Consider first the case where there is no saturation. In this case $\beta = 0$, $\gamma = \sigma_a$ and $\alpha = (1 + \sigma_a^2)^{-1/2}$. Substitution into Eq. (3.12) gives

$$\langle e'^2 \rangle = \frac{2 \langle 1 - \cos(\Delta\phi) \rangle}{\sqrt{1 + \sigma_a^2}} + 2 \left(1 - \frac{1}{\sqrt{(1 + \sigma_a^2)}} \right), \quad (3.14)$$

which can be compared with Eq. (2.12). For small amplitude errors, i.e., $\sigma_a \rightarrow 0$, then

$$\langle e'^2 \rangle \rightarrow 2 \langle 1 - \cos(\Delta\phi) \rangle + \sigma_a^2 = \langle e^2 \rangle, \quad \sigma_a \rightarrow 0, \quad (3.15)$$

i.e., the same result as for Eq. (2.10), as expected. As explained in Section 2.3, however, saturation will be significant when scaling becomes necessary, so the case of scaling but no saturation (Eq. (3.14)) is not pertinent and not considered further. When saturation is present, in the limit $\sigma_a \rightarrow 0$, $\beta \rightarrow 0$, $\alpha \rightarrow 1 - \gamma^2/2$ and $\gamma \rightarrow \sigma_a$, and referring to Eq. (3.12) leads to Eq. (3.15), as expected.

3.3 Saturation

Saturation is a signal-dependent, non-additive, or nonlinear phenomenon, since the effect of adding an error to a spectral amplitude sample depends on the value of that sample.

Therefore, although the noise is considered additive, its effect is non-additive. The approach taken is to determine the *equivalent* additive errors in the amplitude in order to determine the image mse. These equivalent errors are signal-dependent, so the primary difference to the analysis presented in Chapter 2 is that one needs to consider not the density of errors $P(\Delta|F|)$, but the joint density of the errors and the amplitudes $P_{\Delta|F|,|F|}(\Delta|F|, |F|)$. The result is that the equivalent additive errors, denoted $\Delta|F|_s$ where the subscript s denotes saturation, and $|F|$ are not independent and $\Delta|F|_s$ is not zero-mean. The joint density is therefore needed to calculate the quantities β and γ defined in the previous section in order to calculate the image mse.

Since the Fourier amplitude $|F(u, v)|$ is considered as a deterministic variable, $P(|F|)$ refers to the actual distribution (histogram) of Fourier amplitudes of a particular image in the following analysis, even though the term density (or pdf) is used. The effect of saturation can be described as follows. Consider an error $\Delta|F|$ applied to the amplitude $|F|$ that saturates at zero so that the resulting amplitude $|\hat{F}|_s$ is given by

$$\begin{aligned} |\hat{F}|_s &= |F| + \Delta|F| && \text{for } \Delta|F| > -|F| \\ &= 0 && \text{for } \Delta|F| < -|F|. \end{aligned} \quad (3.16)$$

The new amplitude is denoted by $|\hat{F}|_s$ to distinguish it from $|\hat{F}|$ as described by Eq. (2.4). The amplitude of the reconstructed image can then be written in terms of the equivalent additive error $\Delta|F|_s$ as

$$|\hat{F}|_s = |F| + \Delta|F|_s, \quad (3.17)$$

where

$$\begin{aligned} \Delta|F|_s &= \Delta|F| && \text{for } \Delta|F| > -|F| \\ &= -|F| && \text{for } \Delta|F| < -|F|. \end{aligned} \quad (3.18)$$

The form of Eq. (3.17) implies that Eq. (3.12) can be used to calculate $\langle e'^2 \rangle$ as long as β and γ (and therefore α) are calculated using Eqs. (3.7) and (3.8) with $\Delta|F|$ replaced by $\Delta|F|_s$. To make this clear, the values so calculated are denoted by α_s , β_s and γ_s . It is easily seen from Eq. (3.18) that $\Delta|F|_s$ and $|F|$ are dependent, and that $\Delta|F|_s$ is not zero mean.

The problem now reduces to calculating β_s and γ_s for given densities $P(\Delta|F|)$ and $P(|F|)$. Calculation of β_s and γ_s requires determining $\langle |F| \Delta|F|_s \rangle$ and $\langle (\Delta|F|_s)^2 \rangle$, which require the density $P(\Delta|F|_s)$ and the joint density $P_{\Delta|F|_s,|F|}(\Delta|F|_s, |F|)$. Using Eq. (3.17), the joint density $P_{|\hat{F}|_s,|F|}(|\hat{F}|_s, |F|)$ can be used instead of $P_{\Delta|F|_s,|F|}(\Delta|F|_s, |F|)$, as described below. It is required to obtain these densities in terms of the densities $P(\Delta|F|)$ and $P(|F|)$ which characterise the spectral amplitude errors and the image, and which are assumed known.

In the following, expressions for $P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(|\hat{F}|_s, |F|)$ and $P(\Delta|F|_s)$ are derived formally, and then a more heuristic derivation of $P(\Delta|F|_s)$ is given. These expressions are then used to derive expressions for $\langle |F| \Delta|F|_s \rangle$ and $\langle (\Delta|F|_s)^2 \rangle$.

Using Eqs. (3.17) and (A.36), the density $P(\Delta|F|_s)$ can be expressed as

$$P(\Delta|F|_s) = \int_{-\infty}^{\infty} P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(|F| + \Delta|F|_s, |F|) d|F|. \quad (3.19)$$

It is noted from Eq. (3.16) that saturation affects only the negative components of $|\hat{F}|$, and therefore $P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(x > 0, |F|)$ and $P_{|\hat{\mathbf{F}}|, |\mathbf{F}|}(x > 0, |F|)$ are identical. Furthermore, $P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(x = 0, |F|)$ is an impulse with amplitude equal to the sum of $P_{|\hat{\mathbf{F}}|, |\mathbf{F}|}(x, |F|)$ over $x \leq 0$, and $P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(x < 0, |F|) = 0$, since all negative values of $|\hat{F}|$ are removed by saturation. Combining these results gives

$$\begin{aligned} P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(|\hat{F}|_s, |F|) &= P_{|\hat{\mathbf{F}}|, |\mathbf{F}|}(|\hat{F}|_s, |F|) & |\hat{F}|_s > 0 \\ &= \left[\int_{-\infty}^0 P_{|\hat{\mathbf{F}}|, |\mathbf{F}|}(x, |F|) dx \right] \delta(|\hat{F}|_s) & |\hat{F}|_s = 0 \\ &= 0 & |\hat{F}|_s < 0, \end{aligned} \quad (3.20)$$

where $\delta(\cdot)$ is the impulse function. Example plots of $P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(|\hat{F}|_s, |F|)$ and $P_{|\hat{\mathbf{F}}|, |\mathbf{F}|}(|\hat{F}|, |F|)$ are shown in Fig. 3.2. In the figure, the image Fourier amplitude and amplitude error distributions are represented by Rayleigh and normal distributions, respectively.

Since $\Delta|F|$ and $|F|$ are independent,

$$\begin{aligned} P_{|\hat{\mathbf{F}}|, |\mathbf{F}|}(|\hat{F}|, |F|) &= P_{\Delta|\mathbf{F}|, |\mathbf{F}|}(|\hat{F}| - |F|, |F|) \\ &= P(|F|) P_{\Delta|\mathbf{F}|}(|\hat{F}| - |F|), \end{aligned} \quad (3.21)$$

and Eq. (3.20) can be expressed as

$$\begin{aligned} P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(|\hat{F}|_s, |F|) &= P_{|\mathbf{F}|}(|F|) P_{\Delta|\mathbf{F}|}(|\hat{F}|_s - |F|) & |\hat{F}|_s > 0 \\ &= P_{|\mathbf{F}|}(|F|) \delta(|\hat{F}|_s) \int_{-\infty}^{-|F|} P_{\Delta|\mathbf{F}|}(x) dx & |\hat{F}|_s = 0 \\ &= 0 & |\hat{F}|_s < 0. \end{aligned} \quad (3.22)$$

Using Eqs. (3.19) and (3.22), the density $P_{\Delta|\mathbf{F}|_s}(x)$ can now be expressed in terms of $P_{|\mathbf{F}|}(x)$

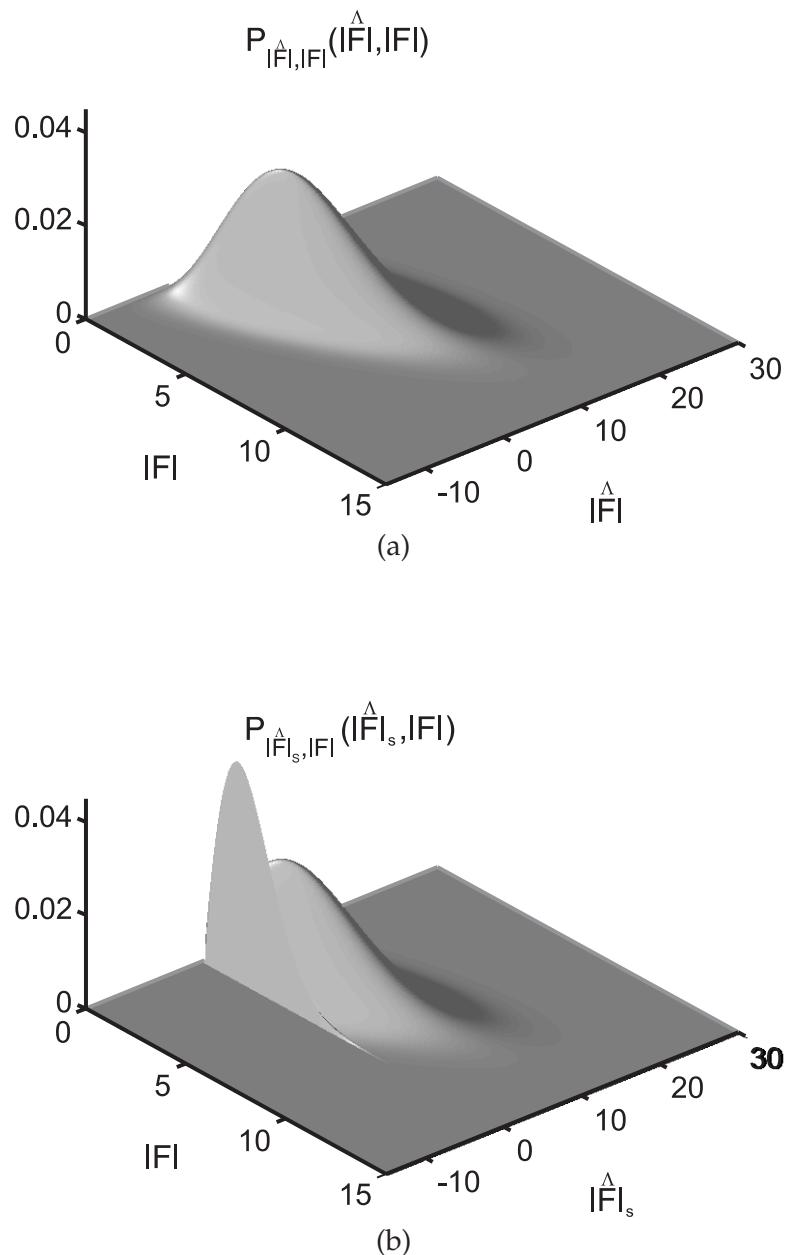


Figure 3.2: Plots of the joint density functions (a) $P_{|\hat{F}|, |F|}(|\hat{F}|, |F|)$ and (b) $P_{|\hat{F}|_s, |F|}(|\hat{F}|_s, |F|)$.

and $P_{\Delta|F|}(x)$ as

$$\begin{aligned}
 P_{\Delta|F|_s}(x) &= \int_{-\infty}^{\infty} P_{\Delta|F|_s,|F|}(x+y, y) dy \\
 &= \int_{-\infty}^{\infty} P_{|\hat{F}|,|F|}(x+y, y) dy \\
 &= \int_{-\infty}^{\infty} P_{|F|}(y) P_{\Delta|F|}(x) dy \\
 &= P_{\Delta|F|}(x), \\
 &= \left(\int_{-x^+}^{\infty} + \int_{-x^-}^{-x^+} \right) P_{|\hat{F}|_s,|F|}(x+y, y) dy \\
 &= \int_{-x^+}^{\infty} P_{\Delta|F|}(x) P_{|F|}(y) dy \\
 &\quad + \int_{-x^-}^{-x^+} P_{|F|}(y) \delta(x+y) \int_{-\infty}^{-y} P_{\Delta|F|}(z) dz dy \\
 &= P_{\Delta|F|}(x) \int_{-x^+}^{\infty} P_{|F|}(y) dy + P_{|F|}(-x) \int_{-\infty}^x P_{\Delta|F|}(z) dz \\
 &= P_{\Delta|F|}(x) [1 - Q_{|F|}(-x)] + P_{|F|}(-x) Q_{\Delta|F|}(x), \\
 &= 0
 \end{aligned} \tag{3.23}$$

$x > 0,$

$-|F|_{\max} \leq x \leq 0,$

$x < -|F|_{\max},$

where $|F|_{\max}$ is the maximum value of $|F|$, $-x^\pm = -x \pm \epsilon$ for $0 < \epsilon \rightarrow 0$, and $Q_x(x)$ is the cumulative distribution function of x . Equation (3.23) gives the required result, i.e., $P_{\Delta|F|_s}(x)$ in terms of $P_{\Delta|F|}(x)$ and $P_{|F|}(x)$.

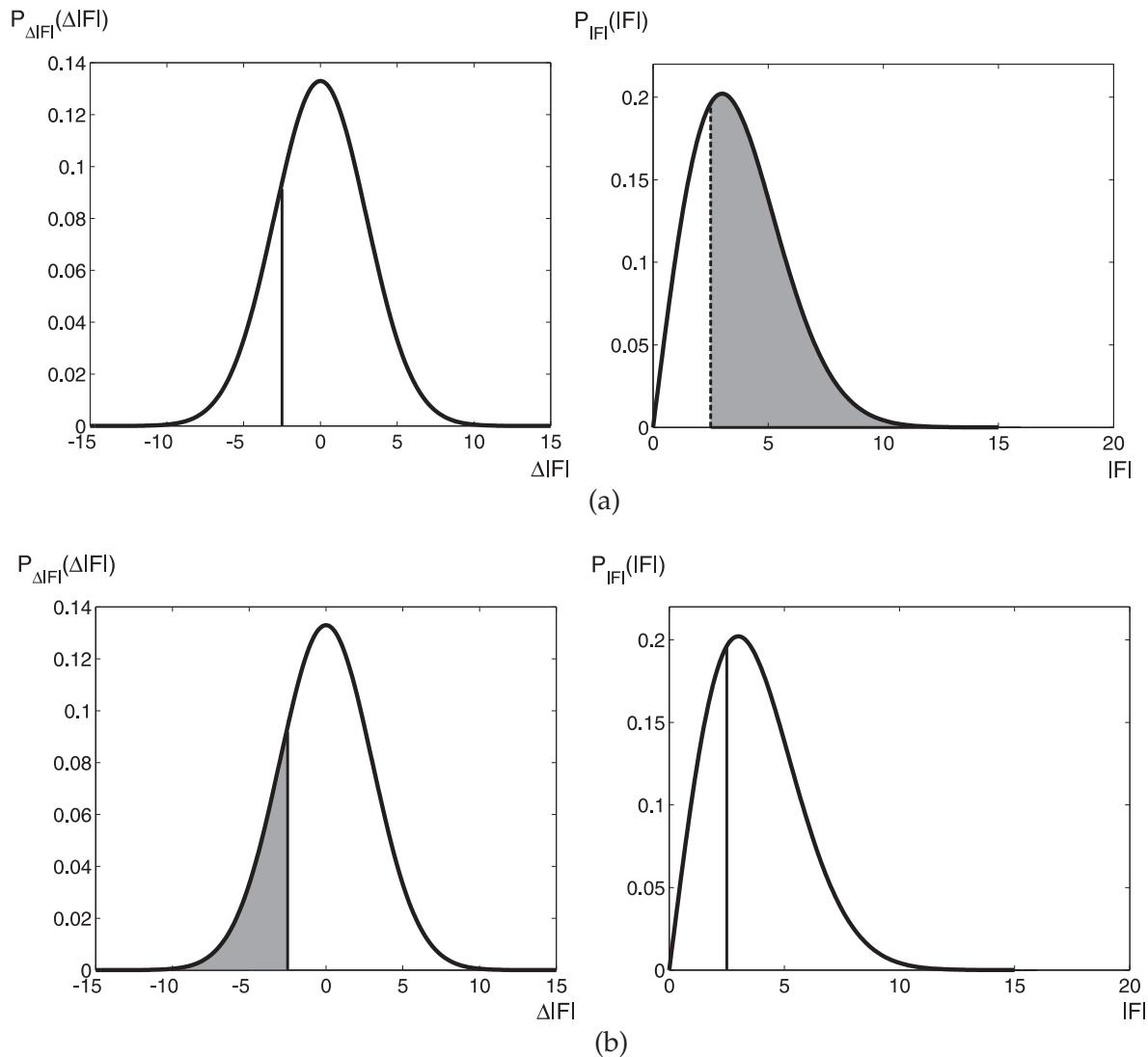
Equation (3.23) can be derived in a more heuristic fashion as follows. Equation (3.18) shows that a particular value of $\Delta|F|_s$ can be considered as the sum of two components, one that arises as a result of saturation and one that does not, i.e.,

$$P_{\Delta|F|_s}(x) = P_{\text{no sat}}(x) + P_{\text{sat}}(x). \tag{3.24}$$

When there is no saturation $\Delta|F|_s = \Delta|F|$ and referring Fig. 3.3 (a) shows that

$$\begin{aligned}
 P_{\text{no sat}}(x) &= \int_0^{\infty} P_{\Delta|F|}(x) P_{|F|}(|F| > -x) d|F|, \\
 &= P_{\Delta|F|}(x) \int_{-x^+}^{\infty} P_{|F|}(y) dy, \\
 &= P_{\Delta|F|}(x) [1 - Q_{|F|}(-x)]. \tag{3.25}
 \end{aligned}$$

Referring to Eq. (3.18), when saturation occurs $\Delta|F|_s = -|F|$, so given that $\Delta|F|_s = x$,

Figure 3.3: Contributions to $P_{\Delta|F|_s}(-2.5)$ (a) without saturation and (b) with saturation.

$|F| = -x$, and since $\Delta|F| \leq x$ for saturation and referring to Fig. 3.3 (b) shows that

$$\begin{aligned} P_{\text{sat}}(x) &= \int_{-\infty}^0 P_{|\mathbf{F}|}(-x) P_{\Delta|\mathbf{F}|}(\Delta|F| < x) d\Delta|F|, \\ &= P_{|\mathbf{F}|}(-x) \int_{-\infty}^x P_{\Delta|\mathbf{F}|}(y) dy, \\ &= P_{|\mathbf{F}|}(-x) Q_{\Delta|\mathbf{F}|}(x). \end{aligned} \quad (3.26)$$

Using Eqs. (3.24) to (3.26) gives Eq. (3.23) for $P_{\Delta|\mathbf{F}|_s}(x)$.

The effect of saturation is illustrated in Fig. 3.4 which shows the density functions for the variables involved. For demonstration purpose, the Fourier amplitude $|F|$ and amplitude error $\Delta|F|$ are represented by Rayleigh and Gaussian distributions, respectively, which are shown in Fig. 3.4(a) and Fig. 3.4(b). The densities are discretised into 50 bins and represented as histograms in Fig. 3.4. The histograms for $|\hat{F}|$ and $|\hat{F}|_s$ are generated by simulation and the values of $\Delta|F|_s$ calculated and their histogram also generated, as shown in Fig. 3.4 (c) - (e). The densities $P_{|\hat{\mathbf{F}}|}(x)$ and $P_{|\hat{\mathbf{F}}|_s}(x)$ are then calculated numerically from $P_{|\mathbf{F}|}(x)$ and $P_{\Delta|\mathbf{F}|}(x)$ using

$$P_{|\hat{\mathbf{F}}|}(x) = \int_{-\infty}^{\infty} P_{|\mathbf{F}|}(y) P_{\Delta|\mathbf{F}|}(x-y) dy, \quad (3.27)$$

and

$$\begin{aligned} P_{|\hat{\mathbf{F}}|_s}(x) &= P_{|\hat{\mathbf{F}}|}(x), & x > 0, \\ &= \left[\int_{-\infty}^0 P_{|\hat{\mathbf{F}}|}(y) dy \right] \delta(x), & x = 0, \end{aligned} \quad (3.28)$$

respectively, and are plotted as the dashed curves in Fig. 3.4 (c) and (d). $P_{\Delta|\mathbf{F}|_s}(x)$ is calculated numerically using Eq. (3.23) and is plotted as the curve in Fig. 3.4(e). The analytical results are consistent with the simulation results, thus verifying Eq. (3.23).

To calculate the α_s and β_s that are needed to calculate $\langle e'^2 \rangle$, $\langle \Delta|F|_s |F| \rangle$ and $\langle \Delta|F|_s^2 \rangle$ are required. These can now be calculated using the densities derived above.

Using the relationships derived above, $\langle \Delta|F|_s |F| \rangle$ is evaluated as follow. First, using Eq. (A.32) $\langle \Delta|F|_s |F| \rangle$ can be expressed as

$$\langle \Delta|F|_s |F| \rangle = \iint_{-\infty}^{\infty} xy P_{\Delta|\mathbf{F}|_s, |\mathbf{F}|}(x, y) dx dy. \quad (3.29)$$

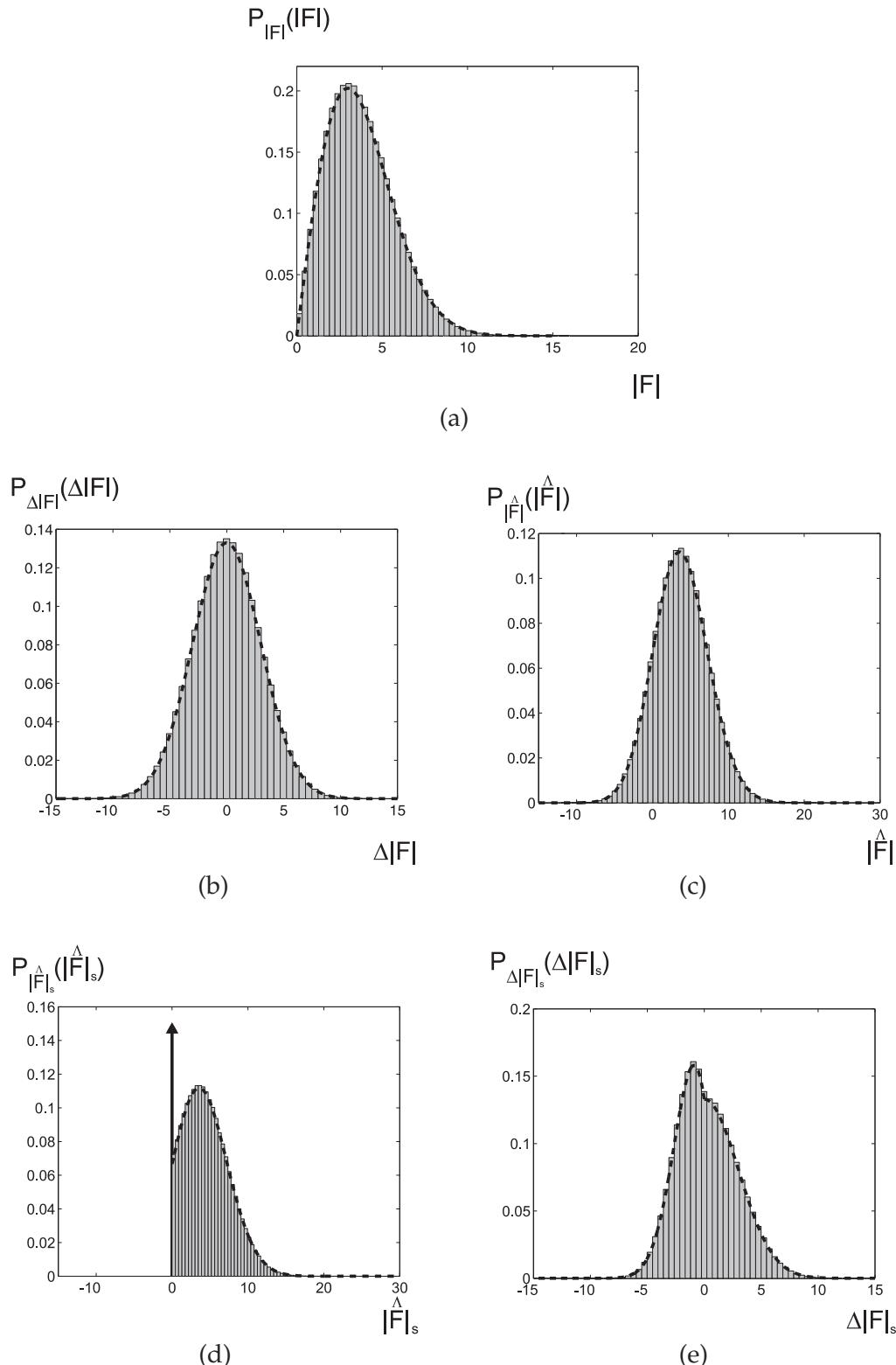


Figure 3.4: Simulation of pdfs for saturation as described in the text. The density functions (a) $P(|F|)$ (Rayleigh), (b) $P(\Delta|F|)$ (Gaussian), (c) $P(|\hat{F}|)$, (d) $P(|\hat{F}|_s)$, and (e) $P(\Delta|\hat{F}|_s)$.

Referring to Eq. (3.17), i.e., $P_{\Delta|\mathbf{F}|_s, |\mathbf{F}|}(x, y) = P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(x + y, y)$ gives

$$\begin{aligned}\langle \Delta|F|_s |F| \rangle &= \iint_{-\infty}^{\infty} xy P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(x + y, y) dx dy \\ &= \int_0^{\infty} \int_0^{\infty} (z - y) y P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(z, y) dz dy,\end{aligned}\quad (3.30)$$

since $|\hat{F}|_s > 0$ and $|F| > 0$. Using Eq. (3.22), $\langle \Delta|F|_s |F| \rangle$ is now given by

$$\begin{aligned}\langle \Delta|F|_s |F| \rangle &= \int_0^{\infty} \left[\left(\int_{0^+}^{\infty} + \int_{0^-}^{0^+} \right) (z - y) y P_{|\hat{\mathbf{F}}|_s, |\mathbf{F}|}(z, y) dz \right] dy \\ &= \int_0^{\infty} y P_{|\mathbf{F}|}(y) \left[\int_{0^+}^{\infty} (z - y) P_{\Delta|\mathbf{F}|}(z - y) dz \right] dy \\ &\quad + \int_0^{\infty} \left\{ \int_{0^-}^{0^+} (z - y) y P_{|\mathbf{F}|}(y) \delta(z) \left[\int_{-\infty}^{-y} P_{\Delta|\mathbf{F}|}(w) dw \right] dz \right\} dy \\ &= \int_0^{\infty} y P_{|\mathbf{F}|}(y) \left[\int_{-y^+}^{\infty} z P_{\Delta|\mathbf{F}|}(z) dz \right] dy - \int_0^{\infty} y^2 P_{|\mathbf{F}|}(y) \left[\int_{-\infty}^{-y} P_{\Delta|\mathbf{F}|}(w) dw \right] dy \\ &= \int_0^{\infty} y P_{|\mathbf{F}|}(y) \left[\int_{-y^+}^{\infty} z P_{\Delta|\mathbf{F}|}(z) dz - y Q_{\Delta|\mathbf{F}|}(-y) \right] dy.\end{aligned}\quad (3.31)$$

Using Eq. (3.23), $\langle \Delta|F|_s^2 \rangle$ is evaluated as

$$\begin{aligned}\langle \Delta|F|_s^2 \rangle &= \int_{-\infty}^{\infty} x^2 P_{\Delta|\mathbf{F}|_s}(x) dx \\ &= \left(\int_{0^+}^{\infty} + \int_{-\infty}^0 \right) x^2 P_{\Delta|\mathbf{F}|_s}(x) dx \\ &= \int_{0^+}^{\infty} x^2 P_{\Delta|\mathbf{F}|}(x) dx + \int_{-\infty}^0 x^2 \left[P_{\Delta|\mathbf{F}|}(x) (1 - Q_{|\mathbf{F}|}(-x)) + P_{|\mathbf{F}|}(-x) Q_{\Delta|\mathbf{F}|}(x) \right] dx \\ &= \langle \Delta|F|^2 \rangle + \int_{-\infty}^0 x^2 [P_{|\mathbf{F}|}(-x) Q_{\Delta|\mathbf{F}|}(x) - P_{\Delta|\mathbf{F}|}(x) Q_{|\mathbf{F}|}(-x)] dx.\end{aligned}\quad (3.32)$$

Equations (3.31) and (3.32) give the desired result, i.e., β_s and γ_s^2 in terms of $P_{\Delta|\mathbf{F}|}(x)$ and $P_{|\mathbf{F}|}(x)$.

Equations (3.31) and (3.32) are verified by simulation as follows. First, using a Rayleigh distribution for $|F|$ and a Gaussian distribution (with normalised variance σ_a) for $\Delta|F|$ as described above, β_s and γ_s^2 were calculated numerically using Eqs. (3.31), (3.32), (3.7) and (3.8) and are shown as the curves in Fig. 3.5 as a function of σ_a . Second, β_s and γ_s were calculated by simulation and are shown by the filled circles in Fig. 3.5. In the simulation, 250000 samples were drawn from $P(|F|)$ and $P(\Delta|F|)$, $\Delta|F|_s$ calculated by Eq. (3.18), and β_s and γ_s^2 calculated by Eqs. (3.7) and (3.8). Good agreement is evident, verifying Eqs. (3.31)

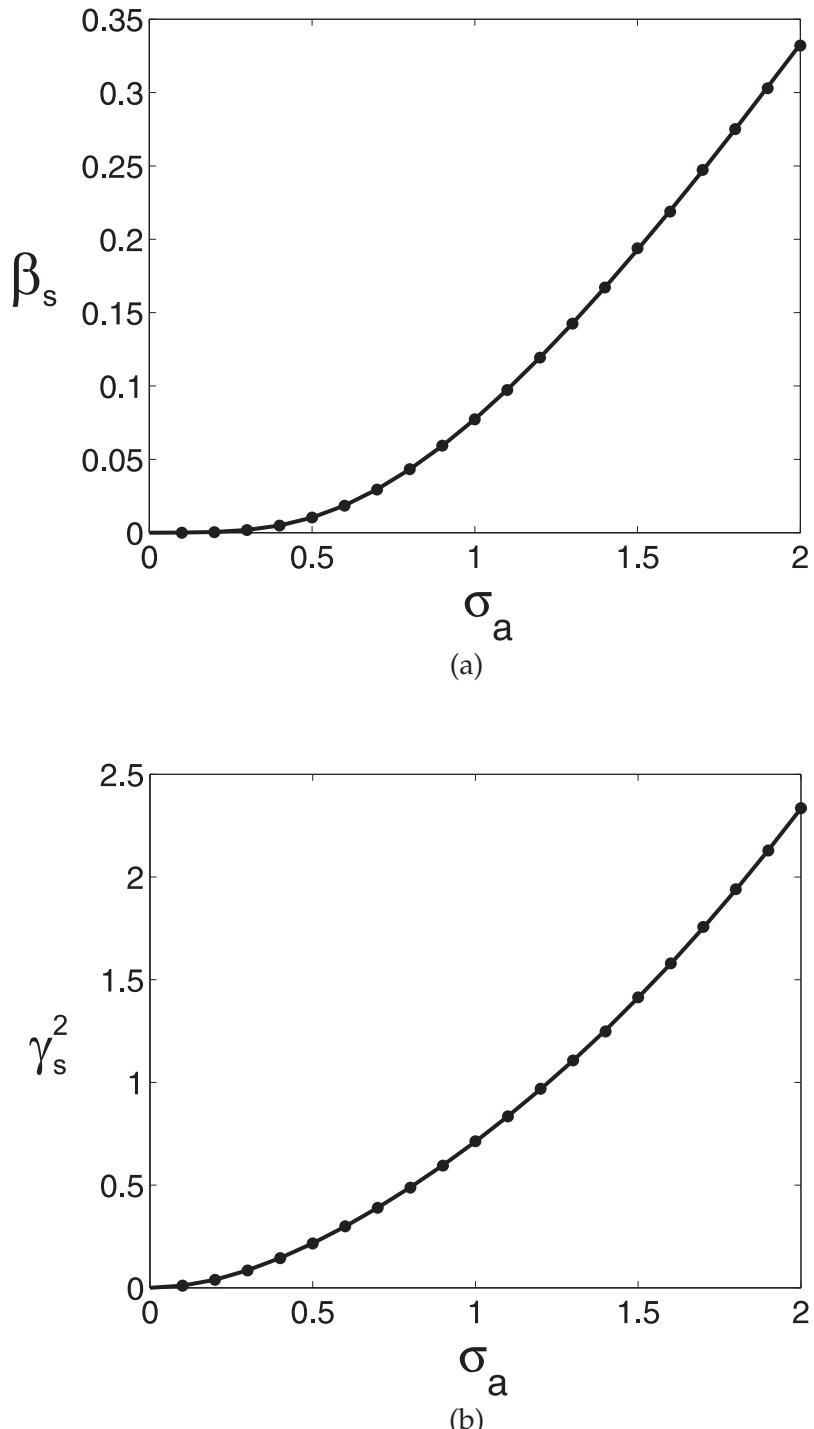


Figure 3.5: (a) β_s calculated using Eqs. (3.7) and (3.31) (—) and by simulation (·), and (b) γ_s^2 calculated using Eqs. (3.8) and (3.32) (—) and by simulation (·), as described in the text.

and (3.32).

In summary then, given an image spectral amplitude distribution $P(|F|)$ and a spectral amplitude error distribution $P(\Delta|F|)$, Eqs. (3.12), (3.11), (3.7), (3.8), (3.31) and (3.32) allow the mse $\langle e'^2 \rangle$ to be calculated in the presence of saturation. Example calculations for images are presented in the next section.

3.4 Results

The theory developed in this chapter is illustrated by both numerical evaluation of the expressions derived above and by simulation with particular images. For small amplitude errors, the results presented in Section 2.4 allow relationship between $\langle e^2 \rangle$ and σ_a and σ_ϕ to be calculated that depend only on the distribution of phase errors. However, for the case of large amplitude errors, the relationship between $\langle e'^2 \rangle$ and σ_a and σ_ϕ depends also on the distribution of the amplitude errors and the distribution of spectral amplitudes of the original image. Therefore, in order to present a set of results, the same image as that used in Chapter 2 (Fig. 2.6) and normally distributed amplitude errors are used. The spectral amplitude distribution for this image is shown in Fig. 3.6. Uniformly distributed phase errors are used as this allows the image error to be presented as a function of $\sigma_{\phi u}$. As shown in Chapter 2, the effect of normally distributed phase errors is very similar to that of uniformly distributed errors if $\sigma_{\phi n}$ is replaced by $\tilde{\sigma}_{\phi n}$ (Eq. (2.22) or Fig. 2.2 (b)).

