NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

The Design and Optimisation of Quasioptical Telescopes

A thesis submitted for the Degree of
Doctor of Philosophy

Presented by
John Lavelle

Department of Experimental Physics
NUI Maynooth
County Kildare
Ireland

Research Supervisor
Dr. Créidhe O'Sullivan

Head of Department
Professor John Anthony Murphy

July 2008

Contents

Abstract

The main focus of this thesis is the analysis and optimisation of systems that operate in the terahertz and submillimetre wavebands. Analysis is carried out on the MBI and ALMA interferometers, and on the HIFI instrument on the Herschel Space Observatory (HSO). MBI is a novel instrument designed to test the technique of bolometric interferometry. ALMA is a 50 element hetrodyne interferometer, currently being constructed in Chile. It is demonstrated that in both MBI and ALMA, the coupling of the signal to the detector horns may be improved by imposing a phase on the field from the sky; a number of possible configurations are considered. HIFI is a high resolution spectrometer, one of three instruments on the HSO. Simulations of the lens antennas used to detect the radiation in the higher frequency channels in HIFI were carried out. A number of methods used to determine the phase centre of the beam from lens antennas are described, and use is made of the same optimisation techniques as for the MBI and ALMA work. As the beams propagating in these systems can be analysed accurately using the paraxial approximation, Gaussian Beam Mode Analysis can be used to simulate the field, and is the main analytical tool used here.

Methods of beam shaping are investigated to gain an insight into how coupling may be increased in these systems, and also to design Diffractive Optical Elements (DOEs) for use at terahertz and submillimetre wavelengths in general. The standard methods of the Gerchberg-Saxton Algorithm, and unidirectional optimisation using Differential Evolution and Simulated Annealing are applied to design DOEs. A novel approach using Gaussian Beam Mode Analysis is described. Here, the Gaussian Beam mode coefficients describing a field are optimised to achieve a desired amplitude distribution at a specified plane or planes. This approach was found to achieve highly optimal results, and has a number of benefits over the other methods.

Acknowledgements

My supervisor during the course of my research was Dr. Créidhe O'Sullivan. I could not have asked for a more dedicated and helpful supervisor, and very much appreciate the time she devoted to me. I would also like to thank Professor Anthony Murphy for inviting me to be part of the group, and Dr. Neil Trappe, for helping me to understand some of the fundamentals of the subject.

My fellow postgraduates in the group, Massimo Candotti, Tim Finn, Karen Foley, Emily Gleeson, Ronan Mahon and Rob May, helped to make my time in Maynooth very enjoyable, and also helped with my work. I had many useful discussions with Rob and Ronan about diffractive optics, which has influenced some of the work described in this thesis.

Thanks to my parents and my sisters, Anne and Clare, for their support and encouragement during the writing of this thesis.

Finally, I would like to acknowledge the financial support that I received from Enterprise Ireland and Science Foundation Ireland.

Acronyms

ALMA	Atacama Large Millimetre Array
BS	Beam Splitter
CMB	Cosmic Microwave Background
COA	Common Optic Assembly
CSA	Calibration Source Assembly
DE	Differential Evolution
DFT	Discrete Fourier Transform
DOE	Diffractive Optical Element
FFT	Fast Fourier Transform
FWHM	Full Wave Have Maximium
GBM	Gaussian Beam Mode
GBMSO	Gaussian Beam Mode Set Optimisation
GSA	Gerchberg Saxton Algorithm
HDPE	High-Density Polyethylene
HEB	Hot Electron Bolometer
HIFI	Heterodyne Instrument for the Far Infrared
HSO	Herschel Space Observatory
LO	Local Oscillator
MA	Mixer Assemblies
MBI	Millimetre-wave Bolometric Interferometer
MSE	Mean Squared Error
PO	Physical Optics
PPL	Phased Locked Loop
RAM	Random Access Memory
SA	Simulated Annealing
SIS	Superconductor/Insulator/Superconductor
SRON	Space Research Organization Netherlands
TE	Transverse Electric
TM	Transverse Magnetic
UV	Ultra Violet
VNA	Vector Network Analyser

Chapter 1

Introduction

1.1 Introduction

This thesis is concerned with the design and analysis of quasioptical systems, and astronomical instrumentation in particular, operating at millimetre- and submillimetre wavelengths.

In this chapter, I introduce some ideas and techniques used throughout the thesis. More detail and in-depth discussion is given, where required, in later chapters. Firstly I introduce the motivation behind making astronomical observations at these wavelengths. Then I describe system components that are commonly used in quasioptical telescopes, and some techniques used to model beam propagation through them. Finally I described the layout of this thesis.

1.2 Astronomy in the Terahertz Waveband

'Millimetre and submillimetre-wave', 'far-infrared' and 'terahertz' are some of the names for the region of the electromagnetic spectrum with which this thesis is concerned. Millimetre-wave optics deals with radiation of around 1mm wavelength, or a frequency of around 300GHz, and terahertz with radiation at around 1THz, or a wavelength of 0.3mm. There is no precise cut-off for the upper and lower limits of these terms, and there is a large overlap in the regions they define; however, they all refer to the region of the electromagnetic spectrum between the optical - where the wavelength can generally be considered to be negligible, and therefore geometric optics can be used to model the propagation of the field - and the radio - where the wavelength is long, and therefore diffraction effects dominate, necessitating the use of electromagnetics to model propagation. Systems built for these wavelengths tend to use components that are somewhat larger than the wavelength, and therefore their beams tend to be reasonably well collimated, and diffraction effects do not dominate, but are important.

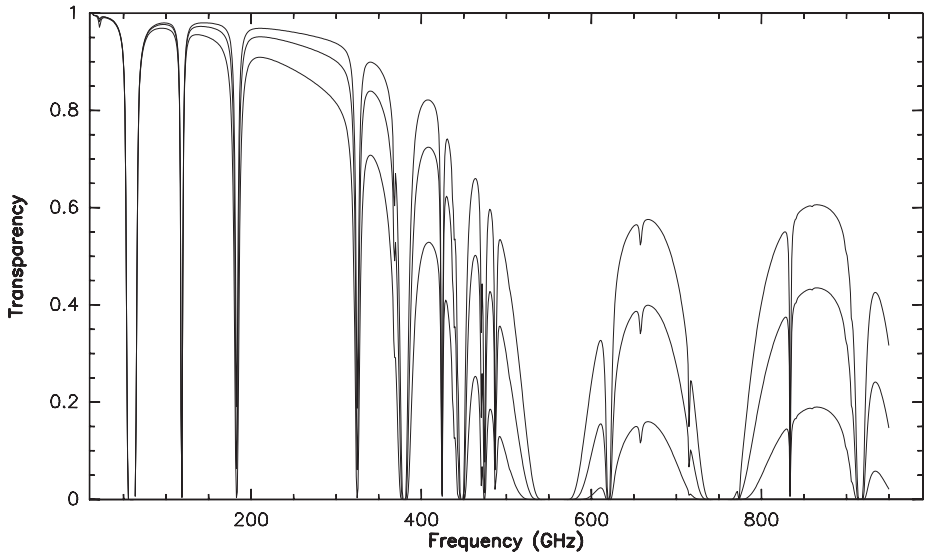


Figure 1.1: Zenith transmission as a function of frequency under typical atmospheric conditions at the ALMA site in Chile. The three curves correspond to 25, 50 and 75 percentiles[1].

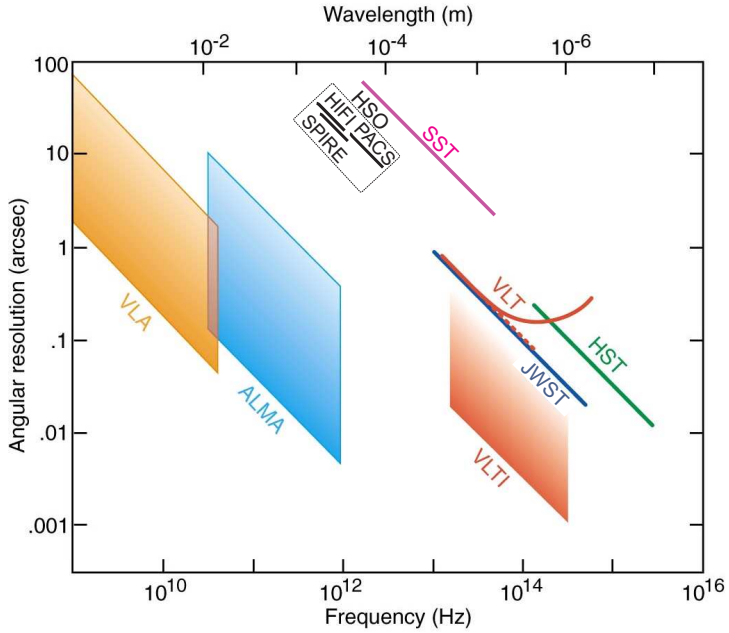


Figure 1.2: The range of angular resolution achieved as a function of operating frequency for the interferometers: Atacama Large Millimeter Array (ALMA), Very Large Telescope Interferometer and Very Large Array (VLA) (for the interferometers, the angular resolution is determined by the length of the baseline, which can be varied). The angular resolution as a function of operating frequency of the telescopes: Herschel Space Observatory (HSO), Spitzer Space Telescope (SST), James Webb Space Telescope (JWST), Hubble Space Observatory (HSO) and Very Large Telescope (VLT) (solid lines correspond to the seeing-limited case and the dashed line to the diffraction limited case with adaptive optics)[1].

The main constraint in the development of terahertz systems is the detection and generation of the radiation. While the technology for generation and detection of radio, microwave and visible radiation is well established, its application at terahertz frequencies is difficult, due to the fundamental limitations of these devices; quasioptical devices occupy a middle ground between *short* wavelength microwave components, and *long* wavelength optical components.

Fundamentally, the reason why telescopes are developed for use at different wavelengths across the electromagnetic spectrum is because, while the underlying electromagnetic wave theory simply scales with wavelength, dramatically different phenomena dominate radiation generation in the different regions of the electromagnetic spectrum[2]. Millimetre and submillimetre astronomy allow phenomena to be observed that would be impossible at other wavelengths. At submillimetre and millimetre-wavelengths it is possible to observe the "cool universe"; objects with temperatures of around 5-50K have emission spectra that peak at these wavelengths. Broadband thermal radiation from small dust grains - typically re-radiated as absorbed shorter-wavelength radiation - is the most common continuum emission process in this band[3]. Stars emit at much higher temperatures than this; however, the broadband flux from distant galaxies is diminished in the UV and optical, both due to redshift and obscuration by internal dust. This dust produces a large peak in the rest-frame far-infrared, which when redshifted greatly enhances the millimetre and submillimetre emission from these objects[3]. From various absorbtion lines at these frequencies, it is possible to resolve the fine-scale kinematic details in young stellar disks that can potentially form planets, and in old ejected stellar envelopes that are forming dust grains and enriching the interstellar medium with carbon, oxygen and nitrogen.

The millimetre and submillimetre regions lie at the crossroads between the radio and infrared/optical wavebands and contain both synchrotron emission, extending from the radio, and thermal emission, extending from the optical/infrared. They also contain the peak of the microwave background, and technically, they have the huge advantage that heterodyne techniques can be used.

Aside from the challenge of manufacturing receivers that operate at these frequencies, the other major obstacle to submillimetre astronomy is that moisture in the atmosphere absorbs much of the radiation by the time it reaches sea level. The Atacama Large Millimeter Array (ALMA) telescope is located at Llano de Chajnantor Observatory in the Atacama desert in northern Chile (5000m altitude)[4]. This is a very dry site, making it excellent for submillimetre astronomy. Figure 1.1 shows the atmospheric transmission at this location. The Herschel and Spitzer space telescopes will allow observations at millimetre and submillimeter wavelengths outside of the atmospheric windows that ALMA and other ground-based telescopes

will observe at. As shown in Figure 1.2, the combination of existing and planned facilities will provide complete wavelength coverage from long wavelength radio to UV. These new instruments will allow observation at much greater resolution and sensitivity than has been previously achieved.

1.3 Quasioptical System Components

1.3.1 Waveguides

Wires are the standard way of transferring signals in electronic devices. This is not possible at terahertz frequencies however; as the frequency of the electromagnetic radiation propagating in the wire increases, a greater proportion of it is radiated outside the wire, until, at terahertz frequencies wires act as antennas. As a result, little of the signal power transmitted to the wire would arrive at the output and unwanted signals may also be coupled to the wire. Wires act as antennas at terahertz frequencies because, whereas, at low frequencies the voltage along a wire is effectively constant due to the wire being small compared with the signal wavelength, at higher frequencies the voltage varies along it; this voltage varies periodically with time, radiating some of the signal power into space. A solution is to propagate the field in free space or a dielectric and confine it with a reflective material. At visible wavelengths this is achieved using fibre optics, and at microwave and terahertz frequencies waveguides are used.

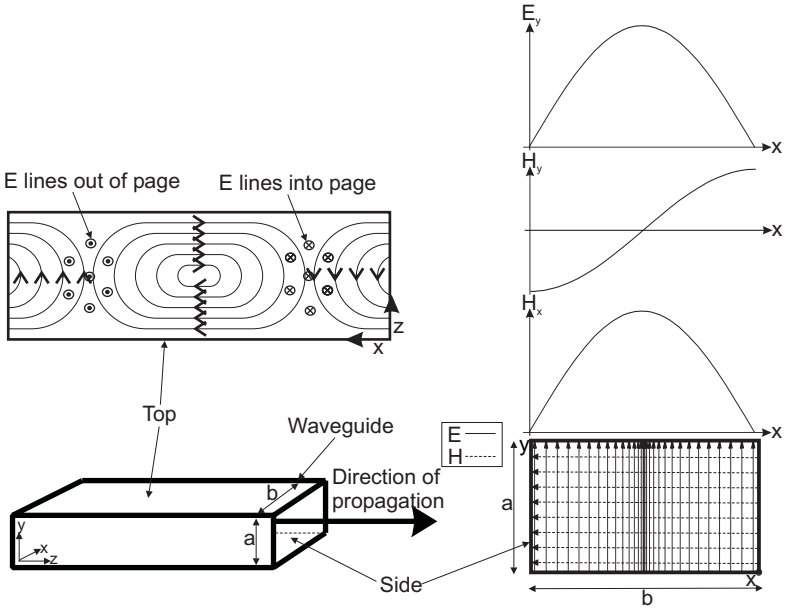


Figure 1.3: Variation of the field components for the TE_{10} mode in a rectangular waveguide.

A waveguide consists of free space surrounded by metal, down which the electromagnetic wave travels. Waveguides are used to transmit signals at terahertz

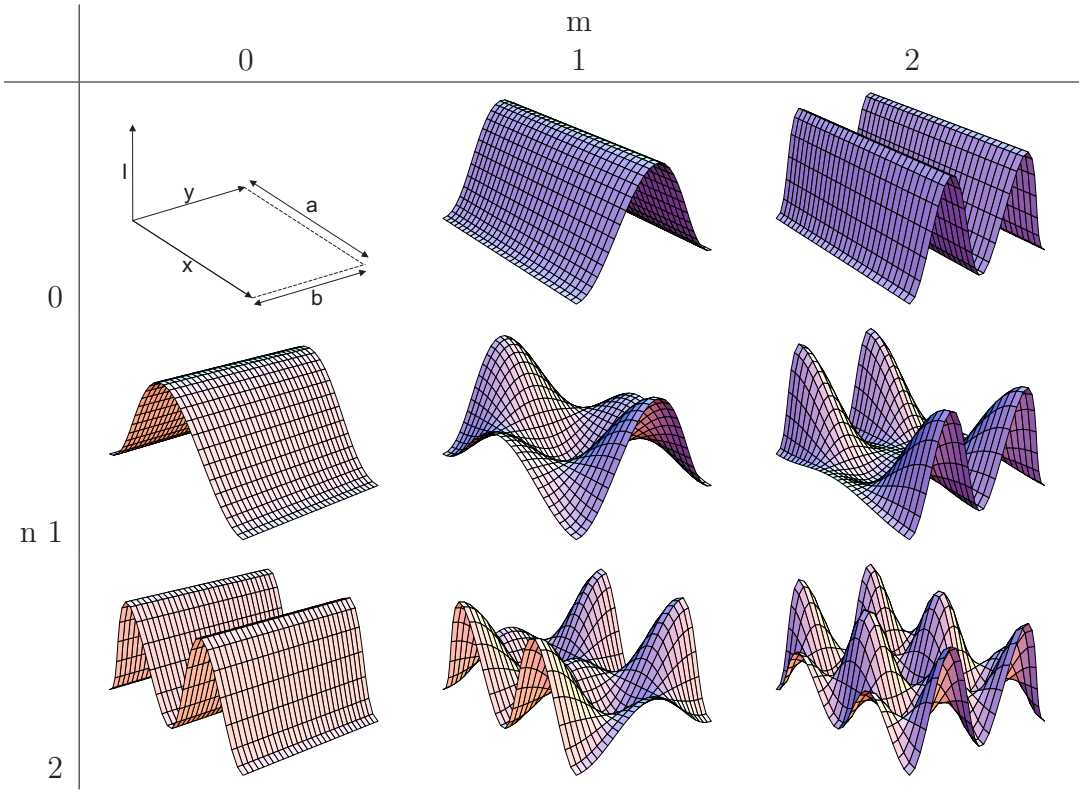


Figure 1.4: Plots of the intensity as a function of distance x, y for $TE_{m,n}$ modes.

frequencies, but their application presents a number of challenges. Generally single mode propagation is desirable, but this requires that waveguides have dimensions of the order of a wavelength. In addition to the problem of fabricating such minute waveguides, ohmic losses are greater at these, higher, frequencies. At microwave frequencies, the metal sides of the waveguide can be considered to be totally reflective, but at terahertz frequencies the penetration depth of the radiation is greater, leading to greater surface resistances and losses.

Rectangular waveguides are either transverse electric (TE) or transverse magnetic (TM) types. In TE waveguides the electric field is perpendicular to the axis of propagation and the magnetic field is parallel, and vice versa for TM waveguides. The electric field component parallel to a metal wall is zero at the surface. The solutions of Maxwell's equations for TE waves, with mode numbers m and n, in a rectangular metal waveguide, are of the general form

$$E_x = \frac{i\omega\mu_0 H_0 n\pi}{k_c^2 b}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi x}{b}\right)\exp[i(\omega t-\beta z)] \tag{1.1}$$

$$E_y = \frac{i\omega\mu_0 H_0 m\pi}{k_c^2 b}\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi x}{a}\right)\exp[i(\omega t-\beta z)] \tag{1.2}$$

$$E_z = 0, \tag{1.3}$$

where

$$\beta^2 = \omega^2\mu_0\epsilon_0 - k_c^2, \tag{1.4}$$

$$k_c^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2, \tag{1.5}$$

a and b are the height and width of the waveguide (as shown in Figure 1.3) ϵ_0 and μ_0 are the permittivity and permeability of free space, t is the time, ω is 2π times the signal angular frequency and H_0 is the maximum magnetic field in the guide. Figure 1.4 shows plots of these TE modes. TE_{00} can not exist, as this would require a non-zero electric field along the guide wall. The modes maintain the same form as they propagate down the waveguide. The total field is a linear sum of all the modes propagating down the waveguide; therefore, if the field is multimoded, the total field varies as a function of propagation distance down the waveguide, causing inconsistant coupling with the detector located in the waveguide. It is desirable to have a single moded signal, for most purposes. Figure 1.3 shows the field variation for a TE_{01} mode. Plots of TE modes are shown in Figure 1.4.

1.3.2 Feed-Horns

A feed-horn is used to couple free space radiation to a waveguide, where it can then be detected. The principle of reciprocity is employed to characterize antennas (such as a feed-horn), whereby the antenna is modeled as a radiation emitter. Essentially, a feed-horn is a waveguide that is flared to increase the surface area, thereby increasing the directivity of the beam that radiates from it. Two common types of horn, shown in Figure 1.5, have circular and rectangular apertures.

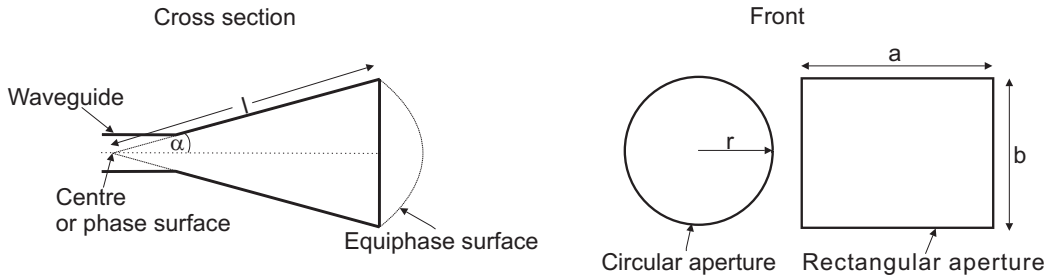


Figure 1.5: Feed horn of slant length l_S, axial length l_a, with a circular aperture of radius r or rectangular aperture with length a and width b.

The number of modes that a feed-horn will support is dictated by the dimensions, of the waveguide cross section. A single-moded hollow circularly-symmetric corrugated horn supports the circularly symmetric HE_{11} mode given by

$$E(r) = \begin{cases} J_o\left(\frac{2.405r}{a}\right)\exp\left(\frac{-i\pi r^2}{\lambda R_h}\right) & r \leq a \\ 0 & r > a, \end{cases} \tag{1.6}$$

where a is the aperture radius of the horn, r is the radial distance in the guide, J_0

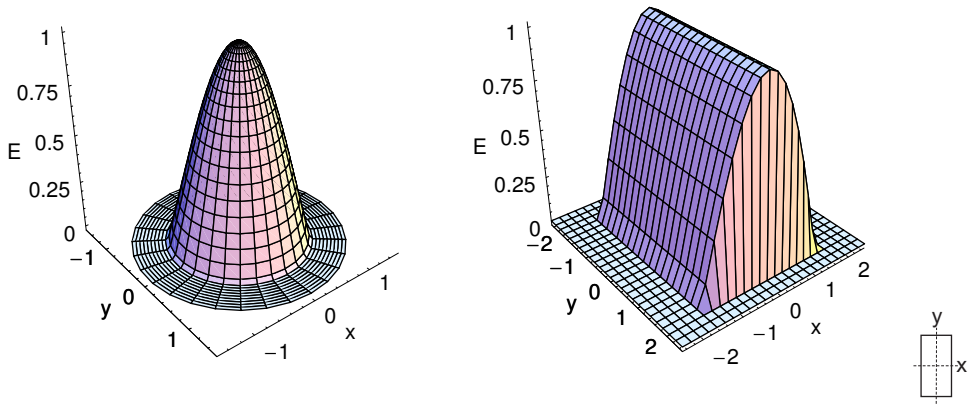


Figure 1.6: The field at the horn aperture of *left* a corrugated conical horn and *right* a smooth-walled rectangular horn.

is the Bessel function of the zeroth order, 2.405 is the first zero of J_0, and R_h is the slant length. In smooth-walled horns, undesirable surface currents flow, which reduce the symmetry and increase the cross-polarisation of the field they produce. A corrugated horn has narrow grooves, each a quarter-wavelength deep, cut into the wall. These corrugations cause cancellation of currents flowing along the wall, as the currents in adjacent grooves are 180° out of phase[5].

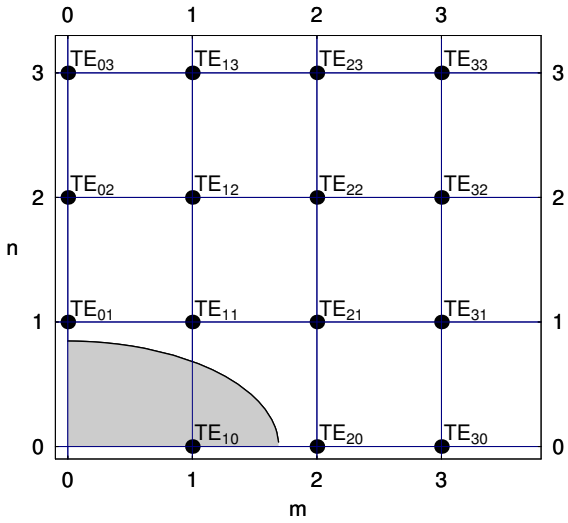


Figure 1.7: In a waveguide, only the modes in the shaded region, given by the inequality of Equation 1.8, propagate in the waveguide.

The field from the aperture of a typical smooth-walled rectangular feed-horn (TE_{10} mode), of length and width a and b, is shown in Figure 1.6 *right* and is given by

$$E(x,y) = \begin{cases} E_0 \cos\left(\frac{\pi x}{a}\right) \exp\left(\frac{i\pi(x^2+y^2)}{\lambda R_h}\right) & \text{within aperture} \\ 0 & \text{outside aperture.} \end{cases} \tag{1.7}$$

A rectangular feed-horn supports the same modes that propagate in its waveguide

Feed Type	Power Coupling
Corrugated circular horn	98%
Corrugated square horn	98%
Smoothed-walled circular horn	91%
Rectangular feed (TE_{10} mode)	88%

Table 1.1: The percentage of power contained in a fundamental Gaussian for a number of feeds[6].

section. The modes that can propagate are related to the waveguide's dimensions and must satisfy the inequality $k_0 > k_c$, where k_c is given in Equation 1.5, therefore

$$m \leq \sqrt{\left(\frac{2a}{\lambda}\right)^2 - \left(\frac{na}{b}\right)^2}. \tag{1.8}$$

The solution to this equation is plotted in Figure 1.7 for a rectangular horn with $a = 2.54$mm, $b = 1.26$mm and $\lambda = 3$mm. Only the modes in the region defined by the inequality can propagate, and therefore in this case, only the TE_{10} mode propagates. The far-field of the horn aperture field is given by its Fourier transform

$$f_x(\theta, \phi) = \iint_{Ap} E_x \exp(ik(x \sin\theta \cos\phi + y \sin\theta \sin\phi))dS \tag{1.9}$$

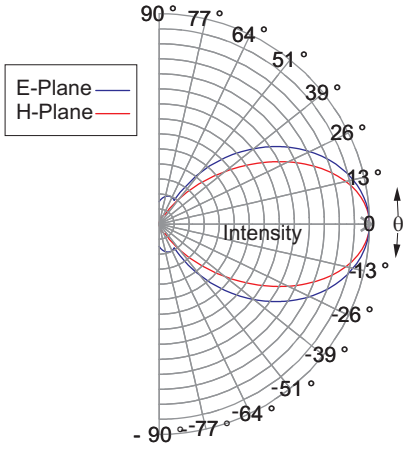


Figure 1.8: The co-polar far-field pattern in the E-plane and H-plane from a waveguide, with $a = 2.56$mm and $b = 1.26$mm and $\lambda = 3$mm.

$$f_y(\theta, \phi) = \iint_{Ap} E_y \exp(ik(x \sin\theta \cos\phi + y \sin\theta \sin\phi))dS, \tag{1.10}$$

where ϕ and θ are the far-field spherical coordinates, and x and y are the Cartesian coordinates at the aperture Ap. The co-polar far-field distribution $F(\theta, \phi)$ is related

to the components of Equations 1.9 and 1.10 by the following paraxial approximation

$$F(\theta,\phi) \propto (1+\cos\theta)(f_x(\theta,\phi)\hat{i} + f_i(\theta,\phi)\hat{j}), \tag{1.11}$$

yielding in this case

$$\begin{aligned} F(\theta,\phi) \propto & (1+\cos\theta)\sqrt{\frac{4}{ab}}\iint_{Ap} \sin\left(\frac{\pi x}{a}\right) \\ & \times \exp(ik(x\sin\theta\cos\phi + y\sin\theta\sin\phi)\exp(-ij\phi))dS\hat{j}. \end{aligned} \tag{1.12}$$

A plot of the field from a waveguide given by this Equation is shown in Figure 1.8. The abrupt change in the E-plane aperture field causes the higher side-lobes in the far-field's E-plane, compared to the H-plane. Side-lobes in the far-field are undesirable for astronomy applications, as weak signals from objects in the field of view can become contaminated with strong sources further off-axis. The Gaussicity of a number of horns feeds is given in Table 1.1.

1.3.3 Mirrors

Sir Isaac Newton is credited with constructing the first practical reflecting telescope, the Newtonian, which consisted of a spherical primary reflector and a secondary flat mirror, used to reflect the light to the observer. This was an improvement on earlier designs, where the observer viewed the light at the focus of the primary, thereby blocking the light. In the Newtonian design, however, only the secondary is in the path of the light incident on the primary, and consequently a smaller proportion of it is lost. As manufacturing techniques improved, parabolic and hyperbolic surfaces were used in place of spherical surfaces, to bring the light to an exact focus. Mirrors have a number of advantages over lenses: they do not suffer from effects such as chromatic aberration, large absorption losses, surface reflections and cross-coupling between the polarisations.

Many configurations of mirrors now exist. The Cassegrain, shown in Figure 1.9 *bottom*, is a common choice, and is used in the Herschel, ALMA and MBI telescopes, discussed in this thesis. It consists of a parabolic primary mirror, with a hole in the centre to permit the light reflected from the hyperbolic secondary mirror to come to a focus behind it. Figure 1.9 *bottom* shows how an off-axis parabolic mirror can be used to focus the light, without the secondary mirror or detector blocking it. Off-axis parabolic mirrors are more difficult to fabricate and distort the wavefront than on-axis configurations. Figure 1.9 shows the field in terms of ray optics and wavefronts. In the conceptual limit, as wavelength $\lambda \to 0$, the light can be modelled accurately as rays; however, at longer wavelengths the concept of wavefronts is more

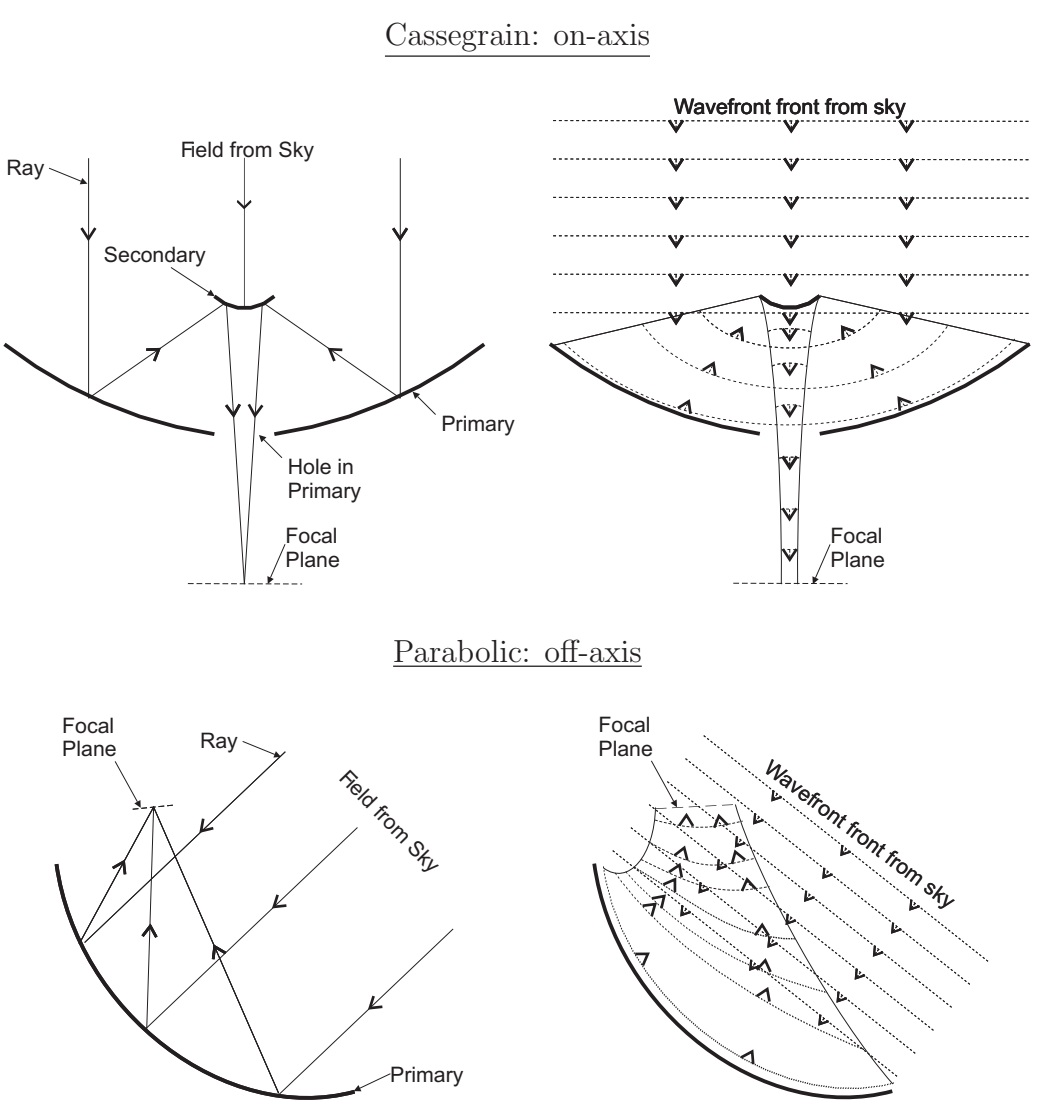


Figure 1.9: Cassegrain and parabolic off-axis telescopes in terms of ray optics and wavefronts.

useful, as diffraction effects must be modelled.

Two important factors when characterising a telescope are the amount of light it can collect and its angular resolution. The amount of light collected is proportional to the area of the primary mirror, while the angular resolution is limited by diffraction and is approximately equal to λ/D, where λ is the telescope's operating frequency and D is the diameter of the primary[7]. It is therefore desirable to build telescopes with large primary mirrors, to increase their angular resolution and sensitivity; however, their cost increases with roughly the cube of the diameter for a ground-based telescope[8] and much more for a space-based telescope. At terahertz frequencies telescopes with diameters of the order of meters are tens's of thousands times larger than the operating wavelength[9], and therefore fabrication of the mirror surface is a much more demanding task, as manufacturing tolerances on the mirror surface are proportional to the operating wavelength.

As of June 2008, the top two largest optical reflector telescopes in the world are the Large Binocular Telescope, with two 8.4m diameter mirror (equivalent to a single 11.8m diameter mirror), and the Southern African Large Telescope, with a diameter of 11m. Larger telescopes are planned for the future, but their diameters are constrained by the cost and technical challenges that these telescopes impose. An interferometer overcomes some of these problems by using an array of apertures. Compared to a monolithic dish of diameter D, equal to the length of a two-element interferometer baseline, B, the interferometer has an angular resolution (fringe spacing) of roughly twice that of the monolithic dish. The increase in angular resolution over a filled dish is because whereas a filled dish is dominated by spacings that are much smaller than the aperture diameter an interferometer is not[10].

A number of optical interferometer telescopes, such as CHARA[11] and the Palomar Testbed Interferometer[12], have recently been built. These ground-based telescopes have served as demonstrators of the technologies needed to build NASA's Space Interferometry Mission, for example, which will very accurately measure the distances and positions of stars and detect Earth-like planets in other solar systems. In Chapter 4 the design of a novel *bolometric* interferometer is described. Lenses were designed in order firstly to adjust the position of the feed-horn field's waist location and size, thereby avoiding truncation at the secondary mirror, and secondly to improve the coupling of the telescope field to the detector arrays.

1.4 Modelling of Quasioptical Systems

The design of telescopes operating at submillimetre-wavelengths is often based on geometric (ray) methods, and consequently, they tend to take the same form as visible wavelength telescopes. However, to accurately model the propagation of a

beam of radiation through such a telescope, diffraction effects must be taken into account. Scalar and paraxial approximations of the wave equation may be employed to model diffraction; however, to achieve a more accurate analysis the vector nature of the radiation must be modelled and a more rigorous application of Maxwell's equations must be employed. Vector Physical Optics (PO) (as opposed to scalar physical optics) very accurately models fields which interact with surfaces that are relatively flat compared to the wavelength, which is typically the case for quasioptical systems. Telescopes with large surface areas relative to the wavelength are very computationally demanding to model using PO. (The PO technique is described in Chapter 6)

The quasioptical systems analysed here occupy a middle ground between typical visible wavelength optical systems - where the wavelength is sufficiently small compared to the optical components for the radiation to be modelled as rays - and systems where the wavelength is close to the component dimension - where a more rigorous application of Maxwell's equations is required. In addition, the assumption that the radiation propagated is a *paraxial* beam can also be applied. By paraxial we mean that the radiation essentially propagates as a beam in a particular well-defined direction, with some diffraction taking place. A paraxial beam - which can be represented as a scalar distribution and where the cross-polar component can be considered to propagate independently of the co-polar - allows a simplification of Maxwell's equations to an elegant and computationally efficient formulation that describes the beam propagation, known as Gaussian Beam Modes (GBMs). Firstly, I shall consider Fresnel and Fraunhofer diffraction, which are also simplifications of Maxwell's equations.

1.4.1 The Fourier Transform and Fraunhofer Diffraction

In the Fresnel approximation (for a beam not spreading too far laterally, i.e. paraxial, and not too close to a truncating aperture), the electric field at a point (x, y, z) is given by

$$E(x, y, z) = \frac{e^{ikz}}{i\lambda z} \iint_{\xi\eta} E(\xi, \eta, 0) e^{i\frac{k}{2z}((x-\xi)^2+(y-\eta)^2)} d\xi d\eta, \tag{1.13}$$

where ξ and η are the coordinates in the aperture plane (at $z = 0$) and k is the wavenumber. In the far-field of the aperture this expression can be simplified to

give the Fraunhofer diffraction integral[13, 14].

$$E(x,y,z) = \frac{e^{ikz}e^{i\frac{k(x^2+y^2)}{2z}}}{i\lambda z}\iint_{\xi\eta} E(\xi,\eta,0)e^{i\frac{2\pi}{\lambda z}(x\xi+y\eta)}d\xi d\eta, \tag{1.14}$$

$$= \frac{e^{ikz}e^{i\frac{k(x^2+y^2)}{2z}}}{i\lambda z}\iint_{\infty} E_n(\xi,\eta)e^{i\frac{2\pi}{\lambda z}(x\xi+y\eta)}d\xi d\eta, \tag{1.15}$$

where E_n is the near-field which is non-zero over some region. This distribution is in fact the Fourier transform of the near-field distribution, as follows:

$$E_f(k_x,k_y) \propto \mathcal{F}\{E_n(\xi,\eta)\}, \tag{1.16}$$

where $k_x(=\frac{x}{\lambda z})$ and $k_y(=\frac{y}{\lambda z})$ are the spatial frequencies in the x and y directions. The inverse transform propagates the far-field back to the near-field. Often, due to the symmetry of the beams propagating in optical systems, the problem may be analysed in 1D; the transforms here are therefore often given in 1D.

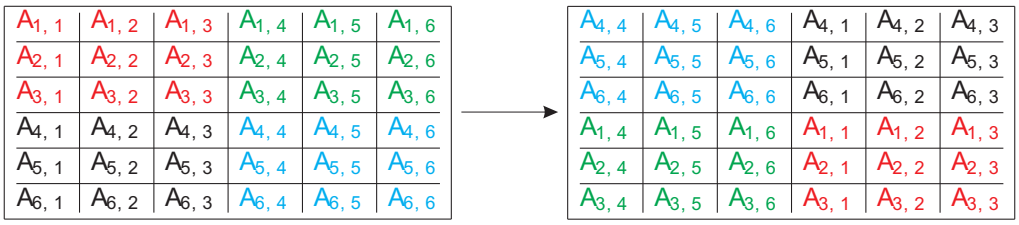


Figure 1.10: The DFT of a field places the zero frequency element in the top left element of the array. The four quadrants must be rearranged to give the far-field pattern.

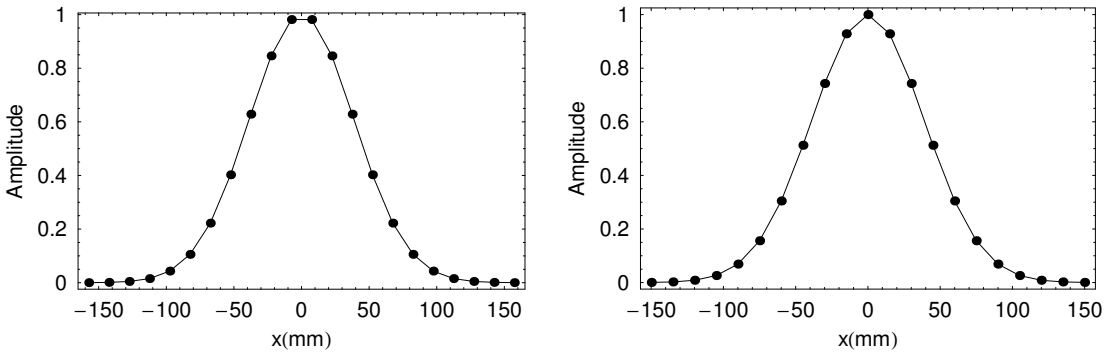


Figure 1.11: A Gaussian function sampled with *left* an even number of samples and *right* an odd number of samples.

The Fourier transform is generally computed with a discrete set of samples points using the Discrete Fourier Transform (DFT). In the far-field the angular separation

between the points is given by

$$\Delta\theta = \sin^{-1}\left(\frac{\lambda}{N\Delta x}\right), \tag{1.17}$$

where N is the number of samples and Δx is the sampling interval in the near-field. The DFT places the lowest frequency terms at the ends of the transform. The zero frequency term must be rearranged to place it at the centre. The elements in the DFT array are cyclical; therefore, elements at opposite ends of the array rows and columns are in sequence. In the case of a 2D array, corresponding to a 2D field, the four quadrants must be swapped, as shown in Figure 1.10. In the case of a 1D array, corresponding to a 1D field, the 2 halves are swapped. Equivalent to swapping the quadrants is rotating the element columns and rows by half the length of the array[15]. Some care must be taken when applying the DFT, or spurious results can occur:

- Figure 1.11 shows a 1D Gaussian function sampled with *left* an even number of samples and *right* an odd number of samples. The DFT generally places the zero frequency element at the beginning of the array; therefore, for a field sampled with an even number of points, it is not possible to rotate the transformed field in order that the zero frequency term appears in the centre. Here, an odd number of samples is used allowing the array to be rotated, in order that the zero frequency term appears in the centre of the array.

- According to Nyquist–Shannon sampling theory, the field should be sampled at twice the highest spatial frequency contained in the field, to avoid aliasing.

Equation 1.17 shows how the sampling in the far-field is related to the extent of the field transformed by the DFT. Increasing the width of the field increases the resolution of the transform. The width of the field can be effectively increased by padding the field with zeros.

1.4.2 Gaussian Beam Modes

The beam illustrated in Figure 1.12 is an x-cut of the first of an infinite series of GBMs of free-space propagation, which can be used to describe beams having localised cross sectional power. This, "fundamental", mode has the form of a Gaussian, with amplitude

$$\psi = \psi_0 \exp\left(-\frac{(x^2+y^2)}{w^2}\right), \tag{1.18}$$

It is a solution of the paraxial wave-equation. This Gaussian beam maintains its Gaussian profile as it propagates, only the scale w and the radius of curvature R of

the beam changes. The variation of w with distance is hyperbolic. The minimum value of w is called the beam waist radius w_0, and it represents a "quasi-focus".

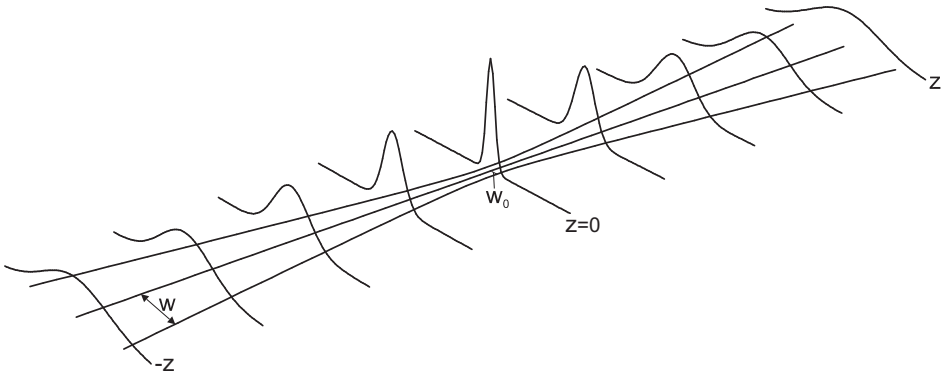


Figure 1.12: A Gaussian beam with beam radius w and waist radius w_0 as a function of propagation distance z.

This fundamental Gaussian beam formulation was used extensively to model lasers beams, as radiation emitted by a laser typically has a high Gaussicity (i.e. similar to Gaussian). It has also proven very useful for modelling millimetre-wave systems, because the radiation emitted by most quasi optical sources is also highly Gaussian, as shown in Table 1.1. The interaction with components such as mirrors and lenses can be modelled simply by computing the effect on the radius of curvature and beam radius waist position of the Gaussian beam, as discussed later.

In order that beams can be modelled more accurately - to account for the percentage of power not accounted for by the fundamental mode and to more accurately model the effects of components which can truncate the field and otherwise affect the amplitude profile - higher order Gaussian Beam Modes (GBMs) are used. The field at any distance z is given as a sum of these modes in which each mode has a weight, given by the mode coefficient.

Following [16], a derivation of GBMs follows. The scalar version of the wave equation is given as

$$\nabla^2\psi + k^2\psi = 0, \tag{1.19}$$

where $k = 2\pi/\lambda$, ψ represents a scalar field and λ the wavelength. The complex field can be written as

$$\widetilde{\psi}(x, y, z) \equiv \widetilde{u}(x, y, z)\exp(-ikz), \tag{1.20}$$

where the term $\exp(-ikz)$ is written separately for convenience. The $\exp(-ikz)$ term has a spatial period of one wavelength λ in the z direction, and the spatial dependence of $\widetilde{\psi}(x, y, z)$ is primarily due to the $\exp(-ikz)$ variation (Equations 1.23 and 1.24 below). Substituting this into the wave equation yields the reduced wave

equation

$$\frac{\partial^2 \widetilde{u}}{\partial x^2} + \frac{\partial^2 \widetilde{u}}{\partial y^2} + \frac{\partial^2 \widetilde{u}}{\partial z^2} - 2ik\frac{\partial \widetilde{u}}{\partial z} = 0. \tag{1.21}$$

In the paraxial approximation we assume that

$$|\tfrac{\partial^2 \widetilde{u}}{\partial z^2}| << |2ik\tfrac{\partial \widetilde{u}}{\partial z}|, \tag{1.22}$$

$$|\tfrac{\partial^2 \widetilde{u}}{\partial z^2}| << |\tfrac{\partial^2 \widetilde{u}}{\partial x^2}| \text{ and} \tag{1.23}$$

$$|\tfrac{\partial^2 \widetilde{u}}{\partial z^2}| << |\tfrac{\partial^2 \widetilde{u}}{\partial y^2}|. \tag{1.24}$$

Therefore, the second partial derivative in z can be dropped yielding the paraxial wave equation

$$\frac{\partial^2 \widetilde{u}}{\partial x^2} + \frac{\partial^2 \widetilde{u}}{\partial y^2} - 2ik\frac{\partial \widetilde{u}}{\partial z} = 0. \tag{1.25}$$

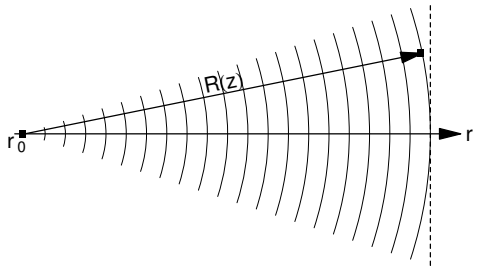


Figure 1.13: A spherical waveform from a *real* point source.

The general solution to the exact wave equation - which corresponds to a uniform spherical wave diverging from a source point $r_0(x_0, y_0, z_0)$, as shown in Figure 1.13 - may be written in the form

$$\widetilde{u}(r; r_0) = \frac{\exp[-ik\rho(r, r_0)]}{\rho(r, r_0)}, \tag{1.26}$$

where $\widetilde{u}(r; r_0)$ is the field at point r due to a source at point r_0, and where the distance $\rho(r, r_0)$ from the source point r_0 to the observation point r is given by

$$\rho(r, r_0) = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}. \tag{1.27}$$

The distance $\rho(r, r_0)$ can be expanded in a power series to give

$$\rho(r, r_0) = z - z_0 + \frac{(x - x_0)^2 + (y - y_0)^2}{2(z - z_0)} + \ldots. \tag{1.28}$$

In the Fresnel approximation, all terms higher than the quadratic are dropped from

$\rho(r, r_0)$, yielding a "paraxial-spherical wave" solution, as follows

$$\begin{aligned}\widetilde{u}(x,y,z) &= \frac{1}{z-z_0}\exp\left(-\frac{ik((x-x_0)^2+(y-y_0)^2)}{2(z-z_0)}\right), \\ &= \frac{1}{R(z)}\exp\left(-\frac{ik((x-x_0)^2+(y-y_0)^2)}{2R(z)}\right),\end{aligned} \tag{1.29}$$

where $R(z)$ the radius of curvature, and (x_0, y_0, z_0) is the origin, as shown Figure 1.13. The validity of this approximation is discussed in Reference [16]. This formulation does not accurately describe a true physical beam, as it extends to infinity in the transverse direction. As shown below, the use of *complex* source coordinates can resolve this.

A paraxial beam propagating in the direction z can be written in the form

$$\widetilde{u}(x,y,z) = \frac{1}{\widetilde{q}(z)}\exp(-\frac{ik(x^2+y^2)}{2\widetilde{q}(z)}), \tag{1.30}$$

where $\widetilde{q}(z)$ is the *complex* radius and is given by

$$\widetilde{q}(z) = \widetilde{q_0} + z - z_0. \tag{1.31}$$

The term $1/\widetilde{q}(z)$ as it appears in Equation 1.30 is complex and is separated into its real and imaginary parts as follows:

$$\frac{1}{\widetilde{q}(z)} \equiv \frac{1}{\widetilde{q_r}(z)} - i\frac{1}{\widetilde{q_i}(z)}. \tag{1.32}$$

Substituting into Equation 1.30 and rewriting in standard notation we have

$$\begin{aligned}\widetilde{u}(x,y,z) =& \frac{1}{\widetilde{q}(z)}\exp(-\frac{ik(x^2+y^2)}{2q_r(z)} - \frac{ik(x^2+y^2)}{2q_i(z)}) \\ \equiv& \frac{1}{\widetilde{q}(z)}\exp(-\frac{ik(x^2+y^2)}{2R(z)} - \frac{i(x^2+y^2)}{w^2(z)}),\end{aligned} \tag{1.33}$$

where $R(z)$ is the radius of curvature, and $w(z)$ is the beam radius such that

$$\frac{1}{\widetilde{q}(z)} \equiv \frac{1}{\widetilde{R}(z)} - i\frac{\lambda}{\pi w^2(z)}. \tag{1.34}$$

The variation with distance z is given by Equation 1.31.

This *Gaussian* beam expression is only the lowest solution in a infinite family of higher-order solutions to the paraxial wave equation. Two commonly used, orthogonal, mode sets are Gaussian-Hermite functions (which are expressed in rectangular coordinates), and Gaussian-Laguerre functions (which are in cylindrical co-

ordinates). The 2D Gaussian-Hermite modes are given by

$$\psi_{m,n}(w_0,x,y,z) = \sqrt{\frac{2^{-m-n+1}}{\pi w^2 m! n!}} \tag{1.35a}$$

$$\times H_m\left(\frac{\sqrt{2}x}{w}\right) H_n\left(\frac{\sqrt{2}y}{w}\right) \tag{1.35b}$$

$$\times \exp(-\frac{x^2+y^2}{w^2}) \tag{1.35c}$$

$$\times \exp(-ikz) \tag{1.35d}$$

$$\times \exp(-\frac{i\pi\left(x^2+y^2\right)}{R\lambda}) \tag{1.35e}$$

$$\times \exp(i\phi_0(m+n+1)) \tag{1.35f}$$

Equation 1.35a normalises the power of each mode, and Equation 1.35b gives the form of the modes H_m and H_n (the Hermite polynomials of degree m and n). Equation 1.35c gives the Gaussian amplitude profile, which is modified by Equation 1.35b. Equation 1.35d gives the phase due to propagation by distance z. Equation 1.35e gives the phase due to the mismatch of the flat reference plane and the equiphase-front of the beam with radius of curvature R, and Equation 1.35f gives the phase that is dependant on mode number, called the "phase slippage" ϕ. The parameters w, R and ϕ vary with z. The beam radius w gives a measure of the transverse extent of the beam: it is the radius at which the amplitude has fallen to $1/e$ of its on-axis value and is given by

$$w = \sqrt{\left(\frac{z^2\lambda^2}{\pi^2 w_0^4}+1\right) w_0^2}. \tag{1.36}$$

Therefore, at $z = 0$mm: $w = w_0$, and the waist size increases hyperbolically with z. The radius of curvature is given by

$$R = z\left(\frac{\pi^2 w_0^4}{z^2\lambda^2}+1\right), \tag{1.37}$$

and the phase slippage is given by

$$\phi = (m+n+1)\phi_0 \,\text{and}\, \phi_0 = \tan^{-1}\left(\frac{z\lambda}{\pi w_0^2}\right). \tag{1.38}$$

$\psi_{m,n}(w_0,x,y,z)$ can be separated into its x and y components as $\psi_{m,n}(w_0,x,y,z) = \psi_m(w_0,x,z)\psi_n(w_0,y,z)$ with

$$\psi_m(w_0,x,z) = \sqrt{\frac{2^{\frac{1}{2}-n}}{\sqrt{\pi} w n!}} H_m\left(\frac{\sqrt{2}x}{w}\right) \exp[-\frac{x^2}{w^2} - ikz - \frac{i\pi x^2}{R\lambda} + i\frac{\phi_0(2m+1)}{2}] \tag{1.39}$$

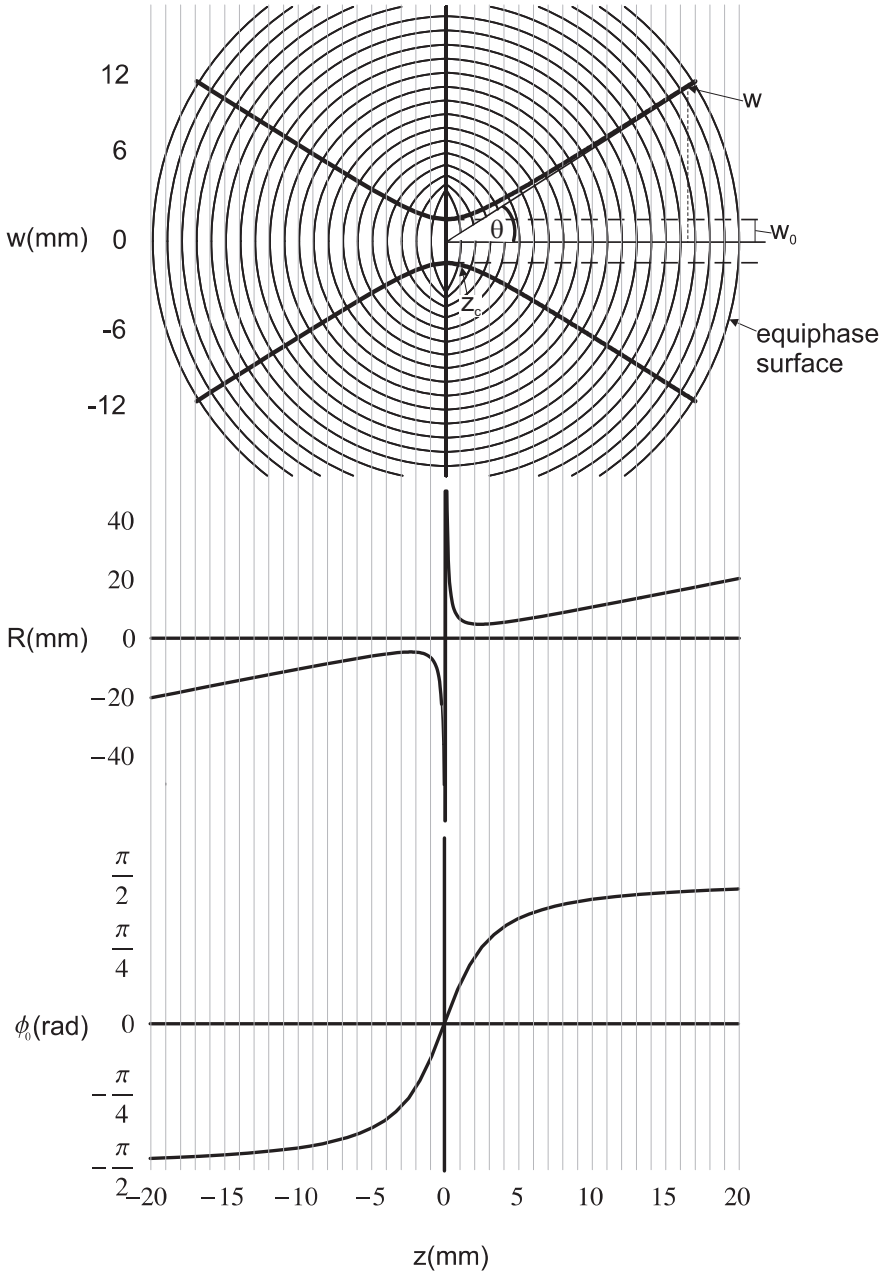


Figure 1.14: *Top* a diagram showing how w changes as a function of z; the spherical equiphase surface shown is approximated by a parabolic one in GBMs. *Middle* and *bottom* plots of the Gaussian beam mode parameters R and ϕ_0 as a function of propagation distance z.

and

$$\psi_n(w_0, y, z) = \sqrt{\frac{2^{\frac{1}{2}-n}}{\sqrt{\pi} w n!}} H_n\left(\frac{\sqrt{2}y}{w}\right) \exp[-\frac{y^2}{w^2} - ikz - \frac{i\pi y^2}{R\lambda} + i\frac{\phi_0(2n+1)}{2}]. \quad (1.40)$$

The Gaussian-Hermite functions provide a complete basis set of orthogonal functions, which obey the orthonormality condition

$$\int_{-\infty}^{\infty} \psi_m(w_0, x, y, z)\psi_n(w_0, x, y, z)dx = \delta_{m,n} \quad (1.41)$$

independent of z[16]. Therefore, they can be used as a basis set to expand an arbitrary paraxial beam $E(x, y, z)$, in the form

$$E(x, y, z) = \sum_m \sum_n A_{n,m}\psi_{n,m}(w_0, x, y, z), \quad (1.42)$$

where $A_{n,m}$ are the mode coefficients, which are given by

$$A_{n,m} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(x, y, z)\psi^*_{n,m}(w_0, x, y, z)dxdy. \quad (1.43)$$

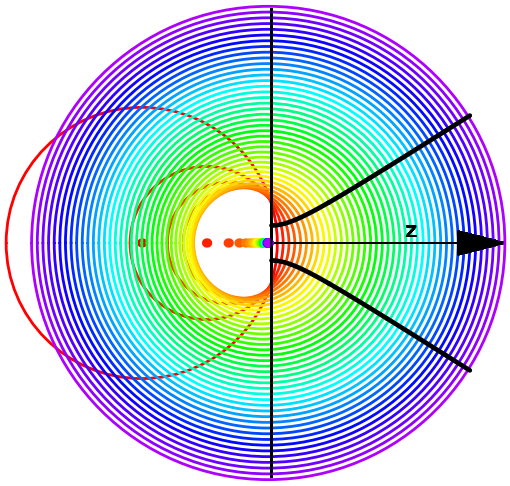


Figure 1.15: The circles which define the spherical phase of a propagating beam. The centres of the circles are indicated with a dot, where the dot has the same colour as the corresponding circle. The radius of the circles R becomes asymptotically closer to z as z increases, and the centre of the circle shifts closer to the waist location.

Although an infinite number of Gaussian-Hermite modes form the basis set, for practical applications less than 20 modes will reconstruct most beams in quasioptical systems satisfactorily. The choice for w_0 and z of the modes is arbitrary, and the waist radius of the modes need not necessarily coincide with that of the field. How-

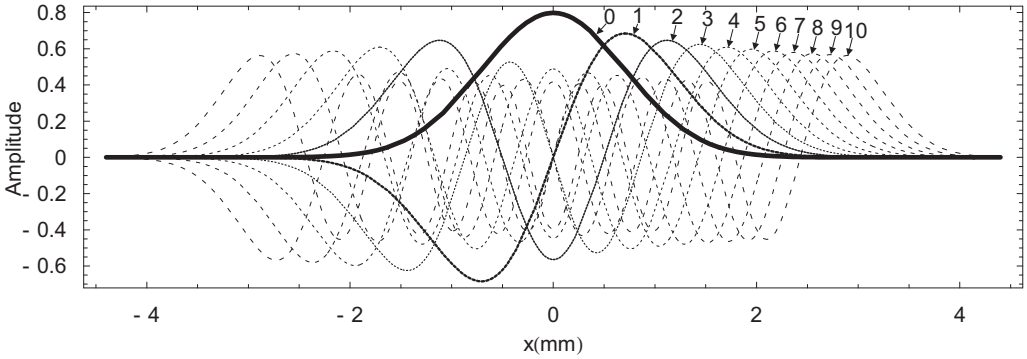


Figure 1.16: The first 10 Hermite polynomials. The even modes are symmetric about the x-axis, the odd modes are asymmetric.

ever, choosing z such that the phase radius of curvature, R, of the modes matches that of the field to be reconstructed, and w such that the highest order mode is similar in extent to the field, reduces the number of modes required to adequately reconstruct the field.

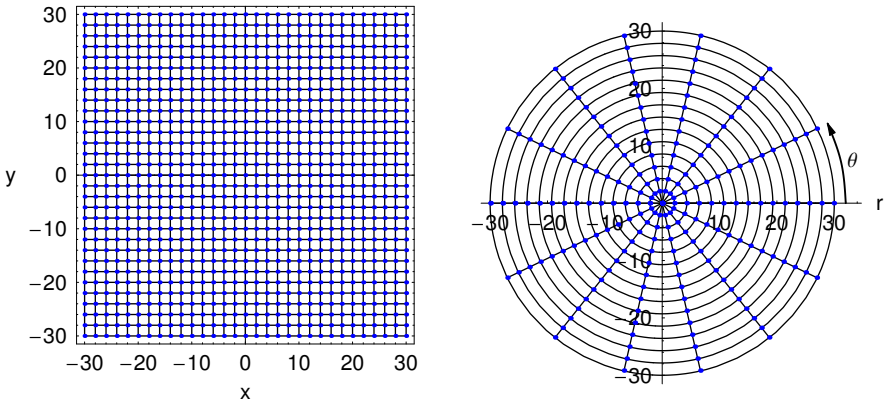


Figure 1.17: *Left* the Cartesian coordinate system for the Gaussian-Hermite polynomials, and *right* the polar coordinate system for the Gaussian-Laguerre polynomials.

Figure 1.14 shows how parameters w, R and ϕ_0, given by Equations 1.36, 1.37 and 1.38 respectively, change with propagation distance z. w varies hyperbolically with z, $\phi_0 = \pi/2$ in the limit $z \to \infty$, and $\phi = -\pi/2$ in the limit $z \to -\infty$. R is indeterminate at $z = 0$. As z goes from 0 to ∞, R varies rapidly from ∞ to a minimum and then increases more slowly to ∞. The spherical phase-front, determined by R, shown in Figure 1.14 is approximated by a parabolic phase. R can alternatively be given in terms of the "confocal distance" by

$$R(z) = z\left[1 + \left(\frac{z_c}{z}\right)^2\right], \tag{1.44}$$

where z_c is the confocal distance, given by

$$z_c = \frac{\pi w_0^2}{\lambda}. \tag{1.45}$$

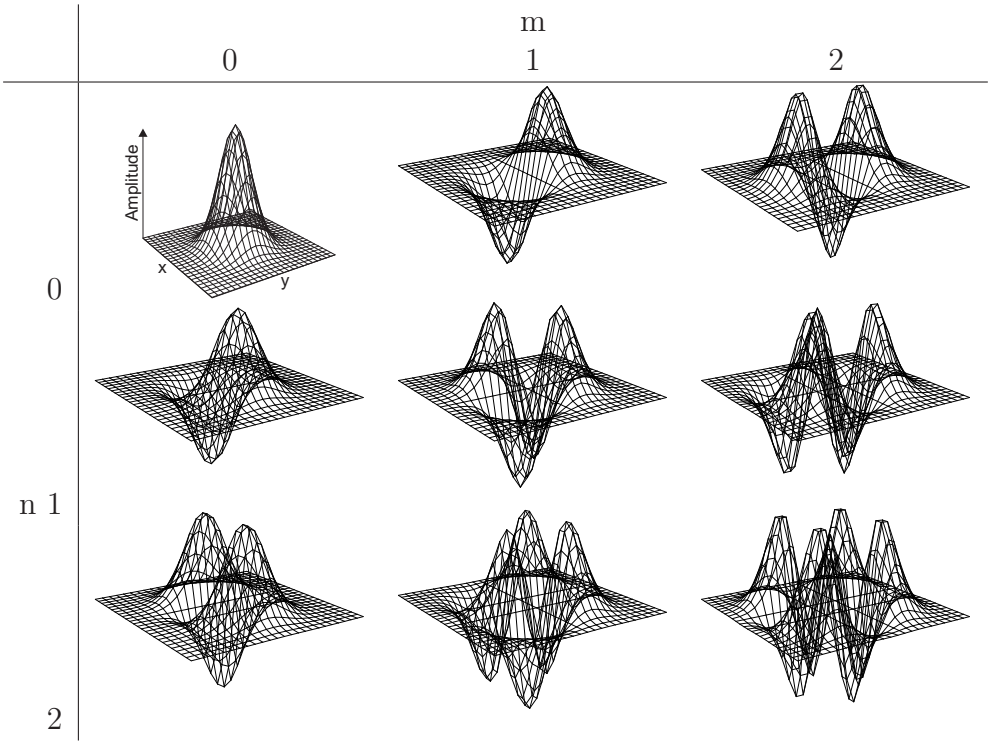


Table 1.2: Two dimensional Gaussian-Hermite polynomials, $H_{m,n}(x, y)$.

The phase slippage can also be given in terms of the confocal distance by

$$\phi_0 = \arctan \frac{z}{z_c}. \tag{1.46}$$

In Figure 1.15 the circles and their centres defining the spherical phase-front of a beam are shown. Figure 1.15 illustrates that as the radius of curvature approaches z, at large distances from the waist, the centre of the circle defining the radius at z becomes closer to the waist location. The beam radius w, the radius of curvature R, and the phase slippage ϕ_0, given in Equations 1.36, 1.44 and 1.46 respectively, change appreciably between the beam waist, at $z = 0$, and the confocal distance, z_c. The near-field is defined by $z << z_c$ and the far-field by $z >> z_c$. The region around z_c is called the transition region. A linear increase in w occurs in the far-field region[6]. The angle of divergence at $z = \infty$, shown in Figure 1.14 *top*, is given by

$$\theta_0 = \lim_{z\to\infty} \left[\tan^{-1}\left(\frac{w(z)}{z}\right)\right] = \tan^{-1}\left(\frac{\lambda}{\pi w_0}\right) \approx \frac{\lambda}{\pi w_0}. \tag{1.47}$$

Figure 1.16 shows the first 10 Hermite modes in one dimension. The modes are indexed 0, 1, 2,..., where 0 is the fundamental. The mode number is equal to the number of zero crossings of the function. Plots of the modes in two dimensions are shown in Table 1.2; they are the product of the one dimensional field in the x and y directions. Gaussian-Hermite polynomials are expressed in Cartesian coordinates

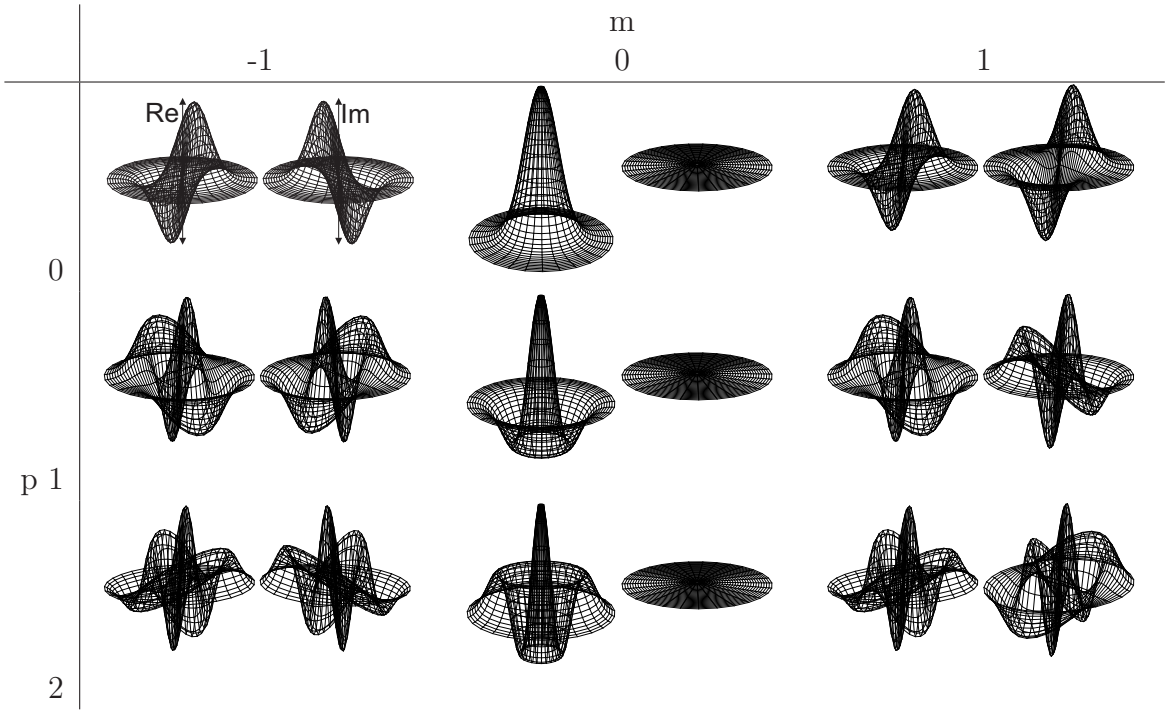


Table 1.3: Gaussian-Laguerre polynomials L_{pm}, in two dimensions. The modes are shown in pairs of real and imaginary parts.

(Figure 1.17 *left*) and have rectangular symmetry, which is appropriate for use with electric fields of rectangular symmetry (e.g. the field from a rectangular horn). However, for feeds which have circular symmetry (e.g. the field from a conical horn) a mode set of the same circular symmetry is more appropriate, as fields can be reconstructed with fewer modes. The, circularly symmetric, Gaussian-Laguerre polynomials, shown in Table 1.3, are one such mode set. These are indexed with $m, \ldots, -2, -1, 0, 1, 2, 3, \ldots$ and $p, 0, 1, 2, \ldots$. The Laguerre modes are given by

$$\psi_{pm} = \sqrt{\frac{2}{\pi}}\sqrt{\frac{p!}{(p+|m|)!}}\frac{2^{\frac{|m|}{2}}\left(\frac{r}{w}\right)^{|m|}}{w}L_{p|m|}\left(\frac{2\theta^2}{w^2}\right)\exp(im\phi - \frac{r^2}{w^2}), \qquad (1.48)$$

where $L_{p|m|}$ is the Laguerre polynomial of index p and m, and r and θ are the polar coordinates (Figure 1.17 *right*).

Many other basis sets exist, including: plane waves, spherical waves and Gabor modes, and these can all be used to describe a field[17]. Consideration of the nature of the field and the surfaces with which it interacts will determine the most suitable choice of mode set.

If system components are too small relative to the beam size, they can truncate the field, causing power from the fundamental mode to scatter into higher order modes. The level of truncation may be estimated by assuming that the beam is essentially Gaussian. The percentage of power that falls within radius r_e, shown in

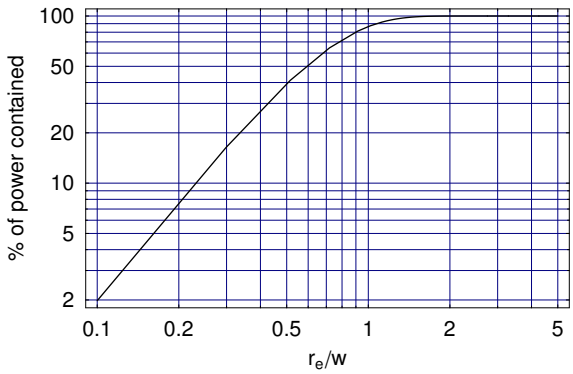


Figure 1.18: The power contained with radius r_e/w of a Gaussian beam.

Figure 1.18, is given by

$$100 \times \int_0^{r_e} E(r)^2 2\pi r dr, \tag{1.49}$$

where $E(r)$ is a Gaussian beam normalised so $\int_0^\infty E(r)^2 2\pi r dr = 1$. Therefore, a beam which is incident on a mirror where the ratio of aperture radius to beam width is $1r_e/w$ will capture 86.47% of the power, with a ratio of $2r_e/w$, this values rises to 99%, and at $2r/w$, 99.96% of the power is reflected. Despite these three high values, a sharp-edged aperture with ratio of aperture radius to beam radius of $2r/w$, produces an intensity variation of the field (i.e. ripples) of $\approx \pm 17\%$ in the near-field, and an aperture radius of $2.3r_e/w$ is required to reduce near-field ripples to the $\approx \pm 1\%$ level, and reduce the far-field side lobes[16].

1.4.3 ABCD Matrices

With Gaussian beam mode analysis, the beam radius of a propagating beam at any plane can be determined using ABCD matrices. It can be shown that the complex $\widetilde{q}$ parameter can be transformed through an arbitrary paraxial system by exactly the same rules that apply to purely geometric optics[16, 6]. This is due to the similarity between the complex beam parameter q and the ray parameter r. In geometric optics, the position r and slope r' of a ray at the input and output of a system element are related by

$$\begin{bmatrix} r_{out} \\ r'_{out} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} r_{in} \\ r'_{in} \end{bmatrix}, \tag{1.50}$$

so

$$r_{out} = A.r_{in} + B.r'_{in} \text{ and} \tag{1.51}$$

$$r'_{out} = C.r_{in} + D.r'_{in}. \tag{1.52}$$

The radius of curvature is defined as $R = r/r'$; therefore, by combining Equation 1.51 and 1.52, the radius of curvature for geometric optics at the output, is given by

$$R_{out} = \frac{A.R_{in} + B}{C.R_{in} + D}. \tag{1.53}$$

Similarly, for Gaussian beam modes, the complex parameter $\widetilde{q}_{out}$ at the output of a system element is given by

$$\widetilde{q}_{out} = \frac{A.\widetilde{q}_{in} + B}{C.\widetilde{q}_{in} + D}. \tag{1.54}$$

From Equation 1.34, the beam radius w and the radius of curvature R are given by

$$w = \left[\frac{\lambda}{\pi \Im\{-1/\widetilde{q}\}}\right]^{\frac{1}{2}} \text{ and} \tag{1.55}$$

$$R = \left[\Re\left\{\frac{1}{\widetilde{q}}\right\}\right]^{-1}. \tag{1.56}$$

Examples of transformation matrices for various system elements are given in Table 1.4.

Free space	Curved interface	Thin lens	Pair of thin lenses
$\begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ \frac{n_2-n_1}{n_2 R} & \frac{n_1}{n_2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}$	$\begin{bmatrix} -\frac{f_2}{f_1} & f_1+f_2 \\ 0 & \frac{f_2}{f_1} \end{bmatrix}$

Table 1.4: Examples of Transformation matrices. Free space: L is the propagation distance. Curved interface: from index n_1 to n_2 ($R > 0$ if concave to left). Thin lens: focal length f. Pair of thin lenses: first f_1 and then f_2, separated by the sum of their focal distances.

To determine the effect of a succession of elements, an overall system matrix can be calculated as a multiplication of the appropriate 2×2 matrices in reverse order to the direction of propagation, to give a 2×2 matrix that represents the whole system. The overall matrix for the system, shown in Figure 1.19, is given by

$$M = \begin{bmatrix} 1 & L_3 \\ 0 & 1 \end{bmatrix} . \begin{bmatrix} 1 & 0 \\ -\frac{1}{f_2} & 1 \end{bmatrix} . \begin{bmatrix} 1 & L_2 \\ 0 & 1 \end{bmatrix} . \begin{bmatrix} 1 & 0 \\ -\frac{1}{f_1} & 1 \end{bmatrix} . \begin{bmatrix} 1 & L_1 \\ 0 & 1 \end{bmatrix} \tag{1.57}$$

To determine the value of w and R at the output of the system, the complex beam parameter $\widetilde{q}$ is calculated using Equation 1.54. With this value of $\widetilde{q}$, w, ϕ and R are given by Equations 1.55 and 1.56.

ABCD matrices can be used to calculate w and R as a beam propagates through an optical system. These are the same for all modes. If the effects of truncation can be ignored ($D > 4w$), then a beam can be reconstructed at any plane in the system using the initial mode coefficients and the appropriate values of w, ϕ and R in the

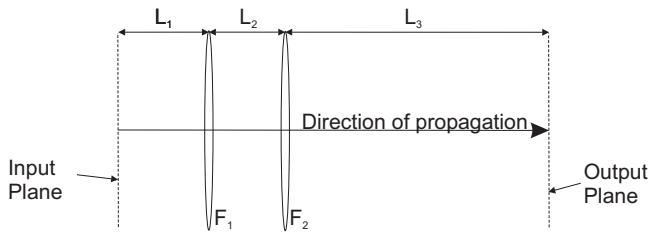


Figure 1.19: The system corresponding to the ABCD matrix of Equation 1.57, with of lenses focal length F_1 and F_2 and free space propagation distance L_1, L_2 and L_3.

Gaussian-Hermite mode expression. If truncation is likely to be important then the mode coefficients must be re-evaluated using Equation 1.43 and integration limits corresponding to the size of the truncating aperture.

In this thesis, frequent use is made of the GBM analysis of quasioptical systems.

1.5 Thesis Outline

A summary of the work discussed in this thesis follows:

Chapter 2 investigates some of the standard techniques used in beam shaping, where one uses a Diffractive Optical Element (DOE) to achieve some desired far-field intensity distribution. Firstly, repeating structures called Dammann gratings are designed to convert an incident Gaussian beam to an array of beams in the far-field. Next, techniques are investigated that can be used to transform a given input beam to the desired far-field intensity distribution, or at least a close approximation to it. The iterative Gerchberg-Saxton Algorithm (GSA), and the global optimisation algorithms Simulated Annealing (SA) and Differential Evolution (DE) are used to convert a Gaussian field to the desired far-field intensity distributions.

Chapter 3 introduces a novel method of DOE design, where the shape of the DOE is found by optimising the coefficients of a Gaussian-Hermite mode set. Typically, regardless of the method used, an exact transform from input field to target far-field intensity distribution cannot be achieved using only one phase transforming surface. A method whereby two surfaces are used to impose a phase on the field to achieve an exact transform is investigated. This mode set optimisation method is then extended to control the beam as it propagates through the near-field region.

Chapter 4 concerns the Millimetre-wave Bolometric Interferometer (MBI) telescope. MBI is a novel type of interferometer that uses bolometers, rather than the standard heterodyne techniques used at low frequecies. In particular I was involved in the design of the Cassegrain quasioptical beam combiner. In the initial design, the field from the input feed-horns was truncated at the secondary of this Cassegrain telescope; feed-horns lenses are designed to reduce this truncation. Conical horns couple the radiation from the sky (interference fringes) to the detector bolometers;

lenses are designed to increase this coupling. A number of alternative approaches to increasing the coupling are also investigated. The optimisation techniques of Chapters 2 and 3 were used in the lens design.

Chapter 5 concerns the telescopes of the Atacama Large Millimeter Array (ALMA). A DOE is designed to improve the coupling of the signal from the sky to the detector horn. The relationship between the location of the DOE in the signal path and the improvement in coupling is investigated. It is demonstrated that the beam shaping techniques of Chapters 2 and 3 may be employed to further increase the coupling by using *two* phase transforming surfaces.

Chapter 6 concerns the Heterodyne Instrument for the Far Infrared (HIFI) on the Herschel Space Observatory (HSO), and describes the modelling of the radiation propagation through the optics using a PO software package. The author was involved in modelling a high frequency channel that uses a lens antenna to couple the radiation from the telescope, rather than the waveguides used in the lower frequency channels. Separate software was used to simulate the far-field from the lens antenna. However, the near-field, and therefore, the position of the so called phase centre of the lens antenna field was required for the PO model. A number of techniques used to determine this location were investigated. Lens antennas are relatively poorly characterised compared to waveguides, and in order to gain a better understanding of them, a ray tracing simulation of a lens antenna fabricated at NUI Maynooth, was compared with a measured field.

Finally Chapter 7 summarises the main conclusions from this work.

Chapter 2

Beam Shaping Part I: Fourier Techniques

2.1 Chapter Outline

In this chapter and Chapter 3, I introduce standard beam shaping techniques, which, although standard, are not commonly used at millimeter and sub-millimetre wavelengths. One possible application would be to create multiple local oscillator beams from a single input beam. More applications maybe found for beam shaping in the future, as commercial applications for millimetre and sub-millimetre are developed. The techniques explored in this chapter and Chapter 3 are employed in Chapters 4 and 5 to improve the coupling of the detector horns to the signal from the sky. In Section 2.3, I design periodic structures, called diffraction gratings, to split an incident Gaussian beam into multiple beams in the far-field. The gratings designed are simple binary structures, giving a limited degree of control over the far-field pattern. In Section 2.4, I use the standard GSA technique to design DOEs that convert an incident Gaussian field to top-hat amplitude distribution, a first-order Gaussian-Hermite amplitude distribution and a set of three Gaussian beams in the far-field. In Section 2.6, I design DOEs using global optimisation algorithms (explained in Section 2.5), where the phase imposed on the Gaussian field at the DOE is optimised to achieve the specified target far-field pattern (or at least an approximation to it). The ease of finding a solution and the profile of the solution is dependent on the amplitude distribution of the input beam and desired output profile; the target distributions used here were chosen to compare the different techniques for a variety of target distributions.

The beam shaping is carried out in one-dimension for convenience; it is shown how the one-dimensional solutions can be directly transformed to two-dimensional solutions which are separable in rectangular coordinates.

2.2 Introduction

In this section, Diffractive Optical Elements (DOEs) - which transform a field distribution at a focal-plane to a prescribed intensity distribution in the far-field - are designed. The light incident on the DOE is assumed to be coherent and monochromatic. DOEs impose a phase on the field (using a dielectric material, or a profiled reflector), thereby creating the desired interference pattern at some distance from the DOE. Continuous phase functions may be approximated by a stepwise profile (binary optics), in order to reduce the number of lithographic masks required, making them more straightforward to fabricate[18]. This simplification of the profile of the grating is especially important at short wavelengths, where the features of the DOE are very small. In this chapter, DOEs for use at millimetre and submillimetre wavelengths are considered. At these longer wavelengths the feature size of the DOEs are larger, and are therefore more straight forward to fabricate - using a milling machine for example - as the DOEs have smoothly varying surfaces.

The design of DOEs is essentially the same problem as phase retrieval, and the same algorithms that are used to design DOEs are used to recover the phase of a field from intensity measurements at two planes[19]. The phase of a field is often required in many areas including microscopy, wavefront sensing, astronomy and crystallography[19]. Phase can be recovered from intensity measurements using the holographic technique - which involves the addition of a reference wave to the wave of interest in the recording plane. The complete field may be determined from the resulting hologram[20]. This method has been successfully demonstrated at millimetre wavelengths[21]; however, it requires the use of additional equipment. Here, phase retrieval is investigated when only intensity data and no reference wave is available. For beam shaping, the problem is to determine what phase distribution, when imposed on the input field, will produce the required far-field distribution. In the far-field beam shaping carried out here, a DOE imposes a phase on a Gaussian input beam at its waist in order to obtain the specified intensity pattern in the far-field, as shown in Figure 2.1; however, the methods described here can be used for any arbitrary input field.

Although the same algorithms that are used to design DOEs are also used to recover the phase of a field from intensity measurements, there are some differences in the nature of the two problems. With phase retrieval one is generally interested in retrieving the phase from noisy data, and a solution is known to exist. However, with DOE design - where one wishes to transform the input beam to a target profile at another plane - an exact solution may not exist[22]. The design of DOEs is a non-linear ill-posed inverse problem, and is therefore difficult. Due to this ill-posed nature, a solution may not exist, and if it does it may not be unique or stable[23].

Analytical methods based on geometric transforms may be applied to the design

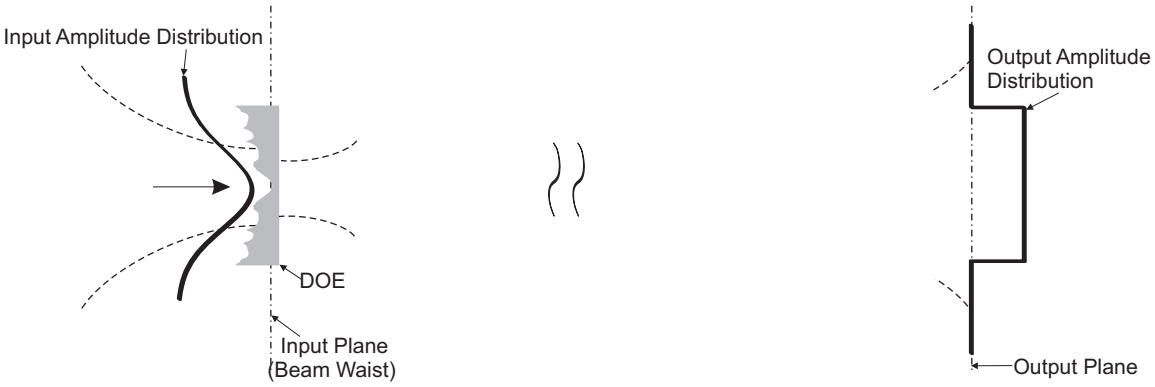


Figure 2.1: A Gaussian beam is focused on to the DOE, which transforms it to a top-hat distribution in the far-field. This is idealised; in reality it is not always possible to achieve an exact transform.

of beam shaping elements with profile distributions that are much larger in scale than the wavelength of radiation incident on them[24]. However, the dimensions of the elements considered here are comparable with the wavelength. Despite this, at these wavelengths geometrical transformations - in which diffraction effects are ignored - are commonly used in DOE design, to give an approximation to the target far-field intensity pattern[25]. A common approach to DOE design which takes diffraction into account is to use the Gerchberg-Saxton Algorithm (GSA) and related algorithms[25, 19, 18]. These, bidirectional, algorithms involve iterative Fourier transformation back and forth between the object and Fourier domains. The unidirectional global optimisation algorithms Simulated Annealing (SA) and Differential Evolution (DE) have also been used to design DOEs. They vary the parameters which encode the phase imposed on the input field, in order to find the solution which results in the best match between the target and calculated distributions - as quantified by some merit function. They are called unidirectional because they only propagate the field in the forward direction (from near- to far-field) - as opposed to the bidirectional GSA, which propagates the field in both directions. Global optimisation algorithms are required due to the non-linear nature of the problem. Local search algorithms such as the quasi-Newton method and simplex method find less optimal solutions. Whereas SA operates on a single set of parameters, DE uses a population of parameters, and tends to produce more optimal results. In this chapter, DOE design using all these techniques is further investigated, and a novel method of DOE design in which a Gaussian beam mode set is optimised, using DE, is introduced.

2.3 Diffraction Gratings

When a beam is transmitted or reflected by any periodic structure, it is diffracted into multiple orders. Here, it is demonstrated that periodic structures, called diffraction gratings, can be used to split a Gaussian beam incident on the grating to multiple beams in the far-field. The period of the grating (the length of the cell - i.e. the smallest profile section that is repeated) determines the angular separation between the orders. The angle θ_m of each output beam of order m is given by the grating equation as follows

$$\sin(\theta_m) = \frac{m\lambda}{P}, \tag{2.1}$$

where P is the period of the structure, m is the order number and λ is the wavelength. This grating equation gives the angular separation of each of the orders, but it does not give the relative power in each; this is determined by the profile of the cells. Continuous cell profiles can be designed to give a large degree of control of the relative power in the diffraction orders. A combination of phase and attenuation/blocking features can be incorporated into the grating, allowing further design freedom to create the desired output distribution; however, the feature size (the smallest element size of the grating) is limited by the manufacturing process. Here, only binary phase-only gratings will be considered, which are approximations to continuous profiles.

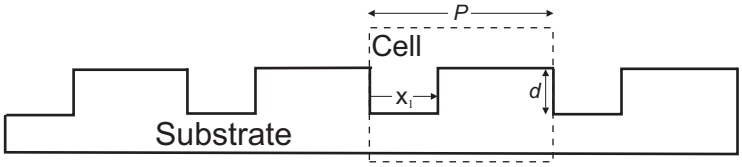


Figure 2.2: A one-dimensional section of a simple phase grating, with free parameter x_1.

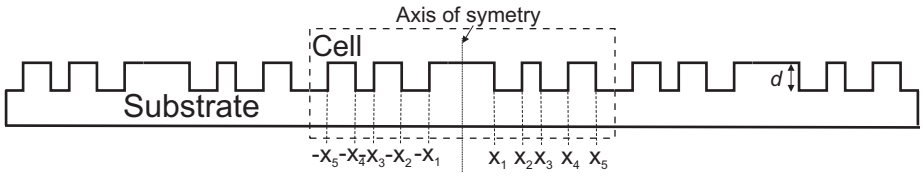


Figure 2.3: A one-dimensional section of a simple phase grating, with free parameters x_1, x_2, x_3, x_4 and x_5.

A simple transmissive binary grating, of period P, shown in Figure 2.2, which transforms a uniform incident field, is analysed here. The phase element of depth d imposes a phase $\phi = 2\pi d(n-1)/\lambda$, where n is the refractive index. Here, d is chosen such that it imposes a phase $\phi = \pi$ rad on the field. A binary phase grating is best understood through Fourier optics theory, and specifically convolution theory, as follows. The cell of Figure 2.2 is two offset rect() functions with a phase

constant applied to the second. To define the unit cell, the offset can be achieved in mathematical notation as follows:

$$\zeta(x) = \text{rect}(x/x_1) * \delta(x + x_1/2) + [\text{rect}[x/(P - x_1)] * \delta(x - (P - x_1)/2)]e^{i\phi}, \quad (2.2)$$

where $*$ denotes convolution. The grating can be generated from this unit cell by convolving with the comb function as follows[18]

$$\text{grating} = \zeta(x) * comb(x/P). \quad (2.3)$$

Using the identity

$$\mathcal{F}\{f * g\} = k\left[\mathcal{F}(f)\right] \cdot \left[\mathcal{F}(g)\right], \quad (2.4)$$

the far-field is given by the multiplication of $\mathcal{F}\{comb(x/P)\}$ and $\mathcal{F}\{\zeta(x)\}$ times a constant, k. The Fourier transform of a comb function is also a comb function[26], corresponding to the diffraction orders at the angles given by Equation 2.1. The Fourier transform of the unit cell, $\mathcal{F}\{\zeta(x)\} = F(\xi)$, multiples with the comb function to give the output field. $F(\xi)$, therefore, forms an envelope function that gives the amplitude and relative phase of the diffracted orders. Since the unit cell $\zeta(x)$ is two offset rect() functions and the Fourier transform of a rect() function is a sinc function, the transform of $\zeta(x)$ is the sum of two sinc functions with offset phase terms. The far-field is given by

$$\begin{aligned} F(\xi) &= \frac{x_1}{P}\frac{\sin(\pi x_1 \xi)}{\pi x_1 P}\exp(I2\pi\xi P/2) \\ &\quad + \frac{P - x_1}{P}\frac{\sin(\pi(P - x_1)\xi)}{\pi(P - x_1)\xi}\exp[i2\pi\xi(P - x_1)/2]. \end{aligned} \quad (2.5)$$

This simplifies to

$$F(\xi) = \frac{1}{2}\frac{\sin(\pi x_1 \xi)}{\pi x_1 \xi}[\exp(i\pi x_1 \xi) - \exp(-i\pi x_1 \xi)]. \quad (2.6)$$

This function is then multiplied by the comb function, and is therefore only defined where the comb function is non-zero, as

$$|F(\xi)|_{q=\frac{m}{P}} = \frac{\sin(m\pi/2)}{m\pi 2}[i\sin(m\pi/2)], \quad (2.7)$$

which is zero for all even values of m. For odd values of m, the diffraction efficiency from the grating is given by

$$\eta_m = \left|\zeta\left(\frac{m}{P}\right)\right|^2 = \left(\frac{2}{m\pi}\right)^2. \quad (2.8)$$

N	η_1	x_1	x_2	x_3	x_4
1	66.4	0.368			
2	77.4	0.132	1.480		
3	65.7	0.122	0.644	0.496	
4	66.3	0.100	0.136	0.370	0.498

Table 2.1: The number of solutions, S_N, that satisfy Equations 2.12 and 2.13 is given by Equation 2.14. The transition points of the solutions with the largest η_1 are given.

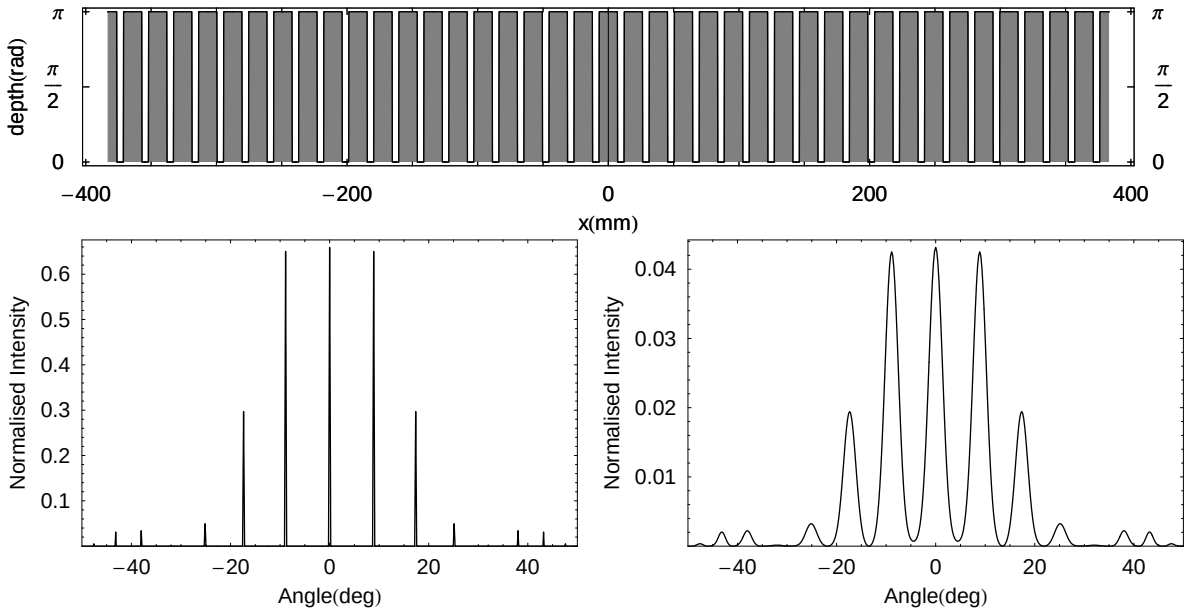


Figure 2.4: A grating is designed with $N = 1$ transition points at $x_1 = 0.368$ and $\lambda = 3$mm, which creates a far-field with a uniform intensity in the central three diffraction orders: *top* the profile of the grating, *bottom left* the far-field from the grating with a uniform field incident on the grating, *bottom right* the far-field from the grating with a field with Gaussian amplitude distribution of radius 20mm and flat phase-front incident on the grating.

In Reference [27] Dammann gratings are designed, in which the grating is binary, but each cell has a number of elements. The gratings have the form shown in Figure 2.3, where the cell profile is defined using transition points to give symmetric cells. The transition points $x_1, x_2, x_3, \ldots, x_N$ are free parameters that are optimised to control the relative power in the beams. The amplitudes of the diffraction orders from a Dammann grating are given, by Fourier transform, as follows

$$\begin{aligned} \mu_0 &= 2\sum_{k=0}^{N}(-1)^k(x_{k+1} - x_k), \\ \mu_n &= \frac{1}{n\pi}\sum_{k=0}^{N}(-1)^k[\sin(2\pi n x_{k+1}) - \sin(2\pi n x_k)]. \end{aligned} \tag{2.9}$$

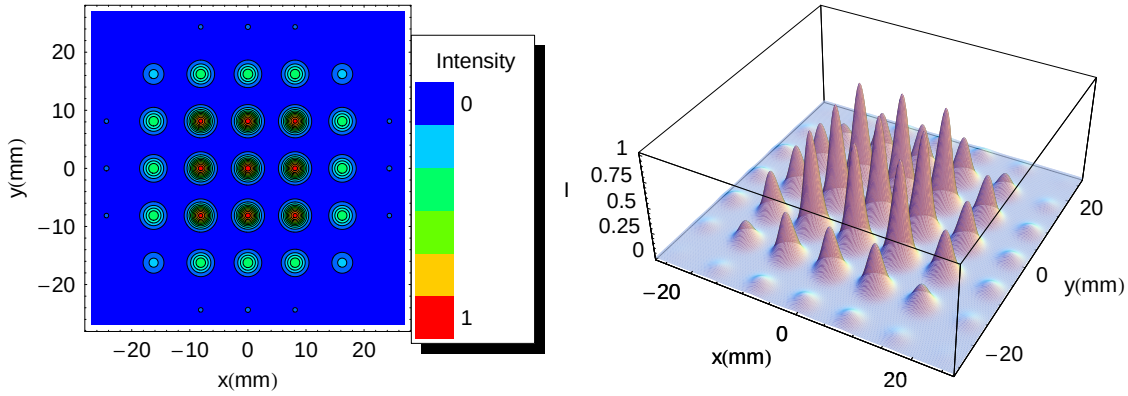


Figure 2.5: The two-dimensional far-field generated by the grating given by $\zeta(x, y) = \zeta(x)\zeta(y)$: *left* contour plot of intensity, *right* three-dimensional plot of intensity.

