# A COMPUTATIONAL STUDY OF GABOR ZONE PLATE GAMMA RAY HOLOGRAPHY

by

**CLARE ESTELLE JACKSON**

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Physics and Astronomy
The University of Birmingham
December 2000

# ABSTRACT

Gamma ray zone plate holography is a new technique with applications to Nuclear Medicine. Unlike other tomographic techniques, three-dimensional images can be reconstructed from just one projection.

The history of zone plate holography is reviewed and the differences between this technique and conventional holography are outlined. Sources of error in the recorded hologram are reviewed and methods for the assessment of image quality are given.

Three image reconstruction techniques are described and compared. These techniques are convolution, deconvolution and the CLEAN algorithm.

Simulated diffraction is the main image reconstruction method which has previously been used to reconstruct images from zone plate holograms. This method is a form of convolution reconstruction. Several variations on this technique are introduced and compared. Matched filtering is also investigated and compared with the simulated diffraction based methods.

An approximate Fourier Wiener filter is used to reconstruct the images by deconvolution. Several different versions of this filter are discussed and compared.

The CLEAN algorithm reconstruction is an iterative method which is based on either convolution or deconvolution.

The three methods are tested using both computer generated and real gamma ray zone plate holograms.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1

# INTRODUCTION

## 1.1 Background to the Thesis

Gabor zone plate holography can be used to image gamma rays emitted from radiopharmaceuticals. A three-dimensional image of the source can be reconstructed from just one projection. Tomography is commonly used for three-dimensional imaging, for example, x-ray CT (Computed Tomography), PET (Positron Emission Tomography) or SPET (Single Photon Emission Tomography). In tomography a large number, for example 64 [75], readings are taken at different angles around the patient. In general, readings are taken at as many projection as there are pixels per side in the image. The patient therefore receives a large dose of radiation and it takes a long time to reconstruct the image. As only one reading at one angle needs to be taken for zone plate holography, this technique could provide faster image acquisition at a reduced dose to the patient.

A zone plate is placed in front of and parallel to a gamma ray detector. The gamma ray detector is a gamma (Anger) camera with its collimator removed. Figure 1.1 shows this arrangement. Gamma rays emitted by a radiopharaceutical cast a shadow of the zone plate on the detector. The shadow will be called a hologram as it contains three dimensional information about the distribution of the radiopharmaceutical, (i.e. the source). An image of the source can then be reconstructed from the hologram.

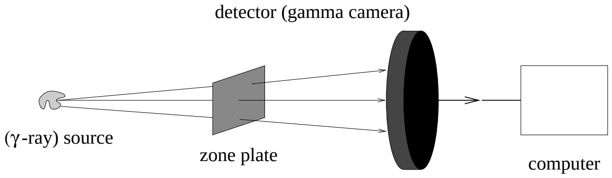


Figure 1.1: Equipment used to record a hologram

If the hologram is reduced to an appropriate size, recorded on photographic film and developed, it can be reconstructed optically by shining light back through it. This reconstruction can be simulated using a computer. Computer based reconstruction has the advantage that other techniques, not based on simulated diffraction, can be used. Zone plate holography research has taken place at the University of Birmingham since 1979 but the image reconstruction techniques still need improvement.

The aim of this work was to find new reconstruction techniques which will improve the image quality. Techniques were applied to a range of computer generated holograms to test their effectiveness for ideal error-free data. The reconstruction methods were then tested on computer generated holograms to which different types of noise were added. Images were then finally reconstructed from real gamma ray zone plate holograms.

In Chapter 2, a short summary of the history of holography is given, from its invention by Gabor in 1948 to the present day. The history of zone plate holography is discussed in more detail with particular emphasis on the work that has been done at the University of Birmingham. Reconstruction techniques used to image zone plate holograms of neutrons, white light and infra red sources can also be used for holograms of gamma ray sources so the work that has been done on these projects is also covered in depth.

In Chapter 3, expressions are derived for hologram formation for both zone plate and conventional holography. The similarities and differences between the two techniques are discussed. Algorithms describing standard optical image reconstruction are given and equations for the reconstructed image of a point source are derived for each type of hologram. Issues specific to imaging gamma ray sources using zone plate holography are highlighted.

In Chapter 4, other sources of error inherent in the recording and processing of zone plate holograms of gamma ray sources are discussed. Methods for the assessment of reconstructed image quality are listed in Chapter 5.

Three different reconstruction techniques are outlined in Chapters 6, 7 and 8. Theoretical zone plate holograms of point sources are calculated using the formulae derived in Chapter 3. The quality of the images produced by each technique are assessed using the techniques given in Chapter 5.

The three techniques are compared further in Chapters 9 and Chapter 10. Images reconstructed from theoretical holograms of extended objects are shown in Chapter 9 and from real holograms of gamma ray sources in Chapter 10.

All of this work is discussed in Chapter 11 and conclusions are drawn. Suggestions for further work are given.

## 1.2 Original Contribution

Gamma ray zone plate holography has been compared with one type of conventional holography and the differences have been highlighted.

Most of the work done on zone plate holography in the past has concentrated on reconstruction methods based on simulated diffraction.

Simulated diffraction is a convolution reconstruction method. Different convolution and matched filter image reconstruction methods have been compared. Convolution with a truncated Fresnel function was found to give better results than simulated diffraction for computer generated holograms of point sources. This method was then used to reconstruct images from a selection of computer generated and real holograms.

Fourier Wiener filtering and the CLEAN algorithm are commonly used to reconstruct two-dimensional images. These techniques have been used to reconstruct three-dimensional images from zone plate holograms. Different approximate Wiener filters have been tested on computer generated point sources and the most effective one has then been used to reconstruct images from a selection of computer generated and real holograms. These images were then compared with those obtained by convolution.

The advantages and disadvantages of the three techniques, convolution, deconvolution and the CLEAN algorithm have been discussed.

# Chapter 2

# LITERATURE REVIEW

## 2.1 History of Holography

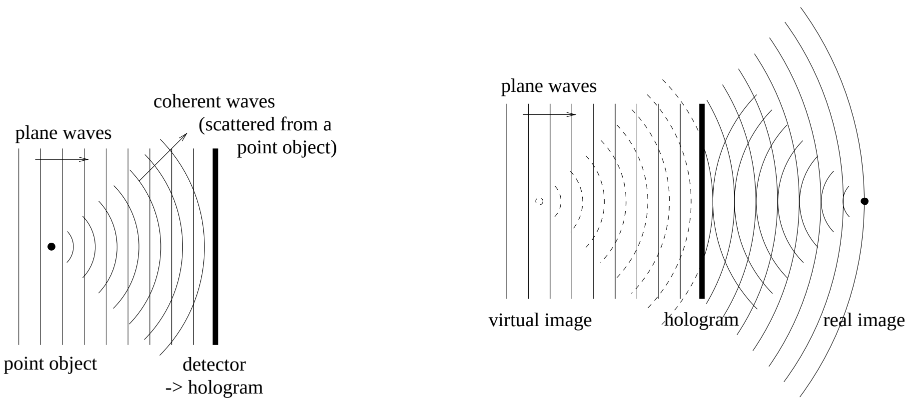
Holography is a way of recording information from a three-dimensional object in two dimensions. The idea is based on the Huygens-Fresnel Principle which states that

> Every point of a wave-front may be considered as a centre of secondary disturbance which gives rise to secondary wavelets, and the wave-front at any later instance may be regarded as the envelope of those wavelets. The secondary wavelets mutually interfere [14].

This means that, at any point in space, a wave contains the whole of the information about its source. If the phase and amplitude of a wave could be recorded, the source could be reconstructed in three dimensions. Most photographic processes just record the intensity or amplitude of a wave. Holography provides a way of recording the phase as well.

In 1971 the British / Hungarian scientist Gabor was awarded the Nobel Prize in Physics for his discovery of holography. It is sometimes argued that the Russian scientist Denisyuk should have also been awarded the Nobel Prize for his invention of reflection holograms [23] which turned holography into a useable technique. His work will be discussed in more detail later.

Gabor had invented holography twenty-four years beforehand when he was investigating ways to improve the resolving power of electron microscopes. He built on work that had been done by others developing two-stage microscopy techniques, for example, Boersch [13]. The resolution of electron microscopes was limited by the spherical aberration of electron lenses which were provided by magnetic fields. Gabor's paper in Nature in 1948 entitled 'a new microscopic principle' [30] described how micrographs could be obtained without electron lenses.



(a) Hologram formation (b) Hologram reconstruction

Figure 2.1: Gabor in-line hologram

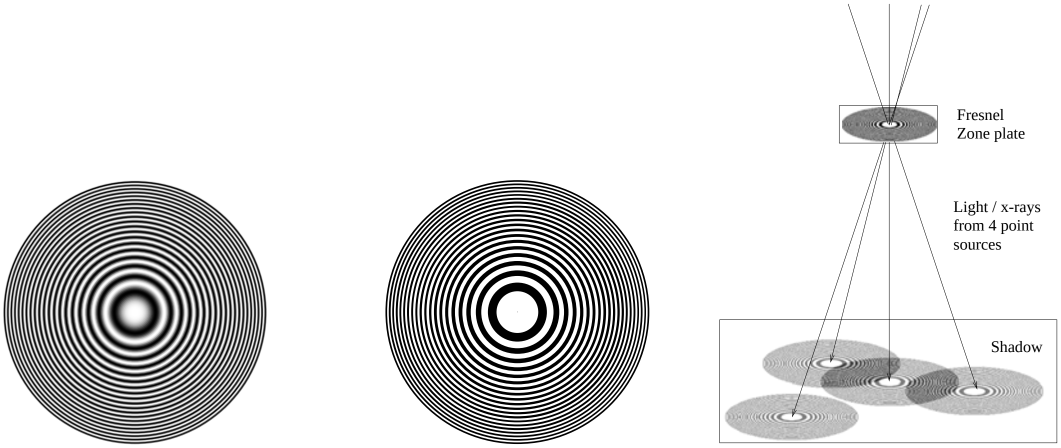
Gabor's method had two stages. In the first stage, the object was illuminated by an electron beam. This beam was partially scattered from the object. The scattered wave then interfered with the primary beam and the interference pattern was recorded on a photographic plate. In the second stage, coherent light was shone through the developed photograph. A real and a virtual image of the object were obtained. This method is illustrated in Figure 2.1. As it was not possible at the time to produce suitable beams of electrons, Gabor demonstrated the feasibility of his technique by using visible light instead of electrons. He used a mercury-arc lamp as his source. The light produced by this lamp was 30-40% coherent. This meant that Gabor could only use very small objects for his holograms. An explanation of the difference between coherent and incoherent light is given in Appendix A.

In his next paper [31], Gabor developed the theory behind his method. He also discussed the impurities in the reconstructed image. The main problem was the presence of the real image in-line with the virtual image. The real image was referred to as the twin image. Only poor quality holograms of two-dimensional objects were produced.

Gabor later called his method 'holography' after the Greek word holos meaning 'whole' and gramma meaning 'message' [32]. He had invented a method of recording the phase as well as the amplitude of a wave, the whole of the wave was preserved.

In 1950, Rogers noticed the similarity between a Fresnel zone plate and the hologram of a point source [71] and said that a hologram could be considered as a zone plate with a complex pattern.

Zone plate holography was first suggested in 1961 by Mertz and Young [60]. Fresnel zone plate imaging was suggested as a technique for use in x-ray cameras. Neither refracting nor reflecting materials were available to provide focusing at x-ray wavelengths and zone plate imaging provided a way round this problem. They illustrated their technique by replacing the lens of a normal camera with a Fresnel zone plate.



(a) Gabor zone plate (b) Fresnel zone plate (c) Shadow casting

Figure 2.2: Zone plates and shadow casting. The shade of gray indicates the transmission of the plate. Black = 0% transmission and white = 100% transmission

A Fresnel zone plate is a binary zone plate. This means that its transmission function at any point is either zero or one. As is discussed in Chapter 3, a Gabor hologram of a point source has the following transmission function $T(r)$

$$T(r) = \frac{1}{2}\left(1 + \cos\left(\frac{\pi r^2}{r_1^2}\right)\right) \tag{2.1}$$

where $r$ is the radial distance from the centre of the plate and $r_1$ is a constant.

A zone plate with the transmission function $t(r)$ is called a Gabor zone plate. A Fresnel zone plate is a binary approximation of a Gabor zone plate. It consists of alternating opaque and transparent concentric rings which are called zones. The outer radii of the transparent concentric zones are proportional to the square root of 1, 2, 3 etc... A Fresnel zone plate and a Gabor zone plate of the same dimensions can be seen in Figures 2.2(b) and 2.2(a). The hologram of a point source would be as large as the detector. The Fresnel and Gabor zone plates, however, are truncated at a maximum radius. Light or x-rays would cast a shadow of the zone plate on the detector as is illustrated in Figure 2.2(c).

If the light source is considered to consist of many point sources, the pattern recorded on the photographic film can be seen as the superposition of the shadows from each of these sources. The photograph would then be an approximation of a Gabor hologram of an equivalent source. This zone plate hologram could then be reconstructed in the same way as a Gabor hologram, by shining coherent light through it. The advantage of this technique is that the hologram is formed by a shadow casting process and so is independent of wavelength or coherence length.

As was explained in Chapter 1, the aim of the work presented here was to find new image reconstruction techniques for zone plate holograms of gamma ray sources. The history of zone plate holography will be discussed in detail in Section 2.2. For the purpose of this report, this general history of holography could end here. For the sake of completeness, however, the major advances made in holography since 1961 will now be discussed.

In 1962, Leith and Uptanieks [52] saw similarities between Gabor's technique and the research they had been doing into radar. They described Gabor's two-step process from a communication theory viewpoint and saw the twin image as an aliasing problem. They suggested that a carrier frequency in the form of an off-axis reference beam could be used to improve image quality. The laser had been invented in 1960 and in 1964, Leith and Upatnieks made the first laser transmission hologram of 3D objects using their off-axis method. Laser transmission holograms produced clear images with realistic depth but could only be viewed with laser light.

Also in 1962, Denisyuk found a way to make white-light reflection holograms [23]. He combined holography with a form of colour photography invented in 1891 by the Nobel Laureate Lippmann. He placed a sheet of photosensitive emulsion between the reference beam and the object. The emulsion had a thickness large compared to the wavelength of visible light. Interference between the reference beam and the beam reflected from the object caused standing waves inside the emulsion. When the emulsion was developed, its density followed the intensity pattern of the standing wave. Denisyuk holograms are often termed volume holograms since they are recorded over the thickness of the emulsion and not just in two dimensions. If monochromatic light is used to produce the hologram, white light can be used in the reconstruction. The hologram will only reflect the fraction of the light with the same wavelength as the original monochromatic source to form the holographic image. The remaining wavelengths will pass through the hologram having no effect on the image. If more than one wavelength is used, colour images are obtained.

Another significant advance was holographic interferometry. Powell and Stetson published the first paper on this in 1965 [67]. Small distortions between two holographic exposures are seen as contours on the reconstructed image. Holographic interferograms are particularly useful in non-destructive testing, fluid flow analysis and quality control.

Benton invented the rainbow hologram in 1968 [4]. This was a transmission hologram which could be viewed with white light. Hologram formation was by a two-step off-axis technique. The real image from a first hologram was used as the object for a second hologram. The first hologram was, however, masked down to a narrow horizontal slit. This resulted in a reconstructed image in one or several colours depending on the viewpoint.

The first moving three-dimensional images were made by Cross in 1972 [18]. Cross combined white light transmission holography with conventional cinematography. He recorded a sequence of holograms in narrow strips along the photographic film. As the hologram was viewed from left to right, different images were seen giving the illusion of movement.

In 1974, Foster [74] invested a way for duplicating holograms mechanically. He used an embossing process similar to the one used to make audio-disks. Rainbow holograms were turned into reflection holograms by an aluminium backing and could be mass produced at a very low cost.

Holography has many applications. Holographic lenses are now used in supermarkets to direct laser light onto the bar codes during checkout. Holograms are also used on some credit cards to make them difficult to copy. Holography can be used for archival recording of fragile museum pieces. A pulsed laser hologram was even used to record Lindow Moss, a 2300 year old Iron Age man unearthed from Lindow Moss, a peat bog in Cheshire [40].

## 2.2 History of Zone Plate Holography

### 2.2.1 General History

As mentioned in section 2.1, the first paper on zone plate holography was written by Mertz and Young in 1961. Incoherent radiation sources can be imaged using this technique so it has found many applications.

In 1972, Barret used a lead Fresnel zone plate to image gamma rays emitted from a thyroid phantom [3]. The detector he used was a gamma camera. Gamma cameras are used in nuclear medicine and are usually fitted with a pinhole aperture or a parallel hole collimator which only lets a tiny fraction ($\approx$ 0.1 - 0.01 %) of the incident gamma rays through. The advantage of a zone plate is that only half of the gamma rays are stopped. The holograms were reconstructed optically.

Fresnel zone plates were again used for gamma ray imaging by Rogers et. al. in 1973 [72]. An image intensifier was used instead of a gamma camera. Zone plate holography is basically an on-axis method and Rogers discovered that the virtual image in line with the real image produced artifacts in the reconstruction. He suggested that an off-axis zone plate could be used to improve the image.

In 1977, Ceglio used zone plates to study the x-ray emission of laser-produced plasmas [20] and [19]. Other applications have included imaging of fission gamma rays produced by reactor fuel and focusing x-rays in 5eV synchrotron sources.

One problem with zone plates is that Moiré zone plate patterns are produced when several zone plate patterns overlap. Moiré patterns are zone plates themselves and focus light. Jaroszewicz wrote a paper on this effect [44].

Jones and Kirz introduced phase zone plates [45] [47]. Instead of opaque zones, phase zone plates have zones which cause a $\pi$ phase shift. This meant that the intensity of the light focussed by the plate was increased by a factor of 4 .

In 1974, McDonald et. al. showed how gamma ray Fresnel zone plate holograms could be reconstructed computationally [53]. He used a two-dimensional fast Fourier transform algorithm. One of the advantages of computer reconstruction is that the image can be processed.

### 2.2.2 The Zone Plate Holography Project at the University of Birmingham

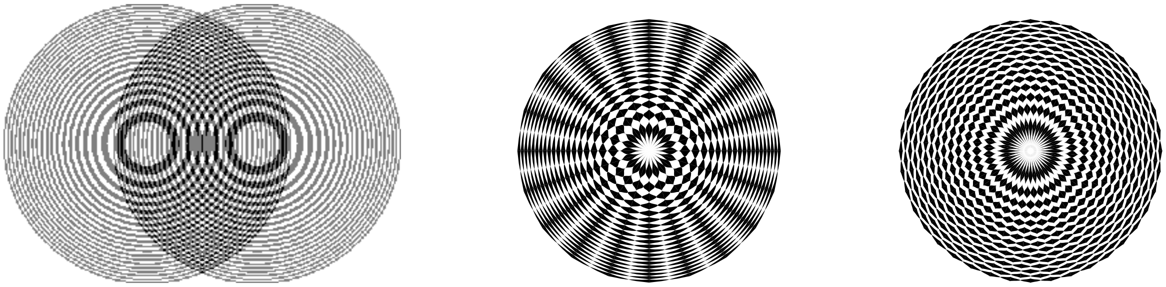
Research into zone plate holography has been undertaken at the University of Birmingham since 1979. Projects have been completed in neutron holography, white light holography and infra red holography. Work is now concentrated on gamma ray holography. The work presented here concerns the computer reconstruction of gamma ray holograms. The algorithms, however, could be applied to holograms formed by other types of radiation. A brief description of current and completed work follows.

#### Neutrons

The first work on neutron zone plate holography was done by Pink in 1979 as part of his MSc [66] under the supervision of Professor Beynon. Neutron holography was investigated as an alternative to neutron tomography for use in non-destructive testing. The zone plate was made of 1mm thick aluminium. The opaque zones were made by using $FeCl_3$ to etch grooves in the aluminium. These grooves were then filled with a gadolinium containing paint. Cold neutrons (with energies less than $\approx 0.025$eV) from a research reactor were scattered from test objects. The hologram was recorded using a gadolinium converter screen with x-ray film behind it. The holograms were reconstructed optically by reducing them in size and shining laser light through them. This work was published in a paper in 1980 [10].

Mast did further work in this field and submitted for his PhD in 1983 [57]. He used thermal neutrons from a radiotherapy beam but the rest of the experimental setup was the same as that used by Pink. The holograms were digitised and reconstructed using a computer. The first algorithm used was an unmodified two-dimensional fast Fourier transform (2DFFT) as described by Singleton [79]. Another single 2DFFT method, based on the Rayleigh-Sommerfeld equation was then used. Results from both computer reconstruction methods were compared with those from the optical reconstruction. The unmodified 2DFFT method produced a poor quality reconstruction. The Rayleigh-Sommerfeld method, however, produced images of a similar quality to those obtained by optical reconstruction. Mast also used the Monte-Carlo code MORSE [25] to make simulated holograms. This work was published in 1983 [7]. Another article was written by Professor Beynon in 1986 [6].

The next student to work on this project was Mathews who completed his PhD in 1988 [58]. The equipment he used was the same as that used by Mast. He used two zone plates of opposite polarity to form two holograms. Opposite polarity means that one plate had opaque zones where the other had transparent zones. One hologram was subtracted from the other and the result was reconstructed. This improved the image. He again used the Rayleigh-Sommerfeld 2DFFT algorithm to reconstruct the holograms. He compared a polar coordinate $(\mathrm{r}, \theta)$ version of this technique with a Cartesian coordinate $(\mathrm{x}, \mathrm{y})$ version. He also applied zone plate holography to Positron Emission Tomography (PET). This is gamma ray holography and will therefore be discussed later.



(a) Moiré zone plate pattern produced by two overlapping Fresnel zone plates

(b) Hampton court (binary) Gabor zone plate

(c) Rotated binary Gabor zone plate

Figure 2.3: Binary zone plate patterns

Kirk finished his PhD in 1992. He started by analysing Moiré zone plate patterns which appear when Fresnel zone plates overlap. Moiré patterns act as zone plates and focus light themselves. A Moiré pattern is shown in Figure 2.3(a). As Moiré patterns have a similar effect to zone plates themselves, Kirk investigated the possibility of making a pure Gabor zone plate by overlapping several Fresnel zone plates.

Fresnel zone plates also contribute to noise in the image. A Gabor zone plate has one focal point a distance $f$ from the plate. An equivalent Fresnel zone plate would have a series of focal points at $f, \frac{f}{3}, \frac{f}{5}, \frac{f}{7}$ etc. If the hologram was reconstructed optically, the presence of these extra focal points would reduce the quality of the image. Fresnel zone plates had always been used because of the difficulty of creating a Gabor zone plate with a smoothly varying transmission function.

Kirk argued that the width of the zones in a binary zone plate could be varied azimuthally as well as radially so that the average transmission at any radius would be the same as for a Gabor zone plate. The result was the Hampton court zone plate, so called because it resembled the Hampton Court clock. Mast and Mathews showed that the binary Gabor plate only had one focal point and, hence, put much of this work on a firmer theoretical footing. A paper was later published on this zone plate design [9]. In this paper, the plate was called a binary Gabor zone plate. The design, which is shown in Figure 2.3(b), was patented. The binary Gabor zone plate lead to artifacts in the image like the spokes of a bicycle wheel. The rotated binary Gabor zone plate, shown in Figure 2.3(c), was an attempt to solve this problem.

### White Light

Woodgate researched white light zone plate holography for his PhD which he finished in 1996 [92]. Just as Gabor used light from a mercury arc lamp to demonstrate his technique for obtaining electron micrographs, the experiments with white light were done with the aim of finding new techniques to improve neutron holography. Neutron zone plate holograms contain errors due to the limited number of neutrons which form the hologram. Such statistical errors do not apply when white light is used.

Many of the experiments were performed jointly by Woodgate and Caplan, another PhD student who submitted her thesis in 1997 [16]. At the same time, Strange, was researching different types of coded aperture imaging and antennas. In his PhD thesis, which was completed in 1996, he discussed methods for the computational reconstruction of zone plate holograms [83].

A CCD camera was placed in a light-tight box with a translucent screen at one end. The camera was focussed on the opposite side of the screen to which the hologram was formed. The hologram was digitised by a SUN video card and stored on a computer. Woodgate wrote postscript programs to draw the zone plate patterns. Zone plates patterns could then be laser printed and photographed onto film. The final zone plate was produced on the developed film. Light was reflected off objects of which holograms were formed.

All the holograms were reconstructed computationally. Woodgate used Mast's Rayleigh-Sommerfeld simulated diffraction reconstruction technique. He pointed out the similarity between this technique and correlating the hologram with a suitably scaled Gabor zone plate pattern. This correlation technique was shown to produce images nearly the same as those produced by the diffraction technique. A Welch window [68] was used with the correlation technique to improve image quality. He also tried to reconstruct the holograms by deconvolution but this gave poor results. He suggested that further work could be to produce an iterative reconstruction algorithm such as the one used by Helstrom in 1967 [37] or Fleming in 1984 [28]. Experimental holograms were compared to simulated holograms produced using the Monte Carlo code MCNP [15].

Strange looked at a number of other reconstruction methods including matrix inversion, matrix Wiener filtering and single value decomposition. Single value decomposition was used in order to remove artifacts from images by removing certain frequencies. Strange used these techniques to reconstruct images from small simulated holograms. They could, however, not be used for larger experimentally produced holograms because the amount of computer memory (RAM) that would be needed to perform the reconstruction was incredibly high. Strange also designed an approximate binary Gabor zone plate, made from a series of lines instead of smooth curves. This was needed so that the plate could easily be cut from a piece of metal and used for future work in gamma ray holography.

Caplan used the correlation technique to reconstruct her holograms. She investigated the effects of varying the number of zones in the zone plate and the frequency of their azimuthal variation on the quality of the reconstructed image. She also investigated the variation of the point spread function with position. If the detector is small, not all of the zone plate shadow is recorded. Caplan performed experiments to look at this effect.

### Infra Red

Infra red holography could be used to image the temperature of the body in three dimensions. Regions of cancerous tissue often have a higher temperature and therefore could be located in this way. The first student to work on this project was Kuo who finished his PhD in 1997 [49]. Shen also worked on this project and finished his PhD in 2000 [77].

Zone plates for infra red holography are produced in the same way as the ones used in the white light experiments. Neutron and white light holography detected radiation scattered from objects. Infra red holography images heat producing sources, for example a tungsten-filament lamp or an electric heater. The hologram is recorded by using a small detector attached to a X-Y plotter which moves it across the plane in which the hologram is formed. The human body emits radiation in the far infra red region. A suitable detector could, however, not be found to record far infra red so experiments were performed in the near infra red. The detector used was a InGaAs detector of 0.2mm diameter. A system was developed to read the detector at intervals so the two-dimensional hologram could be recorded as the plotter moved across the plane.

Kuo used Mast's diffraction algorithm to reconstruct the holograms. He reconstructed a uniform hologram and then subtracted this from the reconstruction of the real hologram to remove artifacts due to digitisation. He also used a modified Welch window to remove noise from the image. Thresholding was used to remove the background noise from the image.

Shen used correlation to reconstruct the holograms. He has also reconstructed the images by correlating it with the appropriate binary Gabor zone plate pattern. He used a sine as well as a cosine zone plate to form two holograms which can be combined to give a hologram produced by complex (real and imaginary) zone plate. The reconstruction of this complex hologram showed improvement over those from a single zone plate.

### Gamma Rays

Mathews applied zone plate holography to PET. In PET, two gamma rays from positron annihilation are detected on opposite sides of the positron source by two PET cameras. A path through the source, along which the annihilation must have taken place, can then be constructed. Mathews imagined there were Fresnel zone plates infront of each camera and subtracted every gamma ray count which would have hit an opaque zone of one of the plates. The resulting holograms were then reconstructed as normal. The images produced using this technique were of comparable quality to those produced by standard back projection. The holography based technique produced the images in a fraction of the time. This work was published in a paper in 1988 [8].

Caplan discussed the difference between white light and gamma ray holography in her PhD thesis. In order to attenuate the gamma rays, zone plates used for gamma ray holography have to be several millimetres thick. This leads to the additional problem of vignetting where gamma rays travel through the zone plate at an angle and are only partially absorbed. This means that gamma ray holograms are generally of poorer quality than those produced by white light.

Perks' project was to image gamma rays in the energy range 140keV to 511keV and he submitted his thesis in 1998 [65]. He used the gamma emitting isotopes $^{99m}$Tc, $^{18}$F, and $^{22}$Na which are used in nuclear medicine. The detector he used was a gamma camera with the collimator removed. The first zone plate was a 3cm lead approximate binary Gabor zone plate which was commissioned by British Nuclear Fuels Ltd for studies into gamma ray imaging. This was later replaced by a 1mm sintered tungsten plate for imaging low energy gammas (140keV). An Ohio Nuclear series 100 gamma camera was first used to record holograms. These holograms were then compared to ones taken using a state of the art Toshiba GCA-7200A gamma camera at St Thomas' Hospital, London and using an ADAC gamma camera at the Queen Elizabeth Medical Centre, Birmingham. The Toshiba GCA-7200A camera holograms gave better images.

Experimental holograms were again compared with holograms simulated using MCNP. The MCNP file for the simulation was written by Dhaba'an, a research assistant in the applied physics group. It included a representation of the approximate binary Gabor zone plate. Perks used Mast's diffraction algorithm for all his reconstructions.

Rew finished her PhD in gamma ray zone plate holography in 1999 [70]. She concentrated on gamma ray energies used in Nuclear Medicine in general and in BNCT in particular. Research is being done into BNCT at the University of Birmingham. BNCT is a treatment for a type of brain tumour called Glioblastoma multiforme. The patient is injected with a boron compound. This compound concentrates to a higher degree in brain tumour cells than in normal brain tissue. When the head is irradiated with thermal neutrons, the following reaction occurs.

$$\mathrm{n} + {}^{10}\mathrm{B} \rightarrow {}^{7}\mathrm{Li} + \alpha + \gamma$$

The alpha particles ($\alpha$) produced in the reaction kill the tumour cells in which they are concentrated. The gamma rays ($\gamma$) have an energy of 478keV and can potentially be imaged using zone plate holography. This would potentially provide a method to monitor the treatment while it is taking place.

A 11-12mm thick sintered tungsten zone plate would be needed to effectively stop 478keV gamma rays. Rew used a series of 1mm plates placed together. A 4mm plate, for example, would have a minimum transmission of $\approx 40\%$. The tungsten plate was cut by a laser and so the original design can be accurately reproduced. She set up a gamma ray holography lab at the University of Birmingham with a GE Maxi Camera 535 which has a higher efficiency than the Ohio camera used by Perks but a lower efficiency than the Toshiba GCA-7200A camera which he also used. BNCT involves beams of neutrons so the effects of neutron activation in the camera and scattering off the walls in the treatment room were also investigated.

Takhar [84] and George [33] are now both working on the gamma ray holography project. They will be using an ADAC Forte gamma camera as well as the GE Maxi Camera 535 to record further zone plate holograms of gamma ray sources. It is hoped that zone plate holograms will be recorded during BNCT patient trials and that the reconstruction techniques presented here will be tested using this data.

## 2.3 Summary

Many advances have been made in holography since Gabor's first paper in 1948. One advance was Mertz and Young's invention of zone plate holography in 1961. Zone plate holography is a method of imaging incoherent radiation sources in three dimensions and so has found many applications. One application is radioisotope imaging in nuclear medicine. It could also be used to image the 478keV gammas emitted during Boron Cancer Therapy.

Reconstruction algorithms developed for zone plate imaging of neutrons, light and infra red can also be used for gamma ray imaging. All the previous PhD projects on zone plate holography completed at the University of Birmingham have therefore been reviewed. Although many techniques have already been tried, the reconstruction still needs improvement.

This work suggests some new reconstruction techniques and tries them on computer generated holograms and some of the limited number of actual gamma ray holograms available at this point. Takhar and George will continue the work on this project using an ADAC Forte gamma camera and a GE Maxi Camera 535 to record more holograms. The aim of their work will be to develop gamma ray zone plate holography to the stage where it can be used to image patients undergoing BNCT.

# Chapter 3

# COMPARISON OF ZONE PLATE HOLOGRAPHY WITH CONVENTIONAL HOLOGRAPHY

Holography is a very broad field and it is beyond the limits of this project to analyse all different types of hologram. The type of hologram which is most similar to zone plate holograms is an in-line Gabor hologram and this is the type which will therefore be considered.

In order to derive the equations for hologram formation and reconstruction, approximations from diffraction theory must be used. These approximations will be reviewed in this Chapter. As in Chapter 2, the starting point of this analysis will be the Huygens-Fresnel principle. The Huygens-Fresnel principle is, as Born and Wolf [14] state, "... the most powerful approximate method and is adequate for the treatment of the majority of problems encountered in instrumental optics" and can be stated as follows.

> Every point on a wavefront serves as a source of spherical secondary wavelets of the same frequency as the primary wave. The optical field at any point beyond an obstruction is the superposition of all such wavelets reaching that point. [35]

The Kirchhoff-Fresnel diffraction formula is a mathematical expression of the Huygens Fresnel principle and is the formula used to analyse a large number of holographic systems. This formula is based on the Kirchhoff boundary conditions which will be given later in equations 3.12 and 3.13. These boundary conditions ignore edge effects at an aperture and thus are only valid for apertures which are larger than a wavelength. This equation can, therefore, not be used to describe very near field diffraction. In this case, the Green's function solution should be used.

The Kirchhoff-Fresnel diffraction theory can be applied to Gabor in-line holography and so is the theory which will be discussed here. An outline of the derivation of this formula will first be given to show where the approximations are made and therefore under which conditions the formula is valid. The formation of a hologram of an opaque pattern on a transparent sheet will then be analysed. This transparency will be assumed to be illuminated by a planar coherent light source and that it is parallel to both the source and the film on which the hologram is recorded. This system will be referred to as a general holographic system as the specific cases of Gabor in-line holography and zone plate holography can be analysed using the relationships found for this, more general case. Both these cases will be analysed and, finally, factors affecting the quality of the reconstructed image will be discussed.

## 3.1 Derivation of the Kirchhoff-Fresnel Diffraction Formula (Proof of the Huygens-Fresnel Principle)

The Huygens-Fresnel principle can be proven by showing that secondary waves from every point on a primary wavefront give the same optical field as the primary wavefront would have done. This is done in Born and Wolf by dividing the primary wave into zones, making assumptions about the directional variation of the amplitude of the secondary waves from each zone and approximately summing the contributions from the secondary waves.

A more thorough proof of the Kirchhoff-Fresnel diffraction formula is also given by Born and Wolf. The physical significance of the assumptions made in this derivation is easier to understand and so this is the method which will be discussed here. The waves will be assumed to be monochromatic.

As described in appendix B, the amplitude of an electromagnetic wave, $\Psi(x, y, z, t)$, satisfies the scalar wave equation.

$$\nabla^2\Psi(x,y,z,t) - \frac{1}{c^2}\frac{\partial^2\Psi(x,y,z,t)}{\partial t^2} = 0 \tag{3.1}$$

where $c$ is the speed of light. $\Psi(x, y, z, t)$ can be separated into a time dependent and a space dependent part. If we have monochromatic light, the time dependent part of the wave just has one frequency and $\Psi(x,y,z,t)$ can be written as

$$\Psi(x,y,z,t)=\psi(x,y,z)e^{-i\omega t} \tag{3.2}$$

where $\psi(x,y,z)$ is the spatial part of the variation and $\omega$ is the frequency of the wave. Substituting this into equation 3.1 gives

$$\nabla^2\psi(x,y,z)+\frac{\omega^2}{c^2}\psi(x,y,z)=0 \quad \text{where} \quad c=\frac{1}{\sqrt{\mu_0\epsilon_0}} \tag{3.3}$$

$\mu_0$ is the permeability of free space and $\epsilon_0$ is the permittivity of free space. This is the time independent wave equation (Helmholtz equation).

Consider a wave with space dependent part $\psi(x,y,z)$. It will satisfy the Helmholtz equation.

$$(\nabla^2+k^2)\psi=0 \qquad \text{where} \qquad k=\frac{\omega}{c} \tag{3.4}$$

Let $V$ be a volume bounded by a surface $S$ and let $P$ be a point within it. Let $\psi(x,y,z)$ have continuous first and second order partial derivatives within the volume and on the surface. If $\varphi(x,y,z)$ is any other function which satisfies the same continuity requirements as $\psi(x,y,z)$, Green's theorem would give

$$\int\int\int_V(\psi\nabla^2\varphi-\varphi\nabla^2\psi)dV=-\int\int_S\left(\psi\frac{\partial\varphi}{\partial\boldsymbol{n}}-\varphi\frac{\partial\psi}{\partial\boldsymbol{n}}\right)dS \tag{3.5}$$

where $\frac{\partial}{\partial\boldsymbol{n}}$ denotes differentiation along the inward normal to the surface $S$.

In particular, if $\varphi(x,y,z)$ also satisfies the Helmholtz equation, i.e.

$$(\nabla^2+k^2)\varphi=0 \tag{3.6}$$

then substituting equations 3.4 and 3.6 into equation 3.5, the left hand side of equation 3.5 vanishes giving

$$\int\int_S\left(\psi\frac{\partial\varphi}{\partial\boldsymbol{n}}-\varphi\frac{\partial\psi}{\partial\boldsymbol{n}}\right)dS=0 \tag{3.7}$$

One solution to the Helmholtz equation is $\varphi(x,y,z)=\frac{e^{iks}}{s}$ where $s$ is the distance from point $P$ to the point $(x,y,z)$. This gives a singularity at $s=0$ so $P$ has to be excluded from the region of integration. This is done by surrounding $P$ by a sphere of radius $\epsilon$. The integration is then between surface $S$ and the surface $S'$ of the sphere. This is illustrated

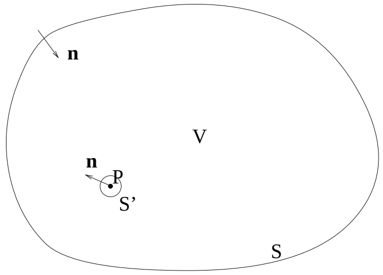


Figure 3.1: Region of integration

in Figure 3.1.

Equation 3.7 then becomes

$$\int\!\!\int_S \left\{\psi\frac{\partial}{\partial \boldsymbol{n}}\left(\frac{e^{iks}}{s}\right)-\frac{e^{iks}}{s}\frac{\partial\psi}{\partial \boldsymbol{n}}\right\}dS+\int\!\!\int_{S'} \left\{\psi\frac{\partial}{\partial \boldsymbol{n}}\left(\frac{e^{iks}}{s}\right)-\frac{e^{iks}}{s}\frac{\partial\psi}{\partial \boldsymbol{n}}\right\}dS'=0 \qquad (3.8)$$

and then

$$\begin{aligned}\int\!\!\int_S \left\{\psi\frac{\partial}{\partial \boldsymbol{n}}\left(\frac{e^{iks}}{s}\right)-\frac{e^{iks}}{s}\frac{\partial\psi}{\partial \boldsymbol{n}}\right\}dS &= -\int\!\!\int_{S'} \left\{\psi\frac{e^{iks}}{s}\left(ik-\frac{1}{s}\right)-\frac{e^{iks}}{s}\frac{\partial\psi}{\partial \boldsymbol{n}}\right\}dS' \\ &= -\int\!\!\int_{\Omega} \left\{\psi\frac{e^{ik\epsilon}}{\epsilon}\left(ik-\frac{1}{\epsilon}\right)-\frac{e^{ik\epsilon}}{\epsilon}\frac{\partial\psi}{\partial \boldsymbol{s}}\right\}\epsilon^2 d\Omega \qquad (3.9)\end{aligned}$$

where $d\Omega$ is an element of the solid angle and $\epsilon$ is the small radial displacement of $S'$ from point $P$. The integral over $S$ is independent of $\epsilon$ and so the right hand side can be replaced by its limit as $\epsilon \to 0$. This limit is $4\pi\psi(x,y,z)$. This gives

$$\psi(x,y,z)=\frac{1}{4\pi}\int\!\!\int_S \left\{\psi\frac{\partial}{\partial \boldsymbol{n}}\left(\frac{e^{iks}}{s}\right)-\frac{e^{iks}}{s}\frac{\partial\psi}{\partial \boldsymbol{n}}\right\}dS \qquad (3.10)$$

which is one form of the integral theorem of Helmholtz and Kirchhoff.

Kirchhoff considered the propagation of a wave from a point source $P_0$ through an opaque screen with an opening in it and determined the light disturbance at a point $P(x,y,z)$ behind the screen. This is illustrated in Figure 3.2. The surface in Kirchhoff's integral is formed by the opening A, a portion of the non-illuminated side of the screen B and a large sphere of radius $R$ centred on $P$, C.

$$\psi(x,y,z)=\frac{1}{4\pi}\left[\int\!\!\int_{\mathrm{A}}+\int\!\!\int_{\mathrm{B}}+\int\!\!\int_{\mathrm{C}}\right]\left\{\psi\frac{\partial}{\partial \boldsymbol{n}}\left(\frac{e^{iks}}{s}\right)-\frac{e^{iks}}{s}\frac{\partial\psi}{\partial \boldsymbol{n}}\right\}dS \qquad (3.11)$$

Kirchhoff then applied the following boundary conditions.

$$\text{on A:} \qquad \psi=\psi^i \qquad \frac{\partial\psi}{\partial \boldsymbol{n}}=\frac{\partial\psi^i}{\partial \boldsymbol{n}} \qquad (3.12)$$

$$\text{on B:} \qquad \psi=0 \qquad \frac{\partial\psi}{\partial \boldsymbol{n}}=0 \qquad (3.13)$$

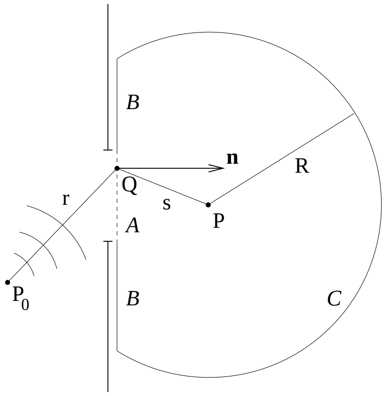


Figure 3.2: Diffraction behind an aperture - surface used to derive the Kirchhoff-Fresnel diffraction formula

where $\psi^i$ is the value relating to the incident field. These boundary conditions arise from the continuity of the electric and magnetic fields and the gradients of these fields across opening A and the opaque nature of surface B. The following expressions are known for $\psi^i$ and $\frac{\partial \psi^i}{\partial \boldsymbol{n}}$ as the point source at $P_0$ emits spherical waves and the distance $P_0Q = \boldsymbol{r}$.

$$\psi^i = \frac{Ae^{ik\boldsymbol{r}}}{\boldsymbol{r}} \qquad \frac{\partial \psi^i}{\partial \boldsymbol{n}} = \frac{Ae^{ik\boldsymbol{r}}}{\boldsymbol{r}} \left[ ik - \frac{1}{\boldsymbol{r}} \right] \cos(\boldsymbol{n}, \boldsymbol{r}) \tag{3.14}$$

$\cos(\boldsymbol{n}, \boldsymbol{r})$ is the cosine of the angle between the vector $\boldsymbol{r}$ and the normal to the opening $\boldsymbol{n}$ and $A$ is a constant. The wave field on the spherical portion of the surface, C, remains to be considered. If it is assumed that the source begins to radiate at some time $t = t_0$, at a time t, the light will have reached a maximum distance from the source of $c(t - t_0)$. If the radius $R$ is chosen to be further away from the source than this, the light will not have reached the surface region C and the integral over this region will vanish. As the integral over B also vanishes, only the integral over A need therefore be considered. The integral becomes

$$\psi(x, y, z) = -\frac{iA}{2\lambda} \int\int_{\mathrm{A}} \frac{e^{ik(\boldsymbol{r}+\boldsymbol{s})}}{\boldsymbol{rs}} [\cos(\boldsymbol{n}, \boldsymbol{r}) - \cos(\boldsymbol{n}, \boldsymbol{s})] dS \tag{3.15}$$

which is the Kirchhoff-Fresnel diffraction formula. $\lambda = \frac{2\pi}{k}$ is the wavelength of the waves.

As the integration moves over all the points in the aperture, $\boldsymbol{r} + \boldsymbol{s}$ will change by many wavelengths so the factor $e^{\boldsymbol{r}+\boldsymbol{s}}$ will oscillate rapidly. If the distances $\boldsymbol{r}$ and $\boldsymbol{s}$ are, however, large compared to the dimensions of the aperture, then the factor $[\cos(\boldsymbol{n}, \boldsymbol{r}) - \cos(\boldsymbol{n}, \boldsymbol{s})]$ will not vary appreciably. This factor may be replaced by $2\cos\delta$ where $\delta$ is the angle between the line $P_0P$ and the normal to the screen. If the distances between $P_0$ and $P$

and the origin are called $\boldsymbol{r'}$ and $\boldsymbol{s'}$ respectively, the factor $\frac{1}{\boldsymbol{rs}}$ may be replaced by $\frac{1}{\boldsymbol{r's'}}$. This gives

$$\psi(x,y,z) = -\frac{iA}{\lambda}\frac{\cos\delta}{\boldsymbol{r's'}}\int\int_{\mathrm{A}} e^{ik(\boldsymbol{r}+\boldsymbol{s})}dS \tag{3.16}$$

Any other surface which approximately fills the aperture could have been chosen instead of A above. If a portion, $W$, of the incident wave front from $P_0$ is chosen instead, a portion C of a cone with its vertex at $P_0$ can be used to complete the surface. If the radius of curvature of the wave is sufficiently large, the integral over C can be neglected. This surface is illustrated in Figure 3.3 On $W$, $\cos(\boldsymbol{n},\boldsymbol{r_0}) = 1$. If a new variable, $\chi$, is

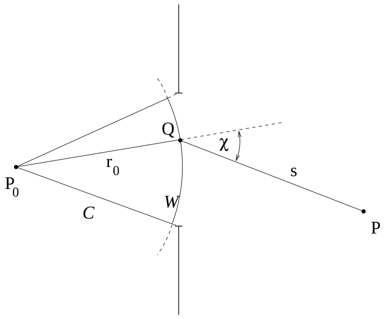


Figure 3.3: Diffraction behind an aperture - surface used to prove the Huygens-Fresnel principle

introduced where $\chi = \pi - (\boldsymbol{r_0},\boldsymbol{s})$, equation 3.15 becomes

$$\psi(x,y,z) = -\frac{i}{2\lambda}\frac{Ae^{ikr_0}}{r_0}\int\int_{W}\frac{e^{iks}}{s}\left(1+\cos(\chi)\right)dS \tag{3.17}$$

where $r_0$ is the radius of curvature of the wave front. This can be written as

$$\psi(x,y,z) = \frac{Ae^{ikr_0}}{r_0}\int\int_{W}\frac{e^{iks}}{s}K(\chi)dS \quad \text{where} \quad K(\chi) = -\frac{i}{2\lambda}\left(1+\cos(\chi)\right) \tag{3.18}$$

It can been seen that every element of the wave front $W$ acts as a source of secondary waves contributing to the field at $P$. The amplitude of these waves is determined by $K(\chi)$, the *inclination factor* which describes the variation with direction of the amplitude of the secondary waves. $K(\chi)$ is a maximum in the original direction of propagation and decreases to zero when $\chi = \frac{\pi}{2}$. The factor $-i$ included in $K(\chi)$ can be also written as $e^{\frac{-i\pi}{2}}$ and may be accounted for by assuming that the secondary waves oscillate a quarter of a period out of phase with the primary wave. The other factor $\frac{1}{\lambda}$ also included in this term may be explained if the secondary waves are assumed to have an amplitude $\frac{1}{\lambda}$ times that of the primary wave.

To conclude, the Huygens-Fresnel principle can be mathematically obtained from the differential wave equation when some assumptions are made. These assumptions are that Kirchhoff's boundary conditions for the optical field behind an aperture apply and that the dimensions of the aperture are small compared to the distance of the source and the point where the field is to be calculated from the aperture. Assumptions must also be made about the amplitude and phase of the secondary waves. The amplitude of each secondary wave is assumed to depend upon the angle between the direction of propagation of the primary wave and that of the secondary wave.

## 3.2 Formation of a General Hologram

L. M. Soroko [81] derived a formula for the diffraction pattern produced by a coherent planar light source a distance $p$ behind a transparency in a plane a distance $q$ in front of the transparency. The transparency is a thin sheet of material, for example photographic film, with a varying transmittance function. This variation, or pattern, could take the form of a picture or some writing and is the two-dimensional object of which the hologram is to be made. Gabor's early holograms were of such objects [30]. Soroko's derivation will be summarised here.

The amplitude of the diffraction pattern, $u(\boldsymbol{x})$, will depend upon the amplitude transmittance function of the transparency, $t(\boldsymbol{\alpha})$, as well as the source amplitude distribution function, $s(\boldsymbol{\epsilon})$. $\boldsymbol{x}(x, y)$, $\boldsymbol{\epsilon}(\epsilon, \eta)$ and $\boldsymbol{\alpha}(\alpha, \beta)$ are vectors which give the position in the hologram, source and transparency planes respectively. These planes are perpendicular to the $z$ axis. This is illustrated in Figure 3.4.

The extended source can be considered as the sum of many coherent point sources. The waves from each point source will be diffracted through the aperture as described by the simplified version of the Kirchhoff-Fresnel diffraction formula given in equation 3.16. As the waves pass through a transparency rather than an aperture, the wave amplitude will be multiplied by the amplitude transmittance of the transparency, $t(\boldsymbol{\alpha})$, at each point.

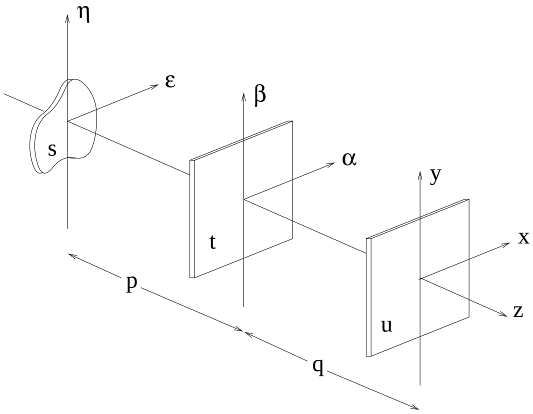


Figure 3.4: Relative positions of source, transparency and hologram

The distances $r$, $s$, $r'$ and $s'$ are given by.

$$\begin{aligned} r^2 &= (\alpha-\epsilon)^2+(\beta-\eta)^2+p^2 & \qquad r'^2 &= \epsilon^2+\eta^2+p^2 \\ s^2 &= (\alpha-x)^2+(\beta-y)^2+q^2 & \qquad s'^2 &= x^2+y^2+q^2 \end{aligned} \tag{3.19}$$

For equation 3.15 to hold, the distances $p$ and $q$ must be much greater than the distances $\alpha, \epsilon, \beta, \eta$. In this case,the following approximations can be used.

$$r \approx p + \frac{(\alpha-\epsilon)^2}{2p} + \frac{(\beta-\eta)^2}{2p} = p + \frac{(\boldsymbol{\alpha}-\boldsymbol{\epsilon})^2}{2p} \qquad r' \approx p \tag{3.20}$$

$$s \approx q + \frac{(\alpha-x)^2}{2q} + \frac{(\beta-y)^2}{2q} = q + \frac{(\boldsymbol{\alpha}-\boldsymbol{x})^2}{2q} \qquad s' \approx q \tag{3.21}$$

Substituting into equation 3.16 and using $k = \frac{2\pi}{\lambda}$, each point in the source therefore has the contribution

$$\psi(x,y,q) = \frac{s(\boldsymbol{\epsilon})}{i\lambda}\frac{\cos\delta}{pq}e^{i\frac{2\pi(p+q)}{\lambda}} \int_{\mathrm{A}} t(\boldsymbol{\alpha}) e^{i\frac{\pi}{\lambda p}(\boldsymbol{\alpha}-\boldsymbol{\epsilon})^2} e^{i\frac{\pi}{\lambda q}(\boldsymbol{\alpha}-\boldsymbol{x})^2} d\boldsymbol{\alpha} \tag{3.22}$$

$\delta$ is the angle between the line between the source point and the point in the diffraction pattern and the normal to the transparency. Since $p$ and $q$ are much larger than $\boldsymbol{\epsilon}$ and $\boldsymbol{x}$, the approximation $\cos\delta = 1$ can be used for all points in the source. Integration over all points in the source gives

$$u(\boldsymbol{x}) = \frac{1}{i\lambda}e^{i\frac{2\pi(p+q)}{\lambda}} \int \frac{s(\boldsymbol{\epsilon})}{p} \int \frac{t(\boldsymbol{\alpha})}{q} e^{i\frac{\pi}{\lambda p}(\boldsymbol{\alpha}-\boldsymbol{\epsilon})^2} e^{i\frac{\pi}{\lambda q}(\boldsymbol{\alpha}-\boldsymbol{x})^2} d\boldsymbol{\alpha} d\boldsymbol{\epsilon} \tag{3.23}$$

and this can be rewritten as

$$u(\boldsymbol{x}) = \frac{1}{i\lambda pq}e^{i\frac{2\pi(p+q)}{\lambda}} \left\{ \left[ \left[ s(\boldsymbol{x}) \otimes e^{i\frac{\pi \boldsymbol{x}^2}{\lambda p}} \right] \cdot t(\boldsymbol{x}) \right] \otimes e^{i\frac{\pi \boldsymbol{x}^2}{\lambda q}} \right\} \tag{3.24}$$

where $\otimes$ denotes a convolution. The amplitude of the light in the hologram plane is therefore given by the following three consecutive operations.

1. Convolution of the light source amplitude distribution and a function $F_1(\boldsymbol{x}) = e^{i\frac{\pi \boldsymbol{x}^2}{\lambda p}}$.
2. Multiplication of this result by the amplitude transmittance function of the transparency.
3. Convolution of the result with a second function $F_2(\boldsymbol{x}) = e^{i\frac{\pi \boldsymbol{x}^2}{\lambda q}}$.

This is illustrated in Figure 3.5.

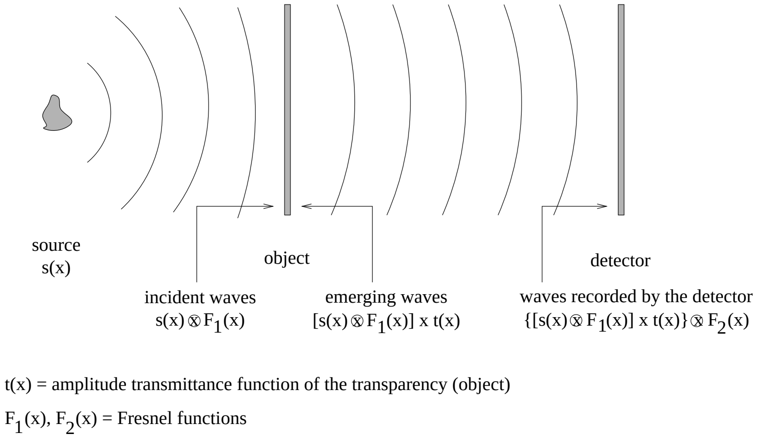


Figure 3.5: Three consecutive operations forming a general hologram

## 3.3 Gabor In-Line Holography (of a Two-Dimensional Object)

Gabor made a hologram of a transparency which had a clear background and opaque writing. The writing was the two-dimensional object of which the hologram was made. He illuminated the transparency with a collimated beam of monochromatic light and recorded its Fresnel diffraction pattern on a photographic plate [34]. The film was developed and the image reconstructed by shining the monochromatic light back through it. The writing was easily readable in the reconstructed image. Gabor won the Nobel Prize for this work which he published in a paper in 1949 [31]. The relationship between the transmittance of a developed hologram and the intensity of light the film was exposed to is given in appendix C.

In the case of Gabor in-line holography, the light source is collimated and so emits plane waves, therefore, $r = p$ and $s(\boldsymbol{\epsilon}) = 1$ over all of $\boldsymbol{\epsilon}$. The $\frac{1}{p}$ term can be replaced by

$A$, the amplitude of the monochromatic plane waves. Equation 3.23 then becomes

$$u(\boldsymbol{x}) = Ae^{i\frac{2\pi p}{\lambda}} \frac{1}{i\lambda q} e^{i\frac{2\pi q}{\lambda}} \int t(\boldsymbol{\alpha}) e^{i\frac{\pi}{\lambda q}(\boldsymbol{\alpha}-\boldsymbol{x})^2} d\boldsymbol{\alpha} \tag{3.25}$$

which can be written as

$$u(\boldsymbol{x}) = \frac{C}{2\pi i} \int t(\boldsymbol{\alpha}) 2\omega_0^2 e^{i\omega_0^2(\boldsymbol{\alpha}-\boldsymbol{x})^2} d\boldsymbol{\alpha} \quad \text{where} \quad C = Ae^{i\frac{2\pi(p+q)}{\lambda}} \quad \text{and} \quad \omega_0^2 = \frac{\pi}{\lambda q} \tag{3.26}$$

The amplitude of the light in the hologram plane is therefore given by the convolution of the amplitude transmittance function of the transparency with the Fresnel function $z(\boldsymbol{x})$.

$$u(\boldsymbol{x}) = \frac{C}{2\pi i} t(\boldsymbol{x}) \otimes z(\boldsymbol{x}) \qquad \text{where} \qquad z(\boldsymbol{x}) = 2\omega_0^2 e^{i\omega_0^2 \boldsymbol{x}^2} \tag{3.27}$$

The Fresnel function has the Fourier transform $Z^F(\boldsymbol{\omega})$ where

$$Z^F(\boldsymbol{\omega}) = 2\pi i e^{-i\frac{\boldsymbol{\omega}^2}{4\omega_0^2}} \tag{3.28}$$

and $\boldsymbol{\omega} = \boldsymbol{\omega}(\omega_x, \omega_y)$. Soroko showed that equation 3.27 can be written in four forms as follows.

$$\begin{aligned} u(\boldsymbol{x}) &= \frac{C}{2\pi i} t(\boldsymbol{x}) \otimes z(\boldsymbol{x}) && \text{first form} \\ u(\boldsymbol{x}) &= \frac{C}{i(2\pi)^2} \int T^F(\boldsymbol{\omega}) Z^F(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}\boldsymbol{x}} d\boldsymbol{\omega} && \text{second form} \\ u(\boldsymbol{x}) &= \frac{C}{4\pi^2} Z^{*F}(2\omega_0^2 \boldsymbol{x}) \int t(\boldsymbol{\alpha}) z(\boldsymbol{\alpha}) e^{-i(2\omega_0^2 \boldsymbol{x})\boldsymbol{\alpha}} d\boldsymbol{\alpha} && \text{third form} \\ u(\boldsymbol{x}) &= \frac{C}{4\pi^2} Z^{*F}(2\omega_0^2 \boldsymbol{x}) \left[T^F(2\omega_0^2 \boldsymbol{x}) \otimes Z^F(2\omega_0^2 \boldsymbol{x})\right] && \text{fourth form} \end{aligned} \tag{3.29}$$

where $T^F(\boldsymbol{\omega})$ is the Fourier transform of $t(\boldsymbol{x})$. These different expressions can be used to analyse the diffraction pattern. First, however, some characteristic lengths need to be introduced. The Fresnel function can be approximated as a series of concentric circles. A Fresnel zone plate can be used to diffract light and the size of the first Fresnel zone is a measure of diffraction. A Fresnel zone plate is a series of alternating transparent and opaque concentric rings which are called zones. Rays travelling to the focal point a distance $q$ behind the plate via a point on the outer edge of the $m$th zone will reach the focal point a out of phase by $\frac{m\lambda}{2}$ with a ray that passed through the centre of the plate. The radius of the $m$th zone, $R_m$, can be shown to be

$$R_m = \sqrt{m\lambda q} \; [36] \tag{3.30}$$

The radius of the central Fresnel zone, $l$, is therefore given by

$$l = \sqrt{\lambda q} \tag{3.31}$$

The size of the finest detail in the object, $\delta$, is related to the maximum spatial frequency, $\Omega_0$, in the spatial structure of the object by

$$\delta = \frac{1}{\Omega_0} \tag{3.32}$$

The following limiting cases will be considered.

1. The size of the object (opaque regions on the transparency) is much less than the size of the central Fresnel zone.
2. The size of the central Fresnel zone is much less than the size of the smallest detail in the object.
3. The size of the first Fresnel zone exceeds the size of the smallest detail in the object but this zone is less than the object itself.

### 3.3.1 The Central Fresnel Zone is much larger than the Object

If the size of the object is $a$, the following relationship can be used.

$$a \ll l \qquad \text{so} \qquad \omega_0^2 a^2 \ll 1 \qquad \text{and} \qquad z(\boldsymbol{\alpha}) \approx 2\omega_0^2 \tag{3.33}$$

The third representation of equation 3.29 then becomes

$$u(\boldsymbol{x}) = \frac{C}{4\pi^2} 2\omega_0^2 Z^{*F}(2\omega_0^2 \boldsymbol{x}) \int t(\boldsymbol{\alpha}) e^{i(2\omega_0^2 \boldsymbol{x})\boldsymbol{\alpha}} d\boldsymbol{\alpha} \tag{3.34}$$

which can be written as

$$u(\boldsymbol{x}) = C \frac{\omega_0^2}{2\pi^2} Z^{*F}(2\omega_0^2 \boldsymbol{x}) T^F(2\omega_0^2 \boldsymbol{x}) \tag{3.35}$$

and as $Z^{*F}(2\omega_0^2 \boldsymbol{x}) = -2\pi i e^{i\omega_0^2 \boldsymbol{x}^2}$, we have

$$u(\boldsymbol{x}) = \frac{C}{2\pi i} z(\boldsymbol{x}) T^F(2\omega_0^2 \boldsymbol{x}) \tag{3.36}$$

In this case, therefore, the diffraction pattern corresponds to the Fourier transform of the transmittance function of the transparency multiplied by the Fresnel function. Another, notably simpler method to obtain this formula is to first expand the terms in $(\boldsymbol{\alpha}-\boldsymbol{x})^2$ in equation 3.26 giving.

$$u(\boldsymbol{x})=\frac{C}{2\pi i}\int t(\boldsymbol{\alpha})2\omega_0^2 e^{i\omega_0^2\boldsymbol{x}^2}e^{-i2\omega_0^2\boldsymbol{\alpha}\boldsymbol{x}}e^{i\omega_0^2\boldsymbol{\alpha}^2}d\boldsymbol{\alpha} \tag{3.37}$$

Born and Wolf [14] state that when the size of the aperture/object is so small that quadratic and higher order terms in $\boldsymbol{\alpha}$ can be neglected. This is referred to as *Fraunhofer diffraction.* *Fresnel diffraction* describes the case when the quadratic terms cannot be neglected. Neglecting the quadratic terms in $\boldsymbol{\alpha}$ in equation 3.37 gives

$$u(\boldsymbol{x})=\frac{C}{2\pi i}2\omega_0^2 e^{i\omega_0^2\boldsymbol{x}^2}\int t(\boldsymbol{\alpha})e^{-i2\omega_0^2\boldsymbol{\alpha}\boldsymbol{x}}d\boldsymbol{\alpha} \tag{3.38}$$

It can be clearly seen that this is the same result as that in equation 3.34. Fraunhofer diffraction is also called *far-field diffraction* and the resulting hologram is called a *Fourier hologram.* Fourier holograms are mainly used to holographically store information and in digital optical computing [85].