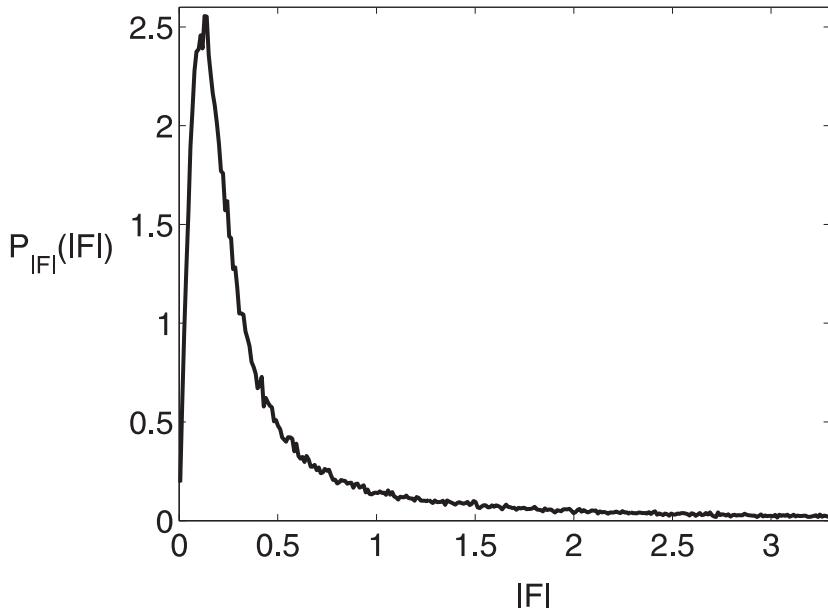


Figure 3.6: Spectral amplitude distribution (the top 5% of the amplitudes are not displayed) of the image used (Fig. 2.6) for the simulations.

The effect of amplitude errors only is addressed first. Using $P(\Delta|F|)$ and $P(|F|)$ as given above, Eqs. (3.12), (3.31), (3.32), (3.7) and (3.8) are used to calculate $\langle e'^2 \rangle$ for $0 < \sigma_a < 2$ and $\sigma_\phi = 0$ and the results are shown as the solid line in Fig. 3.7(a). The relationship Eq. (2.12) between $\langle e^2 \rangle$ and σ_a is shown by the dashed line in Fig. 3.7(a) for comparison. As anticipated, $\langle e^2 \rangle$ overestimates the error $\langle e'^2 \rangle$, and this becomes significant if $\sigma \gtrsim 0.3$. The errors were also calculated by simulation as described in Section 2.4 using the image in Fig. 2.6 with saturation incorporated and the error $\langle e'^2 \rangle$ calculated by scaling the reconstructed images to the original image. The image mses were calculated and averaged over 10 sets of noise signals and are shown as the filled circles in Fig. 3.7(a). The simulation results match the theoretical values well. Note that for phase errors only, scaling and saturation have no effect so that the relationship between $\langle e'^2 \rangle$ and σ_ϕ is the same as for $\langle e^2 \rangle$ and σ_ϕ (Fig. 2.3(b)).

The relationship Eq. (3.13) that describes equivalent phase and amplitude errors can written in the form

$$\alpha_s(1 + \beta_s) = \frac{\sin(\sqrt{3}\sigma_{\phi u})}{\sqrt{3}\sigma_{\phi u}} \quad (3.33)$$

for uniformly distributed phase errors. This relationship was calculated in terms of σ_a and $\sigma_{\phi u}$ for the distributions $P(\Delta|F|)$ and $P(|F|)$ described above as follows. For each value of σ_a , α_s and β were calculated using the equations in the previous section and the lhs of Eq. (3.33) matched with values of the rhs calculated using values of $\sigma_{\phi u}$. The resulting relationship is shown as the dashed line in Fig. 3.7(b). The equivalent relationship based on $\langle e^2 \rangle$ is shown by the solid line in Fig. 3.7(b), and is seen to underestimate the equivalent amplitude errors. The figure shows that, in fact, very large amplitude errors are required to mimic the effect of large phase errors.

To evaluate the effects of both amplitude and phase errors on the reconstructed images, images were generated with different mixtures of amplitude and phase errors to give the same value of $\langle e'^2 \rangle$ in a similar way as described in Chapter 2 (Fig. 2.6). The results are shown in Fig. 3.8. The original image is the same as shown in the left column of Fig. 2.6. The images in each column of Fig. 3.8 have the same value of $\langle e'^2 \rangle$, which increases from left to the right with the values 0.25, 0.5, 0.75, 1, and 1.5. The images in the top and bottom rows contain amplitude errors only ($\sigma_\phi = 0$) and phase errors only ($\sigma_a = 0$), respectively. As noted in Section 3.2, when both amplitude and phase errors are present it is not possible to treat the error as the sum of two terms that are due to the amplitude and phase errors only. This is because the term $e_\phi'^2$ in Eq. (3.12) contains the effects of both amplitude and phase errors. It is therefore not possible to generate images rigorously analogous to those in the center row of Fig. 2.6. Two sets of images with both amplitude and phase errors were generated as follows. The first set of images is generated by determining $\sigma_{\phi u}$ from

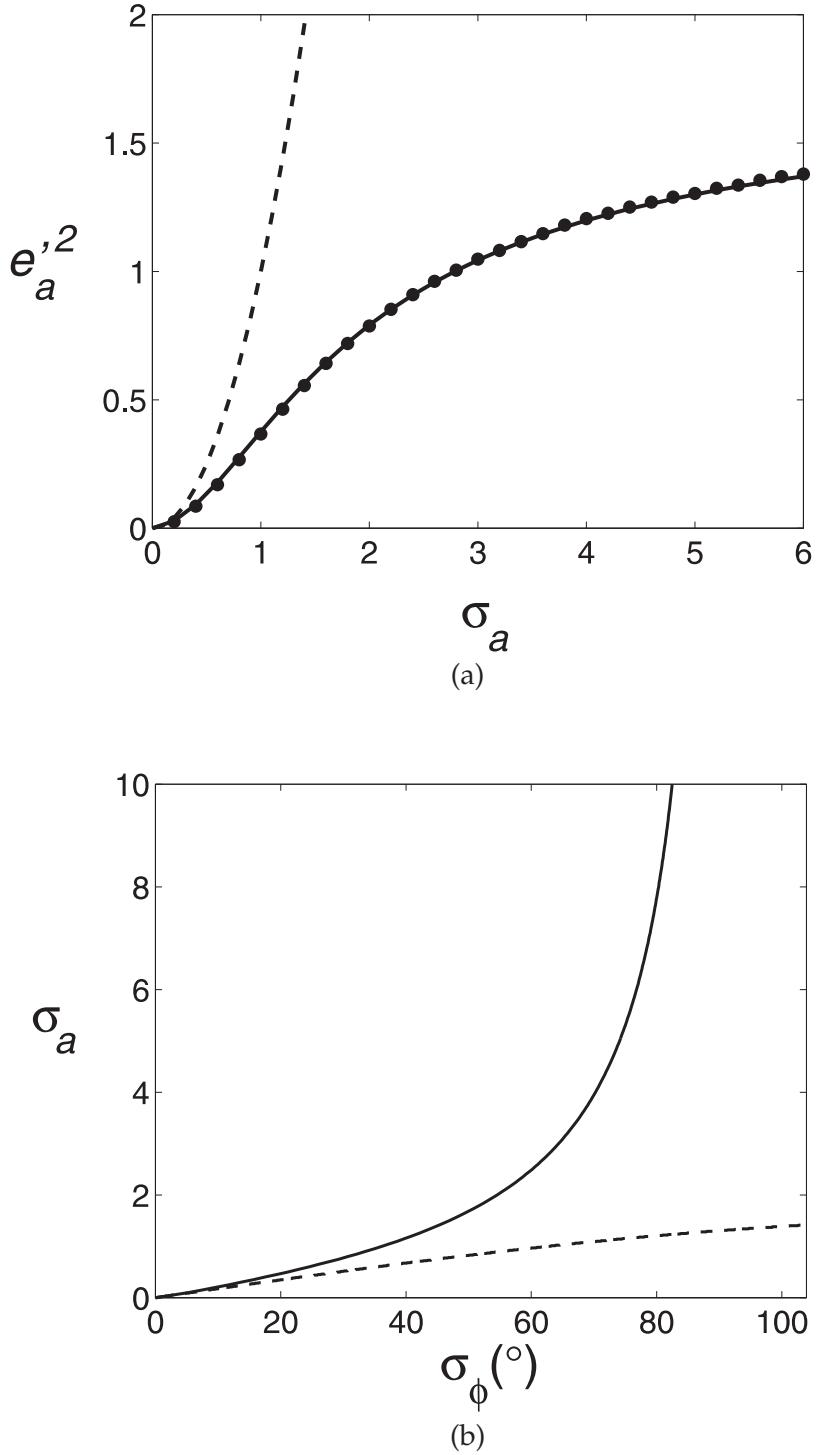


Figure 3.7: (a) The mse $\langle e'^2 \rangle$ versus σ_a for normally distributed amplitude errors (for $\sigma_\phi = 0$) calculated using the spectral amplitude distribution shown in Fig. 3.6 and the equations in the previous section (—). The filled circles show the results of simulations as described in the text. The dashed line shows $\langle e^2 \rangle$ versus σ_a using Eq. (2.10) (for $\sigma_\phi = 0$). (b) Equivalent normally distributed amplitude errors and uniformly distributed phase errors based on $\langle e'^2 \rangle$ calculated using Eq. (3.13) (—), and based on $\langle e^2 \rangle$ using Eq. (2.25) (— —).



Figure 3.8: Images reconstructed with a variety of normally distributed amplitude errors and uniformly distributed phase errors as described in the text. The mse $\langle e'^2 \rangle$ is identical in each column with the values 0.25, 0.5, 0.75, 1 and 1.5, from left to right. The top row is for amplitude errors only and the bottom row is for phase errors only. The middle two rows are for both amplitude and phase errors as described in the text.

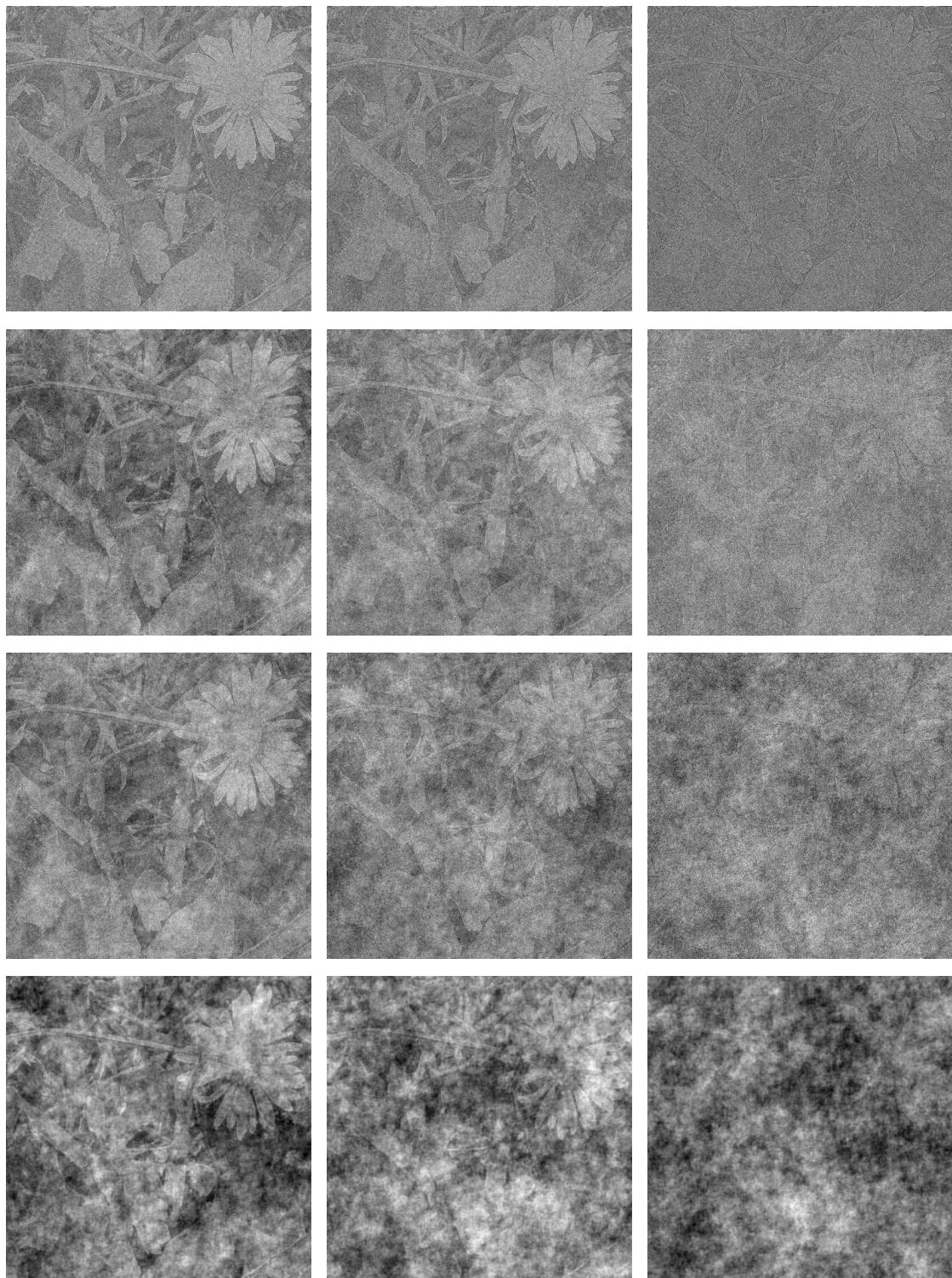


Figure 3.8: (continued).

$2(1 - \langle \cos(\Delta\phi) \rangle) = \langle e'^2 \rangle / 2$ (i.e., solving Eq. (3.12) for $\sigma_{\phi u}$ with $\sigma_a = 0$ for an error $\langle e'^2 \rangle / 2$), and then, using this value of $\sigma_{\phi u}$, solving Eq. (3.12) for σ_a with an error of $\langle e'^2 \rangle$. The resulting images are shown in the second to top row in Fig. 3.8. For the second set of images, σ_a is determined from $2[1 - \alpha(1 + \beta)] = \langle e'^2 \rangle / 2$ (i.e., Eq. (3.12) is solved for σ_a with $\sigma_{\phi u} = 0$ with the left hand side (lhs) equal to $\langle e'^2 \rangle / 2$), and the value of σ_a is then substituted into Eq. (3.12) which is solved for $\sigma_{\phi u}$ with the lhs equal to $\langle e'^2 \rangle$. This results in $e'^2 = e_a'^2$. The resulting images are shown in the third to top row in Fig. 3.8. Inspection of the two center rows in Fig. 3.8 shows that although the images are similar, the upper images have a more amplitude error character and the lower images a more phase error character. Reference to Eq. (3.12) shows the reason for this. For the lower images, a larger value of $(1 - \langle \cos(\Delta\phi) \rangle)$ is required than for the upper row since in the latter case the amplitude error as well contributes to e'^2 . Inspection of Fig. 3.8 shows that the images in a single column (a fixed value of $\langle e'^2 \rangle$) are of quite similar quality although the images with (large) phase errors are slightly worse, depending on what is important to the observer. For the smaller errors, these results show that $\langle e'^2 \rangle$ corrects somewhat for the observation from Fig. 2.6 that phase errors give a slightly larger perceived error than do amplitude errors, due to $\langle e^2 \rangle$ overestimating the error due to amplitude errors. Overall then, the model described here is a realistic model of spectral amplitude and phase errors of arbitrary magnitude. The results show that $\langle e'^2 \rangle$ is a reasonably good metric of image quality, and that amplitude and phase errors have similar effects on the overall quality of reconstructed images.

The image differences are calculated for a subset of the reconstructed images in Fig. 3.8 in order to more clearly see the form of the errors and are shown in Fig. 3.9. They are displayed in the same way as in Fig. 2.7. The same kinds of characteristics are seen as in Fig. 3.8. Amplitude errors tends to affect the whole image uniformly and reduce the contrast and the errors fluctuate rapidly from pixel to pixel. For large amplitude errors however, some structure is seen in the difference image. Inspection of the figure shows that the errors still fluctuate rapidly from pixel to pixel. However, on a larger scale the errors tend to take the opposite sign to those of the original image. This is a result of the overall contrast tending to zero for large amplitude errors. For phase errors, the results are identical to those shown in Fig. 2.7. Random objects rather than points appear in the reconstructed image as the phase error increases, and the errors are more correlated from pixel to pixel and vary slowly over the image. Although less obvious, the contrast decreases when the phase errors are large.

It is informative to study the behaviour of $\langle e^2 \rangle$ and $\langle e'^2 \rangle$ for large amplitude and phase errors. Referring to Eq. (3.17), $|F| + \Delta|F|_s \geq 0$ so that

$$|F|\Delta|F|_s \geq -|F|^2, \quad (3.34)$$

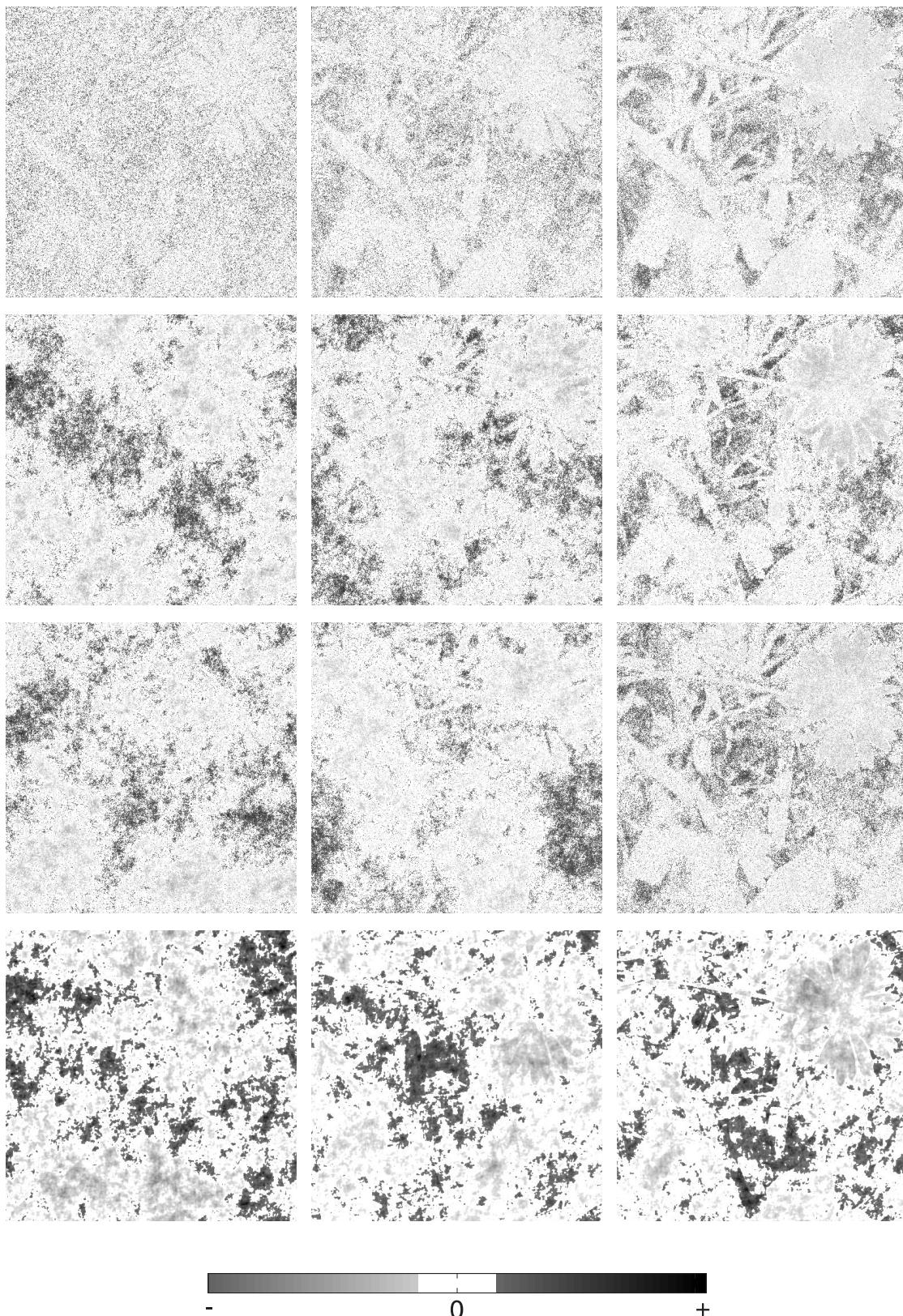


Figure 3.9: The image differences $\Delta f(x, y)$ (Eq. (2.29)) for the images shown in Fig. 3.8 with $\langle e'^2 \rangle$ values of 0.25, 0.75 and 1.5, arranged in the same way as in Fig. 3.8.

and taking the expected value over the noise signals and integrating over u and v gives

$$\frac{\langle |F| \Delta |F|_s \rangle}{|R|^{-1} \iint_R |F(u, v)|^2 du dv} \geq -1 \implies \beta_s \geq -1. \quad (3.35)$$

Equations (3.12) and (3.35) then show that $\langle e'^2 \rangle$ is bounded by $\langle e'^2 \rangle < 2$ for amplitude errors only. This is a key advantage over $\langle e^2 \rangle$ which is unbounded for large σ_a , since these are both normalised errors. For phase errors only that are zero-mean, the largest value of $\langle e'^2 \rangle$ occurs for uniformly distributed phase errors, and evaluation of Eq. (3.12) shows that $\langle e'^2 \rangle$ (as well as $\langle e^2 \rangle$) is also bounded by $\langle e'^2 \rangle \leq 2$ (see Fig. 2.3(b)). Therefore, both amplitude and phase errors give the same maximum mse, although very large amplitude errors would be required. Note that values of $\langle e'^2 \rangle$ larger than 2 can be obtained but this would require an uncentered phase error distribution. Such a phase bias would be unusual, and if present would normally be removed as part of some preprocessing.

The relationship between e'_a and σ_a depends on the amplitude error distribution. To assess this dependence, the relationship was calculated for both normally and uniformly distributed amplitude errors for the image shown in Fig. 2.6 using Eq. (3.12) and the results are shown as the two curves in Fig. 3.10 as a function of σ_a . The errors were also calculated by simulation (averaged over 10 noise signals) and are shown as the symbols in Fig. 3.10. The results show that the relationship between $\langle e'^2 \rangle$ and σ_a is quite insensitive to the form of the distribution of amplitude errors.

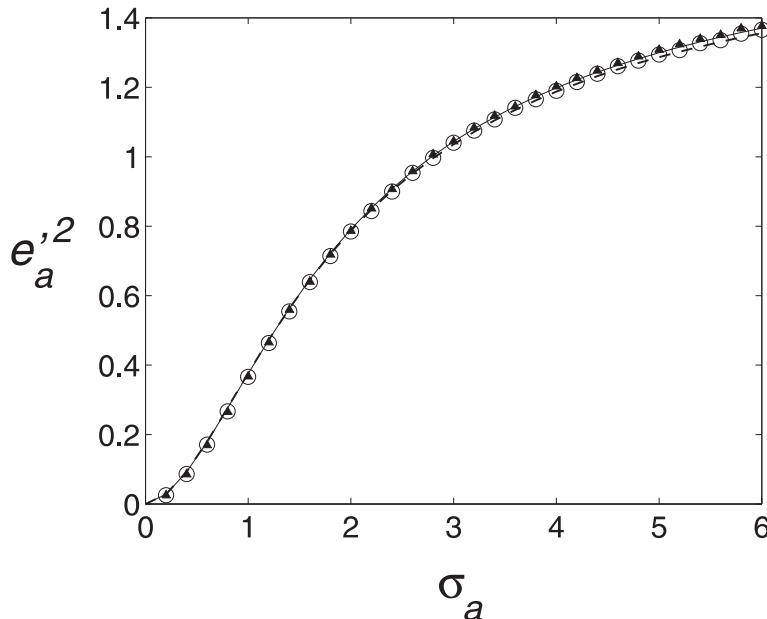


Figure 3.10: The mse e'_a^2 versus σ_a for normally (—) and uniformly (---) distributed amplitude errors calculated using Eq. (3.12). Simulation results are shown by \blacktriangle and \circ for normal and uniform errors, respectively.

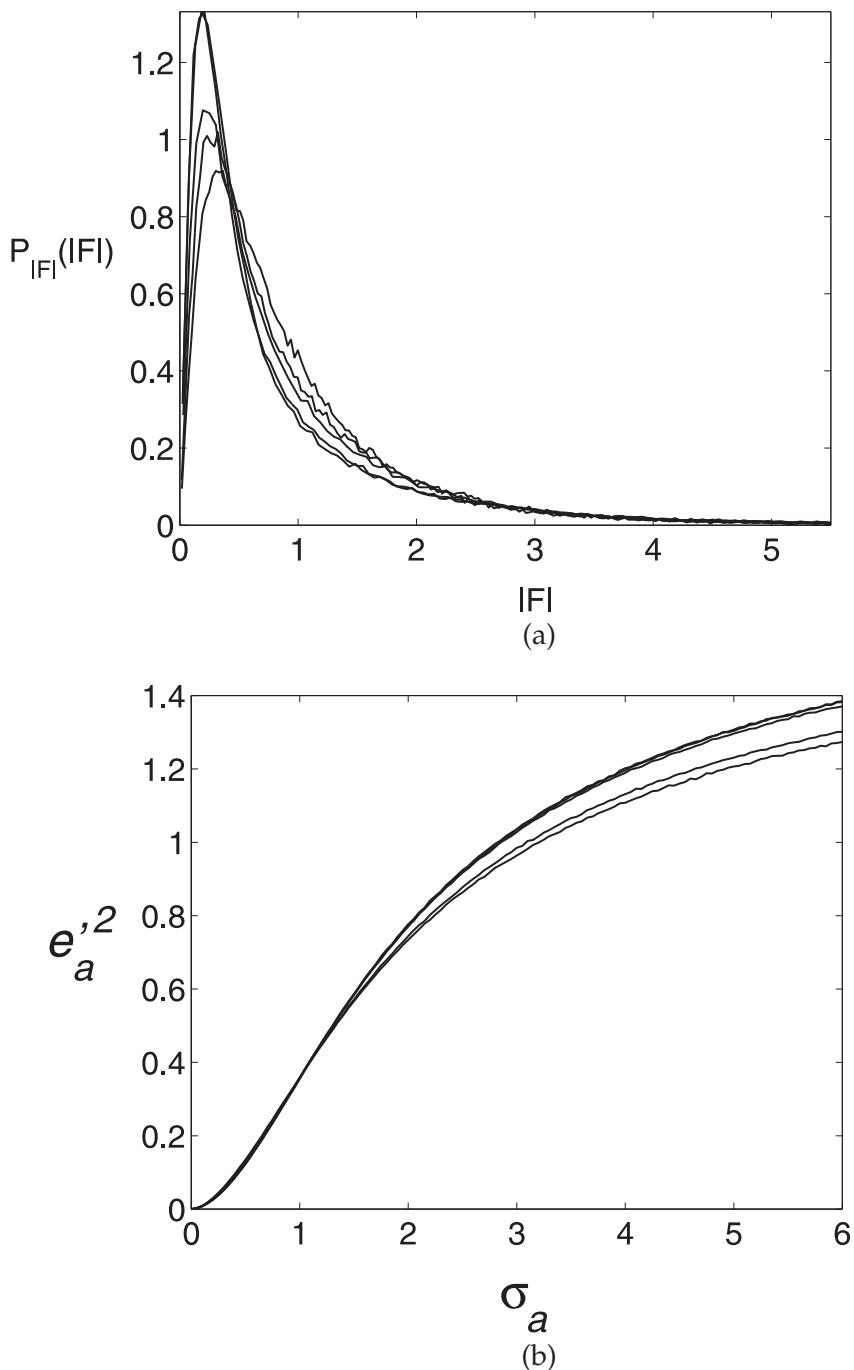


Figure 3.11: (a) Spectral amplitude distributions of the 5 images (each image normalised such that the mean spectral amplitude is 1 and the top 5% of the amplitudes are not displayed), and (b) the relationship between $e_a'^2$ and σ_a for the 5 images.

The relationship between $e_a'^2$ and σ_a also depends on the image spectral amplitude distribution. To assess this dependence, the relationship was calculated for five different images using Eq. (3.12) and normally distributed amplitude errors. The amplitude distributions for the five images are shown in Fig. 3.11 (a) and the corresponding plots of $e_a'^2$ versus σ_a are shown in Fig. 3.11 (b). Inspection of the figure shows that although there are variations in the relationship for large σ_a , the variations are not particularly large.

3.5 Discussion

The error metric $\langle e'^2 \rangle$ developed here corrects for the effects of added noise energy when large amplitude errors are present and is a good representation of the error in most anticipated technical and visual applications. Scaling tends not to affect the contrast or structure in the image. Therefore, a mean square error minimised over contrast independent scaling would appear to be an appropriate metric to compare images in terms of visual interpretability. The model developed here correctly represents the effects of large amplitude errors and saturation by a generalisation of the model presented in Chapter 2. As a result of saturation, the noise model is nonlinear (non-additive). An expression for the image error is obtained by mapping to an equivalent additive error. The error depends on, in addition to the magnitude of the amplitude and phase errors, the distributions of the amplitude and phase errors, and the spectral amplitude distribution of the original image. The expression for the image mse is a complicated function of the distributions of the amplitude errors and the spectral amplitudes of the original image. These expressions are verified by numerical simulation. Calculations of the errors for a variety of images and error distributions show that the dependence on these distributions is quite weak however for typical images. Therefore, the relationship between the image mse $\langle e'^2 \rangle$ and the variance of the amplitude errors σ_a (Fig. 3.7 (a)) can be taken as generally applicable. Likewise, the relationship between amplitude errors σ_a and phase errors σ_ϕ that produce the same image error (Fig. 3.7 (b)) can also be taken as generally applicable. This relationship shows, in particular, that very large amplitude errors are required to produce the same effect as phase errors with standard deviations greater than about 60° . The effects of amplitude and phase errors are similar to those observed in Chapter 2. Amplitude errors produce image errors that are uncorrelated from pixel to pixel and reduce the image contrast. Phase errors produce image errors that are spatially correlated and tend to introduce visual artifacts in reconstructed images.

Chapter 4

Image Fourier amplitude distributions

4.1 Introduction

In the previous chapter, the probability density function of the image Fourier amplitudes took on significance with respect to the effect of large amplitude errors. In the next chapter it will be seen that the image Fourier amplitude distribution is also important in quantitating phase dominance.

The circularly averaged power spectrum of natural images studied in Chapter 6 describes how the mean of the spectral amplitudes over angle varies with radius in Fourier space. Those results are relevant to visual processing, the special characteristics of natural images, and have some relevance for topics such as image compression. The distribution of image Fourier amplitude is therefore of significance in a variety of problems in Fourier imaging and image processing.

A knowledge of the form of pdf of signals is useful in a variety of signal and image processing applications. Signal and image pdfs are important in optimising operations such as coding, compression and quantisation. For example, in image compression using the discrete cosine transform (DCT), the pdf of the DCT coefficients is used to design the quantiser for particular compression-quality tradeoffs [96, 97]. Similarly, the pdf of speech signal amplitudes has been the subject of numerous studies with applications in coding, recognition and enhancement [98, 99]. Furthermore, calculation of Bayesian estimates often uses of image pdfs as a prior model [100]; for example, in remote sensing, image recovery [101], and x-ray crystallography [102].

Two approaches to determining signal pdfs are generally used. In one, standard probability distributions are tried and selected based on their goodness of fit to sample distribu-

tions. This approach has been used to determine pdfs for image DCT coefficients [97, 103, 104] and speech signal amplitudes [99, 105]. Gaussian, Laplacian, generalised Gaussian, and generalised Gamma distributions are commonly used. Distributions with more free parameters fit sample distributions better, and an important question is whether the improved fit is sufficient to justify the increased number of parameters. Distributions determined in this way can be satisfactory for many purposes, but they are empirical in nature and not based on any underlying model of the image. The second approach is to derive the form of the pdf from a model of the structure of the image. This approach has a number of advantages in that it embodies the underlying physics of the image formation, and parameters of the distribution may be related directly to parameters or characteristics of the image. It has been used, for example, to derive distributions of synthetic aperture radar image amplitudes based on the physics of the imaging process [106, 107]. This latter approach is used here to derive a model for the Fourier amplitude distribution of images.

A model for the pdf of image spectral amplitudes is derived, in this chapter, based on a simple model of images. This provides a more satisfying model than do currently used *ad hoc* distributions that are not based on a model for the image but are simply a parameter fit to various standard probability density functions. The analysis also provides a model for the spectral amplitude pdf as a function of spatial frequency, as well as some insights into the effect of correlations in images on the distribution of spectral amplitudes.

Using a simple model consisting of independent, nonoccluding objects and incorporating the power spectrum behaviour described in Chapter 6, an analytical form of the Fourier amplitude distribution is derived in Section 4.2. A comparison of this distribution with data from various images is presented in Section 4.3. The results are discussed in Section 4.4.

4.2 Spectral amplitude distributions

Consider an image $f(x, y)$ and its Fourier transform $F(u, v)$, which is decomposed into the amplitude $|F(u, v)|$ and phase $\phi(u, v)$ as defined previously. The objective here is to derive a suitable distribution for the Fourier amplitude. The amplitude $|F(u, v)|$ is treated for a particular image for various values of (u, v) as a random variable denoted A , and the desired probability density function is denoted $P(A)$.

A commonly used model of images of the natural environment is a collage of occluding objects with a wide range of sizes and shapes [79, 82]. Furthermore, it is shown in Chapter 6 that the property of occlusion has little effect on the power spectral behaviour. Therefore, an image model consisting of nonoccluding objects is used here due to the simplicity that it imparts. Consider an image that is made up of a large number N of nonoccluding objects

randomly placed in the image space, i.e.,

$$f(x, y) = \sum_{n=1}^N f_n(x - x_n, y - y_n), \quad (4.1)$$

where $f_n(x, y)$ is the amplitude of the n -th object (when positioned at the origin), that is located at position (x_n, y_n) in the image. The Fourier transform of the image is then given by

$$F(u, v) = \sum_{n=1}^N F_n(u, v) \exp(i2\pi(ux_n + vy_n)), \quad (4.2)$$

where $F_n(u, v)$ is the Fourier transform of $f_n(x, y)$.