The total power in Equation 2.9 has been normalised to equal 1, as follows

$$\sum_{n=-\infty}^{n=+\infty} |\mu_n|^2 = 1. \tag{2.10}$$

Therefore, the figure of merit (the efficiency), that gives the fraction of light power directed into the central $2N + 1$ diffraction orders, can be given by

$$\eta_1 = \sum_{n=-N}^{n=N} |\mu_n|^2 = |\mu_0|^2 + 2\sum_{n=1}^{N} |\mu_n|^2. \tag{2.11}$$

The objective of Dammann Grating design is to achieve

$$|\mu_0|^2 = |\mu_n|^2 \text{ for } n = 1, 2, \ldots, N, \tag{2.12}$$

while maximizing η_1, with

$$0 \leq x_k \leq x_{k+1} \leq 1/2. \tag{2.13}$$

No analytical method exists for solving Equations 2.12 and 2.13, but solutions have been found by searching the parameter space of $x_1, x_2, x_3, \ldots, x_N$[27]. It has been shown [27] that the number of unique solutions is given by

$$S_N = 2^{N-1}. \tag{2.14}$$

The computational task of finding unique solutions becomes more onerous with more transition points, but global optimisation methods may be used to reduce the computation time required to find a solution with a high efficiency, η_1. The transition points of the solutions with the highest η_1 are given in Table 2.1.

Figure 2.4 *top* shows a grating with one transition point and 40 cells, used to create three beams. A grating is designed with three output beams and with the first diffraction order occurring at 9° off-axis. The period of the cell of the grating, given

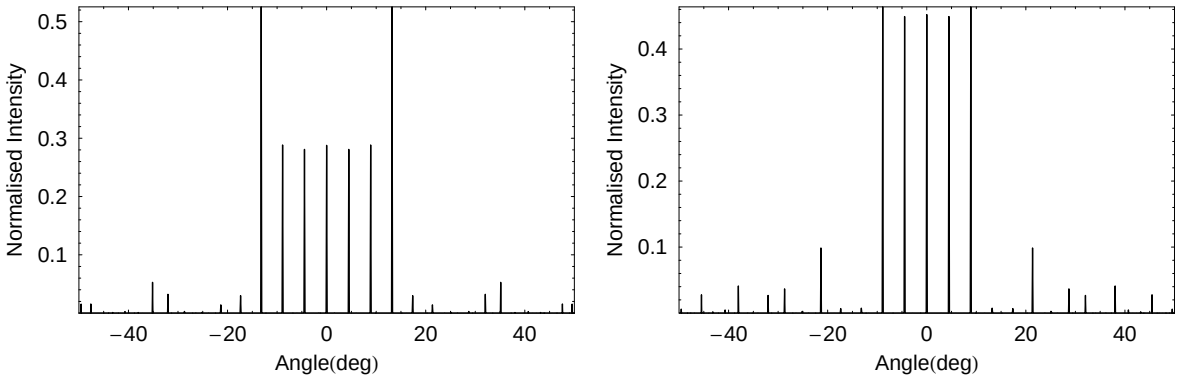


Figure 2.6: The far-fields from two gratings with $N = 2$ transition points, which create a far-field distribution with a uniform intensity at the central five diffraction orders: *left* with transition points $x_1 = 0.242$ and $x_2 = 0.414$, giving an efficiency of $\eta_1 = 48.2\%$; and *right* with transition points $x_1 = 0.132$ and $x_2 = 1.480$, giving an efficiency of $\eta_1 = 77.4\%$.

by Equation 2.1, for the first diffraction order located at 9° is 19.2mm. As there is only one transition point, only one solution exists[27] which satisfies Equations 2.12 and 2.13: $x_1 = 0.368$. Figure 2.4 *left* shows a plot of the diffraction orders when the grating is illuminated by a uniform distribution, given by the FFT of the field at the grating. The three central orders are of uniform intensity, but there is significant power in the higher orders, reducing the efficiency to 66.4%. In practice, at millimetre wavelengths the gratings are typically illuminated with a Gaussian beam with a flat phase-front. The far-field pattern resulting from a non-uniform distribution incident on the grating is that of Figure 2.4 *bottom left* convolved with the far-field pattern of the incident field. The far-field divergence angle of the incident Gaussian beam with a flat phase-front is given by $\theta_0 = \tan^{-1}\left(\frac{\lambda}{\pi w_0}\right)$, where w_0 is the radius of the field incident on the DOE (i.e. the waist); thus, the far-field from the grating is comprised of Gaussian distributions of divergence angle of θ_0, each centred on a diffraction order, and each with a peak intensity proportional to the power of the diffraction orders shown in Figure 2.4 *bottom left*. The far-field pattern of a 20mm-radius Gaussian beam with a flat phase-front incident on the grating is shown in Figure 2.4 *bottom right*; the beams are separated by 9°, and each have a divergence angle of 2.7°. The efficiency of the grating (i.e. the percentage of power contained in the three central beams) is equal to the efficiency for a grating illuminated with a uniform distribution: 66.4%.

From the one-dimensional grating, $\zeta(x)$, which gives the far-field distribution $F(\eta)$, a two-dimensional grating with symmetry in the x and y directions can be constructed from

$$\zeta(x, y) = \zeta(x)\zeta(y), \tag{2.15}$$

and this gives the far-field distribution,

$$F(\xi, \eta) = F(\xi)F(\eta). \tag{2.16}$$

Using Equation 2.15, the two-dimensional far-field distribution generated from the one-dimensional grating of Figure 2.4 *right* is found, and is shown in Figure 2.5.

A grating with more transition points can create a far-field distribution with a larger number of diffraction orders satisfying Equations 2.12 and 2.13. Figure 2.6 shows the far-field of Dammann gratings whose cell has $N = 2$ transition points; the two solutions given in [27], with efficiencies of 48.2% and 77.4%.

The binary gratings only allow a limited amount of control over the far-field distribution, resulting in low efficiencies. Next, DOEs will be designed with unlimited phase freedom. These DOEs allow a much greater degree of control over the far-field pattern, allowing approximations to arbitrary target far-field patterns to be achieved.

2.4 Bidirectional Optimisation (GSA)

A common method used in DOE design is the iterative approach of the GSA[23, 19]. The GSA was originally intended to reconstruct phase from intensity measurements at two planes[19]. The algorithm propagates a field between the near-field and the far-field, replacing the amplitude with the target amplitude at each plane, while retaining the phase. The intensity distribution of the current solution has been shown to become closer to the target with each iteration. The geometry used in the GSA is shown in Figure 2.7.

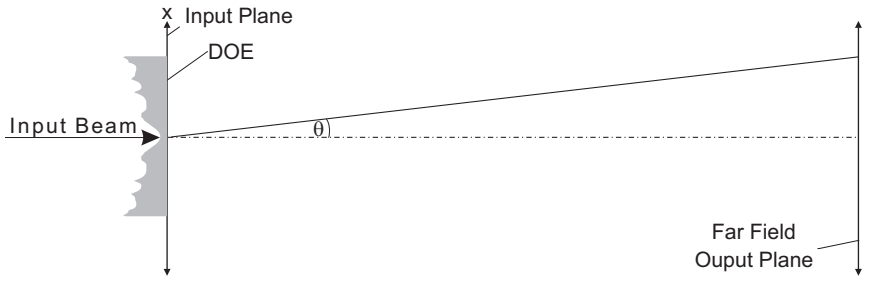


Figure 2.7: The geometry of the DOE design problem to transform an input beam to a required intensity distribution in the far-field.

I will call the target near- and far-field amplitude distributions $T_n(x)$ and $T_f(\theta)$ respectively; for example, for a DOE that is designed to convert an input Gaussian beam with a flat phase-front incident on the DOE to a top-hat amplitude distribution in the far-field, the near-field target distribution is $T_n(x) = \exp[-(x/w)^2]$ and the far-field is $T_f(\theta) = \text{rect}(\theta)$. The calculated far-field distribution is $E_f(\theta)$, with

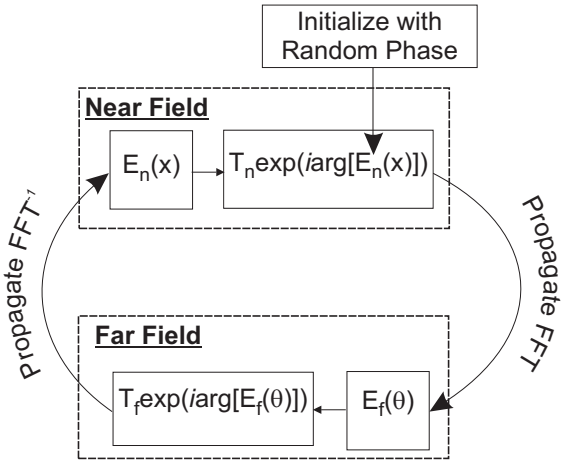


Figure 2.8: The GSA propagates between the near- and far-fields. At each plane the phase of the transformed field is kept and the amplitude is replaced with the target amplitude, T_n or T_f.

phase $\phi_f(\theta) = \arg[E_f(\theta)]$, the calculated near-field distribution is $E_n(x)$ with phase $\phi_n(x) = \arg[E_n(x)]$. The GSA involves the following steps, which are shown in Figure 2.8:

1. Some initial near-field phase distribution, $\phi_n(x)$, is chosen.
2. The near-field amplitude distribution is set to be the target (i.e. known) input distribution as follows: $|E_n(x)| \exp[i\phi_n(x)] \rightarrow T_n(x) \exp[i\phi_n(x)]$.
3. $T_n(x) \exp(i\phi_n(x))$ is transformed to the far-field to give $E_f(\theta)$.
4. The far-field amplitude distribution is replaced with the target distribution as follows: $|E_f(\theta)| \exp[i\phi_f(\theta)] \rightarrow T_f(\theta) \exp[i\phi_f(\theta)]$.
5. $T_f(\theta) \exp(i \arg[E_f(\theta)])$ is transformed to the near-field, $E_n(x)$.
6. Proceed to step 2.

The loop is terminated after a set number of iterations, or when an acceptable solution has been found. The phase that converts $T_n(x)$ to $E_f(\theta)$ is given by $\arg[E_n(x)]$, from the last iteration of the GSA; it is this phase that defines the surface of the DOE.

The Mean Squared Error (MSE) is a metric used to evaluate how close a field pattern is to that prescribed. The MSE of near- and far-field solutions, $E_n(x)$ and $E_f(\theta)$, have been shown to reduce the with each iteration[20]. The MSE of the N-element array, x_{1i} where the desired array is x_{2i}, is given by

$$\mathrm{MSE}(x_1, x_2) = \frac{1}{N} \sum_{i=1}^{N} (x_{1i} - x_{2i})^2 . \tag{2.17}$$

The MSE between the prescribed and trial amplitude distributions in the near-field, $\mathrm{MSE_n}$, and the far-field, $\mathrm{MSE_f}$, is given by

$$\begin{aligned} \mathrm{MSE_n} &= \mathrm{MSE}(|E_n|, |T_n|), \\ \mathrm{MSE_f} &= \mathrm{MSE}(\left|E_f\right|, \left|T_f\right|), \end{aligned} \tag{2.18}$$

where $|E_n|$ and $\left|E_f\right|$ are the trial amplitudes in the near- and far-fields, and T_n and T_f are the target amplitudes in the near- and far-fields. The MSE of the current solution is not required for the operation of the GSA, but it is one of many merit functions that are useful for evaluating its progress.

2.4.1 Proof that the MSE Must Decrease with each Iteration

The following is a proof[20] that with each iteration of the GSA, the MSE of the near-field and far-field distributions decreases. Figure 2.9 shows the amplitude in the near- and far-field at the 100^{th} iteration to find a DOE that converts a Gaussian input amplitude to a super Gaussian amplitude distribution (defined in Section 2.4.7). In the far- to near-field transform of the GSA, $\left|E_f(\theta)\right|$ is replaced with the target far-field amplitude distribution, $T_f(\theta)$, while retaining the phase, to give the new near-field estimate, as follows

$$E_n(x) = \mathcal{F}^{-1}\{T_f(\theta)\exp(i\phi_f(\theta))\}, \tag{2.19}$$

where $\mathcal{F}$ denotes the DFT operation (steps 4 and 6). Alternatively, the modified far-field can be given as $E_f(\theta)$ plus an "error", given by the difference between the obtained and target fields as follows,

$$\mathrm{err_f} = E_f(\theta) - T_f(\theta)\exp[i\phi_f(\theta)]. \tag{2.20}$$

($E_f(\theta)$ has the same phase angle as $T_f(\theta)\exp[i\phi_f(\theta)]$, $\phi_f(\theta)$.) The field transformed to the near-field is then given by

$$\begin{aligned} E_n(x) &= \mathcal{F}^{-1}\{T_f(\theta)\exp[i\phi_f(\theta)]\} \\ &= \mathcal{F}^{-1}\{E_f(\theta) - \mathrm{err_f}(\theta)\}. \end{aligned} \tag{2.21}$$

The far-field "error", $\mathrm{err_f}(\theta)$, corrects the amplitude of $E_f(\theta)$ while retaining the phase. Similarly, the "error" in the near-field is given by

$$\mathrm{err_n} = E_n(\theta) - T_n(\theta)\exp[i\phi_n(\theta)], \tag{2.22}$$

The notation used here is summarised in Table 2.2. Plots of the near-field and far-field amplitude distributions with the "error" are shown in Figure 2.9. By splitting $E_n(x)$ and $E_f(\theta)$ into these component parts and transforming them to the far- and near-field respectively, we can prove that the total error across the field decreases with each iteration, as follows.

Near-Field	Direction	Far-Field
$T_n(x)\exp[i\phi_n(x)]$	$\rightarrow$	$E_f(\theta) = \mathcal{F}\{T_n(\theta)\exp[i\phi_n(\theta)]\}$
$E_n(x) - \mathrm{err_n}(x)$	$\rightarrow$	$E_f(\theta) = E_n^{\mathcal{F}}(\theta) - \mathrm{err_n^{\mathcal{F}}}(\theta)$
$E_n(x) = \mathcal{F}\{T_f(x)\exp[i\phi_f(x)]\}$	$\leftarrow$	$T_f(\theta)\exp[i\phi_f(\theta)]$
$E_n(x) = E_f^{\mathcal{F}}(x) - \mathrm{err_f^{\mathcal{F}}}(x)$	$\leftarrow$	$E_f(\theta) - \mathrm{err_f}(\theta)$

Table 2.2: In the GSA at each plane, the amplitude of the calculated field (E_n in the near-field and E_f in the far-field) is replaced with the target amplitude field (T_n in the near-field and T_f in the far-field). The transform of E_n and E_f to each plane can also be analysed in terms E_n and E_f minus the respective correction fields $\mathrm{err_n}$ and $\mathrm{err_f}$.

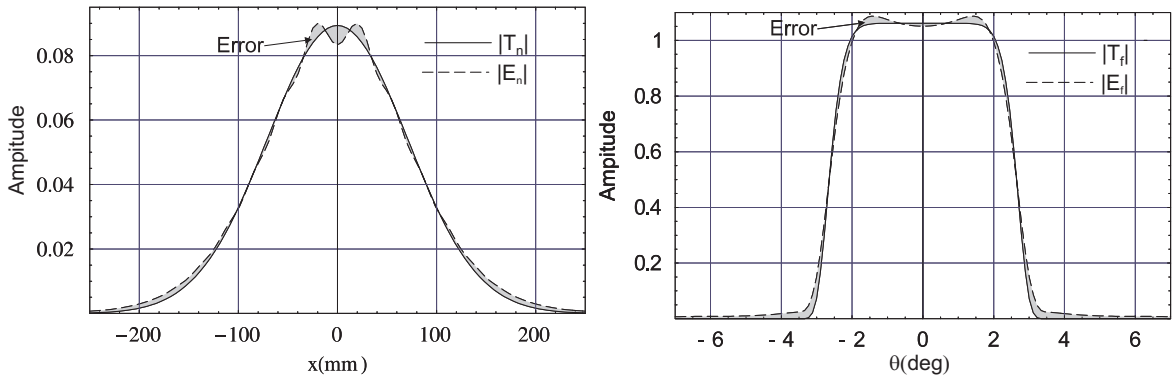


Figure 2.9: The result of the GSA after 100 iterations, for a DOE that converts a 100mm-radius Gaussian to a 2° radius super-Gaussian of order 10. The amplitude of the target field, the solution found, and the difference between the two (the "error" field), are shown. *Left* the near-field and *right* the far-field.

The transform of Equation 2.21 is now considered for an arbitrary point in the far-field, θ_1, and an arbitrary point in the near-field, x_1. Figure 2.10 *left* shows vectors $E_f(\theta_1)$, $T_f(\theta_1)\exp[i\phi_f(\theta)]$ and $\mathrm{err_f}(\theta_1)$. The vector $\mathrm{err_f}(\theta_1)$ is colinear with $E_f(\theta_1)$, as they have the same phase angle, $\phi_f(\theta_1)$. Figure 2.10 *right* shows the near-field at an arbitrary point x_1. The fields $E_f(\theta)$ and $\mathrm{err_f}(\theta)$, when transformed to the near-field, give $E_f^{\mathcal{F}}(x_1)$ and $\mathrm{err_f^{\mathcal{F}}}(x_1)$ respectively. The vector $E_f^{\mathcal{F}}(x_1)$ is equal in amplitude to $T_n(x_1)$ and $E_f^{\mathcal{F}}(x_1)+\mathrm{err_f^{\mathcal{F}}}(x_1) = E_n(x_1) = T_n(x_1)+\mathrm{err_n}(x_1)$. According to Parseval's theorem

$$\sum_{\text{all points}} \left|\mathrm{err_f}(\theta)\right|^2 \Delta\theta = \sum_{\text{all points}} \left|\mathrm{err_f^{\mathcal{F}}}(x)\right|^2 \Delta x, \tag{2.23}$$

where Δx and $\Delta\theta$ are the sampling intervals in the near- and far-fields. Therefore,

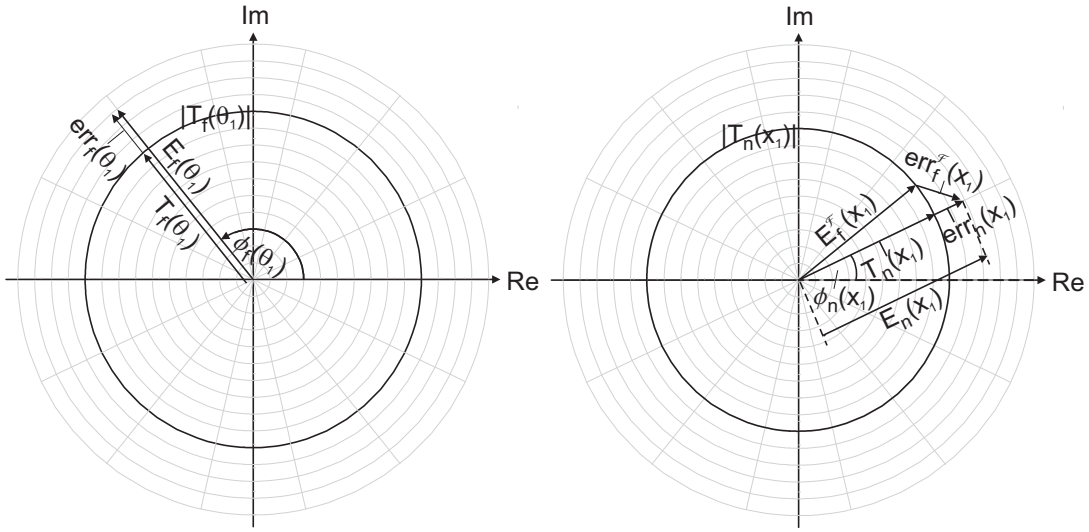


Figure 2.10: *Left* the field at an arbitrary point in the far-field and *right* an arbitrary point in the near-field. The far-field, $T_f(\theta)\exp[i\phi_f(\theta)] = E_f(\theta) - err_f(\theta)$ is transformed to the near-field to give $E_f^{\mathcal{F}}(x_1) - \mathrm{err_f}^{\mathcal{F}}(x_1) = T_n(x_1)\exp[i\phi_n(x_1)] + \mathrm{err_n}(x_1) = E_n(x_1)$. According to Parseval's theorem $\int_{-\infty}^{\infty}\left|\mathrm{err_f}^{\mathcal{F}}(\theta)\right|^2 d\theta = \int_{-\infty}^{\infty}\left|\mathrm{err_f}(x)\right|^2 dx$, and therefore from the diagram $\mathrm{err_n}(x) < \mathrm{err_f}(x)$ averaged across the field - since $\mathrm{err_n}(x_1) \leq \mathrm{err_f}^{\mathcal{F}}(x_1)$.

from Figure 2.10 *right*, since $err_n(x)$ must be less than or equal to $err_n^{\mathcal{F}}(\theta)$ at each point, the power of the error field must decrease or remain constant with each iteration.[20]

In the case of the Gaussian-to-super-Gaussian transformation of Figure 2.9, as shown in Figure 2.11 *left*, initially with each iteration $\mathrm{MSE_n}$ and $\mathrm{MSE_f}$ decrease. After a number of iterations the algorithm stagnates, and no further decrease can be made. This occurs when

$$\arg[\mathcal{F}^{-1}\{T_f(\theta)\exp(i(\phi_f(\theta))\})] \approx \arg[\mathcal{F}^{-1}\{E_f(\theta)\}] \tag{2.24}$$

and

$$\arg[\mathcal{F}\{T_n(x)\exp(i\phi_n(x)\})] \approx \arg[\mathcal{F}\{E_n(x)\}], \tag{2.25}$$

over the region of the respective target distributions, i.e. the GSA stagnates when the transform of the corrected amplitude field creates a field with the same phase distribution or nearly the same phase distribution as the uncorrected transformed field (they are colinear at each point).

For the initial iterations of the GSA Equations 2.24 and 2.25 are not true. As illustrated in Figure 2.11 *left*, as the GSA progresses $\mathrm{MSE_n}$ and $\mathrm{MSE_f}$ reduce and become equal in magnitude; the GSA then stagnates. Figure 2.11 *right* shows the field at x_1 when the field has stagnated; $\mathrm{err_n}(x_1)$ and $\mathrm{err_f}^{\mathcal{F}}(\mathrm{x_1})$ are both equal in magnitude and colinear; $E_n(x_1)$ and $\mathrm{err_f}^{\mathcal{F}}(\mathrm{x_1})$ are also colinear.

At the global minimum, $\arg[E_f^{\mathcal{F}}(x)]$ and $\arg[\mathrm{err_n}(x)]$ will be equal; however, the

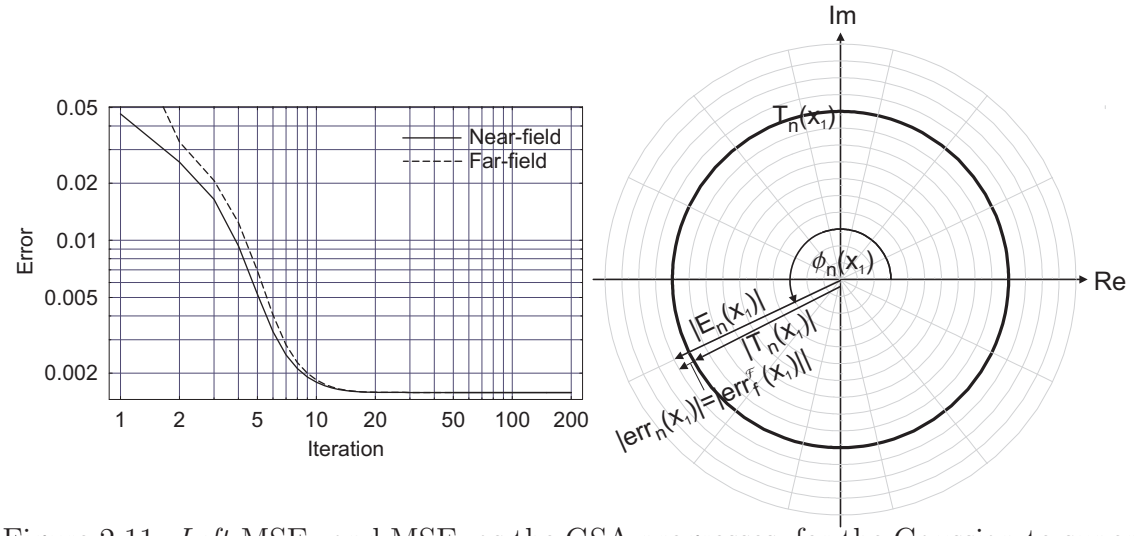


Figure 2.11: *Left* $\mathrm{MSE_n}$ and $\mathrm{MSE_f}$ as the GSA progresses, for the Gaussian to super Gaussian transform of Figure 2.9. With each iteration of the GSA, the MSE in the near- and far-fields decrease until $\mathrm{MSE_n} = \mathrm{MSE_f}$ and the algorithm stagnates. *Right* the field at x_1 when the GSA has stagnated. $\mathrm{err_n}(x_1)$ and $\mathrm{err_f^{\mathcal{F}}}(x_1)$, are equal in magnitude and colinear; $E_n(x)$ is also colinear with both $\mathrm{err_n}(x_1)$ and $\mathrm{err_f^{\mathcal{F}}}(x_1)$.

GSA may get stuck in a local minimum where Equation 2.24 and 2.25 are true. The initial phase distribution determines if the solution given by the GSA is a global or local minimum, but there is no way to know which beforehand.

2.4.2 Padding and Sampling Interval of the FFT for the GSA

The FFT - discussed in Chapter 1 - is used to transform between the near- and far-fields. Its use with the GSA is now discussed using the example of an input Gaussian transformed to a target top-hat amplitude distribution. In practice, when applying the FFT to propagate a field, consideration must be given to the amount of padding (where zero's are added to each end of the array to effectively increase its length) and the sampling rate in both the near- and far-fields, in order to reduce aliasing. When propagating between the near- and far-field in the GSA, the estimated near-field distribution is held in array A_N; this field is modified and transformed to give a second array A_F, the estimated far-field distribution. If, for example, there is not a sufficient amount of padding in A_N, A_F will be under-sampled. Below, the sampling interval is determined that gives an equal number of samples across the extent of the fields after they have been Discrete Fourier or inverse Discrete Fourier transformed, for a given total number of samples in the array.

In order to reduce aliasing, the number of samples for near- and far-field distributions should be approximately equal; therefore, the sampling interval and the amount of padding is chosen such that the number of sample points, N', between 0mm and the extent of the Gaussian in the near-field, Cw_0 (w_0 is the beam radius

and C is a constant), and the number of samples between 0° and the extent of the top-hat, θ_0, are equal, i.e.

$$N' = \frac{\theta_0}{\Delta\theta} = \frac{Cw_0}{\Delta x}, \tag{2.26}$$

where Δx and $\Delta\theta$ are the sampling intervals in the near- and far-fields respectively. A judgement needs to be made as to what the extent of the input Gaussian (adjusted by C) is set to in relation to the extent of the top-hat, because, whereas a top-hat has a definite extent, a Gaussian has an amplitude of zero only at $x = \pm\infty$. Figure 2.12 shows an example of a near-field Gaussian target distribution and a top-hat far-field target distribution of similar extent.

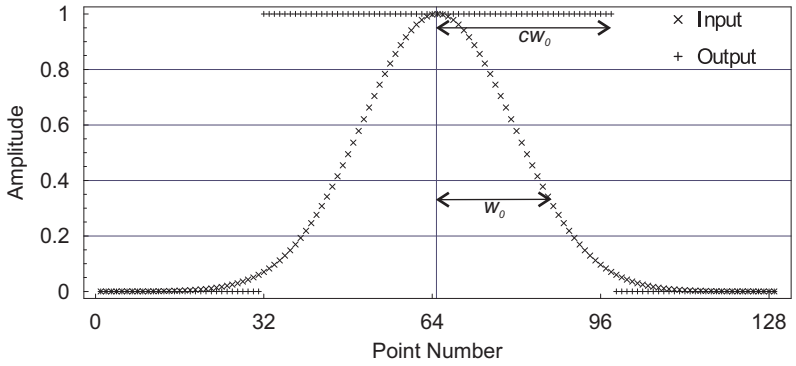


Figure 2.12: When propagating in the GSA the amount of padding is chosen such that the near- and far-field distributions (shown) have a similar extent.

The sampling in the far-field is related to the extent of the field transformed by the Discrete Fourier Transform (DFT) as follows

$$\Delta\theta = \sin^{-1}\left(\frac{\lambda}{N\Delta x}\right), \tag{2.27}$$

where N is the number of samples, λ is the wavelength and Δx is the sampling interval; the extent of the field is given by $N\Delta x$. Increasing the width of the field increases the resolution of the transform. The width of the field can effectively be increased by padding with zeros.

By substituting Equation 2.27 into Equation 2.26, then rearranging, the number of samples, N, is given by

$$N = \csc\left(\frac{\theta_0\Delta x}{Cw_0}\right)\frac{\lambda}{\Delta x}. \tag{2.28}$$

Now, for a given sampling interval in the near field, Δx, Equation 2.28 gives the number of samples in the array required so that the sampling interval is equal to $\Delta\theta$ when the field is Fourier transformed, and vice versa for the inverse transform. FFT algorithms generally operate fastest when $N = 2^n$, where n is an integer; therefore, the value of N is rounded up to the nearest value such that $N = 2^n$ (actually this

is not quite true, as it is rounded to $N = 2^n + 1$ because, as shown in Chapter 1, it is necessary to have an odd number of samples in the array and the last sample is dropped when the FFT is performed). Therefore, Equation 2.28 is used to find a value of Δx for which $N = 2^n + 1$, where n is set to give the required number of samples.

For target far-field distributions other than a top-hat transform, a similar judgement is made for the scale of the target far-field distribution and making it a similar size to the target near-field distribution. The sampling interval is then determined in the same way as was done for the top-hat above.

2.4.3 Initial Phase

The phase of the solution found by the GSA is dependent on the starting phase, as it is a *local* optimisation method. The choice of starting phase for the GSA is not obvious. Here, multiple trials of the GSA are used, each with a quasi-random initial phase distribution. When the GSA is initialized with an asymmetric initial phase distribution, a solution with an asymmetric phase is generally returned; whereas, if a symmetric initial phase distribution is used it should return a solution with a symmetric phase. As symmetric near- and far-field target profiles are sought here, symmetric initial phase distributions are used in order to achieve solutions which have a symmetric phase.

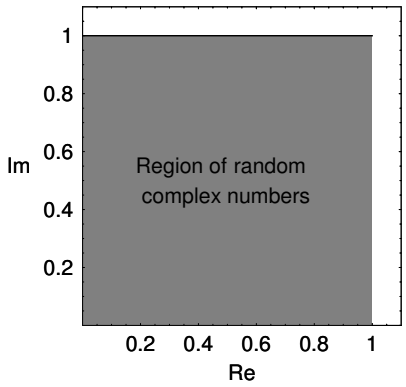


Figure 2.13: randC is a random complex number in the region shown.

Symmetric spherical phase distributions with a random radii of curvature, and the phase of multimoded Gaussian-Hermite distributions with random complex mode coefficients are used to initialise the GSA, and are generated as follows

$$\text{phase} = \begin{cases} R_\psi(x) = \sum_{n=0}^{10} \frac{1}{1+\text{randC}} \psi_{2n}(x) & \text{if} \quad 0 \leq \text{rand}_\text{j}(0,1) < 0.25 \\ R_S(x) = \frac{\text{randC}}{x^2/10^4} & \text{if} \quad 0.25 \leq \text{rand}_\text{j}(0,1) < 0.5 \\ R_\psi(x) + R_S(x) & \text{if} \quad 0.5 \leq \text{rand}_\text{j}(0,1) \leq 1, \end{cases} \tag{2.29}$$

where j is the index of the GSA trial, $\text{rand}_\text{j}(0,1)$ is a random number between 0 and

1, randC is as random complex number in the region shown in Figure 2.13, whose real and imaginary parts lie between 0 and 1, $R_S(x)$ is a random parabolic phase and $R_\psi(x)$ is given by the phase of the sum of even Gaussian-Hermite modes, $\psi_n(x)$, of index n. The initial phase is defined in this way, rather than with random complex numbers, in order to make aliasing, due to under sampling of the field when the FFT is performed, less likely - as the spherical and multimoded phase distribution have a limited spatial frequency - while giving a diverse range of initial phase distributions.

2.4.4 Aliasing Test

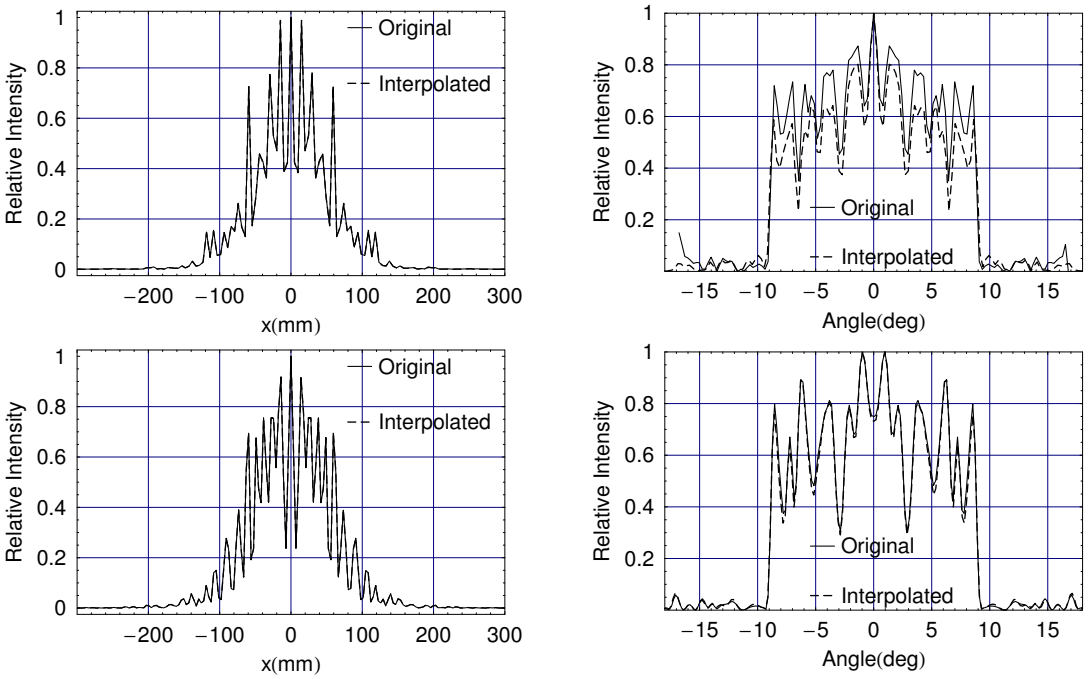


Figure 2.14: The solutions from the GSA after 100 iterations, initialised with a random phase distribution, for a $w = 100$mm input Gaussian beam and a top-hat target with a radius of 9°; the original array and the more finely sampled interpolated data are shown, *Top* the near- and far-field solutions using 128 points in the FFT, *bottom* the near- and far-field solutions using 256 points in the FFT.

Figures 2.14 shows the near- and far-field intensities corresponding to a solution for a DOE that converts an input Gaussian to a 9° top-hat after 100 iterations of the GSA, using 128 and 256 samples. The phase of the solution is interpolated - using a low order polynomial - and resampled at a quarter of the original sampling interval before performing the FFT, in order to verify that aliasing is not occurring. As shown in Figure 2.14 *top right*, where 128 sample points are used in the FFT, there is a discrepancy between far-field from using the original phase distribution and that from the interpolated phase. This occurs because aliasing is occurring in the less finely sampled field, as the Nyquist criterion has not been meet. Increasing the number of points used to 256 reduces the discrepancy, as shown in Figures 2.14

bottom right. This verification is done for all the DOEs designed here using the FFT.

2.4.5 Example 1: Top-Hat with a 2° Radius

A DOE will now be designed to transform a $\lambda = 3$mm 100mm-radius Gaussian amplitude distribution to a top-hat with a radius of 2°, as shown in Figure 2.15. The field is transformed between the near- and far-fields in the GSA using the FFT with N, the number of samples in the FFT, set to $2^8 = 256$, with a sampling interval of 7.82mm in the near-field and 0.083° in the far-field. Quasi-random initial phase distributions are used in multiple trails of the GSA for the reasons explained in 2.4.3.

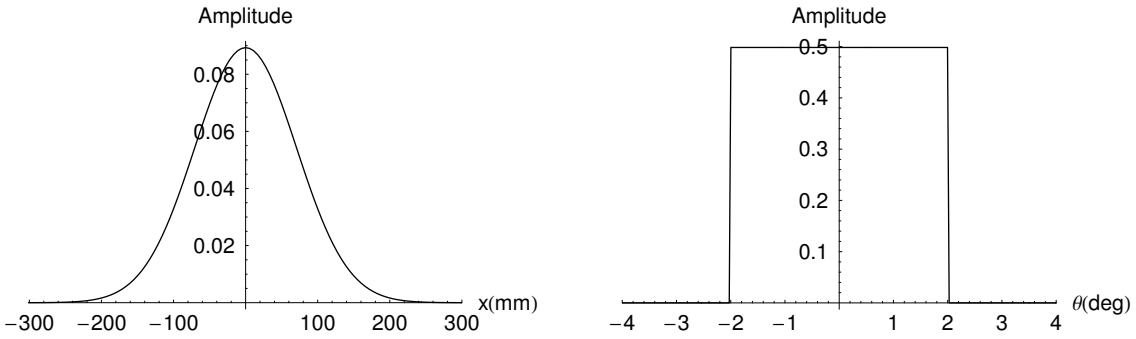


Figure 2.15: *Left* the normalised target input field: a Gaussian with a radius of 100mm. *Right* the normalised target far-field: a top-hat with a radius of 2°.

Figure 2.16 *left* shows the progress of the GSA for 100 different random starting phase distributions, generated using Equation 2.29. Initially, the MSE_f for each trial reduces rapidly with each iteration. Each trial follows a different path. Some of the trials stagnate (i.e. the decrease in MSE_f is very small for a number of iterations), before reducing rapidly and stagnating again. After 100 iterations all of the trials have one of two MSE_f's, which relate to two unique amplitude distributions (shown in Figures 2.17 and 2.18). Figure 2.18 is the more optimal solution, as it has the lower MSE; the phase of this solution is shown in Figure 2.19. The phase of Figure 2.19 *left* would be imposed on the Gaussian input field to give the amplitude distribution of Figure 2.18 *right* in the far-field.

As illustrated in Figure 2.20 *left*, if a Gaussian amplitude field is used as the input field to the DOE, the far-field given by $\mathcal{F}\{T_n(x)\exp(i\arg[E_n(x)])\}$, shown in Figure 2.18 *right*, would result (the same applies to the result in Figure 2.17). If however the near-field amplitude, $\left|E_n(x)\right|$, resulting from the last iteration of the GSA was used as the input, the target top-hat distribution, $T_f(\theta)$, would result (the far-field would given by $\mathcal{F}\{E_n(x)\} = T_f(\theta)\exp[i\phi_f(\theta)]$). The target top-hat distribution could therefore be achieved by absorbing some of the input Gaussian field to achieve the amplitude distribution $\left|E_n(x)\right|$ to give $T_f(\theta) = \mathcal{F}\{E_n(x)\}$ in the

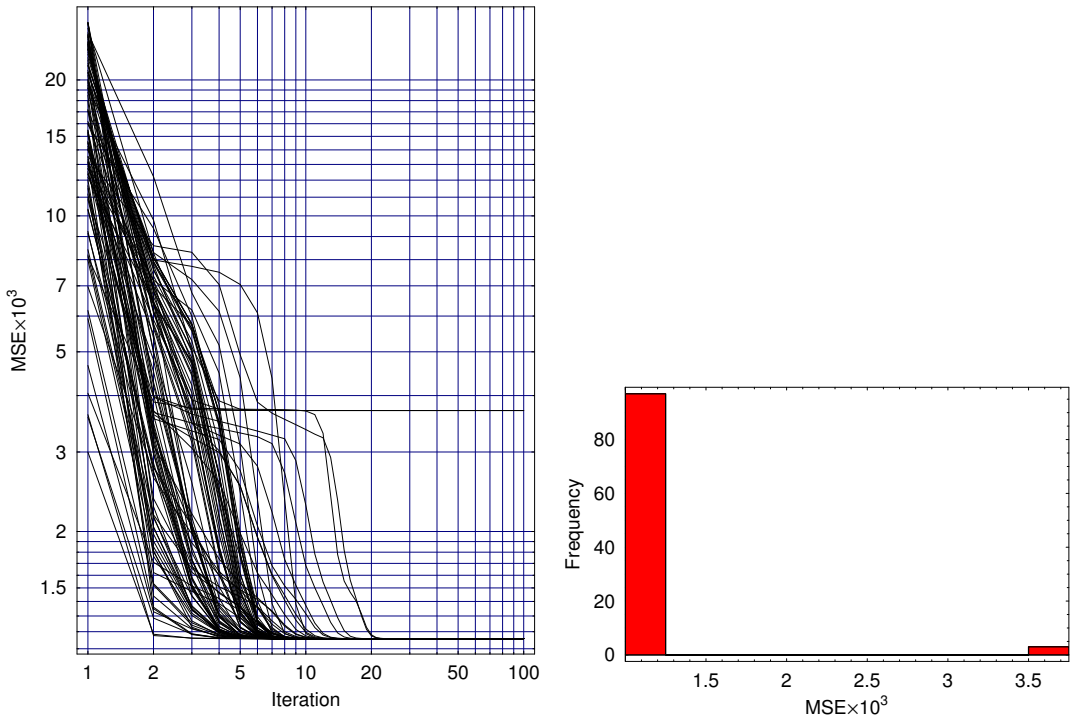


Figure 2.16: A 100mm-radius Gaussian was transformed to a 2° radius target top-hat. *Left* The progress of the GSA (the MSE_f decreases or stagnates as the GSA progress; the lower MSE_f the more optimal the solution). *Right* a histogram of the MSE_f of the solutions from the GSA.

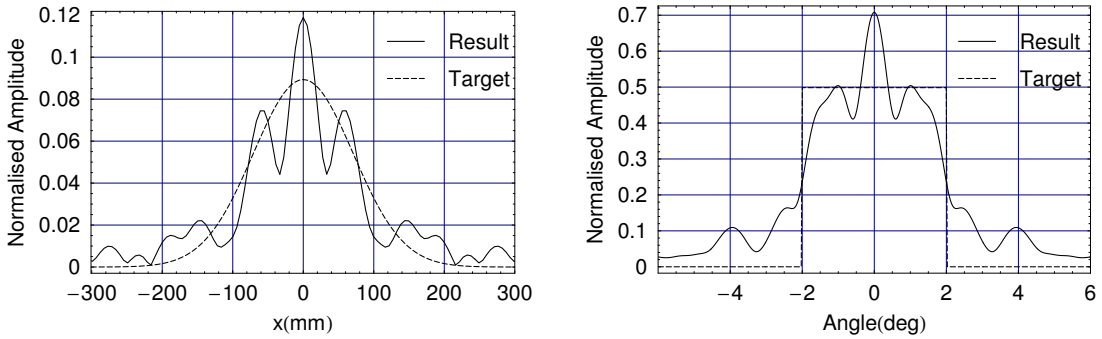


Figure 2.17: *Left* the near-field and *right* the far-field for the "worst" GSA solution of Figure 2.16, which transforms a Gaussian with 100mm-radius to a top-hat with a width of 2°.

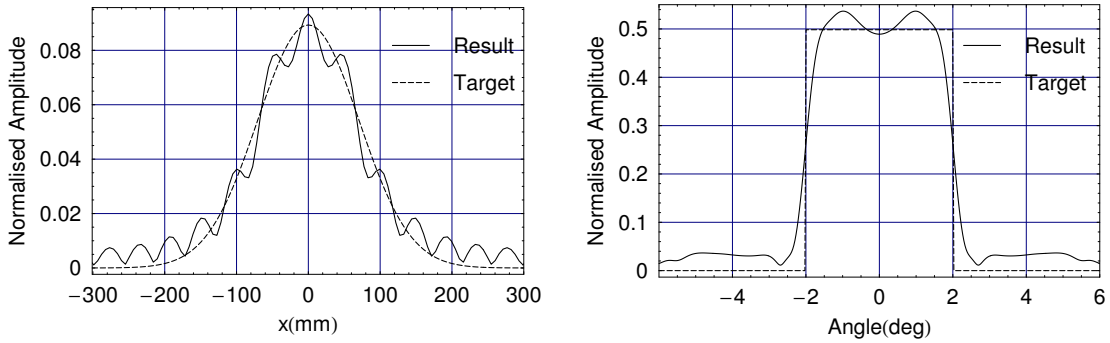


Figure 2.18: *Left* the near-field amplitude and *right* the far-field amplitude distribution, for the "best" solution of Figure 2.16 which transforms a Gaussian with 100mm-radius to a top-hat with a width of 2°.

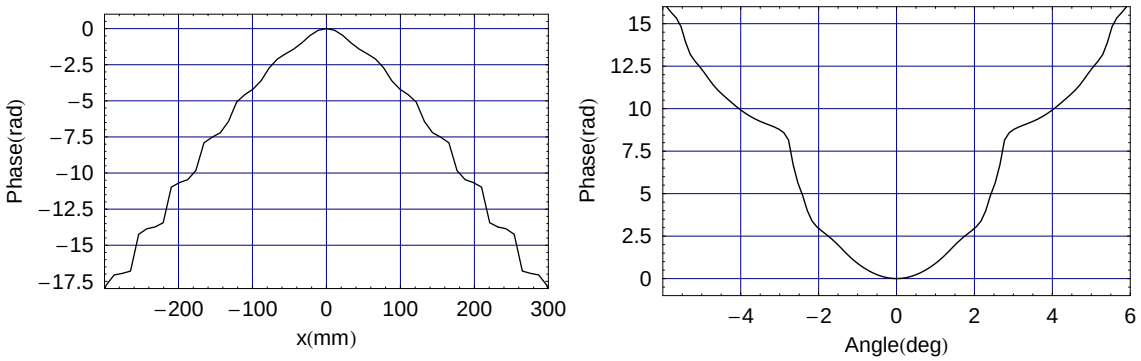


Figure 2.19: *Left* the near-field phase and *right* the far-field phase distribution, for the "best" solution of Figure 2.16 which transforms a Gaussian with 100mm-radius to a top-hat with a width of 2°.

far-field, provided that the extent of $E_n(x)$ is not greater than the DOE aperture[22]. This however necessitates the loss of some of the power of the input beam and the addition of an absorption mask to the DOE.

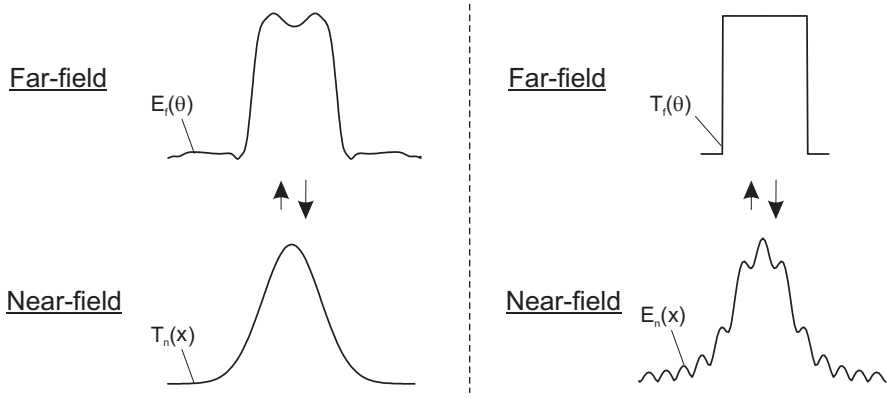


Figure 2.20: A diagram showing the far-field resulting from different input fields for the same solution for the phase given by the GSA, $\phi(x)$. As shown on the *left*, if the Gaussian $T_n(x)$ is used as the input then only the approximation $E_f(\theta)$ to the target far-field distribution, $T_f(\theta)$, is achieved. However, as shown on the *right*, if $\left|E_n(x)\right|$ is used as the input with a DOE which impose a phase, $\phi(x)$, then the target top-hat amplitude distribution, $T_f(\theta)$ is achieved.

Each of the *amplitude* distributions of the two solutions found by the GSA do not correspond to a unique *phase* (DOE) distribution. An infinite number of phase distributions can be combined with in a given amplitude distribution, as any constant phase added to a phase distribution will result in the same amplitude distribution in the far-field. After subtracting the constant phase from each of the solutions, there are just two solutions for each amplitude distribution. Each of these two solutions are mirror images of each other, with the symmetry along the x-axis. When the target and input fields are centrosymmetric, the far-field from a DOE given by the phase of a solution and the mirror of this phase are identical[20].

2.4.6 Example 2: Top-Hat with a 9° Radius

A DOE is now designed to convert a Gaussian to a top-hat with a 9° radius, increased from the 2° radius used previously. Figures 2.21 shows that there is a spectrum of solutions returned by 1000 trials of the GSA, each using 1000 iterations. Figure 2.21 *right* shows that the ordered MSEs almost form a continuum; however, some of the solutions arrive at the same local minima, as observed by the quantisation of the ordered trials. Figure 2.22 shows the progress of the best 200 trials; many of the individual trials arrive at their lowest value after more iterations than for the 2° top-hat, as the GSA was slower to converge to a solution, and the lowest overall final MSE_f value is lower than that for the 2° top-hat. Only 2.3% of the solutions have the most optimal, lowest, value of MSE_f: 0.167; and so a lot of computational effort is required to obtain this solution. Figure 2.23 and 2.24 show the near- and far-field amplitude and phase distributions for the best solution (i.e. the solution with the lowest MSE_f) of the 1000 trials of the GSA.

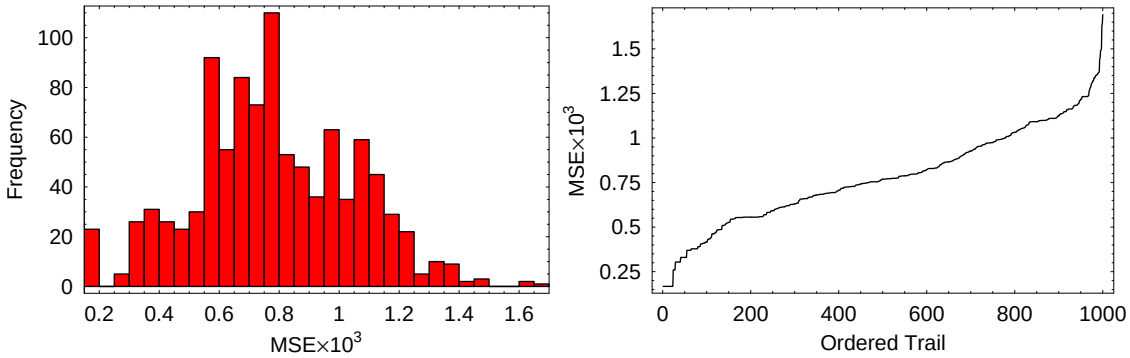


Figure 2.21: The results of the GSA for a DOE that converts a 100mm-radius Gaussian amplitude distribution to a 9° top-hat. *Left* a histogram MSE_f for 1000 trials. *Right* a plot of MSE_f of the final value of the trials, ordered from lowest to highest.

The radius of the top-hat, therefore, influences the solutions returned by the GSA. Figure 2.25 *left* shows a plot of the value of MSE_f after 100 iterations with the radius of the target top-hat varied from 1° to 20°, using 10 trials, with a quasi-random phase distribution, for each value of radius and a 25mm-radius input Gaussian. The range of MSE_fs increases and the lowest MSE_f decreases as the radius of the target top-hat increases. The nonlinearity of the problem increases with increasing top-hat radius, as evidenced by the multiple local minima found. (If the GSA was a global optimisation algorithm, we would expect the final MSE_f to be the same for all trials after a large number of iterations; however, a range of values of MSE_f are found, as the GSA is a *local* optimisation algorithm. A high number of final values of MSE_f indicates that the optimisation problem is highly non-linear, as the GSA finds the minimum of many different local regions of parameter space.) However, a solution with a lower MSE_f can be attained for top-hat fields with larger

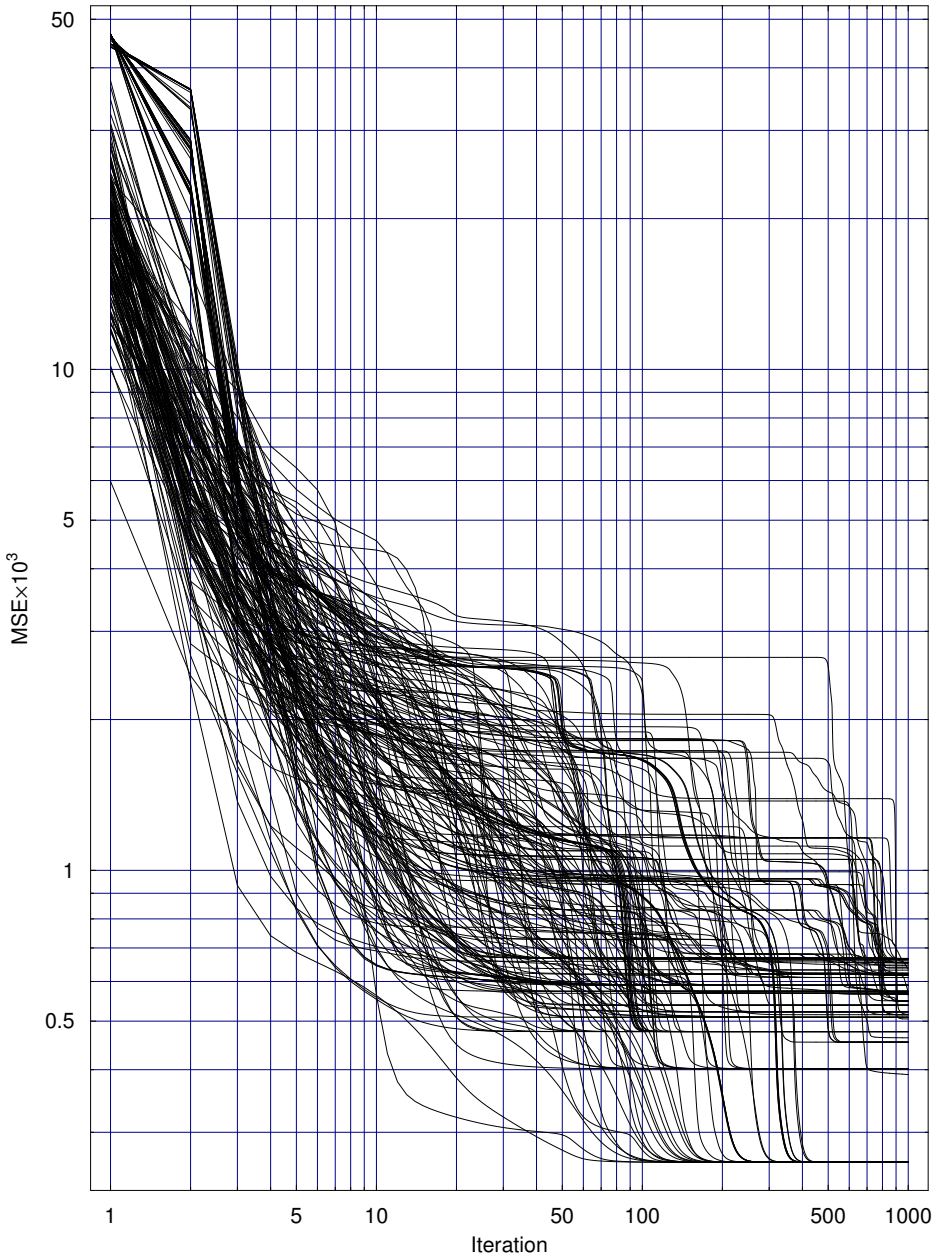


Figure 2.22: A plot of the best 200 MSE_fs as the GSA progresses, for a DOE that converts a 100mm-radius Gaussian to a 9° top-hat.

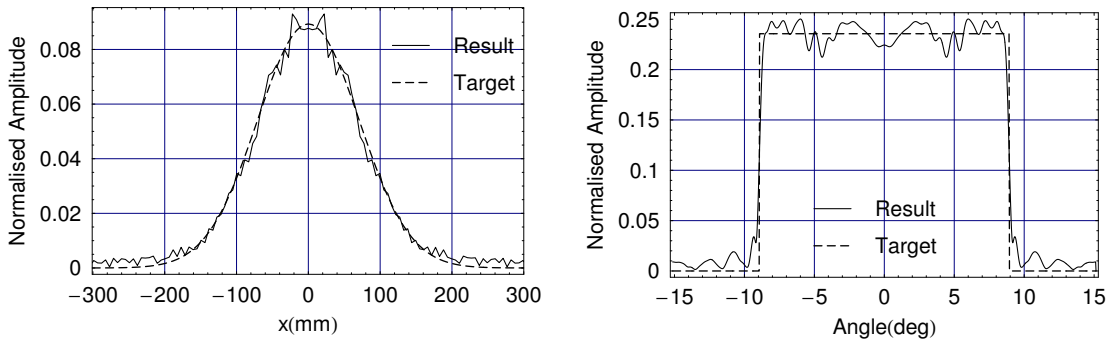


Figure 2.23: The amplitude of the solution with the lowest MSE_f for a Gaussian to 9° radius top-hat beam converter. *Left* near-field amplitude, *right* the far-field amplitude.

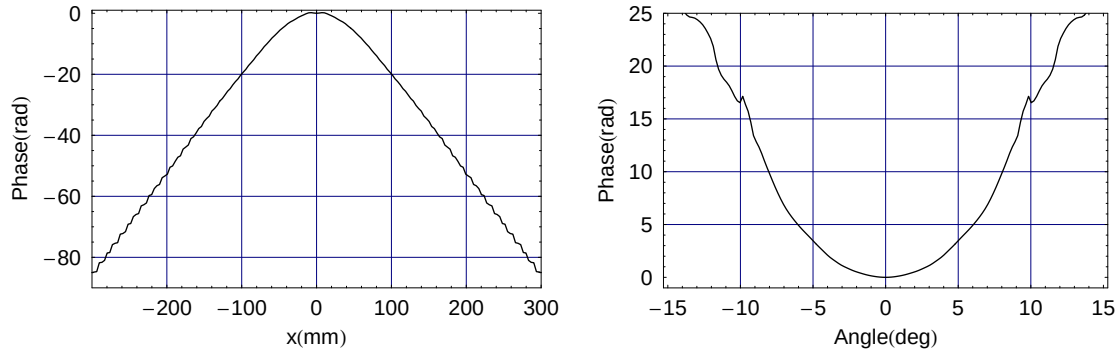


Figure 2.24: The phase of the solution with the lowest MSE_f for a Gaussian to 9° radius top-hat beam converter. *left* near-field phase, *right* the far-field phase.

radii. The computation time required to find a global optimum increases with increasing target top-hat radius. However, there is a limit to the maximum radius of top-hat that can be attained with this method, because when the radius of the top-hat increases above around 10°, the paraxial assumption of the FFT is no longer valid.

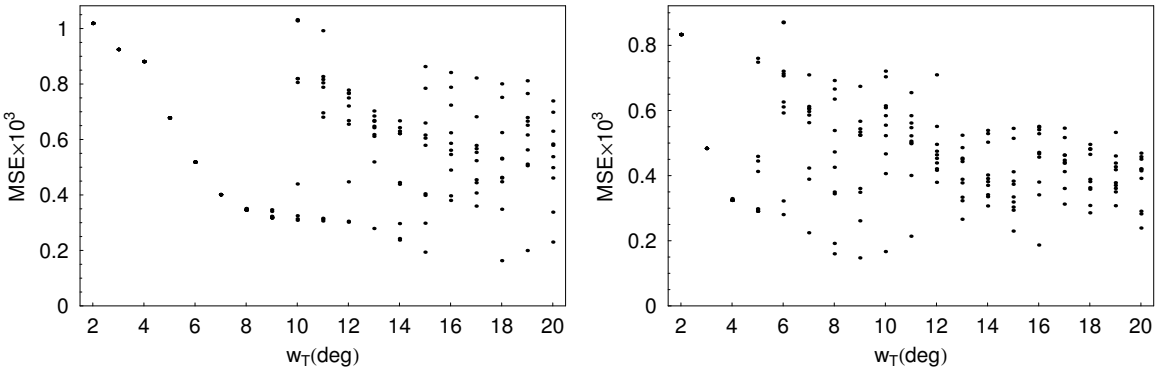


Figure 2.25: Plots of MSE_f after 100 iterations of the GSA as a function of the target top-hat beam radius, using 10 trials with quasi-random starting phases for each increment of beam width. *Left* with a 25mm-radius input Gaussian beam, *right* a 50mm-radius Gaussian beam.

Like Figure 2.25 *left*, Figure 2.25 *right*, shows a plot of the value of MSE_f after 100 iterations with the radius of the target top-hat varied from 1° to 20°, using 10 trials, with a quasi-random phase distribution, except in this case radius of the input Gaussian has been increased from 25mm to 50mm. The far-field pattern for the 50mm input Gaussians are identical to the far-field pattern with a 25mm input Gaussian with twice the target top-hat radius. For any transform

$$r \propto tan^{-1}\left(\frac{\lambda}{\pi w_0}\right), \tag{2.30}$$

where r is a measurement of the extent of the far-field target distribution and w_0 is the input Gaussian radius. Therefore, rather than increase the radius of the target top-hat distribution in order to achieve a solution with a lower MSE, the radius of

the input Gaussian may be increased.

2.4.7 Example 3: Super-Gaussian

A super-Gaussian amplitude distribution is more easily attained than a top-hat[18, 25]. The super-Gaussian is defined by

$$\exp\left[-\left(\frac{x}{\bar{w}}\right)^n\right], \tag{2.31}$$

where n is the order, and $\bar{w}$ is the radius of the super-Gaussian - equal to the radius of the top-hat that it is approximating. Figure 2.26 *right* shows super-Gaussians of a range of orders. As the order is increased it becomes closer to a top-hat, and in the limit, $n \to \infty$, it is a top-hat. Figure 2.26 *left* shows plots of the value $\mathrm{MSE_f}$ for each of ten trials at the final iteration of the GSA, for a DOE that converts a Gaussian to a super-Gaussian, over a range of orders. $\mathrm{MSE_f}$ tends to increase with increasing order. Figure 2.27 shows a plot of the range of values of $\mathrm{MSE_f}$s returned by the GSA for a DOE that converts a Gaussian to a 20th order super-Gaussian with an equivalent top-hat radius of 9°. Figure 2.28 shows a plot of the progress of the GSA trials. Figure 2.29 shows the near- and far-field amplitude distributions corresponding to the lowest $\mathrm{MSE_f}$ solution, and Figure 2.30 shows the phase of this solution. The corresponding two-dimensional amplitude distribution is shown in Figure 2.31.

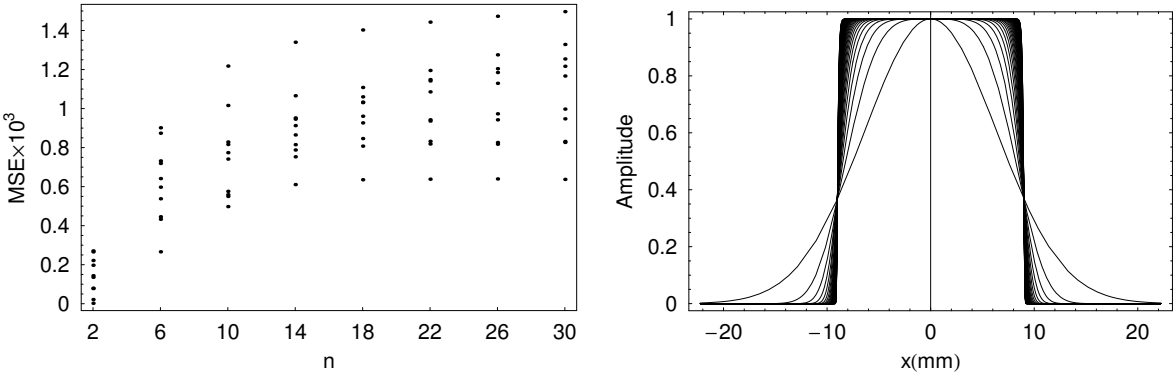


Figure 2.26: *Left* A plot of $\mathrm{MSE_f}$ for a DOE that converts a Gaussian to a super-Gaussian: $\exp\left[-\left(x/\bar{w}\right)^n\right]$, of order n, where $\bar{w}$, the equivalent top-hat radius, is 9° (there are ten trials for each value of n). *Right* a super-Gaussian with orders 2 to 100 in steps of 2.

A Gaussian to super-Gaussian transform is more accurate than a Gaussian to top-hat transform with an equivalent radius; solutions with lower MSE are returned for the Gaussian to super-Gaussian transform. As the corners of the super-Gaussian become less sharp (lower order), the MSE becomes lower. It is more difficult to achieve a sharp amplitude transition, and it is done at the cost of a less-uniform

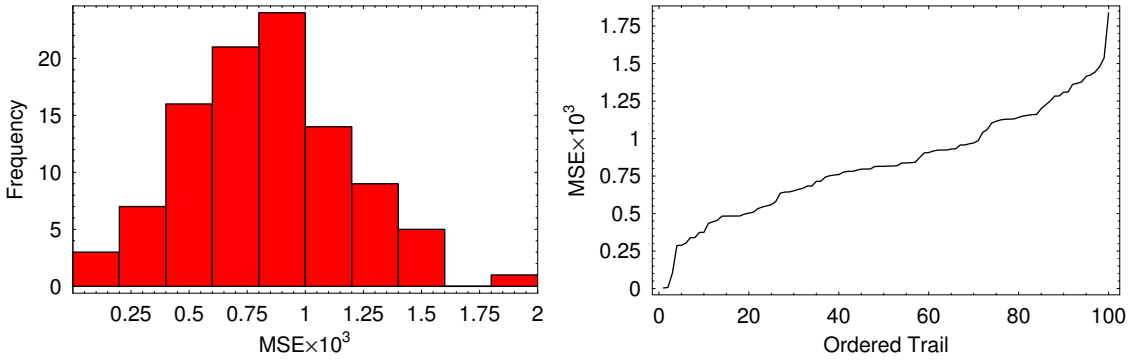


Figure 2.27: *Left* A histogram of MSE_f and *right* MSE_f ordered from smallest to largest. The MSE_f values are for a Gaussian that is converted to a 20th order super-Gaussian equivalent to a 9° top-hat.

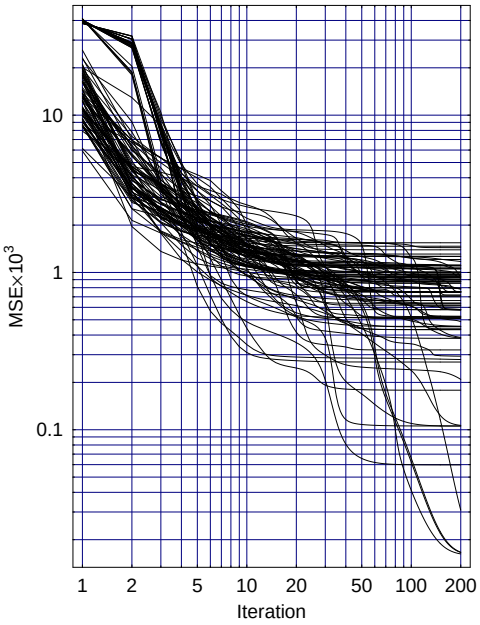


Figure 2.28: The progress of 100 trials of the GSA, for a DOE that converts a Gaussian to a super-Gaussian of order 20.

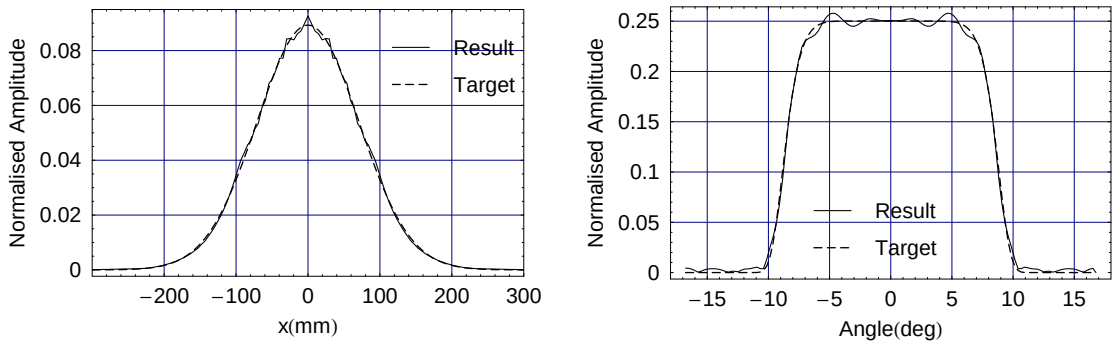


Figure 2.29: The amplitude distributions in the near- and far-fields corresponding to the solution with the lowest MSE_f for a Gaussian to 20th order super-Gaussian transform.

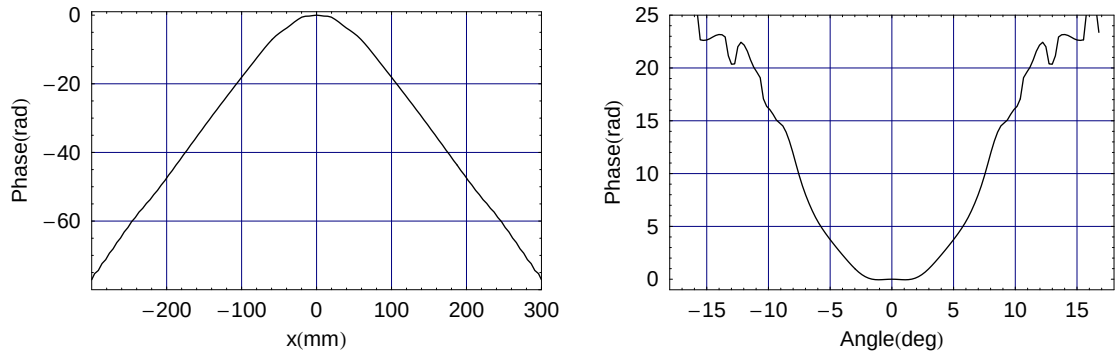


Figure 2.30: The phase of the fields in the near- and far-fields corresponding to the solution with the lowest $\mathrm{MSE_f}$ for a Gaussian to 20th order super-Gaussian transform.

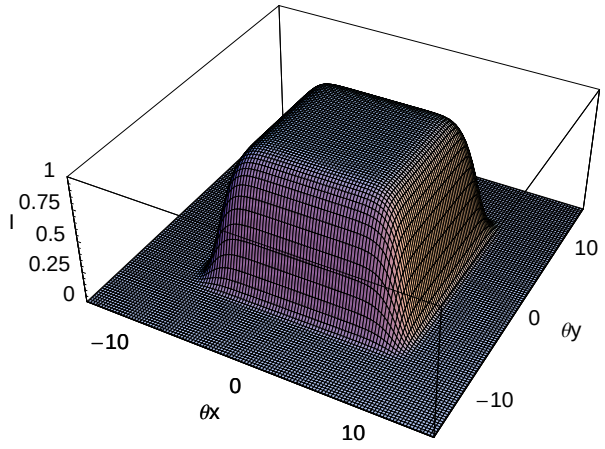


Figure 2.31: The corresponding two-dimensional far-field relative intensity distribution of the one-dimensional distribution shown in Figure 2.29.

amplitude distribution. A super-Gaussian is a suitable target profile if a uniform far-field intensity profile is required; however, the resulting far-field distribution from a DOE optimised with a *top-hat* target profile is closer to a top-hat distribution of equivalent radius than if a *super-Gaussian* were used as the target distribution instead. The choice of whether to use a super-Gaussian or a top-hat amplitude distribution as the target function, therefore, depends on the application.

2.4.8 Example 4: First-Order Gaussian-Hermite

A DOE which converts a 100mm-radius input Gaussian amplitude distribution to the amplitude distribution of a first-order Gaussian-Hermite in the far-field is now considered. As shown in Figure 2.32, a first-order Gaussian-Hermite mode amplitude distribution has both positive and negative values, but may equivalently be given in terms of absolute values and a 180° phase shift. Here, the phase is a free parameter to be optimised, and therefore the *absolute value* of the first-order Gaussian-Hermite distribution is taken as the target, T_f.

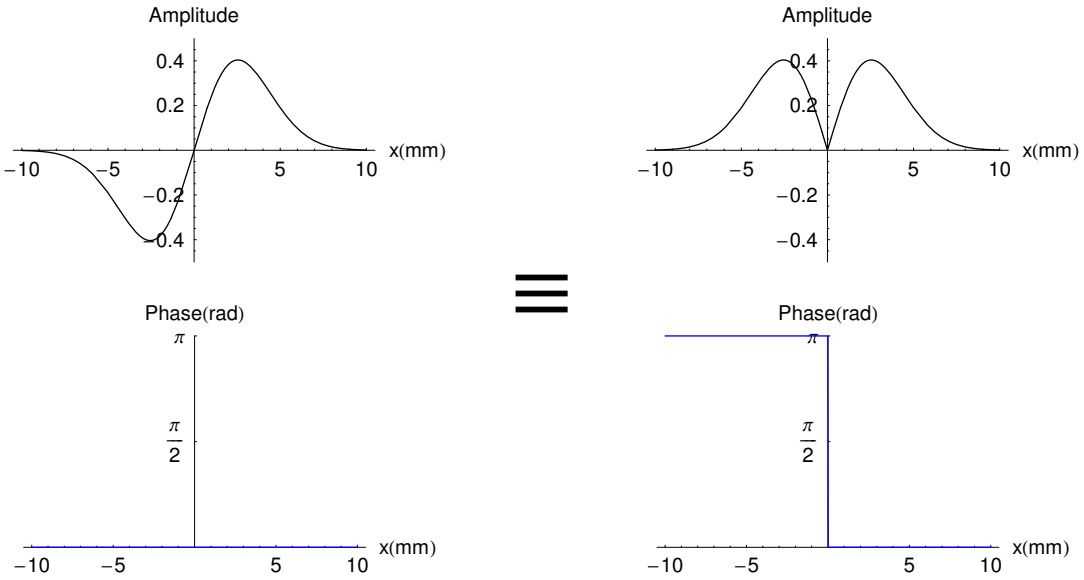


Figure 2.32: The first-order Gaussian-Hermite, shown on the *left*, does not have a symmetric amplitude distribution. It can be forced to have a symmetric distribution if the amplitude is restricted to positive values and the phase is considered, as shown on the right.

Figure 2.33 shows the $\mathrm{MSE_f}$ of the solution returned by the GSA for a target 1^{st} order Gaussian-Hermite amplitude distribution of radius w. A minimum occurs at approximately 2°; the minimum $\mathrm{MSE_f}$ increases with radius and then decreases for larger radii. However, it is not possible to determine exactly at what radius the minimum MSE occurs, given the sampling and number of trials used in the GSA. Although the minimum $\mathrm{MSE_f}$ tends to be lower at larger target radii, at larger radii

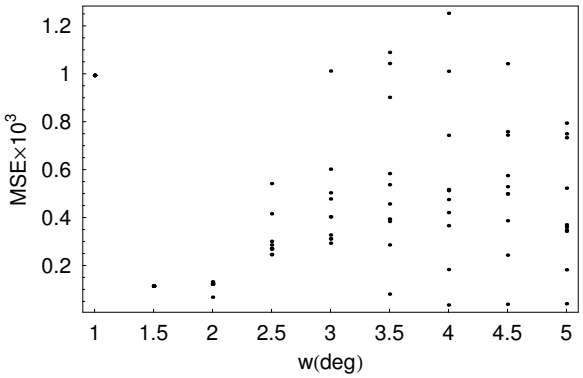


Figure 2.33: A DOE is designed to convert a 100mm Gaussian to a 1st order Gaussian-Hermite amplitude distribution. Final MSE_fs are plotted against the target radius, w in degrees, to find the one that will achieve the lowest MSE_f. 10 trials, with different random initial phases, are used for each step in radius.

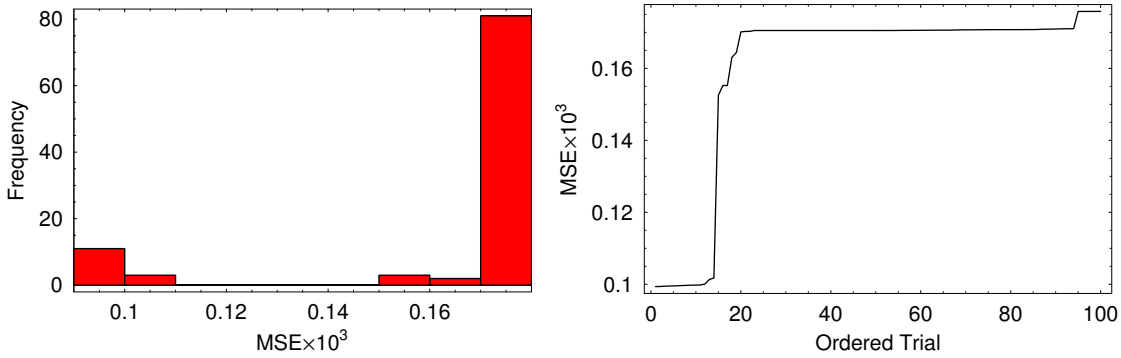


Figure 2.34: The MSE_f of the solution given by the GSA with 100 iterations for a DOE that converts a 100mm input Gaussian to a 2° radius 1st order Gaussian-Hermite amplitude distribution in the far-field. *Left* A histogram of MSE_f values and *right* MSE_f values ordered from smallest to largest MSE.

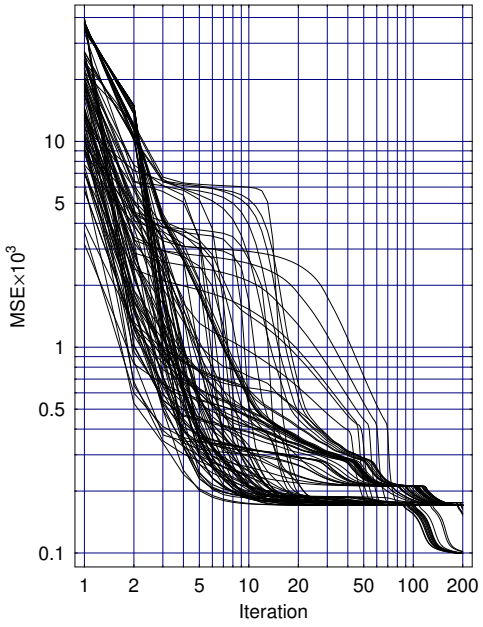


Figure 2.35: The progress of the GSA of 100 trials and 200 iterations for a DOE that converts a Gaussian to a first-order Gaussian-Hermite.

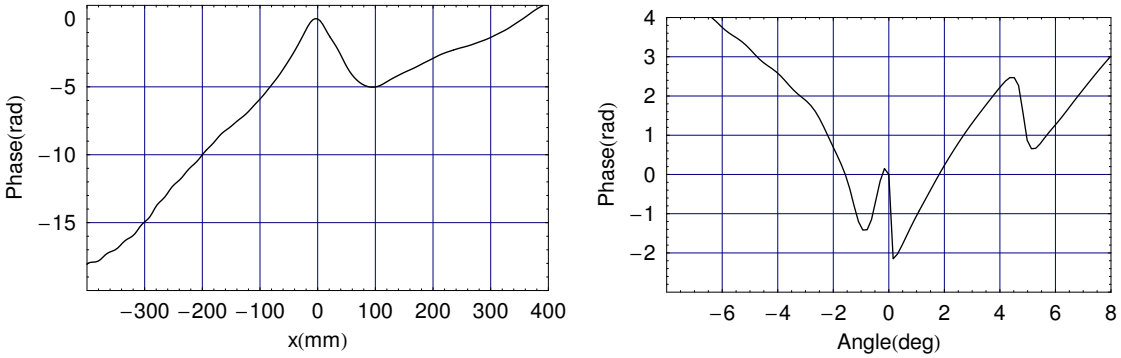


Figure 2.36: The GSA phase of the solution with the lowest $\mathrm{MSE_f}$ of 100 trials of the GSA, each with 200 iterations, for a DOE that converts a Gaussian to a first-order Gaussian-Hermite. *Left* near-field phase, *Right* far-field phase.

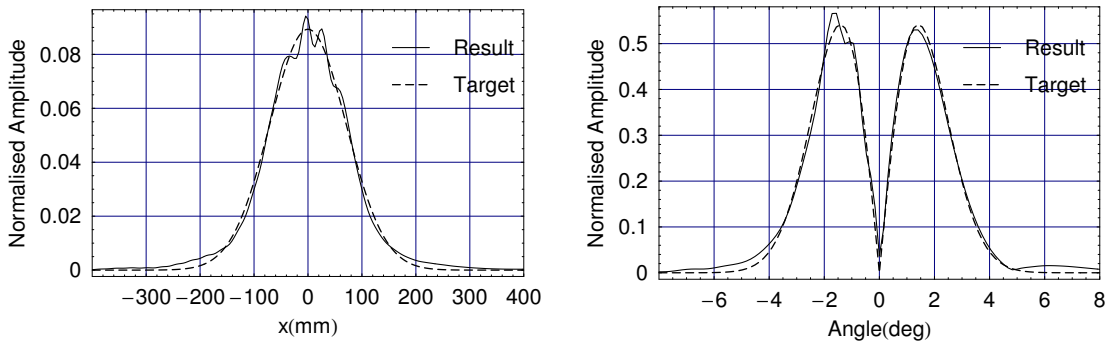


Figure 2.37: The GSA amplitude of the solution with the lowest $\mathrm{MSE_f}$ of 100 trials after 200 iterations for a DOE that converts a Gaussian to a first-order Gaussian-Hermite. *Left* near-field amplitude. *Right* far-field amplitude.

it is close to the non-paraxial regime of the FFT, and therefore the first minimum at 2° is chosen as the target for which a DOE is designed.

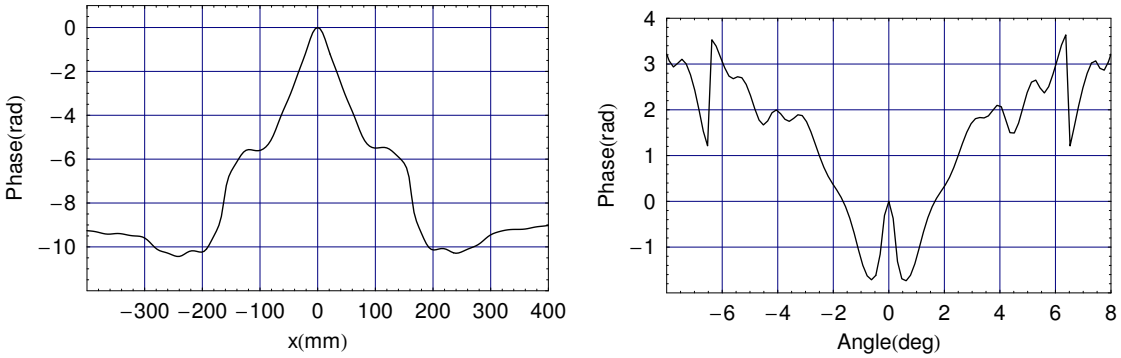


Figure 2.38: The GSA solution with the 20th lowest MSE_f of 100 trials after 200 iterations for a DOE that converts a Gaussian to a first-order Gaussian-Hermite. *Left* the near-field phase, *right* the far-field phase. (This solution corresponds to the symmetric solution with the lowest MSE_f.)

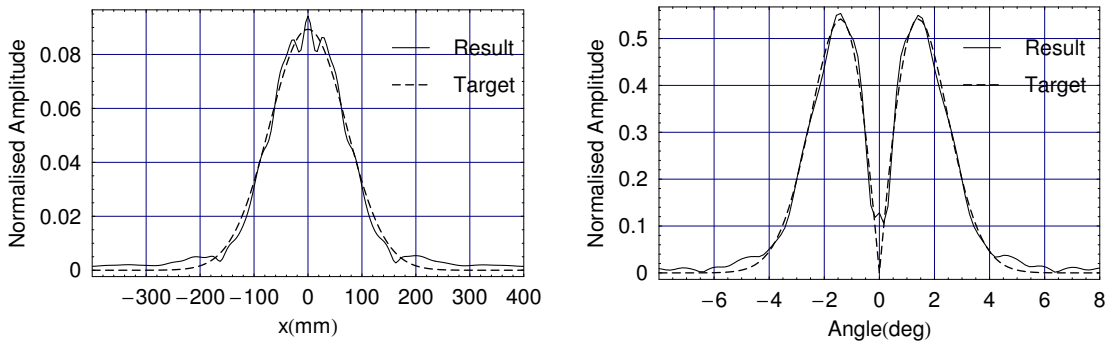


Figure 2.39: The GSA solution with the 20th lowest MSE_f of 100 trials and 200 iterations for a DOE that converts a Gaussian to a first-order Gaussian-Hermite. *Left* Near-field amplitude. *right* the far-field amplitude (absolute value).

The values of MSE_f at the final iteration of 100 trials of the GSA, each with 100 iterations, is shown in Figure 2.34. The progress of each of the trials is shown in Figure 2.35. The solution with the lowest MSE_f is shown in Figures 2.36 and 2.37. Although the GSA was initialised with a symmetric phase distribution, this solution is asymmetric. The asymmetry occurs because the GSA is iterative; small numerical errors which cause a small asymmetry in the field can become amplified with subsequent iterations. Figure 2.38 and 2.39 show the amplitude and phase of the solution with the 20th lowest MSE_f returned from the 100 trials: the solution with the lowest MSE that is symmetric. For this problem the *asymmetric* solution has a lower MSE than the *symmetric* one. The amplitude at 0° is closer to the target amplitude of 0 for the asymmetric solution than the symmetric solution. As the GSA progresses, most of the solutions stagnate at a symmetric solution after between 20 and 100 iterations; the solution then becomes asymmetric and a further decrease in MSE_f is achieved with subsequent iterations, until it stagnates again.

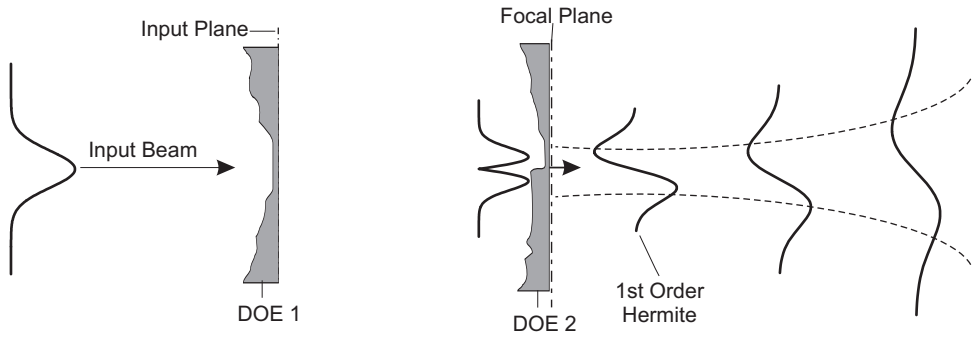


Figure 2.40: DOEs are used to create a 1st order Gaussian-Hermite mode. DOE 1 imposes a phase on the field to create the amplitude distribution of a first Gaussian-Hermite at DOE 2. DOE 2 corrects the phase to achieve a first-order Gaussian-Hermite with a waist a DOE 2.

The amplitude distribution of the solution found by the GSA, shown in Figure 2.37, is close to the amplitude distribution of the target far-field first-order Gaussian-Hermite amplitude distribution. However, although this solution has the amplitude profile of a first-order Hermite mode, it is not a true Hermite mode, as it does not have the correct phase distribution. In order to create a true Gaussian to Gaussian-Hermite mode converter the phase in the far-field must be corrected. Figure 2.40 shows how this can be achieved using two DOEs. The first DOE imposes the phase given by the GSA, shown in Figure 2.36, on the input Gaussian. The second DOE corrects the phase to achieve a 1st order Hermite mode with a flat phase-front. The distance between the DOEs can be reduced by means of a Fourier transforming lens located after the first DOE, using a converging input field or adding the phase function of the lens onto the second DOE[25]. The shape of the second DOE is given by the complex conjugate of the far-field from the first DOE plus a spherical phase to account for the difference between the phase on a plane and a surface times the $\pi/2\times$signum function (where $\pi/2\times$signum function is shown in Figure 2.32 *bottom right*).

2.4.9 Example 5: Three Gaussian Beams

In Section 2.3 diffraction gratings were employed to split an input Gaussian beam into an array of Gaussian beams in the far-field; the GSA is now employed to design a DOE that splits a Gaussian beam into three beams in the far-field. The target field is given by

$$T_f(x) = \frac{\psi_0(x/s - w/5) + \psi_0(x/s) + \psi_0(x/s + w/5)}{\sqrt{3s}} \tag{2.32}$$

where $\psi_0(x)$ is a Gaussian beam, w is the radius of the far-field beams and s is the scale factor, which describes how wide the output Gaussian beams are compared to the input; the beams are separated by $w/5$ and normalised by the $1/\sqrt{3s}$ term.

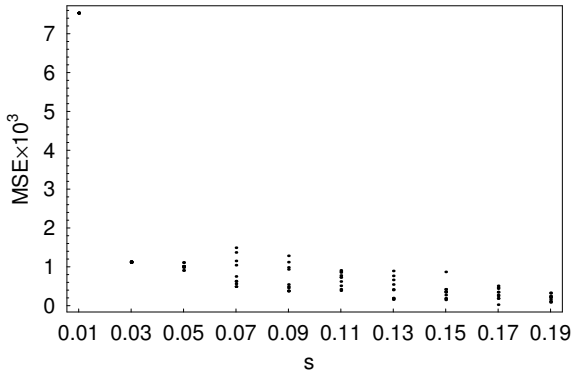


Figure 2.41: The MSE_f of the solutions returned for a DOE that converts a Gaussian beam to three Gaussian beams; MSE_f is given as a function of the scale factor, s (Equation 2.32).

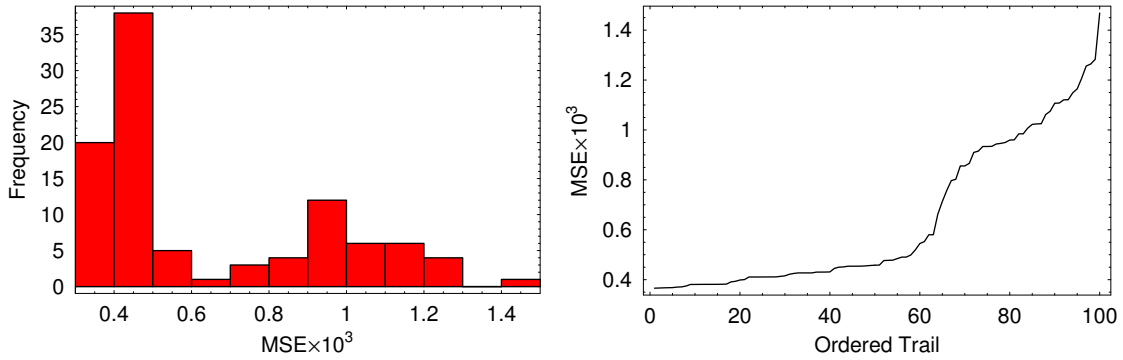


Figure 2.42: MSE_f of solutions returned for a DOE that converts a Gaussian to three equally spaced Gaussian beams, using 100 iterations for each of the 100 trials. *Left* A a histogram of MSE_f and *right* MSE_f ordered from smallest to largest.

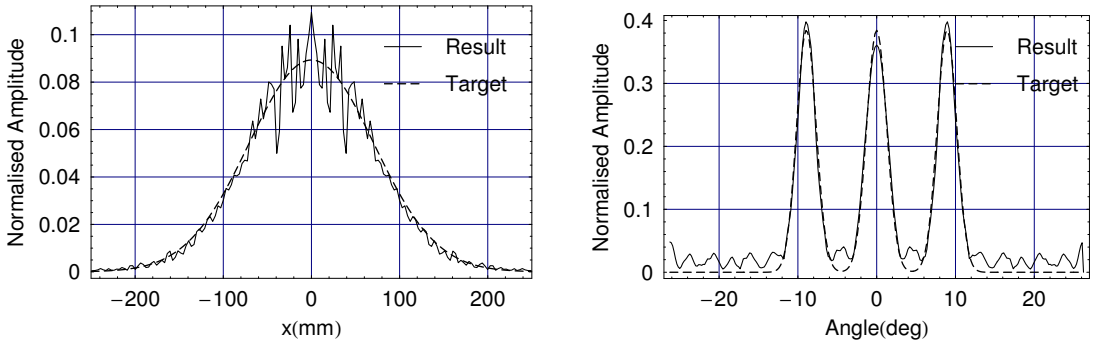


Figure 2.43: The solution with the lowest MSE_f of 100 GSA trials, using 100 iterations each, which transforms a Gaussian into three equally spaced Gaussian beams. *Left* the near-field amplitude, *right:* the far-field amplitude.

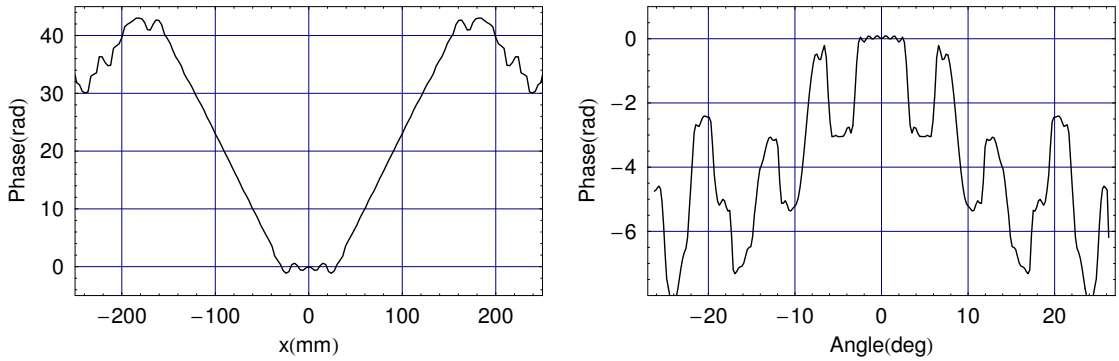


Figure 2.44: The solutions with the lowest MSE_f of 100 GSA trials, using 100 iterations, which transforms a Gaussian beam into three equally spaced Gaussian beams. *Left* the near-field phase which defines the profile of the DOE, *right* the far-field phase.

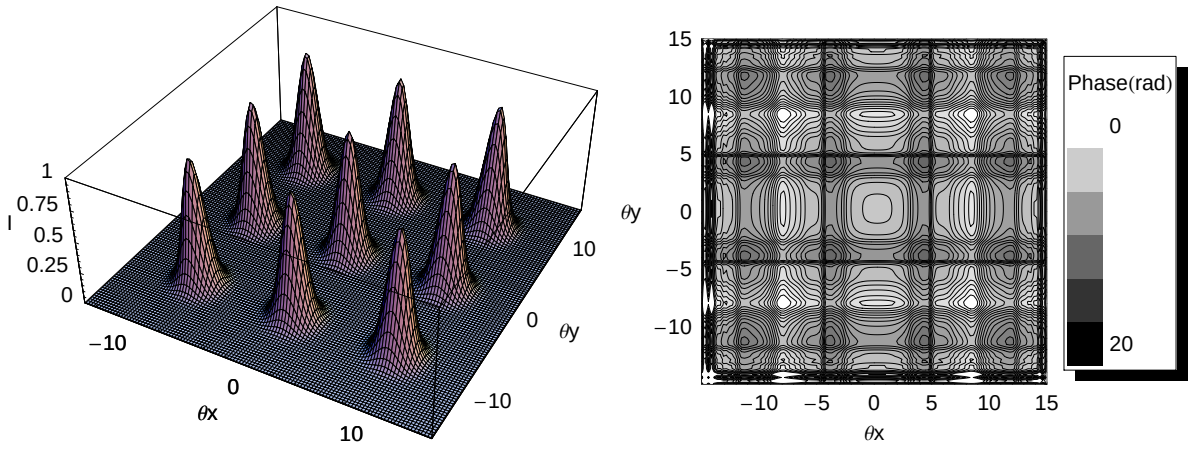


Figure 2.45: The 1D solution for three equally-spaced beams can be directly converted to give a 2D array of beams. *Left* 2D far-field intensity, *right* 2D far-field phase.

Figure 2.41 shows MSE_f as a function of the scale of the field for a Gaussian input field with a radius of 100mm. The plot shows that MSE_f tends to decrease as s increases. The scale $s = 0.09$ is chosen for the DOE because for scales greater than this the solution is non-paraxial, as the extent of the field would be significantly larger than the paraxial limit of $\approx 10°$. To achieve a solution with a lower MSE, the radius of the input Gaussian beam could be increased to reduce the size of the far-field distribution. The MSEs of the solution given by the GSA with 100 trials, each with 100 iterations, is shown in Figure 2.42. The solution with the lowest MSE_f is shown in Figures 2.43 and 2.44. This one-dimensional solution is related to the two-dimensional solution shown in Figure 2.45, by Equation 2.16. The unlimited phase freedom of the GSA allowed more optimal far-field distributions to be achieved than for the Dammann Gratings shown in Section 2.3.