### 3.3.2 The Central Fresnel Zone is much smaller than the Smallest Detail in the Object

The following relationship applies here.

$$l \ll \delta \qquad \text{or} \qquad \omega_0 \gg \Omega_0 \tag{3.39}$$

This means that the maximum frequency in $Z^F(\boldsymbol{\omega})$ is much smaller than $\omega_0$ and so $Z^F(\boldsymbol{\omega}) \approx 2\pi i$. The second representation of equation 3.29 can therefore be simplified.

$$u(\boldsymbol{x})=C\frac{2\pi i}{i(2\pi)^2}\int T^F(\boldsymbol{\omega})e^{i\boldsymbol{\omega}\boldsymbol{x}}d\boldsymbol{\omega}=\frac{C}{2\pi}\int T^F(\boldsymbol{\omega})e^{i\boldsymbol{\omega}\boldsymbol{x}}d\boldsymbol{\omega}=Ct(\boldsymbol{x}) \tag{3.40}$$

The diffraction pattern is simply the amplitude transmittance function of the transparency multiplied by a factor $C$ which depends upon the amplitude and phase of the light source. This is known as the geometric shadow region. Holograms recorded in this region are called *shadowgrams.* As will be discussed in Section 3.4, zone plate holograms are shadowgrams. Shadowgrams are also recorded in coded mask imaging techniques [59].

### 3.3.3 The Central Fresnel Zone is Larger than the Smallest Detail in the Object but Smaller than the Object Itself

The following relationship applies in this case

$$\delta < l < a \tag{3.41}$$

and equation 3.29 is still valid. This is the general Fresnel diffraction case. Gabor recorded his holograms in this region. As was stated in section 3.3.1 this is called *Fresnel diffraction* and is also called *near-field diffraction*. Holograms recorded in this region are called Fresnel holograms. Most conventional holograms used to record and display three-dimensional images are of this type.

Gabor also recorded his holograms in this region. The light source he used was a mercury arc lamp. This lamp produces green light of wavelength 546nm as well as blue light of wavelength 435nm [51]. If, for example, the green light was used, and the photographic plate was placed 1 metre behind the transparency, the size of the central Fresnel zone would be.

$$l = \sqrt{546 \times 10^{-9} \times 1} = 7.39 \times 10^{-4}\text{m} \approx 0.7\text{mm} \tag{3.42}$$

So the smallest detail in the object would have to be less than 0.7 mm and the object itself would have to be bigger than this. If, for example, the transparency was about 5cm square and had writing on it, it would satisfy these conditions. The exact parameters used for Gabor's experiment are not known here but, with this wavelength of light it can be seen that the hologram was probably in this region.

## 3.4 Zone Plate Holography of Two-Dimensional Sources

Figure 3.4 illustrates the equipment needed for gamma ray holography if the transparency was replaced with a sintered tungsten zone plate and the photographic film with a gamma camera with the collimator removed. The source would, of course, be a gamma ray source instead of a coherent light source. Gamma rays have frequencies between $10^{18}$ and $10^{22}$Hz [93]. This means that the longest wavelength would be $\lambda \approx \frac{10^8}{10^{18}} = 0.1\text{nm}$. If the gamma camera was 1 metre behind the zone plate, for example, the size of the central

Fresnel zone of the waves would be.

$$l = \sqrt{0.1 \times 10^{-9} \times 1} = 1 \times 10^{-5}\mathrm{m} = 10\mu\mathrm{m} \tag{3.43}$$

One possible application of this technique is to nuclear medicine. Here, the gamma ray sources would be several centimetres in size. So, even if the sources produced coherent rays, the wavelength of gamma rays would definitely put any experiment in the geometric shadow region as described in Section 3.3.2. The fact that gamma ray sources do not produce coherent waves further ensures that no diffraction is seen.

Each point in the source casts a shadow of the zone plate on the detector. The hologram is the sum of these shadows. First, a planar source will be considered. This source will be a distance $p$ behind the zone plate, as in Figure 3.4. It will be assumed that the shadow is the same size and shape for point sources at the same distance $p$ behind the zone plate, i.e. the shadow is shift invariant for points in the same plane. For this to be the case, the distances $p$ and $q$ have to be large compared with the size of the zone plate and the distance of the point source from the central axis of the imaging system (which passes through the centre of the zone plate and of the detector). The zone plate also must be assumed to be infinitely thin. Using the same notation as before, each shadow would be magnified by a factor $M$ where

$$M = 1 + \frac{q}{p} \tag{3.44}$$

A point source with intensity $S_i(\boldsymbol{\epsilon_i})$ at position $\boldsymbol{\epsilon_i}(\epsilon_i, \eta_i)$ will give a shadow, the centre of which is at $\boldsymbol{x_i}(x_i, y_i)$ where

$$x_i = -\left(\frac{q}{p}\right)\epsilon_i = L\epsilon_i \qquad\qquad y_i = -\left(\frac{q}{p}\right)\eta_i = L\eta_i \tag{3.45}$$

A definition of the intensity of a source of electromagnetic waves is given in appendix B. If the zone plate has an intensity transmittance function $T(\boldsymbol{\alpha})$, the detector is of size $D$ wide by $D$ high and $p, q \gg D$, the source will cast a shadow of intensity $U_i(\boldsymbol{x})$ given by.

$$U_i(x,y) = \frac{D^2}{4\pi(p+q)^2} S_i(\epsilon_i, \eta_i) T\left(\frac{x}{M} - L\epsilon_i, \frac{y}{M} - L\eta_i\right) \tag{3.46}$$

The factor $\frac{D^2}{4\pi(p+q)^2}$ is an approximation of the fraction of the radiation which is emitted by the isotropic source which is recorded by the detector. As $p, q \gg D$, the curvature of a sphere of radius $p + q$ over the detector area is small enough to be neglected.

If the extended source is seen as the sum of many point sources, the shadow, $U(\boldsymbol{x})$, cast by the extended source of intensity $S(\boldsymbol{\epsilon})$ is the integral over all these point sources.

$$U(\boldsymbol{x}) = \frac{D^2}{4\pi(p+q)^2} \int S\left(\boldsymbol{\epsilon}\right) T\left(\frac{\boldsymbol{x}}{M} - L\boldsymbol{\epsilon}\right) d\boldsymbol{\epsilon} \tag{3.47}$$

Two new variables, $\rho$ and $\sigma$ will now be introduced where $\rho = L\epsilon$ and $\sigma = L\eta$. Substituting $\boldsymbol{\rho}(\rho, \sigma) = L\boldsymbol{\epsilon}(\epsilon, \eta)$ into equation 3.47 gives

$$U(\boldsymbol{x}) = \frac{D^2}{4\pi(p+q)^2} \int S\left(\frac{\boldsymbol{\rho}}{L}\right) T\left(\frac{\boldsymbol{x}}{M} - \boldsymbol{\rho}\right) d\boldsymbol{\rho} \tag{3.48}$$

The resulting shadow is therefore a convolution between the source distribution magnified by a factor $L$ and the intensity transmittance of the zone plate magnified by a factor $M$.

$$U(\boldsymbol{x}) = \frac{D^2}{4\pi(p+q)^2} S\left(\frac{\boldsymbol{x}}{L}\right) \otimes T\left(\frac{\boldsymbol{x}}{M}\right) \tag{3.49}$$

Looking at equation 3.45, note that as $p$ and $q$ are both positive numbers, $L$ is negative. This means that the source distribution is flipped horizontally and vertically before it is scaled and convolved with the zone plate pattern.

## 3.5 Comparison of Gabor and Zone Plate Holograms of Single Points

Comparing equation 3.27 to 3.49, it can be seen that the amplitude of a Gabor hologram and the intensity of a zone plate hologram of two-dimensional sources can both be approximated as the convolution between two functions. For simplicity, the constant multiplication factors will be ignored here. The amplitude of a Gabor hologram is the convolution between the amplitude transmittance of the transparency (the object) and a Fresnel function whereas the intensity of a zone plate hologram is the convolution between the source distribution and the intensity transmittance of the zone plate. If the intensity transmittance of the zone plate was also a Fresnel function, the resulting zone plate hologram would be similar to a Gabor hologram of an equivalent coherent light source. As is stated in appendix C, intensity transmittance is the square of the amplitude transmittance.

If conventional holographic reconstruction techniques are to be used, the ideal zone plate to use would give a hologram which was as similar to a conventional hologram as possible. It would therefore have an intensity transmittance which was a Fresnel function.

$$T(\boldsymbol{x}) = 2\omega_{zp}^2 e^{i\omega_{zp}^2 \boldsymbol{x}^2} \tag{3.50}$$

Where $\omega_{zp}$ is a constant. This function, however, has both real and imaginary parts. The intensity transmittance function can only be real and have a value between 0 (opaque) and 1 (transparent). The best possible alternative is therefore

$$T(\boldsymbol{x}) = \frac{1}{2}\left[1 + \cos\left(\omega_{zp}^2 \boldsymbol{x}^2\right)\right] = \frac{1}{2}\left[1 + \Re\left\{e^{i\omega_{zp}^2 \boldsymbol{x}^2}\right\}\right] \tag{3.51}$$

where $\Re\{f\}$ is the real part of a function $f$. $[1 + \cos\left(\omega_{zp}\boldsymbol{x}^2\right)]$ has zeros at $\boldsymbol{x} = \frac{\sqrt{\pi}}{\omega}, \frac{\sqrt{3\pi}}{\omega}, \frac{\sqrt{5\pi}}{\omega} \ldots \frac{\sqrt{(2n+1)\pi}}{\omega} = r_0, \sqrt{3}r_0, \sqrt{5}r_0 \ldots \sqrt{2n+1}r_0$ and so the transmittance function can be written as

$$T(\boldsymbol{x}) = \frac{1}{2}\left[1 + \cos\left(\frac{\pi \boldsymbol{x}^2}{r_0^2}\right)\right] \tag{3.52}$$

a zone plate with this transmittance function is called a Gabor zone plate.

### 3.5.1 Zone Plate Hologram of a Point Gamma Ray Source

If the source is a point source of unit intensity at point $(\epsilon = 0, \eta = 0)$, the source distribution can be described by a delta function, $S(\boldsymbol{\epsilon}) = \delta(\boldsymbol{\epsilon})$. If a Gabor zone plate is used, $T(\boldsymbol{x}) = \frac{1}{2}\left[1 + \cos\left(\frac{\pi \boldsymbol{x}^2}{r_0^2}\right)\right]$ and equation 3.47 gives the intensity of the hologram as

$$U(\boldsymbol{x}) = \frac{D^2}{4\pi(p+q)^2}\frac{1}{2}\left[1 + \cos\left(\frac{\pi \boldsymbol{x}^2}{M^2 r_0^2}\right)\right] \tag{3.53}$$

This can be written as

$$U(\boldsymbol{x}) = \frac{D^2 M^2 r_0^2}{8\pi^2(p+q)^2}\frac{1}{2}\left[2\omega_{zp}^2 + 2\omega_{zp}^2 \cos(\omega_{zp}^2 \boldsymbol{x}^2)\right] \qquad \text{where} \qquad \omega_{zp} = \frac{\sqrt{\pi}}{M r_0} \tag{3.54}$$

A zone plate hologram of a point source therefore contains a constant background term minus the real part of a Fresnel function of spatial frequency $\omega_{zp}$.

### 3.5.2 Gabor Hologram of an Opaque Point in the Middle of an Otherwise Transparent Transparency

In equation 3.26, the amplitude of the hologram was given as

$$u(\boldsymbol{x}) = \frac{C}{2\pi i}\int t(\boldsymbol{\alpha}) 2\omega_0^2 e^{i\omega_0^2(\boldsymbol{\alpha}-\boldsymbol{x})^2} d\boldsymbol{\alpha} \quad \text{where} \quad C = Ae^{i\frac{2\pi(p+q)}{\lambda}} \quad \text{and} \quad \omega_0^2 = \frac{\pi}{\lambda q} \tag{3.55}$$

If the transparency were completely transparent, i.e. $t(\boldsymbol{\alpha}) = 1$ over all of $\boldsymbol{\alpha}$, equation 3.55 would become

$$u(\boldsymbol{x}) = \frac{C}{2\pi i} \int 2\omega_0^2 e^{i\omega_0^2(\boldsymbol{\alpha}-\boldsymbol{x})^2} d\boldsymbol{\alpha} = C = Ae^{i\frac{2\pi(p+q)}{\lambda}} \tag{3.56}$$

If the transparency contains a single opaque point at $\boldsymbol{\alpha} = 0$ the amplitude of the hologram would no longer contain the contribution from this point. The point will be assumed to be square and of size $\delta_p \times \delta_p$ and that $\delta_p$ is so small that the waves from one side of the source are the same as those from the other side. This gives

$$u(\boldsymbol{x}) = C\left[1 - \frac{\delta_p^2\omega_0^2}{\pi i} e^{i\omega_0^2\boldsymbol{x}^2}\right] \tag{3.57}$$

The intensity of the hologram is therefore

$$U(\boldsymbol{x}) = u(\boldsymbol{x})u^*(\boldsymbol{x}) = \frac{C^2\delta_p^2}{\pi}\left[\frac{\pi}{\delta_p^2} + \frac{\delta_p^2\omega_0^4}{\pi} - 2\omega_0^2\sin(\omega_0^2\boldsymbol{x}^2)\right] \tag{3.58}$$

A Gabor hologram of a point therefore contains a constant background term plus the imaginary part of a Fresnel function of spatial frequency $w_0$.

## 3.6 Optical/Convolution Image Reconstruction

Both types of hologram can be reconstructed in the way in which a Gabor hologram is formed. A Gabor hologram is already recorded on photographic film, the film then has to be developed. The resulting amplitude transmittance function of the film will be assumed to be linearly related to the intensity recorded. If $\Gamma$ is a constant which depends upon the way the film was developed, the amplitude transmittance function will be

$$t_G(\boldsymbol{\alpha}) = G_1 - G_2 2\omega_0^2\sin(\omega_0^2\boldsymbol{\alpha}^2) \quad \text{where}$$

$$G_1 = \Gamma C^2\left(1 + \frac{\delta_p^4\omega_0^4}{\pi^2}\right) \quad \text{and} \quad G_2 = \Gamma\frac{C^2\delta_p^2}{\pi} \tag{3.59}$$

The relationship between the amplitude transmittance of photographic film and the intensity of radiation the film was exposed to is discussed in appendix C. In zone plate holography, the hologram is recorded using a gamma camera. It is then stored as two-dimensional pixel array on a computer. This can be printed out onto a transparency. Again, it will be assumed that the resulting amplitude transmittance function is linearly related to the intensity recorded. $\Upsilon$ will be introduced as the constant of linearity.

$$t_Z(\boldsymbol{\alpha}) = Z_1 + Z_2 2\omega_{zp}^2 \cos(\omega_{zp}^2 \boldsymbol{\alpha}^2) \quad \text{where}$$
$$Z_1 = \Upsilon \frac{D^2}{8\pi(p+q)^2} \quad \text{and} \quad Z_2 = \Upsilon \frac{D^2 M^2 r_0^2}{16\pi^2(p+q)^2} \tag{3.60}$$

As was shown in Section 3.3, shining a beam of plane coherent light waves through a transparency gives a pattern in the hologram plane equivalent to the convolution between the amplitude transmittance of the transparency and a Fresnel function. The first form of equation 3.29 gives

$$u_G(\boldsymbol{x}) = \frac{C_r}{2\pi i} t_G(\boldsymbol{x}) \otimes 2\omega_r^2 e^{i\omega_r^2 \boldsymbol{x}^2} \quad \text{and} \quad u_Z(\boldsymbol{x}) = \frac{C_r}{2\pi i} t_Z(\boldsymbol{x}) \otimes 2\omega_r^2 e^{i\omega_r^2 \boldsymbol{x}^2} \tag{3.61}$$

where $C_r$ and $\omega_r$ are constants dependent on the wavelength of the light and the separation of the hologram and the detector.

$$\omega_r^2 = \frac{\pi}{\lambda_r q_r} \qquad C_r = A_r e^{i\frac{2\pi(p_r+q_r)}{\lambda_r}} \tag{3.62}$$

$\lambda_r$ is the wavelength of the light used to reconstruct the image, $q_r$ is the hologram to detector distance and $p_r$ is the source to hologram distance. As plane waves are used, $p_r$ just gives the phase of the waves as they reach the hologram and have no effect on the intensity of the reconstructed image. $A_r$ is the amplitude of the plane waves.

The convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of their individual Fourier transforms [68]. The Fresnel function of spatial frequency $\omega_r$ has the following Fourier transform.

$$Z^F(\boldsymbol{\omega}) = 2\pi i e^{-i\frac{\boldsymbol{\omega}^2}{4\omega_r^2}} \tag{3.63}$$

where $\boldsymbol{\omega} = \boldsymbol{\omega}(\omega_x, \omega_y)$. From this, the following Fourier transform pairs can be deduced.

$$\Re\left\{2\omega_r^2 e^{i\omega_r^2 \boldsymbol{x}^2}\right\} \Leftrightarrow \Re\left\{2\pi i e^{-i\frac{\boldsymbol{\omega}^2}{4\omega_r^2}}\right\} \quad \text{therefore} \quad 2\omega_r^2 \cos(\omega_r^2 \boldsymbol{x}^2) \Leftrightarrow 2\pi \sin\left(\frac{\boldsymbol{\omega}^2}{4\omega_r^2}\right)$$
$$\Im\left\{2\omega_r^2 e^{i\omega_r^2 \boldsymbol{x}^2}\right\} \Leftrightarrow \Im\left\{2\pi i e^{-i\frac{\boldsymbol{\omega}^2}{4\omega_r^2}}\right\} \quad \text{therefore} \quad 2\omega_r^2 \sin(\omega_r^2 \boldsymbol{x}^2) \Leftrightarrow 2\pi \cos\left(\frac{\boldsymbol{\omega}^2}{4\omega_r^2}\right) \tag{3.64}$$

where $\Re\{f\}$ is the real part and $\Im\{f\}$ the imaginary part of a function $f$.

To summarise, expressions for the amplitude transmittance of both Gabor and zone plate holograms of on-axis point objects have been derived. It has been shown that, if the image is reconstructed by shining light back through the hologram, the image seen at a given distance behind the transparency is given by the convolution of the amplitude transmittance function of the hologram with a Fresnel function $z(\boldsymbol{x}) = 2\omega_r^2 e^{i\omega_r^2\boldsymbol{x}^2}$.

The Fourier transform of the convolution of two functions is equal to the product of their individual Fourier transforms. The Fourier transforms of the real and imaginary parts of a Fresnel function have been found and can be used to evaluate the reconstructed image from both types of hologram. Expressions for the reconstructed image will be found first when $\omega_r = \omega_0$ and $\omega_r = \omega_{zp}$ and then when $\omega_r \neq \omega_0$ and $\omega_r \neq \omega_{zp}$. In the first case, which will be called matched convolution, an image of the point should be obtained. In the second case, unmatched convolution, there should be no discernible image.

### 3.6.1 Matched Convolution

**Gabor Hologram:** $\omega_r = \omega_0$

First, a Gabor hologram of a point will be investigated. It will be assumed that the hologram stretches to ±∞ in the $x$ and $y$ directions. The constant term will be ignored as this will just give a constant term in the reconstructed image. An expression will be found for the contribution to the reconstructed image from the sin term when $\omega_r = \omega_0$. The aim is to find the amplitude of the reconstructed image, $u_{G2}(\boldsymbol{x})$ .

$$u_{G2}(\boldsymbol{x}) = C_g 2\omega_0^2 \sin(\omega_0^2\boldsymbol{x}^2) \otimes 2\omega_0^2 e^{i\omega_0^2\boldsymbol{x}^2} \quad \text{where} \quad C_g = -\frac{C_r}{2\pi i}G_2 \tag{3.65}$$

Taking the Fourier transform of each side gives

$$\begin{aligned} U_{G2}^F(\boldsymbol{\omega}) &= C_g(2\pi)^2\left[\cos\left(\frac{\boldsymbol{\omega}^2}{4\omega_0^2}\right)\sin\left(\frac{\boldsymbol{\omega}^2}{4\omega_0^2}\right) + i\cos^2\left(\frac{\boldsymbol{\omega}^2}{4\omega_0^2}\right)\right] \\ &= C_g(2\pi)^2\left[\frac{1}{2}\sin\left(\frac{\boldsymbol{\omega}^2}{2\omega_0^2}\right) + i\frac{1}{2}\left\{1+\cos\left(\frac{\boldsymbol{\omega}^2}{2\omega_0^2}\right)\right\}\right] \end{aligned} \tag{3.66}$$

then, taking the inverse Fourier transform gives the following expression for the amplitude of the reconstructed image.

$$u_{G2}(\boldsymbol{x}) = C_g\left[\pi\omega_0^2\cos\left(\frac{\omega_0^2}{2}\boldsymbol{x}^2\right) + i\left\{\frac{1}{2}\delta(0) + \pi\omega_0^2\sin\left(\frac{\omega_0^2}{2}\boldsymbol{x}^2\right)\right\}\right] \tag{3.67}$$

where $\delta(\boldsymbol{x})$ is the Dirac delta function.

The imaginary part of the reconstructed image therefore contains a point at $\boldsymbol{x} = \mathbf{0}$ as well as the imaginary part of a Fresnel function of frequency $\frac{\omega_0}{\sqrt{2}}$ . The real part just contains a real part of a Fresnel function, also of frequency $\frac{\omega_0}{\sqrt{2}}$. If an optical reconstruction is used, only the intensity will be recorded. This is given by

$$U_{G2}(\boldsymbol{x}) = C_g^2\left[(\pi\omega_0^2)^2\cos^2\left(\frac{\omega_0^2}{2}\boldsymbol{x}^2\right) + \left\{\frac{1}{2}\delta(\mathbf{0}) + \pi\omega_0^2\sin\left(\frac{\omega_0^2}{2}\boldsymbol{x}^2\right)\right\}^2\right] =$$

$$C_g^2\left[(\pi\omega_0^2)^2 + \left(\frac{\delta(\mathbf{0})}{4}\right)^2\right] \quad (3.68)$$

The intensity of the reconstructed image therefore contains just a point at $\boldsymbol{x} = \mathbf{0}$ as well as a constant background term, as would be expected.

### Zone Plate Hologram: $\omega_r = \omega_{zp}$

The analysis here is very similar to that for the Gabor hologram. Again, the constant background term will be ignored. The aim is to find the amplitude of the reconstructed image, $u_{Z2}(\boldsymbol{x})$.

$$u_{Z2}(\boldsymbol{x}) = C_z 2\omega_{zp}^2\cos(\omega_{zp}^2\boldsymbol{x}^2) \otimes 2\omega_{zp}^2 e^{i\omega_{zp}^2\boldsymbol{x}^2} \quad \text{where} \quad C_z = \frac{C_r}{2\pi i}Z_2 \quad (3.69)$$

Taking the Fourier transform of each side gives

$$U_{Z2}^F(\boldsymbol{\omega}) = C_z(2\pi)^2\left[\sin^2\left(\frac{\boldsymbol{\omega}^2}{4\omega_{zp}^2}\right) + i\sin\left(\frac{\boldsymbol{\omega}^2}{4\omega_{zp}^2}\right)\cos\left(\frac{\boldsymbol{\omega}^2}{4\omega_{zp}^2}\right)\right]$$

$$= C_z(2\pi)^2\left[\frac{1}{2}\left\{1 - \cos\left(\frac{\boldsymbol{\omega}^2}{2\omega_{zp}^2}\right)\right\} + i\frac{1}{2}\sin\left(\frac{\boldsymbol{\omega}^2}{2\omega_{zp}^2}\right)\right] \quad (3.70)$$

then, taking the inverse Fourier transform gives the following expression for the amplitude of the reconstructed image.

$$u_{Z2}(\boldsymbol{x}) = C_z\left[\left\{\frac{1}{2}\delta(\mathbf{0}) - \pi\omega_{zp}^2\sin\left(\frac{\omega_{zp}^2}{2}\boldsymbol{x}^2\right)\right\} + i\pi\omega_{zp}^2\cos\left(\frac{\omega_{zp}^2}{2}\boldsymbol{x}^2\right)\right] \quad (3.71)$$

The intensity is therefore given by

$$U_{Z2}(\boldsymbol{x}) =$$

$$C_z^2\left[\left\{\frac{1}{2}\delta(\mathbf{0}) - \pi\omega_{zp}^2\sin\left(\frac{\omega_{zp}^2}{2}\boldsymbol{x}^2\right)\right\}^2 + (\pi\omega_{zp}^2)^2\cos^2\left(\frac{\omega_{zp}^2}{2}\boldsymbol{x}^2\right)\right] =$$

$$C_z^2\left[(\pi\omega_{zp}^2)^2 + \left(\frac{\delta(\mathbf{0})}{4}\right)^2\right] \quad (3.72)$$

Again, the intensity of the reconstructed image therefore contains just a point at $\boldsymbol{x} = \mathbf{0}$ as well as a constant background term.

### 3.6.2 Unmatched Convolution

**Gabor Hologram:** $\omega_r \neq \omega_0$

The frequency of the Fresnel function in a plane a distance $q_r$ behind the transparency is given by $\omega_r = \sqrt{\frac{\pi}{\lambda q_r}}$. As discussed in the previous Section, when $\omega_r = \omega_0$ the reconstructed image has a point in the correct position and a constant background. Here, a formula will be derived for the reconstructed image when $\omega_r \neq \omega_0$, that is, in a plane other than the focal plane. As in Section 3.6.1, the aim is to calculate the amplitude of the reconstructed image, $u_{G2}(\boldsymbol{x})$.

$$u_{G2}(\boldsymbol{x}) = C_g 2\omega_0^2 \sin(\omega_0^2 \boldsymbol{x}^2) \otimes 2\omega_r^2 e^{i\omega_r^2 \boldsymbol{x}^2} \tag{3.73}$$

Taking Fourier transforms of both sides gives

$$\begin{aligned} U^F_{G2}(\boldsymbol{\omega}) &= C_g (2\pi)^2 \left[ \cos\left(\frac{\boldsymbol{\omega}^2}{4\omega_r^2}\right) \sin\left(\frac{\boldsymbol{\omega}^2}{4\omega_0^2}\right) + i\cos\left(\frac{\boldsymbol{\omega}^2}{4\omega_r^2}\right) \cos\left(\frac{\boldsymbol{\omega}^2}{4\omega_0^2}\right) \right] \\ &= C_g 2\pi^2 \Bigg[ \sin\left(\left\{\frac{1}{4\omega_0^2} + \frac{1}{4\omega_r^2}\right\}\boldsymbol{\omega}^2\right) - \sin\left(\left\{\frac{1}{4\omega_0^2} - \frac{1}{4\omega_r^2}\right\}\boldsymbol{\omega}^2\right) + \\ &\qquad i\left\{ \cos\left(\left\{\frac{1}{4\omega_0^2} + \frac{1}{4\omega_r^2}\right\}\boldsymbol{\omega}^2\right) + \cos\left(\left\{\frac{1}{4\omega_0^2} - \frac{1}{4\omega_r^2}\right\}\boldsymbol{\omega}^2\right) \right\} \Bigg] \end{aligned} \tag{3.74}$$

Taking the inverse Fourier transform gives the following expression for the amplitude of the reconstructed image

$$\begin{aligned} &u_{G2}(\boldsymbol{x}) = \\ &C_g \Bigg[ 2\pi \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2+\omega_0^2}\right) \cos\left(\left\{\frac{\omega_r^2\omega_0^2}{\omega_r^2+\omega_0^2}\right\}\boldsymbol{x}^2\right) - 2\pi \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2-\omega_0^2}\right) \cos\left(\left\{\frac{\omega_r^2\omega_0^2}{\omega_r^2-\omega_0^2}\right\}\boldsymbol{x}^2\right) + \\ &i\left\{ 2\pi \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2+\omega_0^2}\right) \sin\left(\left\{\frac{\omega_r^2\omega_0^2}{\omega_r^2+\omega_0^2}\right\}\boldsymbol{x}^2\right) + 2\pi \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2-\omega_0^2}\right) \sin\left(\left\{\frac{\omega_r^2\omega_0^2}{\omega_r^2-\omega_0^2}\right\}\boldsymbol{x}^2\right) \right\} \Bigg] \\ &= 2\pi C_g \left[ a_g e^{i a_g \boldsymbol{x}^2} - b_g e^{-i b_g \boldsymbol{x}^2} \right] \quad \text{where} \quad a_g = \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2+\omega_0^2}\right) \quad b_g = \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2-\omega_0^2}\right) \end{aligned} \tag{3.75}$$

The intensity of the image is therefore given by

$$\begin{aligned} &U_{G2} = (2\pi C_g)^2 \left[ a_g^2 + b_g^2 - 2a_g b_g \cos\left((a_g+b_g)\boldsymbol{x}^2\right) \right] = \\ &(2\pi C_g)^2 \left[ \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2+\omega_0^2}\right)^2 + \left(\frac{\omega_r^2\omega_0^2}{\omega_r^2-\omega_0^2}\right)^2 - 2\left(\frac{\omega_r^4\omega_0^4}{\omega_r^4-\omega_0^4}\right) \cos\left(\left\{\frac{2\omega_r^2\omega_0^4}{\omega_r^4-\omega_0^4}\right\}\boldsymbol{x}^2\right) \right] \end{aligned} \tag{3.76}$$

which, as expected, has a singularity when $\omega_r = \omega_0$.

### Zone Plate Hologram: $\omega_r \neq \omega_{zp}$

The same method as before gives the following expression for the amplitude of the reconstructed image, $u_{Z2}(\boldsymbol{x})$.

$$
\begin{aligned}
&u_{Z2}(\boldsymbol{x}) = \\
&C_z \left[ 2\pi \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 - \omega_{zp}^2} \right) \sin \left( \left\{ \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 - \omega_{zp}^2} \right\} \boldsymbol{x}^2 \right) - 2\pi \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 + \omega_{zp}^2} \right) \sin \left( \left\{ \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 + \omega_{zp}^2} \right\} \boldsymbol{x}^2 \right) + \right. \\
&\left. i \left\{ 2\pi \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 - \omega_{zp}^2} \right) \cos \left( \left\{ \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 - \omega_{zp}^2} \right\} \boldsymbol{x}^2 \right) + 2\pi \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 + \omega_{zp}^2} \right) \cos \left( \left\{ \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 + \omega_{zp}^2} \right\} \boldsymbol{x}^2 \right) \right\} \right] \\
&= 2\pi C_z \left[ i a_z e^{i a_z \boldsymbol{x}^2} + i b_z e^{-i b_z \boldsymbol{x}^2} \right] \quad \text{where} \quad a_z = \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 + \omega_{zp}^2} \right) \quad b_z = \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 - \omega_{zp}^2} \right) \qquad (3.77)
\end{aligned}
$$

and the intensity is therefore

$$
\begin{aligned}
&U_{G2} = (2\pi C_z)^2 \left[ a_z^2 + b_z^2 + 2 a_z b_z \cos(((a_z + b_z)\boldsymbol{x}^2) \right] = \\
&(2\pi C_z)^2 \left[ \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 + \omega_{zp}^2} \right)^2 + \left( \frac{\omega_r^2 \omega_{zp}^2}{\omega_r^2 - \omega_{zp}^2} \right)^2 + 2 \left( \frac{\omega_r^4 \omega_{zp}^4}{\omega_r^4 - \omega_{zp}^4} \right) \cos \left( \left\{ \frac{2\omega_r^4 \omega_{zp}^2}{\omega_r^4 - \omega_{zp}^4} \right\} \boldsymbol{x}^2 \right) \right] \qquad (3.78)
\end{aligned}
$$

which again has a singularity when $\omega_{zp} = \omega_r$.

To compare the variation in intensity of the reconstructed image with hologram to detector image, an example Gabor hologram will be analysed. The hologram will be assumed to have been recorded with a transparency to photographic film separation of $q$ =1m. The wavelength of light will be set as $\lambda = 546$nm which is the wavelength of light emitted by the mercury arc lamp Gabor used in his experiments. These two values give $\omega_0 = 2399\text{m}^{-1}$. The wavelength of light is used in the reconstruction will also be set to 546nm. The constant $C_g$ will be set to $C_g = 10^{-6}$. To compare a zone plate hologram to this, the parameters $\omega_{zp}$ and $C_z$ will be set equal to the equivalent parameters for the Gabor hologram, that is, $\omega_{zp} = \omega_0$ and $C_z = C_g$. These may be artificial values for a zone plate hologram, but the recorded hologram would have to be reduced in size to diffract light if the image were to be reconstructed optically. This would give the same values for the constants $a$ and $b$ in both cases, i.e: $a = a_g = a_z$ and $b = b_g = b_z$.

Figure 3.6 shows the variation of the constants $a$ and $b$ as the hologram to detector distance varies from 0m to 2m in intervals of 0.01m. $b$ tends to infinity as $q_r \rightarrow$1m.

Figure 3.7 shows how the intensity at $\boldsymbol{x} = \boldsymbol{0}$ varies with hologram to detector distance for both types of hologram. The intensity tends to infinity in both cases as $q_r \rightarrow$1m.

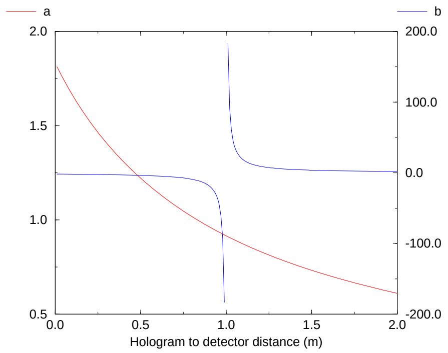
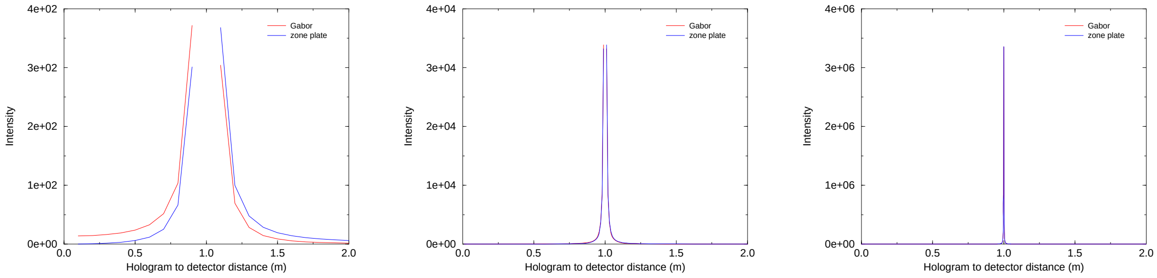


Figure 3.6: Variation in the values of $a$ and $b$ with hologram to detector distance



(a) interval = 0.1m (b) interval = 0.01m (c) interval = 0.001m

Figure 3.7: Variation of on-axis intensity with hologram to detector distance for ideal Gabor and zone plate holograms of a single point

Values have been calculated at intervals of 0.1m (Figure 3.7(a)), 0.01m (Figure 3.7(b)) and 0.001m (Figure 3.7(c)). The largest values of intensity are therefore at 0.9m and 1.1m, 0.99m and 1.01m, and 0.999m and 1.001m respectively. A real hologram does not stretch to infinity in the $x$ and $y$ directions, however, so the intensity in the focal plane would not be infinity in this case.

## 3.7 Comparison of Holograms of more than One Point

### 3.7.1 Zone Plate Hologram of two Point Gamma Ray Sources

Referring back to equations 3.47 and 3.52, a Gabor zone plate will give a hologram with the following intensity.

$$U(\boldsymbol{x}) = \frac{D^2 M^2 r_0^2}{8\pi^2 (p+q)^2} \int S(\boldsymbol{\epsilon}) \frac{1}{2} \left[ 2\omega_{zp}^2 + 2\omega_{zp}^2 \cos\left(\omega_{zp}^2 (\boldsymbol{x} - LM\boldsymbol{\epsilon})^2\right)\right] d\boldsymbol{\epsilon} \tag{3.79}$$

A hologram of two point gamma-ray sources at $\boldsymbol{\epsilon} = \boldsymbol{o_1}$ and $\boldsymbol{\epsilon} = \boldsymbol{o_2}$ and both at the same distance $p$ from the zone plate will be given by

$$U(\boldsymbol{x}) = \frac{D^2M^2r_0^2}{8\pi^2(p+q)^2}\frac{1}{2}\left[4\omega_{zp}^2 + 2\omega_{zp}^2\cos\left(\omega_{zp}^2\left(\boldsymbol{x}-LM\boldsymbol{o_1}\right)^2\right) + 2\omega_{zp}^2\cos\left(\omega_{zp}^2\left(\boldsymbol{x}-LM\boldsymbol{o_2}\right)^2\right)\right] \tag{3.80}$$

If this hologram is again printed out onto a transparency, its transmittance function will be

$$t_Z(\boldsymbol{\alpha}) = 2Z_1 + Z_2 2\omega_{zp}^2\cos\left(\omega_{zp}^2\left(\boldsymbol{\alpha}-LM\boldsymbol{o_1}\right)^2\right) + Z_2 2\omega_{zp}^2\cos\left(\omega_{zp}^2\left(\boldsymbol{\alpha}-LM\boldsymbol{o_2}\right)^2\right) \tag{3.81}$$

where $Z_1$ and $Z_2$ are as in equation 3.60. This is simply the superposition of holograms from each individual point.

### 3.7.2 Gabor Hologram of two Opaque Points

Referring back to Section 3.5.2, a hologram of two opaque points at $\boldsymbol{\alpha} = \boldsymbol{o_1}$ and $\boldsymbol{\alpha} = \boldsymbol{o_2}$ in an otherwise transparent transparency would have the following amplitude

$$u(\boldsymbol{x}) = C\left[1 - \frac{\delta_p^2\omega_0^2}{\pi i}e^{i\omega_0^2(\boldsymbol{x}-\boldsymbol{o_1})^2} - \frac{\delta_p^2\omega_0^2}{\pi i}e^{i\omega_0^2(\boldsymbol{x}-\boldsymbol{o_2})^2}\right] \tag{3.82}$$

where $\delta_p^2$ is the size of each point as before. The corresponding intensity would be

$$\begin{aligned} U(\boldsymbol{x}) = \frac{C^2}{\pi}\Bigg[\pi + \frac{2\delta_p^4\omega_0^4}{\pi} - 2\delta_p^2\omega_0^2\sin\left(\omega_0^2\left(\boldsymbol{x}-\boldsymbol{o_1}\right)^2\right) - 2\delta_p^2\omega_0^2\sin\left(\omega_0^2\left(\boldsymbol{x}-\boldsymbol{o_2}\right)^2\right) \\ + \frac{2\delta_p^4\omega_0^4}{\pi}\cos\left(\omega_0^2\left\{\left(\boldsymbol{x}-\boldsymbol{o_1}\right)^2 - \left(\boldsymbol{x}-\boldsymbol{o_2}\right)^2\right\}\right)\Bigg] \end{aligned} \tag{3.83}$$

The developed film would therefore have the following transmittance function

$$\begin{aligned} t_G(\boldsymbol{\alpha}) = 2G_1 - G_2 2\omega_0^2\sin\left(\omega_0^2\left(\boldsymbol{\alpha}-\boldsymbol{o_1}\right)^2\right) - G_2 2\omega_0^2\sin\left(\omega_0^2\left(\boldsymbol{\alpha}-\boldsymbol{o_2}\right)^2\right) + \\ G_3 2\omega_0^2\cos\left(\omega_0^2\left\{-2\left(\boldsymbol{o_1}-\boldsymbol{o_2}\right)\boldsymbol{x} + \boldsymbol{o_1}^2 - \boldsymbol{o_2}^2\right\}\right) - \Gamma C^2 \qquad \text{where} \qquad G_3 = \Gamma\frac{C^2\delta_p^4\omega_0^2}{\pi^2} \end{aligned} \tag{3.84}$$

$G_1$ and $G_2$ are as in equation 3.59. The second and third terms correspond to holograms of the individual points but the fourth term comes from interference between the contributions from both points. The relative magnitude of this contribution compared to those from the individual points is given by

$$F = \frac{G_3}{G_2} = \frac{\Gamma C^2\delta_p^4\omega_0^2}{\pi^2}\frac{\pi}{\Gamma C^2\delta_p^2} = \frac{\delta_p^2\omega_0^2}{\pi} = \frac{\delta_p^2}{\lambda q} \tag{3.85}$$

If, as in Section 3.3.3, the wavelength of the light is set to $\lambda = 546$nm and the distance of the image behind the hologram to $q = 1$ m, this gives $F = 1.8\times10^6\delta_p^2$. Therefore, as long as each point has $\delta_p < 0.7$mm, this contribution will be smaller than that from the individual points.

The interference term has a cosine variation. The frequency of this variation is given by $2\omega_0^2(\boldsymbol{o_1} - \boldsymbol{o_2})$ m$^{-1}$. Using the same value for $\omega_0$, this gives $3.7\times10^6(\boldsymbol{o_1} - \boldsymbol{o_2})$m$^{-1}$. So, if $(\boldsymbol{o_1} - \boldsymbol{o_2}) = 0.01$m, for example, zeros in this term will be separated by 86$\mu$m. The size of the central Fresnel zone ($l = \sqrt{\lambda q}$) for the terms from individual points will be 740$\mu$m.

Unlike a zone plate hologram, which is simply the superposition of the holograms from each point in the source, a Gabor hologram also contains terms from the mutual interference of waves from different points in the source. This Section showed the terms for two points but holograms of more than two points or extended objects will contain further terms. The magnitude and frequency of these contributions compared to the magnitudes of the Fresnel functions from the individual points will depend upon the separation of the points as well as the frequency of light used and the distance of the transparency from the photographic plate.

Investigations could be made to determine whether adding some of the mutual interference terms to a zone plate hologram to make it more like an equivalent conventional hologram would improve the image reconstruction. As these terms depend upon the object of which the hologram was made, this would have to be an iterative procedure. A first approximation of the object would have to be found initially, the corresponding mutual interference terms could then be calculated and used to improve the image. This process could then be repeated until the image quality reached an acceptable level.

## 3.8 Hologram of a Three-Dimensional Object

### 3.8.1 Zone Plate Holography

The factors $M$ and $L$ which were given in equations 3.44 and 3.45 depend upon the source to zone plate distance $p$ as well as the fixed zone plate to detector distance $q$. A zone plate hologram of a source extended in the $z$ direction as well as in directions perpendicular to this axis, $(\epsilon, \eta)$, will have varying values of $M$ and $L$. Instead of $M$ and $L$, $M(p)$ and

$L(p)$ will be used as these values are functions of $p$.

An extra complication is that a gamma-ray from a point which is not on the surface nearest to the zone plate will have to travel through the rest of the source before travelling through the air to the zone plate. The attenuation of this gamma-ray by the source must therefore be considered. The linear attenuation coefficient, $\mu_l$ for gamma-rays is defined by

$$I = I_0 e^{-\mu_l t} \tag{3.86}$$

where $I$ is the number of transmitted photons, $I_0$ is the number without an absorber and $t$ is the thickness of the absorber. This can also be given as the mass attenuation coefficient $\mu_m$ or the mean free path $\lambda_{mf}$ where

$$\mu_m = \frac{\mu_l}{\rho} \quad \text{and} \quad \lambda_{mf} = \frac{1}{\mu_l} \tag{3.87}$$

and $\rho$ is the density of the medium. Intensity of radiation is the rate of transfer of energy across unit areas by the radiation. For a monoenergetic source, this is proportional to the number of photons being emitted per second. $I$ in equation 3.86 can therefore be replaced by $s(\epsilon)_a$, the intensity after travelling through the source and $I_0$ by $s(\epsilon)_b$, the intensity before travelling through the source.

One commonly used radiopharmaceutical is $^{99m}$Tc (140keV gamma-rays) [75]. The effect of this attenuation can be estimated using this example. $^{99m}$Tc is usually obtained from a Mo/Tc generator as a salt dissolved in water. In nuclear medicine, this is then injected into a person for imaging purposes. Mass attenuation coefficients [50] for 140keV gamma-rays in tissue as well as in water are therefore given in Table 3.1. The density of water is 1000kgm$^{-3}$, of air 1.29kgm$^{-3}$ [89] [21] and the density of tissue will be assumed to be the same as that of water. Linear attenuation coefficients were calculated using these values and can be seen in Table 3.2.

| Radioisotope<br>Gamma-ray energy | $^{99m}$Tc<br>140keV |
|---|---|
| Air | 0.0134 |
| $H_2O$ | 0.0149 |
| Tissue | 0.0144 |

Table 3.1: Mass attenuation coefficients (m$^2$kg$^{-1}$)

| Radioisotope | $^{99\mathrm{m}}$Tc |
|---|---|
| Gamma-ray energy | 140keV |
| Air | 0.0173 |
| $H_2O$ | 14.9 |
| Tissue | 14.4 |

Table 3.2: Linear attenuation coefficients ($\mathrm{m}^{-1}$)

If a phantom contains $^{99\mathrm{m}}$Tc and is 0.01m thick, for example, 140keV gamma-rays will be attenuated to a factor $e^{-14.9\times0.01} = 0.86$ times the original intensity in travelling from the back to the front of the phantom. This means that 14% of the gamma-ray intensity will be lost. If the phantom is 5cm thick, 53% is lost. If 140keV gamma-rays travel through 1m of air, only 1.7% will be lost.

An expression for the resulting hologram is given by equation 3.47 but integrated over all distances $p$ in the $z$-direction of sources from the zone plate. The attenuation of gamma-rays which pass through the source also need to be considered and this will be represented by a factor $\gamma(\epsilon, \eta, p, S(\epsilon, \eta, p))$ which depends upon the entire source distribution, $S(\epsilon, \eta, p)$, as well as the position of the point, $(\epsilon, \eta, p)$.

$$U(\boldsymbol{x}) = \frac{D^2}{4\pi} \int \frac{1}{(p+q)^2} \int \gamma(\boldsymbol{\epsilon}, p, S(\boldsymbol{\epsilon}, p)) S\left(\boldsymbol{\epsilon}, p\right) T\left(\frac{\boldsymbol{x}}{M(p)} - L(p)\boldsymbol{\epsilon}\right) d\boldsymbol{\epsilon} dp \tag{3.88}$$

The attenuation term adds an extra degree of complexity to the problem of reconstructing an image of the source. It implies that the whole of the source distribution needs to be known before an expression for the contribution from one point in the source can be found. In practice, an image of the source could be calculated first, neglecting attenuation, so the location of the source could be found. The attenuation could be included in a second stage to correct the amplitude of the image. This approach is often used in SPECT which is a well established imaging technique for nuclear medicine [75].

Another factor to consider is the scaling of images from zone plate holograms in the $z$ direction if they are reconstructed optically. As was shown in Section 3.6.2, $\omega_r$ should be set equal to $\omega_{zp}$ for an image of a point to be obtained. If a Gabor zone plate is used $\omega_r$ and $\omega_{zp}$ are given by

$$\omega_r = \sqrt{\frac{\pi}{\lambda_r q_r}} \quad \text{and} \quad \omega_{zp} = \frac{p\sqrt{\pi}}{(p+q)r_0} \tag{3.89}$$

where $\lambda_r$ is the wavelength of the light used in the reconstruction, $r_0$ is the first zone radius of the zone plate and $q_r$ is the point where the reconstructed image will be found. So, when $\omega_r = \omega_{zp}$, $q_r$ is given by

$$q_r = \frac{(p+q)^2 r_0^2}{\lambda_r p^2} \tag{3.90}$$

$q_r$ is therefore not a linear function of $q$. This will produce an undesirable scaling effect in the $z$ direction in the reconstructed image.

### 3.8.2 Gabor In-Line Holography

Gabor used the experimental arrangement described in Section 3.3 to make holograms of two-dimensional objects. In order to compare this technique to zone plate holography for three-dimensional imaging, a series of transparencies will have to be used as the model of the object to be imaged. The transparencies will be assumed not to refract or attenuate the light waves travelling through it and the opaque regions will completely absorb any light incident on them.

The hologram from a series of $N$ thin transparencies will be considered. Each transparency will be separated a distance $\Delta q$ from its neighbours with the closest one a distance $q_{min}$ from a sheet of photographic film and the furthest one at a distance $q_{max} = q_{min} + \Delta q(N-1)$ . The transparencies are illuminated by coherent plane waves. This is illustrated in Figure 3.8. In Section 3.2, the amplitude of the light in the hologram plane for a general hologram was shown to be given by the convolution of the light source distribution with one Fresnel function, multiplication of this by the amplitude transmittance function of the transparency and then convolution of this with a second Fresnel function. Consider the case when $N$=2, i.e. there are only two transparencies. If the amplitude of the plane waves at the first transparency is given by $Ae^{i\frac{2\pi p}{\lambda}}$ and the waves emerging from the first transparency are seen as a source for a hologram of the second transparency, referring back to equation 3.24, the amplitude of the light in the hologram plane will be given by

$$u(\boldsymbol{x}) = Ae^{i\frac{2\pi p}{\lambda}} \frac{1}{i\lambda\Delta q q_{min}} e^{i\frac{2\pi(\Delta q + q_{min})}{\lambda}} \left\{ \left[ \left[ t(\boldsymbol{x}, q_{min} + \Delta q) \otimes e^{i\frac{\pi \boldsymbol{x}^2}{\lambda \Delta q}} \right] \cdot t(\boldsymbol{x}, q_{min}) \right] \otimes e^{i\frac{\pi \boldsymbol{x}^2}{\lambda q_{min}}} \right\} \tag{3.91}$$

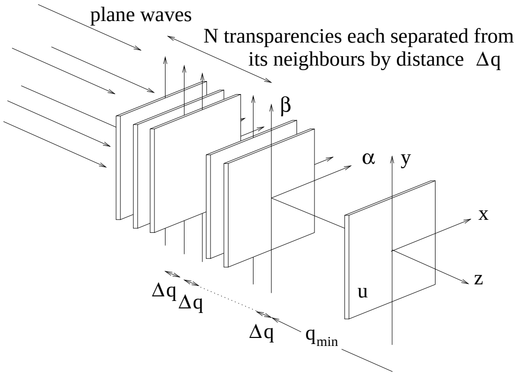


Figure 3.8: Relative positions of the transparencies and the hologram

For $N$ transparencies, this becomes

$$u(\boldsymbol{x}) = Ae^{i\frac{2\pi p}{\lambda}}\frac{1}{i\lambda}\frac{1}{q_{min}(\Delta q)^{N-1}}e^{i\frac{2\pi(q_{min}+\Delta q(N-1))}{\lambda}}\Bigg\{\Bigg[\Bigg[\Bigg[\Bigg[\cdots\Bigg[\Bigg[\Bigg[\Bigg[\Bigg[t(\boldsymbol{x}, q_{min}+\Delta q(N-1))\otimes$$

$$e^{i\frac{\pi\boldsymbol{x}^2}{\lambda\Delta q}}\Bigg]\cdot t(\boldsymbol{x}, q_{min}+\Delta q(N-2))\Bigg]\otimes e^{i\frac{\pi\boldsymbol{x}^2}{\lambda\Delta q}}\Bigg]\cdot t(\boldsymbol{x}, q_{min}+\Delta q(N-3))\Bigg]\otimes e^{i\frac{\pi\boldsymbol{x}^2}{\lambda\Delta q}}\Bigg]\cdots$$

$$\cdot\, t(\boldsymbol{x}, q_{min}+\Delta q)\Bigg]\otimes e^{i\frac{\pi\boldsymbol{x}^2}{\lambda\Delta q}}\Bigg]\cdot t(\boldsymbol{x}, q_{min})\Bigg]\otimes e^{i\frac{\pi\boldsymbol{x}^2}{\lambda q_{min}}}\Bigg]\Bigg\} \qquad (3.92)$$

It would be interesting to see what happens as $\Delta q \longrightarrow 0$ and $N \longrightarrow \infty$ but in this case, Kirchhoff's diffraction formula would no longer be valid. The intensity which will be recorded on the photographic film is, again, given by $U(\boldsymbol{x}) = \Re\{u(\boldsymbol{x})\}^2 + \Im\{u(\boldsymbol{x})\}^2$. It is obvious that, unlike zone plate holography, this is not simply the sum of the contributions from each plane parallel to the detector.

As mentioned before, however, this is not a type of hologram which has been used. It is impractical as light would be stopped by the object pattern on one transparency before it reached the next transparency and little detail about the side nearest the detector would be obtained. Leith and Upatnieks off-axis holograms [52], which give good quality images, involved scattering an object beam and an angled reference beam off the front of a solid object. Information from the side of the object viewed from the direction of the photographic plate is then recorded. This technique, however, only gives holograms which contain information about the surface of the object. Zone plate holography is truly three-dimensional with points throughout the volume of a three-dimensional radiation source contributing to the hologram.

## 3.9 Summary

In this Chapter, zone plate and conventional holography were discussed. Born and Wolf's derivation of the Kirchhoff-Fresnel diffraction formula was summarised and Soroko's derivation of the amplitude of a general hologram was reviewed. Using these formulae, a particular type of conventional holography, Gabor in-line holography, was analysed in detail and compared to zone plate holography. Holograms of single points were shown to be similar. Equations for the optical image reconstruction of holograms of single points were derived and the origins of the signal and noise terms were shown. The similarities and differences between zone plate and Gabor holograms of two points were derived and an attempt was made to quantify these differences. It was concluded that these differences will have an even greater effect in holograms of extended objects/sources. Holograms of three-dimensional objects/sources were compared and the differences were indicated. The effect of attenuation of gamma-rays by the source material was discussed and an attempt was made to find the magnitude of this effect. An undesirable scaling effect in the $z$ direction of an optically reconstructed zone plate hologram was found.

# Chapter 4

# INHERENT ERRORS ASSOCIATED WITH GAMMA RAY ZONE PLATE HOLOGRAPHY

The equipment used for zone plate holography was described in Section 3.4. A gamma camera is used to record the hologram and this affects the quality of the hologram. The pixel size must also be small enough for the frequencies in the hologram to be correctly recorded. The image is reconstructed using a computer. The accuracy of the digitisation and of the floating point arithmetic will also have an effect. Radioactive decay is a random process and, therefore, the quality of the hologram will improve with the time for which it is recorded. In this Chapter an attempt will be made to analyse and quantify the effect of each of these factors on the quality of a hologram.

## 4.1 Gamma Camera

The components of a gamma camera detector head can be seen in Figure 4.1. A lead collimator is usually attached. The most common type of collimator is a parallel hole

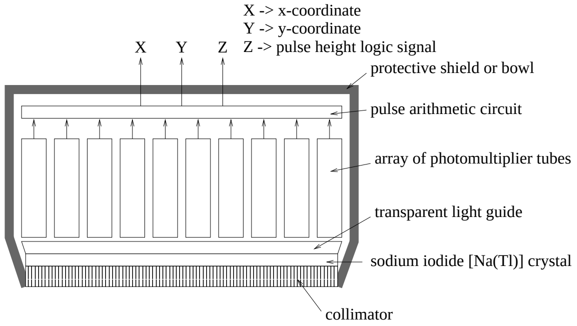


Figure 4.1: Diagram of a gamma camera detector head

collimator and this ensures that only gamma rays incident at right angles to the detector head will reach the crystal. The crystal is thallium activated sodium iodide NaI(Tl). When a gamma ray is absorbed by the crystal, molecules in the crystal are excited and then decay back to their ground state by emitting visible light. The total light emitted by NaI(Tl) depends upon the gamma ray energy absorbed. This relationship is close to linear over most of the significant energy range.

The crystal is optically coupled to an array of photomultiplier (PM) tubes. Each PM tube converts the scintillation light from the crystal into an electrical pulse. The total charge in the pulse is proportional to the mean number of photons incident on the tube. A pulse arithmetic circuit then calculates the position (X,Y) and energy (Z) of the gamma ray which caused the pulse using the signals from all the photomultiplier tubes.

The Ohio-nuclear inc. series 100 radioisotope camera, for example, contains 37 photomultiplier tubes and has a 33.7cm diameter crystal which is 1.27cm thick. Despite the small number of photomultiplier tubes, the pulse arithmetic circuit can calculate the position of a gamma ray hitting the crystal to a high degree of accuracy. The FWHM of the line spread function at the surface of the crystal is 0.5cm for 140keV gammas [63]. The spatial resolution for the cameras which have been used in work on gamma-ray holography at the University of Birmingham are shown in Table 4.1.

| manufacturer | model | spatial resolution (FWHM-m) |
|---|---|---|
| Ohio-Nuclear inc. | Series 100 radioisotope camera | $5\times10^{-3}$ [63] |
| GE | Maxicamera 535 | $5\times10^{-3}$ [65] |
| ADAC | Vertex gamma camera | $3.7\times10^{-3}$ [65] |
| | Forte gamma camera | $3.4\times10^{-3}$ [78] |
| Toshiba | GCA-7200A | $3.8\times10^{-3}$ [17] |

Table 4.1: Spatial resolution of example gamma cameras

Gamma rays react with the crystal molecules in two main ways, photoelectric absorption and Compton scattering. Photoelectric absorption results in the whole of the gamma ray energy being absorbed by the crystal. The emitted light and hence the pulse amplitude will be proportional to the gamma ray energy. Pulses from this interaction contribute to a photopeak in the pulse height spectrum as can be seen in Figure 4.2. Compton scattering results in the partial absorption of gamma ray energy depending upon the angle of scatter. The pulse amplitude in this case will depend upon the fraction of the energy absorbed. This gives rise to the Compton continuum in the pulse height spectrum. The positions at which the photons which were absorbed directly depend upon the position of the source. The positions at which photons were scattered and the fraction of energy that was absorbed in each case provides little useful information. The gamma camera allows a window to be placed over the photopeak. Only the pulses within this window are then accepted and recorded. Rew [70] studied the effect of this window on reconstructed image quality in her PhD thesis.

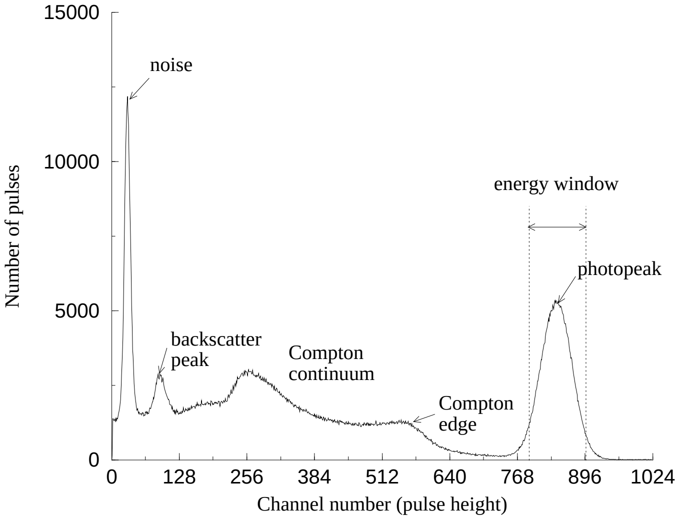


Figure 4.2: Example pulse height spectrum from a NaI(Tl) scintillation detector

The pulses from the photopeak are output as a list of (X,Y) positions. The data is stored in a square pixel array which usually contains between 64×64 and 1024×1024 pixels. When a gamma ray is detected, the closest pixel to its calculated position is incremented by one. Each array element therefore contains the number of gamma rays which hit the detector crystal at that pixel position. The pixel array forms an image of the gamma ray intensity incident on the detector.

## 4.2 Floating Point Arithmetic

A digital computer is only capable of representing a finite set of numbers. The computers used in this project are Sun Solaris SPARC Ultra 2 and use the IEEE 754 binary standard to store real numbers. This standard specifies floating point representations for single and double precision numbers. In general, the floating point representation can be written as follows

$$\pm \left(d_0 + d_1\beta^{-1} + \ldots + d_{p-1}\beta^{-(p-1)}\right)\beta^e, (0 \leq d_i < \beta) \tag{4.1}$$

$d_0, d_1, \ldots, d_{p-1}$ is called the significand and has $p$ digits, $\beta$ is the base and is assumed to be even and $e$ is the exponent. The values of these parameters for single and double precision IEEE 754 binary standard numbers are shown in Table 4.2. The number of bits required to store each part of the number are also shown. An extra bit is needed to store the sign (positive or negative) of the number. The exponent is biased. In the case of single precision, this means that the exponent is given by the stored number minus 126. In double precision numbers, 1022 is subtracted [41]. Floating point numbers are usually normalised. This means that the most significant bit of the significand, $d_0$, is non-zero. The IEEE binary standard has $\beta$ =2 so $d_0$ is 1 for normalised numbers and there is no need to waste a bit of memory storing it. Only $p-1$ bits are therefore needed to store the significand, not $p$.

| Parameter | Single precision | Double precision |
|---|---|---|
| $\beta$ | 2 | 2 |
| $p$ | 24 | 53 |
| $e_{max}$ | +127 | +1023 |
| $e_{min}$ | -126 | -1022 |
| Exponent width (bits) | 8 | 11 |
| Significand width (bits) | 23 | 52 |
| Format width (bits) | 32 | 64 |

Table 4.2: IEEE standard for single and double precision floating point numbers

The C header files float.h and limits.h were used to acquire specific information about the limiting values of real and integer numbers using Sun workstations. These values can be seen in Table 4.3. It can be seen that the values for floating-point numbers correspond to the IEEE standard.

