THE UNIVERSITY *of* EDINBURGH

An Application of Optical Interference to Dynamic Position Measurement in Three Dimensions

James D.C. Allan

Ph.D. Thesis
University of Edinburgh
1987

Abstract

This thesis is concerned with the measurement of the positions of points and bodies moving in trajectories in three dimensions, and the use of a new technique of optical interference which allows such measurements to be made dynamically.

A variety of existing techniques for both static and dynamic three-dimensional position measurement are discussed, and the design of the new interferometer is introduced.

The geometry of points, curves and surfaces in three dimensions is examined, with emphasis on the intersection of the point loci represented by the coordinate output of measuring instruments. The coordinates output by the interferometer represent surface loci which are quadric surfaces. A method of calculating the position and orientation of a body using three quadric surface intersection curves is presented.

Diffraction of monochromatic light at an aperture is considered and it is shown that an interferometer working by division of wavefront can be used to obtain continuous information about the movement of the source of radiation, with that source free to move in up to three dimensions. A lens may be used to produce a compact instrument based on these principles. The diffraction integral equations are modified to incorporate the effect of a lens in the diffraction field. It is shown that even complex lenses can be represented by a few parameters in the diffraction equations.

From the evaluation of these diffraction integrals, it is shown how the movement of interference fringes provides a coordinate output and how this is related to the locus of the radiation source. A method of obtaining very high resolution measurements of interference fringe pattern movement is presented. The interferometer was built and tested and the above theory verified in practice in a series of optical bench tests.

The implementation of a system which uses this interferometer to measure the dynamic performance of industrial robots is considered. The optimum positions for the instruments are derived, and the method of designing the interferometer to give the required resolution is presented.

Contents

List of Figures

Chapter 1

Introduction

1.1 Summary

This thesis is concerned with the measurement of the positions of points moving in three dimensions and the employment of a technique of optical interference to obtain these measurements. This optical approach admits the possibility of a non-contact, three dimensional measurement system capable of dynamic operation in monitoring the position and orientation of a moving object. The instrument itself is an interferometer and has no moving mechanical parts.

The subject matter covered includes: a review of the requirements of robot performance testing and the instruments which may be used in such tests, the geometrical aspects of the location of points and bodies, the theoretical aspects of diffraction and interference in the interferometer, and methods of sensing interference fringe movement. An instrument has been built to test the theory, and the experimental results are reported. An application in which the position and orientation of a body are to be measured is considered and the solution for optimal location of several instruments with respect to the swept volume of the body is obtained. Finally the potential development of the instrument is described and applications are discussed.

1.2 Overview of Contents

Chapter 1

Many physical principles have been applied to problems of position measurement, and some of these are apparent from the examples of existing position measurement devices described in this chapter. Techniques, for both static and dynamic measurement are discussed here, the background to the discussion being the requirement to be able to make measurements of the performance of industrial robots. The need for a system which can make accurate dynamic measurements of a robot's position in three dimensions was the original motivation for the design of the interferometer described in this thesis. Chapter 1 concludes with a brief description of the operation of this instrument.

Chapter 2

One perspective which establishes a basis for comparison between different approaches is geometry. The classification of instruments used in measurement by their underlying physics is less important than an understanding of the geometry of the space in which they operate. A classification may constitute a more or less exhaustive list. The geometry, as considered in Chapter 2, draws together the similarities between some apparently quite diverse techniques. It is shown that the rather unusual geometry of point location, by the interferometric method introduced in Chapter 1, may be used in essentially the same way as the geometry of some familiar devices.

Chapter 3

The ability to sense the position of a *point* in space of up to three dimensions, by whatever means, is a first step. A frequent requirement is that information be obtained about the location of a *body*, an assemblage of points, with respect to a reference frame. Chapter 3 addresses the relationship between information about a point and information about a body. It is shown that knowledge about a number of space curves and surfaces which pass through points on a body can be used to determine the position and orientation of the body. The conclusions of Chapter 3 anticipate the later application of the interferometer in practical situations. The approach is more general than this however and may be applied in other circumstances with other instruments and other geometries.

Chapter 4

With the basic requirements for point and object location laid down in Chapters 2 and 3, Chapter 4 introduces the physical background to the interferometer. Using the Rayleigh-Sommerfeld formulation for diffraction, expressions for electric field and irradiance at a distant observation point are given, employing paraxial, Fresnel and Fraunhofer approximations. This is for the case of plane waves incident on rectangular slit apertures in a non-axial direction. The stringent conditions required for these approximations to be valid apply to the situation in which diffraction of incident radiation at an aperture is observed in free space beyond the aperture.

The introduction of a set of refracting elements to form a lens behind the aperture could conceivably complicate the situation: the paths of diffracted rays are composed of many straight line segments at different refracted angles at each

surface of the lens, each "ray" with a different route to its image point. Ignoring the surfaces of the lens and the positions of these surfaces and treating the lens as a device which carries out a particular, idealised operation on incident rays allows the electric field and irradiance at the image surface to be obtained with no more difficulty than in the lensless case. This lens model is also given in Chapter 4.

Chapter 5

In Chapter 5 the idealised lens is employed in the determination of electric field and irradiance due to diffraction of incident radiation at two parallel slits placed in front of the lens. With incident off-axis plane waves this gives results similar to those obtained with the approximations of the previous chapter. With incident spherical wavefronts from a point source the interference fringes of the irradiance pattern are again shown to be present and to be functions of the source position (in terms of their period and amplitude). The amplitude of these fringes can be seen to be at variance with the incident plane wave case due to the variation of diffraction pattern overlap with source position.