Consider first the case where the objects are circularly symmetric, i.e., $f_n(x, y)$ is a function only of $\sqrt{x^2 + y^2}$. The Fourier transform $F_n(u, v)$ then depends only on $\rho = \sqrt{u^2 + v^2}$, where (ρ, φ) are the polar coordinates in Fourier space, so that

$$F(u, v) = \sum_{n=1}^N F_n(\rho) \exp(i2\pi(ux_n + vy_n)). \quad (4.3)$$

For a given (u, v) the right-hand-side of Eq. (4.3) is a 2-D random walk of N steps in the complex plane. Each step has length $|F_n(\rho)|$ and angle (direction) $\varphi_n(u, v) = \phi\{F_n(\rho)\} + 2\pi(ux_n + vy_n)$. The real part of Eq. (4.3) is

$$\text{Real}\{F(u, v)\} = \sum_{n=1}^N |F_n(\rho)| \cos(\varphi_n(u, v)). \quad (4.4)$$

Since the positions (x_n, y_n) are uniformly distributed, unless (u, v) is very small, so is $\varphi_n(u, v)$. Therefore, $\langle \cos(\varphi_n) \rangle = 0$ and $\text{var}\{\cos(\varphi_n)\} = 1/2$, where $\text{var}\{\cdot\}$ denotes the variance. By the central limit theorem (Section A.2.11), for large N , $\text{Real}\{F(u, v)\}$ is normally distributed with mean zero and variance

$$\sigma^2(\rho) = \frac{1}{2N} \sum_{n=1}^N |F_n(\rho)|^2. \quad (4.5)$$

The same applies to the imaginary part of $F(u, v)$. Referring to Section A.2.10, for fixed ρ , the Fourier amplitude is therefore Rayleigh distributed, i.e.,

$$P(\rho; A) = \frac{A}{\sigma^2(\rho)} \exp(-A^2/2\sigma^2(\rho)). \quad (4.6)$$

Note that if all the objects are point-like with amplitudes b_n , i.e.,

$$f(x, y) = \sum_{n=1}^N b_n \delta(x - x_n, y - y_n), \quad (4.7)$$

where $\delta(x, y)$ is the two-dimensional delta function, then the Fourier transform is

$$F(u, v) = \sum_{n=1}^N b_n \exp(i2\pi(ux_n + vy_n)), \quad (4.8)$$

and the Fourier amplitude is Rayleigh distributed over *all* Fourier space, i.e.,

$$P(A) = \frac{A}{\sigma^2} \exp(-A^2/2\sigma^2), \quad (4.9)$$

where $\sigma^2 = (1/2N) \sum_{n=1}^N b_n^2$.

Most images, of course, do not contain only circularly symmetric objects. However, in most images, objects occur in many orientations. Even in images that have a predominance of features with particular orientations, there are generally a reasonable number of different such orientations. In addition, the Fourier transform averages over all the objects in an image, and the amplitude distribution over all angles φ in Fourier space is of interest. Hence, except in the most extreme cases, the effect of the asymmetry of individual objects on the amplitude distribution is expected to be extremely small. Therefore, it is assumed, to a first approximation, that Eq. (4.6) applies to general images.

Using Eq. (4.6) shows that for fixed ρ , the Fourier intensity $I = A^2$ is exponentially distributed, i.e.,

$$P(\rho; I) = \frac{1}{2\sigma^2(\rho)} \exp(-I/2\sigma^2(\rho)). \quad (4.10)$$

Using Eq. (4.10), the circularly averaged spatial power spectrum $\langle I(\rho) \rangle_\varphi$ of the image as a function of spatial frequency ρ is calculated and given as

$$\begin{aligned} \langle I(\rho) \rangle_\varphi &= \int_{-\infty}^{\infty} I P(\rho; I) dI = \int_0^{\infty} \frac{I}{2\sigma^2(\rho)} \exp\left(\frac{-I}{2\sigma^2(\rho)}\right) dI \\ &= 2\sigma^2(\rho). \end{aligned} \quad (4.11)$$

Equation (4.11) relates the circularly averaged power spectrum $\langle I(\rho) \rangle_\varphi$ to the ρ -dependence of the variance in Eq. (4.6). For natural images, the former is shown in Chapter 6 to take the form

$$\langle I(\rho) \rangle_\varphi = k\rho^{-\gamma}, \quad (4.12)$$

where k is a scaling constant and $\gamma > 0$ is constant for individual images and varies between approximately 1 and 3. This property, together with the relationship (Eq. (4.11)),

allows the ρ -dependence of the variance $\sigma^2(\rho)$ to be modelled as

$$\sigma^2(\rho) = \frac{k}{2} \rho^{-\gamma}. \quad (4.13)$$

This then gives a model of the intensity distribution as

$$P(\rho; I) = \frac{\rho^\gamma}{k} \exp(-\rho^\gamma I/k), \quad (4.14)$$

where γ and k are parameters. The density function for the amplitude is then

$$P(\rho; A) = \frac{2A\rho^\gamma}{k} \exp(-\rho^\gamma A^2/k). \quad (4.15)$$

The distribution $P(A)$ can now be calculated as follows. The distribution of intensities $P(I)$ in the region $\rho_1 < \rho < \rho_2$ of Fourier space is given by

$$\begin{aligned} P(I) &= \int_{\rho_1}^{\rho_2} P(\rho; I) 2\pi\rho d\rho / \int_{\rho_1}^{\rho_2} 2\pi\rho d\rho \\ &= \frac{2}{k(\rho_2^2 - \rho_1^2)} \int_{\rho_1}^{\rho_2} \rho^{\gamma+1} \exp(-I\rho^\gamma/k) d\rho. \end{aligned} \quad (4.16)$$

It is necessary to impose limits on ρ due to divergences at $\rho = 0$ and $\rho = \infty$. Substituting $t = I\rho^\gamma/k$ into Eq. (4.16) gives

$$\begin{aligned} P(I) &= \frac{2k^{2/\gamma}}{\gamma(\rho_2^2 - \rho_1^2) I^{1+2/\gamma}} \int_{I\rho_1^\gamma/k}^{I\rho_2^\gamma/k} t^{2/\gamma} \exp(-t) dt \\ &= \frac{2k^{2/\gamma}}{\gamma(\rho_2^2 - \rho_1^2) I^{1+2/\gamma}} \left[\Gamma\left(1 + \frac{2}{\gamma}, \frac{\rho_1^\gamma I}{k}\right) - \Gamma\left(1 + \frac{2}{\gamma}, \frac{\rho_2^\gamma I}{k}\right) \right], \end{aligned} \quad (4.17)$$

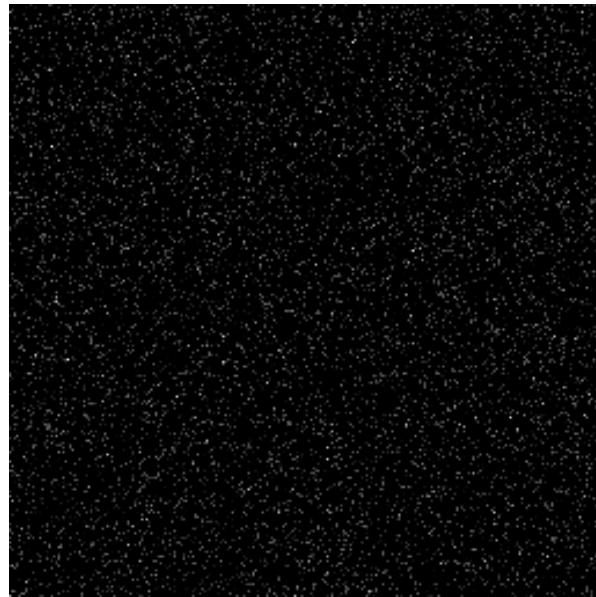
where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function (Section A.3.3). The density function for the amplitude is then given by

$$P(A) = \frac{2(c-1)A^{1-2c}}{(b^{c-1} - a^{c-1})} [\Gamma(c, aA^2) - \Gamma(c, bA^2)], \quad (4.18)$$

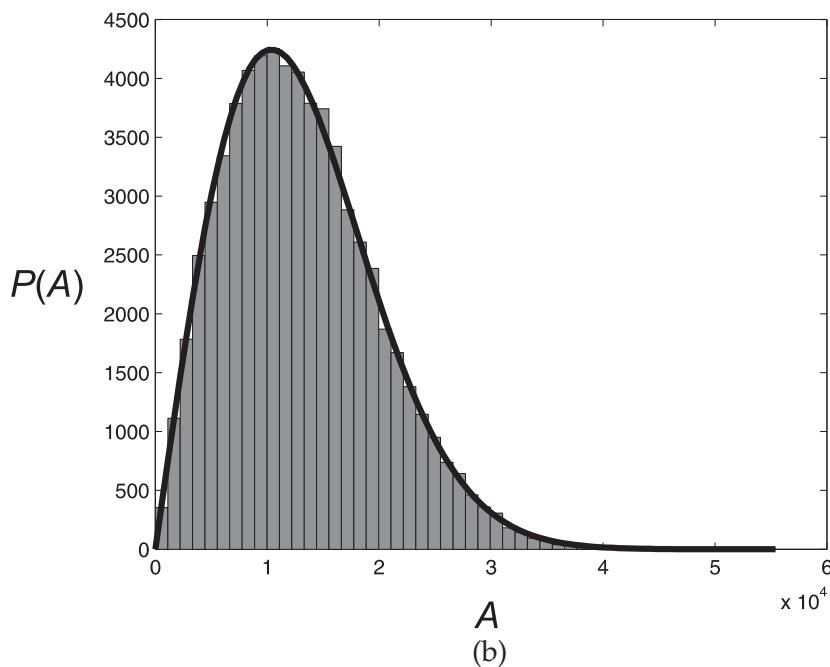
where $a = \rho_1^\gamma/k$, $b = \rho_2^\gamma/k$ and $c = 1 + 2/\gamma$. Equation (4.18) is then the desired model of the density function $P(A)$ with parameters (a, b, c) .

4.3 Results

Distribution functions are calculated here for various images to illustrate the theory derived in the previous section. To illustrate the Rayleigh distribution of amplitudes for an image composed of point-like objects, Eq. (4.9), a 256×256 pixel image with 10000 point



(a)



(b)

Figure 4.1: (a) A sample random point image, and (b) its amplitude histogram and the Rayleigh density function Eq. (4.9) (—).

objects with amplitudes chosen randomly from a uniform distribution and randomly positioned in the image was generated. An example of such an image is shown in Fig. 4.1 (a), and the histogram of its Fourier amplitudes is shown in Fig. 4.1 (b). The Rayleigh distribution, Eq. (4.9), with appropriate scaling, is also shown in Fig. 4.1 (b) and is seen to match the histogram almost exactly.

Although the final result derived above is a model of the spectral amplitude distribution $P(A)$ over all of Fourier space, an intermediate result is the amplitude distribution as a function of spatial frequency $P(\rho; A)$, given by Eq. (4.15). This latter density function could be useful, for example, as a prior distribution in image reconstruction problems. The distribution Eq. (4.15) was therefore compared with data from the image shown in Fig. 4.2. The image is 512×512 pixels and the DFT was calculated and partitioned into 10 annuli each of 20 spatial frequency samples in width, and the data within each annulus divided into 50 amplitude bins. The center radius of each annulus is denoted $\bar{\rho}$, and the histogram of the Fourier amplitudes A in each annulus, $h(\bar{\rho}; A)$, is shown in Fig. 4.3 (a). Since the number of samples in each annulus is proportional to $2\pi\bar{\rho}$, $h(\bar{\rho}; A)$ is compared to $2\pi\bar{\rho}P(\rho; A)$ which is plotted in Fig. 4.3 (b), with appropriate scaling. Comparing Fig. 4.3 (a) and (b) shows good qualitative agreement. The resolution in Fig. 4.3 (a) is limited due to a trade-off between annuli width, bin size within each annulus, and the number of samples per bin, particularly around low frequency region, however. As a result, the fall-off in the histogram for small A is not apparent.



Figure 4.2: Image used to calculate the Fourier amplitude distributions.

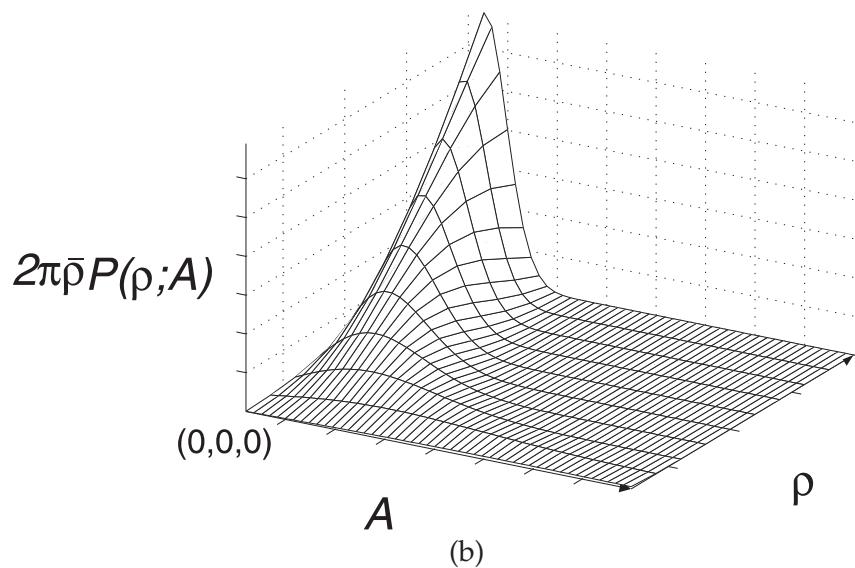
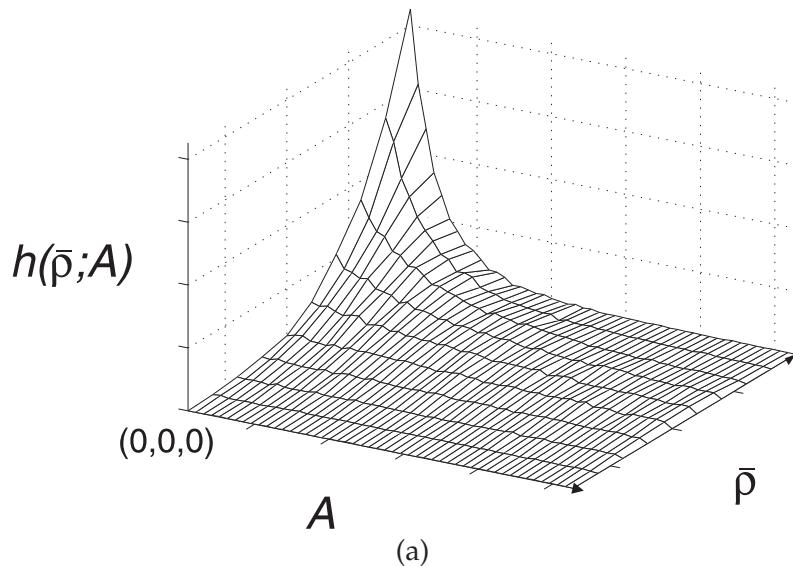


Figure 4.3: (a) The histogram of the Fourier amplitudes $h(\bar{\rho}; A)$ of the image shown in Fig. 4.2, and (b) the density $2\pi\bar{\rho}P(\rho; A)$ from Eq. (4.15)).

The accuracy of the spectral amplitude distribution, Eq. (4.18), was assessed by comparing it to histograms from a variety of images. This was done by calculating the parameters (a , b , c) in Eq. (4.18) using maximum likelihood (ML) and comparing the resulting pdf with image histograms. Twenty images were used for the experiments, each of size 512×512 pixels and 256 gray levels. The images included outdoor scenes and urban scenes. The Fourier transform of each image was calculated using the discrete Fourier transform (DFT) and the spectral amplitude samples are denoted A_i . For each image, the parameters a , b and c were determined by maximising the loglikelihood, i.e.,

$$(a, b, c) = \arg \max_{(a', b', c' > 1)} \left[\sum_i \log P_{\mathbf{a}' \mathbf{b}' \mathbf{c}'} (A_i) \right], \quad (4.19)$$

where the sum is over all the amplitudes. The constraint $c > 1$ ensures that $\gamma > 0$. The normalised rmse

$$e = \left(\frac{\sum_i [\alpha P(A_i) - h(A_i)]^2}{\sum_i [h(A_i)]^2} \right)^{1/2} \quad (4.20)$$

between the density function $P(A_i)$ and the amplitude histogram $h(A_i)$ was calculated as a measure of the goodness of fit, where α is a normalising constant. The histogram for each image was computed by dividing the range of amplitude values into 200 equally spaced bins. Because the spectrum contains a very small number of very large amplitudes, the largest 1% of the amplitudes were not included in the histogram since they would preclude the assignment of sufficient bins with a sufficient number of samples in each and most of the amplitudes would be placed in a few bins. The mean value of the normalised rmse over the 20 images was 0.05 with a range of 0.02 to 0.07, indicating good agreement. The values of c obtained were between 1 and 1.5, giving values of the exponent of A , $1 - 2c$, in Eq. (4.18) between -1 and -2 .

One of the images used is shown in Fig. 4.2, and its amplitude histogram and the corresponding density function is shown in Fig. 4.4. The density function is seen to be a good representation of the amplitude distribution. To prevent the bulk of the amplitude values being compressed into a small region at the low amplitude end of the plot, the top 5% of the amplitude values are not shown in Fig. 4.4.

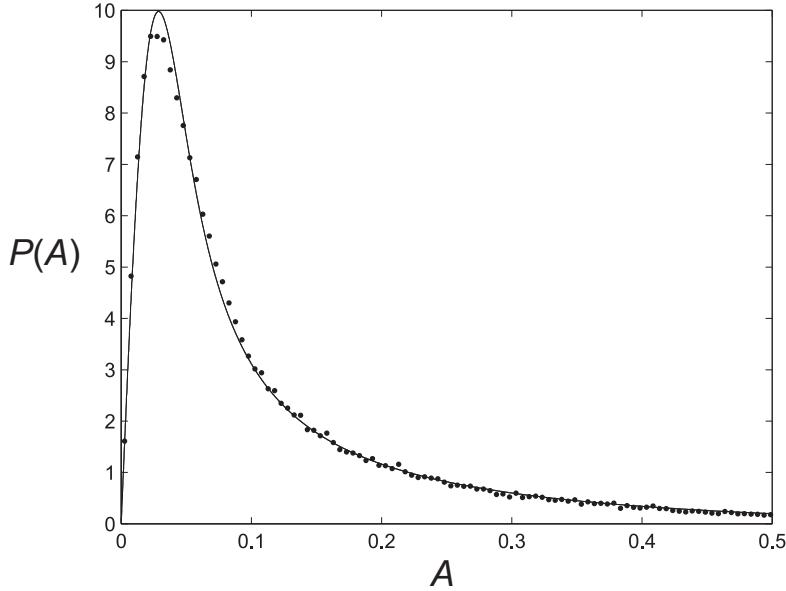


Figure 4.4: The Fourier amplitude histogram of the image shown in Fig. 4.2 (· · ·) and the density function Eq. (4.18) (—).

4.4 Discussion

Using an image model consisting of independent objects together with the observed behaviour of the circularly averaged power spectra with spatial frequency, expressions for distributions of image spectral amplitudes have been derived. The amplitude distribution as a function of spatial frequency, Eq. (4.15), shows good qualitative agreement to image spectral amplitude data. The overall distribution of spectral amplitudes, Eq. (4.18), fits well to histograms of the spectral amplitudes of a variety of images.

For images with uncorrelated pixel values the spectral amplitudes are Rayleigh distributed. The presence of extended objects in images introduces correlations between pixel values. Correlations narrow the spectrum around the origin of Fourier space, and since the spectral amplitudes are generally larger at smaller spatial frequencies, this increases the proportion of larger amplitudes. This characteristic is illustrated in Fig. 4.5. The figure shows the amplitude distributions for the point object image in Fig. 4.1(a) and for the image in Fig. 4.2, scaled so that their leading edges (low amplitude region) overlap. The longer tail of the image distribution relative to the Rayleigh distribution is evident.

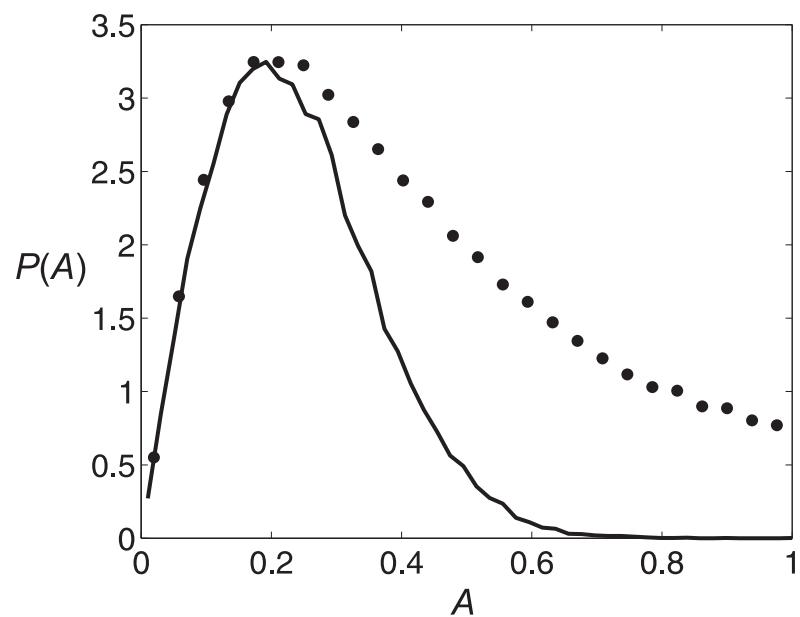


Figure 4.5: Spectral amplitude distributions of a random point image (Fig. 4.1 (a)) (—) and the image in Fig. 4.2 (···).

Chapter 5

Phase dominance

5.1 Introduction

Phase dominance refers to the general observation that loss of image spectral phase information tends to lead to a less recognisable image than does loss of the spectral amplitude information, which implies that the phase contains more information than the amplitude. Phase dominance generally refers to the effect on an image of replacing its Fourier amplitude or phase by that from another image. The general observation is that replacing the phase tends to have a much larger effect than does replacing the amplitude. Although phase dominance has been known for some time in many fields, most studies so far have been based only on a qualitative and subjective analysis, and no quantitative basis for this phenomenon has emerged. It is not clear if phase dominance is a purely objective phenomenon or if it is related to visual perception. Phase dominance appears to be a general phenomenon, but there remain questions around whether genuine counter-examples (i.e., amplitude dominance) exist.

The phenomenon phase dominance is studied in this chapter. Examples of phase dominance are shown and characteristics for different kinds of images are illustrated in Section 5.2. In Section 5.3, the mse that results from exchanging the Fourier amplitude or phase is evaluated theoretically. In Section 5.4, two reported counter-examples to phase dominance are critically examined. Conclusions are presented in Section 5.5.

5.2 Examples of phase dominance

Phase dominance, for images, is commonly demonstrated as follow [26, 28, 27]. Two unrelated images $f(x, y)$ and $g(x, y)$ are used to reconstruct two images $f_a g_\phi(x, y)$ and $f_\phi g_a(x, y)$

as

$$\begin{aligned} f_a g_\phi(x, y) &= \mathcal{F}^{-1} \{ |F(u, v)| \exp(i\phi \{ G(u, v) \}) \} \\ f_\phi g_a(x, y) &= \mathcal{F}^{-1} \{ |G(u, v)| \exp(i\phi \{ F(u, v) \}) \}, \end{aligned} \quad (5.1)$$

and the reconstructed images are compared with the original images. It is generally found that $f_a g_\phi(x, y)$ resembles $g(x, y)$ more than it does $f(x, y)$, and $f_\phi g_a(x, y)$ resembles $f(x, y)$ more than it does $g(x, y)$, and this fact is referred to as phase dominance. If the opposite is the case it is sometimes referred to as amplitude dominance. The images $f(x, y)$ and $g(x, y)$ must of course be unrelated to ensure independence of their Fourier amplitudes and phases.

An example is shown in Fig. 5.1, where two synthetic images (Fig. 5.1 (c) and (d)) are constructed by exchanging the Fourier amplitude and phase of the two original images (Fig. 5.1 (a) and (b)). By visual inspection, it is obvious that the synthetic images resemble the images that contribute the phase information. To make the comparison to the original images quantitative, the mse between each reconstructed image and each original image is calculated and are listed in Table 5.1. The results support the visual observations. In fact, the differences between the mses are quite dramatic, indicating that phase dominance is a strong phenomenon.

In these examples, the amplitudes and phases are exchanged at the same position in Fourier space. To exchange amplitudes and/or phases at different positions in Fourier space would destroy most of the information content. This has little significance for the phases but, as seen in previous chapters, the Fourier amplitudes fall off with increasing spatial frequency in a fairly characteristic fashion. Therefore, this characteristic fall-off is transferred from one image to the other when the amplitudes are exchanged. This is expected to have a lesser effect on the reconstructed images if the characteristic fall-off is similar for both images. The circularly averaged power spectra of the two images in Figs. 5.1(a) and 5.1(b) are shown in Fig. 5.2, and are seen to be similar in the fall-off with spatial frequency.

As described in Chapter 4, images made up of many randomly positioned points have a flat power spectrum. Phase dominance for such images is illustrated in Fig. 5.3 where two images containing randomly positioned points are used to reconstruct images with their Fourier amplitudes and phases exchanged. Phase dominance is evident in the reconstructed images and the mses are listed in Table 5.1.

To assess the effects of two images having a different fall-off of their power spectra, the two images shown in Figs. 5.4(a) and 5.4(b), where Fig. 5.4(a) and Fig. 5.1(a) are identical, were used to reconstruct images with their Fourier amplitudes and phases exchanged. The circularly averaged power spectra of Fig. 5.4(b) are shown in Fig. 5.2(c). Phase dominance

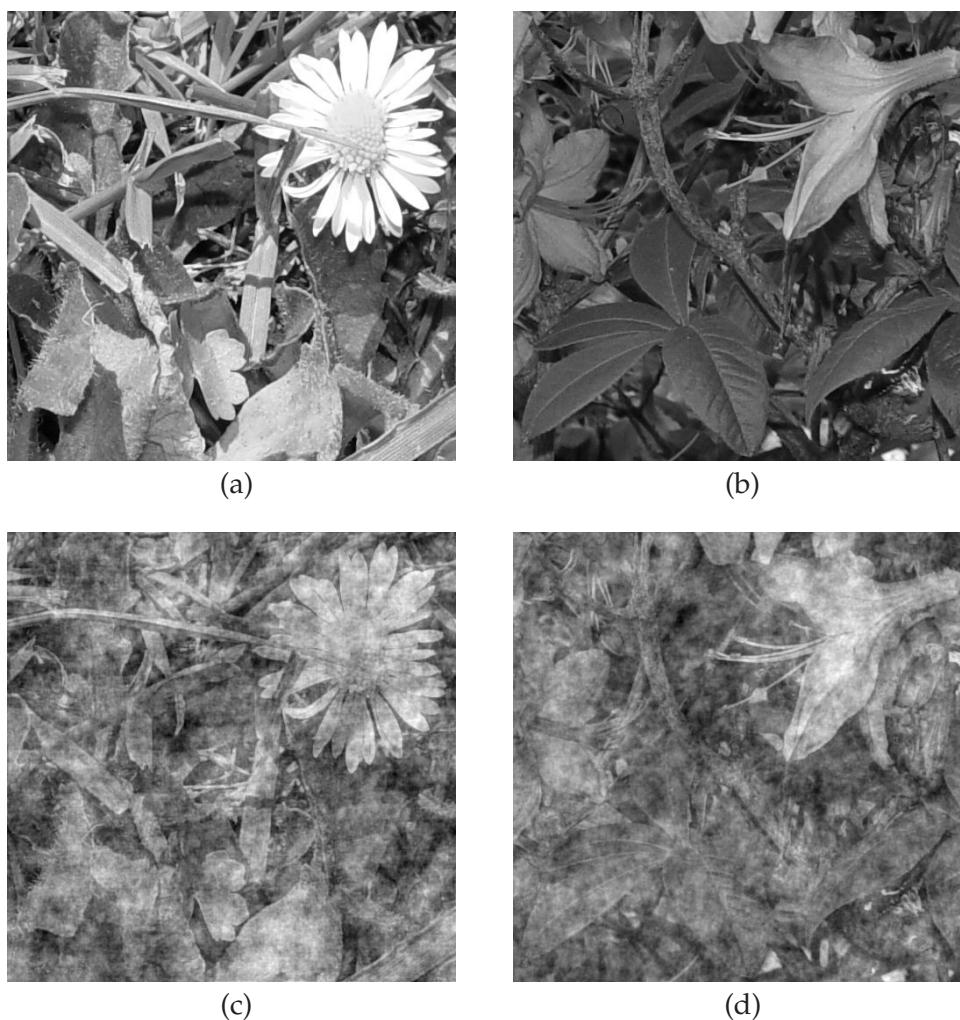


Figure 5.1: Illustration of phase dominance. (a) and (b): original images. (c) and (d): composite images calculated using the spectral phase of (a) and (b), respectively.

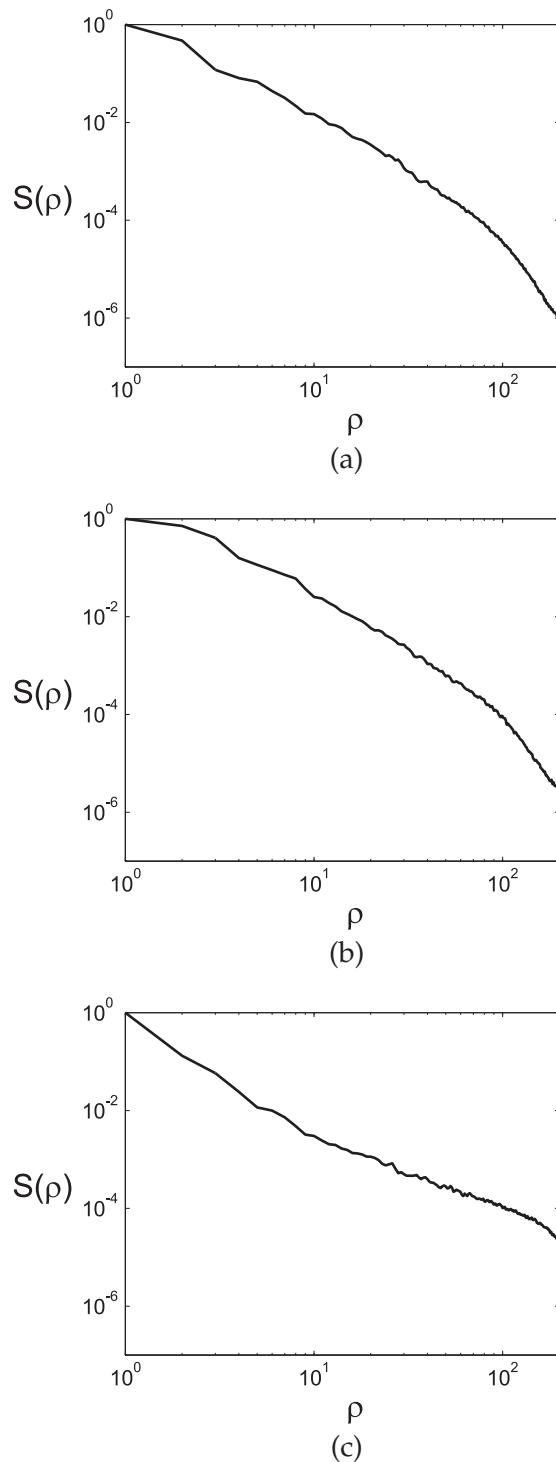


Figure 5.2: Circularly averaged power spectrum for the testing images (a) Fig. 5.1 (a) , (b) Fig. 5.1 (b) and (c) Fig. 5.4 (b)).

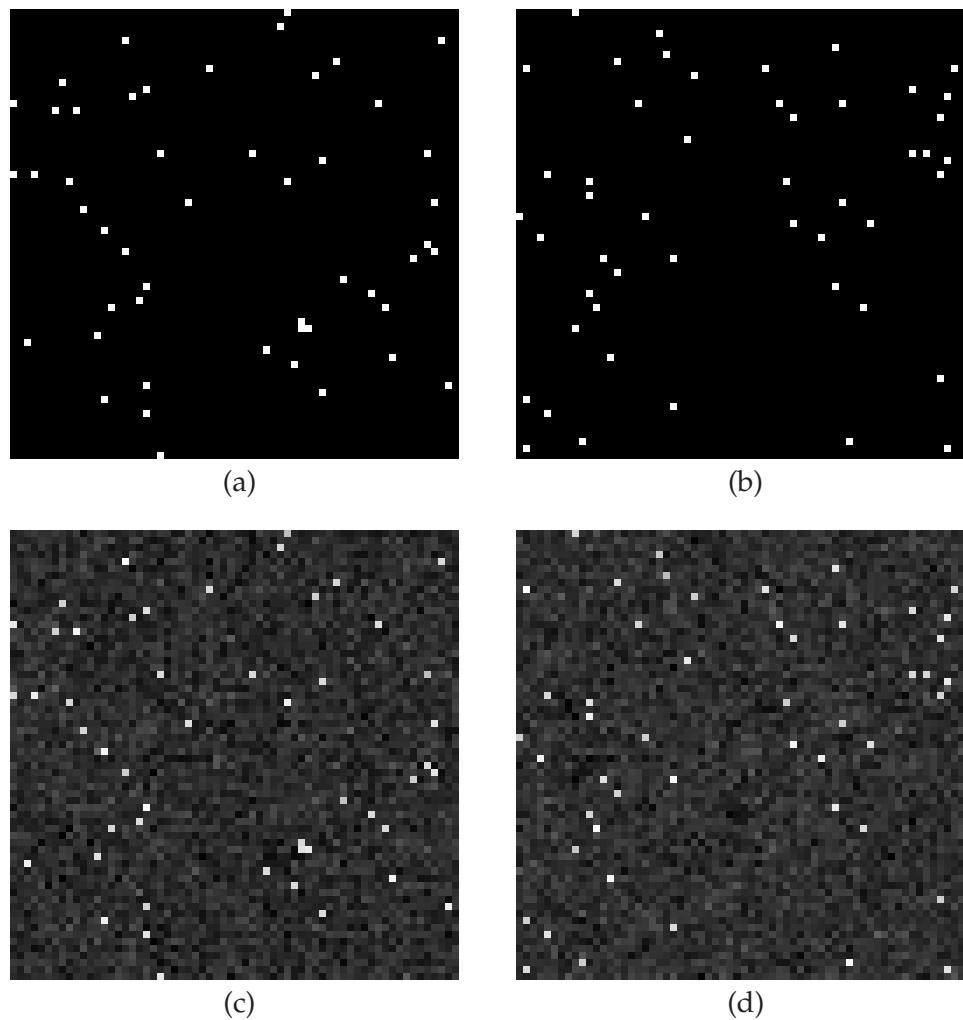


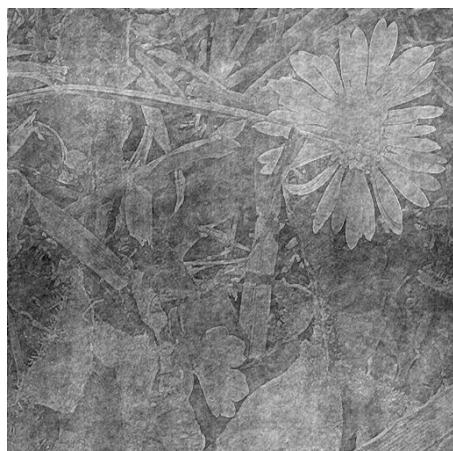
Figure 5.3: Illustration of phase dominance with randomly positioned points images (a) and (b): original images. (c) and (d): composite images calculated using the spectral phase of (a) and (b), respectively.