The minimum and maximum positive normalised floating-point numbers for given values of $\beta$, $p$ and $e$ are respectively

$$\text{minimum} = \left(1 + 0\beta^{-1} + \ldots + 0\beta^{-(p-1)}\right)\beta^e = \beta^e \quad \text{and}$$

$$\text{maximum} = \left((\beta-1) + (\beta-1)\beta^{-1} + \ldots + (\beta-1)\beta^{-(p-1)}\right)\beta^e = \beta^{e+1} \tag{4.2}$$

| | min value | max value |
|---|---|---|
| Single precision (float) | 1.17549e-38 ($2^{-126}$) | 3.40282e+38 ($2^{128}$) |
| Double precision (double) | 2.22507e-308 ($2^{-1022}$) | 1.79769e+308 ($2^{1024}$) |
| Character (char) | -128 | 127 |
| Integer (int) | -2147483648 | 2147483647 |
| Long (long) | -2147483648 | 2147483647 |
| Unsigned integer (unsigned) | 0 | 4294967295 |

Table 4.3: Limiting values of different number formats

When a real number is rounded to one of these numbers, the maximum possible difference between the real number and the floating-point number is $\left(\frac{\beta}{2}\right)\beta^{-p}\beta^{e}$. A way of measuring the error is to find the difference between the real number and the floating point number divided by the real number. This is called the relative error. This error therefore lies in the following range

$$\left(\frac{1}{2}\right)\beta^{-p} \leq \text{relative error} \leq \left(\frac{\beta}{2}\right)\beta^{-p} \tag{4.3}$$

The maximum value of the error, $\left(\frac{\beta}{2}\right)\beta^{-p}$ is called the machine precision, $\epsilon$. $\beta$ =2 and $p$ =24 for single precision so $\epsilon$ =5.960464e-8. For double precision, $\beta$ =2 and $p$ =53 so $\epsilon$ = 1.110223e-16.

NAG routines [62] are a useful tool for manipulating data. These routines can, for example, perform fast Fourier transforms, invert matrices and solve sets of simultaneous equations. NAG fortran double precision numerical library mark 16A was the version installed on the computers used in this project so it was decided to perform all computations using double precision numbers.

To investigate the errors involved in arithmetic operations, i.e: addition, subtraction, multiplication and division, two square arrays of size $N \times N$ were filled with random double precision numbers. The numbers had the largest possible number of significant figures. Each number in the second array was added to the corresponding number in the first array and then was subtracted from it. The differences between the resulting numbers and the numbers in the first array were found. These errors were then divided by the corresponding numbers from the first array to give relative errors. The sum of the magnitudes $\Sigma_m$ was found for array sizes from $N$ =2 to $N$ =1024. $\Sigma_m$ was found to be the same when a NAG routine was used as when the program was written in C. The average relative error should be the same for each array size so the sum of the errors would be expected to increase as $N^2$. The operations needed to calculate $\Sigma_m$ from the relative errors will introduce errors into its value.

This procedure was also performed for multiplication and division. That is, random numbers were first multiplied and then divided by other random numbers and then the relative errors in the results were found. Functions of the form $y = Ax^B$ were fitted to the data using DWPOL [90], where $y$ is $\Sigma_m$ and $x$ is $N$. This can be seen in Figure 4.3.

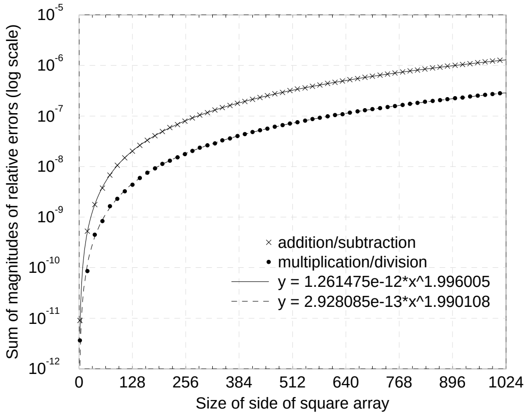


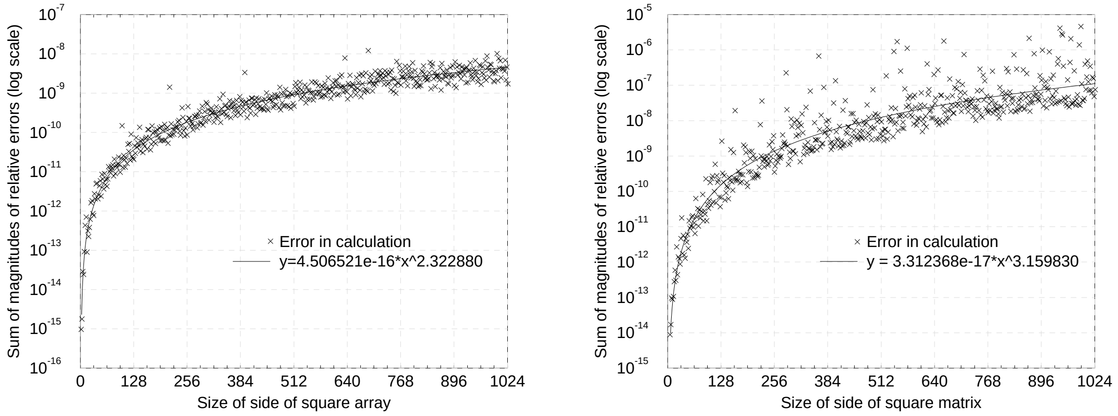
Figure 4.3: Errors associated with basic arithmetic operations

| | $\ln(A)$ | $B$ | $\Sigma_{i_{min}}^{i_{max}}(y_i - y_i^{fitted})^2$ |
|---|---|---|---|
| addition/subtraction | -27.3907388 | 1.99600502 | 0.52941 |
| multiplication/division | -28.8592574 | 1.99010809 | 1.9109 |

Table 4.4: Equations fit to errors from basic arithmetic operations

The values of $B$ given in Table 4.4 are very close to 2, so the errors vary as $N^2$ as predicted. Using the fitted functions to predict the relative errors involved in one addition and one subtraction, the error of for an addition followed by a subtraction was calculated to be 1.271608e-12 and the error for a multiplication followed by a division was 2.928085e-13 . As stated before, the machine precision for double precision numbers is 1.110223025e-16. These results show that arithmetic operations reduce this accuracy.

Two NAG routines which are used in this project are two-dimensional fast Fourier transforms (2D FFT) and matrix inversion. If the two-dimensional array of which the Fourier transform is to be found has size $N \times N$ and the matrix also has size $N \times N$, the errors in these operations could, again, be expected to vary as $N^2$. The array and matrix were again filled with random numbers. The Fourier transform of the array was found and then the inverse Fourier transform. The difference between the result and the original array was then calculated. The matrix was inverted and the result and the original matrix multiplied. The difference between this and an identity matrix was found. The results can be seen in Figure 4.4.



(a) 2D FFT/inverse 2D FFT (b) matrix inversion/multiplication

Figure 4.4: Errors in example calculations performed using NAG routines

| | $\ln(A)$ | $B$ | $\Sigma_{i_{min}}^{i_{max}}(y_i - y_i^{fitted})^2$ |
|---|---|---|---|
| 2D FFT/inverse 2D FFT | -35.3358361 | 2.32288049 | 111.89 |
| matrix inversion/multiplication | -37.9462832 | 3.15983044 | 675.38 |

Table 4.5: Equations fit to errors from Fourier transforms and matrix inversion

As can be seen from Table 4.5, $\Sigma_m$ due to 2D FFTs increases as $N^{2\frac{1}{3}}$ whereas $\Sigma_m$ due to a matrix inversion followed by a matrix multiplication increases as $N^3$. With both of these operations, therefore, the relative error per number will increase as the size of the array or matrix increases.

Hologram pixel values are often stored as integers between 0 and 255. In this case, the accuracy of the pixel values is much less than double precision so double precision arithmetic will not noticeably decrease the accuracy. Expected holograms for given source distributions are used in the reconstruction algorithms, however, and these can be calculated to double precision. It is here that the accuracy of floating point arithmetic should be considered.

## 4.3 Radioactive Decay

The fundamental law of radioactive decay [48] gives the activity of a radioactive source $A(t)$ with half life $T_{\frac{1}{2}}$ as

$$A(t) = -\left.\frac{dN(t)}{dt}\right|_{\text{decay}} = \lambda N(t) = \frac{\ln(2)}{T_{\frac{1}{2}}} N(t) \tag{4.4}$$

where $N(t)$ is the number of radioactive atoms at time $t$ and $\lambda$ is called the decay constant. If $N(0)$ is the number of atoms at time $t = 0$, $N(t)$ is given by

$$N(t) = N(0)e^{-\lambda t} \tag{4.5}$$

The probability, $p$, that a single atom will decay in a time interval $t = 0$ to $t = \Delta t$ is therefore

$$p = 1 - e^{-\lambda \Delta t} \tag{4.6}$$

Each atom has the same probability of decay. The expected counts in time $\Delta t$ is given by the number of atoms in the source multiplied by the probability $p$. The number of atoms at $t = 0$, $N(0)$ can be calculated using equation 4.4.

$$\text{Expected count in } \Delta t = \overline{N(t) - N(t + \Delta t)} = \frac{A(t)}{\lambda}\left(1 - e^{-\lambda \Delta t}\right) = \overline{N_{\Delta t}} \tag{4.7}$$

Usually, the counting interval $\Delta t$ is much shorter than the half life of the radioactive source. In this case, $p \ll 1$ and the Poisson distribution can be used to describe the number of counts recorded experimentally [48]. The probability $P(N_{\text{actual}})$ that a number of counts $N_{\text{actual}}$ are recorded is given by

$$P(N_{\text{actual}}) = \frac{\left(\overline{N_{\Delta t}}\right)^{N_{\text{actual}}} e^{-\overline{N_{\Delta t}})}}{N_{\text{actual}}!} \tag{4.8}$$

The standard deviation of this distribution, $\sigma$, is given by

$$\sigma = \sqrt{\overline{N_{\Delta t}}} \tag{4.9}$$

The higher the number of counts recorded, the less the effect these statistical fluctuations will have on a hologram.

## 4.4 Sampling Theory

A signal is called bandlimited if its Fourier transform is zero outside a bounded region in the frequency plane. If the Fourier transform is $F(\omega_x, \omega_y)$, it is bandlimited if

$$F(\omega_x, \omega_y) = 0 \quad \text{for} \quad |\omega_x| > \omega_x^0 \quad |\omega_y| > \omega_y^0 \tag{4.10}$$

The Nyquist theory states that if you sample a signal at a rate $\omega_x^s$ in the $x$-direction and a rate $\omega_y^s$ in the $y$-direction, the signal will be undersampled if these frequencies are less than twice the maximum frequency in the signal [43], i.e.

$$\omega_x^s < 2\omega_x^0 \quad \omega_y^s < 2\omega_y^0 \tag{4.11}$$

In this Section, a more thorough investigation of the effects of sampling will be given. The effects of sampling and undersampling on the Fourier transform of the data stored in a hologram will be demonstrated.

The gamma camera is used to record the hologram as an array of $N \times N$ pixels. This process can be thought of as sampling the hologram at a series of points separated by a distance $\Delta_p$, equal to the pixel size. If the intensity of the radiation incident on the detector is given by $h(x, y)$, the series of values which would ideally be recorded would be $h_{j_x^h j_y^h}$ where

$$h_{j_x^h j_y^h} = h(x, y)\delta(x - j_x^h \Delta_p, y - j_y^h \Delta_p) = h(j_x^h \Delta_p, j_y^h \Delta_p) \tag{4.12}$$

where $j_x^h$ and $j_y^h$ vary between $-\frac{N}{2}$ and $\frac{N}{2} - 1$. The sampled signal $\tilde{h}(x, y)$ can therefore be described by

$$\tilde{h}(x, y) = \sum_{j_y^h = -\frac{N}{2}}^{\frac{N}{2}-1} \sum_{j_x^h = -\frac{N}{2}}^{\frac{N}{2}-1} h(x, y)\delta(x - j_x^h \Delta_p, y - j_y^h \Delta_p) \tag{4.13}$$

The two-dimensional *comb* or *shah* $Ш(x, y, a, b)$ function is given by

$$Ш(x, y, a, b) = |a||b| \sum_{j_y^h = -\infty}^{\infty} \sum_{j_x^h = -\infty}^{\infty} \delta\left(x - j_x^h a, y - j_y^h b\right) \quad [69] \tag{4.14}$$

In the region $\frac{-N\Delta_p}{2} \leq x \leq \left(\frac{N}{2} - 1\right)\Delta_p$ and $\frac{-N\Delta_p}{2} \leq y \leq \left(\frac{N}{2} - 1\right)\Delta_p$, $\tilde{h}(x, y)$ is therefore described by

$$\tilde{h}(x, y) = \Delta_p^2 h(x, y) Ш\left(x, y, \Delta_p, \Delta_p\right) \tag{4.15}$$

A two-dimensional Fourier transform (2D FT) can be used to find the spatial frequencies in the hologram and to discover if it is under sampled or not. The 2D FT of $\text{ш}(x, y, a, b)$, $\text{III}\left(\omega_x, \omega_y, \frac{2\pi}{a}, \frac{2\pi}{b}\right)$, is given by

$$\text{III}\left(\omega_x, \omega_y, \frac{2\pi}{a}, \frac{2\pi}{b}\right) = F\{\text{ш}(x, y, a, b)\} = \sum_{\mu_y=-\infty}^{\infty} \sum_{\mu_x=-\infty}^{\infty} \delta\left(\omega_x - \frac{2\pi\mu_x}{a}, \omega_y - \frac{2\pi\mu_y}{b}\right) \text{[87]} \tag{4.16}$$

Taking the limit as $N \to \infty$, the Fourier transform of the sampled hologram, $\tilde{H}_\infty(\omega_x, \omega_y)$ will be given by

$$\tilde{H}_\infty(\omega_x, \omega_y) = H(\omega_x, \omega_y) \otimes \text{III}\left(\omega_x, \omega_y, \frac{2\pi}{\Delta_p}, \frac{2\pi}{\Delta_p}\right) \tag{4.17}$$

where $\otimes$ denotes a convolution. The Fourier transform of the sampled signal is therefore equivalent to the sum of a series of Fourier transform patterns each shifted by a distance $\frac{2\pi}{\Delta_p}$ in the $\omega_x$ or $\omega_y$ direction with respect to each other. If the hologram is under sampled, these spectra overlap. For this to happen, the highest frequency in the $x$ or $y$ direction in the hologram, $\omega_{max}$, would have to be greater than $\frac{\pi}{\Delta_p}$. This is called the Nyquist limit [56].

The continuous two-dimensional Fourier transform (2D FT)$F(\omega_x, \omega_y)$ of a function $f(x, y)$ is given by

$$F(\omega_x, \omega_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i(\omega_x x + \omega_y y)} dx dy \tag{4.18}$$

The discrete Fourier transform of the $N \times N$ array of hologram pixels can be calculated using a NAG routine. The two-dimensional discrete Fourier transform $D_{k_x k_y}$ (2D DFT) of the hologram pixels $h_{j_x j_y}$ is given by

$$D_{k_x k_y} = \frac{1}{N} \sum_{j_y=0}^{N-1} \sum_{j_x=0}^{N-1} h_{\left(j_x - \frac{N}{2}\right)\left(j_y - \frac{N}{2}\right)} e^{\left[-2\pi i\left(\frac{j_x k_x}{N} + \frac{j_y k_y}{N}\right)\right]} \tag{4.19}$$

where $k_x$, $k_y$, $j_x$ and $j_y$ vary between 0 and $N - 1$. The NAG routine calculates this using the Fast Fourier Transform method [68]. As can be seen in equation 4.18, the continuous Fourier transform is given by an integral over all space ($-\infty \leq x \leq \infty$, $-\infty \leq y \leq \infty$). Equation 4.19 can be re-arranged so it can be compared with the 2D FT. $k_x$, $k_y$, $j_x$ and $j_y$ need to vary between $-\frac{N}{2}$ and $\frac{N}{2} - 1$ as in equation 4.13. Using this range of indices reflects the fact that $x$, $y$, $\omega_x$ and $\omega_y$ can be negative as well as positive. Substituting

$k'_x = k_x - \frac{N}{2}$, $k'_y = k_y - \frac{N}{2}$ and $j'_x = j_x - \frac{N}{2}$, $j'_y = j_y - \frac{N}{2}$ into equation 4.19 gives

$$D_{\left(k'_x+\frac{N}{2}\right)\left(k'_y+\frac{N}{2}\right)} = \frac{1}{N} e^{-i\pi(k'_x+k'_y)} \sum_{j'_y=-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{j'_x=-\frac{N}{2}}^{\frac{N}{2}-1} e^{-i\pi(j'_x+j'_y)} h_{j'_x j'_y} e^{\left[-2\pi i\left(\frac{j'_x k'_x}{N}+\frac{j'_y k'_y}{N}\right)\right]} \tag{4.20}$$

The size of a side of the detector, $P$, is given by $P = N\Delta_p$. The 2D DFT above can be written as

$$D(k'_x, k'_y) = \frac{\Delta_p}{P}(-1)^{k'_x+k'_y} \sum_{j'_y=-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{j'_x=-\frac{N}{2}}^{\frac{N}{2}-1} (-1)^{j'_x+j'_y} h_{j'_x j'_y} e^{\left[-2\pi i\left(\frac{j'_x k'_x \Delta_p}{P}+\frac{j'_y k'_y \Delta_p}{P}\right)\right]} \tag{4.21}$$

If the size of the detector, $P$, remains constant. As $N \to \infty$, the above equation will become

$$D(k'_x, k'_y) = \frac{\Delta_p}{P}(-1)^{k'_x+k'_y} \sum_{j'_y=-\infty}^{\infty} \sum_{j'_x=-\infty}^{\infty} (-1)^{j'_x+j'_y} h_{j'_x j'_y} e^{\left[-2\pi i\left(\frac{j'_x k'_x \Delta_p}{P}+\frac{j'_y k'_y \Delta_p}{P}\right)\right]} \tag{4.22}$$

Comparing this to equation 4.16 and noting that $x = j'_x\Delta_p$ and $y = j'_y\Delta_p$, the following sequence for calculating the two-dimensional Fourier transform can be devised.

**Algorithm for calculating 2D Fourier transform**

1. Multiply the hologram array values $h_{j'_x j'_y}$ by $(-1)^{j'_x+j'_y}$ where $-\frac{N}{2} \leq j'_x \leq \left(\frac{N}{2}-1\right)$ and $-\frac{N}{2} \leq j'_y \leq \left(\frac{N}{2}-1\right)$.

2. Calculate the 2D DFT.

3. Multiply the result by $(-1)^{k'_x+k'_y}$ where $-\frac{N}{2} \leq k'_x \leq \left(\frac{N}{2}-1\right)$ and $-\frac{N}{2} \leq k'_y \leq \left(\frac{N}{2}-1\right)$.

4. The result $D(k'_x, k'_y)$ tends to $F\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$ convolved with $\text{III}\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y, \frac{2\pi}{\Delta_p}, \frac{2\pi}{\Delta_p}\right)$ as $N \to \infty$ where $F(\omega_x, \omega_y)$ is the 2D FT of a function $f(x, y)$ and

$$f(x,y) = \begin{cases} h(x,y) & \text{if } -\frac{P}{2} \leq x \leq \frac{P}{2} - \Delta_p \text{ and } -\frac{P}{2} \leq y \leq \frac{P}{2} - \Delta_p \\ 0 & \text{if } x < -\frac{P}{2}, x > \frac{P}{2} - \Delta_p, y < -\frac{P}{2}, \text{ or } y > \frac{P}{2} - \Delta_p \end{cases} \tag{4.23}$$

5. As the 2D FT is convolved with the shah function, it will be repeated at $\frac{2\pi}{P}k'_x = \mu_x \frac{2\pi}{\Delta_p}$ and $\frac{2\pi}{P}k'_y = \mu_y \frac{2\pi}{\Delta_p}$ where $\mu_x$ and $\mu_y$ are integers. This means that the separation of adjacent copies of the 2D FT in both directions, $\Delta k$ will be $N$. Aliasing will therefore occur if $F\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$ has non-zero values when $|k'_x|$ or $|k'_y|$ is greater than $\frac{N}{2}$

As discussed in Section 3.6, a Gabor zone plate hologram of an on-axis point source has the following variation

$$v(x,y) = 2\omega_{zp}^2 \cos(\omega_{zp}^2(x^2+y^2)) \quad \text{where} \quad \omega_{zp} = \frac{p\sqrt{\pi}}{r_0(p+q)} \tag{4.24}$$

and $r_0$ is the first zone radius, $p$ is the separation of the source and zone plate and $q$ is the separation of the zone plate and detector. We aim to discover when a hologram is under sampled. The exact magnitude of this function in the hologram and the constant term which is also present can be neglected for the purpose of this analysis.

Referring back to equation 3.64, $v(x,y)$ has the following Fourier transform.

$$V(\omega_x,\omega_y) = 2\pi \sin\left(\frac{\omega_x^2+\omega_y^2}{4\omega_{zp}^2}\right) \tag{4.25}$$

As in equation 4.23, only the part of $v(x,y)$ which is incident on the detector is recorded. This is similar to multiplying $v(x,y)$ by a 'box' function which truncates it at the edges of the detector. The Fourier transform of a box function is a sinc function. The Fourier transform of $v(x,y)$ is therefore convolved with this sinc function. An expression for the resulting Fourier transform will be derived here. This effect is often called 'ringing' and can be lessened by using a window function. Kuo [49] investigated the application of several different windows to images from zone plate holograms as part of his PhD.

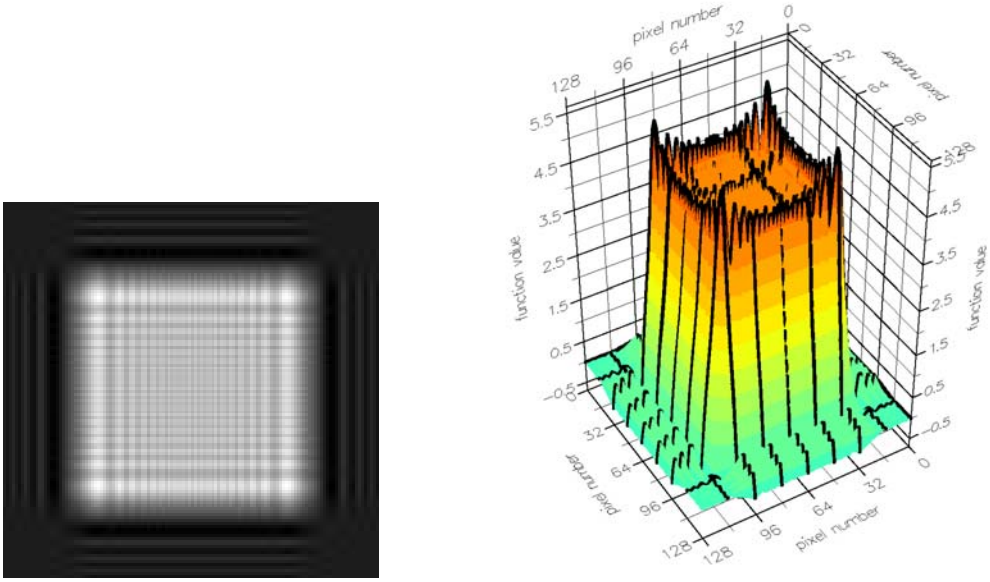
This part of $v(x,y)$ recorded by the detector is $f(x,y)$. The Fourier transform of $f(x,y)$ is given by

$$\begin{aligned} F(\omega_x,\omega_y) &= V(\omega_x,\omega_y)\frac{1}{4}E(\omega_x,\omega_y) = \\ & 2\pi\sin\left(\frac{\omega_x^2+\omega_y^2}{4\omega_{zp}^2}\right)\frac{1}{4}\left\{\mathrm{Erf}\left[\frac{1}{2\omega_{zp}}\left(\left(\frac{P}{2}-\Delta_p\right)2\omega_{zp}^2-\omega_x\right)\frac{1-i}{\sqrt{2}}\right]-\right. \\ & \left.\mathrm{Erf}\left[\frac{1}{2\omega_{zp}}\left(\left(-\frac{P}{2}\right)2\omega_{zp}^2-\omega_x\right)\frac{1-i}{\sqrt{2}}\right]\right\}\left\{\mathrm{Erf}\left[\frac{1}{2\omega_{zp}}\left(\left(\frac{P}{2}-\Delta_p\right)2\omega_{zp}^2-\omega_x\right)\frac{1-i}{\sqrt{2}}\right]-\right. \\ & \left.\mathrm{Erf}\left[\frac{1}{2\omega_{zp}}\left(\left(-\frac{P}{2}\right)2\omega_{zp}^2-\omega_y\right)\frac{1-i}{\sqrt{2}}\right]\right\} \text{ [80]} \end{aligned} \tag{4.26}$$

where Erf$[x]$ is the error function and

$$\begin{aligned} \mathrm{Erf}\left[K\left(\frac{1-i}{\sqrt{2}}\right)\right] &= \frac{2}{\sqrt{\pi}}\int_0^{K\left(\frac{1-i}{\sqrt{2}}\right)} e^{-t^2}dt = (1-i)\int_0^{\sqrt{\frac{2}{\pi}}K} e^{i\frac{\pi}{2}u^2}du \\ &= \int_0^{\sqrt{\frac{2}{\pi}}K} \cos\left(\frac{\pi}{2}u^2\right)+\sin\left(\frac{\pi}{2}u^2\right)du + i\int_0^{\sqrt{\frac{2}{\pi}}K}\sin\left(\frac{\pi}{2}u^2\right)-\cos\left(\frac{\pi}{2}u^2\right)du \\ &= \left[\mathrm{FresnelC}\left(\sqrt{\frac{2}{\pi}}K\right)+\mathrm{FresnelS}\left(\sqrt{\frac{2}{\pi}}K\right)\right]+i\left[\mathrm{FresnelS}\left(\sqrt{\frac{2}{\pi}}K\right)-\mathrm{FresnelC}\left(\sqrt{\frac{2}{\pi}}K\right)\right] \end{aligned} \tag{4.27}$$

FresnelC$(x)$ and FresnelS$(x)$ are the cosine and sine Fresnel integrals and can be calculated using NAG routines. Using the relationship in equation 4.27, the real and imaginary parts of $E(\omega_x, \omega_y)$ can be calculated. As before, $\omega_x = \frac{2\pi}{P}k'_x$ and $\omega_y = \frac{2\pi}{P}k'_y$. Figure 4.5 shows the real part of $E(\omega_x, \omega_y)$ for $N = 128$, $\Delta_p = 0.4296$, $p = 22$, $q = 110$ and $r_0 = 1.02$.



(a) 128x128 pixel array, black=-7.057721e-01, white=5.651405e+00

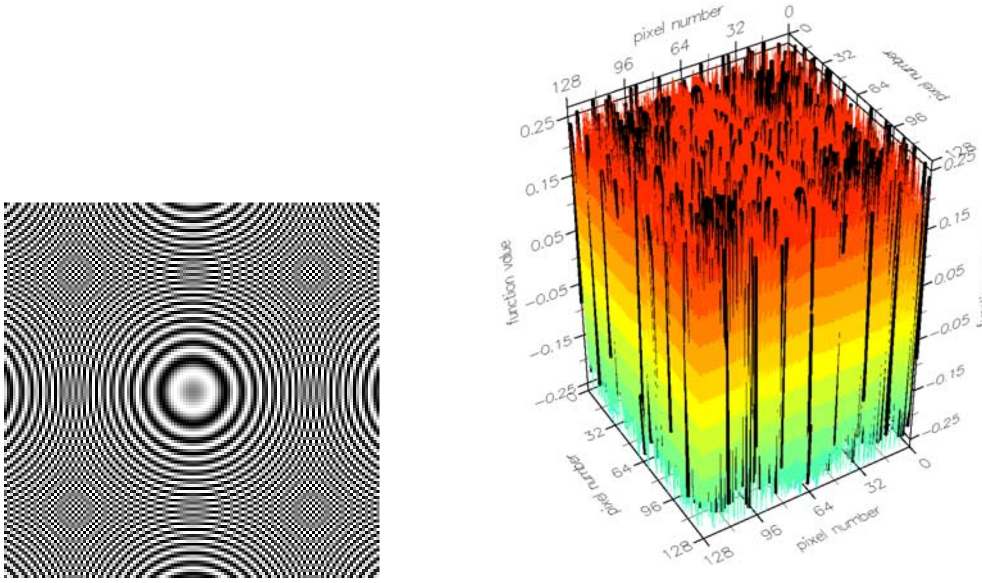
(b) x,y=pixel number, z=amplitude

Figure 4.5: Calculated Fourier transform of the effect of having a finite detector area $\left(E\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)\right)$ - real part, for $-64 \leq k'_x \leq 63$ and $-64 \leq k'_y \leq 63$

Figure 4.6 on page 72 shows the corresponding values of $V\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$ and the real part of $F\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$. In this case, $E\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$ becomes zero when $|k'_x|$ and $|k'_y|$ are less than $\frac{N}{2}$ so the pixel size is sufficiently small for the frequencies in the hologram to be correctly recorded.

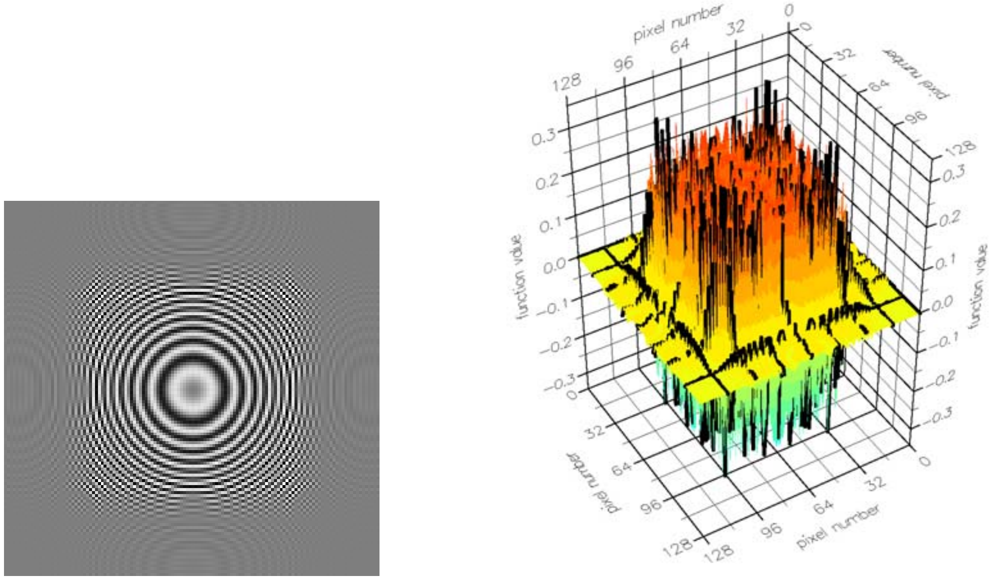
Figure 4.7 on page 73 shows the real and imaginary parts of $E(\frac{2\pi}{P}k'_x, 0)$ over values of $k'_x$ from -256 to 255. This function is shown for the above values and also when $\Delta_p$ is increased by a factor 2. The figure illustrates that, when $\Delta_p$ =0.8592, the hologram is under sampled.

The zone plates used experimentally, however, have a varying transmission up to a maximum radius $R$, after which the transmission has a minimum value. Considering the case where the minimum transmission is zero, the hologram will have a variation $v_R(x, y)$



(a) 128x128 pixel array, black=-2.659753e-01, white=2.659750e-01

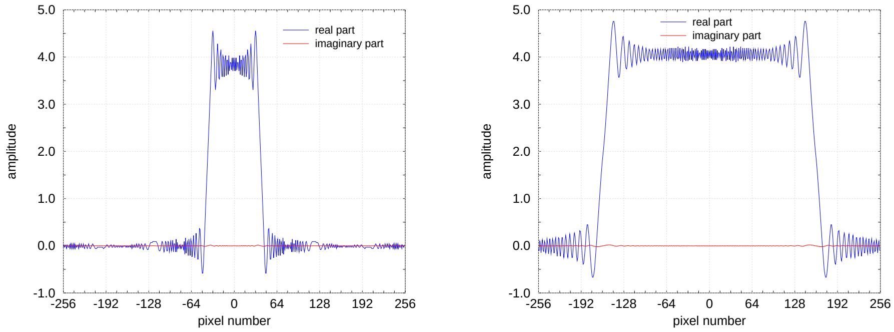
(b) x,y=pixel number, z=amplitude



(c) 128x128 pixel array, black=-3.438793e-01, white=3.496619e-01

(d) x,y=pixel number, z=amplitude

Figure 4.6: Calculated Fourier transform of an infinite Gabor zone plate ($V\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$) and calculated Fourier transform of the portion of the zone plate recorded by the detector ($F\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$) for -64$\leq k'_x \leq$63 and -64$\leq k'_y \leq$63



(a) $\Delta_p$ =0.4296 (b) $\Delta_p$ =0.8592

Figure 4.7: Effect of the finite detector size, $E\left(\frac{2\pi}{P}k'_x, 0\right)$, for -256$\leq k'_x \leq$255

where

$$v_R(x,y) = \begin{cases} v(x,y) & \sqrt{x^2+y^2} \leq R \\ 0 & \sqrt{x^2+y^2} > R \end{cases} \tag{4.28}$$

Considering $v_R(x,y)$ to be $f(x,y)$ multiplied by a circular aperture function $a(x,y)$. $v_R(x,y)$, $f(x,y)$ and $a(x,y)$ are all radially symmetric so they can be written as functions of the radius $r$.

$$v_R(r) = f(r)a(r) \quad (where) \quad a(r) = \begin{cases} 1 & r \leq R \\ 0 & r > R \end{cases} \tag{4.29}$$

The 2DFT $F(s)$ of a radially symmetric function $f(r)$ can be written in terms of $r$ rather than $x$ and $y$ as follows.

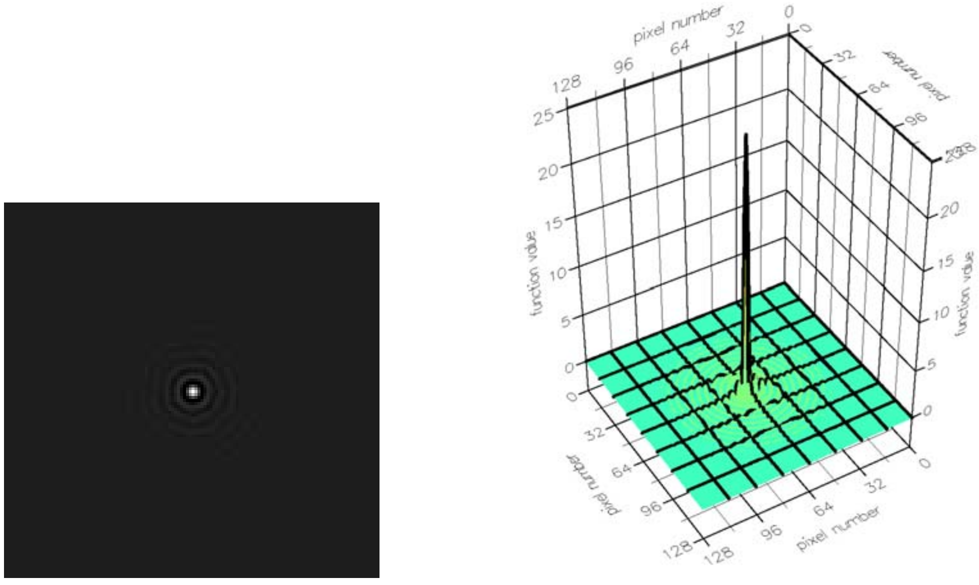
$$F(s) = 2\pi \int_0^\infty f(r) J_0(sr) r dr \text{ [69]} \tag{4.30}$$

$J_0(x)$ is a zeroth order Bessel function of the first kind. This way of writing the 2DFT is called a *Fourier-Bessel Transform* or *Hankel Transform*. The Hankel transform of $a(r)$, $A(s)$, is given by

$$A(s) = \frac{R}{s} J_1(Rs) \quad \text{where} \quad s = \sqrt{\omega_x^2 + \omega_y^2} \text{ [69]} \tag{4.31}$$

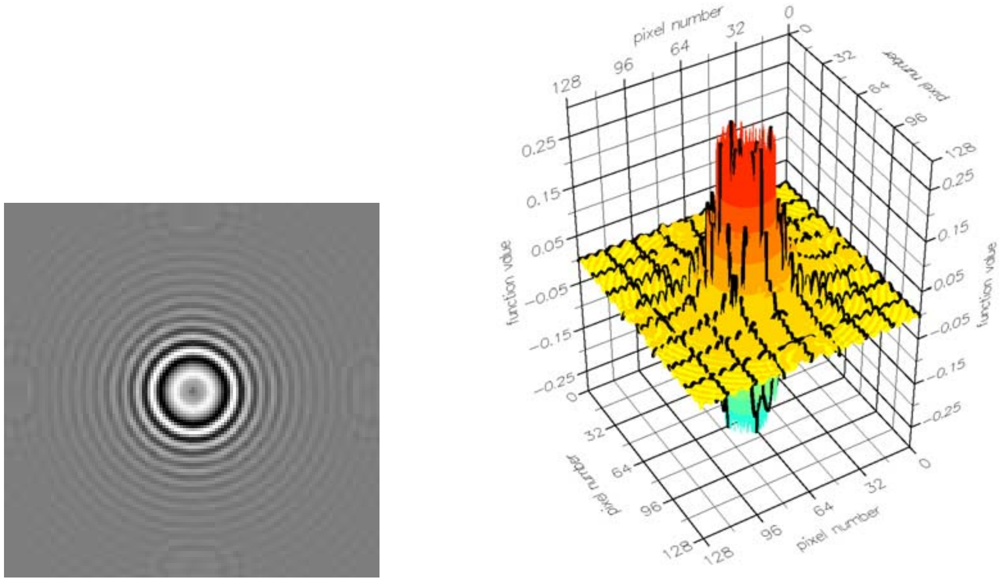
and $J_1(x)$ is a first order Bessel function of the first kind. This function can be calculated using a NAG routine. The Fourier transform of $v_R(x,y)$, $V_R(\omega_x, \omega_y)$ is therefore given by

$$V_R(\omega_x, \omega_y) = F(\omega_x, \omega_y) \otimes A(\omega_x \omega_y) \tag{4.32}$$



(a) 128x128 pixel array, black=-3.275953e+00, white=2.513274e+01

(b) x,y=pixel number, z=amplitude



(c) 128x128 pixel array, black=-2.964582e-01, white=3.044598e-01

(d) x,y=pixel number, z=amplitude

Figure 4.8: Fourier transform of a circular aperture of radius equal to the maximum radius of the zone plate $\left(A\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)\right)$ and of the truncated Gabor zone plate $\left(V_R\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)\right)$ - real part, for -64$\leq k'_x \leq$63 and -64$\leq k'_y \leq$63

With the parameters used for Figure 4.6 and $R = 32 \times \Delta_p$, the values for $A\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$ and $V_R\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$ are as shown in Figure 4.8.

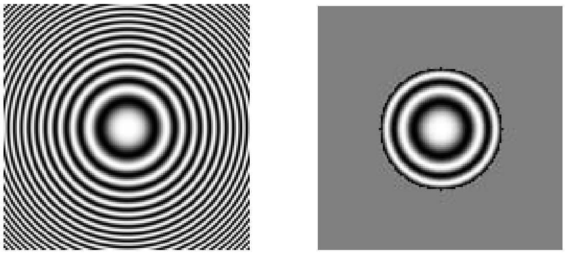
In this case, convolution with the circular aperture, $A\left(\frac{2\pi}{P}k'_x, \frac{2\pi}{P}k'_y\right)$, makes the Fourier transform of the hologram decrease to near zero at an even lower frequency.

In summary, the Fourier transform of a zone plate hologram of a single on-axis point source can be analysed to see whether the pixel size is sufficiently small for the frequencies to be correctly recorded. Using equations 4.26, 4.27, 4.31 and 4.32, the expected Fourier transform for a hologram sampled at a high enough frequency can be calculated. The Fourier transform obtained using the algorithm outlined on page 69 can then be compared with the calculated transform. If the hologram is under sampled, the two transforms will be obviously different. To illustrate this, the Fourier transforms calculated using the equations and the algorithm are compared in Figures 4.9 and 4.10. The hologram in Figure 4.9 has the parameters used in Figure 4.7(a) and Figure 4.10 those used in Figure 4.7(b).

## 4.5 Summary

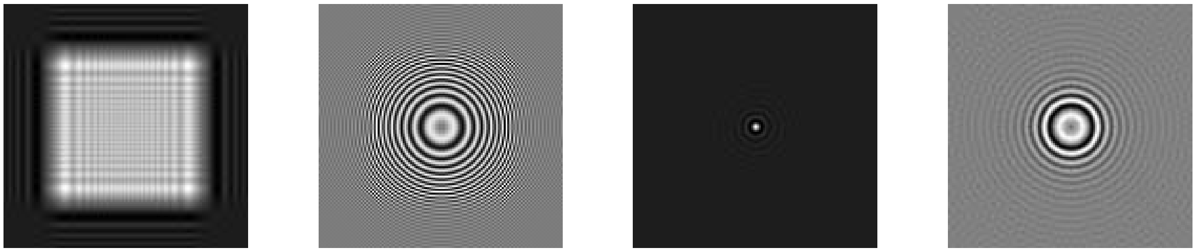
In this Chapter, sources of error due to the equipment used to detect, store and process zone plate holograms of gamma ray sources have been listed. The way that the gamma camera, floating point arithmetic and the statistics of radioactive decay contribute to errors in the recorded hologram have been explained.

The difference between continuous and discrete Fourier transforms has been outlined and an expression for the calculation of the discrete Fourier transform from the equivalent continuous transform has been derived. Formulae describing the effect of the limited detector size and the truncation of the zone plate pattern at a maximum radius have also been found. These expressions can be used to test if the hologram is under sampled.



(a) $f\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black= -1.678e-01 white= 1.678e-01

(b) $v_R\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black= -1.677e-01 white= 1.679e-01

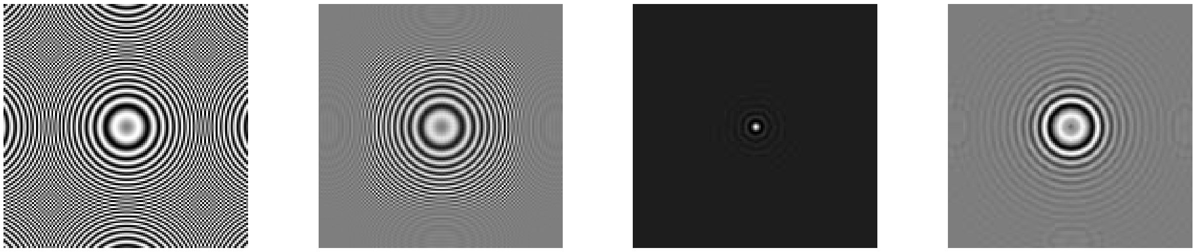


(c) $E\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-7.058e-01 white=5.651e+00

(d) 2D DFT of $f\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black=-3.147e-01 white=3.295e-01

(e) 2D DFT of $a\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black=-3.282e-01 white=2.507e+01

(f) 2D DFT of $v_R\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black=-3.023e-01 white=3.001e-01



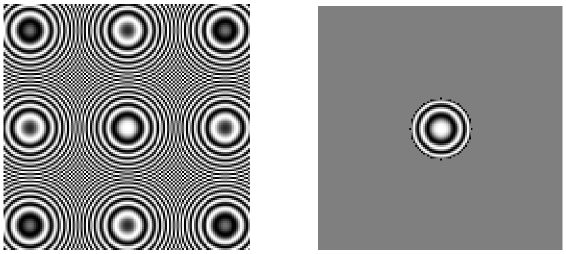
(g) $V\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-2.660e-01 white=2.660e-01

(h) $F\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-3.439e-01 white=3.497e-01

(i) $A\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-3.276e+00 white=2.513e+01

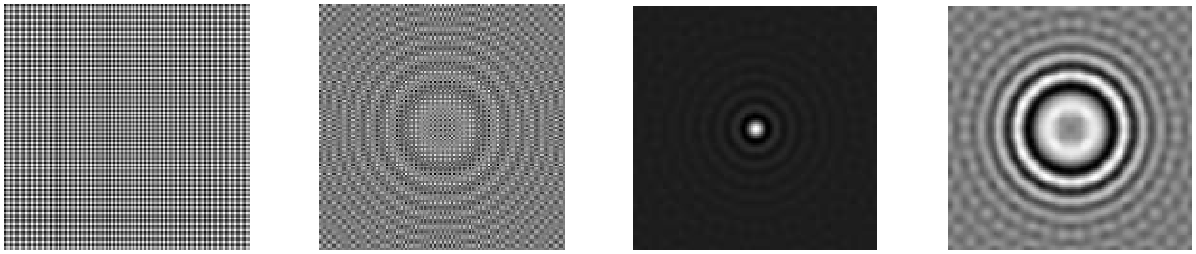
(j) $V_R\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-2.965e-01 white=3.045e-01

Figure 4.9: Comparison of actual and calculated Fourier transforms for a hologram with $N = 128$, $\Delta_p = 0.4296$, $p = 22$, $q = 110$, $r_0 = 1.02$ and $R = 13.7472$



(a) $f\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black= -1.678e-01 white= 1.678e-01

(b) $v_R\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black= -1.676e-01 white= 1.678e-01

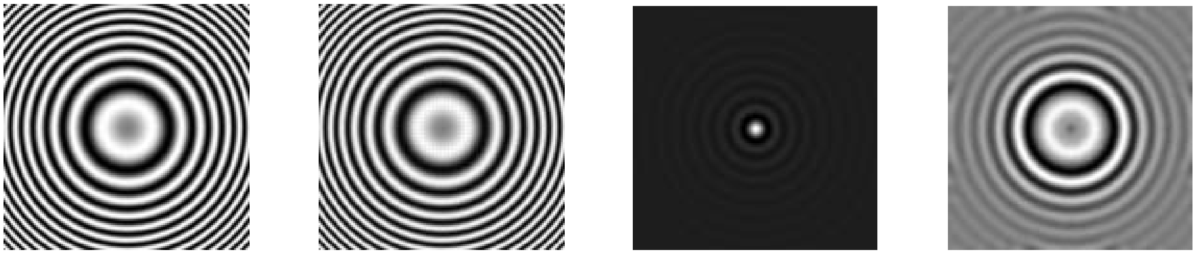


(c) $E\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=3.705e+00 white=4.305e+00

(d) 2D DFT of $f\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black=-4.846e-01 white=5.199e-01

(e) 2D DFT of $a\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black=-8.152e-01 white=6.227e+00

(f) 2D DFT of $v_R\left(\Delta_p j'_x, \Delta_p j'_y\right)$ black=-7.733e-02 white=8.086e-02



(g) $V\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-6.649e-01 white=6.649e-01

(h) $F\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-7.113e-01 white=7.109e-01

(i) $A\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-8.263e+00 white=6.283e+01

(j) $V_R\left(\Delta_p k'_x, \Delta_p k'_y\right)$ black=-7.463e-02 white=7.520e-02

Figure 4.10: Comparison of actual and calculated Fourier transforms for a hologram with $N = 128$, $\Delta_p = 0.8592$, $p = 22$, $q = 110$, $r_0 = 1.02$ and $R = 13.7472$

# Chapter 5

# ASSESSMENT OF RECONSTRUCTED IMAGE QUALITY

The aim of this work is to discover and evaluate methods for reconstruction of images from zone plate holograms. Reconstruction of a perfect image is made difficult by the sources of error outlined in Chapter 4 and these errors will result in different artifacts in the reconstructed image for each method. Subjective judgement of image quality, that is, "I think image A looks better than image B", may give a reasonable assessment of the effectiveness of a reconstruction method but is subject to the personal preferences of the human observer. A method for measurement of image quality is needed so the reconstruction methods can be compared.

ICRU report 54 entitled "Medical Imaging - The Assessment of Image Quality" [64] states that there are three kinds of physical parameters which are fundamental to image system specification.

1. Large-scale (macro) system transfer function (characteristic curve) which measures the relationship between system input, e.g., exposure quanta, and the output image, e.g., optical density.
2. Spatial resolution properties.
3. Noise properties.

Each of these parameters will be considered in turn.

## 5.1 System Transfer Function

This gives a measure of the response of the imaging system to source distributions containing different spatial frequencies.

### 5.1.1 Relationship between the Source Distribution and the Output Image

How is the input signal determined so the output signal can be compared with it? To answer this question, the equations which describe the formation of zone plate holograms will first be reviewed.

Equation 3.46 gave the relationship between source distribution, $S(\boldsymbol{\epsilon}, p)$, and hologram intensity $U(\boldsymbol{x})$ as

$$U(\boldsymbol{x}) = \frac{D^2}{4\pi} \int \frac{1}{(p+q)^2} \int \gamma(\boldsymbol{\epsilon}, p, S(\boldsymbol{\epsilon}, p)) S(\boldsymbol{\epsilon}, p)\, T\left(\frac{\boldsymbol{x}}{M(p)} - L(p)\boldsymbol{\epsilon}\right) d\boldsymbol{\epsilon} dp \tag{5.1}$$

where $T(\boldsymbol{\epsilon})$ is the intensity transmittance of the zone plate, $\gamma(\boldsymbol{\epsilon}, p, S(\boldsymbol{\epsilon}, p))$ is a factor which depends upon the source distribution, and as was explained in Chapter 3, the other factors and constants depend upon the arrangement of the equipment.

To simplify matters, a simple two-dimensional source will be used to illustrate the relationship between the source distribution and the output image. The source will be a distance $p_0$ from the zone plate. As the source is two-dimensional, the emitted gamma-rays will not have to travel through other parts of the source. If gamma-ray attenuation due to air is assumed to be negligible, the approximation $\gamma(\boldsymbol{\epsilon}, p_0, S(\boldsymbol{\epsilon}, p_0)) = 1$ can be used. If $S'(\boldsymbol{x}) = S(\frac{\boldsymbol{x}}{L(p_0)}, p_0)$ and $T'(\boldsymbol{x}) = T(M(p_0)\boldsymbol{x})$ are scaled versions of the source distribution and zone plate pattern respectively, the hologram intensity, $U(\boldsymbol{x})$, can be written as a convolution.

$$U(\boldsymbol{x'}) = C(p_0, q, D) S'(\boldsymbol{x'}) \otimes T'(\boldsymbol{x'}) \tag{5.2}$$

$C(p_0, q, D)$ is a constant containing the factors in equation 5.1.

The four sources of error discussed in Chapter 4 were the gamma camera, floating point numbers, the fluctuations due to radioactive decay and under sampling. The fluctuations due to radioactive decay can be considered as noise added to the source strength term, this noise will be called $N_s$. The gamma camera and floating point numbers contribute to errors, or noise, in the recorded hologram intensity, this noise will be called $N_a$. Equation 5.2 then becomes

$$U(\boldsymbol{x'}) = C(p_0, q, D) \left[S'(\boldsymbol{x'}) + N_s(\boldsymbol{x'})\right] \otimes T'(\boldsymbol{x'}) + N_a(\boldsymbol{x'}) \tag{5.3}$$

If the hologram is under sampled, this will lead to artifacts in the reconstructed image. This effect is different from a random fluctuation, however, and so cannot be included in a noise term.

This work does not involve the improvement of zone plate holograms by changing the equipment used. The aim of this project is to find and compare different reconstruction methods. This can be done using real zone plate holograms or computer generated holograms. Computer-generated holograms can be used to test if a reconstruction method would work if no noise was present and, if so, if it still works when different types of noise are added. In this case, the size, shape and position of the source are precisely known and so reconstructed images can be compared with a theoretical perfect image of the source.

It is, however, extremely difficult, if not impossible, to accurately simulate all of the effects described above. Future and current work on the zone plate holography project must aim to produce an imaging system that can be used clinically and, therefore, methods must work for zone plate holograms of sources in-vivo. The quality of an image from a real hologram can not be compared with a theoretical perfect image as this is unknown. The hologram which would have been generated by the reconstructed source can, however, be estimated using equation 5.1. This estimated hologram can then be compared with the actual hologram.

If a method does not work for a computer-generated hologram, it most probably will not work for a real one. This work will, therefore, first concentrate on computer-generated holograms. Techniques which are seen to work well will then be applied to the few real holograms that are presently available.

### 5.1.2 Optical and Modulation Transfer Functions (OTF/MTF)

If a reconstruction technique reconstructs a point source $S(\boldsymbol{x}, p) = \delta(\boldsymbol{x_0}, p_0)$ as a spot $H(\boldsymbol{x_0}, p_0)$ and the spot has the same shape no matter where the point source is, i.e. what $(\boldsymbol{x_0}, p_0)$ is, the system is spatially invariant. In zone plate holography, the spot is approximately the same for a constant value of $p_0$ and any value of $\boldsymbol{x_0}$.

A commonly used measure of the relationship between system input and the output image for a spatially invariant imaging system is the optical transfer function (OTF) and its magnitude, the modulation transfer function (MTF). The MTF gives the relative amplitude of the output signal as a function of the spatial frequency of a sinusoidal input. If $H^F_{p_0}(\boldsymbol{\omega})$ is the two-dimensional Fourier transform of $H(\boldsymbol{x_0}, p_0)$, the optical transfer function is the normalised frequency response [43]

$$\mathrm{OTF}(\boldsymbol{\omega}) = \frac{H^F_{p_0}(\boldsymbol{\omega})}{H^F_{p_0}(\mathbf{0})} \tag{5.4}$$

The modulation transfer function is therefore

$$\mathrm{MTF}(\boldsymbol{\omega}) = \frac{|H^F_{p_0}(\boldsymbol{\omega})|}{|H^F_{p_0}(\mathbf{0})|} \tag{5.5}$$

## 5.2 Spatial Resolution

The spatial resolution can be investigated by finding the output image when the source is a known, simple object. This information can then be used to predict the resolution for more complex sources.

### 5.2.1 Point Spread Function (PSF)

As was mentioned in Section 5.1.2, a point source $S(\boldsymbol{x}, p) = \delta(\boldsymbol{x_0}, p_0)$ at point $(\boldsymbol{x_0}, p_0)$ will have a contribution $H(\boldsymbol{x_0}, p_0)$ to the reconstructed image. If the reconstruction technique works well, $H(\boldsymbol{x_0}, p_0)$ should be located at $(\boldsymbol{x_0}, p_0)$ but will have probably been blurred by the imaging and reconstruction processes. The degree of blurring may vary between different positions and different reconstruction techniques. The reconstruction techniques used in this work assume that the shadow of the zone plate is the same size and shape for point sources in the same plane (same value of $p$). It could therefore be assumed that the point spread function is the same for sources in the same plane. Indeed, the work done by Caplan [16] showed that this was a good approximation.

The degree of blurring can be measured by fitting a function to the point spread function. The central limit theorem states that the mean $X_n$ of $n$ measurements follows a Gaussian distribution, whatever the distribution of the individual measurements. A Gaussian curve is, therefore, often fitted to experimental measurements to find the variance of the measurements about their mean. The point spread function is a result of a non-ideal imaging system and is due to many factors. Random measurement errors play their part but so does the image reconstruction technique used. In fact, this work is concerned with the differences in blurring between image reconstruction techniques.

It is difficult to decide on a better function to fit to the point spread functions so, Gaussian functions will be used. A Gaussian curve can be fitted to a line of data through the middle of the reconstructed point in each of the three directions $(x, y, p)$ separately. A one-dimensional Gaussian curve is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-|x-\mu|^2}{2\sigma^2} \right\} \tag{5.6}$$

where $p(x)$ is the amplitude of the point at position $x$, $\mu$ is the position of the centre of the point and $\sigma^2$ is the variance. The variance of this fitted curve is a measure of the spatial resolution.

### 5.2.2 Line Spread Function (LSF) and Edge Spread Function (ESF)

The line spread function and the edge spread function are the responses of the system to a line and an edge respectively. The LSF could be investigated by placing a line source in a plane parallel to the detector but the ESF would be more difficult to determine because an uniform source which entirely filled a portion of the field of view would be needed. The edge would be the edge of this extended source.

## 5.3 Noise

For simplicity, the difference between a reconstructed image and the ideal image will be referred to as “noise”. As discussed before, noise comes from three main sources, statistical fluctuations of radioactive decay, the spatial resolution of the gamma camera and errors intrinsic to the reconstruction process and storage of the hologram using a computer.

### 5.3.1 Variance of an Image of a Uniform Source

The image of a uniform source should be uniform. The variance of the intensity of this image gives a measure of output noise. This, however, does not take the spatial correlation of the noise into account. As was shown in equation 5.3, there are two types of noise, source noise $N_s$ and recording noise $N_a$. The variance of an image of a uniform source gives no indication of the relative effects of these two different types of noise.

### 5.3.2 Wiener Spectrum (W)

The Wiener spectrum measures the noise power as a function of spatial frequency. It is given by the Fourier transform of the noise autocorrelation function. The total noise $N_t(\boldsymbol{x'})$ is given by the image minus the ideal image. If $\mathcal{R}\{\}$ is a function describing the reconstruction process, the total noise for the simplified case in equation 5.3 will be

$$N_t(\boldsymbol{x'}) = \mathcal{R}\left\{C(p_0, q, D)\left[S'(\boldsymbol{x'}) + N_s(\boldsymbol{x'})\right] \otimes T'(\boldsymbol{x'}) + N_a(\boldsymbol{x'})\right\} - S'(\boldsymbol{x'}) \qquad (5.7)$$

The noise autocorrelation function,$A_N(\boldsymbol{x'})$, is given by

$$A_N(\boldsymbol{x'}) = \int N_t(\boldsymbol{\tau})N_t(\boldsymbol{x'} + \boldsymbol{\tau})d\boldsymbol{\tau} \qquad (5.8)$$

The Wiener-Khintchine theorem states that the autocorrelation of a function $F(\boldsymbol{x})$ is simply given by the Fourier transform of the absolute square of $F^F(\omega)$ where $F^F(\boldsymbol{\omega})$ is the Fourier transform of $F(\boldsymbol{x})$ [69]. If $N_t^F(\boldsymbol{\omega})$ is the Fourier transform of $N_t(\boldsymbol{x'})$, equation 5.8 can be written as

$$A_N(\boldsymbol{x'}) = \mathcal{F}\left\{|N_t^F(\boldsymbol{\omega})|^2\right\} \qquad (5.9)$$

where $\mathcal{F}\{\}$ denotes a Fourier transform. The Wiener spectrum $\mathrm{W}(\boldsymbol{\omega})$ is simply given by

$$\mathrm{W}(\boldsymbol{\omega}) = |N_t^F(\boldsymbol{\omega})|^2 \qquad (5.10)$$

The amount of noise at different frequencies can be seen. This spectrum is easy to calculate when the exact source distribution and therefore the ideal image is known. Noise due to statistical fluctuations as well as that due to the gamma camera can be simulated and added to a computer generated hologram. The relative contribution of the two types of noise can therefore be compared for images reconstructed using different methods.

### 5.3.3 Difference Metrics

#### Mean Squared Error (MSE) and Signal to Noise Ratio (SNR)

A commonly used measure of the error between an output image and the ideal image is given by the mean squared error. If the ideal image is $S'(\boldsymbol{x'})$, as before, and the reconstructed image of the source is $S_R(\boldsymbol{x'})$, the mean squared error is given by

$$\mathrm{MSE} = \frac{1}{N_p}\sum\left|S_R(\boldsymbol{x'}) - S'(\boldsymbol{x'})\right|^2 \qquad (5.11)$$

where $N_p$ is the total number of pixels and the sum is over all of them. This metric depends upon the scale of $S'(\boldsymbol{x'})$, you will get a larger error for a larger source activity. It is therefore better to use a normalised error NMSE.

$$\mathrm{NMSE} = \frac{\sum |S_R(\boldsymbol{x'}) - S'(\boldsymbol{x'})|^2}{\sum |S'(\boldsymbol{x'})|^2} \tag{5.12}$$

this is called the normalised mean-square error metric. Its square root, $E$ is called the normalised root-mean-square error (NRMSE).

The signal to noise ratio (SNR) is defined in decibels as

$$\mathrm{SNR} = 10 \log_{10} \left( \frac{1}{\mathrm{NMSE}} \right) \tag{5.13}$$

**Fienup Error Metric (FEM)**

If the reconstructed image is shifted by a few pixels with respect to the ideal image or is multiplied by a constant factor, the image will look good but the MSE will be huge. This is a major limitation of mean-squared error metrics, they fail to differentiate between different types of error.

Fienup [27] derived an expression for a metric which is invariant to a shift and multiplication by a constant factor. This error metric will be referred to as the Fienup error metric (FEM).

$$\mathrm{FEM}^2 = 1 - \frac{\max_{\boldsymbol{x_s}} |r_{S'S_R}(\boldsymbol{x_s})|^2}{r_{S'S'}(\mathbf{0}) r_{S_R S_R}(\mathbf{0})} \tag{5.14}$$

$r_{S'S_R}(\boldsymbol{x_s})$ is the cross correlation of $S'(\boldsymbol{x'})$ and $S_R(\boldsymbol{x'})$, $r_{S'S'}(\boldsymbol{x_s})$ is the autocorrelation of $S'(\boldsymbol{x'})$ and $r_{S_R S_R}(\boldsymbol{x_s})$ is the autocorrelation of $S_R(\boldsymbol{x'})$. $\max_{\boldsymbol{x_s}} |r_{S'S_R}(\boldsymbol{x_s})|$ is the maximum value of $r_{S'S_R}(\boldsymbol{x_s})$ for all values of $\boldsymbol{x_s}$.

## 5.4 Summary

In this Chapter, methods for assessing image quality were reviewed. Measures of image quality based on the system transfer function as well as spatial resolution and signal to noise ratios were given. These methods depend upon having an ideal reference image with which the reconstructed image can be compared. It is therefore easier to first assess the effectiveness of image reconstruction methods by using computer generated holograms.

The OTF/MTF can be calculated to investigate the frequency response of zone plate holography using theoretical point sources. The PSF can be calculated to compare the resolution of the different reconstruction techniques. It is felt that LSF and ESF are more suited to experimental measurement. MSE, NMSE and FEM error metrics can be used to measure the amount of error in the images.

When images are reconstructed from real holograms, an ideal reference image is not available. The reason for recording the hologram is to determine the distribution of gamma-ray sources from the reconstructed image. An estimation of the hologram that would have been generated by a reconstructed source distribution can, however, be calculated and this can be compared with the recorded hologram.