Detection of the fringe pattern position on the lens image surface is considered. Knowledge of the position of this pattern gives information about the position of a point source (or the orientation of incident plane radiation). For a high resolution arrangement with fine interference fringes, a means of sensing fringe movement using a diffraction grating is described and the sensed power transmitted by that grating is related to a locus of the source. For a point source this locus is a hyperboloid of revolution with foci at the centres of the two slit apertures.

Chapter 6

To test the foregoing theory some experiments were carried out. These are described in Chapter 6. Suitable slits were obtained and set up with a lens and means to observe the generated interference patterns. Using a low power gas laser, both plane and spherical incident wavefronts were produced and the effect of source movement on the position of interference patterns was observed. Within the limitations of the equipment this was found to be in good agreement with prediction. The movement of coarse interference patterns was monitored with a video camera. For fine fringe patterns, movement was detected using a diffraction grating, the power transmitted through the grating being detected with a photomultiplier.

To obtain the direction as well as the magnitude of pattern movements a beam splitter was used to produce two similar interference patterns at two focal surfaces of the same lens. With gratings positioned in quadrature the output signals as the pattern moved were two sine waves, superimposed on DC levels, and 90 degrees out of phase. Other tests were carried out to establish the robustness of the optical arrangement.

Chapter 7

Having ascertained that the principles involved can be embodied in a working, single-coordinate laboratory system, attention is given in Chapter 7 to the implementation of the instrument and its geometry in practice. The most general case, monitoring the position of a body moving with six degrees of freedom, is considered as an example. This case can be identified with problems of industrial inspection and with the dynamic testing of machinery such as industrial

robots.

The method of obtaining the optimum positions of the "receivers" for the given circumstances is derived. (This method of solution can also be applied to other devices for three dimensional position measurement). In this example of an implementation, due consideration is given to the necessary engineering design of a supporting space frame for the receivers.

In the example, the known information is the working volume to be covered by the instrument and the resolution of measurement required within this volume. It is shown how the various parameters of the optical system are determined, in order to meet this requirement. This is, in effect, tailoring the design of the interferometer for the task.

Chapter 8

In the final chapter the design of the interferometer is reviewed and improvements and changes to its components are discussed. For the experimental part of this work, components were chosen for their availability, and flexibility of application in a laboratory environment. Development for applications requiring reduced size and weight of the instrument is quite possible. A compact, solid-state design is presented. Several applications are discussed and calculated values for the resolution of point position measurements over various ranges are given.

1.3 Performance Measurement in Robotics

Accurate performance measurements of industrial robots are required:

1. for the calibration of machines during manufacture and in the field,

2. for use as part of the design cycle in improving both the static and dynamic characteristics of new designs of robot,

3. for purposes of comparison between the products of different manufacturers.

These requirements for both static and dynamic performance measurements are particularly demanding, and before proceeding to describe the capabilities of various existing measuring instruments and the interferometer which is the main subject of this thesis, some aspects of metrology in robotics are discussed.

1.3.1 Static Tests

Much useful information can be gained about a robot's performance by a series of static tests which measure the repeatability and accuracy of the machine over its work envelope. With suitable algorithms, the measured position data (on the accuracy of the robot) obtained from these procedures can then be used to calibrate the machine, eliminating a large part of its kinematic error.

In the *off-line* programming of an industrial robot, a series of commands defining the required sequence of positions of the machine is given, usually with position and orientation of the end-effector defined in cartesian coordinates and Euler angles (or roll, pitch and yaw) respectively. From this input data in $x, y, z, \phi, \theta, \psi$ the robot control computer calculates the joint angles or joint displacements in the machine, corresponding to the desired input coordinates.

This is the inverse kinematics calculation, and to carry it out the controller uses stored values of various kinematic parameters of the machine. This kinematic model is illustrated in Figure 1.1, which shows a three axis machine. The model comprises parameters of four types: S is joint length, θ is the joint angle,