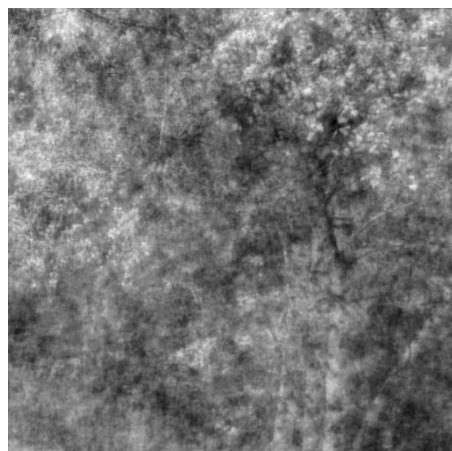
(a)



(b)



(c)



(d)

Figure 5.4: Illustration of phase dominance with images with distinctly different slope for the fall off of their power spectra (a) and (b): original images. (c) and (d): composite images calculated using the spectral phase of (a) and (b), respectively.

Table 5.1: Mse calculated using Eq. (3.1) for images generated by retaining either the Fourier amplitude or Fourier phase as described in the text.

Original Image	Amplitude retained		Phase retained	
	a	b	a	b
Fig. 5.1	1.55	1.66	0.39	0.39
Fig. 5.3	2.01	2.03	0.41	0.41
Fig. 5.4	1.95	1.81	0.67	0.67

is still evident as can be seen in Figs. 5.4(c) and 5.4(d) although it is not as strong as in Figs. 5.1(c) and 5.1(d). This is shown by the mses listed in Table 5.1 that do not show as large a difference than those for Fig. 5.1.

5.3 Theory

In order to understand the basis of phase dominance, the expected mse is determined when either the Fourier amplitude or phase in an image is replaced by the amplitude or phase from another image. It is assumed that the overall behaviour of amplitude with spatial frequency is the same in both images, and the energy is taken to be the same in both images (by scaling the second image if necessary). Exchanging the amplitude or phase is equivalent to adding errors to the amplitude or phase so that the results from the previous chapters can be used to calculate the mse.

For the case where the phase is replaced, the mse is easily calculated. Since no particular value of the Fourier phase is distinguished, it is uniformly distributed on $(-\pi, \pi)$. As a result of wrapping, the phase error introduced by replacing the phase of an image by the phase of another image is also uniformly distributed on $(-\pi, \pi)$. Recalling the results of Chapter 2, the expected mse, denoted here by E_p , is then

$$E_p = 2. \quad (5.2)$$

For the case where the Fourier amplitude is replaced the expected mse, denoted here by E_a , is given by

$$\begin{aligned} E_a &= \frac{\langle (\Delta|F|)^2 \rangle}{\langle |F|^2 \rangle} \\ &= \int_{-\infty}^{\infty} x^2 P_{\Delta A}(x) dx / \int_0^{\infty} x^2 P_A(x) dx, \end{aligned} \quad (5.3)$$

where $\Delta A = A_1 - A_2$, and A_1 and A_2 are the Fourier amplitudes of the two images. For a particular point in Fourier space, $P_{\Delta A}(x)$ is given by

$$\begin{aligned} P_{\Delta A}(x) &= P_{A_1}(x) \odot P_{A_2}(-x) \\ &= \int_{-\infty}^{\infty} P_{A_1}(x') P_{A_2}(x+x') dx'. \end{aligned} \quad (5.4)$$

Consider first the case of random point images for which, as shown in Chapter 4, the Fourier amplitudes are Rayleigh distributed, i.e.,

$$\begin{aligned} P_A(x) &= \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) & x \geq 0 \\ &= 0 & x < 0. \end{aligned} \quad (5.5)$$

Therefore, $P_{\Delta A}(x)$ is given as

$$\begin{aligned} P_{\Delta A}(x) &= \frac{1}{\sigma^4} \exp\left(-\frac{x^2}{2\sigma^2}\right) \int_0^{\infty} (xx' + x'^2) \exp\left(-\frac{1}{\sigma^2}(xx' + x'^2)\right) dx' & x \geq 0 \\ &= \frac{1}{\sigma^4} \exp\left(-\frac{x^2}{2\sigma^2}\right) \int_{-x}^{\infty} (xx' + x'^2) \exp\left(-\frac{1}{\sigma^2}(xx' + x'^2)\right) dx' & x < 0. \end{aligned} \quad (5.6)$$

Note that $P_{\Delta A}(x) = P_{\Delta A}(-x)$, and evaluating Eq. (5.6) for $x \geq 0$ using Eqs. 3.462.5 and 3.462.7 of [108] gives

$$P_{\Delta A}(x) = \frac{|x|}{4\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) + \frac{\sqrt{\pi}}{8\sigma} \exp\left(-\frac{x^2}{4\sigma^2}\right) \left[1 - \text{erf}\left(\frac{|x|}{2\sigma}\right)\right] \left(2 - \frac{x^2}{\sigma^2}\right). \quad (5.7)$$

For such an image then, using Eqs. 5.3, 5.7, and Eq. 3.461.3 of [108] and Mathematica [109], gives

$$\begin{aligned} E_a &= \frac{1}{\sigma^2} \int_0^{\infty} x^2 P_{\Delta A}(x) dx \\ &= \frac{1}{4\sigma^4} \int_0^{\infty} x^3 \exp\left(-\frac{x^2}{2\sigma^2}\right) dx + 2\sqrt{\pi} \left[\int_0^{\infty} (x^2 - 2x^4) \exp(-x^2) \text{erfc}(x) dx \right] \\ &= (4 - \pi)/2 \approx 0.42. \end{aligned} \quad (5.8)$$

Referring to Table 5.1 for point images (Fig. 5.3) shows that the mse value of 0.41 agrees with this analysis. Note that for Fig. 5.1 the mse value of 0.39 is also close to the analytical value, even though the amplitudes are not Rayleigh distributed. For Fig. 5.4, the mse is slightly higher at 0.67. This might be expected since in this case the Fourier amplitudes for the two images are somewhat differently distributed.

Consider now the case of general images where the Fourier amplitudes are not Rayleigh distributed, but that the average fall-off of the amplitudes with spatial frequency is the

same for both images. As shown in Chapter 4, to a good approximation the Fourier amplitudes can be considered to be Rayleigh distributed in a thin resolution shell $\rho \pm \Delta\rho$. Therefore, the mse can be calculated by summing the (unnormalised) squared error over the resolution shells and normalising the result. The parameter σ^2 of the Rayleigh distribution now depends on ρ , i.e., $\sigma^2(\rho)$, and the energy in each resolution shell is $2\sigma^2(\rho)$. The mse, denoted \tilde{E}_a , is then

$$\tilde{E}_a = \int_0^\infty E_a(\rho) 2\sigma^2(\rho) 2\pi\rho d\rho / \int_0^\infty 2\sigma^2(\rho) 2\pi\rho d\rho, \quad (5.9)$$

where $E_a(\rho) = (4 - \pi)/2$ is the mse at resolution ρ , so that

$$\tilde{E}_a = (4 - \pi)/2. \quad (5.10)$$

Hence, the mse ≈ 0.42 for *any* spatial frequency dependence of the Fourier amplitudes, as long as this is the same for both images. This explains the mse value of 0.39 for Fig. 5.1 in Table 5.1.

For two images with different spatial frequency dependence of the Fourier amplitudes, exchanging the Fourier amplitudes will transfer energy between different resolutions and the resulting mse is expected to be larger. This is consistent with the larger value of 0.67 for Fig. 5.4 in Table 5.1. This value is still much smaller than the value of 2 that results when the phase are exchanged.

Overall then, this analysis shows that the mse expected when the phases are exchanged is 3-5 times that expected when the amplitudes are exchanged. This provides a rigorous explanation of the widely observed phenomenon of phase dominance.

5.4 Phase dominance counter-examples

The above analysis indicates that phase dominance appears to be a general phenomenon. A number of authors have, however, shown examples of images that supposedly demonstrate amplitude dominance. Two such examples are those of Juvells *et al.* [110] and Lohmann *et al.* [111]. In the next two subsections each of these two examples are examined in order to explain these observations.

5.4.1 Counter-example of Juvells *et al.*

Juvells *et al.* [110] studied phase dominance based on Eq. (5.1), and using four test images. Various synthetic images were constructed by swapping the Fourier amplitudes and phases between two of these images. Their results are assessed here.

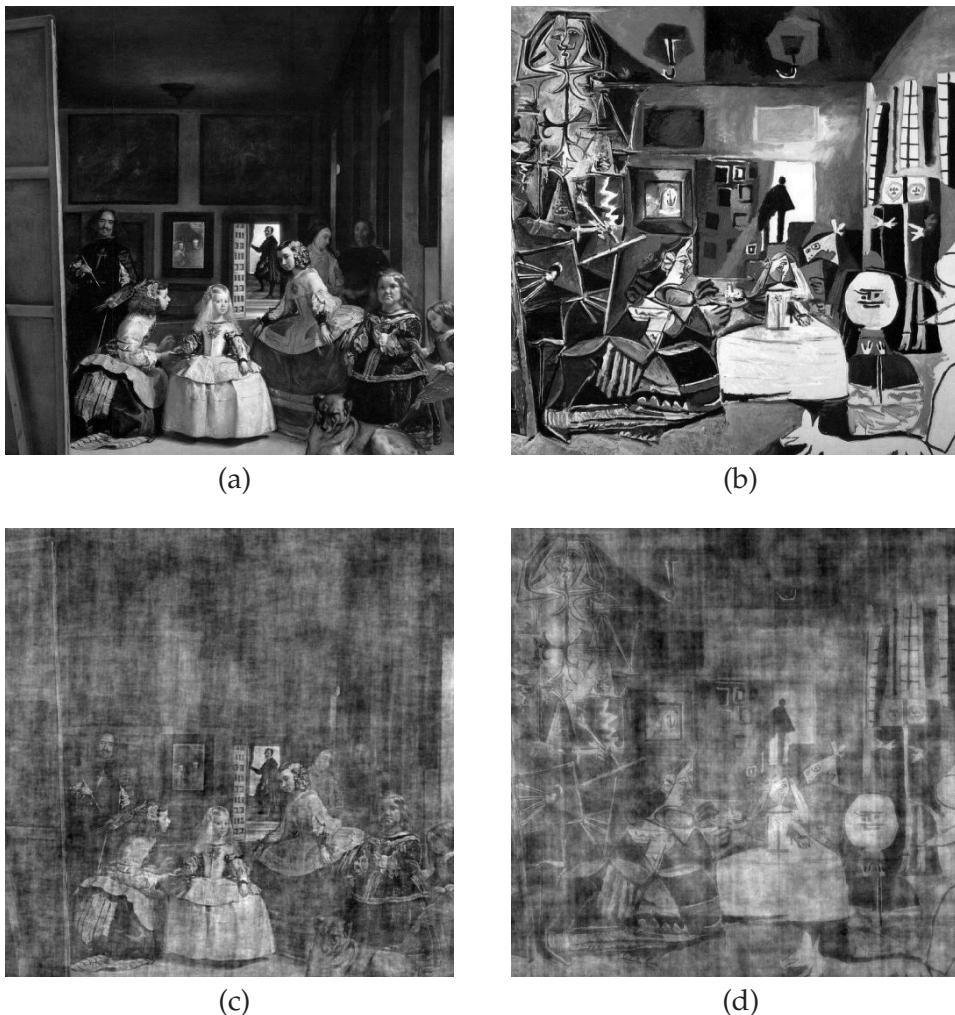


Figure 5.5: The example of Juvells *et al.* repeated here. (a) and (b) Original images. (c) and (d): Reconstructed images calculated using the spectral phase of (a) and (b), respectively.

In their first experiment, the amplitudes and phases of two images shown in Figs. 5.5 (a) and (b) were swapped. The same experiment was conducted here and the reconstructed images are shown in Figs. 5.5 (c) and 5.5 (d). The results clearly show phase dominance. The mses values listed in Table 5.2 are consistent with this observation.

For their second experiment, the Fourier amplitude and phase of Figs. 5.6 (a) and (b) were exchanged, and repeating this calculation gives the images shown in Figs. 5.6 (c) and (d). Juvells *et al.* state that although Fig. 5.6 (d) resembles Fig. 5.6 (b), i.e., phase dominance, Fig. 5.6 (c) resembles Fig. 5.6 (a) less so, indicating some degree of amplitude dominance. However, inspection of Fig. 5.6 (c) still shows evidence of many of the features in Fig. 5.6 (a). For example, the dress of the girl in the foreground, the painting on the back wall, and the edge of the board on the left are all evident in Fig. 5.6 (c), although the contrast is reduced by the diagonal streaking. More striking is that the doorway in the background with the person's head, shoulders and arm in Fig. 5.6 (a) are quite evident in Fig. 5.6 (c). Also, there is little evidence of the rectangles in Fig. 5.6 (b) in Fig. 5.6 (c). In summary, Fig. 5.6 (c) shows more phase dominance character than amplitude dominance character, despite the very structured form of Fig. 5.6 (b). This is borne out by the calculated mses listed in Table 5.2

Inspection of Fig. 5.6 (c) show that it contains streaks oriented at $\pm 45^\circ$ to the horizontal axis. This is due to Fig. 5.6 (b) exclusively containing objects oriented at these angles, thus its transform being dominated by spatial frequencies at these angles. To see the effects of these spatial frequencies on Fig. 5.6 (c), they were removed by filtering with the filter shown in Fig. 5.7 that reduces the spectral components within the angle range $\pm(40^\circ - 50^\circ)$. The result is shown in Fig. 5.6 (e) and is quite dramatic. Once the streaks in Fig. 5.6 (c) are removed, much of the character of Fig. 5.6 (a) is recovered. Finally, Fig. 5.6 (f) shows an image reconstructed using the phase from Fig. 5.6 (a) and an amplitude that is circularly symmetric and decreases monotonically with spatial frequency as $|F(u, v)| \propto \rho^{-1}$. The resemblance to Fig. 5.6 (a) is quite striking although the amplitude function used is "bland."

Table 5.2: Mse calculated using Eq. (3.1) for images generated to study the counter-examples of phase dominance provided by Juvells *et al.* [110].

Original Image	Amplitude retained		Phase retained	
	a	b	a	b
Fig. 5.5	1.61	1.79	0.52	0.52
Fig. 5.6	1.79	1.70	0.55	0.55
Fig. 5.8	1.98	3.49	1.90	1.90

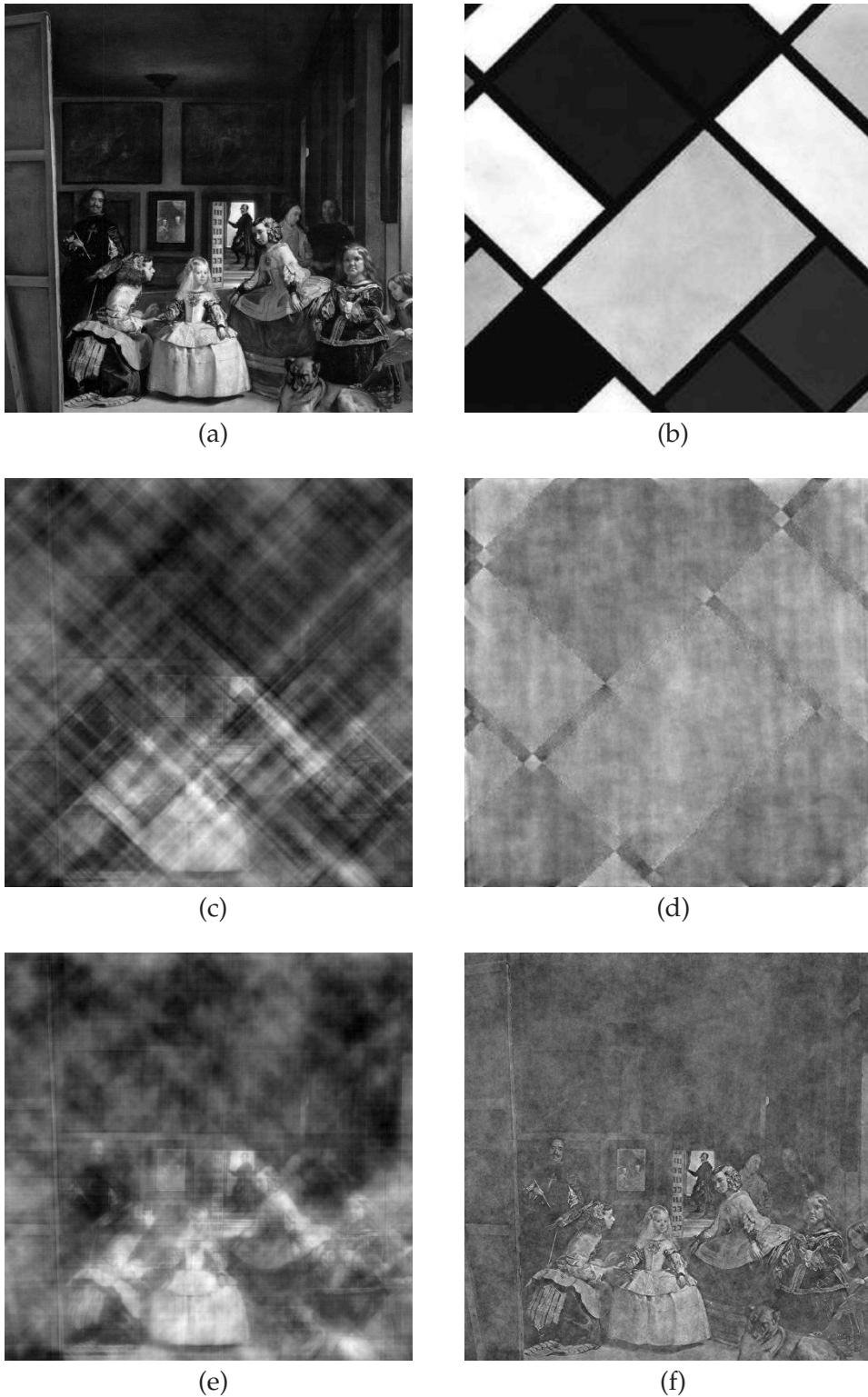


Figure 5.6: Images generated to study counter-examples given by Juvells *et al.* [110]. (a) and (b) Original images. (c) and (d) Images reconstructed by exchanging the Fourier amplitude and phase of images (a) and (b). (e) Reconstructed image with with phase of (a) and amplitude of (b) with the diagonal spatial frequencies filtered out as described in the text. (f) Image reconstructed with phase of (a) and amplitude function $|F(u,v)| \propto \rho^{-1}$.

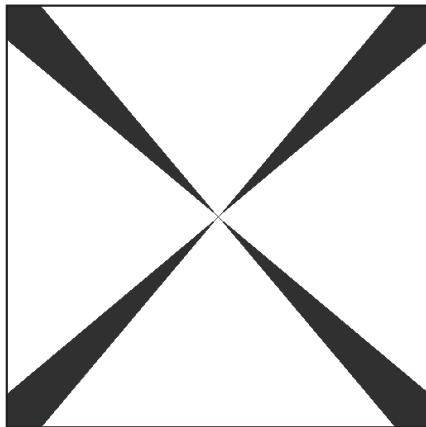


Figure 5.7: Filter applied to Fig. 5.6 (c) to produce Fig. 5.6 (e).

In their third experiment, Juvells and *et al.* [110] exchanged the amplitude and phase of Fig. 5.5 (a) and an image containing periodic vertical stripes. These two images are shown in Figs. 5.8 (a) and (b). This calculation has been repeated and the resulting images are shown in Figs. 5.8 (c) and (d). Fig. 5.8 (c) resembles the test bar and they argue that this demonstrates amplitude dominance. However, although Fig. 5.8 (c) shows vertical stripes, the amplitude of the stripes are unrelated to those in Fig. 5.8 (c). This is borne out by the mse value, listed in Table 5.2 which shows that the errors are similar whether the amplitude or phase is retained. The Fourier amplitude of the test bar image is shown in Fig. 5.9 where it is seen that the amplitude is zero at almost all positions in Fourier space. Therefore, when this amplitude function is used almost all of the spectral phase information is also lost when the image is reconstructed. This is therefore not a fair test of phase dominance. To allow the phase information to be retained, an image was reconstructed using the phase from Fig. 5.8 (a), the non-zero amplitude from Fig. 5.8 (b) together with the zero amplitudes from Fig. 5.8 (b) replaced by a constant equal to $0.01a$ where a is the maximum value of the Fourier amplitudes from Fig. 5.8 (b). The reconstructed image is shown in Fig. 5.8 (e) and is seen to slightly resemble Fig. 5.8 (a), demonstrating some phase dominance. Similarly, Fig. 5.8 (f) is reconstructed using the phase of Fig. 5.8 (b) and a constant amplitude, and the result resembles Fig. 5.8 (b), showing phase dominance.

In summary, careful experiments using the image from Juvells *et al.* show that they, in fact, show phase dominance.

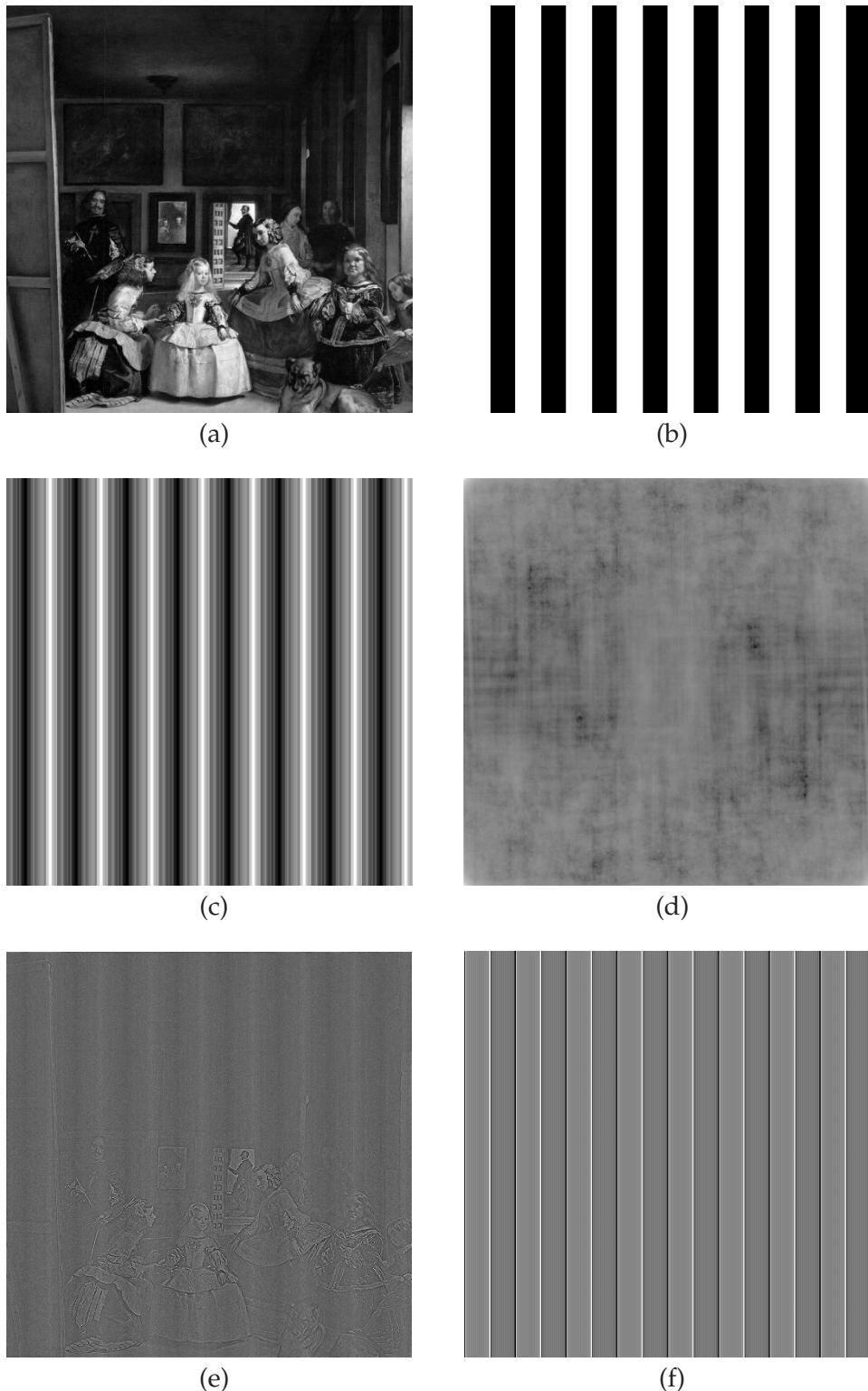


Figure 5.8: Images from Juvells *et al.* [110]. (a) and (b) Original images. (c) and (d) Images reconstructed by exchanging the Fourier amplitude and phase of the two original images, where (d) is plotted with log scale to enhance the contrast. (e) Image reconstructed from the phase of (a) and amplitude from (b) with the zero amplitudes replaced by a constant. (f) Image reconstructed from the phase of (b) and a constant amplitude.

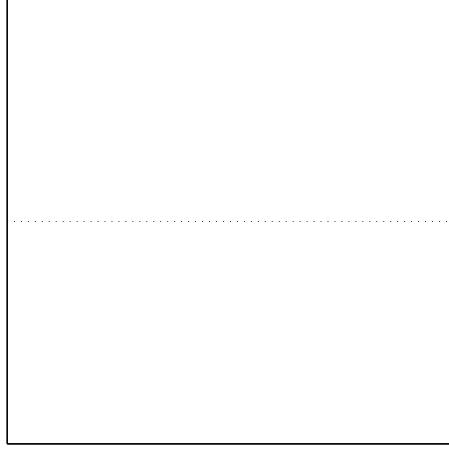


Figure 5.9: The Fourier amplitude of Fig. 5.8 (b), where white and black denote zero and non zero values, respectively.

5.4.2 Counter-examples of Lohmann *et al.*

Lohmann *et al.* [111] studied the importance of Fourier amplitude and phase with a somewhat different approach using one image. They generated the images $f_a(x, y)$ and $f_\phi(x, y)$ given by

$$\begin{aligned} f_a(x, y) &= \mathcal{F}^{-1} \{ |F(u, v)| \exp(i\Phi(u, v)) \} \\ f_\phi(x, y) &= \mathcal{F}^{-1} \{ A(u, v) \exp(i\phi \{ F(u, v) \}) \}, \end{aligned} \quad (5.11)$$

where $A(u, v)$ and $\Phi(u, v)$ are amplitude and phase functions that are independent of $|F(u, v)|$ and $\phi \{ F(u, v) \}$, respectively. They assessed the relative importance of the amplitude and phase by examining the resemblance of $f_a(x, y)$ and $f_\phi(x, y)$ to $f(x, y)$. If $f_\phi(x, y)$ resembles $f(x, y)$ more than does $f_a(x, y)$, then they described this as phase dominance. If $f_a(x, y)$ resembles $f(x, y)$ more than does $f_\phi(x, y)$, then they described it as amplitude dominance. However, a difficulty with this approach is the question of how to choose the functions $\Phi(u, v)$ and $A(u, v)$? Different choices could lead to different conclusions. They could be chosen, in an attempt to be unbiased, to be constant, but this may introduce particular characteristics into the images. They could be chosen to be random, but then what distribution should they be drawn from? What should the average spatial frequency structure be? It is obviously important that $\Phi(u, v)$ and $\phi \{ F(u, v) \}$ are independent, and that $A(u, v)$ and $|F(u, v)|$ are independent. The approach using Eq. (5.1) is more satisfactory than using Eq. (5.11) since it avoids the difficulty of choosing independent amplitude or phase functions.

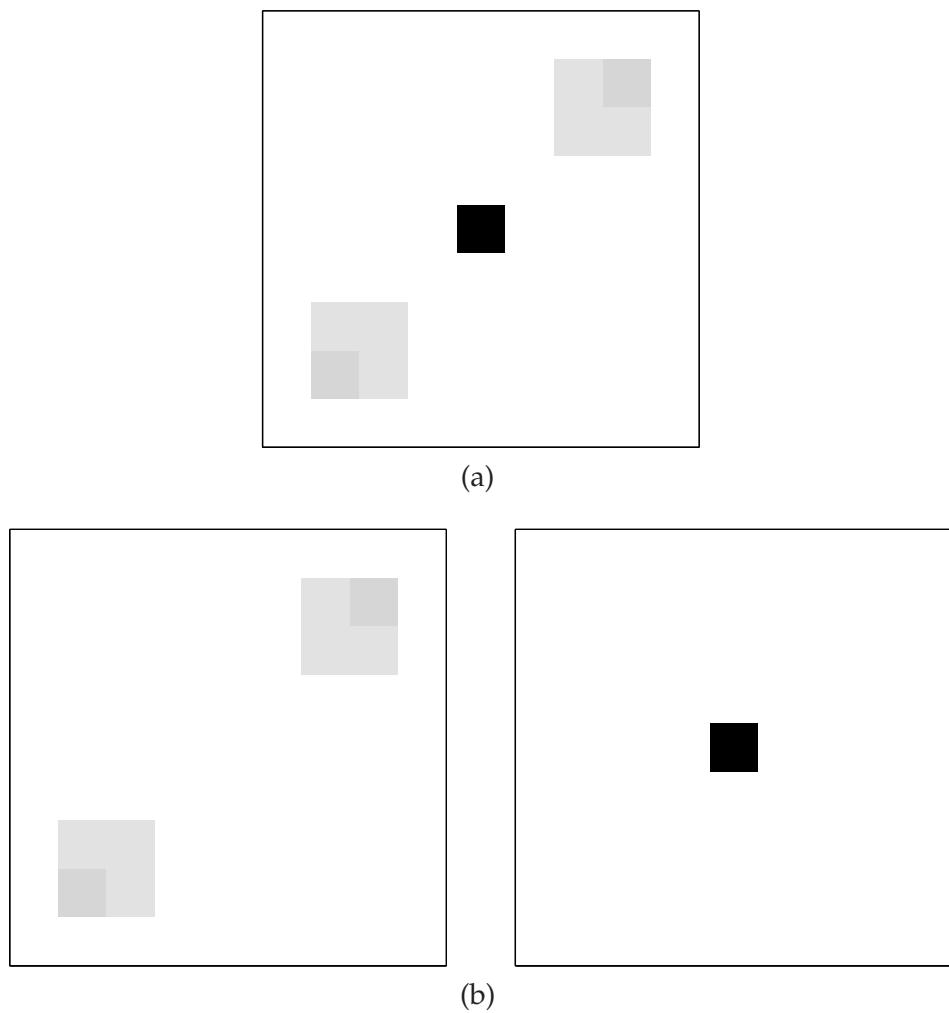


Figure 5.10: (a) The test images $f_1(x, y)$ and (b) the test image $f_2(x, y)$ (left: real part, right: imaginary part).

In their results they presented images that apparently show either strong amplitude or strong phase dominance. The images they used have a strong central peak whose amplitude is equal to the integrated value of the remainder of the image. Since their images have a large number of pixels, the amplitude of the central peak greatly exceeds the amplitude of the other parts of the image. Their example images are centrosymmetric, and images of the same kind are shown in Fig. 5.10. Images in Fig. 5.10 were studied, and a small number of pixels is used (9×9) so that the whole image can be displayed, i.e., the central peak and the remaining part of the image are displayed at the correct relative amplitudes.

They considered two images. In one both the central peak and the remainder of the image are real (similar to the image shown in Fig. 5.10(a)), and in the other the central peak is pure imaginary and the remainder of the image is real (similar to the image shown in Fig. 5.10(b)). These two images are denoted $f_1(x, y)$ and $f_2(x, y)$, respectively. In their study, the relative importance of amplitude and phase is studied by constructing images using Eq. (5.11) and using constant functions for $A(u, v)$ and $\Phi(u, v)$ given by

$$\begin{aligned} A(u, v) &= \langle |F(u, v)|^2 \rangle^{1/2} \\ \Phi(u, v) &= 0. \end{aligned} \quad (5.12)$$

The images $f_{1a}(x, y)$ and $f_{1\phi}(x, y)$, respectively, calculated in this way are shown in Fig. 5.11. The mses of the reconstructed images are calculated and are listed in Table 5.3. Inspection of Fig. 5.11 shows that $f_{1a}(x, y)$ is a perfect reconstruction of $f_1(x, y)$. However, $f_{1\phi}(x, y)$ is a reasonable reconstruction considering the energy in the central peak (mse of 0.10). The latter observation is not apparent in Lohmann [111] since the central peak in the reconstructions was not shown.

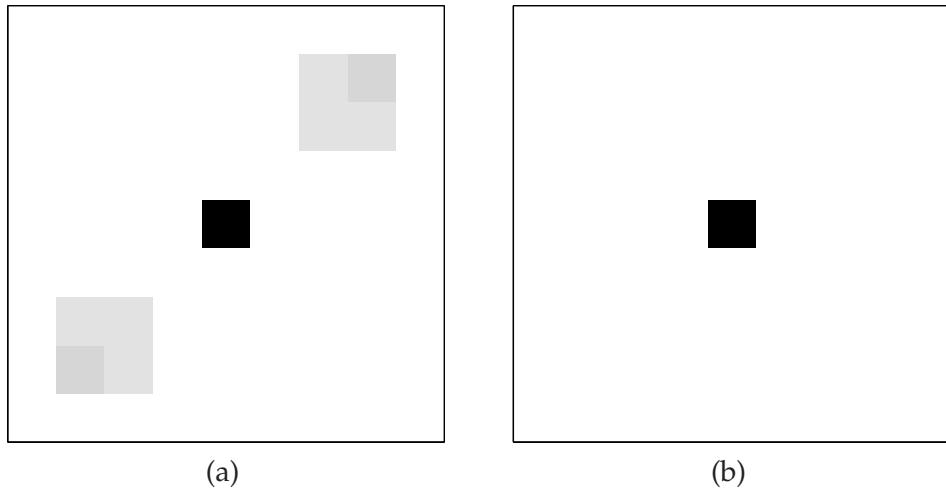


Figure 5.11: Reconstructed images (a) $f_{1a}(x, y)$ and (b) $f_{1\phi}(x, y)$, for $\phi(u, v) = 0$.