# Chapter 6

# IMAGE RECONSTRUCTION BY CONVOLUTION

In this Chapter the method of reconstruction by convolution will be discussed. Reconstruction by convolution with a Gabor zone plate pattern has been used by many people at the University of Birmingham to reconstruct images from binary Gabor zone plate holograms. Mast [57], Mathews [58], Kirk [46], Woodgate [92], Strange [83], Kuo [49], Caplan [16], Perks [65], Rew [70] and Shen [77] have all used this technique and they referred to it as the correlation reconstruction method. This method involved correlating the hologram with an appropriately scaled zone plate pattern to reconstruct the image in one plane parallel to the zone plate. The same effect is obtained if the hologram is convolved with a version of the same zone plate pattern which has been flipped about lines through the centre in both the $x$ and $y$ directions. As the zone plate is symmetrical in both the $x$ and $y$ directions, the flipped plate is identical to the original plate. Correlation and convolution, therefore, produce the same result. Thinking of the technique as convolution rather than correlation makes the relationship between the Fourier transforms of the hologram, zone plate and image more straightforward. The form of convolution reconstruction which has been used the most is simulated diffraction. This method will be outlined and different convolution reconstruction techniques will be suggested. An improvement to the convolution technique, which will be called *Normalised Convolution Reconstruction*, will also be presented here.

An expression for the relationship between source distribution, $S(\epsilon, p)$, and hologram intensity, $U(\boldsymbol{x})$, was given in equation 3.46 as follows.

$$U(\boldsymbol{x}) = \frac{D^2}{4\pi} \int \frac{1}{(p+q)^2} \int \gamma(\epsilon, p, S(\epsilon, p)) S(\epsilon, p)\, T\left(\frac{\boldsymbol{x}}{M(p,q)} - L(p,q)\epsilon\right) d\epsilon dp \tag{6.1}$$

$T(\boldsymbol{x})$ was the intensity transmittance of the zone plate. As was explained in Chapter 3,

$\gamma(\boldsymbol{\epsilon}, p, S(\boldsymbol{\epsilon}, p))$ is a factor which depends upon the position of a point in the source distribution, $\boldsymbol{\epsilon}, p$, as well as the rest of the source distribution in relation to it. This work will concentrate on methods that do not include this factor. That is, it will be assumed that $\gamma(\boldsymbol{\epsilon}, p, S(\boldsymbol{\epsilon}, p)) = 1$ over all of $\boldsymbol{\epsilon}, p$. This approximation can be used if the source is two dimensional or sufficiently small so that the gamma rays do not have to pass through other parts of the source to reach the detector. This factor can also be neglected if the source has a low linear attenuation coefficient. Linear attenuation coefficients were discussed in Section 3.8.1. If a large, dense, source was used, this factor would have to be considered. Equally, if this technique was used in nuclear medicine to image brain tumours, the gamma rays would have to pass through the brain tissue and the skull before reaching the detector and this would also have to be accounted for. For the simple computer generated sources and real sources which have been used in this work, however, this approximation is sufficient.

The gamma camera records the hologram as a square array of $N \times N$ pixels where $N$ is usually a power of 2, that is, 64, 128, 256, 512 or 1024. The hologram is also to be reconstructed in a series of $N_p$ planes at distances $p_0, p_1, p_2 ... p_{N_p - 1}$ from the zone plate and parallel to the zone plate and gamma camera. The space in which the object lies can be thought of not only as a series of planes but also as voxels which increase in size with distance from the camera. Each voxel corresponds to a region of space from which a gamma ray passing through the centre of the zone plate will be detected in one particular pixel by the gamma camera. The source distribution can be represented by average values for each voxel and the zone plate transmission function can be represented by pixels of the appropriate size. This is illustrated in Figure 6.1.

The images generally have 360×360 pixels in each ($x$,$y$) plane. To reconstruct a truly three dimensional image, the voxels would have to have a similar size in the $z$ direction as in each ($x$,$y$) plane. This would mean reconstructing 360 planes in the $z$ direction, or some other similar number, and would require a lot of computation. In the work presented here, the image is reconstructed in a small number of parallel planes rather than a set of voxels which completely fill the field of view of the detector. This is a sensible approach as most of the sources which will be used are two dimensional and placed parallel to the detector. We can then see if the image just appears in the correct plane or if there are artifacts in the surrounding planes.

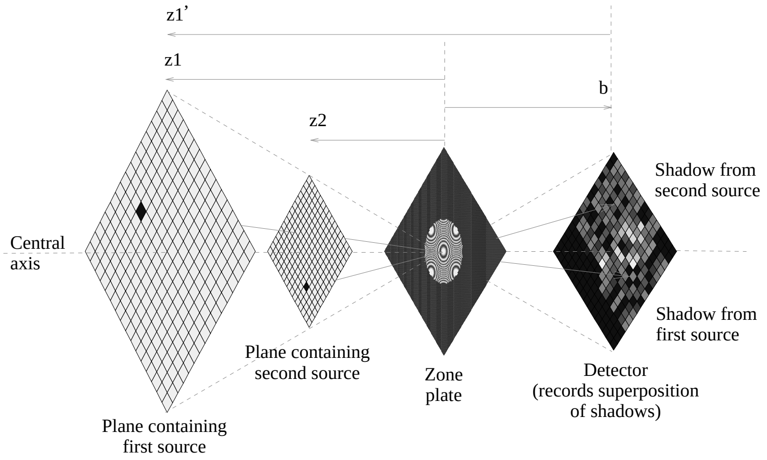


Figure 6.1: Illustration of field of view of detector with zone plate attached

Equation 6.1 can be written as a summation rather than a continuous integral. This helps us to understand the value stored in each hologram pixel.

$$U_{i'j'} = \sum_{k=0}^{N_p-1} C^k \sum_{i,j=-\frac{N}{2}}^{\frac{N}{2}-1} S^k_{ij} T^k_{(i'-i)(j'-j)} \quad \text{where} \quad U_{ij} = U(i\Delta_p, j\Delta_p) \quad C^k = \frac{D^2}{4\pi}\frac{1}{(p_k+q)^2}$$

$$S^k_{ij} = S(i\frac{\Delta_p}{L(p_k,q)}, j\frac{\Delta_p}{L(p_k,q)}, p_k) \quad \text{and} \quad T^k_{ij} = T(i\frac{\Delta_p}{M(p_k,q)}, j\frac{\Delta_p}{M(p_k,q)}) \tag{6.2}$$

The factors $D$, $q$, $L(p,q)$ and $M(p,q)$ are as defined in Chapter 3. As $-\frac{N}{2} \leq i', j' \leq \frac{N}{2}-1$, values of $T^k_{ij}$ outside $-\frac{N}{2} \leq i,j \leq \frac{N}{2}-1$ are needed. This problem is overcome by padding $T^k_{ij}$ with zeros to twice its original size, that is

$$\text{padded} \quad T^k_{ij} = \begin{cases} T^k_{ij} & -\frac{N}{2} \leq i,j \leq \frac{N}{2}-1 \\ 0 & -N \leq i,j \leq -\frac{N}{2}-1 \\ 0 & \frac{N}{2} \leq i,j \leq N-1 \end{cases} \tag{6.3}$$

In practice, three fast Fourier transforms are used to perform a convolution. The Fourier transforms are performed in the way described in Section 4.4. When a theoretical hologram is calculated, both $T^k_{ij}$ and $S^k_{ij}$ are padded as above. If this is not done, the convolution becomes circular. In this case, if $i = -\frac{N}{2}-1$, for example, this value does not exist so the value at $i = \frac{N}{2}-1$ is taken instead.

## 6.1 Optical Reconstruction - how Convolution Algorithms can be used to Simulate Diffraction

In Section 3.3.3, the diffraction pattern produced when plane waves are shone through a transparency with transmittance $t(\boldsymbol{x})$ was given as

$$u(\boldsymbol{x}, q_r) = \frac{C_r^q}{2\pi i} t(\boldsymbol{x}) \otimes z^{q_r}(\boldsymbol{x}) \qquad \text{where}$$
$$z^{q_r}(\boldsymbol{x}) = 2\omega_r^2 e^{i\omega_r^2 \boldsymbol{x}^2} \qquad C_r = A_r e^{i\frac{2\pi(p_r+q_r)}{\lambda_r}} \qquad \text{and} \qquad \omega_r^2 = \frac{\pi}{\lambda_r q_r} \tag{6.4}$$

when $\delta < l < a$. As was stated before, $\delta = \frac{1}{\Omega_0}$ where $\Omega_0$ is the maximum spatial frequency in $t(\boldsymbol{x})$, $l = \sqrt{\lambda_r q_r}$ is the size of the central Fresnel zone, and $a$ is the total size of the pattern represented by $t(\boldsymbol{x})$. As before, $q_r$ is the distance from the transparency to the detector, $\lambda_r$ is the wavelength of the light used in the reconstruction and $A_r$ and $p_r$ are constants representing the amplitude and phase of the light.

If the hologram were recorded using an ideal Gabor zone plate, it will contain a constant term plus a Fresnel function of frequency $\omega_{zp} = \sqrt{\pi} M(p, q) r_0$. As was shown in Section 3.8.1, if $\delta < l < a$, the object distribution that was in plane $p$ will be reconstructed a distance $q_r$ behind the hologram where

$$q_r = \frac{(p+q)^2 r_0^2}{\lambda_r p^2} \tag{6.5}$$

If the image was reconstructed using a ruby laser, the wavelength of light would be $\lambda_r$ =633nm [73]. Values which are commonly used for gamma-ray zone plate holograms are $p$ =0.2m, $q$ =0.2m and $r_0$ =0.01m. If the hologram was printed out on film the same size as the detector, the distance of the image behind the film would be $q_r$ =632m and the radius of the central Fresnel zone would be $l$ =0.02m. The hologram would have to be reduced in size to focus the image in a more reasonable position. If the hologram was reduced by a factor $R_s$, this would give

$$q_r = \frac{(p+q)^2 r_0^2}{\lambda_r p^2 R_s^2} \tag{6.6}$$

If diffraction is simulated using a computer by convolving the hologram with the appropriate $z^{q_r}(\boldsymbol{x})$ pattern for one $(x, y)$ plane at a time, the value of $\lambda_r$ is not important. In this case, $\omega_r$ is set equal to the expected value of $\omega_{zp}$ for an object in the appropriate image plane. Equations for the calculation of $\omega_{zp}$ were given in Section 3.5.1. As in equation 6.2, we can write the equation for the reconstructed image in plane $q_r$, $V_{ij}^{q_r}$, as a summation.

$$V_{i'j'}^{q_r} = \sum_{i,j=-\frac{N}{2}}^{\frac{N}{2}-1} U_{ij} Z_{(i'-i)(j'-j)}^{q_r} \tag{6.7}$$

$U_{ij}$ represents the hologram pixel values and $Z_{ij}^{q_r}$ represents the Fresnel function ($z^{q_r}(\boldsymbol{x})$) used for the reconstruction. Both $U_{ij}$ and $Z_{ij}^{q_r}$ are padded as in equation 6.3.

To illustrate reconstruction be simulated diffraction, a computer generated hologram of a point source with unit strength per unit area will be used. The zone plate hologram values will be those in Table 6.1. These are typical parameters for a zone plate hologram. The parameters of the appropriate reconstructing Fresnel function can be calculated the using the equations given in Section 3.6.1.

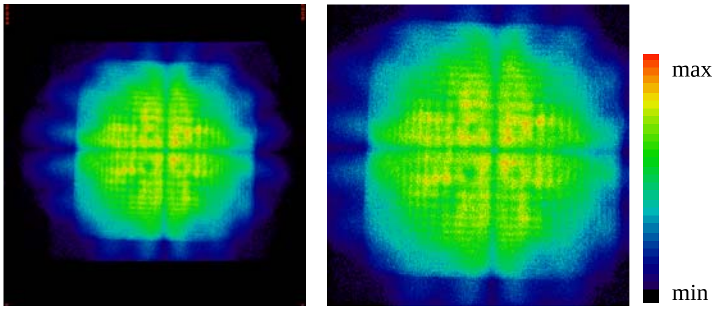
| | | | |
|---|---|---|---|
| First zone radius ($r_1$) | 1.02e-02m | Source → zone plate distance ($z$) | 0.22m |
| Detector pixel size ($\Delta$) | 1.074e-03m | Zone plate → detector distance ($b$) | 0.36m |
| Number of pixels ($n_x \times n_y$) | 360 × 360 | Number of zones ($n = \sqrt{\frac{R}{r_1}}$) | 30 |

Table 6.1: Hologram parameters used for the theoretical hologram used in most of this Chapter

These parameters are typical for actual gamma ray zone plate holograms. $R$ is the maximum radius of the zone plate, it is totally opaque outside this radius. Although the detector stores data in a 512×512 pixel array, it is only the central 360×360 which represent the active region of the detector. A real experimental hologram is shown in Figure 6.2 to illustrate this. The point can be reconstructed by convolving the zone plate pattern with the Fresnel function. For simplicity, the parameters of the reconstructing light are set so $\frac{C_r^q}{2\pi i} 2\omega_r^2 = 1$. The reconstructed point can be seen in Figure 6.3.

## 6.2 Using the Zone Plate Pattern as a Matched Filter

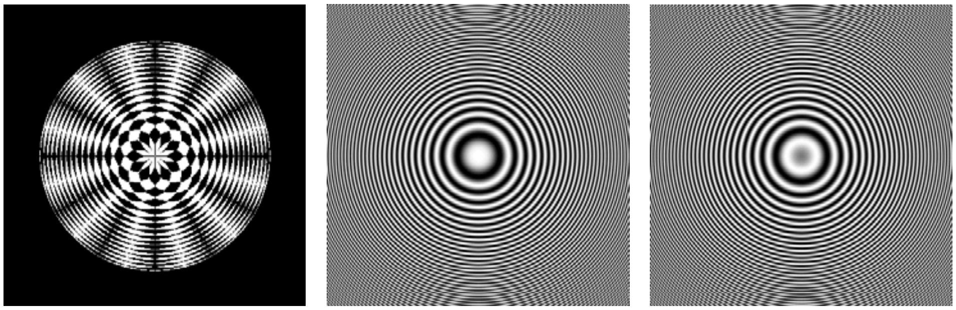
A pure Gabor zone plate (GZP) is not used to record zone plate holograms, a binary plate is used instead. As was mentioned in Chapter 2, this plate has the same average intensity transmittance at each radius as the pure Gabor plate but the plate is either transparent or opaque at each point. The advantage that this has over the pure plate is that it is much easier to manufacture. Holes just have to be cut in a metal sheet of constant thickness.



(a) a hologram as recorded by the gamma camera

(b) the central 360×360 pixels

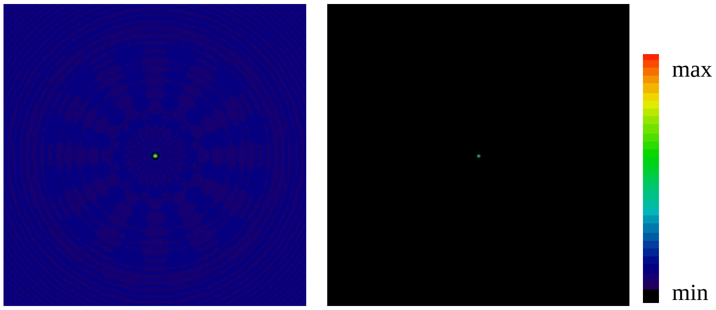
Figure 6.2: Illustration of the useful hologram area



(a) simulated hologram of point source 0≤pixel value≤1.1754e-13

(b) Fresnel function - real part -1≤pixel value≤1

(c) Fresnel function - imaginary part - 1≤pixel value≤1



(d) reconstructed point - real part -2.2890e-10≤pixel value≤1.7175e-09

(e) reconstructed point - intensity 1.9132e-28≤pixel value≤2.9498e-18

Figure 6.3: Theoretical reconstruction of a point source by simulated diffraction

Another advantage of a binary Gabor zone plate over a pure Gabor zone plate is that it transmits more light if it is used as a lens. If the zone plate patterns are recorded on photographic film so that the amplitude transmittance of the film is the same as the zone plate, the total intensity transmittance will be greater for the binary plate than for the pure Gabor plate lens.

The amplitude transmittance of a pure Gabor zone plate, $t(r)_{\mathrm{GZP}}$ is

$$t(r)_{\mathrm{GZP}} = \frac{1}{2}\left[1+\cos\left(\frac{\pi r^2}{r_0^2}\right)\right] \tag{6.8}$$

where, as in Chapter 3, $r$ is the radial distance from the centre of the plate and $r_0$ is a constant. A binary Gabor zone plate (BGZP) has the same average transmittance for each radius as the the pure plate. The amplitude transmittance for a binary Gabor zone plate is either zero or one at any one position. Whether it is zero or one depends upon the number of sectors in the plate. This is illustrated in Figure 6.4. The average amplitude and intensity transmittance of the plate is unaffected by the number of zones. To simplify matters, a binary plate with just one zone will be considered. In this case, the amplitude transmittance of a binary Gabor zone plate, $t(r,\theta)_{\mathrm{BGZP}}$ is

$$t(r,\theta)_{\mathrm{BGZP}} = \begin{cases} 1 & \theta \le 2\pi t(r)_{\mathrm{GZP}} \\ 0 & \theta > 2\pi t(r)_{\mathrm{GZP}} \end{cases} \tag{6.9}$$

The average intensity transmittance (AIT) of a pure Gabor zone plate is given by

$$\begin{aligned} \mathrm{AIT}_{\mathrm{GZP}} &= \frac{\int_0^R \int_0^{2\pi} t(r)^2_{\mathrm{GZP}} r\,d\theta\,dr}{\pi R^2} = \frac{\int_0^R \int_0^{2\pi} \left[\frac{1}{2}\left[1+\cos\left(\frac{\pi r^2}{r_0^2}\right)\right]\right]^2 r\,d\theta\,dr}{\pi R^2} \\ &= \frac{2\pi \int_0^R \left[\frac{1}{2}\left[1+\cos\left(\frac{\pi r^2}{r_0^2}\right)\right]\right]^2 r\,dr}{\pi R^2} \end{aligned} \tag{6.10}$$

where $R$ is the maximum radius of the lens. The average intensity transmittance of a binary Gabor zone plate is given by

$$\begin{aligned} \mathrm{AIT}_{\mathrm{BGZP}} &= \frac{\int_0^R \int_0^{2\pi} t(r,\theta)^2_{\mathrm{BGZP}} r\,d\theta\,dr}{\pi R^2} = \frac{\int_0^R \int_0^{2\pi t(r)_{\mathrm{GZP}}} r\,d\theta\,dr}{\pi R^2} \\ &= \frac{2\pi \int_0^R \frac{1}{2}\left[1+\cos\left(\frac{\pi r^2}{r_0^2}\right)\right] r\,dr}{\pi R^2} \end{aligned} \tag{6.11}$$

If the maximum radius of the lens, $R$, is set to $R = \sqrt{n} r_0$ where $n$ is an integer, the integrals in equations 6.10 and 6.11 simplify to the following numbers

$$\mathrm{AIT}_{\mathrm{GZP}} = \frac{3}{8} \qquad \mathrm{AIT}_{\mathrm{BGZP}} = \frac{1}{2} \tag{6.12}$$

Both zone plate patterns would focus light at the same position. As the binary zone plate pattern has a higher average intensity transmittance, the intensity of light at the focal point could be expected to be brighter than the intensity if a pure Gabor plate was used. Further properties of binary zone plates are discussed by Beynon and Strange [11].

A linear filter, $L(\boldsymbol{x})$, can be used to reconstruct an image from the hologram by convolving the hologram with the filter. Objects which were at different distances ($p$) behind in front of the zone plate can be reconstructed by using a zone plate pattern scaled to the size of the shadow cast by a point at a distance $p$.

$$V^{p}_{i'j'} = \sum_{i,j=-\frac{N}{2}}^{\frac{N}{2}-1} U_{ij} L^{p}_{(i'-i)(j'-j)} \tag{6.13}$$

Again, $U_{ij}$ represents the hologram values and $V^{p}_{ij}$ the reconstructed image values at a distance $p$ in front of the zone plate. The similarity of this equation to equation 6.7 may be noted. As Jain [43] proves, the linear filter which will produce an image with the highest signal to noise ratio is simply the hologram of a single point. In this case, this will be a magnified version of the binary plate pattern. Jain calls the optimum linear filter a matched filter.

If the binary plate pattern is to be used in a reconstruction algorithm, it needs to be generated using an algorithm. An example of a binary zone plate can be seen in Figure 2.3. Figure 6.4 illustrates how the pattern is generated. If the expected hologram

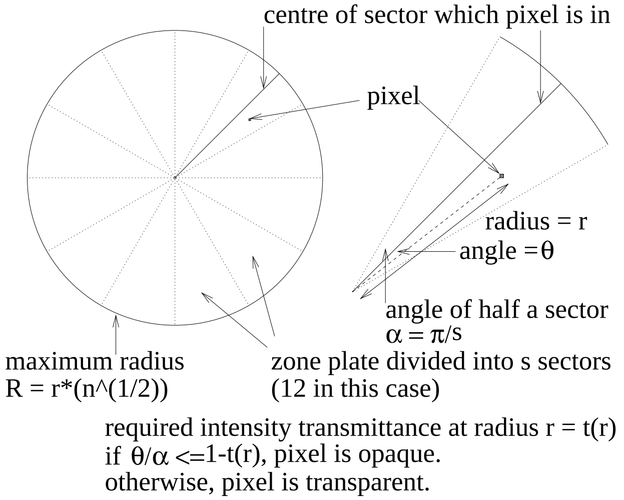
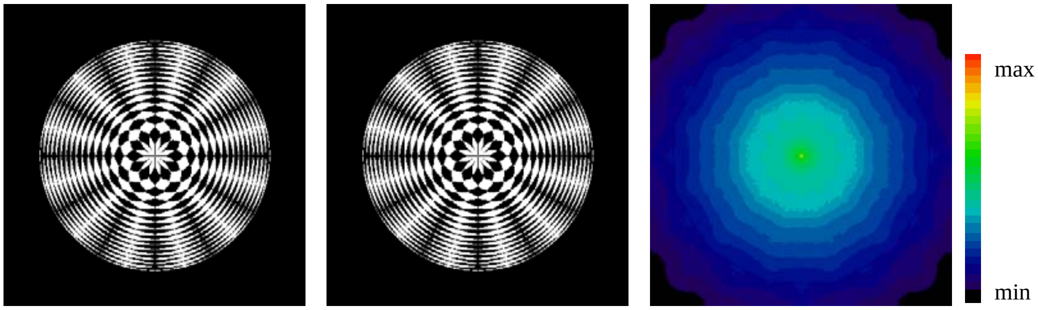


Figure 6.4: Calculation of the binary zone plate pattern

of a point source is convolved with itself, the point should be reconstructed. The hologram from Figure 6.3 was reconstructed using a matched filter and the result can be seen in Figure 6.5. The reconstructed point in Figure 6.5(c) sits on top of an error term which steadily increases towards the centre of the image. Fenimore and Cannon [26] discovered a similar effect when using a matched filter to reconstruct images generated using uniformly redundant arrays (URAs). URAs are binary patterns which, like zone plates, can be used to obtain extra information about light sources by detecting the shadow cast by a source through the pattern. Again, like zone plates, they have an autocorrelation which approximates a delta function. Fenimore and Cannon removed the 'pyramid' error term from their images by changing the values assigned to white and black pixels in their matched filter. This new filter was said to be 'balanced'. The balanced filter (new $L^p_{ij}$) can be calculated from the old filter (old $L^p_{ij}$) in the following way.



(a) simulated hologram of point source $0 \leq$pixel value$\leq 1.1754\text{e-}13$

(b) matched filter $0 \leq$pixel value$\leq 1$

(c) reconstructed point $4.7487\text{e-}11 \leq$pixel value$\leq 3.4568\text{e-}09$

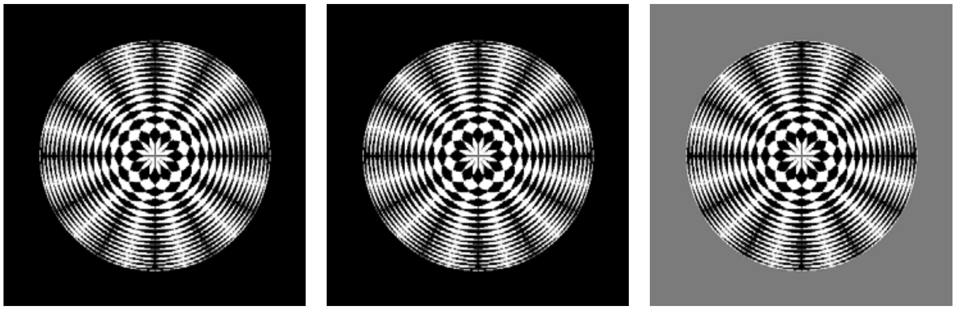
Figure 6.5: Theoretical reconstruction of a point source using a matched filter

$$\text{new } L^p_{ij} = \begin{cases} 1 & \text{if old } L^p_{ij} = 1 \\ -\frac{\rho}{1-\rho} & \text{if old } L^p_{ij} = 0 \end{cases} \tag{6.14}$$

$\rho$ is the ratio of transparent (white) pixels to total pixels. Fenimore and Cannon compared this to Brown's mismatch method where the new value of $L^p_{ij}$ is set to $= -1$ if its old value was zero. Fenimore and Cannon's method is the same as this if the number of white and black pixels are equal. A binary zone plate has approximately equal numbers of white and black pixels for $\sqrt{x^2+y^2} \leq R$ but all the pixels are opaque (black) outside this region. A modified mismatch method is, therefore, suggested. This method uses Brown's method for $\sqrt{x^2+y^2} \leq R$, but, outside $R$, all pixels are set to zero. That is,

$$\text{new } L^p_{ij} = \begin{cases} \begin{cases} 1 & \text{if old } L^p_{ij} = 1 \\ -1 & \text{if old } L^p_{ij} = 0 \end{cases} & \text{if } \sqrt{x^2+y^2} \leq R \\ 0 & \text{if } \sqrt{x^2+y^2} > R \end{cases} \tag{6.15}$$

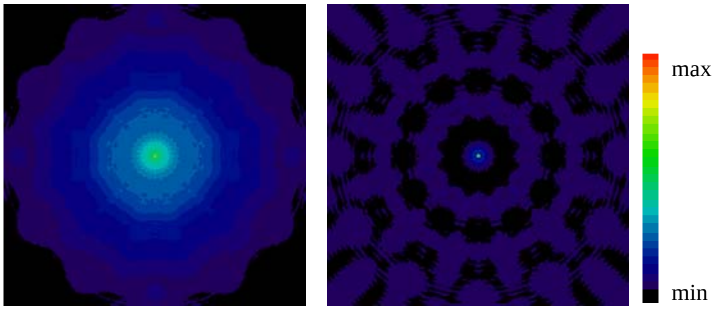
Reconstructed points using both these methods are shown in Figure 6.6. Values taken in a



(a) simulated hologram of point source 0≤pixel value≤1.1754e-13

(b) filter for balanced convolution -0.29353≤pixel value≤1

(c) filter for modified mismatch -1≤pixel value≤1



(d) reconstructed point using balanced convolution -2.1048e-10≤pixel value≤3.4568e-09

(e) reconstructed point using modified mismatch -2.0558e-10≤pixel value≤3.4568e-09

Figure 6.6: Theoretical reconstruction of a point source using balanced convolution and modified mismatch methods

diagonal line through the centre of points reconstructed using the matched filter, balanced convolution and modified mismatch methods are compared in Figure 6.7. The modified mismatch method is closest to zero around the point and, therefore, appears to be the best one of these filters. These methods will be compared using the image assessment methods from Chapter 5 in Sections 6.5, 6.6 6.7 and 6.8.

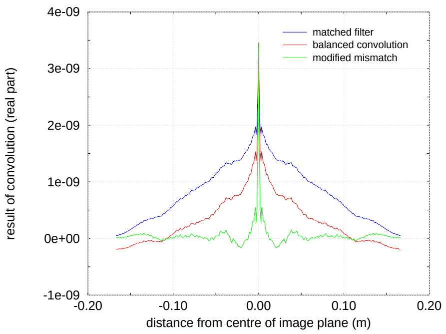


Figure 6.7: Diagonal line taken through images of points reconstructed using filters based upon the binary zone plate pattern

## 6.3 Hybrid Method - using $R$ to improve Simulated Diffraction

As the zone plate pattern is truncated at a maximum radius $R$, it makes sense to truncate the Fresnel function used in simulated diffraction at the same radius. The results are shown in Figure 6.8. A diagonal line of data taken through the real part of the reconstructed point is shown together with the same data from the standard simulated diffraction reconstruction in Figure 6.9 Looking at this figure, it appears that the effect is negligible for a point source. Again, these methods will be compared using the image assessment methods from Chapter 5 in Sections 6.5, 6.6 6.7

## 6.4 Normalised Convolution

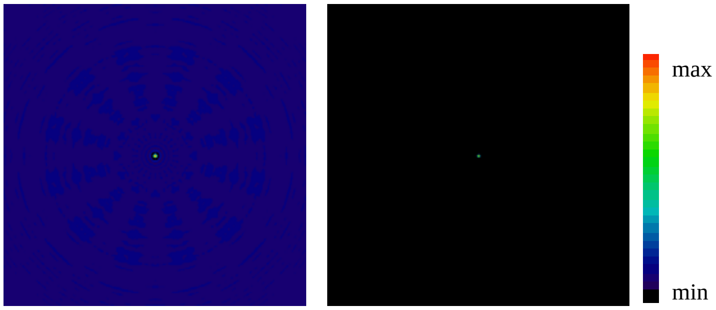
If a point is reconstructed in one plane using a linear filter, the value of the convolution where the point is reconstructed depends upon the portion of the zone plate shadow which is actually ‘seen’ by the detector.

In Section 3.4 the term $\frac{D^2}{4\pi(p+q)^2}$ was used to approximate the fraction of the emitted radiation which was recorded by the detector. As before, $D^2$ is the area of the detector. If the detector had $N \times N$ pixels, and the hologram was recorded in each of these pixels, the count in each pixel would be $\frac{D^2}{4\pi(p+q)^2 N^2} S_{ij}^p \tau$ where $S_{ij}^p$ is the appropriate source intensity and $\tau$ is the data acquisition time. However, the size of the pixels varies with the source to zone plate distance $p$. It, therefore, makes sense to calculate source strength per unit area instead of total source strength in each pixel. That is, the value reconstructed in

(a) simulated hologram of point source $0 \leq$ pixel value $\leq$ 1.1754e-13

(b) Fresnel function - real part $-1 \leq$ pixel value $\leq 1$

(c) Fresnel function - imaginary part - $1 \leq$ pixel value $\leq 1$



(d) reconstructed point - real part $-2.1476\text{e-}10 \leq$ pixel value $\leq 1.7175\text{e-}09$

(e) reconstructed point - intensity $7.0481\text{e-}29 \leq$ pixel value $\leq 2.9498\text{e-}18$

Figure 6.8: Theoretical reconstruction of a point source by convolution with a truncated Fresnel function

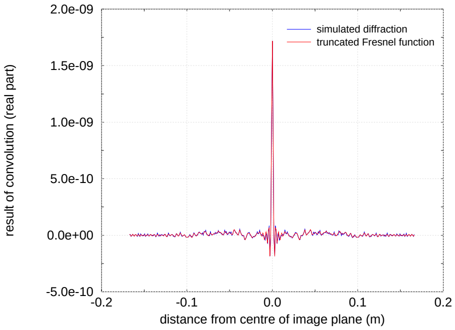
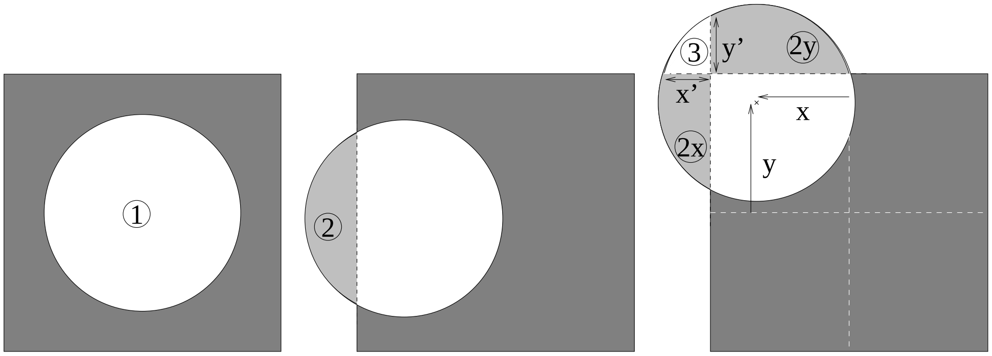


Figure 6.9: Diagonal line taken through images of points reconstructed using filters based upon the binary zone plate pattern

each pixel will be $\frac{S^p_{ij}\tau}{\Delta_p^2}$ where $\Delta_p$ is the size of a pixel in plane $p$.

If the point is reconstructed using a filter based upon the binary zone plate pattern, the value of the convolution at the point ($V^p_{ij}$) will be the number of the pixels which recorded a value (i.e: the number of transparent pixels in the correctly scaled zone plate pattern) times the count in each pixel. The aim is to recover $S^p_{ij}\tau$ from this information. The hologram is, however, only recorded in pixels corresponding to holes in the zone plate pattern. The zone plate pattern could be totally recorded by the detector, overlap in the $x$ or $y$ direction or both. This is illustrated in Figure 6.10. We will assume that $p$ and $q$ have been chosen so that the whole shadow fits on the detector when it is in the centre, i.e: that it can't overlap two opposite edges of the detector at once. If $x$ and $y$ are distances from the centre of the detector, $x'$ and $y'$, which are shown in Figure 6.10, are given by



(a) all the shadow is recorded

(b) the shadow overlaps in one direction

(c) the shadow overlaps in both directions

Figure 6.10: Illustration of the zone plate pattern overlapping the edge of the detector

$$x' = \sqrt{R^2 - \left(\frac{D}{2} - y\right)^2} - \frac{D}{2} + x \qquad y' = \sqrt{R^2 - \left(\frac{D}{2} - x\right)^2} - \frac{D}{2} + y \tag{6.16}$$

The areas ①, $②_x$, $②_y$ and ③ can be calculated as follows

$$\begin{aligned}
① &= \pi R^2 \\
②_x &= R^2 \arccos\left(\frac{\frac{D}{2} - x}{R}\right) - \left(\frac{D}{2} - x\right)\sqrt{R^2 - \left(\frac{D}{2} - x\right)^2} \\
②_y &= R^2 \arccos\left(\frac{\frac{D}{2} - y}{R}\right) - \left(\frac{D}{2} - y\right)\sqrt{R^2 - \left(\frac{D}{2} - y\right)^2} \\
③ &= R^2 \arcsin\left(\sqrt{x'^2 + y'^2}\right) - \frac{\sqrt{x'^2 + y'^2}}{4}\sqrt{4R^2 - x'^2 - y'^2} + \frac{x'y'}{2}
\end{aligned} \tag{6.17}$$

The area of the zone plate shadow seen by the detector, $A^p_{ij}$, is then calculated as follows

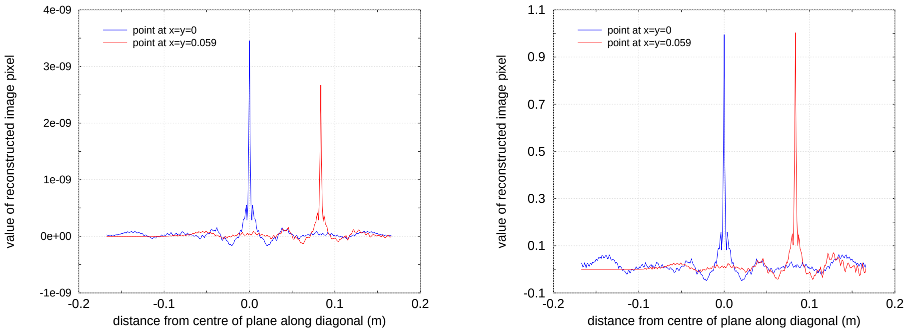
$$A^p_{ij} = \begin{cases} ① - ②_x - ②_y + ③ & \text{if } x' \text{ and } y' \text{ are both real and positive} \\ ① - ②_x - ②_y & \text{if } x' \text{ and } y' \text{ are both real and negative} \\ ① - ②_x & \text{if } x' \text{ is a complex number and } y' \text{ is real and negative} \\ ① - ②_y & \text{if } y' \text{ is a complex number and } x' \text{ is real and negative} \\ ① & \text{if } x' \text{ and } y' \text{ are both complex numbers} \end{cases} \tag{6.18}$$

A hologram generated using a binary plate has approximately the same number of 'white' as 'black' pixels inside the maximum radius of the plate shadow so will assume that the area of the detector which the radiation hits is approximately half the above value. Using this approximation, the value of the convolution for the reconstructed point can be written as

$$V^p_{ij} = \frac{A^p_{ij}}{8\pi(p+q)^2} S^p_{ij}\tau \quad \text{and so} \quad \frac{S^p_{ij}\tau}{\Delta^2_p} = \frac{8\pi(p+q)^2 V^p_{ij}}{A^p_{ij}\Delta^2_p} \tag{6.19}$$

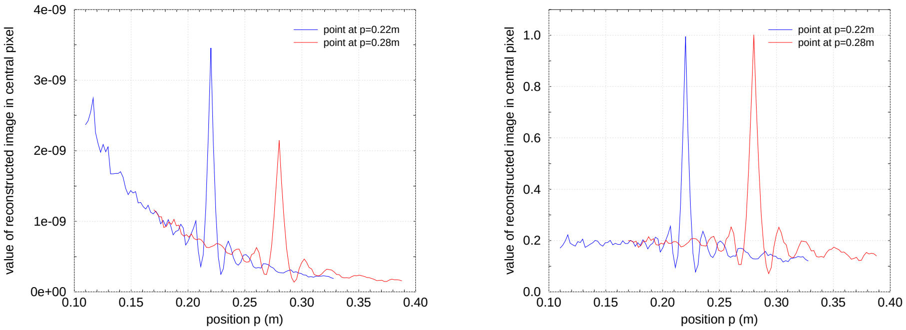
The above equation depends upon $p$ and therefore corrects for different distances in the $z$ ($p$) direction. Figure 6.11 shows the effect of the normalisation in the $z$ as well as the $(x, y)$ direction. As before, holograms were simulated for point source with $\frac{S^p_{ij}\tau}{\Delta^2_p} = 1$ and the points were reconstructed using the modified mismatch method. When the reconstruction was not normalised, the reconstructed points had varying amplitude. Normalisation made the amplitude of each point $= 1$ within a small margin of error. The errors result from the number of black and white pixels in the zone plate shadow not being exactly equal for all positions. Normalisation of simulated diffraction or convolution with a truncated Fresnel function is, however, more complicated. Simulated diffraction relies upon the Fresnel function extending to infinity in the $x$ and $y$ directions. As it is truncated by either the size of the detector or the maximum radius of the zone plate, the calculated amplitudes are affected. As the Fresnel function is smoothly varying between $+1$ and $-1$, so we would expect a reconstructed point to have a smaller amplitude than if it were reconstructed using a binary filter. What this amplitude would be is, however, difficult to calculate.

To overcome this problem, the theoretical reconstructed image amplitude of an on-axis point will first be calculated. This will be done in the appropriate plane using the appropriate reconstruction method. This amplitude, $V^{p,sd}_{00}$, will be used to calibrate the method. Using this value, $A^p_{ij}$ can be replaced with a correctly scaled version for this



(a) points separated in $x$, $y$ - unnormalised convolution

(b) points separated in $x$, $y$ - normalised convolution



(c) points separated in $z$ - unnormalised convolution

(d) points separated in $z$ - normalised convolution

Figure 6.11: Magnitudes of theoretical reconstructed points in different positions for both unnormalised and normalised convolution

method, $A'^{p}_{ij}$ as follows.

$$A'^{p}_{ij} = \frac{V^{p,sd}_{00}}{V^{p}_{00}} A^{p}_{ij} = \frac{8(p+q)^2 V^{p,sd}_{00}}{R^2} A^{p}_{ij} \qquad (6.20)$$

## 6.5 Comparison of Reconstructed Points using Error Metrics

The values of the difference metrics MSE, NMSE and FEM, which were defined in Chapter 5, are be given in Table 6.2 for points with $\frac{S^{p}_{ij,\tau}}{\Delta^{2}_{p}} = 1$ reconstructed using techniques in this Chapter. The parameters for the theoretical hologram are those given in Table 6.1. The ideal reconstructed image would be a single point with amplitude $= 1$. The sum of squares of expected image values is therefore also $= 1$. This means that

NMSE=360×360×MSE in this case. As only a single point has been reconstructed, the signal depends upon this single value whereas any non-zero value in another pixel will contribute to the noise. For this reason, the errors are high. The signal to noise ratio (SNR) is negative as the errors are so high and, therefore, will not be considered here. Normalisation increases the error for two reasons. First, it increases all the values in the reconstructed image and second, it increases the values towards the edges more than those in the centre. Giving the reconstructed point the correct amplitude improves the signal but it also increases the unwanted error values in the other pixels. As the number of pixels containing the signal is much less than those only containing only errors, this increases the MSE and NMSE values. FEM should be constant when the image is multiplied by a constant value but the normalisation is not constant. As the values in the unnormalised images are so low, they are practically zero in comparison to the expected value which is equal to one. This is why the NMSE values are all equal to one and the MSE values are all $= \frac{1}{360\times360}$ for the unnormalised images. The central point signal is multiplied by a smaller number than the edges and so FEM also decreases when the image is normalised. Despite all of this, convolution with a truncated Fresnel function appears to be the best method for point sources and, if the intensity is taken rather than the real part, this provides an obvious improvement. Later results in Chapter 9, however, show that the real part gives better results than the intensity for other source objects. Woodgate [92] also found this

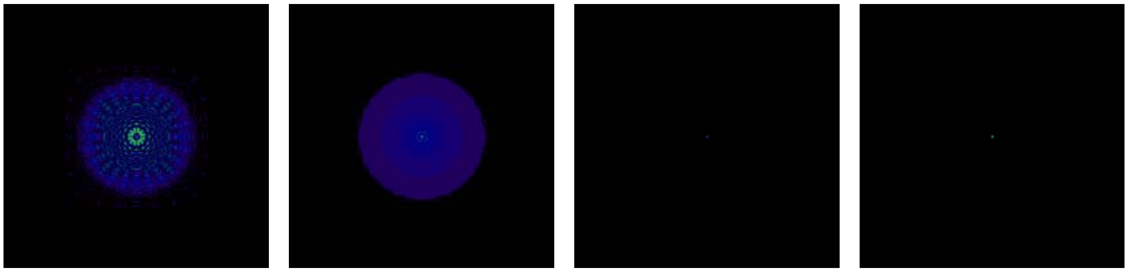
| | MSE | NMSE | FEM |
|---|---|---|---|
| simulated diffraction (real part) | 7.7160e-06 | 1.0000 | 0.92297 |
| simulated diffraction (real part) - normalised | 1.3788e-04 | 17.870 | 0.94701 |
| simulated diffraction (intensity) | 7.7160e-06 | 1.0000 | 0.67548 |
| simulated diffraction (intensity) - normalised | 1.6136e-05 | 2.0912 | 0.67650 |
| matched filter | 7.7160e-06 | 1.0000 | 0.99985 |
| matched filter - normalised | 0.075924 | 9839.8 | 0.99990 |
| balanced convolution | 7.7160e-06 | 1.0000 | 0.99914 |
| balanced convolution - normalised | 0.011298 | 1464.2 | 0.99932 |
| modified mismatch | 7.7160e-06 | 1.0000 | 0.96552 |
| modified mismatch - normalised | 0.00035885 | 46.506 | 0.97914 |
| truncated Fresnel (real part) | 7.7160e-06 | 1.0000 | 0.91805 |
| truncated Fresnel (real part) - normalised | 0.00012676 | 16.428 | 0.94262 |
| truncated Fresnel (intensity) | 7.7160e-06 | 1.0000 | 0.65636 |
| truncated Fresnel (intensity) - normalised | 1.4776e-05 | 1.9150 | 0.65695 |

Table 6.2: Error metric values for the different convolution reconstruction techniques

to be the case.

## 6.6 Comparison of Modulation Transfer Functions

The modulation transfer functions (MTF) for the images of point sources from Section 6.5 are shown in Figure 6.12. The MTFs are only shown for the unnormalised images as there is little visible difference between the normalised and unnormalised cases.

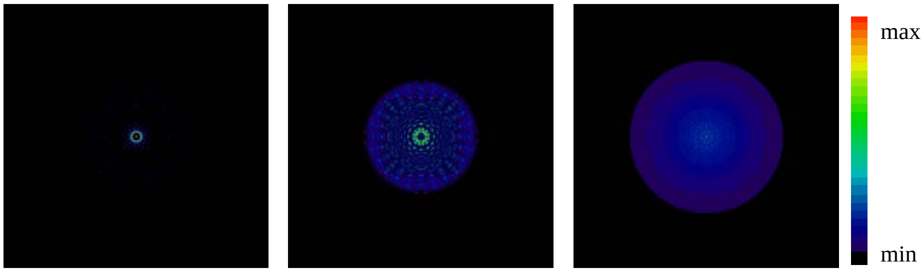


(a) simulated diffraction (real part) 8.7426e-06≤pixel value≤9.2842

(b) simulated diffraction (intensity) 5.1016e-09≤pixel value≤1.0000

(c) matched filter 1.4973e-10≤pixel value≤1.0000

(d) balanced convolution 7.0275e-09≤pixel value≤1.0000



(e) modified mismatch 7.9172e-08≤pixel value≤1.0979

(f) truncated Fresnel (real part) 2.1824e-07≤pixel value≤7.6554

(g) truncated Fresnel (intensity) 5.5643e-04≤pixel value≤1.0000

Figure 6.12: Modulation transfer functions

The MTF for an ideal image of a point would be uniform and equal to one for all frequencies. Finding the error metric for the difference between the MTFs and this uniform array is another measure of the quality of each image. The greater the error, the worse the image is. As the MTF depends only on the frequencies in the reconstructed image and not on its amplitude, the difference between the normalised and unnormalised cases is small. The unnormalised images have slightly lower error metric values. As in Section 6.5, convolution with the truncated Fresnel function appears to be the most effective method overall. Modified mismatch is the best matched filter method.

| | NMSE | FEM |
|---|---|---|
| simulated diffraction (real part) | 0.86086 | 0.84955 |
| simulated diffraction (real part) - normalised | 0.86800 | 0.84539 |
| simulated diffraction (intensity) | 0.95640 | 0.67520 |
| simulated diffraction (intensity) - normalised | 0.96940 | 0.67772 |
| matched filter | 0.99993 | 0.99982 |
| matched filter - normalised | 0.99994 | 0.99986 |
| balanced convolution | 0.99972 | 0.99907 |
| balanced convolution - normalised | 0.99966 | 0.99896 |
| modified mismatch | 0.99208 | 0.96133 |
| modified mismatch - normalised | 0.99238 | 0.96508 |
| truncated Fresnel (real part) | 0.89939 | 0.89448 |
| truncated Fresnel (real part) - normalised | 1.0002 | 0.89278 |
| truncated Fresnel (intensity) | 0.93566 | 0.65636 |
| truncated Fresnel (intensity) - normalised | 0.96026 | 0.65830 |

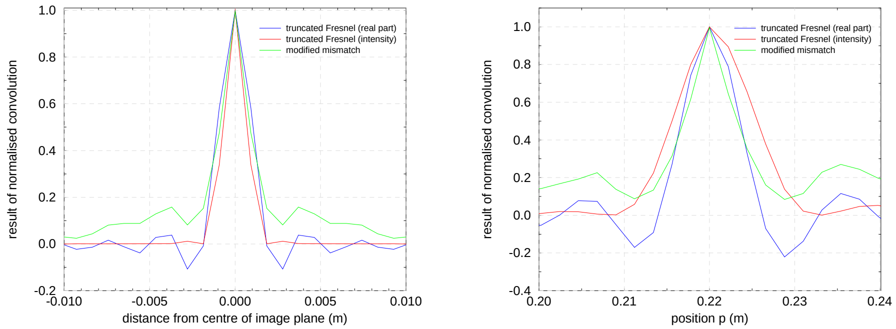
Table 6.3: Comparison of modulation transfer function error metric values for the different convolution reconstruction techniques

## 6.7 Comparison of Point Spread Functions

In Chapter 5, it was suggested that a point spread function could be measured by fitting a one-dimensional Gaussian function to the reconstructed point in each of the three directions separately. As usual, the theoretical hologram used had the parameters given in Table 6.1 The function to be fitted in direction $x'$, $p(x')$, is given by

$$p(x') = A \exp\left\{\frac{-|x'-\mu|^2}{2\sigma^2}\right\} \tag{6.21}$$

As modified mismatch filtering and convolution with the truncated Fresnel function are the best binary and diffraction-like methods respectively, it was decided to just analyse these two techniques. The point spread function (PSF) should be the same for both $x$ and $y$ directions so it was decided to just fit the function along the diagonal of the reconstructed image plane. $\sigma$ for the point spread function in the $x$ and $y$ directions will be the same as $\sigma$ for the diagonal if it is assumed that the PSF is circularly symmetric. The reconstructed points are normalised and examples can be seen in Figure 6.13. The Gaussian was fitted to the peaks using the nonlinear regression routine from the statistical package SPSS. The goodness of fit is indicated by the R-squared value. R-squared is the ratio of the variation predicted by the fitted equation to the total variation. An ideal fit would, therefore, have an R-squared value of one and, if the equation did not predict the variation at all, the R-squared value would be zero. The values of $A$, $\mu$, $\sigma$ and R-squared



(a) PSF along diagonal ($(x, y)$ plane)

(b) PSF in z direction

Figure 6.13: Comparison of modified mismatch reconstruction with reconstruction by convolution with a truncated Fresnel function

can be seen in Table 6.4. Modified mismatch and the real part of the truncated Fresnel reconstruction have similar values of $\sigma$, with $\sigma \approx 0.8$mm in the $x$ and $y$ directions and $\sigma \approx 3$mm in the $z$ direction. Taking the intensity rather than the real part of the truncated Fresnel reconstruction improves the resolution in the $x$ and $y$ directions but makes it worse in the $z$ direction. The values of the resolution will, obviously, be different for different hologram parameters. For example, as the field of view of the detector decreases for distances closer to the zone plate, the resolution increases. This technique will be tried for different sources in Chapters 9 and 10. From these data, the indication is that the $z$ resolution is worse than the $x$ and $y$ resolution.

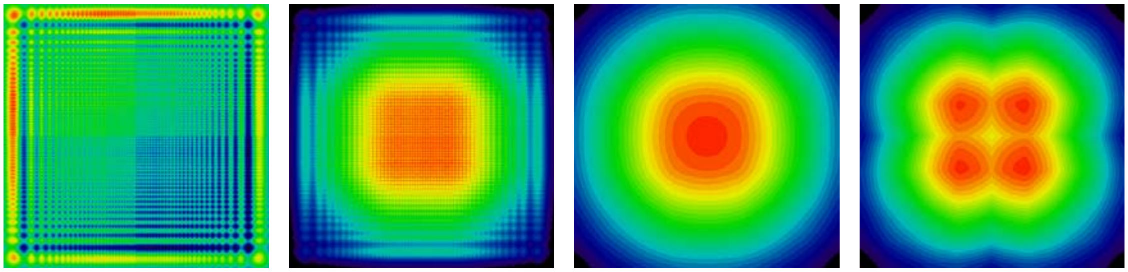
| | $A$ | $\mu$ | $\sigma$ | R-squared |
|---|---|---|---|---|
| $(x, y)$ diagonal | | | | |
| truncated Fresnel (real part) | 1.036 | 0.000 | 8.088e-04 | 0.9747 |
| truncated Fresnel (intensity) | 1.001 | 0.000 | 6.289e-04 | 0.9995 |
| modified mismatch | 9.539e-01 | 0.000 | 8.482e-04 | 0.9592 |
| $z$ direction | | | | |
| truncated Fresnel (real part) | 1.043 | 2.201e-01 | 2.661e-03 | 0.9646 |
| truncated Fresnel (intensity) | 9.888e-01 | 2.206e-01 | 4.248e-03 | 0.9969 |
| modified mismatch | 8.905e-01 | 2.201e-01 | 3.186e-03 | 0.9570 |

Table 6.4: Point spread functions for the modified mismatch and truncated Fresnel techniques

## 6.8 Comparison of Images of Uniform Sources

In Section 6.7 images of point sources were compared. In Chapter 5 the variance of an image of an uniform source was given as a measure of the quality of an imaging system. A hologram of a two dimensional uniform source with unit strength per unit area and the parameters from Table 6.1 was generated. The image from this hologram was then reconstructed in the correct plane using all the different convolution reconstruction techniques. The unnormalised images can be seen in Figure 6.14 and the normalised images in Figure 6.15.

As the Fienup Error Metric (FEM) is invariant to a multiplication by a constant factor, the FEM value for images of uniform source is directly related to the variance. The FEM values are given in Table 6.5. The NMSE values have been included for completeness. The NMSE values for the unnormalised images are, again, mostly equal to one. This

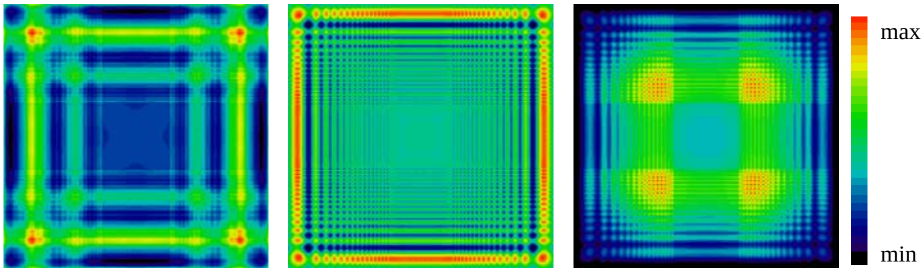


(a) simulated diffraction (real part) -2.4418e-07≤pixel value≤3.1041e-07

(b) simulated diffraction (intensity) 1.8397e-14≤pixel value≤4.8748e-12

(c) matched filter 1.4101e-05≤pixel value≤8.9434e-05

(d) balanced convolution -5.0411e-06≤pixel value≤3.1316e-05



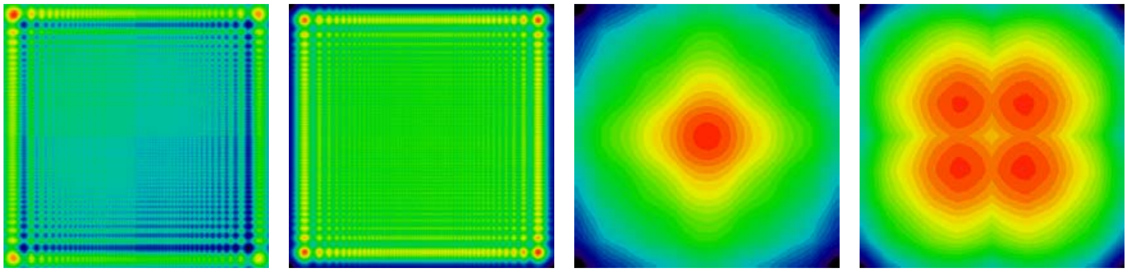
(e) modified mismatch -1.2397e-06≤pixel value≤4.0877e-07

(f) truncated Fresnel (real part) -1.7583e-07≤pixel value≤2.7305e-07

(g) truncated Fresnel (intensity) 1.0591e-12≤pixel value≤3.5645e-16

Figure 6.14: Unnormalised images of uniform sources

is because the amplitudes of the unnormalised images are much lower than the expected amplitudes. The FEM value for matched filtering is very low compared to the other values but this is because the image does not have a negative minimum pixel value whereas all the other images do. Looking at the FEM values for the other methods, the matched filter methods give a better reconstruction than either the real part of the simulated diffraction or the truncated Fresnel function reconstructions.

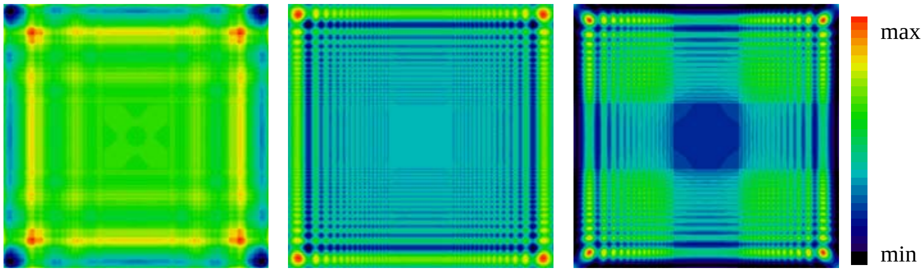


(a) simulated diffraction (real part) - $367.51 \leq$ pixel value $\leq 537.96$

(b) simulated diffraction (intensity) $9.9791e+04 \leq$ pixel value $\leq 2.4923e+06$

(c) matched filter $1.6266e+04 \leq$ pixel value $\leq 2.5872e+04$

(d) balanced convolution - $5820.0 \leq$ pixel value $\leq 9094.3$



(e) modified mismatch - $1075.0 \leq$ pixel value $\leq 263.63$

(f) truncated Fresnel (real part) $-264.64 \leq$ pixel value $\leq 498.70$

(g) truncated Fresnel (intensity) $1231.9 \leq$ pixel value $\leq 6.8174e+05$

Figure 6.15: Normalised images of uniform sources

| | NMSE | FEM |
|---|---|---|
| simulated diffraction (real part) | 1.00000 | 0.96067 |
| simulated diffraction (real part) - normalised | 1.1414e+04 | 0.94017 |
| simulated diffraction (intensity) | 1.00000 | 0.24160 |
| simulated diffraction (intensity) - normalised | 2.1738e+12 | 0.049007 |
| matched filter | 0.99989 | 0.11246 |
| matched filter - normalised | 4.7221e+08 | 0.0078773 |
| balanced convolution | 0.99997 | 0.33785 |
| balanced convolution - normalised | 3.2357e+07 | 0.26374 |
| modified mismatch | 1.0000 | 0.19566 |
| modified mismatch - normalised | 1.0329e+05 | 0.27680 |
| truncated Fresnel (real part) | 1.00000 | 0.81775 |
| truncated Fresnel (real part) - normalised | 1.1856e+04 | 0.82739 |
| truncated Fresnel (intensity) | 1.00000 | 0.25435 |
| truncated Fresnel (intensity) - normalised | 7.8546e+10 | 0.16362 |

Table 6.5: Comparison of error metric values for images of uniform sources for the different convolution reconstruction techniques

## 6.9 Summary

Methods to perform simulated diffraction using convolution and reconstruction using linear filters have been outlined. Three different filters based on the binary zone plate pattern were given and tested on a computer generated hologram of a point source. Modified mismatch was found to be the best method for this hologram but the matched filter method gave better results for a hologram of a two dimensional uniform source.

A hybrid method, combining the maximum radius of the zone plate pattern with simulated diffraction was introduced. This method was found to give the best results for the hologram of a point source.

A method for normalising the reconstructed image was also derived, enabling points to be reconstructed with the correct amplitude.

Although convolution with the actual zone plate pattern could be expected to give better images than convolution with a different, ideal zone plate, simulated diffraction produced better results for a point in a single plane than convolution with the binary plate pattern.

# Chapter 7

# IMAGE RECONSTRUCTION BY DECONVOLUTION

In this Chapter the method of reconstruction by deconvolution using Fourier transforms will be discussed. Reconstruction by deconvolution was used by Woodgate [92] but as he did not use a Wiener filter, his reconstructed images were of poor quality. Deconvolution using a matrix Wiener filter was used by Strange [83] on holograms generated using MCNP [15]. The reconstructed images were good but the technique could only be used on holograms with a maximum of 64×64 pixels due to the limitations of the computers available. The use of a parametric Wiener filter will be explained here and techniques for optimising the output will be given. As in the Chapter on reconstruction by convolution (Chapter 6), an attempt will be made to normalise the reconstructed image.

## 7.1 Expression of the hologram formation process using Fourier transforms and using matrix notation

The equation for hologram formation was written as the following summation in Chapter 6.

$$U_{i'j'} = \sum_{k=0}^{N_p-1} C^k \sum_{i,j=-\frac{N}{2}}^{\frac{N}{2}-1} S_{ij}^k T_{(i'-i)(j'-j)}^k \quad \text{where} \quad U_{ij} = U(i\Delta_p, j\Delta_p) \quad C^k = \frac{D^2}{4\pi}\frac{1}{(p_k+q)^2}$$

$$S_{ij}^k = S(i\frac{\Delta_p}{L(p_k,q)}, j\frac{\Delta_p}{L(p_k,q)}, p_k) \quad \text{and} \quad T_{ij}^k = T(i\frac{\Delta_p}{M(p_k,q)}, j\frac{\Delta_p}{M(p_k,q)}) \quad (7.1)$$

$S(\boldsymbol{\epsilon}, p)$ represented the source distribution and $T(\boldsymbol{x})$ the intensity transmittance of the zone plate pattern. $U_{i'j'}$ are pixel values and $-\frac{N}{2} \leq i', j' \leq \frac{N}{2} - 1$. The contribution of objects in a single plane at a distance $p$ from the zone plate, $U^p$, to the hologram is given by the convolution of $S^p$ with $T^p$ multiplied by a constant, $C^p$.

$$U^p = C^p\,(S^p \otimes T^p) = H^p \otimes S^p \quad \text{where} \quad H^p = C^p T^p \quad (7.2)$$

From the convolution theorem [69] states that the Fourier transforms of $U^p$, $H^p$ and $T^p$ are related in the following way

$$U^{p^F} = H^{p^F} S^{p^F} \tag{7.3}$$

where $A^F$ is the Fourier transform of a variable $A$.