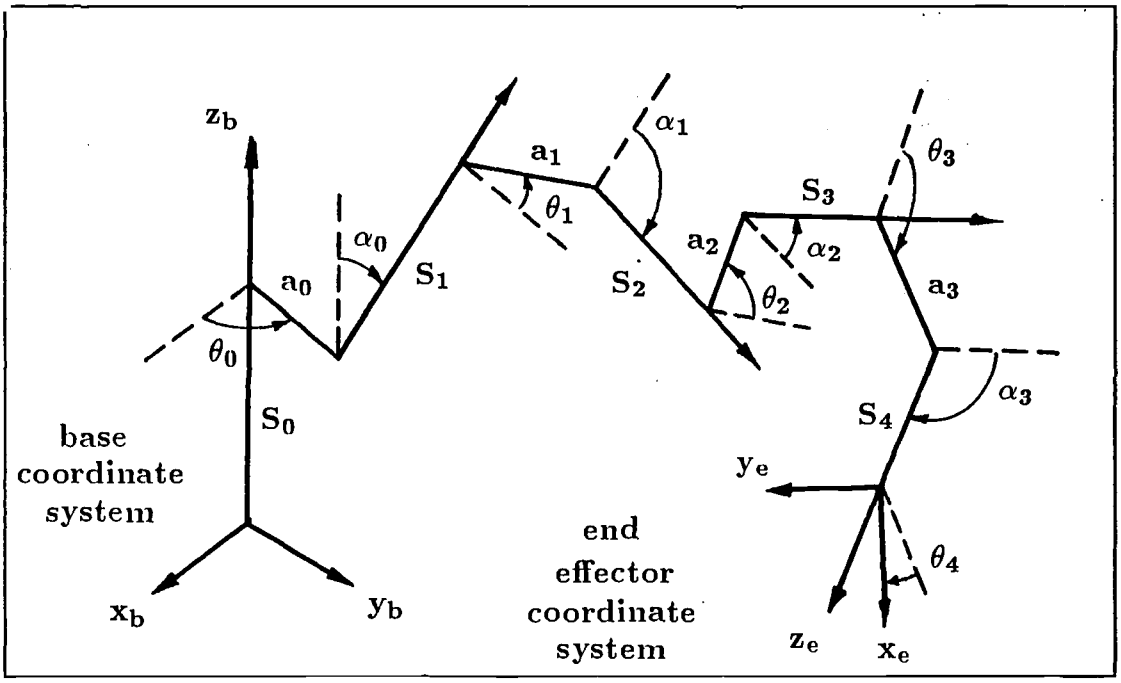


Figure 1.1: Kinematic Parameters of a Three Degree of Freedom Robot

a is the link length, the length of the common perpendicular between adjacent joint axes and α is the angle between adjacent joint axes. For a rotary joint, θ is variable and for a prismatic joint the length S is variable. For an n degree of freedom robot, and including the parameters required for the transformation to the base and end-effector coordinate systems, there are $4n+6$ kinematic parameters in the model. For details of this see [Featherstone] and also [Paul] (who uses a different nomenclature).

If these parameters are taken as the nominal values for the machine, i.e. calculated from the nominal design dimensions, then the accuracy of the machine's performance will almost inevitably be poor. There are several reasons for this. Firstly, the build-up of tolerances is hard to avoid in the manufacture and assembly of machined components. This can result in considerable cumulative errors in the constant parameters of the robot. Secondly and potentially

more seriously, the joint position transducers can be assembled to the machine with an error in their zero position, putting a constant bias on the variable joint parameter. Thirdly, the kinematic model assumes that the links are rigid bodies and no account is taken of the deflection of the arm under load. Other factors affecting accuracy are backlash, compliance and kinematic inaccuracy in the drive trains, and the expansion and contraction of material with temperature.

The first two of these problems can be overcome using a calibration procedure which derives the correct kinematic model for each individual robot. The procedure involves moving the robot to a number of different known positions in its working volume and recording the joint variables of the robot at each position. The set of simultaneous kinematic equations obtained in this way can then be solved for the true values of the kinematic parameters of the robot. The model is then stored in EPROM in that particular robot's controller. This approach is now used by some robot manufacturers to improve the accuracy of their products. One published account of a calibration procedure is that of [Whitney *et al.*].

Having established the computational method of calibration, the remaining problem is *how* to obtain accurate data on the robot's true positions during calibration. The accuracy with which these true positions must be measured is determined by the error which is ultimately acceptable in the machine after calibration. The measuring instrument itself must have an accuracy which is an order of magnitude better than the acceptable tolerance on the parameters being measured. In other words, the tolerances of the measurement system must be negligible compared to the measured value tolerances. If the position of a machine is intended to be correct to $\pm 1mm$ say, then an instrument with an error of its own of that order is quite inadequate. The measured error from the

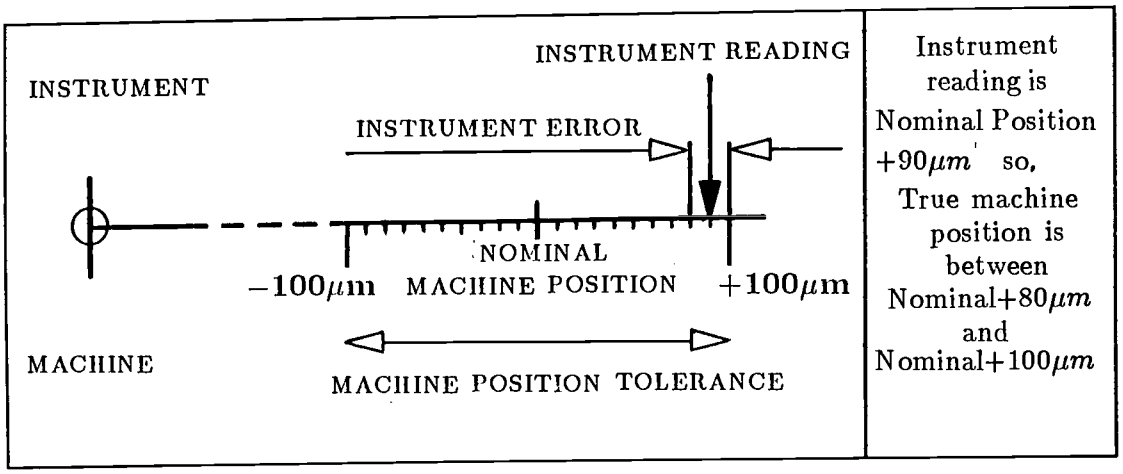


Figure 1.2: Machine and Instrument Error

theoretical position could be due to either the machine or the instrument. See Figure 1.2.

In robotics the current generation of robots have a positional *repeatability* which is typically $\pm 100\mu m$ but an *inaccuracy* which can be $\pm 5mm$ or worse. This poor accuracy is due mainly to errors in the kinematic model of each machine as discussed above.

Since it is required to be able to verify repeatability as well as accuracy of robots, the tolerance of $\pm 100\mu m$ (representative of robots of the size of the Puma 560 and ASEA IRb6), on the repeated return to the same programmed position, is the quantity which the system must be able to measure. (As Whitney points out, the effects of thermal expansion are of this order, and so the figure of $\pm 100\mu m$ repeatability will be difficult to improve upon). The instrument used for making the measurements must therefore have a $\pm 10\mu m$ error at most, i.e. be accurate to $\pm 10\mu m$ over its working volume. This ensures that the contribution of instrument error to the measurement reading is acceptable (i.e. one tenth of the tolerance on the quantity being measured). With the addition of suitable, precisely located fixtures, a measurement system with sufficient accuracy to

check the repeatability of a robot could also give the precise information on inaccuracy necessary for deriving the true kinematic model for calibration.