2.5 Global Optimisation Methods

The unidirectional global optimisation algorithms Simulated Annealing (SA) and Differential Evolution (DE) are now used to design DOEs. Whereas bidirectional algorithms, such as the GSA, iterate between the near- and far-fields, with unidirectional optimisation the field is transformed from the near- to far-field only. They vary the parameters which encode the phase imposed on the input field, in order to find the solution which results in the best match between the target and calculated distributions, as quantified by the merit function. *Global* rather than *local* optimisation algorithms are used due to the nonlinear nature of the problem; local search algorithms, such as the quasi-Newton method and simplex method, are much more likely to find sub-optimal solutions. The GSA is a local search method and finds a local minimum from the starting phase; the final solution therefore is dependent on the initial phase distribution. Global optimisation algorithms search for a global optimum and are less dependent on the initial phase. Whereas SA operates on a single set of parameters, DE uses a population of parameters and would be expected to produce more optimal results.

2.5.1 Simulated Annealing

Simulated Annealing is a global optimisation algorithm that is analogous to the annealing process of molten substances which coalesce into a minimum energy configuration when cooled slowly[28]. The atomic energies, E, of the substance are distributed according to a Boltzmann probability distribution

$$P(E) \sim \exp\left(\frac{-E}{kT}\right) \tag{2.33}$$

where k is the Boltzmann constant and T is the equilibrium temperature. If the substance is cooled slowly - maintaining thermal equilibrium - Boltzmann's equation shows that the energy state of a molten substance is gradually lowered until finally the lowest energy state is attained. When implemented to solve a computational optimisation problem, energy is equated to the value of the merit function. The algorithm is initialised with random parameter values. At each step a trial is generated with a randomly chosen parameter value close to the current value. If the trial is downhill (i.e. the merit function has a lower, more optimal, value than the current value) the new parameter value is accepted unconditionally. If the merit function has a higher value, it is accepted (if a uniform distrubted random number is smaller than) according to the following equation

$$\Theta = \exp\left(-\frac{d\beta}{T}\right) [29], \tag{2.34}$$

where d is the difference between the uphill, current, merit function value (e.g. MSE) and the merit function value of the base point (the current lowest value); T, the "temperature", decreases with successive iterations and β is a constant. The acceptance probability Θ decreases as d increases and/or as T decreases. The optimal value of the parameter β is dependent on the problem, and is found empirically. T should be lowered at a sufficiently low rate so that the algorithm does not get trapped in a local minimum, but high enough so that the computation time is not excessively long. It is the Simulated Annealing's acceptance of parameters with merit function values which are higher than the base point that enables it to escape from local minima.

2.5.2 Differential Evolution

Evolutionary Algorithms (EAs) - which include Genetic Algorithms (GAs), evolutionary programming, Evolutionary Strategies (ESs) and genetic programming - use mechanisms inspired by biological evolution. Differential Evolution (DE), used here, is a type of ES. The distinction between GAs and ESs is that GAs perform logical operations on bit strings, whereas ESs perform arithmetic operations on floating point numbers and can be used with continuous functions.[29]

GAs have been applied to optimisation problems in many fields and have proven to be effective in solving complex high-dimensional problems, including electromagnetic and diffractive optics applications[25, 30, 31, 18, 31, 32, 33]. GAs perform the following basic steps, which are illustrated in Figure 2.46

1. **Initialize** the starting population of parameters for a function f.

2. **Encode** the solution parameters as "genes".

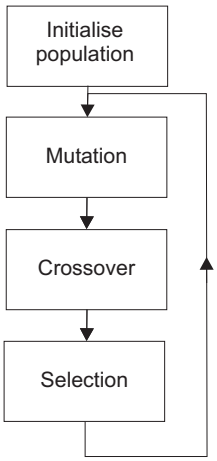


Figure 2.46: GAs evolve "fitter" individuals through mutation, crossover and selection.

3. Create a string of the genes to form a "chromosome".

4. Evaluate and assign **fitness** values to individuals in the population by evaluating f.

5. Perform **reproduction** through the fitness-weighted selection of individuals from the population.

6. Perform **recombination** and **mutation** to produce members of the next generation.

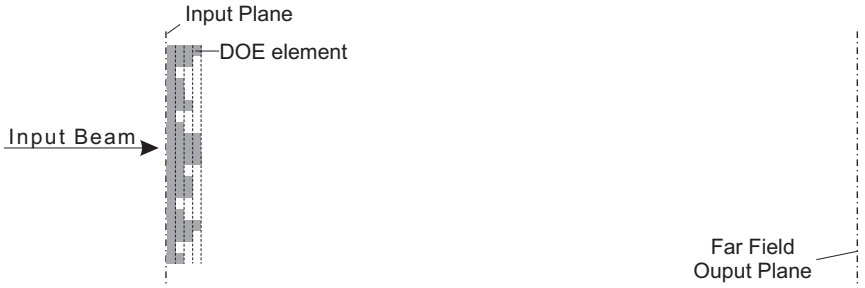


Figure 2.47: A transmission phase grating in which the DOE elements are limited to discrete values.

GAs perform well where each DOE element is limited to a discrete set of values, as shown in Figure 2.47. To work with a continuous profile, the thickness of the phase elements can be encoded as an integer string of q bits and converted to base 10 by

$$x^{'} = (< b_{15} \ldots b_0 >)_2 = \left(\sum_{k=0}^{q-1} a_k 2^k \right)_{10} \tag{2.35}$$

and then converted to a real number in the appropriate range by

$$x = b_L + \frac{b_U - b_L}{2^q - 1} x', \tag{2.36}$$

where b_U and b_L are the upper and lower limits of the range of phase depths[34]. The problem with this approach is that much computational effort may be required to change, say, the parameter value $01111_2 (x = 15_{10})$, to the optimal value of $10000_2 (x = 16_{10})$, as 5 bits must *simultaneously* be flipped to make this improving move, preventing incremental improvement. Thus $x = 15$ is one of many local minima even if the object function is unimodal[29]. GAs, therefore, are more suited to DOEs with a discrete number of levels, rather than the optimisation of the continuous gratings considered here. GAs operate on bit strings with logical operations, as opposed to the arithmetic operations of DEs. The two main benefits of DE over GAs are: firstly, the DE algorithm is less complicated, and secondly it provides greater control over the distribution of the mutant vector.

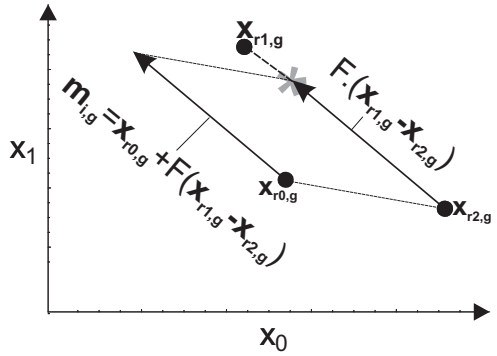


Figure 2.48: Differential mutation in which a vector operation is performed on parameters of index r from two randomly selected population members. The weighted differential, $F.(\mathbf{x}_{r1,g} - \mathbf{x}_{r2,g})$, is added to the base vector, $\mathbf{x}_{r0,g}$, to produce the mutant, $\mathbf{m}_{i,g}$.

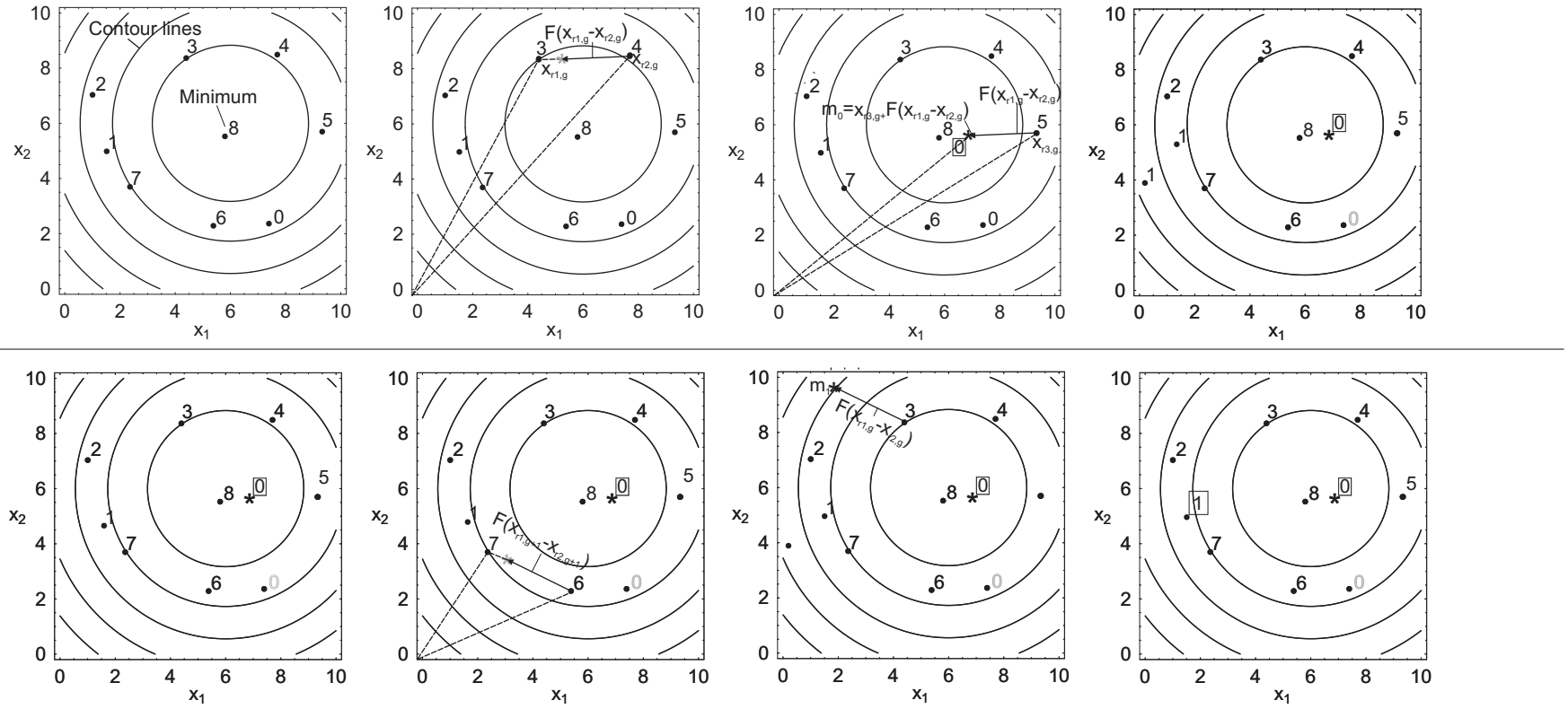
DE maintains a population of individuals (50 individuals are used in the optimisations carried out here), each of which consist of the parameters that encode the phase of the DOE as follows

$$P_{\mathbf{x},g} = (\mathbf{x}_{i,g}), \quad i = 0, 1, \ldots, N_p - 1, \quad g = 0, 1 \ldots g_{max}, \tag{2.37}$$

where $\mathbf{x}_{i,g}$ is the population member of index i, and g is the generation to which the individual and population belongs. The population is comprised of the individuals

$$\mathbf{x}_{i,g} = (x_{j,i,g}), \quad j = 0, 1, \ldots, D - 1, \tag{2.38}$$

where j is the parameter index, and D is the total number of parameters. Arithmetic operations are performed on the individuals, treating them as vectors with their origin at 0. The individuals are therefore referred to as vectors.



(a)	(b)	(c)	(d)
(e)	(f)	(g)	(h)

Figure 2.49: The mutation and selection procedure of DE. (Not shown is the crossover part of DE, which occurs with a given probability after mutation is performed.) The contour plots show the value of the merit function as a function of parameters x_1 and x_2, where a lower value is more optimal. A population of vectors with these parameters is shown, *Top* generation g and *bottom* generation $g+1$. (a) The DE population is initialized with random vectors in the region of the allowed parameter space. All the vectors are assigned an index. (b) Vectors 3 and 4 are randomly selected and labelled $\mathbf{x_{r1,g}}$ and $\mathbf{x_{r2,g}}$. The difference vector $\mathbf{x_{r1,g}} - \mathbf{x_{r2,g}}$ is calculated and scaled by F, a number between 0 and 1, to give $F \times (\mathbf{x_{r1,g}} - \mathbf{x_{r2,g}})$. (c) **Mutation:** another vector, 5, is selected at random and labelled $\mathbf{x_{r3,g}}$. This vector is added to $F \times (x_{r1,g} - x_{r1,g})$ to give m_0. (d) **Selection:** $\mathbf{m_0}$ competes against vector 0. The function of the new vector $\mathbf{m_0}$ has a lower (more optimal) value than the function of vector 0 and so replaces it in the next generation. (e), (f) & (g) A new population is mutated and selected as was done in (a) (b) & (c). (h) This time the trial vector, $\mathbf{m_1}$, loses, and so vector 1 survives to the next generation.

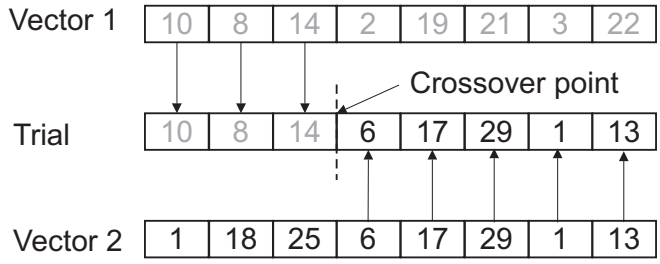


Figure 2.50: One point crossover. Each string represents a vector of parameters. Each vector contributes a contiguous series of parameter values to the trial vector. The crossover point is randomly chosen[29].

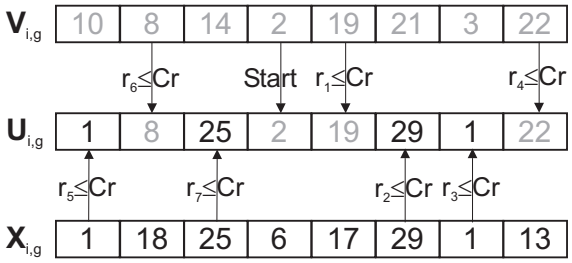


Figure 2.51: Uniform crossover: the individual parameters from two members of the population are selected with a given probability to contribute to the mutant (explained further in the text).

The first step in DE is the **initialization** of the population with random real-valued numbers in the allowed range. DE then carries out **mutation** on the population. An intermediate population of mutant vectors $P_{\mathbf{m},g}$ is created:

$$\begin{aligned} P_{\mathbf{m},g} &= (\mathbf{m}_{i,g}), \quad i = 0, 1, \ldots, N_p - 1, \quad g = 0, 1 \ldots g_{max}, \\ \mathbf{m}_{i,g} &= (v_{j,i,g}), \quad j = 0, 1, \ldots, D_{-1}. \end{aligned} \tag{2.39}$$

The new population $P_{\mathbf{m},g}$ is created as follows (illustrated in Figure 2.48). Two vectors from the current population are chosen at random and the difference between them is calculated:

$$\mathbf{d} = \mathbf{x}_{1,g} - \mathbf{x}_{2,g}. \tag{2.40}$$

This difference vector is then scaled by a factor F

$$F.\mathbf{d} = F.(\mathbf{x_{1,g}} - \mathbf{x_{2,g}}). \tag{2.41}$$

The scale factor F is generally a number between 0 and 1; a standard value of 0.6 is used here[35] as an intermediate value. A third vector from the population is added to create the mutant vector $\mathbf{m}_{i,g}$

$$\mathbf{m}_{i,g} = \mathbf{x}_{r0,g} + F.(\mathbf{x}_{1,g} - \mathbf{x}_{2,g}). \tag{2.42}$$

The mutant vector then competes with a vector from the existing population.

Figure 2.49 (a)-(h) shows two examples for a function of two parameters, x_1 and x_2, that we wish to minimise using DE. The algorithm starts with a vector population of random numbers within the viable parameter region. Perturbations are then performed on these vectors as described above, in order to explore more of the parameter space. As shown in Figure 2.49 (b) the difference vector $\mathbf{x_{r1,g}} - \mathbf{x_{r2,g}}$ is calculated from two randomly chosen vectors $\mathbf{x_{r1,g}}$ and $\mathbf{x_{r2,g}}$ and multiplied by a scale vector F, creating a new vector. Another randomly selected vector, $\mathbf{x_{r0,g}}$ is added to this to yield m_0, shown in Figure 2.49 (c). The vectors are indexed and vectors of the same index compete against each other. Vector 0 competes with the mutant $\mathbf{m_0}$ and loses because the function of the new vector is lower. $\mathbf{m_0}$ is marked so that it is labelled 0 in the next generation. Figure 2.49 (e)-(g) shows the next iteration in which another three vectors are randomly selected to create m_1. As shown in Figure 2.49 (h), this time the mutant has a higher (less optimal) function value than vector 1 of the same index, and so it is the mutant that is eliminated.

The second important part of DE, which is also implemented in GAs, is **recombination**. Recombination in EAs mimics biological recombination, in which two chromosomes are paired up and exchange some portion of their DNA. Recombination can be discrete or continuous. In continuous or arithmetic recombination new trial vectors are given by a linear combination of existing vectors. This type of recombination cannot be applied to symbolic or binary variables.

DE uses discrete recombination, also known as crossover. Figure 2.50 shows a simple type of crossover, called single point crossover, where each vector contributes a continuous series of parameter values to the new trial vector. The type of crossover implemented in DE is uniform crossover, as shown in Figure 2.51, and is given by

$$\mathbf{u}_{i,g} = U_{j,i,g} = \begin{cases} m_{j,i,g} & if(rand_j(0,1) \leq C_r \text{ or } j = j_{rand})x_{j,i,g} \\ \text{otherwise.} \end{cases} \tag{2.43}$$

The crossover probability, $C_r \in [0, 1]$ controls the fraction of parameters that are passed from each of the two randomly selected vectors to the trial vector. For two vectors A and B, with $C_r = 0.8$ on average the trial vector will inherit 20% of its parameters from vector A and 80% from B. It is equally probable, however, that B is selected first and the percentage that each vector contributes to the trial is reversed, therefore using C_r is equivalent to using $1 - C_r$. The standard value of C_r used here is 0.5[35], so each vector in the randomly selected pair of vectors each have equal probability of contributing to the trial vector.

2.6 Unidirectional Optimisation using Global Optimisation

The global optimisation methods described above are now applied to the DOE design problem.

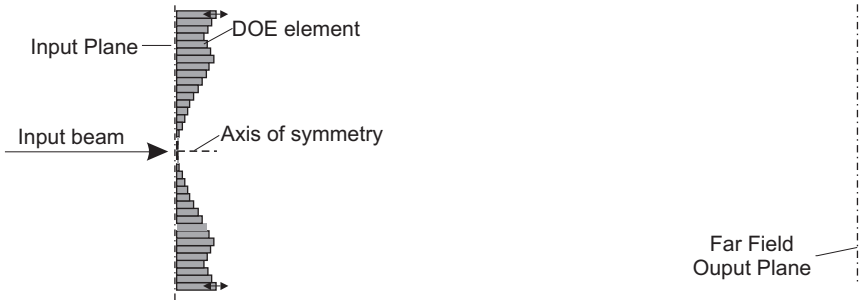


Figure 2.52: In DOE design using unidirectional optimisation, the global optimisation algorithm optimises the depth of each element of the DOE.

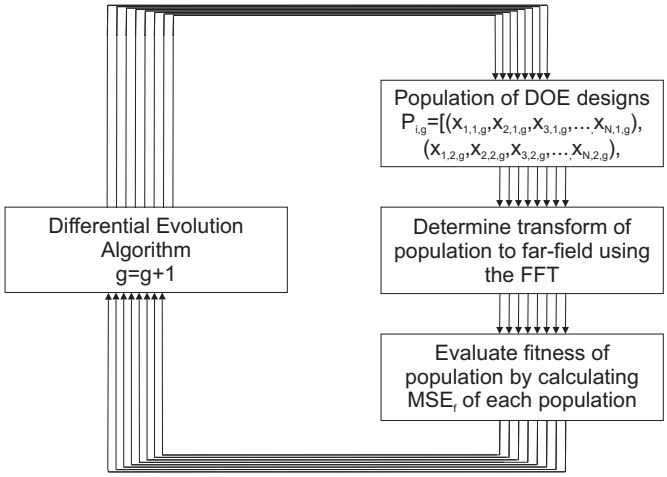


Figure 2.53: The DE algorithm designs the DOE by finding the far-field of a population of DOE designs and generating a new population based on the fitness of the current population.

The DOE is divided into a number of discrete elements, $\phi(x_1), \phi(x_2), \ldots, \phi(x_N)$, as shown in Figure 2.52. The DOE imposes a phase $\phi(x)$ on the input field with amplitude distribution $T_n(x)$. Here, a 100mm-radius Gaussian beam with a flat phase-front is used as the input field, to obtain both 2° and 9° top-hat amplitude distributions in the far-field, to enable direct comparison with the GSA results obtained previously. For simplicity, here, the optimisation is restricted to the 1D case. The far-field is given by

$$\text{far-field} = \mathcal{F}\{T_n(x)\exp(i\phi(x))\}, \tag{2.44}$$

and both DE and SA are used to design the DOE. DE optimises a random population of DOE designs, as shown in Figure 2.53. The phase angle imposed on the beam

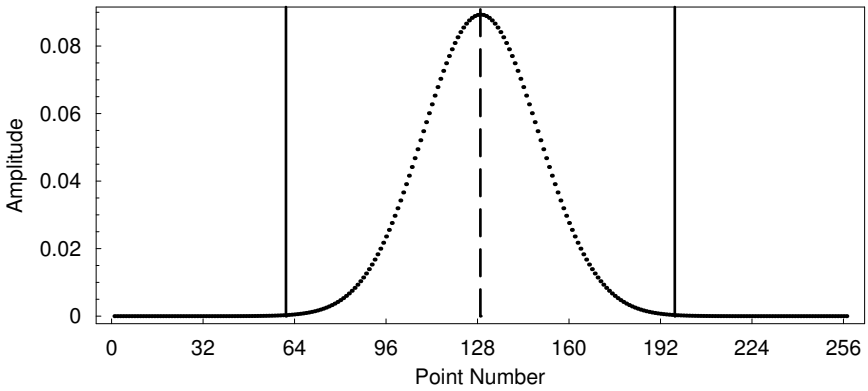


Figure 2.54: The amplitude array samples which are transformed to the far-field using the FFT. The solid vertical lines enclose the region where the amplitude is greater than the on axis amplitude times 3×10^{-2}. Only the phase within this region is encoded for optimisation by the DE. Each DE parameter controls the elements either side of the dashed line.

by each element is proportional to the depth of the element. The optimisation algorithm's parameters ($x_{j,i,g}$) in Equation 2.38 are the phases imposed by each element. As a *symmetric* DOE is sought, each parameter in the optimisation controls the thickness of two elements equidistant from the axis of symmetry (except the parameter that controls the central element). As very little power is contained in the beam far from the axis of symmetry, DOE elements in this region have very little effect on the far-field; therefore, only DOE elements contained in the region between the axis of symmetry and where the amplitude of the field drops to 3×10^{-2} times the on axis values are parametrised, as shown in Figure 2.54. The field outside of this region is padded with zeros.

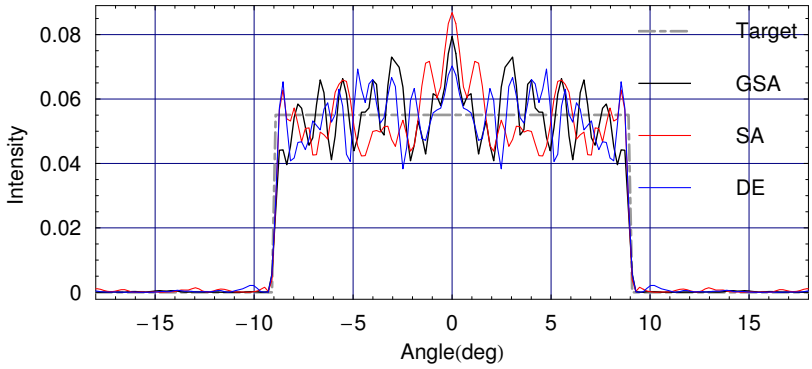


Figure 2.55: The far-field intensity distribution of a DOE that converts a $\lambda = 3$mm 100mm Gaussian field to a target 9° radius top-hat using the GSA, SA and DE.

SA and DE are used to optimise the phase angle imposed by each of the elements by minimising the MSE_f. Random initial values of phase parameters in the range $-\pi \leq \phi(x) \leq \pi$ are used to start the optimisation. Table 2.3 shows MSE_f of the optimisation for a DOE that converts a 100mm-radius Gaussian to top-hat amplitude distributions with 2° and 9° radii. The MSE_f for the GSA for the same target distributions is shown for comparison. The GSA result is the best of 100

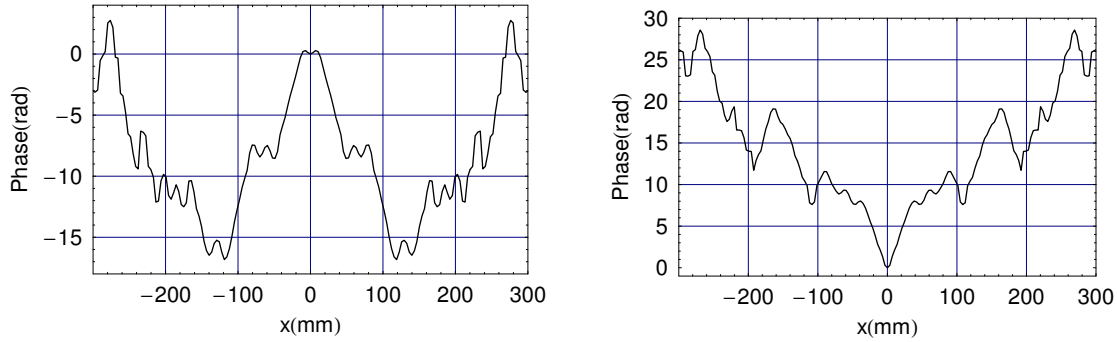


Figure 2.56: The near-field phase distribution, $\phi(x)$, as result of the optimisation to convert a 100mm-radius Gaussian to a target 9° radius top-hat using: *Left* SA and *right* DE.

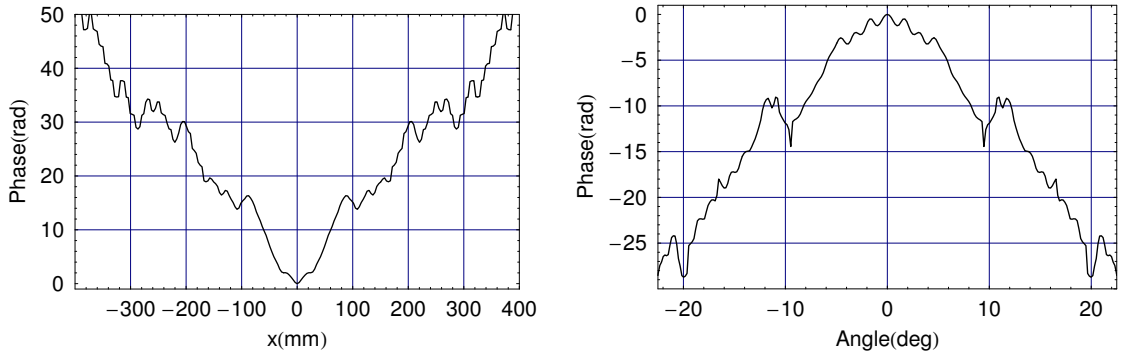


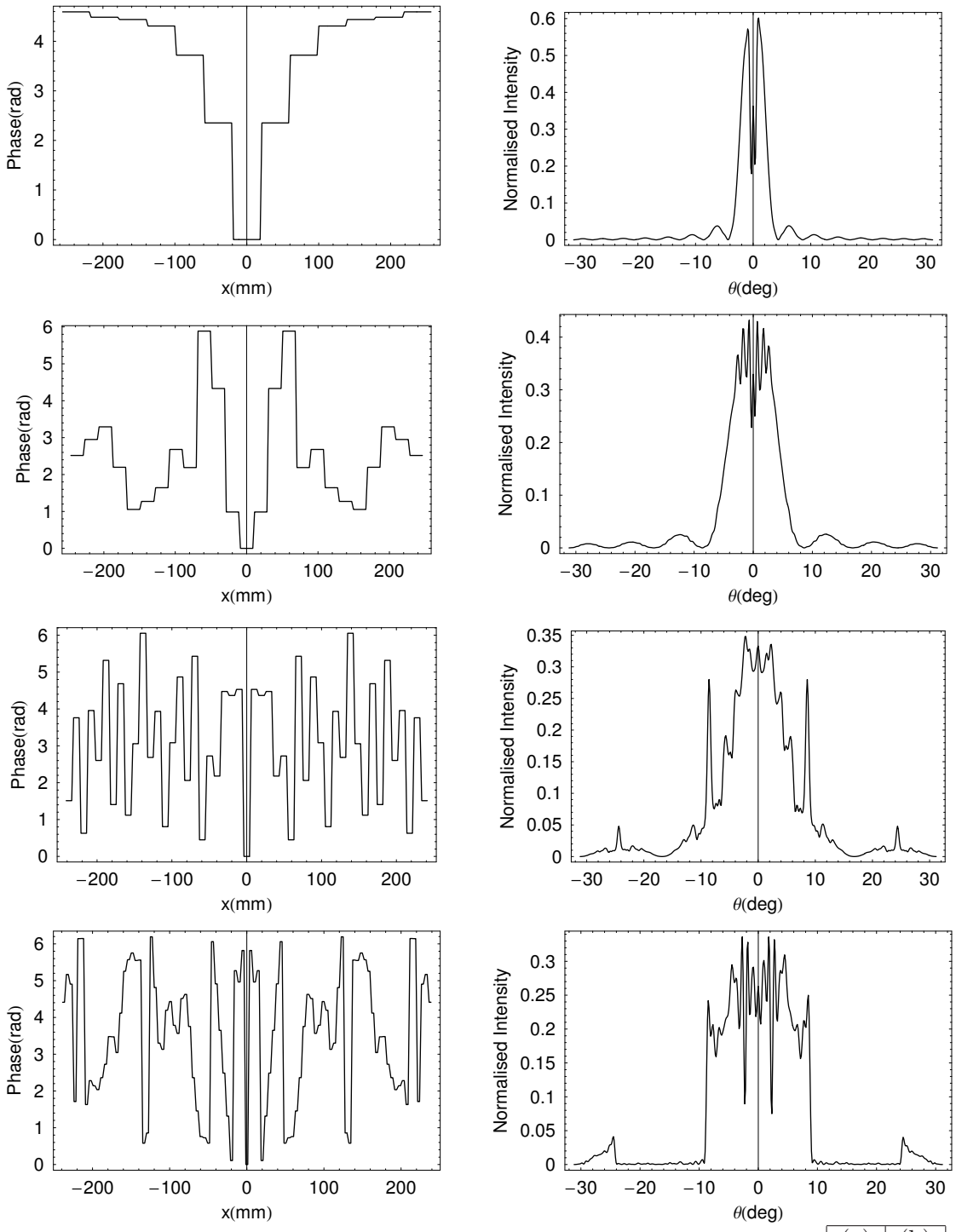
Figure 2.57: The solution for Gaussian which is transformed into a 9° top-hat field, for the best of 100 trials of the GSA, each with 100 iterations. *Left* the phase of the solution in the near-field and *right* the phase of the solution in the far-field.

Radius	GSA	DE	SA
2°	2.264×10^{-3}	2.264×10^{-3}	2.264×10^{-3}
9°	0.379×10^{-3}	0.407×10^{-3}	0.442×10^{-3}

Table 2.3: MSE_f of optimisations to transform a 100mm Gaussian beam into a top-hat amplitude distribution with a radius of 2° and 9°, using the GSA, DE and SA.

trials, each with 100 iterations and the SA and DE optimisations are limited to 1000 iterations. For the 2° radius target top-hat, the phase profile found by SA and DE are identical to the GSA solution shown in Figure 2.18. With a 9° radius target top-hat, the MSE_f of the solutions returned by the GSA, DE and SA optimisations are lower than that of the 2° target top-hat. Figure 2.55 shows the far-field distribution for the 9° target distribution, using GSA, DE and SA, where 256 points and the same sampling interval have been used in FFT for all three. The results of the optimisations are similar; the GSA gives the most optimal result, then DE, and SA gives a field which has the highest MSE_f. Figure 2.56 shows the near-field phase distributions of the SA and DE solutions, and Figure 2.57 shows the phase in the near- and far-field of the phase of the solution returned by the GSA. A smoother solution is more desirable, as smoother DOEs are more straightforward to fabricate using a milling machine. Also, the scalar approximation, used here for determining their profile, is invalid if the feature sizes of the DOEs are much less than a wavelength[22]. The Gaussian to top-hat transform was optimised using DE and SA with a range of sampling intervals. When the sampling interval was smaller than used above, the optimisation was found to suffer from aliasing; and, a less optimal far-field intensity distribution was found. Decreasing the sampling interval by increasing the size of the array used in the FFT from 256 to 512 or 1024 did not result in a more optimal solution.

The effect of feature size of the DOE on the far-field intensity pattern was investigated. There is a limit of the feature size that can be fabricated using lithographic techniques. In the optimisations described above, each discrete element, $\phi(x_1), \phi(x_2), \ldots, \phi(x_N)$, in Equation 2.44 was free to take any value in the optimisation. By simply using fewer points in the FFT to increase the feature size, aliasing occurs. As an alternative, the feature size was increased while fixing the numbers of points in the FFT to 512 points. This was achieved by giving blocks of elements the same phase; these blocks were optimised, where before each sample point in the FFT was. The far-field intensity distribution is given by Equation 2.44. This approach has the advantage that fewer parameters are optimised. Figure 2.58 shows the phase distribution found using DE with 1000 iterations and the far-field intensity distribution from the DOEs with features sizes of 39.6mm, 19.8mm, 9.9mm and 4.9mm, relating to blocks of 16, 8, 4 and 2 sample points in the FFT; the optimisation gives



(a)	(b)
(c)	(d)
(e)	(f)
(g)	(h)

Figure 2.58: *Left* the phase imposed by a DOE and *right* the far-field from the DOE, optimised using DE with 1000 iterations with a 100mm input Gaussian and a 9° target top-hat distribution, with feature sizes of (a) & (b) 39.6mm, (c) & (d) 19.8mm, (e) & (f) 9.9mm and (g) & (h) 4.9mm.

MSEs of 7.4×10^{-3}, 4.98×10^{-3}, 3.1×10^{-3} and 1.1×10^{-3} respectively. Therefore, larger blocks resulted in less optimal far-field intensity distributions.

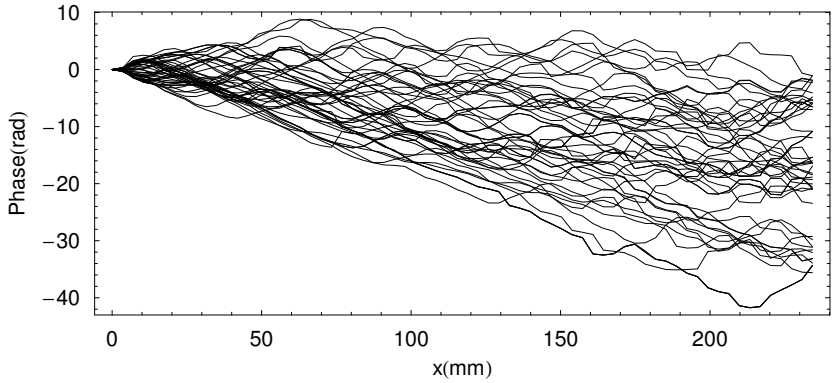


Figure 2.59: The 50 phase distributions - shown as a function of off-axis distance - with the lowest MSE_f of 100 trials of the GSAs, each with 100 iterations. They are used as initial phase distributions to initialise the DE and SA optimisations.

In Reference [22] a method is described for using a genetic algorithm to design DOEs with binary elements where, rather than using random values for the initial phase, the result of the GSA is used to initialise the genetic algorithm. In this way, the population of "fit" parameters is used to initialise the genetic algorithm which improves on the design given by the iterative procedure alone. Here, the phase distributions given by the GSA are used to initialise the algorithm. Figure 2.59 shows the 50 phase distributions with the lowest MSE of 100 trials of the GSA alone. These are used to initialise the DE and SA optimisation; however, no decrease in the MSE could be achieved from that of the GSA. Further attempts to reduce the MSE were made by initialising the algorithms with perturbations of the GSA solutions; whereby a lower number of GSA solutions where selected, multiple copies made, and perturbations of these copies made by adding random phase distributions, as shown in Figure 2.60. In separate trials the range of random values was varied and both wrapped and unwrapped phase distributions were used. However, no decrease in final MSE_f could be achieved over the best GSA solution. Only if the GSA was restricted to about 10 iterations would the optimisation return a solution with a lower MSE_f than was returned when random phase values were used to initialise the optimisation.

DE typically performed better, or gave the same result as SA for the various different target fields that were sought. This concurs with the literature comparing SA and DE[31]. Population based optimisation algorithms, such a DE, generally out perform optimisation algorithms that operated on only a single point in parameter space for each step, such as SA, because they are more effective in exploring the parameter space. DE has less of a tendency to converge prematurely than SA; although, this is at the cost of longer computation times. Therefore, DE will be used in the next section, which describes a novel algorithm beam shaping algorithm

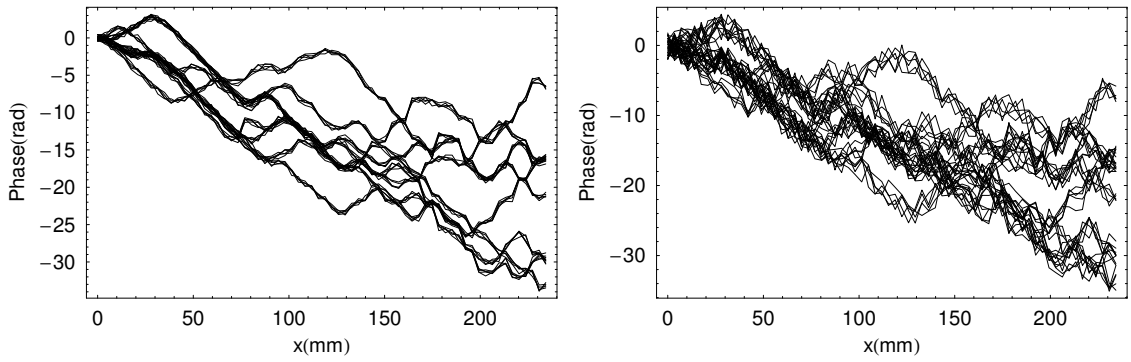


Figure 2.60: Random values are added to the phase given by the GSA trials in the range: *left* ± 0.1 and *right* $\pm\pi$.

based on GBMs.

2.7 Conclusion

In this chapter diffractive optical elements (DOEs) were designed to transform an input Gaussian beam to specific intensity distributions in the far-field. Far-field amplitude distributions corresponding to a top-hat, a first-order Gaussian-Hermite and three equi-spaced Gaussian beams were used as examples to test different optimisation techniques. The bidirectional Gerchberg-Saxton Algorithm (GSA) as well as unidirectional techniques based on simulated Annealing (SA) and differential evolution (DE) were investigated. It was possible to find good DOE solutions in all cases.

For a near- to far-field transform for which the optimisation is not too nonlinear (for example, a Gaussian to top-hat beam with a small top-hat radius), the GSA and unidirectional optimisations find the same optimum. As shown in Figure 2.25, increasing the scale of a top-hat target amplitude distribution tends to result in a more accurate approximation to it; however, increasing the scale also makes the optimisation task more onerous, because of the increased nonlinearity of the problem. This nonlinearity causes the GSA result to be very dependent on the initial phase, and it tends to find a suboptimal solution. The "brute force" method of using quasi-random initial phase distributions was used to initialise multiple trials of the GSA, in order to a obtain a more optimal solution. Whereas more optimal solutions could be found as the target scale of the top-hat was increased, a highly optimal solution occurred at *intermediate* scales for the Gaussian to Gaussian-Hermite and multiple beam splitter problems.

The advantage of unidirectional optimisation using global optimisation algorithms over the GSA is that it is more flexible, as arbitrary constraints may be placed on the target field, although, it has the disadvantage that it tends to find less optimal solutions than the GSA while using the "brute force" approach. DE gave

more optimal results than SA, but required longer computation times. Using DE with a larger population size and a larger number of iterations gives more optimal results, but increases the computation time. This approach is more suited for use with a multiple processor computer, rather than the single processor computer that the optimisation were carried out here.

Chapter 3

Beam Shaping Part II: Gaussian Beam Mode Techniques

3.1 Introduction

In Chapter 2, the GSA and unidirectional optimisation using DE were discussed. The GSA quickly converges to a solution, but is liable to get stuck in a local minimum. Unidirectional optimisation using DE is more flexible, as it allows arbitrary constraints to be placed on the solution, but is computationally intensive. Here, I describe an approach based on GBMA, which I developed with the aim of overcoming the problems associated with unidirectional optimisation and the GSA. This technique retains the advantage of DE (that arbitrary constraints or freedoms may be placed on the field) while reducing the computation time.

In Section 3.2, the technique used to design a DOE is described. The technique involves finding the mode coefficients of the Gaussian beam mode set that best describes the desired field in the near- and far-field. In Section 3.3 the relationship between the solution given by the GSA and the mode set optimisation technique are discussed. In Section 3.4 I show how grating solutions can be improved by absorbing radiation, thereby altering the amplitude as well as the phase of the input beam. It is generally not possible to obtain an exact transform of the target near-field to far-field distribution using a single DOE; in Section 3.5 I show how an exact transform can be achieved by using two DOEs. In Section 3.6, I develop the techniques to designed DOEs to control the field in the near-field region, to achieve the prescribed amplitude distributions at multiple planes.

3.2 Gaussian Beam Mode Set Optimisation

An alternative beam shaping algorithm to the GSA is now investigated. The phase retrieval problem can be recast in terms of Gaussian beam modes as: What set

of mode coefficients describes a field with the desired intensity in the near- and far-fields? In this way the near-field is given as a sum of GBMs by

$$\text{Near}(x) = \sum_{n=0}^{n_{max}} A_n e^{i\phi_n} \psi_n(w_0, x, z = 0) \tag{3.1}$$

where A_n and ϕ_n are the amplitude and phase of the n$^{\text{th}}$ GBM's mode coefficient, n_{max} is the total number of modes used and $\psi_n(w_0, x, z)$ are the Gaussian-Hermite modes in many of the examples tried here. The far-field is given by

$$\text{Far}(x) = \sum_{n}^{n_{max}} A_n e^{i\phi_n} \psi_n(w_0, x, z = \infty). \tag{3.2}$$

The power in the fields given by Equations 3.1 and 3.2 is given by

$$P = \sum_{n=0}^{n_{max}} A_n^2. \tag{3.3}$$

The normalised fields are given by $E_n(x) = \text{Near}(x)/\sqrt{P}$ and $E_f(x) = \text{Far}(x)/\sqrt{P}$. The phase that gives the DOE profile is given by

$$\phi_{DOE,n}(x) = \arg\left[E_n(x)\right]. \tag{3.4}$$

The mode coefficients are optimised to find the values that give the best match to the prescribed near- and far-field distributions. An advantage of this approach, over unidirectional optimisation using FFTs (described in Chapter 2), is that the number of parameters to optimize is greatly reduced, while still allowing a diverse range of phase distributions to be searched.

In [36] and [37], phase retrieval from intensity measurements was achieved by searching for the mode coefficients of a modelled field - which was given by the sum of Gaussian-Laguerre modes - that best describe the measured field. In [36] the mode coefficients were recovered from a simulated measured field by using the Levenberg-Marquardt algorithm to find the mode set that minimised the MSE between the intensity of the measured and modelled fields. This mode fitting method involves estimating the waist and propagation distance, z, of the modes. I successfully recovered phase from a simulated measured field using this method, with the Newton method in place of the Levenberg-Marquardt algorithm. In the case of phase retrieval from a simulated measured field, the GSA converged to a solution after several iterations, and mode coefficients were successfully recovered for all the trial fields attempted. However, the mode fitting method gave poor results compared to the GSA when applied to beam shaping problems. For phase retrieval from measured data, a set of mode coefficients which give the measured field exactly

(apart from the noise component) is known to exist. This is not the case with beam shaping, and the problem is much more non-linear. With beam shaping, the choice of w_0 and z of the mode set is not as obvious as for phase retrieval from a measured field, where w_0 and z can be estimated, for example, from the dimensions of the feed and propagation distance. Here, w_0 (and z if necessary) are optimised along with the mode coefficients, and as the Newton method and Levenberg-Marquardt algorithm can find suboptimal solutions for beam shaping, a more robust optimisation method is used.

Here, a global optimisation algorithm is used to optimise the mode coefficients. The global optimisation maximises the quality of the trial fields, $E_f(x)$ and $E_n(x)$, where the quality is defined using a merit function such as the MSE. Another merit function used here is the overlap integral of the trial and target fields. The overlap integral of two amplitude distributions, $U_a(x)$ and $U_b(x)$ in one dimension, is given by

$$\bar{\eta}(U_a, U_b) = \int_{-\infty}^{\infty} U_a(x) U_b(x) dx. \tag{3.5}$$

For two identical normalised amplitude distributions, i.e. where $U_a(x) = U_b(x)$, $\eta = 1$. Where $U_a(x) \neq U_b(x)$ then $\eta < 1$. In the near- and far-fields, η is given by

$$\eta_n = \bar{\eta}\left(\left|E_n(x)\right|, T_n(x)\right) \tag{3.6}$$

and

$$\eta_f = \bar{\eta}\left(\left|E_f(x)\right|, T_f(x)\right) \tag{3.7}$$

respectively, where $T_n(x)$ is the target normalised field amplitude incident on the DOE, and $T_f(x)$ is the target normalised far-field amplitude. The scale $T_f(x)$ can be optimised, along with the mode coefficients, in order to allow more freedom to find a better solution, as shown in Figure 3.1. DE is used to find the mode set that maximises $\eta = \eta_n + \eta_f$. Even though it is η that is optimised, the figures of merit η_n^2 and η_f^2 are quoted in the examples shown later, as they are more meaningful; they can be interpreted as the power coupling of the target to the optimised field. As described in Chapter 1, in Gaussian Beam Mode Analysis (GBMA), the radius of the GBMs as a function of position along the axis of propagation is given by

$$w(z) = \left[w_0^2 \left(1 + \left(\frac{\lambda z}{\pi w_0^2} \right)^2 \right) \right]^{\frac{1}{2}}, \tag{3.8}$$

where w_0 is the waist and z is the propagation distance from the waist position. The beam radius at $z = \infty$ is equal to ∞; however, for convenience, during the optimisation the radius of the modes is kept constant, i.e. $w(z = \infty) = w(z =$

$0) = w_0$. At $z = \infty$ the radius of curvature $R = \infty$, which is equal to the radius of curvature at $z = 0$. The difference between the mode set at the near- and far-fields is a $\pi/2$ rad phase slippage between the modes. The actual distribution can be obtained by simply scaling appropriately. As was done with the GSA in Chapter 2,

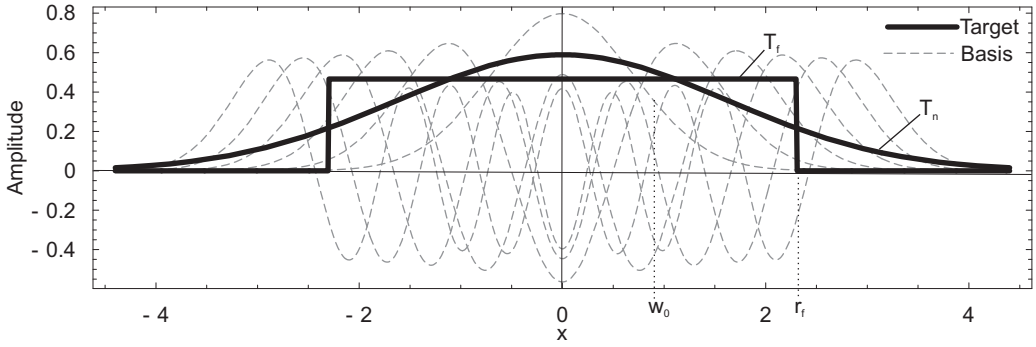


Figure 3.1: The target input field, $T_n(x) = \exp(-x^2/w^2)$, and the target output field $T_f(x) = \text{rect}(x)$, with the basis GBMs.

I use the example of a far-field with a top-hat, amplitude distribution. The top-hat has radius r_f, which is related to the radius, w, of the mode set. Figure 3.1 shows $T_n(x)$ and $T_f(x)$ with a Gaussian-Hermite basis set containing the first eight even modes. Intuitively, the mode set radius should be such that the near- and far-fields are similar in size to the target field, while the extent of the highest order mode of the basis set is not significantly larger than that of the fields, as shown in Figure 3.1. Along with the mode coefficients and r_f, w_0 is also optimised in order that the optimum radius of mode set can be found to best describe both $T_n(x)$ and $T_f(x)$. The optimisation is initialised with random values of r_f (if the radius of the top hat is not fixed) and the waist radius of the basis mode set, w_0, in a range close to that shown in Figure 3.1.

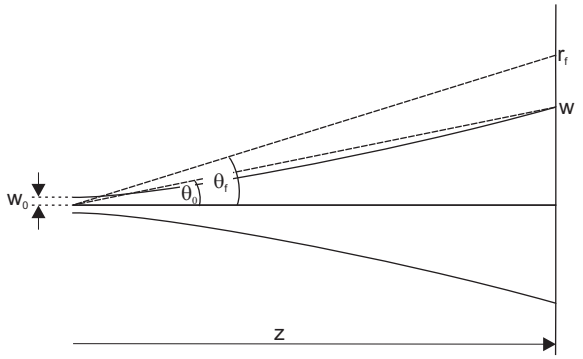


Figure 3.2: In the GBMSO where the scale of the target distribution is free, the parameter r_f, which specifies the scale of target distribution, is initially defined relative to the far-field beam radius, w, which is set to w_0, as shown in Figure 3.1. While this is convenient during the GBMSO, it is more meaningful to convert r_f to the angular coordinate, θ_f, only after the optimisation. r_f can be converted to θ_f during the optimisation if a particular value of θ_f is required.

After the optimisation the top-hat radius is converted to an angular coordinate

as follows. The radius of the top-hat, relative to the optimum mode set radius r_f/w_0 is known. Since the far-field diversion angle of a fundamental is given by

$$\theta_0 = \tan^{-1}\left(\frac{\lambda}{\pi w_0}\right), \tag{3.9}$$

the actual angular radius of the top-hat in the far-field θ_f (Figure 3.2) is given by

$$\frac{\theta_f}{\theta_0} = \frac{r_f}{w_0} \tag{3.10}$$

or

$$\theta_f = \theta_0 \frac{r_f}{w_0} = \frac{\lambda}{\pi w_0}\frac{r_f}{w_0} = \frac{r_f \lambda}{\pi w_0^2} \tag{3.11}$$

As stated previously, during the optimisation, the waist radius of the modes, w_0, and the radius of the target far-field amplitude distribution, r_f, are kept as free parameters to allow more freedom to find an optimal solution; however, for certain applications, a particular value of θ_f may be required - in this case r_f is not set to any value, as θ_f is dependent on w_0 as well as r_f. It is given by rearranging Equation 3.11 as follows:

$$r_f = \tan(\theta_f)\pi w_0^2/\lambda. \tag{3.12}$$

In order to achieve a particular value of θ_f, during the optimisation r_f is given by Equation 3.12

In summary, to design a DOE, SA or DE is used to optimise the amplitude of the mode coefficients $A_1, A_2, \ldots, A_{max}$, the phase of the mode coefficients $\phi_1, \phi_2, \ldots, \phi_{max}$, the radius of the modes, w_0, and the radius of the target far-field amplitude distribution, r_f, by maximising $\eta = \eta_n + \eta_f$. In this case the radius of the target field r_f is kept as a free parameter, and after the optimisation is converted to an angle. Keeping the scale of the far-field distribution as a free parameter allows more freedom for a more optimal solution to be found.

3.2.1 The Progress of the DE Optimisation

A DOE is designed to convert a Gaussian distribution to a top-hat field, in order to compare the performance of the MSE and η merit functions and to demonstrate the operation of the DE algorithm. The coefficients of the first 6 even Gaussian-Hermite modes are optimised. A Gaussian field with a 30mm-radius is used as the input beam and a top-hat function with a variable radius is the target output. The normalised top-hat field is given by

$$T_f(x) = \mathrm{rect}(r_f) = \begin{cases} \sqrt{1/(2r_f)} & \text{if } x < r_f \\ 0 & \text{otherwise,} \end{cases} \tag{3.13}$$

where r_f is the radius of the top-hat field.

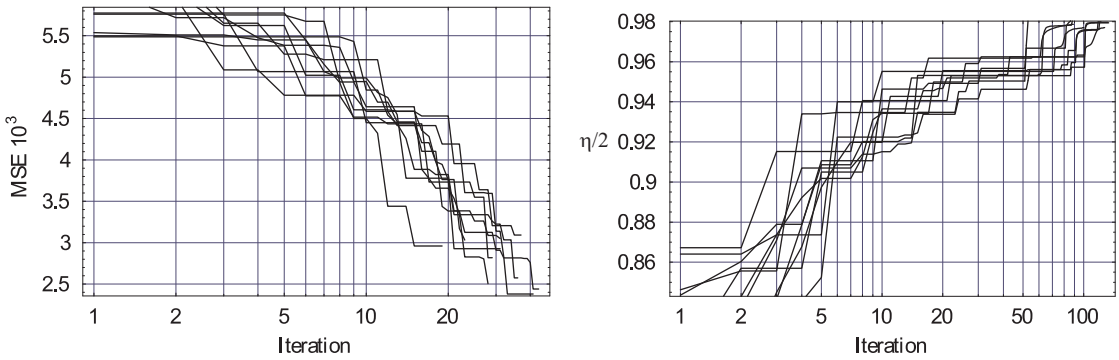


Figure 3.3: Progress plot with 10 trials *left* using MSE as a merit function without normalising the trial field (MSE is given as a function of the iteration number), and *right* using η as a merit function ($\eta/2$ is given as a function of the iteration number), for a DOE to convert a Gaussian input amplitude field to a top-hat using the 0-10 even Gaussian-Hermite modes.

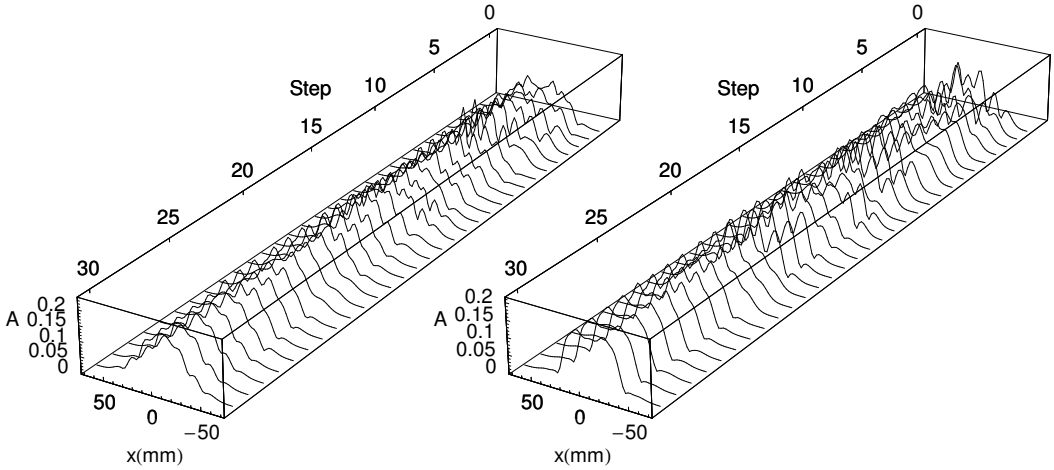


Figure 3.4: The progress of the DE optimisation where the merit function η is used, showing the amplitude of the field for the fittest parameters for the first 30 steps. *Left* near-field, *right* far-field.

The DE algorithm was initialised with random values assigned to the amplitude and phase of 8 even Gaussian-Hermite modes, and the optimisation was terminated when it converged, or when 100 iterations were reached. Initial phase values were chosen in the range $0 \rightarrow 2\pi$, and amplitude values were chosen in the range $0 \rightarrow 1$. The optimisation was run 10 times with a different set of random initial parameters for each trial, using η, the MSE where the trial field is normalised and the MSE and where it is not normalised as the merit function. Figure 3.3 shows a progress plot for the trials using the η merit function and the MSE where the trial field is not normalised. Using the MSE as a merit function and not normalising the trial field, there is a range of final values, due to the optimisation converging prematurely and giving a less optimal result, whereas when η is used as the merit function, all the trials converge on the same optimal value. Using the MSE merit function with the field normalised was found to give equivalent results to that where the η merit

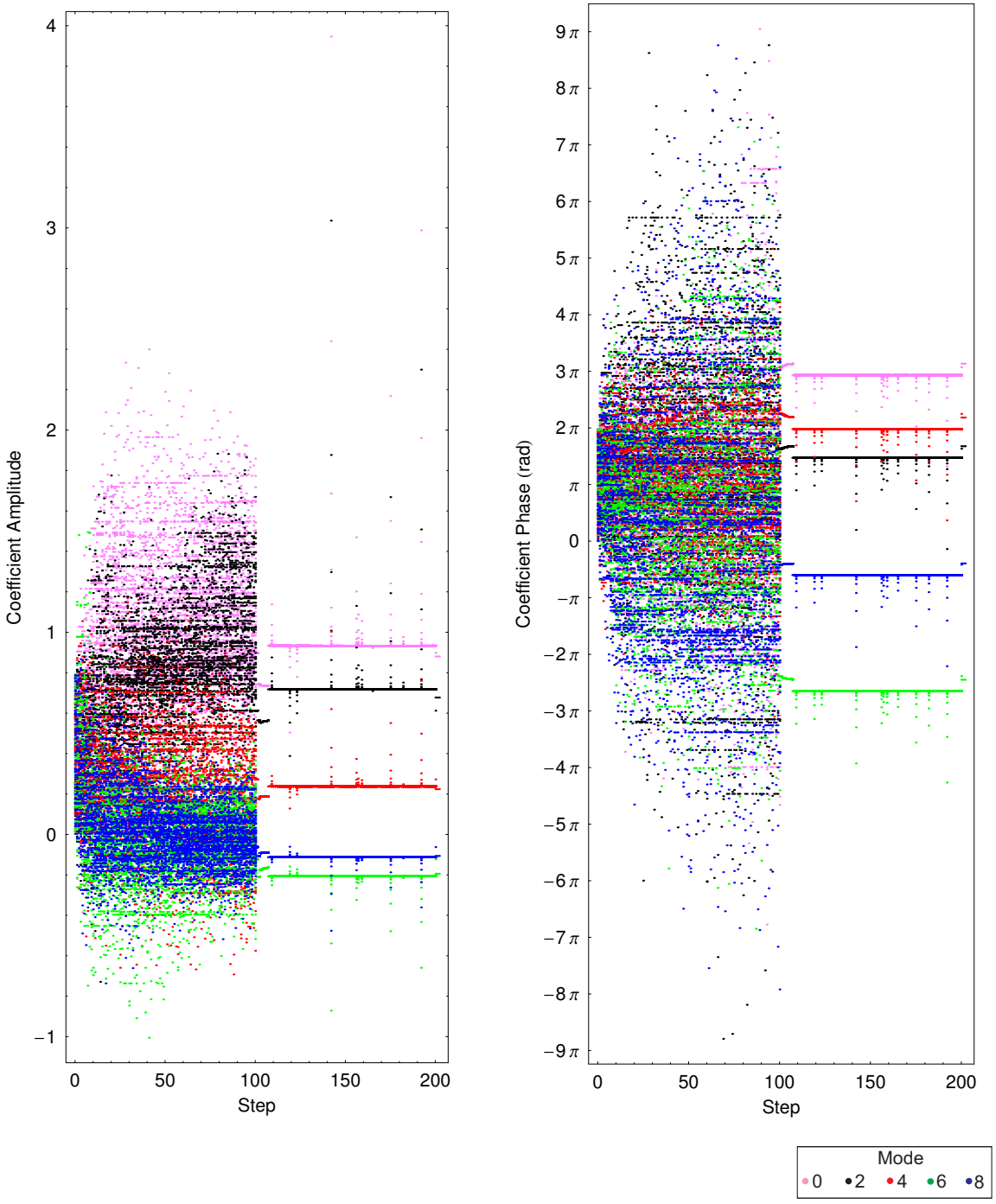


Figure 3.5: The parameter values at each step of the DE optimisation, to transform a Gaussian input field to a top-hat distribution, using the first 8 even modes and optimising *left* the amplitude and *right* the phase of the coefficient. The amplitude coefficient parameters are given random initial values in the range $0 \rightarrow 0.8$ and the phase coefficient parameters are given random values in the range $0 \rightarrow 2\pi$. The field is normalised when the merit function (the overlap integral) is evaluated. After 100 iterations the most optimal parameters are selected and a local optimisation is carried out.

function was used. The η merit function was chosen for the optimisation carried out here. Figure 3.4 shows how the near- and far-field amplitude distributions evolve as the optimisation progresses.

Figure 3.5 shows the parameter values at each step of the DE optimisation. The parameters are initially given random values, and each subsequent step is a perturbation of the previous step. A number of regions are explored simultaneously by the population of parameter values. In the plot the parameters (mode coefficients) are represented by different colours. The points are clustered around particular regions of parameter space, corresponding to parameter values of high fitness. As the optimisation progresses a region of higher fitness is found, and the points then cluster around that new region. There is a general drift of each parameter towards a particular value during the course of the optimisation. At a predetermined number of steps (in this case 100 steps) the fittest parameters are selected, and the local minimisation technique, the Quasi-Newton method, is applied. In the optimisations carried out below, 300 steps are used.

3.2.2 Example 1: Top-Hat with a Variable Radius

A DOE is now designed to convert a Gaussian input field of wavelength $\lambda = 3mm$ to a top-hat distribution. A 100mm-radius Gaussian field is converted to a top-hat with a radius that is kept as a free parameter to be optimised during the DE. The optimisation was carried out with even Gaussian-Hermite modes in the ranges $0 \rightarrow 2$, $0 \rightarrow 6$, $0 \rightarrow 10$, $0 \rightarrow 14$, $0 \rightarrow 18$, $0 \rightarrow 22$, $0 \rightarrow 26$ and $0 \rightarrow 30$. The DE optimisation was terminated after 300 steps, and the result of this was optimised using the local minimisation Quasi-Newton method. Figure 3.6 *left* shows the final value of the merit function η for each of the optimisations. A peak of $\eta = 1.986$ occurred for $0 \rightarrow 18$ modes used in the optimisation (for a perfect match $\eta = 2$). The value of r_f found by the optimisation corresponded to a top-hat radius of $\theta_f = 4.1°$. Figure 3.6 *right*, shows that the time taken to complete the optimisation increases exponentially with the increasing number of modes. Figure 3.7 *left* shows that the size of the optimal mode set used to describe both the Gaussian and top-hat distributions tends to decrease with the size of the mode set used to describe the field. This is expected, as the extent of the mode set should be similar to the extent of the distributions they are describing, as shown in Figure 3.1. As shown in Figure 3.7 *right*, the radius of the top-hat found by the optimisation remains approximately constant relative to the radius of the Gaussian distribution, regardless of the number of modes used. However, as shown in Figure 3.8, when r_f is converted to the far-field angular radius, θ_f, using Equation 3.11, the scale of the optimum radius tends to increase as a function of the maximum number of modes used in the optimisation, because θ_f is inversely proportional to w_0^2.

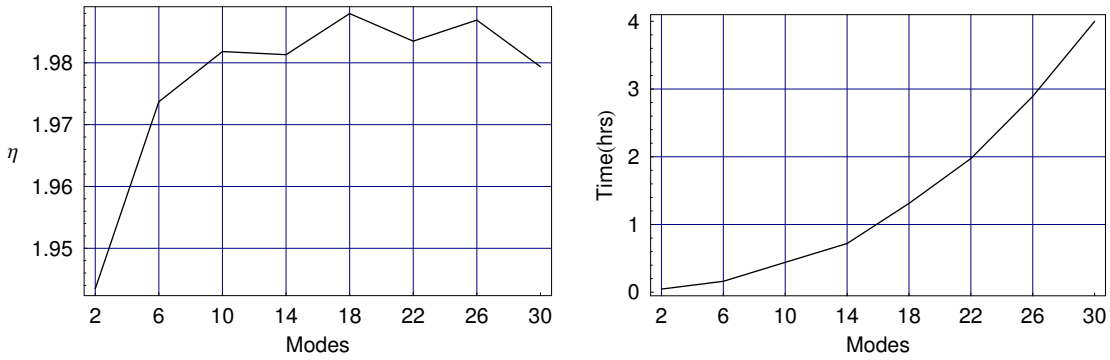


Figure 3.6: *Left* the merit function η as a function of the number of modes used in the optimisation. *Right* the Computation time as a function of the number of modes used in the optimisation.

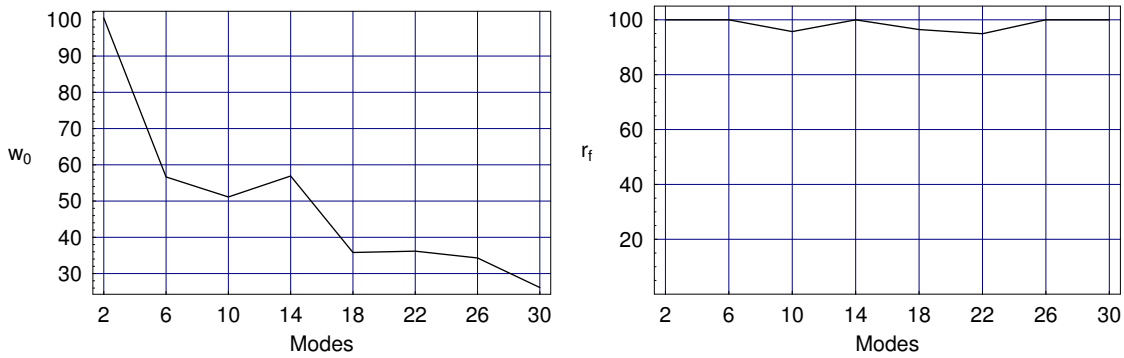


Figure 3.7: *Left* waist radius, w_0 (in mm), of the mode set found by the optimisation discussed in the text, as a function of the maximum mode order used in the GBMSO *right* the radius of the top-hat field, r_f (in mm), found by the optimisation as a function of the maximum mode order used in the GBMSO.

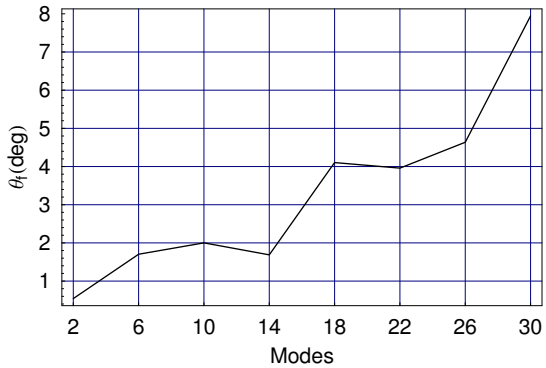


Figure 3.8: The radius of the top-hat field in angular coordinates given by Equation 3.11 using the parameters found by the optimisation shown in Figure 3.7, as a function of the maximum mode order used in the GBMSO.

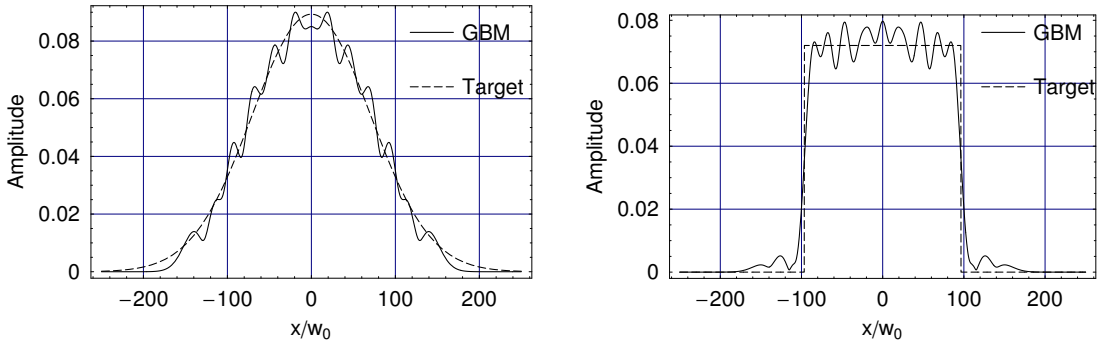


Figure 3.9: *Left* the near-field and *right* the far-field amplitude of the field corresponding to the optimised mode set.

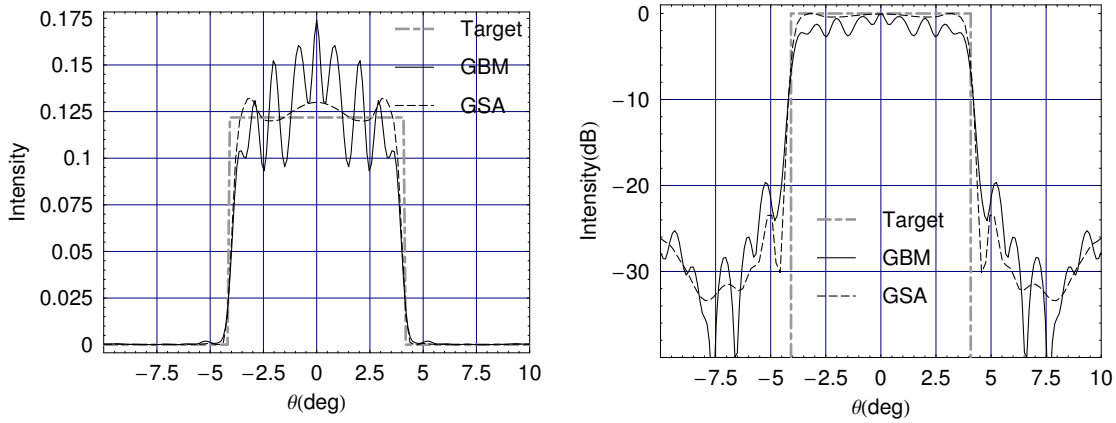


Figure 3.10: The far-field pattern from a DOE that converts a Gaussian beam to a top-hat with $\lambda = 3mm$, using the best of 100 trials of the GSA and GBMSO. *Left* plot of linear intensity and *right:* Log plot of intensity.

The far-field amplitude distribution shown in Figure 3.9 *right* would result if the amplitude of the near-field solution, $|E_n(x)|$, (shown in Figure 3.9 *left*) is used as the input. However, the DOE is designed for a *Gaussian* input field. In order to determine the actual far-field from the DOE given by the phase of the near-field solution, the amplitude distribution is replaced by the 100mm input Gaussian distribution, $T_n(x)$, and transformed to the far-field using the FFT as follows:

$$\text{far-field} = \mathcal{F}\{T_n(x)\exp(i\arg[E_n(x)]\}. \tag{3.14}$$

The resulting field is shown in Figure 3.10. The difference between the field described by the GBMs in the near-field, $E_n(x)$, and the target Gaussian distribution causes a corresponding deviation in the far-field compared to the optimised solution given by $E_f(x)$. This far-field intensity distribution is much less uniform than that returned by the GSA. The solution returned by the GSA has a large spherical phase component, which the limited number of GBMs used in the GBMSO cannot describe exactly.

3.2.3 Example 2: 5° Radius Top-Hat with a Variable GBM Waist Position

In order that the GBMs can construct the spherical phase component of the field, the reference plane, z, (i.e. the waist location) of the GBMs is optimised, along with the amplitude and phase of the mode coefficients and w_0, as shown in Figure 3.11: GBMSO is carried out with a variable reference plane, z, for a DOE that converts a 100mm-radius Gaussian to a 5° radius top-hat field. Figure 3.12 *left* shows the merit function η for optimisations using even modes in the ranges $0 \rightarrow 2$, $0 \rightarrow 6$, $0 \rightarrow 10$, $0 \rightarrow 14$, $0 \rightarrow 18$, $0 \rightarrow 22$ and $0 \rightarrow 26$; initialised with random values of w_0 in the range 25mm $\rightarrow$ 35mm and z in the range -1200mm $\rightarrow -700$mm, giving values of w in the range 39.9mm $\rightarrow$ 52.2mm at the input plane (at the DOE). η tended to increase with the number of modes used and peaked for 22 modes at $\eta = 1.991$. Figure 3.12 *right* shows that the evaluation time increased with the number of modes used in the optimisation. Figure 3.13 shows the amplitude and phase of the mode coefficients returned for each range of modes.

Figure 3.14 shows the values of w_0 and z found by the optimisations. The large degree of variation in the values of w_0 and z returned by the optimisation may be explained by the fact that a given phase and amplitude distribution can be described well by different mode sets, if a sufficiently large number of modes are available.

Figures 3.15-3.18 show the results of the optimisation using 22 modes. Figure 3.15 shows the amplitude and phase of the GBMs returned by the optimisation. The GBMs are able to construct a Gaussian in the near-field almost exactly, with

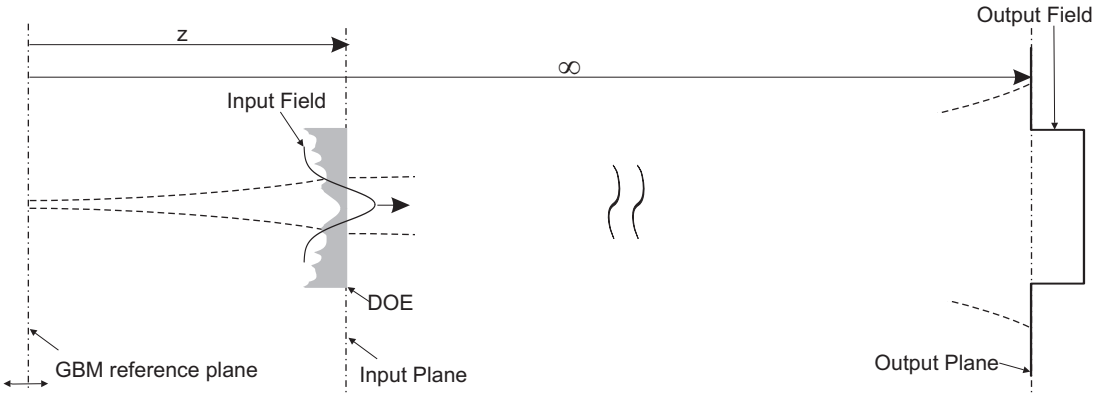


Figure 3.11: The incoming Gaussian beam is transformed to a top-hat distribution in the far-field. The position of the reference plane is a parameter to be optimised in the GBMSO.

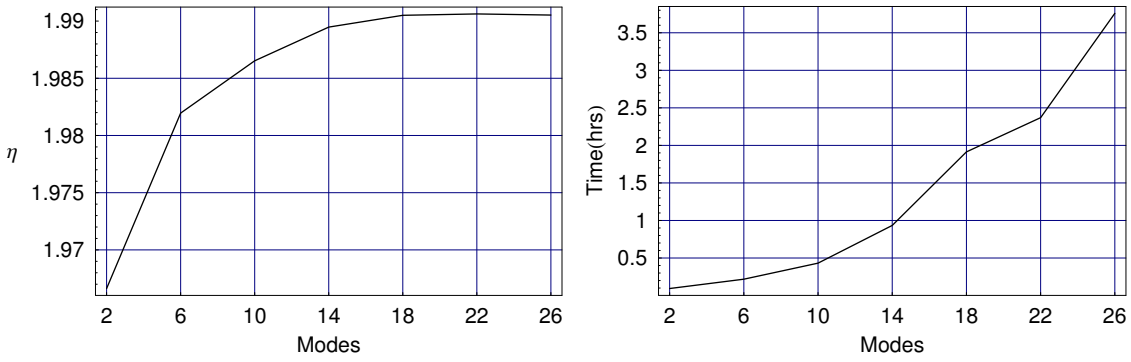


Figure 3.12: The best result of the optimisations to transform a Gaussian input field to a top-hat with a 5° radius. *Left* the merit function η as a function of the number of modes used in the optimisation. *Right* the Computation time as a function of the number of modes used in the optimisation.

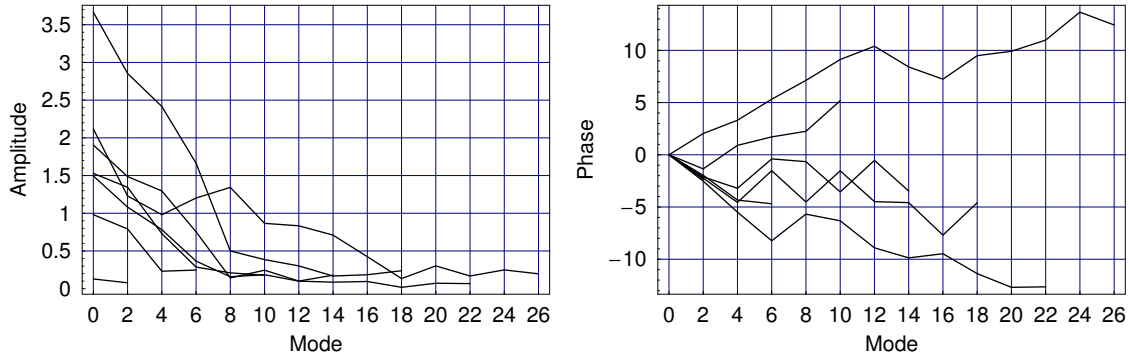


Figure 3.13: *Left* the amplitude and *right* the phase of the optimised mode coefficients for a DOE that converts a 100mm Gaussian to a 5° radius top-hat using even modes in the ranges: $0 \rightarrow 2$, $0 \rightarrow 6$, $0 \rightarrow 10$, $0 \rightarrow 14$, $0 \rightarrow 18$, $0 \rightarrow 22$ and $0 \rightarrow 26$.

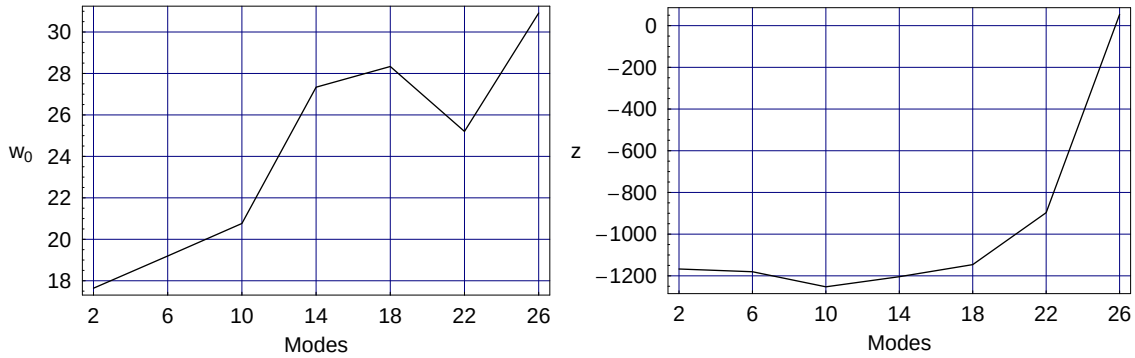


Figure 3.14: The results of the optimisation to transform a Gaussian input field to a top-hat with a 5° radius. *Left* waist radius, w_0 (in mm) of the mode set found by the optimisation as a function of the maximum mode order used in the GBMSO, *right* propagation distance, z (in mm), to the waist location, as a function of the maximum mode order used in the GBMSO.

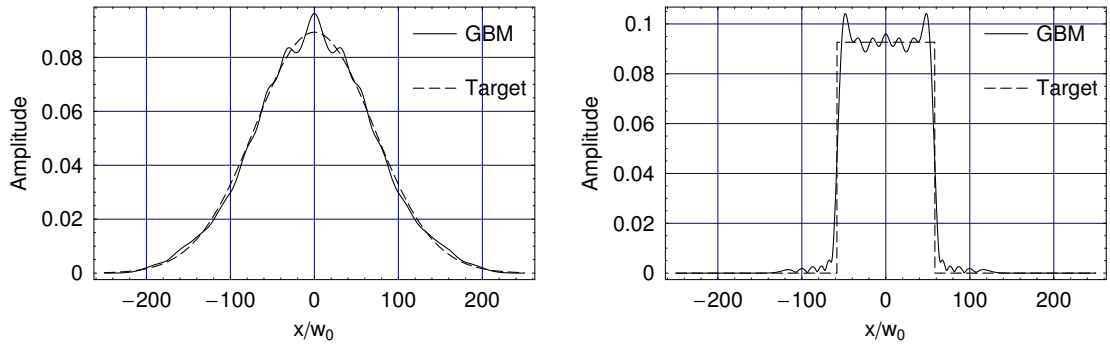


Figure 3.15: The results of the optimisation to transform a Gaussian input field to a top-hat with a 5° radius. *Left* the amplitude of the GBMs in the near-field $|E_n|$ and *right* the amplitude of the GBMs in the far-field $\left|E_f\right|$.

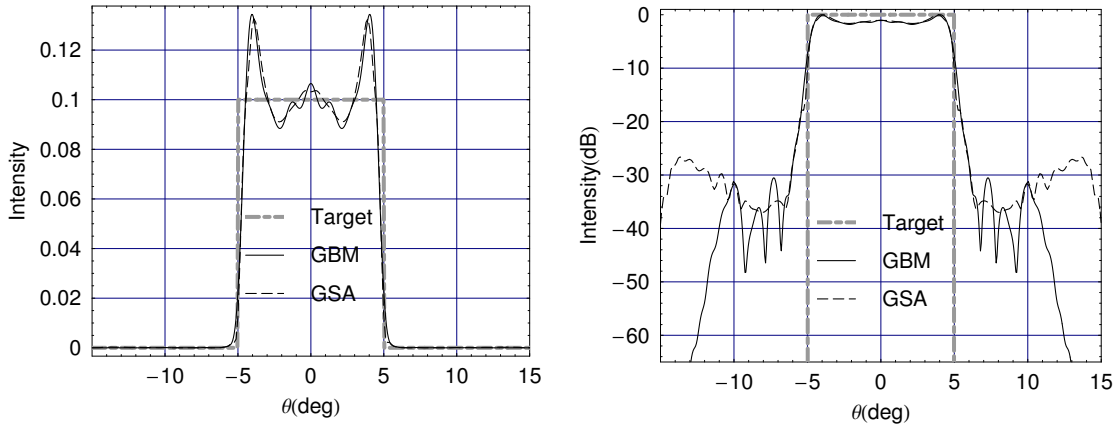


Figure 3.16: The far-field pattern from a DOE that converts a Gaussian beam to a 5° radius top-hat with $\lambda = 3mm$, using the best of 100 trials of the GSA and GBMSO using even Gaussian-Hermite modes in the range $0 \rightarrow 10$. For the GBMSO $\eta_f^2 = 0.979$ compared to $\eta_f^2 = 0.982$ for the GSA.

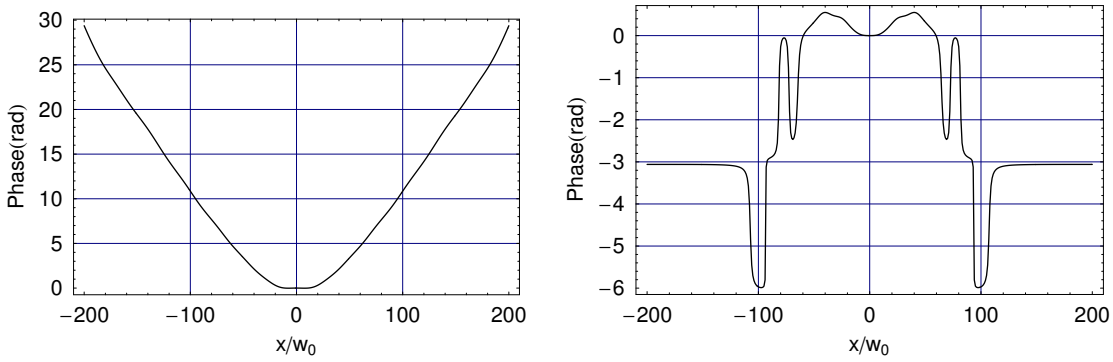


Figure 3.17: The phase in *left* the near-field giving the DOE profile and *right* the far-field of a DOE that converts a 100mm Gaussian to a top 5° radius top-hat using GBMSO.

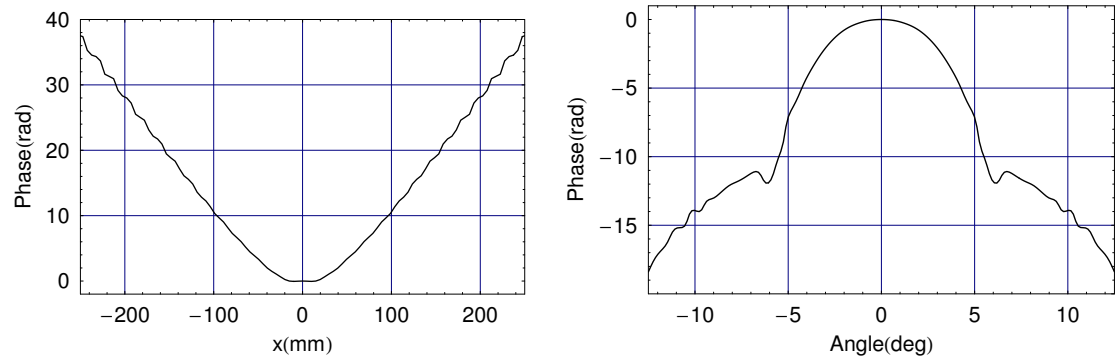


Figure 3.18: The solution with the lowest MSE for 100 trials of the GSA, each with 100 iterations for a DOE that converts a 100mm Gaussian to a top 5° radius top-hat. The phase in *left* the near-field giving the DOE profile and *right* the far-field.

a merit function $\eta_n^2 = 0.998$ and $\eta_f^2 = 0.983$ in the far-field. The similarity of the field given by the GBMs to the near-field Gaussian resulted in a far-field pattern, using the FFT (see Equation 3.14), which was close to that given by $\left|E_f\right|$. This far-field pattern is shown in Figure 3.16, for which $\eta_f^2 = 0.978$. This compares to $\eta_f^2 = 0.983$ for the best of 100 trials of the GSA, each with 100 iterations. High spatial frequency components of the near-field tend to create a far-field with more off-axis power. As the spatial frequency is limited in the GBMSO by the number of modes used, the resulting solutions tend to have less unwanted off-axis power. The phase of the GBMSO solution in the near-field, shown in Figure 3.17, is very similar to that of the GSA solution shown in Figure 3.18, however, the phase of the GBMSO solution is slightly smoother than that of the GSA. The GBMSO solution could not describe the higher frequency components of the solution found by the GSA due to the limited number of modes used, resulting in a slightly less optimal solution as quantified by the merit function. However, a smoother solution is generally more desirable for manufacturing.

3.2.4 Example 3: 9° Radius Top-Hat with a Variable GBM Waist Position

Figures 3.19-3.22 show the results of the GBMSO for a DOE that converts a Gaussian input beam to a 9° radius top-hat field, using the first 22 even modes - again using z as a free parameter in the optimisation. In this case, $\eta_n^2 = 0.9994$ and $\eta_f^2 = 0.9872$, and the FFT in Equation 3.14 gave $\eta_f^2 = 0.9854$. This is marginally better than the result of the best of 100 trials of the GSA, also shown in Figure 3.20, where the $\eta_f^2 = 0.9851$. Comparing the phase of the solutions of the GBMSO and the GSA shown in Figures 3.21 and 3.22, the phase of the solution returned by GBMSO is much smoother than that returned by the GSA; it would therefore be much simpler to manufacture.

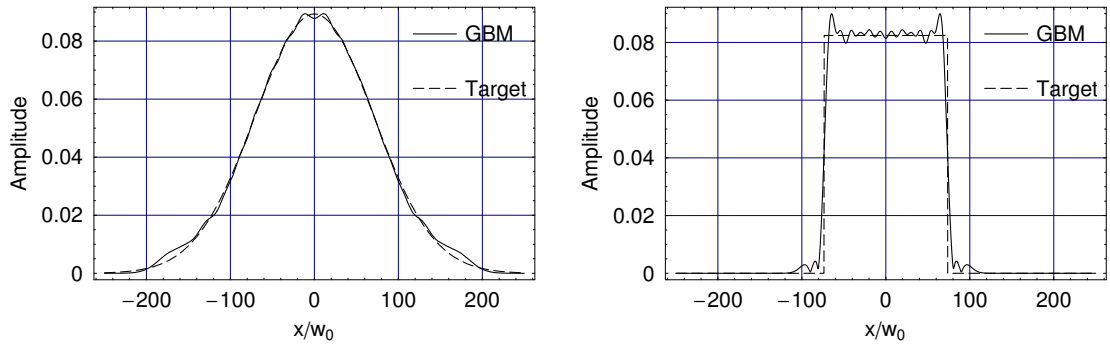


Figure 3.19: The results of the optimisations to transform a Gaussian input field to a top-hat with a 9° radius. *Left* the amplitude of the GBMs in the near-field $|E_n|$ and *Right* the amplitude of the GBMs in the far-field $|E_f|$.

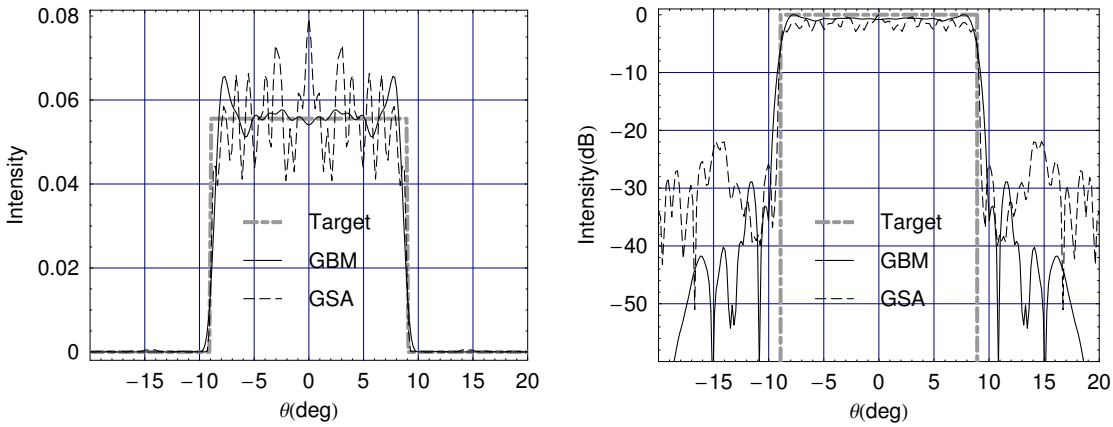


Figure 3.20: The far-field pattern from a DOE that converts a Gaussian to a 9° radius top-hat with $\lambda = 3$ using the best of 100 trials of the GSA, and GBMSO using even Gaussian-Hermite modes in the range $0 \rightarrow 10$. For the GBMSO $\eta_f^2 = 0.9854$, compared to $\eta_f^2 = 0.9851$ for the GSA.

The GSA performed poorly because the optimisation from the Gaussian to top-hat beam becomes more nonlinear as the radius of the top-hat is increased. With

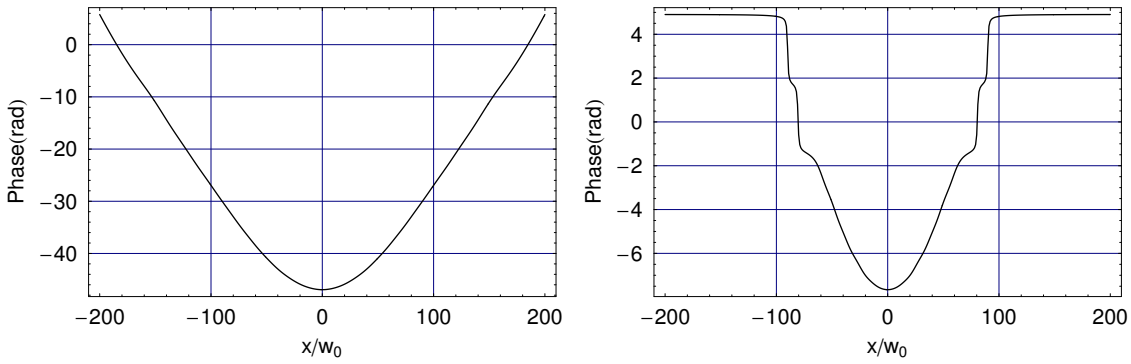


Figure 3.21: The solution for a Gaussian beam is which transformed into a 9° top-hat amplitude distribution, using GBMSO. *Left* the phase of the solution in the near-field, and *right* the phase of the solution in the far-field.

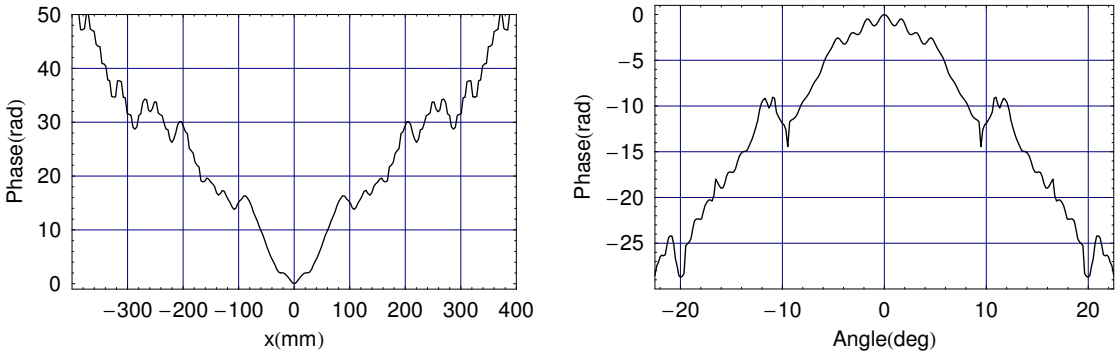


Figure 3.22: The solution for a Gaussian beam which is transformed into a 9° top-hat field, for the best of 100 trials of the GSA, each with 100 iterations. *Left* the phase of the solution in the near-field and *right* the phase of the solution in the far-field.

GBMSO a global optimisation algorithm was used allowing it to find a more optimal solution than the GSA, which is a local optimisation algorithm. For this transform the GBMSO performed better than either the GSA or unidirectional optimisation.

3.2.5 Example 4: 1st Order Gaussian-Hermite, using Even Modes

A DOE is now designed which converts a Gaussian beam to a first-order Gaussian-Hermite amplitude distribution in the far-field, where radius of the Gaussian-Hermite is a parameter to be optimised. Whereas previously, using only even Gaussian-Hermite modes was the logical choice, as both the Gaussian and top-hat are symmetric, here, the choice of whether to use even or odd modes is not obvious. In Chapter 2 it was shown how a first order Gaussian-Hermite amplitude is not symmetric when considering only amplitude, but is symmetric if described in terms of absolute of amplitude and phase (the phase is antisymmetric).

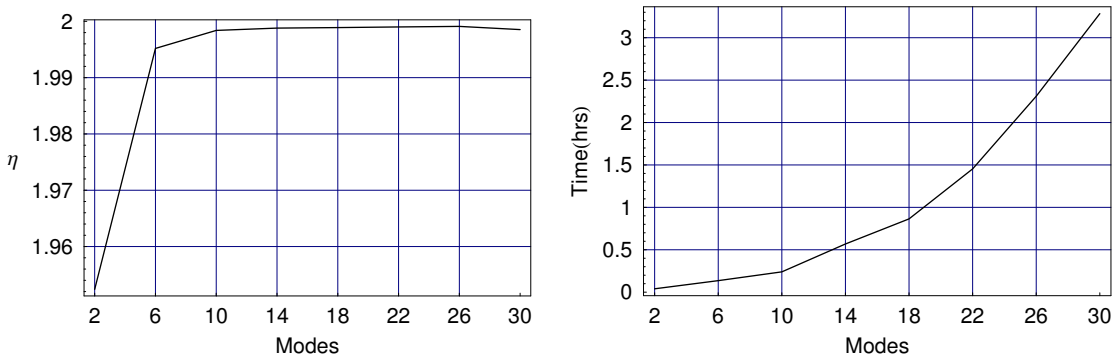


Figure 3.23: The results of the DE optimisation to convert a 100mm-radius Gaussian to a 1st order Gaussian-Hermite beam, the radius of which is a free parameter. *Left* the merit function η as a function of modes used in the optimisation. *Right* the time taken for the optimisation.

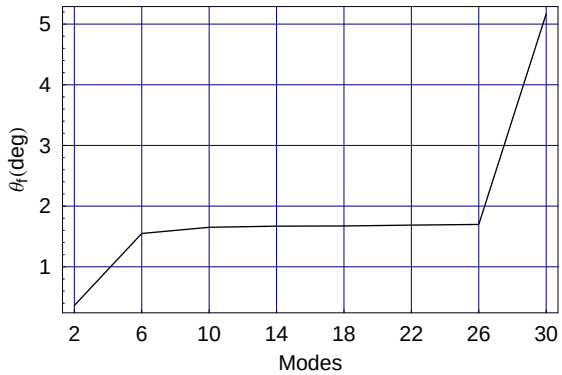


Figure 3.24: The far-field radius in degrees of the 1st order Gaussian-Hermite found by the GBMSO optimisation, as a function of the highest order mode used in the optimisation.