If $U^p$ and $S^p$ were written as column vectors instead of two-dimensional arrays, the hologram formation process could be written in matrix form. This will be explained using the following simple example. A 2×2 source distribution and a 2×2 zone plate transmittance function will be used.

$$C^p = \frac{1}{4} \quad T^p = \boxed{\begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix}} \quad S^p = \boxed{\begin{matrix} 5 & 6 \\ 7 & 8 \end{matrix}} \quad \text{so} \quad H^p = \boxed{\begin{matrix} 0.25 & 0.5 \\ 0.75 & 1 \end{matrix}} \quad \text{and} \quad U^p = \boxed{\begin{matrix} 15 & 10 \\ 13 & 8 \end{matrix}} \tag{7.4}$$

If $U^p$ and $S^p$ are written as column vectors, the convolution by $T^p$ and multiplication by $C^p$ can be combined into a 'hologram formation matrix' $[H^p]$.

$$\boldsymbol{U^p} = [H^p]\boldsymbol{S^p} \qquad \begin{pmatrix} 15 \\ 13 \\ 10 \\ 8 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.75 & 0.25 \\ 0 & 1 & 0 & 0.75 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ 7 \\ 6 \\ 8 \end{pmatrix} \tag{7.5}$$

## 7.2 Deconvolution using Fourier and standard (matrix) Wiener filters

Deconvolution involves the inversion of the convolution equation, it tries to undo the action of convolution. If Fourier transforms were used, equation 7.3 could be inverted and $S^p$ could be calculated in the following way.

$$S^p = \quad \text{inverse Fourier transform of} \quad S^{p^F} = \frac{U^{p^F}}{H^{p^F}} \tag{7.6}$$

Similarly, if matrix notation were used, $[H^p]$ could be inverted to calculate $\boldsymbol{S^p}$.

$$\boldsymbol{S^p} = [H^p]^{-1}\boldsymbol{U^p} \tag{7.7}$$

This will only work, however, for an ideal, noise free imaging system. As both reconstruction techniques involve division of one set of data by another, small errors in the data will lead to large errors in the output. This is particulary the case when the problem is ill-conditioned. Ill-conditioning means that small perturbations in $H^p$ or $U^p$ lead to large differences in $S^p$. This was been shown to be the case for zone plate holography when Woodgate [92] used deconvolution without a Wiener filter and obtained poor results. A Wiener filter is used to optimise the output for a noisy imaging system.

Andrews and Hunt [1] state that there are two different types of Wiener filter, the Fourier Wiener Filter, which uses Fourier transforms, and the standard Wiener filter which uses the matrix representation. Each method assumes that the signal is not perfect but has noise added to it and that the signal (hologram) and noise are uncorrelated. The sources of error in a zone plate hologram were discussed in Chapter 4. In Chapter 5 these errors were classified as source noise $N_s$ and recording noise $N_a$. The signal recorded from object plane $p$ is not $U^p$ but a noisy version $\tilde{U^p}$ given by

$$\tilde{U^p} = C^p\left([S^p + N_s^p] \otimes T^p\right) + N_a^p = H^p \otimes S^p + H^p \otimes N_s^p + N_a^p \tag{7.8}$$

As the Wiener filter assumes that the signal and noise are uncorrelated, we have to replace this by a single error term $E^p$.

$$\tilde{U^p} = H^p \otimes S^p + E^p \tag{7.9}$$

The reconstructed image, $\tilde{S^p}$, should be as close to $S^p$ as possible. Fourier transforms and a filter can be used to achieve this. The filter, $\Phi$, is designed to minimise the sum of squared differences between $\tilde{S^p}$ and $S^p$ in the following way.

$$\tilde{S^p}^F = \frac{\tilde{U^p}^F \Phi^p}{H^{p^F}} \tag{7.10}$$

The Fourier Wiener filter is the filter which optimises $\tilde{S^p}$. It is derived in appendix D.

$$\text{Fourier Wiener filter} \quad \Phi^p = \frac{\left|U^{p^F}\right|^2}{\left|U^{p^F}\right|^2 + \left|E^{p^F}\right|^2} \tag{7.11}$$

Inserting this into equation 7.10 gives

$$\tilde{S^p}^F = \frac{\tilde{U^p}^F}{H^{p^F}} \frac{\left|U^{p^F}\right|^2}{\left|U^{p^F}\right|^2 + \left|E^{p^F}\right|^2} = \frac{\tilde{U^p}^F H^{p^{F*}}}{\left|H^{p^F}\right|^2 + \left|\frac{E^{p^F}}{S^{p^F}}\right|^2} \tag{7.12}$$

$H^{p^{F*}}$ is the complex conjugate of the Fourier transform of $H^p$. The parametric Fourier Wiener filter includes a parameter $\gamma$ which can be varied to increase or decrease the effect of filtering.

$$\text{parametric Fourier Wiener filter} \quad \Phi^p = \frac{\left|U^{p^F}\right|^2}{\left|U^{p^F}\right|^2 + \gamma \left|E^{p^F}\right|^2} \tag{7.13}$$

The standard (matrix) Wiener filter, $[\Phi^p]$, is the filter which minimises the sum of squared differences between $\tilde{\boldsymbol{S}}^{\boldsymbol{p}}$ and $\boldsymbol{S}^{\boldsymbol{p}}$ when it is calculated using the inverse of matrix $[H^p]$.

$$\tilde{\boldsymbol{S}}^{\boldsymbol{p}} = [\Phi^p][H^p]^{-1}\tilde{\boldsymbol{U}}^{\boldsymbol{p}} \tag{7.14}$$

Using a similar derivation to that used for the Fourier Wiener filter, $[\Phi^p]$ is found to be

$$\text{Wiener filter} \quad [\Phi^p] = \left([H^p]^{*T}[H^p] + [\phi_{S^p}]^{-1}[\phi_{E^p}]\right)^{-1}[H^p]^{*T}[H^p] \tag{7.15}$$

$[\phi_{S^p}]$ and $[\phi_{E^p}]$ are the signal and noise covariance matrices. If each pixel value in the signal is assumed to be uncorrelated with other values in the signal and the same can be said for the noise, the off-diagonal elements in these matrices can be ignored. The diagonal elements are the expected variances of each pixel value.

There is also a parametric Wiener filter which is given by

$$\text{parametric Wiener filter} \quad [\Phi^p] = \left([H^p]^{*T}[H^p] + \gamma[\phi_{S^p}]^{-1}[\phi_{E^p}]\right)^{-1}[H^p]^{*T}[H^p] \tag{7.16}$$

where, again, $\gamma$ is a variable which can be changed to increase or decrease the effect of the filter.

The matrix Wiener filter has the advantage that more than one plane of data could be included in the matrix equation. If a Fourier Wiener filter is used, the image planes have to be reconstructed one at a time and information from sources in other planes is just treated as a contribution to the noise term. The following simple example shows how two planes could be reconstructed simultaneously using a matrix Wiener filter. One of the sets of data is the one which was used in equation 7.4.

$$\begin{gathered}
S^{p_1} = \boxed{\begin{matrix} 5 & 6 \\ 7 & 8 \end{matrix}} \quad H^{p_1} = \boxed{\begin{matrix} 0.25 & 0.5 \\ 0.75 & 1 \end{matrix}} \quad S^{p_2} = \boxed{\begin{matrix} 8 & 6 \\ 3 & 4 \end{matrix}} \quad H^{p_2} = \boxed{\begin{matrix} 1.5 & 2 \\ 1 & 0.5 \end{matrix}} \\
U^{p_1} = \boxed{\begin{matrix} 15 & 10 \\ 13 & 8 \end{matrix}} \quad U^{p_2} = \boxed{\begin{matrix} 22 & 11 \\ 5.5 & 2 \end{matrix}} \quad U^{p_1+p_2} = \boxed{\begin{matrix} 37 & 21 \\ 18.5 & 10 \end{matrix}}
\end{gathered} \tag{7.17}$$

This can be written in matrix form.

$$\begin{pmatrix} 1 & 0.5 & 0.75 & 0.25 & 0.5 & 2 & 1 & 1.5 \\ 0 & 1 & 0 & 0.75 & 0 & 0.5 & 0 & 1 \\ 0 & 0 & 1 & 0.5 & 0 & 0 & 0.5 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.5 \end{pmatrix} \begin{pmatrix} 5 \\ 7 \\ 6 \\ 8 \\ 8 \\ 3 \\ 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 37 \\ 18.5 \\ 21 \\ 10 \end{pmatrix} \tag{7.18}$$

The two planes could be reconstructed by simply inverting the matrix. In the example above, the matrix happens to be singular but this is due to the numbers which were used. The matrix Wiener filter provides a way to use a singular matrix for reconstruction. $[H]$ may be singular, but $[H]^{*T}[H] + \gamma[\phi_S]^{-1}[\phi_E]$ can be inverted if $\gamma$ is large enough.

## 7.3 Application of Wiener filters to zone plate holography

If the hologram parameters given in Table 6.1 are used, the hologram is a 360×360 pixel array. If matrix representation were used, this would require inversion of a $360^2 \times 360^2$ matrix or larger. As the Fourier Wiener filter is far more practical, it is the one which will be investigated in this Chapter.

The Fourier Wiener filter relies upon knowing the magnitude of the ratio of the Fourier transform of the error term to the Fourier transform of the signal from the source, $\left|\frac{E^{pF}}{U^{pF}}\right|$, for all frequency values. Two possible ways of estimating this will be presented.

### 7.3.1 Using the power spectral density

Press et. al. [68] suggest that the power spectral density, $\left|\tilde{U^p}^F\right|^2$ should be plotted to give an indication of the relative contribution of errors to the hologram. This can be done by assuming that the power spectral density of the hologram is approximately equal to the power spectral density of the signal from the source plus the power spectral density of the noise.

$$\left|\tilde{U^p}^F\right|^2 \approx \left|U^{pF}\right|^2 + \left|E^{pF}\right|^2 \tag{7.19}$$

The power spectral density of the noise is also called the Wiener spectrum. The Wiener spectrum was mentioned as a tool for analysing the errors in an image in Chapter5. Press et. al. illustrate how $\left|\frac{E^{pF}}{U^{pF}}\right|$ is found using the graph which is shown in Figure 7.1. The contribution of the signal to the power spectrum should first be estimated. A mathematical model should then be fitted to this contribution. Subtracting this from the power spectrum will give an estimation of the noise term.

Figure 7.2 shows a theoretical ideal hologram with the parameters given in Table 6.1 and the same hologram with 50% additive random noise taken from an uniform distri-

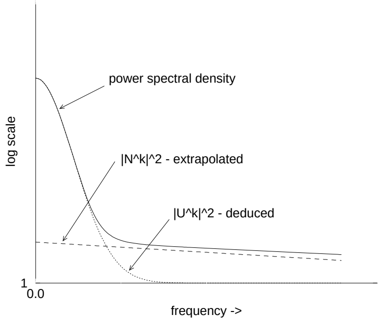
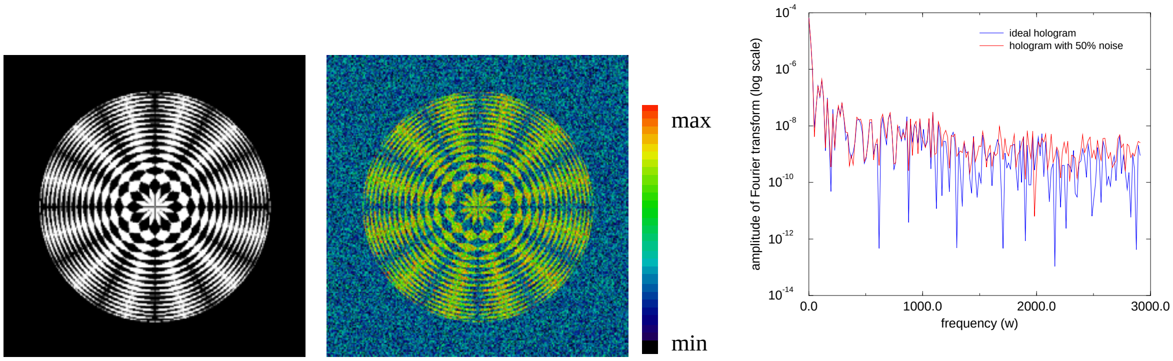


Figure 7.1: Fitting a 'signal model' and a 'noise model' to the power spectrum - taken from Numerical Recipes by Press et. al.

bution. This is artificial a hologram would not have negative pixel values but it is just designed to illustrate the effect of adding noise. The power spectra, which were taken as lines at $y = 0$ out from the centre of the two-dimensional Fourier transforms of the holograms, are shown. The effect of adding noise to the hologram can be clearly seen but it would be very difficult to find a function which fitted the spectrum of the ideal hologram. It would therefore be impractical to apply this method for finding $\left|\frac{E^{pF}}{U^{pF}}\right|$ to zone plate holography.



(a) ideal hologram 0≤pixel value≤1.1754e-13

(b) hologram with 50% noise -5.8770e-14≤pixel value≤1.7631e-13

(c) power spectra

Figure 7.2: A theoretical 'ideal' hologram and the same hologram with 50% noise added, together with their power spectra

## 7.3.2 Finding a constant which gives the minimum error

As it is not obvious how to separate the signal from the noise, an approximate Wiener filter must be found. Two possibilities are given below.

### Assuming $\left|E^{pF}\right|^2$ is constant

One kind of noise in recorded signals is *band-limited white Gaussian noise* [39]. It has equal power per hertz in some band of frequencies and a Gaussian distribution of amplitudes if a large number of measurements of its amplitude are made. In the discussion of sources of "noise" in the hologram in Chapter 4, some sources of error were listed which definitely could not be described as white noise. However, assuming $\left|E^{pF}\right|^2$ is constant for all frequencies is a useful approximation. If it is assumed that equation 7.19 can be used, equation 7.12 for the reconstructed signal $\tilde{S^p}^F$ becomes

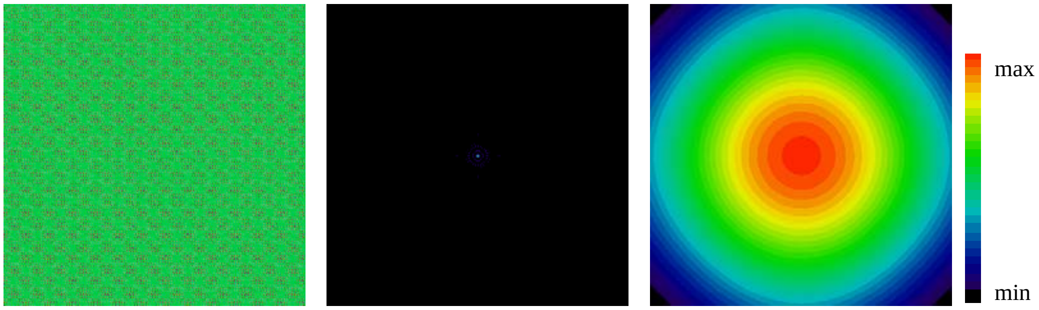
$$\tilde{S^p}^F = \frac{\tilde{U^p}^F}{H^{pF}} \left( \frac{\left|\tilde{U^p}^F\right|^2 - K1}{\left|\tilde{U^p}^F\right|^2} \right) \tag{7.20}$$

where $K1 = \left|E^{pF}\right|^2$ is a constant. $K1$ can be varied between zero and the maximum value of $\left|\tilde{U^p}^F\right|^2$. If $\left|\tilde{U^p}^F\right|^2 - K1$ becomes negative for any frequency, $S^{pF}$ should be set to zero for that frequency. The effect of this filter is illustrated in Figure 7.3. The hologram used had the parameters given in Table 6.1 and 10% additive random noise taken from an uniform distribution. The variation of the NMSE and FEM values with $K1$ are also shown in Figure 7.3. These error metric values were defined in Chapter 5 and are a measure of the difference between the reconstructed image and the source object. This Wiener filter will be referred to as a $K1$ Wiener filter.

### Assuming $\left|\frac{E^{pF}}{S^{pF}}\right|^2$ is constant

If $\left|\frac{E^{pF}}{S^{pF}}\right|^2$ were constant, this would mean that the errors would be highly correlated with the source term. As a large proportion of the errors arise from statistical fluctuations due to radioactive decay, this is a sensible approximation. Substituting $\left|\frac{E^{pF}}{S^{pF}}\right|^2 = K2$ into equation 7.12 gives

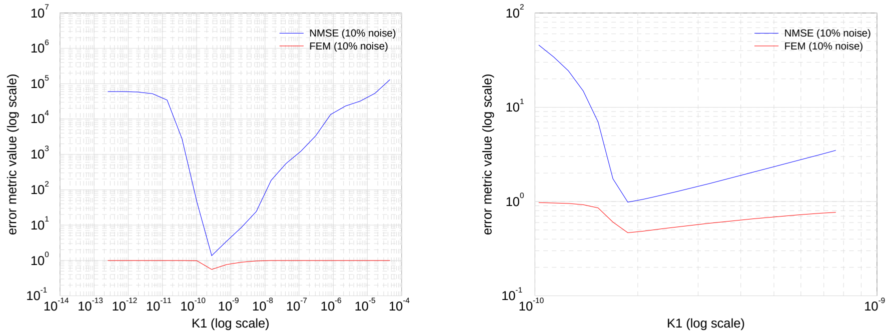
$$\tilde{S^p}^F = \frac{\tilde{U^p}^F H^{pF*}}{\left|H^{pF}\right|^2 + K2} \tag{7.21}$$



(a) reconstruction of hologram with 10% noise, $K1$ is too small -2.1374≤pixel value≤2.0476

(b) reconstruction of hologram with 10% noise, optimal $K1$ -0.039582≤pixel value≤1.0389

(c) reconstruction of hologram with 10% noise, $K1$ is too large -0.046716≤pixel value≤1.0001



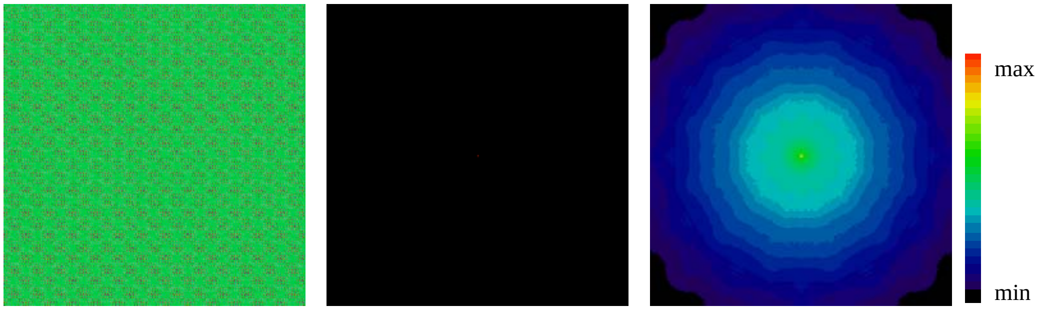
(d) variation of NMSE and FEM metric values with $K1$ for holograms with 10% noise

(e) variation of NMSE and FEM metric values with $K1$ for holograms with 10% noise - close up of minima

Figure 7.3: Effect of varying $K1$ on the reconstructed image from a theoretical noisy hologram

$K2$ is a constant which can be varied between zero and infinity. However, it only makes sense to increase it up to the maximum value of $\left|H^{p^F}\right|^2$. The effect of this filter is illustrated in Figure 7.4. Again, the hologram used had the parameters given in Table 6.1 and 10% additive random noise taken from an uniform distribution. The variation of the NMSE and FEM values with $K2$ is also shown in Figure 7.4.

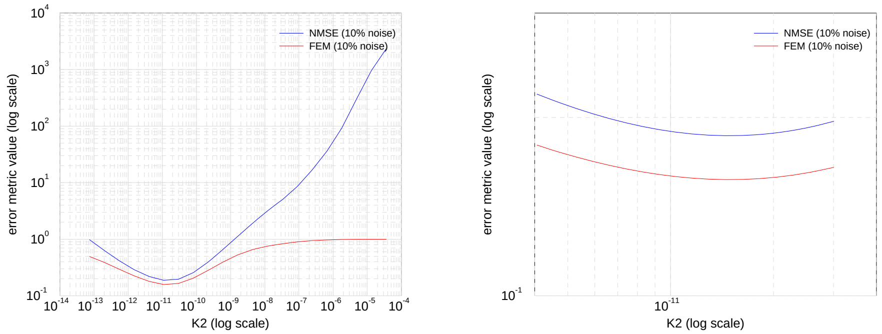
Comparing Figures 7.3 and 7.4, it can be seen that this method gives lower error values than the previous method. This method will be referred to as $K2$ Wiener filtering.



(a) reconstruction of hologram with 10% noise, $K2$ is too small $-2.1366 \leq$ pixel value $\leq 2.0468$

(b) reconstruction of hologram with 10% noise, optimal $K2$ $-0.0067784 \leq$ pixel value $\leq 0.99978$

(c) reconstruction of hologram with 10% noise, $K2$ is too large $0.013306 \leq$ pixel value $\leq 0.99999$



(d) variation of NMSE and FEM metric values with $K2$ for holograms with 10% noise

(e) variation of NMSE and FEM metric values with $K2$ for holograms with 10% noise - close up of minima

Figure 7.4: Effect of varying $K2$ on the reconstructed image from a theoretical noisy holograms

## 7.4 Normalisation

### 7.4.1 To Account for the Effect of $K1/K2$ on the Magnitude of the Reconstructed Point

If the image from an ideal hologram of a point source was reconstructed using deconvolution, a Wiener filter would not be needed as there would be no errors. As $H^p$ is equivalent to the hologram of a central point source with unit magnitude in plane $p$, reconstructing a hologram of such a source would involve dividing the Fourier transform of $H^p$ by itself. This will produce an array full of ones. Depending upon the way that the inverse discrete

Fourier transform is defined in the subroutine used, taking the inverse Fourier transform of an array full of ones will reconstruct the point source with either unit amplitude or with amplitude $= N$. So, in this one case, the amplitude will either be correct or can be corrected by dividing the output pixel values by $N$.

If a Wiener filter is used, the filter value will have a large effect on the amplitude. For example, if an ideal hologram with the parameters given in Table 6.1 is reconstructed without using a Wiener filter, the reconstructed point has an amplitude =1. If a $K2$ Wiener filter is used and with $K2 = 1$, the reconstructed point has an amplitude =2.1896e-09.

A way to normalise the reconstructed images to the correct amplitude is to calculate the ideal hologram from a unit source in the centre of the reconstruction plane and then reconstruct this using the same Wiener filter as was used to reconstruct the image from the hologram. The pixel values in the reconstructed image can then be normalised by dividing them by the amplitude of the point reconstructed from the ideal hologram. The two Wiener filters are then given by the following equations.

$$\begin{aligned}\text{normalised } \tilde{S}^p &= \frac{\mathcal{F}^{-1}\left\{\frac{\tilde{U^p}^F {H^p}^{F*}}{\left|{H^p}^F\right|^2}\left(\frac{\left|\tilde{U^p}^F\right|^2 - K1}{\left|\tilde{U^p}^F\right|^2}\right)\right\}}{\text{maximum of } \mathcal{F}^{-1}\left\{\frac{\left|{H^p}^F\right|^2 - K1}{\left|{H^p}^F\right|^2}\right\}} \\ \text{or normalised } \tilde{S}^p &= \frac{\mathcal{F}^{-1}\left\{\frac{\tilde{U^p}^F {H^p}^{F*}}{\left|{H^p}^F\right|^2 + K2}\right\}}{\text{maximum of } \mathcal{F}^{-1}\left\{\frac{\left|{H^p}^F\right|^2}{\left|{H^p}^F\right|^2 + K2}\right\}}\end{aligned} \tag{7.22}$$

where $\mathcal{F}^{-1}$ denotes an inverse Fourier transform.

### 7.4.2 To account for part of the zone plate hologram falling outside the detector

Equation 6.18 in Chapter 6 gave normalisation constants $A^p_{ij}$ which accounted for the fact that, for points sources in some positions, the zone plate shadow is not wholly recorded by the detector but overlaps the edge. This means that, for some frequencies, the amplitude of the Fourier transform is reduced. This, in turn, decreases the amplitude of the reconstructed point. Although the relationship between reconstructed amplitude and overlap is not as straightforward as that for convolution, it has been found that multiplying the image amplitude by $\frac{A^p_{00}}{A^p_{ij}}$ is an effective way to normalise the output.

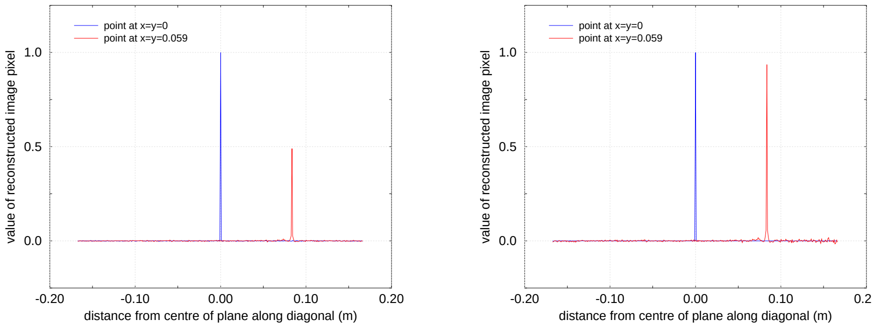
If the image from an ideal, noiseless hologram of a point source is reconstructed using deconvolution, no Wiener filter will be needed to reconstruct the plane containing the source. However, if the image is reconstructed in different planes, the errors due to the difference between the hologram and the theoretical shadow used for the reconstruction will give rise to large errors in the image. For this reason, a theoretical hologram with 1% additive noise taken from a uniform distribution was used to investigate the effect of normalisation. Again, the hologram had the parameters given in Table 6.1. The optimum $K2$ Wiener filter was found for the plane containing the source and then this was used for all the planes.

Multiplying the image amplitude by $\frac{A^p_{00}}{A^p_{ij}}$ does help normalise the image amplitude for different $(x, y)$ positions. However, by comparing Figure 7.5 to Figure 6.11, it can be seen that the normalisation method used in convolution is more effective. The normalised $K2$ Wiener filter appears to have better $z$ resolution than the unnormalised filter but this may be because it had a higher optimal value of $K2$. The values for the normalised case were found to be $K2 =$7.66e-14 for the point at $z =$0.22m and $K2 =$4.65e-14 for the point at $z =$0.28m. For the unnormalised case, these values were $K2 =$4.45e-17 and $K2 =$4.20e-14 respectively. Better results may have been obtained by finding the optimal Wiener filter for each plane individually but this would have been very time consuming.

## 7.5 Effect of a mis-aligned zone plate

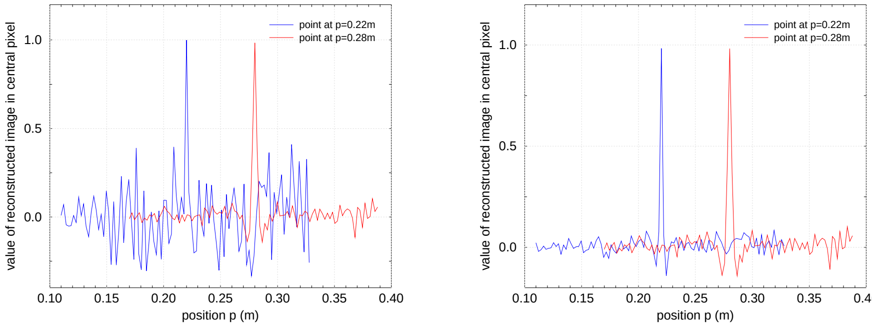
In this Chapter, so far, only methods using the binary zone plate pattern have been discussed. In Chapter 6, techniques using both the binary plate and Fresnel functions were discussed. The techniques which used the binary plates were not found to work as well as the ones which used Fresnel functions. Although the binary plate is the obvious pattern to use for deconvolution, other possibilities should also be investigated.

One disadvantage of using the binary zone plate pattern is that it needs to be aligned correctly. When a zone plate hologram is recorded, it is possible that the zone plate may not be correctly aligned with the detector. The zone plate may be rotated by an angle up to that of half a sector with respect to the expected alignment. This is illustrated



(a) points separated in $x$, $y$ - unnormalised $K2$ Wiener filter

(b) points separated in $x$, $y$ - normalised $K2$ Wiener filter

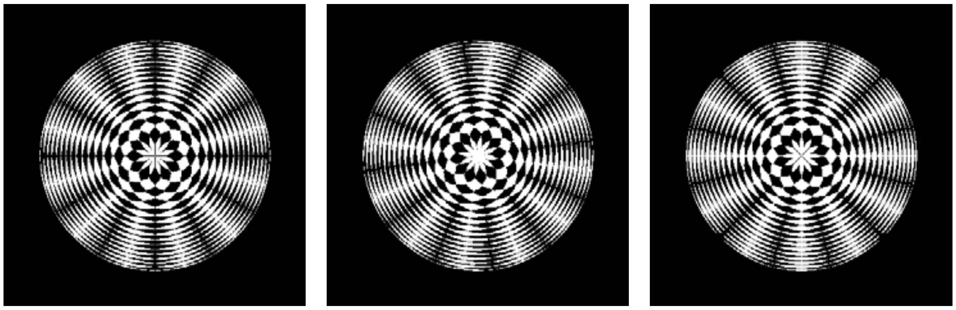


(c) points separated in $z$ - unnormalised $K2$ Wiener filter

(d) points separated in $z$ - normalised $K2$ Wiener filter

Figure 7.5: Magnitudes of theoretical reconstructed points in different positions for both unnormalised and normalised $K2$ Wiener filtering of a hologram containing 1% additive noise taken from a uniform distribution

in Figure 7.6. The hologram with 1% noise used in the previous Section is aligned with Figure 7.6(a). It was reconstructed using a optimal $K2$ Wiener filter and the zone plate shadows in each of Figures 7.6(a), 7.6(b) and 7.6(c). A line of data was then taken across the diagonal of each reconstructed image and this can be seen in Figure 7.7. From this Figure, it is clear that the zone plate being mis-aligned has a large effect on the reconstructed image.



(a) alignment of hologram

(b) zone plate pattern rotated by a quarter of a sector

(c) zone plate pattern rotated by half a sector

Figure 7.6: How the expected zone plate can be mis-aligned with the actual zone plate

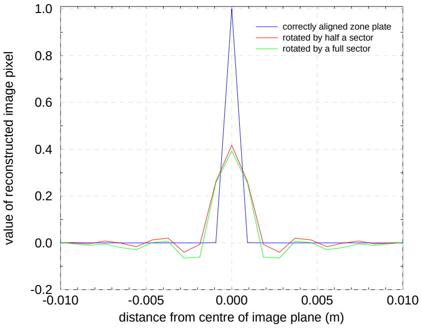
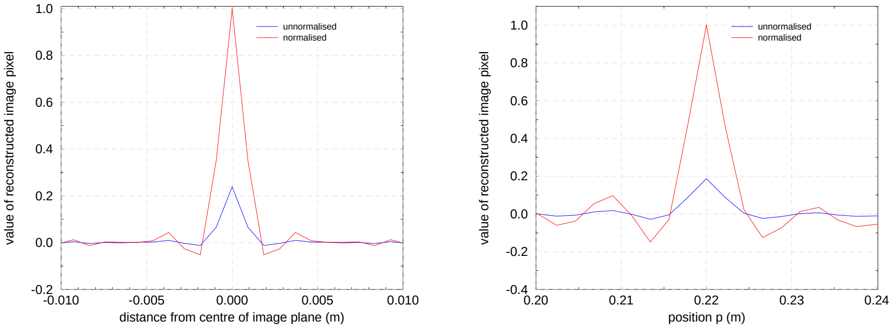


Figure 7.7: Effect of a mis-aligned zone plate on the point spread function

## 7.6 Using a truncated Gabor zone plate with a Wiener filter

Pure Gabor zone plates were discussed in Chapter 3. They are the type of zone plate which would make a zone plate hologram as much like a conventional optical hologram as possible. Binary zone plates are an approximation of this type of plate. Gabor zone plates are radially symmetric so the alignment of the plate is not an issue. For this reason, it is suggested that a Gabor plate could be used for the reconstruction as well as the binary plate pattern. If poor results were obtained when the binary plate was used, this could be used to check that this was not due to mis-alignment of the zone plate. The zone plate should be truncated at the same radius as the binary plate and have all the same parameters. This method will be referred to as Gabor Wiener filtering. As before, the optimal filter could be a $K1$ or a $K2$ filter. Again, a $K2$ Wiener filter is used here as it was found to produce better results.

Point spread functions for this method and a hologram with 1% additive noise and the parameters in Table 6.1 are shown in Figure 7.8.



(a) PSF along diagonal ($(x, y)$ plane)

(b) PSF in $z$ direction

Figure 7.8: Point spread functions for both unnormalised and normalised $K2$ Gabor Wiener filtering of a hologram containing 1% additive noise taken from a uniform distribution

## 7.7 Comparison of reconstructed points using error metrics

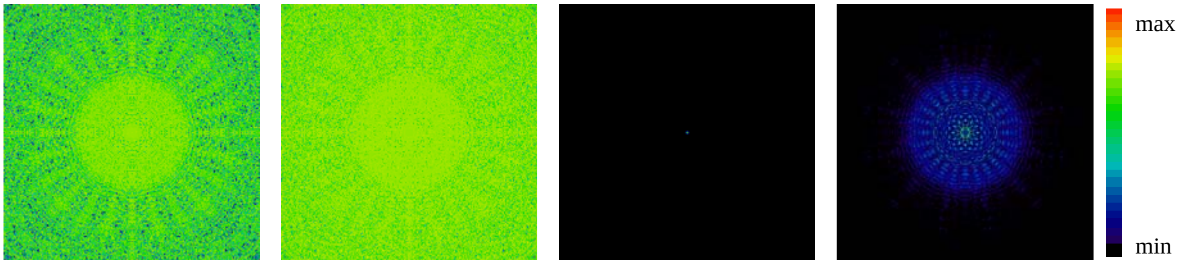
This Section can be compared with Section 6.5 for convolution reconstruction. For a single point source, the MSE value is just a constant multiple of the NMSE value. Therefore, only the NMSE and FEM values are given here. The image from the hologram with 1% noise used in the previous Sections was reconstructed using an optimal $K2$ Wiener filter and an optimal $K2$ Gabor Wiener filter. The NMSE and FEM values for both normalised and unnormalised reconstructions can be seen in Table 7.1. Comparing this to Table 6.2 for the convolution reconstruction, the $K1$ and $K2$ binary Wiener filter methods have far better FEM and NMSE values than any of the convolution methods. The Gabor $K2$ Wiener filter has better FEM values than all the matched filter methods but the Gabor $K1$ Wiener filter has worse FEM values than most of the convolution techniques.

| | NMSE | FEM |
|---|---|---|
| $K1$ Wiener filter | 0.075864 | 0.061159 |
| $K1$ Wiener filter - normalised | 0.14610 | 0.12703 |
| $K2$ Wiener filter | 0.012717 | 0.011437 |
| $K2$ Wiener filter - normalised | 0.029303 | 0.028450 |
| Gabor $K1$ Wiener filter | 0.99990 | 0.99988 |
| Gabor $K1$ Wiener filter - normalised | 9545.9 | 0.99990 |
| Gabor $K2$ Wiener filter | 0.82416 | 0.81856 |
| Gabor $K2$ Wiener filter - normalised | 6.7331 | 0.87070 |

Table 7.1: Error metric values for the different Wiener filter reconstruction techniques

## 7.8 Comparison of Modulation Transfer Functions

This Section can be compare with Section 6.6 for the convolution methods. The modulation transfer functions (MTF) are shown in Figure 7.9 for the different deconvolution reconstruction methods. As in Section 6.6, the NMSE and FEM values for the difference between the MTFs and an array filled with ones have been calculated. Again, the $K2$ binary Wiener filter has the lowest NMSE and FEM values. The values are lower for the unnormalised images than for the normalised images. All the NMSE and FEM values for the binary Wiener filters are lower than those for all the convolution techniques. The Gabor $K2$ Wiener filter has lower values than most of the convolution methods and the $K1$ Wiener filter has higher values than most.



(a) $K1$ Wiener filter $0.0033100 \leq$ pixel value $\leq 1.3276$

(b) $K2$ Wiener filter $0.091849 \leq$ pixel value $\leq 1.3170$

(c) Gabor $K1$ Wiener filter $4.1660\text{e-}10 \leq$ pixel value $\leq 1.0000$

(d) Gabor $K2$ Wiener filter $3.5934\text{e-}05 \leq$ pixel value $\leq 5.2537$

Figure 7.9: Modulation transfer functions

| | NMSE | FEM |
|---|---|---|
| $K1$ Wiener filter | 0.080598 | 0.059786 |
| $K1$ Wiener filter - normalised | 0.12929 | 0.11638 |
| $K2$ Wiener filter | 0.010950 | 0.0093764 |
| $K2$ Wiener filter - normalised | 0.024107 | 0.022967 |
| Gabor $K1$ Wiener filter | 0.99988 | 0.99978 |
| Gabor $K1$ Wiener filter - normalised | 0.99985 | 0.99962 |
| Gabor $K2$ Wiener filter | 0.73924 | 0.70838 |
| Gabor $K2$ Wiener filter - normalised | 0.72471 | 0.70598 |

Table 7.2: Comparison of modulation transfer function error metric values for the different deconvolution reconstruction techniques

## 7.9 Comparison of point spread functions

As in Section 6.7 for the convolution methods, one-dimensional Gaussian functions were fitted to the reconstructed points. The function fitted in direction $x'$, $p(x')$, was given by

$$p(x') = A \exp\left\{\frac{-|x' - \mu|^2}{2\sigma^2}\right\} \tag{7.23}$$

This was fitted along the $(x, y)$ diagonal of the plane containing the image of the point source and also in the $z$ direction. The points which were analysed were reconstructed by using an optimal normalised $K2$ Wiener filter and Gabor Wiener filter. The images have to be reconstructed in a series of planes to find the $z$ resolution and it would have been very time consuming to do this for each deconvolution method. The results can be seen in Table 7.3. Comparing these results to the equivalent ones for the convolution methods given in Table 6.4, both Wiener filter methods appear to have better resolution in both the $(x, y)$ and $z$ directions. This will be investigated later with different holograms.

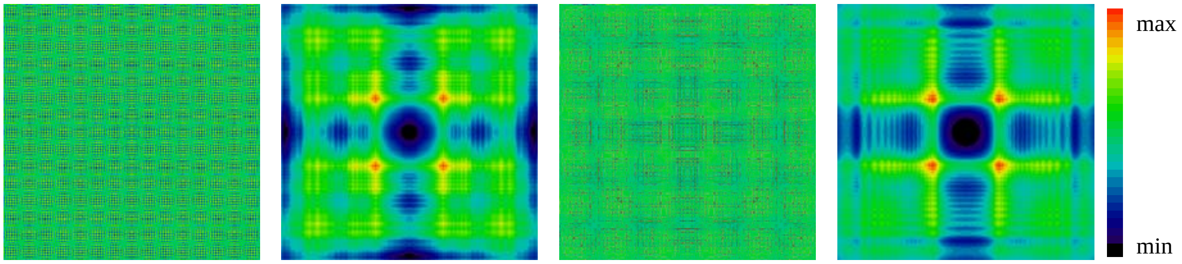
| | $A$ | $\mu$ | $\sigma$ | R-squared |
|---|---|---|---|---|
| $(x, y)$ diagonal | | | | |
| $K2$ Wiener filter | 1.000 | 0.000 | 2.446e-04 | 1.000 |
| Gabor $K2$ Wiener filter | 1.005 | 0.000 | 6.319e-04 | 0.9882 |
| $z$ direction | | | | |
| $K2$ Wiener filter | 9.847e-01 | 2.199e-01 | 9.488e-04 | 0.9670 |
| Gabor $K2$ Wiener filter | 1.010 | 2.200e-01 | 1.732e-03 | 0.9919 |

Table 7.3: Point spread functions for the different Wiener filter techniques

## 7.10 Comparison of Images of Uniform Sources

Images of two dimensional uniform sources were reconstructed using the deconvolution methods. This was done for the convolution methods in Section 6.8. The same hologram was used here as in Section 6.8. Unlike the hologram of the point source, no noise was added. Part of the shadow cast by a two dimensional uniform source would not have been recorded by the hologram. The hologram, therefore, already contained inherent errors and required Wiener filtering.

The unnormalised images can be seen in Figure 7.10 and the normalised images in Figure 7.11. The NMSE and FEM values are given in Table 7.4. The $K2$ Wiener filter has lower values than th $K1$ Wiener filter and, unexpectedly, the Gabor $K2$ Wiener filter has lower values than the standard $K2$ Wiener filter. In general, the NMSE and FEM values are comparable to those for the convolution methods.



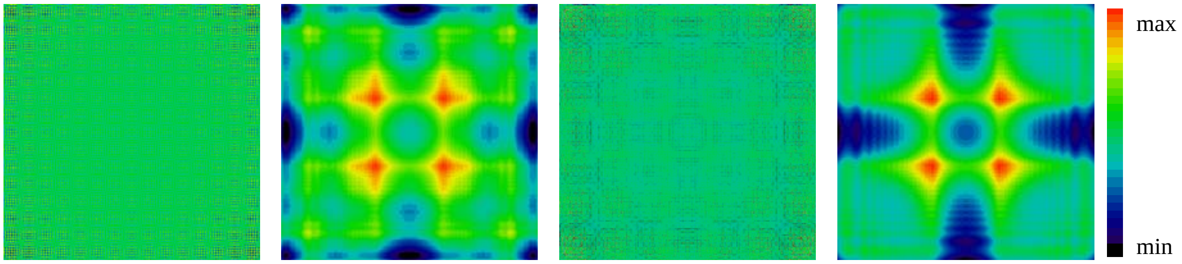
(a) $K1$ Wiener filter -974.84≤pixel value≤1036.0

(b) $K2$ Wiener filter -2.9269≤pixel value≤5.3785

(c) Gabor $K1$ Wiener filter -1545.5≤pixel value≤1600.5

(d) Gabor $K2$ Wiener filter -2.5816≤pixel value≤5.7461

Figure 7.10: Unnormalised images of uniform sources



(a) $K1$ Wiener filter -1.4818e+04≤pixel value≤1.4849e+04

(b) $K2$ Wiener filter -349.33≤pixel value≤1366.5

(c) Gabor $K1$ Wiener filter -2395.5≤pixel value≤2881.5

(d) Gabor $K2$ Wiener filter -70.949≤pixel value≤2310.2

Figure 7.11: Normalised images of uniform sources

| | NMSE | FEM |
|---|---|---|
| $K1$ Wiener filter | 1.0498e+05 | 0.99998 |
| $K1$ Wiener filter - normalised | 8.2221e+06 | 0.99999 |
| $K2$ Wiener filter | 1.7815 | 0.67093 |
| $K2$ Wiener filter - normalised | 3.8289e+05 | 0.22447 |
| Gabor $K1$ Wiener filter | 1.8204e+05 | 0.99927 |
| Gabor $K1$ Wiener filter - normalised | 2.3337e+05 | 0.99866 |
| Gabor $K2$ Wiener filter | 1.3856 | 0.51682 |
| Gabor $K2$ Wiener filter - normalised | 1.2132e+06 | 0.12472 |

Table 7.4: Comparison of error metric values for images of uniform sources for the different deconvolution reconstruction techniques

## 7.11 Summary

Methods of reconstruction by deconvolution have been outlined. The need for a Wiener filter has been explained and different methods for finding the optimal filter have been given. A way to normalise the image amplitude has also been given. A method using a Gabor zone plate pattern has been introduced to overcome effects caused by mis-alignment of the zone plate. Error metric values, modulation transfer functions and point spread functions have been calculated and shown to be better than those for convolution method. This technique, however, has the disadvantage of having to find an optimal value of $K1$ or $K2$. The method giving the most favourable results for theoretical point sources was the $K2$ Wiener filter.

# Chapter 8

# IMAGE RECONSTRUCTION USING THE CLEAN ALGORITHM

The CLEAN algorithm was invented by Högbom in 1974 for use in high-resolution radio interferometry [38]. The CLEAN algorithm was one of a number of techniques which Högbom presented in his paper. To quote Högbom

> The procedure that appears to work the best at the time of writing has become known as "CLEAN". It is an iterative procedure that operates in the map plane and which uses the known shape of the dirty beam to distinguish between real structure and sidelobe disturbances in the dirty map...

No explanation is given for the procedure being called "CLEAN". In this case, the dirty map would be the hologram and the dirty beam would be the zone plate pattern, that is, the shadow cast by a point source. The method works by subtracting the contribution of each point to the hologram a point at a time until very little remains.

The CLEAN algorithm has been used in microscopy [61], [29], and radioastronomy [22], [55] for image reconstruction. There are many variations on this basic algorithm and these are outlined on the 'Mathworld' web site [69]. This algorithm is, however, usually used to reconstruct two-dimensional images in one plane. The version of the algorithm used here performs the following iteration procedure to reconstruct an image in any number of planes in the $z$ direction.

1. Set all pixel values in the reconstructed image planes to zero.
2. Reconstruct the image in all the planes by normalised convolution. The resulting image will be referred to as the hologram image. As the reconstruction is normalised, the hologram image pixel values should roughly correspond to the source intensity per unit area for each position.

3. Find the maximum hologram image pixel value and then increment the corresponding reconstructed image pixel by a given amount $\Delta_c$.
4. Subtract the contribution that a source in this position with strength $\Delta_c$ would have had from the hologram data.
5. Go back to step number 2 and repeat this process until the maximum value of the hologram image is less than $\Delta_c$.

An alternative to this sequence would be to replace the word convolution by deconvolution in the above sequence. Potentially, any of the methods outlined in Chapters 6 and 7 could be used. As the truncated Fresnel function and $K2$ Wiener filter methods were found to be the most promising, these are the ones that will be used in this work.

The convolution or deconvolution reconstruction may not reconstruct all the image pixels with the correct amplitude. If it did, this technique would not be needed. What is important, however, is that the image pixel with the maximum amplitude corresponds to a position where there is a source. If $\Delta_c$ is too high, there is a risk that the pixel will be given too high an amplitude and too much will be subtracted from the hologram for that position. If this happens, the behaviour of the algorithm will be unpredictable. A higher value of $\Delta_c$ will mean the image is reconstructed in less time but a lower value should give a more accurate image.

## 8.1 Assessment of Image Quality

To compare this technique with convolution and deconvolution, holograms of point sources with the same parameters as those in Chapters 6 and 7 were reconstructed using both the convolution based and the deconvolution based CLEAN algorithms. In each case, the reconstructed points were in the correct pixels and all the other pixel values were still zero. This was still the case when the reconstruction planes were placed $1\times10^{-5}$m apart. The accuracy of the amplitude of the reconstructed point depended on the value of $\Delta_c$ with a lower value of $\Delta_c$ giving a more accurate value. Finding the FEM and NMSE error metric values is therefore meaningless here. A Gaussian can also not be fitted to a single pixel value surrounded by zeros.

As only one pixel value is changed by each iteration, the time this reconstruction method takes is much longer than convolution and deconvolution methods. The time it takes is the number of iterations multiplied by the time for one iteration. The number of iterations depends upon the value of $\Delta_c$ and the size and strength of the source. The time taken for one iteration is slightly more than the time taken to reconstruct the same image by either convolution or deconvolution, whichever one is used.

## 8.2 Summary

The application of the CLEAN algorithm to the reconstruction of three-dimensional images has been explained. This algorithm can be based on either a convolution or a deconvolution reconstruction method. The results for theoretical point sources are incredibly good but the CLEAN algorithm reconstruction takes many times longer to run than convolution or deconvolution reconstruction. The time needed depends upon the size and strength of the source as well as the number of pixels in the hologram, the number of image planes and the amount by which an image pixel is incremented by one iteration.

This algorithm appears to work well for point sources. This could have been expected as it was originally designed to image point sources. It has only be shown to work well for extended sources when given a reasonable starting model [69] but no a priori knowledge is used to reconstruct the images in Chapters 9 and 10. The algorithm used in this work is similar to the basic Högbom CLEAN algorithm but is used for images in more than one plane. There are many variations of the CLEAN algorithm which have been shown to achieve greater speed and produce more realistic images [69] but investigating all these different possibilities is beyond the scope of the work presented here.

# Chapter 9

# RECONSTRUCTION OF COMPUTER GENERATED HOLOGRAMS

In this Chapter some more theoretical holograms will be reconstructed using the three reconstruction methods which were explained in Chapters 6, 7 and 8. The convolution method which will be used is convolution with a truncated Gabor zone plate as this was found to give the best results for the point sources in Chapter 6. $K2$ Wiener filtering will be the deconvolution method used here as this was the best method for point sources in Chapter 7. First, the reconstruction techniques will be tested on images from an ideal theoretical hologram of a simple, one pixel thick cross source. The parameters of this hologram will, again, be those given in Table 6.1. The zone plate dimensions and the pixel size in this table are the same as those of the real experimental holograms in Chapter 10.

The theoretical holograms will be calculated using the relationship between source distribution given by equation 6.2. As in equation 3.88, this relationship assumes that the source to zone plate ($p$) and zone plate to detector ($q$) distances are much larger than the dimensions of the detector. In this case, the distance a gamma ray travels from a point in the source to the detector is approximately equal to $(p + q)$ no matter where the point is is the field of view. In practise, for the parameters commonly used for zone plate holography, $p$ and $q$ would be too small for this to be a good approximation. The work in this Chapter, however, will just attempt to give an indication of whether the methods should work or not and will not include all of the factors which affect a zone plate hologram. The efficiency of the gamma camera and the fact that the zone plate is several millimetres thick, for example, are other factors which will not be included here.

The effect of adding other sorts of errors to an ideal hologram will, however, be investigated. These are the errors which were mentioned in Chapter 4 and they will be measured using the error metrics given in Chapter 5. Images from a hologram of two different cross sources in two different planes will then be reconstructed. Next, the reconstruction methods will be used to reconstruct images from ideal computer generated holograms of the same object but where the size of the object is different in different holograms. Finally, the methods will be used to reconstruct images from holograms of the same object but with different parameters.

## 9.1 Hologram of a Simple Two-Dimensional Cross Source Placed Parallel to the Zone Plate

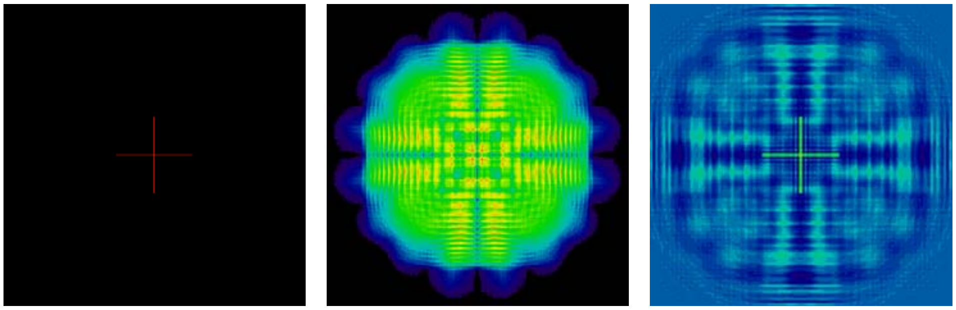
### 9.1.1 Reconstruction of Images from an Ideal Hologram

An ideal hologram of the cross source shown in Figure 9.1(a) was calculated using the parameters in Table 6.1. This hologram was reconstructed using the different methods in five planes each separated by 2cm where the image should appear in the middle plane. First, the image was reconstructed using normalised convolution and normalised $K2$ Wiener filtering. As was explained in Chapter 8, the CLEAN algorithm needs to be based on another image reconstruction technique. As reconstruction using the CLEAN algorithm takes many times longer, it seemed sensible to reconstruct the image using the other methods and then decide which one of these methods the CLEAN algorithm should be based on.

| | convolution - real part | convolution - intensity | $K2$ Wiener filter |
|---|---|---|---|
| FEM | 9.8571e-01 | 9.9657e-01 | 9.4923e-01 |
| NMSE | 5.1198e+02 | 1.1767e+05 | 2.9741e+01 |

Table 9.1: Error metrics for the normalised image reconstruction of a hologram of the cross source

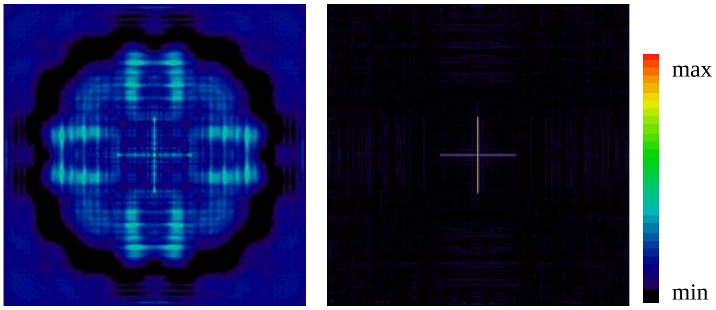
To find the optimum value of $K2$, $K2$ was first set equal to the maximum value of the square of the magnitude of the Fourier transforms of ideal holograms of points in each plane. This was used as a starting point as we can be fairly certain that the errors do not have a greater contribution than the maximum frequency contribution of the signal. The image was reconstructed using this value of $K2$. The hologram that would have been generated by this reconstructed image was then calculated. The FEM error metric for



(a) source object 0≤pixel value≤1

(b) theoretical hologram - 4.4806e-26≤pixel value≤1.6103e-11

(c) convolution - real part -1.7719≤pixel value≤4.8903



(d) convolution - intensity 1.6982e-05≤pixel value≤26.159

(e) $K2$ Wiener filter -0.062004≤pixel value≤1.5615

Figure 9.1: Normalised reconstruction by convolution and $K2$ Wiener filter

the difference between this calculated hologram and the actual hologram were then found. Next, $K2$ was divided by ten and the whole process was repeated and the new FEM value was found. $K2$ continued to be divided by ten until the FEM value stopped decreasing and increased. The previous value of $K2$ was then assumed to be the optimal value.

First, images were reconstructed using normalised convolution and deconvolution. The images reconstructed in the middle plane can be seen in Figure 9.1 and all the reconstructed image planes can be seen in Figures 9.2 and 9.3. Table 9.1 gives the FEM and NMSE error metric values for the difference between the source object and the image over all five planes. The error metric values confirm what is obvious from looking at the Figures, $K2$ Wiener filtering gives a better image than convolution. In the calculations performed in Chapter 3, it was predicted that the intensity of a pure Gabor zone plate hologram of a point source should give a point image but that the real part would have other terms in the background. The intensity was also found to give a better reconstruction for binary Gabor zone plate holograms of point sources in Chapter 6. For this cross source, however, the real part gives a better reconstruction.

| | convolution - real part | convolution - intensity | $K2$ Wiener filter |
|---|---|---|---|
| FEM | 9.7860e-01 | 9.9390e-01 | 9.0609e-01 |
| NMSE | 3.4022e+02 | 6.5848e+04 | 1.5379e+01 |

Table 9.2: Error metrics for the partially normalised image reconstruction of a hologram of the cross source

As was explained in Chapter 6, for some points, the whole of the zone plate pattern is not recorded by the gamma camera. This was accounted for in the normalisation of the convolution reconstruction. Although this helps points towards the edge of the hologram have the correct magnitude, it also amplifies errors increasing with distance from the centre. In Chapter 7, an attempt was also made to account for this 'overlap' problem in the Wiener filtering methods. The results were, however, not as satisfactory as it was more difficult to assess the effect that overlapping has in this case. Partially normalised reconstructions where the overlap factor was not included were therefore calculated.

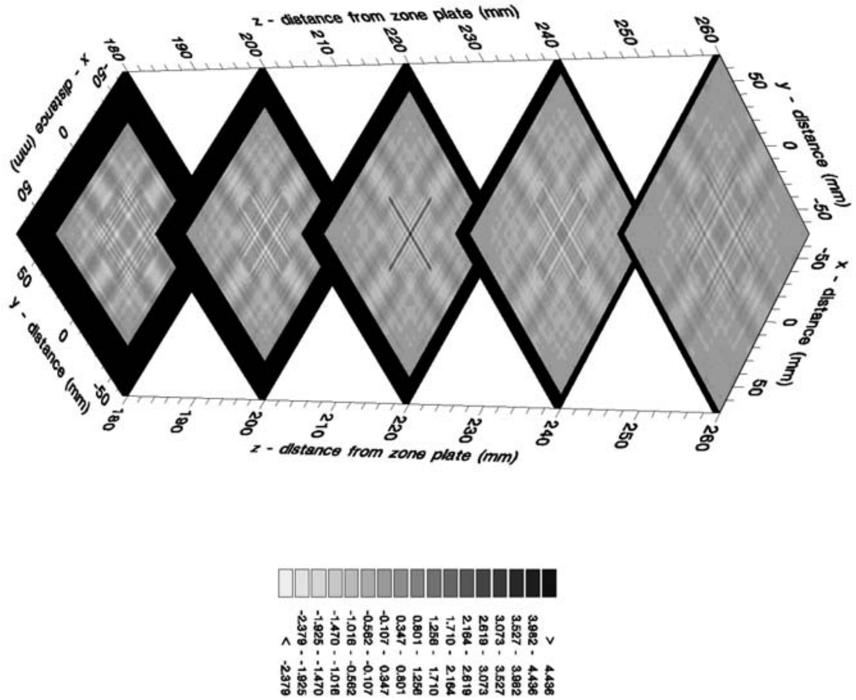
The images reconstructed in the middle plane can be seen in Figure 9.4 and all the reconstructed image planes can be seen in Figures 9.5 and 9.6. Table 9.2 gives the FEM and NMSE error metric values for the difference between the source object and the image over all five planes. The Figures show an obvious difference for the intensity of the convolution but the real part of the convolution and the $K2$ Wiener filter look very similar. The FEM and NMSE error metric values are, however, all lower than those for the fully normalised reconstruction. The partial normalisation method, therefore, appears to be better.

| | CLEAN (convolution - real part) | CLEAN ($K2$ Wiener filter) |
|---|---|---|
| FEM | 1.1594e-01 | 5.4048e-03 |
| NMSE | 1.1922e-01 | 5.0331e-02 |

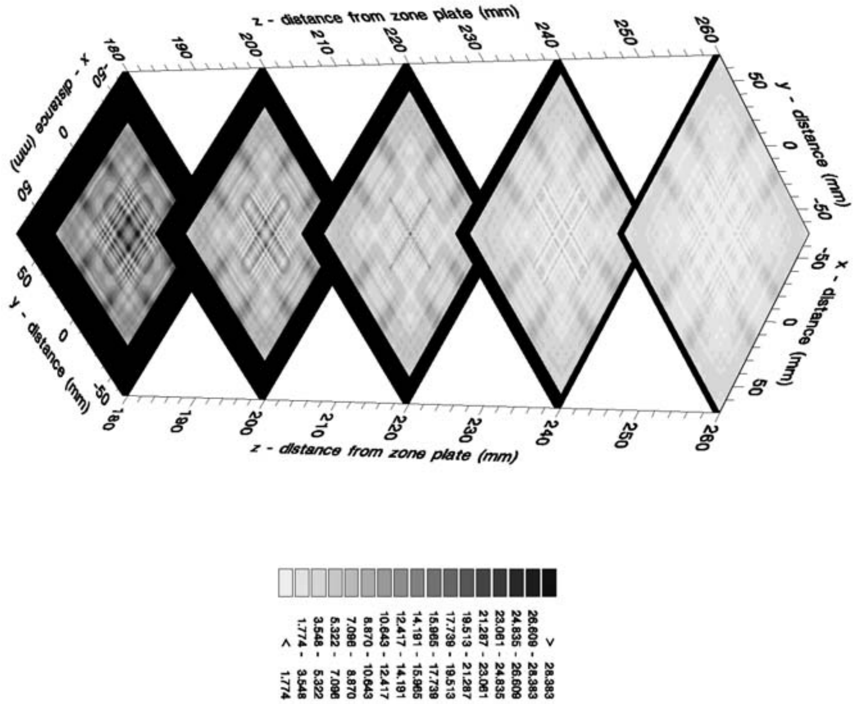
Table 9.3: Error metrics for the partially normalised image reconstruction of a hologram of the cross source using the CLEAN algorithm

The partially normalised convolution (real part) and $K2$ Wiener filter methods were used as the basis for the CLEAN algorithm. $\Delta$, the increment by which a pixel value is increased by one iteration, was set to the maximum value of the image reconstructed by convolution or deconvolution divided by ten. $\Delta$ for convolution was greater than that for deconvolution so more iterations would have been needed. Deconvolution also takes longer because there are more arithmetic operations involved. The convolution based CLEAN algorithm reconstruction ran overnight and was finished by the morning. When the reconstruction using the deconvolution based CLEAN algorithm had not finished two days later, $\Delta$ was decreased to half of the maximum value of the image reconstructed by deconvolution.

The results for the object plane can be seen in Figure 9.7 and in all five planes in Figure 9.8. The CLEAN algorithms mainly reconstructed points in the correct plane, most of the values in the other three planes remained equal to zero. When straightforward convolution and deconvolution were used, this was not the case. The corresponding FEM and NMSE error metric values are given in Table 9.3.