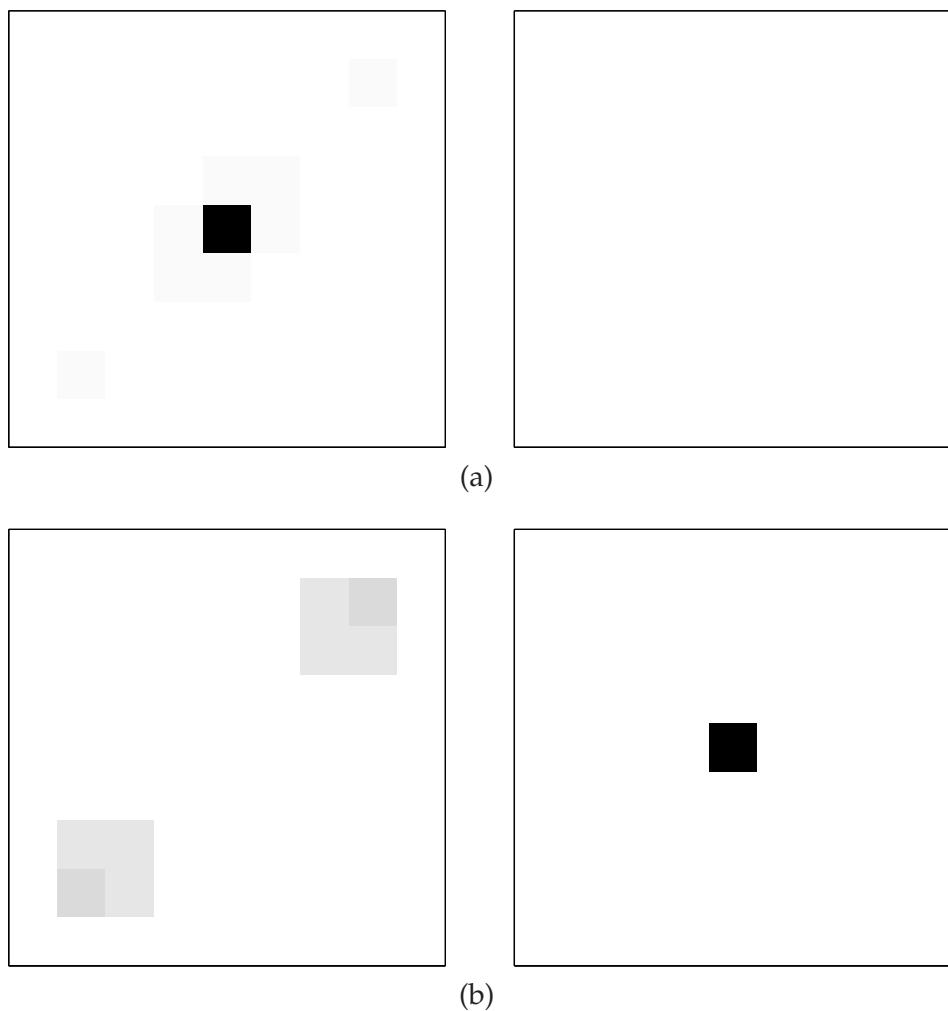


Figure 5.12: Reconstructed images (a) $f_{2a}(x, y)$ and (b) $f_{2\phi}(x, y)$, for $\phi(u, v) = 0$. Left: real part, right: imaginary part.

Reconstructions from $f_2(x, y)$ are shown in Fig. 5.12. Comparison of Fig. 5.12 with Fig. 5.10(b) shows that $f_{2a}(x, y)$ is not a good reconstruction of $f_2(x, y)$ (the peak appears in the real part rather than in the imaginary part, leading to a very high mse of 2), but that $f_{2\phi}(x, y)$ is an almost perfect reconstruction (mse of 0.006), i.e., phase dominance. These are, essentially, the results obtained by Lohmann [111], and on the face of it, $f_1(x, y)$ shows amplitude dominance and $f_2(x, y)$ phase dominance.

Table 5.3: Mse calculated using Eq. (3.1) for images generated by retaining either the Fourier amplitude or Fourier phase for the examples described in the text. (^a denotes that the phase is also retained)

Original image	Amplitude retained	Phase retained	Figure
$f_1(x, y)$	0.00 ^a	0.12	5.11
$f_1(x, y)$	2	0.12	5.13 (a)
$f_1(x, y)$	2.02	0.12	5.13 (b)
$f_2(x, y)$	1.96	0.006	5.12
$f_2(x, y)$	0.12	0.006	5.14 (a)
$f_2(x, y)$	2.07	0.006	5.14 (b)
$f_1(x, y)$	1.54	0.51	5.15
$g(x, y)$	1.94	0.51	5.15

However, further investigation shows that these results are misleading. A real and centrosymmetric image produces a real and even signal in the both x and y directions, i.e., the 2-D Fourier transform of such image is real (phases are 0 or π). Adding a strong central peak in the image domain is equivalent to adding a large positive real constant to the whole Fourier transform resulting in a real and positive Fourier transform, i.e., $\phi_{f_1}(u, v) = 0$ for all u and v where $\phi_f(u, v)$ denotes $\phi \{ \mathcal{F} \{ f(x, y) \} \}$. Therefore, the choice $\phi(u, v) = 0$ does not change the phase, i.e., $\phi(u, v) = \phi_f(u, v)$, and this choice does not therefore shed any light on the amplitude/phase dominance of $f_1(x, y)$. The problem is that the phase function chosen is not independent of true phase function of the image. The problem can be further illustrated by replacing $\phi(u, v) = 0$ by $\phi(u, v) = \pi/2$. The reconstructed images for $f_1(x, y)$ and $f_2(x, y)$ are shown in Figs. 5.13 (a) and 5.14 (a), respectively. Comparison of Figs. 5.13 (a) and 5.14 (a) with Figs. 5.11 (b) and 5.12 (b), respectively, shows that in this case both $f_1(x, y)$ and $f_2(x, y)$ exhibit phase dominance (taking into account the real and imaginary parts of the reconstruction). This is supported by the corresponding mses (Table 5.3). Since the choice $\phi(u, v) = \pi/2$ is constant, as is the true phase, it could introduce special properties into the reconstructed images. Therefore, images were also calculated with the phase $\phi(u, v)$ set equal to a random value for each sample in Fourier space. The reconstructed images are shown in Figs. 5.13 (b) and 5.14 (b), for $f_1(x, y)$ and $f_2(x, y)$, respectively. Inspection of Figs. 5.13 (b) and 5.14 (b) shows that they bear no resemblance to

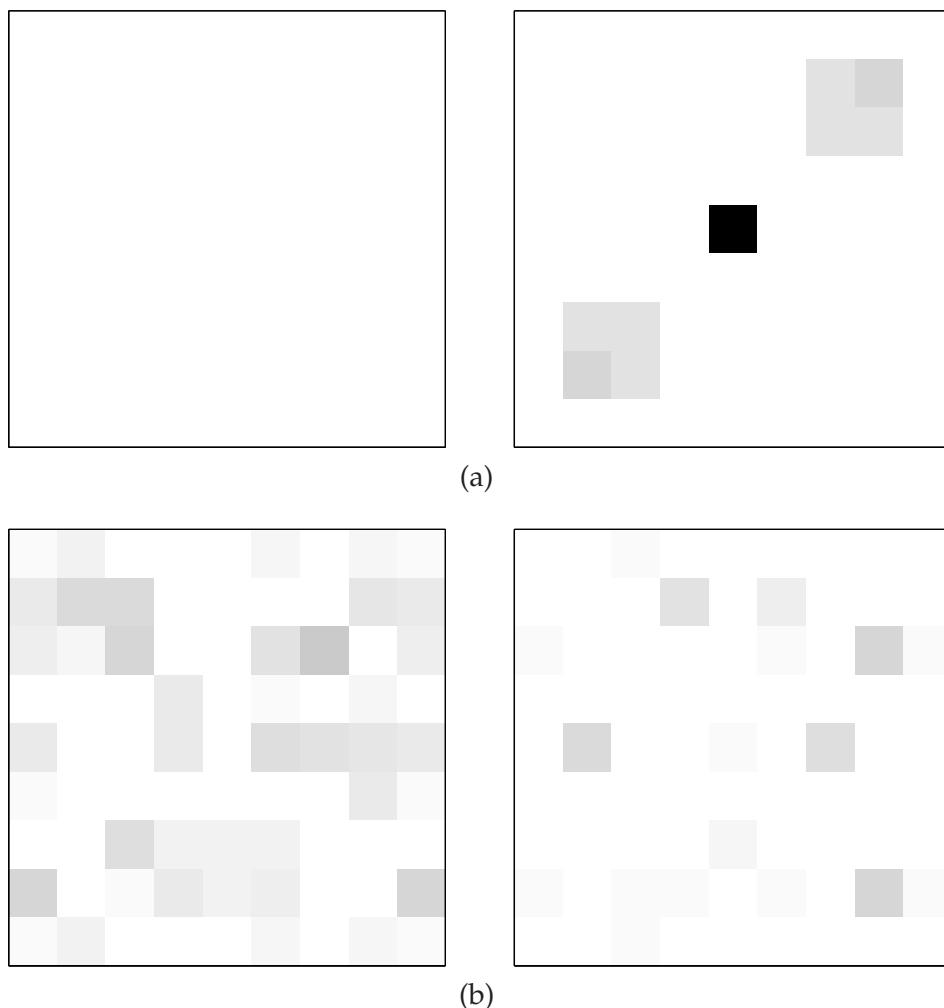


Figure 5.13: Reconstructed images $f_{1a}(x, y)$ for (a) $\phi(u, v) = \pi/2$ and (b) $\phi(u, v)$ random. Left: real part; right: imaginary part.

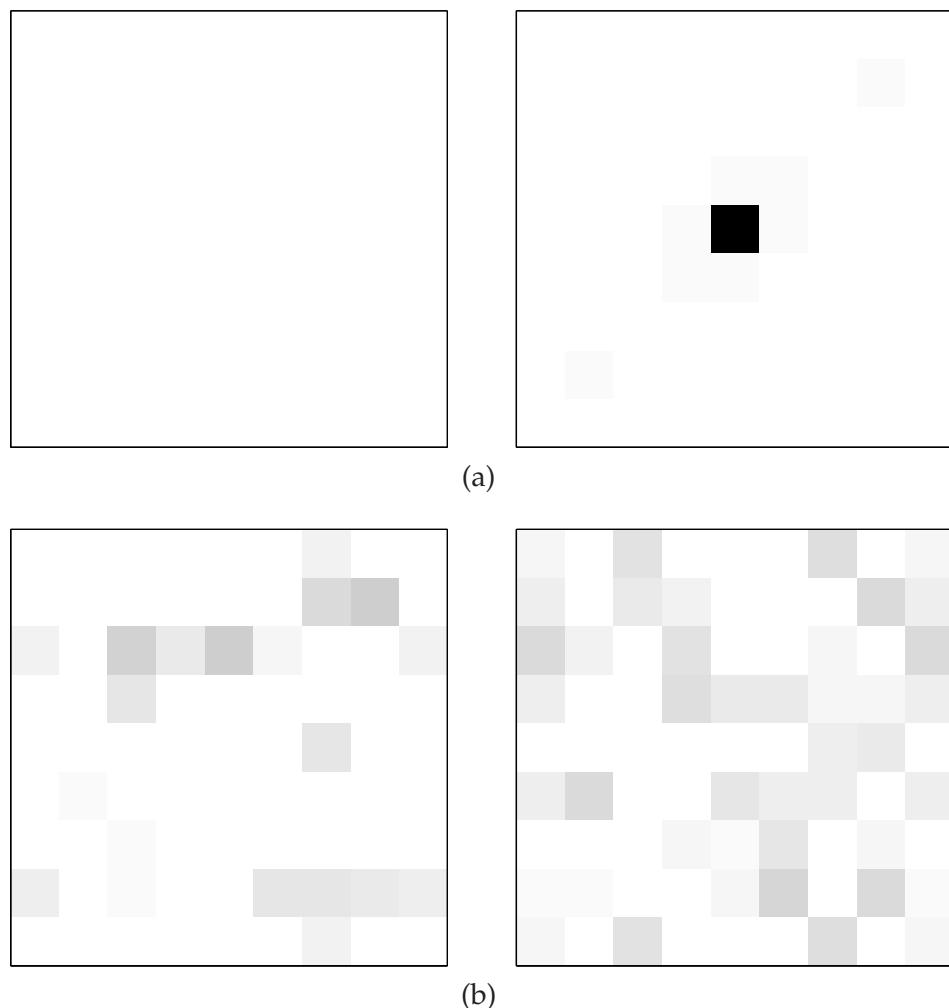


Figure 5.14: Reconstructed images $f_{2a}(x, y)$ for (a) $\phi(u, v) = \pi/2$ and (b) $\phi(u, v)$ random. Left: real part; right: imaginary part.

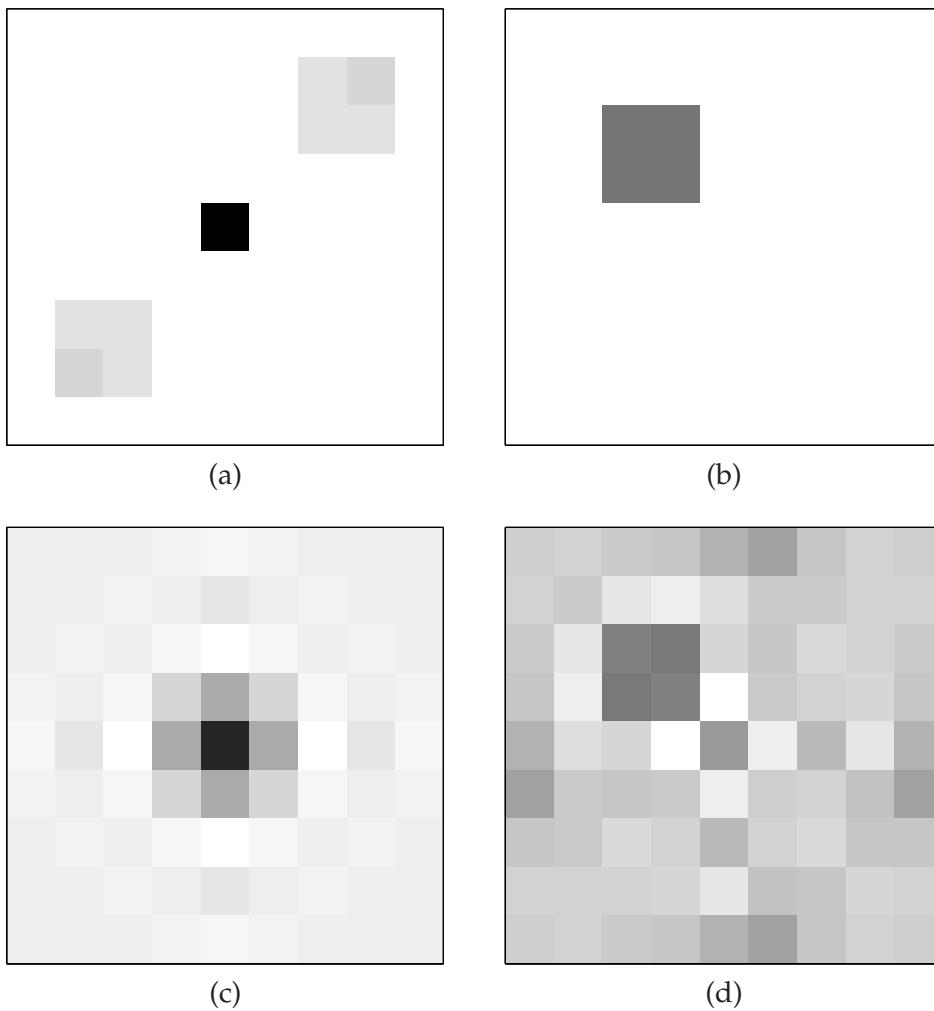


Figure 5.15: The test images (a) $f_1(x, y)$ and (b) $g(x, y)$, and the reconstructed images (c) $f_{1\phi}g_a(x, y)$ and (d) $f_{1a}g_\phi(x, y)$.

the original images (also reflected by the high mses listed in Table 5.3), and so there is no indication of amplitude dominance.

As described earlier, the relative importance of amplitude and phase is better studied using two images as in Eq. (5.1), since it is easier to obtain independent amplitude and phase functions. For example, consider the image $g(x, y)$ shown in Fig. 5.15(b), and the images $f_{1\phi}g_a(x, y)$ and $f_{1a}g_\phi(x, y)$ calculated using Eq. (5.1) and shown in Figs. 5.15(c) and 5.15(d), respectively. Even though the image has the strong central peak which tends to dominate the phase in Fourier space, both $f_1(x, y)$ and $g(x, y)$ show phase dominance. In fact the mses obtained when retaining the amplitude are almost 4 times those obtained when retaining the phase (Table 5.3).

5.5 Discussion

Phase dominance appears to be a general phenomenon and it is difficult to construct counter-examples. The theoretical analysis of the mse that results when the Fourier amplitude or phase are exchanged between two images shows clearly that the fundamental basis of phase dominance. It would be useful, however, to analyse the mse in the case where the two images have different spatial frequency behaviours. Phase dominance is sometimes presented as a “mysterious” phenomenon. This analysis shows that there is no mystery; exchanging the phase simply introduces much more error in an image than does exchanging the amplitude. There is probably little justification for invoking characteristics of the human visual system to explain phase dominance.

The analysis of two reports of apparent counter-examples to phase dominance shows that, in fact, they are erroneous. Their results are due to either not removing only amplitude or phase information, not adding independent unrelated amplitude or phase information, or incorrectly evaluating the resulting images.

It is interesting to consider the relationship between phase dominance and phase retrieval. Reconstruction of an image from its Fourier amplitude (phase retrieval) is usually effective under fairly mild conditions [21, 19, 23], indicating that there is substantial information in the amplitude. This might appear, at first sight, to be inconsistent with phase dominance. This is not the case however. First, phase dominance implies only that reconstruction from the Fourier phase should be easier than reconstruction from the Fourier amplitude. It is well known that this is the case, although reconstruction from the phase is of lesser practical significance. Second, successful reconstruction from the Fourier amplitude requires at least some independent *a priori* information on the image, indicating that the amplitude alone has insufficient information to define an image.

Chapter 6

Power spectra of natural images

6.1 Introduction

This chapter is concerned with studies of the power spectra of natural images. As described in Section 1.7.3, the term natural image refers to images of the natural environment that the human visual system encounters and to which it has, presumably, adapted. The term power spectrum refers to the amplitude squared of the Fourier transform of an image, as a function of spatial frequency, i.e., $|F(u, v)|^2$. Although the power spectrum of an individual image has some angular structure, there is no consistent angular structure over an ensemble of images. The power spectra of natural images generally, however, do have a characteristic behaviour with the magnitude of the spatial frequency. The term power spectrum as used here refers to the angular average of the amplitude squared of the Fourier transform of an image as a function of the magnitude of spatial frequency, i.e., $S(\rho) = \langle |F(u, v)|^2 \rangle_\varphi$.

The characteristic of natural images addressed here is the observation that the circularly averaged spatial power spectrum on a log–log scale decreases linearly as spatial frequency increases, with a slope of ≈ 2 . Hence, the power spectrum $S(\rho)$ is a power law function of spatial frequency, i.e.,

$$S(\rho) = k\rho^{-\gamma}, \quad (6.1)$$

where k is a constant and $\gamma \approx 2$. Example power spectra of three 512×512 pixel natural images, plotted on a log–log scale, are shown in Fig. 6.1(a), and the ensemble averaged power spectrum over 40 natural images is shown in Fig. 6.1(b). Linearity of the log–log power spectra with a slope of about -2 is evident in the figure. To understand this phenomenon in detail, relevant questions one may ask are then: (1) What are the salient

statistical characteristics of natural images, and (2) What is it about these statistics that produces the characteristic dependence of the power spectrum on spatial frequency?

The first question, although complex, can probably be answered in part as follows, as has been recognised by a number of investigators [75, 84, 79]. A visual scene is generally made up of opaque objects, and the objects have a wide range of sizes and shapes. Furthermore the luminosity of individual objects is relatively constant, i.e., the variation in the luminosity across a single object is much less than the variation between objects. In other words, partially transparent objects and objects with continuously graded luminosity tend to be relatively uncommon in the natural environment. Since the objects are opaque, they *occlude* other objects, or parts of objects, behind them. In this sense then, a simple model of the visual environment is a collage of occluding, constant intensity objects with a wide range of sizes and shapes. There is evidence that the distribution of the object sizes is self-similar [112, 79], which is referred as *scaling* in this thesis. Collections of such objects are scale invariant, i.e., they have the same characteristics when viewed at different scales, and the object sizes follow a power law distribution. They can sometimes be described as fractals, which are loosely defined as patterns made up of parts that are similar in some sense to the whole. Since objects in a scene generally occur at a wide range of distances, and many objects (e.g. rocks) span a wide range of sizes, when captured in images they span a wide range of angular scales. It is not unreasonable therefore that images might exhibit self-similarity. An important consequence of this observation is that such a scene will contain many sharp *amplitude edges*, with a characteristic range of spacings between them, at the boundaries between the occluding constant intensity regions. This simple model of the natural visual environment has been discussed and utilised by a number of authors [79, 113, 81, 83].

The second question then becomes: "Which of the characteristics of occlusion, edges, and/or scaling are the primary source of the dependence of the power spectrum on spatial frequency?" This question has been investigated by a number of authors, although it is fair to say that a clear answer has not yet emerged. Issues revolve around the relative importance of scaling and edges, and the range of spatial frequencies over which power law behaviour of the power spectrum is present [79, 113, 114, 115, 81, 83].

An important paper on this topic is by Ruderman [79]. By analysis of a set of natural images, he showed that these images generally conform to the model described above. His results showed scaling (of the correlation function) in the spatial domain. He showed that a power law power spectrum with exponent $\gamma = 2 - \eta$ gives a power law correlation function of the image with exponent $-\eta$. For his image dataset he found $\eta \approx 0.19$. The images in the dataset were segmented by hand into "objects" and the correlation between pixels within and between objects evaluated. The results showed that the pixel correlations be-

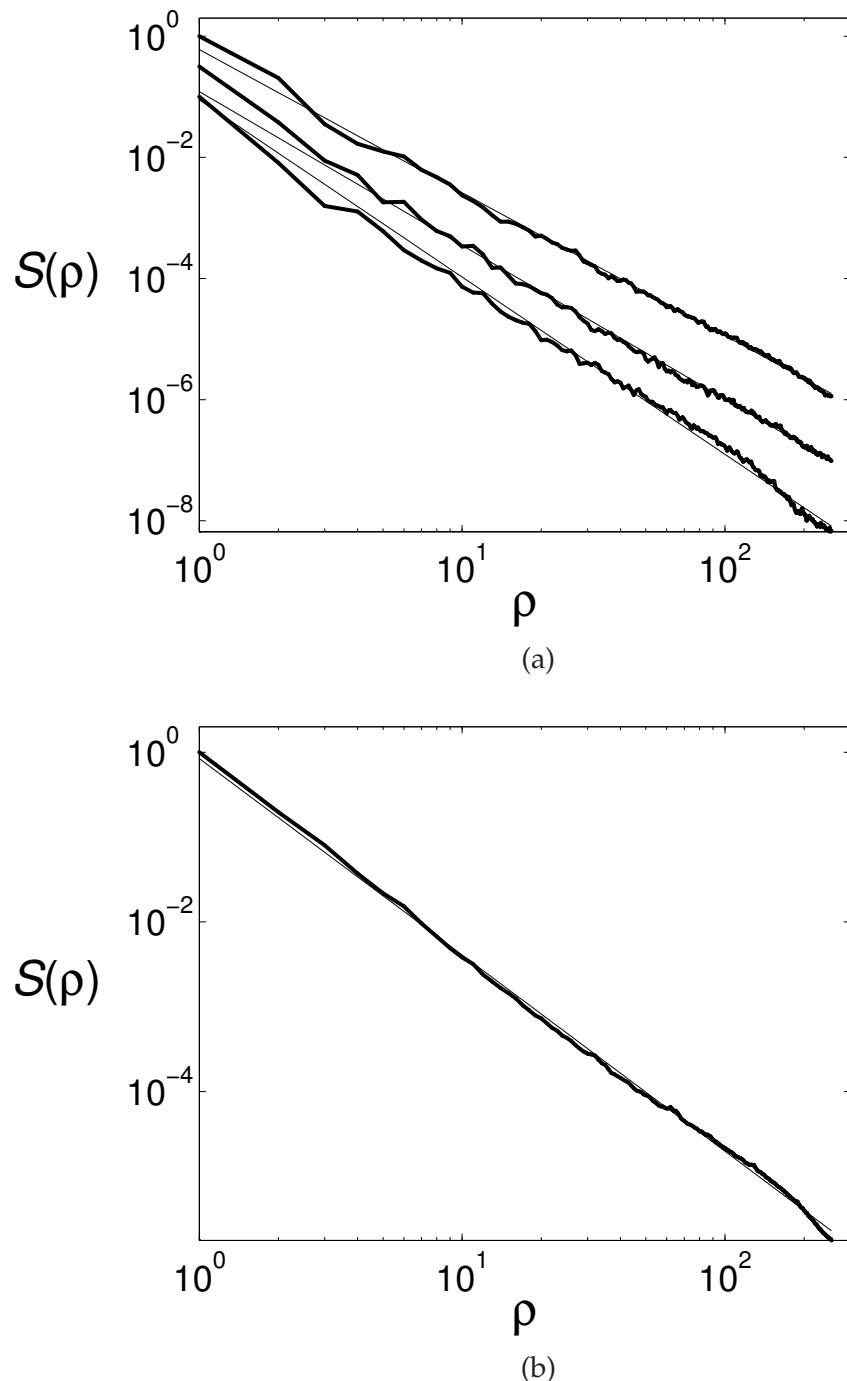


Figure 6.1: Circularly averaged power spectrum (thick curve) and a linear fit of (a) three images and (b) ensemble over 40 images. The slopes are 2.3, 2.5 and 2.9 in (a) and 2.3 in (b).

tween objects were significantly smaller than those within objects, and both were relatively independent of pixel separation. This observation is consistent with a model consisting of independent, occluding, constant luminosity objects. The probability that two pixels belong to the same object showed a weak power law component in pixel separation. He also considered an image model of occluding disks with a power law size (radius) distribution and showed that this leads to a power law correlation, and a power law power spectrum.

In a subsequent paper, Balboa *et al.* [113] argue that it is the characteristic spectral properties of edges, rather than scaling in the image domain, that leads to a power law power spectrum. They show that a variety of object sizes leads to spectral components that fall off with spatial frequency with slopes between 0 and -3 giving a total slope in the vicinity of -2 . They show an example of a model image with an exponential object size distribution that gives a power law spectrum. They conclude that the power law behaviour of the power spectrum is due primarily to the presence of edges (due to occlusion) in natural images and that a power law distribution of object sizes is not necessary.

Subsequent discussion by these authors [114, 115] provides supporting evidence for each viewpoint, the discussion centering around the range of spatial frequencies over which the spectrum is studied and the effect of the finite size of an image. The precise effects of occlusion, edges, and scaling on the behaviour of the power spectrum are therefore not yet resolved [81, 83]. Although this topic has been the subject of considerable study, as described above, there is no systematic study that isolates the effects of these different characteristics on image spectra. Of particular importance is practical versus theoretical effects. For example, although scaling of images maybe be theoretically necessary to obtain power law spectra, are deviations from power law spectra significant if images are not self-similar? Also, full scale invariance is not practical because of divergences for object sizes a for which $a \rightarrow 0$ and $a \rightarrow \infty$. For a power law size distribution as $a \rightarrow 0$ the image is totally covered in tiny objects, and as $a \rightarrow \infty$ the image contains one large object. It is necessary then to impose upper and lower limits on the object sizes and the image is then not strictly self-similar. Furthermore, the property of occlusion and constant luminosity objects is often invoked as the source of edges in images, however edges can be obtained in other ways. The property of occlusion as a necessary ingredient is therefore not clear.

A simple image model consisting of randomly placed disks is used here to perform a systematic study of the effects of occlusion, edges and object size distribution and luminosity on the behaviour of the power spectrum by numerical simulation. The methods used are described in Section 6.2 and the results presented in Section 6.3. A theoretical analysis is made for the case of nonoccluding disks in Section 6.4. A comparison of Ruderman's results [79] with the results of the analysis for images containing nonoccluding disks (Section 6.4) is presented in Section 6.5. Finally, conclusions are discussed in Section 6.6.

6.2 Modelling natural images with synthetic images

An image model that consists of disks placed at random in an image frame of 512×512 pixels is used in the numerical simulations, although parameters are presented in terms of an image whose extent is normalised to unity. A variety of different kinds of image are used that are generated using different objects and different size distributions.

Two different kinds of object (disk) amplitude are used. One of these has constant amplitude and the other has a variable amplitude that varies as a raised cosine that drops smoothly from its maximum value at the center to zero at the edges as

$$l(r) = \frac{L}{2} \left(1 + \cos \left(\frac{\pi r}{a} \right) \right), \quad (6.2)$$

where $l(r)$ is the amplitude, r is distance from the center of the disk, a is the disk radius and L is the maximum amplitude (Fig. 6.2).

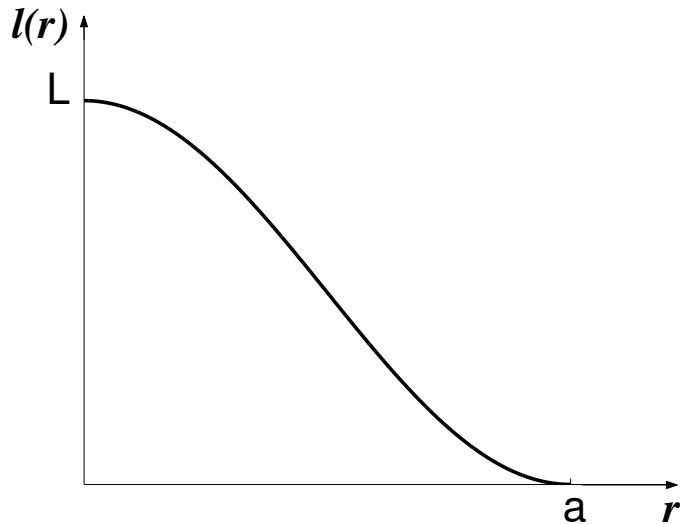


Figure 6.2: Amplitude function for variable amplitude objects.

Two different object size distributions are considered, power law and exponential, which are defined by a probability density $P(a)$ of disk radii a . In practical images, lower and upper bounds of disk radii, denoted by a_1 and a_2 (normalised to an image width of unity) respectively, are necessary. The lower bound corresponds to the maximum resolution of the observer, and the upper bound should not allow objects much larger than the image frame. For self-similar images the radius distribution is given by

$$\begin{aligned} P(a) &= k_p(\alpha, a_1, a_2) a^{-\alpha} && \text{for } a_1 < a < a_2 \\ &= 0 && \text{otherwise,} \end{aligned} \quad (6.3)$$

where the normalising constant $k_p(\alpha, a_1, a_2)$ is given by

$$\begin{aligned} k_p(\alpha, a_1, a_2) &= \frac{\alpha - 1}{a_1^{-\alpha+1} - a_2^{-\alpha+1}} && \text{for } \alpha \neq 1 \\ &= \frac{1}{\ln(a_2) - \ln(a_1)} && \text{for } \alpha = 1. \end{aligned} \quad (6.4)$$

The range $1 \leq \alpha \leq 3$ is used in the simulations and $\alpha > 0$ is referred to as the *scaling exponent*. For the exponential radius distribution $P(a)$ is given by

$$\begin{aligned} P(a) &= k_e(d, a_1, a_2) \exp(-a/d) && \text{for } a_1 < a < a_2 \\ &= 0 && \text{otherwise,} \end{aligned} \quad (6.5)$$

where

$$k_e(d, a_1, a_2) = [d(\exp(-a_1/d) - \exp(-a_2/d))]^{-1} \quad (6.6)$$

and d is the *correlation length*.

Both *occluding* and *nonoccluding* images were generated, using opaque and transparent objects, respectively. Opaque objects occlude the objects behind them. Although “behind” does not apply to two-dimensional images, a projection of a three-dimensional scene is simulated by adding disks to the image sequentially such that an added disk occludes the pixels that it occupies, i.e., occludes the parts of disks added previously that occupy the same pixels. Transparent objects are considered transparent luminous sources and the amplitude at any image pixel is the sum of the amplitudes of the objects that intersect that pixel. The characteristics of occluding and nonoccluding and constant and variable amplitude disks are illustrated in Fig. 6.3. The disk amplitude, or maximum amplitude L , was uniformly distributed on the range 0 to 255. For nonoccluding images, the amplitude was rescaled to the range 0 – 255 after the all disks were placed in the image frame.

A sufficient number of disks must be included in an image to cover the image frame completely and to obtain a statistically stationary ensemble. The number of disks required depends on α or d and was set such that the total disk area is ten times the image frame area, which was found to be sufficient.