In the solution shown here, GBMSO was carried out without making the distance to the reference plane, z, a variable, because trials in which z was a variable did

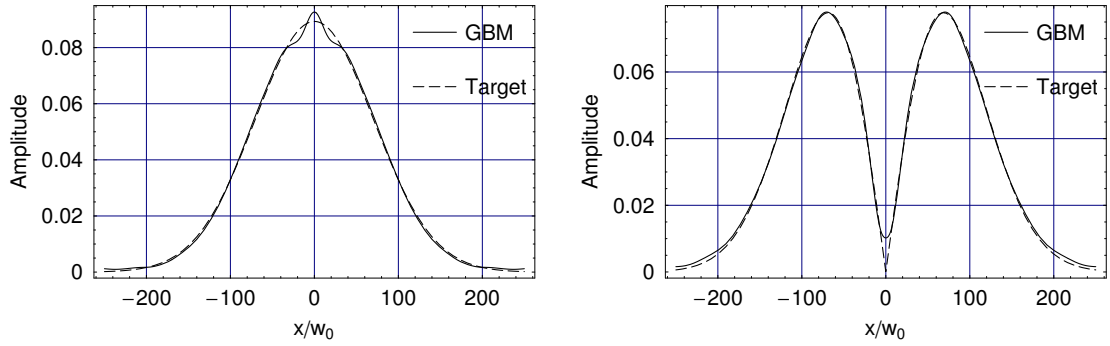


Figure 3.25: The results of the optimisation to transform a Gaussian input field to a 1st order Gaussian-Hermite amplitude distribution. *Left* the amplitude of the GBMs in the near-field $|E_n|$ and *right* the amplitude of the GBMs in the far-field $|E_f|$.

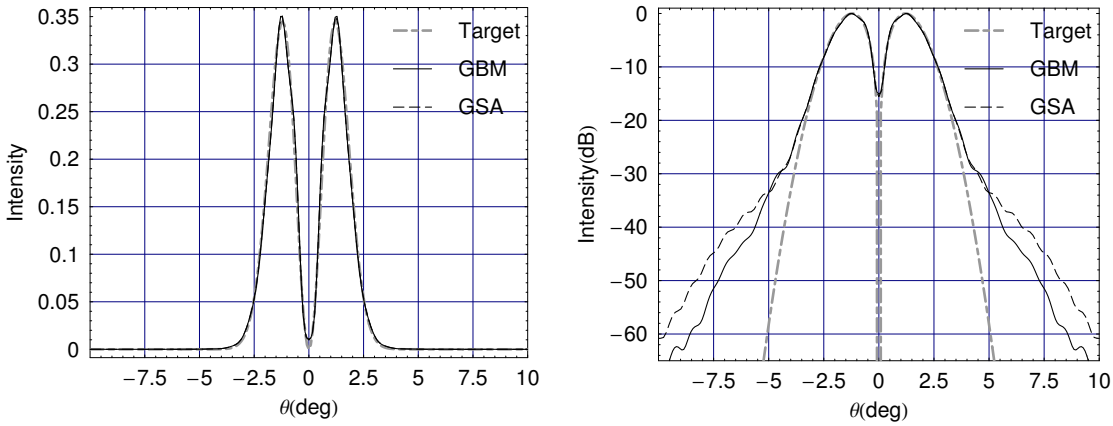


Figure 3.26: The far-field of a DOE designed to convert a 100mm-radius Gaussian to a 1st order Gaussian-Hermite amplitude distribution, given by Equation 3.14. *Left* linear scale, *right* log scale.

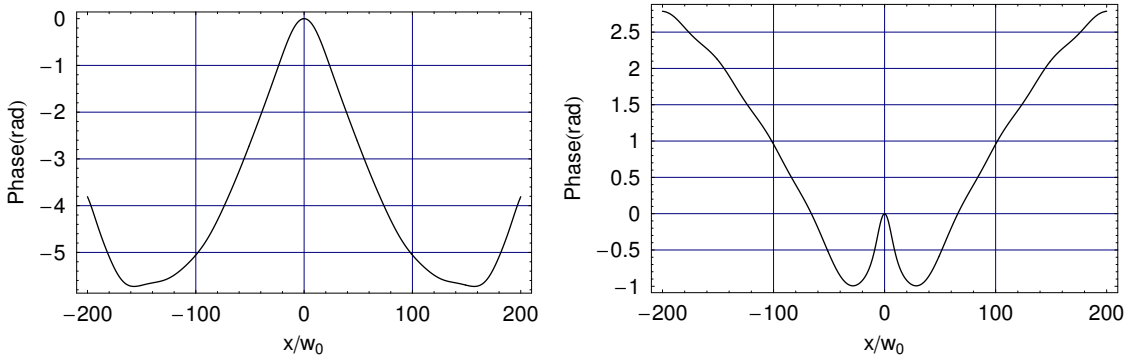


Figure 3.27: The phase given by *left* $\arg(E_n)$, the near-field phase distribution and *right* $\arg(E_f)$, the far-field phase distribution.

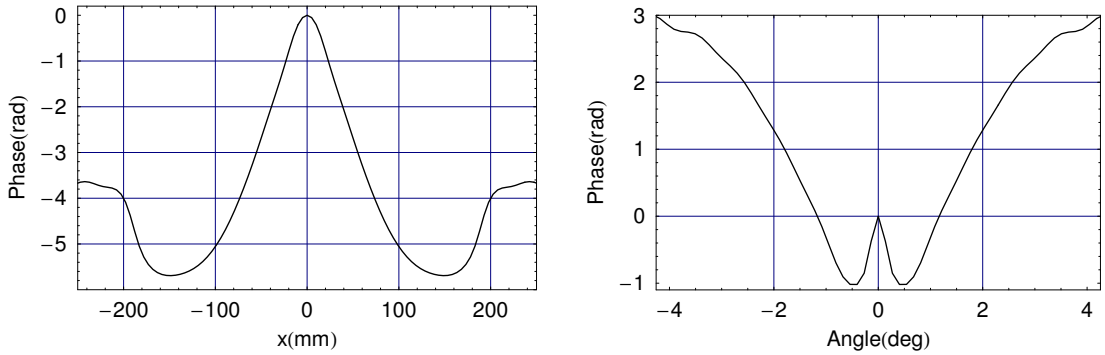


Figure 3.28: The phase of the best of 100 trials of the GSA, each with 100 iterations, for a DOE that converts a 100mm-radius Gaussian distribution to a Gaussian-Hermite amplitude distribution with a scale found by GBMSO to be optimum.

not yield an improved solution (i.e. a solution with a greater value of η) over the solution where z was fixed; this is because the optimal phase distribution does not have a larger spherical component. Figure 3.23 *left* shows the value of the merit function, η, found by optimising even modes in the ranges $0 \rightarrow 2$, $0 \rightarrow 6$, $0 \rightarrow 10$, $0 \rightarrow 14$, $0 \rightarrow 18$, $0 \rightarrow 22$, $0 \rightarrow 26$ and $0 \rightarrow 30$. The merit function, η, tended to increase with the number of modes used, and peaked when using even modes in the range $0 \rightarrow 26$ at $\eta = 1.9991$. Figure 3.23 *right* shows that the evaluation time increased with the number of modes used in the optimisation. The far-field radius of the optimised solution, given as a divergence angle, as a function of the number of modes used in the optimisation, is shown in Figure 3.24. The far-field angular scale of the 1$^{\text{st}}$ order Gaussian-Hermite found by the optimisations was constant at 1.61°, except where $0 \rightarrow 2$ and $0 \rightarrow 30$ modes were used. Where $0 \rightarrow 30$ modes were used, the GBMSO found a solution at a larger scale that was not possible when a smaller number of modes were used. This solution had a more complicated near- and far-field phase distribution.

The solution found by using the modes in the range $0 \rightarrow 26$ was selected for further analysis, as it had the highest merit function, η. As shown in Figure 3.25, the near- and far-field solutions are very close to the target near- and far-field amplitude distributions. Figure 3.26 and 3.27 show the amplitude and phase of the field when a Gaussian is used as the input amplitude distribution, and Fourier transformed to the far-field using Equation 3.14 - where $\eta_f^2 = 0.9968$. The scale of the 1$^{\text{st}}$ order Gaussian-Hermite found from the GBMSO was then used in the GSA for comparison; all of the 100 trials of the GSA, each with 100 iterations, gave a solution with $\eta_f^2 = 0.9953$. The phase of this solution, shown in Figure 3.28, is very similar to that found by GBMSO. The scale found by GBMSO enabled the GSA to find a good solution; the GSA cannot optimise the scale of the target far-field solution. Using only even modes in the GBMSO ensured that only a symmetric solution could be returned - unlike the GSA, where an asymmetric solution can be returned, as

shown in Chapter 2. There is a scale of Gaussian-Hermite that a symmetric solution is most optimal and another scale where a asymmetric solution is optimal. The anti-symmetric solution is described well by the odd Gaussian-Hermite modes, as they also have a π phase shift at x=0mm.

3.2.6 Example 6: 1st Order Gaussian-Hermite, using Odd Modes

GBMSO is now used to optimise only the odd modes for the Gaussian to 1st order Gaussian-Hermite amplitude distribution transform, where the scale of the target distribution is free. Figure 3.29 shows the near- and far-field amplitude distributions, using 43 odd modes. As the odd modes are zero at $x = 0$, the near-field construction of the target Gaussian beam using GBMs must be zero at $x = 0$. The near-field phase, shown in Figure 3.30, has a π phase shift at $x = 0$. The phase returned by the GSA is also asymmetric. The intensity distribution in the far-field, given by the FFT of a Gaussian amplitude distribution combined with the near-field phase solution is shown in Figure 3.31; η_f^2 was found to be equal to 0.977.

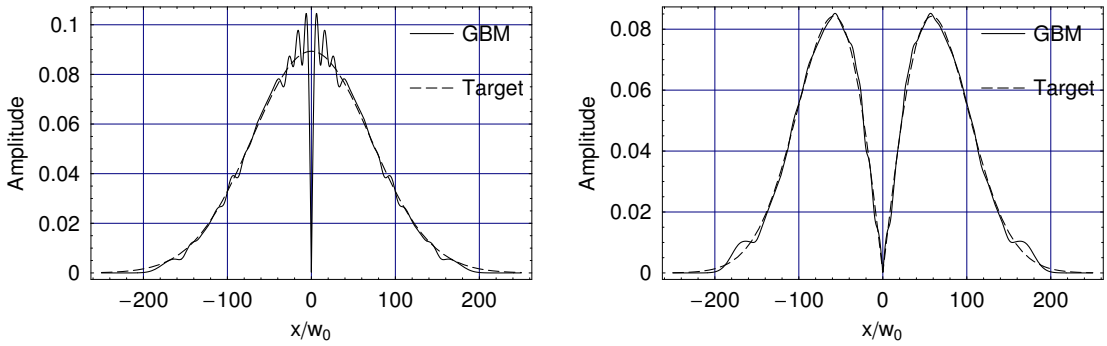


Figure 3.29: The results of the optimisation to transform a Gaussian input field to a 1st order Gaussian-Hermite amplitude distribution using only odd modes. *Left* the amplitude of the GBMs in the near-field $|E_n|$ and *right* the amplitude of the GBMs in the far-field $\left|E_f\right|$

3.2.7 Example 7: Three Gaussian Beams

In this example, GBMSO is employed to design a DOE that converts a Gaussian to three equally spaced Gaussian beams. The target field is given by

$$T_f(x) = \frac{\psi_0(w, x/s - w/5, 0) + \psi_0(w, x/s, 0) + \psi_0(w, x/s + w/5, 0)}{\sqrt{3s}}, \qquad (3.15)$$

where the beams $\psi_0(w, x, z)$ are separated by $w/5$ and the x coordinate is scaled by s. The beams are normalised by the $1/\sqrt{3s}$ term.

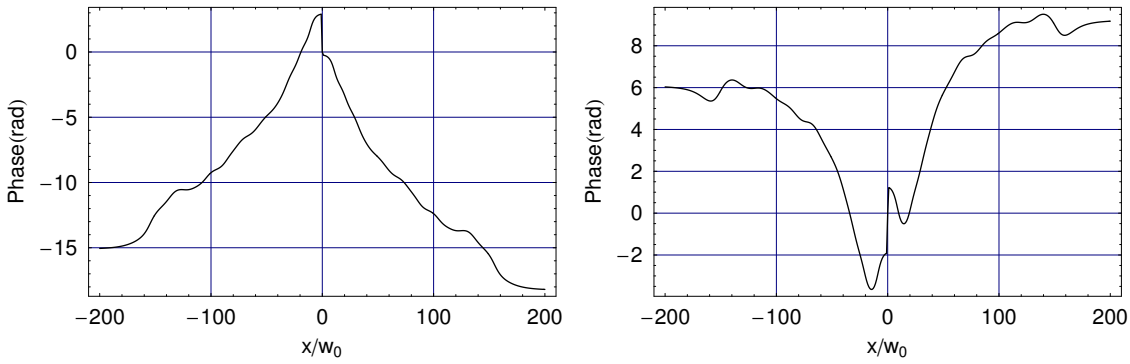


Figure 3.30: *Left* the near-field phase distribution given by $\arg(E_n)$ and *right* the far-field phase distribution given by $\arg(E_f)$.

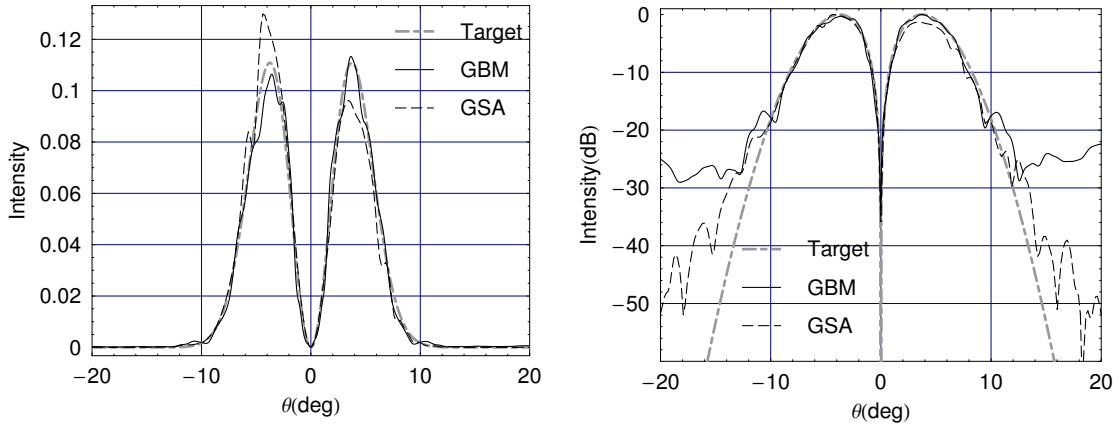


Figure 3.31: The far-field from a DOE to convert a 100mm-radius Gaussian to a 1st order Gaussian-Hermite amplitude distribution, given by Equation 3.14. *Left* linear scale, *right* log scale (only odd modes were used).

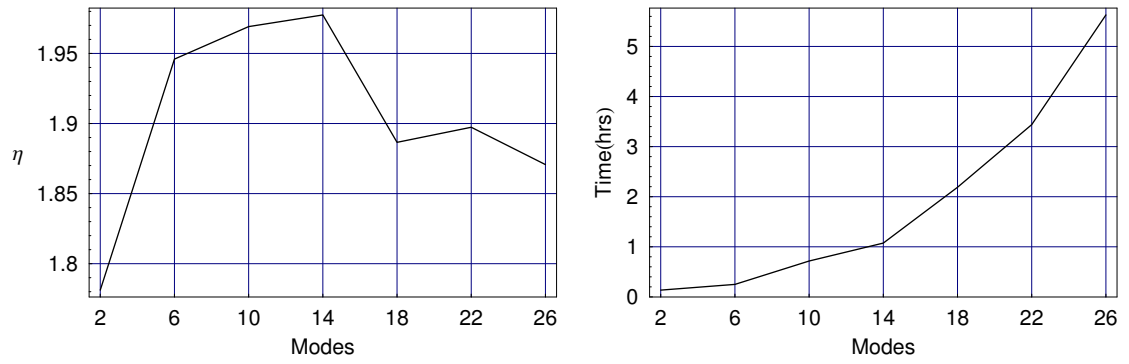


Figure 3.32: The merit function η for trials using even modes in the ranges $0 \rightarrow 2$, $0 \rightarrow 6$, $0 \rightarrow 10$, $0 \rightarrow 14$, $0 \rightarrow 18$, $0 \rightarrow 22$ and $0 \rightarrow 26$.

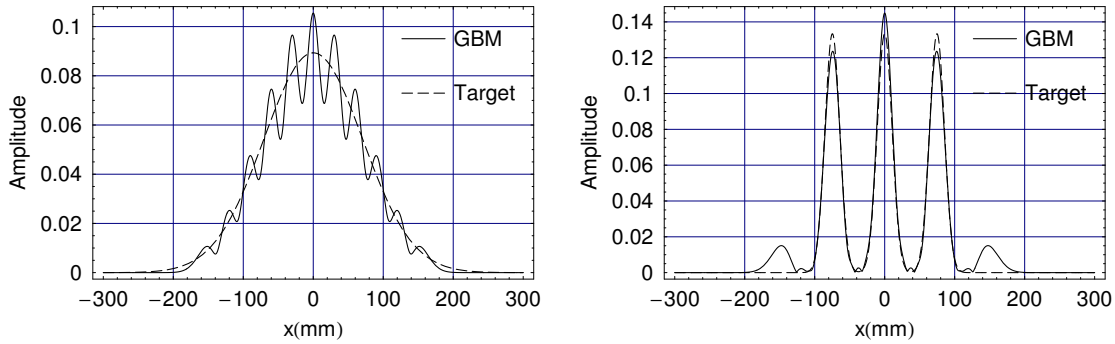


Figure 3.33: The amplitude of the field, found using GBMSO with 14 even modes, *left* the near-field, *right* the far-field.

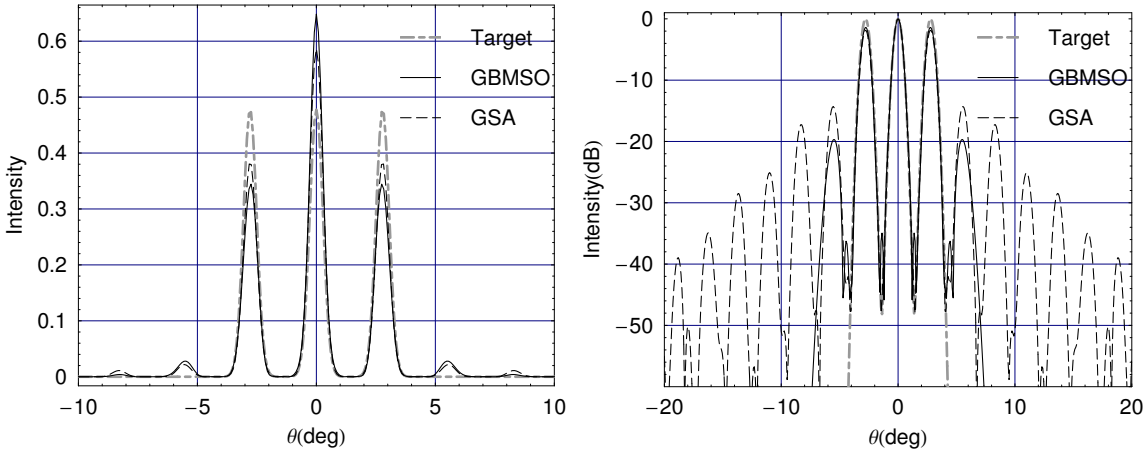


Figure 3.34: The intensity in the far-field obtained by illuminating the DOE with an input Gaussian beam. The DOE found using GBMSO (using $0 \rightarrow 14$ even modes) and found using the GSA are compared. The GSA solution was found using the scale of the output beams found by the GBMSO. *Left* linear scale, *right* log scale.

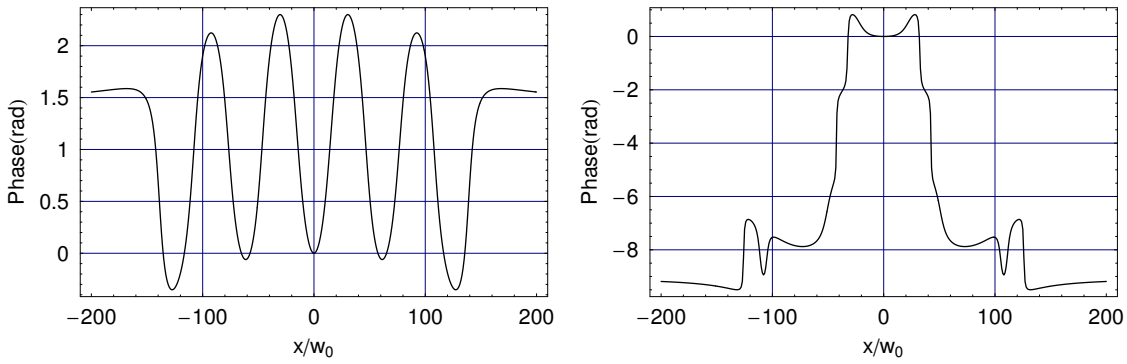


Figure 3.35: *Left* the near-field phase, which gives the DOE surface, of the solution found using GBMSO using $0 \rightarrow 14$ even modes, *right* the far-field phase.

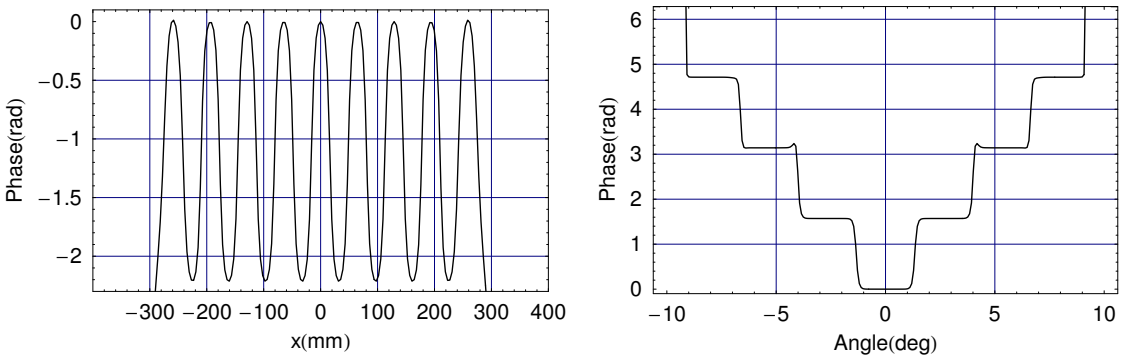


Figure 3.36: The phase of the solution in, *left* the near-field and *right* the far-field, found by the GSA, that gives the far-field distribution shown in Figure 3.34.

The GBMSO was carried out with and without keeping the distance to the reference plane a variable; as keeping z a free parameter did not give an improved solution, the solution found by keeping z fixed to zero is given here. Figure 3.32 shows η, the coupling, as a function of the number of modes and the evaluation time of the DE optimisations. Figures 3.33-3.35 show the results for the best solution, which optimises even modes in the range $0 \rightarrow 14$. For the GSA $\eta_f^2 = 0.938$ and for the GBMSO $\eta_f^2 = 0.932$. The best scale, s, found by GBMSO is used to give the field shown in Figure 3.34, using the best of 100 trials of the GSA. The near-field phase returned by the GSA is shown in Figure 3.36. Although η was smaller for the GBMSO solution than for the GSA, the GBMSO found a scale that enabled the GSA to return a highly optimal result. Scaling the solution is not possible with the GSA and setting the scale to an arbitrary fixed value is likely to give less optimal results. The phase given by GBMSO optimisation has the advantage over that of the GSA in that it has larger features.

3.3 The GSA using GBMs

The GSA was implemented in Chapter 2 using the FFT to propagate between the near- and far-fields. Here, the algorithm will be implemented using GBMs to transform the fields, instead of the FFT. An advantage of the GSA over GBMSO is that, although it can get stuck in a local minimum, using a large number of modes does not greatly increase the time required to find a solution. The relationship between the solution found using GBMSO and that found by the GSA is discussed here.

As is the case where the GSA was implemented using the FFT, the algorithm propagates between the near-field and the far-field, where, at each plane, the amplitude is replaced by the target near- and far-field amplitude distributions, $T_n(x)$ and $T_f(x)$, while retaining the phase. The GSA is given an initial phase to start the algorithm. Using GBMs, each iteration of the GSA is carried out as follows. In

order to propagate from the near- to far-field, the near-field mode coefficients are calculated; they are given by

$$\mathrm{CN_n} = \int_{-\infty}^{\infty} T_n(x) \exp(i \arg[E_n(x)]) \times \psi_n^*(w_0, x, z = 0) dx, \tag{3.16}$$

where $\psi_n(w_0, x, z)$ is the Gaussian-Hermite of index n and $E_n(x)$ is the near-field from the previous iteration, given by Equation 3.19. The far-field is then given by

$$E_f(x) = \sum_n \mathrm{CN_n} \psi_n(w_0, x, z = \infty). \tag{3.17}$$

Next, $T_f(x) \exp(i \arg[E_f(x)])$ is propagated to the near-field, by first calculating the far-field mode coefficients as follows:

$$\mathrm{CF_n} = \int_{-\infty}^{\infty} T_f(x) \exp(i \arg[E_f(\theta)]) \times \psi_n^*(w_0, x, z = \infty) dx. \tag{3.18}$$

The near-field is then given by

$$E_n(x) = \sum_n \mathrm{CF_n} \psi_n(w_0, x, z = 0). \tag{3.19}$$

The GSA continues with the next iteration where the near-field mode coefficients are calculated using Equation 3.16.

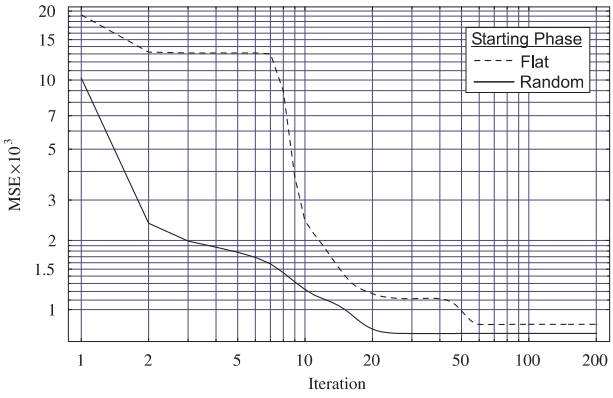


Figure 3.37: The progress of the GSA using GBMs for two different initial phase distributions for a Gaussian input and a top-hat output.

As a test example, the GSA was implemented using the first 50 even GBMs to propagate between the near- and far-fields for a DOE that converts an input Gaussian beam into a top-hat. The size of the near-field input Gaussian and far-field output top-hat were chosen arbitrarily; the solution can be scaled for the required size of input field, as was the case for GBMSO. Figure 3.37 shows $\mathrm{MSE_{nf}}$ (where

MSE_{nf} is the sum of the MSE between the target and obtained fields in the near-field and the far-field) as the GSA progresses, for both flat and random initial phase distributions. As was the case where FFTs are used to propagate the field, the plot shows the characteristic stagnation of the GSA, where the decrease in MSE_{nf} is very small for a number of iterations.

3.3.1 Optimisation of the Initial Phase of the GSA

Rather than using a random phase distribution to start the GSA, an attempt was made to find a phase distribution that returned the smallest MSE_{nf}. The *initial phase distribution*, used as an input to the GSA, was optimised using SA and DE, in order to attempt to improve the solution returned from the GSA, as illustrated in Figure 3.38. The amplitude and phase of the mode coefficients A_n and ϕ_n, with $CN_n = A_n e^{i\phi_n}$, are optimised as before except here, rather than using them to define the near- and far-fields directly, they are used to define the initial phase distribution for the GSA. MSE_{nf} of the field from the final iteration is used as the merit function. The GSA is now implemented in GBMs in one-dimension. An arbitrarily scaled input Gaussian and top-hat were used as the target input and output fields, and the first 20 even GBMs were used to propagate between the near- and far-fields. The mode set that defined the initial phase was optimised, using DE and SA, but neither optimisation procedure could improve on the random input field. The reason for this may be found by examining the plot of Figure 3.39, showing MSE_{nf} of the GSA as it progresses, with the initial phase defined using the first 20 even modes. The plots show MSE_{nf} from the GSA as a function of two arbitrarily selected mode amplitude and phase coefficient parameters, with the other modes set to random values of amplitude and phase. The plots show that MSE_{nf} remains at a constant value over a range of input phase values defined by the mode coefficients, and only jumps to a different value when the phase is sufficiently different, because the GSA finds the same local minimum for similar starting phase values.

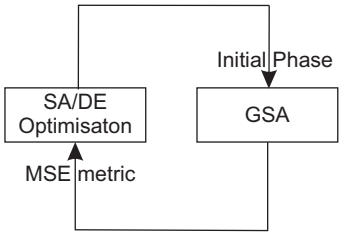


Figure 3.38: The initial phase of the GSA is optimised using SE or DE.

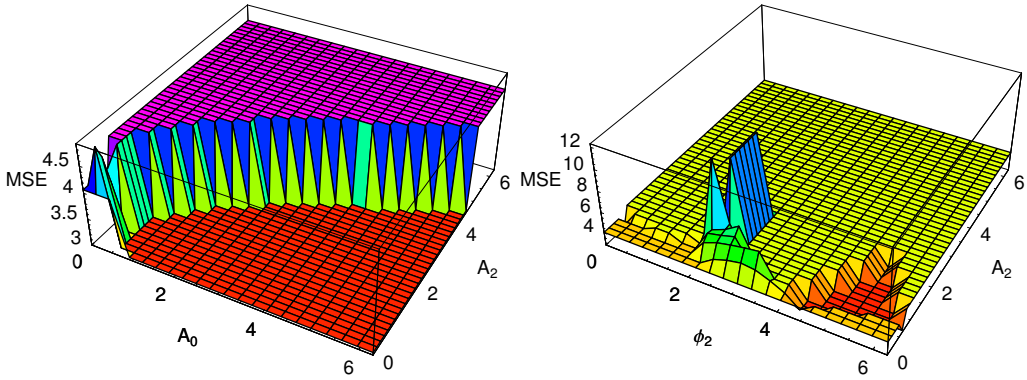


Figure 3.39: MSE_{nf} from the GSA propagating between the near- and far-fields using the first 20 even Gaussian Beam modes. The initial phase of the GSA is set by using a fixed set of random complex mode coefficients and varying *left*: A_0 and A_2, the zeroth and second order mode coefficient amplitudes and *right*, A_2 and ϕ_2, the second order mode coefficient amplitude and phase simultaneously. The plots are flat because GSA tends to go to the same local minimum for similar initial phases, which is the case when similar values of mode coefficients are used as the input.

3.3.2 Value of Mode Coefficients as GSA Progresses

In Section 3.2, GBMSO was used to find the mode set that best describes the near-field and far-field amplitude distributions. Here the near- and far-field mode set, CN_n and CF_n, found using the GSA propagating with GBMs are compared with those found using GBMSO. The amplitude of the mode coefficients in the near- and far-field as the GSA progresses are shown in Figure 3.40; they are observed to converge towards a single set of modes that describe the target near-field and far-field amplitude distributions; however, the algorithm typically stagnates, and does not find near- and far-field mode sets that are exactly equal. This also occurs with the phase of the mode coefficients. For particular target near- and far-field distributions, solutions were found with MSE_{nf} equal to zero, and here the value of the near-field and far-field mode coefficients were exactly equal, and the target near- and far-fields were described exactly. This leads to the question: Will MSE_{nf} be reduced if the mode coefficients are set to values between those in the near- and far-fields?

An arbitrarily scaled input Gaussian and top-hat were selected, and the first 20 even mode coefficients were used by the GSA to propagate between the near- and far-fields. Figure 3.41 *left* shows a plot of MSE_{nf} using the mean of the amplitude and phase of the mode coefficients in the near- and far-field, CN_n and CF_n. A random initial phase was used. MSE_{nf} at the final iteration of the GSA has indeed been reduced - in this case by a factor of 2.4. Figure 3.41 *right* shows plots of the GSA MSE_{nf}, when two different sets of random initial phases were used. The

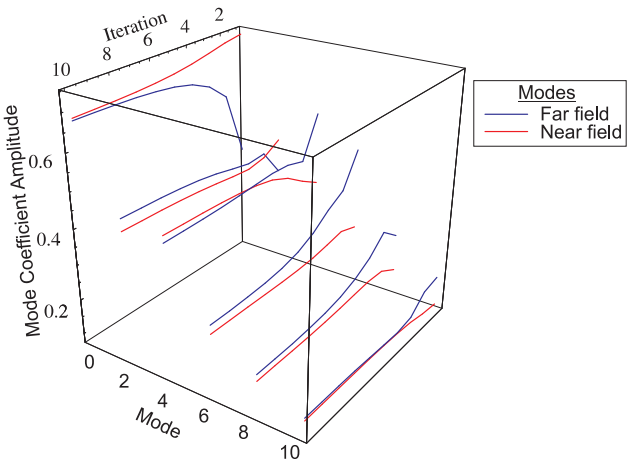


Figure 3.40: Even mode coefficients in the range $0 \rightarrow 10$ are used to propagate between the near- and far-fields, for a target top-hat distribution. The plot shows the value of the amplitude of the coefficients as the GSA progresses. The GSA is started with random values for the mode coefficients. The near- and far-field mode coefficients converge as the GSA progresses.

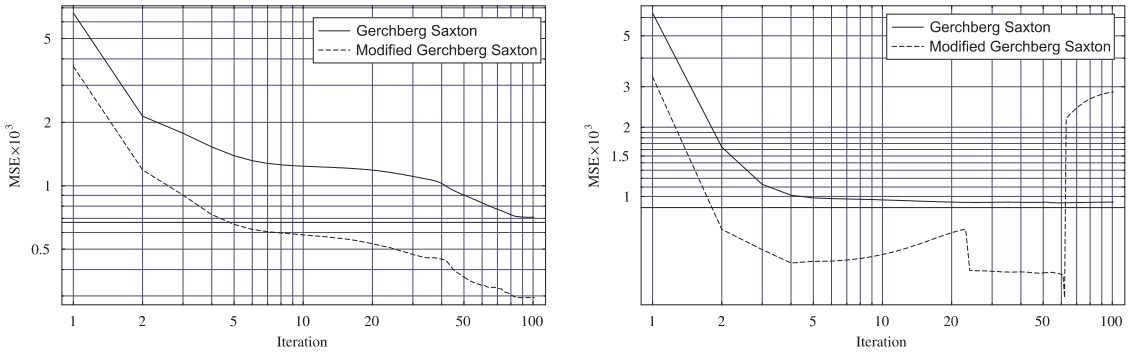


Figure 3.41: *Left* plot of $\mathrm{MSE_{nf}}$ for the GSA, propagating fields using the first 40 even GBMs, converting a Gaussian input to a top-hat field distribution. The GSA is modified by using mode coefficients which are the average of the near-field and far-field coefficients. *Right* for this random input phase, averaging the phase of the mode coefficients in the near- and far-fields returns a spurious phase coefficient and results in $\mathrm{MSE_{nf}}$ increasing above that of the unmodified technique.

plot on the left shows the modified GSA has a smaller $\mathrm{MSE_{nf}}$ than the unmodified GSA during this evaluation. For the plot on the right, $\mathrm{MSE_{nf}}$ of the modified GSA jumps above that of the unmodified. This occurred because of phase wrapping of either the near- or far-field modes, giving a spurious value for the coefficient's phase. This can be rectified by either taking averages of the real and imaginary parts of the coefficients, or unwrapping their phase first before calculating the mean. When this is done the modified GSA consistently gave a significantly lower a $\mathrm{MSE_{nf}}$ than the unmodified GSA, for all trials attempted. However, although the average MSA has been reduced, the far-field from a DOE given by the FFT of the phase of the GBM reconstruction using the average of $\mathrm{CN_n}$ and $\mathrm{CF_n}$ is of poorer quality than the far-field where the modes $\mathrm{CN_n}$ are used to give the phase of the DOE.

3.3.3 Example: Top-Hat

The optimisation of the initial phase was unsuccessful, and so the author has resorted to a "brute force" method of using multiple random initial trial phases, as used in Chapter 2. The amplitude and phase of the coefficients for the initial field of the GSA were given random values in the range $0 \rightarrow 1$ for the amplitude, and $0 \rightarrow 2\pi$ for the phase of the mode coefficients.

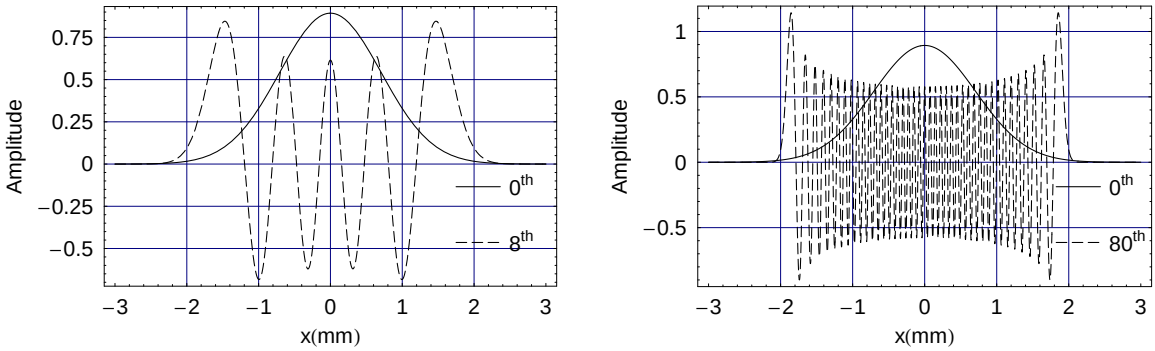


Figure 3.42: GMSO finds values of w_0 such that the highest mode has approximately the same extent as the input Gaussian amplitude distribution. *Left* an example where the highest order mode is the 8^{th}, and this is shown with the 0^{th} order (Gaussian) input field. *Right* 80 modes were used to propagate the field in the GSA using GBMSO. The value of w_0 is chosen such that the highest order has the same extent as the 0^{th} order input field.

Section 3.2.2 shows an example of a Gaussian transformed to a top-hat using GBMSO, where the radius of the target top-hat, r_f, and the radius of the modes, w_0, used in the optimisation are free parameters. Figures 3.7 *left* and *right* show the values found for r_f and w_0 as a function of the number of modes used in the optimisation. r_f was independent of the number of modes used, and the values of r_f found by the optimisations were approximately equal to the radius of the input Gaussian. The optimisations find values of w_0 such that the extent of the highest

order mode is similar to the extent of the input Gaussian. Equation 3.11 shows that the scale of the far-field distribution in angular coordinates is proportional to r_f/w_0^2. Therefore, by using a low number of modes, the GBMSO finds a solution where the far-field pattern has a small scale in terms of its angular extent.

For this example, GBMSO was first used to find optimal values of w_0 and r_f with 8 modes. The optimisation found a value of w_0 for which the 8^{th} order mode was similar to the extent of the input Gaussian distribution, as shown in Figure 3.42 *left*. Where the w_{in} is the radius of the input Gaussian, with 8 modes GBMSO found a solution where $r_f = w_{in}$ and $w_0 = 0.58 \times w_{in}$, giving $r_f/w_0^2 = 3.0$ (which is proportional to the far-field top-hat radius given as a divergence angle (θ_f) by Equation 3.11). The GSA was then used to propagate the field composed of 8 modes with these values of r_f and w_0. Using the same value of w_0 and r_f and the same number of modes in both the GSA and the GBMSO allows direct comparison of the mode coefficients and near- and far-field phase distributions that each found. The GSA was evaluated for 1000 trials, each with 50 iterations, and with a new set of random coefficients to define the initial phase for each trial. Figure 3.43 shows the amplitude and phase of the mode coefficients (*top left* and *top right*) and near- and far-field amplitude distributions (*middle left* and *middle right*) of two solutions returned by the 1000 trials, where one field is the complex conjugate of the other. As was the case where FFTs were used to propagate the field in Chapter 2, the solution and its complex conjugate were returned, so essentially there was only one unique solution. The results of the GBMSO for the same near- and far-field target distributions and the same value of w_0 are shown with those of the GSA; the mode coefficients found by GBMSO lay between both the amplitude and phase of the near- and far-field mode coefficients, CN_n and CF_n (given by Equations 3.16 and 3.18), given at the last iteration of the GSA. Figures 3.43 *middle left* and *middle right* show both the amplitude of the reconstruction of $T_n(x)\exp(i\phi_{DOE,n}(x))$ and $T_f(x)\exp(i\phi_{DOE,f}(x))$, and the amplitude $E_n(x)$ and $E_f(x)$. The reconstruction of $T_n(x)\exp(i\phi_{DOE,n}(x))$ and especially $T_f(x)\exp(i\phi_{DOE,f}(x))$ is poor. With a large number of modes (and w_0 fixed to the same value to achieve the same scale of far-field distribution), $T_n(x)\exp(i\phi_{DOE,n}(x))$ and $T_f(x)\exp(i\phi_{DOE,f}(x))$ can be reconstructed exactly. However, when the higher number of modes was used to propagate the field in the GSA, the GSA returned phase distributions (shown in Figures 3.43 *bottom left* and *bottom right*) that were very similar to those returned when only 8 modes were used to propagate the field. Using more than 8 modes to propagate the field in the GSA increased the amount of off-axis power in the far-field, E_f, but did not increase the number of solutions found. Apart from an offset, the phase in the near- and far-field given by the sum of modes using both CN_n and CF_n GSA and the GBMSO are identical at each plane over the region

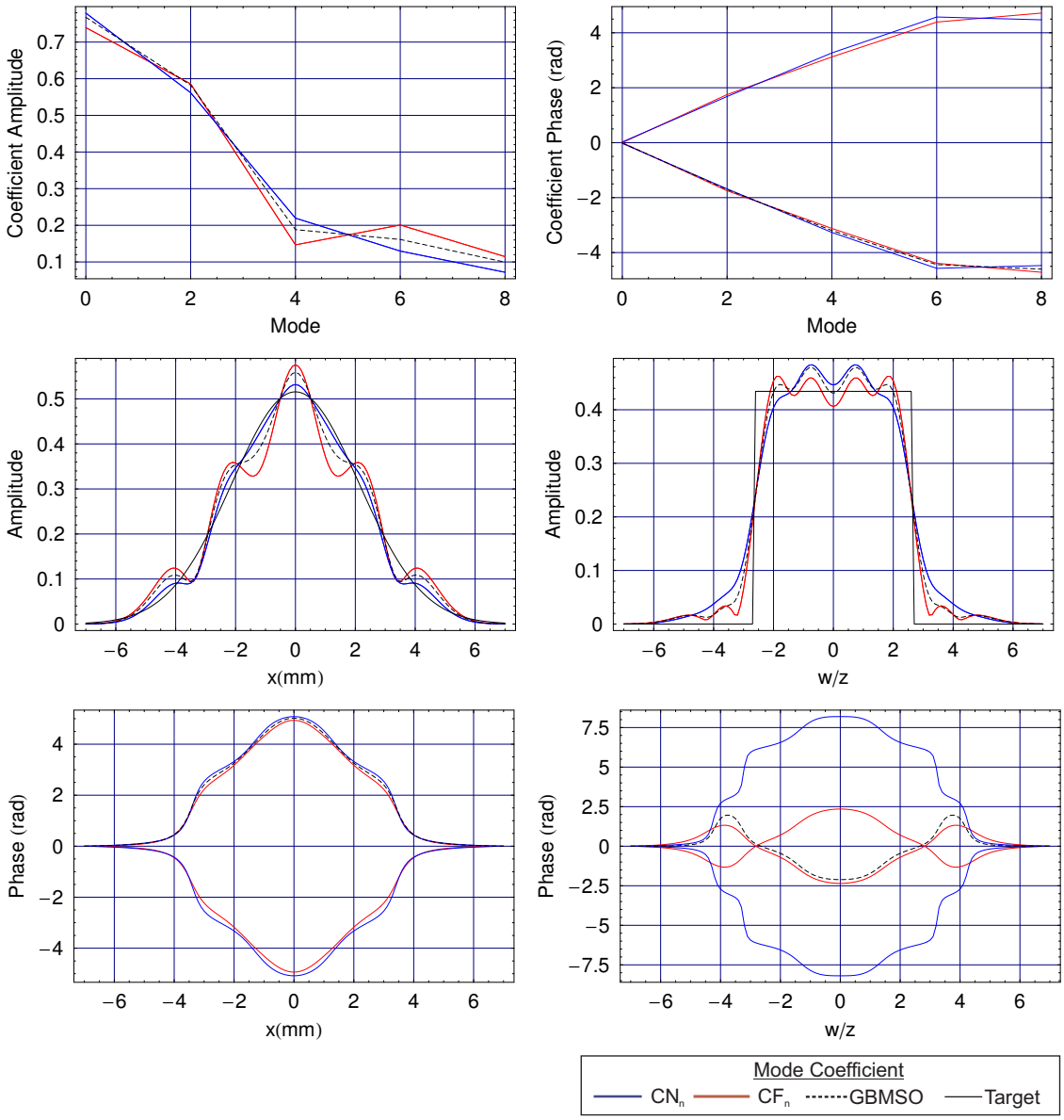


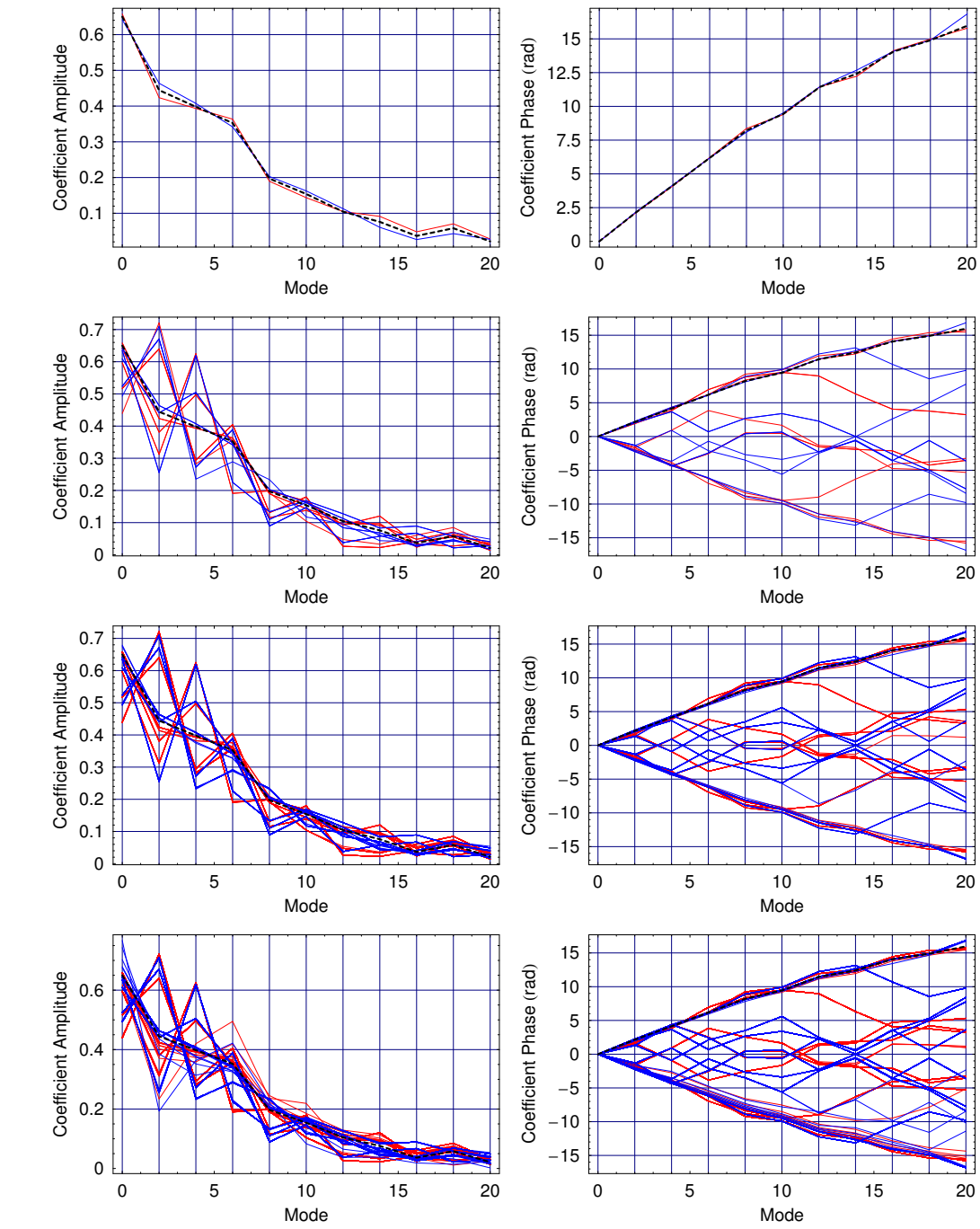
Figure 3.43: The output from the GSA for a Gaussian to top-hat transform, using the first 8 even modes to propagate the field, and 1000 random initial phases, shown with the results of the GBMSO, which used the same value of r_f, w_0 and the same number of modes. The plots show the near- and far-fields reconstructed with both the near- and far-field mode coefficients at each plane. *Top* the mode coefficients, *middle* the amplitude in *left* the near-field and *right* the far-field, *bottom* the phase in *left* the near-field giving the DOE profile and *right* the far-field. The field given by the modes CN_n is the reconstruction of $T_n(x)\exp(i\arg[E_n(x)])$ in the near-field, and $E_f(x)$ in the far-field. The field given by the modes CF_n is the reconstruction of $T_f(x)\exp(i\arg[E_f(x)])$ in the far-field, and $E_n(x)$ in the near-field.

were there is significant power in the field. For an iteration of the GSA to make an improvement, a significant change in the phase of $E_n(x)$ over the extent of $T_n(x)$, and in the phase of $E_f(x)$ over the extent of $T_f(x)$ from the previous iteration must be made (a change in phase outside the extent of the target distribution has no effect on the field at the conjugate plane, propagated using Equations 3.17 and 3.19). However, with this large value of w_0, the high order modes contribute very little power to $E_n(x)$ and $E_f(x)$ over the extent of $T_n(x)$ and $T_f(x)$ respectively, because the extent of the modes is much larger than the target distributions. When the field is propagated to the opposite plane, most of the power in the higher order modes is contained outside the extent of the target distribution. Therefore, the higher order modes are less important for a small scale of far-field target distribution.

The Gaussian to top-hat, transform was carried out again with r_f equal to the radius of the input Gaussian, and the number of modes increased to even modes in the range $0 \rightarrow 20$, and the value of w_0 set to the same value as that found by the GBSMO, $w_0 = 0.4 \times w_{in}$. This smaller value of w_0 gave a far-field distribution with a larger angular extent, with $r_f/w_0^2 = 6.3$, than the previous transform where $0 \rightarrow 8$ modes were used. Figure 3.44 shows the solution returned from the GSA for 1000 trials with 200 iterations in each. Instead of just two possible solutions, as was the case when $0 \rightarrow 8$ modes were used to propagate the field, a number of different solutions are returned by the GSA. Figure 3.45 *right* shows that $\mathrm{MSE_{nf}}$ ordered from lowest to highest is not continuous, due to different initial phases giving the same solution. Figure 3.44 shows plots for the best 500 trials. Increasing the number of trials included resulted in only a small number of additional solutions, as many are identical.

The Gaussian to top-hat transform was carried out with the number of modes used to propagate the field increased to even modes in the range $0 \rightarrow 80$, and the value of w_0 set such that the 80^{th} order mode was similar in extent to the extent of the input Gaussian, as shown in Figure 3.42 *right*. With a small value of w_0, the far-field distribution is larger in term of its angular extent. Figure 3.46 *right*, shows that the ordered histogram of the MSE of the final iteration of the trials is continuous, as each trial arrives at a different solution. Figure 3.47 shows that there is a larger variation in the quality of the solutions, but that the amplitude distribution of the best solution (shown in black) is closer to the target than when 8 or 20 modes were used.

We have seen that using a smaller value of w_0 enables the mode set to better reconstruct the target distributions. By using a large value of w_0, which gives a small scale for the far-field amplitude distribution, the target field can be reconstructed, but only with a large number of modes. However, when these modes are propagated to the conjugate plane, each high order mode does not contribute a significant



(a)	(b)
(c)	(d)
(e)	(f)
(g)	(h)

Figure 3.44: Plots of the mode coefficient values in the near- and far-field for the GSA using 20 modes and 1000 trials, shown with the modes found by GBMSO for comparison. The modes coefficients found by the GBMSO lie between the near- and far-field mode coefficients of the trial with the lowest MSE. The phase of the fields given by the mode coefficients of both the GSA and GBMSO are identical over the region of the target distributions in the near- and far-fields. The trials were ordered from best (lowest MSE) to worst (highest MSE) (a) & (b) show the best trial, (c) & (d) the best 10 trials, (e) & (f) the best 100 trials and (g) & (h) the best 500 trials.

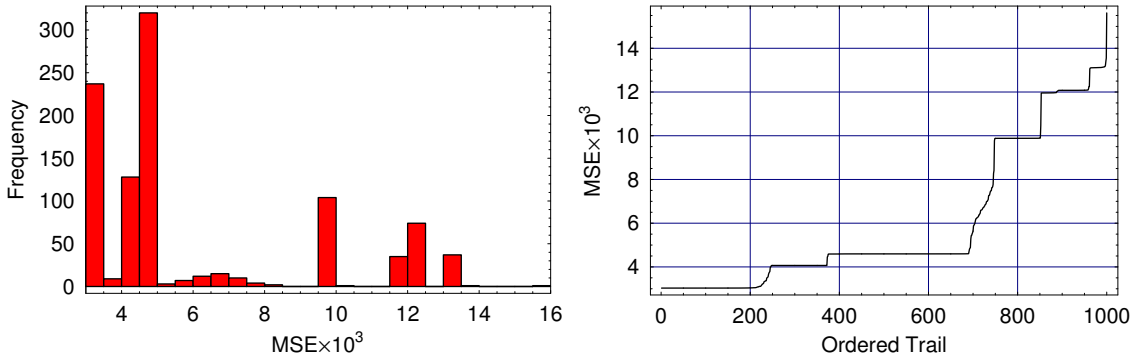


Figure 3.45: MSE_{nf} of the output from the GSA for 1000 trials, where 20 modes were used to propagate the field. *Left* histogram of MSE *right* MSE ordered from smallest to largest MSE.

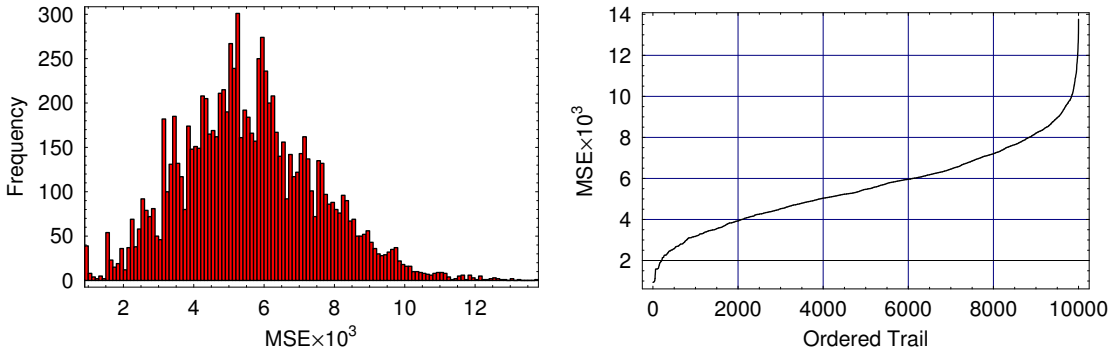


Figure 3.46: The MSE of the output from the GSA for 1000 trials, each with 200 iterations, using 0-80 even Gaussian-Hermite modes. *Left* histogram of MSE values. *Right* MSE ordered from smallest to largest MSE.

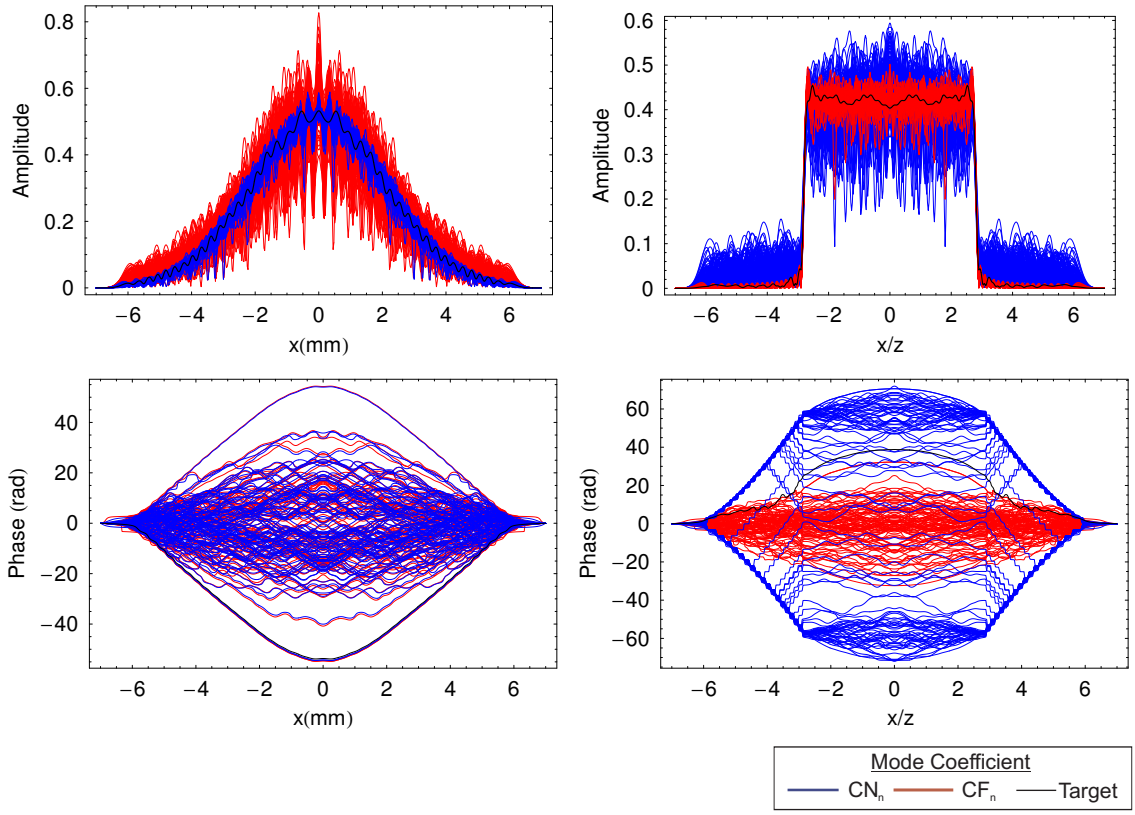


Figure 3.47: The near- and far-field amplitude and phase of the solutions given by the GSA for a Gaussian to top-hat transform using the first 80 Gaussian-Hermite modes and 1000 trials, each with 200 iterations. The black line shows the solution with the lowest MSE. *Top:* the amplitude in *left* the near-field *right* the far-field. *Bottom* the phase in *left* the near-field giving the DOE profile *right* the far-field . The field given by the modes CN_n is the reconstruction of $T_n(x)\exp(i\arg[E_n(x)])$ in the near-field, and $E_f(x)$ in the far-field. The field given by the modes CF_n is the reconstruction of $T_f(x)\exp(i\arg[E_f(x)])$ in the far-field, and $E_n(x)$ in the near-field. There is a large variation in the quality of the solutions returned by the GSA.

amount of power to the field in the region of the target distribution, due to their large extent. Much of the power of $E_f(\theta)$ is therefore contained outside of the target distribution, and the higher order modes have little effect on the phase over the extent of the target distributions. As a significant deviation in the phase is necessary to improve the solution of the GSA, it is unable to reduce the MSE further within a relatively small number of iterations. With a small value of w_0, giving a large scale for the far-field amplitude distribution, a high number of modes which have a similar extent to the target distributions can be used to reconstruct them, and when these modes are propagated to the conjugate plane their extent is still similar to the target distribution at that plane. Less of the power of $E_n(x)$ and $E_f(x)$ therefore falls outside of $T_n(x)$ and $T_f(x)$, giving less unwanted off-axis power. When a large number of modes contribute to the power in $E_n(x)$ and $E_f(x)$ over the extent of $T_n(x)$ and $T_f(x)$, $\phi_{DOE,n}(x)$ and $\phi_{DOE,f}(x)$ are more likely to vary with each iteration, and reduce the MSE; although, with the higher number of modes, the optimisation is more nonlinear, and the GSA may stagnate for a number of iterations. And with a larger number of modes contributing a significant amount of power to $E_n(x)$ and $E_f(x)$, it is more likely that a combination of mode coefficients can be found that describe the target distributions accurately.

3.3.4 The Operation of GBMSO Compared to the GSA

Whereas with each iteration of the GSA all of the MSE arises either at the near-field or the far-field plane, with GBMSO the total error is split between the two planes, as this minimises the error at both planes. Provided that a sufficient number of modes are used in the optimisation, GBMSO finds the same phase distribution as the GSA, despite $\left|E_n(x)\right|$ and $\left|E_f(\theta)\right|$ (the amplitude distributions given by the sum of modes in the near- and far-field) being significantly different from the amplitude distributions of the GSA solutions at the same planes. Section 3.3.3 shows an example where both the GSA and GBMSO optimise using 8 modes that have the same radius, allowing a direct comparison between the phase distributions that each find; both optimisations find the same phase distribution over the extent of the target distribution.

Here, the operation of GBMSO is analysed for an input field which has an amplitude distribution $T_n(x)$ and a flat phase-front and a target far-field $T_f(\theta)$. When the DOE designed using GBMSO is used to transform the field $T_n(x)$ to the far-field, the phase $\phi_{DOE,n}(x) = \arg[E_n(x)]$ is imposed on $T_n(x)$, to give the far-field as follows:

$$R_f(\theta) = \mathcal{F}\{T_n(x)\exp[i\phi_{DOE,n}(x)]\}. \tag{3.20}$$

$T_n(x)\exp[i\phi_{DOE,n}(x)]$ can be given by $E_n(x)$ plus the error, where the error is

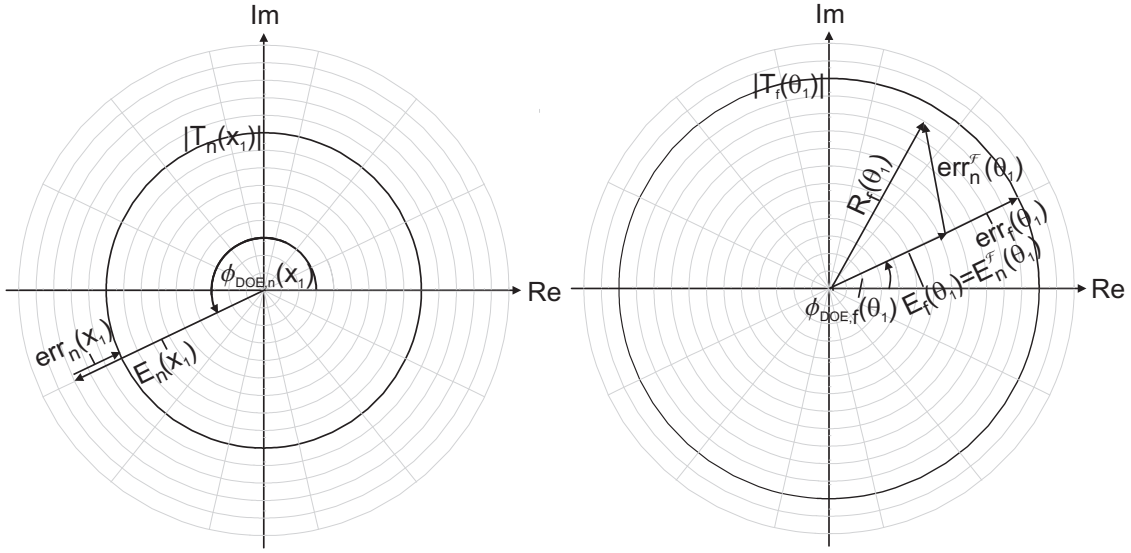


Figure 3.48: The near-field phase of the solution, $\phi_{DOE,n} = \arg[E_n(x)]$, given by the optimisation, is used to transform the input field which has an amplitude distribution $T_n(x)$ and a flat phase-front to the far-field to give the field $R_f(\theta)$. *Left* the optimisation finds a field with and amplitude distribution $|E_n(x)|$. The "error" at the near-field plane, $err_n(x)$, is the difference between the $|E_n(x)|$ and the target field at the same plane, $T_n(x)$. *Left* $err_n(x)$ is necessarily colinear with $E_n(n)$. *Right* for a *suboptimal* set of mode coefficients $err_n(x)$ and $E_n(n)$ transformed to the far-field to give $err_n^{\mathcal{F}}(x)$ and $E_n^{\mathcal{F}}(n)$ are out of phase as shown. However, when the field is optimised, these two components are colinear, as shown in Figure 3.49.

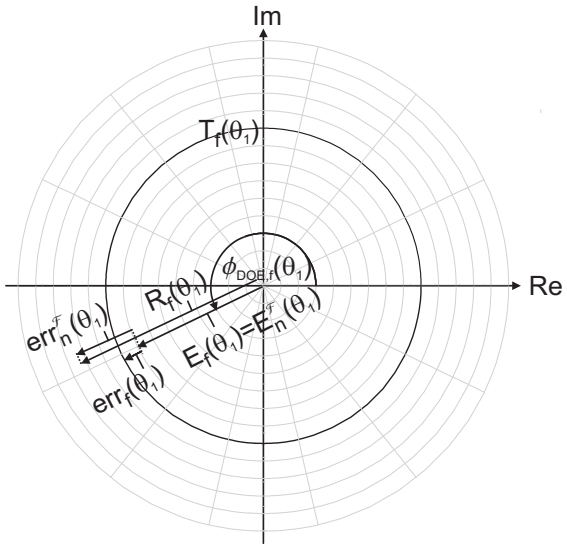


Figure 3.49: Where $\phi_{DOE,n}$ is the field given by an *optimal* set of mode coefficients, the components $err_n^{\mathcal{F}}(x)$ and $E_n^{\mathcal{F}}(n)$ are colinear when transformed to the far-field.

given by the difference between the target and obtained fields, as follows:

$$\mathrm{err_n}(x) = T_n(x)\exp[i\phi_{DOE,n}(x)] - E_n(x). \tag{3.21}$$

The far-field is then given by

$$\begin{aligned} R_f(\theta) &= \mathcal{F}\{T_n(x)\exp[i\phi_{DOE,n}(x)]\} \\ &= \mathcal{F}\{E_n(x) + \mathrm{err_n}(x)\} \\ &= \mathcal{F}\{E_n(x)\} + \mathcal{F}\{\mathrm{err_n}(x)\}. \end{aligned} \tag{3.22}$$

Figure 3.48 shows the transform from near- (*left*) to far-field *right* for a point x_1 in the near-field and θ_1 in the far-field. At the near-field plane, $E_n(x)$ and $\mathrm{err_n}(x)$ are necessarily colinear. Both GBMSO and the GSA minimises the error and find a phase where $E_n(x)$ and $\mathrm{err_n}(x)$ are still colinear over the extent of the target distributions when transformed to the far-field, as shown in Figure 3.49.

3.4 Phase and Absorption Gratings

Thus far, the DOEs have been restricted to phase only. Here, DOEs will be investigated that both impose a phase on the input field and absorb some of the radiation across the field, to give the required amplitude distribution. This amplitude and phase freedom allows any amplitude distribution to be achieved in the far-field, but necessitates the loss of some of the power of the beam. GBMSO allows near-field solutions to be found with a specified trade-off between the power absorbed and the accuracy of the transform to the target far-field profile.

For the GSA where FFTs are used to propagate the field, if the amplitude of the near-field solution, $|E_n(x)|$ (rather than the given input distribution, $T_n(x)$), is used as the input, then the exact target $T_f(\theta)$ will be produced in the far-field. Given that the DOE is designed for an input field with a Gaussian distribution, in addition to imposing a phase on the field given by $\phi_{DOE,n}(x) = \arg[E_n(x)]$, some of the input field must also be absorbed, so that its amplitude distribution is $\left|E_n(x)\right|$. Figure 3.50 shows a Gaussian to top-hat transformation, in which $E_n(x)$ is normalised such that $\left|E_n(x)\right|$ is equal to $\left|T_n(x)\right|$ for at least one point (here at $x = 0$), so that there is no unnecessary loss of radiation[22].

In GBMSO the merit function can be weighted to give a solution that is similar to $\left|T_n(x)\right|$ - rather than a solution with the amplitude distribution given by the GSA, $\left|E_n(x)\right|$, which would give the exact target distribution, $T_f(\theta)$, but would necessitate a larger loss of power. Absorption can then be used to give a field at the DOE with an amplitude and phase equal to the amplitude and phase of the near-field GBMSO solution.

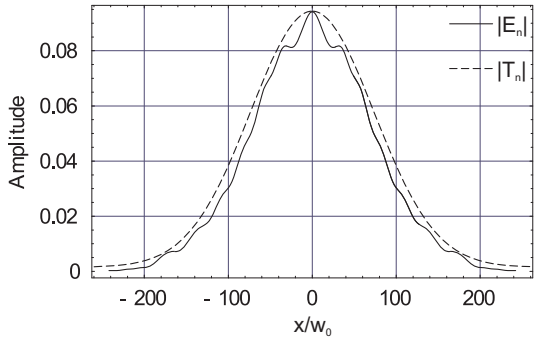


Figure 3.50: Some of the input field $\left|T_n(x)\right|$ is absorbed to give $\left|E_n(x)\right|$ where everywhere $\left|E_n(x)\right| \leq \left|T_n(x)\right|$, in order to achieve an exact far-field transform.

If the merit function is weighted with a constant, α_n, as follows: $\alpha_n\eta_n + \eta_f$, where $\alpha_n > 1$, $\int_{-\infty}^{\infty} \mathrm{err_n}(x)dx$ is reduced. In the limit where α_n provides sufficient weight, $\int_{-\infty}^{\infty} \mathrm{err_n}(x)dx$ is reduced to zero, and therefore, $\int_{-\infty}^{\infty} \mathrm{err_n^{\mathcal{F}}}(\theta)d\theta$ is also zero, provided that a sufficient number of modes are used to describe the near-field. For the solutions returned by GBMSO, $\phi_{DOE,n}(x)$ and $\phi_f(\theta)$ have been found to be independent of α_n; this is the case because $\mathrm{err_n^{\mathcal{F}}}(\theta)$ has the same phase distribution as $E_f(\theta)$, and $\mathrm{err_f^{\mathcal{F}}}(x)$ has the same phase distribution as $E_n(x)$, so $\phi_{DOE,n}(x)$ is the same regardless of the weighting (Section 3.6.1 shows that the coupling between two fields is maximised if they have the same phase). Using GBMs to propagate the field in the GSA returns the near-field mode coefficients for the near-field, $\mathrm{CN_n}$, and for the far-field, $\mathrm{CF_n}$. Provided there are a sufficient number of modes used, the field given by the reconstruction with coefficients $\mathrm{CN_n}$ constructs the target near-field exactly, and similarly the far-field reconstruction with coefficients $\mathrm{CF_f}$ constructs the target far-field exactly. GBMSO finds a solution between $\mathrm{CN_n}$ and $\mathrm{CN_f}$. If a sufficiently large value of α_n is used to weight the merit function, $\alpha_n\eta_n + \eta_f$ in the GBMSO, the mode coefficients returned by the GBMSO are equal to the near-field mode coefficients returned by the GSA (i.e. $A_n = \mathrm{CN_n}$). By weighting the merit function used in GBMSO, a continuum of solutions may be returned between $\mathrm{CN_n}$ and $\mathrm{CN_f}$, where, provided a sufficient number of modes have been used, the reconstruction with modes $\mathrm{CN_n}$ gives a field with an amplitude distribution equal to $T_n(x)$ and the reconstruction with $\mathrm{CF_n}$ gives a field with an amplitude distribution equal to $T_f(x)$. Increasing the weight of α_n reduces $\int_{-\infty}^{\infty} \mathrm{err_n}(x)^2dx$, while increasing $\int_{-\infty}^{\infty} \mathrm{err_f}(x)^2dx$. Regardless of the weight used, $\mathrm{err_n^{\mathcal{F}}}(\theta)$ is collinear with $E_n^{\mathcal{F}}(\theta)$ at each point across the field. This has a practical application where the DOE can both impose a phase on the field and control the transmission across it, i.e. the DOE has a complex refractive index[25]. Using the field where η_f is weighted with a large value gives a far-field amplitude distribution, $\left|E_f(x)\right|$, which is equal to $\left|T_f(\theta)\right|$ and the near-field amplitude distribution, $\left|E_n(x)\right|$, is poorer approximation to $\left|T_n(x)\right|$. The DOE would impose the phase, $\arg[E_n(x)]$, on the field and absorb some of

the field incident on the DOE with an amplitude distribution $T_n(x)$ to achieve an amplitude distribution given by $\left|E_n(x)\right|$. There is a trade-off between the energy that is absorbed by the DOE and the accuracy of the transform. By weighting the merit function this trade-off can be controlled. GBMSO has the advantage over the GSA using FFTs that the spatial extent of the near- and far-field solutions is limited - whereas, the GSA using FFTs gave a solution, $E_n(x)$, with some of the power outside the target distribution, due to the arbitrarily high spatial frequency component of the far-field[25]. Therefore, when absorbing the amplitude to convert $\left|T_n(x)\right|$ to $\left|E_n(x)\right|$, some of the field is truncated; whereas, with GBMSO the spatial extent of the near- and far-field distributions is limited by the number of modes used in the optimisation, and less of the power can occur outside of the DOE aperture.

3.5 Beam Shaping Using Multiple Surfaces

It was demonstrated above how a DOE can be used to modify the amplitude and phase of an input field to give a desired output field. Absorbing some of the input field necessitates the loss of some of the power. If $|E_n(x)|$ given by the GBMSO deviates sufficiently from $T_n(x)$, as was the case in the beam splitting example, a large proportion of the input field needs to be absorbed in order to achieve the conversion $\left|T_n(x)\right| \rightarrow \left|E_n(x)\right|$. Here, two successive phase imposing surfaces are used for beam shaping in order to create a field that converts $T_n(x)$ to $T_f(\theta)$ exactly, without any loss of power.

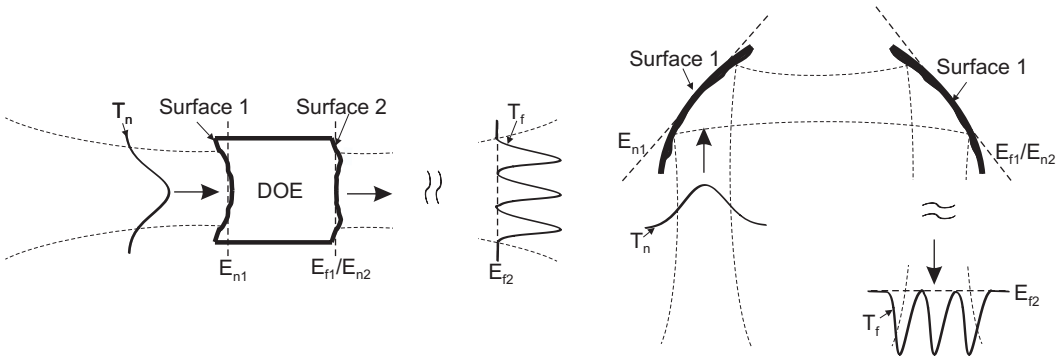


Figure 3.51: Beam shaping using two surfaces: *Left* in transmission, using a dielectric material and *right* in reflection, using mirrors. The figure shows an example of a DOE that converts an incident Gaussian beam to three beams in the far-field.

The beam shaping could be done in two stages: the first DOE would be used to convert the input field, $T_n(x)$, to the amplitude distribution $\left|E_n(x)\right|$ (rather than $\left|E_f(\theta)\right|$) at the second phase imposing surface. The second DOE would then impose a phase on its input field, $E_n(x)$, to give $\left|E_f(\theta)\right| = T_f(\theta)$ in its far-field. The problem with this approach is that the first DOE may not be able to convert $T_n(x)$ to $\left|E_n(x)\right|$ in the same way as it was not possible to convert $T_n(x)$ to $\left|E_f(\theta)\right|$. The approach

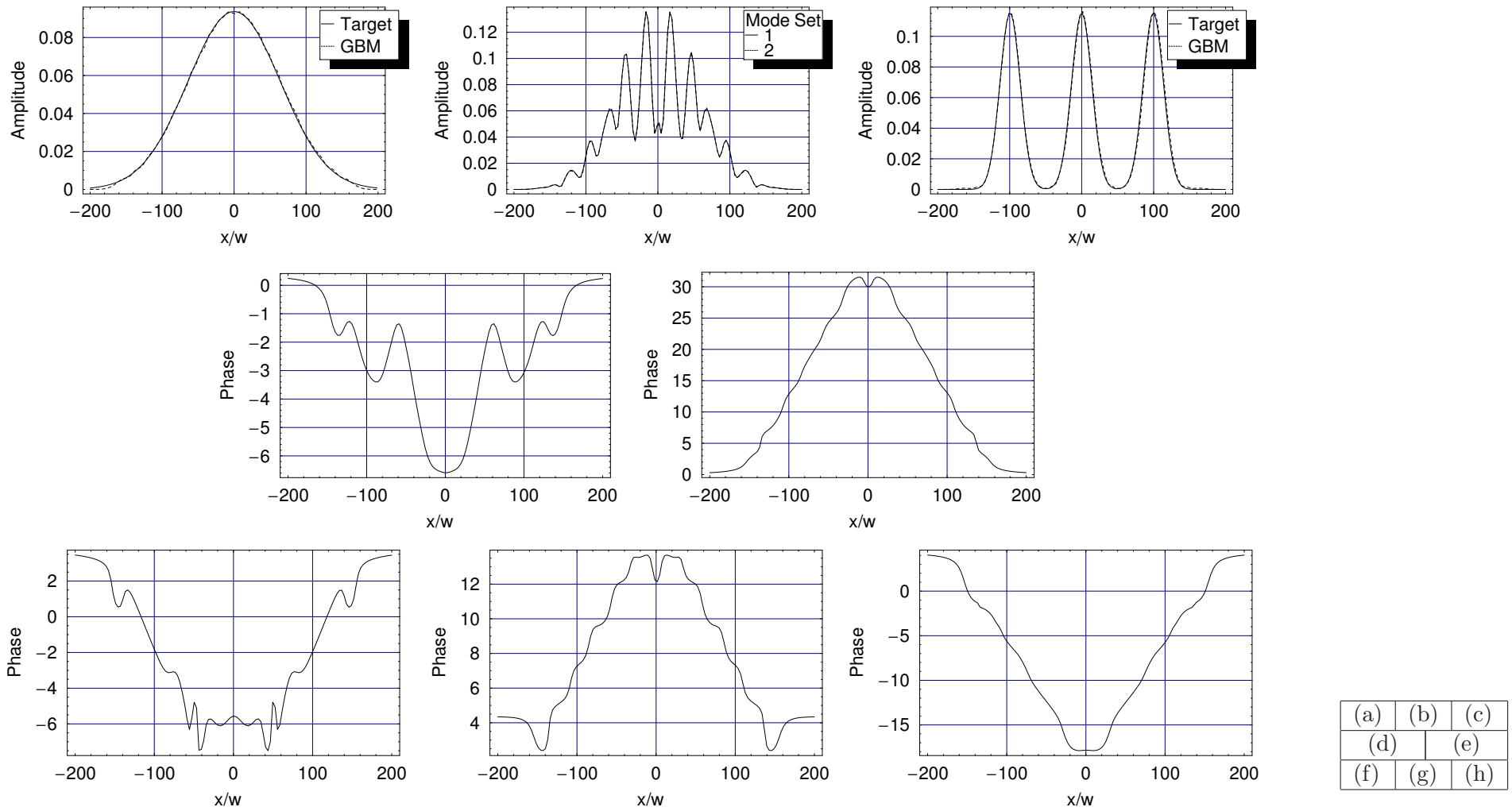


Figure 3.52: The amplitude at (a) surface 1, (b) surface 2 and (c) the output. The phase at (d) surface 1, (e) the output. (f) and (g) are the phase distributions of the far and near-fields of mode sets 1 and 2 respectively. (h) is the phase required at surface 2 to obtain the amplitude distribution of (c).

taken here, shown in Figure 3.51, is to allow the first DOE to create *any* far-field amplitude which, when used as the input to the second DOE, can be used to create the far-field pattern $T_f(\theta)$. Figure 3.51 shows how beam shaping using two surfaces is achieved with on-axis or off-axis elements.

To achieve the desired profile, instead of using the single mode set as was used previously in GBMSO, two mode sets are now optimised. E_{n1} and E_{f1} are the near- and far-fields of mode set 1 and E_{n2}, and E_{f2} are the near- and far-fields of mode set 2. With reference to Figure 3.51, E_{n1} is the field at the input, surface 1, and E_{f1} is the field at surface 2; E_{n2} is also located at surface 1, and E_{f2} is the field at the output, i.e. in the far-field of surface 2. The merit function, which is maximised using DE, is given by

$$\text{merit} = \bar{\eta}(|E_{n1}|, T_n) + \bar{\eta}(\left|E_{f1}\right|, |E_{n2}|) + \bar{\eta}(\left|E_{f2}\right|, T_f). \tag{3.23}$$

The merit function is equal to 3, when the target input function is equal to the near-field of mode set 1, the amplitude profile of the far- and near-field of mode sets 1 and 2, respectively, are equal, and the far-field amplitude of mode set 2 is equal to the target output amplitude distribution.

The method is used to convert a single Gaussian beam to three equally spaced Gaussian beams. The target far-field distribution was the same as that using a single set of GBMs to convert a Gaussian to three beams, shown in Section 3.2.7. The width of the input and output target fields are parameters which were also optimised. Using the first even 30 GBMs the optimiser found the solution given in Figure 3.52. DE was used to optimise the two mode sets and it achieved a merit function equal to 3. The near-field amplitude distribution of mode set 1, $\left|E_{n1}(x)\right|$, was exactly equal to $\left|T_n(x)\right|$ and the far-field of mode set 2, $\left|E_{f2}(x)\right|$, was exactly equal to the target amplitude distribution $\left|T_f(x)\right|$, the amplitude profile of mode set 1 and 2 matched exactly at the intermediate plane. Using this second surface introduced an extra degree of freedom that allowed an exact solution to be found.

3.6 Beam Control

Thus far in this chapter, the input field was transformed to the prescribed far-field pattern, without regard to the evolution of the field in the intermediary, near-field, region. In this section, the distribution of the field is considered as it propagates in the near-field region, and DOEs are designed to control it by extending the GBMSO algorithm. The purpose of this section is to demonstrate the flexibility of the GBMSO algorithm, which allows novel DOE designs to be created.

3.6.1 Mode Converter

Higher order unimodal Gaussian-Hermite and Gaussian-Laguerre beams have applications in optical systems, because they keep the same profile, only changing their scale as they propagate in free space[22]. For example, first order Gaussian-Hermite beams are employed as 3D optical tweezers, in which dielectric particles are manipulated by means of radiation pressure, scattering force or dipole force[38, 39, 40]. Previously in this chapter, a 1st order Gaussian-Hermite amplitude distribution was created in the far-field of a DOE; however, the phase distribution differed from a 1st order Gaussian-Hermite, and so the beam diffracted more rapidly than would a 1st order Gaussian-Hermite mode - where the phase is flat at the waist and spherical down-beam. The phase distribution may be corrected by a second DOE, which imposes a phase on the beam in the far-field[25]. The Gaussian-Hermite-type mode would then propagate as a pure 1st order Gaussian-Hermite, as defined in Gaussian beam mode theory - with its beam radius expanding as it propagates with distance along the z axis from the second phase flattening DOE, according to Equation 3.8. Here, in order to reduce the complexity of the system, only one DOE is used to transform the Gaussian to a first order Gaussian-Hermite, as shown in Figure 3.53.

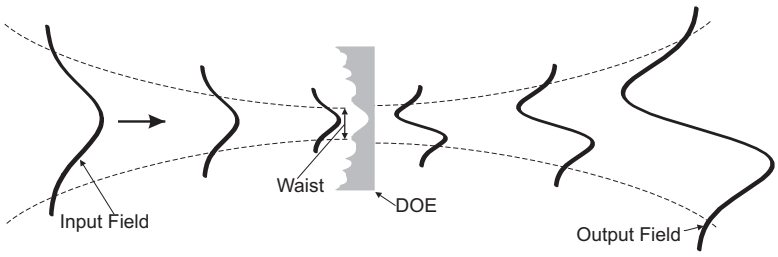


Figure 3.53: DOEs are designed to convert the input field to Gaussian-Hermite modes.

The DOE is designed by maximising the coupling, given by

$$\bar{\eta}(w) = \left| \int_{-\infty}^{\infty} \psi_0(w, x, 0)[\exp(i\phi_{\mathrm{DOE}}(x))\psi_n^*(w, x, z = 0)]dx \right| \tag{3.24}$$

where $\psi_0(w, x, 0)$ is the Gaussian input field with a flat phase, $\psi_n(w, x, z = 0)$ is the Gaussian-Hermite mode of index n and beam radius w, and $\phi_{\mathrm{DOE}}(x)$ is the phase distribution imposed by the DOE. Either the input or the output may be scaled with x, to allow more freedom to maximise $\bar{\eta}(w)$. Note that the field is not propagated before $\bar{\eta}(w)$ is calculated, as was done previously. The coupling integral can be performed at any plane, and therefore, it is not necessary to propagate the field before $\bar{\eta}$ is calculated. $\bar{\eta}^2$ gives the amount of power converted into the Gaussian-Hermite mode of order n. 100% power conversion of one mode to another can never be achieved with just one phase-only DOE, because only the phase of the input field

is altered. For 100% power conversion, the phase *and* amplitude of input and output fields must be matched (requiring the use of two phase imposing surfaces, as shown in Section 3.5); however, it is typically possible to scatter a significant amount of power from the input to the target output mode.

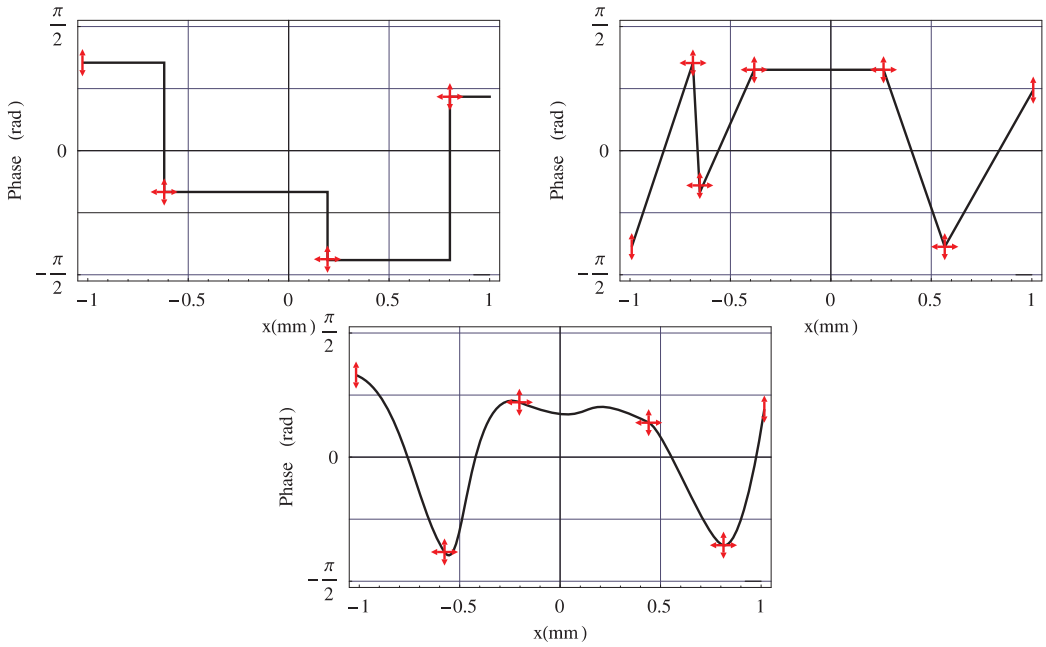


Figure 3.54: Examples of phase interpolation schemes used, with polynomial interpolation of order: *top left* 0, *top right* 1, *bottom* 3.

In Equation 3.24, the phase, $\phi_{\mathrm{DOE}}(x)$, is optimised, in order to maximise $\bar{\eta}$. In order to reduce the number of parameters to be optimised, the phase is defined at a small number of points, and $\phi_0(w, x, 0)$ is determined by interpolating between these points. A number of phase interpolation schemes, shown in Figure 3.54, were considered. The position of the points indicated by red crosses were optimised, in order to maximise coupling with the target field, using Equation 3.24. The phase between the points is found by interpolating using polynomials of order 0, 1 and 3. Using polynomial interpolation of order 0 proved to be most efficient in reducing the time it took for the optimisation algorithm to converge.

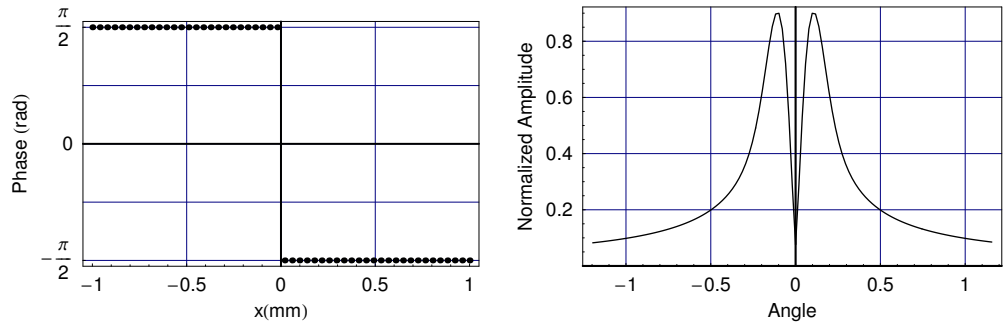


Figure 3.55: *Left* the phase distribution to convert a Gaussian to a 1st order Gaussian-Hermite, *right* the amplitude in the far-field.

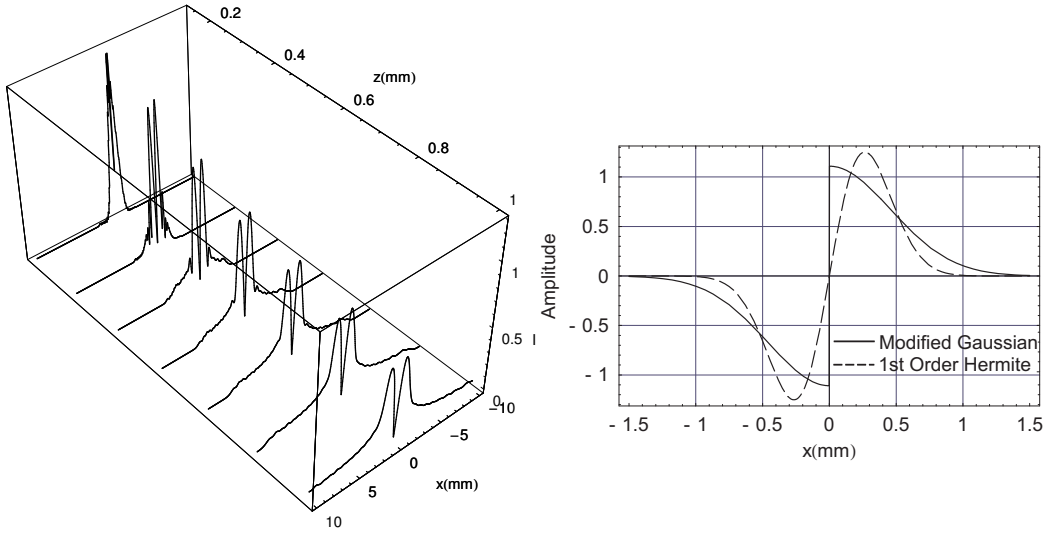


Figure 3.56: *Left* a Fresnel simulation of the propagation of the Gaussian field with the phase of Figure 3.55 *left* imposed on it at z=0mm. The intensity of the field at number of planes is shown in the near-field region. At $z = 0$mm, the amplitude distribution is not exactly equal to a Gaussian, because of the instability of the Fresnel integral when propagating small distances. *Right* a Gaussian with a π phase shift form 0 to $-\infty$ has 82.7% coupling with a 1$^{\text{st}}$ order Gaussian-Hermite.

Simulated annealing with zero order interpolation was used to design the DOE profile shown in Figure 3.55, which converts a Gaussian to an approximately first order Gaussian-Hermite distribution; 82.7% of the power was converted. The solution found by the optimisation imposes a phase given by $-\frac{\pi}{2}\,\mathrm{sgn}(x)$, where sgn is the signum function, on the input Gaussian field. Figure 3.56 shows that the field rapidly evolves from a Gaussian to an intensity distribution similar to a 1$^{\text{st}}$ order Gaussian-Hermite as it propagates.

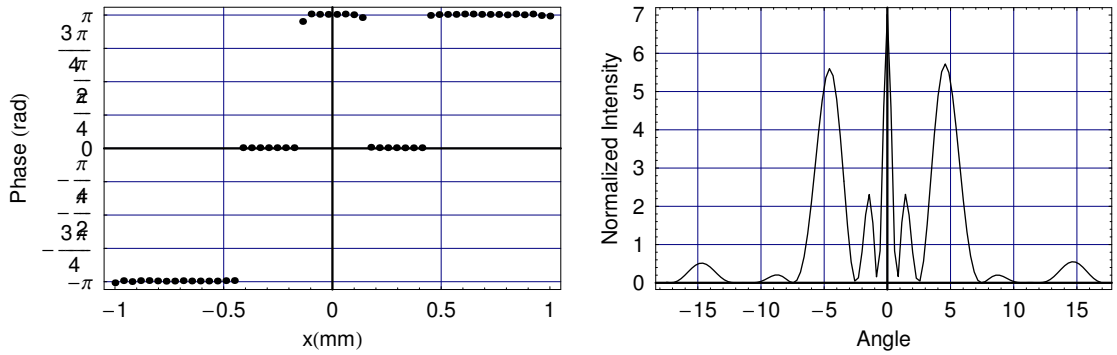


Figure 3.57: *Left* the phase distribution to convert a Gaussian to a 4$^{\text{th}}$ order Gaussian-Hermite, *right* the intensity in the far-field.

Figure 3.57 shows the result of the optimisation to transform a Gaussian to a 4$^{\text{th}}$ order Gaussian-Hermite in which 75.2% of the power was converted. Figure 3.58 *left* shows a plot of the fraction of the power converted from a Gaussian field to a Gaussian-Hermite mode of order between 1 and 46, and Figure 3.58 *right* shows the radius of the higher order Gaussian-Hermite at which this was achieved. The

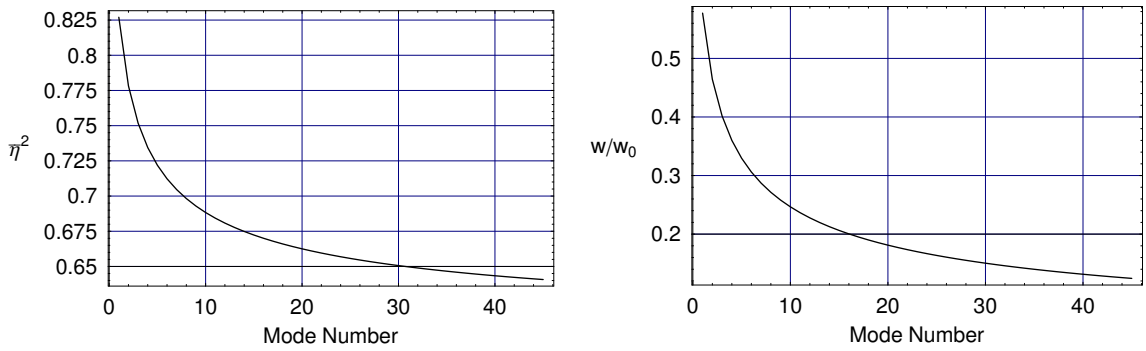


Figure 3.58: The results for a DOE that converts a fundamental Gaussian mode to a higher order Gaussian-Hermite mode. *Left* the power coupled by the higher order Gaussian-Hermite to the field from the converter DOE mode. *Right* the radius of the Gaussian-Hermite w for which $\bar{\eta}^2$ is maximised.

conversion efficiency decreases with increasing order number.

It is observed that all of the solutions found by the optimisation algorithm match the phases of the input and target functions (the π rad phase shift for $x < 0$ is equivalent to multiplying the amplitude by -1). It follows that in order to maximise the power coupling between the detector horn and sky field in a telescope on a plane where a phase can be imposed on the field, the phases of the two fields should be matched at that plane. Matching the phases of the two fields maximises coupling, because determining $\bar{\eta}$ involves the integration of the product of two complex functions: the field from the horn and the complex conjugate of the field from the sky. One function is multiplied by another complex function, $\exp(i\phi_{\mathrm{DOE}}(x))$, representing the DOE - which is of unit magnitude. $\phi_{\mathrm{DOE}}(x)$, the phase of the DOE, is a free parameter. It is now shown that, when the amplitude is fixed and the phase is a free parameter, the coupling of two fields is maximised when the phases are matched. The goal is to maximise

$$\bar{\eta} = \int_{-\infty}^{\infty} h(x)e^{i\phi_h(x)}e^{i\phi_{\mathrm{DOE}}(x)}\left(s(x)e^{i\phi_s(x)}\right)^* dx, \tag{3.25}$$

where $\phi_{\mathrm{DOE}}(x)$ is the phase imposed by the DOE (which is free to take any distribution), $h(x)e^{i\phi_h(x)}$ is the field from the horn and $s(x)e^{i\phi_s(x)}$ is the field from the sky at the DOE. The integral involves the product of the two fields first, as shown in Figure 3.59 *left*, the magnitude of the product of two complex numbers is independent of the phases of the complex numbers, $h(x)e^{i\phi_h(x)}$ times $s(x)e^{i\phi_s(x)}$, as follows:

$$\begin{aligned} \left|P(x)\right| = \left|p(x)e^{i\phi_p(x)}e^{i\phi_{\mathrm{DOE}}(x)}\right| &= \left|h(x)e^{i\phi_h(x)}e^{i\phi_{\mathrm{DOE}}(x)}s(x)e^{-i\phi_s(x)}\right| \\ &= \left|h(x)s(x)e^{i(\phi_h(x)-\phi_s(x)+\phi_{\mathrm{DOE}}(x))}\right| \\ &= h(x)s(x). \end{aligned} \tag{3.26}$$

The integral of this function over x is the sum of infinitesimal elements $P(x_1)\Delta x + P(x_2)\Delta x + ... + P(x_n)\Delta x$, i.e.

$$\int_{-\infty}^{\infty} P(x)dx = p(x_1)e^{i(\phi_p(x_1)+\phi_{\mathrm{DOE}}(x_1))}\Delta x + p(x_2)e^{i(\phi_p(x_2)+\phi_{\mathrm{DOE}}(x_2))}\Delta x + ... + p(x_n)e^{i(\phi_p(x_n)+\phi_{\mathrm{DOE}}(x_n))}\Delta x, \quad (3.27)$$

where $\phi_p(x) = \phi_h(x) + \phi_s(x)$ and $p(x) = h(x)s(x)$. As shown in Figure 3.59 *right*, where a complex number is added to another varying from 0 to 2π, matching the phases of each of the elements (by choosing an appropriate value for $\phi_{\mathrm{DOE}}(x)$) $\phi_{\mathrm{DOE}}(x)$ such that, $\phi_{\mathrm{DOE}}(x) = \phi_p(x)$ achieves the maximum possible value of $\bar{\eta}$, which is given by

$$\bar{\eta} = \sum_{x=-\infty}^{\infty} \left|h(x)e^{i\phi_h(x)}\right| \left|s(x)e^{i\phi_s(x)}\right| \Delta x. \quad (3.28)$$

Therefore, $\bar{\eta}$ is maximised when the DOE is used to match the phases of the two fields.