(a) real part



(b) intensity

Figure 9.2: Normalised reconstruction by convolution

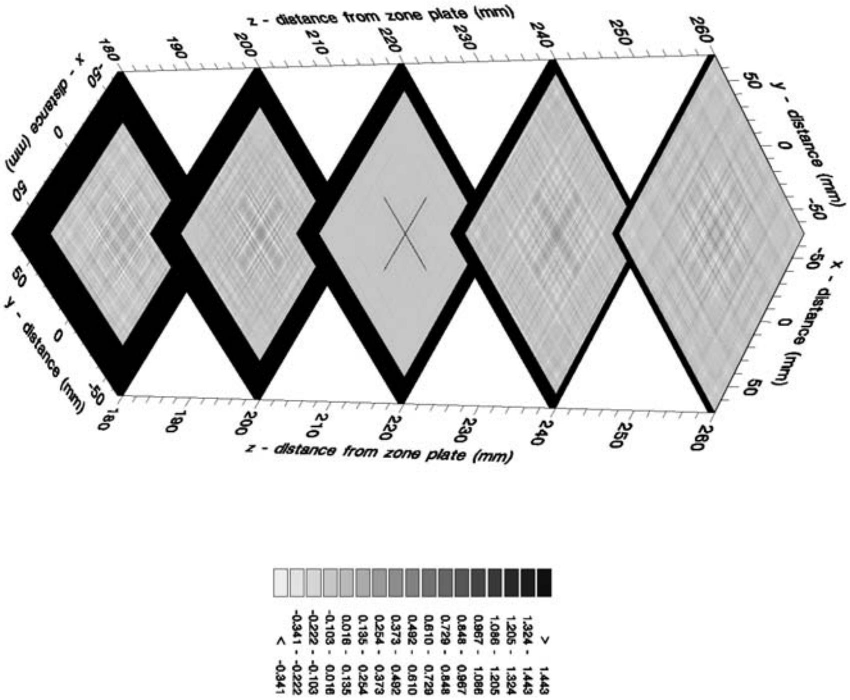
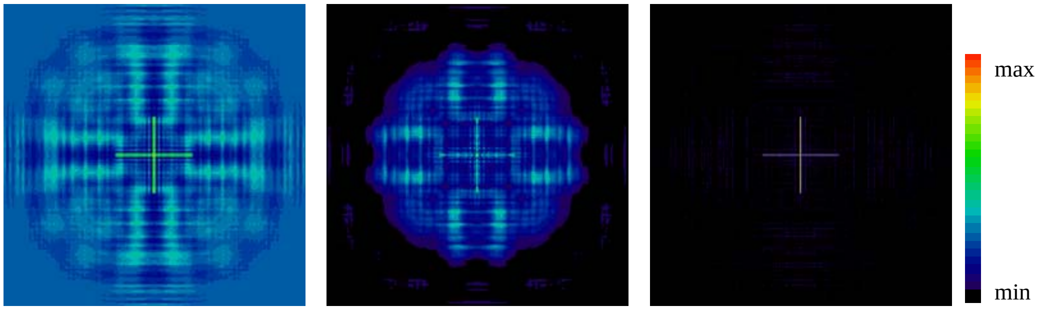


Figure 9.3: Normalised reconstruction by $K2$ Wiener filter

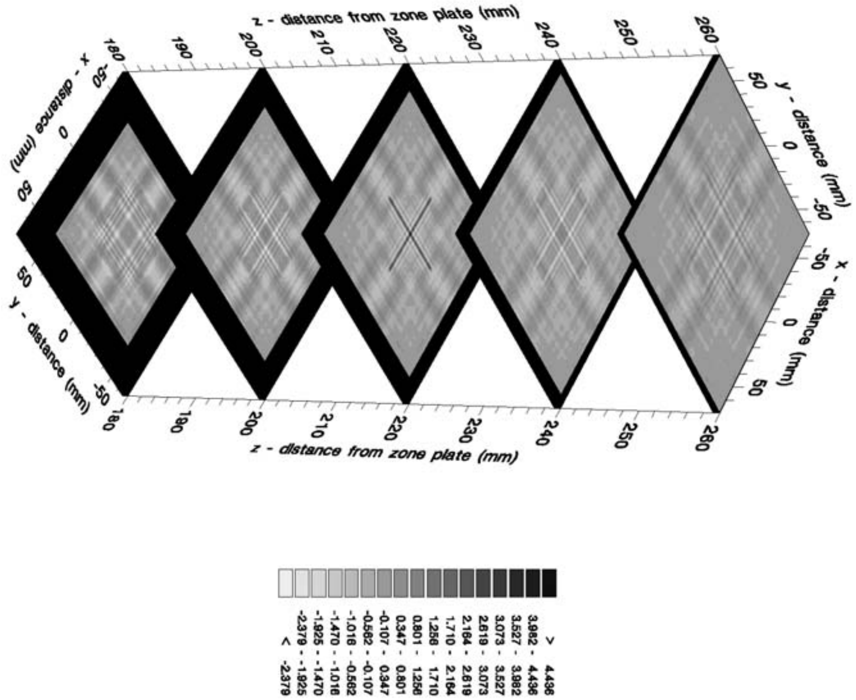


(a) convolution - real part $-1.7719 \leq$ pixel value $\leq 4.8903$

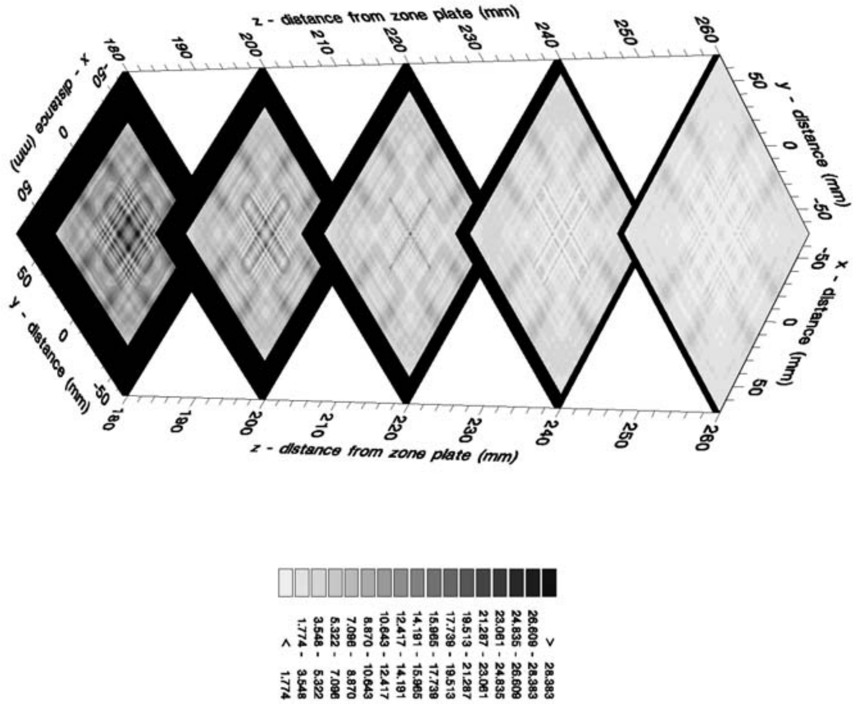
(b) convolution - intensity $8.4712\text{e-}06 \leq$ pixel value $\leq 26.159$

(c) $K2$ Wiener filter $-0.061567 \leq$ pixel value $\leq 1.5615$

Figure 9.4: Partially normalised reconstruction by convolution and $K2$ Wiener filter



(a) real part



(b) intensity

Figure 9.5: Partially normalised reconstruction by convolution

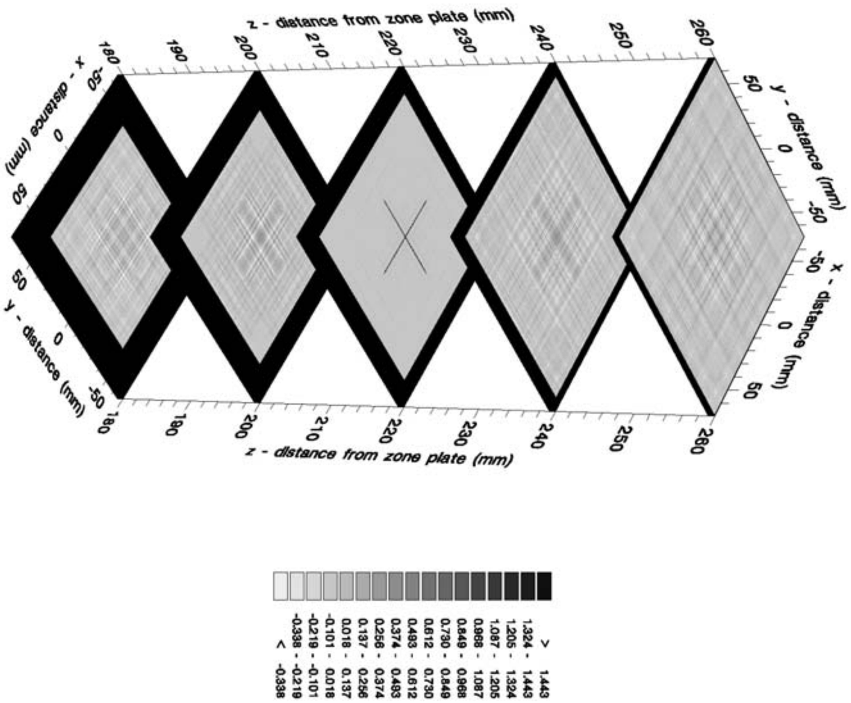
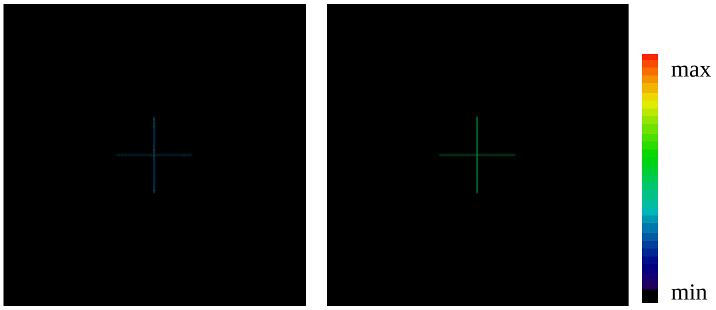


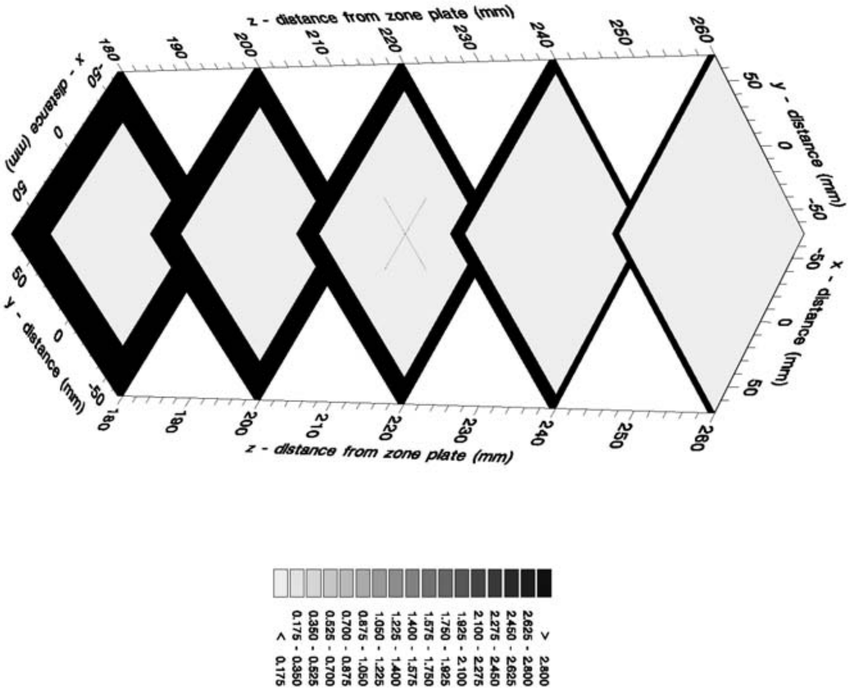
Figure 9.6: Partially normalised reconstruction by $K2$ Wiener filter



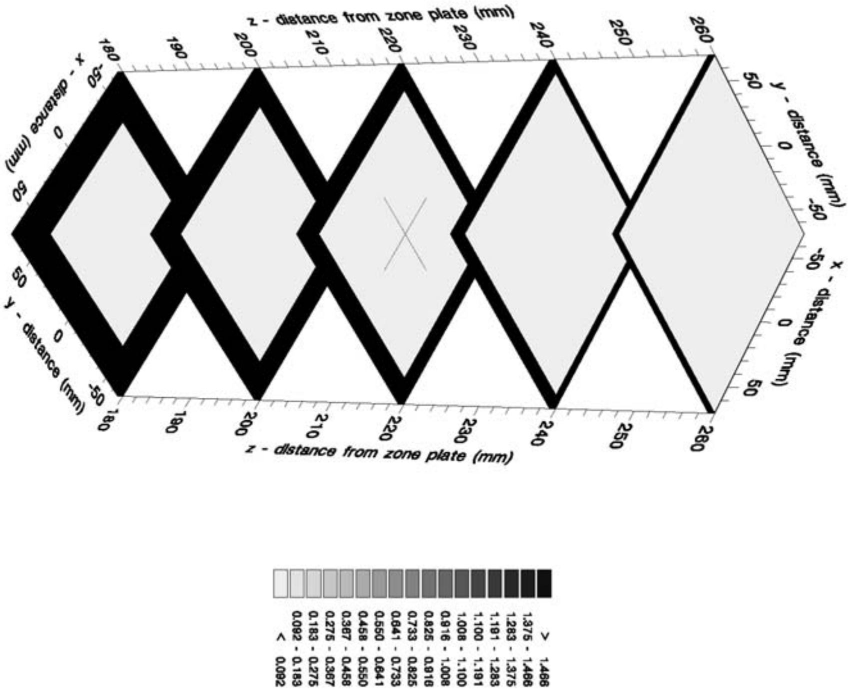
(a) CLEAN algorithm (convolution) $0 \leq$ pixel value $\leq 2.9754$

(b) CLEAN algorithm (deconvolution) $0 \leq$ pixel value $\leq 1.5578$

Figure 9.7: Partially normalised reconstruction using the CLEAN algorithm



(a) algorithm based on convolution



(b) algorithm based on deconvolution

Figure 9.8: Partially normalised reconstruction using the CLEAN algorithm

### 9.1.2 Reconstruction of Images from Holograms with Added Errors

Four different types of error were added to the ideal hologram of the cross source to see the effect it had on the image. These errors will be numbered ①, ②, ③ and ④. The errors included in each one are listed below.

① The source magnitude was replaced by a number taken from a Poisson distribution with a mean of 100 for each pixel containing the source. As a Poisson distribution has a standard deviation equal to the square root of the mean, the standard deviation of this distribution would have been 10 [24]. This was intended to simulate a source of magnitude 100 with ten percent noise due to statistical fluctuations. The theoretical hologram was then calculated using this source.

② The ideal hologram was convolved with a two-dimensional Gaussian function with FWHM equal to 0.38cm which is the spatial resolution of a Toshiba GCA-7200A gamma camera given in Table 4.1. This was in order to simulate the fact that the gamma camera stores the data in pixels which are smaller than the resolution of the camera.

③ Noise taken from a uniform distribution between zero and ten percent of the maximum ideal hologram value was added to the ideal hologram. This was to simulate errors due to floating point arithmetic and counts from background radiation. The minimum value of the noise was zero so that none of the hologram values would become negative. This is sensible because the output of the gamma camera does not contain negative values.

④ The thickness of the cross source was increased from one to five pixels. This was to see if the reconstruction techniques worked as well for a source which covered a larger number of pixels next to each other.

Images were reconstructed from these holograms in the same five planes as for the ideal hologram (Section 9.1.1) using convolution, deconvolution and the CLEAN algorithm. Corresponding error metric values to those given for the ideal hologram can be seen in Tables 9.4 and 9.5. Values for the intensity of the convolution reconstruction are not included as this was found to give a worse reconstruction than the real part in Section 9.1.1.

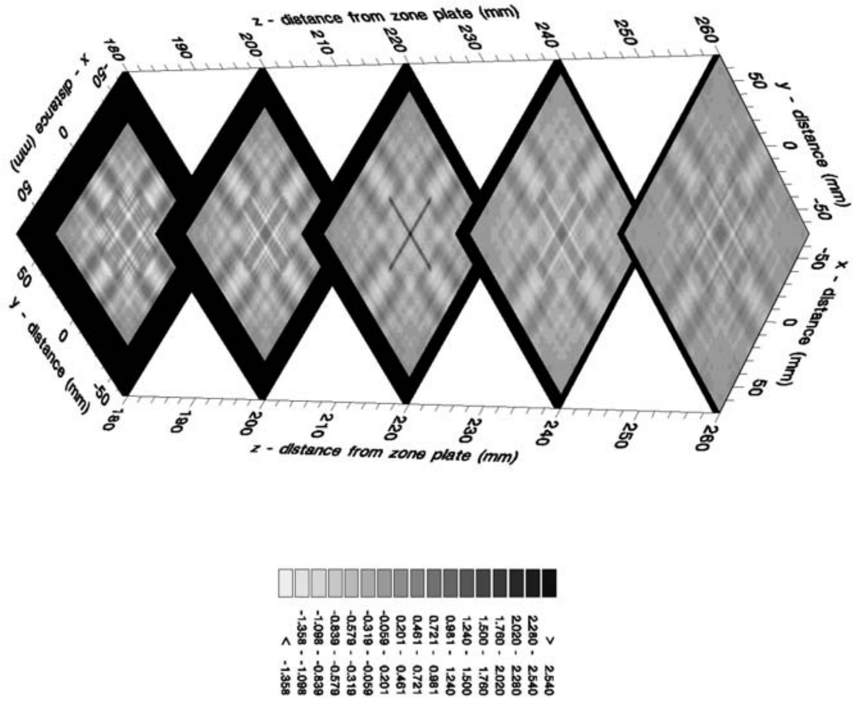
All of the reconstructed images will not be shown here as those from holograms ① and ③ do not vary noticeably from those for the ideal hologram. In these cases, the CLEAN reconstructions also only reconstructed points in the correct plane. When images were reconstructed from holograms ② and ④ using the CLEAN algorithms, points were also reconstructed in the planes next to the image plane. The partially normalised reconstructed image in all five planes can be seen in Figures 9.9 and 9.10 for hologram ② and Figures 9.11 and 9.12 for hologram ④.

| | | convolution - real part | $K2$ Wiener filter |
|---|---|---|---|
| ① | FEM | 9.7860e-01 | 9.0638e-01 |
| | NMSE | 3.4022e+02 | 1.5428e+01 |
| ② | FEM | 9.9214e-01 | 9.7394e-01 |
| | NMSE | 2.6256e+02 | 5.9946 |
| ③ | FEM | 9.7913e-01 | 9.4273e-01 |
| | NMSE | 3.5004e+02 | 2.6171e+01 |
| ④ | FEM | 9.8499e-01 | 9.2735e-01 |
| | NMSE | 1.2793e+03 | 2.6447e+01 |

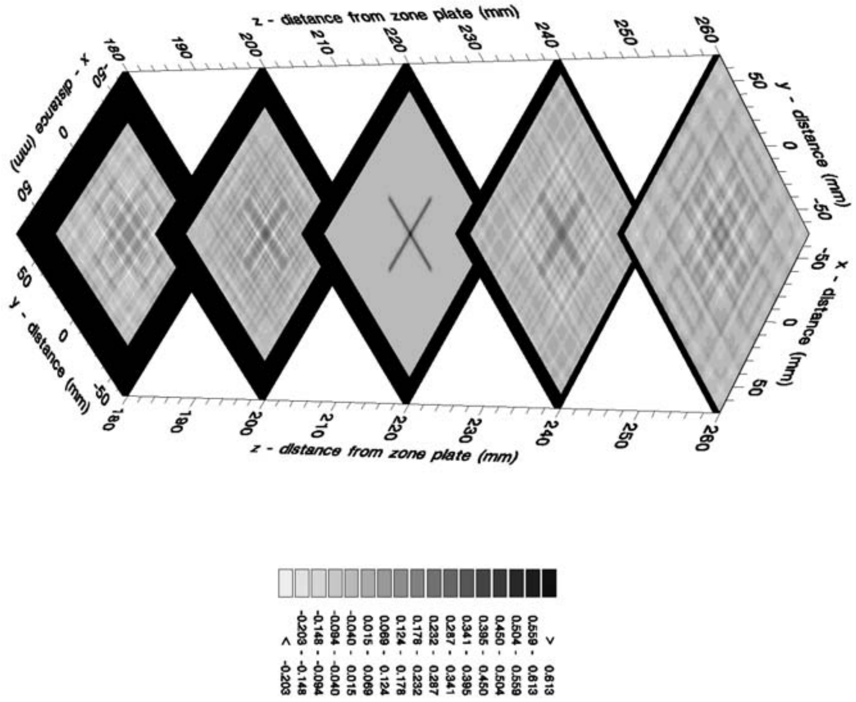
Table 9.4: Error metrics for the partially normalised image reconstruction of noisy holograms of the cross source using convolution and deconvolution

| | | CLEAN (convolution - real part) | CLEAN ($K2$ Wiener filter) |
|---|---|---|---|
| ① | FEM | 8.0714e-02 | 2.2204e-16 |
| | NMSE | 9.1916e-02 | 4.4941e-02 |
| ② | FEM | 8.3049e-01 | 5.1763e-01 |
| | NMSE | 8.9780e-01 | 7.0649e-01 |
| ③ | FEM | 1.1231e-01 | 2.2204e-16 |
| | NMSE | 1.2033e-01 | 4.7384e-02 |
| ④ | FEM | 9.0816e-01 | 8.3823e-02 |
| | NMSE | 2.2752 | 8.9783e-02 |

Table 9.5: Error metrics for the partially normalised image reconstruction of noisy holograms of the cross source using CLEAN algorithms

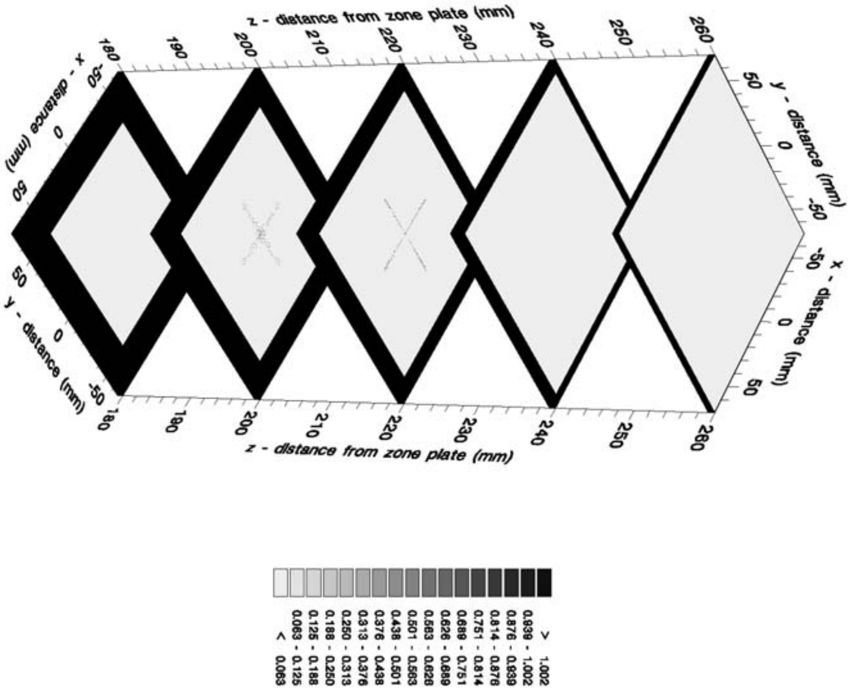


(a) convolution - real part

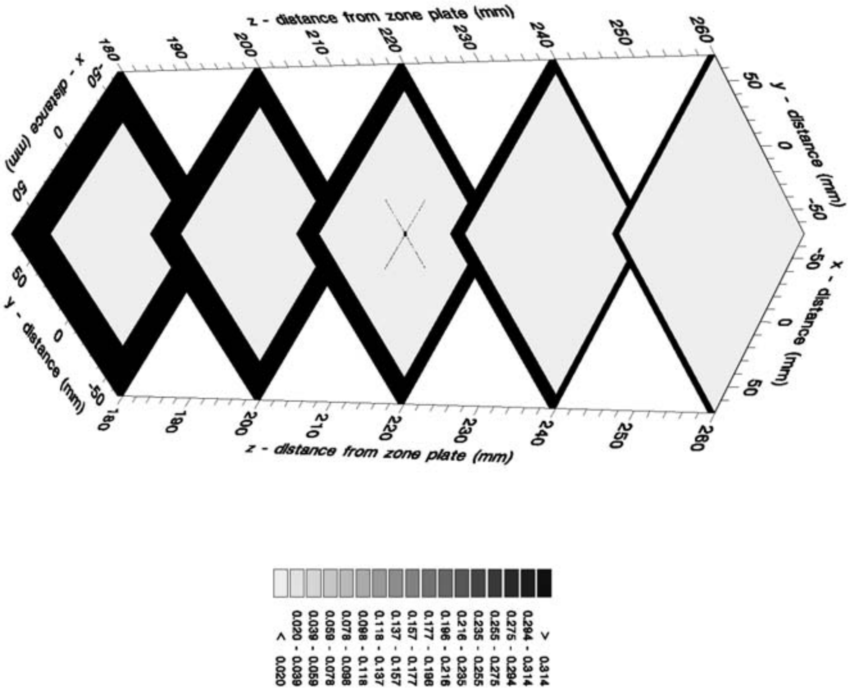


(b) $K2$ Wiener filter

Figure 9.9: Partially normalised reconstruction for hologram ② by convolution and $K2$ Wiener filter

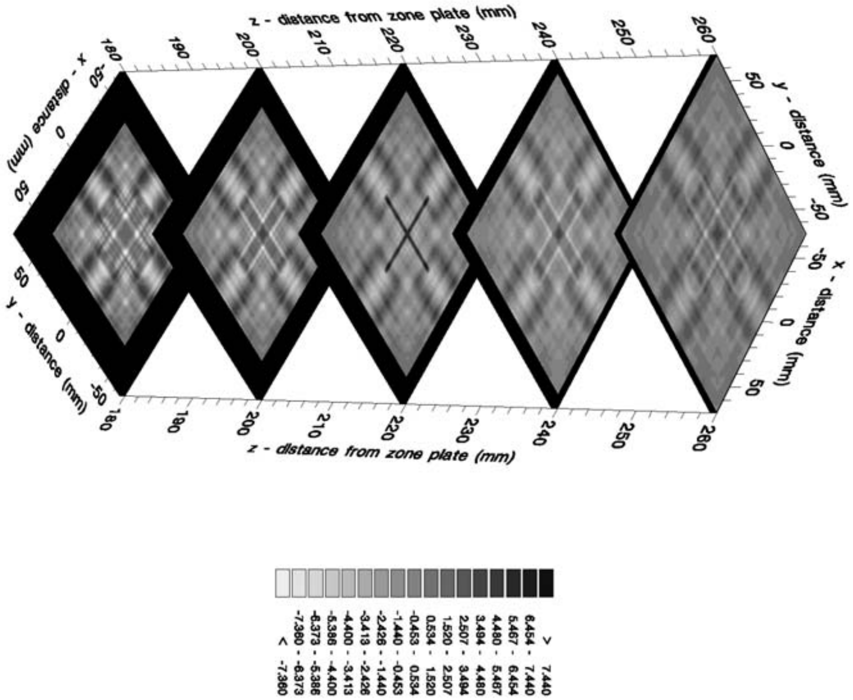


(a) algorithm based on convolution

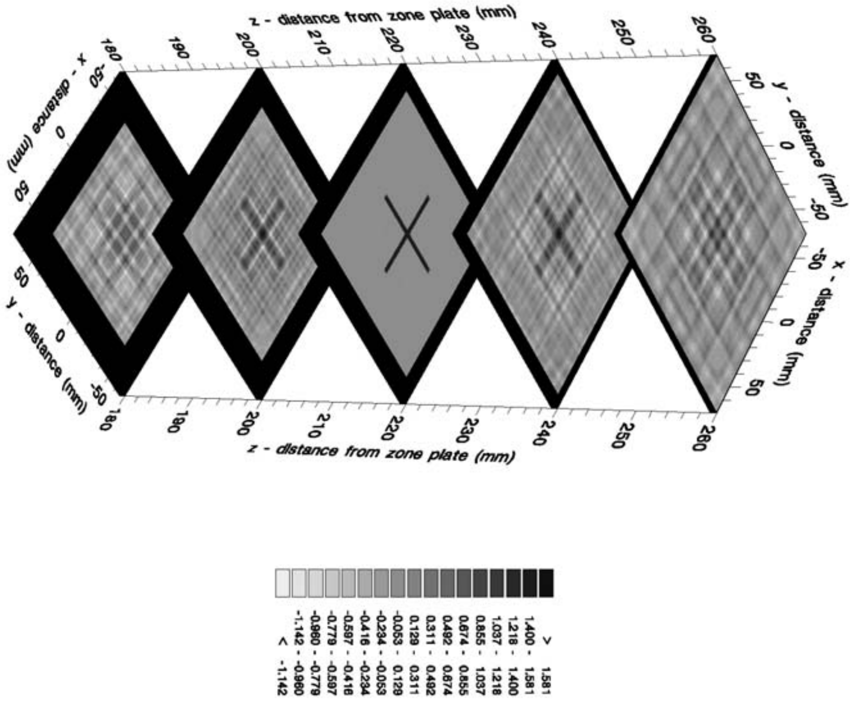


(b) algorithm based on deconvolution

Figure 9.10: Partially normalised reconstruction for hologram ② using CLEAN algorithms



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.11: Partially normalised reconstruction for hologram ④ by convolution and $K2$ Wiener filter

In summary, errors representing statistical fluctuations in the source strength (hologram ①) and recording noise or counts due to background radiation (hologram ③) had less of an effect on the reconstructed image than the limited spatial resolution of the gamma camera (hologram ②). This was just done for one set of hologram parameters and one theoretical source object. The magnitudes of each of these errors would depend upon the exact details of the experiment, for example, source strength. These images, however, give an indication of the sort of effect these errors have.

In the case of $K2$ Wiener filtering, this effect can be lessened by deconvolving the hologram with a zone plate pattern which is also blurred by the same Gaussian function instead of an ideal zone plate pattern. For a partially normalised reconstruction, this reduced the FEM error metric value to 9.5980e-01 but, strangely, increases the NMSE error to 2.6744e+02. If convolution is used, the blurred zone plate pattern can be used to calculate the magnitude of the reconstruction of an ideal point source instead of the pure zone plate pattern and can hence improve the normalisation. Again, this reduces the FEM error value to 9.9156e-01 but increases the NMSE value to 1.4858e+03.

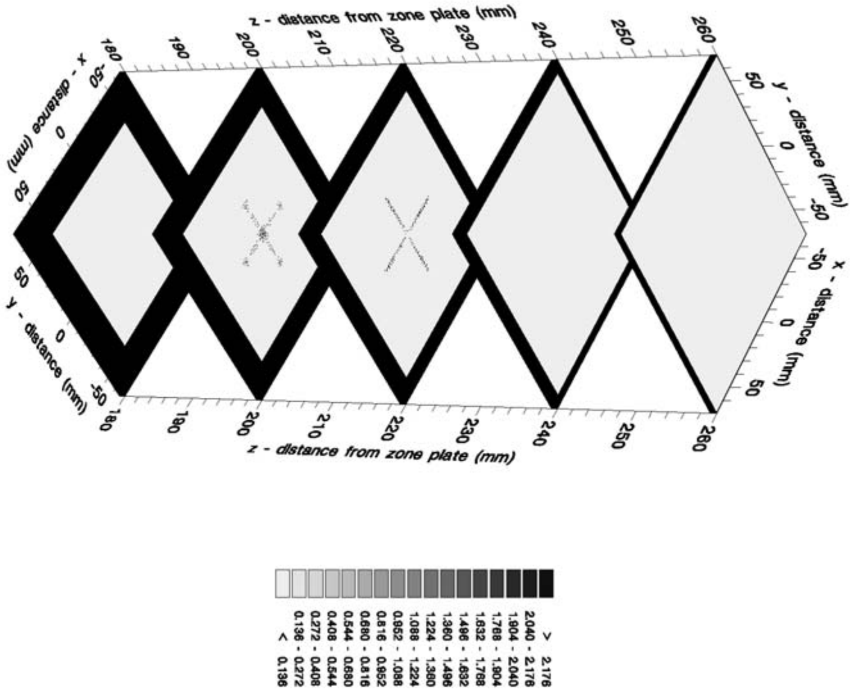
From Table 9.5 it can be seen that the CLEAN algorithm based on deconvolution gave consistently lower error metric values than the CLEAN algorithm based on convolution. The algorithm based on deconvolution will, therefore, be used for the rest of the holograms in this Chapter.

## 9.2 Reconstruction of Images from Holograms of Objects of Different Sizes with the Same Parameters

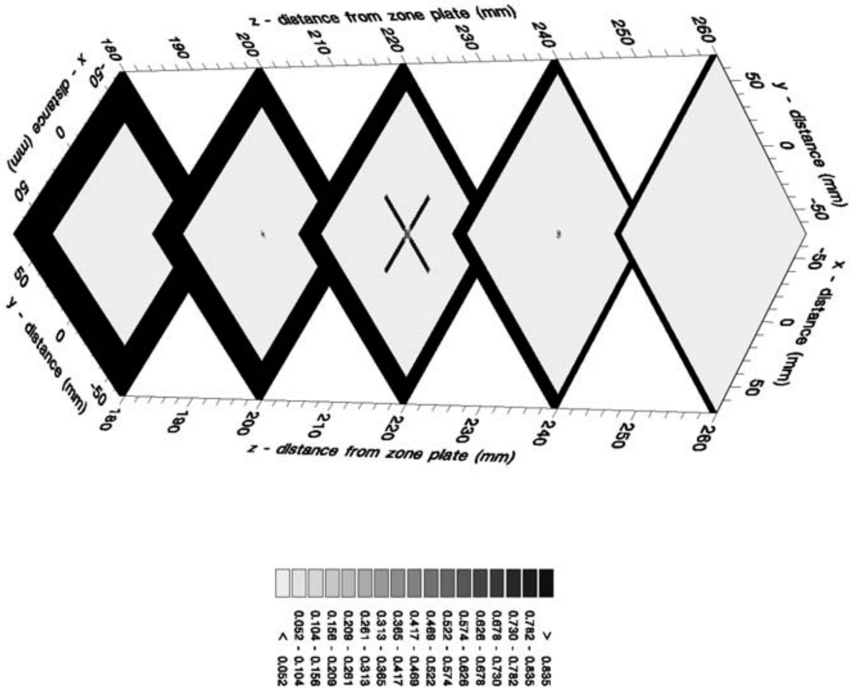
The same object was used for each of these four holograms. It was some text ‘CEJ’ in the centre on the central plane. This was chosen because it is only slightly more complicated than a cross source and because its size can easily be changed by changing the font size. The parameters for the hologram were, again, those in Table 6.1. Four holograms were generated with increasing font size and will be represented by $\boxed{1}$, $\boxed{2}$, $\boxed{3}$ and $\boxed{4}$. The size of the object of which the holograms were calculated increase from hologram $\boxed{1}$ to hologram $\boxed{4}$.

Figures 9.13 and 9.14 show the images from hologram $\boxed{1}$ in five planes each separated by 2cm. The object should appear in the middle plane. Figures 9.16 and 9.15 show the images from hologram $\boxed{2}$, Figures 9.17 and 9.18 those from hologram $\boxed{3}$ and Figures 9.20 and 9.19 show the image from hologram $\boxed{4}$. The corresponding error metric values are given in Table 9.6.

The error metric values for images reconstructed using the CLEAN algorithm increase as the size of the object increases. The NMSE values also increase with object size for convolution and deconvolution. There is no obvious trend in the FEM values for convolution and the FEM values decrease with object size for deconvolution.

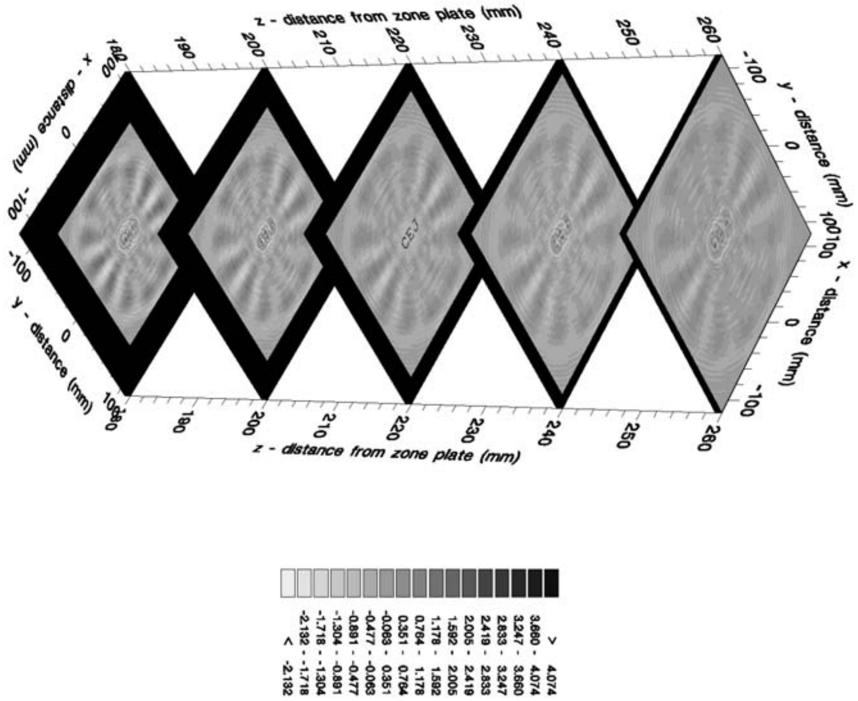


(a) algorithm based on convolution

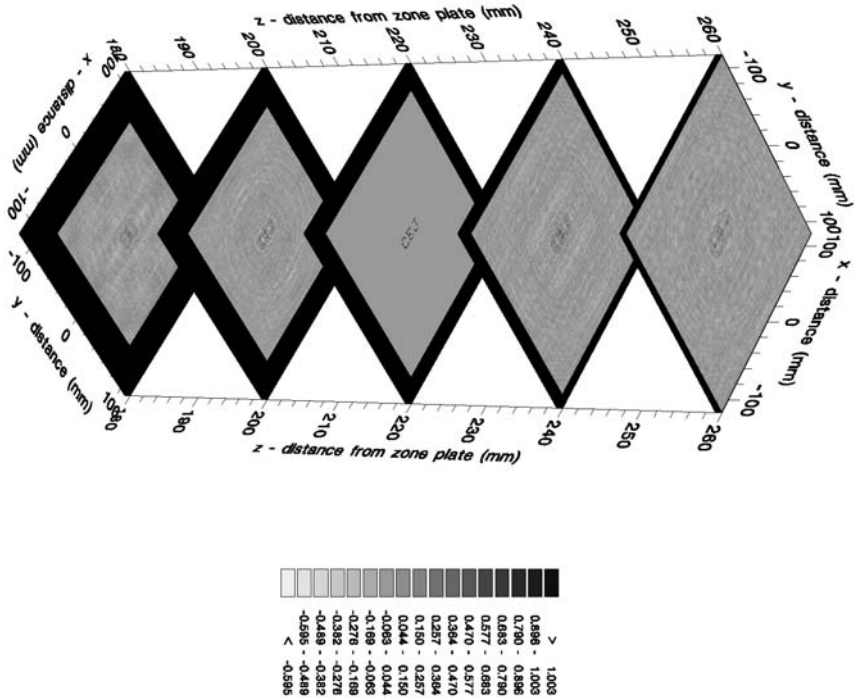


(b) algorithm based on deconvolution

Figure 9.12: Partially normalised reconstruction for hologram ④ using CLEAN algorithms



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.13: Partially normalised reconstruction of hologram $\boxed{1}$ by convolution and $K2$ Wiener filter

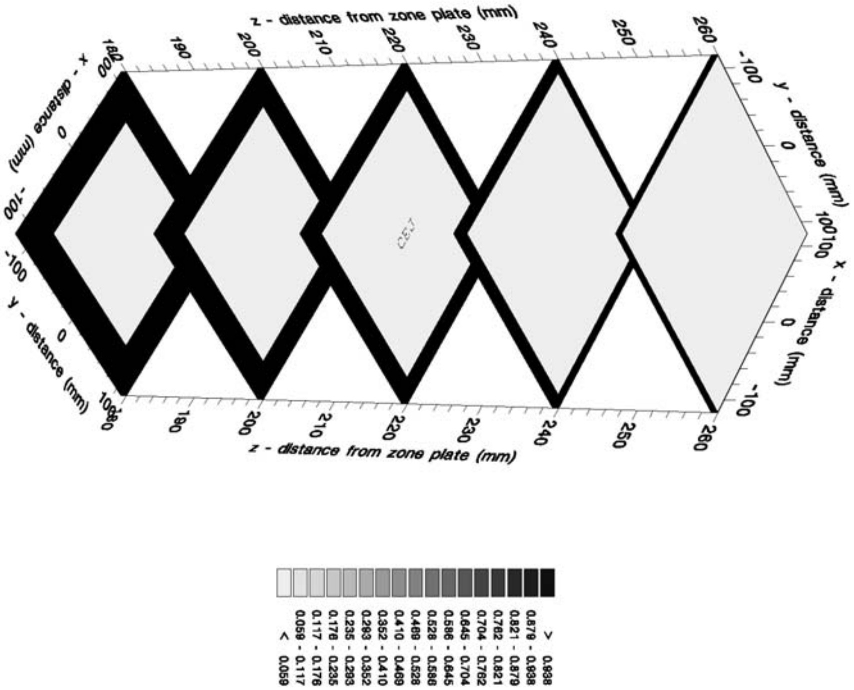


Figure 9.14: Partially normalised reconstruction of hologram $\boxed{1}$ using the deconvolution based CLEAN algorithm

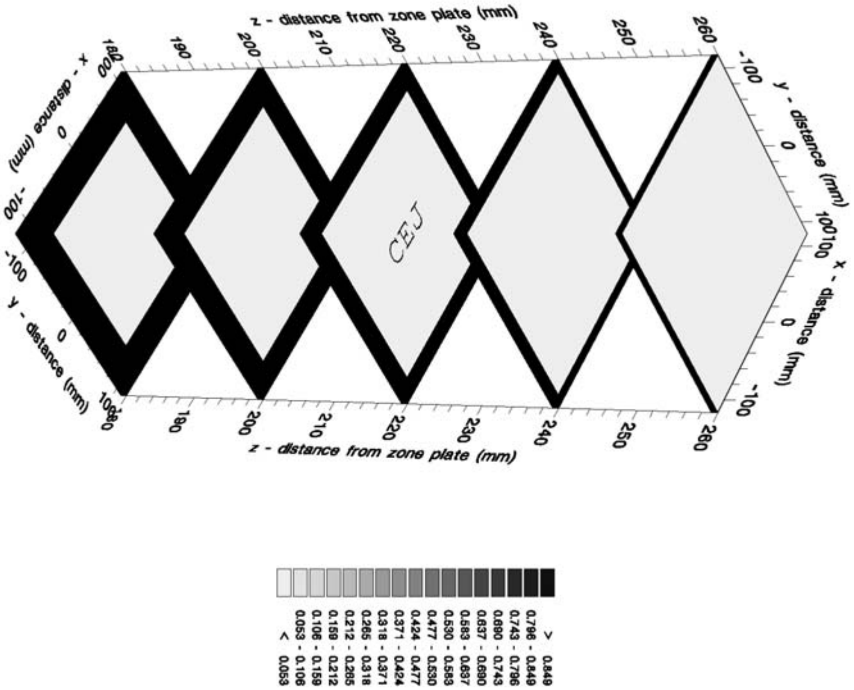
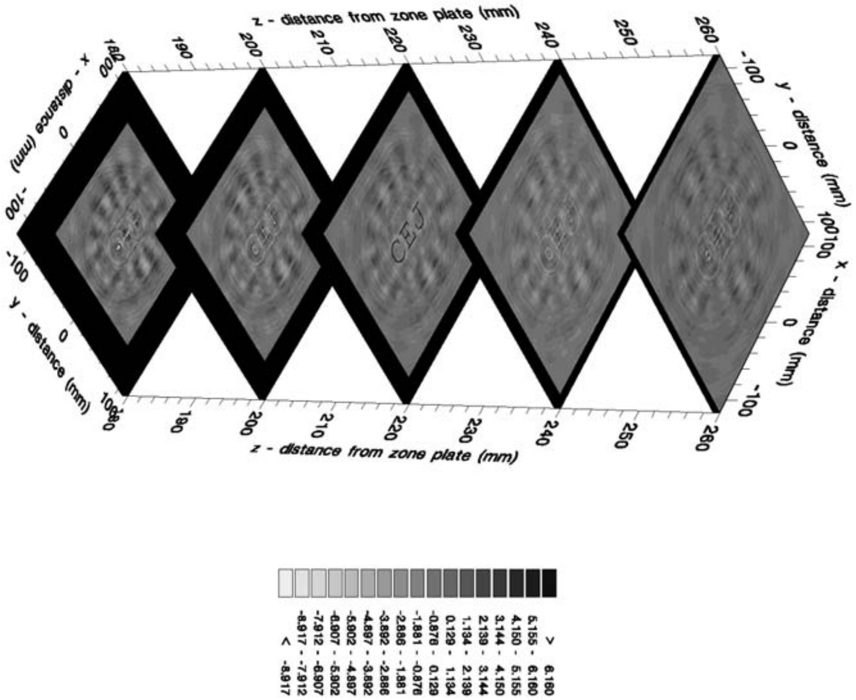
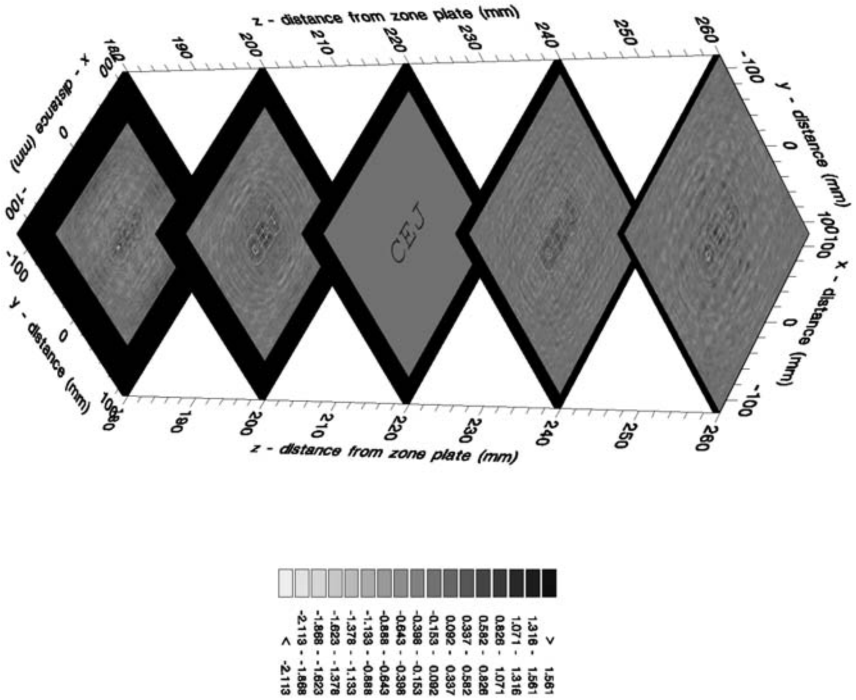


Figure 9.15: Partially normalised reconstruction of hologram $\boxed{2}$ using the deconvolution based CLEAN algorithm

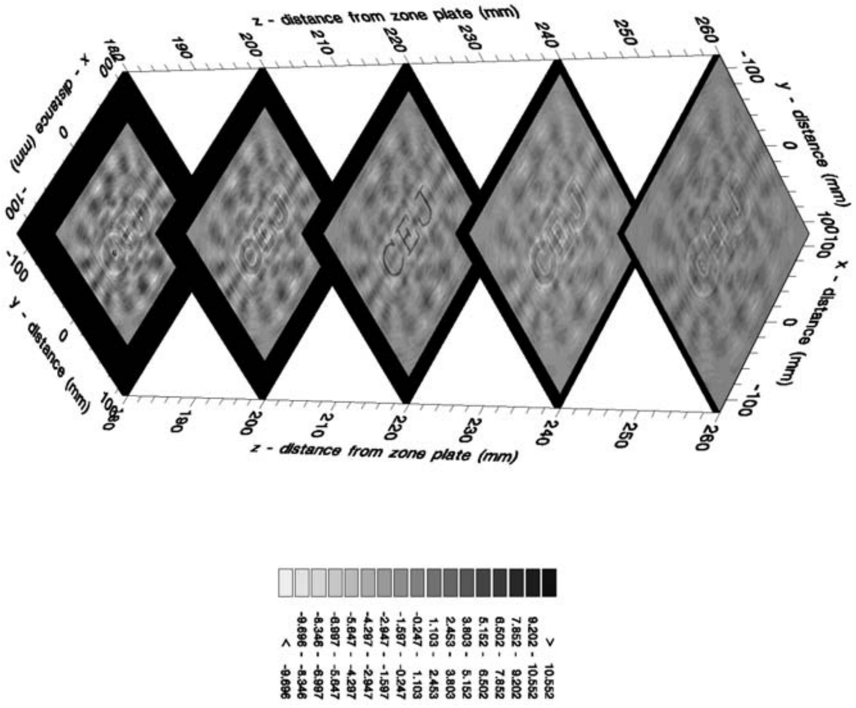


(a) convolution - real part

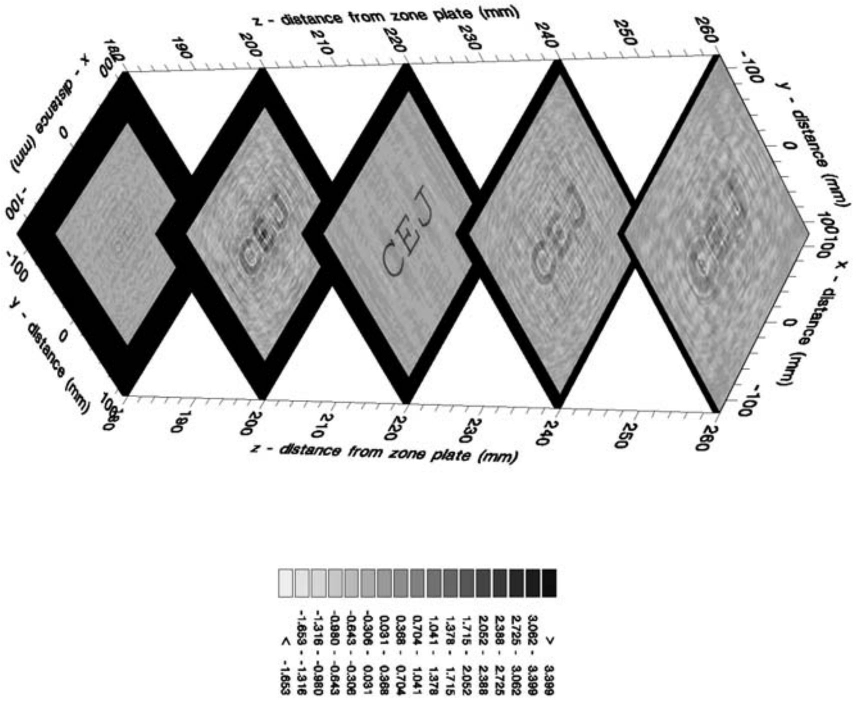


(b) $K2$ Wiener filter

Figure 9.16: Partially normalised reconstruction of hologram $\boxed{2}$ by convolution and $K2$ Wiener filter



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.17: Partially normalised reconstruction of hologram $\boxed{3}$ by convolution and $K2$ Wiener filter

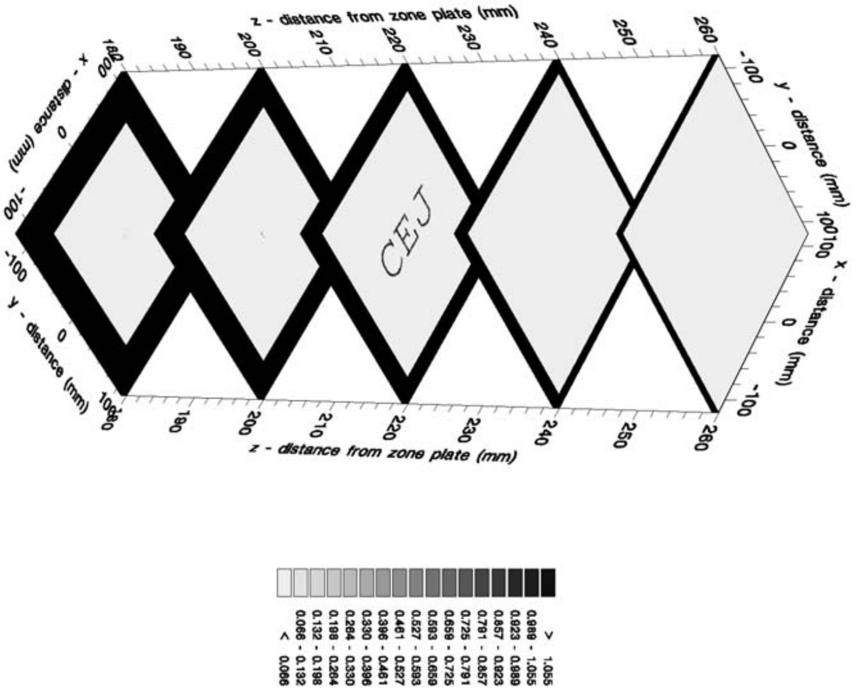


Figure 9.18: Partially normalised reconstruction of hologram $\boxed{3}$ using the deconvolution based CLEAN algorithm

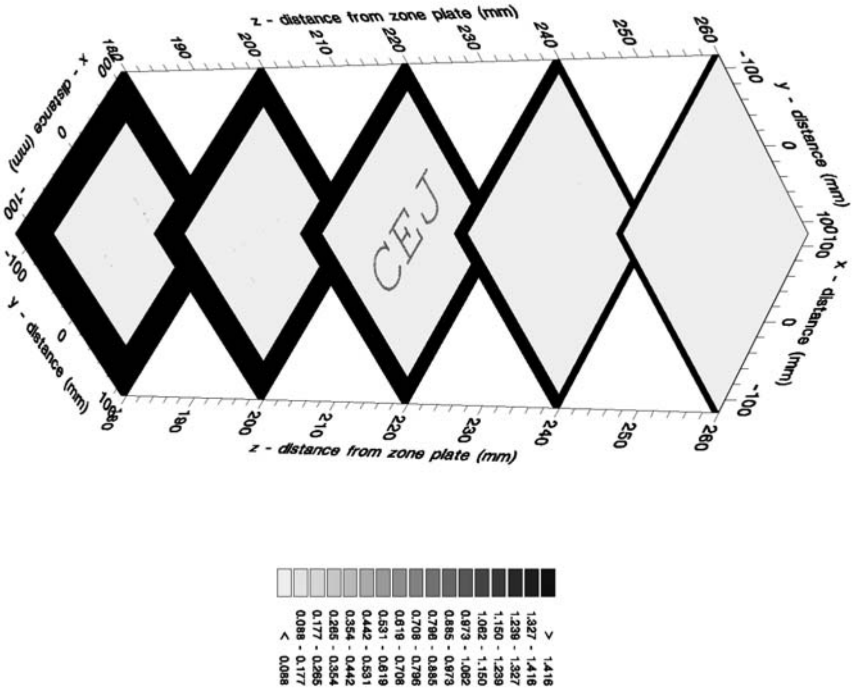
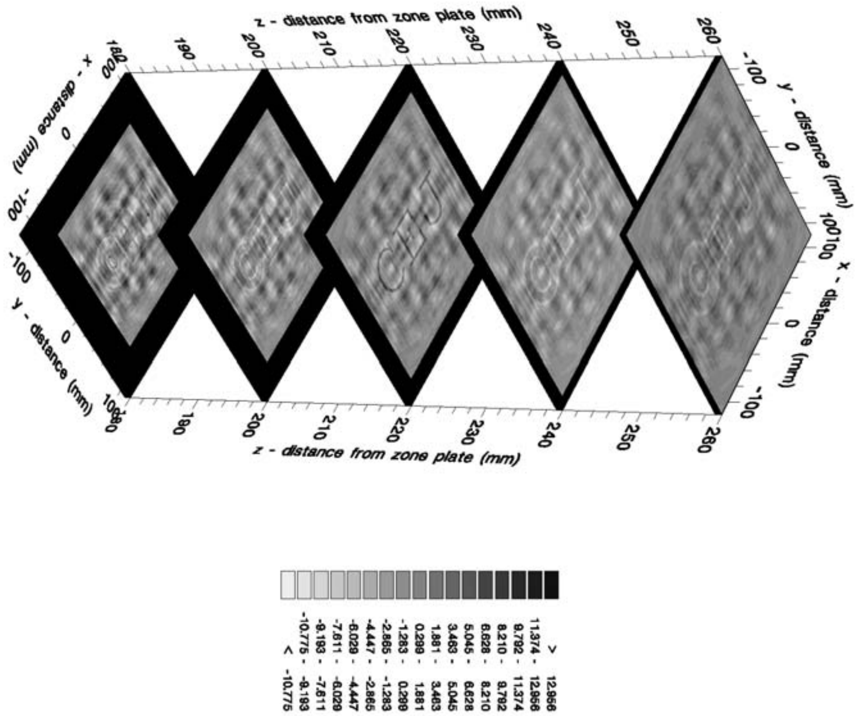


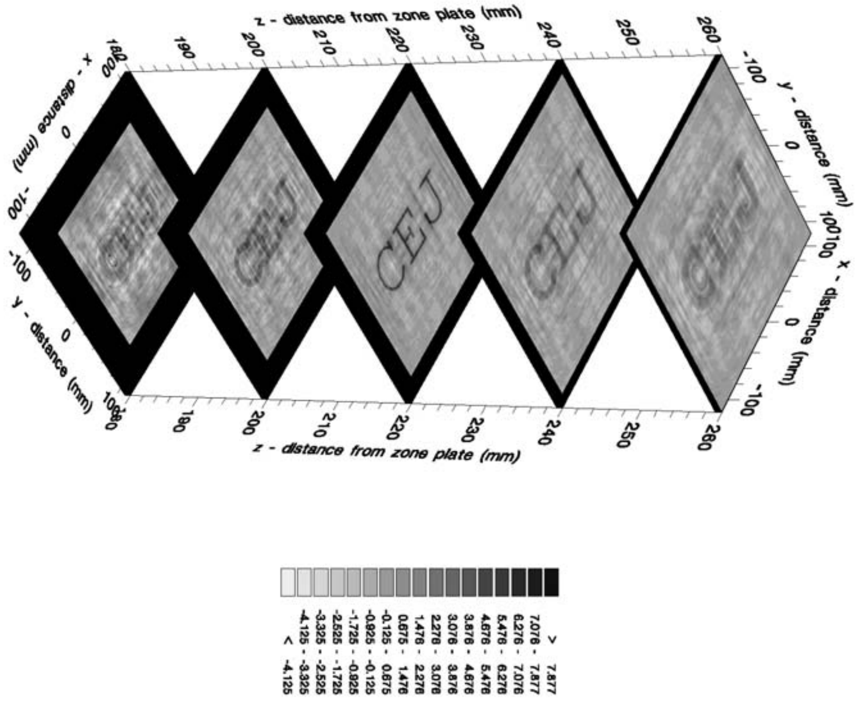
Figure 9.19: Partially normalised reconstruction of hologram $\boxed{4}$ using the deconvolution based CLEAN algorithm

| | | convolution - real part | $K2$ Wiener filter | CLEAN algorithm |
|---|---|---|---|---|
| $\boxed{1}$ | FEM | 9.8473e-01 | 9.3670e-01 | 4.4409e-16 |
| | NMSE | 4.6307e+02 | 1.6746e+01 | 1.1014e-05 |
| $\boxed{2}$ | FEM | 9.7042e-01 | 9.1376e-01 | 4.3816e-04 |
| | NMSE | 5.8324e+02 | 1.9859e+01 | 1.0008e-02 |
| $\boxed{3}$ | FEM | 9.8100e-01 | 9.0504e-01 | 4.7211e-02 |
| | NMSE | 1.0852e+03 | 5.3713e+01 | 5.0374e-02 |
| $\boxed{4}$ | FEM | 9.8895e-01 | 8.9557e-01 | 6.2602e-02 |
| | NMSE | 1.3118e+03 | 2.5649e+02 | 6.3961e-02 |

Table 9.6: Error metrics for the partially normalised image reconstruction of holograms of the text source object using convolution, deconvolution and the CLEAN algorithm



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.20: Partially normalised reconstruction of hologram $\boxed{4}$ by convolution and $K2$ Wiener filter

## 9.3 Reconstruction of Images from Holograms of the Same Object with Different Parameters

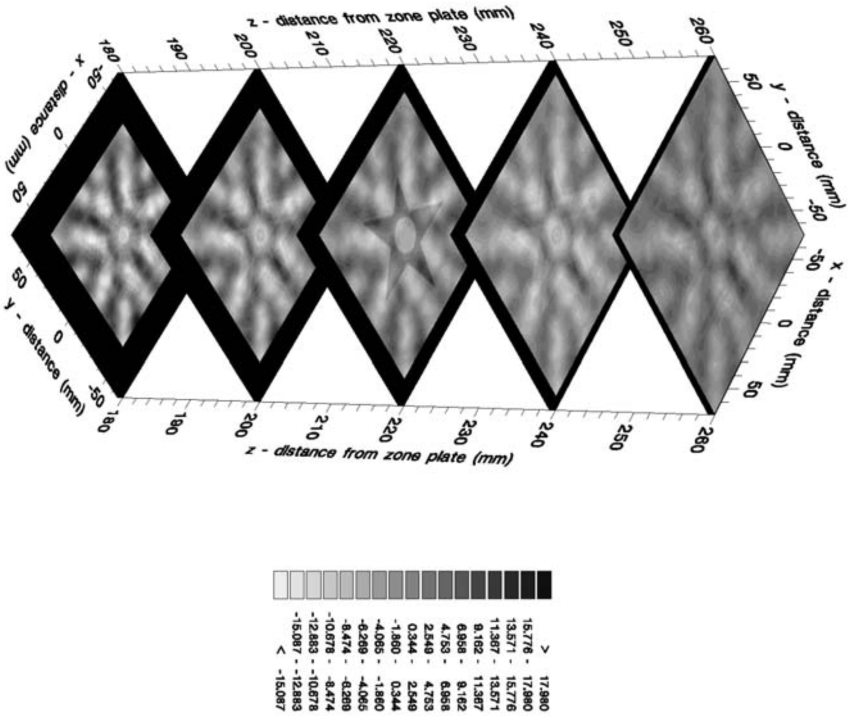
A five pointed star with a hole in the middle was used here. This was chosen as it is a more extended object than the ones used in the previous sections. Five holograms were generated of the object. The first hologram had the parameters from Table 6.1 in the other four holograms either the zone plate to detector distance or the source to zone plate distance was changed. The hologram parameters for each of the five holograms are given in Table 9.7.

| Parameter | holograms | value | Parameter | holograms | value |
|---|---|---|---|---|---|
| First zone radius | all | 1.02e-2m | Detector pixel size | all | 1.074e-3m |
| Source→zone plate | ➊➍➎ | 0.22m | Zone plate→detector | ➊➋➌ | 0.36m |
| | ➋ | 0.16m | | ➍ | 0.23m |
| | ➌ | 0.28m | | ➎ | 0.44m |
| Number of pixels | all | 360×360 | Number of zones | all | 30 |

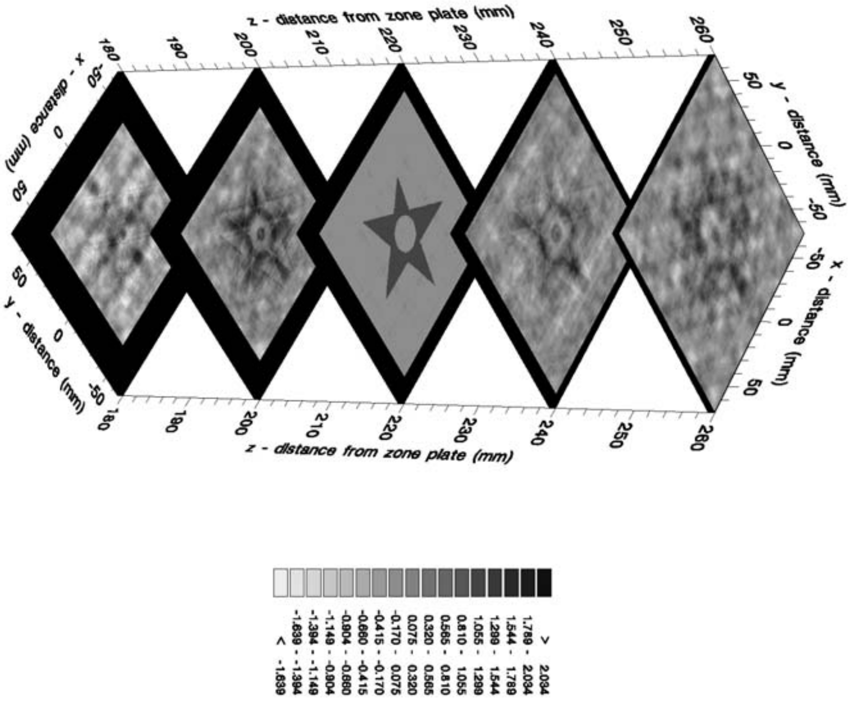
Table 9.7: Hologram parameters used for the holograms of the star object

Figures 9.21 and 9.22 show the images reconstructed from hologram ➊ and Figures 9.24 and 9.23 those from hologram ➋. Figures 9.25 and 9.26 show the images from hologram ➌, Figures 9.28 and 9.27 those from hologram ➍ and Figures 9.29 and 9.30 those from hologram ➎. The NMSE and FEM values are given in Table 9.8.