1.3.2 Dynamic Tests

The dynamic performance of robots is also of interest. In particular, the behaviour of machines attempting to follow particular three-dimensional trajectories can yield useful information on the performance of various control schemes in the robot controller. Obtaining measurements of the robot's position and orientation while it is in motion is however a more difficult problem than the acquisition of measurements in a static situation.

To obtain position and orientation of the end-effector, at least three points on the end-effector must be monitored, (see Chapter 3). To obtain a good record of the machine's trajectory, points on that trajectory (for each of the three target points) must be measured at a frequency which is high enough to give a good approximation to the continuous path of the end-effector. From the trajectory position data, velocity and acceleration profiles can be derived. This information can be used to evaluate the performance of the control system, and can be compared with the programmed trajectory or the output of a computer simulation of the machine's dynamic behaviour.

Initially, useful information on a robot's dynamic performance can be obtained with measuring equipment which is less accurate than that used for the static measurements described above. (This is born out by some of the work being done at Unimation, using a system described later in this chapter). To obtain more detailed information on dynamic behaviour, for instance on vibrations and small perturbations from the programmed trajectory at the end-effector, more precise dynamic instruments are required.

1.3.3 Performance Standards

To overcome the vagueness and ambiguity in the specifications[1] given by robot manufacturers for the performance of their products, attempts are being made to establish standards of terminology, common test criteria and universally recognised test procedures. The adoption of one standard would allow industrial users to make meaningful comparisons between the performance claims of different robot manufacturers.

The Ford Motor Company have produced a set of standards against which robots can be measured prior to acceptance by that company. The Ford Robot and Automation Applications Consulting Centre (RAACC) standard [Ford] includes tests for reliability, repeatability, accuracy, overshoot and settling time, cycle time, power consumption, start-up software, and input/output response. For the tests which require independent measurements of end-effector position, the standard also specifies the types of non-contact measurement instrumentation to be used. The test equipment specified for static tests has much higher accuracy than that to be used in dynamic tests.

Another standard in preparation is that of the International Standards Organisation (ISO). At present the working document [ISO] is in two parts covering performance criteria and testing methods. The particular instruments to be used for the tests are not specified, but the total uncertainty of measurement due to factors such as instrument error, external influences (e.g. temperature), and computation errors is required to be less than 25% of the robot's repeatability.

In Japan the Ministry of International Trade and Industry (MITI), Agency of Industrial Science and Technology has sponsored a "Study of Standardiza-

[1] See for instance the criticism in the preface of [Ford].

tion of Industrial Robots" [Ozaki, Ito and Inagaki]. The scope of this study ranges from standardized ways of describing the configuration of robot arms and their controllers, through specifying performance criteria, to reliability and maintenance considerations. Osaki *et al.* recognise from the start the difficulties inherent in obtaining dynamic performance data.

In Europe a Tripartite Group for Robot Performance Testing was set up under the auspices of the UK Department of Trade and Industry (DTI), the French Ministry of Industry and Research (MTR) and the German Ministry for Research and Technology (BMFT). In addition to their work on standards, one result of these tripartite studies was the establishment of a requirement for an instrument system capable of carrying out three-dimensional dynamic performance measurements of moving robots. The accuracy specified for such a device was given as $\pm 10 \mu m$ over $1 m^3$, [Davey]. The specification also requires the system to have the ability to measure position and orientation, to be portable, to have automated collection and processing of data, and to apply no disturbing load to the end effector.

Some general comments can be made on the above investigations into standards. The purpose of such standards is in three main areas. The first of these is to standardize terminology. Clearly this is important if meaningful comparisons are to be made between the features and capabilities of different robots. The second function is to produce a set of standard tests. These are a means of putting some figures on a machine's ability to perform particular tasks, for example, the positional repeatability which a robot can achieve when approaching a point from different directions. The third function defines (as part of the test procedure) the accuracy of the instrumentation or even the type of instrument to be used in the tests.

The following two sections survey static and dynamic instruments for obtaining the necessary independent measurements of the robot's position. Some of the methods were designed with robots as their application area and some are adaptations or proposed adaptations of existing instruments and techniques to this particular problem. Some of the techniques described in the section on dynamic systems can also be used to obtain the static measurements. Both the following sections concentrate on the capabilities of the instruments themselves, rather than test results for specific robots.

1.4 Review of Static Systems

1.4.1 Electrical Contacts

A system tried and ultimately rejected by the robot manufacturer Unimation Europe Ltd makes use of an accurate fixture on which the robot to be calibrated is located. The fixture has a number of electrical contact points or surfaces at precisely known locations and the robot is equipped with a probe which it moves to make electrical contact with each of these in turn. As each contact is made, the joint angles of the robot are recorded. With a sufficient number of measurements at contact points distributed throughout the robot's working volume, the kinematic model of the machine can be derived, although in this case the objective was restricted to nulling the angle transducers. No figures are available for the accuracy of this procedure.

Some observations can be made. The accuracy of a robot calibration would depend on the accuracy of the fixtures and the electrical probe. There can be a short time delay between contact being made and the machine coming to a stop (or the readings being taken), introducing an error into the readings. Also,

the robot may in fact apply a force to the contact on the fixture and cause a small deformation in the fixture, in the probe or in the robot arm itself. Thus to operate at all, a system requiring contact to be made can in fact distort the measurements it is designed to obtain.