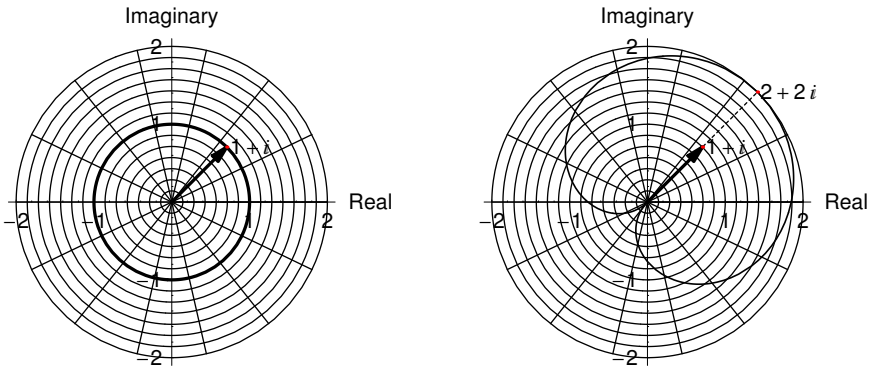


Figure 3.59: *Left* the product of the complex number $e^{i\pi/4}$ and another complex number of magnitude 1 and phase angle varying from 0 to 2π rad yields a complex number of the same magnitude for all phase angles. *Right* the sum of a complex number $e^{i\pi/4}$ and another complex number of magnitude 1 and phase angle from 0 to 2π rad is a maximum when the phase angles are equal.

3.6.2 Beams that are Laterally Confined over Large Distances

So called "diffraction-free beams", in which the amplitude does not change in form or scale while propagating, were identified in 1987[41]. These beams have an amplitude cross-section of a Bessel function, as shown in Figure 3.60, and are given by

$$E(r, z) = E_0 \exp(ik_z z)J_0(k_r r), \quad (3.29)$$

where J_0 is the zeroth-order Bessel function, k_z, and k_r are the longitudinal and radial components of the free-space wave vector, related by $k_0^2 = k_z^2 + k_r^2$, where k_0 is the free space wavenumber.

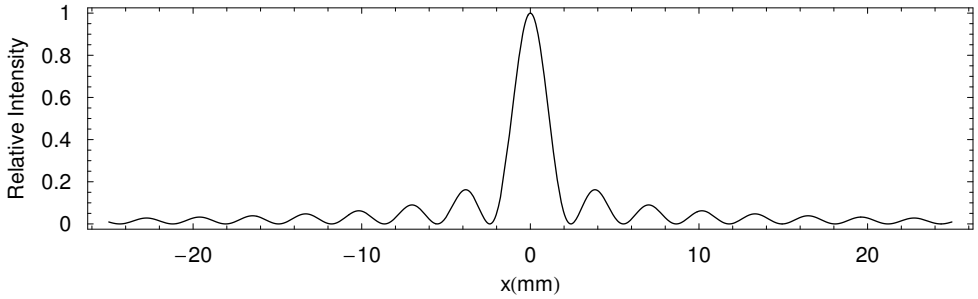


Figure 3.60: A zero order Bessel function, which is the intensity distribution of a cross-section of a Bessel beam.

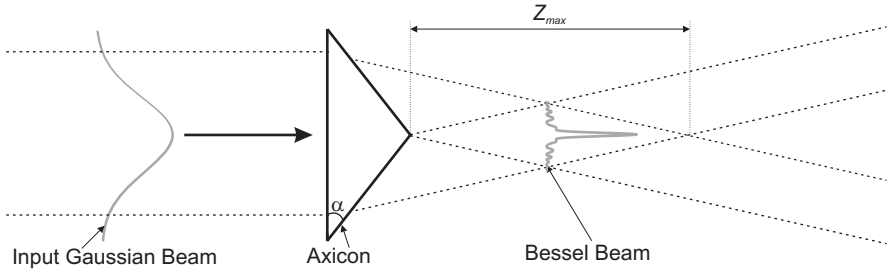


Figure 3.61: The input Gaussian is transformed into a pseudo-Bessel beam within the region Z_{max}.

Such beams have been used to provide a very large depth-of-field in imaging systems and to trap particles in each of their rings, using radiation pressure[42]. Each ring of a two dimensional Bessel beam has equal power[43]. Ideal Bessel beams have infinite extent and power; but experimentally generated pseudo-Bessel beams can only approximate them, having finite extent and power. Pseudo-Bessel beams have been generated using dielectric conical shaped lenses called axicons[44, 45]. Figure 3.61 shows that axicons produce a pseudo-Bessel beam in the region Z_{max} (in the overlap region of the two beams produced by the axicon), after which it diffracts rapidly. Axicons produce Bessel beams when illuminated with a uniform distribution of infinite extent; however, practical axicons have a finite aperture size, and Gaussian beams are generally used to illuminate them, rather than plane waves.

An axicon was designed and fabricated and the field it produced was measured by colleagues at NUI Maynooth[45]. The HDPE axicon had a slant angle $\alpha = 20°$ (as shown in Figure 3.61) and a radius of 30mm. It was illuminated with a 78mm-radius Gaussian beam of wavelength $\lambda = 3$mm with a flat phase-front. The field was simulated in [45] using Gaussian-Laguerre modes. Here, the field is propagated using the Fresnel integral. The plots of the intensity distribution shown in Figures 3.62 and 3.63 agree closely with the GBM simulation and the measured field. The field maintains a central spot of radius 6mm in the "non-diffracting" region.

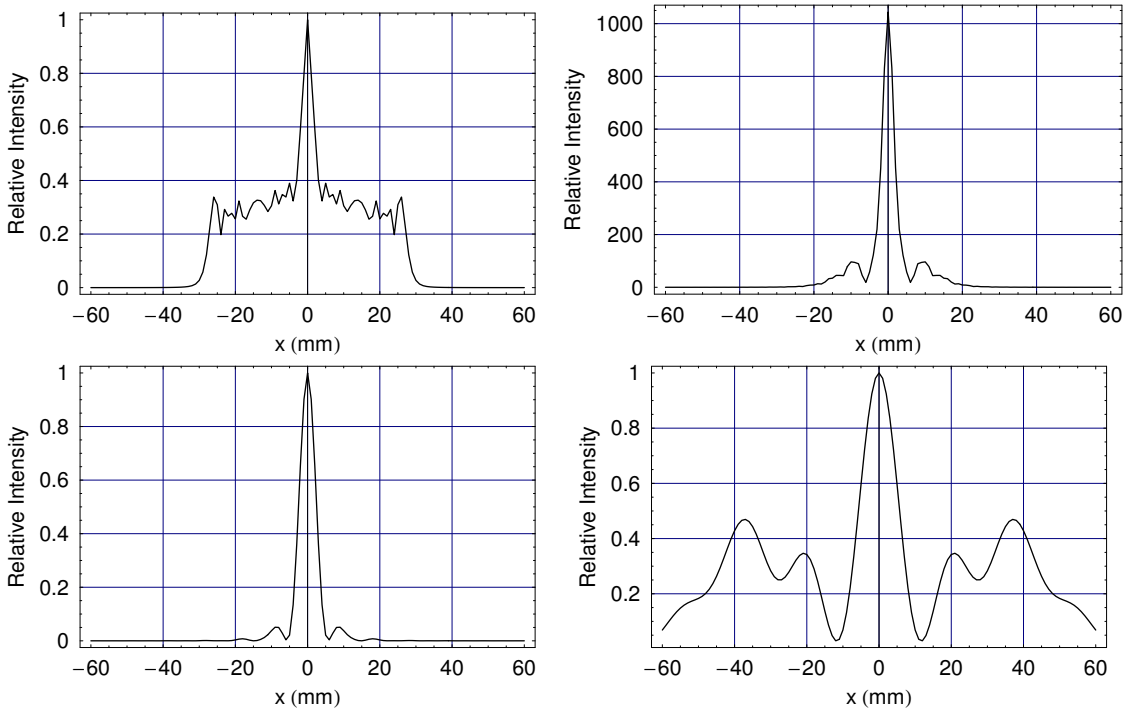


Figure 3.62: A cut across of an axicon beam simulated using Fresnel diffraction. *Top left* at 5 mm from axicon, the central spot is beginning to form. *Top right* at 50 mm from axicon an intense central spot forms. *Bottom left* the central spot is maintained at 100mm from the axicon. *Bottom right* at 300 mm from the axicon the beam has significantly dissipated.

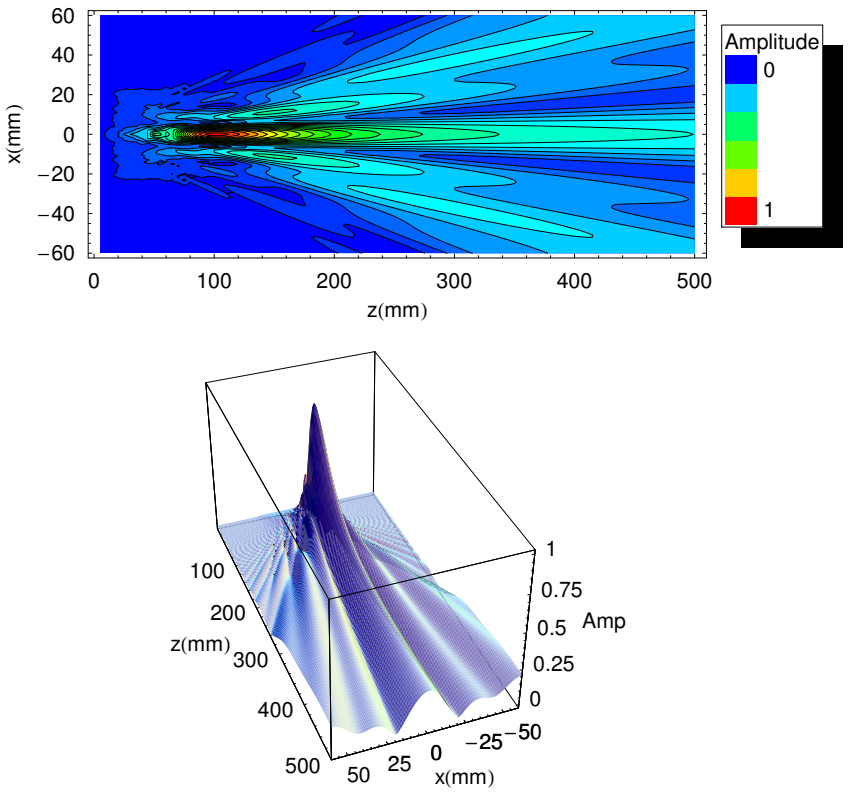


Figure 3.63: A cut across of an axicon beam simulated using Fresnel diffraction. z is the distance from the axicon. *Top* contour plot of the amplitude and *bottom* 3D plot of amplitude.

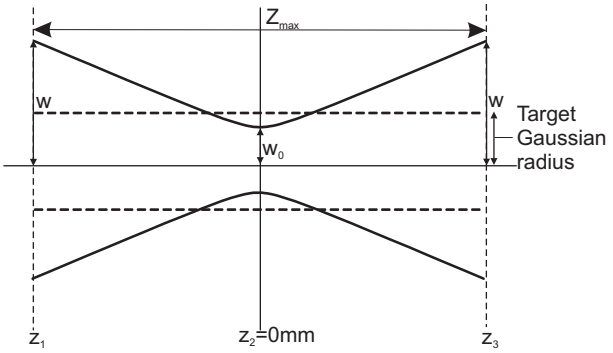


Figure 3.64: The plane z_2 is set to $z = 0$mm to minimises the diffraction of the mode set (see text for details).

Here, GBMSO is used to design other DOEs that create "non-diffracting" beams, optimised for Gaussian input fields, in one-dimension. The optimisation finds a beam which is collimated over a region by maximising $\bar{\eta}$ which is calculated over a number of planes as follows:

$$\bar{\eta} = \int_{-\infty}^{\infty} \psi_0(w,x,0) \left(\frac{\left| \sum_{n=0}^{n_{max}} A_n e^{i\phi_n} \psi_n(w,x,z_1) \right| + ... + \left| \sum_{n=0}^{n_{max}} A_n e^{i\phi_n} \psi_n(w,x,z_n) \right|}{\sqrt{\sum_{n=0}^{n_{max}} A_n^2}} \right) dx, \tag{3.30}$$

where $\psi_0(w,x,0)$ is a fundamental Gaussian and $\psi_n(w,x,z_n)$, at plane $z = z_n$, are Gaussian-Hermite GBMs with mode index n, A_n and ϕ_n are the amplitude and phase of the mode coefficients, and n_{max} is the highest order mode used to reconstruct the field. The mode set is thus optimised to achieve a Gaussian field of constant radius in the region $Z_{max} : z_1 \rightarrow z_3$. In order to reduce the computation time, the optimisation is carried out by maximising $\bar{\eta}$ at just three planes z_1, z_2 and z_3. In this case $z_2 = 0$, $z_1 = -z_3 = Z_{max}/2$, as shown in Figure 3.64. The optimum choice for the waist, w_0, of the modes used in Equation 3.30 is now considered. Figure 3.65 *top* shows both the 60^{th} order mode (the highest order mode in the mode set used to reconstruct the field) with w_0 chosen such that it has a similar extent to the fundamental Gaussian required at $z = 0$mm. This choice of w_0 allows a good reconstruction of the field at the plane $z = 0$mm; however, at $z = 100$mm the mode set has diffracted greatly and it will therefore be difficult to construct the target field at this plane. Figure 3.65 *bottom* shows the 60th mode when the value of w_0 is chosen to minimise w, the radius of the modes, at $z = 100$mm. The red line in Figure 3.66 *left* and *right* shows the value of w_0 for which w is a minimum at a given propagation distance z. w_0 is chosen such that it minimises w at $Z_{max}/2$, as diffraction is greatest at this, furthest, plane. The equation for w_0 that gives the minimum value of w as a function of propagation distance, z, is derived as follows. w, the radius at distance

z is given by

$$\begin{aligned} w &= \sqrt{\left(\frac{z^2\lambda^2}{\pi^2 w_0^4} + 1\right) w_0^2}[6] \\ w^2 &= w_0^2 + \left(\frac{\lambda z}{\pi}\right)^2 w_0^{-2}. \end{aligned} \tag{3.31}$$

Differentiating both sides w.r.t. w_0 gives

$$\frac{\mathrm{d}w^2}{\mathrm{d}w_0} = 2w_0 - \left(\frac{\lambda z}{\pi}\right)^2 2w_0^{-3}. \tag{3.32}$$

w^2 (and therefore w) is a minimum when $\frac{\mathrm{d}w^2}{\mathrm{d}w_0} = 0$,

$$\begin{aligned} 0 &= 2w_0 - \frac{2\lambda z w_0^{-3}}{\pi} \\ w_0 &= \sqrt{\frac{z\lambda}{\pi}}. \end{aligned} \tag{3.33}$$

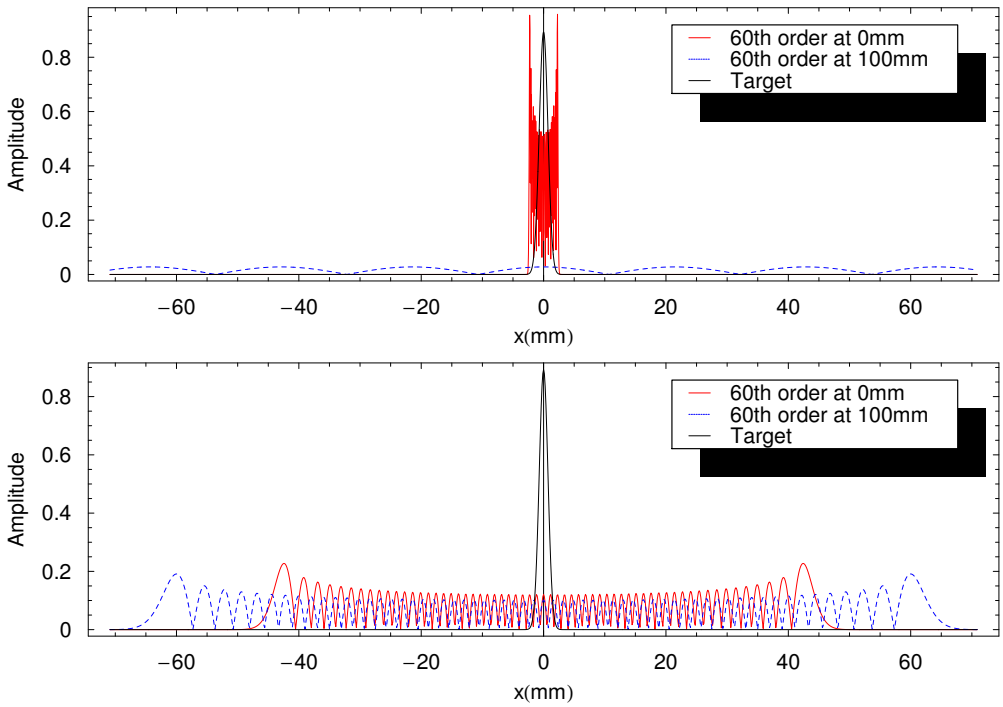


Figure 3.65: *Top* with w_0 chosen such that the highest order (60^{th} order) GBM has a similar extent to the fundamental Gaussian at z=0mm, at z=100mm it has diffracted greatly. The "target" is the desired amplitude profile at all planes in the range 0 to 100mm. *Bottom* w_0 is chosen to minimise diffraction at $z = 100$mm.

By using only real valued mode coefficients, the number of parameters in the optimisation was reduced by two, and was found to produce equally good results as when both the amplitude and phase of the mode coefficients were optimised. It also

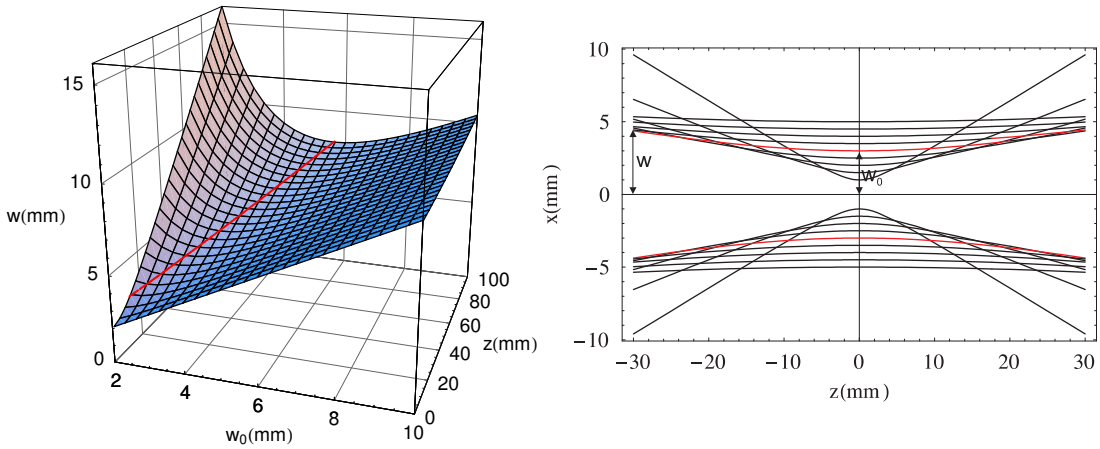


Figure 3.66: *Left* Gaussian beam width, w, w_0 as function of propagation distance, z, where $\lambda = 1$mm. *Right* the radius of a Gaussian, w, for a number of values of w_0. At $z = 30$mm the minimum value of w is 4.4 mm, and that occurs when the waist, w_0, is 3.1mm.

has the benefit that the merit function only needs to be calculated at planes z_2 and z_3, and not at z_1, because the field at z_1 is equal to the complex conjugate of the field at z_3 and has an identical amplitude profile. Using only real valued coefficients results in a field with a flat phase-front at z_2 .

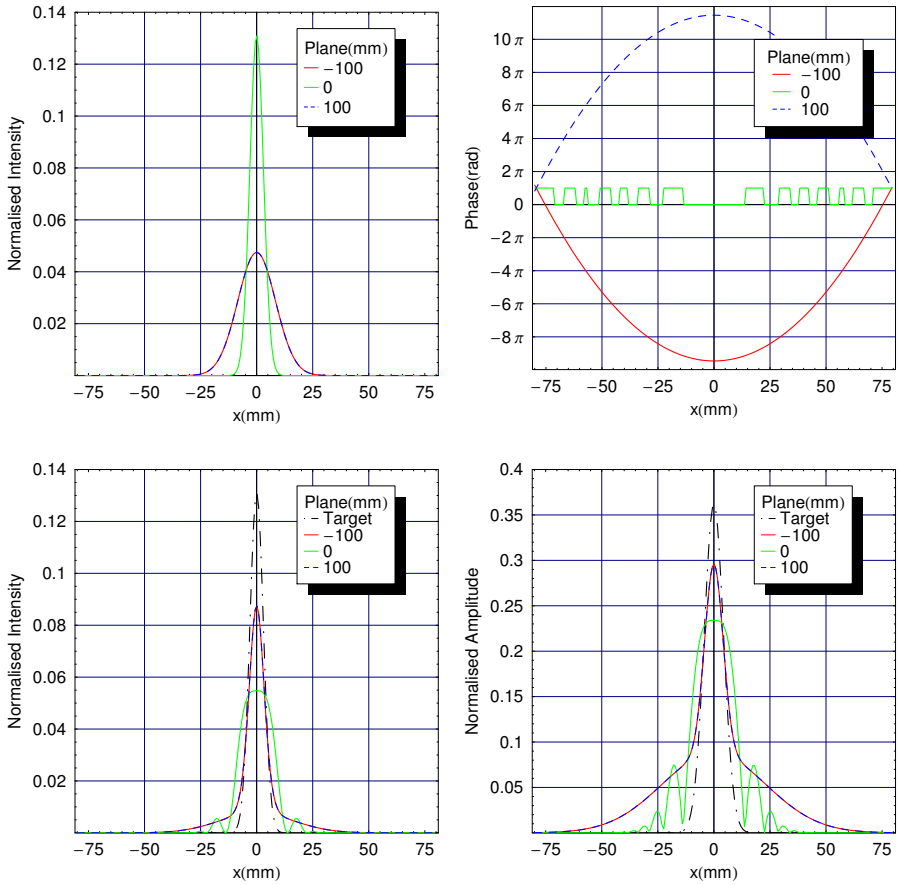


Figure 3.67: Plots for a field optimised to have a 6mm-radius Gaussian intensity distribution at $z = \pm 100$mm - i.e. a "non-diffracting" beam over a region of 200mm. *Top left* the intensity distribution of a fundamental Gaussian mode with a 6mm waist (equal to the target Gaussian radius in the optimisation) at -100mm, 0mm and 100mm, shown for comparison. *Top right* the phase of the field optimised with a target Gaussian of radius 6mm at $z = \pm 100$mm. *Bottom left* the intensity of the field optimised with a target Gaussian of radius 6mm at $z = \pm 100$mm. *Bottom right* the amplitude of the field optimised with a target Gaussian of radius 6mm at $z = \pm 100$mm.

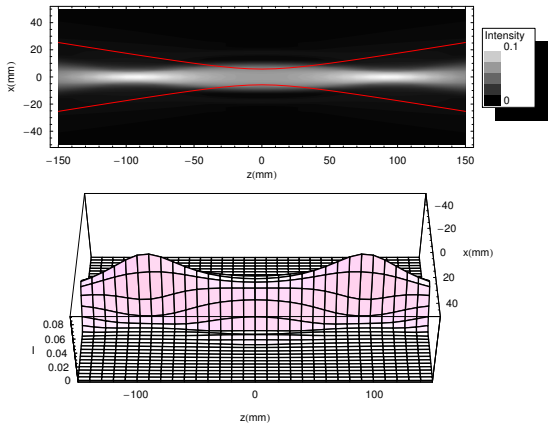


Figure 3.68: Plot of intensity showing the propagation of the beam shown in Figure 3.67. The red line (*top*) shows the radius of a propagating Gaussian beam with a waist of 6mm at $z = 0$mm for comparison.

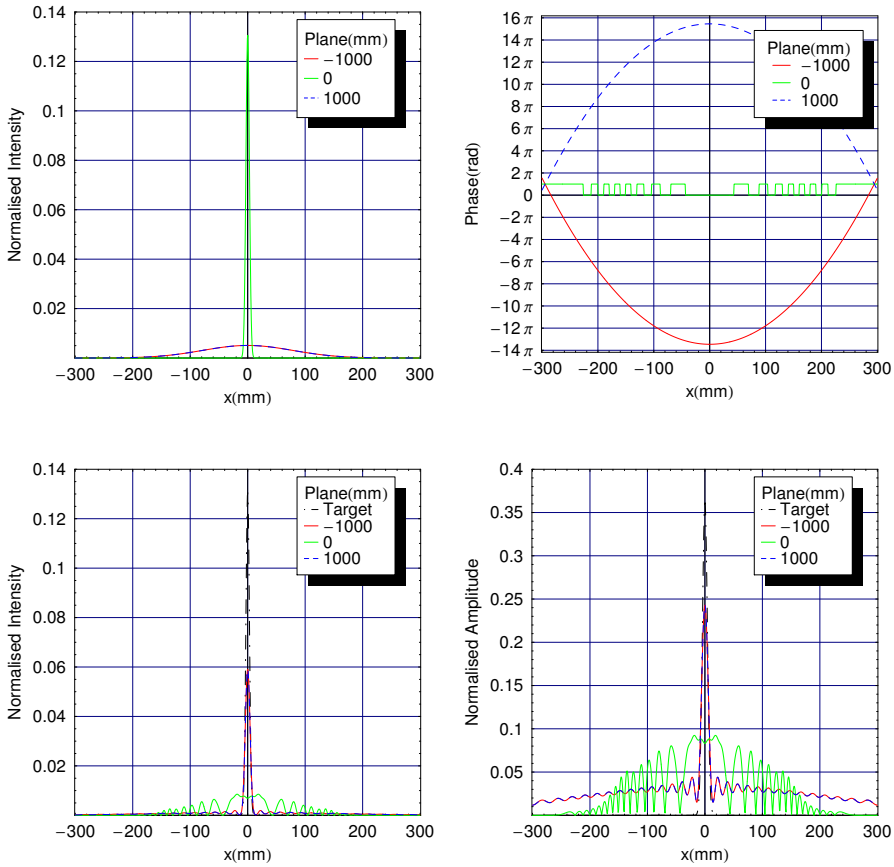


Figure 3.69: Plots for a field optimised to have a 6mm-radius Gaussian distribution at $z = \pm 1000$mm. *Top left* plot of the amplitude of a fundamental Gaussian mode with a 6mm waist radius at $z = 0$mm (equal to the target Gaussian radius in the optimisation) at $z = -1000$mm, 0mm and 1000mm. *Top right* the phase of the field optimised with a target Gaussian of radius 6mm at $z = \pm 1000$mm. *Bottom left* the intensity of the field optimised with a target Gaussian of radius 6mm at $z = \pm 1000$mm. *Bottom right* the amplitude of the field optimised with a target Gaussian of radius 6mm at $z = \pm 1000$mm.

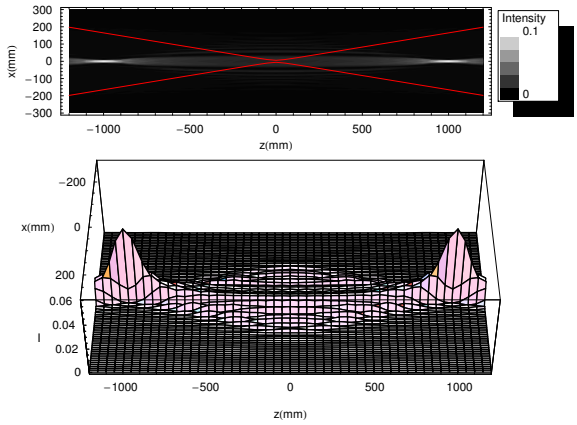


Figure 3.70: The propagation of the beam shown in Figure 3.69. The red line (*top*) shows the radius of a propagating Gaussian beam with a waist of 6mm at $z = 0$mm for comparison.

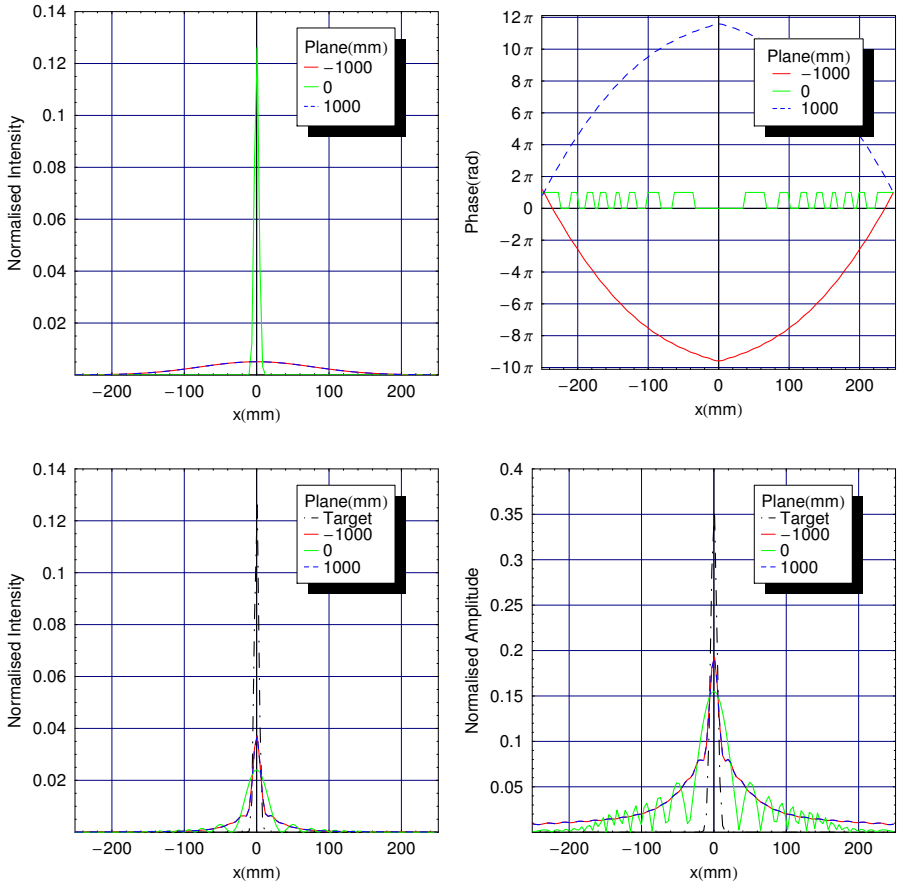


Figure 3.71: Plots of a field optimised to have a 6mm-radius Gaussian distribution at $z = \pm 1000$mm and a constant intensity at $x = 0$ between propagation distances $z = \pm 1000$mm. *Top left* a fundamental Gaussian mode with a 6mm waist radius shown at -1000mm, 0mm and 1000mm, for comparison. *Top right* the phase of the field optimised with a target Gaussian of radius 6mm (equal to the target Gaussian radius in the optimisation) at $z = \pm 1000$mm. *Bottom left* the intensity of the field optimised with a target Gaussian of radius 6mm at $z = \pm 1000$mm. *Bottom right* the amplitude of the field optimised with a target Gaussian of radius 6mm at $z = \pm 1000$mm.

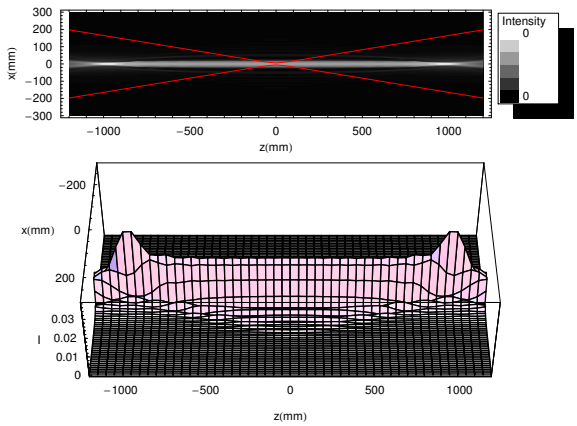


Figure 3.72: The propagation of a beam shown in Figure 3.71. The red line (*top*) shows the radius of a propagating Gaussian beam with a waist of 6mm at $z = 0$mm for comparison.

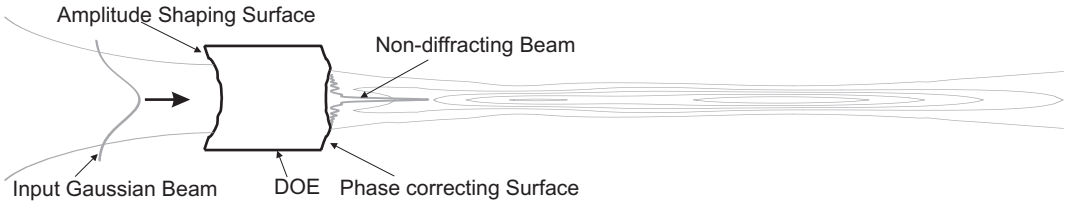


Figure 3.73: This DOE uses two surfaces to produce a non-diffracting beam; the first surface creates the required amplitude profile on the second surface which imposes the required phase.

Figure 3.67 and 3.68 show the solution found using DE and optimising the first 60 even mode coefficients with λ=3mm and a target Gaussian of radius 6mm at $z = \pm 100$mm. The field is similar to that of the target Gaussian at z_1 (-100mm) and z_3 (100mm), and the field has a form similar to a Bessel function at $z = 0$mm. Figure 3.69 and 3.70 show the solution for a target Gaussian at $z = \pm 1000$mm. As the target planes are separated, the field between the target planes becomes less focussed (as observed by the difference in peak intensity between the target and optimised fields), and between the two planes, z_1 and z_3, the field also becomes more dispersed. To reduce this dispersion, the merit function can be evaluated at more planes with Equation 3.30. However, calculating the merit function at multiple planes increases the computation time. An alternative is to perform the merit another function calculation along $x = 0$mm over the "non-diffracting" region; so the optimisation will search for a solution which also maintains on axis power. This merit function is given by

$$\bar{\eta_2} = \bar{\eta} + \frac{\int_{z_1}^{z_2} \psi_0(w, x=0, z=0) \left| \sum_{n=0}^{n_{max}} A_n e^{i\phi_n} \psi_n(w, x=0, z) \right| dz}{\sqrt{\int_{z_1}^{z_2} \left| \sum_{n=0}^{n_{max}} A_n e^{i\phi_n} \psi_n(w, x=0, z) \right|^2 dz}}, \tag{3.34}$$

where $\bar{\eta}$ is given by Equation 3.30. Figure 3.71 and 3.72 show the solution for a target Gaussian at $z = \pm 1000$mm, with constant on-axis power maintained throughout the propagation, optimised using Equation 3.34. Although less power is contained in the target Gaussian, the spot is maintained throughout the region between the two planes.

Thus far "non-diffracting" beams have been found without regard to how they may be generated in practice. For far-field beam shaping, the amplitude of the optimised near-field solution is replaced by a Gaussian distribution, to determine the far-field achievable in practice. For these "non-diffracting" beams, however, illuminating the DOE (given by the phase of the solution at the input plane) with a Gaussian gave poor results over the plane that it was intended to be "non-diffracting", and it was found that the amplitude of the optimised solution was also required to

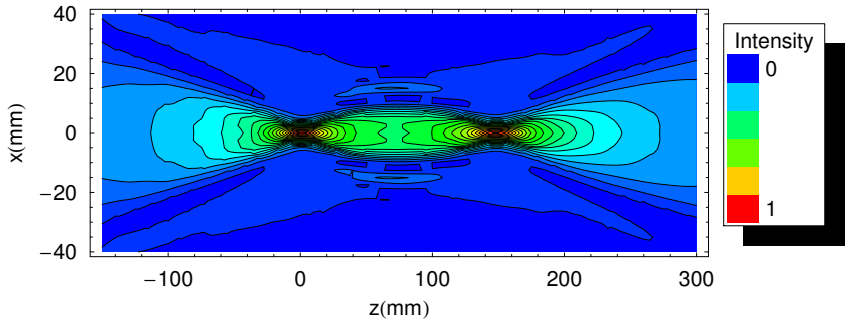


Figure 3.74: The propagation of a field optimised to have a Gaussian amplitude distribution, of any radius, at $z = -150$mm, and a 3.5mm Gaussian amplitude distribution at $z = 0$mm and 150mm.

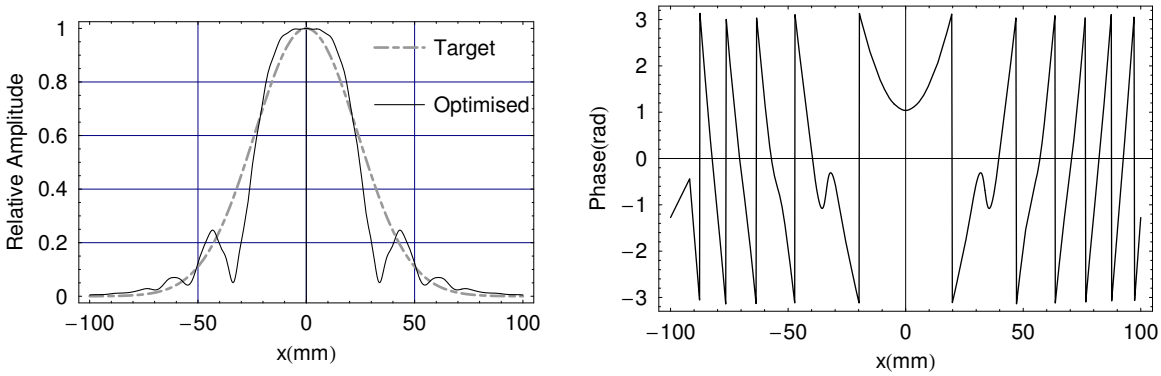


Figure 3.75: *Left* the field at $z = 150$mm where it is optimised to have a Gaussian amplitude profile of any radius (optimised), shown with this optimised Gaussian (target). *Right* the phase of the field at $z = -150$mm, which gives the shape of the DOE.

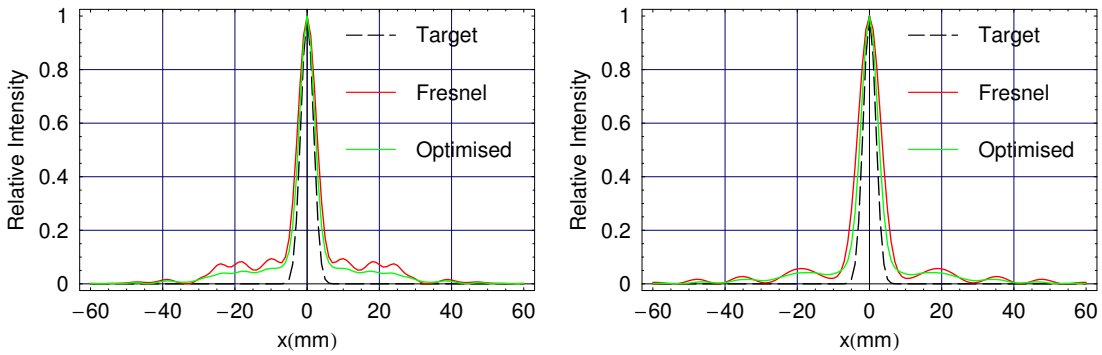


Figure 3.76: The optimised solution of Figure 3.75 is replaced with a Gaussian and propagated a distance of *left* 150mm (the plane at $z = 0$mm of Figure 3.74) and *right* 300mm (the plane at $z = 150$mm of Figure 3.74).

produce a reasonably "non-diffracting" beam.

The amplitude of the optimal solution may itself be produced by beam shaping which creates the input field for a second DOE. The *amplitude* profile given by the optimised solution at the input plane may be produced by beam shaping using GBMSO (as described in Section 3.2) for input to a second DOE, the second DOE would then impose the required phase on the field to give a field with both the amplitude and phase of the solution given by the GBMs at the input plane. Two surfaces would therefore be required to create the beam. They may be integrated into a single DOE as shown in Figure 3.73. The first surface of the DOE focuses onto the second surface, creating the amplitude distribution of the optimised solution and the second surface imposes a phase onto the solution, so that the field produced has the correct amplitude *and* phase. However, using two surfaces to perform the beam shaping is more complicated than a single surface DOE.

DOEs that produce "non-diffracting" beams using a single surface are now investigated. In order to give the optimisation more freedom to find a solution, the radius, w, of the input field, $\psi_0(w, x, 0)$, is made a free parameter. The merit function is given by

$$\bar{\eta_2} = \int_{-\infty}^{\infty} \psi_0(w, xz) \frac{\left| \sum_{n=0}^{n_{max}} A_n e^{i\phi_n} \psi_n(w, x, z_1) \right|}{\sqrt{\sum_{n=0}^{n_{max}} A_n^2}} dx + \bar{\eta}, \tag{3.35}$$

where $\psi_0(w, x, 0)$ is a Gaussian distribution of radius w and $\bar{\eta}$ is given in Equation 3.30. The amplitude and phase of the first 60 even mode coefficients are optimised to give a Gaussian amplitude profile of any radius at $z = -150$mm and a Gaussian with a radius of 3.5mm at 0mm, and 150mm. The propagation of the optimised field is shown in Figure 3.74. Figure 3.75 *left* shows the amplitude and phase of the field at $z = -150$mm where it was optimised to have a Gaussian amplitude profile of any radius. The phase distribution of the field shown in Figure 3.75 *right* is imposed on a Gaussian field with a flat phase-front and the beam radius returned by the optimisation, in order to analyse its performance. Figure 3.76 shows the field propagated distances of 150 and 300mm, where the Fresnel integral is used to propagate the field. Using a Gaussian input produces a field similar to that of the optimised solution. This is a significant improvement on the previous result when the target was replaced by a Gaussian input. The optimised solution has a larger beam radius than the target Gaussian. This is because there is a trade-off between the amount of power in the region outside the spot and the radius of the spot size; a larger optimised spot size allows better power coupling with the target central spot.

3.6.3 Beams that Avoid Obstacles in their Path

Here, GBMSO is employed to design a DOE which eliminates the on-axis power over a specified region. Possible applications of these DOEs include:

- optical tweezers to trap particles in the region with no power,
- the avoidance of obstructions in quasioptical systems and
- the avoidance obstructions between a transmitter and receiver in line-of-sight communications equipment.

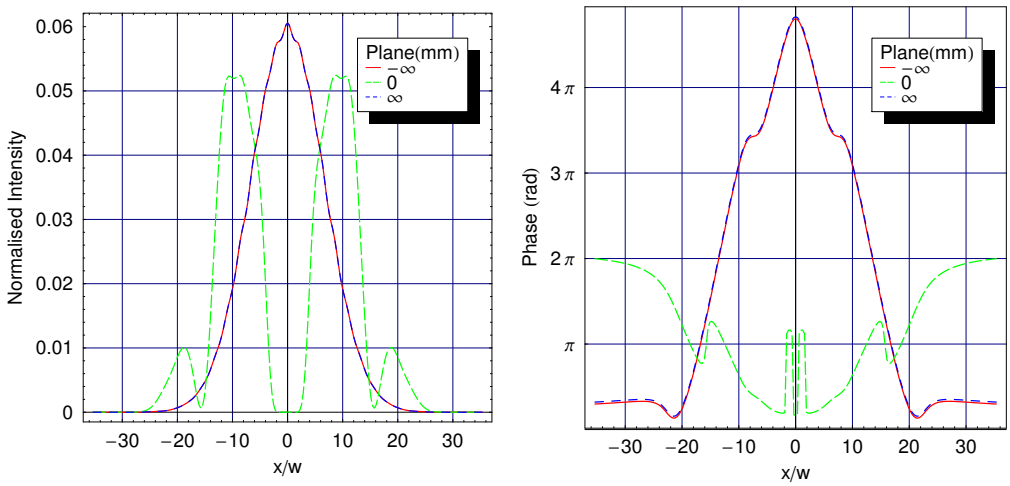


Figure 3.77: A symmetric DOE is designed to create a beam which is Gaussian at $z = \pm\infty$ and to avoid an obstacle in the region $2\text{mm} < x < -2\text{mm}$ at $z = 0\text{mm}$. *Left* the intensity and *right* the phase at $z = 0\text{mm}$ and $z = \pm\infty$.

The optimisation finds a solution by maximising the following merit function:

$$\eta_a(z=\infty) = \int_{-\infty}^{\infty} \frac{\psi_0(\bar{w},x,0)\left|\sum_{n=0}^{n_{max}} A_n e^{i\phi_n}\psi_n(w,x,z=\infty)\right|}{\sqrt{\sum_{n=0}^{n_{max}} A_n^2}}dx \tag{3.36}$$

$$\eta_b(z=0) = \sqrt{1-\left(\int_{lim_1}^{lim_2} \frac{\psi_0(\bar{w},x,0)\left|\sum_{n=0}^{n_{max}} A_n e^{i\phi_n}\psi_n(w,x,z=0)\right|}{\sqrt{\sum_{n=0}^{n_{max}} A_n^2}}dx\right)^2} \tag{3.37}$$

$$\bar{\eta} = \eta_a(z=\infty) + \eta_b(z=0) \tag{3.38}$$

where lim_1 to lim_2 is the region over which no power is required. During the optimisation, the mode set does not expand with distance z, as w is fixed. Equation 3.36 is the merit function quantifying the similarity of the far-field to a Gaussian beam. The field at $z = \infty$ is equal to the complex conjugate of the field at $z = -\infty$,

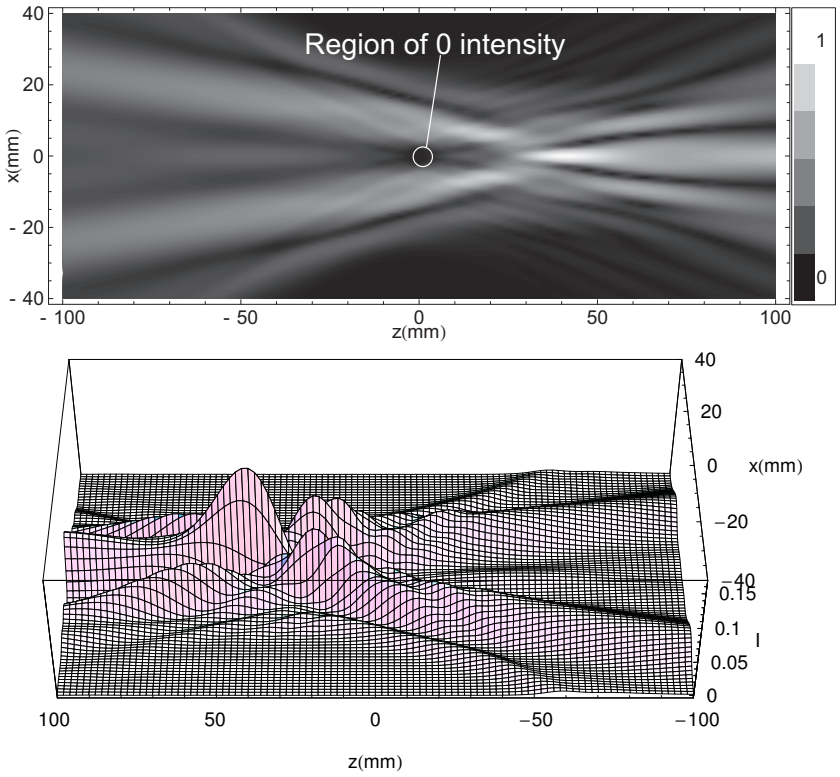


Figure 3.78: A symmetric DOE is designed to create a beam which has a Gaussian distribution at $z = \pm\infty$ and to avoid an obstacle in the region $2\text{mm} > x > -2\text{mm}$ at $z = 0\text{mm}$. The plots show the propagation of the field near $z = 0\text{mm}$.

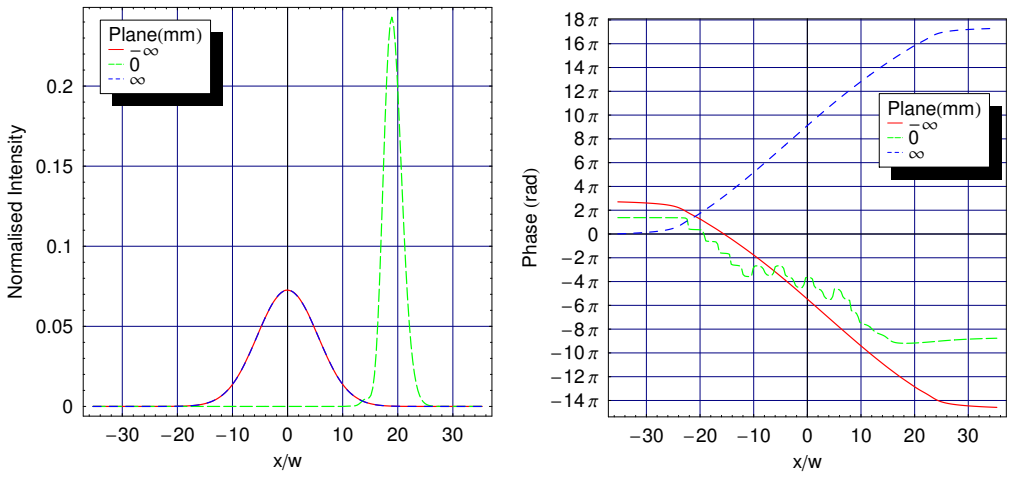


Figure 3.79: The propagation of a field which is Gaussian at $z = \pm\infty$, and at z=0mm has no power in the region $x < 10\text{mm}$. *Left* intensity, *right* phase.

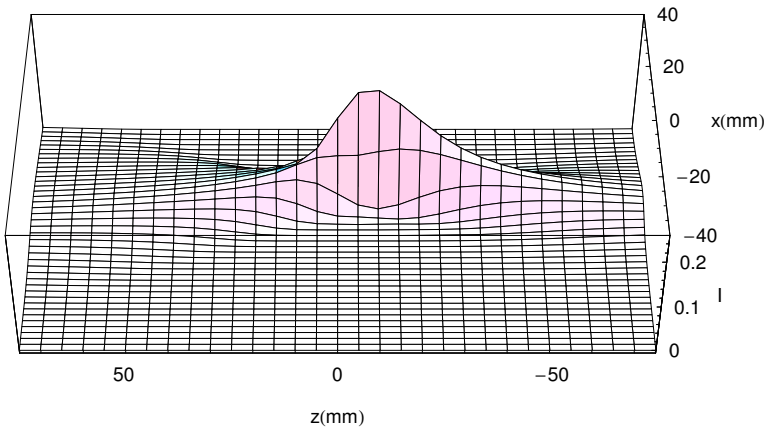


Figure 3.80: The propagation of a field which has a Gaussian distribution at $z = \pm\infty$, and at z=0mm has no power in the region $x < 10$mm.

and therefore (as only even GBMs are used in the optimisation) the amplitude profile and merit function are identical at $z = \pm\infty$. As a result, instead of calculating the merit function at $z = \pm\infty$, it is only necessary to calculate the merit function at one of these planes. Equation 3.37 gives the square root of the power outside the region lim_1 to lim_2. The width of the target Gaussian, $\bar{w}$, is not fixed but used as a parameter to be optimised in order to allow more freedom to find a good solution. $\bar{\eta}$, in Equation 3.38, is equal to a maximum value of two when the target Gaussian at $z = \pm\infty$ is achieved and no power exists in the region lim_1 to lim_2 at $z = 0$.

Figures 3.77 and 3.78 shows the result of the optimisation, using the first 20 even modes, where $lim_1 = -2$ and $lim_2 = 2$. The final merit function was $\bar{\eta} = \eta_a + \eta + b = 2$, indicating that the GBMs describe the field exactly at $z = \pm\infty$, with no power between lim_1 and lim_2 at $z = 0$. Figures 3.79 and 3.80 show the field for a DOE that eliminates power at $z = 0$ in the region $lim_1 = 10$ to $lim_2 = \infty$, so the field is required to go to one side of the region. To optimise this field, both odd and even modes are required, as the field is asymmetric. The first 20 modes are used and 100% coupling is achieved at both planes, with no power over the region $lim_1 = 10$ to $lim_2 = \infty$ at $z = 0$.

3.7 Conclusion

In this Chapter, DOEs were designed that transformed a Gaussian beam to a top-hat amplitude distribution, 1st order Gaussian-Hermite amplitude distribution and three separate beams in the far-field in one-dimension. The GBMSO technique was shown to achieve highly optimal solutions. The advantage of unidirectional optimisation using global optimisation algorithms over the GSA is that it is more flexible, as arbitrary constraints may be placed on the target field. GBMSO finds highly optimal solutions, while allowing more control over the target field (for example, by allowing

the scale of the target field to be optimised) and phase (for example, by restricting the mode set to only even modes, to achieve a symmetric phase).

GBMA allows an intuitive explanation as to why target far-field intensity distributions with small scale features are difficult to attain. For a particular scale of far-field distribution, the ratio the target field distribution to the beam waist radius, r_f/w_0^2, is fixed (the far-field angular radius, θ_f, is proportional to r_f/w_0^2, see Equation 3.11). The modes that describe the near-field distribution are required to interfere in the required way to describe the far-field target distribution, but interfere destructively outside of the extent of the target distribution. If the target to waist ratio r_f/w_0^2 is small in comparison with the extent of the modes, it is unlikely that such mode coefficients can be found, and especially unlikely for higher order modes, with their large spatial extent.

For the Gaussian to top-hat amplitude distribution conversion, increasing the scale of the far-field distribution always allows a more optimal solution to be obtained. However, for the Gaussian to 1$^{\text{st}}$ order Gaussian-Hermite and three beams amplitude distribution conversion, an intermediate scale can be found where both the near and far-field distributions can be described well.

For large far-field targets, a larger number of modes can be used to contribute to the intensity distribution, as r_f/w_0^2 is larger. And with a larger number of modes, there is a greater number of possible combinations, and therefore and a greater likelihood that a particular combination accurately describes both the near- and far-field target distributions. However, due to the greater number of modes that are required to be optimised, the problem is highly non-linear.

GBMSO reduced the number of parameters to be optimised compared with the number required in unidirectional optimisation. Reducing the number of parameters does not necessarily, in itself, make the problem less nonlinear and less computationally demanding. GBMSO gave more optimal solutions than unidirectional optimisation, which was more likely to get stuck in a local minimum. Whereas unidirectional optimisation operates in the forward direction - transforming from the near- to far-field only - and the GSA operates in both directions - iterating between the near- and far-field - GBMSO operates in *both* directions *simultaneously* (i.e. changing the value of a mode coefficient changes *both* the near- and far-field distributions). And whereas, the GSA modifies the amplitude of the solution field at each iteration to give an improved solution, unidirectional optimisation modifies the phase. GBMSO can modify *both* the amplitude and phase, allowing the algorithm more freedom to jump out of a local minimum than the GSA or unidirectional optimisation. Practical application of the GBMSO algorithm to astronomical systems will be described in later chapters.

With unidirectional optimisation using the FFT to transform to the far-field,

trial phase distributions can have arbitrarily high spatial frequencies, and therefore much of the power in the far-field may be directed away from the target. This results in the optimisation being more likely to get stuck at local optima. GBMSO is much more efficient in this respect, as the power of the trial near- and far-field distributions are constrained by the extent of the modes.

It was shown that the GBMSO technique could be extended to control the field at multiple planes in the near-field region. DOEs were designed that create fields that are confined over large distances. DOEs were also designed that avoid obstacles. The approach is more intuitive and allows novel fields to be created where an analytical approach may be difficult or intractable.

In addition to the work demonstrated here, DOEs can be designed to create radially symmetric intensity distributions using Gaussian-Laguerre modes, instead of the Gaussian-Hermite modes. DOEs can also be optimised to control fields at multiple wavelengths. The work done on optimisation techniques here could be used to design entire diffractive optical systems from scratch by optimising the parameters of ABCD matrices, for example.

Chapter 4

The Optical Beam Combiner For MBI

4.1 Introduction

The Millimetre-wave Bolometric Interferometer (MBI) is a novel, ground-based, *bolometric* interferometer - designed to demonstrate the technique of bolometric interferometry at millimetre wavelengths, and to search for faint polarisation signals from the CMB[46]. The main advantage of interferometry: high resolution, comes at a cost: a lower collecting area relative to a filled dish. Bolometers have high sensitivity and are commonly used at millimetre wavelengths. In MBI the two technologies are combined; the resulting instrument has capabilities that allow it to be used for a wide range of astronomical observations that would not be possible using more standard techniques[47]. The telescope is being developed under a NASA programme, by a consortium of Brown University, Cardiff University, The University of Manchester, Northwestern University, the University of Richmond and the University of Wisconsin.

In this chapter, I will describe the MBI instrument - and in particular the design of the optical beam combiner, with which I was involved. The challenge for the MBI project is to combine interferometry - a technique generally used at lower frequencies - with the bolometric detectors - used to give higher sensitivity at millimetre wavelengths. To achieve the high sensitivity required, MBI requires careful design and analysis, as well as the control of systematic errors. The optimisation of the beam combiner involves the use of lenses, whose performance must be modelled and well understood if the MBI instrument is to achieve its goals. I will begin by outlining the particular advantages offered by interferometry, before describing the work on the MBI beam combiner design. Then I describe the work that I carried out to reduce diffraction of the field from the feed horns and improve the coupling of the signal field to the detector horns.

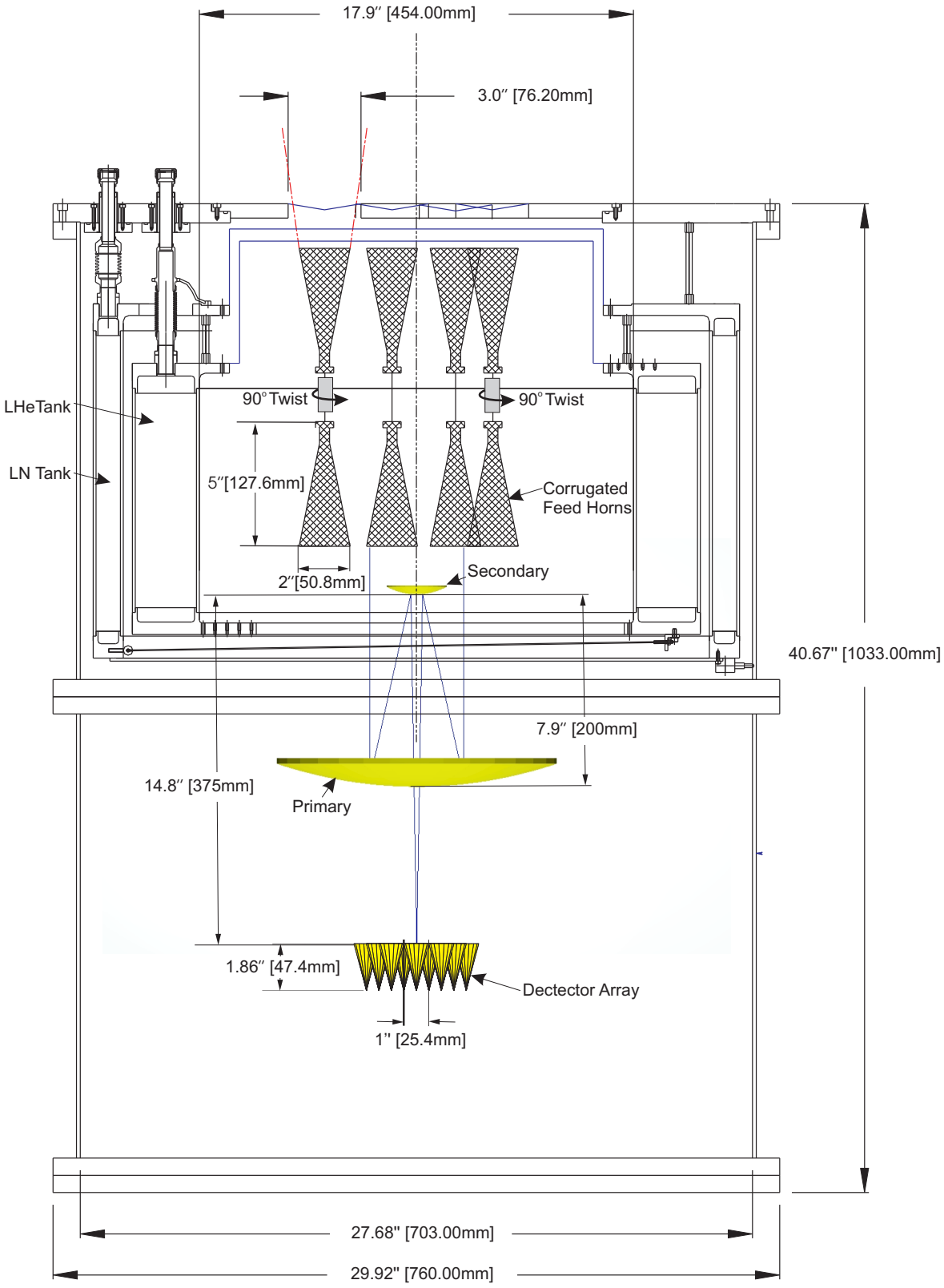


Figure 4.1: The MBI Cassegrain beam combiner.

4.2 Overview of MBI

An interferometer, as opposed to a single-dish telescope, has the advantage that it instantaneously performs a differential measurement, allowing the detection of faint signals. The fringes produced by each of the individual baselines of an interferometer sample the sky with positive and negative weights, thereby eliminating the requirement of mechanical chopping or rapid scanning. Additionally, power spectrum measurements of the CMB are made more straightforward, because the sampling of each baseline is very fine in Fourier space.

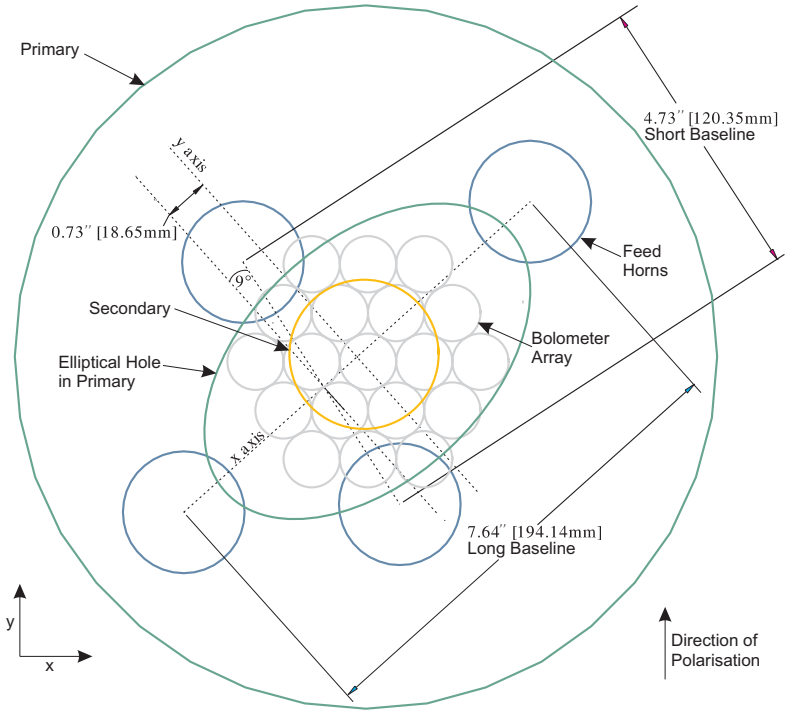


Figure 4.2: Top view of MBI-4 telescope. Dimensions are given in inches and millimetres.

Interferometers allow the direct measurement of Stokes parameters, without differencing the signal from separate detectors. One of the main challenges of the measurement of the CMB polarisation is the separation of the very weak B from the relatively much stronger E modes (as the two modes can leak into each other). An interferometer can separate the E and B modes more effectively than a single-dish experiment can. MBI will serve as a testbed for future space-borne millimetre wave instruments. MBI could also be used to test the measurement of the polarised foregrounds at millimetre wavelengths, in the next-generation of Cosmic Microwave Background (CMB) polarisation missions.

MBI is currently (March 2008) in the assembly and testing phase of its development. In future versions of the instrument, the number of baselines will be increased in order to enhance sensitivity. This initial four aperture version of MBI (MBI-4) may be expanded to eight apertures (MBI-8); ultimately the aim is to build a 64-

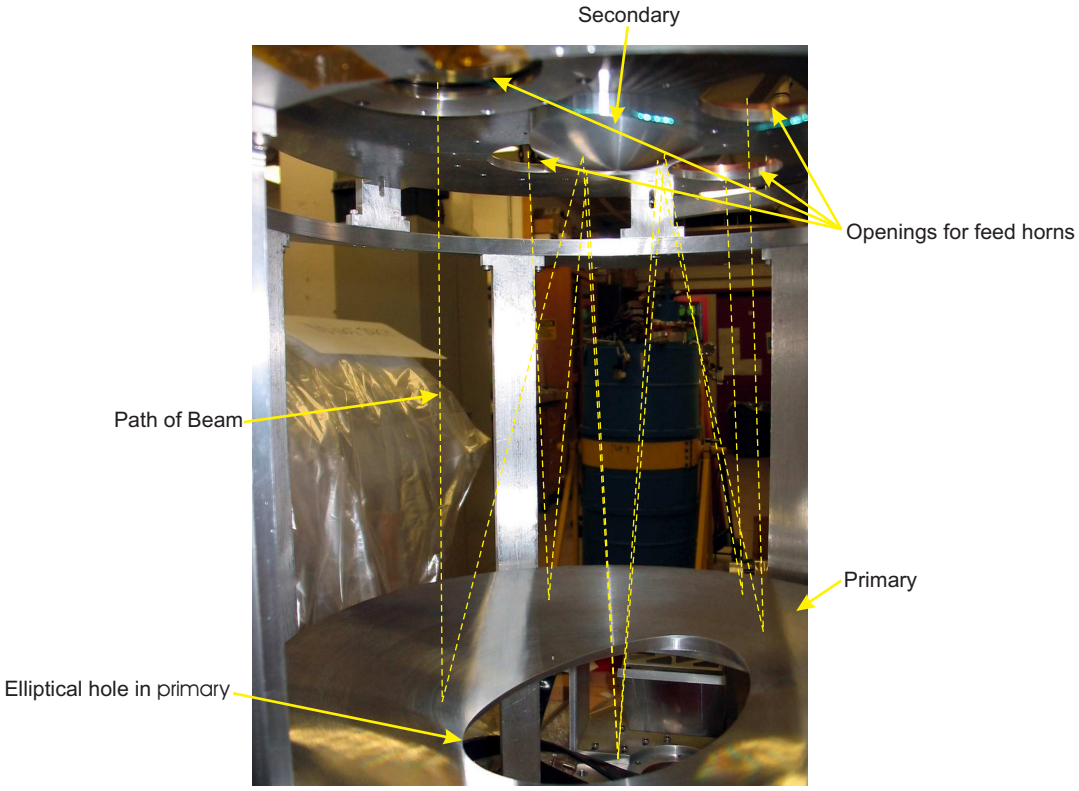


Figure 4.3: Photograph of the MBI telescope.

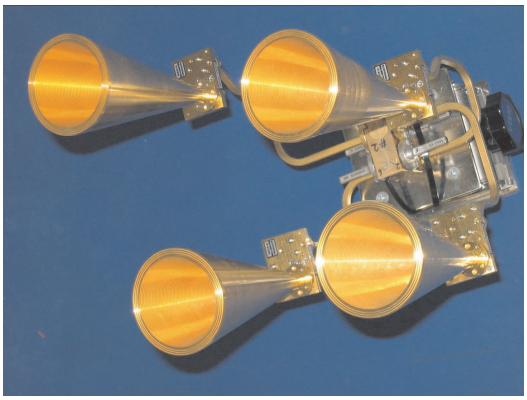


Figure 4.4: MBI-4 partially assembled for room temperature tests. Visible are the four corrugated feed-horns connected to the beam combiner.

Instrument	λ (mm)	Sensitivity ($mK\sqrt{s}$)	Days to reach ($1\mu K/synth.beam$)	θ (degrees)	ϕ (degrees)
MBI-4	3	1.9(1.2)	42(17)	8	1.4
MBI-64	3	0.7(0.5)	5.7(2.9)	8	1.0

Table 4.1: The sensitivity is the NET for each beam to polarised fluctuations, $(Q^2 + U^2)^{1/2}$, in a 2.7 K blackbody. Detector noise and background photon noise for ground-based observations are included. Figures in parentheses are the background-limited case. Sensitivity calculations assume equivalent temperature of atmosphere, optics and CMB is 20 K Rayleigh-Jeans, optical efficiency is 40%, and bandwidth is 20%. The field-of-view (FOV) of each interferometer element is denoted by θ and the size of the synthesized beam by ϕ. MBI-4 is a sparse array with 32 synthesized beams on the sky. MBI-64 is a close-packed array with 64 synthesized beams[48].

element interferometer (MBI-64). MBI will demonstrate the use of bolometers to detect the interference pattern while achieving the sensitivity levels predicted in Table 4.1. Cooled bolometers are the highest-sensitivity broadband detectors at millimetre and submillimetre wavelengths; they can achieve the fundamental noise limit imposed by photon noise from the CMB itself.

The simplicity of the MBI optical system eliminates numerous systematic problems that plague imaging optical systems: corrugated feed-horns have extremely low side lobes and have highly predictable and symmetric beam patterns; there are no reflections from optical surfaces between the horns and the sky, which can induce spurious instrumental polarisation in imaging optics; it does not suffer from aberrations from off-axis mirrors; lastly, the arrays are maintained at cryogenic temperatures to minimize losses.

Figure 4.1 shows a diagram of the MBI instrument, and Figure 4.2 shows a schematic of the instrument viewed from above, with the four feed-horns and 19 detector horns in an hexagonal array. Figure 4.3 shows photographs of the instrument, manufactured at Brown University. The detector horns sample the interference pattern from the telescope at the focal-plane. The four horns of the two baselines are shown in the photograph of Figure 4.4 which are arranged to give roughly uniform u-v coverage in the l-space range 50 - 100. The antennas are corrugated feeds that operate at 90 GHz and whose beams have a FWHM = 7°. The horns view the sky through individual windows. Filters, placed in front of the feed-horns, are used to reduce the optical load on the cryogenics and define the detector passbands. The cryostat uses 17 litres of liquid nitrogen and 25.7 litres of liquid helium, cooling the optics in the cryostat to 4K for 50 hours. The field from the sky is coupled to the detectors by 19 smooth-wall conical horns arranged in a hexagonal array in the combiner focal-plane. The detectors are spider web bolometers manufactured by JPL, cooled to 340mK. The cryostat is mounted in an altitude-azimuth mount. The phase in each of the four inputs is sequentially modulated between $-90°$ and

+90°, using ferrite-based modulators.

The optical beam combiner was designed after the feed-horn and detector array design had been fixed. This, together with the restrictions imposed by the size of the cryostat, meant that truncation at the secondary mirror would be a problem, as the secondary mirror must be small enough to fit between the corrugated horns. But if the beam width at the secondary mirror could be reduced by moving the beam waist position forward to the aperture of the corrugated horn, then the waist size would be increased, reducing the rate of the diffraction of the beam as it propagates from the horn. The beam width at the secondary mirror would consequently be reduced. The waist position can be fixed to the horn aperture by a phase-flattening lens. This would also reduce the frequency-dependence of the horn phase centres. Lenses could also be used to increase the coupling of the detectors to the telescope field by using phase matching lenses at the detector plane. In this chapter I will describe the design of both corrugated and detector feed-horn lenses for the MBI beam combiner.

4.3 Feed-Horn Lenses

4.3.1 Feed-Horns

The feed-horns are of the circular corrugated (scalar horn) type. These horns produce a truncated Bessel beam at the aperture. Here, this field is approximated with a Gaussian beam. Without a lens, the waist radius of the horn beam is given by

$$w_0 = \frac{w}{\sqrt{\pi^2 w^4/(\mathrm{R}^2\lambda^2) + 1}}, \tag{4.1}$$

where R, the radius of curvature, is equal to the slant length of the horn (as shown in Figure 4.5) and w, the radius at the horn, is 0.6435 times the aperture radius[6]. There is a power coupling of 98% between the fundamental Gaussian beam representation of the horn and the more accurate truncated Bessel function description of the field from a scalar horn[6]. The slant length of the horn (Figure 4.5) is 130mm, and the aperture radius is 25.4mm giving a radius, w, of 16.1mm at the aperture. The distance the waist is located behind the horn aperture z_p is given by

$$\Delta z = \frac{R}{1 + (\lambda R/\pi(0.6435\ \mathrm{a})^2)^2}, \tag{4.2}$$

where a is the aperture radius[6]. These formulae give a waist radius of $w_0 =$ 7.56mm, which is located 101.3mm behind the horn aperture.

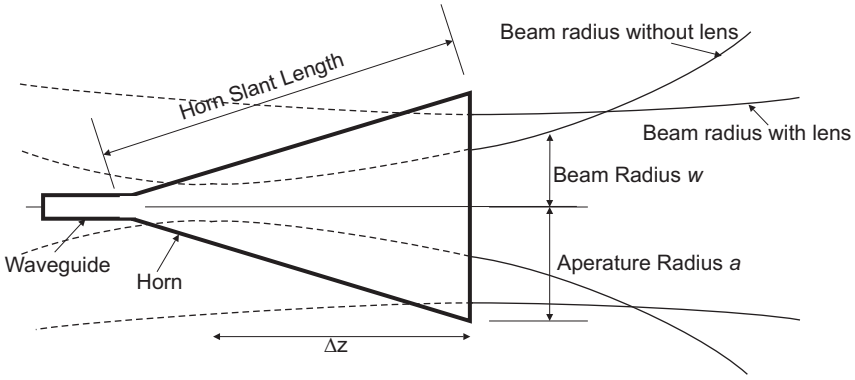


Figure 4.5: Without a lens, the waist of the field produced by the horn is at a virtual location a distance Δz behind the aperture; with the lens, the waist is at the aperture, and the beam expands less rapidly as it propagates.

Location	Radius (mm)	
	Without lens	With lens
Waist	7.7	16.1
Primary mirror (100mm-radius)	83	23
Secondary mirror (32mm-radius)	89	15
Focal plane	416	198

Table 4.2: The value of the radius of the Gaussian beam approximation of the horn field as it propagates through the, as system predicted by Zemax.

The beam width at the focal-plane (where the detectors are located) is

$$w = \frac{\lambda f}{\pi w_0}. \tag{4.3}$$

The equivalent focal length of the system is $f = 3000$mm, giving $w = 416$mm at the focal-plane. Table 4.2 shows the radius of the fundamental Gaussian beam representation of the field as it propagates through the system, as determined by the software package Zemax (using the same approximation as where ABCD matrices are used with GBMs). The field is significantly truncated at the secondary mirror.

4.3.2 Design of Feed Horns Lenses

Lenses are now designed to flatten the phase at the mouth of the feed-horns and so reduce the truncation at the secondary mirror. The usual procedure for designing lenses at quasioptical wavelengths uses geometric techniques. Although it may be expected that the resulting beam would deviate considerably from that predicted due to diffraction, geometric-optics assumptions are surprisingly good when compared with experimental results for lenses designed in this way[49]. Geometric optics is used for the initial design of the lenses here[49]. Three types of refractive lens design are considered, each of which consist of a curved and a planar surface.

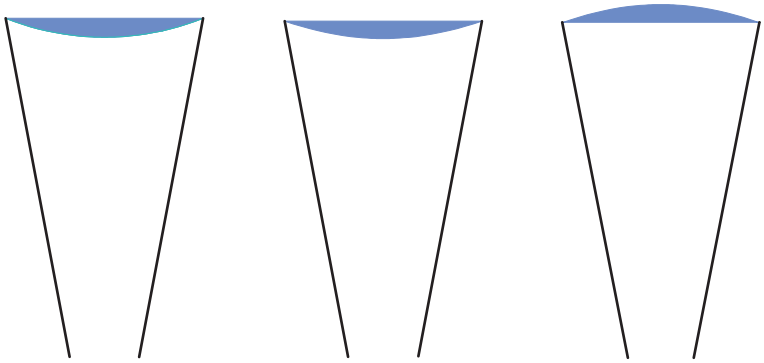


Figure 4.6: Horn lens design options for use on the MBI feed-horns. *Middle* horn with hyperbolic lens, *left* horn with spherical lens, *right* horn with plano-convex lens. (Drawn to scale.)

Geometric optics models radiation as rays of constant power, and uses Snell's law of refraction to determine the change of ray direction on entering a lens. Snell's law, which follows from Fermat's principle, states that the total time for travel by a ray must be the same for adjacent ray paths. Modelling the radiation as rays of constant power allows the power to be tracked, but does not account for diffraction.

Three types of lens, shown in Figure 4.6, are analysed here: the horn facing spherical, the hyperbolic and the plano-convex.

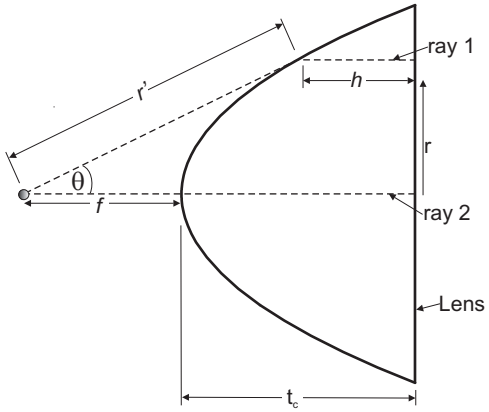


Figure 4.7: The time for the rays (shown as dashed lines) to travel to the back of the lens is the same for each ray shown.

Figure 4.7 shows a single-surface lens - so called because refraction takes place only at the first surface and passes normally through the second surface. As the incoming beam is assumed to be rotationally symmetric, the lens will be also. The equation for the lens surface is derived by equating the optical-path length from the source (which is modelled as a point source) to all points in the aperture plane, with the optical-path length along the axis. Referring to Figure 4.7, the following equation may be given

$$r' + nh = f + nt_c, \tag{4.4}$$

where f is the focal length, n is the refractive index of the dielectric lens material

and t_c is the central thickness of the lens. The surface of the lens is given by

$$r' = \frac{(n-1)f}{ncos(\theta) - 1}. \tag{4.5}$$

This equation describes the hyperbolic surface of revolution. In Cartesian form the shape of this lens (shown in Figure 4.7) is described by

$$h(r) = \frac{f\left(\sqrt{\frac{(n+1)r^2}{f^2(n-1)} + 1} - 1\right)}{n+1} \tag{4.6}$$

where $h(r)$ is the depth of the lens and r is the distance from the axis of symmetry[6].

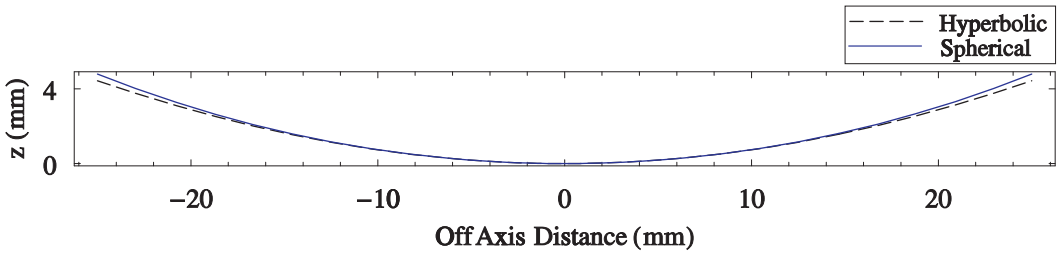


Figure 4.8: Comparison between spherical and hyperbolic lens shapes.

In our case, the lens must have a focal length such that a beam with a phase radius of curvature of R_{in} is converted to a beam of phase radius $R_{out}(=\infty)$[6]. In this case, it can be shown that[6]

$$\frac{1}{R_{out}} = \frac{1}{R_{in}} - \frac{1}{f} \tag{4.7}$$

The field has a radius of curvature of 130mm at the horn's aperture. Substituting $R_{in} = 130$mm and $R_{out} = \infty$ for this lens gives $f = 130$mm. For a hyperbolic lens made of HDPE with a refractive index of $n = 1.52$ the central thickness of the lens is 4.43mm. The profile for this hyperbolic lens is shown in Figure 4.8. A model of the horn and lens is shown in Figure 4.9. The hyperbolic lens can be approximated by a spherical surface, given by

$$h_r(r) = f(n-1)\left(1 - \sqrt{1 - \frac{r^2}{f^2(n-1)^2}}\right), \tag{4.8}$$

where r is the distance from the axis of symmetry, f is the focal length and n is the refractive index of the lens. A spherical lens, which approximates a hyperbolic surface, may be more straightforward to manufacture; however, this type of lens does not satisfy Fermat's principle for appreciable off-axis distances[6]. The phase shift due to the spherical lens, as a function of distance from the axis of symmetry,

is given by

$$\Delta\phi_s(r) = -\frac{2(n-1)\pi\left(t_c - z_s(r)\right)}{\lambda}. \tag{4.9}$$

Figure 4.10 shows that the phase error due to the spherical lens approximation ($\Delta\phi_s(r) - \Delta\phi_h(r)$, where $\Delta\phi_h(r)$ is the phase due to a hyperbolic lens) is negligible for off-axis distances of less than 10mm; most of the power that the horn couples to is within this region. For a spherical lens made of HDPE with a refractive index of n = 1.52 the central thickness is 4.79mm; its profile is shown in Figure 4.8.

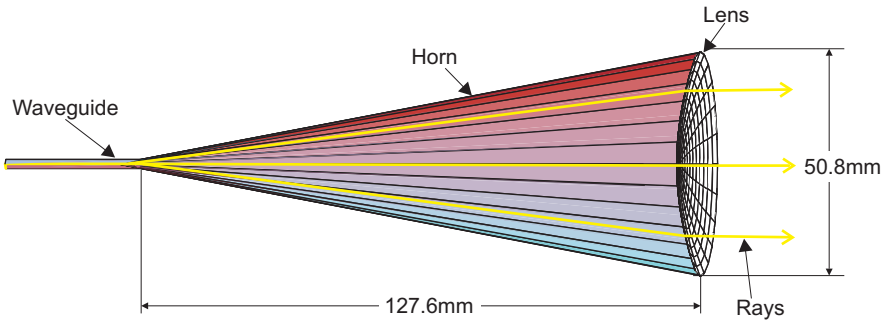


Figure 4.9: MBI feed-horn cross section with a hyperbolic HDPE lens with its curved surface facing the source. (Drawn to scale.)

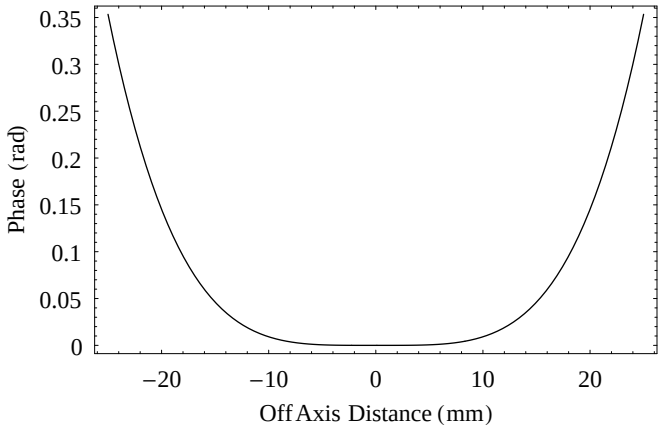


Figure 4.10: Phase error introduced by using a spherical approximation to a hyperbolic lens.

The third type of lens analysed here is one where the surface facing the source is planar, as shown in Figure 4.6 (c). This is considered to be a two surface lens, since neither surface coincides with the equiphase surface of the beam, and therefore both surfaces redirect the rays[6]. The equation which describes this surface is derived in the same way as the lens with its curved surface facing toward the source. The surface of this lens is described by

$$h(r) = \frac{f + (n-1)t_c - \sqrt{f^2 + r^2}}{n\sqrt{\frac{r^2}{n^2\left(f^2+r^2\right)-r^2}} + 1 - 1}. \tag{4.10}$$

The refractive index for HDPE n= 1.52, r is the off axis distance and f=130mm.

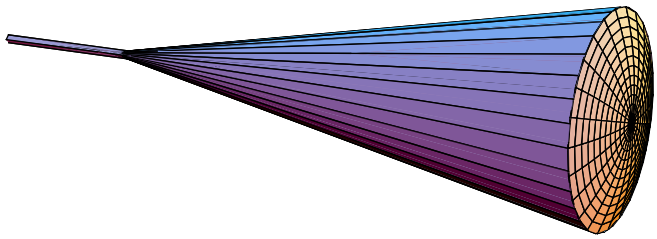


Figure 4.11: Horn with a plano-convex HDPE lens with its curved surface facing away from the source.

	Hyperbolic	Spherical	Plano-Convex
c	0.015	0.015	0.015
K	-2.310	0.000	-1.977

Table 4.3: Parameters describing the lens sag for a hyperbolic, spherical and plano-convex lens.

In optical systems, this lens has the advantage that neither surface coincides with an equiphase-front, and therefore power reflected from the surface does not couple to the incident beam; however here, because the lens in contact with the horn, any reflected radiation is coupled with the horn, regardless of the lens' shape. Also, this type of lens does not interfere with the waveguide modes, as it does not protrude into the horn[6]. Figure 4.11 shows a model of the horn with this plano-convex lens.

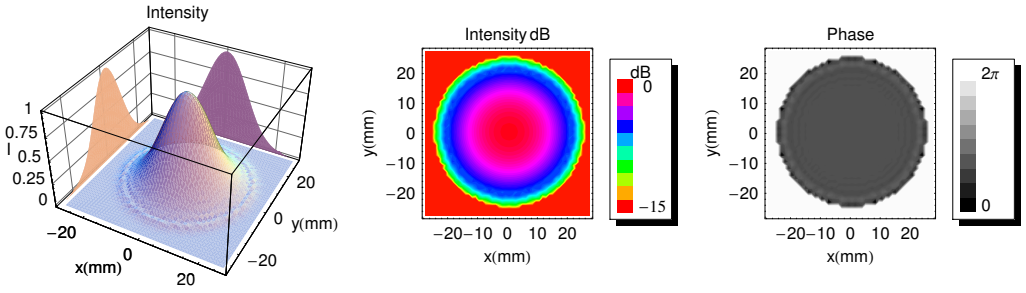
The designs considered, in order of increasing complexity, are the spherical, hyperbolic and plano-convex lens (shown in Figure 4.6). The plano-convex lens is the preferred option from an optical point of view. However, due to physical constraints in the cryostat, the curved surface must face toward the source.

The lens design was analysed using scalar physical optics in the commercial software package Zemax[50]. Zemax was used with a fundamental Gaussian approximation to describe the horn beam. The lens is defined in Zemax by its sag, given by

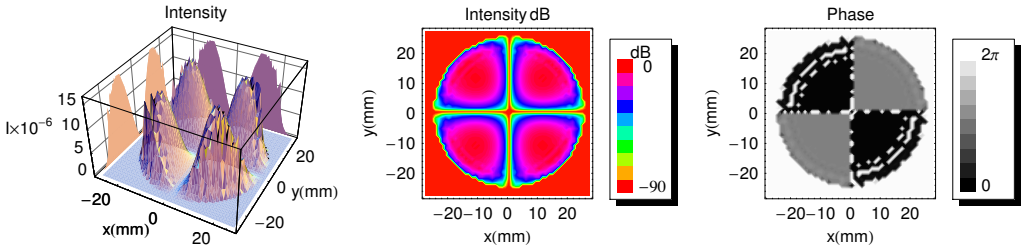
$$h(r) = \frac{cr^2}{1 + \sqrt{1 - (1 + K)c^2r^2}}, \tag{4.11}$$

where c is the curvature (the reciprocal of the radius of curvature), r is the radial coordinate in lens units and K is the conic constant. The parameters, found by a least square fit routine, for each of the lenses are given in Table 4.3. The co- and cross-polar near-field and far-field of the horn with this hyperbolic lens, as predicted using Zemax (which models the effects of diffraction) are shown in Figure 4.12. Ideally there would be no cross-polar component to the field; the cross-polar component calculated here contaminates the orthogonal polarisation.

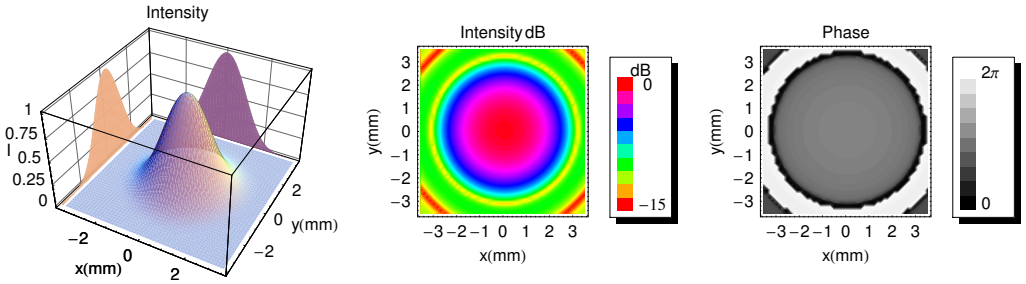
Table 4.2 lists the beam radius at the primary and secondary mirrors and the focal-plane, both with and without feed-horn lenses; with the lenses, the radius had



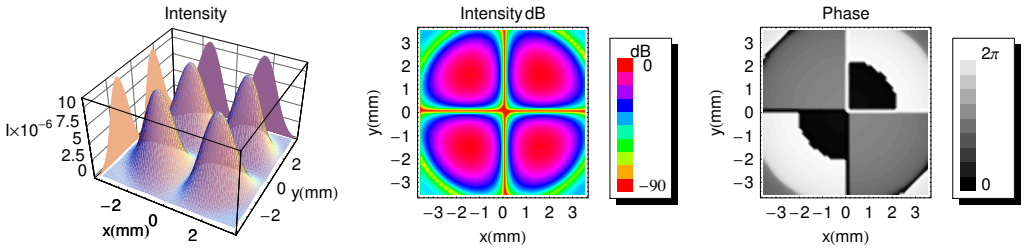
(a)



(b)



(c)



(d)

Figure 4.12: The far-field radiation pattern of the horn with a hyperbolic lens, which has the parameters shown in Table 4.3 (a) normalised co-polar and (b) normalised cross-polar field immediately after the lens. (c) Co-polar and (b) Cross-polar far-fields. The far-field was calculated at a distance from the horn where 1mm$\equiv 2.3^{\circ}$.

Figure 4.13: Feed-horn with lens.

been reduced at each plane, reducing truncation at the secondary mirror. Figure 4.13 shows a photograph of one of the feed-horns with the spherical lens, actually manufactured. Grooves have been machined onto the surface, to simulate an antireflection coating and reduce the amount of reflected power[49]. The spherical lens has been chosen over the hyperbolic because it is simpler to manufacture.

4.4 Detector Horn Lenses

In addition to the feed-horn lenses, other lenses are designed to improve the coupling of the telescope field at the focal-plane to each of the detector horns. Using phase imposing surfaces at a given plane, the best possible coupling between the field from the detector horns and field from the secondary is achieved when the phases of the two fields are matched. Lenses were designed to flatten the field from the sky and the field from the horn, using both surfaces of the lens, thereby matching them. Using the thin lens approximation, lenses can be considered to create a phase delay proportional to their thickness; therefore, in order to flatten a phase-front, the lenses have the same profile as the phase of the field. The first design analysed here uses two separate lenses; one to flatten the phase of the field from the telescope and another to flatten the phase of the field from the horn. A further lens design option was analysed where the detector lenses have a single phase-flattening surface, to flatten the phase from both the detector horns and the telescope field. The calculated increase in coupling is the same whether the phase of the horn and telescope field are flattened using one or two surfaces; however, both designs are analysed here in order to compare them for ease of manufacture. Figure 4.14 shows a schematic of the individual HDPE lenses used to flatten the phase of the detector horns. Also shown is a lens to flatten the phase of the field from the telescope which may be made from

a continuous piece of HDPE. The design of the lenses to flatten the field from the 200mm baseline feeds is described in Section 4.4.2. The actual field incident on the focal-plane, however, will be due to two baselines, with four horns. The design of lenses that flatten the field due to both baselines is described in Section 4.4.3.

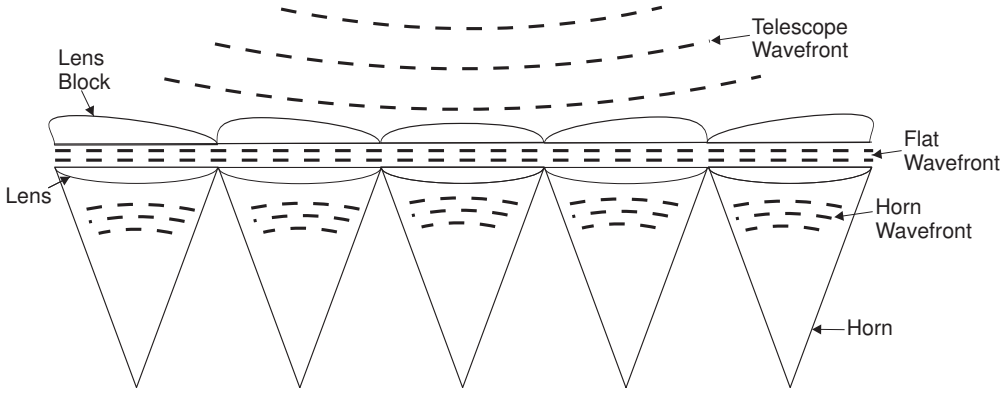


Figure 4.14: Array of horns with integrated lens for flattening the phase of the field from the telescope and individual lenses to flatten the phase of the field from the horns.

4.4.1 Lenses to Flatten the Field from the Detector Horns

The lenses to flatten the phase-front of the individual detector horns are designed using the same techniques as was used to design the feed-horn lenses shown in Section 4.13. The detector horns have a radius if 12.7mm and a slant length of 54.3mm. The conic constant and eccentricity of this lens for each of the detector horns is 0.039 and 0.000 respectively. Spherical source-facing lenses (profile shown in Figure 4.15) were chosen because they are easier to clamp to the lens, with the filters on the front. A photographs of the detector horns with the detector horn lenses is shown in Figures 4.16 and 4.17.

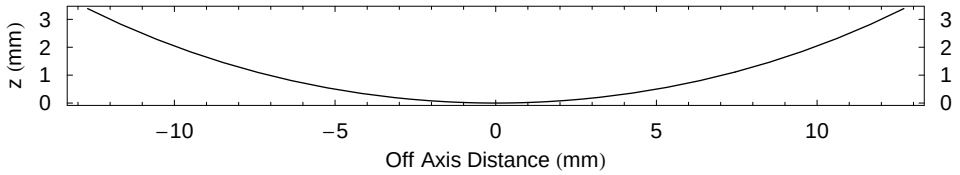


Figure 4.15: Profile of the spherical detector horn lens, of central thickness 3.4mm.

4.4.2 Lens to Flatten the Field from a Single Baseline

The interference pattern at the focal-plane is due all the baselines shown in Figure 4.2. For a preliminary investigation, lenses to flatten the phase of the field from only the two longest baselines will now be considered. For simplicity five detector horns are considered, which are aligned with the feed-horns as shown schematically in Figure 4.18.

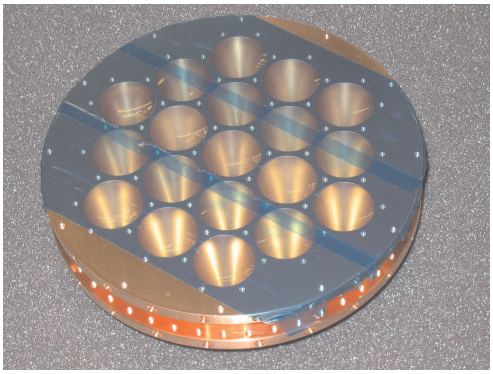
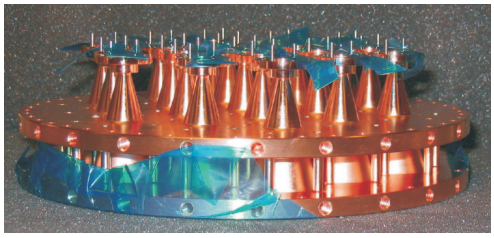


Figure 4.16: *Left* lenses and holder for detector horn array, *right* bolometer horn array: top view.