The only conclusion that can be drawn from these results is that the images look better when the object is smaller in the field of view of the detector (holograms ➌ and ➎) than when it is larger (holograms ➋ and ➍). Holograms ➌ and ➎ have lower error metric values for the images reconstructed using the CLEAN algorithm. However, there is no obvious trend in the results for convolution or deconvolution.



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.21: Partially normalised reconstruction of hologram ❶ by convolution and $K2$ Wiener filter

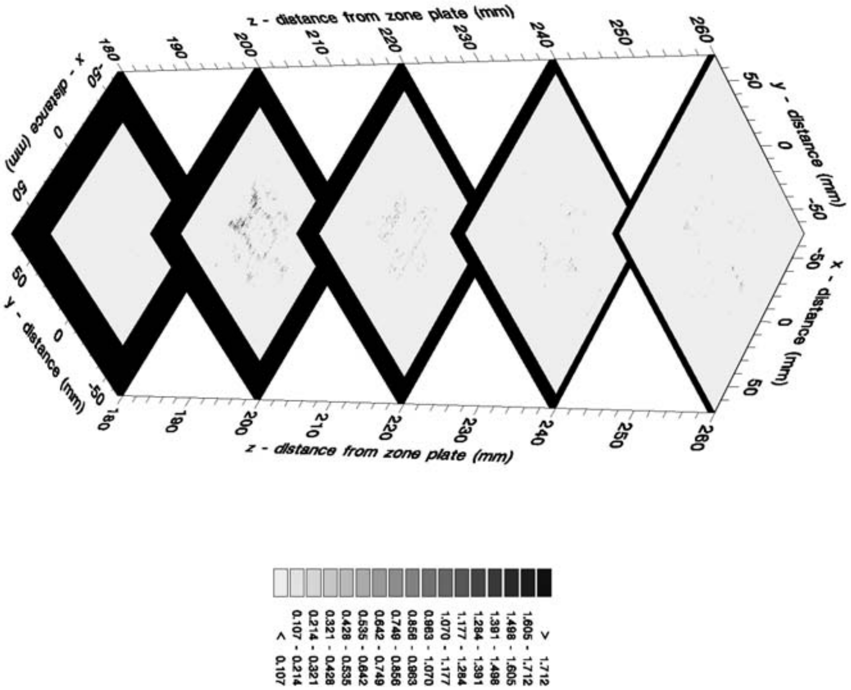


Figure 9.22: Partially normalised reconstruction of hologram ❶ using the deconvolution based CLEAN algorithm

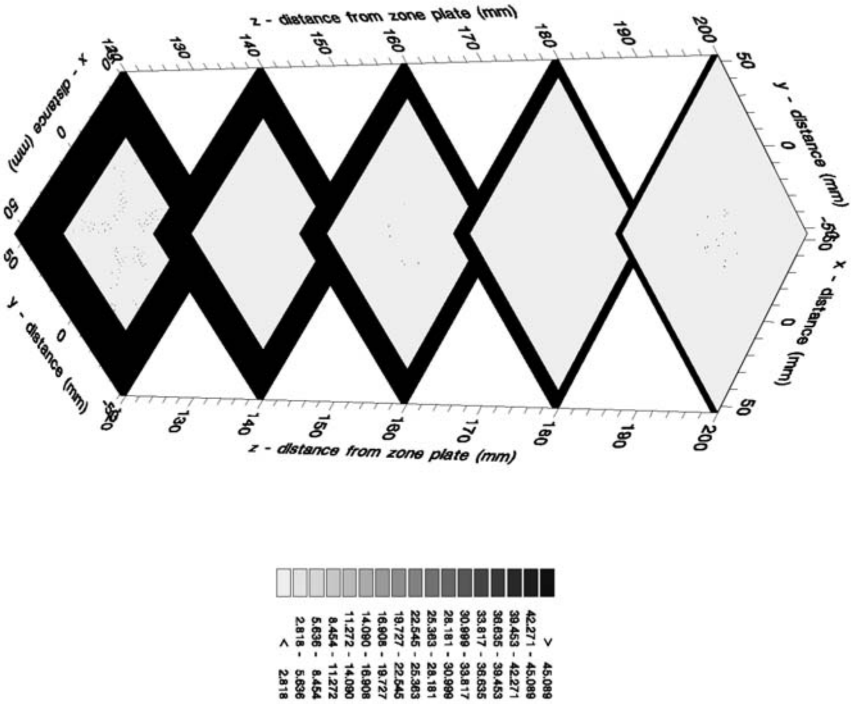
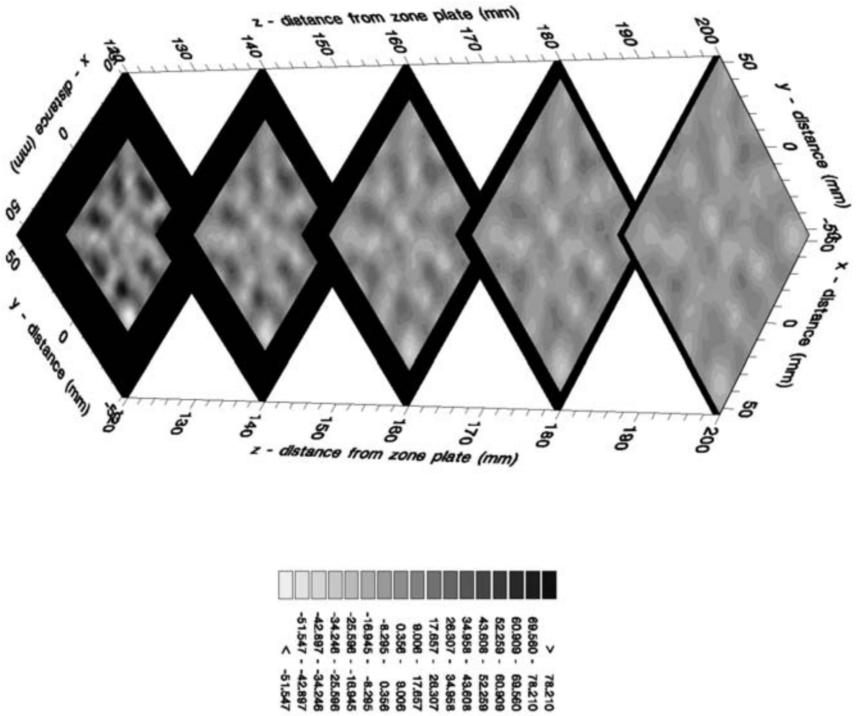
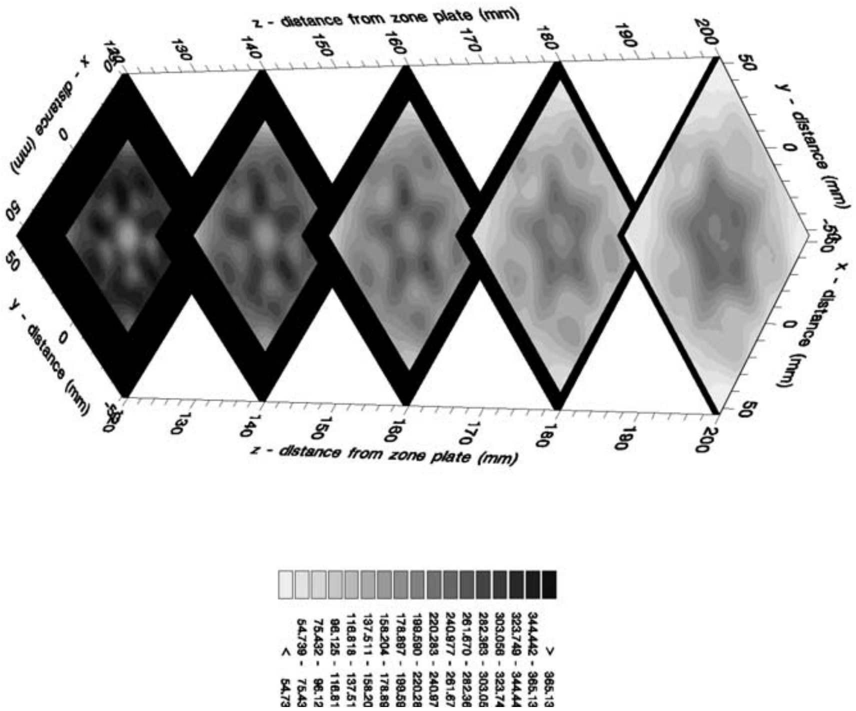


Figure 9.23: Partially normalised reconstruction of hologram ❷ using the deconvolution based CLEAN algorithm

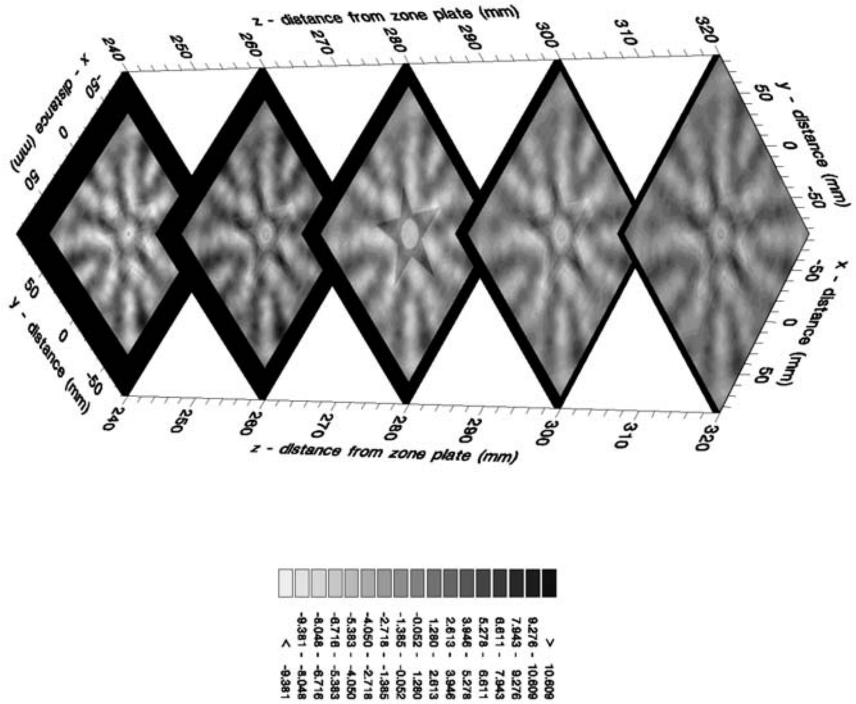


(a) convolution - real part

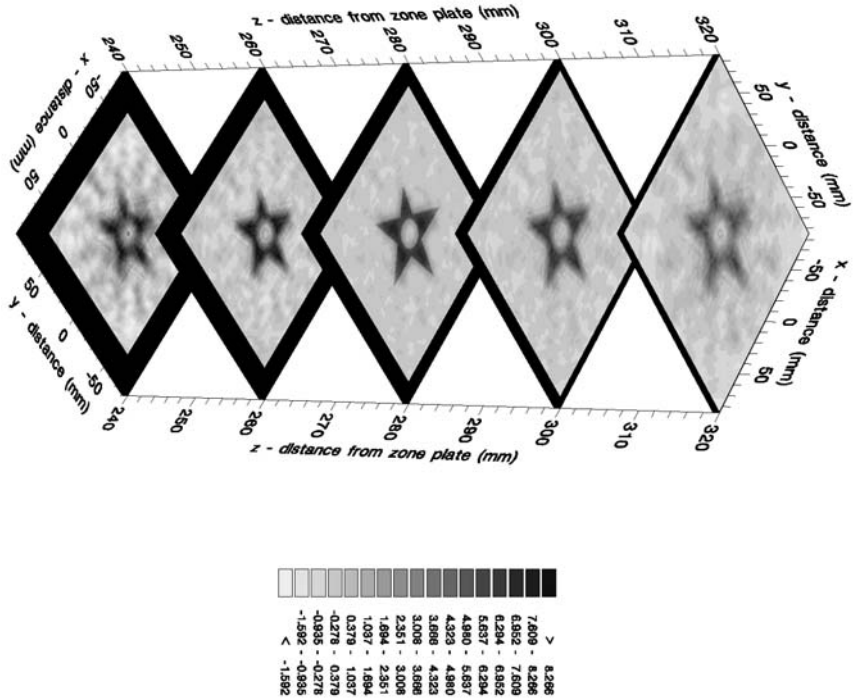


(b) $K2$ Wiener filter

Figure 9.24: Partially normalised reconstruction of hologram ❷ by convolution and $K2$ Wiener filter



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.25: Partially normalised reconstruction of hologram ❸ by convolution and $K2$ Wiener filter

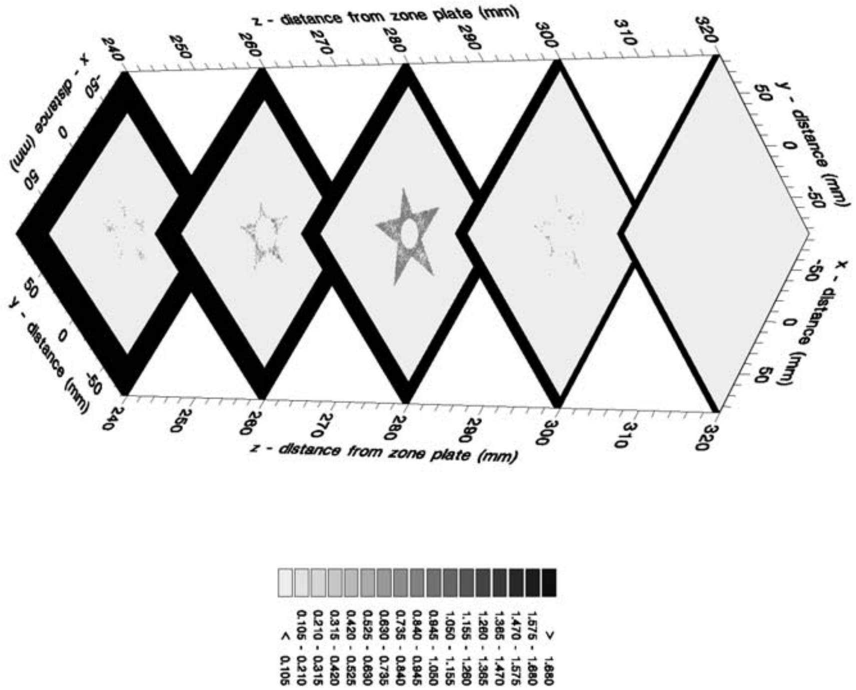


Figure 9.26: Partially normalised reconstruction of hologram ❸ using the deconvolution based CLEAN algorithm

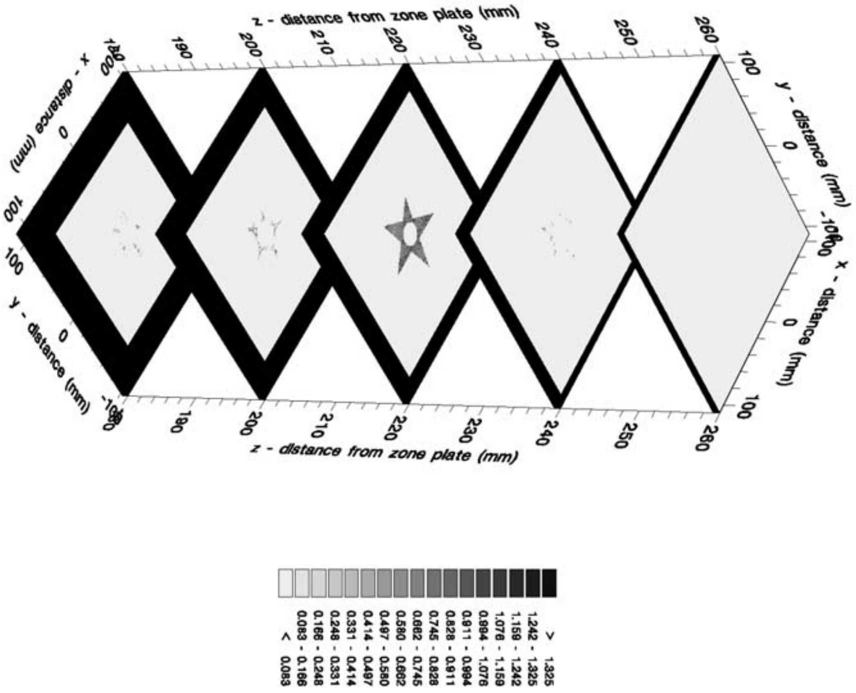
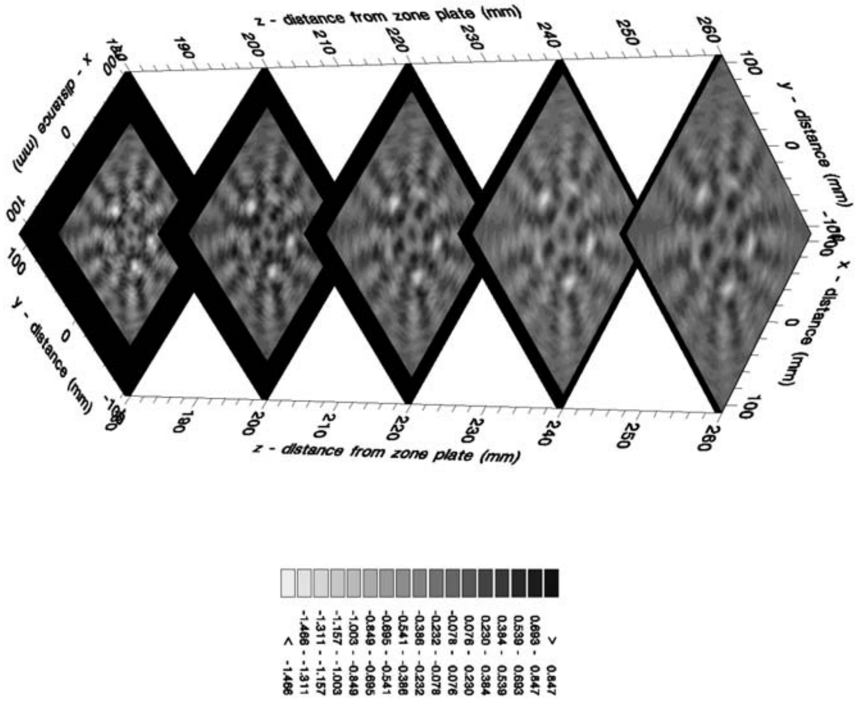
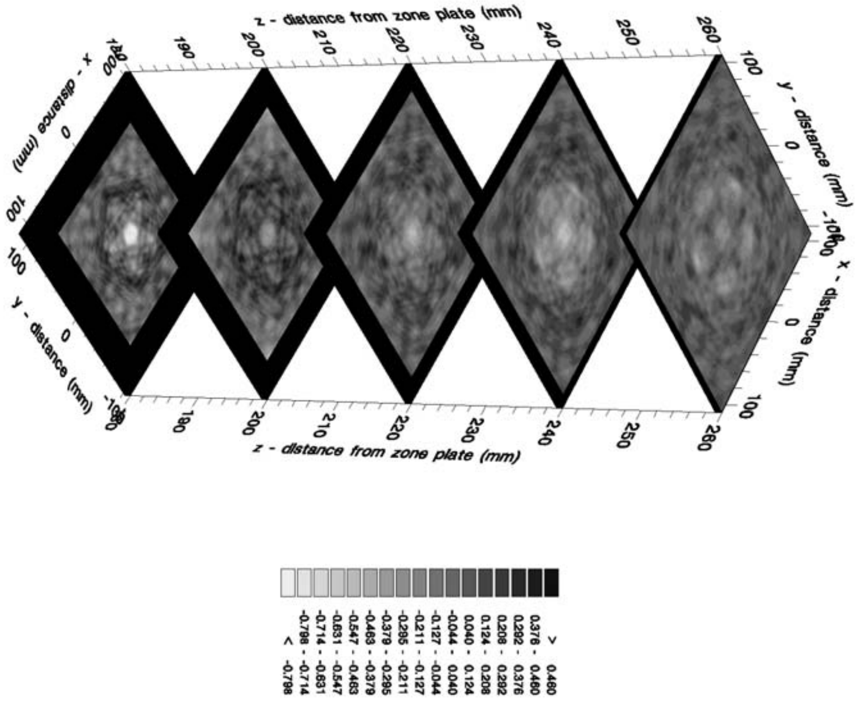


Figure 9.27: Partially normalised reconstruction of hologram ❹ using the deconvolution based CLEAN algorithm

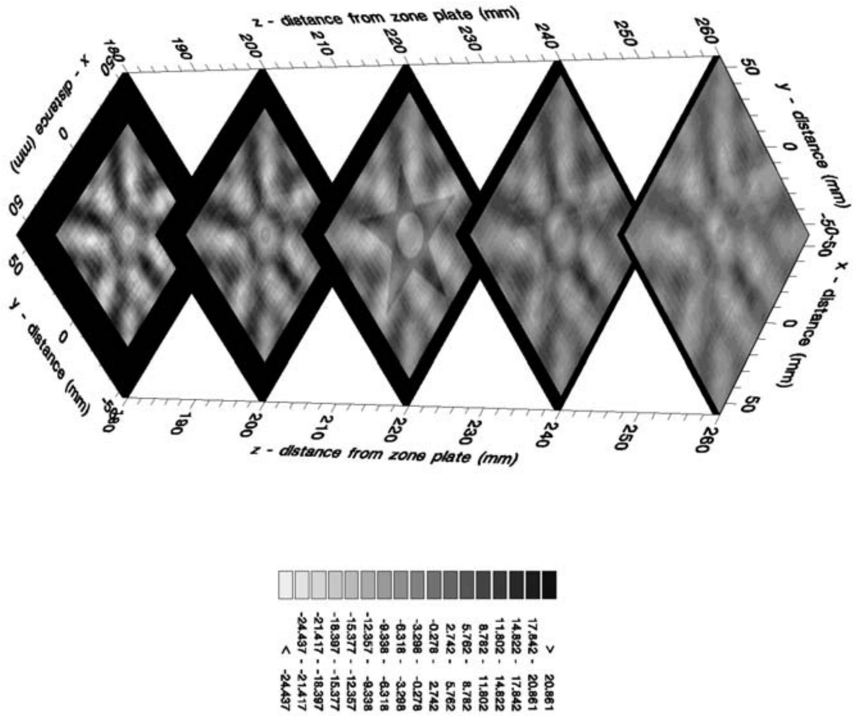


(a) convolution - real part

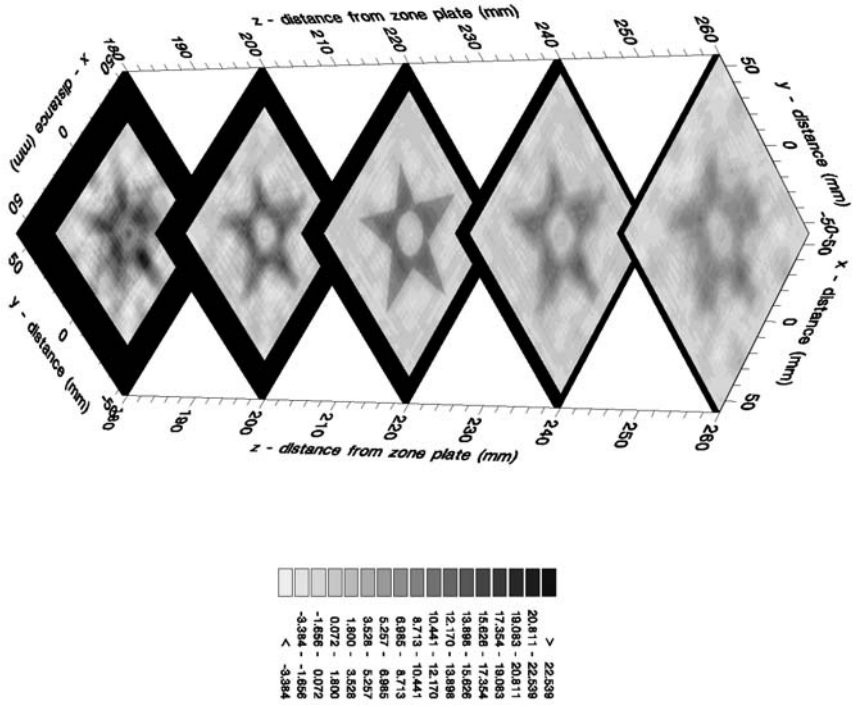


(b) $K2$ Wiener filter

Figure 9.28: Partially normalised reconstruction of hologram ❹ by convolution and $K2$ Wiener filter



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.29: Partially normalised reconstruction of hologram ❺ by convolution and $K2$ Wiener filter

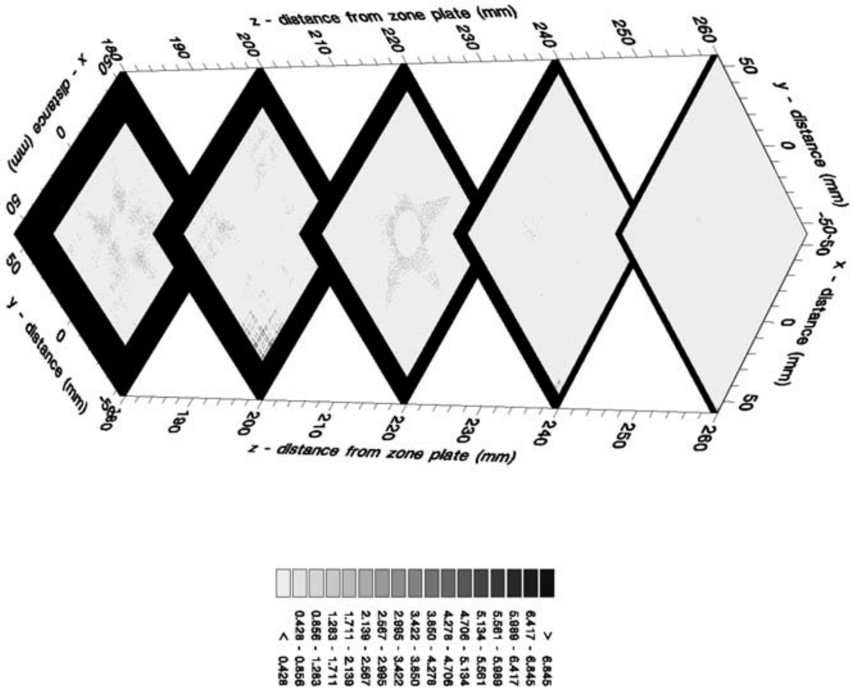


Figure 9.30: Partially normalised reconstruction of hologram ❺ using the deconvolution based CLEAN algorithm

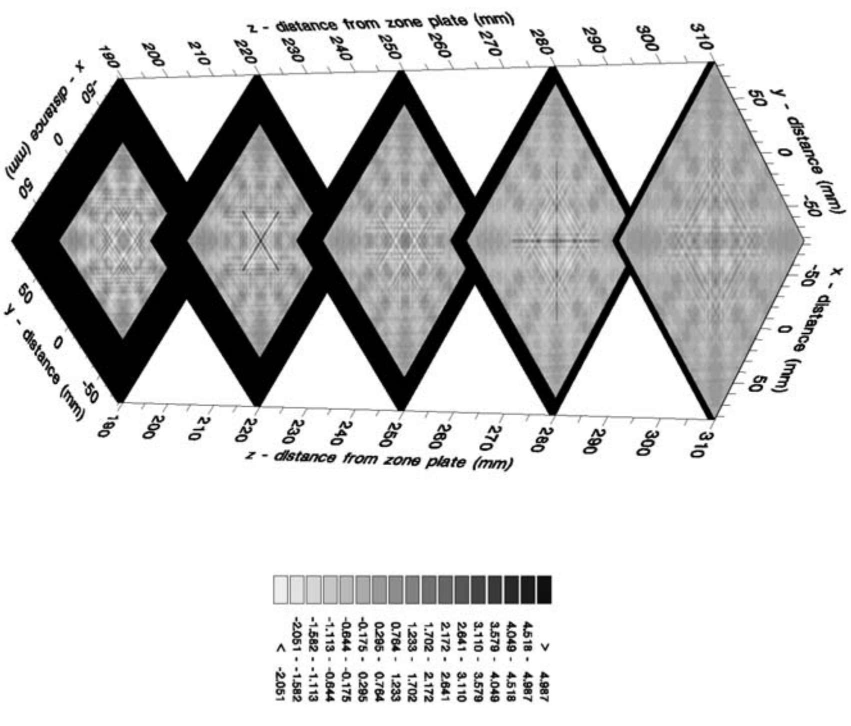
| | | convolution - real part | $K2$ Wiener filter | CLEAN algorithm |
|---|---|---|---|---|
| ❶ | FEM | 9.9169e-01 | 9.5303e-01 | 9.9695e-01 |
| | NMSE | 2.6142e+03 | 2.6967e+01 | 7.7337e+00 |
| ❷ | FEM | 9.4255e-01 | 9.4457e-01 | 9.8985e-01 |
| | NMSE | 1.5465e+04 | 1.8333e+06 | 6.3258e+01 |
| ❸ | FEM | 9.8946e-01 | 8.5427e-01 | 3.6092e-01 |
| | NMSE | 1.2280e+03 | 2.5173e+02 | 3.7700e-01 |
| ❹ | FEM | 9.9185e-01 | 9.8631e-01 | 2.0121e-01 |
| | NMSE | 1.6990e+01 | 8.0404e+00 | 2.0157e-01 |
| ❺ | FEM | 9.9049e-01 | 8.0550e-01 | 9.7744e-01 |
| | NMSE | 4.7154e+03 | 9.0645e+02 | 8.0226e+00 |

Table 9.8: Error metrics for the partially normalised image reconstruction of holograms of the star shaped source using convolution, deconvolution and the CLEAN algorithm

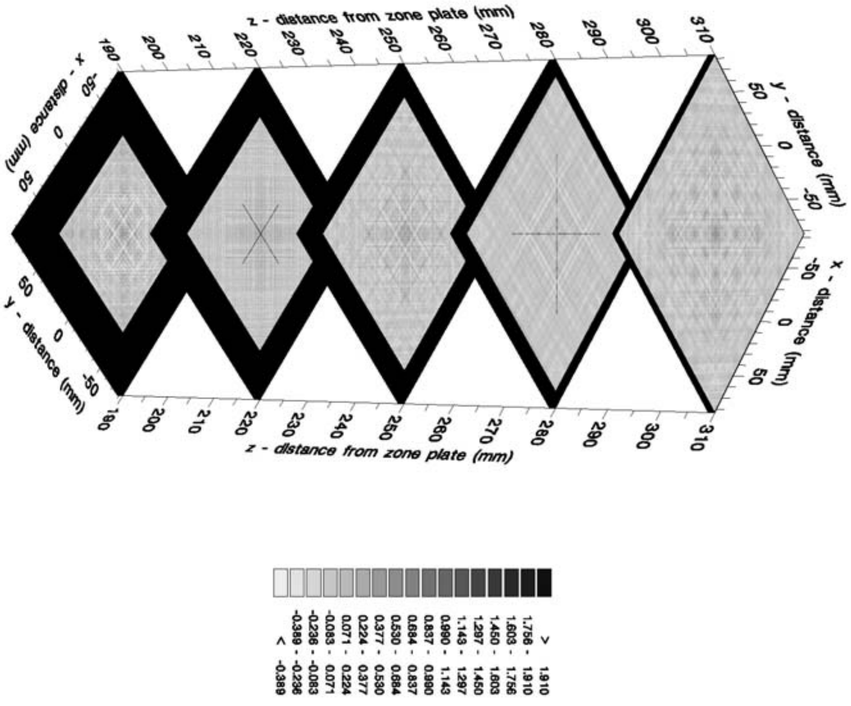
## 9.4 Reconstruction of Two Cross Sources in Different Planes

The first one of the two cross sources used here is the same as the one in Section 9.1.1 and has the same parameters, it is positioned at 0.22m behind the zone plate. The other cross is a diagonal one and is positioned 0.28m from the zone plate. Image planes were reconstructed at 0.19m, 0.22m, 0.28m and 0.31m. The reconstructed images in all five planes can be seen in Figures 9.31 and 9.32 and the corresponding error metric values are in Table 9.9.

The image from the CLEAN algorithm showed the two crosses mainly in the correct planes. Very little was reconstructed in the surrounding planes. Convolution and deconvolution gave almost as much in the surrounding planes as in the planes which actually contained the sources.



(a) convolution - real part



(b) $K2$ Wiener filter

Figure 9.31: Partially normalised reconstruction for hologram of two cross sources by convolution and $K2$ Wiener filter

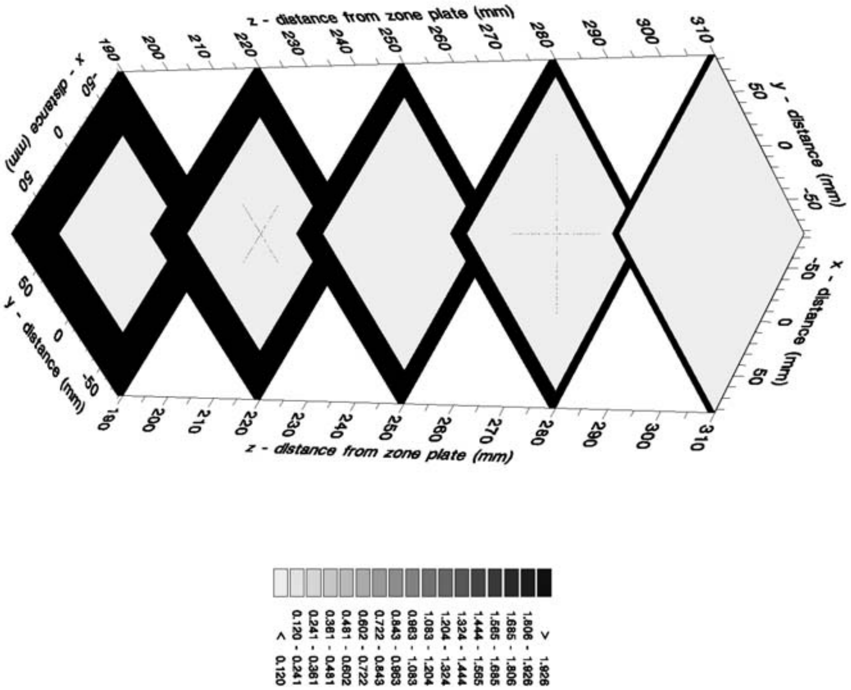


Figure 9.32: Partially normalised reconstruction for hologram of two cross sources using the CLEAN algorithm

| | convolution | deconvolution | CLEAN (deconvolution) |
|---|---|---|---|
| FEM | 9.8034e-01 | 9.0700e-01 | 3.8379e-01 |
| NMSE | 2.3815e+02 | 1.5315e+01 | 3.8505e-01 |

Table 9.9: Error metrics for the partially normalised image reconstruction of a hologram of two cross sources

## 9.5 Summary

The three reconstruction techniques have been tested for a variety of computer generated holograms. Thee reconstructed images had lower error metric values when the overlap factor (described in Section 6.4) was not included. These partially normalised reconstruction methods were then used to reconstruct the rest of the images in the Chapter. The CLEAN algorithm was found to work better when it was based on deconvolution than when it was based on convolution so this method was used.

The effect of different types of error, increasing object size and different hologram parameters were investigated. Although a more thorough investigation would be required to draw definite conclusions, these results gave some indications of the sources of error in the reconstructed images. These results suggest that the finite resolution of the gamma camera is a major contributing factor to errors in the reconstructed image for these reconstruction methods. They also indicate that the errors in the image are generally larger for objects which are larger in the field of view of the detector.

# Chapter 10

# RECONSTRUCTION OF REAL HOLOGRAMS

The zone plate holograms which will be used in this Chapter were recorded by Perks as part of his PhD project [65]. Perks recorded holograms using a Toshiba GCA-7200A gamma camera at St Thomas' hospital, London. Since then, Rew [70] has recorded many holograms using a GE Maxicamera 535 in the Physics and Astronomy department of the University of Birmingham. The holograms recorded using the GE camera are, however, of poorer quality than those recorded on the Toshiba camera. The spatial resolution was given in Table 4.1 for all these cameras. The GE camera has a spatial resolution of 5mm and the Toshiba camera has a spatial resolution of 3.8mm. Although the spatial resolution of the GE camera is only slightly worse than that of the Toshiba camera, the GE camera needs to re-adjusted and this has made the images worse than could be expected.

Takhar [84] and George[33] are now working on their PhDs on this project and have started taking images using an ADAC Forte gamma camera which is also in the Physics and Astronomy department. The images from this camera are comparable to those from the Toshiba camera. It will be left to them to test these reconstruction techniques further on real holograms taken using the ADAC camera. Suggestions for further work they can do to test the reconstruction techniques are given in Chapter 11.

## 10.1 Cross Source Placed Parallel to the Zone Plate

Perks recorded two holograms of a cross source made of four capillary tubes filled with technicium 99m. The bungs holding the Technicium in the tubes were all placed together in the centre of the cross so there was a region of no activity here. The two holograms had the same source to zone plate distance (0.22m) but different zone plate to detector distances (0.23m and 0.44m). The other parameters can be seen in Tables 10.1 and 10.3.

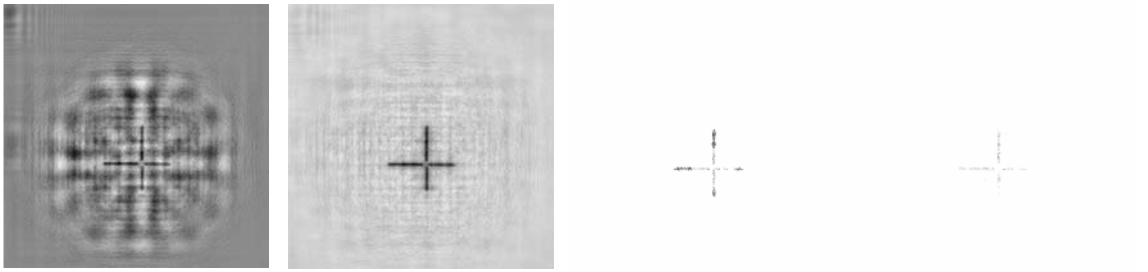
Note that the pixel size, first zone radius and number of zones in the zone plate are all the same as those used for the computer generated holograms.

Figure 10.1: Cross source recorded using the collimator (taken from Perks' thesis)

Figure 10.1 shows an image of the cross source taken using the gamma camera with the collimator still attached. This picture was taken from Perks' thesis. The region of no activity in the centre of the cross can not be seen here.

| | | | |
|---|---|---|---|
| First zone radius ($r_1$) | 1.02e-02m | Source → zone plate distance ($z$) | 0.22m |
| Detector pixel size ($\Delta$) | 1.074e-03m | Zone plate → detector distance ($b$) | 0.44m |
| Number of pixels ($n_x \times n_y$) | 512 × 512 | Number of zones ($n = \sqrt{\frac{R}{r_1}}$) | 30 |

Table 10.1: Parameters for the first hologram of the cross source



(a) convolution - real part - 3.7730e+12$\leq$pixel value$\leq$4.5012e+12

(b) $K2$ Wiener filter -1.0095e+12$\leq$pixel value $\leq$4.5933e+12

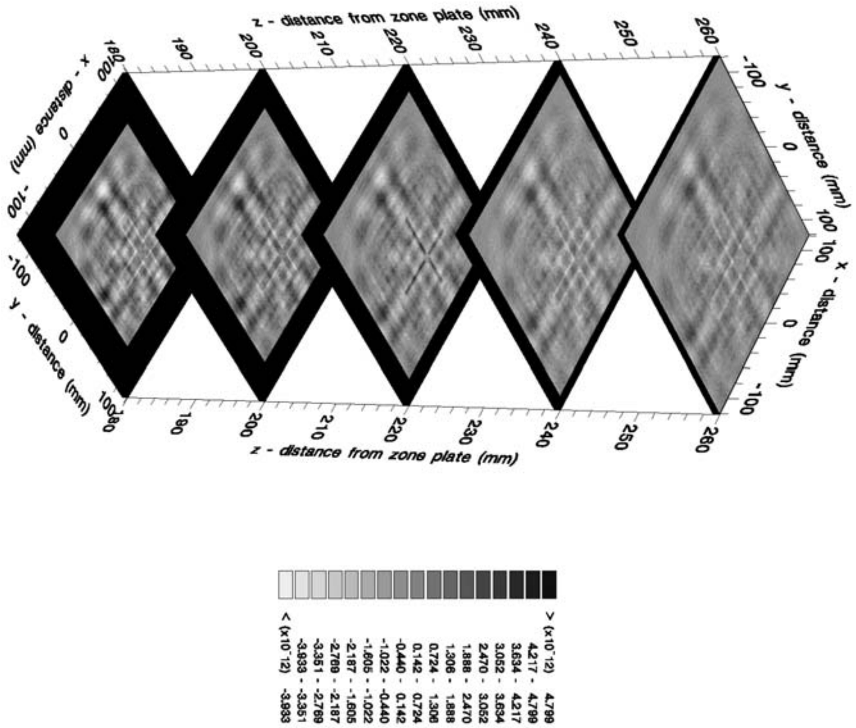
(c) CLEAN algorithm (convolution)0$\leq$pixel value $\leq$1.6626e+12

(d) CLEAN algorithm (deconvolution)0$\leq$pixel value $\leq$8.5754e+11

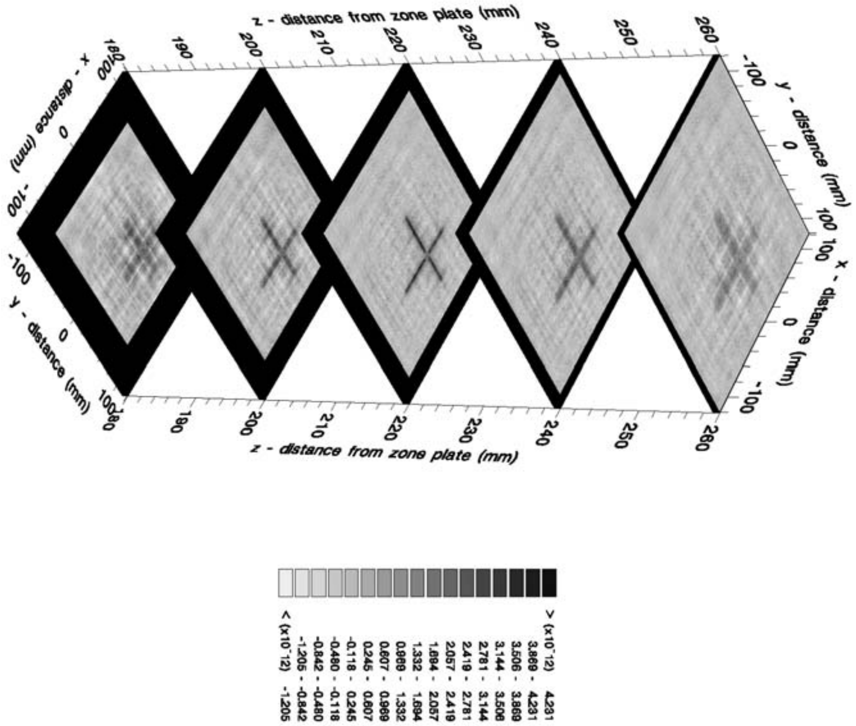
Figure 10.2: Partially normalised reconstruction for the first real hologram of the cross source in the object plane: white = minimum, black = maximum pixel value

Figures 10.2, 10.3 and 10.3 show the partially normalised reconstructions of the image of the cross for the first hologram. This technique was shown to be better than the fully normalised reconstruction in Chapter 9. The deconvolution based CLEAN algorithm gave better results than the convolution based CLEAN algorithm in Chapter 9. This method was then used for the remaining computer generated holograms. It takes a long time to reconstruct images using the CLEAN algorithm so it did not seem sensible to try both techniques for all the computer generated holograms. However, as there are only a limited number of real holograms available, both techniques will be tried here.

Figure 10.2 shows the image in the object plane and Figures 10.3 and 10.4 show the image in all five planes. Table 10.2 gives the corresponding error metric values. As the object is not exactly know the error values were calculated by comparing the actual hologram to the theoretical hologram which would have been produced by the reconstructed source. The deconvolution reconstruction gives the best FEM error metric value and the convolution based CLEAN reconstruction gives the best NMSE error metric value in this case. The convolution based CLEAN reconstruction only reconstructed points in the images plane but the points were incremented by a large value in each iteration to make the reconstruction be completed in a reasonable amount of time. The points were incremented by the maximum value of the image reconstructed by convolution or deconvolution divided by ten. If a smaller value had been chosen, the reconstruction would have taken longer but the image may have been better.

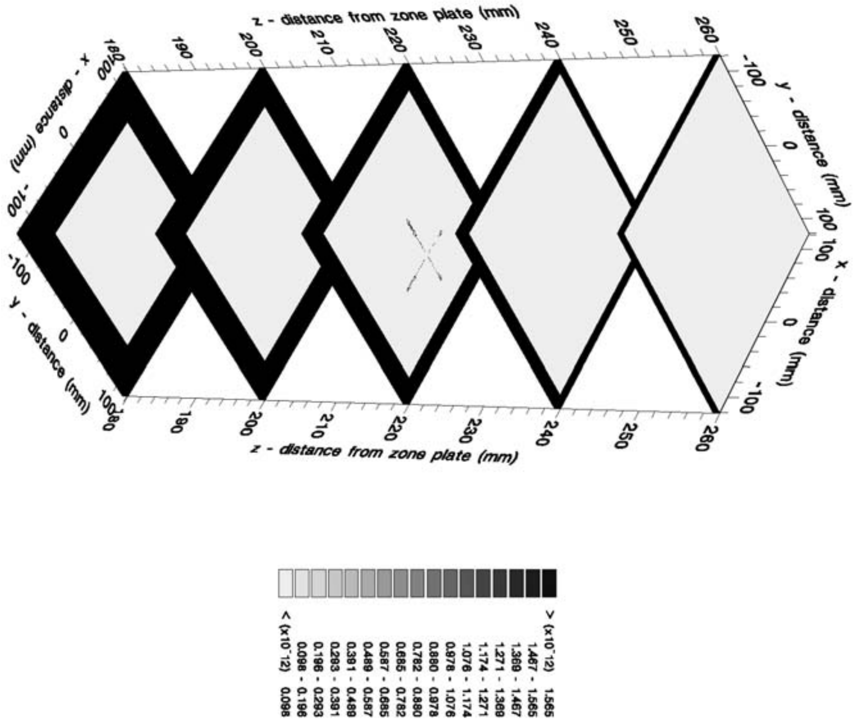


(a) convolution - real part

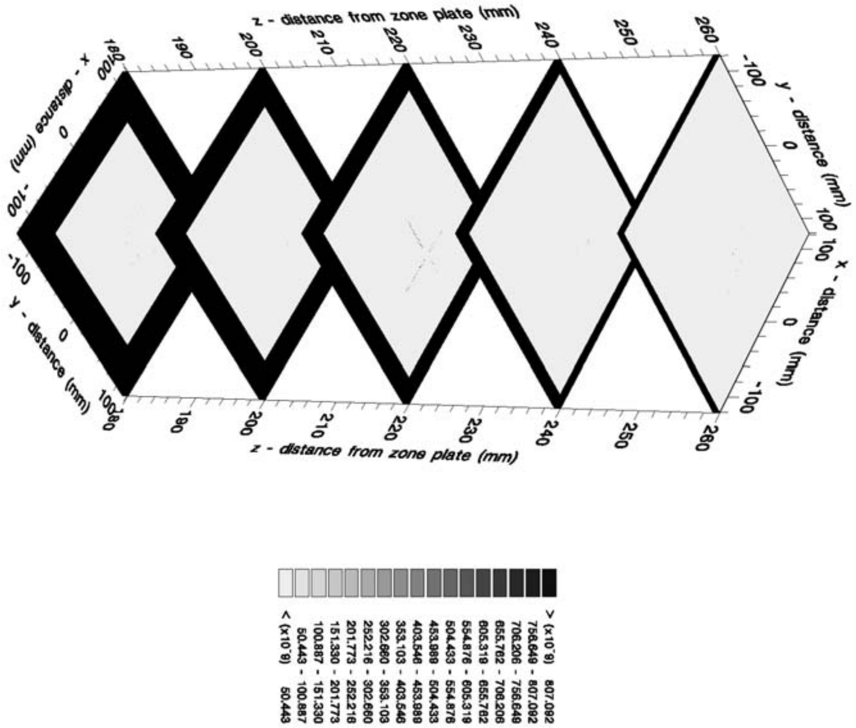


(b) $K2$ Wiener filter

Figure 10.3: Reconstruction by convolution and $K2$ Wiener filter of the first real hologram of the cross source in all five planes



(a) algorithm based on convolution



(b) algorithm based on deconvolution

Figure 10.4: Reconstruction using the CLEAN algorithm of the first real hologram of the cross source in all five planes

| | convolution | deconvolution | CLEAN (convolution) | CLEAN (deconvolution) |
|---|---|---|---|---|
| FEM | 3.2774e-01 | 1.9731e-02 | 3.033e-02 | 5.172e-01 |
| NMSE | 1.0187e+03 | 1.3105e+03 | 2.654e-02 | 8.7956e-01 |

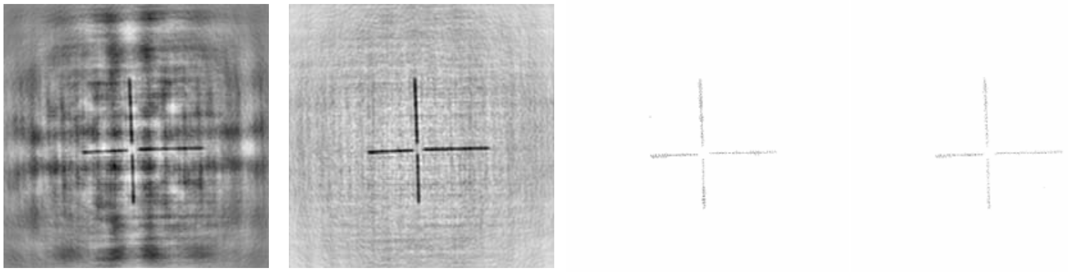
Table 10.2: Error metrics for the partially normalised image reconstruction of the first real hologram of the cross source

The second hologram of this source has a larger zone plate to detector distance (see Table 10.3. This means that the field of view of the camera through the zone plate is smaller for any given distance. This is why the cross appears to be larger in the image planes. Figure 10.5 shows the image in the object plane and Figures 10.6 and 10.7 show the image in all five planes. Table 10.4 gives the corresponding error metric values.

The convolution based CLEAN method again only reconstructed the image in the correct plane whereas the deconvolution based CLEAN method gave a few stray points in other planes. For this hologram, both CLEAN methods gave almost equally low error values which were better than either the straight convolution or deconvolution.

| | | | |
|---|---|---|---|
| First zone radius ($r_1$) | 1.02e-02m | Source → zone plate distance ($z$) | 0.22m |
| Detector pixel size ($\Delta$) | 1.074e-03m | Zone plate → detector distance ($b$) | 0.23m |
| Number of pixels ($n_x \times n_y$) | 512 × 512 | Number of zones ($n = \sqrt{\frac{R}{r_1}}$) | 30 |

Table 10.3: Parameters for the second hologram of the cross source



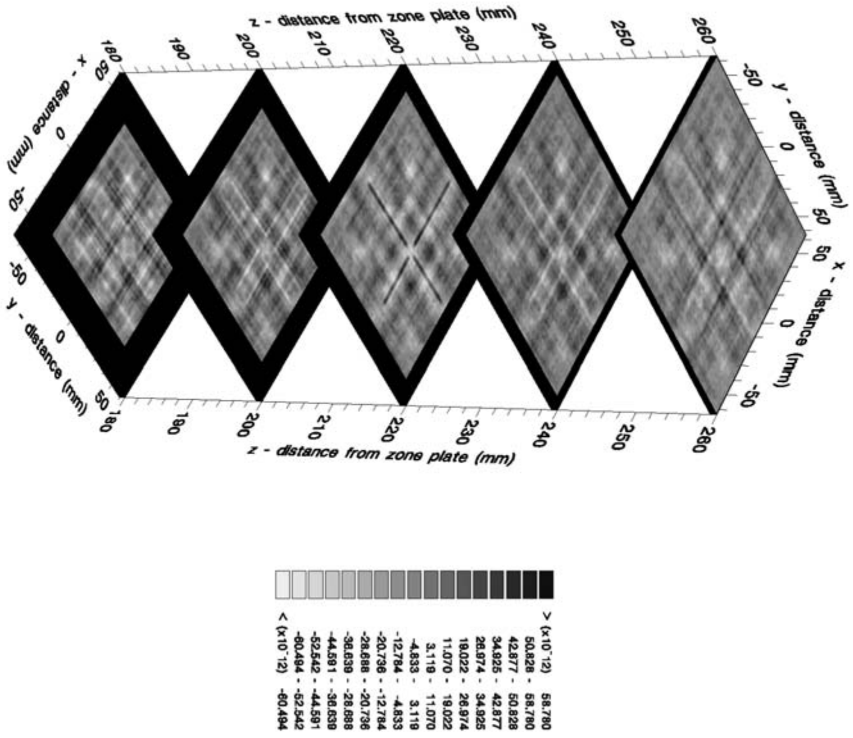
(a) convolution - real part - 5.7391e+13≤pixel value≤6.4241e+13

(b) $K2$ Wiener filter -2.5615e+13≤pixel value ≤7.0304e+13

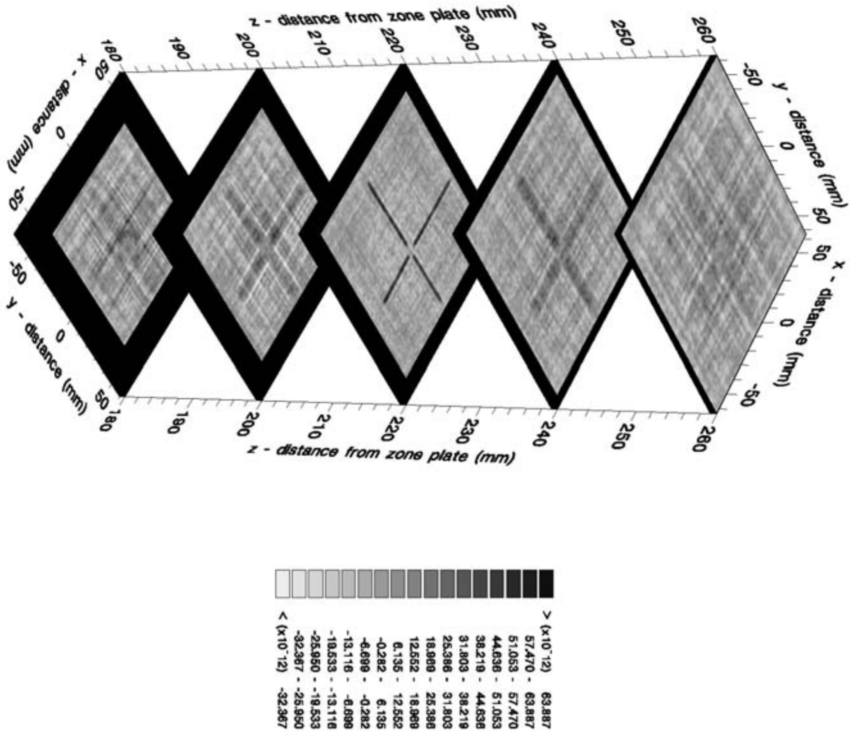
(c) CLEAN algorithm (convolution)≤pixel value 0≤1.3727e+13

(d) CLEAN algorithm (deconvolution)≤pixel value 0≤8.2294e+12

Figure 10.5: Partially normalised reconstruction for the second real hologram of the cross source in the object plane: white = minimum, black = maximum pixel value

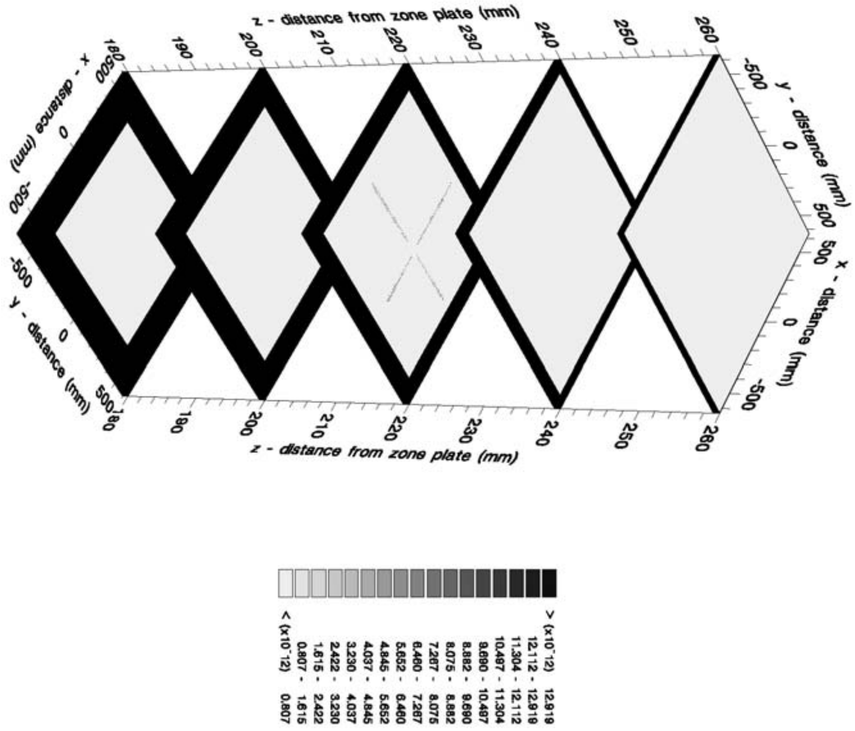


(a) convolution - real part

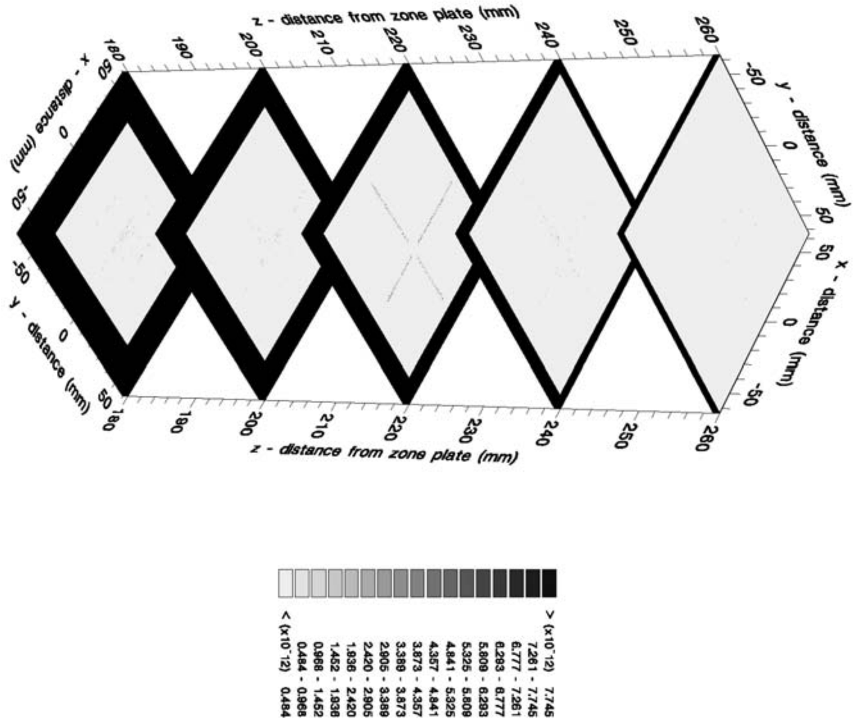


(b) $K2$ Wiener filter

Figure 10.6: Reconstruction by convolution and $K2$ Wiener filter for the second real hologram of the cross source in all five planes



(a) algorithm based on convolution



(b) algorithm based on deconvolution

Figure 10.7: Reconstruction using the CLEAN algorithm for the second real hologram of the cross source in all five planes

| | convolution | deconvolution | CLEAN (convolution) | CLEAN (deconvolution) |
|---|---|---|---|---|
| FEM | 1.2855e-01 | 6.1953e-02 | 1.3809e-02 | 1.3007e-02 |
| NMSE | 2.1426e+04 | 5.9493e+03 | 3.7175e-01 | 4.7110e-01 |

Table 10.4: Error metrics for the partially normalised image reconstruction of the second real hologram of the cross source

## 10.2 Three-Dimensional Heart Phantom

Figure 10.8 shows a diagram of the heart phantom. This, again, is taken from Perks' thesis. The phantom is a cylinder which is joined at one end. Its walls are hollow and were filled with Technicium 99m via a hole in the joined end which was then plugged.