1.4.2 Proximity Sensors

Another technique which requires fixtures to be positioned in the robot's working volume has been used by Citroen, Peugeot, and Unimation. The system operates as follows. Machined cubes are positioned in the workspace, and a trihedron carrying eddy current proximity sensors is fitted to the robot's end-effector. See Figure 1.3. The robot is programmed to move from one cube to the next, positioning the trihedron close to the corners of the cubes. The system used by Unimation was built by McMaster at Cranfield Institute of Technology, and uses Karmen Sciences eddy current sensors which have an accuracy of $\pm 10\mu m$.

With six proximity sensors in position, this type of system can measure repeatability both in position and orientation. No contact occurs between the moving and stationary parts of the system. Thus no force is imparted to the robot by the fixture, and readings of position are not distorted by the compliance of the machine. Furthermore, with suitable data capture equipment, measurements of overshoot and damping time can be made; see for instance [Fohanno].

1.4.3 Coordinate Measurement Machine

Perhaps the most obvious way to obtain three dimensional static position measurements at the required accuracy of $\pm 10\mu m$ is to use an industrial coordinate measurement machine of the type commonly used for the inspection of the di-

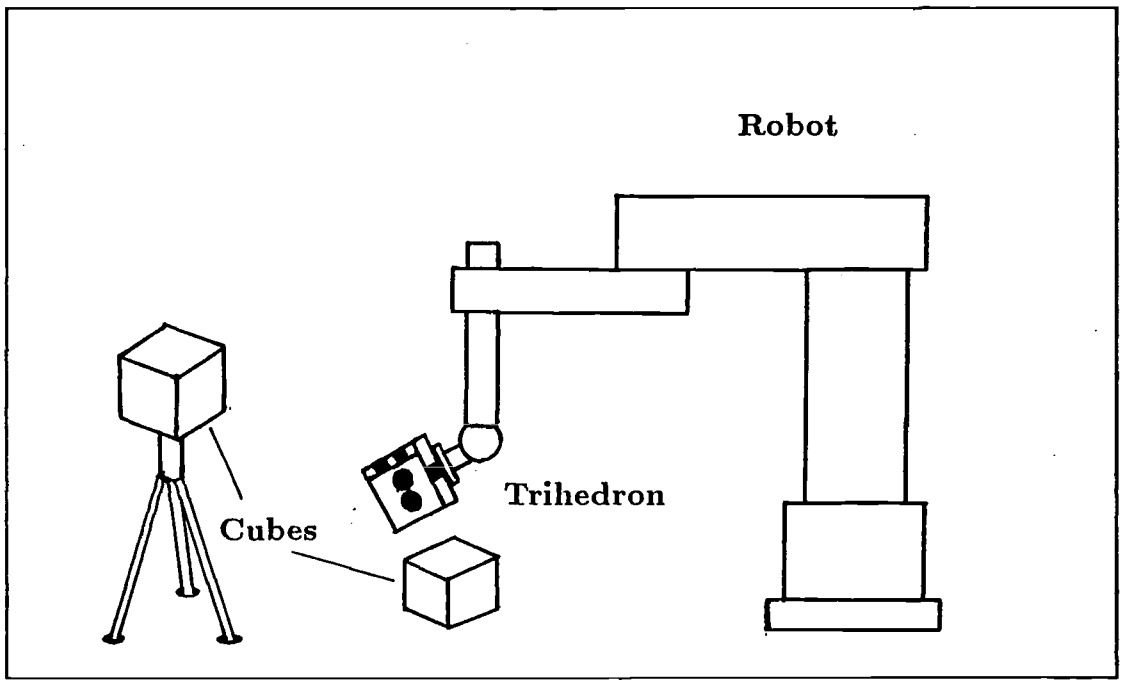


Figure 1.3: Trihedron and Proximity Sensors

mensions of manufactured parts. The nearest approach to this in large scale testing has been at the Institut für Produktionstechnik und Automatisierung (IPA), Stuttgart, where over 20 industrial robots have been tested using a prismatic three axis coordinate measuring instrument with an accuracy of $\pm 100\mu m$. See [Brodbeck and Schiele] and also [Warnecke and Brodbeck]. The arrangement of the test stand built at IPA is shown in Figure 1.4.

To avoid the effects of contact with the robot the three axis machine is fitted with a head incorporating three inductive sensors, mutually at right angles. To obtain a measurement of the robot's position, the three axis device is moved until the head sensors are positioned around a metal sphere fitted to the robot's end effector. The three axes are then adjusted until the inductive transducer readings indicate that the sphere is centred in the head. The coordinates of the three axis device then give the position of the centre of the sphere.