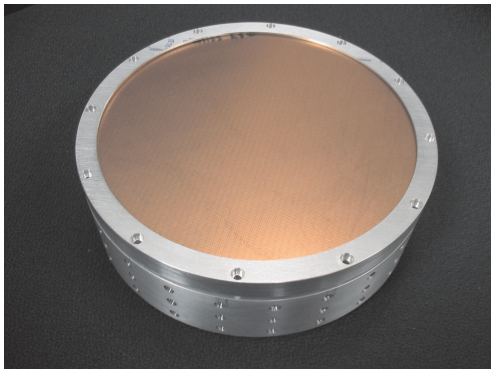


Figure 4.17: *Left* bolometer horn array: bottom view, *right* bolometer unit with filters and filter clamp.

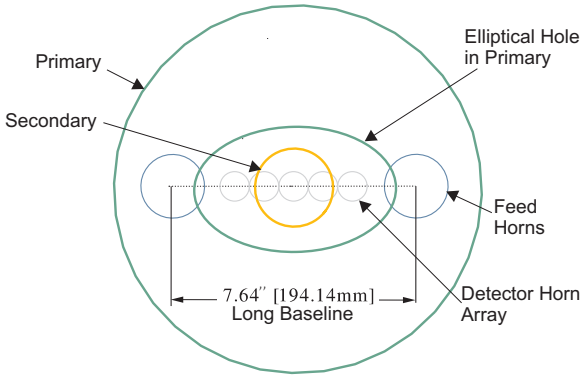


Figure 4.18: Lenses to flatten the phase of the field from the single "long" baseline at the five detector horns, which are in alignment with the feed-horns, are designed.

The lens is designed directly from the phase of the field, so a finely sampled field is required. The resolution of the field calculated if Zemax is used is limited by the amount of RAM available on the computer on which it is run. The large guard band that Zemax requires to propagate the field using its physical optics routine and the large area of the interference pattern at the focal-plane, put a lower limit of 5.53mm on the sampling interval for the computer, with 4GB of RAM, on which it is run. Initially, as this is an insufficient sampling rate to design a lens with a smooth profile, and for simplicity, the field is approximated by two tilted fundamental Gaussian beams to simulate the interference pattern at the focal-plane for one base line, as shown in Figure 4.19; later in this chapter, the re-sampled field predicted by Zemax (after propagation through the optical system) will be used to define the lens shape.

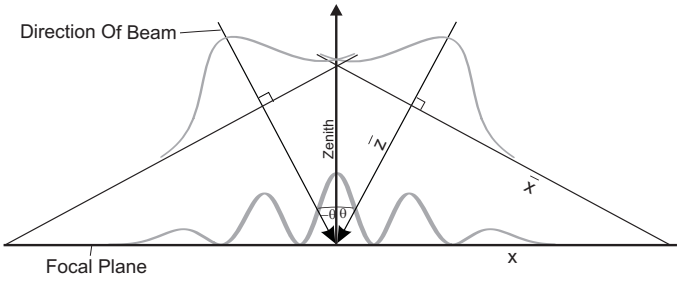


Figure 4.19: An illustration of the interference pattern created by the field from the two feed-horns of a single baseline.

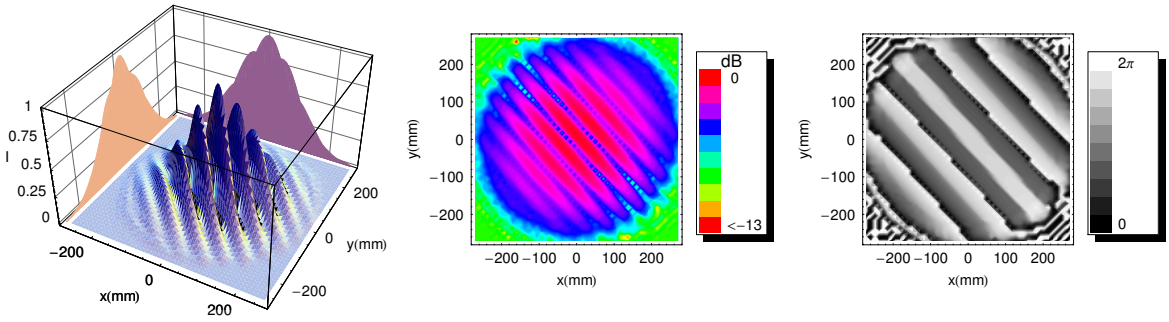


Figure 4.20: The field at the focal-plane for the long base line. (This field has the same orientation for the x and y axis as is shown in Figure 4.2.) *Left* normalised intensity with intensity projected onto the x and y axis, *middle* the intensity in dB, *right* the phase.

Due to the symmetry of this simplified system, the field from both feed-horns have equal radii of curvature and the angle that the beams make with the zenith are θ and $-\theta$, as illustrated in Figure 4.19. Zemax calculated θ to be 1.18° by ray tracing with a single ray from the feed-horn to the focal-plane. Using a fundamental GBM to simulate the field from the field Zemax gives the radius of the beams and their radii of curvature at the centre of the focal-plane. These are used to determine the waist radius, w_0, and its location, z, for the Gaussian beam in order to compute

the field due to the interference of the two beams at the focal-plane. The waist is given by

$$w_0 = \frac{w}{\sqrt{1 + \frac{\pi^2 w^4}{R^2 \lambda^2}}}[6] \tag{4.12}$$

and the propagation distance from the waist is given by

$$\bar{z} = \frac{R\left(1 \pm \sqrt{1 - \frac{4\pi^2 {w_0}^4}{R^2 \lambda^2}}\right)}{2}[6] \tag{4.13}$$

These give $w_0 = 2.2$mm and $\bar{z} = 406.3$mm (the lower value given by Equation 4.13). The two beams are then added coherently to give the interference pattern shown in Figure 4.20.

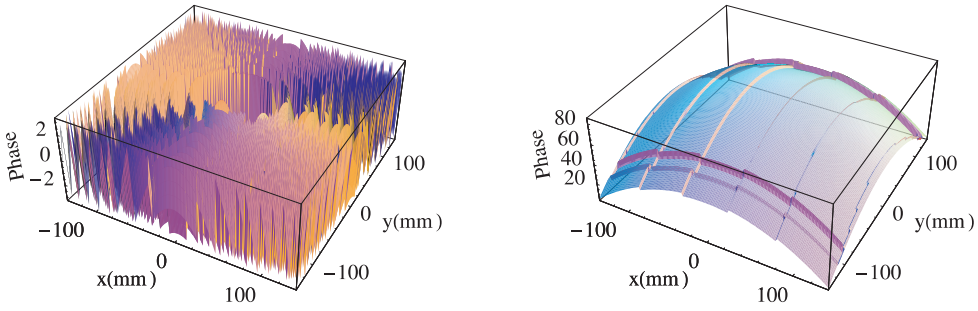


Figure 4.21: Phase due to interference pattern of the 2 beams of the 200mm baseline *left* wrapped phase (radians), *right* unwrapped phase (radians).

Figure 4.21 *left* shows the phase of the interference pattern at the focal-plane, as a result of the interference between the two Gaussian beams; Figure 4.21 *right* shows this phase unwrapped. It is this *unwrapped* phase that is used to design the lens, in order to avoid unnecessary discontinuities in the lens profile. The location of the detector horns (each separated by 25.4mm: the diameter of the horns) and the intensity of the field with which they are to couple, are shown in Figure 4.22, for a cut across the centre of the horns.

In order to flatten the phase, the lens must create a phase delay equal to the phase of the field incident at the focal-plane. The thickness of the lens, $h(x, y)$ is given as a function of the phase of the field $\phi(x, y)$ by

$$h(x, y) = \frac{\lambda\phi(x, y)}{2\pi(n - 1)} \tag{4.14}$$

where the refractive index, n, is 1.52 for HDPE and λ is 3.33mm. Figure 4.23 shows the profile of the lenses across the focal-plane and the phase and unwrapped phase from which this shape was determined. Phase jumps of π are present and cannot be removed by unwrapping the phase, as this only removes phase jumps of 2π. Figure 4.24 *top* shows the shape of the lens, the cylindrical slab common to each lens creates

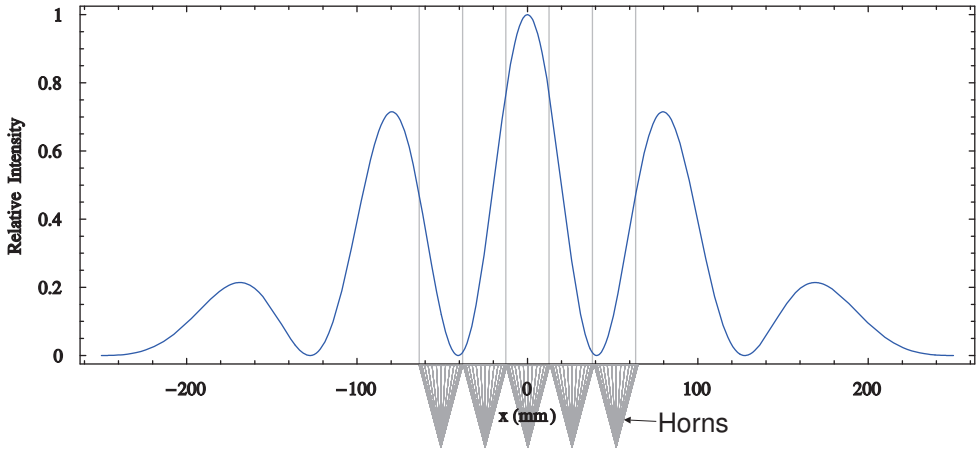


Figure 4.22: Cut across x axis of the 200mm normalised intensity baseline pattern at the focal-plane. The grid lines mark the extent of each of the detector horns.

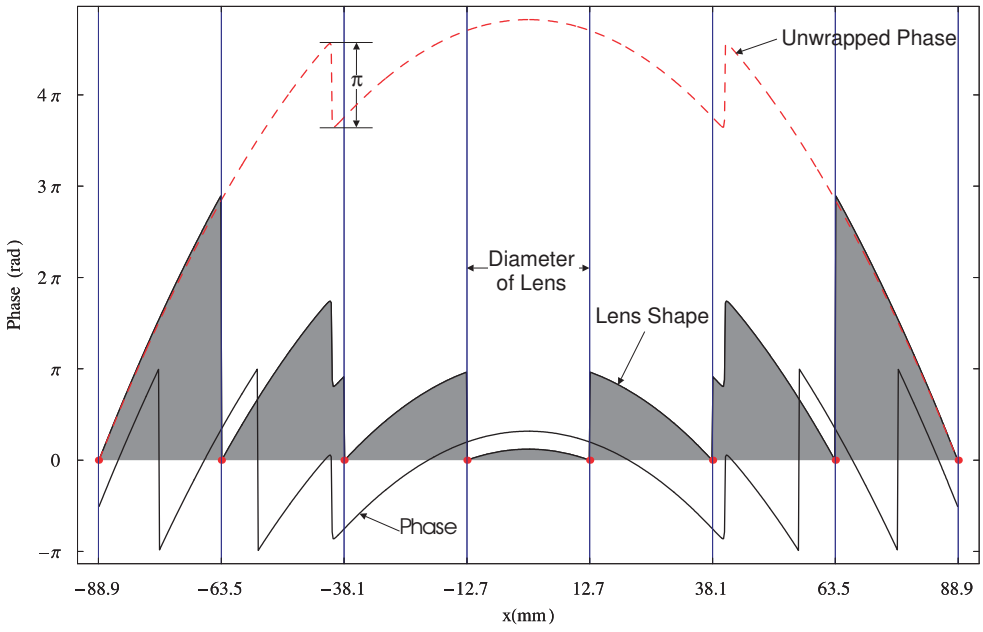


Figure 4.23: Profile of the lens across the centre of the horns for 200mm baseline of Figure 4.18. The vertical lines along the x axis indicate the position of the edge of the horns.

a constant phase delay across the field, and can therefore be removed in order to reduce the thickness of the lenses. The final lens shape to correct the phase for one baseline is shown in Figure 4.24 *bottom*.

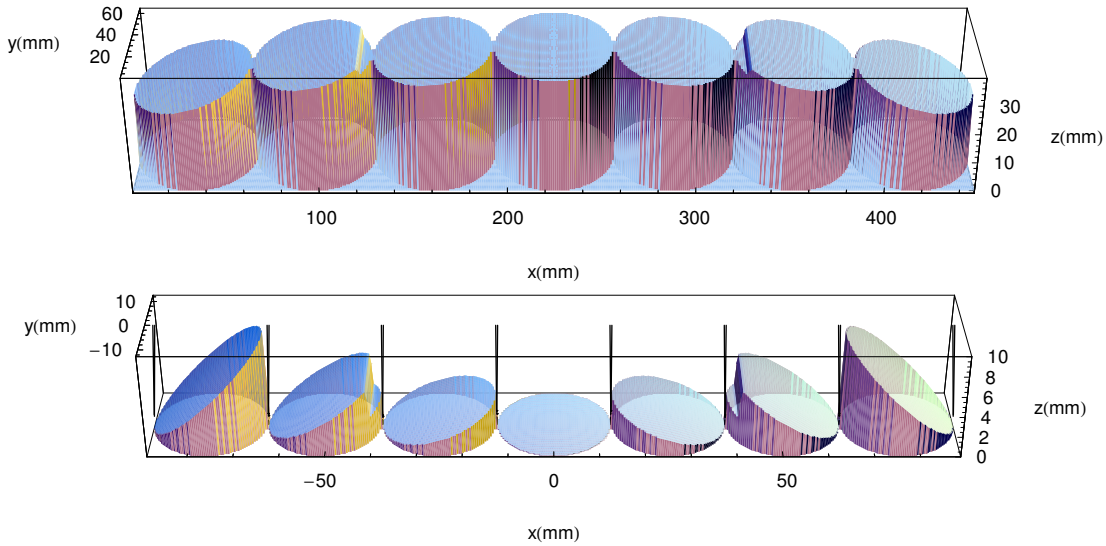


Figure 4.24: *Top* the lenses for the 200mm baseline with cylindrical section remaining, and *bottom* with the cylindrical section removed. Units in are millimetres. The scale of the z axis in relation to the x and y axis has been increased to show the lens profile. The vertical lines emanating from the surface of the lenses show the directions of the two beams which cause the interference pattern; the angle the lens surfaces make with the azimuth is sufficiently small to ignore any possible effect of shadowing due to the discontinuities.

4.4.3 Lens to Flatten the Field from Two Baselines

In the previous section, the field due to the fundamental Gaussian beams predicted by Zemax were used to determine the field at the focal-plane. Now, Zemax is used to propagate the field from the four horns of the two baselines (the longest and the shortest) using Fourier optics. This produces a more accurate field, as it models any distortions of the field due to the mirrors. As mentioned previously, the sampling interval of the field given by Zemax is limited to 5.53mm, and so the field is re-sampled by interpolation in order to achieve a finer sampling. The correct orientation of the detector horns with respect to the rest of the optics and the direction of the beam polarisation shown in Figure 4.2 are used throughout this section. The fields are coherently summed at the focal-plane to give the field due to all four feeds. The intensity and phase for the co-polar and cross-polar patterns are shown in Figure 4.25. Figure 4.26 shows the field at the focal-plane with the location of the array of detector horns superimposed. The numbering of the horns on Figure 4.26(b) is used throughout the rest of this chapter. Figure 4.27 shows the overlay of field due to the long and short baselines, these interfere to give the

field from both together, also shown. The direction cosine of the beams from their point of intersection at the centre of the focal-plane are shown in Table 4.4. The sampling rate of the field is increased by interpolating the field and resampling. The interpolation method used first reconstructs the field using Gaussian-Hermite modes, so that the field is represented as an analytical function, which can then be sampled at an arbitrary rate. The representation of the field in terms of a Gaussian-Hermite basis set is discussed in detail in Chapter 1. The expansion coefficients for a given function, $\mathrm{Ef}(x, y, w_0, 0)$, is given by the overlap integrals

$$A_{m,n} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi^*_{m,n}(x, y, 0, w_0,)\mathrm{Ef}(x, y, 0, w_0)dxdy \tag{4.15}$$

where $\psi_{m,n}(x, y, 0, w_0)$ are the Gaussian-Hermite modes and $\mathrm{Ef}(x, y, 0, w_0)$ is the field from the telescope at the focal-plane[16]. The reconstruction of the field is given by

$$\mathrm{Ef}(x, y) = \sum_{m,n} A_{m,n}\psi_{m,n}(x, y, 0, w_0). \tag{4.16}$$

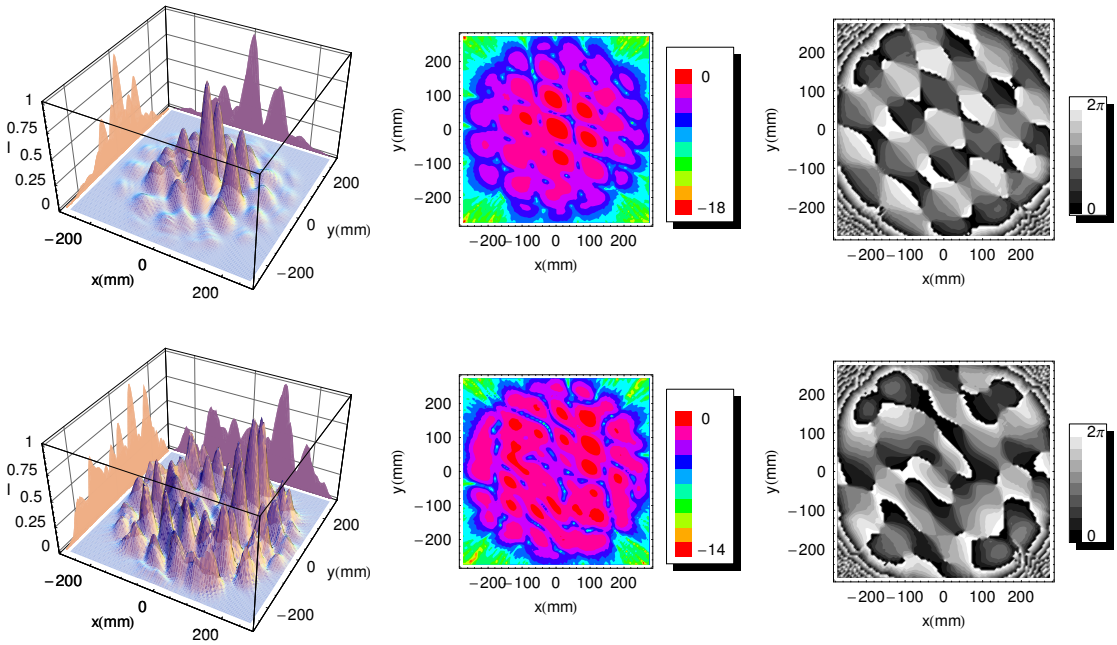


Figure 4.25: *Top* the co-polar, and *bottom* the cross-polar field at the focal-plane. *Left* intensity, *middle* contour plot of intensity (dB), *right* phase. The maximum value of intensity of the co-polar component is 1.66×10^3 greater than the cross-polar maximum.

The interference pattern at the focal-plane is complicated, and therefore it would be expected that a high number of modes would be required to reconstruct it. However, by choosing a mode set that has an appropriate spherical phase-front,

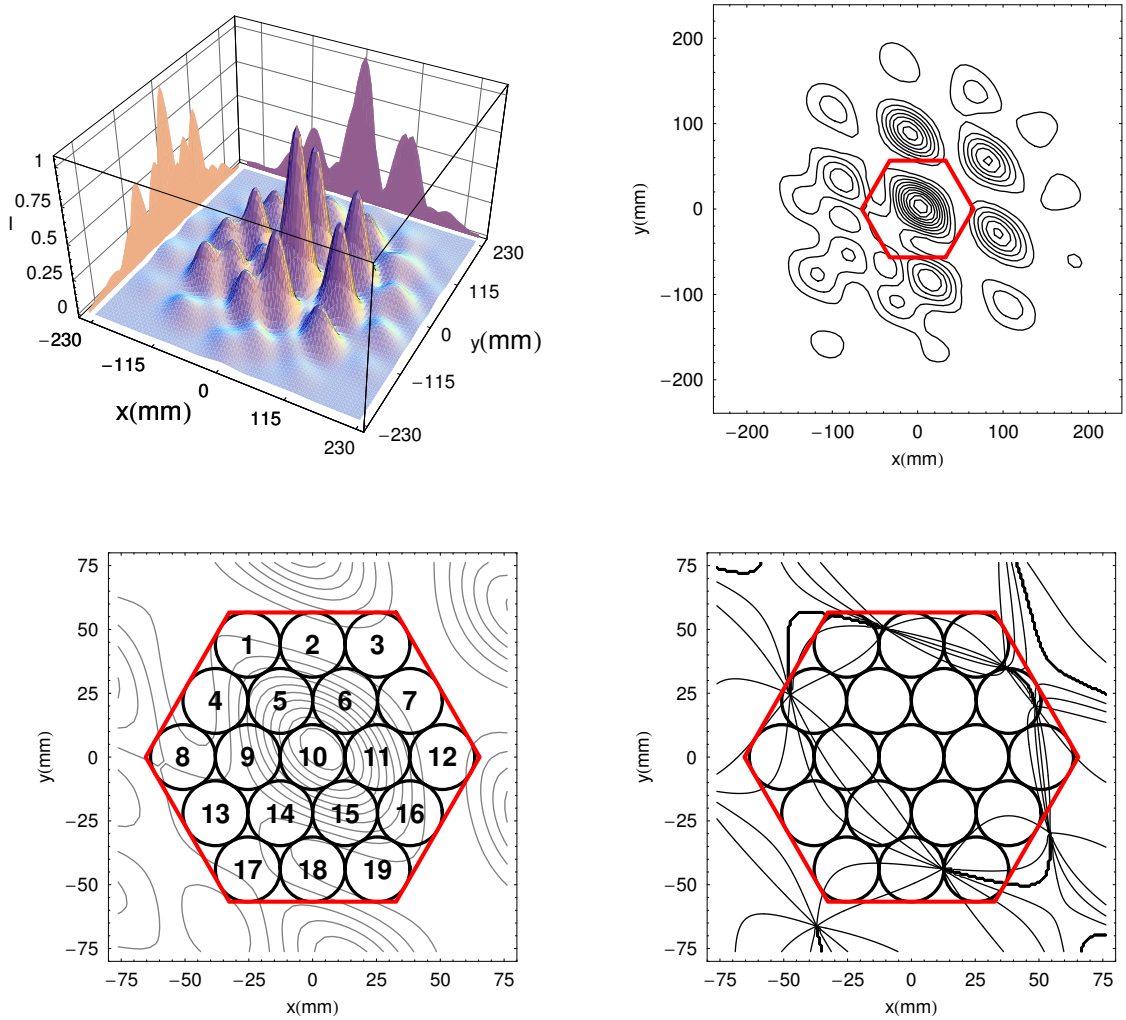


Figure 4.26: The field at the location of the detector horn array. *Top left* Normalised intensity (with projection onto the x and y axes), *top right* contour plot of intensity with outline of detector horn array superimposed, *bottom left* intensity over the detector horn area, *bottom right* contour plot of phase over the detector horn.

99% of the power in the field was reconstructed with an 11×11 array of modes, as illustrated in Figure 4.28. The procedure for determining this mode set is discussed in Chapter 6.

	l	m	n
Upper long baseline	0.0243	0.0213	0.9995
Lower long baseline	-0.0239	-0.0218	0.9995
Upper short baseline	-0.0057	0.0202	0.9998
Lower short baseline	0.0161	-0.0135	0.9998

Table 4.4: The direction cosines (l, m, n) of the beams from the for feed-horns at the centre of the focal-plane, relating to the axis (x, y, z).

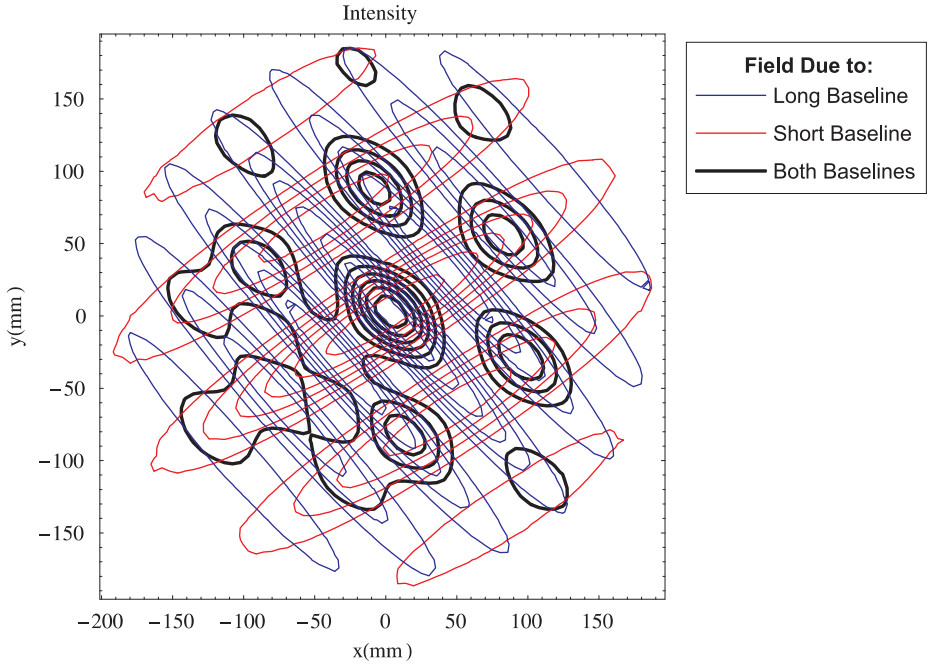


Figure 4.27: The focal-plane intensity due to the short baseline, the long baseline and both baselines combined.

In order to eliminate discontinuities on the surface of the lens, it is necessary to unwrap the phase of the field. The field is therefore unwrapped using the following method: firstly, the phase is sampled and the central column, $x = 0$mm, shown in red in Figure 4.29 *top*, is selected. This line is unwrapped by adding or subtracting 2π if a 2π phase jump occurs, in order to make it continuous, as shown in Figure 4.29 *middle*. Then each row is unwrapped as shown in Figure 4.29 *bottom left*. Each column is then aligned with the unwrapped row at $x = 0$mm, to give the unwrapped field of Figure 4.29 *bottom right*. Note that discontinuities are still present in the phase. These do not occur when the phase of the field from each baseline is unwrapped separately, and arise due to phase vortices in the field caused by the interference of the four beams.

The lenses designed using the unwrapping algorithm described above have discontinuities on the lens surface. The plots of Figure 4.30 show the unwrapped

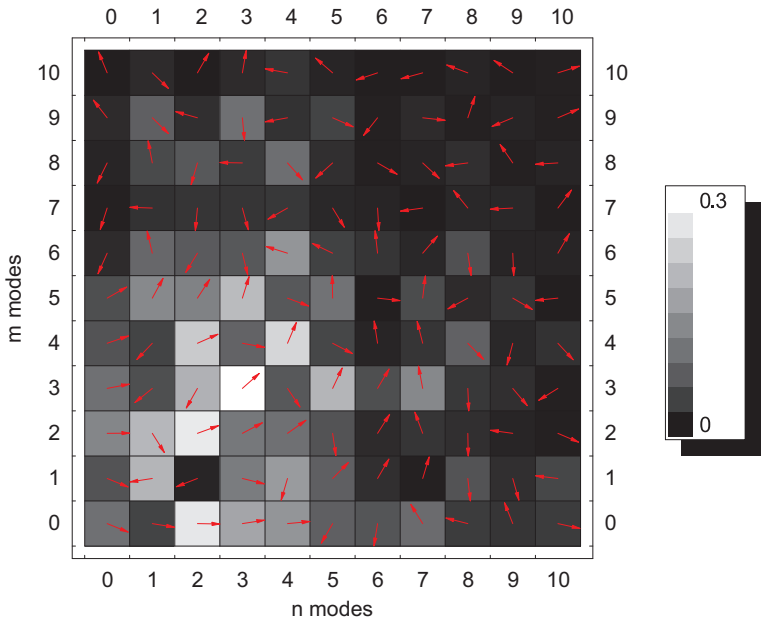


Figure 4.28: Gaussian-Hermite mode coefficients for the reconstruction of the focal-plane field. White represents the largest and black the small value of mode coefficient. The arrows represent the phase of the modes.

phase: unwrapped relative to *left* $x = 0$mm as described above and *right* the line $x = 125$mm. A shift in the locations of the discontinuities is observed. A partial solution, to have fewer discontinuities on the lens surface, is to unwrap the phase at each lens individually; however, discontinuities will still occur if a phase vortex is located on the lens surface.

The location of the phase vortices can be observed clearly in Figure 4.26 *bottom right*. The phase for each lens is unwrapped individually, in order that only the lenses which have a phase vortex on their surfaces contain a discontinuity. The shape of the lenses is shown in Figure 4.31, and a cross section is shown in Figure 4.32

An alternative to using separate lenses to flatten the telescope and detector wavefronts separately is to use one lens surface to match the wavefronts simultaneously, and thereby decrease the cross-polar component due to reflections at each lens surface. In this case the phase-front of the field from the telescope needs to be converted to the spherical phase-front produced by the horn at the aperture. A meniscus lens (Figure 4.34) can be used to do this. The surface facing the horn is spherical to match the phase-front of the horn. This lens is considered to be single-surface because refraction only takes place on the surface that faces the telescope. Figure 4.33 (a) shows this type of lens; its profile is the same as the phase of the interference pattern between the horn and the telescope field.

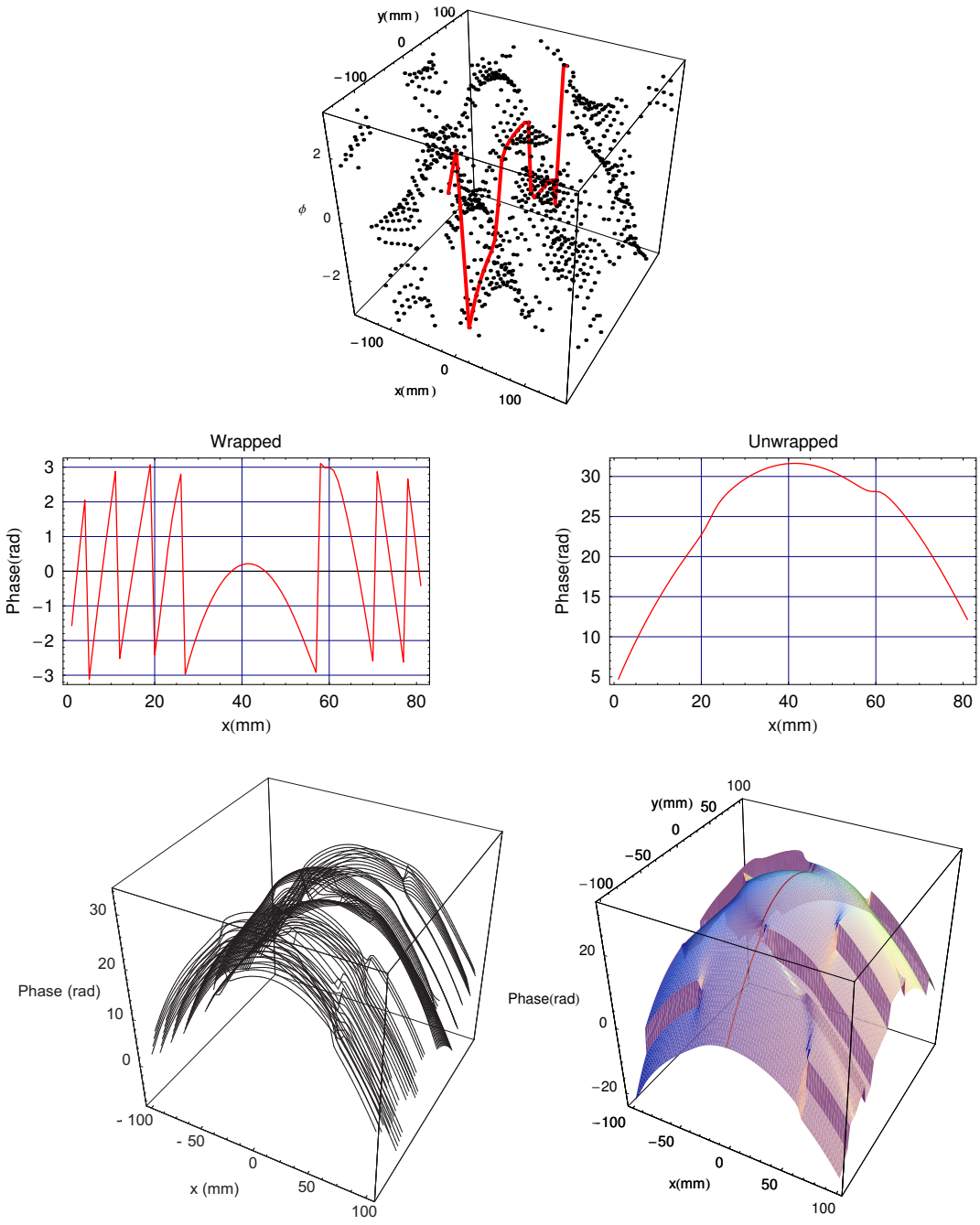


Figure 4.29: Steps of unwrapping of phase *Top* the centre of the sampled field is selected along the line in red, *middle* this column of sampled field data is unwrapped, *bottom left* each row is unwrapped, *bottom right* each row is aligned with the unwrapped column, at x=0mm (shown in red).

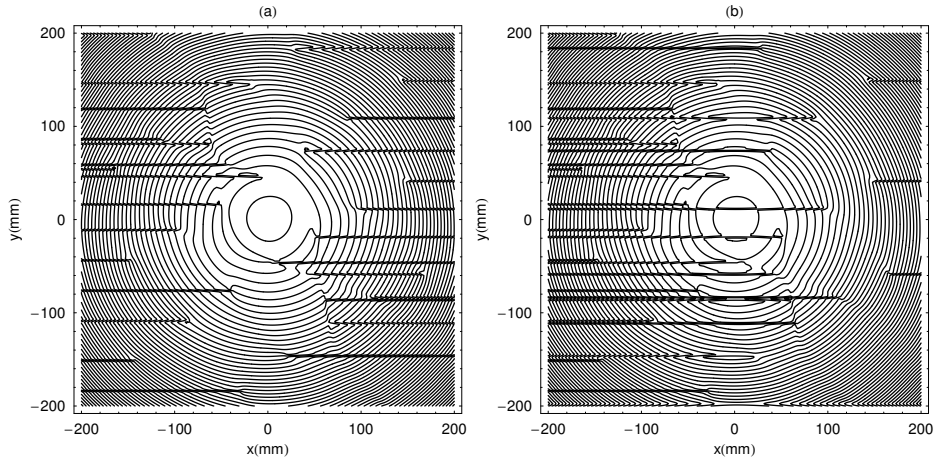


Figure 4.30: The location of the possible discontinuities depends on the column chosen as a reference in Figure 4.29. Two possible solutions are shown here.

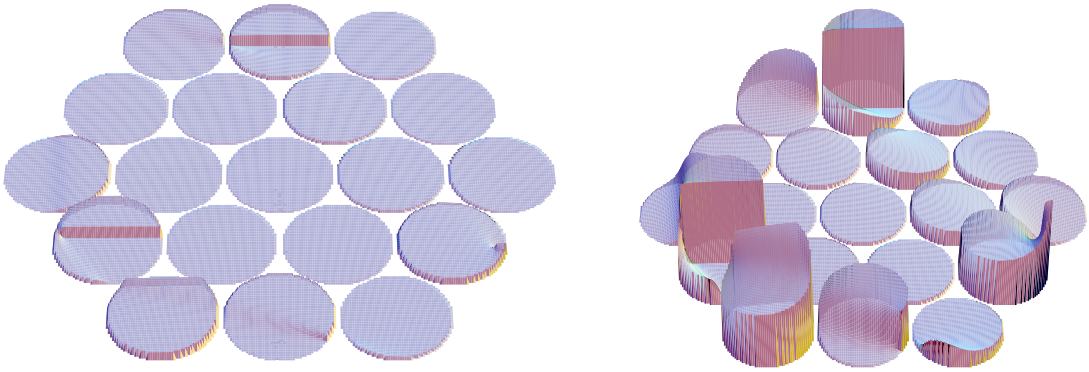


Figure 4.31: Shape of the lenses designed to match the field from the two baselines. *Left* to scale, *right* z axis enlarged. Discontinuities occur on the surface due to phase vortices in the field. The maximum height of the lenses is 6.78mm.

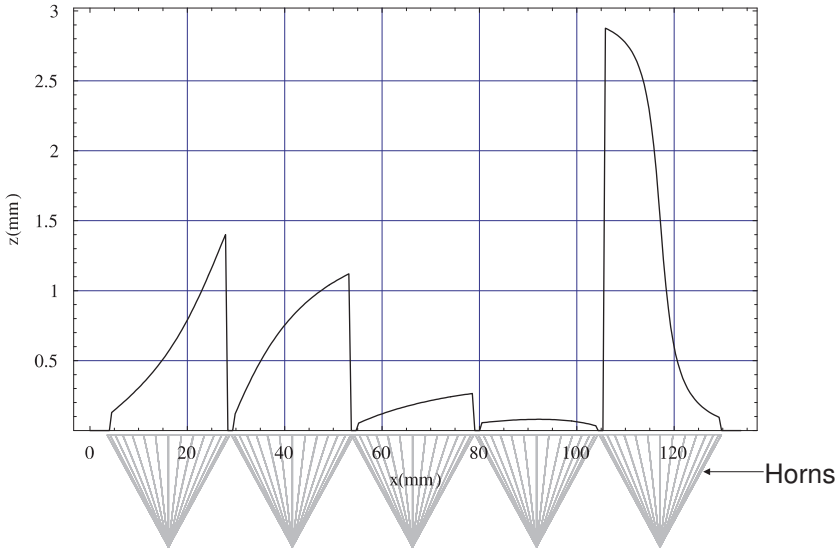
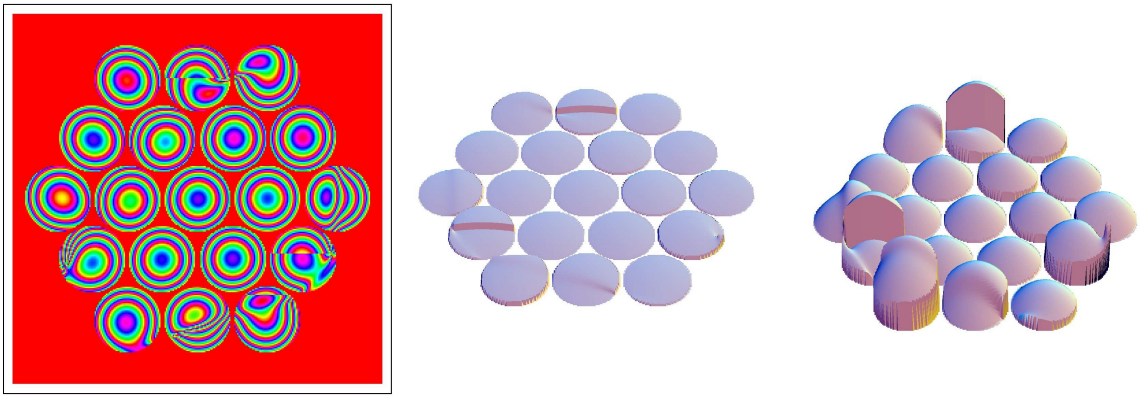
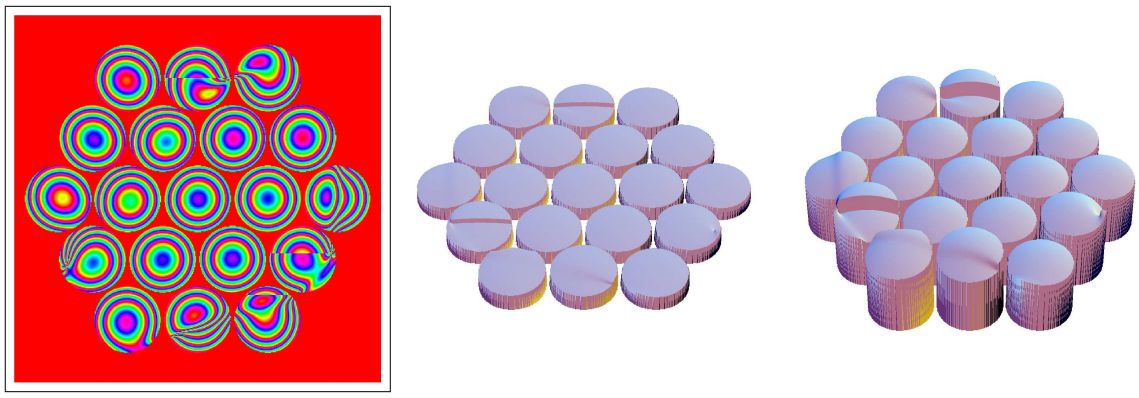


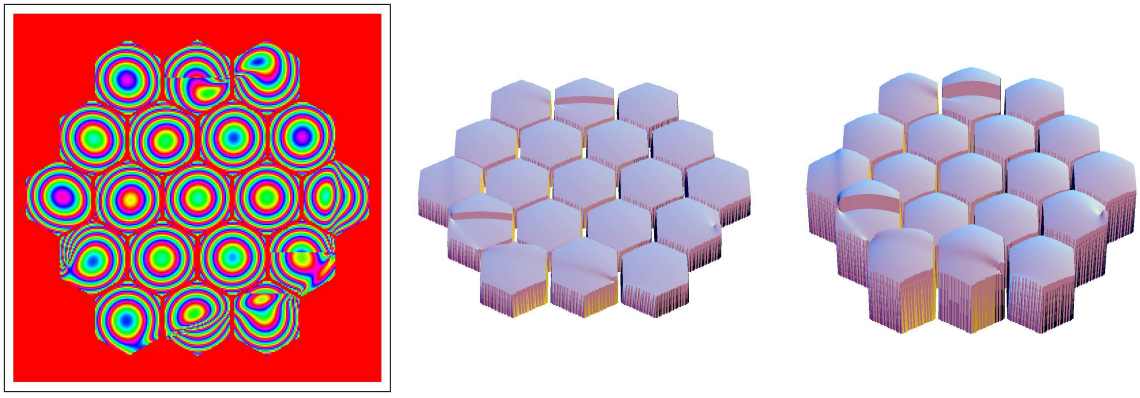
Figure 4.32: Shape of lenses along centre of row of horns 8-12.



(a)



(b)



(c)

Figure 4.33: Shape of single-surface meniscus lenses. (*Left* contour plot, *middle* profile of lens, *right* profile of lens with z axis exaggerated.) ((a) Circular lens with no extended section, (b) Circular lens with 22mm extension to improve coupling, (c) Hexagonal lens with 22mm extension to improve coupling).

4.5 Coupling Analysis

The coupling of the field from the telescope to the detector horns has been analysed for various horn arrangements (the results are summarized in Table 4.5). The coupling of the field from the telescope to the detector horns without a lens is given by the following equation:

$$C = \left| \frac{\int_0^{2\pi} \int_0^{ap} E_H(r,\theta,0) E_I^*(r,\theta,0) r \, dr \, d\theta}{\sqrt{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,0)\right|^2 r \, dr \, d\theta} \sqrt{\int_0^{2\pi} \int_0^{\infty} \left|E_I(r,\theta,0)\right|^2 r \, dr \, d\theta}} \right|^2 , \tag{4.17}$$

where E_H is the horn field and E_I is the field incident on the focal-plane, *ap* is the aperture radius of the horn and the coordinate system is centred on the centre of the horn aperture. The numerator of this equation is the overlap integral of the field incident on the detector horn aperture with the conjugate of the field from the telescope. The denominator normalises the field so that the coupling is given as a fraction of the power coupled by the horn to the entire telescope field.

As the effect of the detector horn lens is to flatten the phase of both the horn and the telescope field, the coupling with the phase flattening lens is given by

$$C = \left| \frac{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,0)\right| \left|E_I(r,\theta,0)\right| r \, dr \, d\theta}{\sqrt{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,0)\right|^2 r \, dr \, d\theta} \sqrt{\int_0^{2\pi} \int_0^{\infty} \left|E_I(r,\theta,0)\right|^2 r \, dr \, d\theta}} \right|^2 , \tag{4.18}$$

and the coupling with just a lens to flatten the phase of the field due to horn at the aperture, is given by

$$C = \left| \frac{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,0)\right| E_I(r,\theta,0) r \, dr \, d\theta}{\sqrt{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,0)\right|^2 r \, dr \, d\theta} \sqrt{\int_0^{2\pi} \int_0^{\infty} \left|E_I(r,\theta,0)\right|^2 r \, dr \, d\theta}} \right|^2 . \tag{4.19}$$

In Equations 4.18 and 4.19 the modulus of the telescope and horn fields are used to model the effect of the lenses. The effect on coupling of the two arrangements is shown in Table 4.5. The lenses to flatten the phase of *both* the telescope field and horn field increase the coupling of the horns by an average of 2.4 times over that of the bare horn, whereas, lenses which flatten the field of just the horn increase the coupling by a factor of 1.7.

	Absolute Coupling		Normalised w.r.t. Ideal					
Horn	Bare	Ideal	Bare	HL	HTL	Circ	Hex	Tilted
1	1.3×10^{-3}	3.5×10^{-3}	0.37	0.59	0.62	0.72	0.78	0.37
2	1.9×10^{-4}	2.3×10^{-3}	0.08	0.14	0.36	0.53	0.59	0.25
3	1.8×10^{-4}	9.8×10^{-4}	0.19	0.29	0.47	0.59	0.66	0.22
4	9.5×10^{-4}	3.5×10^{-3}	0.28	0.45	0.51	0.62	0.68	0.34
5	8.3×10^{-3}	1.9×10^{-2}	0.43	0.71	0.71	0.82	0.89	0.44
6	5.7×10^{-3}	1.4×10^{-2}	0.40	0.65	0.66	0.77	0.83	0.41
7	3.2×10^{-4}	1.5×10^{-3}	0.22	0.38	0.39	0.48	0.54	0.27
8	1.7×10^{-3}	4.6×10^{-3}	0.37	0.63	0.68	0.87	0.87	0.39
9	4.0×10^{-3}	1.1×10^{-2}	0.37	0.61	0.65	0.84	0.84	0.41
10	1.3×10^{-2}	3.0×10^{-2}	0.44	0.71	0.72	0.90	0.90	0.44
11	7.4×10^{-3}	1.8×10^{-2}	0.42	0.69	0.69	0.86	0.86	0.43
12	4.9×10^{-5}	1.2×10^{-3}	0.04	0.08	0.38	0.64	0.64	0.19
13	3.1×10^{-3}	7.5×10^{-3}	0.42	0.70	0.71	0.83	0.90	0.42
14	2.7×10^{-3}	8.2×10^{-3}	0.34	0.56	0.63	0.76	0.83	0.40
15	6.3×10^{-3}	1.5×10^{-2}	0.42	0.68	0.68	0.79	0.85	0.43
16	2.4×10^{-3}	6.0×10^{-3}	0.39	0.63	0.64	0.74	0.79	0.40
17	2.5×10^{-3}	5.9×10^{-3}	0.42	0.69	0.71	0.83	0.90	0.43
18	5.5×10^{-4}	2.8×10^{-3}	0.20	0.33	0.50	0.64	0.71	0.22
19	3.7×10^{-4}	1.9×10^{-3}	0.19	0.31	0.48	0.61	0.69	0.20

Table 4.5: Absolute coupling: (Bare: without any lens, Ideal: maximum possible coupling). Values are also given that are normalised with respect to an ideal horn, which is one that would couple all the radiation incident across its aperture. Bare: horn without a lens, HL: Horn lens to flatten the field from the horn, HTL: Horn and telescope lenses to flatten the phase from both the horn and telescope field, Circ: with a circular lens offset by a vertical distance to maximize coupling, Hex: Hexagonal lens offset by a vertical distance to maximize coupling, tilted: horn without a lens tilted to maximize coupling, as illustrated in Figure 4.37.

An ideal horn - one which would couple *all* the radiation incident across its aperture - is used as a baseline with which to compare the coupling of the other horns; the values of coupling in Table 4.5 are given as a ratio of the ideal coupling. This *ideal* horn would produce a field identical to that incident on in it, in order to couple 100% of the of this incident field. An ideal horn would therefore have a normalised coupling of 1. The coupling of this *ideal* horn is given by

$$C = \left| \frac{\int_0^{2\pi} \int_0^{ap} E_I(r,\theta,0) E_I^*(r,\theta,0) r \, dr \, d\theta}{\sqrt{\int_0^{2\pi} \int_0^{ap} \left|E_I(r,\theta,0)\right|^2 r \, dr \, d\theta} \sqrt{\int_0^{2\pi} \int_0^{\infty} \left|E_I(r,\theta,0)\right|^2 r \, dr \, d\theta}} \right|^2 \tag{4.20}$$

Equation 4.20 is a normalised overlap integral of the field incident on the telescope aperture with the conjugate of the same.

4.6 Misalignment Analysis

4.6.1 Effect of Misalignment on Coupling

The effect of a horizontal offset of the lens, as shown in Figure 4.34, on coupling for a single-surface lens is given by

$$C(r_o, \theta_o) = \left| \frac{\int_0^{2\pi} \int_0^{ap} \left[E_H(r,\theta,0) \exp(-i\phi_H)\right] \left[E_I(r,\theta,0) \exp(-i\phi_I)\right]^* \, dr \, d\theta}{\sqrt{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,0)\right|^2 \, dr \, d\theta} \sqrt{\int_0^{2\pi} \int_0^{\infty} \left|E_I(r,\theta,0)\right|^2 \, dr \, d\theta}} \right|^2 , \tag{4.21}$$

$$\phi_H(r,\theta,0) = \arg\left[E_H(r + r_o, \theta + \theta_o)\right] , \tag{4.22}$$

$$\phi_I(r,\theta,0) = \arg\left[E_I(r + r_o, \theta + \theta_o)\right] , \tag{4.23}$$

where r_o and θ_o are radial and angular offsets respectively. $\phi_H(r,\theta,0)$ and $\phi_I(r,\theta,0)$ are the phase delays created by the horn and telescope lenses respectively. Figure 4.35 shows a plot of coupling with misalignment distance for horn 10; a 20% decrease in coupling occurs for an misalignment of -2mm along the x axis. A similar decrease in coupling is found for the other horns of the array.

Figure 4.36 shows the coupling of the a horn as a function of displacement from the nominal position, across the focal-plane, where only a lens to flatten the phase of the field from each of the individual horns (and not the field from the telescope) is used. This plot was made by performing the overlap integral of Equation 4.18 but with the telescope field shifted with respect to the horn. This is effectively a

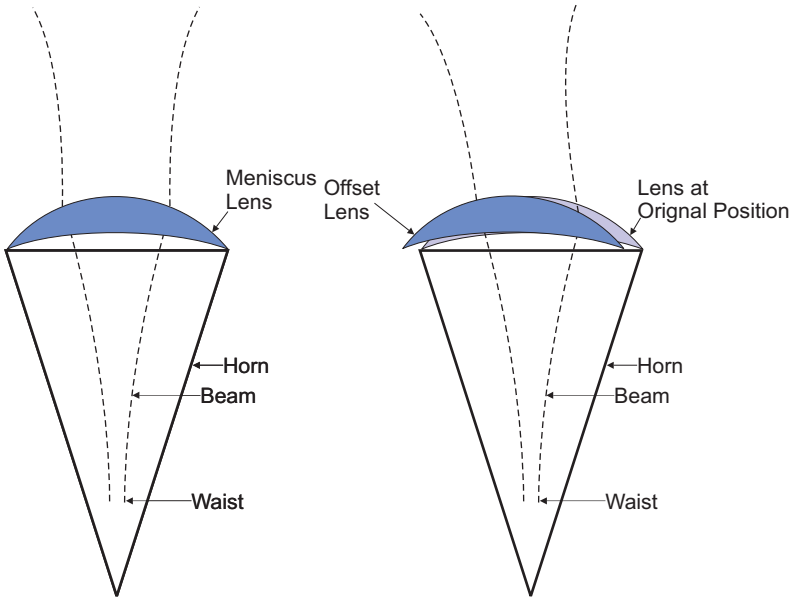


Figure 4.34: *Left* single-surface meniscus lens, *right* misaligned meniscus lens.

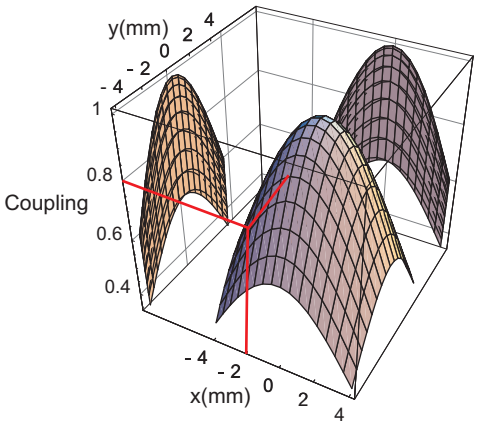


Figure 4.35: Normalized coupling as a function of horizontal misalignment for horn 10, with projections of coupling onto the x and y axes ($x = r\cos(\theta)$ and $y = r\sin(\theta)$). The red line shows the coupling with a -2mm offset along the x axis. The plot has been projected onto the x and y axes.

convolution of the horn with the field, and is therefore the image that would be measured by one of the detector horns if it scanned the field. The red points in Figure 4.36 indicate the nominal positions of the horns. From this contour plot it can be seen that if horns 6, 11 and 15 were displaced, for example, a small distance to the left, they would experience an increase in coupling, whereas horns 5, 9 and 14 would experience a decrease in coupling. The horns on the periphery of the array would experience only a small change in coupling, since they are in a low field region. This plot also shows how the current array will only measure a portion of the interference pattern.

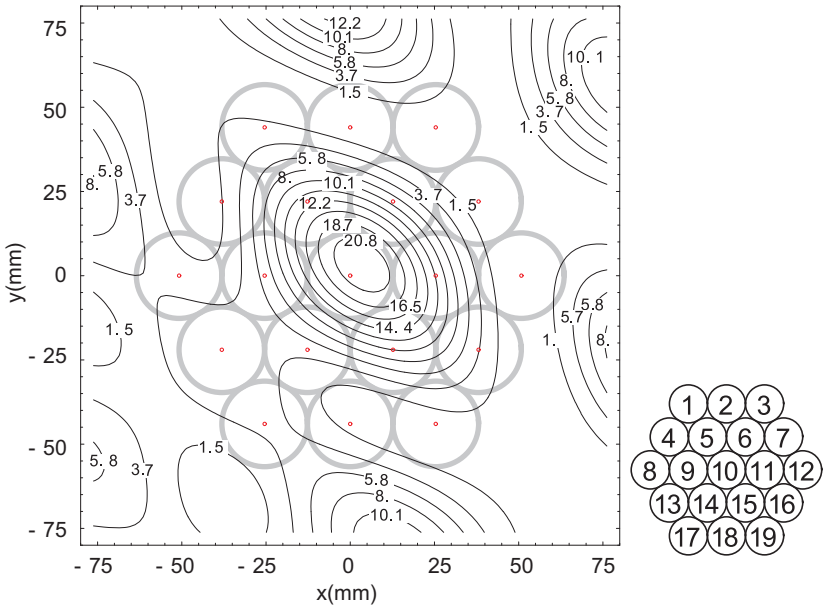


Figure 4.36: The power coupling$\times 10^3$ of a horn to the telescope field as a function of position across the focal-plane. The horn has a lens to correct for the spherical phase-front of its field at the aperture. This then is a plot of the telescope field convolved with the detector horn field. The red dots shows centre of each of the horns of the array indicated by the grey circles.

4.6.2 Optimisation of the Vertical Position of the Detector Lens

The coupling is now analysed for horns which have a horizontal and vertical misalignment and tilt of their meniscus lens. This misalignment normally decreases coupling, but can increase, depending on the telescope field.

Since the beam of the horn expands as it propagates away from the horn, an optimal distance exists at which to match the phase of the beam from the secondary and the beam of the detector horns. This is similar approach as that used to optimised the coupling between the detector horn and telescope field in ALMA (Chapter 5). The variation of coupling with vertical offset of the lens is now investigated to find the optimal distance of the lens from the horn aperture. Figure 4.37 *left* shows the

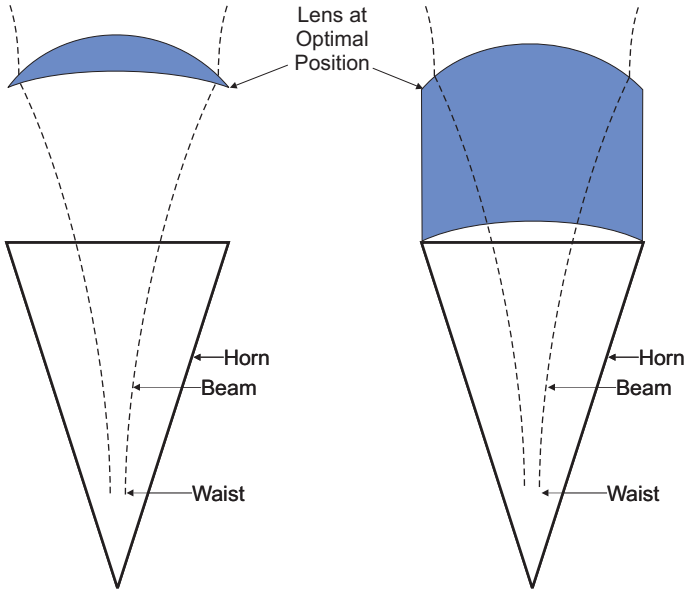


Figure 4.37: Meniscus lens at optimal position for coupling with telescope field: *left* using a thin lens, *right* using a thick lens for easier mounting.

lens at the optimal position; however this lens would be more difficult to mount than one situated at the horn aperture, and therefore may not be practical for this application. Figure 4.37 *right* shows the equivalent lens attached to the horn. The surface of this lens at the horn aperture is curved to match the wavefront of the horn beam at this location, so that it does not focus the beam. The lens of Figure 4.37 *right* is thicker; however, HDPE is a low loss material (< 0.1dB/cm)[6]. The coupling at this optimum distance, for each horn, is given in Table 4.5. These lenses are shown in Figure 4.33 (b).

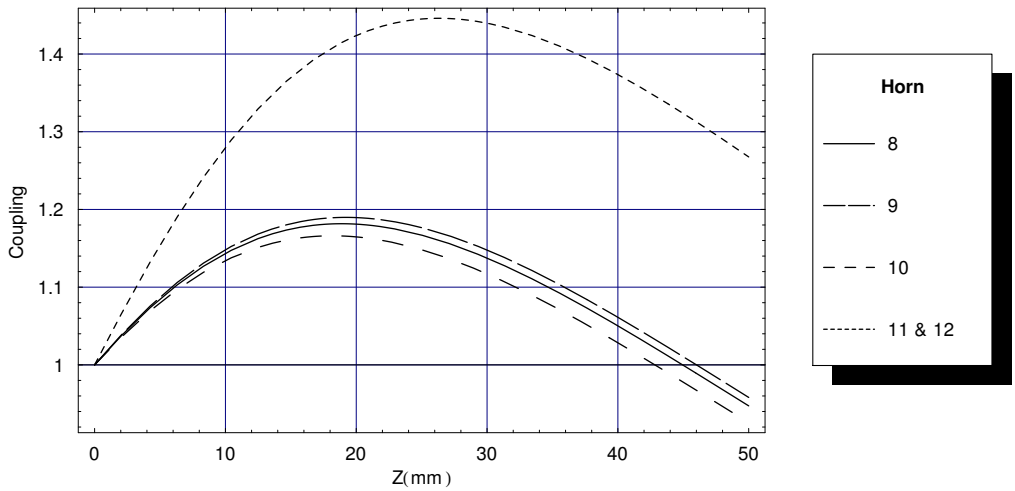


Figure 4.38: Normalized coupling as a function of lens position above the aperture.

The coupling of a vertically "misaligned" single-surface lens is given by

$$C(z) = \left| \frac{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,z)\right| \left|E_I(r,\theta,0)\right| \, dr \, d\theta}{\sqrt{\int_0^{2\pi} \int_0^{ap} \left|E_H(r,\theta,z)\right|^2 \, dr \, d\theta} \sqrt{\int_0^{2\pi} \int_0^{\infty} \left|E_I(r,\theta,0)\right|^2 \, dr \, d\theta}} \right|^2 , \tag{4.24}$$

where z is the distance the horn field has been propagated from the horn aperture and the dependence of the fields on z has been included. Figure 4.38 shows a plot of coupling with propagation distance z. Figure 4.38 shows a plot of coupling with vertical offset of the lens for horns 8 to 12. The maximum coupling occurs at an offset distance of approximately 20mm for horns 8, 9, 11 and 12 and 26mm for horn 10.

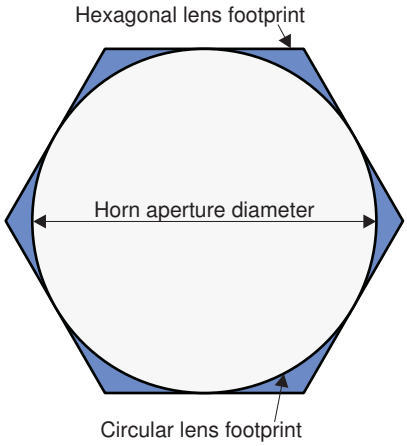


Figure 4.39: The footprint of a circular and hexagonal lens.

Hexagonal lenses allow the telescope field to couple some of the field which would otherwise be lost in the region between the horns. Figure 4.39 shows the footprint of a circular and a hexagonal lens. The area of a hexagon is given by $2\sqrt{3}r^2$, where r is the minimal radius. As illustrated in Figure 4.39, the ratio of the area of a hexagon divided by the area of a circle, πr^2, is $2\sqrt{3}/\pi \approx 1.1$, so around 10% of the power in the horn array area bounded by the hexagonal region in Figure 4.26 is lost in the space between the horns. The hexagonal lenses need to be vertically offset from the horn aperture to improve the coupling of the radiation to the horns, otherwise the radiation incident on the blue region in Figure 4.39 would not couple with the horn and would be lost in the region between the horns. As was the case for the circular meniscus lenses, for the hexagonal lenses the maximum coupling to the field was achieved at a distance of approximately 20mm from the horn. The maximum coupling for each horn, using circular and hexagonal lenses which are offset 20mm from the horn, are listed in Table 4.5. These lenses are shown in Figure 4.33 (c).

4.6.3 Effect of Tilt on Horn Coupling

The coupling of a tilted detector horn to the telescope field without a lens is, in Cartesian co-ordinates, given by

$$C(\theta_x, \theta_y) = \left| \frac{\int_{-cir}^{cir}\int_{-cir}^{cir} E_H(\overline{x},\overline{y}) E_I^*(\overline{x},\overline{y})\, dx, dy}{\sqrt{\int_{-cir}^{cir}\int_{-cir}^{cir} \left|E_H(\overline{x},\overline{y})\right|^2 dx, dy}\sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|E_I(\overline{x},\overline{y})\right|^2 dx, dy}} \right|^2 \tag{4.25}$$

where $\overline{x} = x\cos\theta_x$ and $\overline{y} = y\cos\theta_y$, cir is the radial limit of the circular horn aperture, and θ_x and θ_y are the tilts in the x and y directions. Cartesian co-ordinates are used here for ease of calculating the effect of tilts.

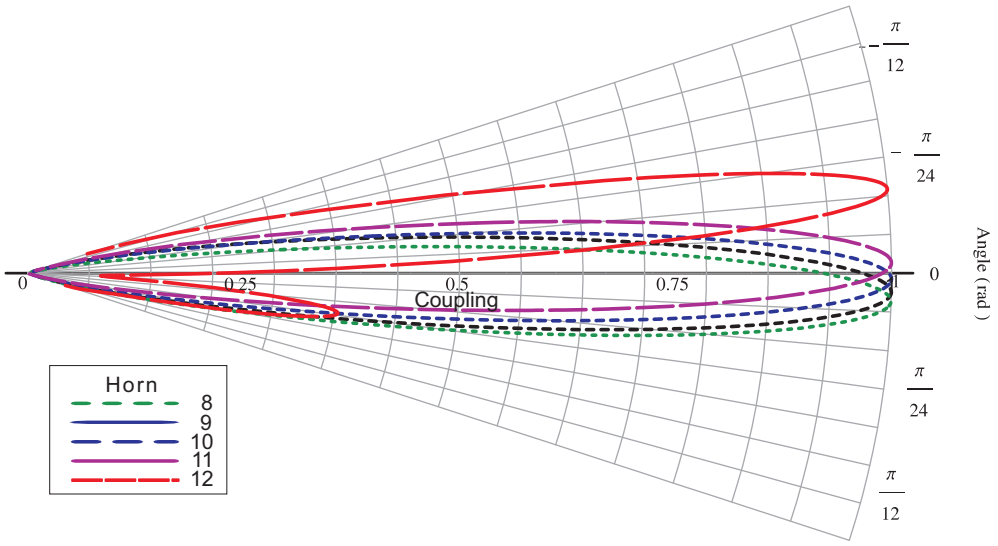


Figure 4.40: Coupling as a function of angle θ_x ($\theta_y = 0$), for horns $8 \rightarrow 12$.

The analysis of the tilted horns is important for two reasons. Firstly, it shows the effect of misalignment and secondly tilting the horns may actually increase the coupling, as the phase-front of the telescope field is curved; the horns could therefore be intentionally tilted, to improve coupling. Tilting the horns to increase coupling may be preferential to using lenses, as lenses increase the level of cross-polarisation. Figure 4.40 shows the normalized coupling as a function of θ_x and $\theta_y = 0$ for horns 8 to 12. It shows both the coupling at the zero tilt angle, and the value of θ_x for which coupling is maximised. Horns $8 \rightarrow 11$ appear to be at approximately the optimal angle for maximum coupling with no tilt. An increase in coupling of 75% can be achieved by rotating horn 12 around the y axis by 5.5°. Coupling as a function of angle for horn 12 has a second local maximum which has important implications for the algorithm to optimise its tilt, as a linear optimisation algorithm may get stuck at a local minimum.

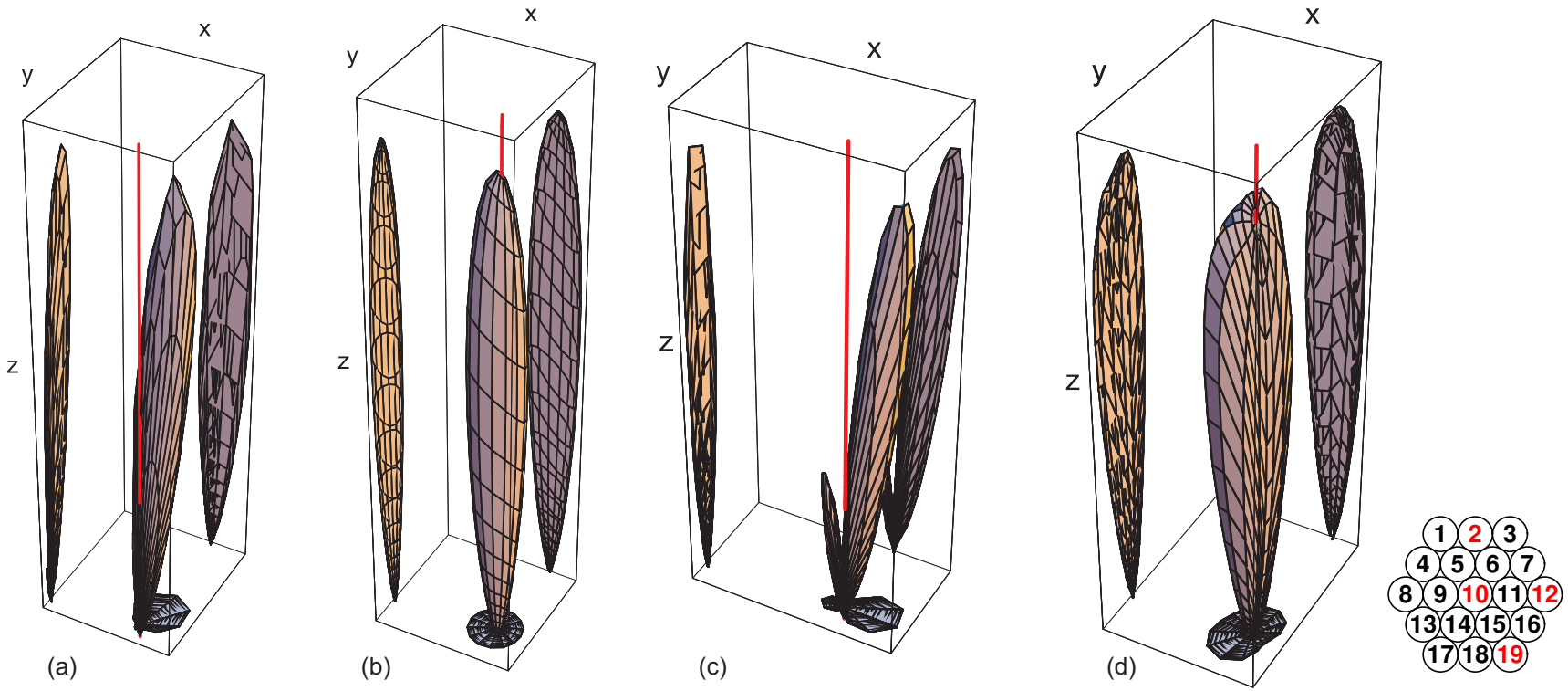


Figure 4.41: Coupling as a function of angle for horn number (a) 2, (b) 10, (c) 12 and (d) 19, with projection of the plot onto the x and y planes. (e) shows the position of these horns. The red lines show the direction of the zenith. The actual coupling for the untilted horns is given by the point at which the red line meets the surface. The horn can be tilted in order to maximise coupling.

So far, the effect of tilt on coupling has been examined with freedom of rotation in one dimension only; however, both θ_x and θ_y need to be varied simultaneously in order to find the maximum coupling. Figure 4.41 shows the coupling of a number of horns as a function of θ_x and θ_y. The vertical line emanating from the centre of the plot is the axis along which, $\theta_x = 0, \theta_y = 0$, i.e. where the horns are not tilted. Figure 4.42 *left* and *right* shows an array of untilted and optimally tilted (for best coupling) horns, respectively. The dashed lines are along the z axis of the untilted horns, and the solid lines are along the z axis of the tilted horns. The optimal coupling was found by optimizing θ_x and θ_y using the Newton method[51]. By using $\theta_x = 0$ and $\theta_y = 0$ as starting values for the variables of the optimization algorithm, the global maximum coupling was found for each. For most horns only a small increase in coupling is achieved by tilting the horn as they are close to the optimal angle with no tilt; however, the coupling is increased by a factor of 4.75 and 3.12 for horns 2 and 12 respectively when they are tilted. A larger array may be used in the future to detect a wider area of the field. Figure 4.30 (b) shows that the phase of the field is more curved away from the origin, therefore tilted horns placed further from the centre in a larger array would result in a large increase in coupling. Tilting the horns should therefore be considered for the future generations of MBI, which have more horns, if the focal-plane is curved.

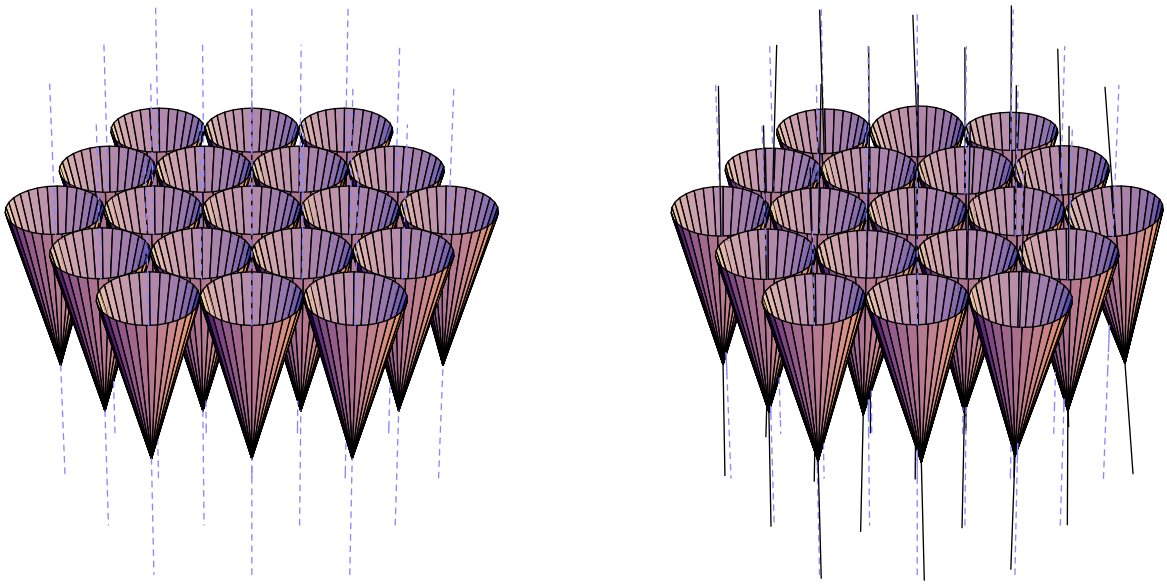


Figure 4.42: *Left* untilted horns, *right* optimally tilted horns.

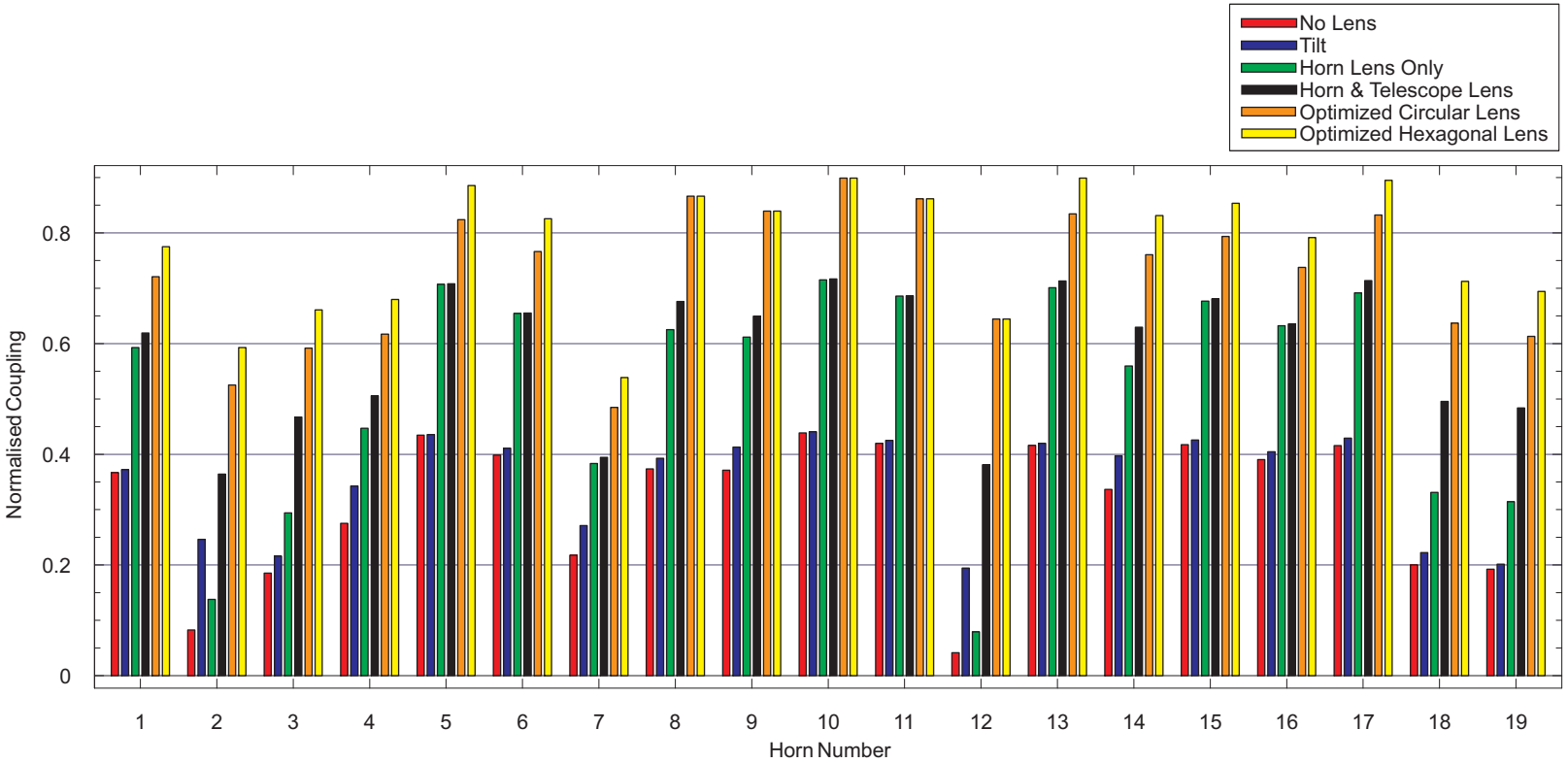


Figure 4.43: A bar chart summarising the results given in Table 4.5.

4.7 Conclusion

In this chapter, the use of lenses in MBI - a novel type of bolometric interferometer - was investigated. Lenses for the feed-horns of the two baselines were designed in order to reduce the diffraction of the beams from the feed-horns, thereby reducing truncation at the secondary mirror. Spherical lenses were chosen over hyperbolic and plano-convex lenses for their simplicity. Lenses were also designed to improve coupling at the array of detector horns. The coupling is maximised when the phase of the field from the detector horns and telescope are matched. The improvement in coupling for a number of configurations was investigated: including lenses to match the telescope field and horn field with two surfaces, including horn phase flattening lenses only, including lenses which match the phase of the field using one surface, including lenses which are vertically offset from the aperture to improve coupling, and finally horns without lenses that are tilted to improve coupling. Table 4.5 and Figure 4.43 summarise the power coupling calculations for these configurations. The best lenses from the point of view of maximising coupling would match the amplitude of the field produced by the horn to the amplitude of the field at the horn aperture as well as their phases. However, this would require amplitude shaping, which is non-trivial (it is discussed in Chapters 2 and 3). Figure 4.33 (a) shows lenses in which the phase of the horn and telescope fields are matched using a single-surface. This is a better design than that shown in Figure 4.11, in which the field is flattened using 2 lenses, as each surface introduces contamination of the orthogonal polarisations. If the lenses are offset, as shown in Figure 4.33 (b), the coupling may be increased further, and if hexagonal offset lenses are used, as shown in Figure 4.33 (c), the coupling is increased further again. The exact scheme for using bolometric interferometers is still under discussion but this work shows what ranges of improvement you might expect under different circumstances.

The designs chosen for fabrication were spherical lenses, for both the detector and horn lenses; they are easier to manufacture than either the hyperbolic or the more complicated profile that matched the phase of the detector horn and sky field. The detector horn array with the lenses and with the array filter is shown in Figure 4.16 *left*. The horn array is made from a single piece of metal, shown in Figure 4.16 *left* and Figure 4.17 *left*. The bandwidth filters, shown in Figure 4.17 *right*, are placed on top of the horn array.

A significant improvement in coupling was found with the lenses; however, they have some disadvantages: cross-coupling of the orthogonal polarisations occurs at the lens interfaces, especially when grooves are used to reduce reflections. In addition, HDPE is birefringent, due to the alignment of its long chain polymers. In future generations of MBI, the array size may be increased. Using tilted horns, as shown in Figure 4.42, for a larger array may be more beneficial than for this 19 el-

ement array, as the phase is approximately spherical, and at the edges significantly deviates from a flat phase.

Chapter 5

Phase Matching For The ALMA Telescope

5.1 Introduction

In this chapter, I describe the work that I carried out to model the ALMA telescope and increase the coupling of the signal from the sky to the telescope, using techniques introduced in Chapter 3.

The Atacama Large Millimeter Array (ALMA) interferometer telescope is the work of an international collaboration, involving the European Southern Observatory (ESO), the Japanese National Institutes of Natural Sciences (NINS) and the National Radio Astronomical Observatory (NRAO) in the USA[52]. The original ALMA design consisted of 64 12m diameter antennas located on a high and dry site, illustrated in Figure 5.1, allowing millimetre and submillimetre observations with the same angular resolution and sensitivity that is achieved at visible and infrared wavelengths[1]. The Chajnantor plateau in Chile was chosen as the location of the telescope array because of its excellent atmospheric transparency, its flat terrain, and because it is large enough to accommodate long baselines and movement of the antennas. The maximum resolution will be ≈ 10 milliarcseconds at the highest frequency and longest baseline, which is $\approx$ 10km. The 64 antennas have now been scaled back to 50 12m diameter Cassegrain telescopes. The remaining 14 antennas may be added later. As sensitivity is directly proportional to the surface area of the antennas, the smaller number will result in a reduction in sensitivity of $\approx 22\%$. However, since integration time goes as $1/N^2$, where N is the number of antennas, the integration time will need to be increased by 1.65 to achieve the same sensitivity as the original 64 antenna array[53].

The telescopes will operate in the frequency range 30GHz to 1THz, in 10 frequency bands, shown in Table 5.1, using 10 dual-polarisation receivers. The signal from each telescope's secondary mirror is fed to 10 self-contained cartridges, each

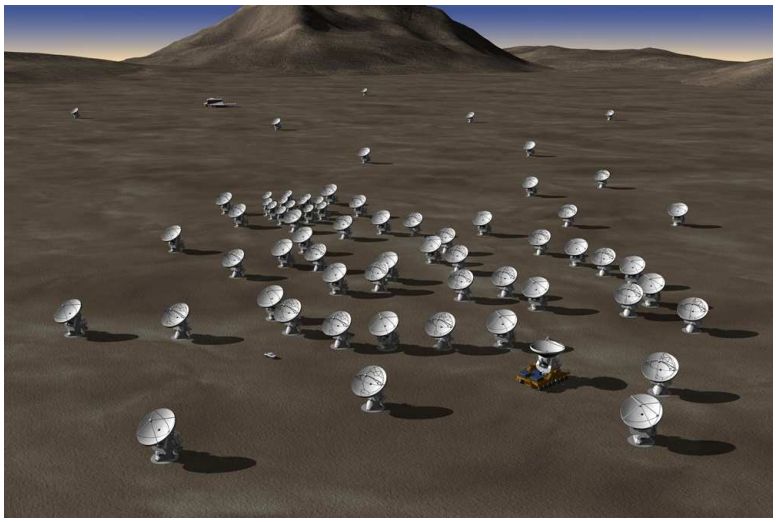


Figure 5.1: Artist's Impression of ALMA Array Configuration.

holding two orthogonally-polarised channels. These cartridges are arranged around the centre of a 970mm diameter dewar, located on the telescope axis. Each band produces a separate beam in the focal-plane, allowing the sky to be viewed simultaneously in the ten frequency bands. Selection of the observing band simply requires re-pointing of the antenna. The use of these cartridges limits how tightly packed the receiver channels can be; however this is not a major problem as the telescope has a wide field of view[54].

A top view of the dewar is shown in Figure 5.2. Most of the optical elements for each band are integrated into a single cartridge. However, for the lowest frequency bands, the optics are located on the top of the dewar, as they are too large to fit into a cartridge. Aberrations, polarisation distortion and vignetting by the hole in the primary mirror would greatly affect some of the cartridges that are far off the telescope axis, so mirrors are used to bring the beam closer to the centre. Atmospheric phase correction is achieved by means of a water vapour radiometer. A mirror is used to bring the beam of the radiometer to a position at the centre of the focal-plane, so that it is close to all of the observing beams[4].

Because of the difference in sizes of the components for the different bands, slightly different configurations are used for each. An important requirement is to minimize cross-polarisation effects and optical losses. This was achieved by minimising the number of optical elements, using reflective rather than transmissive optics wherever possible, maintaining large beam clearances, minimizing angles of incidence on offset reflectors and balancing cross-polarisation between consecutive mirrors. As many of the optical elements as possible are located in the dewar to minimise thermal noise. For bands 1 and 2, there is insufficient space above the dewar to fit the mirrors; consequently, lenses are used instead. For bands 3 and 4, some of the optics are too large to fit inside the dewar, and must be kept at

Band	Lowest (GHz)	Centre (GHz)	Highest (GHz)
1	31.5	38	45
2	67	81	90
3	84	100	116
4	125	144	163
5	163	187	211
6	211	243	275
7	275	323	370
8	385	442	500
9	602	661	720
10	787	868	950

Table 5.1: The Ranges of the 10 Frequency Bands of the ALMA Telescopes.

Label	Distance(mm)
Δz	9.09
D1	44.48
D2	95.91
D3	149.44
D4	191.09
D5	84.59
z_{w1}	44.92

Table 5.2: The parameters for the mixer horn shown in Figures 5.3 and 5.6 [9].

ambient temperature. The optics of the remaining bands can be located inside their cartridges, allowing the optics to be aligned and tested as a single unit.

5.2 ALMA Band 9 Optics

In band 9, the field from the secondary mirror is re-imaged onto the two receiver feed-horn polarisation channels with the two elliptical mirrors M3 and M4, as shown in Figure 5.3 *right*, with the parameters given in Table 5.2[55, 4, 54]. The optics allow frequency independent coupling between the telescope and mixer horn and were designed using geometrical optics. This double-ellipsoidal mirror coupling system, shown in Figure 5.4, re-images the mixer horn aperture to the secondary mirror. The mirrors were designed to cancel aberrations (without the presence of a grid to linearly polarize the beam).

The aim of the work in the section is to see if the coupling between the sky field and horn field can be improved. Firstly, the coupling of the detector corrugated horn field to the sky field will be calculated. A phase plate will then be designed to impose a phase on the sky field to improve the coupling to the detector horn. At first this is done using a one-dimensional analysis. With one surface, maximum coupling is achieved by matching the detector horn field to the sky field. A two-dimensional analysis is then carried out to more accurately model the system. Two surfaces are

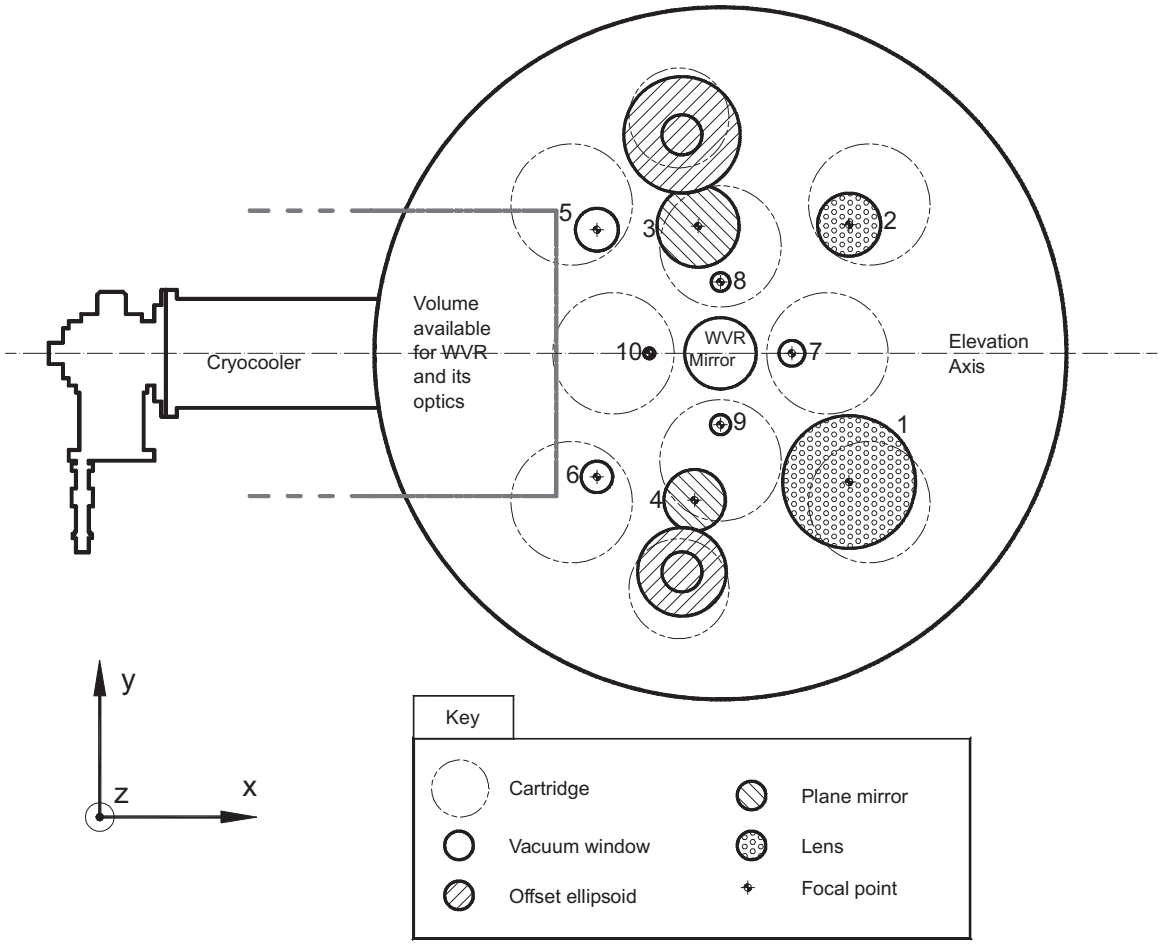


Figure 5.2: Layout of the dewar viewed from above[4].

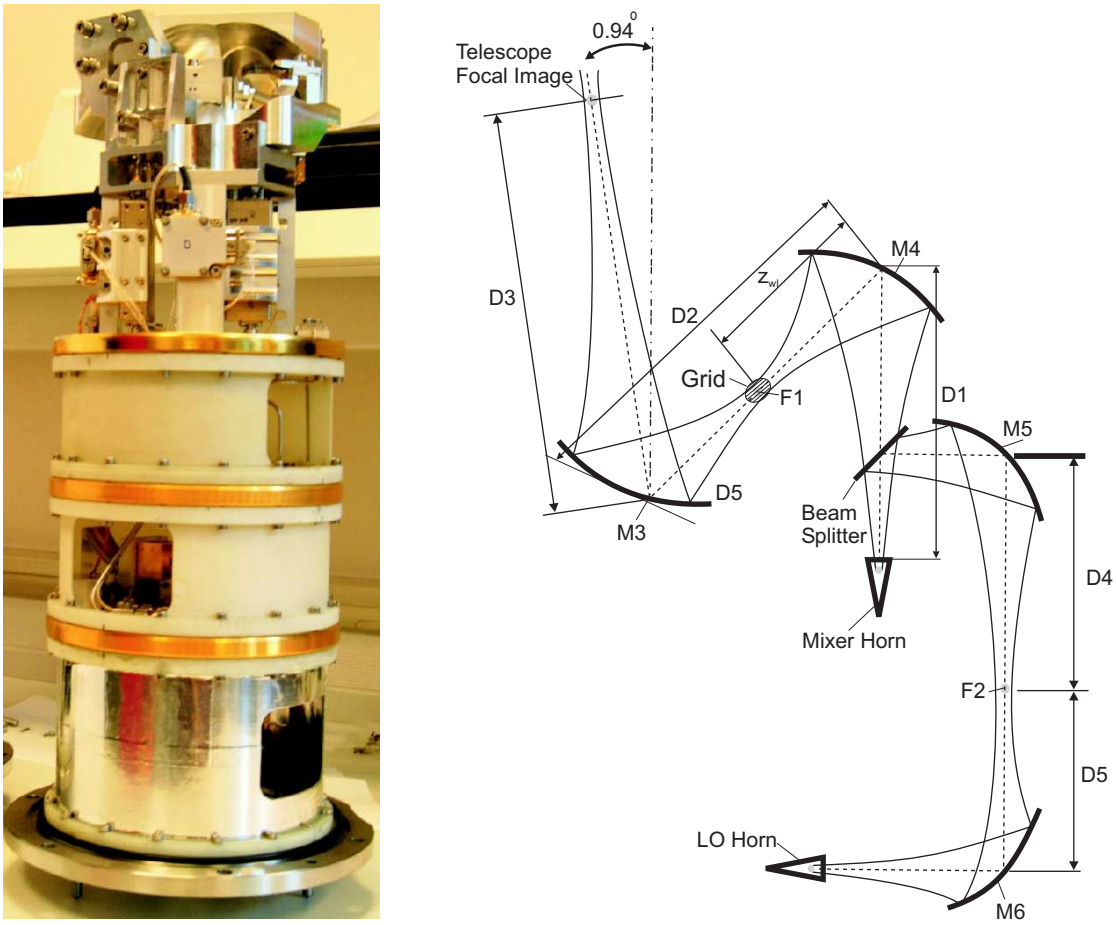


Figure 5.3: *Left* the band 9 cartridge. *Right* ALMA band 9 optical configuration for the vertical polarisation transmitted through the grid. A similar optical system, to that shown after the grid is used to couple the orthogonal, horizontal, polarisation, reflected by the grid.[9] The distances labelled are given in Table 5.2

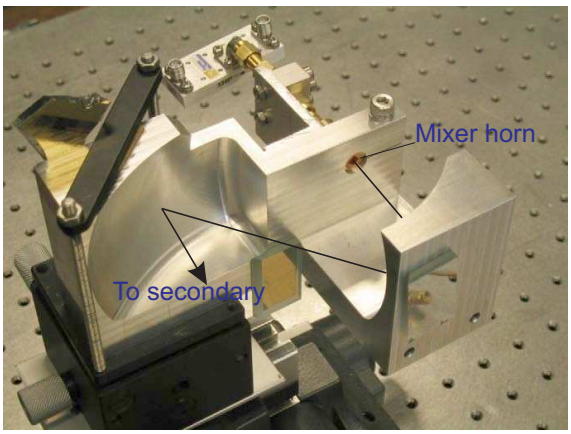


Figure 5.4: A prototype mirrors M3 and M4[55].

then used in order to match both the phase and amplitude of the detector horn and sky fields. This requires beam shaping, as discussed in Chapters 2 and 3.

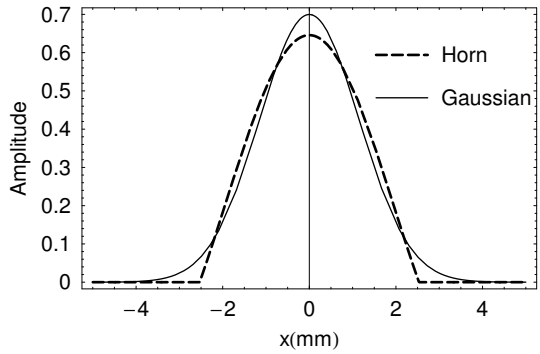


Figure 5.5: The field from the ALMA mixer corrugated horn and its Gaussian approximation.

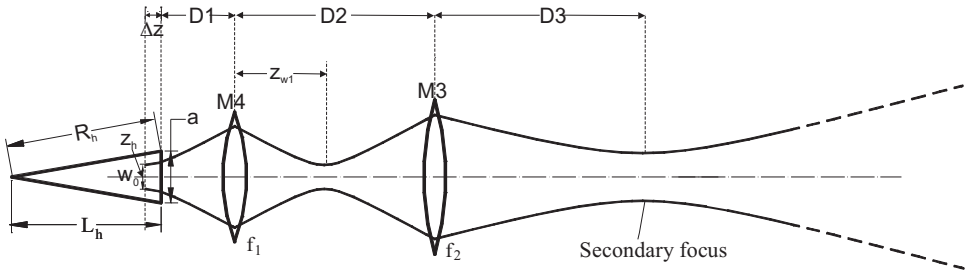


Figure 5.6: The equivalent on-axis layout for band 9 using lenses.

A single-moded hollow circularly-symmetric horn supports the circularly symmetric HE_{11} mode, given by

$$E(r) = \begin{cases} J_o\left(\frac{2.405r}{a}\right)\exp\left(-\frac{i\pi r^2}{\lambda R_h}\right) & r \leq a \\ 0 & r > a, \end{cases} \tag{5.1}$$

where r is the radial distance in the horn guide, J_0 is the Bessel function of zeroth order and a is the aperture radius[6]. The aperture field has a radius of curvature equal to the slant length of the horn R_h. The waist radius is given by

$$w_h = \frac{w}{\sqrt{1 + (\pi w^2/\lambda R_h)^2}}, \tag{5.2}$$

as shown in Chapter 1, and the distance from the horn aperture to the waist (located behind the aperture) is given by

$$\Delta z = \frac{R_h}{1 + (\lambda R_h/\pi w^2)^2}. \tag{5.3}$$

w_h and Δz, therefore, are dependent on the wavelength, λ. The best choice of w for the fundamental Gaussian mode used to model the horn-field is that with

the greatest coupling to the horn-field. The maximum coupling with the horn-field occurs when $w = 0.6435a$, for which the coupling $C = 0.99$ and C^2, the power coupling, is 0.98[6]. The ALMA mixer horn has an aperture diameter of 5.066mm, and a slant length of 15.722mm. Figure 5.5 shows the horn-field amplitude with the fundamental Gaussian that best approximates it. Table 5.3 shows the radius of the Gaussian beam at mirrors M3 and M4, calculated using Equations 5.2 and 5.3 and the ABCD matrix method.

Frequency (GHz)	602	661	720
Wavelength (mm)	0.497	0.453	0.416
Horn waist, w_h	1.115	1.059	1.006
Horn waist offset, Δz	-8.364	-9.089	-9.735
Radius at horn aperture	1.630	1.630	1.630
Radius at M4	7.595	7.381	7.214
Radius at M3	8.457	8.363	8.291

Table 5.3: The fundamental Gaussian parameters of the beam

The corrugated circular feed-horn produces a field that is 98% Gaussian (Figure 5.5), and therefore the propagation of the beam can be modelled with a high degree of accuracy using the fundamental Gaussian mode approximation. The ABCD matrix technique, discussed in Chapter 1, is used to model the beam transformation by the mirrors, by modelling them as on-axis lenses, as shown in Figure 5.6.