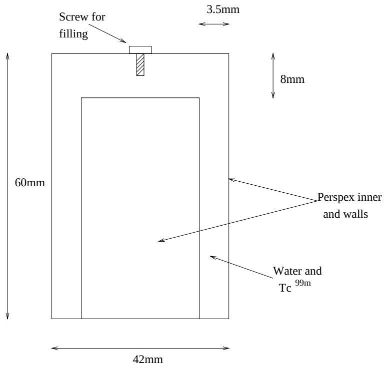


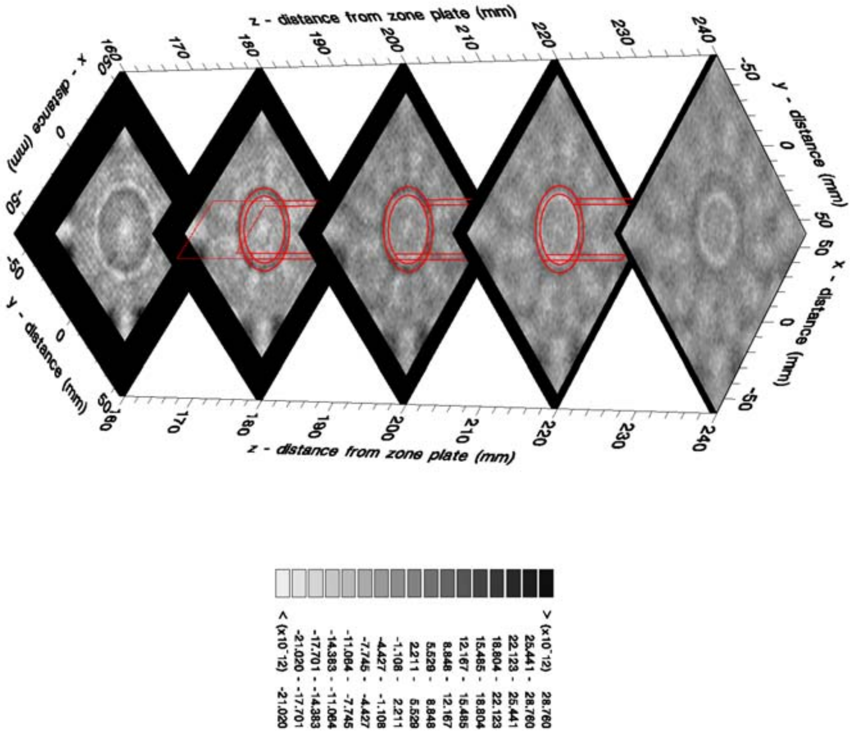
Figure 10.8: Heart phantom (taken from Perks' thesis)

Holograms of this phantom were recorded from two directions, an end view and a side view. The parameters are given in Table 10.5.

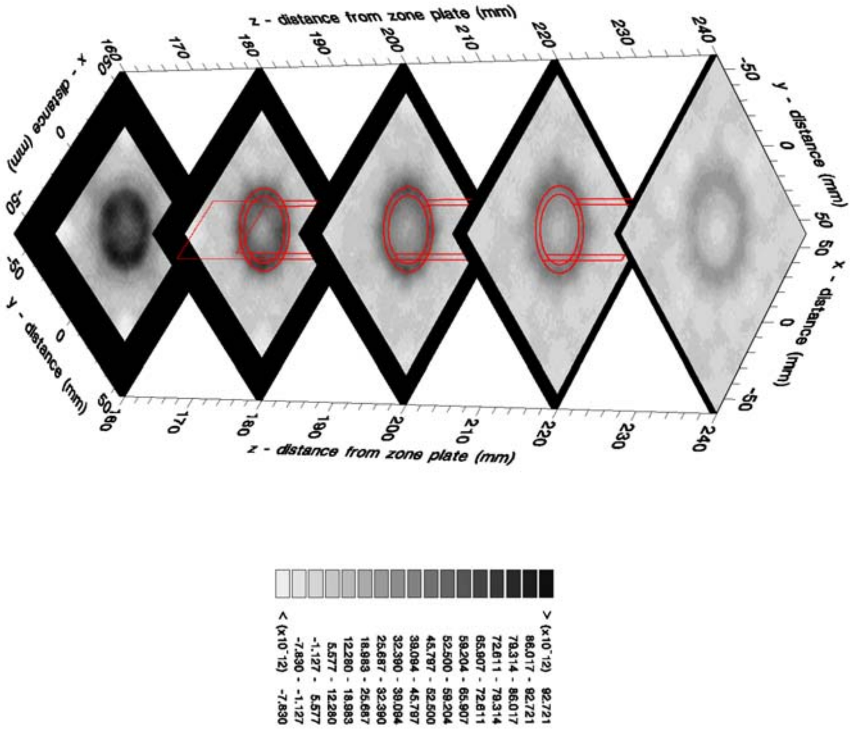
| | | | |
|---|---|---|---|
| First zone radius ($r_1$) | 1.02e-02m | Source → zone plate distance ($z$) | 0.17m→0.23m |
| Detector pixel size ($\Delta$) | 1.074e-03m | Zone plate → detector distance ($b$) | 0.39m |
| Number of pixels ($n_x \times n_y$) | 512 × 512 | Number of zones ($n = \sqrt{\frac{R}{r_1}}$) | 30 |

Table 10.5: Parameters for the holograms of the heart phantom (both end and side views)

Figures 10.9 and 10.10 show the reconstructed images from the end view hologram in planes at 0.16m 0.18m 0.20m 0.22m and 0.24m. Both CLEAN reconstructions gave points in the planes at 0.18m and 0.22m but only the occasional point in the other planes. As the phantom was present in these planes, this is good, but it should have also appeared in the plane at 0.20m. As usual, the straight convolution and deconvolution methods gave an image in all the planes. Lines have been drawn on the reconstructed images to indicate where the heart phantom should be. The error metric values are given in Table 10.6. As with the first hologram of the cross source, deconvolution gave the best FEM value but the CLEAN algorithms gave the best NMSE values.

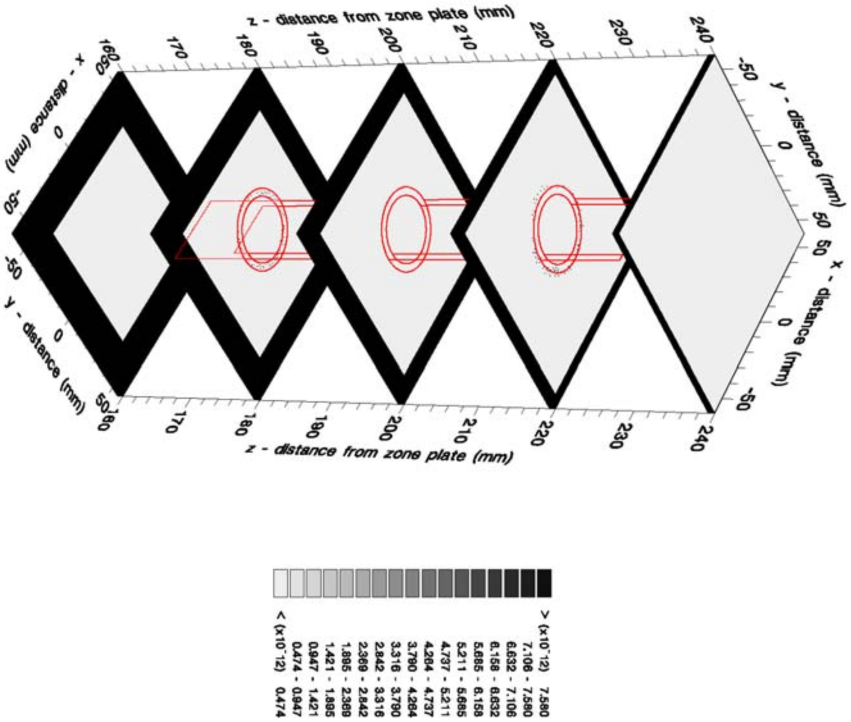


(a) convolution - real part

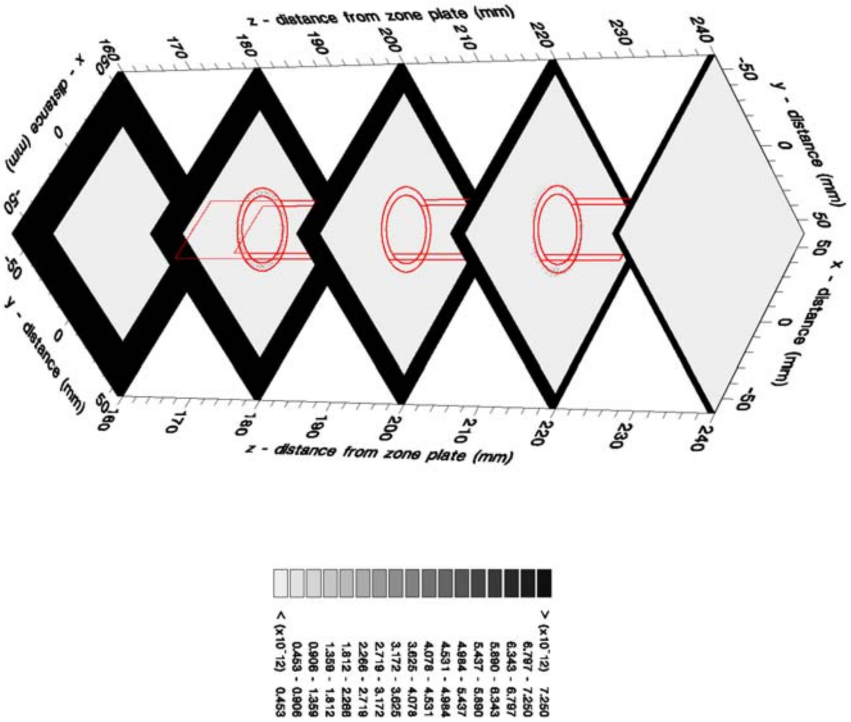


(b) $K2$ Wiener filter

Figure 10.9: Reconstruction by convolution and $K2$ Wiener filter of the end view of the heart phantom



(a) algorithm based on convolution



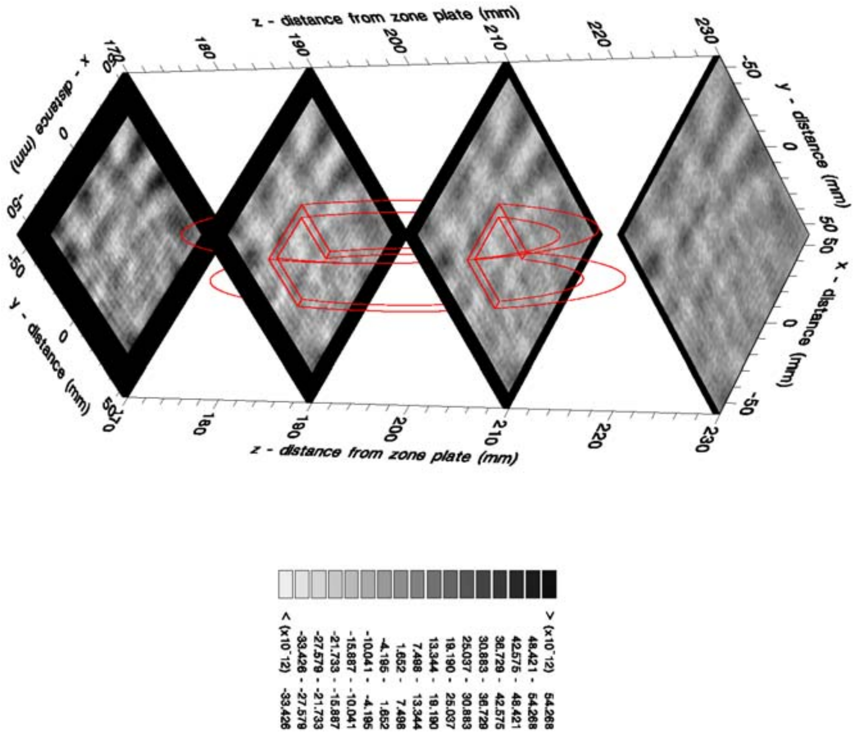
(b) algorithm based on deconvolution

Figure 10.10: Reconstruction using the CLEAN algorithm of the end view of the heart phantom

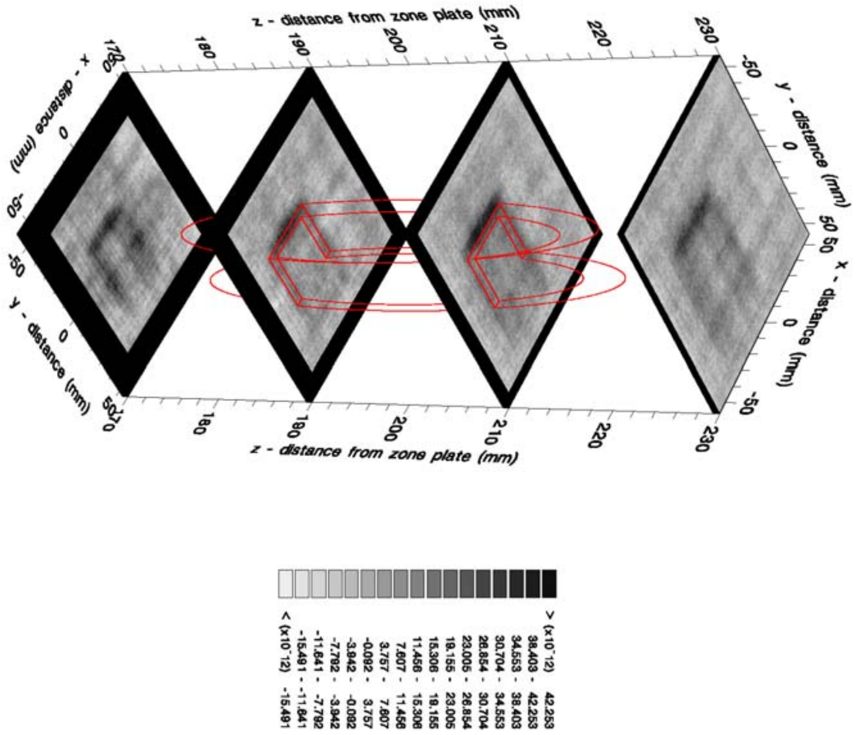
| | convolution | deconvolution | CLEAN (convolution) | CLEAN (deconvolution) |
|---|---|---|---|---|
| FEM | 3.0718e-01 | 1.1603e-02 | 2.2656e-02 | 1.6489e-02 |
| NMSE | 1.1016e+01 | 1.9114e+05 | 5.4889e-01 | 4.3984e-01 |

Table 10.6: Error metrics for the partially normalised image reconstruction of the end view of the heart phantom

Figures 10.11 and 10.12 show the reconstructed images from the side view hologram in planes at 0.17m 0.19m 0.21m and 0.23m. Again, lines have been drawn on the reconstructed image to indicate where the heart phantom should be. Both CLEAN algorithms give the image in just the plane at 0.17m but this is not where the heart phantom is. The FEM values are worse for the CLEAN algorithms than for the straight deconvolution but the NMSE values are, surprisingly, better.

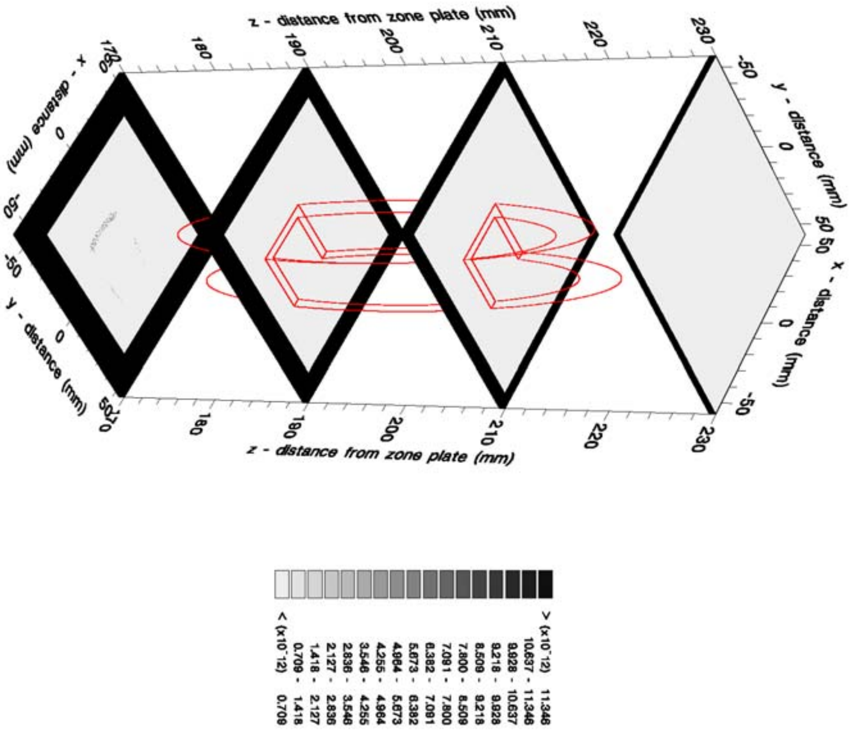


(a) convolution - real part

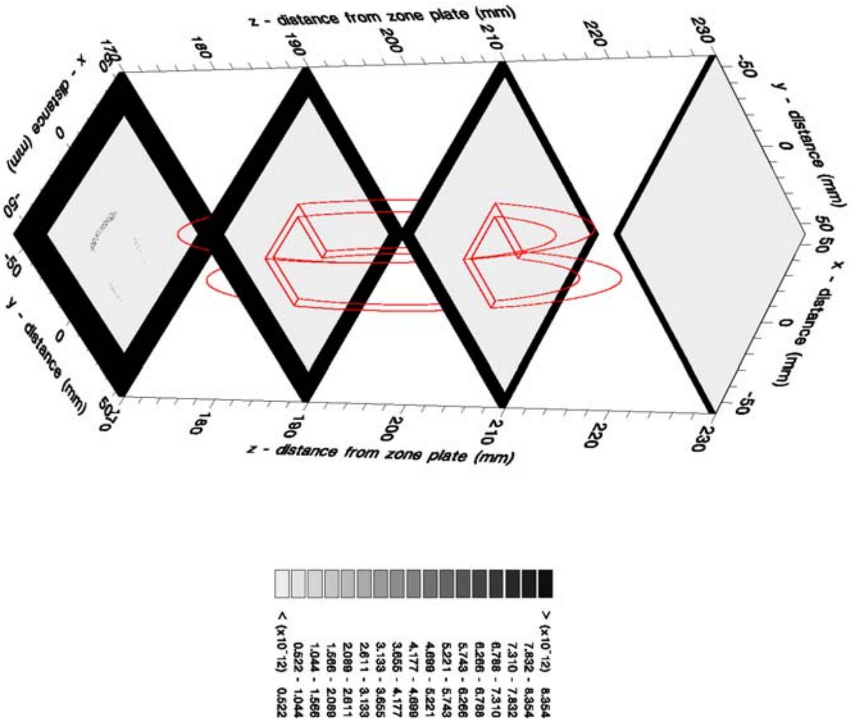


(b) $K2$ Wiener filter

Figure 10.11: Reconstruction by convolution and $K2$ Wiener filter of the side view of the heart phantom



(a) algorithm based on convolution



(b) algorithm based on deconvolution

Figure 10.12: Reconstruction using the CLEAN algorithm of the side view of the heart phantom

| | convolution | deconvolution | CLEAN (convolution) | CLEAN (deconvolution) |
|---|---|---|---|---|
| FEM | 1.8865e-01 | 1.1576e-02 | 1.3243e-01 | 1.3356e-01 |
| NMSE | 1.5983e+03 | 2.9431e+03 | 8.7771e-01 | 8.7949e-01 |

Table 10.7: Error metrics for the partially normalised image reconstruction of the side view of the heart phantom

## 10.3 Summary

The $K2$ Wiener filter (deconvolution) method and the CLEAN reconstruction method have been shown to work better for real holograms than convolution which is the method which has been used previously. The methods, however, work much better for a simple two-dimensional cross than for a three-dimensional heart phantom.

The CLEAN algorithm may be improved by improving the model for hologram formation. This theoretical work should be combined with experimental work to ensure that the normalisation is correct. A point source should be reconstructed with the same amplitude no matter where it is placed in the field of view of the detector. The CLEAN algorithm relies upon being able to accurately predict the hologram which would be recorded from a point source in any position. This predicted hologram is subtracted from the actual hologram when the corresponding point is added to the image. Effects such as vignetting due to the finite thickness of the zone plate (i.e. it is not infinitely thin) could prove to be important. If holograms of point sources in all possible positions can be accurately predicted, the CLEAN algorithm may work better for real three-dimensional sources.

# Chapter 11

# DISCUSSION, CONCLUSIONS AND FURTHER WORK

## 11.1 Discussion of Results

Holography is a technique where information from a three-dimensional object is recorded using a two-dimensional detector. A three-dimensional image can then be reconstructed from this information. The history behind this technique has been summarised in this report. The history of zone plate holography has been discussed in more detail.

Zone plate holography was compared and contrasted with Gabor in-line holography. Expressions for the hologram formation were found for each technique. It was shown that, if a pure Gabor zone plate were used, the resulting hologram would contain similar terms to a Gabor in-line hologram of an equivalent coherent light source. Gabor in-line holograms were shown to contain extra self-interference terms. If images were reconstructed optically from both types of hologram, these differences would result in differences in the images.

Other sources of error in a gamma ray zone plate hologram were discussed. These include errors due to the limited resolution of the gamma camera, statistical fluctuations in the activity of the gamma ray source, floating point arithmetic and under sampling. Methods for the assessment of reconstructed image quality were listed.

Image reconstruction methods based on convolution were outlined. These techniques included simulated diffraction and convolution with a matched filter. A combination of these two methods, convolution with a truncated Fresnel function, was introduced. This technique was based on simulated diffraction and, hence, the Fresnel function was complex. The reconstructed images, therefore, had both real and imaginary parts. The sum of the squares of these two parts gave the intensity of the image. The matched filter methods produced images with just a real part as both the filter and the hologram only had real parts.

Holograms of point sources were compared using the image assessment methods mentioned before. Modified mismatch was found to be the best matched filter method and convolution with a truncated Fresnel function the best overall method for computer generated holograms of point sources. Convolution with a truncated Fresnel function produced good results when either the real part or the intensity was used. It was decided to use this method as the convolution reconstruction method for the rest of the work presented here. One-dimensional Gaussian curves were fitted to the image which was reconstructed from the hologram of a point source with the parameters given in Table 6.1. The standard deviation, $\sigma$, of the Gaussian was $\approx$0.8mm for the real part and $\approx$0.6mm for the intensity in the $x$ and $y$ directions. In the $z$ direction, $\sigma$ was $\approx$2.7mm for the real part and $\approx$4.2mm for the intensity.

Methods of reconstruction based on deconvolution using Wiener filters were then investigated. These techniques attempted to reverse the hologram formation process to recover the source object. A Wiener filter was needed to reduce the effect of errors in the hologram. Two different approximate Wiener filters were given. These were called $K1$ and $K2$ Wiener filters. These filters were based on the binary zone plate pattern. The image quality was shown to decrease dramatically if the zone plate was misaligned. Deconvolution with a pure Gabor zone plate pattern was introduced as a way to check if the zone plate was aligned correctly. Deconvolution using the binary zone plate pattern with a $K2$ Wiener filter was found to be the most promising method. It was decided to use this method as the deconvolution reconstruction method for the rest of the work presented here. One-dimensional Gaussian curves were fitted to the image which was reconstructed from the hologram of a point source with the parameters given in Table 6.1. The standard deviation, $\sigma$, of the Gaussian was $\approx$0.2mm in the $x$ and $y$ directions and $\approx$1mm in the $z$ direction.

The third reconstruction method which was used was the CLEAN algorithm which was an iterative technique and could be based on convolution or deconvolution. The same computer generated holograms were used as for the convolution and deconvolution methods. Even when planes were placed 0.01mm either side of the image plane, the image was still reconstructed in a single pixel in the correct place in the correct plane with both versions of the method. For point sources, therefore, this technique seemed very good although it took many times longer to perform the reconstruction than for straightforward convolution or deconvolution.

All three techniques were tested further using theoretical holograms of extended objects. The real part of the convolution reconstruction proved to be of superior quality to the intensity so only the real part was then used. In general, deconvolution gave a better reconstruction to convolution. The CLEAN algorithm gave better results than either of the other two techniques for holograms of simple objects but did not work so well for larger or more complicated objects. Although a gamma camera can produce images in 1mm×1mm pixels, the intrinsic spatial resolution of the camera is approximately 5mm. This limited resolution of the gamma camera was found to degrade the image more than statistical fluctuations of the activity of the source or counts due to background radiation.

When images were reconstructed from real gamma ray zone plate holograms, a similar effect was seen to that for the theoretical holograms. Images of the simple two-dimensional cross source were reconstructed well but images of the extended three-dimensional heart phantom were not.

Other reconstruction techniques were investigated but these proved to be impractical for reconstructing real zone plate holograms. These techniques are discussed in appendix F.

## 11.2 Conclusions

Different forms of reconstruction by convolution and deconvolution have been given. The method of convolution to be used was chosen as convolution with a truncated Fresnel function. The method of deconvolution was chosen as $K2$ Wiener filtering with the binary zone plate pattern.

Reconstruction by convolution, deconvolution and using the CLEAN algorithm give images of varying quality. Partial normalisation gave better results than full normalisation. If the images are fully normalised, they are normalised in each $(x, y)$ plane as well as in the $z$ direction. Partial normalisation just normalises the images in the $z$ direction. Partial normalisation was therefore the method which was used. Reconstruction by deconvolution takes longer than that by convolution as the optimum Wiener filter value needs to first be found. The CLEAN algorithm takes many times longer than either of the other two methods.

Deconvolution gives a better reconstruction, that is a lower Fienup error metric (FEM) value, for the majority of the theoretical holograms and for all the real holograms which were investigated. The CLEAN algorithms gave lower FEM error metric values for some holograms but not others. Theoretical and real holograms of two-dimensional cross sources were reconstructed mainly in the image plane by the CLEAN algorithm and not in planes placed 2cm either side. The more complicated and the larger the object was, the more likely the CLEAN algorithm was to reconstruct the image in the wrong plane. This was not totally unexpected as this technique is known to work well for point objects but not for extended objects unless a priori knowledge about the source in incorporated into the algorithm [69].

When some of the possible errors in gamma ray zone plate holograms were simulated, errors due to the limited resolution of the gamma camera were found to have the largest effect on the reconstructed image. This effect can be lessened by using a binary zone plate pattern which has been blurred by the same point spread function for the deconvolution reconstruction. For the simple theoretical holograms which were investigated, the resolution of the gamma camera did not appear to have as much of a detrimental effect on the image as increasing the size of the object in the field of view of the camera or complexity (thickness) of the object in pixels. Further investigations would, however, be needed to prove this.

Images reconstructed using the CLEAN algorithm improve if the images are incremented by a smaller amount ($\Delta$) by each iteration. The algorithm has to run overnight with a large value of $\Delta$ to obtain images of cross sources. It took several days to obtain the images of the star object in Chapter 9. This number of results that could be generated has been limited by this.

In summary, the achievements of this project have been as follows.

The formation of zone plate holograms has been studied in depth and compared to an example of conventional holography. The similarities and differences have been highlighted. Other sources of error in gamma ray holography have been listed and methods for image quality assessment have been discussed.

Three reconstruction methods have been described. Images produced using these methods have been compared using the quality assessment methods. The errors in gamma ray holography have been simulated and their relative effects measured.

The existing convolution reconstruction method has been improved by introducing normalisation. Modified mismatch has been shown to give a better point spread function for the matched filter method.

Reconstruction by deconvolution using Fourier transforms has been shown to be effective for real holograms. A method for finding the optimum filter value automatically has been introduced.

The CLEAN algorithm has been used to reconstruct images in three dimensions. A version of this algorithm based on deconvolution rather than convolution has also been used.

## 11.3 Suggestions for Further Work

This work has focussed on the development of new image reconstruction techniques. For most of the time during which this work was being done, very few gamma ray zone plate holograms were available. The holograms recorded by Rew [70] were of poorer quality than the ones previously recorded by Perks [65] which have been used here. This work has therefore concentrated on reconstructing computer generated holograms. This has shown that many of the artifacts in the images are due to the reconstruction methods themselves rather than errors in the experimental holograms.

Takhar [84] and George are now recording holograms of comparable quality to those recorded by Perks. The techniques which have been developed here need to now be tested and improved using real holograms. The normalisation needs to be checked. Holograms of point gamma ray sources should be recorded. These sources should have a long half life so that the activity does not decrease appreciably over the period during which the measurements are taken. The amplitude of the reconstructed point should be the same no matter where the source is placed.

The versions of the CLEAN algorithm used in this work were based upon convolution with a truncated Fresnel function and $K2$ deconvolution with the binary zone plate pattern. The algorithm may work better when it is based upon a different convolution or deconvolution technique.

If the convolution or deconvolution could be speeded up, lower values of $\Delta$ could be used for the CLEAN algorithm. In Chapter 4, a method for calculating the discrete Fourier transform of a Fresnel function was derived. This Fourier transform could be calculated directly instead of being performed by a NAG routine to speed it up. Alternatively, the possibility of performing Fourier transforms electronically instead of computationally could be investigated as a way to speed the process up.

The optimum Wiener filter was found automatically using the method given in Section 9.1.1. Work could be done to discover if the optimum values were, in fact, found or if another value would have produced better results. If the values were not the optimum ones, the technique for finding the value could, perhaps, be improved.

Holograms of simple objects could be recorded where the exact position and size of the object are known. This was not known exactly for Perks' holograms. The expected hologram could then be calculated for an equivalent theoretical source object. This could then be compared with the actual hologram using error metrics. Errors, for example the limited resolution of the gamma camera, could then be added to the calculated hologram to make it more like the actual hologram. When all the major effects had been accounted for, the two holograms would be very similar. Once all the properties and amplitudes of the errors were known, methods could be developed for removing the effects of these errors from the reconstructed image.

In Chapter 3, formulae were derived for the optically reconstructed image from an ideal pure Gabor zone plate hologram of a point source. The same could be done for a hologram which contained errors taken from a certain statistical distribution. The effect of, say, detection noise would then be better understood.

Random noise may be removed from a real hologram by recording two holograms of the same source, one after the other. If the first hologram is correlated with the second, only the parts of the signal which are the same in both holograms will remain. This means that any random noise will be minimised as the correlation between the two sets of noise will be small (if not zero).

To summarise, this work has introduced some new reconstruction techniques and has tried to investigate their effectiveness using computer generated holograms. These techniques need to now be tested on real data. The results obtained from real holograms can then be used to further improve these methods by improving the model of hologram formation. Once the model of hologram formation has been improved, the CLEAN algorithm may work better for real three-dimensional sources.

# Appendix A

# COHERENT AND INCOHERENT LIGHT

This explanation is based on lecture notes from a short course on imaging and digital processing which was given in September 2000 at De Montfort University [42].

Consider an object which is illuminated by perfectly spatially coherent light - in the form of a plane or spherical wave. If $u_{in}(x, y)$ represents the complex amplitude of the wave emerging from the object and $u_{out}(x, y)$ is the complex amplitude of the wave in the image plane, the two are related by

$$u_{out}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{in}(x',y')p(x,y;x',y')dx'dy' \qquad \text{(A.1)}$$

where $p(x, y; x', y')$ represents the complex amplitude at $(x, y)$ in the output due to a unit strength point at $(x', y')$ in the input.

If the optical system is isoplantic, i.e. a simple shift in the object results in a simple shift in the image (scaled by the magnification of the optical system), this reduces to

$$u_{out}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{in}(x',y')p(x-x',y-y')dx'dy' \qquad \text{(A.2)}$$

If we consider the case of narrowband light, its complex amplitude varies with time. Equation A.1 then becomes

$$u_{out}(x,y,t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{in}(x',y',t)p(x,y;x',y')dx'dy' \qquad \text{(A.3)}$$

The instantaneous intensity is defined as

$$I(x,y,t) = |u(x,y,t)|^2 \qquad \text{(A.4)}$$

whereas the time averaged intensity, $\bar{I}(x, y)$, which would be recorded by an optical detector is equal to

$$\bar{I}(x,y) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} I(x,y,t)dt = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |u(x,y,t)|^2 dt \qquad \text{(A.5)}$$

The intensity of the image is given by

$$\begin{aligned}\bar{I}_{out}(x,y) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} &|u_{out}(x,y,t)|^2 dt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y;x',y')p^*(x,y;x',y') \\ &\qquad \times \left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_{in}(x',y',t)u^*_{in}(x'',y'',t)\right] dt \qquad \text{(A.6)}\end{aligned}$$

The term in [ ] is called the *mutual intensity* of the narrowband light, $J(x',y',x'',y'')$

Incoherent light is defined to be such that

$$J(x',y',x'',y'') = \bar{I}(x',y')\delta(x'-x'',y'-y'') \qquad \text{(A.7)}$$

That is, two neighbouring points at $(x',y')$ and $(x'',y'')$ have uncorrelated fields for any $(x',y') \neq (x'',y'')$. Equation A.6 then becomes

$$\bar{I}_{out}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |p(x,y;x',y')|^2 \bar{I}_{in}(x',y')dx'dy' \qquad \text{(A.8)}$$

The quantity $|p(x,y;x',y')|^2$ is the intensity point spread function.

For perfectly incoherent illumination, an optical system is linear in intensity and, if isoplanicity holds, the output (image) intensity is equal to the input (object) intensity convolved with the intensity point spread function. This simplification cannot be made for coherent light and the mutual intensity term is still important.

# Appendix B

# ELECTROMAGNETIC WAVES

Maxwell's equations can be simplified for an infinite vacuum (containing no charges or currents) to the following [82]

$$\nabla \wedge \boldsymbol{E} = -\mu_0 \frac{\partial \boldsymbol{H}}{\partial t} \qquad \text{Faraday's induction law} \tag{B.1}$$

$$\nabla \wedge \boldsymbol{H} = \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \qquad \text{Ampère's circuital law} \tag{B.2}$$

$$\nabla \cdot \boldsymbol{E} = 0 \qquad \text{Gauss's law - electric} \tag{B.3}$$

$$\nabla \cdot \boldsymbol{H} = 0 \qquad \text{Gauss's law - magnetic} \tag{B.4}$$

where $\wedge$ denotes a cross product, $\cdot$ a dot product, $\boldsymbol{E}$ is the electric field, $\boldsymbol{H}$ is the magnetic field, $\mu_0$ is the permeability of free space and $\epsilon_0$ is the permittivity of free space. These equations can be re-arranged to give the following vector wave equations for $\boldsymbol{E}$ and $\boldsymbol{H}$

$$\nabla^2 \boldsymbol{E} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0 \qquad \nabla^2 \boldsymbol{H} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{H}}{\partial t^2} = 0 \tag{B.5}$$

where $c$ is the speed of light and

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \tag{B.6}$$

These equations are valid for each rectangular component of $\boldsymbol{E} = E_x\hat{x} + E_y\hat{y} + E_z\hat{z}$ and $\boldsymbol{H} = H_x\hat{x} + H_y\hat{y} + H_z\hat{z}$. Therefore, the scalar wave equation

$$\nabla^2 V - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = 0 \tag{B.7}$$

is valid for $V = E_x, E_y, E_z, H_x, H_y$ or $H_z$.

For a plane wave travelling in the $\hat{z}$ direction, appropriate solutions to the vector wave equations are

$$\boldsymbol{E} = \boldsymbol{E_0} e^{i(\omega t - kz)} \qquad \boldsymbol{H} = \boldsymbol{H_0} e^{i(\omega t - kz)} \quad \text{where} \quad \frac{\omega}{k} = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \equiv c \tag{B.8}$$

and $\boldsymbol{E_0} = E_{0x}\hat{x} + E_{0y}\hat{y} + E_{0z}\hat{z}$ and $\boldsymbol{H_0} = H_{0x}\hat{x} + H_{0y}\hat{y} + H_{0z}\hat{z}$. These solutions must also satisfy equations B.3 and B.4. Substituting equation B.8 into these equations gives $E_{0z} = 0$ and $H_{0z} = 0$ and

$$\boldsymbol{E} = E_{0x}e^{i(\omega t - kz)}\hat{x} + E_{0y}e^{i(\omega t - kz)}\hat{y} \qquad \boldsymbol{H} = H_{0x}e^{i(\omega t - kz)}\hat{x} + H_{0y}e^{i(\omega t - kz)}\hat{y} \tag{B.9}$$

substituting this into equation B.1 gives

$$H_{0x} = -\sqrt{\frac{\epsilon_0}{\mu_0}}E_{0y} \quad \text{and} \quad H_{0y} = \sqrt{\frac{\epsilon_0}{\mu_0}}E_{0x} \tag{B.10}$$

The time averaged Poynting vector $< \boldsymbol{S} >$ gives the time averaged power flow due to the wave per unit area and also gives the direction of this flow [82]. In this case, it can be shown that

$$< \boldsymbol{S} >= \boldsymbol{E} \wedge \boldsymbol{H}^* = \sqrt{\frac{\epsilon_0}{\mu_0}}\left(E_{0x}^2 + E_{0y}^2\right)\hat{z} = \sqrt{\frac{\mu_0}{\epsilon_0}}\left(H_{0x}^2 + H_{0y}^2\right)\hat{z} \tag{B.11}$$

In his book entitled “Optics” [91], Welford states that

> It is found that all kinds of waves have to be characterised by two different quantities. These are of widely differing physical natures, depending upon the kind of wave, but in all cases there is an *amplitude*, which varies in time and space and gives interference effects, and an *intensity*, which represents the rate of energy transport. With suitable interpretations the complex amplitude and its squared modulus, the intensity, can be used in all cases. All interference experiments and many diffraction experiments can be described in these terms.

Intensity of waves (radiation) is defined as the rate of transfer of energy across unit areas by the radiation [89]. This is given by the magnitude of the time averaged Poynting vector. Intensity, $I$, is therefore given by

$$I = \sqrt{\frac{\epsilon_0}{\mu_0}}\left(E_{0x}^2 + E_{0y}^2\right) = \sqrt{\frac{\mu_0}{\epsilon_0}}\left(H_{0x}^2 + H_{0y}^2\right) = \sqrt{\frac{\epsilon_0}{\mu_0}}|\boldsymbol{E}|^2 = \sqrt{\frac{\mu_0}{\epsilon_0}}|\boldsymbol{H}|^2 \tag{B.12}$$

The wave amplitude, $\psi$, is given by

$$\psi = \sqrt{I} = \left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{4}}|\boldsymbol{E}| = \left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{4}}|\boldsymbol{H}| \tag{B.13}$$

# Appendix C

# PHOTOGRAPHIC FILM

The intensity transmittance of a developed photographic film, $T$, depends upon the intensity of light the film was exposed to. An illustration of the characteristic Hurter-Driffield curve relating intensity transmittance to exposure is shown in figure C.1 [81].

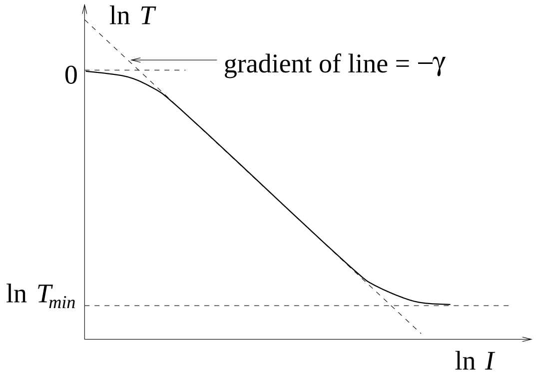


Figure C.1: Characteristic Hurter-Driffield curve, plotted using coordinates convenient in holography. $I$ = exposure (energy incident per unit area), $T$ = intensity transmittance

When the transmittance has its maximum value (= 1), i.e. the photographic film is transparent, $\ln(T_{max}) = 0$. The curve is not linear for low exposures. At high exposures, the curve is also not linear as the film reaches a maximum optical density which corresponds to a minimum transmittance $T_{min}$. The linear part of the curve can be described by the following relationship.

$$T = I^{-\gamma} \tag{C.1}$$

It is convenient if the transmittance of a hologram recorded of photographic film increases as the exposure increases and therefore $\gamma$ must be negative. This is achieved by taking a positive print of the negative hologram. If $\gamma = -2$ this simplifies matters considerably as

$$T = I^2 \quad \text{and} \quad t(x,y) = \sqrt{T(x,y)} = I(x,y) \tag{C.2}$$

where $t(x,y)$ is the amplitude transmittance of the hologram.

# Appendix D

# DERIVATION OF THE OPTIMAL (WIENER) FILTER FOR DECONVOLUTION

This derivation is taken from Numerical Recipes by Press et. al. [68]. Equation 6.2 gave the equation for hologram formation as

$$U_{i'j'} = \sum_{k=0}^{N_p-1} C^k U^k_{i'j'} = \sum_{k=0}^{N_p-1} C^k \sum_{i,j=-\frac{N}{2}}^{\frac{N}{2}-1} S^k_{ij} T^k_{(i'-i)(j'-j)} \tag{D.1}$$

The contribution of objects in a single plane a distance $p$ from the zone plate, $U^p$, to the hologram is given by the convolution of $S^p$ with $T^p$ multiplied by a constant, $C^p$.

$$U^p = C^p \left( S^p \otimes T^p \right) \tag{D.2}$$

Taking Fourier transforms of both sides and rearranging gives

$$S^{p^F} = \frac{1}{C^p} \frac{U^{p^F}}{T^{p^F}} \tag{D.3}$$

The imaging process is, however, not perfect for the reasons outlined in chapter 4. Instead of $U^p$, a 'noisy' version $\tilde{U^p}$ is recorded which includes the error term $E^p$. For simplicity,$E^p$ will be considered to be an additive error.

$$\tilde{U^p} = U^p + E^p = C^p \left( S^p \otimes T^p \right) + E^p \tag{D.4}$$

The convolution theorem gives the following relationship between the Fourier transforms of $U^p$, $S^p$ and $T^p$.

$$\tilde{U^p}^F = C^p S^{p^F} T^{p^F} + E^{p^F} \tag{D.5}$$

Where $A^F$ is the Fourier transform of variable $A$. The aim is to find an optimal filter $\Phi^p$ that will give a reconstructed image, $\tilde{S^p}$, as close to $S^p$ as possible.

$$\tilde{S^p}^F = \frac{1}{C_k} \frac{\tilde{U^p}^F \Phi^p}{T^{p^F}} \tag{D.6}$$

$\Phi^p$ must minimise the sum of squared differences between $\tilde{S^p}$ and $S^p$. From Parseval's theorem,

$$\sum_k \left|\tilde{S^p} - S^p\right|^2 = \sum_k \left|\tilde{S^p}^F - S^{pF}\right|^2 \tag{D.7}$$

this gives

$$\sum_k \left|\tilde{S^p} - S^p\right|^2 = \sum_k \frac{1}{C_k}\left|\frac{\left[U^{pF} + E^{pF}\right]\Phi^p}{T^{pF}} - \frac{U^{pF}}{T^{pF}}\right|^2 \tag{D.8}$$

The Wiener filter assumes that the signal and noise are uncorrelated. In this case, expanding the above formula gives

$$\sum_k \frac{1}{C^p}\left|T^{pF}\right|^{-2}\left\{\left|U^{pF}\right|^2\left|1-\Phi^p\right|^2 + \left|E^{pF}\right|^2\left|\Phi^p\right|^2\right\} \tag{D.9}$$

Differentiating with respect to $\Phi^p$ and setting the result equal to zero gives

$$\Phi^p = \frac{\left|U^{pF}\right|^2}{\left|U^{pF}\right|^2 + \left|E^{pF}\right|^2} \tag{D.10}$$

# Appendix E

# OTHER RECONSTRUCTION METHODS

## E.1 Matrix Wiener Filtering

This method was described in chapter 7. The standard (matrix) Wiener filter was given by equation 7.15 as

$$\text{Wiener filter} \quad [\Phi^p] = \left([H^p]^{*T}[H^p] + [\phi_{S^p}]^{-1}[\phi_{E^p}]\right)^{-1}[H^p]^{*T}[H^p] \tag{E.1}$$

$[H^p]$ is the hologram formation matrix and $[\phi_{S^p}]$ and $[\phi_{E^p}]$ are the signal and noise covariance matrices. The parametric Wiener filter was given by equation 7.16 as

$$\text{parametric Wiener filter} \quad [\Phi^p] = \left([H^p]^{*T}[H^p] + \gamma[\phi_{S^p}]^{-1}[\phi_{E^p}]\right)^{-1}[H^p]^{*T}[H^p] \tag{E.2}$$

Poisson noise was added to simulated holograms so that the matrix Wiener filter could be tested for holograms with known amounts of noise. For a Poisson distribution, the probability of obtaining $r$ events if the mean expected number is $\lambda$, is given by

$$P(r,\lambda) = \frac{\lambda^r e^{-\lambda}}{r!} \tag{E.3}$$

Using a NAG routine to generate random numbers in a Poisson distribution, each pixel value $g_{ij}$ in the hologram was replaced by a random number from a Poisson distribution with $\lambda = g_{ij}$.

Figures E.1 and E.2 show the reconstruction by parametric Wiener filtering with different values of $\gamma$ for two holograms of the same object. The hologram in figure E.1 has 16×16 pixels and the hologram in figure E.2 32×32. Before the Poisson noise was added, each hologram was scaled so all the values lay between 0.0 and 1.0e+04. When $\gamma$ =0.0, equation 7.16 becomes a matrix inversion. It can be seen from figure E.1 that matrix Wiener filtering provides an improvement over matrix inversion.

# GLOSSARY

$D$ length of one size of the detector, 42

$E(\omega_x, \omega_y)$ difference between Fourier transform of $f(x, y)$ and Fourier transform of $v(x, y)$, 71

$K1$ constant used in first approximate Wiener filter, 115

$K2$ constant used in second approximate Wiener filter, 116

$L$ factor source distribution is magnified by, 43

$L(\boldsymbol{x})$ a linear filter, 93

$M$ factor zone plate shadow is magnified by, 42

$S(\boldsymbol{\epsilon})$ intensity of incoherent gamma ray source, 43

$T(r)$ Intensity transmittance function of the zone plate, 17

$T^F(\boldsymbol{\omega})$ Fourier transform of $t(\boldsymbol{x})$, 38

$U(\boldsymbol{x})$ intensity of shadow cast by an incoherent gamma ray source, 43

$U_{ij}$ hologram pixel values, 93

$V_{ij}^p$ reconstructed image values at a distance $p$ in front of the zone plate, 93

$Z^F(\boldsymbol{\omega})$ Fourier transform of the Fresnel function, 38

$\Gamma$ constant relating the amplitude transmittance of photographic film to the total intensity it was exposed to, 45

$\Im\{f\}$ imaginary part of a function $f$, 46

$\Omega_0$ maximum spatial frequency in the object, 39

$\Psi(x,y,z,t)$ amplitude of an electromagnetic wave, 30

$\Re\{f\}$ real part of a function $f$, 46

$\Upsilon$ constant relating the amplitude transmittance of photographic film to the total intensity recorded by the gamma camera, 46

$\delta$ size of the finest detail in the object, 39

$\otimes$ 2D convolution operation, 36

$\psi(x,y,z)$ spatial variation of the amplitude of an electromagnetic wave, 31

$a$ size of the object, 39

$a(x,y)$ circular aperture function, 73

$f(x,y)$ part of $v(x,y)$ recorded by the detector, 70

$h(x,y)$ perfect recorded signal, 67

$l$ radius of the central Fresnel zone, 39

$p$ source to object/zone plate distance, 35

$q$ object/zone plate to detector distance, 35

$s(\boldsymbol{\epsilon})$ source amplitude distribution function, 35

$t(\boldsymbol{\alpha})$ amplitude transmittance function of the transparency, 35

$u(\boldsymbol{x})$ amplitude of the waves incident on the detector, 35

$v(x,y)$ signal from Gabor zone plate hologram of an on-axis point source, 70

$z(\boldsymbol{x})$ Fresnel function, 38

**2D DFT** discrete two-dimensional Fourier transform, 68

**2D FFT** two-dimensional fast Fourier transform, 64

**2D FT** continuous two-dimensional Fourier transform, 68

**BGZP** binary Gabor zone plate, 92

**BNCT** Boron Neutron Capture Therapy, 26

**CT** Computed Tomography, 12

**ESF** edge spread function, 82

**FEM** Fienup error metric, 84

**GZP** pure Gabor zone plate, 90

**LSF** line spread function, 82

**MSE** mean squared error, 83

**MTF** modulation transfer function, 81

**NMSE** normalised mean squared error, 84

**OTF** optical transfer function, 80

**PET** Positron Emission Tomography, 12

**PM tube** photomultiplier tube, 60

**PSF** point spread function, 81

**SNR** signal to noise ratio, 83

**SPET** Single Photon Emission Tomography, 12

**W** Wiener spectrum, 83

**Ш**$(\omega_x, \omega_y, \frac{2\pi}{a}, \frac{2\pi}{b})$ 2D FT of ш$(x, y, a, b)$, 68

**ш**$(x, y, a, b)$ sampling function, 67

**Appendix F – this conference paper has been removed from the electronic version of this thesis due to copyright restrictions.**

# REFERENCES

[1] H. C. Andrews and B. R. Hunt. *Digital Image Restoration*. Prentice-Hall, Inc., 1977.

[2] E. Bacary. LastWave signal and image processing environment. Freeware, http://wave.cmap.polytechnique.fr/soft/LastWave, 2000.

[3] H. H. Barret. Fresnel zone plate imaging in nuclear medicine. *Journal of Nuclear Medicine*, 13:382–385, 1972.

[4] S. A. Benton. Hologram reconstructions with extended incoherent sources. *Journal of the Optical Society of America*, 59:1545–1546, 1969.

[5] G. Beylkin. Wavelets, multiresolution analysis and fast numerical algorithms. A draft of INRIA lectures, 1991.

[6] T. D. Beynon. Neutron holography. *Physics Bulletin, The Institute of Physics*, 37:128–131, 1986.

[7] T. D. Beynon, V. S. Crocker, and H. U. Mast. Zone plate encoded neutron holography. *Neutron Radiography*, pages 945–951, 1983.

[8] T. D. Beynon, M. R. Hawkesworth, T. R. Mathews, and M. A. Odwyer. Positron emission holography - a new method of creating positron camera images. *Nuclear Instruments and Methods in Physics Research Section A - Accelerators, Spectrometers and Associated Equipment*, 273(2–3):908–910, 1988.

[9] T. D. Beynon, I. Kirk, and T. R. Mathews. Gabor zone plate with binary transmittance values. *Optics Letters*, 17(7):544–546, 1992.

[10] T. D. Beynon and A. G. Pink. Neutron holography using fresnel zone plate encoding. *Nature*, 283:749–751, 1980.

[11] T. D. Beynon and R. M. R. Strange. Computational study of diffraction patterns for near-field fresnel and gabor zone plates. *Journal of the Optical Society of America A*, 17(1):101–106, 2000.

[12] Y. J. Bo Peng and X. Zhang. Noise suppression with wavelets in image reconstruction for aperture synthesis. In *Astronomical Data Analysis Software and Systems V*, volume 101 of *ASF Conference Series*, pages 223–226, 1996.

[13] H. Boersch. Zur Bilderzeugung im Mikroskop. *Zeitschrift für Technische Physik*, 19, 1938.

[14] M. Born and E. Wolf. *Principles of Optics.* Permagon Press, 6 (corrected) edition, 1983.

[15] J. F. Briesmeister, editor. *MCNP - A General Monto Carlo N - Particle Transport Code. Version 4B.* Los Alamos National Laboratory, USA, 1997.

[16] S. T. Caplan. *Optimisation of Binary Gabor Zone Plate Encoded Holography Techniques with Visible Wavelengths.* PhD thesis, School of Physics and Astronomy, The University of Birmingham, December 1997.

[17] V. Carus. Correspondance using email, February 2000.

[18] H. J. Caulfield, editor. *Handbook of Optical Holography.* Academic Press, 1979.

[19] N. M. Ceglio. Zone plate coded imaging on a microscopic scale. *Journal of Applied Physics*, 48(4):1563–1565, 1977.

[20] N. M. Ceglio, D. T. Attwood, and E. V. George. Zone plate encoded imaging of laser-produced plasmas. *Journal of Applied Physics*, 48(4):1566–1569, 1977.

[21] H. Cember. *Introduction to Health Physics.* McGraw-Hill, 1996.

[22] T. J. Cornwell. A method of stabilising the CLEAN algorithm. *Astronomy and Astrophysics*, 121:281–285, 1983.

[23] Yu. N. Denisyuk. Photographic reconstruction of the optical properties of an object in its own scattered radiation field. *Soviet Physics Doklady*, 7:543, 1962.

[24] E. R. Dougherty. *Probability and Statistics for the engineering, computing and physical sciences.* Prentice Hall, 1990.

[25] M. B. Emmett. The MORSE monte carlo radiation transport code system. *Oak Ridge National Laboratory*, ORNL-4972, 1975.

[26] E. E. Fenimore and T. M. Cannon. Coded aperture imaging with uniformly redundant arrays. *Applied Optics*, 17(3), 1978.

[27] J. R. Fienup. Invariant error metrics for image reconstruction. *Applied Optics*, 36(32):8352–8357, 1997.

[28] J. S. Fleming and B. A. Goddard. An evaluation of techniques for stationary coded aperture three dimensional imaging in nuclear medicine. *Nuclear Instruments and Methods in Physics Research*, 221:242–246, 1984.

[29] D. L. Fried. Analysis of the CLEAN algorithm and implications for superresolution. *Journal of the Optical Society of America A-Optics Image Science and Vision*, 12(5):853–860, 1995.

[30] D. Gabor. A new microscopic principle. *Nature*, 161:777–778, 1948.

[31] D. Gabor. Microscopy by reconstructed wave-fronts. *Proceedings of the Royal Society of Biology*, A197:454–487, 1949.

[32] D. Gabor. Holography of the "whole picture". *New Scientist*, 29(4):74–78, 1966.

[33] S. George. Developments in gamma ray holography for medical imaging. Course 4 project report, School of Physics and Astronomy, The University of Birmingham, March 2000.

[34] P. Hariharan. *Optical Holography - Principles, techniques and applications.* Cambridge University Press, 2 edition, 1996.

[35] E. Hecht. *Schaum's theory and problems of optics.* McGraw-Hill, 1975.

[36] E. Hecht. *Optics.* Addison-Wesley, 3 edition, 1998.

[37] C. W. Helstrom. Image restoration by the method of least squares. *Journal of the Optical Society of America*, 57:297–303, 1967.

[38] J. A. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astophysics Supplement Series*, 15:417–426, 1974.

[39] P. Horowitz and W. Hill. *The Art of Electronics.* Cambridge University Press, 1980.

[40] Holophile Inc. History of holography. Web page, http://www.holophile.com/history.htm, 1998.

[41] Sun Microsystems Inc. *What every computer scientist should know about floating-point arithmetic.* Sun Microsystems Inc., 1994.

[42] Institute of Mathematical and Simulation Sciences, De Montfort University. *Introduction to Fourier Optics and Optical Processing*, 2000.

[43] A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall International, Inc, University of California, Davis, 1989.

[44] Z. Jaroszewicz. A review of fresnel zone plate moirè patterns obtained by translations. *Optical Engineering*, 31(3), 1992.

[45] A. R. Jones. The focal properties of phase zone plates. *British Journal of Applied Physics*, 2:1789–1791, 1969.

[46] I. Kirk. *An Investigation into Aspects of the Behaviour of Overlapping Fresnel Zone Plate Systems: A New Concept in Zone Plate Design.* PhD thesis, Department of Physics, University of Birmingham, April 1992.

[47] J. Kirz. Phase zone plates for x-ays and the extreme ultra-violet. *Journal of the Optical Society of America*, 64(3):301–309, 1974.

[48] G. F. Knoll. *Radiation Detection and Measurement.* John Wiley & Sons, Inc., 1989.

[49] M. K. Kuo. *Binary Gabor Zone Plate Encoded Holography in the Near-Infrared Region.* PhD thesis, School of Physics and Astronomy, The University of Birmingham, September 1997.

[50] Argonne National Laboratory. *Reactor Physics Constants.* United States Atomic Energy Commission, 2 edition, 1963.

[51] National High Magnetic Field Laboratory. Molecular expressions microscopy primer. Web page, http://www.micro.magnet.fsu.edu/primer/anatomy/sources.html, 1999.

[52] E. N. Leith and J. Upatniecks. Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America*, 52:1123, 1962.

[53] B. MacDonald, L. T. Chang, V. Perez-Mendez, and L. Shirashi. Gamma-ray imaging using a fresnel zone plate aperture, multi-wire proportional chamber detector and computer reconstruction. *IEEE Transactions on Nuclear Science*, NS–21:678–684, 1974.

[54] S. Mallat. *A Wavelet tour of Signal Processing.* Academic Press, 1998.

[55] K. A. Marsh and J. M. Richardson. The objective function implicit in the CLEAN algorithm. *Astronomy and Astrophysics*, 182:174–178, 1987.

[56] A. G. Marshall and F. R. Verdun. *Fourier Transforms in NMR, Optical and Mass Spectrometry.* Elsevier, 1990.

[57] H. U. Mast. *Zone Plate Encoded Neutron Holography.* PhD thesis, Department of Physics, The University of Birmingham, November 1983.

[58] T. R. Mathews. *Studies in Fresnel Zone Plate Encoded Neutron Holography and Synthesised Holographic Encoding.* PhD thesis, Department of Physics, The University of Birmingham, February 1988.

[59] J. L. Matteson, D. E. Gruber, W. A. Heindl, M. A. Pelling, L. E. Peterson, R. E. Rothschild, R. T. Skelton, P. L. Hink, K. R. Slavis, W. R. Binns, T. Tumer, and G. Visser. Recent advances in czt strip detectors and coded mask imagers. *Astronomy and Astrophysics Supplement Series*, 138:575–576, 1999.

[60] L. Mertz and N. O. Young. Fresnel transformations of images. In *Proceedings of the International Conference on Optical Instruments*, pages 305–312, London, 1961. Chapman and Hall.

[61] P. D. Nellist and S. J. Pennycook. Accurate structure determination from image reconstruction in ADF STEM. *Journal of Microscopy*, 190:159–170, 1998.

[62] Numerical Algorithms Group Limited, Oxford, UK. *The NAG Fortran Library Manual, Mark 14*, 1990.

[63] Ohio-nuclear inc. *Series 100 radioisotope camera manual.*

[64] International Commission on Radiation Units and Measurements. Medical imaging - the assessment of image quality. Technical Report 54, ICRU, 1995.

[65] J. R. Perks. *Gamma Ray Imaging using Binary Gabor Zone Plate Holography.* PhD thesis, School of Physics and Astronomy, The University of Birmingham, January 1998.

[66] A. G. Pink. Production of fresnel zone plate encoded neutron holograms and their optical reconstruction. Master's thesis, Department of Physics, The University of Birmingham, October 1979.

[67] R. L. Powell and K. A. Stetson. Interferometric vibration analysis of three-dimensional objects by wavefront reconstruction. *Journal of the Optical Society of America*, 55:612, 1965.

[68] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran, Second Edition.* Cambridge University Press, 1992.

[69] Wolfram Research. Eric weisstein's world of mathematics. Web page, http://www.mathworld.wolfram.com, 2000.

[70] G. A. A. Rew. *Studies in Zone Plate Encoded Holography with High Energy Gamma Rays.* PhD thesis, School of Physics and Astronomy, The University of Birmingham, December 1999.

[71] G. L. Rogers. Gabor diffraction microscopy: the hologram as a generalised zone plate. *Nature*, 166(4214):237, 1950.

[72] W. L. Rogers, K. S. Han, L. W. Jones, and W. H. Beierwalter. Application of a fresnel zone plate to gamma ray imaging. *Journal of Nuclear Medicine*, 13:612–615, 1973.

[73] J. W. Rohlf. *Modern Physics from $\alpha$ to $Z^0$*. John Wiley and Sons, Inc, 1994.

[74] G. Saxby. *Practical Holography.* Prentice Hall, 2 edition, 1994.

[75] P. F. Sharp, H. G. Gemmell, and F. W. Smith, editors. *Practical Nuclear Medicine.* IRL Press, 1989.

[76] Y. Shen. Binary gabor zone plate encoded holography in middle and far infrared region. Paper for the Third International Conference on Optical Information Processing, May 1999.

[77] Y. Shen. *Investigation of Binary Gabor Zone Plate Encoded Holography Technique with Infrared.* PhD thesis, School of Physics and Astronomy, The University of Birmingham, June 2000.

[78] K. Sinclair. Conversation on telephone, January 2000.

[79] R. C. Singleton. An algorithm for computing the mixed radix fast fourier transform. *IEEE Transactions on Audio and Electroacoustics*, AU–17(2):93–103, 1969.

[80] Waterloo Maple Software and the University of Waterloo. Maple V release 3, 1994.

[81] L. M. Soroko. *Holography and Coherent Optics.* Plenum Press, 1980.

[82] D. H. Staelin, A. W. Morgenthaler, and J. A. Kong. *Electromagnetic waves.* Prentice-Hall International, Inc., 1994.

[83] R. M. R. Strange. *Studies in Zone Plate Encoded Holography and Antennas.* PhD thesis, School of Physics and Space Research, The University of Birmingham, 1996.

[84] P. S. Takhar. Development of gamma ray holography for imaging in nuclear medicine. Mid-term report, School of Physics and Astronomy, The University of Birmingham, January 2000.

[85] J. Turunen, A. Vasara, H. Ichikawa, E. Noponen, J. Westerholm, M. R. Taghizadeh, and J. M. Miller. Storage of multiple images in a thin synthetic fourier hologram. *Optics Communications*, 84(5.6):383–392, 1991.

[86] C. E. Unwin, G. A. A. Rew, J. R. Perks, T. D. Beynon, and M. C. Scott. Reconstruction of images from gabor zone plate gamma-ray holography. In *3rd International Conference on Optical Information Processing*, volume 3900 of *Proc SPIE*, pages 345–352, 1999.

[87] V. Čížek. *Discrete Fourier Transforms and their Applications.* Adam Hilger Ltd, 1986.

[88] M. Wax and T. Kailath. Efficient inversion of toeplitz-block toeplitz matrix. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-31(5):1218–1221, 1983.

[89] R. C. Weast, editor. *CRC Handbook of Chemistry and Physics.* CRC Press, Inc., 1984.

[90] D. R. Weaver. Dwpol polynomial fitting program, 1997.

[91] W. T. Welford. *Optics.* Oxford University Press, 3 edition, 1988.

[92] I. K. Woodgate. *Zone Plate Encoded Holography in the Optical Region.* PhD thesis, School of Physics and Space Research, The University of Birmingham, June 1996.

[93] H. D. Young and R. A. Freedman. *University Physics.* Addison-Wesley, 9 edition, 1996.