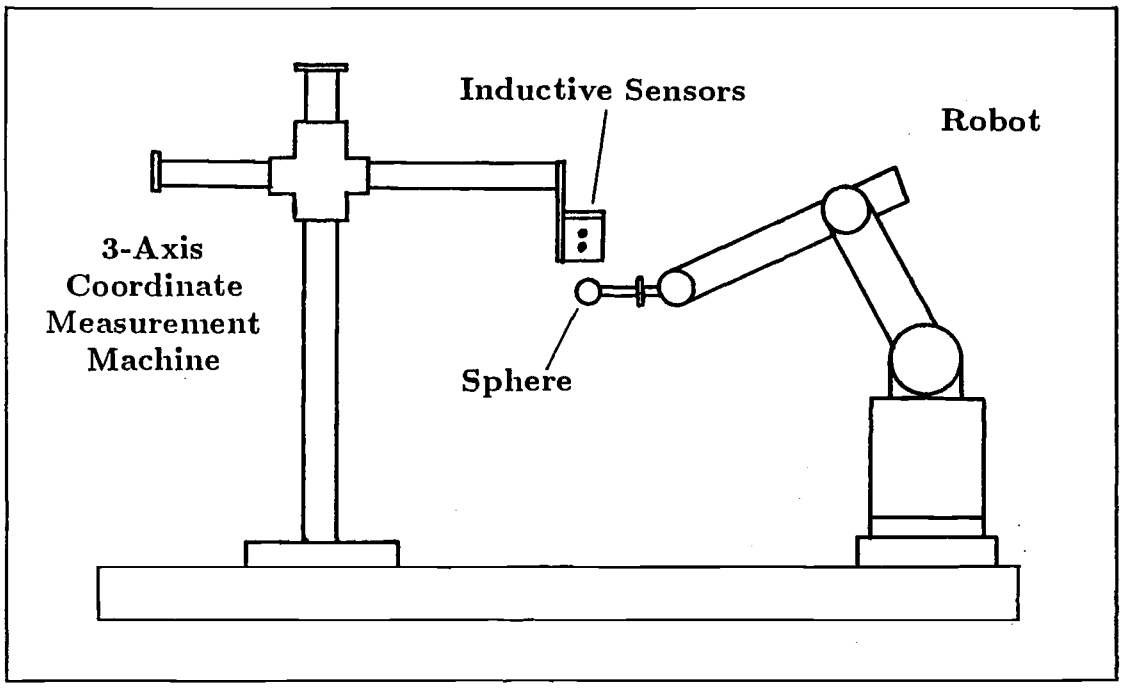


Figure 1.4: IPA Coordinate Measurement Test Stand

1.4.4 Structured Light

A variety of instruments which use structured light are discussed in the section on dynamic systems. One instrument which has been used for essentially static, non-contact measurement is the "Optocator", [Machinery] manufactured by Selcom AB of Sweden. This device, which has been used mounted on a robot to perform inspection tasks on car body components, [Brunk], has sufficient accuracy to be part of a measurement system for establishing the performance of the robot itself.

The instrument works as illustrated in Figure 1.5. A spot of laser light (modulated at $16kHz$) is projected onto the surface of interest. Scattered light from the surface is imaged onto a photosensitive surface and the position of the image spot determines the angular position of the surface spot. Since the original light beam emerged from the device at a known position and angle,

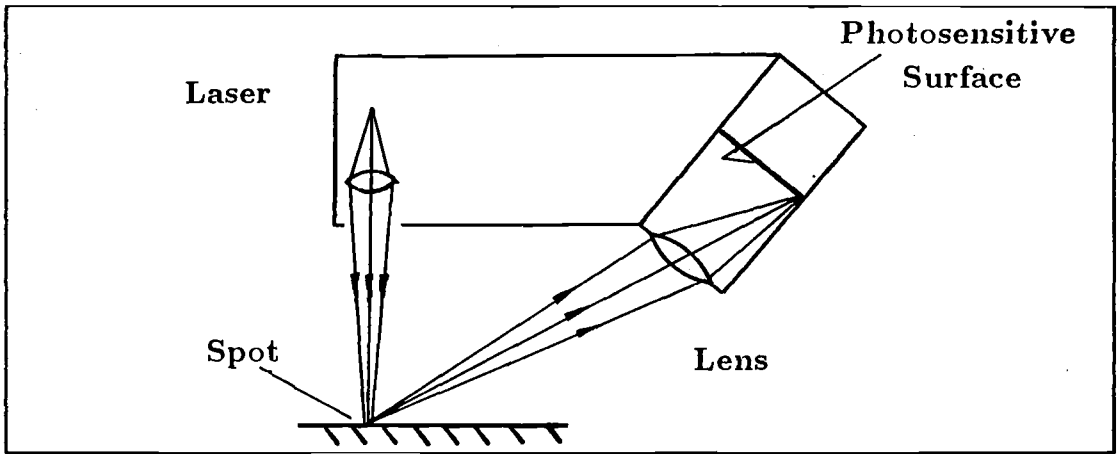


Figure 1.5: Optocator

the range to the surface can be calculated by triangulation. The system can be obtained with various measurement ranges, from $8mm$ to $512mm$ and has a resolution of 0.025% of range. The accuracy claimed by the manufacturers is $\pm 0.1\%$ of measurement range which, for the case of $8mm$ range, means $\pm 8\mu m$, [Selcom].

The accuracy of the system is comparable to that of eddy current proximity sensors, and is sufficient for use in repeatability tests. The size, weight ($0.7kg$) and cost of the device suggest that a multiple sensor, 6 degree of freedom, trihedron-type system based on this technique would not be practical with the device in its present form.