Different objects and size distributions were used to generate six different kinds of image, which are denoted A – F and their generation and characteristics are listed in Table 6.1. The different images selectively include or exclude particular characteristics. Image Type A represents constant amplitude occluding objects with a power law size distribution. Image Type B replaces the power law size distribution with an exponential size distribution, but with the remaining characteristics unchanged, which identifies the importance of the

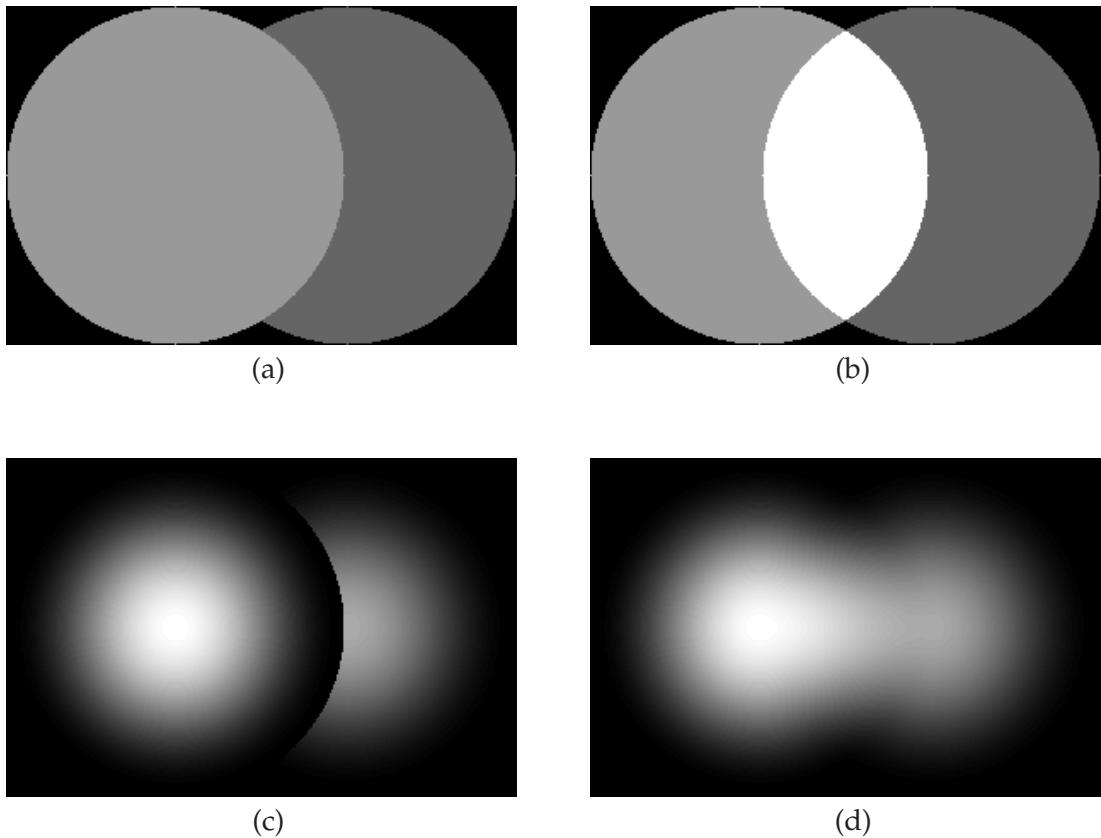


Figure 6.3: Constant amplitude (a) occluding (b) nonoccluding disks, and variable amplitude (c) occluding and (d) nonoccluding disks.

Table 6.1: Image types as described in the text.

Image Type	Objects			Image Properties				
	Amplitude	Type	Size Distribution	Edges	Occluding	Self-Similar	Piecewise Constant	
A	constant	opaque	power	Y	Y	Y	Y	Y
B	constant	opaque	exponential	Y	Y	N	Y	
C	constant	transparent	power	Y	N	Y	Y	
D	constant	transparent	exponential	Y	N	N	Y	
E	variable	opaque	power	Y	Y	Y	N	
F	variable	transparent	power	N	N	Y	N	

size distribution of objects in the image. Image Types C and D use the same kind of disk type and size distributions as Image Type A and B, respectively, but the disks are nonoccluding (transparent), so that the effect of occlusion can be studied. Image Type E contains occluding objects that have variable amplitude. Image Type F uses variable amplitude nonoccluding disks to produce images without edges.

The Fourier transform of the images was calculated using the discrete Fourier transform (DFT). To remove the edge effects due to periodicity of the DFT, the images were constructed such that the periodic extension of an image is continuous at the image boundaries. This was done by wrapping disks that extend out of the image boundary. The ensemble-averaged, circularly averaged power spectrum $S(\rho)$ was calculated as

$$S(\rho) = \langle S_i(\rho) \rangle_i = \langle \langle |F_i(\rho, \varphi)|^2 \rangle_\varphi \rangle_i, \quad (6.7)$$

where $F_i(\rho, \varphi)$ is the Fourier transform of the i -th image in the ensemble, and $\langle \rangle_\varphi$ and $\langle \rangle_i$ denote averaging over φ and i , respectively. The units of spatial frequency ρ are cycles per image. The average over φ was computed over shells in Fourier space of a thickness equal to two DFT sample spacings and dividing by the number of samples within each shell. $S(\rho)$ was calculated on the range $1 < \rho < 256$. Since the image has a finite size of unity, only integral values of ρ are relevant. The power spectra were fitted to the power law dependence

$$S(\rho) = B\rho^{-\gamma} \quad (6.8)$$

by fitting a regression line to the log-log power spectrum data approximately equally spaced in log frequency in the range of $10 < \rho < 200$ and γ calculated from the slope of the regression line.

6.3 Simulation results

Model images of each kind were generated as described above. The normalised minimum and maximum disk radii were $a_1 = 0.004$ and $a_2 = 0.4$. For self-similar images the values $\alpha = 1, 2, 3$ were used and for the exponential size distribution the values $d = 0.04, 0.02, 0.006$ (normalised to an image width of unity) were used. The number of disks used, determined as described in Section 6.2, varied between 200 and 25,000 for these parameter values. Sample images of each kind for these values of the parameters are shown in Fig. 6.4. The image properties as listed in Table 6.1 can be observed in these images. The circularly averaged power spectra were calculated as described above, averaged over 20 images, for each kind of image shown in Fig. 6.4, and are plotted on a log-log scale in Fig. 6.5. Regression lines, calculated as described in Section 6.2, are also shown in Fig. 6.5 and the derived exponents γ are listed in Table 6.2.

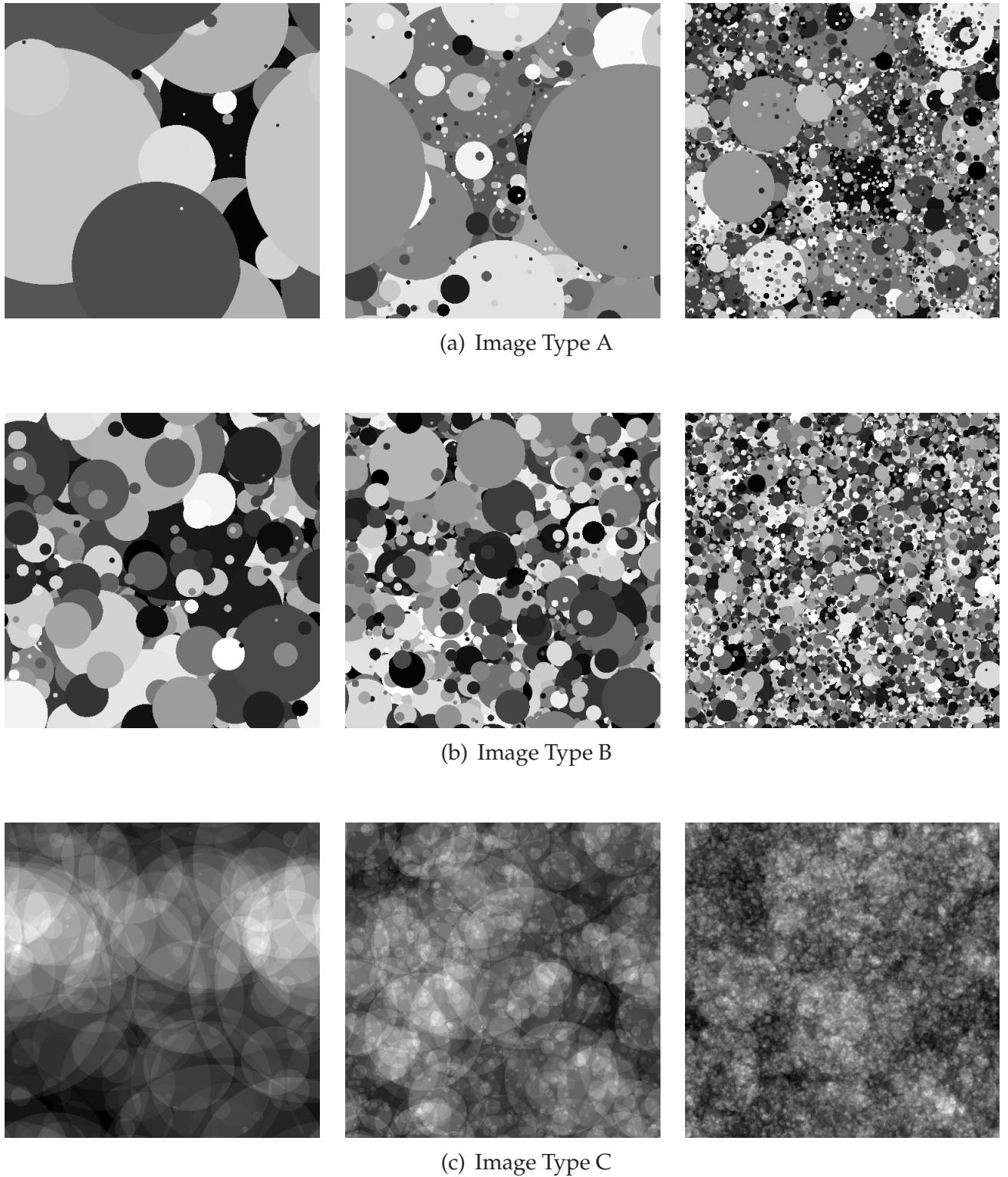
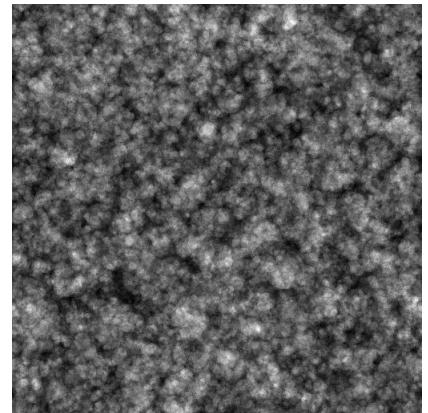
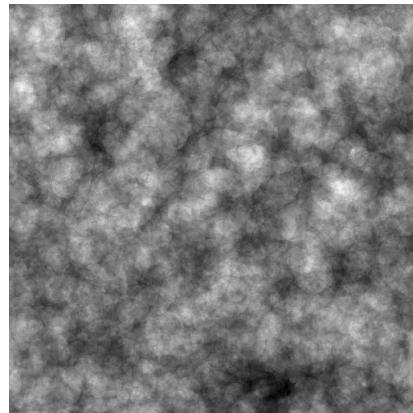
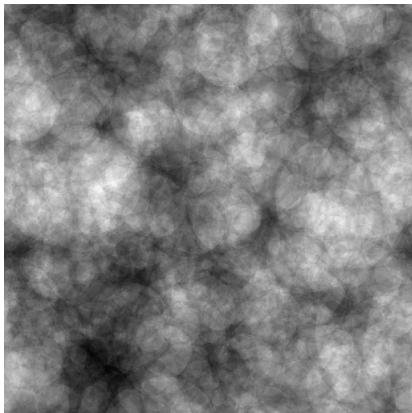
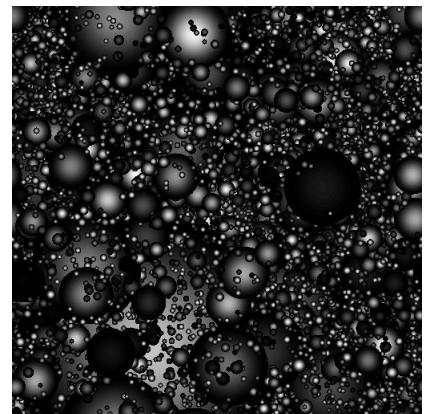
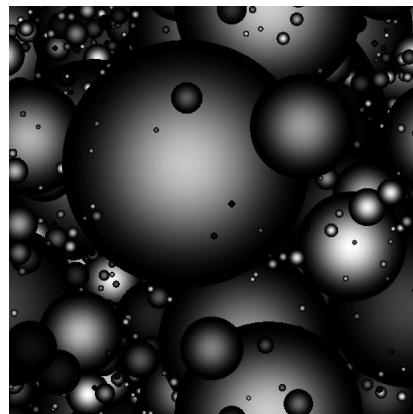
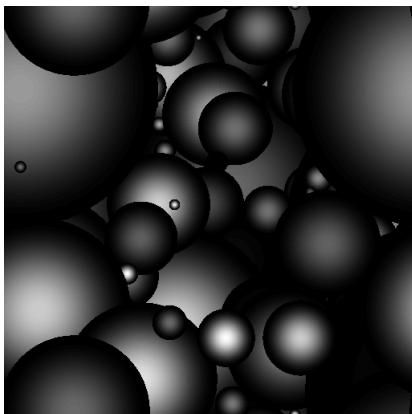


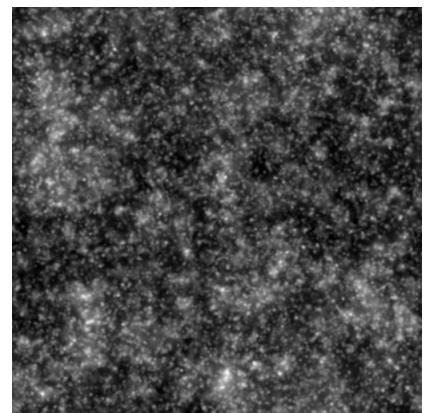
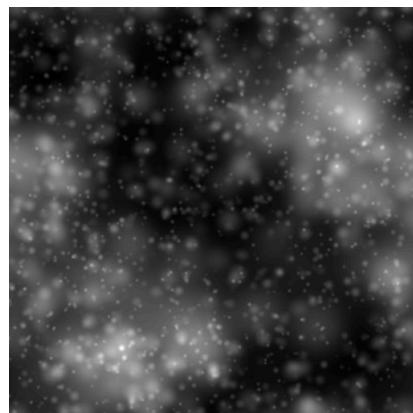
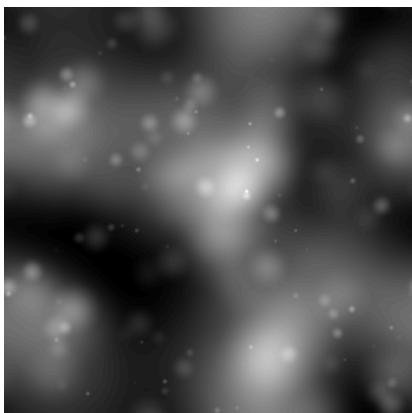
Figure 6.4: Model images of each type as labelled and described in Table 6.1. The columns correspond to $\alpha = 1$ or $d = 0.04$ (left), $\alpha = 2$ or $d = 0.02$ (center), and $\alpha = 3$ or $d = 0.006$ (right).



(d) Image Type D



(e) Image Type E



(f) Image Type F

Figure 6.4: Model images of each type as labelled and described in Table 6.1. The columns correspond to $\alpha = 1$ or $d = 0.04$ (left), $\alpha = 2$ or $d = 0.02$ (center), and $\alpha = 3$ or $d = 0.006$ (right).

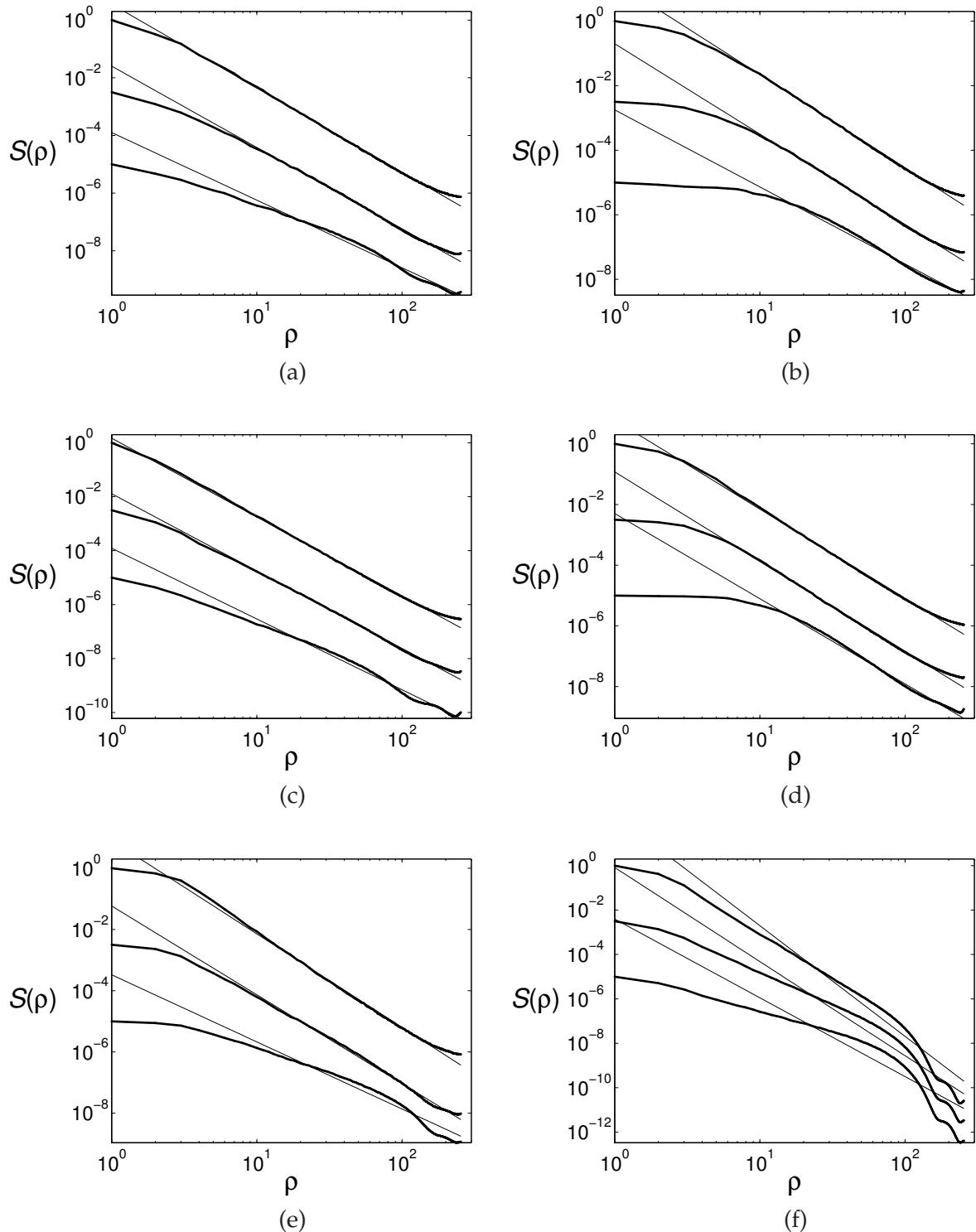


Figure 6.5: Ensemble averaged power spectra (thick lines) for (a) type A, (b) type B, (c) type C, (d) type D, (e) type E, and (f) type F images, for $\alpha = 1$ (top line), $\alpha = 2$ (middle line) and $\alpha = 3$ (bottom line) (for types A, C, E, and F) and for $d = 0.04$ (top line), $d = 0.02$ (middle line) and $d = 0.006$ (bottom line) (for types B and D). Plots have been shifted vertically to eliminate overlap. Linear fits as described in the text are shown by the thin lines.

Table 6.2: Slopes of log–log power spectra.

Image type	$\alpha = 1$ or $d = 0.04$	$\alpha = 2$ or $d = 0.02$	$\alpha = 3$ or $d = 0.006$
A	2.9	2.8	2.3
B	2.8	2.7	2.4
C	2.9	2.8	2.6
D	2.9	2.9	2.8
E	3.0	2.8	2.2
F	4.9	4.2	3.5

The following observations are made by examining Fig. 6.5 and Table 6.2. For image type A, occluding with a power law size distribution, the log–log power spectrum is linear over the entire spatial frequency range, although there is some deviation from linearity at lower spatial frequencies for the larger value of α . For large values of α the size distribution is quite narrow (since a minimum disk radius has been imposed) and the spectrum tends to mimic that of a disk of that size. For image type B, occluding with an exponential size distribution, the log–log power spectrum is linear in the high frequency region but is distinctly nonlinear at lower spatial frequencies. For image types C and D, containing nonoccluding objects with power law and exponential size distributions respectively, the behaviour of the power spectrum is very similar to type A and B images, respectively. For image type E, with variable amplitude occluding objects, the behaviour of the power spectrum is quite similar to that of images of type A and C, although there is more deviation from linearity at lower spatial frequencies, particularly for the larger values of α . For image type F, with variable amplitude transparent objects, the spectrum is quite nonlinear over the whole spatial frequency range. Referring to Table 6.2, the asymptotic slopes γ are in the range 2.3–2.9 for images containing constant amplitude disks (type A, B, C, D images). The slopes are very similar, 2.2–3.0, for type E images that contain occluding variable amplitude disks but do contain edges. For type F images that do not contain edges the log–log power spectra are nonlinear and also fall off much more rapidly with spatial frequency. The slopes tend to decrease with increasing α and decreasing d although the dependence on α or d is quite weak.

6.4 Circularly averaged power spectra of nonoccluding models

In this section, the circularly averaged power spectrum for images containing nonoccluding disks for both power law and exponential size distributions is estimated analytically in order to further understand the numerical results described above. The analytical results here are also compared with those of Ruderman [79].

Consider an image $f(x, y)$ that consists a large number N of nonoccluding disks so that

$$f(x, y) = \sum_{m=1}^N A_m \text{disk}\left(\sqrt{(x - x_m)^2 + (y - y_m)^2} / a_m\right), \quad (6.9)$$

where $\text{disk}(r)$ is the unit disk of unit amplitude, and A_m , a_m , and (x_m, y_m) are the amplitude, radius, and center of the m -th disk, respectively. The two-dimensional Fourier transform of a disk, in polar coordinates, is given by (Eq. (A.8) in Appendix A)

$$\mathcal{F}\{\text{disk}(r)\} = 2\pi \int_0^\infty \text{disk}(r) J_0(2\pi\rho r) r dr, \quad (6.10)$$

where (r, θ) and (ρ, φ) are polar coordinate systems in image and Fourier space, respectively, and the angular variables are suppressed, and $J_n(x)$ is the n -th order Bessel function of the first kind. Evaluation of Eq. (6.10) gives

$$\mathcal{F}\{\text{disk}(r)\} = J_1(2\pi\rho)/\rho = \text{jinc}(\rho), \quad (6.11)$$

which defines the jinc function. The Fourier transform, $F(u, v)$ of the image is then

$$F(u, v) = \sum_{m=1}^N A_m a_m^2 \text{jinc}\left(a_m \sqrt{u^2 + v^2}\right) \exp(-i2\pi(ux_m + vy_m)), \quad (6.12)$$

where (u, v) are Cartesian coordinates in Fourier space. Transforming to polar coordinates in Fourier space, the Fourier transform is

$$F(\rho, \varphi) = \sum_{m=1}^N A_m a_m^2 \text{jinc}(\rho a_m) \exp(-i2\pi\rho(x_m \cos \varphi + y_m \sin \varphi)), \quad (6.13)$$

and the power spectrum is

$$|F(\rho, \varphi)|^2 = \sum_{m=1}^N \sum_{n=1}^N A_m A_n a_m^2 a_n^2 \text{jinc}(\rho a_m) \text{jinc}(\rho a_n) \exp(-i2\pi\rho C_{mn}[\cos(\varphi - \theta_{mn})]), \quad (6.14)$$

where

$$C_{mn} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2} \quad (6.15)$$

and $\tan(\theta_{mn}) = (y_m - y_n)/(x_m - x_n)$. The circularly averaged power spectrum is then

$$\frac{1}{2\pi} \int_0^{2\pi} |F(\rho, \varphi)|^2 d\varphi = \sum_{m=1}^N \sum_{n=1}^N A_m A_n a_m^2 a_n^2 \text{jinc}(\rho a_m) \text{jinc}(\rho a_n) J_0(2\pi\rho C_{mn}). \quad (6.16)$$

The ensemble average $S(\rho)$, normalised by the number of disks, and taking $N \gg 1$, is therefore given by

$$\begin{aligned} S(\rho) &= \left\langle \frac{1}{2\pi N} \int_0^{2\pi} |F(\rho; \varphi)|^2 d\varphi \right\rangle \\ &= \langle A^2 \rangle \langle a^4 \text{jinc}^2(\rho a) \rangle + N \langle A \rangle^2 \langle a^2 \text{jinc}(\rho a) \rangle^2 M(\rho). \end{aligned} \quad (6.17)$$

The function $M(\rho)$ in Eq. (6.17) is independent of the distribution of disk radii and is given by

$$\begin{aligned} M(\rho) &= \langle J_0(2\pi\rho C_{mn}) \rangle \\ &= \int_0^{\sqrt{2}} J_0(2\pi\rho C) P(C) dC, \end{aligned} \quad (6.18)$$

where $P(C)$ is the density function of C_{mn} (given by Eq. (6.15)). For a square image of side length unity, as shown in Fig. 6.6, the maximum value of C_{mn} is $\sqrt{2}$. If the disks were placed at random with a uniform distribution over all locations, then x_m, y_m, x_n and y_n are independent random variables, uniformly distributed on $(0, 1)$.

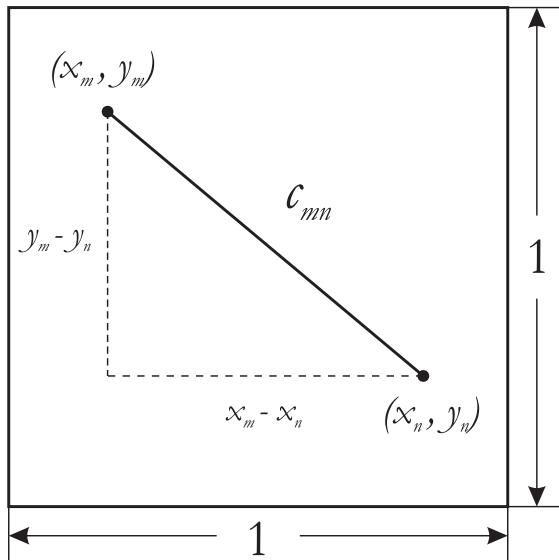


Figure 6.6: Graphical illustration of C_{mn} the distance between the centers of two disks in an image frame of dimensions unity.

$P(C)$ is calculated as follows. First representing \mathbf{C} as a random variable, and setting up new random variables $\mathbf{g}_x, \mathbf{g}_y, \mathbf{z}_x, \mathbf{z}_y$ and \mathbf{w} , which are defined as

$$\begin{aligned}\mathbf{C} &= \sqrt{(\mathbf{x}_m - \mathbf{x}_n)^2 + (\mathbf{y}_m - \mathbf{y}_n)^2} \\ &\equiv \sqrt{\mathbf{g}_x^2 + \mathbf{g}_y^2} \equiv \sqrt{\mathbf{z}_x^2 + \mathbf{z}_y^2} \equiv \sqrt{\mathbf{w}},\end{aligned}\quad (6.19)$$

the density $P(g_x)$ is the convolution of $P(x_m)$ and $P(-x_n)$, which is a triangle function ranging from -1 to 1 given as

$$\begin{aligned}P(g_x) &= P(x_m) \odot P(x_n) \\ &= 1 - |g_x| && \text{for } |g_x| \leq 1 \\ &= 0 && \text{otherwise.}\end{aligned}\quad (6.20)$$

Applying the PDF transformation (Eq. (A.30) in Appendix A.2.5), and using Eqs. (6.19) and (6.20) gives

$$\begin{aligned}P(z_x) &= z_x^{-1/2} - 1 && \text{for } |z_x| \leq 1 \\ &= 0 && \text{otherwise.}\end{aligned}\quad (6.21)$$

$P(g_y)$ and $P(z_y)$ are defined similarly. The density $P(w)$ is the convolution of $P(z_x)$ and $P(z_y)$, i.e.,

$$P(w) = P(z_x) \odot P(z_y) = \int_{-\infty}^{\infty} P_{z_x}(z) P_{z_y}(w-z) dz. \quad (6.22)$$

Substituting Eq. (6.21) into Eq. (6.22) gives

$$P(w) = \int_{0,w-1}^{w,1} \left[(wz - z^2)^{-1/2} + (w-z)^{-1/2} - z^{-1/2} + 1 \right] dz, \quad (6.23)$$

where the intervals $(0, w)$ and $(w-1, 1)$ determine the density $P(w)$ for $0 \leq w \leq 1$ and $1 \leq w \leq 2$, respectively. Evaluating Eq. (6.23) accordingly gives

$$\begin{aligned}P(w) &= \pi - 4\sqrt{w} + w && \text{for } 0 \leq w \leq 1 \\ &= 2[\sin^{-1}(2/w - 1) - 1] + 4\sqrt{w-1} - w && \text{for } 1 \leq w \leq 2.\end{aligned}\quad (6.24)$$

Lastly, applying the pdf transformation (Eq. (A.30) in Appendix A.2.5) gives the density function $P(C)$, where $\mathbf{C} = \sqrt{\mathbf{w}}$, as

$$\begin{aligned}P(C) &= 2\pi C - 8C^2 + 2C^3 && \text{for } 0 \leq C \leq 1 \\ &= 4[\sin^{-1}(2/C^2 - 1) - 1]C + 8C\sqrt{C^2 - 1} - 2C^3 && \text{for } 1 \leq C \leq \sqrt{2},\end{aligned}\quad (6.25)$$

which is plotted in Fig. 6.7. Note that although the density function is represented by two equations for two regions, there is no discontinuity at $C = 1$.

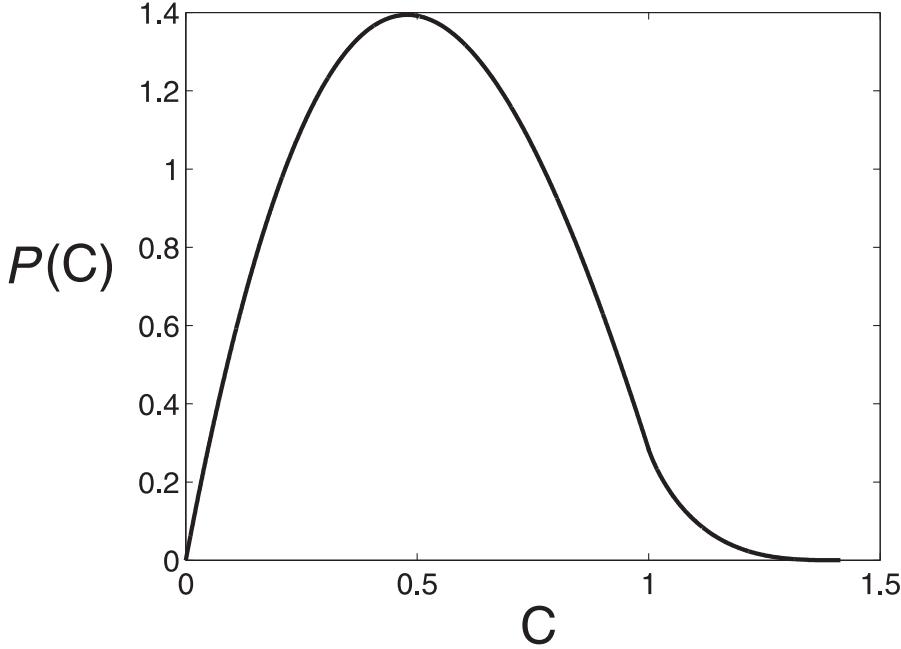


Figure 6.7: The distribution function $P(C)$ (given by Eq. (6.25)).

The density function $P(C)$ of the region 0 to $\sqrt{2}$ makes the integral in Eq. (6.18) difficult to evaluate analytically. However, Eq. (6.18) can be evaluated analytically for $0 \leq C \leq 1$. First let $M(\rho) = M_{0,1}(\rho) + M_{1,\sqrt{2}}(\rho)$, where the subscript denotes the range of integration. Using Eqs. 1.105, 1.109 and 1.106 of Wheelon [116] gives

$$\begin{aligned}
M_{0,1}(\rho) &= \int_0^1 J_0(2\pi\rho C) P(C) dC \\
&= \frac{1}{\rho} J_1(2\pi\rho) - \frac{2}{\pi^2 \rho^2} [J_0(2\pi\rho) s_{1,-1}(2\pi\rho) - J_{-1}(2\pi\rho) s_{2,0}(2\pi\rho)] \\
&\quad + \frac{1}{\pi^2 \rho^2} [(\pi\rho) J_1(2\pi\rho) - J_2(2\pi\rho)] \\
&= I_1(\rho) + I_2(\rho) + I_3(\rho),
\end{aligned} \tag{6.26}$$

where $s_{\mu,\nu}(z)$ is Lommel's function (Eq. (A.61)). The asymptotic form of Eq. (6.26) was studied. Using the asymptotic expansions of Bessel and Lommel's functions (Pgs. 195

and 351 of Watson [117]), each term of Eq. (6.26) is given as

$$\begin{aligned} I_1(\rho) &= \frac{1}{\rho} J_1(2\pi\rho) \\ &= \frac{1}{\pi\rho^{3/2}} \cos(2\pi\rho - 3\pi/4), \quad \rho \rightarrow \infty \\ &= O\left(\rho^{-3/2}\right), \end{aligned} \tag{6.27}$$

$$\begin{aligned} I_2(\rho) &= -\frac{2}{\pi^2\rho^2} [J_0(2\pi\rho)s_{1,-1}(2\pi\rho) - J_{-1}(2\pi\rho)s_{2,0}(2\pi\rho)] \\ &= -\frac{2}{\pi^3\rho^{5/2}} \left[\cos\left(2\pi\rho - \frac{\pi}{4}\right) \left(1 - \frac{1}{(2\pi\rho)^2}\right) - \cos\left(2\pi\rho + \frac{\pi}{4}\right) (2\pi\rho + 1) \right], \quad \rho \rightarrow \infty \\ &= O\left(\rho^{-3/2}\right), \end{aligned} \tag{6.28}$$

and

$$\begin{aligned} I_3(\rho) &= \frac{1}{\pi^2\rho^2} [(\pi\rho)J_1(2\pi\rho) - J_2(2\pi\rho)] \\ &= \frac{1}{\pi^2\rho^2} \left[\rho^{1/2} \cos\left(2\pi\rho - \frac{3\pi}{4}\right) - \frac{1}{\pi\rho} \cos\left(2\pi\rho - \frac{5\pi}{4}\right) \right], \quad \rho \rightarrow \infty \\ &= O\left(\rho^{-3/2}\right), \end{aligned} \tag{6.29}$$

Inspection of Eqs. (6.27) to (6.29) show that

$$M_{0,1}(\rho) = O\left(\rho^{-3/2}\right). \tag{6.30}$$

$M_{1,\sqrt{2}}(\rho)$ is difficult to calculate analytically and so was calculated numerically. Both $M_{0,1}(\rho)$ and $M_{1,\sqrt{2}}(\rho)$ are plotted in Fig. 6.8, and $|M_{0,1}(\rho)|$, $|M_{1,\sqrt{2}}(\rho)|$ and $M(\rho)$ are plotted on a log–log scale in Fig. 6.9. Observing these two figures, it is seen that both terms oscillate almost out of phase, and that $M_{0,1}(\rho)$ is slightly larger in magnitude than $M_{1,\sqrt{2}}(\rho)$ when $\rho \leq 5$ and they are about the same when $\rho > 10$. The slope of the envelope in Fig. 6.9 is $-3/2$, which shows that both terms decay as $\rho^{-3/2}$, which agrees with the asymptotic analysis for $M_{0,1}(\rho)$ above. The envelope of $M(\rho)$ decays linearly on a log–log plot with a slope of -3 (Fig. 6.9). Calculation of $M(\rho)$ for a wide range of values of ρ (Fig. 6.10) showed the same behaviour, giving

$$M_{\max}(\rho) \approx k_1\rho^{-3}, \tag{6.31}$$

where $k_1 \approx 0.065$. Evidently the $\rho^{-3/2}$ behaviours of $M_{0,1}(\rho)$ and $M_{1,\sqrt{2}}(\rho)$ cancel out, giving the residual behaviour ρ^{-3} .