5.2.1 Phase Matching using One Surface in One-Dimension

A phase plate is designed to improve the coupling between the field from the sky and the mixer horn. Initially, a dielectric phase plate will be designed for evaluation purposes. This plate works in transmission by causing a phase delay proportional to its optical depth. Since a dielectric phase plate would introduce losses and cause cross-coupling between the polarisations, ultimately (if the increase in power coupling is considered worthwhile) the profile of the phase plate will be machined into the mirror M3.

The coupling of the horn field, E_H^*, to the sky field, E_S^* in 1D is given by

$$C = \frac{\int_{-\infty}^{\infty} E_H^*(x,z)E_S(x,z)dx}{\sqrt{\int_{-\infty}^{\infty} \left|E_H(x,z)\right|^2 dx}.\sqrt{\int_{-\infty}^{\infty} . \left|E_S(x,z)\right|^2 dx}}, \tag{5.4}$$

The power coupling is given by $|C|^2$, and represents the fraction of incident power flowing from one beam to the other[6]. This calculation can be performed at *any* plane between the horn and the sky. Although the amplitude profile of E_H and E_S vary with z, the phase varies such that Equation 5.4 is independent of z.

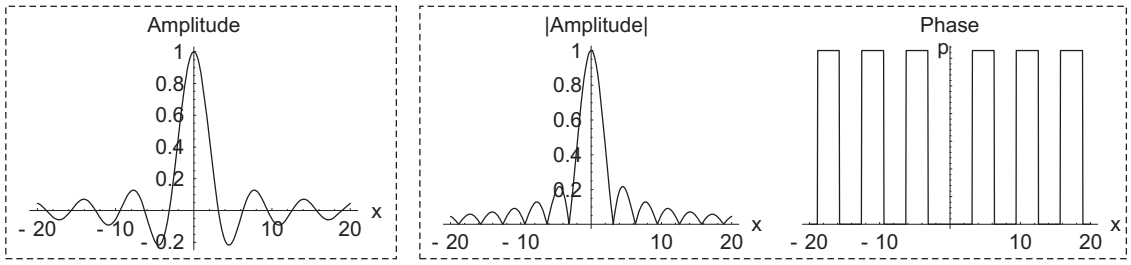


Figure 5.7: *Left* the top-hat amplitude distribution at the secondary mirror has a sinc function distribution when propagated to the focal-plane (in one-dimension). *Right* the same field is shown in terms of absolute amplitude and phase. (Negative amplitude is equivalent to a positive amplitude of equal magnitude combined with a π rad phase shift.)

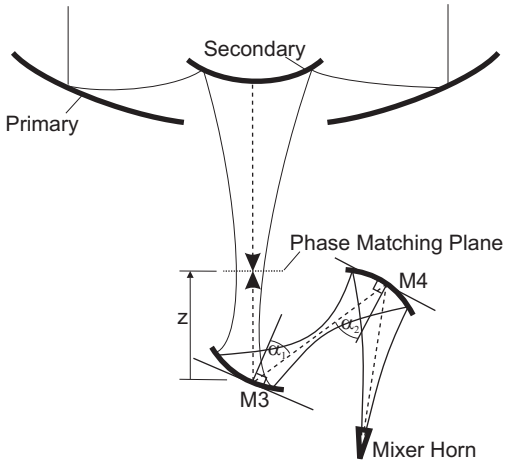


Figure 5.8: The phase plate maximises coupling between the fields from the horn and the secondary mirror by matching their phases.

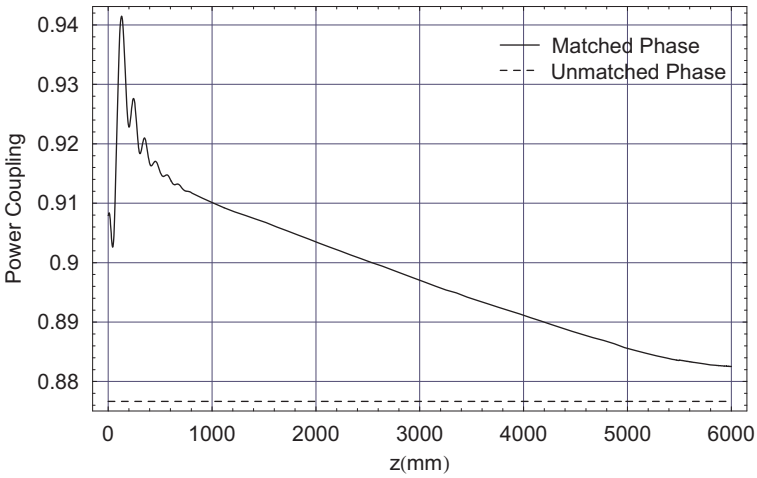


Figure 5.9: The solid line shows the power coupling between the sky and horn field when their phases are matched at a plane located at z, as a function of that plane's distance from M3 (since the two field evolve as they propagate, the shape of the phase plate, required to match the fields, changes with z). The plot is done over the region between M3, located at $z = 0$mm to $z = 6000$mm. A Gaussian approximation of the feed-horn field is used. The dashed line shows the coupling of the fields when the phase is not matched.

The field from the secondary mirror is approximated by a top-hat distribution with an aperture radius of 375mm and a phase radius of curvature of 6m. The ALMA optics were designed to match as closely as possible the phase-front of the field from the secondary mirror to the phase-front of the mixer horn-field. However, as shown in Figure 5.7, the top-hat field from the secondary mirror produces a field with an oscillating amplitude, which is periodically π rad out of phase with the horn-field (Figure 5.5), at the focal-plane. In Chapter 1 it is shown that if *only one surface* is available to impose a phase to optimise the coupling of the two fields, E_H and E_S, the *maximum* coupling that can be achieved is when the phases of the two fields are matched. Although the horn and sky fields are focused to the same plane, and there is therefore no spherical phase component in the phase difference between the horn and sky fields, at the focal-plane the π phase shifts in the field from the secondary mirror reduces the coupling. As the horn and sky field propagate away from the focal-plane the π phase jumps disappear from the sky field; however, the phase of the two fields are still mismatched.

The optimal coupling, i.e. the coupling at the plane where the phases of the fields are matched by a phase plate, was calculated as

$$\bar{C}(z) = \frac{\int_{-\infty}^{\infty} \left|E_H(x,z)\right| \left|E_S(x,z)\right| dx}{\sqrt{\int_{-\infty}^{\infty} \left|E_H(x,z)\right|^2 dx} \sqrt{\int_{-\infty}^{\infty} \left|E_S(x,z)\right|^2 dx}}, \tag{5.5}$$

where z is the distance from mirror M3, as shown in Figure 5.8. Taking the *modulus* of the horn and sky fields in Equation 5.5 models the effect of the phase plate, which flattens the phase. The horn and secondary mirror field are propagated using GBMs. The field at the secondary mirror is modelled using the first 60 even Gaussian-Hermite modes with a radius of curvature of 6m. The fundamental Gaussian predicted by the ABCD analysis above is used to model the horn-field. 99.3% of the power in the top-hat field is reconstructed by the GBMs, and 98% of the power of the field of the horn is reconstructed by the fundamental Gaussian that approximates it (shown in Figure 5.5). Later, the analysis is performed for a more accurate multi-moded reconstruction of the horn.

At M3, without a phase plate the power coupling is 0.876, and with a phase plate power coupling is increased to 0.908. The phase plate will be located between M3 and the secondary mirror, so the effect of phase matching is considered at planes in this region. Figure 5.9 shows the power coupling achieved with the phase plate location in the range z=0mm, where M3 is located, to z=6149.4mm, at the secondary mirror. The profile of the phase plate that would match the phases of the two fields varies with z, because the beams of both fields evolve - and therefore the phase profile that needs to be imposed to match the two fields varies also.

To more accurately determine the effect of a phase plate, the truncated Bessel

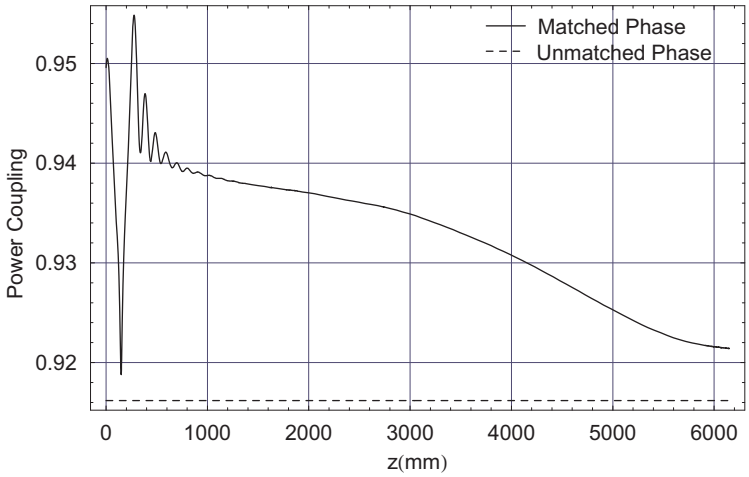


Figure 5.10: The solid line shows the power coupling between the sky and horn field when their phases are matched at a plane located at z, as a function of that plane's distance from M3 (as the two field evolve as they propagate, the shape of the phase plate, required to match the fields, changes with z). The plot is done over the region between M3, located at $z = 0$mm, and the secondary mirror, located at $z = 6149.22$mm. A truncated Bessel function is used to model the field from the horn. The dashed line shows the coupling of the fields when the phase is not matched.

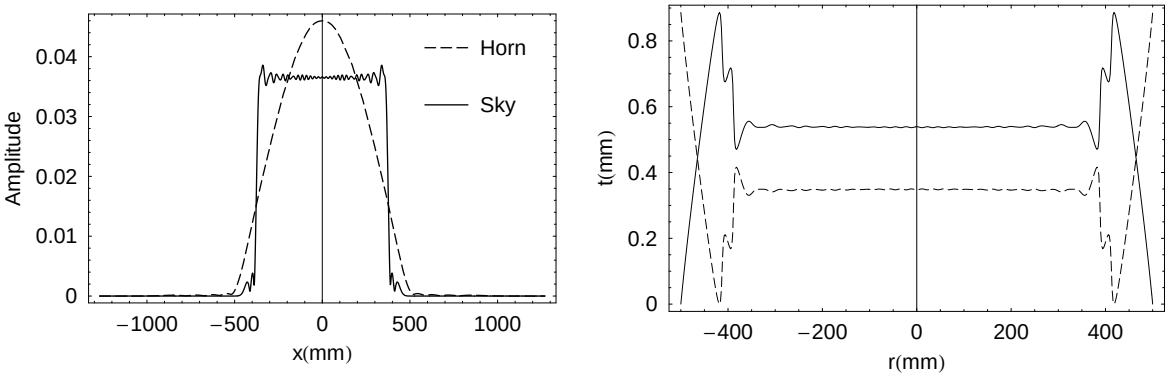


Figure 5.11: *Left* the amplitude of the field from the horn and the sky at the secondary before any phase matching. *Right* the plate profile that matches the phase of the two fields at the secondary mirror, the dashed line is given by Equation 5.8: $\lambda \arg[E_S(x, 6149.44)E_H^*(x, 6149.44)]/(2\pi(n-1))$ and the solid is given by $\lambda \arg[E_S^*(x, 6149.44)E_H(x, 6149.44)]/(2\pi(n-1))$ (refractive index n is= 1.52).

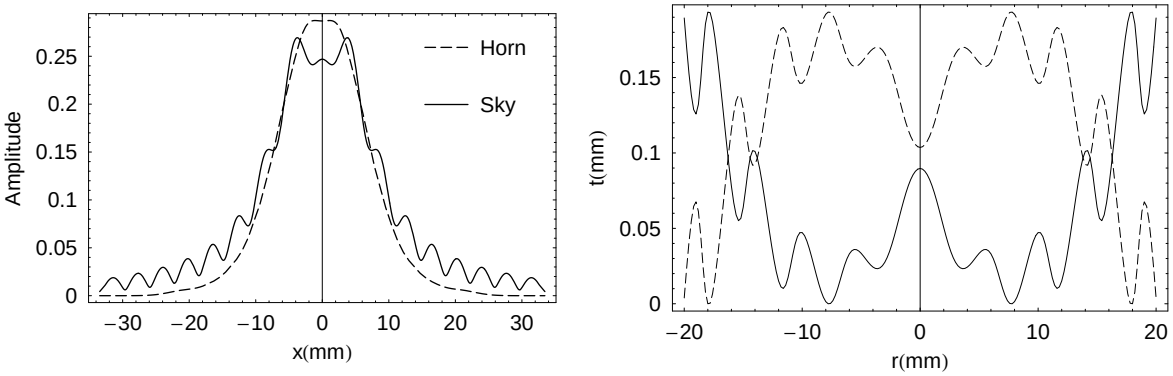


Figure 5.12: *Left* the amplitude of the field from the horn and the sky at M3. *Right* the phase plate profile that matches the phase of the two fields at M3, the dashed line is given by Equation 5.8: $\lambda \arg[E_S(x, 0)E_H^*(x, 0)]/(2\pi(n-1))$ and the solid is given by $\lambda \arg[E_S^*(x, 0)E_H(x, 0)]/(2\pi(n-1))$ (refractive index n is= 1.52).

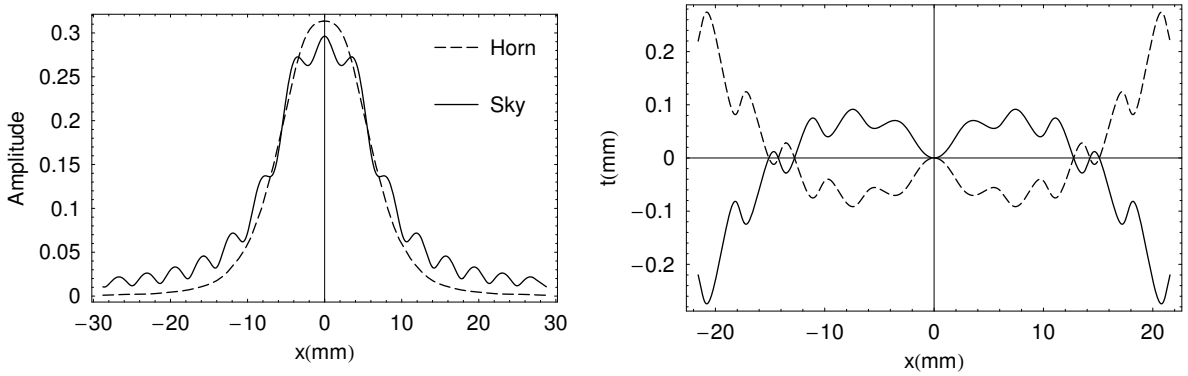


Figure 5.13: *Left* The amplitude of the field over a plane at the distance from M3, $z = 275.22$mm, that maximum coupling is achieved. *Right* the phase plate profile that matches the phase of the two fields. The dashed line is given by Equation 5.8: $\lambda \arg[E_S(x, 275.22)E_H^*(x, 275.22)]/(2\pi(n-1))$ and the solid is given by $\lambda \arg[E_S^*(x, 275.22)E_H(x, 275.22)]/(2\pi(n-1))$ (refractive index n is= 1.52).

function of Equation 5.1 is used in place of the fundamental Gaussian approximation to the horn-field. The horn-field is propagated using the first 20 even Gaussian-Hermite modes, which reconstruct effectively all the power of the horn-field. Figure 5.10 shows the power coupling using this, more accurate, model of the horn-field. The overall shape of the plot is similar, but the coupling is better by a percent or so than for the Gaussian approximation of the horn-field. Again, the power coupling varies as a function of z, as the amplitude profile of the horn and sky fields become more or less similar, due to diffraction. The power coupling varies rapidly and oscillates in the near-field region close to M3. This is due to the amplitude profile of the horn-field and, especially, the field from the secondary mirror varying rapidly.

Figures 5.11 and 5.12 show the amplitude of the fields both at the secondary mirror and M3, along with the phase required to match the phases of the two fields at these planes. It can be observed that the amplitude profile of the two fields are much more alike at M3 than at the secondary mirror; hence the greater coupling achieved when the phases are matched at M3. With phase matching at M3, the power coupling is increased from 0.916 to 0.95. Figure 5.13 *left* shows the amplitude of the fields at a distance a 275.22mm from M3, where a maximum power coupling of 0.956 is achieved. The power coupling at the maximum, minimum and mid frequencies that the cartridge will operate at is shown in Table 5.4. An increase in coupling is achieved across the band; although, the increase is less great at the edges of the band. The phase that matches the horn and sky fields is given by

$$\phi(x) = \arg(E_S^*(x, z)E_H(x, z)), \tag{5.6}$$

or, alternatively, the argument of the complex conjugate

$$\phi(x) = \arg(E_S(x, z)E_H^*(x, z)). \tag{5.7}$$

Frequency	602	661	720
Power coupling, no matching	0.914	0.916	0.913
Power coupling, matched phase	0.939	0.956	0.950

Table 5.4: The power coupling between the field from the sky from the horn where the phases are matched at the optimal position, 275.22mm from M3, and where no phase matching is done.

For a transmission phase plate made from a dielectric, the thickness of the dielectric as a function of off axis distance is given by

$$t(x) = \frac{\lambda\phi(x)}{2\pi(n-1)}, \tag{5.8}$$

and for a reflective plate the profile is given by

$$t(x) = \frac{\lambda\phi(x)}{2\cos(\alpha)2\pi}, \tag{5.9}$$

where λ is the wavelength and $\phi(x)$ is the phase that is to be imposed on the field given by Equation 5.6 or 5.7, n is the refractive index, and α is the angle of the reflected beam as shown in Figure 5.8[22]. Figure 5.13 *right* shows the HDPE phase plates (with a refractive index of 1.52) that match the phases of the two fields at $z = 275.22$mm. The field amplitudes are also shown in Figure 5.13. By restricting the radius, r, over which the phase plate matches the phases to 15mm, its maximum thickness can be reduced to 0.1mm, while only reducing the coupling of the fields to 0.955, as most of the power of the beam is within this region.

5.2.2 Phase Matching using One Surface in Two-Dimensions

Thus far, the analysis has been one-dimensional; a two-dimensional analysis will now be carried out, by assuming circular symmetry of the field propagating in the telescope. Unlike two-dimensional fields that are separable in rectangular coordinates, transforms of circular symmetric fields can generally *not* be found from the one-dimensional transform[13]. An exception is the fundamental, Gaussian, mode, which is both separable in rectangular coordinates *and* circularly symmetric. The horn field is essentially Gaussian, and even the top-hat field from the secondary mirror can be approximated reasonably well with a fundamental Gaussian mode, and so, we might expect the field predicted by the one-dimensional analysis to be close to a one-dimensional cut through the two-dimensional field.

As described in Chapter 1, Gaussian-Laguerre modes, $\psi_p(r)$, as a function of off-axis distance r, have the appropriate circular symmetry, and are now used in place of Gaussian-Hermite modes to propagate the field. Associated Laguerre modes, $\psi_{p,m}(r)$, may be used to reconstruct a field which is not fully circularly symmetric;

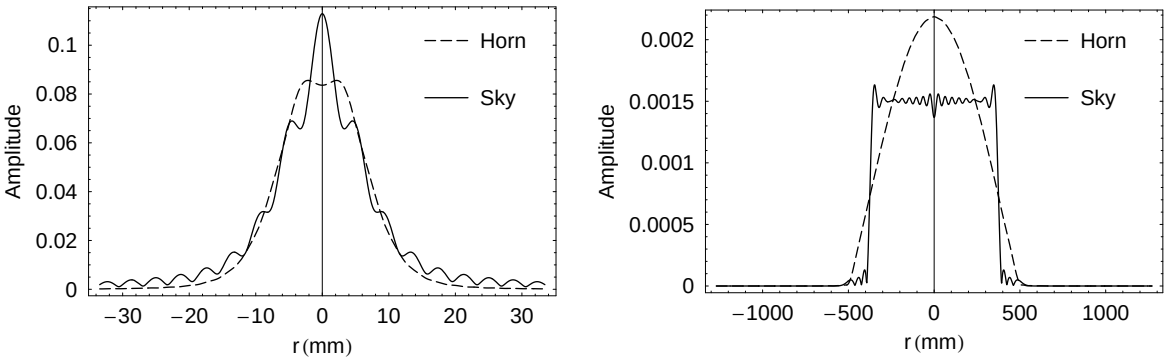


Figure 5.14: A one-dimensional cut through the two-dimensional fields at *left* M3 and *right* the secondary mirror.

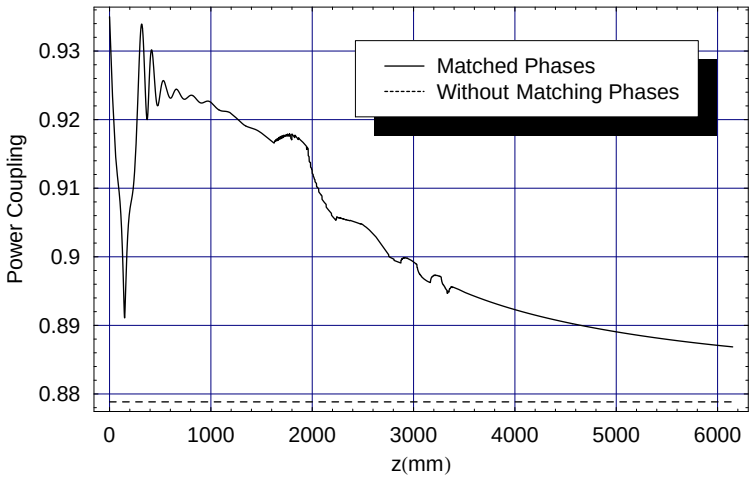


Figure 5.15: The solid line shows the power coupling as a function of phase plate distance from M3, at 0mm, to the secondary mirror location at 6149.44mm, calculated for a phase plate which matches the phases of the two fields. A truncated two dimensional Bessel function is used to model the field from the horn. The dashed line shows the coupling of the fields when the phases are not matched.

however, in this case as the field is assumed to be circularly symmetric, m is set to zero. The mode coefficients of field $E(r)$ are given by

$$A_p = 2\pi \int_0^\infty E(r)\psi_p^*(r) r dr. \tag{5.10}$$

The reconstructed field is then given by the sum of the modes weighted by the appropriate coefficients. The horn field and the field from the secondary mirror are propagated using Gaussian-Laguerre modes in place of the Gaussian-Hermite modes used previously, and using the Bessel function approximation for the horn field. The amplitude of the fields at M3 and the secondary mirror are shown in Figure 5.14; the amplitude of the field from the secondary mirror at M3 differs appreciably from that of the one-dimensional analysis. Figure 5.15 shows the coupling when the phases of the two fields are matched at a plane between M3 and the secondary mirror. This plot is similar to that of the one-dimensional analysis shown in Figure 5.10. With no phase matching the power coupling between the horn and sky fields is 0.88. A peak power coupling of 0.93 occurs if the phases are matched 317mm from M3 (compared to 275.2mm for the one-dimensional case), and a second maximum, with approximately the same coupling, occurs at M3. The phase profiles that match the phase from the sky and horn at M3 are shown in Figure 5.16.

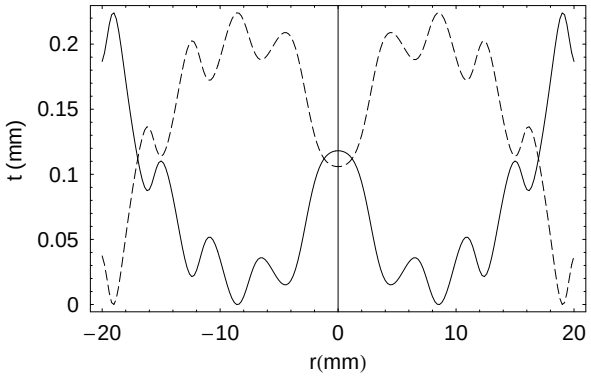


Figure 5.16: A cut through the two-dimensional phase plate profile that matches the phase of the sky and horn fields, the dashed line is given by $\lambda \arg[E_S(r,317)E_H^*(r,317)]/(2\pi(n-1))$, and the solid is given by $\lambda \arg[E_S^*(r,317)E_H(r,317)]/(2\pi(n-1))$. A Gaussian Beam reconstruction of a truncated Bessel function was used to model the horn field.

In this analysis, the horn and sky fields are assumed to be rotationally symmetric. These rotationally symmetric phase plates, of refractive index 1.52, are shown in Figure 5.17; they have a maximum depth of 0.22mm for a 20mm×20mm phase plate.

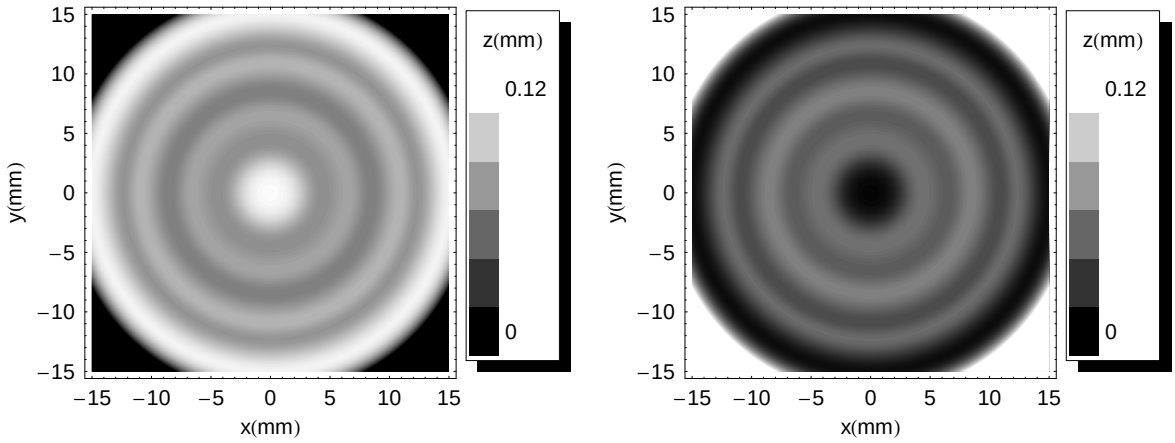


Figure 5.17: Two possible transmission phase plates (of height z) made from a material with a dielectric constant of 1.52, located at M3. The profile is given by Equation 5.8: *Left* $z = \lambda \arg[E_S^*(r,0)E_H(r,0)]/(2\pi(n-1))$, *Right* $z = \lambda \arg[E_S(r,0)E_H^*(r,0)]/(2\pi(n-1))$

5.2.3 Phase Matching using Two Surfaces in One-Dimension

Thus far, only one surface was used to match the phase of the two beams. The feasibility of using both mirrors M3 and M4 is now investigated. In order to maximise coupling of the horn to the field from the sky, both the amplitude and phase must be matched. In order to achieve this, M4 can be used to impose a phase on the field from the horn that produces an amplitude profile as close as possible to the amplitude profile of the field from the sky at M3, where the phase of the two fields are then matched by using M3 to impose an appropriate phase. Alternatively, as is shown in Figure 5.18, it is equally valid to consider M3 as the beam shaping surface and M4 as the phase matching surface. If the horn-field can be exactly transformed to the amplitude profile of the sky-field at M4, then 100% of the power can be coupled. However, as can be observed in Figure 5.19, which shows the horn-field at M4 and the field from the sky-field at M3, the field from the secondary mirror is truncated at M4, which puts an upper limit on the fraction of the power in the telescope field that can be coupled to the horn field. The fraction of power incident from the sky on M3 is given by integrating the intensity of the sky-field over the 20mm-radius mirror as follows:

$$pc = \frac{\int_{-20}^{20} \left|E_S(x, z_h = 140.39)\right|^2 dx}{\sqrt{\int_{-\infty}^{\infty} \left|E_S(x, z_h = 140.39)\right|^2 dx}}, \tag{5.11}$$

where z_h is the distance from the horn and $E_S(x, z_h = 140.39)$ is the field from the sky at M3. The fraction of power incident on M3 is equal to 0.984, giving an upper limit of 98.4% power coupling.

Here, the Gaussian Beam Mode Set Optimisation (GBMSO) beam shaping algorithm, described in Chapter 3, is used for the beam shaping. GBMSO is used in

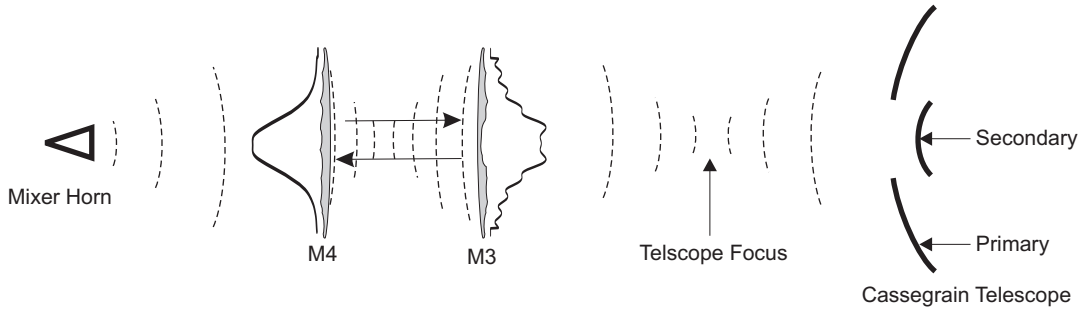


Figure 5.18: The ALMA telescope with mirrors M3 and M4 acting as beam shaping/phase matching surfaces.

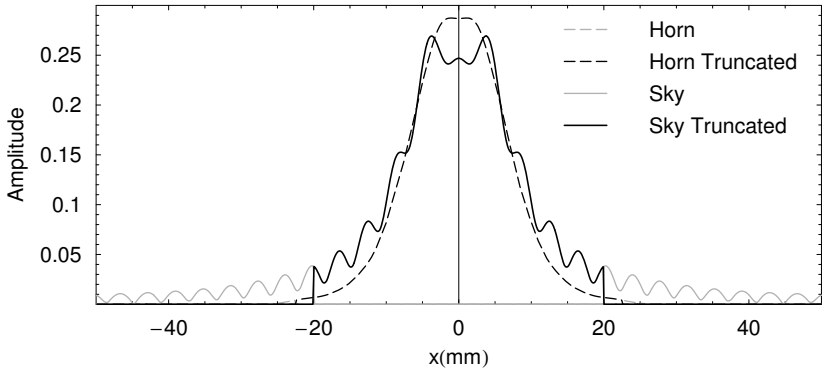


Figure 5.19: The field from the horn at the 22.5mm-radius mirror M4, and the field from the sky at the 20mm-radius mirror M3. The beams are shown with and without the truncation introduced by the mirrors.

preference to the Gerchberg-Saxton Algorithm (GSA) because the GSA transforms between the near- and far-fields, whereas here a near-field to near-field transformation is required; also, GBMSO can limit the maximum spatial frequency of the phase of the solution, giving a smoother profile. Figure 5.19 shows the field from the horn at M4 and the field from the sky at M3. The objective of the GBMSO is to determine the phase profiles that transform the horn-field at M4 to the sky-field at M3 and the sky-field M3 to the horn-field at M4.

As shown in Chapter 3, the merit function used in GBMSO to quantify the similarity of two amplitude distributions, $U_a(x)$ and $U_b(x)$, is given by the overlap integral of the two. Here, the amplitude profile is required over the surface of the mirror only, and the merit function at a given plane is given by

$$\eta(U_a, U_b) = \frac{\int_{-r}^{r} U_a(x)U_b(x)dx}{\sqrt{\int_{-r}^{r} U_a(x)^2 dx}\sqrt{\int_{-r}^{r} U_b(x)^2 dx}}, \tag{5.12}$$

where r is the radius of the mirror. Referring to Figure 5.6, the waist location of the field between the two mirrors is taken as the reference plane, $z = 0$mm. The two mirrors are separated by 95.91mm, with a waist located 42.92mm from M4, and 50.98mm from M3. GBMSO uses Differential Evolution (DE) optimisation (as described in Chapter 2) to find the mode set that best describes both the amplitude

of the horn field at M4 and the amplitude of the sky field at M3. In this way, the phase given by the GBMs at M4, when imposed on the field from the horn, gives a field close to the amplitude distribution of the field from the secondary mirror mirror at M3, and vice versa for the reverse direction. The natural choice for $z = 0$mm, the reference plane for the GBMs, is at the waist location of the telescope field. In this way, when using this location for the waist, the GBMs have a phase-front radius of curvature equal to the radius of curvature of M3 and M4 at their locations, and therefore the GBMSO should give a solution that is close to the shape of the mirrors. The merit functions that quantify how close the field given by the GBMs is to the horn-field at M4 and the sky-field at M3 are

$$\eta_{M4} = \eta(\left|E_{GBM}(x, z = -44.92)\right|, \left|E_H(x, z = -44.92)\right|), \quad (5.13)$$

$$\eta_{M3} = \eta(\left|E_{GBM}(x, z = +50.98)\right|, \left|E_S(x, z = +50.98)\right|), \quad (5.14)$$

where $E_{GBM}(x, z)$ is the field given by the GBMs. The goal is to maximise $\eta = \eta_{M4} + \eta_{M3}$, by optimising the following parameters: the amplitude and phase of the GBM coefficients and the beam waist of the modes, w_0. The optimisation is initialised with values of w_0 in the range $0.4 \rightarrow 1.2$, because, as shown in Figures 5.6 and 5.18, the field has a waist (of 0.885mm) between mirror M3 and M4 and the mode-set of the solution given by the GBMSO is expected to describe the field well using a mode-set waist of this value. The even mode coefficients in the ranges $0 \rightarrow 2$, $0 \rightarrow 4$ and $0 \rightarrow 8$ are optimised giving $\eta = 1.99329$, 1.99543 and 1.99670 respectively. The maximum mode number was restricted to 8 to achieve a solution with a smooth phase distribution, making it more straight-forward to manufacture. The solution at the mirrors M3 and M4, for the even modes in the range $0 \rightarrow 8$, is selected for further analysis because it has the highest value of η. It is shown in Figure 5.20. In Chapter 3, a field with a flat phase was used as the input; however, here the field incident on M4 is not flat, and the phase that M4 imposes on the field is given by the difference between the phase of the horn-field at M4 and the phase of the GBMSO at M4, or equivalently

$$\phi_{M4}(x) = \arg[E_{GBM}(x, z_{GBM} = -44.92)E_H^*(x, z_h = 44.48)]. \quad (5.15)$$

Similarly, the phase that M3 imposes is given by

$$\phi_{M3}(x) = \arg[E_{GBM}(x, z_{GBM} = 50.98)E_S^*(x, z_h = 140.39)]. \quad (5.16)$$

These phase distributions are shown in Figure 5.21. In order to determine the coupling of the field from the horn to the field from the sky with the modified mirrors, the field is propagated from the horn to M3, where the power coupling to

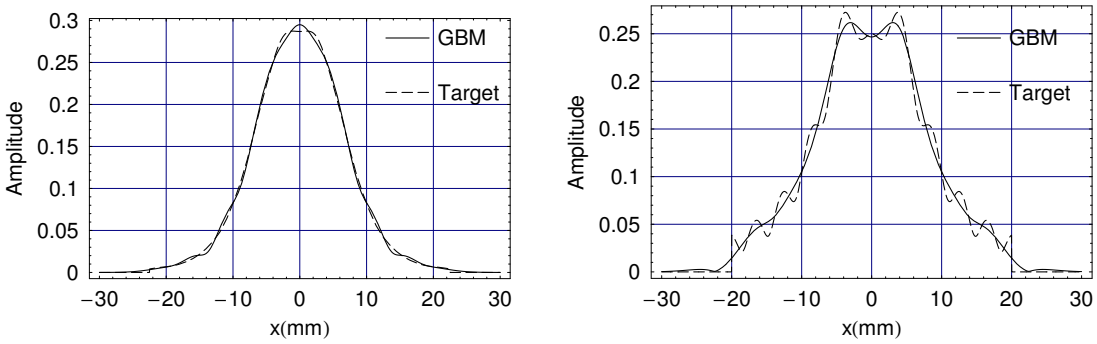


Figure 5.20: *Left* the horn-field at M4 and the GBMSO solution, $\left|E_{GBM}(x, z_{GBM} = -44.92)\right|$, at M4. *Right* the horn-field at M3 and the GBMSO solution, $\left|E_{GBM}(x, z_{GBM} = +50.98)\right|$ at M3.

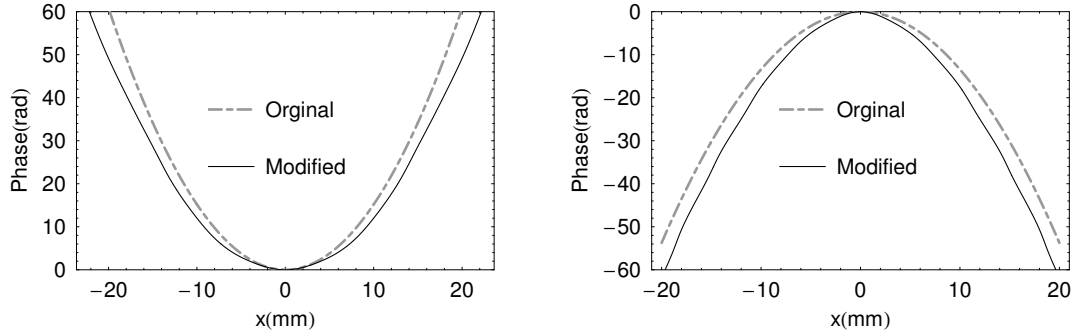


Figure 5.21: The phase imposed by the mirrors (determined by the GBMSO optimisation procedure) and the original unmodified mirror. *Left* M4 and *right* M3

the field from the sky is then calculated. The field immediately after the modified M4 mirror is given by

$$E_{M4}(x) = E_H(x', z_h = 44.48)\exp(i\phi_{M4}(x)). \tag{5.17}$$

The field immediately before M3 is given by the Fresnel integral as follows:

$$E_{M3}(x) = -\frac{i}{\lambda}\frac{e^{ikz}}{z}\int_{-\infty}^{\infty} E_{M4}(x')\exp\left(\frac{ik(x-x')^2}{2z}\right)dx', \tag{5.18}$$

where k is the wavenumber, z is the propagation distance and x' is the coordinate at M4. The field is propagated from M4, using this Fresnel integral, a distance of 95.90mm to M3. The power coupling of the field from the sky to the horn with the modified mirrors is given by

$$pc = \left(\int_{-20}^{20} \frac{[E_{M3}(x)\exp(i\phi_{M3}(x))]E_S(x, z_h = 140.39)}{\sqrt{\int_{-\infty}^{\infty}\left|E_{M3}(x)\right|^2}\sqrt{\int_{-\infty}^{\infty}\left|E_S(x, z_h = 140.39)\right|^2}}dx\right)^2. \tag{5.19}$$

This is the overlap integral of the field from the sky with the field from the horn, including the phase that the modified mirror M3, $\phi_{M3}(x)$, imposes on the fields, which gives a power coupling $pc = 0.978$. The field can then be transformed to the far-field by

$$\text{far-field} = \mathcal{F}\{E_{M3}(x)\exp(i\phi_{M3}(x))\}, \tag{5.20}$$

where $\mathcal{F}\{\}$ is the Fourier transform operator. The far-field intensity distribution, shown in Figure 5.22, is similar to the field at the secondary mirror. Figure 5.23 shows that the phase in the far-field is identical to the phase of the field from the sky at the secondary mirror.

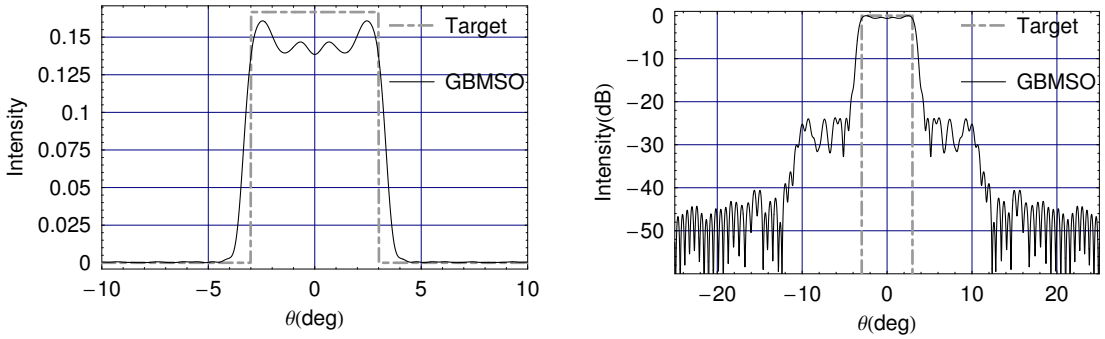


Figure 5.22: The far-field intensity of the field from the horn with beam shaping using M4 and M3 and the target far-field intensity (the field at the secondary mirror). *Left* linear plot and *right* log plot

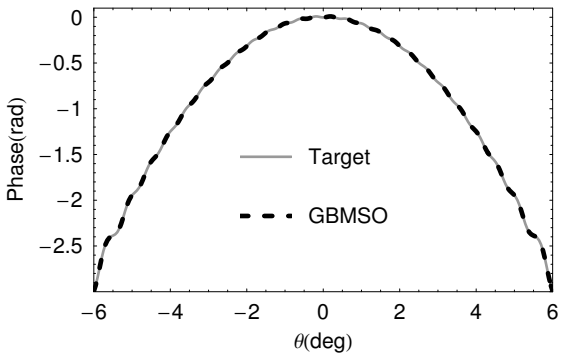


Figure 5.23: The phase in the far-field form the horn with beam shaping and the phase of the field at the secondary mirror. M3 is used as the reference plane.

This work shows that beam shaping can be used to increase the power coupling in band 9 of ALMA and that the modifications required are manufacturable. If this option is pursued, a full 2D determination of the new mirror surfaces can be carried out.

5.3 Conclusion

This chapter has investigated the possibility of improving the efficiency (coupling) of the band 9 optics of the ALMA telescope. Currently, about 88% of the radiation from the sky is coupled to the detector horn. It was shown that profiling the mirror M3 (to match the phases of the sky and horn fields) could increase this to 93%. This has the additional benefit of not introducing any new optical components into the system. This increase in coupling was shown to be dependent on the plane at which the phase matching was done, but in this case no other plane improved significantly on doing the matching at M3 itself. An improvement in coupling is also seen at the edge of the band, as shown in Table 5.4

The same analysis carried out in one dimension (for speed) showed a very similar level of improvement and dependence on the phase-matching plane. The position of the optimum phase-matching plane did differ between the two analyses, but the same conclusions can be drawn. By using both mirrors, both the amplitude and phase of the field from the horn can be controlled to achieve a power coupling of 97.8% (this analysis was carried out in one-dimension for simplicity, but it shows improvements of this magnitude can be expected if beam shaping was used in the telescope). This compares with the maximum possible power coupling of 98.4%; 100% coupling cannot be achieved due to the truncation of the horn-field at M3. In order to couple all the radiation at the design frequency, the mirrors need to be sufficiently large to capture all of the beam from the sky and horn.

Complete control of the field can be achieved with three mirrors: two beam shaping mirrors - as described in Chapter 3 - and one phase matching mirror; al-

lowing a field with any specified amplitude and phase distribution to be produced at the given plane. However, while diffractive elements can be designed to couple the all radiation at a particular frequency, the shape of the mirrors is suboptimal for coupling radiation at other frequencies, restricting their usable bandwidth.

Currently quasioptical systems are designed using geometric techniques; however, the method described above could be extended to optimise the shape of multiple mirrors - using a small number of GBMs, to achieve smooth mirror profiles - for improved performance. Currently telescope mirrors are standard geometric shapes; the improvement in coupling gained by using profiled mirrors is not generally considered worth the extra effort involved in the computation and fabrication. Telescopes designed using geometric techniques also have the advantage that the can be aligned using visible wavelength lasers. As telescopes become more refined, designs based on diffractive optics will allow improvements in coupling and novel designs. Optimisation using GBMs is a useful approach, as it allows the beams to be defined using a small number of parameters, and small numbers of GBMs can be used to design mirrors with smooth profiles.

Chapter 6

The HIFI Instrument On The Herschel Space Observatory

6.1 Introduction

In this chapter, I describe the HIFI instrument and the work I carried out to simulate it, using the software package GRASP (with a model originally built by Massimo Candotti). Next, I describe the work I carried out to simulate the lens antennas used in band 5 of HIFI, using various optic design software packages. Lastly, I compare a number of methods to determine the phase centre location of the simulated lens antenna field. These methods were the standard FWHM method, the far-field phase flattening method, the fundamental Gaussian beam mode method, and also a novel multimoded Gaussian beam fitting method, which I developed.

The Herschel Space Observatory (HSO), when it is launched in 2009, will be the largest infrared space observatory[3]. As the telescope will not be obscured by the atmosphere, it will be able to observe at wavelengths that have not thoroughly previously been explored. It will take four months for the telescope to travel to the second Lagrange point of the Sun-Earth system, 1×10^6km from the earth, where it will spend a nominal mission lifetime of three years. It is 7.5 metres high, 4×4 metres in overall cross section and has a launch mass of around 3.3 tonnes. It will be launched on an Ariane 5 rocket, with the Planck telescope, as shown in Figure 6.1. Herschel has a 3.5 metre diameter Cassegrain design, with three instruments cooled close to absolute zero. The three instruments are:

- Heterodyne Instrument for the Far Infrared (HIFI): a very high resolution heterodyne spectrometer, operating at wavelength ranges between 157-625μm.
- Spectral and Photometric Imaging Receiver (SPIRE): an imaging photometer and an imaging Fourier transform spectrometer, both using bolometer detectors. The photometer provides broadband photometry simultaneously in

bands centered on 250μm, 350μm and 500μm, or spectroscopy with a resolution of at least 100 over the entire wavelength range.

- Photodetector Array Camera and Spectrometer (PACS): an imaging photometer and medium resolution grating spectrometer, covering Herschel's shortest wavelength band: 60-210μm. The 64×32 pixel bolometer array will cover a field-of-view of $1'.75 \times 3'.5$.

This chapter is concerned with the HIFI instrument, which provides continuous frequency coverage from 480 to 1250 GHz in five bands, and coverage of 1410-1910 GHz in a further two bands (shown in Table 6.1) with resolving powers of up to 0.03km/s, and sensitivity close to the theoretical quantum noise limit.

Band	Frequency (GHz)	Mixer Type
1	$480 - 640$	CH/SIS
2	$640 - 800$	CH/SIS
3	$800 - 960$	CH/SIS
4	$960 - 1120$	CH/SIS
5	$1120 - 1250$	LA/SIS
6L	$1410 - 1708$	LA/HEB
6H	$1708 - 1920$	LA/HEB

Table 6.1: The frequency range covered in the seven bands on the HIFI instrument. CH: Corrugated horn antenna, LA: Lens antenna. SIS: Superconductor/Insulator/Superconductor, HEB: Hot Electron Bolometer

6.2 Optical configuration

6.2.1 Overview

The HIFI instrument is housed in a cryostat unit inside the HSO, to cool the telescope thereby reducing thermal noise. The HIFI instrument has four main optical subsystems, which are contained in individual modules: the Common Optic Assembly (COA), the Mixer Assemblies (MA), the Local Oscillator (LO) optics assembly and the Calibration Source Assembly (CSA). The overall layout is shown in Figure 6.2. The radiation from the sky is fed to the COA, after which it is split into the seven individual channels. Five of the mirrors, M3-M5 (shown in Figure 6.3), are common to all the channels. Each of the seven channels is fed to an individual MA - one for each channel. Each channel splits the radiation into two orthogonal polarisations. Each channel has two sub-MAs, shown in Figure 6.4 - one for each polarisation - which focus the radiation onto the mixer horns or lens antennas that feed the detectors.

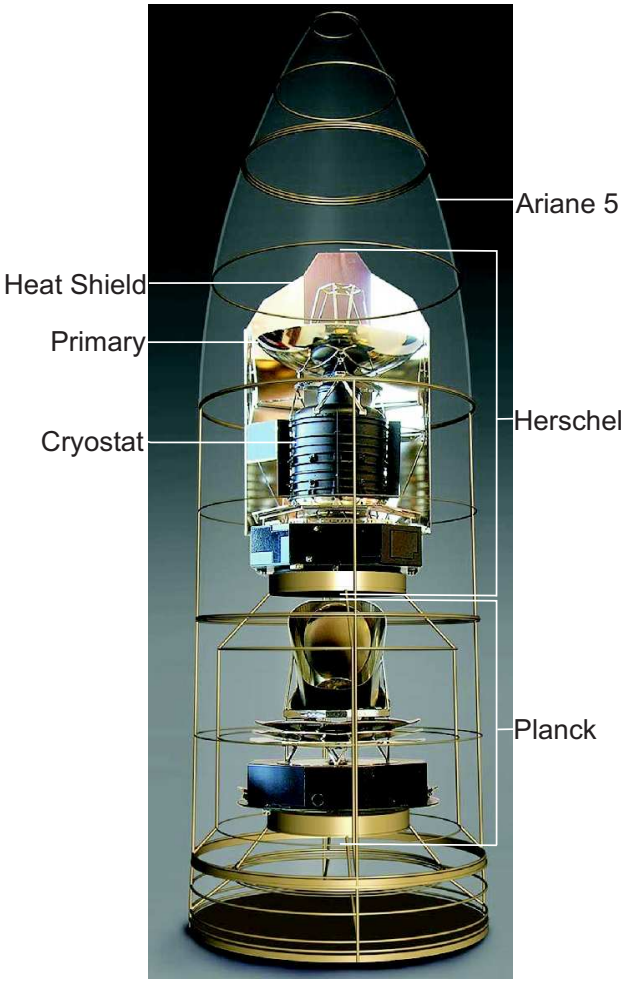


Figure 6.1: Artist's impression of the Herschel and Planck Telescope on the Ariane 5 rocket (Image: ESA)

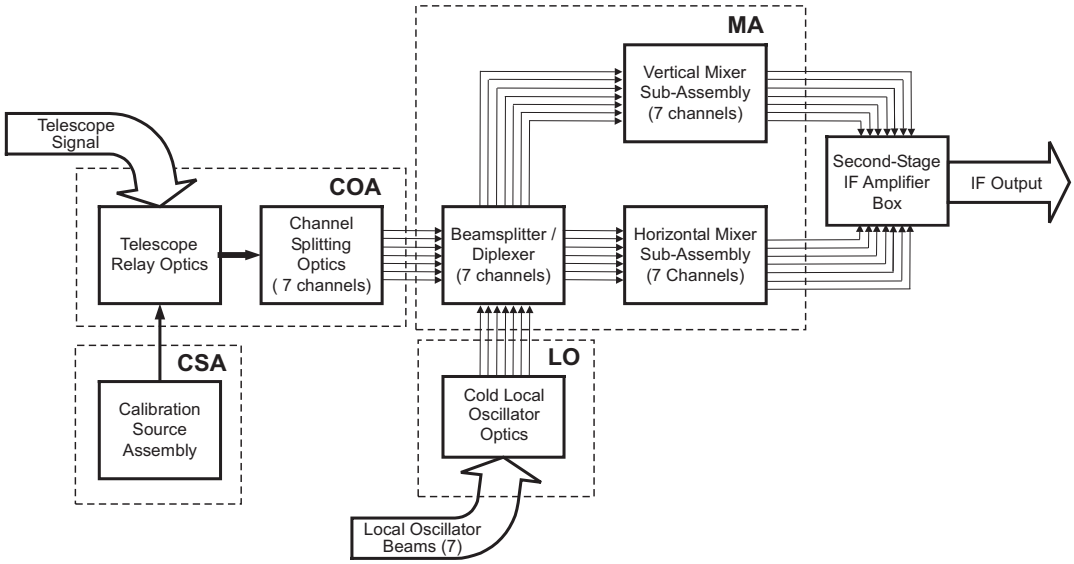


Figure 6.2: An overview of HIFI.

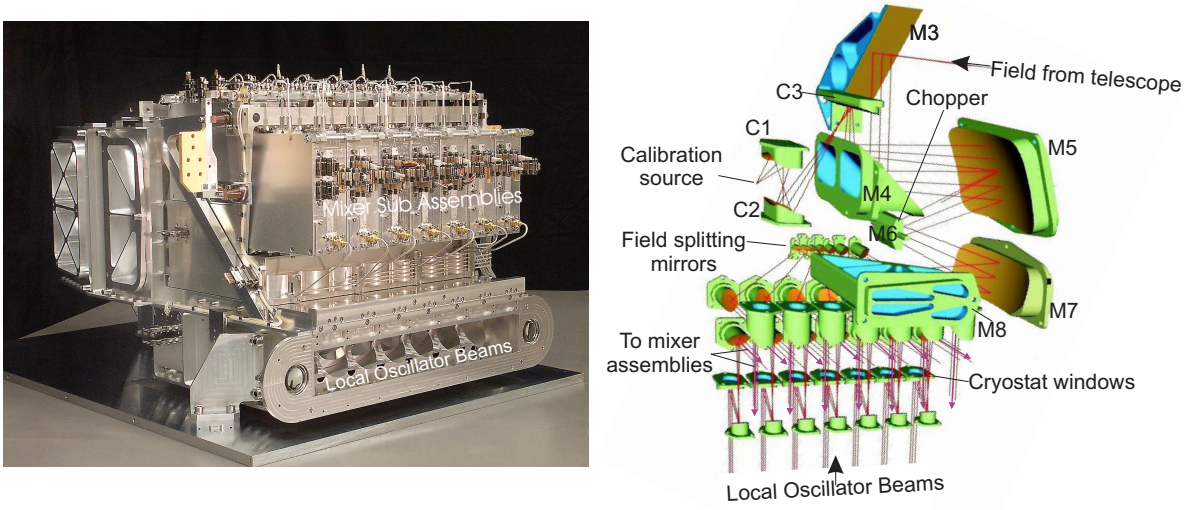


Figure 6.3: The Common Optic Assembly.

HIFI is a single pixel instrument, operating only one of its seven frequency channels at a time. Each channel in HIFI has a unique focal-plane, and its own angular direction towards the sky. The optics relays re-image the radiation from the focal-plane to the diplexers, where the signal is superimposed onto the LO signals. Most observations with HIFI will be made using beam switching. Calibration is achieved by using a focal-plane chopper to switch between an astronomical source and a nearby reference position, with a maximum beam throw of 3 arc-minutes on the sky and chopping frequencies of up to 5Hz. HIFI also has a calibration module: a black-body source with an adjustable temperature in the range of 15-100K.

6.2.2 Common Optics Assembly (COA)

A description of the path of the radiation from the sky through the optics follows[56, 57]. The light incident on the secondary mirror is directed towards the COA at the telescope's focal-plane, shown in Figure 6.3. The COA relays the radiation to the seven separate MAs, each of which corresponds to one of the seven bands covered by HIFI. The first COA mirror in the signal path is M3, which acts as a folding mirror. The signal is then redirected by M4 to mirrors M5, M6 and M7, which act as an Offner system. An Offner system, shown in Figure 6.5, consists of a spherical concave mirror of radius R and a smaller spherical convex mirror of radius $R/2$ that share a common centre of curvature. This provides re-imaging with unit magnification. The mirror with the larger radius, R, has been split in two to reduce mass, resulting in two mirrors M5 and M7. Mirrors C1, C2 and C3 relay the radiation from the calibration source to the Offner system. The flat chopper mirror, M6, is used to select either the calibration or telescope signal, and provides a total scan angular range of about 10 degrees. The flat mirror, M8, folds the beams towards the 7 field

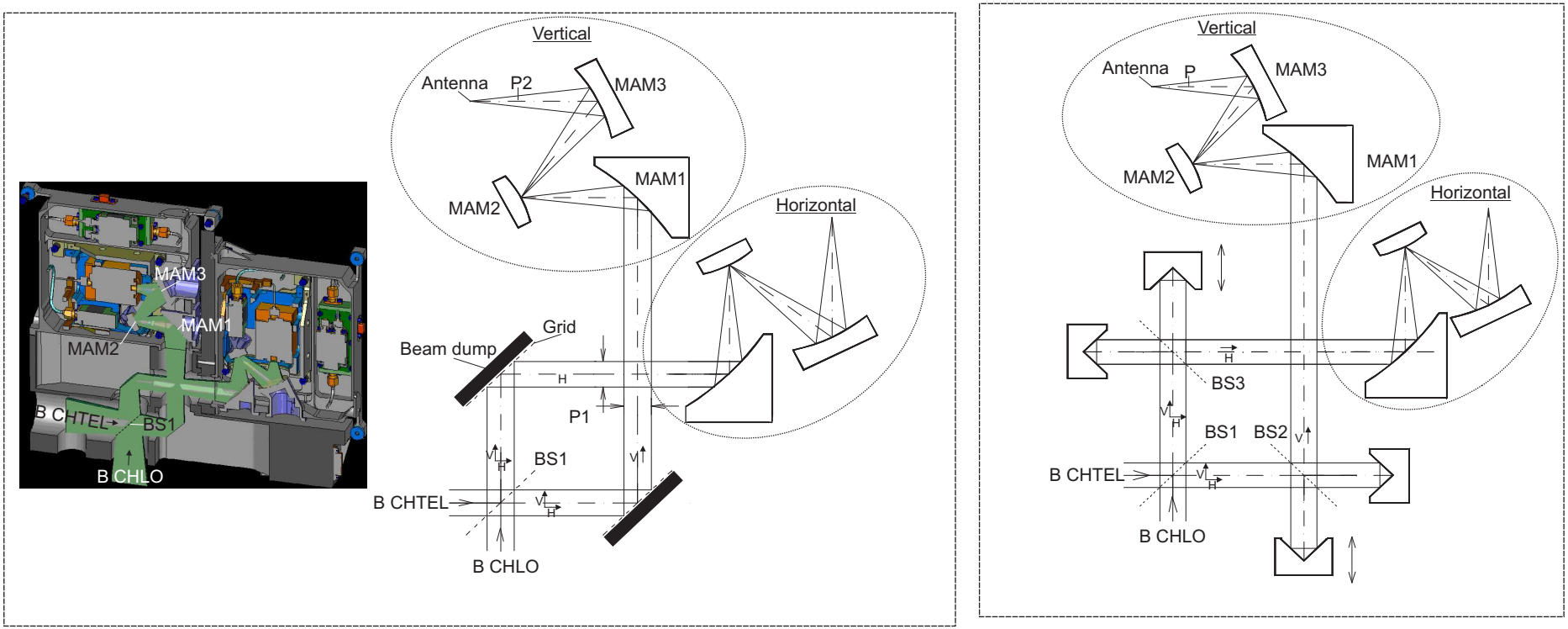


Figure 6.4: Mixer assembly optics of channels: *left* 1, 2 *right* 3 and 4, 5, 6L and 6H. "B CHTEL" refers to the telescope beam and "B CHLO" refers to the LO beam entering the diplexer.[56]

splitting mirrors, where an image of the HSO Cassegrain is located. These mirrors separate the beam into seven beams, 50mm apart, aligned with the optical axes of the 7 MAs. Each of the 7 beams are then re-imaged to their individual MAs. The identical, wavelength independent, LO source units are located outside the cryostat. Each LO beam is relayed to the MAs via the cold LO optics.

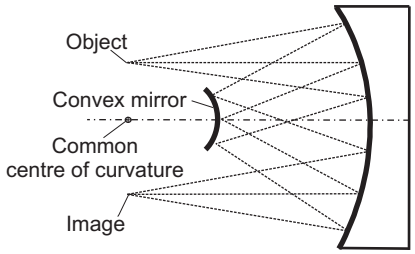


Figure 6.5: Basic optical arrangement of the Offner relay optical system, which provides unit magnification. It consists of a large concave spherical mirror with radius of curvature R, and a small convex spherical mirror with radius of curvature $R/2$. The two mirrors have a common radius of curvature.

6.2.3 Mixer Assembly (MA)

Figure 6.4 shows the two different types of MAs, in which the LO and telescope beams are superimposed using the diplexer. From there, the two combined LO/telescope polarisations each enter the two identical sub-MAs, where they are coupled into the mixer feed-horn or lens antenna and then detected. Channels 1-4 use horns to couple the radiation, whereas the higher frequency channels, 5, 6L and 6H use lens antennas, as horns and waveguides at these frequencies are difficult to manufacture, due to their small dimensions. Figure 6.4 shows the two different configurations for the MAs.

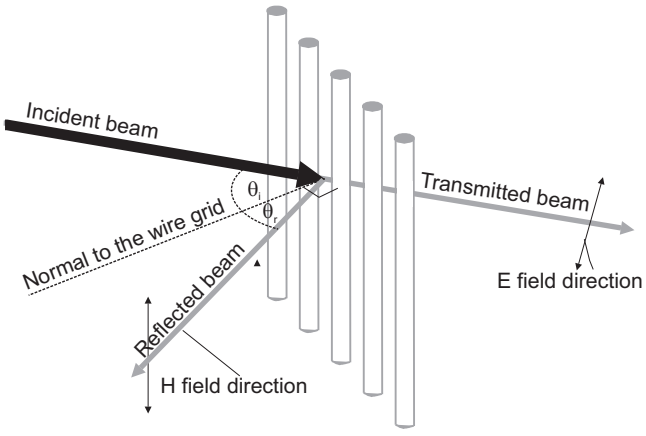


Figure 6.6: The effect of a wire grid on linearly polarised radiation.

For bands 1, 2 the LO and telescope signals are combined with the beam splitter BS1, and focussed onto the antenna using MAM1, MAM2 and MAM3, for both

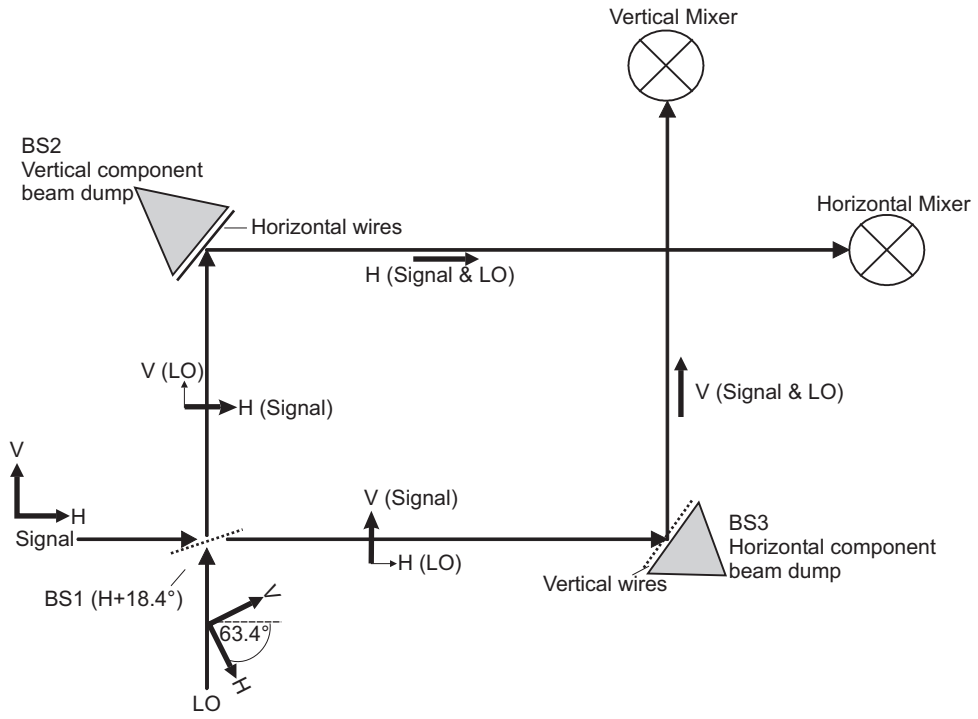


Figure 6.7: The MSA configuration for bands 1 and 2.

vertical and horizontal polarisations. The waist positions are located at P1 and P2 (see Figure 6.4). The field at position P1 is re-imaged to P2 which is also an image of the field at the telescope's secondary mirror.

The vertical and horizontal components of the telescope signal are split by means of a polarising grid, as shown in Figure 6.6[6]. The polarising grids are comprised of parallel wires. A typical polarising grid operates as follows. The incident field can be split into two orthogonal linear polarisations, perpendicular and parallel to the wires. The wire diameters and relative spacing are much less than the wavelength of the incident radiation, and the wires are many wavelengths in length. The parallel component induces a current in the wires, and is reflected. The perpendicular component - which does not induce a current in the wire - is transmitted. The parallel component, therefore, "sees" the grid as a mirror, whereas it is invisible to the perpendicular component. A field of amplitude A incident on the grid with electric vector at an angle θ to the wires is split into the transmitted and reflected components, with amplitudes $A\sin\theta$ and $A\cos\theta$ respectively. The horizontal component of the vertical channel and vertical component of the horizontal channels are dumped by absorbing the field transmitted by the grids.

As shown in Figure 6.7, in HIFI the wire grid polarisers are angled at 18.4° to the plane of the page. Therefore, 90% ($\cos(18.4 \times \pi/180)^2$) of the horizontal signal power is reflected towards BS2, and 10% is transmitted to BS3; whereas, 90% of the vertical component is reflected to BS3, and 10% is transmitted to BS2. The LO is polarised at $45° + 18.4° = 63.4°$, so 50% of the LO power is directed toward BS2,

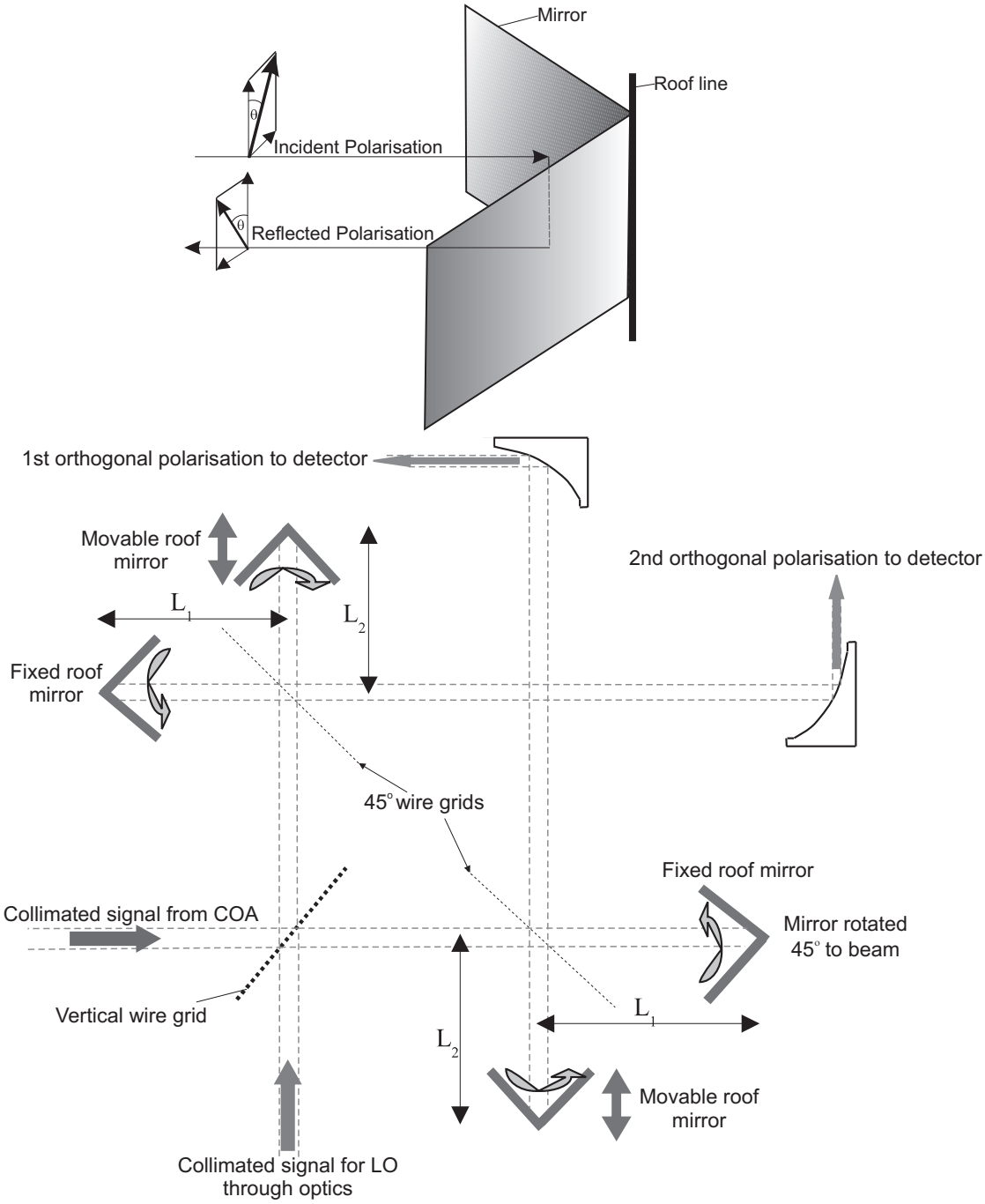


Figure 6.8: *Top:* the roof mirrors of the Martin-Puplett interferometer. *Bottom:* The Martin-Puplett interferometer in the diplexer configuration.[58]

and 50% of the power is directed towards BS3. At BS2, the LO and sky signal are orthogonal and at 18.4° to the horizontal wire grid; the H component is transmitted to the mixer (81% of the original signal power), whereas the unwanted V component is transmitted and dumped. In a similar way, the V component in transmitted and the H component dumped at BS3.

For the higher frequency bands, there is less power available, and dumping the power is therefore avoided. This is achieved by means of a Martin-Puplett interferometer, for channels 3, 4, 5, 6L and 6H. Two identical Martin-Puplett interferometers, shown in Figure 6.8, are used to rotate the polarisation of the component of the field transmitted through the grid, to avoid dumping the more limited LO signal. A roof mirror consisting of two plane mirrors at 90° to each other rotates the polarisation of the incoming beam by 90°. The Martin-Puplett interferometer used in HIFI is a four-port-device, with two input and two output ports, and is capable of broadband operation. Martin-Puplett interferometers are not used in all the bands, because the path length between the beam splitter and the roof mirror must be chosen such that the component reflected by the roof mirror and the component transmitted by the mirror are in phase, and this path length is frequency dependent, resulting in large coupling losses at the edge of the band ($\approx 25\%$).

6.3 Verification of HIFI

It is important to characterise the performance of the telescope modules before they are integrated into the final instrument. The simulated field acts a reference to which measured field can compared. It can then be determined if the as built system is performing as designed. The model can also help to give an understanding as to why the measured and simulated field are not in agreement; a component of the system may be particularly sensitive to tolerancing for example. HIFI was primarily designed using the ray tracing software package CodeV, and the field predicted by CodeV would be expected to deviate significantly from reality, due to the long wavelength of the radiation propagating in the system.

Measurements of the field from the HIFI modules were taken at the Space Research Organization Netherlands (SRON) using a 2D scanner at 480GHz by W. Jellema[59]. A Vector Network Analyser (VNA) is mounted on the scanner, shown in Figure 6.9. The scanner takes measurements across a 2D plane in the x and y directions, and can also take measurements at a number of planes in the z direction; this was achieved with two orthogonal linear stages. The radiation is coupled to a Schottky mixer with a corrugated horn, which allows room temperature measurements. The VNA down-converts the frequencies of the signal using the heterodyne technique, and measures both the magnitude and phase of the signal with a Phase

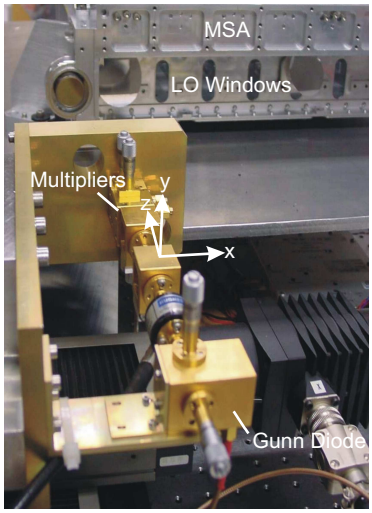


Figure 6.9: The XYZ scanner at SRON.

Locked Loop (PLL). The PLL allows measurements of signals down to around -60dB.

A simple ABCD Gaussian beam mode model (discussed in Chapter 1) was first employed by the author to model the field. This, computationally efficient, method gave a good estimate of the size and the phase radius of curvature of the beam as it propagated through the optics, and allowed a truncation analysis to be performed, as well as an estimate of the coupling of the telescope field to the sky field. As GBMA is a scalar technique, there is assumed to be no cross-coupling between the co- and cross-polar components of the field. However, in reality there is a significant amount of cross-coupling between the two components when they interact with mirror, especially an off-axis mirror. The more rigorous, but computationally intensive, Physical Optics (PO) method was used to simulate the field with the software package GRASP, which can accurately model the co- and cross-polar components of the field. A brief description of the PO method is given below, and a more detailed description of GRASP and the PO method is given in the GRASP manual[60].

6.3.1 The PO Method

PO accurately models the propagation of the field, using an electromagnetic approach; this models the cross-polar as well as the co-polar components of the field.

In PO, the electromagnetic field induces a surface current on the reflective surface, which in turn creates another electromagnetic field. The PO approximation assumes that a mirror can be treated as perfectly conducting, and that a particular point is on an infinite planar surface at a tangent to the scattering surface. Therefore, the radius of curvature must be large relative to the wavelength - a condition that is fulfilled by the optics in HIFI. In the PO approximation, the induced currents are given by

$$J_1^* = 2\hat{\mathbf{n}} \times H^i, \tag{6.1}$$

where J_1^* is the induced electric current, $\widehat{\mathbf{n}}$ is the unit surface normal pointing outward on the illuminated side of the surface, and H^i is the incident magnetic field. The radiated field from a set of induced or equivalent currents is given by

$$\begin{aligned} \mathbf{E} &= -i\omega\left(A^e + \frac{1}{k^2}\nabla(\nabla\cdot A^e)\right) - \frac{1}{\epsilon}\nabla\times A^m \\ \mathbf{H} &= \frac{1}{\mu}\nabla\times A^e - i\omega\left(A^m + \frac{1}{k^2}\nabla(\nabla A^m)\right), \end{aligned} \tag{6.2}$$

where, ϵ and μ are the permittivity and permeability respectively, ω is the angular frequency and k is the wavenumber, distance R is given by $R = r - r'$ where r is the observation point and r' is the integration variable running over the surface; finally, A^e and A^m are the electric and magnetic vector potentials given by

$$\begin{aligned} A^e &= \frac{\mu}{4\pi}\iint_B J^e(r)\frac{e^{-ikR}}{R}ds \\ A^m &= \frac{\epsilon}{4\pi}\iint_B J^e(r)\frac{e^{-ikR}}{R}ds. \end{aligned} \tag{6.3}$$

Applying the differential operators on A^e and A^m in Equation 6.2 yields

$$\begin{aligned} D(J) &= J\left(-\frac{i}{kR} - \frac{1}{k^2R^2} + \frac{i}{k^3R^3}\right) + \left(J\cdot\widehat{R}\right)\widehat{R}\left(-\frac{i}{kR}\right) - \frac{3}{k^2R^2} + \frac{3i}{k^3R^3} \\ E(r) &= \frac{\zeta}{4\pi}\iint_B D(J^e)e^{-ikR}k^2ds - \frac{1}{4\pi}\iint_B J^m\times\widehat{R}\frac{(1+ikR)e^{ikR}k^2}{k^2R^2}ds, \\ H(r) &= \frac{1}{4\pi\zeta}\iint_B D(J^m)e^{-ikR}k^2 + \frac{1}{4\pi}\iint_B J^e\times\widehat{R}\frac{(1+ikR)e^{-ikR}k^2}{k^2R^2}ds, \end{aligned} \tag{6.4}$$

where $\widehat{R} = r - s$ is the vector from a source point s on surface S, and ζ is the impedance of free space, $\sqrt{\mu/\epsilon}$.

The field at a plane can be calculated by numerically evaluating Equations 6.4. As the integration is particularly computationally demanding, the grid of points at which the integral is evaluated should be minimised in order to reduce computation time, while a sufficient number should be used in order that the result is accurate. The best choice grid for performing the integration is dependant on the shape of the scatterer. A natural choice of coordinate system for integrating over a curved surface - such as the shape of the mirrors used in HIFI - is the polar coordinate system shown in Figure 6.10 *left*. However, if the integration integral is evaluated at regular intervals of ϕ and ρ, as shown in Figure 6.10 *left*, there is a bias in the number of points near the centre. In order to compensate for this, instead of having a constant number of integration points at a given angle, N_ρ, and at a given radius,

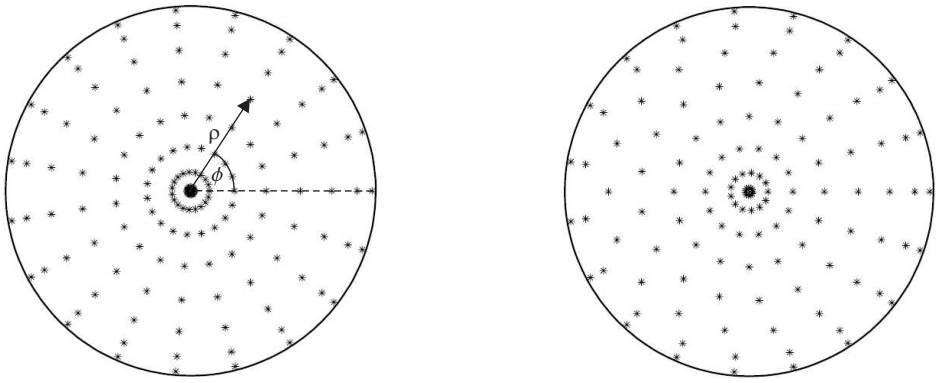


Figure 6.10: Possible integration grids used when evaluating the PO integral with $N_\Phi = 19$ and $N_\rho = 8$. *Left* polar coordinate grid. *Right* the reduced grid used by GRASP.

N_Φ, an N'_Φ which is dependant on ρ is used to reduce the number of points at the centre as shown in Figure 6.10 *right*.

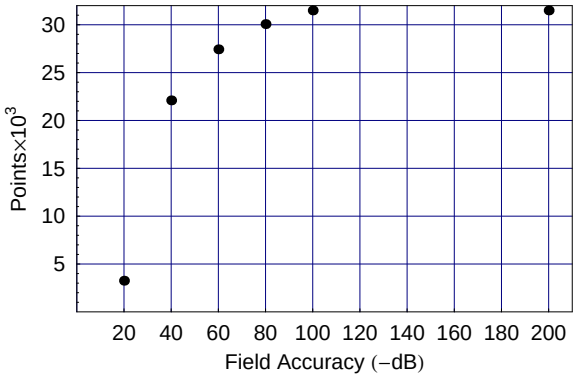


Figure 6.11: The number of integration points required as a function of the specified field accuracy for a offset single-reflector system example of the GRASP manual[60].

GRASP has an automatic convergence procedure, to determine when a sufficient number of integration points have been used to achieve the specified accuracy. The user specifies the "field accuracy", which is the maximum value, given in dB, below which the convergence criteria has been meet (a field accuracy of 80 dB, for example, means that the field error due to convergence is at least 80 dB below the maximum value of the field at any of the field points - corresponding to ripples of 1 dB at the 60 dB level) as follows. The integration density is given by two numbers, N_ρ and N'_Φ, which specify the number of integration points. The integration is performed at number of test points with $N_\rho = N'_\Phi = 10$; N_ρ is then doubled, and the integration is calculated with $N_\rho = 20$ and $N'_\Phi = 10$. Convergence is then tested; convergence is achieved if the fields differ by less than $10^{\text{field accuracy}/20}$ at all the test points. N_ρ is then doubled and compared with the previous calculation in this way, until the convergence criterion is met. The same procedure is then carried out with N_ρ set to 10 and N'_Φ doubled, until the convergence has been achieved. Figure 6.11 shows the

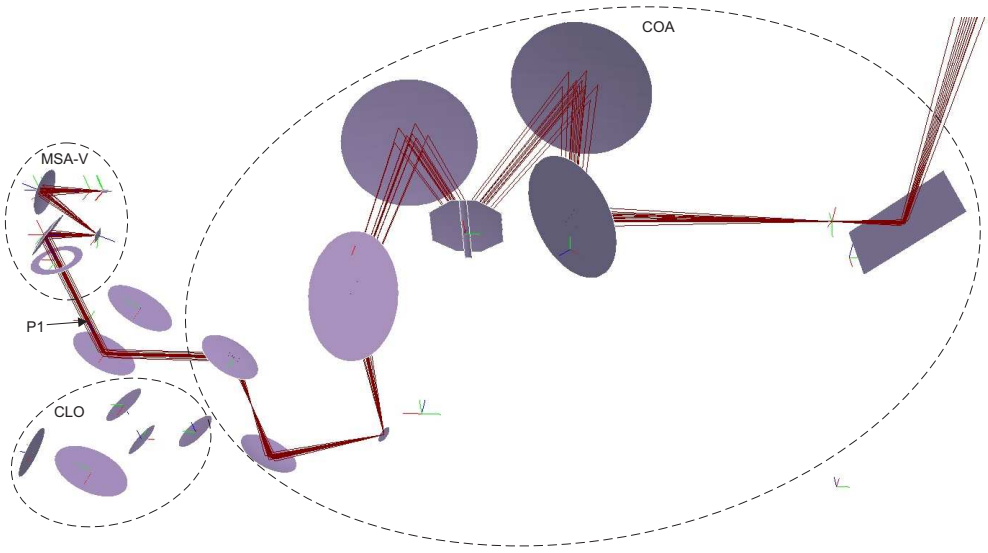


Figure 6.12: A graphical representation of the HIFI optics in GRASP.

number of points required for a given field accuracy, using data from the GRASP manual[60], for a offset single-reflector system. Increasing the accuracy from -20dB to -80dB increases the number of points, and hence the integration time, by a factor of 9.2. A further increase in the field accuracy requires only a small increase in the number of points.

The mirrors are defined as polynomials in GRASP. The location of each mirror (at the chief ray intercept) is defined relative to a global coordinate system. A local coordinate system for each of the mirrors allows rotation and displacement, in order to model the effects of misalignments. Figure 6.12 shows the graphical representation of the HIFI system in GRASP. The field was calculated using PO at the measurement plane, P1 (shown in Figure 6.12), to ensure that it was similar to the measured.

6.4 Lens Antennas

The mixer horns used in channels 1 to 4 of HIFI produce fields that are well characterized and which have a theoretical Gaussicity of 98%, providing a field that couples well to the signal field. The use of waveguides for the higher frequency channels, however, is not feasible. Although, semiconductor lithographic techniques may be employed in the manufacture of the miniature horns that would be required at these higher frequencies, the attenuation of fundamental-mode waveguides increases with frequency, and the effects of both metal tolerances and misalignments become more severe[61] (as the skin depth gets smaller with increasing frequency, and the surface roughness of the walls of the feed-horn and waveguide structures become increasingly more important). Planar antennas, which consist of patches or slots, have the advantage that they are relatively simple to fabricate, compared to waveguides, but suffer from low directivity and loss of power into surface waves.

The solution chosen for the higher frequencies bands in HIFI (5, 6L and 6H) is to use lens antennas instead of horns or slots. A lens antenna is a dielectric lens mounted onto a planar antenna. The main function of the lens, is to focus the radiation onto the detector, thereby increasing its directivity[62]. An elliptical lens, shown in Figure 6.13, redirects rays in order that they are parallel at the radiating aperture when the feed is placed at the focus. Also shown in Figure 6.13 is the hemispherical surface that can be used to approximate the elliptical one, for ease of manufacture. A secondary benefit of using the dielectric lens is that it gives mechanical rigidity to the device. Although the Gaussicity of a lens antenna is less than that of a waveguide ($\sim$94%), they are cheaper to fabricate than the waveguides used at these frequencies, and have a large spectral bandwidth[2].

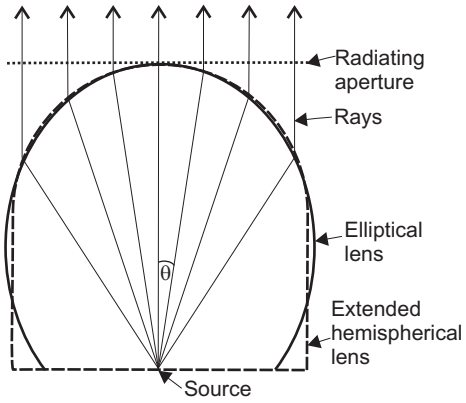


Figure 6.13: A lens antenna typically creates parallel rays at the aperture plane; although the HIFI lens antennas give a field with a quasi-focus a distance in front of the lens.

Lens antennas are commonly used as advanced receivers for terahertz spectroscopy systems, automotive radars, and a number of other industrial and medical applications[63]. They are a popular choice in part due to their capability of being integrated with electronic components such as detecting diodes, local oscillators and mixers[64].

Optical theory requires that the lens electrical size is in the order of several wavelengths, in order to refract the radiation. As a result, the minimum size of the lens is proportional to its operating frequency, and therefore lens radii are rarely smaller than 3λ, making them impractical for use at the lower microwave frequencies, where they are prohibitively large and heavy, suffer from large substrate losses and are expensive. Lens antennas are used predominantly in millimetre-wave imaging and quasioptical systems.

6.4.1 HIFI Lens Antenna

The lens antennas used on HIFI are composed of an extended elliptical silicon-dielectric lens mounted onto a double slot planar feed antenna, which has a very

Parameter	Band 5	Band 6L	Band 6H
Lens Diameter (mm)	5.0	5.0	5.0
Ellipticity (mm)	1.019	1.019	1.019
Extension Length (mm)	0.756	0.756	0.756
Slot Spacing (μm)	44.0	32.0	28.0
Slot Length (μm)	77.4	56.0	50.0
Slot Width (μm)	4.0	4.0	3.6

Table 6.2: The dimensions of the lens antennas and double slot antennas used on HIFI.

similar shape to the extended hemispherical lens of Figure 6.13. The planar feed (double slot) antenna, shown in Figure 6.14, is connected to the active circuitry, which consists of a detecting diode and mixer. The lens antenna used on HIFI focusses the radiation a short distance in front of the lens. The silicon used for the dielectric lens on HIFI has a dielectric constant of 11.68, and is dispersionless over the entire bandwidth of operation[2]. The lens has a $1/4\lambda$ matching layer to reduce reflections.

The author was involved in modelling the field produced by the lens antennas which are used in bands 5, 6L and 6H. Software called PILRAP[65] was used to produce the far-field pattern for the lens antenna. PILRAP treats the antenna as a transmitter, as opposed to a receiver, using ray tracing to determine the field at the lens surface and physical optics to calculate the far-field pattern of both the co- and cross-polar components of the field. The field for the lens antenna for these bands was also calculated in the SRON technical report using PILRAP in [66]; the fields given in this section were verified using those found by the author and transformed to the waist location. Figure 6.15 shows a screen shot of PILRAP with the band 5 parameters. The dimensions of the lens antenna and the feeds are given in Table 6.2. Bands 5, 6L and 6H were analysed; the detailed results for the band 5 lens are given here. Figure 6.15 shows a screen shot of PILRAP with the band 5 parameters.

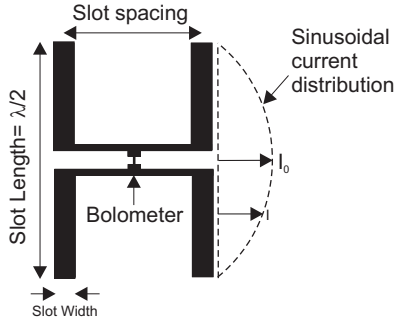


Figure 6.14: The double slot antenna on the back of the lens antenna.

PILRAP produces the field of the lens antenna in polar coordinates and in the

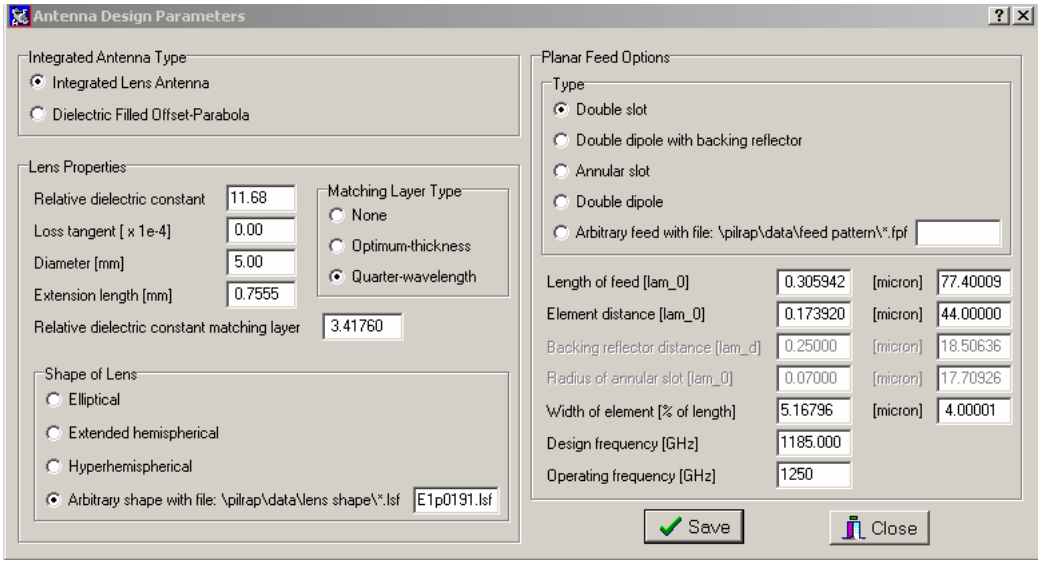


Figure 6.15: Screenshot of PILRAP software, which models lens antennas.

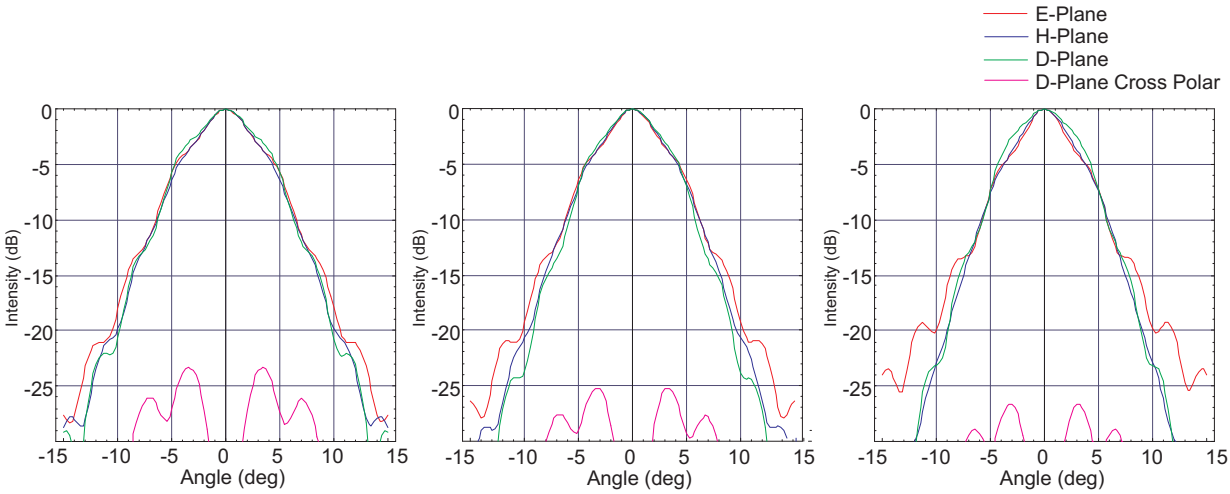


Figure 6.16: The far-field of the band 5 lens antennas at *left* 1120GHz, *middle* 1185GHz and *right* 1250GHz.

Frequency (GHz)	Waist (mm)	Position (mm)	Gaussicity (%)
1120	0.74	14.3	94.43
1150	0.73	14.3	94.71
1185	0.73	14.2	94.90
1220	0.73	14.2	94.89
1250	0.73	14.2	94.70

Table 6.3: The size of the waist, the waist location and Gaussicity of the field across the operational bandwidth of the band 5 lens antenna, as calculated by PILRAP.

far-field. Figure 6.16 shows cuts of the far-field from the band 5 lens antenna given by PILRAP at 1120GHz, 1185GHz and 1250GHz; these far-field patterns agree with [66]. The cross-polar level is less than -23dB across the band. The Gaussicity of the field was found by determining the best fit Guassian to the field. Table 6.3 shows the Gaussicity of the lens antenna field for band 5, calculated by PILRAP across the operational bandwidth; the Gaussicity is in the range 94.43% to 94.90% across the band.

In order to model the propagation of the lens antenna field through the band 5 optics the near- rather than far-field was required. PILRAP used a default reference plane location in order to calculate the far-field. At the design frequency of 1185GHz, the location of the phase centre is found to be 14.2mm in front of the PILRAP reference location: the interface between the elliptical surface and cylindrical sections. In order to correct for this, the spherical phase component of the field, as found by fitting a Gaussian amplitude distribution with a spherical phase-front, was subtracted. The field was then decomposed into associated Gaussian-Laguerre modes with the GBM parameter $z = \infty$, (associated Laguerre, rather than Hermite, polynomials are used as the field is essentially circularly symmetric, reducing the number of modes required to accurately reconstruct the field). With the spherical component removed, when the field, given by the sum of the GBMs, is propagated to $z = 0$, it is located at the phase centre. (This procedure is explained in more detail in Section 6.5)

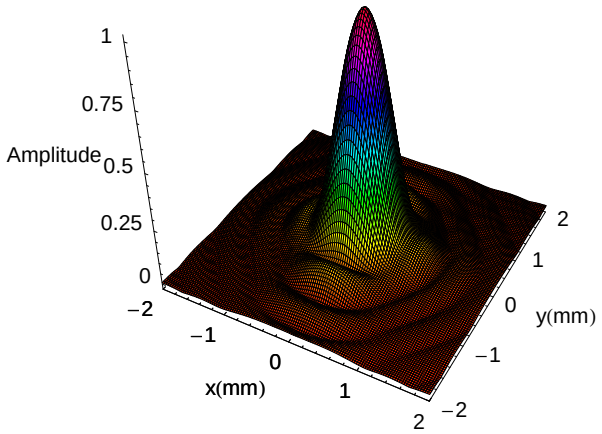


Figure 6.17: The field at the waist position for the band 5 lens antenna at 1185GHz.

The field was also transformed to the waist using the FFT instead of associated Laguerres. For the two-dimensional FFT, the field is required in Cartesian coordinates, so the field reconstructed by the associated Laguerre polynomials was resampled on a Cartesian grid. The field at the phase centre given by the Laguerre polynomial's transformation and FFT were identical, as expected. The field at the

Parameter	HIFI	NUIM
Operating λ	0.253	2.99
Diameter	5.0	30.0
Extension length	0.732	14.0

Table 6.4: The original silicon HIFI lens and the HDPE scale model of the lens used at NUI Maynooth.

phase centre (for an operational frequency of 1185GHz) is shown in Figure 6.17.

6.4.2 Model of a HDPE Lens Antenna

The use of lens antennas in quasioptical systems is relatively new and is not yet fully understood. To gain an understanding of the operation of this type of feed and how it can be modelled, a lens antenna was manufactured by Curran[67] and tested at NUI Maynooth. It was not possible to use an exact model of the band 5 silicon lens antenna used on HIFI, as the VNA, used to measure the field produced by the lens antenna, operates at frequencies of around 100GHz (whereas the band 5 HIFI lens operates at 1185GHz), and silicon is difficult to machine. A HDPE lens was chosen, as it has a relatively large refractive index and can be easily machined. As the refractive index of HDPE is 1.52, and is lower than for silicon (which has a refractive index of 3.14) a relatively longer extension length is required for the HDPE lens than that of the silicon lens in order to focus the field. The dimensions of the HIFI and NUI Maynooth lens are given in Table 6.4. The lens antenna is fed by a waveguide with a cross section of 2.54×1.27mm.

Gaussian Beam Mode Analysis

Here, Gaussian Beam Mode Analysis (GBMA) is first employed to predict the field from a lens antenna. GBMA can take diffraction effects into account and is computationally efficient. With GBMA the field is assumed to be paraxial, i.e. the field does not diverge too greatly as it propagates. The transition between the paraxial and non-paraxial regime is gradual and [68] gives an indication of the applicability of GBMA to four different regimes: The radius of the fundamental Gaussian that approximates the field at the waist location can be used to give an indication of the validity of the paraxial approximation. For $w_0/\lambda \geq 0.9$ the paraxial approximation is valid. For $0.5 \leq w_0/\lambda \leq 0.9$ a first order correction is required to give an accurate description. For $0.25 \leq w_0/\lambda \leq 0.5$ the first order correction is no longer sufficient. For $w_0/\lambda \leq 0.25$ the paraxial approximation introduces significant error.

The Gaussian that couples best with the field at the waveguide is determined by maximising the coupling of a fundamental Gaussian mode to the waveguide field. A maximum coupling of the Gaussian field to the waveguide occurs for $w_0 = 1.2$mm,

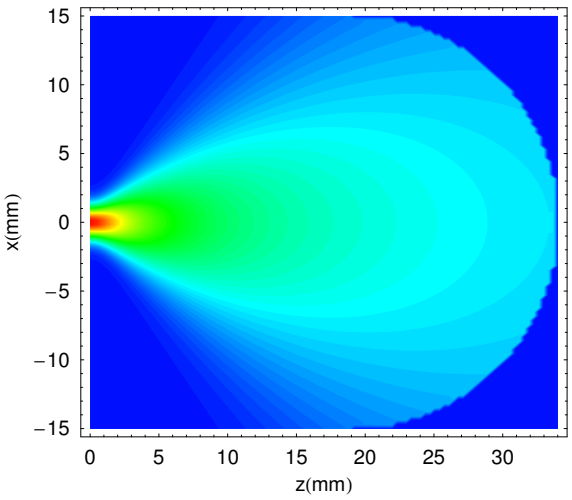


Figure 6.18: The field propagated in the lens using the fundamental GBM approximation.

giving a value of w_0/λ of 0.4. Therefore, the GBMA can only give an approximation of the propagation of the field. The propagated field predicted by GBMA diverges less rapidly than it would do in reality. Therefore, the on axis amplitude decreases more rapidly in the physical field. Figure 6.18 shows the Gaussian approximation to the field propagated within the lens, showing the approximate distribution of the field.