1.4.5 Theodolites

A non-contact static system, which has been used in practice at Renault, employs two theodolites to take bearings of azimuth and elevation to a target point on the robot. See Figure 1.6. In principle, with the relative position and orientation of the theodolites known, the cartesian coordinates of the target point can be calculated by triangulation. Renault do in fact use this technique to calibrate

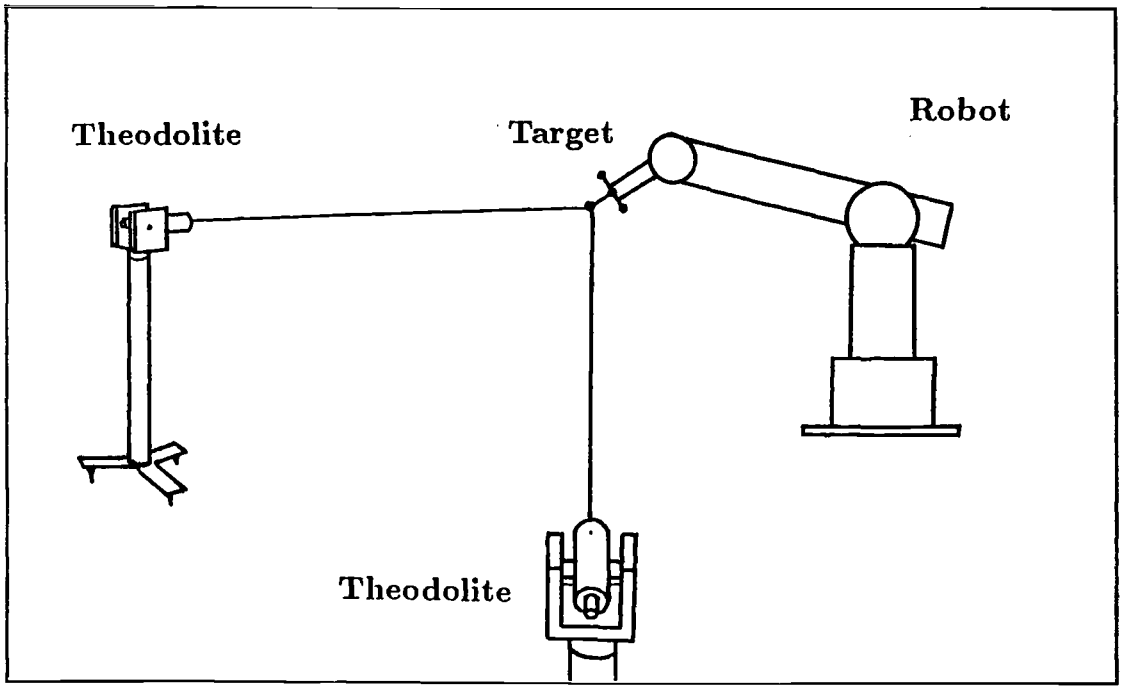


Figure 1.6: Theodolites

their own robots.

An evaluation of the two theodolite method has been carried out by J.P. Desmaret of Renault [Desmaret]. The target used was a set of five polystyrene spheres, $0.5mm$ in diameter arranged as shown in Figure 1.7. This arrangement keeps at least three of the spheres in view from the theodolites for most positions of the end-effector. The spheres appear as small spots of light, the light being transmitted to the spheres by optical fibres inside the mounting pins. Considerable effort is expended in the calibration of the theodolites themselves and in the calculation of their relative positions using rods of known size positioned in the field of view.

The azimuth and elevation reading from a theodolite defines a *line* in space which passes through the sighting point. The lines obtained from two theodolites will not in general intersect, however. The coordinates of the sighting point are

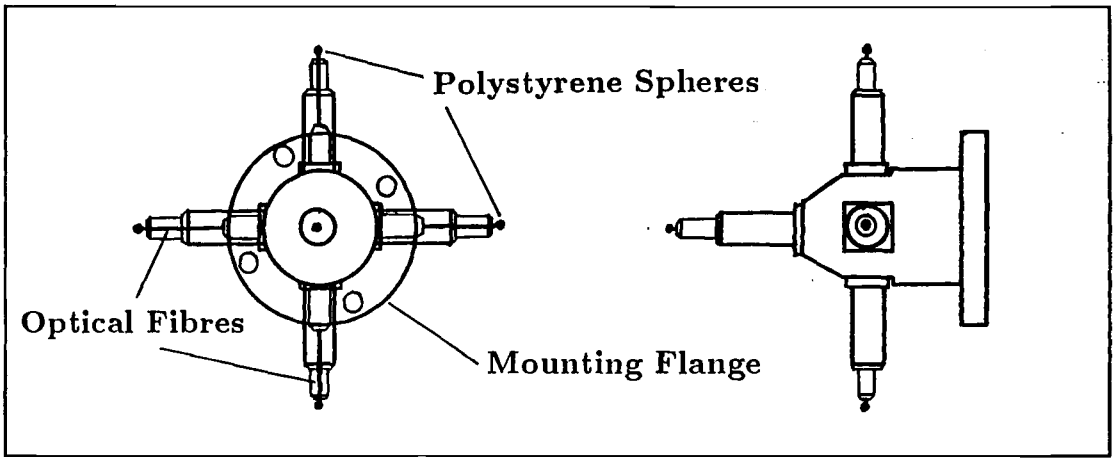


Figure 1.7: Theodolite 5 Point Target

therefore taken to be those calculated for the mid-point of the common normal to these two lines. Sets of precise grid lines were used to test the system and establish its accuracy when used at the range and with the field of view required for robot position measurement. The estimated accuracy of the system is given by Desmaret as $\pm 0.15mm$.

Several factors contribute to the inaccuracy in the system. It is completely dependent on the quality of the calibration procedure which is carried out each time the theodolites are set up. The stability of the instrument mountings and the number and precision of the standard rods used for set-up also affect subsequent readings, as does the quality and visibility of the target points. Finally the system is dependent on manual operation, the eyesight of the operator, and the inevitable subjective component in his alignment of the theodolite with its target.