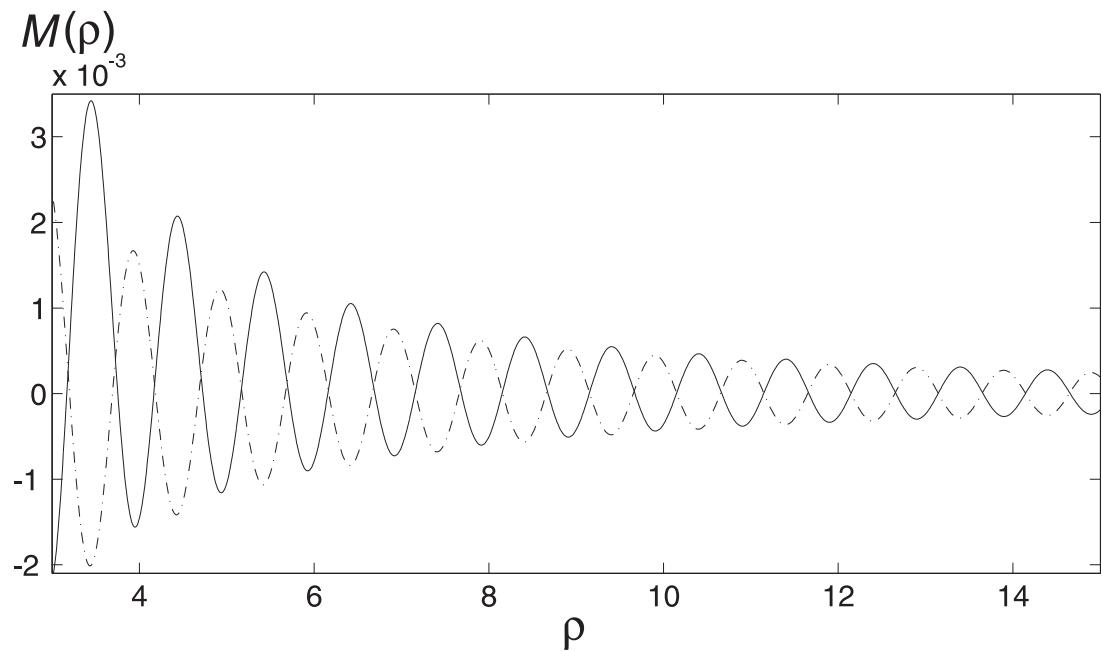


Figure 6.8: $M_{0,1}(\rho)$ (—) and $M_{1,\sqrt{2}}(\rho)$ (---) versus ρ .

$\text{Log}(M(\rho))$

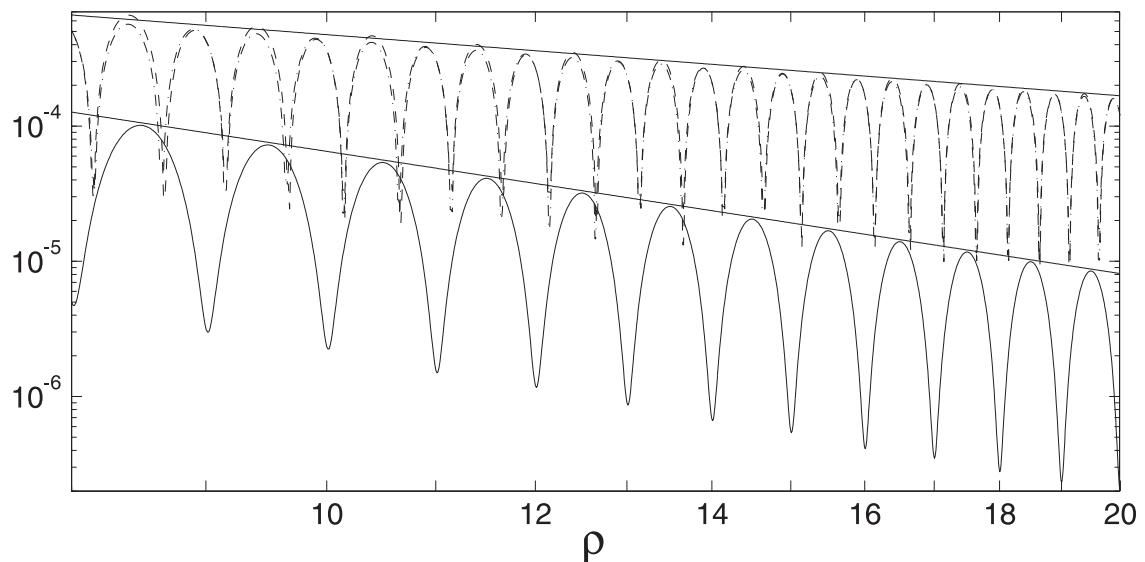


Figure 6.9: $|M_{0,1}(\rho)|$ (---), $|M_{1,\sqrt{2}}(\rho)|$ (- · - · - ·) and $M(\rho)$ (—) versus ρ on a log-log plot.

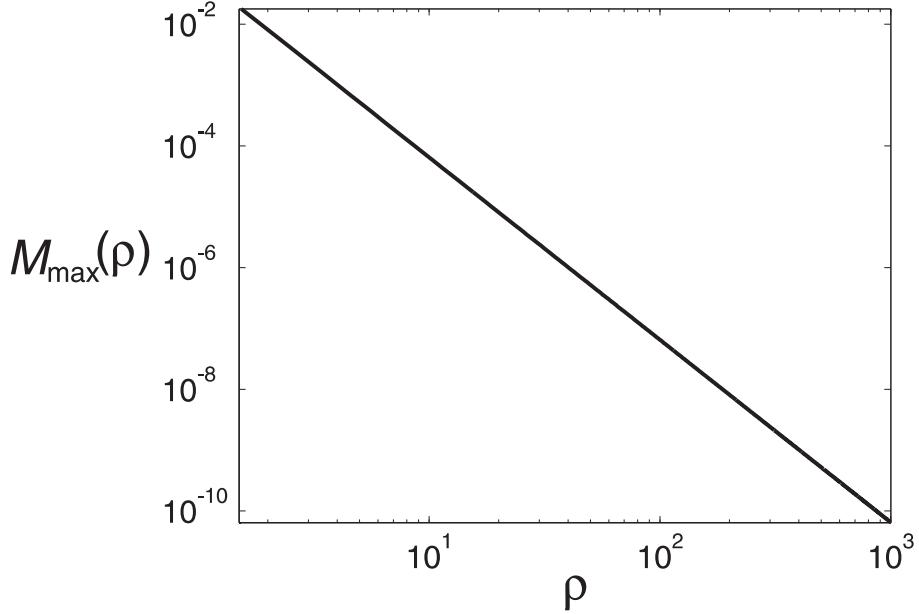


Figure 6.10: Envelope of $M(\rho)$ versus ρ on a log–log plot.

For disk amplitudes uniformly distributed on $(0, 1)$, $\langle A^2 \rangle = 1/3$ and $\langle A \rangle^2 = 1/4$. The quantities $\langle a^4 \text{jinc}^2(\rho a) \rangle$ and $\langle a^2 \text{jinc}(\rho a) \rangle^2$ depend on the form of the distribution of disk radii $P(a)$.

6.4.1 Power law size distribution

Although it is necessary to impose upper and lower limits on a to normalise the density $P(a)$ for a power law size distribution, it is instructive to first ignore the normalisation and evaluate the averages in Eq. (6.17) for a on the interval $(0, \infty)$. For a power law density $\langle a^4 \text{jinc}^2(\rho a) \rangle$ is given by

$$\langle a^4 \text{jinc}^2(\rho a) \rangle_p = k_p \int_{a_1}^{a_2} a^4 \text{jinc}^2(\rho a) a^{-\alpha} da. \quad (6.32)$$

Using Eqs. 6.576.2 and 9.122.1 of Gradshteyn and Ryzhik [108],

$$\begin{aligned} \langle a^4 \text{jinc}^2(\rho a) \rangle_p &\approx \frac{k_p}{\rho^2} \int_0^\infty J_1^2(2\pi\rho a) a^{2-\alpha} da \\ &= \frac{k_p \Gamma(\frac{5-\alpha}{2})}{2^5 \pi^{3-\alpha} \rho^{5-\alpha} \Gamma(\frac{\alpha-1}{2})} F \left[\frac{5-\alpha}{2}, \frac{3}{2}; 3; 1 \right] \\ &= \left(\frac{k_p \Gamma(\alpha)(\alpha^3 - 5\alpha^2 + 7\alpha - 3)}{2^5 \pi^{2-\alpha} (\alpha-2) \cos(\pi\alpha/2) [\Gamma((\alpha+1)/2)]^4} \right) \rho^{\alpha-5} \quad \text{for } 2 < \alpha < 3 \\ &\approx k_2 \rho^{\alpha-5} \quad \text{for } 2 < \alpha < 3, \end{aligned} \quad (6.33)$$

where $\Gamma(\cdot)$ is the gamma function, $F[\cdot, \cdot; \cdot; \cdot]$ is the Gauss' hypergeometric function and the subscript p denotes a power law $P(a)$.

The function $\langle a^2 \text{jinc}(\rho a) \rangle$ for a power law size distribution is given by

$$\langle a^2 \text{jinc}(\rho a) \rangle_p = k_p \int_{a_1}^{a_2} a^2 \text{jinc}(\rho a) a^{-\alpha} da \quad (6.34)$$

and is estimated using $a_1 \rightarrow 0$ and $a_2 \rightarrow \infty$ as follows. Applying Eq. 6.561.14 of Gradshteyn and Ryzhik [108] gives

$$\begin{aligned} \langle a^2 \text{jinc}(\rho a) \rangle_p^2 &\approx \left(\frac{k_p}{\rho} \int_0^\infty J_1(2\pi\rho a) a^{1-\alpha} da \right)^2 \\ &= \left(\frac{k_p ((1-\alpha)\pi^{\alpha-1})}{4 \cos(\pi\alpha/2) [\Gamma((1+\alpha)/2)]^2} \right)^2 \rho^{2\alpha-6} \quad \text{for } 1/2 < \alpha < 3 \\ &\approx k_3 \rho^{2\alpha-6} \quad \text{for } 1/2 < \alpha < 3. \end{aligned} \quad (6.35)$$

Both k_2 and k_3 depend on the limits a_1 and a_2 .

Referring to Eqs. (6.17), (6.31), (6.33) and (6.35) shows that the first term in Eq. (6.17) behaves as $\rho^{\alpha-5}$ and the second term as $N\rho^{2\alpha-9}$. Calculations using relevant values of α , a_1 , a_2 and N show that the first term in Eq. (6.17) dominates over the second term for $\rho > 2$ and $\alpha \leq 2.5$ (both terms diverge as α approaches 3), so that in this approximation $S(\rho)$ behaves as

$$S_p(\rho) \sim \rho^{\alpha-5}. \quad (6.36)$$

Although Eq. (6.36) gives a to the behaviour of $S(\rho)$, a proper calculation requires consideration of the finite limits for a and the averages in Eq. (6.17) cannot then be evaluated analytically. Equation (6.17) was therefore evaluated numerically using $a_1 = 0.004$, $a_2 = 0.4$, $\alpha = 1, 2, 3$ and the relevant values of N . The first term in Eq. (6.17) again dominates and the log-log plot of $S(\rho)$ is shown in Fig. 6.11. Inspection of the figure shows that the log $S(\rho)$ versus log ρ plot is quite linear for $\alpha = 1$ and $\alpha = 2$ with slopes of 3.0 and 2.9 respectively. For $\alpha = 3$, the log-log plot is only approximately linear with a slope of 2.7. Comparison with Fig. 6.5(c) shows that this analysis matches the numerical simulations very well.

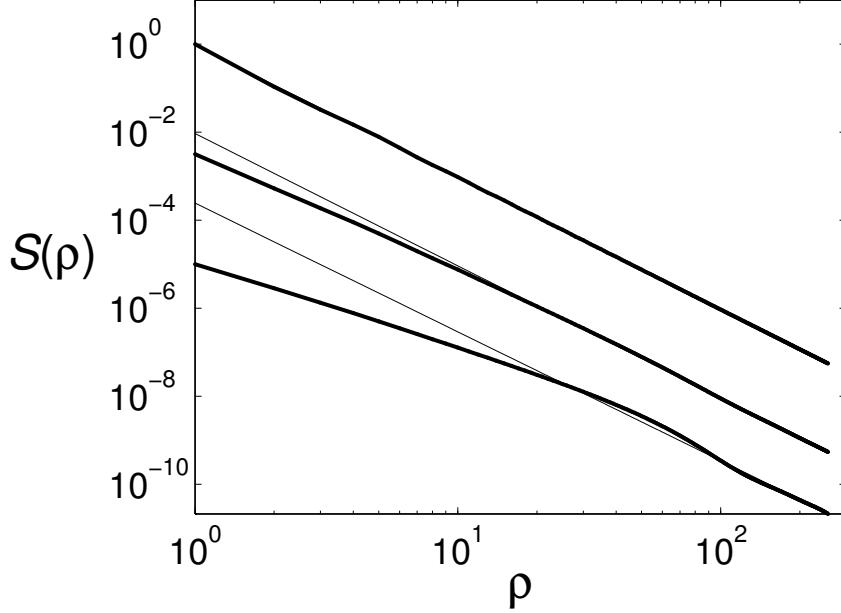


Figure 6.11: Power spectra for images containing nonoccluding disks with a power law size distribution calculated using Eq. (6.17) for \$\alpha = 1\$ (top line), \$\alpha = 2\$ (middle line) and \$\alpha = 3\$ (bottom line). Thin lines show regression lines.

6.4.2 Exponential size distribution

The function $\langle a^4 \text{jinc}^2(\rho a) \rangle$ for an exponential size distribution is given by

$$\langle a^4 \text{jinc}^2(\rho a) \rangle_e = k_e \int_{a_1}^{a_2} a^4 \text{jinc}^2(\rho a) \exp(-a/d) da. \quad (6.37)$$

This density can be normalised for $a_1 = 0$ and $a_2 = \infty$ and is calculated initially using these values. Referring to Eq. 6.626.2 of Gradshteyn and Ryzhik [108],

$$\begin{aligned} \langle a^4 \text{jinc}^2(\rho a) \rangle_e &= \frac{1}{d\rho^2} \int_0^\infty a^2 J_1^2(2\pi\rho a) \exp(-a/d) da \\ &= \frac{3\pi}{d} \int_0^{\pi/2} \frac{\cos^2 \theta}{(1/4d^2 + 4\pi^2\rho^2 \cos^2 \theta)^{5/2}} d\theta, \end{aligned} \quad (6.38)$$

where the subscript e denotes an exponential $P(a)$. Equation (6.38) was evaluated using *Mathematica* [109], which gives

$$\langle a^4 \text{jinc}^2(\rho a) \rangle_e = \frac{2d^2}{\pi(1+x^2)^{5/2}\rho^2} \left[(x^4 - 1) E\left(\frac{x^2}{1+x^2}\right) + (1+x^2) K\left(\frac{x^2}{1+x^2}\right) \right], \quad (6.39)$$

where $x = 4\pi d\rho$, and $K(\cdot)$ and $E(\cdot)$ are the complete elliptic integrals of the first and second kind, respectively. Using Eq. 17.3.26 of Abramowitz and Stegun [118] shows that

$$K\left(\frac{x^2}{1+x^2}\right) \sim 2\ln(2) + \ln(x) \quad x \rightarrow \infty. \quad (6.40)$$

Using Eq. (6.40) and noting that $E(1) = 1$ shows that for large ρ the first term in Eq. (6.39) dominates so that

$$\langle a^4 \text{jinc}^2(\rho a) \rangle_e = (d/2\pi^2)\rho^{-3} \quad \text{for } \rho \gg 1/4\pi d. \quad (6.41)$$

The function $\langle a^2 \text{jinc}(\rho a) \rangle$ for an exponential size distribution is given by

$$\langle a^2 \text{jinc}(\rho a) \rangle_e^2 = k_e \int_{a_1}^{a_2} a^2 \text{jinc}(\rho a) \exp(-a/d) da. \quad (6.42)$$

This is also evaluated, using $a_1 = 0$, $a_2 = \infty$ and Eq. 6.623.1 of Gradshteyn and Ryzhik [108] to give

$$\begin{aligned} \langle a^2 \text{jinc}(\rho a) \rangle_e^2 &= \left(\frac{1}{d\rho} \int_0^\infty a J_1(\rho a) \exp(-a/d) da \right)^2 \\ &= \left(\frac{2\pi}{d(1/d^2 + 4\pi^2\rho^2)^{3/2}} \right)^2 \\ &= (16\pi^4 d^2)^{-1} \rho^{-6} \quad \text{for } \rho \gg 1/2\pi d. \end{aligned} \quad (6.43)$$

Equations (6.41) and (6.43) show that the averages in Eq. (6.17) are power law in ρ only for large spatial frequencies when $\rho \gg 1/4\pi d$ and $\rho \gg 1/2\pi d$ for Eqs. (6.41) and (6.43), respectively. For smaller values of ρ both terms are distinctly non power law. Referring to Eqs. (6.17), (6.31), (6.41) and (6.43) shows that the first and second terms in Eq. (6.17) behave as ρ^{-3} and ρ^{-9} , respectively, for large ρ . The first term dominates over the second so that the overall behaviour is

$$S_e(\rho) \sim \rho^{-3} \quad \text{for } \rho \gg 1/4\pi d. \quad (6.44)$$

For finite limits on a the averages need to be calculated numerically. The first term in Eq. (6.17) dominates over the second for $\rho > 2$ and the results of the calculation for the values of a_1 , a_2 and d used above are shown in Fig. 6.12. Inspection of the figure shows linear behaviour on a log–log plot for $\rho > 1/4\pi d$ with a slope of -3 , but nonlinear behaviour for smaller spatial frequencies. Comparison of Fig. 6.12 with Fig. 6.5(d) shows good agreement.

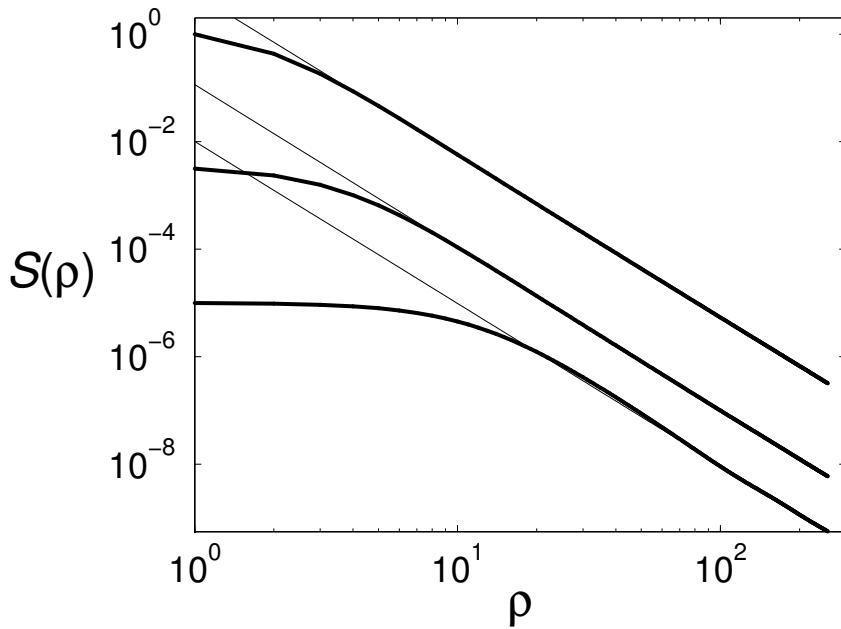


Figure 6.12: Power spectra for images with exponentially distributed disk sizes for $d = 0.04$ (top line), $d = 0.02$ (middle line) and $d = 0.006$ (bottom line) calculated numerically as described in the text. These lines show asymptotic linear fits with slopes of -3.

In summary, the expected behaviour of the circularly averaged power spectrum is expressed in Eq. (6.17). This expression can be evaluated approximately for $a_1 = 0$ and $a_2 = \infty$ for both power law and exponential size distributions (Eqs. (6.36) and (6.44)), showing linear behaviour over all spatial frequencies only for the power law size distribution. Numerical calculation of Eq. (6.17) with finite limits and appropriate parameters for power law and exponential size distributions (Figs. 6.11 and 6.12) show the same kind of behaviour and match the simulation results shown in Figs. 6.5(c) and 6.5(d).

6.5 Comparison with Ruderman's results

Ruderman [79] has calculated the correlation properties of images containing occluding disks with a power law size distribution and, by implication, the behaviour of the power spectrum for such images. The relationship between the power spectrum power law exponent γ and the exponent of the power law object size distribution α can be derived from his analysis, although this was not mentioned specifically. The derivation of the relationship is described below and compared with the results of the analysis for images containing nonoccluding disk as described in Section 6.4. Ruderman first defined the correlation

function $C(x)$ of an image as

$$C(x) = \left\langle \langle \phi(x_0) \phi(x_0 + x) \rangle_\theta \right\rangle_{x_0} \Big|_\phi, \quad (6.45)$$

where the $\phi(x)$ is the pixel value at position x and the averaging is performed over all images ϕ , initial positions x_0 and displacements of length x and angle θ . Using the Fourier relationship between the correlation function and the power spectrum, and applying the power spectrum characteristics of natural images, i.e., $S(\rho) \sim \rho^{-\gamma}$, the correlation function $C(x)$ of natural images behaves as (see appendix of [79])

$$C(x) = -C_1 + C_2 x^{\gamma-2}, \quad (6.46)$$

where C_1 and C_2 are constants. Using several natural images, the boundaries of objects within each image were determined manually, and $C(x)$ calculated for pairs of points both within the same object and in different objects. This showed that the correlation is higher for a pair of points within the same object (than for points in different objects) and is independent of the distance between the two points. He then categorised the correlation into two groups, namely the correlation for points in the same object, $C_{\text{same}}(x)$, and the correlation for points in different objects, $C_{\text{diff}}(x)$, i.e.,

$$C(x) = P_{\text{same}}(x) C_{\text{same}}(x) + [1 - P_{\text{same}}(x)] C_{\text{diff}}(x), \quad (6.47)$$

where $P_{\text{same}}(x)$ is the probability that a pair of points separated by x is within a single object. Since the data showed no particular trend for $C_{\text{same}}(x)$ and $C_{\text{diff}}(x)$, they were replaced by constants,

$$\begin{aligned} C(x) &= C_{\text{diff}} - (C_{\text{diff}} - C_{\text{same}}) P_{\text{same}}(x) \\ &= C_0 P_{\text{same}}(x), \end{aligned} \quad (6.48)$$

where C_0 is a constant, and Eq. (6.48) shows that $P_{\text{same}}(x)$ is proportional to $C(x)$.

He then considered an image made up of occluding circular objects with a power law size distribution with exponent $-\alpha$. Calculating $P_{\text{same}}(x)$ gave

$$P_{\text{same}}(x) \sim x^{3-\alpha}. \quad (6.49)$$

Using Eqs. (6.46), (6.48) and (6.49) shows that the relationship between γ and α is

$$\gamma = 5 - \alpha. \quad (6.50)$$

Interestingly, the relationship (6.50) is identical to that in Eq. (6.36), despite the quite different approaches and the different models (occluding and nonoccluding). This indicates

that occlusion has little effect on the power spectra of the model images, and provides further evidence suggesting that occlusion may not be necessary to obtain power law power spectra in natural images.

6.6 Discussion

The results presented in Section 6.3 allow the effects of various image characteristics on the dependence of the power spectrum on spatial frequency to be determined using this simple model of natural images. For the primary model, Image Type A, of self-similar images containing constant amplitude occluding objects, the log–log power spectra are linear over a wide range of spatial frequencies with a slope of ~ 2.8 . There is some deviation from linearity for the larger value $\alpha = 3$, although this is an artifact of the imposed lower limit of object sizes. In this case, the large value of α produces a narrow size distribution violating self-similarity, and the transform approaches that of a single object. For self-similar images containing nonoccluding, constant amplitude objects there is little difference in the power spectrum behaviour to that for occluding objects. Note that images containing constant amplitude nonoccluding objects do contain edges, and in fact tend to contain more edges than do images with occluding objects, but of smaller amplitude. This indicates therefore that occlusion is not an essential ingredient for the observed power spectrum behaviour, but that edges are important. This is supported by the results shown in Fig. 6.5 (e) for self-similar images that contain occluding, variable amplitude objects. These images contain edges and, despite the fact that the objects are not constant amplitude, the power spectra are relatively linear, particularly for the smaller values of α . The small amount of nonlinearity introduced is probably due to modulation of the image spectrum by the spectrum of the objects themselves. The quite different behaviour of the power spectra of images that do not contain edges is clearly evident in Fig. 6.5 (f). In this case the log–log power spectra are distinctly nonlinear, despite the image being self-similar. For high spatial frequencies the spectra fall off rapidly with frequency as a result of the objects, and hence the images, having a continuous first derivative. The log–log power spectra of images containing objects with an exponential size distribution, whether occluding or nonoccluding, show distinctly nonlinear behaviour at the lower spatial frequencies (Figs. 6.5 (b) and 6.5 (d)), despite these images containing edges.

The results allow the important qualities of natural images that lead to the characteristic behaviour of the power spectra to be isolated. The results indicate that edges are an essential ingredient but occlusion is not, although edges are often a result of occlusion. Constant amplitude objects are not essential although variations in amplitude do introduce some nonlinearity into the log–log power spectrum. Self-similarity is necessary for a power law power spectrum over a wide range of spatial frequencies. Derivations from

self-similarity (due either to limits on the object size or to non power law size distributions) introduce nonlinearity into the log–log power spectrum, particularly at low spatial frequencies. Strictly speaking, these results show only that edges and scaling are sufficient conditions for power law power spectra within this image model. However, the results for images without edges or without scaling suggest that edges and scaling are necessary conditions. Real natural images have a lot of other structure not included in the used image model that might contribute to the shape of power spectra. However, it is likely that the importance of edges and self-similarity carry over into real natural images.

It is interesting to note that for self-similar images both the analysis in the Section 6.4 (Eq. (6.36)) and that of Ruderman [79] (Section 6.5, Eq. (6.50)) suggest a strong dependence of γ on α ($\gamma = 5 - \alpha$). However the simulation results (Table 6.2) show only a weak dependence of γ on α . This is probably due to the restricted range of object sizes used and the finite image frame, as the above results are based on full self-similarity (an infinite range of object sizes). Also, the numerical results using Eq. (6.17) with a finite range of object sizes also show the weak dependence of γ on α . Lastly, the value $\gamma \approx 2.8$ obtained from the simulations is somewhat larger than is typically seen in natural images although this is probably due to the particular simple model that has been used.

Chapter 7

Conclusions and suggestions for future research

In many imaging applications, particularly diffraction imaging but also others, images are reconstructed from their spatial Fourier transforms. A number of topics related to the effects of errors in the magnitude and phase of measured transforms have been investigated. The effects of errors in the amplitude and phase on reconstructed images have been quantitated in terms of the mse. The relative effects of amplitude and phase errors have been identified. Equivalent amplitude and phase errors have been defined. Large amplitude errors lead to an increase in the energy of reconstructed images and to saturation of the amplitude. These effects have been modelled and expressions for the mse derived. The mse in that case depends on the amplitude and phase error distributions, and the image Fourier amplitude distribution, although not sensitively so that the results are generally applicable. Image Fourier amplitude distributions play an important role in these studies and a simple image model has been used to derive a model of these distributions which fits image data well. The phenomenon of phase dominance has been quantitatively explained by analysis of the resulting Fourier amplitude and phase errors, and reported counter-examples have been shown to be erroneous. A simple model of natural images has been used to show that the key characteristics of natural images that lead to the characteristic behaviour of the power spectrum with spatial frequency are edges and self-similarity, and that occlusion is not a necessary ingredient.

There are a number of open questions and extensions that are worthy of further research. In some imaging applications the Fourier amplitude errors can be expected to track the overall behaviour of the amplitude measurements with spatial frequency (as mentioned in Chapter 2). Analysis of the reconstructed image mse, equivalent amplitude and phase errors, and the relative effects of amplitude and phase errors, in this case would be worth-

while. The image Fourier amplitude distributions derived in Chapter 4 take a rather complicated and inconvenient form, and a more convenient approximation would be useful. Extension of the analysis of the mse for phase dominance in the case where the two images have different spatial frequency behaviours would also be worthwhile. The simple models of natural images used in Chapter 6 may be useful for psychophysical experiments to study aspects of visual processing.

Appendix A

Relevant mathematics and probability

A.1 Fourier transform

The Fourier transform of a continuous one dimensional function $f(x)$ is defined as [119]

$$\mathcal{F}\{f(x)\} = F(u) = \int_{-\infty}^{\infty} f(x) \exp(-i2\pi ux) dx, \quad (\text{A.1})$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform operation. In this thesis the upper-case letter is generally used to denote the Fourier transform of a function denoted by the corresponding lower-case letter. Given $F(u)$, the original signal is recovered by the inverse Fourier transform defined as

$$\mathcal{F}^{-1}\{F(u)\} = \int_{-\infty}^{\infty} F(u) \exp(i2\pi ux) du, \quad (\text{A.2})$$

where $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse Fourier transform. The Fourier transform operation is defined for any square-integrable function $f(x)$. A Fourier transform pair is denoted by

$$f(x) \iff F(u). \quad (\text{A.3})$$

The two-dimensional forward and inverse Fourier transforms are defined as

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp(-i2\pi(ux + vy)) dx dy \quad (\text{A.4})$$

and

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) \exp(i2\pi(ux + vy)) du dv, \quad (\text{A.5})$$

where (x, y) and (u, v) are Cartesian coordinates in spatial and Fourier space, respectively. The two dimensional forward and inverse Fourier transform in polar coordinates are easily derived from Eqs. (A.4) and (A.5) and are given by

$$F(\rho, \varphi) = \int_0^\infty \int_0^{2\pi} f(r, \theta) \exp(-i2\pi\rho r \cos(\theta - \varphi)) r \, dr \, d\theta \quad (\text{A.6})$$

and

$$f(r, \theta) = \int_0^\infty \int_0^{2\pi} F(\rho, \varphi) \exp(i2\pi\rho r \cos(\theta - \varphi)) \rho \, d\rho \, d\varphi. \quad (\text{A.7})$$

The two-dimensional Fourier transform of a circularly symmetric function denoted $f(r)$ (also known as the Hankel transform) can be determined by evaluating the θ integration in Eq. (A.6) giving

$$F(\rho) = 2\pi \int_0^\infty f(r) J_0(2\pi\rho r) r \, dr, \quad (\text{A.8})$$

where $J_0(x)$ is zeroth order Bessel function of the first kind Section A.3.1. The Hankel transform is self reciprocal, i.e., the forward and inverse kernels are identical. For example, the Hankel transform of a disk is

$$\mathcal{F}\{\text{disk}(r)\} = 2\pi \int_0^\infty \text{disk}(r) J_0(2\pi\rho r) r \, dr = \text{jinc}(\rho), \quad (\text{A.9})$$

where $\text{disk}(\cdot)$ is defined as

$$\begin{aligned} \text{disk}(r) &= 1 & \sqrt{x^2 + y^2} \leq r \\ &= 0 & \text{otherwise,} \end{aligned} \quad (\text{A.10})$$

and $\text{jinc}(x)$ is defined by Eq. (A.67).

The two-dimensional discrete forward and inverse Fourier transforms (DFT) are defined as [119]

$$F(k) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) \exp\left(-\frac{i2\pi kx}{N}\right) \quad (\text{A.11})$$

and

$$f(x) = \sum_{k=0}^{N-1} F(k) \exp\left(\frac{i2\pi kx}{N}\right) \quad (\text{A.12})$$

for N samples of data.

A.1.1 Properties of Fourier transform

The following theorems are easily derived from the definition of the Fourier transform [119].

Shift theorem

$$\mathcal{F}\{f(x-a, y-b)\} = F(u, v) \exp(i2\pi(ua + vb)), \quad (\text{A.13})$$

where a and b are real constants.

Similarity theorem

$$f(ax, by) \iff \frac{1}{|ab|} F\left(\frac{u}{a}, \frac{v}{b}\right), \quad (\text{A.14})$$

where a and b are constants.

Parseval's theorem

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x, y)|^2 dx dy = \iint_{-\infty}^{\infty} |F(u, v)|^2 du dv. \quad (\text{A.15})$$

Convolution theorem The convolution of two functions $f(x, y)$ and $g(x, y)$ is defined as

$$f(x, y) \odot g(x, y) = \iint_{-\infty}^{\infty} f(x', y') g(x - x', y - y') dx' dy', \quad (\text{A.16})$$

where \odot denotes convolution. The convolution theorem for Fourier transforms states that

$$f(x, y) \odot g(x, y) \iff F(u, v)G(u, v), \quad (\text{A.17})$$

i.e., convolution in one domain corresponds to multiplication in the other domain.

Autocorrelation theorem The (cross) correlation of two functions $f(x, y)$ and $g(x, y)$ is

$$f(x, y) \otimes g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') g(x + x', y + y') dx' dy', \quad (\text{A.18})$$

where \otimes denotes the correlation operator. Correlation is related to convolution by

$$f(x, y) \otimes g(x, y) = f(x, y) \odot g(-x, -y). \quad (\text{A.19})$$

The correlation theorem for Fourier transforms is

$$f(x, y) \otimes g(x, y) \iff F(u, v)^* G(u, v). \quad (\text{A.20})$$

The correlation of a function with itself is called autocorrelation, and it follows that

$$f(x, y) \otimes f(x, y) \iff |F(u, v)|^2, \quad (\text{A.21})$$

where $|F(u, v)|^2$ is the power spectrum.

A.2 Probability

A.2.1 Probability theory

Probability theory is the study of mathematical models of random phenomena. To describe a random model, a sample space \mathcal{S} and events are defined as the set of all possible outcomes and a subset of the sample space \mathcal{S} , respectively. A number $P(A)$, called the probability of the event A , is assigned to each event A , which describes the frequency with which event A occurs. Probability can also be defined to describe the rate of occurrence of multiple events, for example, $P(A, B)$ is the joint probability of the event A and B . The probability must satisfy the following conditions.

1. $0 \leq P(A) \leq 1$.
2. $P(\mathcal{S}) = 1$.
3. If the events A and B are mutually exclusive, i.e., $P(A, B) = 0$, then the probability of either events A or B occurring is $P(A \cup B) = P(A) + P(B)$.