Ray Tracing

Lens antenna models employing Gaussian beam modes and physical optics to model the propagation from the planar antenna to the lens surface can only achieve an approximate prediction of the fundamental Gaussian beam properties, but otherwise a poor prediction of the structure of the beam, and therefore a multimodal GBM analysis was considered not worth pursuing.

The commercial software package Zemax was used to model the lens using ray tracing. Ray tracing assumes that the wavelength of the field is small in comparison with the dimensions of the component. The rays propagate along straight lines, without interfering with each other. Ray tracing is fast, and can therefore be used to model multiple reflections; however, rays are not suited to modelling diffraction. If the surface of the lens is in the far-field of the feed, ray tracing can be employed. There is no exact cut-off for the far-field, but it is considered to be the region where the propagation distance is much greater than the confocal distance, where the confocal distance is defined as follows

$$z_c = \frac{\pi w_0^2}{\lambda}. \tag{6.5}$$

The confocal distance for the field from the waveguide is 1.53mm. As the extension

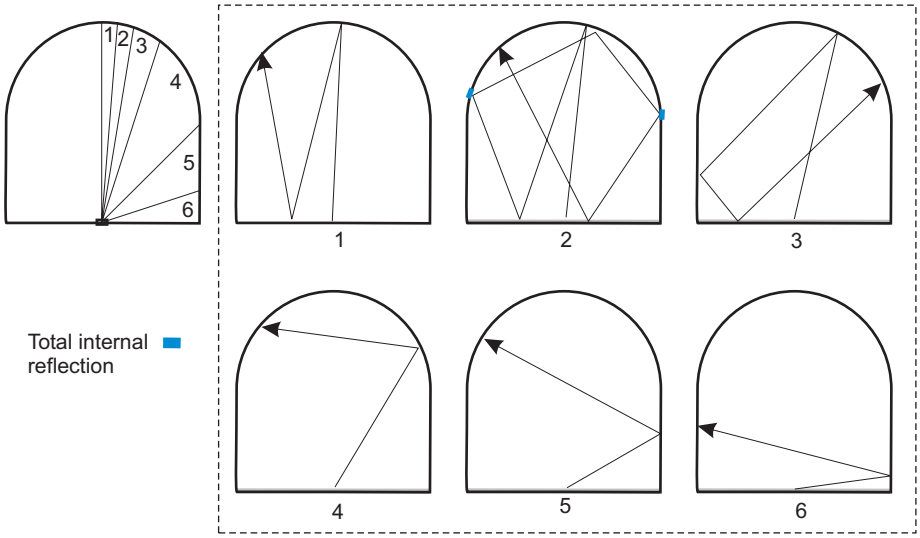


Figure 6.19: The path of a ray depends on the angular domain at which the ray is launched.

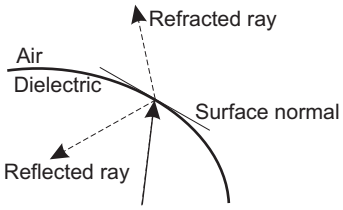


Figure 6.20: The refraction of a ray at a dielectric/air interface.

length of the lens is 19mm, the surface of the lens can be considered to be in the far-field of the feed.

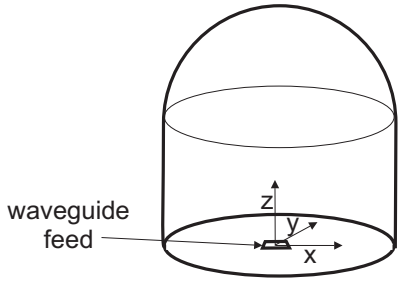


Figure 6.21: The HDPE lens antenna with the waveguide and coordinated system used for the ray tracing.

Firstly, it is useful to split the rays into angular domains, in order to gain an understanding of the path of the rays for a given launch angle[69]. Figure 6.19 shows six different angular domains. When the ray intercepts with a surface it splits into two rays (the refracted and reflected rays), as shown in Figure 6.20, except where the angle of incidence is greater than a critical angle where total internal reflection occurs. This reflected radiation should not be considered a loss, but as interference, because part of the reflected wave can leave the lens after one or more internal reflections, and can therefore affect the radiation pattern. The cases of Figure 6.19

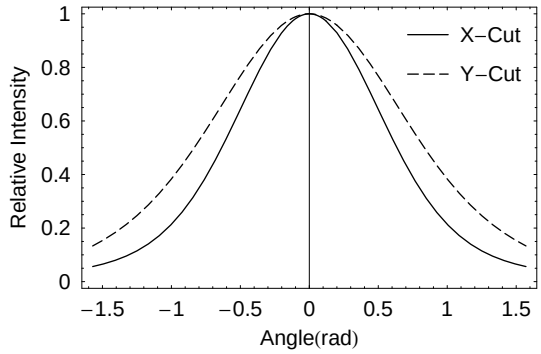


Figure 6.22: Cuts of the far-field from the 2.54×1.27mm rectangular waveguide, which is the lens antenna's feed.

show the reflected ray where there is a significant amount of power in the ray. Rays in domains 3 to 6 do not contribute to the main beam of the lens antenna.

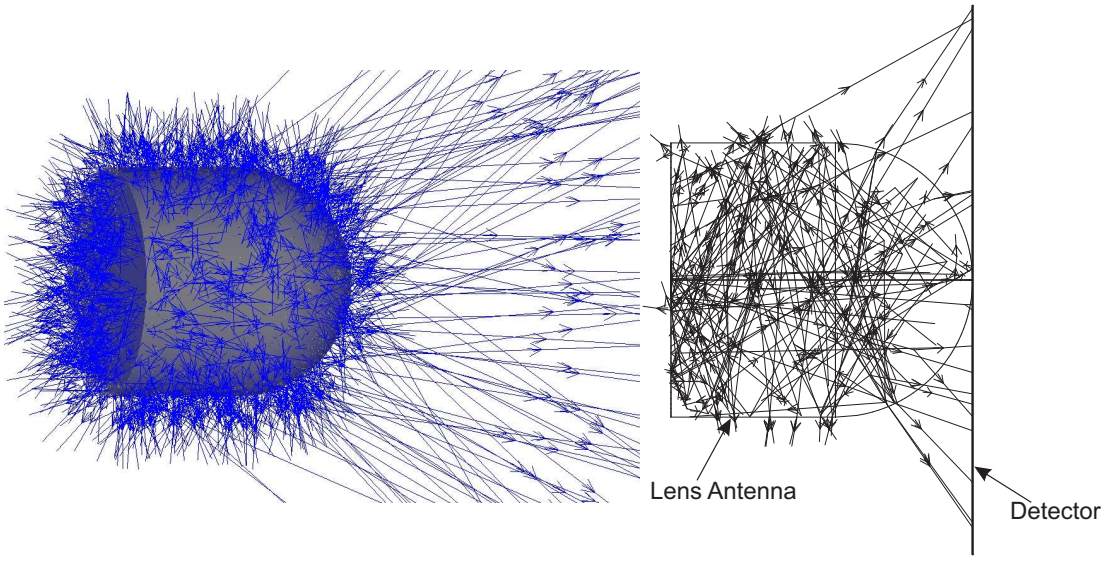


Figure 6.23: The path of a single ray through *left* 3D view, *right* cross section.

The lens is modelled using the ray tracing "non-sequential" mode of Zemax. The non-sequential mode allows the rays to be reflected multiple times within the lens. Modelling the lens in Zemax using the non-sequential mode involves defining the shape of the lens and generating a file which defines the far-field from the source. The lens is defined using the "Standard Lens" 3D object, to create a model of the HDPE lens with the dimensions given in Table 6.4. The field from the waveguide is input as a file containing a list of far-field intensity values of the source over a range of unit vectors, (l, m, n), which relate to directions (x, y, z), as shown in Figure 6.21. The far-field from the waveguide is modelled using the analytical Fourier transform techniques described in Chapter 1 (cuts in the X and Y planes are shown in Figure 6.22), and a tabulated ASCII file in the format file specified in the Zemax manual[50] is generated. In the tabulated file, the far-field intensity values are given for random values of launch angles in the direction of a hemisphere towards the lens antenna,

with random values of l and m in the range $-1 \rightarrow 1$ and $0 \rightarrow 1$ respectively. A 100×100mm detector is placed at the front of the lens antenna, as shown in Figure 6.23. As many rays do not hit the detector, a large number of rays are required; 4×10^6 rays are used in this analysis. Figure 6.23 show the path of the rays generated (by ray splitting as shown in Figure 6.20) from a single input ray. The field at the detector was output to a file, and was then transformed to the far-field using the FFT.

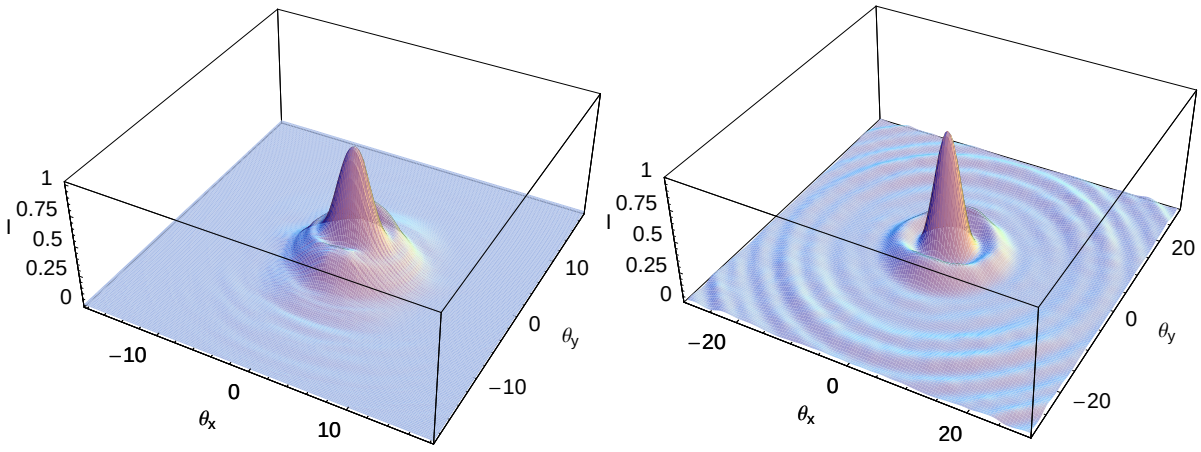


Figure 6.24: The intensity of the field from the lens antenna: *left* measured and *right* simulated.

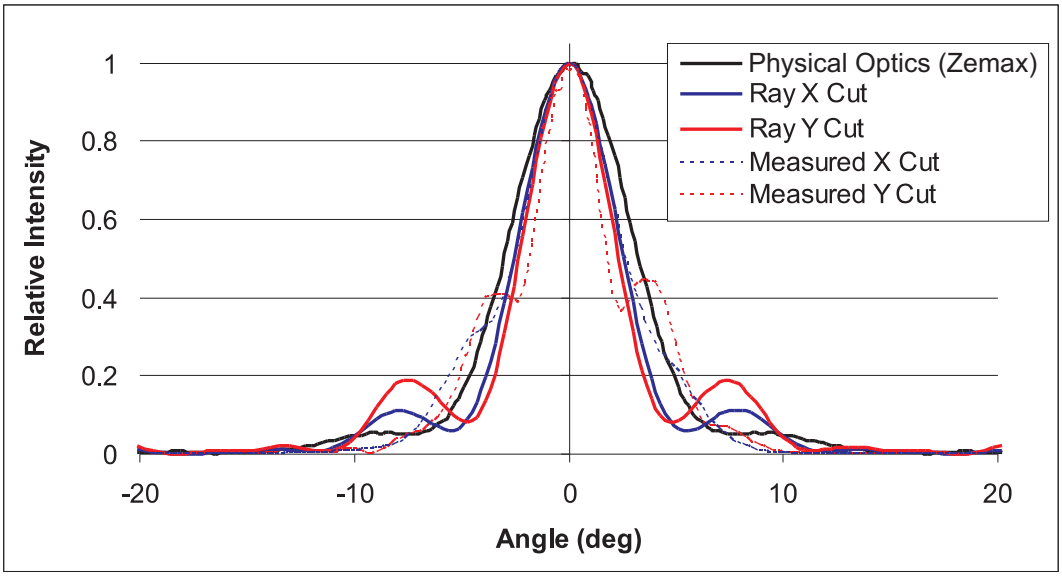


Figure 6.25: The measured and predicted far-field from the HDPE lens antenna.

As well as ray tracing, the field was also propagated using the angular spectrum propagator in Zemax. The angular spectrum propagator is based on Fourier optics and takes diffraction of the beam into account; however, it does not model the reflections inside the lens.

The field from the lens antenna was measured using a VNA mounted onto an XY scanner. The radiation from the lens antenna was coupled to the VNA using a bare waveguide. Figures 6.24 and 6.25 show the simulated and measured far-field intensity distributions. The simulations predict the size of the beam, but not the detailed structure of the beam sidelobes. Small deviations in the location of the feed were found to create large deviations in the predicted far-field pattern. This analysis shows that manufacturing tolerances are critical to create a lens antenna which performs as designed.

6.4.3 Simulation of the Band 5 MSA

The author was involved in simulating the band 5 MSA using GRASP. There was a large discrepancy between the field measured and simulated at P1. The steps taken to determine the cause of this discrepancy are outlined here.

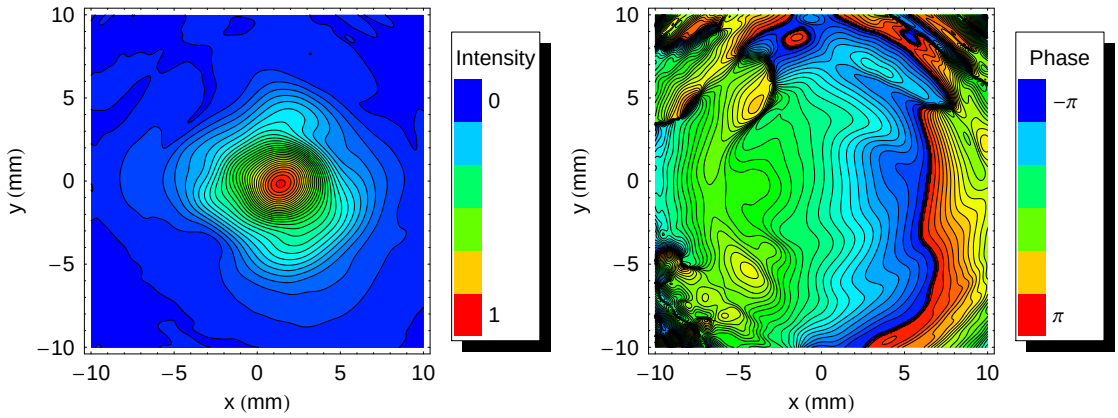


Figure 6.26: *Left* 1190GHz GRASP simulation co-polar intensity and *right* the phase, of the field at P1 (shown in Figure 6.12).

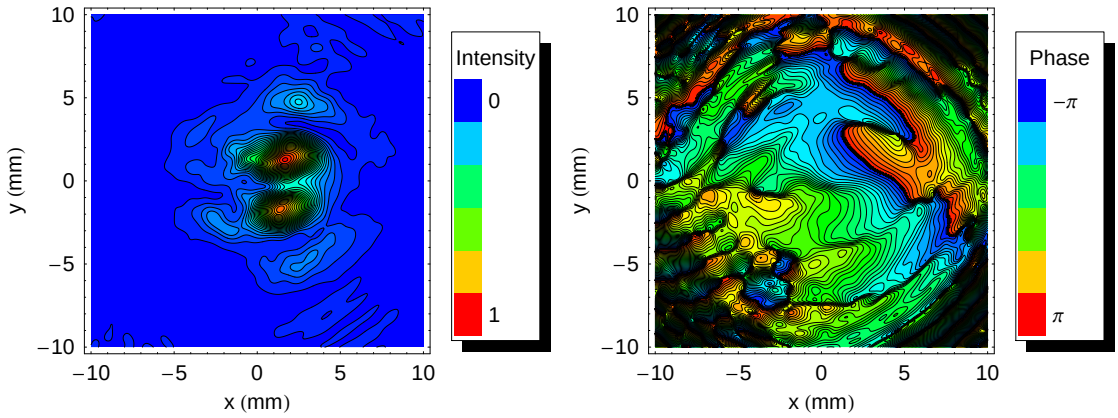


Figure 6.27: *Left* 1190GHz GRASP simulation cross-polar intensity and *right* the phase, of the field at P1 (shown in Figure 6.12).

The lens antenna field, simulated using PILRAP - which has a waist of 0.54mm located 14.4mm in front of the lens antenna reference plane - was used as the input field for GRASP. Figures 6.26 and 6.27 show the field at P1 (location shown in

Parameter	Geometry					
	1	2	3	4	5	6
Dielectric constant	11.68	11.68	11.68	11.68	11.68	11.68
Diameter (mm)	5.00	5.00	5.00	5.00	5.00	5.00
Extension length (mm)	0.73	0.73	0.77	0.77	0.76	0.76
$l/4\lambda$ layer	No	Yes	No	Yes	No	Yes
Ellipticity	1.0124	1.0124	1.0148	1.0148	1.0192	1.0192
Design frequency (GHz)	1560	1560	1560	1560	1560	1560
Operating frequency (GHz)	1619.4	1619.4	1619.4	1619.4	1619.4	1619.4
Gaussicity	94.34	95.15	94.17	95.24	93.54	94.65
Waist radius (mm)	0.40	0.39	0.42	0.41	0.57	0.56
Phase Centre Location	10.80	10.80	11.50	11.20	16.10	15.90

Table 6.5: Trial geometries used to test the effect of deviations from the nominal design on the far-field pattern.

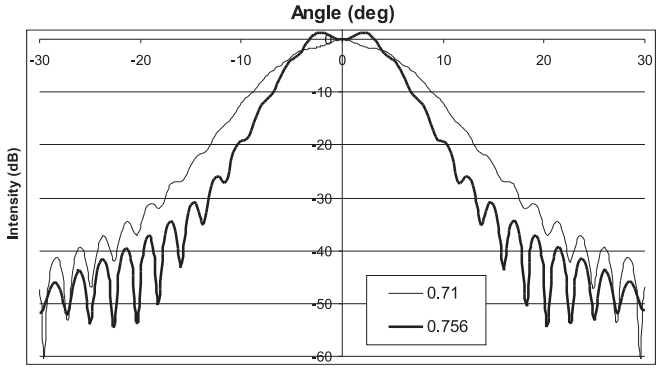


Figure 6.28: Log Co-polar intensity patterns (0° cut) for lens antennas with extensions lengths of 0.756mm (nominal) and 0.710mm.

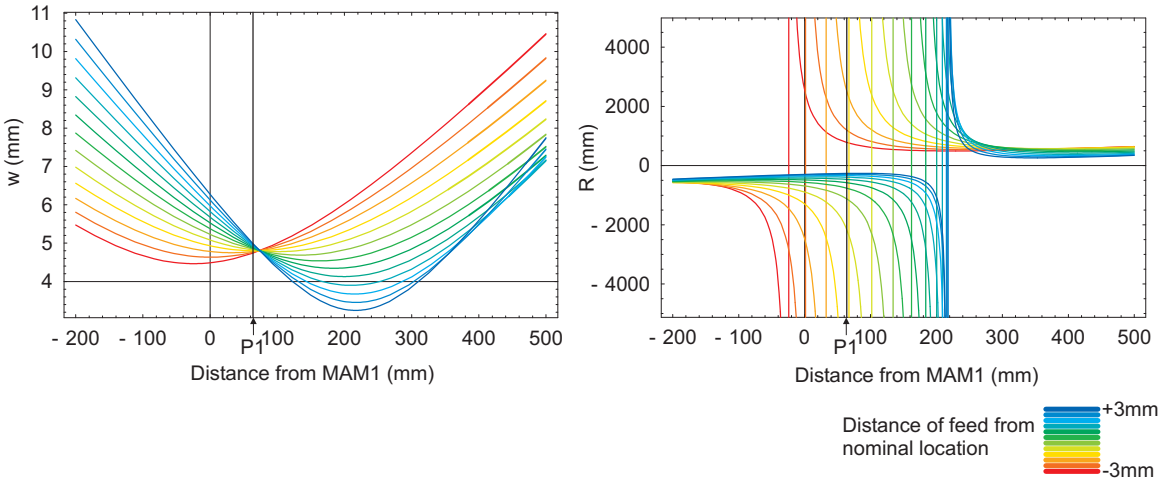


Figure 6.29: *Left* the beam radius in the region close to MAM1 and *right* the radius of curvature of the field, for waist locations in the region ±3mm from the nominal. The waist is located where the radius of curvature changes from $-\infty$ to $+\infty$. The nominal location of the waist is at P1 (shown).

Figure [4]) predicted by GRASP at 1190GHz (the same frequency that the measurements were carried out at). A fundamental Gaussian beam mode fitting procedure applied to the predicted far-field (described in Section 6.5) was used to verify that this simulated field did indeed have a waist at this location. However, there the measured and simulated fields did not match. A number of possible causes of the mismatch between the simulated and measured fields were investigated. The effect of deviations from the nominal shape of the lens was considered as one possible cause. There was uncertainty as to whether the correct geometry lens had been supplied by the manufacturers and if the lens had a $\lambda/4$ matching layer. Table 6.5 shows some of the different geometries simulated. Small deviations in the geometry of the lens resulted in significant deviations in the predicted far-field pattern, and consequently in the waist location. Figure 6.28 shows an example of a lens antenna where the where the extension length of the lens has been changed from the nominal length of 0.756mm to 0.710mm; the far-field intensity pattern from the shorter lens deviates greatly from that of the nominal design.

The effect of this deviation in the waist location of the lens antenna beam on the waist location of the field after MAM1 (nominally ay P1) was investigated. Due to the long computation time of the GRASP simulations, it would take an excessively long time to run multiple simulations for multiple waist locations. Instead, a fundamental Gaussian beam is propagated to the region near P1 using ABCD matrices and GBMA. The radius of the field for a number of waist locations in the region $\pm$3mm from the nominal position is plotted in Figure 6.29. In the lower frequency channels, where feed-horns are used instead of lens antennas, the size of the waist is approximately 1.5mm, and therefore, the location of the waist at P1 was largely insensitive to small deviations in the antenna beam waist location. However, as shown in Figure 6.29, a small deviation in the waist location of lens antenna produces a large deviation in the location of the waist after MAM1, due to the smaller waist produced by the lens antenna.

The lens antenna fields were also found to be very sensitive to the location of the double slot antenna which couples to the radiation at the back of the lens. It was found that the double slot antennas were misaligned, and a new lens antenna with improved tolerances was manufactured. The field produced by this antenna gave much better agreement with the field predicted by PILRAP.

6.5 The Phase Centre and Waist Location

The location of the phase centre of feed-horns and lens antennas is investigated in this section. The phase centre is an inexact concept that can be loosely defined as the point from which radiation is considered to emanate[49]. When designing

or analysing a quasioptical system, the beam from a feed-horn is often represented as a spherical wave. Real horns, however, produce only approximately spherical phase-fronts; therefore, there is no single point from which spherical waves emanate, and here it is the location that best approximates this definition of the phase centre that is sought.

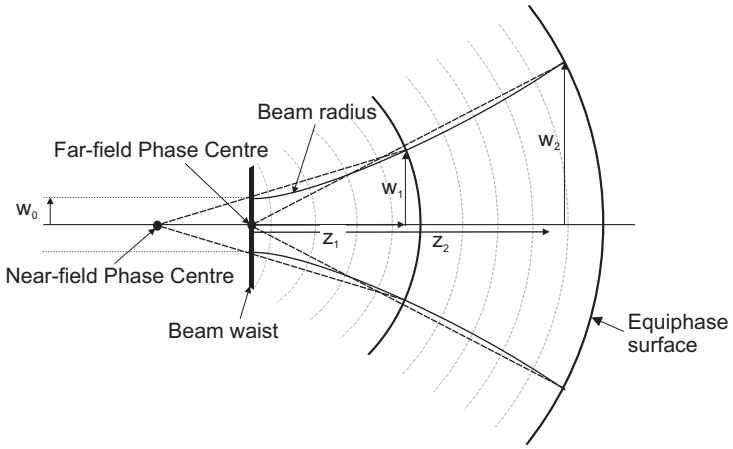


Figure 6.30: The location of the waist and phase centre for a field propagated a large distance to the far-field, and a small distance to the near-field. There is a difference between the locations of the phase centre from the two far-fields.

The phase centre of a beam is generally defined as the centre of curvature of the far-field equiphase surface. However, it is often the near-field that is of interest[70], especially in quasioptical systems, as measurements are often made at this location. In Chapter 1, it was shown that the equiphase surface of a Gaussian beam in the far-field is spherical, with radius R, where R is the distance to the beam waist location. Thus the phase centre and the beam waist are co-located. This is true only in the far-field; in the near-field, the phase centre appears to occur further behind the waist location, as shown in Figure 6.30. In this chapter, I consider the more general definition of the phase centre when the field of interest is not necessarily in the far-field.

A further complication arises if the beam is multimoded; even though these beam modes have identical equiphase radii of curvature, the relative phases of the modes change as the beam propagates. The overall beam will therefore not have the spherical phase-front of the individual modes, nor will it necessarily have the same phase centre.

The phase centre may be at a real or virtual location. For example, the waist of a field incident on a mirror or lens is located at a real point in space. The phase centre from a horn, however, is a virtual one, some distance in the "half space" behind its aperture. In reality, radiation propagates as wave guide modes in the horn. Chapter 5 describes how the optics of the ALMA telescope are designed such that the waist positions of the mirror and sky field are matched, in order to optimise

coupling between them. In this case, both the horn and sky waist positions are real. To maximise the coupling between the field from the sky and a horn, the *real* phase centre location of the sky field and the *virtual* horn field phase centre are matched, in the same way as real waist positions might be. In practice it often does not matter if the phase centre location is a virtual or real one.

Knowledge of the location of the phase centre can be used to reduce the number of GBMs required to reconstruct a field, $E(x, y)$, that is propagating in an optical system. The mode coefficients, A_{nm}, of the constituent modes are given by

$$A_{nm} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(x, y) \psi^*_{nm}(w_0, x, y, z) dx dy, \tag{6.6}$$

where $\psi_{nm}(w_0, x, y, z)$ are, for example, 2D Gaussian-Hermite modes with indices m and n, which have been propagated a distance z from the waist position. The radius of curvature of the modes at a propagation distance, z, is given by

$$R(z) = z \left(1 + \left(\frac{\pi w_0^2}{\lambda z} \right)^2 \right). \tag{6.7}$$

A field with a large radius of curvature requires a large number of modes to describe it, if those modes have a flat phase-front. The accuracy of the analysis is not affected by the number of modes used to propagate the field, provided that a sufficient amount of the power in the field is represented in those modes; however, it is more computationally efficient to use fewer modes. Often a better understanding of the characteristics of the field is gained if it is represented with as few modes as possible. Knowing the phase centre location allows a better choice of mode set radius of curvature, R, (given in Equation 6.7) and a reduction in the number of modes needed to reconstruct the field.

In this section, a number of methods to determine the phase centre or waist location of simulated and measured fields are investigated. The best method to choose may depend on the information given (e.g. the aperture field of the corrugated horn in Section 6.5.1, or the far-field in the case or the HIFI band 5 lens antenna in Section 6.5.2) or whether noise is present (e.g. in the ALMA measured data in Section 6.5.4). The methods investigated, described next, involve finding the minimum FWHM intensity of a beam, flattening the far-field phase of a beam and Gaussian-beam-mode fitting to a beam.

6.5.1 Location of Minimum Beam Width Method

The Fresnel transform was used by Glesson [5] to propagate the field from a 545GHz corrugated conical horn antenna (of aperture radius 4.1mm and slant length 28mm) to the half space behind its aperture. This horn is multi-moded, allowing propagation of the HE_{11}, HE_{21}, HE_{31}, HE_{12}, HE_{02} and HE_{01} hybrid modes. The field at the aperture is given by the incoherent sum of these modes, each of which is likely to have a different phase centre. The overall phase centre is assumed to coincide with the position where the beam is narrowest, i.e. where the full width half maximum (FWHM) of the beam intensity profile is a minimum. Figure 6.31 shows the propagation of the field in the half space behind the aperture of the horn and the location of the minimum FWHM, 11mm behind the horn aperture. Where Gaussian beam modes are used to propagate the field instead (to resolve the issue of instability as $z \rightarrow 0$mm encountered with the Fresnel transform), the minimum FWHM was found 12mm behind the horn aperture. This method of phase centre determination has been used by Gleeson to position horns at the focal-plane of both the Planck and QUaD telescopes[5].

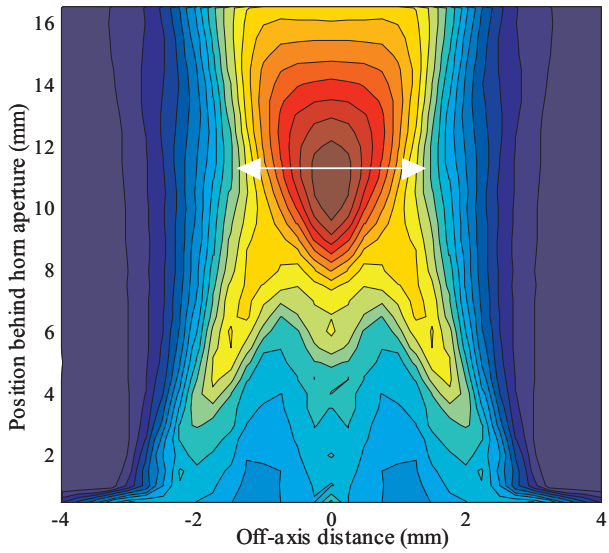


Figure 6.31: Beam intensity pattern in the vicinity of the horn's virtual waist, using the direct Fresnel transformation of the waveguide modes at the aperture. The horizontal line indicates the position of the minimum FWHM. The beam pattern is normalised with respect to the maximum intensity over the range plotted[5].

6.5.2 Far-Field Phase Flattening Method

PILRAP [65] was used to model the far-field from lens antennas, as described earlier in this chapter. PILRAP produces the far-field of the specified lens antenna, but the field at the *phase centre* was required as input into the physical optics software GRASP, in order to model the propagation of the field through the optics of band

5, 6L and 6H of the HIFI instrument on the Herschel telescope. PILRAP gives the output of the field in spherical coordinates, as shown in Figure 6.32. In this section, a $\Phi = 0°$ cut of the far-field produced by the band 5 lens antenna at 1185GHz is used as an example, but in the full analysis a two-dimensional field is used for determining the phase centre of the field for input to GRASP. The far-field cut, $E(\theta)$, given by PILRAP at $\Theta = 0$ is of the form

$$F_{\text{sph}}(\theta) = u(\theta)\exp(-i\phi(\theta)), \tag{6.8}$$

where u and ϕ are the amplitude and phase respectively. This field is propagated to the near-field region using either the FFT or GBMs. The inverse FFT of the far-field distribution gives the field located at the centre of the spherical phase surface, i.e. at the reference plane. GBMs can be used to propagate to any plane in the near-field, by decomposing the far-field into a set of GBM modes with the propagation distance, z, set to ∞; then reconstructing it at $z = 0$, where the phase slippage between the modes is 0; thereby, giving the field at the waist location.

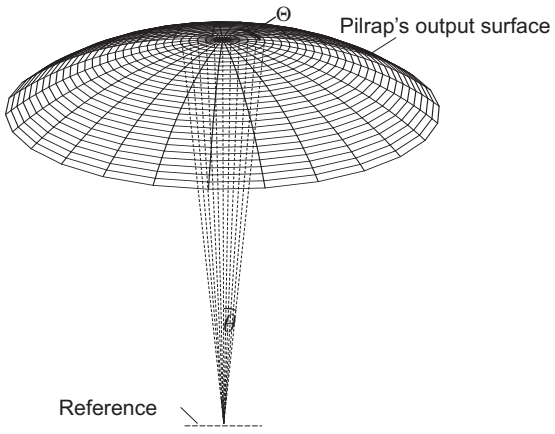


Figure 6.32: The coordinate system used by PILRAP. PILRAP produces the field on a spherical surface, centred on the reference plane.

The equiphase surface of the far-field of the lens antenna, illustrated in Figure 6.33, is approximately spherical, and, by definition, the sphere is centred on the phase centre. The far-field output produced by PILRAP, however, is on a spherical surface centred on a "reference plane", as shown in Figure 6.32. The location of the reference plane can be chosen by the user; by default it is located at the interface between the hemispherical (or elliptical) and cylindrical sections of the lens antenna. If this reference plane were located at the phase centre, then the equiphase surface and PILRAP's output surface coincide and there is no spherical component to the output phase; however, this is generally not the case. The following procedure to obtain the location of the phase centre assumes that if the phase were referenced to the true phase centre, the far-field phase distribution would be flat. The spherical phase component of the output phase is removed so that when the field is propagated

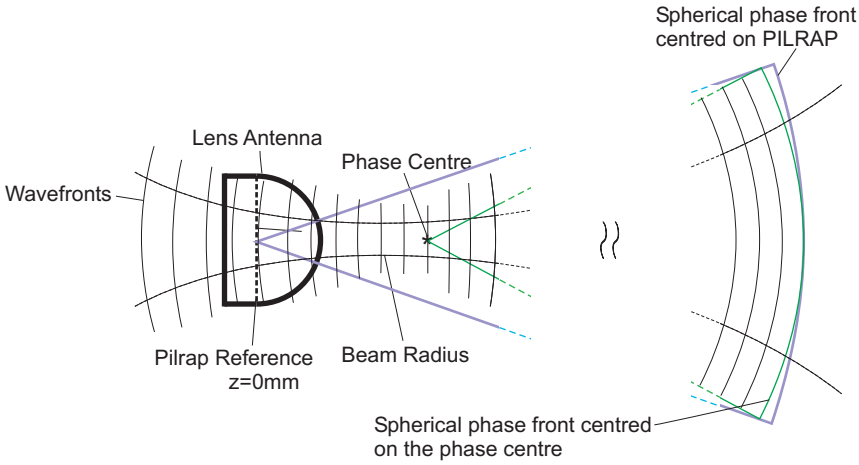


Figure 6.33: The lens antenna used on HIFI produces a field with a real waist located in front of the lens. PILRAP outputs the field is spherical coordinates where the sphere of the coordinate system are centred at the reference plane, $z = 0$mm. If the reference plane was centred on the phase centre of the field, the surface of the put field and the surface of the wavefront would match; however, as the phase centre is located at the reference plane, due to the mismatch of the two surfaces, there is an additional spherical on the field output by PILRAP.

to the near-field, using the Fourier transform or GBMs, it is located at the phase centre. The following steps are taken, in order to produce the lens antenna field located at the phase centre:

1. Use PILRAP to produce a simulated far-field from the lens antenna, with the phase reference located at the reference position.

2. Subtract the best fit spherical phase component from this field, in order to produce a field with no spherical component.

3. Inverse Fourier transform the field, to produce a field at the phase centre.

The field with a spherical phase component subtracted from it is given by

$$E(\theta, \alpha) = u(\theta) \exp[-i(\phi(\theta) - \alpha[\phi(\theta)]^2)]. \tag{6.9}$$

The term $-\alpha[\phi(\theta)]^2$ subtracts a parabolic approximation to a spherical phase from the phase of the original data. α is determined by minimising

$$\text{flatness}(\alpha) = \text{Var}(\phi(\theta) - \alpha[\phi(\theta)]^2), \tag{6.10}$$

where Var() is the variance, calculated using the Newton method. Figure 6.34 shows the phase of the lens antenna in the far-field with the best spherical component removed. The distance between the reference plane and the phase centre is given by

$$z_0 = \frac{2\alpha}{k}, \tag{6.11}$$

where k is the wave number.

Figure 6.35 *left* shows a plot of z_0 as a function of the range of off-axis angles used in Equation 6.10 (over the range of off-axis angles $4.5° \rightarrow 20°$), where the phase centre is calculated to be between $z_0 = 10.2$mm and $z_0 = 19.3$mm. The amplitude of the field is small far off-axis and will have little influence on the location of the phase centre. The upper and lower limits of the range of the phase distributions included in order to determine z_0 are chosen somewhat arbitrarily; the values of z_0 calculated using off-axis angles in the range $4.5° \rightarrow 20°$ are shown in Figures 6.35 *left*.

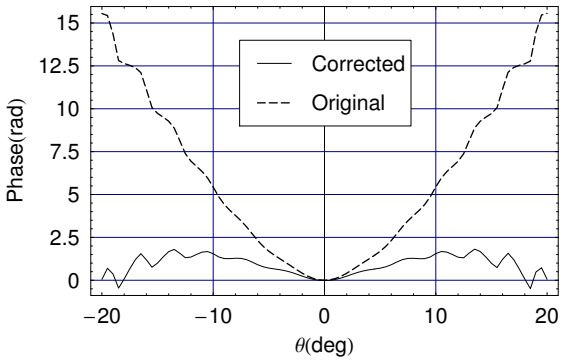


Figure 6.34: Cut at $\Phi = 0°$ of the simulated far-field of the band 5 lens antenna. The original phase output produced by PILRAP and the same phase with the spherical component removed are both shown across the $\Phi = 0°$ cut. The range of off-axis values $0° \rightarrow 20°$ were included when minimising the flatness(α), given by Equation 6.10.

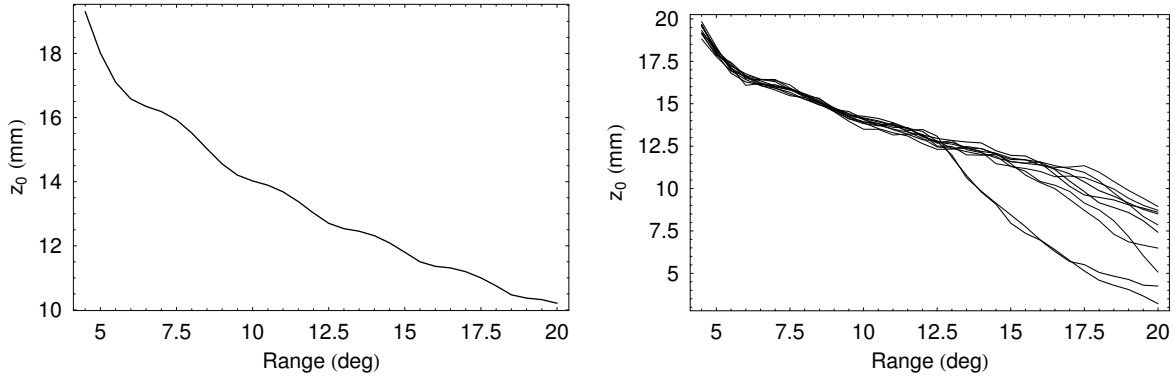


Figure 6.35: z_0 as a function of the range of angles used in Equation 6.10 (with the origin at 0°) for the phase flattening method using Equation 6.10. *Left* without noise. *Right* ten trials with added random noise with a standard deviation of $1 \times 10^{-2} \times$peak amplitude.

Noise was then added to the simulated complex field before it was propagated to the region of the phase centre, to simulate a real measured field. The noise is complex and has a Gaussian amplitude distribution with a standard deviation μ. Figure 6.35 *right* shows the values of z_0 returned by the optimisation for a range of off-axis distances. There is a large degree of variability in the values of z_0 when the

field at large off-axis distances are included. This is because the intensity of the field is lower further off-axis, and therefore the signal to noise ratio is smaller. Figures 6.36 *middle* and *bottom* show the field in the region of the phase centre where μ is 1×10^{-2} and 5×10^{-2} times peak amplitude of the field respectively. The noise significantly affects where the phase centre is predicted to be, if field values too far off-axis are used.

The FWHM method, described above, was carried out on this field for comparison with the phase flattening method. Figure 6.36 *top* shows the field propagated to the region of the phase centre with the location of the phase centre found by the phase flattening method. The FWHM method located the phase centre within the region 11.2mm to 12.5mm in front of the reference plane. It was not possible to locate an exact position of the phase centre, as the FWHM did not vary significantly over this range. The location of the phase centre found using the phase flattening method was close to the region given by the FWHM method, as shown on Figure 6.36 *top*. Figures 6.36 *middle* and *bottom* show the effect of noise on the location of the phase centre determined by this method. With $\mu = 1 \times 10^{-2}$, the region of the FWHM is close to that when no noise was added, however, the size of this region increased when the noise was increased to $\mu = 5 \times 10^{-2}$.

6.5.3 Fundamental Gaussian Beam Fitting Method

The phase flattening method can potentially give erroneous results; this is because phase at large off-axis angles - where there is little power and relatively high noise - can have as much effect on the predicted location of the phase centre as the phase closer to the axis. The FWMH method is an indirect method, as it predicts the location of phase centre based on the intensity of the field as it propagates, it does not use the phase information, and can therefore also potentially give erroneous results. A more reliable approach to determine the phase location is to use *both* the phase and amplitude of the field; the amplitude is used to weight the contribution of the field from large off-axis angles.

As the fields used in quasioptical systems tend to have high Gaussicity, we can easily determine the best fundamental Gaussian beam mode that couples to the field. In this way the determination of the location of the phase centre is an optimisation problem, in which the coupling of the field E_f to the fundamental Gaussian mode ψ_0 is maximised by varying the fundamental Gaussian's parameters z and w_0. The one-dimensional coupling of the fields in integral form is given by

$$C_0(w_0, z) = \frac{\int_{-\infty}^{\infty} E_f(x)\psi_0^*(x, w_0, z)dx}{\sqrt{\int_{-\infty}^{\infty} \left|E_f(x)\right|^2 dx}\sqrt{\int_{-\infty}^{\infty} \left|\psi_0(x, w_0, z)\right|^2 dx}}. \tag{6.12}$$

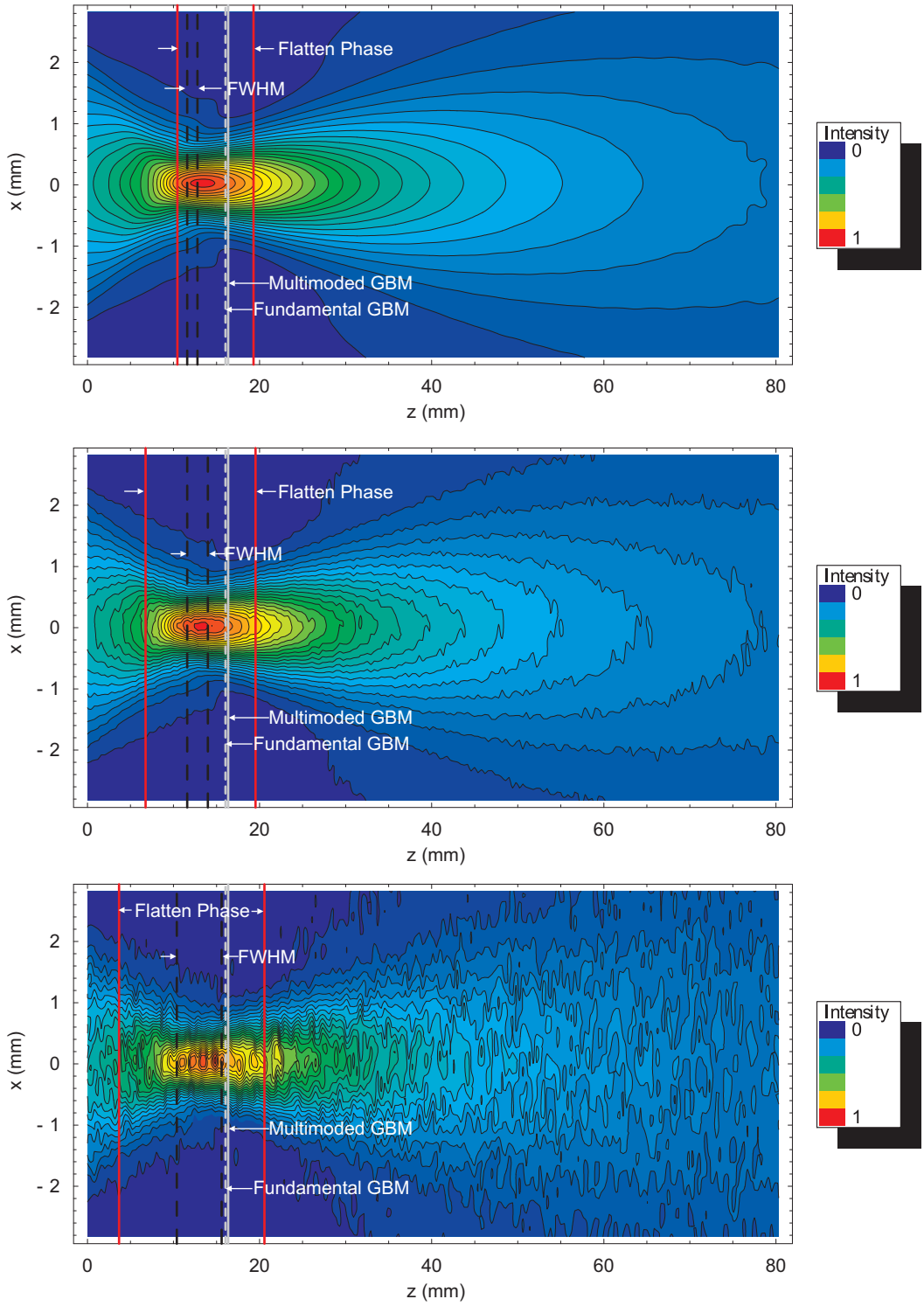


Figure 6.36: The intensity of the field in the region of the phase centre. The vertical lines show the ranges of the location of the phase centre found by the FWHM, phase flattening, fundamental GBM fitting and multimoded GBM fitting methods. *Top* No noise added in far-field. *middle* $\mu = 1 \times 10^{-2}$ times peak amplitude in the far-field. *Bottom* μ is increased to 5×10^{-2} times peak amplitude of the field.

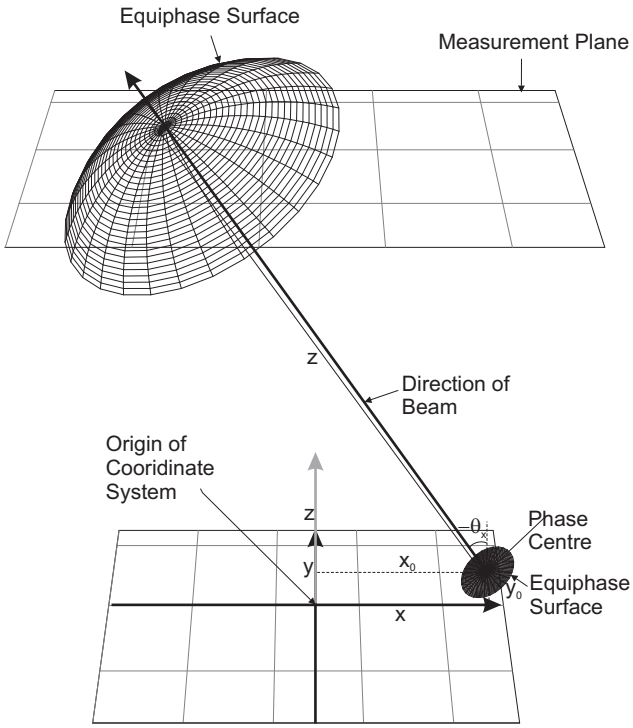


Figure 6.37: With fundamental Gaussian beam fitting to measurement data, the beam position and direction is determined relative to the origin of the coordinate system.

Finding only the two parameters w_0 and z that give the maximum coupling allows the fundamental Gaussian to be fit to an *on-axis symmetric* beam, which is often the case for simulated data. However, where measured field data is used the beam is likely to be off-set and/or tilted with respect to the measurement plane, as shown in Figure 6.37. Additional parameters are required to give the fundamental Gaussian mode freedom to fit to this field; these are the offset and the tilts in both the x and y directions. Thus x, y and z are replaced by x', y' and z', to give

$$x' = (x + x_0)\cos(\theta_x) \tag{6.13}$$

$$y' = (y + y_0)\cos(\theta_y) \tag{6.14}$$

$$z' = z + (x + x_0)sin(\theta_x) + (y + y_0)\sin(\theta_y), \tag{6.15}$$

where x_0 and y_0 are the offsets in the x and y directions, and θ_x and θ_y are the rotation about the x and y axes respectively. The optimisation now involves maximising the six parameters that affect the coupling, $C_0(w_0, x_0, y_0, z, \theta_x, \theta_y)$. These parameters are very useful for understanding the essential characteristics of the beam; the distance to the waist location is given by z and the coupling gives a measure of how Gaussian-like the field is.

The fundamental Gaussian beam mode fitting method, described above, in effect weights the phase with a Gaussian amplitude, which is the correct weighting to fit to

a *Gaussian* field. For a non-Gaussian beam *with an essentially spherical down-beam phase*, the maximum coupling to the fundamental Gaussian mode is achieved when the phases of the fields are matched, and so the correct value of z, the distance to the waist location, is found (this is shown in Chapter 3, where the coupling of two field is maximised when their phases are matched). If the phase of the field is essentially spherical away from the waist, and since *each* higher order mode has the *same* radius of curvature for a given value of z, the value of z found by fundamental mode fitting is therefore also the best choice for calculating the higher order mode coefficients, as follows

$$A_{m,n} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_f(x,y)\psi^*_{m,n}(w_0,x',y',z')dxdy, \tag{6.16}$$

where $\psi_{m,n}$ are the Gaussian-Hermite modes, and $A_{m,n}$ are mode coefficients of index m,n. In this way the phase of each Gaussian mode has the same radius of curvature as the field, thus reducing the number of modes required to describe the field to a given accuracy.

The method of fundamental GBM fitting is applied to the cut of the lens antenna field simulated by PILRAP as discussed above. The mode fitting method is carried out across a plane, so the PILRAP field ($F_{\text{sph}}(\theta)$), which is given in spherical coordinates, is mapped onto a Cartesian coordinate frame, by first interpolating (using a polynomial) the field, in order to resample it at regular intervals. The field is then given in terms of off-axis distance on a plane as follows,

$$F_{\text{cart}}(x) = F_{\text{sph}}\left[\tan^{-1}\left(\frac{x}{z}\right)\right]\exp\left(-\frac{i\pi x^2}{\lambda R}\right), \tag{6.17}$$

where x is the off-axis distance and z is the propagation distance. The term $\exp\left(-\frac{i\pi x^2}{\lambda R}\right)$ imposes an additional spherical phase on the field to account for the difference between the phase at a plane and a spherical surface. The beam fitting in this lens antenna case was done at $z = 1000$mm, in the far-field, and therefore $R \approx z$. Only the parameters z and w_0 are optimised, as it is a symmetric field. The best fit Gaussian had a propagation a distance, z, of 1016.0mm, giving $z_0 = 16.0$mm. As was done for the FWHM and phase flattening methods, noise with a standard deviation of $\mu = 1\times10^{-2}$ and 5×10^{-2} was added to the field before the phase beam fitting was done. As shown in Figures 6.36 *middle* and *bottom*, the noise was found to have no affect on the location found for z_0. This is because the noise had little correlation with the best fit Gaussian mode.

6.5.4 Multi-Moded Fitting Method

In this section, a multi-moded beam fitting approach is considered. The analysis is carried out for a top-hat field with a flat phase-front propagated a distance to the near-field region. The method is then applied to the PILRAP field, before applying it to a measured field from the ALMA focal-plane.

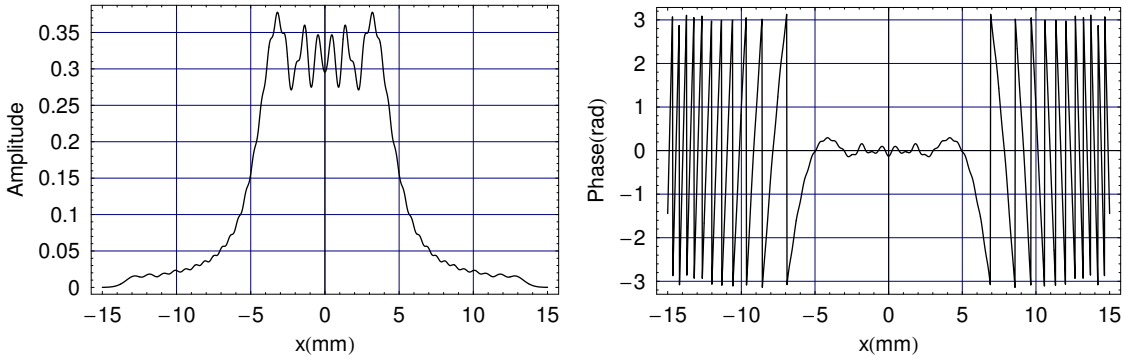


Figure 6.38: The amplitude and phase of a top-hat field of radius 5mm and with a flat phase-front propagated 10mm from the waist location.

For the fundamental Gaussian beam mode fitting method, described above, the field to which the fitting is done is assumed to have a spherical phase-front. This is not the case for non-Gaussian beams. For a beam containing higher order modes, as the beam propagates away from the waist location, the modes lose equality of on-axis phase; consequently, the beam acquires a *non*-spherical phase-front at down-beam distances, in spite of the fact that the individual beam-modes' own phase-fronts remain spherical[71, 72, 73]. This effect is shown in Figure 6.38, where a top-hat field with a wavelength of 0.45mm and a radius of 5mm, has been propagated to a plane $z = 10$mm away from the waist, producing a non-spherical phase-front. In order to account for this deviation from a spherical phase-front, a multimoded approach is taken; and instead of calculating the power coupling to just the fundamental mode, the total maximum power coupling of higher order modes is determined as a function of the parameters w_0 and z. Initially, the problem is considered for a simulated symmetric one-dimensional field, $E_f(x)$, where x', y', θ_x and θ_y (given in Equations 6.13, 6.14 and 6.15) are zero.

The total power of the reconstructed field is given by

$$P(w_0, z) = \sum_{n=0}^{n_{max}} \left| A_n(w_0, z) \right|^2, \tag{6.18}$$

where n_{max} is the order of the highest order mode used and

$$A_n(w_0, z) = \int_{-\infty}^{\infty} E_f(x) \psi_n^*(w_0, z) dx. \tag{6.19}$$

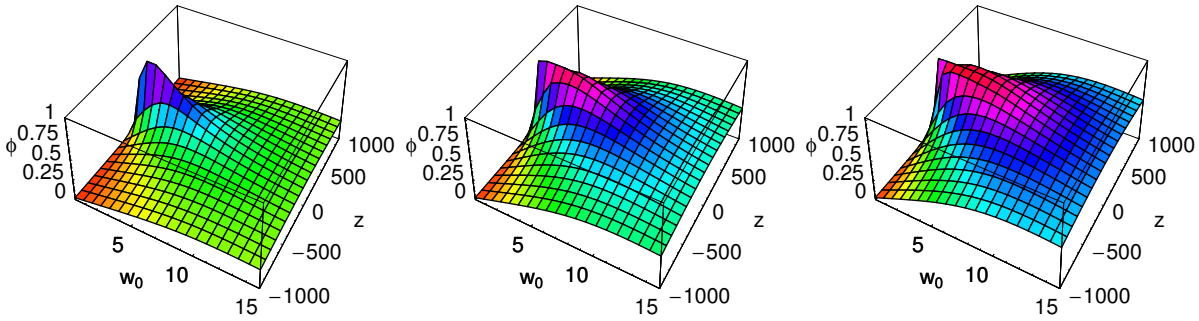


Figure 6.39: The power coupling to a top-hat field propagated a distance from the waist location with w_0 and z given by Equation 6.18 for *left* the fundamental, *middle* 0-4 modes and *right* 0-8 modes.

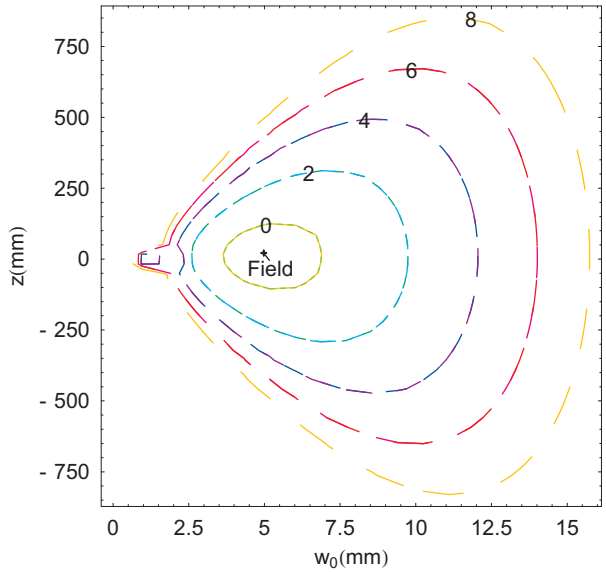


Figure 6.40: The 0.95 contour of the power coupling to a top-hat field with a flat phase-front propagated a distance $z = 10$mm from the waist location, using Gaussian-Hermite modes of waist radius $w_0 = 5$mm ($z = 10$, $w_0 = 5$ is marked "field"). The contours are labelled with the maximum mode number used (only even modes contribute). $z = 10$, $w_0 = 5$ is expected to have the highest coupling, but it becomes increasingly more difficult to discriminate this point as the number of modes used in Equation 6.18 is increased.

For a symmetric field, the odd modes can be neglected, as they do not couple with the field. As z deviates from the distance to the waist location, the total power contained in the modes, $P(w_0, z)$, tends to decrease, as the spherical component of the field deviates from the spherical component of the modes, $n > n_{max}$ modes would be required to accurately reconstruct the field. Maximum power coupling is achieved when z is equal to the distance to the waist location. Plots of power coupling as a function of w_0 and z, the waist position and radius of the mode set used to reconstruct the field, are shown in Figure 6.39, using: the 0 order mode, the $0 \rightarrow 4$ order modes and $0 \rightarrow 8$ order modes. As the number of modes is increased, the total power contained in them increases; however, the region in parameter space of maximum $P(w_0, z)$ also becomes less well defined, making it more difficult to distinguish the exact location of maximum coupling, and therefore the waist location. This is also shown in Figure 6.40, where the contour for 0.95 power coupling encircles an increasingly large region of the parameter space as the number of modes is increased. This shows that, for the purpose of reconstructing the field, it becomes less critical to choose z such that it is exactly equal to the waist location, and the choice of w_0 becomes more arbitrary. In the limit $n_{max} \rightarrow \infty$ a power coupling of 1 is always achieved, regardless of what value of w_0 and z is chosen, i.e. with enough modes any width of mode set with any radius of curvature can be used to reconstruct a field.

Considering the definition of the phase centre again, it is the point from which the spherical waves appear to originate if the field is at a large distance from the phase centre, or for Gaussian quasioptical beams it is the location where the phase-front is flat. The phase distributions of fields from real feeds, however, tend not to be flat, even at the "waist location" - but the algorithm that searches for the phase centre should find the closest approximation to it. Assuming that the field is flat at the waist location, when the parameter z in Equation 6.19 is set to the distance to the waist location, the mode coefficients with this mode set will be purely real; however, if z is *not* equal to the distance to the waist location the extra phase information is encoded as *complex* mode coefficients, because the phase-front of the modes and field no longer match; this observation will now be used to determine z more accurately, by overcoming the problem of the waist location being less well defined as the number of modes used to give $P(w_0, z)$ is increased.

I alter the way $P(w_0, z)$ is defined, by considering only the *real* part of the mode coefficients, as follows

$$P_{Re}(w_0, z) = \sum_{n=0}^{n_{max}} \Re\{A_n(w_0, z)\}^2. \tag{6.20}$$

The maximum of $P_{Re}(w_0, z)$ is now sought instead of $P(w_0, z)$; the reason for redefin-

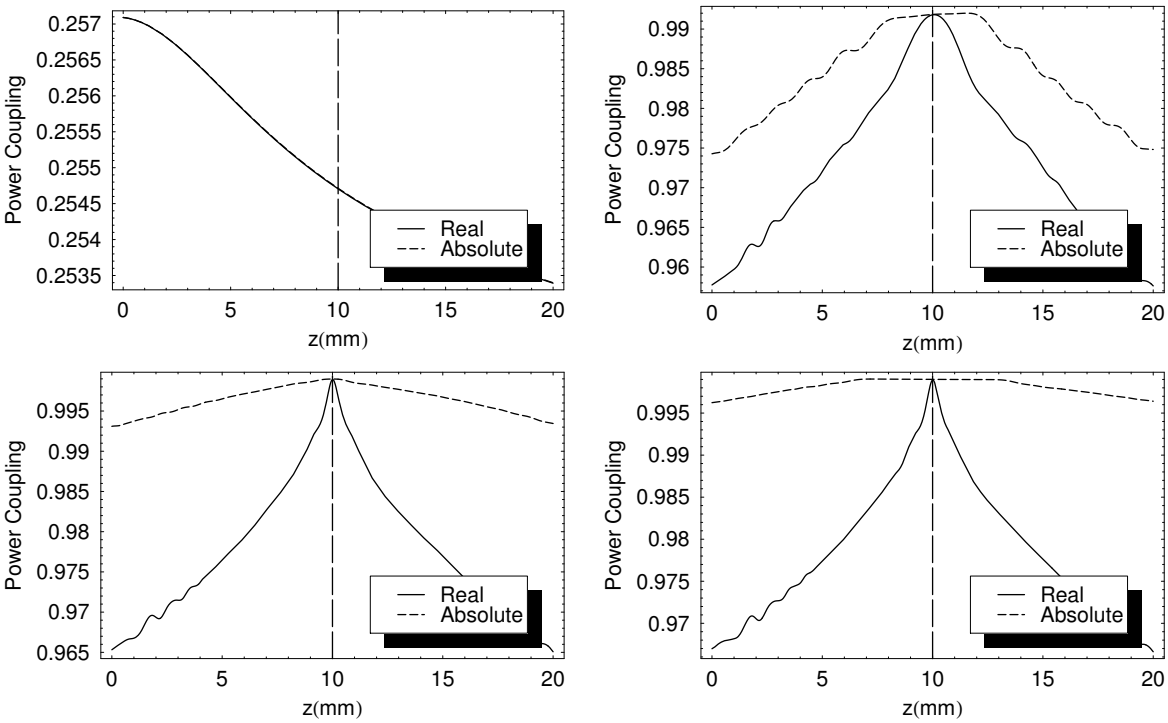


Figure 6.41: The variation of total power coupling for a finite mode set coupling to the field shown in Figure 6.38, as a function of GBM parameter z, and with $w_0 = 1$. The coupling is calculated using the absolute value as shown in Equation 6.18, and the real part only as shown in Equation 6.20. *Top left* only the fundamental mode is used; the real and absolute values are equal, *top right* 0-30 modes are used, *bottom left* 0-60 modes are used and *bottom right* 0-90 modes are used. (The actual value of z is 10mm.)

Modes	Power	w_0(mm)	z(mm)
0	0.8952	5.038	9.595
0-30	0.9947	1.278	10.027
0-60	0.9989	1.009	10.009

Table 6.6: The power coupling to the field shown in Figure 6.38 using the real part of the mode coefficient only - were w_0 and z are found by maximising $P_{Re}(w_0, z)$, using the simulated annealing method (see Section 2).

ing the power coupling in this way will become clear later. A problem is introduced by using $P_{Re}(w_0, z)$ instead of $P(w_0, z)$, because it varies sinusoidally with z, with period λ. This continuous phase variation as the modes propagate, makes the optimisation problem highly nonlinear. A constant phase may be subtracted from the modes in order to maximise the real part of the mode coefficients as follows,

$$P_{Re}(w_0, z) = \sum_{n=0}^{n_{max}} \Re\left\{ A_n(w_0, z) \exp\left(-i \arg\left[\overline{A_n(w_0, z)} \right] \right) \right\}^2, \tag{6.21}$$

where $\overline{A_n(w_0, z)}$ is the mean of the mode coefficients.

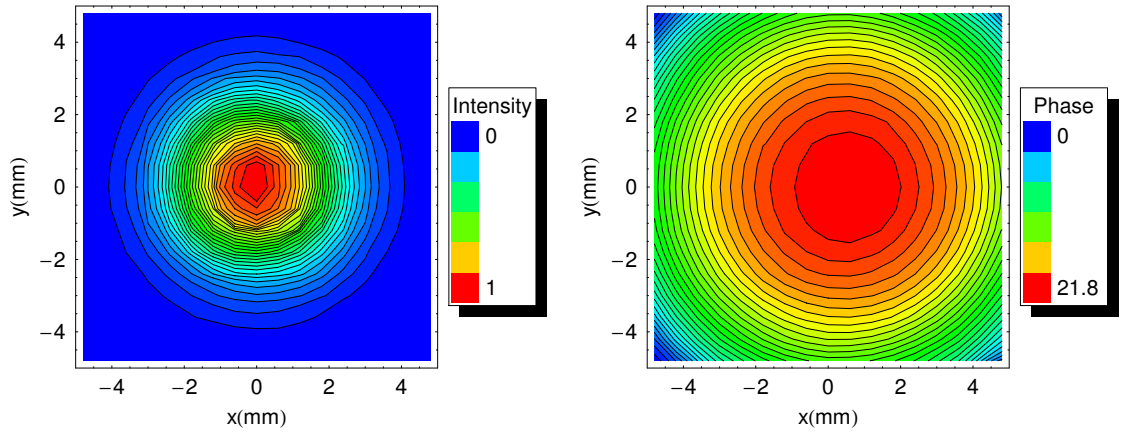


Figure 6.42: The relative intensity and phase in radians of measurements at the focal-plane of band 9 of the ALMA telescope at 654GHz.

Figure 6.41 shows $P_{Re}(w_0, z)$ (the power coupling using the real part of the mode coefficients, calculated with Equation 6.20) for the field shown in Figure 6.38 - a top-hat distribution propagated 10mm using the even modes in the range $0 \rightarrow 60$ and with the GBM parameter w_0 set to 1. The results found using Equation 6.20 are compared to those using the full amplitude coefficients (Equation 6.18). The modes sets used in Equations 6.18 and 6.20 have the same w_0, i.e. the plots show $P_{Re}(w_0 = 1, z)$ and $P(w_0 = 1, z)$ as a function of z. Figure 6.41 shows that when $P_{Re}(w_0, z)$ is used as the merit function the waist location can be defined more exactly than when $P(w_0, z)$ is used. This is because, when the phase of the modes does not match the phase of the field, the coefficients become complex, and the real

n \ m				
	$A_{0,0}$	$A_{0,1}$	$A_{0,2}$	$A_{0,3}$
	$A_{1,0}$	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$
	$A_{2,0}$	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$
	$A_{3,0}$	$A_{3,1}$	$A_{3,2}$	$A_{3,3}$

Table 6.7: The values of the modes $A_{m,n}$, in each of the four arrays shown, contained in the four squares, were optimised.

Modes	w_0(mm)	z(mm)	x_0(mm)	y_0(mm)	θ_x(deg)	θ_y(deg)
$A_{0,0}$	1.186	23.361	0.417	0.163	-0.016	0.004
$A_{0,0} \rightarrow A_{1,1}$	1.186	23.360	0.322	0.161	-0.016	0.004
$A_{0,0} \rightarrow A_{2,2}$	1.280	23.366	0.866	0.714	-0.016	0.005
$A_{0,0} \rightarrow A_{3,3}$	1.216	23.360	0.332	0.154	-0.017	0.004

Table 6.8: The GBM parameters found by multi-mode GBM fitting for the ALMA measurement data, optimising the modes in the ranges $A_{0,0} \rightarrow A_{m,n}$, as shown in Figure 6.7.

part is smaller; consequently, $P_{Re}(w_0, z)$ can only be equal to 1 when z is equal to the distance to the waist location. When only the fundamental mode is used, as shown in Figure 6.41 *top left*, the value of z that maximises $P_{Re}(w_0 = 1, z)$ is not close to the waist location, as might be expected, because the value of w_0 is suboptimal. Using more modes, the maximum of $P(w_0 = 1, z)$ becomes *less* well defined, but significantly the maximum of $P_{Re}(w_0 = 1, z)$ becomes *better* defined.

The results of the optimisation of $P_{Re}(w_0, z)$ with parameters w_0 and z (rather than just z, as was done above) are shown in Table 6.6. By using just the fundamental mode, the distance found to the waist was close to the actual 10mm distance. Increasing the number of modes allowed the exact distance to the waist location to be found.

Figure 6.36 *top* shows the location of the phase centre found by the multimoded GBM optimisation using the even modes in the range $0 \rightarrow 10$, where $z_0 = 14.3$mm. As was expected, this value is very close to that of the fundamental GBM fitting method, as the field produced by PILRAP is in the far-field. Figures 6.36 *middle* and *bottom* show the location for a field with added noise; as was the case for the fundamental GBM fitting, the noise had no effect on the location found for z_0 with this method.

As another example, the phase centre corresponding to a field measurement of band 9 ALMA, described in Chapter 5, is given here. The field at the focal-plane measured at SRON by Andrey Baryshev was found at 645GHz, and is shown in Figure 6.42. The position of the scanner relative to the mirror M3 was unknown, and was found using this fundamental GBM fitting method. The optimisation was

	Modes used in the optimisation			
Modes	$A_{0,0}$	$A_{0,0} \rightarrow A_{1,1}$	$A_{0,0} \rightarrow A_{2,2}$	$A_{0,0} \rightarrow A_{3,3}$
$A_{0,0}$	0.991 - 0.076i	0.991	0.995	0.815 + 0.001i
$A_{0,1}$		0.002	0.001i	0.369 + 0.001i
$A_{0,2}$			0.002-0.004i	0.103 - 0.003i
$A_{1,3}$				-0.020 + 0.004i
$A_{1,0}$		0.085 -0.001i	0.004- 0.001i	0.367 - 0.003i
$A_{1,1}$		0.001 -0.001i	0.001-0.002i	0.181 + 0.001i
$A_{1,2}$			-0.002-0.006i	0.032 - 0.003i
$A_{1,3}$				-0.023 - 0.004i
$A_{2,0}$			-0.006+0.001i	0.097 + 0.002i
$A_{2,1}$			0.001-0.003i	0.031 + 0.004i
$A_{2,2}$			0.044	0.026 - 0.008i
$A_{2,3}$				0.013 + 0.001i
$A_{3,0}$				-0.027 + 0.001i
$A_{3,1}$				-0.026 - 0.004i
$A_{3,2}$				0.012
$A_{3,3}$				0.015 + 0.004i
Total power	0.989	0.989	0.991	0.993

Table 6.9: The waist location is found by maximising the total power in the real part of the mode coefficients. This was done for the arrays of modes in the ranges $A_{0,0} \rightarrow A_{m,n}$ shown in Figure 6.7. The complex coefficients are calculated by performing the overlap integral shown in Equation 6.16, with the parameters w_0, z, x_0, y_0, θ_x and θ_y, found by the optimisation (given in Table 6.8).

carried out with Simulated Annealing using the real part of the mode coefficients in two dimensions. The following GBM parameters, shown in Figure 6.37, are optimised: w_0, z, the offset in the x direction x_0, the offset in the y direction y_0, the tilt in the x direction with respect to the azimuth θ_x and the tilt in the y direction with respect to the azimuth θ_y. Table 6.8 shows the parameters found using just the fundamental mode and also using the modes in the ranges $A_{0,0} \rightarrow A_{1,1}$, $A_{0,0} \rightarrow A_{2,2}$ and $A_{0,0} \rightarrow A_{3,3}$, as shown in Table 6.7. The variation in the offsets in the x and y directions, x_0 and y_0 found when higher order modes are used, occurs because the modes allow similar power coupling to be achieved when the mode set is offset laterally by small distances, as is achieved when there is no such offset. If x_0 and y_0 are required, Gaussian-Laguerre modes (described in Chapter 1) could be used in place of Gaussian-Hermite modes. These modes, of index p and m, are circularly symmetric for all integer values of $p \geq 0$ with $m = 0$. Calculating the power coupling as above would give x_0 and y_0 relative to a symmetric mode set (these symmetric modes would not couple to the asymmetric part of the field and the maximum value of $P_{Re}(w_0, z)$ would be better defined). A low number asymmetric modes (with index $|p| > 0$) could be used to describe some of the asymmetry of the field while still allowing x_0 and y_0 to be determined, but there is a trade-off between how much of the asymmetry is described and the accuracy of the determination of x_0 and y_0. The parameters z and the tilts are fairly independent of the number of modes used, and the fundamental mode gave an excellent estimation of the position of the waist location (even though this fundamental mode approximation coupled to less than 99% of the power).

6.6 Conclusion

This chapter describes the optics of HIFI. The higher frequency channels use lens antennas to couple the radiation from the telescope, rather than the waveguides used in the low frequency channels. These devices bring new challenges. Tolerencing of the lens antenna itself and the alignments of the lens antenna in the system were found to be much more critical than for waveguides, due to the small size of the feed and the small waist that the lens antenna itself produces.

The phase centre of the HIFI simulated lens antenna field was required in order to model beam propagation through the full channel. The phase centre is generally defined as the centre of curvature of the phase-front of a beam; however, real feeds generally do not produce fields with spherical phase-fronts, and only the location that best approximates this definition of a phase centre may be found. In this chapter, four methods for finding the phase centre were compared. The FWHM method uses the indirect approach of finding the location of the narrowest part of

the beam, and assumes that this is co-located with the phase centre. It does not use the phase information, and so it may be a crude estimation of the phase centre location. The far-field phase flattening method is direct, but the position of phase centre varied considerably with the off-axis range of the phase which was flattened. Both methods were found be susceptible to noise. The fundamental GBM fitting method gives more accurate results as it weights the phase with the amplitude of the field. The multimoded GBM fitting method does not make any assumptions about the field (e.g. that it has a high Gaussicity). The fundamental and multimoded GBM fitting approach were found to find a similar value for the location of the phase centre, for the examples (from HIFI and ALMA) carried out here. Both the fundamental and multimoded methods were found to be robust when noise was added to the field.

Chapter 7

Conclusion

7.1 Chapter 1

The main concerns of this thesis are the analysis and optimisation of quasioptical telescopes operating at millimetre and submillimeter wavelengths. In particular, the HIFI instrument on the Herschel Space Observatory, the ALMA interferometer and the MBI telescope were analysed. Many of the astronomical phenomena that can be observed in this wavelength regime, and discussed in Chapter 1, cannot be seen at other wavelengths. Chapter 1 outlined some of the problems concerned with observing at these wavelengths, and also described some of the components used in these telescopes. Due to the long wavelengths that such telescopes operate at, the effects of diffraction on the radiation propagating in them must be taken into account. Techniques used to model the radiation propagating in millimetre-wave telescopes were described. An efficient technique, which models the field as a sum of, orthogonal, Gaussian Beam Modes (GBMs), was described in more detail, as it was used throughout this thesis.

7.2 Chapter 2

Chapters 5 and 4 describe how power coupling between the horn and the sky field of the MBI and ALMA telescopes may be increased by altering the current design. To gain a deeper understanding as to how power coupling can be increased, beam shaping techniques - where a phase is imposed on a beam to achieve a desired far-field intensity distribution - were investigated in Chapters 2 and 3.

In Chapter 2, the standard technique of the Gerchberg-Saxton Algorithm (GSA) was used to transform a Gaussian beam with a flat phase-front incident on a transmissive Diffractive Optical Element (DOE) to a prescribed far-field target intensity distribution. The design of a suitable DOE is in effect a phase retrieval problem. The GSA is an iterative algorithm, which has been proven to give an improved so-

lution with each iteration, but it can "get stuck" and give a suboptimal near-field phase distribution (i.e. required DOE surface). Whether the final solution of the GSA is a global or local optimum depends on what phase distribution is used to initialise the algorithm. As the phase distribution which achieves a globally optimal solution could not be known before hand, multiple trials with different random phase distribution were used to initialise the GSA. Three basic test cases were used to test the algorithms; these were the transformation of an input Gaussian amplitude distribution to a top-hat, a first-order Gaussian-Hermite distribution and three equi-spaced Gaussian beams.

In addition to allowing freedom of the phase distribution, freedom of the scale of the target distribution can allow a more optimal solution to be achieved. Generally, more optimal solutions can be found at larger scales; however, the nonlinearity of the problem tends to increase as the scale of the target distribution is increased, resulting in a greater likelihood that the GSA stagnates at a sub-optimal phase distribution. For certain phase distributions, highly optimal solutions could be found for intermediate scales of far-field target distributions; however, finding the exact scale is computationally expensive, as it requires many trials of the GSA.

The global optimisation unidirectional algorithms Simulated Annealing (SA) and Differential Evolution (DE) were applied to the DOE design problem. Whereas the GSA is a local search algorithm, DE and SA are global search algorithms, as they are capable of "jumping out" of local minima, but they are not guaranteed to find a globally optimal solution for the DOE design problem. With unidirectional optimisation, the input field is combined with trial phases and transformed in the forward direction only, i.e. to the far-field. The phase values at discrete points along the beam were encoded as parameters, which were free to take any value. SA and DE generate trial parameters which encode the phase of the DOE. The field given by this phase distribution combined with the amplitude of the field incident on the DOE (i.e. a Gaussian with a flat phase-front) is then transformed to the far-field using the FFT, where a metric determines the quality of the trial solution. For both SA and DE, whether or not the trial solution is accepted as a base point with which to generate the trial solutions for next iteration is dependant on the quality of the far-field distribution of the current iteration and a random factor. The specific acceptance criteria are different for each, and whereas SA operates on a single set of parameters encoding a single phase distribution, DE operates on a population of parameters. For the test Gaussian to top-hat transform, where the scale of the top-hat was small, DE, SA and the best of multiple trials of the GSA arrived at the same solution; this solution is assumed to be a globally optimal solution (however, it has not been proven that a more optimal solution does not exist). DE out-performed SA when the scale of the target output distribution was small; however, at large scales,

the best of multiple trials of the GSA out-performed both optimisations using both DE and SA.

The application of SA and DE to beam shaping is summarized a follows

SA: Optimises a single set of parameters which encode the phase. It performs well where the optimisation is not too non-linear.

DE: Optimises a population of parameters which encode the phase. It tends to be slower than SA, but outperforms SA where the optimisation is more non-linear.

7.3 Chapter 3

In Chapter 3, DOEs were designed by optimising a field described by the sum of GBMs. Here the mode coefficients that best describe both the near and far-field target distributions were found. This method I refer to as Gaussian Beam Mode Set Optimisation (GBMSO). The shape of the DOE is given by the phase of the near-field solution. As this optimisation was non-linear, DE was used. An advantage to using an generic optimisation algorithm like DE over the GSA is that the scale of the target distribution could be optimised to find the best solution. For certain transforms, making the GBM reference plane (the distance from the waist of the modes) a free parameter allowed the GBMs to construct the near-field distribution accurately with fewer modes, and the optimisation found a highly optimal solution more reliably. For certain near- to far-field transforms, keeping the scale of the target distribution a free parameter also allowed highly optimal solutions to be found. In the case of the Gaussian to first-order Hermite distribution transform, at a particular scale of target distribution, an asymmetric phase is more optimal than a symmetric one. The GBMSO approach allows more control over the phase distribution, for example, by using only even modes, so the most optimal *symmetric* phase can be found for any given scale of target intensity distribution.

Where the scale of the target intensity distribution is small, the optimisation is more linear, and GBMSO and the GSA find very similar phase distributions. Where GBMSO and the GSA find approximately the same optimum, the solutions given by GBMSO tend be slightly less optimal than those given by the GSA, due to the limited spatial frequencies of the modes. This can be an advantage however, as the DOEs given by the GBMs are smoother, and therefore easier to manufacture. The far-field from DOEs designed using GBMSO also have the advantage over those designed using the GSA with FFTs, that the fields they produce less unwanted off-axis power, and higher order modes are more likely to be truncated by components in an optical system and be lost in the system anyway. By comparing the size of mode set used in the optimisation with the scale of the target distribution, Gaussian Beam

Mode Analysis gives an intuitive explanation as to why the quality of the transform is dependant on the scale of the target distribution.

Unidirectional optimisation using DE or SA was much more likely to find sub optimal solutions than GBMSO. This is because with unidirectional optimisation using the FFT to transform to the far-field, trial phase distributions can have arbitrarily high spatial frequencies, and therefore much of the power in the far-field may be directed away from the target. As this off-axis power does not contribute to the field in the region of the target distribution, it does not affect the figure of merit used to determine the quality of the field, and therefore it is more difficult to incrementally improve the solution. This results in the optimisation being more likely to get stuck at a local extremum. GBMSO is much more efficient in this respect, as the power of the trial far-field distributions are constrained by the extent of the modes. Also, GBMSO has more freedom to make an improving move because it can control both the amplitude and phase of the trial field during the optimisation, whereas unidirectional optimisation controls only the phase.

The performance of the algorithms discussed here depend on the target distributions; however, the following general conclusions are drawn:

GSA: Has the advantage that it is a simple algorithm to implement, and quickly converges to a solution for a particular target distribution; however, it is highly dependent on starting phase for certain target distributions.

Unidirectional: Has the advantage that arbitrary constraints or freedoms may be placed on solutions (e.g. the scale of the target amplitude distribution may be optimised); however, large computation power may be required, as it is liable to get stuck in a suboptimal solution.

GBMSO: Like unidirectional optimisation, it has the advantage that arbitrary constraints or freedoms may be placed on solutions. It generally gives good solutions with smooth phase distributions, while being computationally efficient. It is also more intuitive that the other algorithms. However, poor solutions are achieved where the optimal solution cannot be described by a low number of modes. In cases where GBMSO cannot find a good solution but the scale is optimised, this scale can be used as the target scale for the GSA.

DOEs were designed to convert the fundamental Gaussian-Hermite mode to a higher order mode. It was found that where any phase may be imposed on the input mode, maximum coupling of the input mode to the output mode is achieved when the phases of the modes are matched. This involves imposing a π phase shift where the higher order mode is negative. This approach maximises coupling of the mode, but other, unwanted, modes couple to the output beam.

The GBMSO method was expanded to design DOEs which control the near-field, demonstrating that it is an efficient and flexible method. This was done by optimising to achieve the desired target distribution at multiple planes. In order to allow more freedom to achieve a good design, the optimisation should be as unconstrained as possible, for example by allowing the radius of the input field to be optimised. This optimisation approach allows DOE's to be designed that would be difficult to achieve using analytical techniques.

7.4 Chapter 4

Chapter 4 concerns a novel bolometric interferometer called MBI. A quasioptical beam combiner (Fizeau combiner) is used to interfere radiation from the interferometric horns and the interference pattern generated at the combiner focal-plane is coupled to an hexagonal array of bolometers with conical corrugated horns. The beam combiner has a Cassegrain telescope design. The beam from the sky incident on the combiner had a small waist, and the resulting diffraction and spreading of the beam resulted in significant truncation at the combiner secondary. Feed horn lenses were designed and subsequently manufactured for the interferometer horns to reduce this effect. A second set of lenses were designed to improve the coupling of the interferometer field to the detector horns. Where a phase is imposed at a single surface, maximum coupling is achieved when the phases of the field from the telescope and the detector horns are matched. In this case, HDPE lenses, placed at each detector horn aperture, were used to impose a phase on the field. The phases of the fields can be matched by flattening the phase from the horn and sky separately, by using two surfaces, or by integrating them into one surface. The result of the optimisations are summarised in Figure 4.43. The improvement in coupling depended on the location of the horn on the focal-plane; typically an improvement of around 33% power coupling was achieved, for lenses which flatten only the phase from the detector horns. For many of the horns, flattening the sky field, in addition to the field from the detector horns, did not significantly improve the power coupling, while for some of the horns there was a large improvement. The detector lenses were made a fitted to the telescopes by the MBI collaboration and the instrument is now collecting data. The first fringes were measured in the spring of 2008. A number of alternative lens configurations were considered. Hexagonal lenses offset a distance from the horn aperture can be used to couple significantly more radiation than lenses placed on the aperture. The coupling may be improved while avoiding the use of lenses by tilting the horns. For simplicity, individual spherical lenses were used to flatten the phase of the detector horns at their aperture, but not the field due to the interferometer. These other design options could be considered for future

generations of MBI, which will have detectors over a larger area of the focal-plane.

7.5 Chapter 5

Chapter 5 concerns the ALMA interferometer. The interferometer will be comprised of 50 telescopes, operating in the frequency range 30GHz to 1THz, in 10 frequency bands. Alterations to the reference design of band 9 in order to improve coupling of the detector horns to the sky field were considered. The coupling was increased in the software model of the telescope by using the same method as Chapter 4: the phases of the horn and sky fields were matched. The intention was to alter the shape of one of the two mirrors in the signal path between the detector horn and the secondary mirror to improve coupling. An initial design consisted of a transmissive DOE to match the phases of the horn and sky fields. The improvement in coupling was found to be dependent on the plane at which the phase matching was done. The variation in the improvement in power coupling as a function of the location of the plane at which the fields are matched is because the field amplitude profile of a field changes as it propagates. The optimum coupling is achieved at the plane where the amplitude profiles are most similar.

In band 9, there are two mirrors in the signal path between the secondary mirror and the feed-horn. By using both mirrors to impose a phase at two planes, there is more freedom to control the field, and the power coupling can be increased above that achieved where only one surface imposes a phase. One mirror imposes a phase on the field from the detector horn, and the phases of the horn and sky fields are matched at the second. Alternatively, the roles of the mirrors can be swapped. GBMSO is used to perform beam shaping in order to match both the amplitude and phase of the beams. For this application, GBMSO has a number of advantages over the GSA: it can be used to do beam shaping between one mirror and another where the second mirror is not necessarily in the far-field of the first; a mode set with a radius of curvature which matches that imposed by the mirror can be used so that the beam shaping gives a phase profile close to that of the mirrors; and by using a small number of modes can be used to give a smooth mirror profile.

7.6 Chapter 6

Chapter 6 concerns the HIFI instrument on the Herschel Space Observatory. HIFI is a very high resolution heterodyne spectrometer, operating at $480 - 1920$GHz. It is a complicated instrument, consisting of many mirrors, which keep the signal field quasi collimated, and is split into 7 bands. In the 4 lowest frequency bands, horns are used to couple the radiation to the detector, whereas in the three highest frequency bands

lens antennas - which consist of a dielectric lens mounted onto a double slot antenna - are used. These lens antennas are less well understood than horns. The field in band 5 which used lens antenna was modelled using the PO software GRASP. The measured field at the mixer sub assembly did not match that simulated. A simulation of the lens antenna field was used to model possible manufacturing errors. The problem was eventually found to be due to a misalignment of the double slot antenna at the back of the lens antenna. A HDPE lens antenna, fabricated at NUI Maynooth, was modelled using ray tracing. As was the case with the model HIFI lens antenna, the model did not match the measured field exactly, and like the HIFI lens antenna, this was found to be due to manufacturing tolerances.

The software used to simulate the lens antenna gave as its output the lens antenna located far-field; however, the field at the waist location was required for input to the GRASP model. The waist location was found by several methods: determining the location of the FWHM, subtracting the spherical component of the far-field phase, fundamental GBM fitting and multimoded GBM fitting. The methods were compared and the effects of noise on the waist location was investigated; the GBM fitting methods were found to be the most robust when noise was added.

In summary, this thesis investigated algorithms for designing beam shaping elements in quasioptical astronomical telescopes. Practical applications of the algorithms to three current quasioptical telescopes were described in detail.

GBMSO has been shown to achieve highly optimal results and could be further developed to optimise more complex systems, since standard components such as lenses and mirrors can be modelled very efficiently using Gaussian Beam Modes. The location and focal length of lenses and mirrors, modelled using ABCD matrices, could be optimised to achieve a desired field at the output. In addition, the effect of standing waves and aberrations may be modelled during the optimisation, using related techniques that have been developed recently[57, 58].

Bibliography

[1] R. Bachiller et al. Science with ALMA. Publicity Document.

[2] R. Cheville et al. *Sensing with Terahertz Radiation*. Springer, 2003.

[3] G. L. Pilbratt. Herschel mission: status and observing opportunities. In *Proceedings of the dusty and molecular universe: a prelude to Herschel and ALMA*, 2005.

[4] J. W. Lamb et al. ALMA receiver optics design. Technical Report ALMA Memo 362, ESA, 2001.

[5] E. Gleeson. *Single and Multi-moded Corrugated Horn Design for Cosmic Microwave Background Experiments*. PhD thesis, NUI Maynooth, Co. Kildare, Ireland, 2004.

[6] P. Goldsmith. *Quasioptical Systems*. IEEE Press, 1997.

[7] B. F. Lane. *High-Precision Infra-Red Stellar Interfermetry*. PhD thesis, California Institute of Technology, Pasadena, California, USA, 2003.

[8] W. Bagnuolo. CHARA technical report, expanding (U,V)-plane coverage with outriggers. Technical report, Center for High Angular Resolution Astronomy, Georgia State University, Atlanta GA, 1997.

[9] M. Candotti. *Design, Analysis And Optimisation Of Quasi-Optical Systems For Sub-Mm Astronomy Instrumentation*. PhD thesis, NUI Maynooth, Co. Kildare, Ireland, 2007.

[10] P. T. Timbie et al. The Einstein Polarization Interferometer for Cosmology (EPIC) and the Millimeter-wave Bolometric Interferometer (MBI). *Elsevier Science*, 50:999–1008, 2006.

[11] H. McAlister et al. First results from the CHARA array. an interferometric and spectroscopic study of the fast rotator alpha Leonis Regulus. *The Astrophysical Journal*, 628:439–452, 2005.

[12] M. Colavita et al. The Palomar testbed interferometer. *The Astrophysical Journal*, 510:505–521, 1999.

[13] J. Goodman. *Introduction to Fourier Optics*. Roberts & Co, 2005.

[14] E. Hecht. *Optics*. Addison Wesley, 1987.

[15] M. Nixon et al. *Feature Extraction and Image Processing*. Newnes, 2002.

[16] A. Siegman. *Lasers*. University Science Books, 1986.

[17] S. Withington. Astronomical telescopes and instrumentation. Book from course at the optical, infrared and millimetre space telescope conference, Glasgow 2004, 2004.

[18] D. O'Shea et al. *Diffractive Optics: Design, Fabrication, and Test*. SPIE, 2003.

[19] J. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21:2758–2769, 1982.

[20] R. Gerchberg et al. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237246, 1972.

[21] R. J. Mahon. Digital holography at millimetre wavelengths. *Optics Communications*, Volume 260:469–473, 2006.

[22] V. Soifer et al. *Methods for Computer Design of Diffractive Optical Elements*. Wiley, 2002.

[23] H. Kim et al. Iterative fourier transform algorithm with regularization for the optimal design of diffractive optical elements. *Optical Society of America*, 21:2353–2365, 2004.

[24] C. Salema et al. *Solid Dielectric Horn Antennas*. Artech House, 1998.

[25] F. Wyrowski et al. *Diffractive Optics for Industrial and Commercial Applications*. Vch Verlagsgesellschaft Mbh, 1998.

[26] R. Bracewell. *The Fourier Transform and its applications*. McGraw Hill, 2000.

[27] H. Dammann et al. Coherent optical generation and inspection of two-dimensional periodic structures. *Opt. Acta 24*, 24:505515, 1977.

[28] E. Aarts. *Simulated Annealing: Theory and Applications*. Springer, 1987.

[29] K. Price et al. *Differential Evolution*. Springer, 2005.

[30] J. Johnson et al. Genetic algorithms in engineering electromagnetics. *IEEE Antennas and Propagation Magazine*, 39:7–21, 1997.

[31] C. Rocha-Alicanoa. Differential evolution algorithm applied to sidelobe level reduction on a planar array. *International Journal Electronics Communications*, 61:286–290, 2006.

[32] E. Johnson et al. Microgenetic-algorithm optimization methods applied to dielectric gratings. *Optical Society of America*, 12:1152–1160, 1995.

[33] J. Jiang et al. A rigorous unidirectional method for designing finite aperture diffractive optical elements. *Optics Express*, 7:237–242, 2000.

[34] J. Jiang. *Rigorous analysis and design of diffractive optical elements*. PhD thesis, The University of Alabama, Huntsville, USA, 2000.

[35] S. Wolfram. *The Mathematica Book*. Wolfram Media Incorporated, 2004.

[36] K. Isaak et al. Phase retrieval at millimetre and submillimetre wavelengths using a Gaussian-beam formulation. In *Fourth International Symposium on Space Terahertz Technology*, 1993.

[37] A. Cutolo et al. Characterization of the transverse modes in a laser beam: Analysis and application to a Q-switched Nd:YAG laser. *Applied Optics*, 31:2723–2733, 1992.

[38] T. Meyrath et al. A high frequency optical trap for atoms using hermite-gaussian beams. *Optical Society of America*, 8:2843–2851, 2005.

[39] D. et al McGloin. Advanced micromanipulation using bessel beams. In *Conference Lasers and Electro-Optics Europe. CLEO/Europe.*, page 710, 22-27 June 2003.

[40] J. Arlt et al. Atom guiding along non-diffracting light beams. In *Quantum Electronics and Laser Science Conference*, pages 82–83, May 1999.

[41] J. Durnin. Diffraction free bessel beams. *Journal of the Optical Society of America*, 58:651–654, 1987.

[42] L. Paterson et al. Light-induced cell separation in a tailored optical landscape. *Applied Physics Letters*, 87(12):123901, 2005.

[43] V. Garces-Chavez et al. Simultaneous micromanipulation in multiple planes using a self-reconstructing light beam. *Nature*, 419:145–147, 2002.

[44] S. Monk el al. The generation of bessel beams at millimetre-wave frequencies by use of an axicon. *Optics Communications*, 170:213–215, 1999.

[45] N. Trappe et al. The quasi-optical analysis of bessel beams in the far infrared. *Infrared Physics and Technology*, 46:233–247, 2005.

[46] J. Kim. *The Millimeter-wave Bolometric Interferometer*. PhD thesis, Brown University, Providence, Rhode Island, USA, 2006.

[47] I. S. Ohta et al. The first astronomical mm and submm observations with multi-fourier transform interferometer. *Infrared and Millimeter Waves and 13th International Conference on Terahertz Electronics*, 2:359–360, 2005.

[48] G. Tucker. MBI publicity document. Brown University.

[49] A. D. Olver et al. *Microwave Horns and Feeds*. IEEE, 1994.

[50] Zemax Development Corporation. *Optical Design Program, User's Guide*. ZEMAX, February 2008.

[51] C. T. Kelley. *Solving Nonlinear Equations with Newton's Method*. Society for Industrial Mathematics, 1987.

[52] T. Wiklind. Science with Herschel and ALMA: complementarity with ELTs. In *Proceedings of the International Astronomical Union*, volume 1. Cambridge University Press, 2005.

[53] J. Turner et al. ALMA scientific advisory committee. ASAC report, European Southern Observatory, Garching, Germany, 2005.

[54] M. Carter et al. ALMA front-end optics. Technical report, ESA, April 2004.

[55] A. Baryshev et al. ALMA band 9 optical layout. Technical Report ALMA Memo 394, ESA, 2001.

[56] B. Kruizinga et al. HIFI main optics design and analysis. Technical report, TNO - TPD, Deft, the Netherlands, 2001.

[57] T. Finn. *Quasioptical Verification of HIFI for the Herschel Space Observatory and Standing Waves in Quasioptical Systems*. PhD thesis, NUI Maynooth, Co. Kildare, Ireland, 2007.

[58] N. Trappe. *The Quasi-Optical Verification of HIFI*. PhD thesis, NUI Maynooth, Co. Kildare, Ireland, 2002.

[59] W. Jellema et al. A near-field measurement system facility at 480 GHz. Technical report, SRON, 2000.

[60] K. Pontoppidan. *GRASP9 Technical Description*. Ticra, TICRA Engineering Consultants, Copenhagen, Denmark, 2005.

[61] M. van der Vorst et al. Effect of internal reflections on the radiation properties and input impedance of integrated lens antennascomparison between theory and measurements. *IEEE Transactions On Microwave Theory And Techniques*, 49, No. 6:1118–1125, June 2001.

[62] B. Chantraine-Bars et al. A new accurate design method for millimeter-homogeneous dielectric substrate lens of arbitrary shape. *IEEE Transactions On Antennas And Propagation*, 53, No. 3:1069–1082, 2005.

[63] A. V. Borksin et al. Advanced design of an elliptical lens antenna for mm-wave and sub-mm wave receivers. In *European Microwave Conference*, volume 3, pages 1413 – 1416, Amsterdam, 2004.

[64] A. Neto et al. Reflections inside an elliptical dielectric lens antenna. *IEE Proc.-Microw. Antennas and Propagation*, 145, No. 3:243–247, 2004.

[65] M. van der Vorst. *Integrated Lens Antennas for Submillimetrewave Applications*. PhD thesis, Eindhoven University of Technology, The Netherlands, 1999.

[66] B. Jackson. Quasioptical lens and MSA optical designs for bands 5, 6L, and 6H. Technical Report Draft 4.3, SRON, 2003.

[67] G. Curran. Quasi-optical design of the HIFI intrument for the Herschel. Master's thesis, NUI Maynooth, Co. Kildare, Ireland, 2002.

[68] P. Goldsmith. Non-paraxial gaussian beams. *Applied Optics*, 29:1940–1946, May 1990.

[69] X. Wu et al. Design and characterization of single and multiple-beam mm-wave circularly polarized substrate lens antennas for wireless communications. *IEEE Transactions On Microwave Theory And Techniques*, 49, NO. 3:2408–2411, 2001.

[70] J. Murphy et al. Phase centers of horns antennas using gaussian beam mode analysis. *IEEE Transactions on Antennas & Propagation*, 38:1306–1310, 1990.

[71] R. Wylde et al. Gaussian beam-mode analysis and phase-centers of corrugated feed horns. *IEEE Transactions On Microwave Theory And Techniques*, 41:1691–1699, 1993.

[72] J. Yang et al. Calculation of phase centers of feeds for reflectors when the phase variations are large. In *Antennas and Propagation Society International Symposium*, volume 4, pages 2050–2053. IEEE, June 1998.

[73] Z. Huijuan. The effects of near-field factors on rectangular horn antenna's phase center. In *Antennas, Propagation & EM Theory, 2006. ISAPE '06. 7th International Symposium on*, pages 1–4, Oct. 2006.