A.2.2 Conditional probability and Bayes' theorem

The conditional probability $P(A|B)$, the probability of the event A occurring given that event B has occurred, is defined by

$$P(A|B) = \frac{P(A, B)}{P(B)}, \quad (\text{A.22})$$

provided that $P(B) > 0$.

Given $P(A)$, $P(B)$ and $P(B|A)$, the conditional probability $P(A|B)$ can be calculated from

Bayes' theorem as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (\text{A.23})$$

If the events A and B are statistically independent, i.e.,

$$P(A, B) = P(A)P(B), \quad (\text{A.24})$$

then $P(A|B) = P(A)$.

A.2.3 Random Variables

A random variable is a function $x(x)$ that assign a probability to each possible outcome x . Unlike other mathematical variables, a random variable does not predict/describe the actual outcome of a particular experiment, but rather describes the possibility of each outcome for any experiment that fits the same criteria. A random variable x satisfies the following,

1. The set $\{x \leq x'\}$ is an event for every $x' \in \mathbb{R}$.
2. $P(x = \infty) = P(x = -\infty) = 0$.

Random variables can be discrete or continuous, depending on the nature of the outcome.

For a discrete random variable x , the probability of outcome x is denoted by

$$P_x(x) = P_x(x = x). \quad (\text{A.25})$$

For a discrete random variable, $P_x(x)$ must satisfy the following two conditions,

1. $0 \leq P_x(x) \leq 1$.
2. $\sum_x P_x(x) = 1$.

Continuous random variables are described by the cumulative distribution function defined as

$$Q_x(x) = P_x(x \leq x), \quad (\text{A.26})$$

or the probability density function (PDF) defined as

$$P_x(x) = \frac{dQ_x(x)}{dx}. \quad (\text{A.27})$$

A probability density function must satisfy

1. $\int_{x_1}^{x_2} P_{\mathbf{x}}(x) dx = P(x_1 < \mathbf{x} \leq x_2).$
2. $\int_{-\infty}^{\infty} P_{\mathbf{x}}(x) = 1.$
3. $Q_{\mathbf{x}}(x) = \int_{-\infty}^x P_{\mathbf{x}}(x') dx'.$

A.2.4 Expected value and Variance

The expected value μ_x of a random variable \mathbf{x} is defined as

$$\mu_x = \langle x \rangle = \int_{-\infty}^{\infty} x P_{\mathbf{x}}(x) dx, \quad (\text{A.28})$$

where $\langle \cdot \rangle$ denotes the expected value operator, and the variance is

$$\sigma^2(x) = \langle (x - \langle x \rangle)^2 \rangle = \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 P_{\mathbf{x}}(x) dx. \quad (\text{A.29})$$

A.2.5 PDF of a function of a random variable

Given two random variables \mathbf{x} and \mathbf{y} where $\mathbf{y} = g(\mathbf{x})$, the PDF $P_{\mathbf{y}}(y)$ is given by [120]

$$P_{\mathbf{y}}(y) = \sum_{i=1}^n \frac{P_{\mathbf{x}}(x_i)}{|g'(x_i)|}, \quad (\text{A.30})$$

where x_i are the roots of $y = g(x)$ and $g'(x_i) = \left. \frac{dy}{dx} \right|_{x=x_i} \neq 0$.

A.2.6 Joint probability density function

Given two continuous random variables \mathbf{x} and \mathbf{y} , the joint cumulative distribution function is defined as

$$Q_{\mathbf{x},\mathbf{y}}(x, y) = P_{\mathbf{x},\mathbf{y}}(\mathbf{x} \leq x, \mathbf{y} \leq y), \quad (\text{A.31})$$

and the probability joint density function $P_{\mathbf{x},\mathbf{y}}(x, y)$ is defined as

$$P_{\mathbf{x},\mathbf{y}}(x, y) = \frac{d^2 Q_{\mathbf{x},\mathbf{y}}(x, y)}{dx dy}. \quad (\text{A.32})$$

The conditional density and Bayes' theorem for density functions are given by

$$P(x | y) = \frac{P_{\mathbf{x},\mathbf{y}}(x, y)}{P(y)} = \frac{P(y | x) P(x)}{P(y)}. \quad (\text{A.33})$$

In an experiment with multiple random variables, the density function of each individual

random variable is called the marginal density, defined as

$$P(x) = \int_{-\infty}^{\infty} P_{\mathbf{x},\mathbf{y}}(x, y) dy. \quad (\text{A.34})$$

Two random variables \mathbf{x} and \mathbf{y} are statistically (mutually) independent if and only if

$$P_{\mathbf{x},\mathbf{y}}(x, y) = P(x)P(y). \quad (\text{A.35})$$

A.2.7 Density function for the sum of 2 random variables

Given two random variable \mathbf{x} and \mathbf{y} , the density of $\mathbf{z} = \mathbf{x} + \mathbf{y}$ is given by

$$P_{\mathbf{z}}(z) = \int_{-\infty}^{\infty} P_{\mathbf{x},\mathbf{y}}(z - y, y) dy. \quad (\text{A.36})$$

If \mathbf{x} and \mathbf{y} are independent, then

$$\begin{aligned} P_{\mathbf{z}}(z) &= \int_{-\infty}^{\infty} P_{\mathbf{y}}(y) P_{\mathbf{x}}(z - y) dy, \\ &= P_{\mathbf{x}}(x) \odot P_{\mathbf{y}}(y), \end{aligned} \quad (\text{A.37})$$

i.e., the density of the sum of two independent variables is the convolution of the two density functions.

A.2.8 Covariance and Correlation

The covariance between two random variables \mathbf{x} and \mathbf{y} , with expected values μ_x and μ_y is defined as

$$\text{cov}(\mathbf{x}, \mathbf{y}) = \langle (\mathbf{x} - \mu_x)(\mathbf{y} - \mu_y) \rangle = \langle \mathbf{x}\mathbf{y} \rangle - \mu_x\mu_y. \quad (\text{A.38})$$

The correlation ρ_{xy} , also called correlation coefficient, between two random variables \mathbf{x} and \mathbf{y} with expected values μ_x and μ_y and standard deviations σ_x and σ_y is defined as

$$\rho_{xy} = \frac{\text{cov}(\mathbf{x}, \mathbf{y})}{\sigma_x \sigma_y} = \frac{\langle \mathbf{x}\mathbf{y} \rangle - \mu_x\mu_y}{\sigma_x \sigma_y}. \quad (\text{A.39})$$

A positive correlation indicates an increasing relationship and the negative correlation means a decreasing relationship between the two variables. The two variables are uncorrelated if the correlation is zero. If two random variables \mathbf{x} and \mathbf{y} are independent,

then

$$\langle \mathbf{x}\mathbf{y} \rangle = \langle \mathbf{x} \rangle \langle \mathbf{y} \rangle, \quad (\text{A.40})$$

which means that the covariance is zero and \mathbf{x} and \mathbf{y} are uncorrelated. However, the reverse is not always true.

A.2.9 Gaussian distribution

The PDF of a Gaussian or normally distributed random variable is

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad (\text{A.41})$$

where μ and σ are the mean and the standard deviation, respectively. Uncorrelated normally distributed random variables are independent.

A.2.10 Rayleigh distribution

The Rayleigh PDF is given by

$$P(x) = \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right). \quad (\text{A.42})$$

Rayleigh distribution often arises when two random variables \mathbf{x} and \mathbf{y} are independent, identically and normally distributed with zero mean. The variable $\rho = \sqrt{\mathbf{x}^2 + \mathbf{y}^2}$ is then Rayleigh distributed.

A.2.11 Central limit theorem

The Central limit theorem states that for mutually statistically independent random variables $\mathbf{x}_1, \dots, \mathbf{x}_n$, each with zero mean and finite variances $\sigma_1^2, \dots, \sigma_n^2$, the density of their sum

$$\mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n, \quad (\text{A.43})$$

with appropriate normalisation, approaches the normal distribution

$$P(x) \simeq \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (\text{A.44})$$

as n tends to infinity and the individual variances are small compared to $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. The Central limit theorem validates the use of Gaussian distributions in many situations.

A.2.12 Maximum likelihood and Maximum *a posteriori* estimation

In many applications one needs to estimate certain parameters from a set of observed values of a random variable. Such estimates can be defined in various ways. Let $\hat{\theta}$ denote an estimate of the parameter θ , given observed data \mathbf{x} . The likelihood function is the conditional probability function of x_i given θ

$$L(x_i; \theta) = P(x_i | \theta), \quad (\text{A.45})$$

and it can be interpreted as given that θ is the true value of the unknown parameter what is the probability that the outcome x_i is observed. The maximum likelihood estimate of θ given a set of observed data $\{x_1, x_2, \dots, x_N\}$ is defined as

$$\hat{\theta}_{ML} = \arg \max_{\theta} [L(x_1, x_2, \dots, x_N; \theta)], \quad (\text{A.46})$$

which gives the estimate $\hat{\theta}$ such that the likelihood function is maximised, where the likelihood function is

$$L(x_1, x_2, \dots, x_N; \theta) = \prod_{n=1}^N P(x_n | \theta). \quad (\text{A.47})$$

It is often convenient to maximise the log likelihood

$$\log L(x_1, x_2, \dots, x_N; \theta) = \sum_{n=1}^N P(x_n | \theta) \quad (\text{A.48})$$

An alternative estimates is the maximum *a posteriori* estimate which is obtained by maximising the posterior density $P(\theta|x)$, i.e.,

$$\hat{\theta}_{MAP} = \arg \max_{\theta} [P(\theta|x)]. \quad (\text{A.49})$$

Using Bayes' theorem (Eq. (A.23)), Eq. (A.49) becomes

$$\hat{\theta}_{MAP} = \arg \max_{\theta} \left[\frac{P(x|\theta)P(\theta)}{P(x)} \right]. \quad (\text{A.50})$$

$P(x)$ is a normalising constant and is independent of θ so that

$$\hat{\theta}_{MAP} = \arg \max_{\theta} [P(x|\theta)P(\theta)]. \quad (\text{A.51})$$

The density $P(\theta)$ is known as a prior density since it contains independent information on the parameter θ . The MAP estimate can be more useful than the ML estimate if reliable

prior information on θ is available.

A.2.13 Minimum mean squared error estimate

The minimum mean squared error estimate (MMSE) is defined as

$$\begin{aligned}\hat{\theta}_{MMSE} &= \arg \min_{\theta} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\hat{\theta} - \theta)^2 P(x, \theta) dx d\theta \right] \\ &= \int_{-\infty}^{\infty} \theta P(\theta) P(x|\theta) d\theta.\end{aligned}\quad (\text{A.52})$$

If the data x are Gaussian distributed, then $\hat{\theta}_{MMSE} = \hat{\theta}_{ML}$.

A.3 Special Functions

A number of special function are used throughout this thesis. Their definitions and relevant properties are stated here.

A.3.1 Bessel functions of the first kind

The n -th order Bessel function of the first kind is defined as

$$J_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-ni\theta + ix \sin \theta) d\theta. \quad (\text{A.53})$$

The asymptotic expansion for large x is (Page 195 of [117])

$$J_n(x) \sim \left(\frac{2}{\pi x} \right)^{\frac{1}{2}} \left[\cos \left(x - \frac{(-2n-1)\pi}{4} \right) \right], \quad x \rightarrow \infty. \quad (\text{A.54})$$

A.3.2 Complete elliptic integrals

The complete elliptic integral of the first kind is defined as [108]

$$K(m) = \int_0^1 [(1-t^2)(1-mt^2)]^{-1/2} dt = \int_0^{\pi/2} (1-m \sin^2 \theta)^{-1/2} d\theta \quad (\text{A.55})$$

and of the second kind as

$$E(m) = \int_0^1 (1-t^2)^{-1/2} (1-mt^2)^{1/2} dt = \int_0^{\pi/2} (1-m \sin^2 \theta)^{1/2} d\theta. \quad (\text{A.56})$$

A.3.3 Gamma function

The Gamma function is defined as [108]

$$\Gamma(z) = \int_0^\infty \exp(-t)t^{z-1} dt, z > 0, \quad (\text{A.57})$$

and the incomplete gamma function $\Gamma(\alpha, x)$ as

$$\Gamma(\alpha, x) = \int_x^\infty \exp(-t)t^{\alpha-1} dt. \quad (\text{A.58})$$

A.3.4 Hypergeometric function

The generalised hypergeometric function is defined as

$${}_A\text{F}_B [a_1, a_2, \dots, a_A; b_1, b_2, \dots, b_B; z] = \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \cdots (a_p)_n}{(b_1)_n (b_2)_n \cdots (b_p)_n} \frac{z^n}{n!}, \quad (\text{A.59})$$

where $(a)_n = a(a+1)(a+2)\cdots(a+n-1)$ and $(a)_0 = 1$. The special case ${}_2\text{F}_1 [\cdot, \cdot; \cdot; \cdot]$, called Gauss' hypergeometric function, where the subscripts are often suppressed and denoted as $\text{F} [\cdot, \cdot; \cdot; \cdot]$, is defined as

$$\text{F} [a, b; c; z] = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!}. \quad (\text{A.60})$$

A.3.5 Lommel's function

Lommel's function $s_{\mu,\nu}(z)$ (Pg. 346 of Watson [117]) is defined as

$$s_{\mu,\nu}(z) = \frac{z^{\nu+1}}{(\mu-\nu+1)(\mu+\nu+1)} {}_1\text{F}_2 \left[1; \frac{1}{2}(\mu-\nu) + \frac{3}{2}, \frac{1}{2}(\mu+\nu) + \frac{3}{2}; -\frac{1}{4}z^2 \right]. \quad (\text{A.61})$$

The asymptotic expansion (Page 351 of [117]) for large $|z|$ where $|\arg z| < \pi$ and $\mu \pm \nu$ are not odd positive integers is

$$s_{\mu,\nu}(z) \sim z^{\mu-1} \left\{ 1 - \frac{(\mu-1)^2 - \nu^2}{z^2} + \frac{[(\mu-1)^2 - \nu^2][(3\mu-3)^2 - \nu^2]}{z^4} - \dots \right\}. \quad (\text{A.62})$$

A.3.6 Probability integral

The error function, also referred to as probability integral, is the integral of the Gaussian distribution defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt. \quad (\text{A.63})$$

The complementary error function is given by

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt = 1 - \operatorname{erf}(x). \quad (\text{A.64})$$

A.3.7 Sinc function

The sinc function is defined as

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \quad (\text{A.65})$$

The sinc function frequently occurs in optical systems with square apertures, since $\operatorname{rect}(x) \iff \operatorname{sinc}(u)$, where

$$\begin{aligned} \operatorname{rect}(x) &= 1, & |x| < 1/2, \\ &= 0, & |x| > 1/2. \end{aligned} \quad (\text{A.66})$$

A.3.8 Jinc function

The jinc function is defined as

$$\operatorname{jinc}(x) = \frac{J_1(2\pi x)}{x} \quad (\text{A.67})$$

where $J_1(\cdot)$ is the First order Bessel function of the first kind. The jinc function appears frequently in imaging system with circular apertures since it is the Hankel transform of a disk.

References

- [1] S. A. Akhmanov and S. Y. Nikitin, *Physical optics* , Oxford University Press, New York, USA, (1997).
- [2] E. Hecht, *Optics*, 3rd ed. , Addison-Wesley, Reading, Massachusetts, USA, (1998).
- [3] D. C. Giancoli, *Physics for scientists and engineers*, 3rd ed. , Prentice Hall, New Jersey, USA, (2000).
- [4] J. W. Goodman, *Introduction to Fourier optics*, 2nd ed. , McGraw-Hill, New York, USA, (1996).
- [5] S. G. Lipson, H. Lipson, and D. D. Tannhauser, *Optical physics*, 3rd ed. , Cambridge University Press, New York, USA, (1995).
- [6] R. E. Blahut, *Theory of remote image formation* , Cambridge University Press, Cambridge, UK, (2004).
- [7] J. C. Dainty and J. R. Fienup, *Image Recovery: Theory and Application*, Chap. 7: Phase Retrieval and Image Reconstruction for Astronomy, p. 231 , Academic Press, Orlando, Florida, USA, (1987).
- [8] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and synthesis in radioastronomy*, 2nd ed. , John Wiley & Sons, (2001).
- [9] J. C. Curlander and R. N. McDonough, *Synthetic aperture radar: systems and signal processing* , John Wiley & Sons, New York, USA, (1991).
- [10] H. N. Chapman, A. Barty, S. Marchesini, A. Noy, S. P. Hau-Riege, C. Cui, M. R. Howells, R. Rosen, H. He, J. G. H. Spence, U. Weierstall, T. Beetz, C. Jacobsen, and D. Shapiro, "High-resolution *ab initio* three-dimensional x-ray diffraction microscopy," *J. Opt. Soc. Am. A* **23**(5), 1179–1200 (2006).
- [11] J. Drenth, *Principles of Protein X-Ray Crystallography* , Springer-Verlag, New York, USA, (1994).

- [12] W. H. Hsiao and R. P. Millane, "Effects of Fourier-plane amplitude and phase errors on image reconstruction. I. Small amplitude errors," *J. Opt. Soc. Am. A* **24**(10), 3180–3188 (2007).
- [13] Z. P. Liang and P. C. Lauterbur, *Principles of magnetic resonance imaging: a signal processing perspective*, IEEE Press, New York, USA, (2000).
- [14] E. L. O'Neill and A. Walther, "The question of phase in image formation," *Optica Acta* **10**, 33–40 (1963).
- [15] A. Walther, "The question of phase retrieval in optics," *Optica Acta* **10**, 41–49 (1963).
- [16] Y. M. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," *Opt. Commun.* **30**, 304–308 (1979).
- [17] R. G. Lane, "Recovery of complex images from fourier magnitude," *Opt. Commun.* **63**(1), 6–10 (1987).
- [18] R. G. Lane, W. R. Fright, and R. H. Bates, "Direct pahse retrieval," *IEEE Trans. Acoust., Speech, and Signal Process. ASSP-35*, 520–525 (1987).
- [19] M. H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Trans. Acoust., Speech, and Signal Process. ASSP-30*(2), 140–154 (1982).
- [20] R. H. T. Bates and M. J. McDonnell, *Image Restoration and Reconstruction*, Oxford University Press, Oxford, UK, (1986).
- [21] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Am. A* **7**(3), 394–411 (1990).
- [22] R. W. Grechberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* **35**(2), 237–246 (1972).
- [23] J. R. Fienup, "Phase retrieval algorithms - a comparison," *Appl. Opt.* **21**(15), 2758–2769 (1982).
- [24] V. Elser, "Phase retrieval iterated projections," *J. Opt. Soc. Am. A* **20**(1), 40–55 (2003).
- [25] R. P. Millane, "Iterative Projection Algorithms for Solving Inverse Problems," *Proc. OCEANS 2003* pp. 2714–2719 (2003).
- [26] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE* **69**(5), 529–541 (1981).
- [27] R. P. Millane and W. H. Hsiao, "On apparent counterexamples to phase dominance," *J. Opt. Soc. Am. A* **20**(4), 753–756 (2003).

- [28] L. N. Piotrowski and F. W. Campbell, "A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase," *Perception* **11**, 337–346 (1982).
- [29] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Trans. Acoust., Speech, and Signal Process.* **ASSP-28**(6), 672–680 (1980).
- [30] P. L. V. Hove, M. H. Hayes, J. S. Lim, and A. V. oppenheim, "Signal reconstruction from signed Fourier transform magnitude," *IEEE Trans. Acoust., Speech, and Signal Process.* **ASSP-31**(5), 1286–1293 (1983).
- [31] T. S. Huang, J. W. Burnett, and A. G. Deczky, "The importance of phase in image processing filters," *IEEE Trans. Acoust., Speech, and Signal Process.* **ASSP-23** (1975).
- [32] W. A. Pearlman and R. M. Gray, "Source coding of the discrete Fourier transform," *IEEE Trans. Inf. Theory* **IT-24**(6), 683–692 (1978).
- [33] W. H. Hsiao and R. P. Millane, "Effects of Fourier amplitude and phase errors on image reconstruction," in *Proc. IVCNZ'02*, D. N. Kenwright, ed., pp. 221–225 , Auckland, New Zealand, (2002).
- [34] W. H. Hsiao, E. Ip, and R. P. Millane, "Effects of Fourier amplitude and phase errors on interpretability of images," in *Proc. IVCNZ'03*, D. Bailey, ed., pp. 175–180 , Palmerston North, New Zealand, (2003).
- [35] A. W. Lohmann, G. Weigelt, and B. Wirnitzer, "Speckle masking in astronomy: triple correlation theory and applications," *Appl. Opt.* **22**, 4028–4037 (1983).
- [36] R. H. Bates, "Aspects of the Erlangen-bispectrum," *Optik* **76**, 23–26 (1987).
- [37] P. Negrete-Regagnon, "Object recovery from the bispectral phase in astronomical imaging," *J. Opt. Soc. Am. A* **15**(7), 1787–1796 (1998).
- [38] K. T. Knox and T. J. Thompson, "Recovery of images from atmospherically degraded short exposure photographs," *Astrophys. J. , Lett. ED.* **193**, 45–48 (1974).
- [39] K. T. Knox, "Image retrieval from astronomical speckle pattern," *J. Opt. Soc. Am. A* **66**, 1236–1239 (1976).
- [40] R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," *Opt. Eng.* **21**, 829–832 (1982).
- [41] R. G. Paxman and J. R. Fienup, "Optical misalignment sensing and image reconstruction using phase diversity," *J. Opt. Soc. Am. A* **5**(6), 914–923 (1988).
- [42] R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," *J. Opt. Soc. Am. A* **9**(7), 1072–1085 (1992).

- [43] M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius, and K. Urban, "Electron microscopy image enhanced," *Nature* **392**(6678), 768–769 (1998).
- [44] B. Freitag, S. Kujawa, P. M. Mul, J. Ringnalda, and P. C. Tiemeijer, "Breaking the spherical and chromatic aberration barrier in transmission electron microscopy," *Ultramicroscopy* **102**(3), 209–214 (2005).
- [45] A. J. Koster and H. Zandbergen, "Electron microscopy: Cutting the cost of high-resolution microscopy," *Nat. Mater.* **4**(12), 885–886 (2005).
- [46] B. Fultz and J. M. Howe, *Transmission Electron Microscopy and Diffractometry of Materials*, Springer-Verlag, (2001).
- [47] U. Weierstall, Q. C. J. C. H. Spence, M. R. Howells, and A. Isaacson, "Image reconstruction from electron and X-ray diffraction patterns using iterative algorithms: experiment and simulation," *Ultramicroscopy* **90**(2-3), 171–195 (2002).
- [48] R. W. Gerchberg and W. O. Saxton, "Phase determination from image and diffraction plane pictures in electron-microscope," *Optik* **34**(3), 275–284 (1971).
- [49] R. W. Gerchberg and W. O. Saxton, "Practical algorithm for determination of phase from image and diffraction plane pictures," *Optik* **35**(2), 237–246 (1972).
- [50] M. R. Teague, "Deterministic phase retrieval - a green-function solution," *J. Opt. Soc. Am. A* **73**(11), 1434–1441 (1983).
- [51] S. C. Mayo, P. W. Miller, S. W. Wilkins, T. J. Davis, D. Gao, T. E. Gureyev, D. Paganin, D. J. Parry, A. Pogany, and A. W. Stevenson, "Quantitative X-ray projection microscopy: phase-contrast and multi-spectral imaging," *J. Microsc.* **207**, 79–96 (2002).
- [52] S. Bajts, A. Barty, K. A. Nugent, M. McCartney, M. Wall, and D. Paganin, "Quantitative phase-sensitive imaging in a transmission electron microscope," *Ultramicroscopy* **83**(1-2), 67–73 (2000).
- [53] G. H. Stout and L. H. Jensen, *X-ray structure determination: a practical guide*, 2nd ed., John Wiley & Sons, New York, USA, (1989).
- [54] A. Hofmann, *The physics of synchrotron radiation*, Cambridge University Press, Cambridge, UK, (2004).
- [55] D. Sayre, "The squaring method: a new method for phase determination," *Acta Crystallogr. A* **5**, 60–65 (1952).
- [56] C. Giacovazzo, *Direct methods in crystallography*, Academic Press, New York, USA, (1980).

- [57] M. M. Woolfson, "Direct methods—from birth to maturity," *Acta Crystallogr. A* **43**, 593–612 (1987).
- [58] D. W. Green, V. M. Ingram, and M. F. Perutz, "The structure of haemoglobin IV. Sign determination by the isomorphous replacement method," in *Proc. R. Soc. London Ser. A* **255**, pp. 287–307 , (1954).
- [59] D. M. Blow, "The structure of haemoglobin VII. Determination of phase angles in the non-centrosymmetric [100] zone," in *Proc. R. Soc. London Ser. A* **247**, pp. 302–336 , (1958).
- [60] W. A. Hendrickson, J. L. Smith, and S. Sheriff, "Direct phase determination base on anomalous scattering," *Meth. Enzymol.* **115**, 41–55 (1985).
- [61] I. K. Robinson, I. A. Vartanyants, G. J. Williams, M. A. Pfeifer, and J. A. Pitney, "Reconstruction of the shapes of gold nanocrystals using coherent x-ray diffraction," *Phys. Rev. Lett.* **87**(19), 1995,505 (2001).
- [62] J. S. Wu, U. Weierstall, and J. C. h. Spence, "Diffractive electron imaging of nanoparticles on a substrate," *Nat. Mater.* **4**(12), 912–916 (2005).
- [63] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, and D. Sayre, "Reconstruction of a yeast cell from X-ray diffraction data," *Acta Crystallogr. A* **62**, 248–261 (2006).
- [64] J. C. H. Spence and R. B. Doak, "Single molecule diffraction," *Phys. Rev. Lett.* **92**(19), 198,012 (2004).
- [65] J. C. H. Spence, K. Schmidt, J. S. Wu, G. Hembree, U. Weierstall, B. Doak, and P. Fromme, "Diffraction and imaging from a beam of laser-aligned proteins: resolution limits," *Acta Crystallogr.* **61**, 237–245 (2005).
- [66] M. J. Tovée, *An introduction to the visual system* , Cambridge University Press, New York, USA, (1996).
- [67] B. A. Wandell, *Foundations of vision* , Sinauer Associates, Sunderland, Massachusetts, USA, (1995).
- [68] J. D. Mollon and J. K. Bowmaker, "The spatial arrangement of cones in the primate fovea," *Nature* **360**(6405), 677–679 (1992).
- [69] P. D. Gowdy and C. M. Cicerone, "The spatial arrangement of the L and M cones in the central fovea of the living human eye," *Vision Res.* **38**(17), 2575–2589, (1998).
- [70] S. Otake, P. D. Gowdy, and C. M. Cicerone, "The spatial arrangement of L and M cones in the peripheral human retina," *Vision Res.* **40**(6), 677–693 (2000).

- [71] S. H. C. Hendry and R. C. Reid, "The koniocellular pathway in primate vision," *Annu. Rev. Neurosci.* **23**, 127–153 (2000).
- [72] A. J. R. White, S. G. Solomon, and P. R. Martin, "Spatial properties of koniocellular cells in the lateral geniculate nucleus of the marmoset *Callithrix jacchus*," *J. Physiol.* **531**(1), 203–218 (2001).
- [73] P. R. Martin, A. J. R. White, A. K. Goodchild, H. D. Wilder, and A. E. Sefton, "Evidence that blue-on cells are part of the third geniculocortical pathway in primates," *Eur. J. Neurosci.* **9**, 1536–1541 (1997).
- [74] M. Mishkin, L. G. Ungerleider, and K. A. Macko, "Object vision and spatial vision - 2 cortical pathways," *Trends Neurosci.* **6**(10), 414–417 (1983).
- [75] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**(12), 2379–2394 (1987).
- [76] H. B. Barlow, *Sensory Communication*, Chap. Possible principles underlying the transformation of sensory messages, pp. 217–234 , Cambridge: M.I.T. Press, (1961).
- [77] D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," *Ophthalmic Physiol. Opt.* **12**(2), 229–232 (1992).
- [78] A. van der Schaaf and J. H. van Hateren, "Modelling the power spectra of natural images: Statistics and information," *Vision Res.* **36**, 2759–2770 (1996).
- [79] D. L. Ruderman, "Origins of scaling in natural images," *Vision Res.* **37**(23), 3385–3398 (1997).
- [80] D. J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vision Res.* **37**(23), 3367–3383 (1997).
- [81] R. P. Millane, S. Alzaidi, and W. H. Hsiao, "Scaling and power spectra of natural images," in *Proc. IVCNZ'03*, D. G. Bailey, ed., pp. 148–153 , Palmerston North, New Zealand, (2003).
- [82] W. H. Hsiao and R. P. Millane, "Effects of occlusion, edges and scaling on the power spectra of Natural Images," *J. Opt. Soc. Am. A* **22**, 1789–1797 (2005).
- [83] R. P. Millane and W. H. Hsiao, "Natural images and visual processing," *Optics & Photonics News* **16**(12) (2005).
- [84] D. J. Field, "What the statistics of natural images tell us about visual coding?" in *Proc. SPIE*, B. E. Rogowitz, ed., vol. 1077 of *Human Vision, Visual Processing, and Digital Display*, pp. 269–276 , (1989).

- [85] D. C. Knill, D. J. Field, and D. Dersten, "Human discrimination of fractal images," *J. Opt. Soc. Am. A* **7**(6), 1113–1123 (1990).
- [86] R. S. Powers and J. W. Goodman, "Error rates in computer-generated holographic memories," *J. Opt. Soc. Am.* **14**(7), 1690–1701 (1975).
- [87] J. Bucklew and N. C. Gallagher, "Comprehensive error models and a comparative study of some detour-phase holograms," *J. Opt. Soc. Am.* **18**(16), 2861–2869 (1979).
- [88] R. J. Read, "Structure-factor probabilities for related structures," *Acta Crystallogr. A* **46**, 900–912 (1990).
- [89] J. R. Fienup and A. M. Kowalczyk, "Phase retrieval for a complex-valued object by using a low-resolution image," *J. Opt. Soc. Am. A* **7**(3), 450–458 (1990).
- [90] A. B. Watson, J. Hu, and J. F. M. III, "Digital video quality metric based on human vision," *J. Electron. Imaging* **10**(1), 20–29 (2001).
- [91] B. Girod, *Digital images and human vision*, Chap. Chap. 15 , MIT Press, Cambridge, MA, (1993).
- [92] J. R. Fienup, "Invariant error metrics for image reconstruction," *Appl. Opt.* **36**(32), 8352–8357 (1997).
- [93] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- [94] W. H. Hsiao and R. P. Millane, "Effects of Fourier amplitude and phase errors on image reconstruction," in *Proc. SPIE*, P. J. Bones, M. A. Fiddy, and R. P. Millane, eds., vol. 5562 of *Image Reconstruction from incomplete data III*, pp. 27–37 , (2004).
- [95] R. P. Millane and W. H. Hsiao, "Effects of amplitude and phase errors on perceived images," in *International Workshop on Noncrystallographic Phase Retrieval* , Cairns, Australia, (2003).
- [96] W. K. Pratt, *Digital Image Processing* , John Wiley & Sons, New York, USA, (1978).
- [97] R. L. Joshi and T. R. Fischer, "Comparison of generalized Gaussian and Laplacian modeling in DCT image-coding," *IEEE Signal Process. Lett.* **2**, 81–82 (1995).
- [98] D. L. Richards, "Statistical properties of speech," *Proc. Inst. Elec. Eng.* **111**, 941–949 (1964).
- [99] S. G. and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.* **10**, 204–207 (2003).

- [100] A. Mohammad-Djafari, "From deterministic to probabilistic approaches to solve inverse problems," in *Proc. SPIE*, A. Mohammad-Djafari, ed., vol. 3459 of *Bayesian Inference for Inverse Problems*, pp. 2–11 , (1998).
- [101] J. C. Ye, K. J. Webb, C. A. Bouman, and R. P. Millane, "Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework," *J. Opt. Soc. Am. A* **16**, 2400–2412 (1999).
- [102] S. Baskaran and R. P. Millane, "Bayesian image reconstruction from partial image and aliased spectral intensity data," *IEEE Trans. Image Process.* **8**, 1420–1434 (1999).
- [103] F. Muller, "Distribution shape of 2-dimensional DCT coefficients of natural images," *Electron. Lett.* **29**, 1935–1936 (1993).
- [104] J. H. Chang, "Image probability distribution based on generalized Gamma function," *IEEE Signal Process. Lett.* **12**, 325–328 (2005).
- [105] J. W. Shin, J. H. Chang, and N. S. Kim, "Statistical modeling of speech signals based on generalized Gamma distribution," *IEEE Signal Process. Lett.* **12**, 258–261 (2005).
- [106] J. K. Jao, "Amplitude distribution of composite terrain radar clutter and the k-distribution," *IEEE Trans. Antennas Propag.* **AP-32**, 1049–1062 (1984).
- [107] E. Conte and M. Longo, "Characterization of radar clutter as a spherically invariant random process," *IEE Proc.-F Radar and Signal Process.* **134**, 191–196 (1987).
- [108] I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, *Table of integrals, series and products*, 4th ed. , Academic Press, New York, USA, (1980).
- [109] Wolfram Research, Inc., *Mathematica*.
- [110] I. Juvells, S. vallmitjana, A. Carnicer, and J. campos, "The role of amplitude and phase of the Fourier transform in the digital image processing," *Am. J. Phys.* **59**(8), 744–748 (1991).
- [111] A. W. Lohmann and D. M. and G. Shabtay, "Significance of phase and amplitude in the Fourier domain," *J. Opt. Soc. Am. A* **14**(11), 2901–2904 (1997).
- [112] D. L. Ruderman and W. Bialek, "Statistics of natural images: scaling in the woods," *Phys. Rev. Lett.* **73**(6), 814–817 (1994).
- [113] R. M. Balboa, C. W. Tyles, and N. M. Grzywacz, "Occlusions contribute to scaling in natural images," *Vision Res.* **41**(7), 955–964 (2001).
- [114] N. M. Grzywacz, R. M. Balboa, and C. W. Tyler, "Response to letter to the editor," *Vision Res.* **42**(25), 2803–2805 (2002).

- [115] D. L. Ruderman, "Letter to the editor," *Vision Res.* **42**(25), 2799–2801 (2002).
- [116] A. D. Wheelon and H. T. Robacker, *Tables of Summable Series and Integrals Involving Bessel Functions: Part 2 - A table of integrals involving Bessel functions*, Holden-Day, San Francisco, USA, (1968).
- [117] G. N. Watson, *A treatise on the theory of Bessel Functions*, 2nd ed. , Cambridge University Press, Cambridge, UK, (1966).
- [118] M. Abramowitz and I. Stegun, *Handbook of Mathematical functions* , Dover Publications, New York, USA, (1964).
- [119] R. N. Bracewell, *The Fourier transform and its applications* , McGraw-Hill, New York, USA, (1978).
- [120] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. , McGraw-Hill, New York, USA, (1984).