1.4.6 Photogrammetry

At the UK National Engineering Laboratory the use of photogrammetry for making position measurements of robots has been explored, [Welsh]. A robot

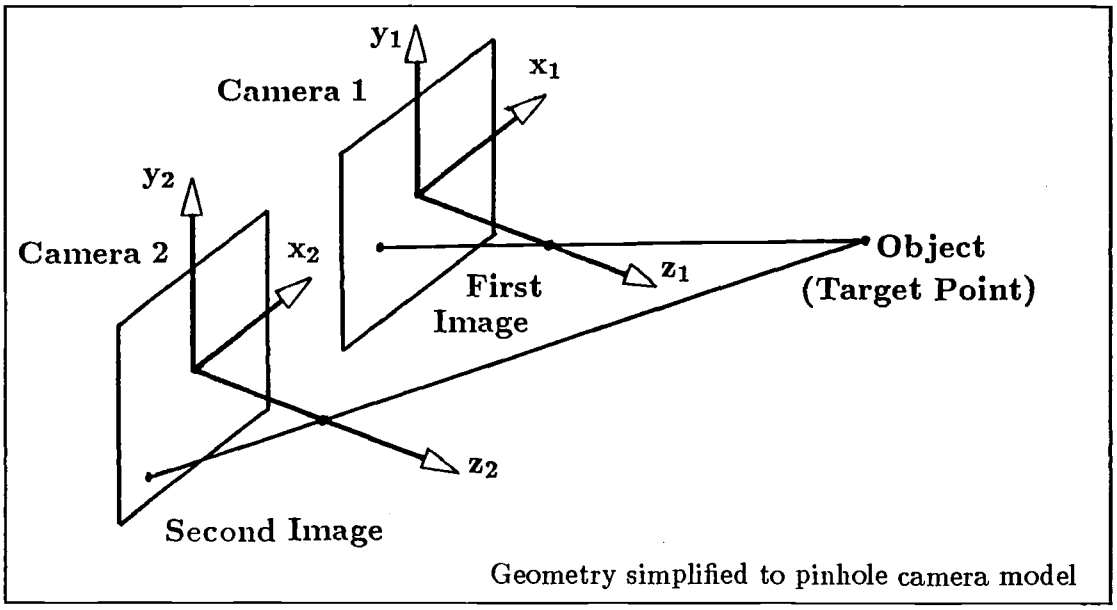


Figure 1.8: Photogrammetry

carrying markers or targets is photographed in various positions by two or more cameras. The film is developed and is then the subject of manual measurements to locate the coordinates of the targets as seen from each camera. As with the two theodolite system, the observed measurements define two lines in space which do not, in reality, intersect at the target point. The geometry is illustrated in Figure 1.8. With the NEL system these lines have generally been found to be between $10\mu m$ and $100\mu m$ apart at their common normal. The accuracy achievable using two cameras can be from 1 part in 10 000 up to 1 part in 25 000. With multiple cameras 1 part in 10^5 can be obtained.

The photogrammetric method shares a number of problems with the theodolite technique. These include the setting-up and calibration problems of the instruments, the sharpness of the target and the subjectivity of the manual measurements which are made. In addition, factors such as the film grain size and the accuracy of the equipment used to make measurements on the photo-

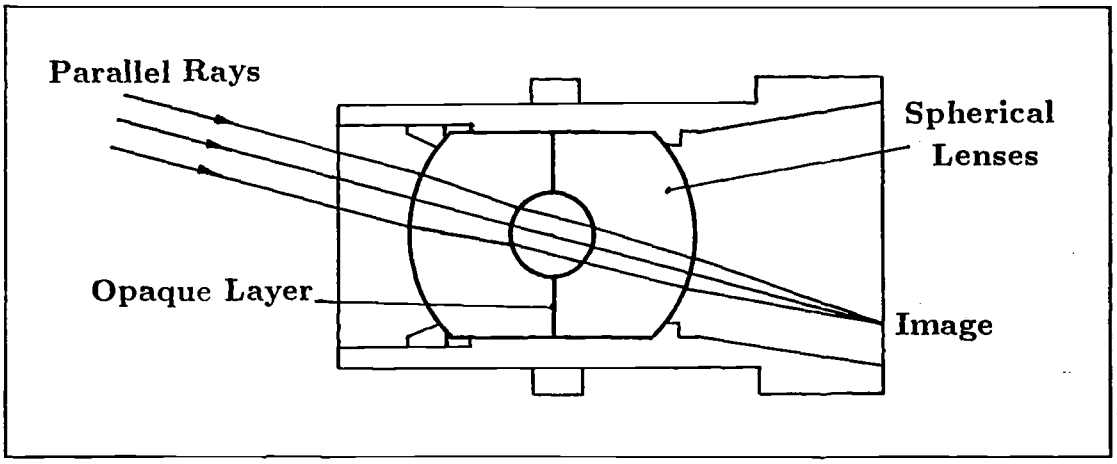


Figure 1.9: NPL Centrax Lens

graphic negative affect the overall accuracy. Other sources of inaccuracy specific to photogrammetry are focus error and the geometric aberrations present in the camera lenses.

These latter problems have been tackled at the UK National Physical Laboratory, where a new design of lens, the Centrax lens shown in Figure 1.9, has been developed specifically for photogrammetry [Burch and Forno (1)]. A point source of light in the object field is imaged by this lens as a circular spot surrounded by concentric rings. The lens has an large depth of field, from $300mm$ to infinity, and in theory is distortion-free. With a three camera system and a suitably accurate measuring microscope, it is anticipated that this lens will allow an accuracy of 1 part in 10^6, i.e. 1 micron across a 1 meter field, [Burch and Forno (2)]. To date, 2 in 10^6 has been achieved.

1.4.7 Fresnel Zone Plates

The most precise non-contact static system yet developed was conceived at the National Physical Laboratory, [Gates *et al.*], and has been used as a means of calibrating industrial three axis coordinate measurement machines. It consists

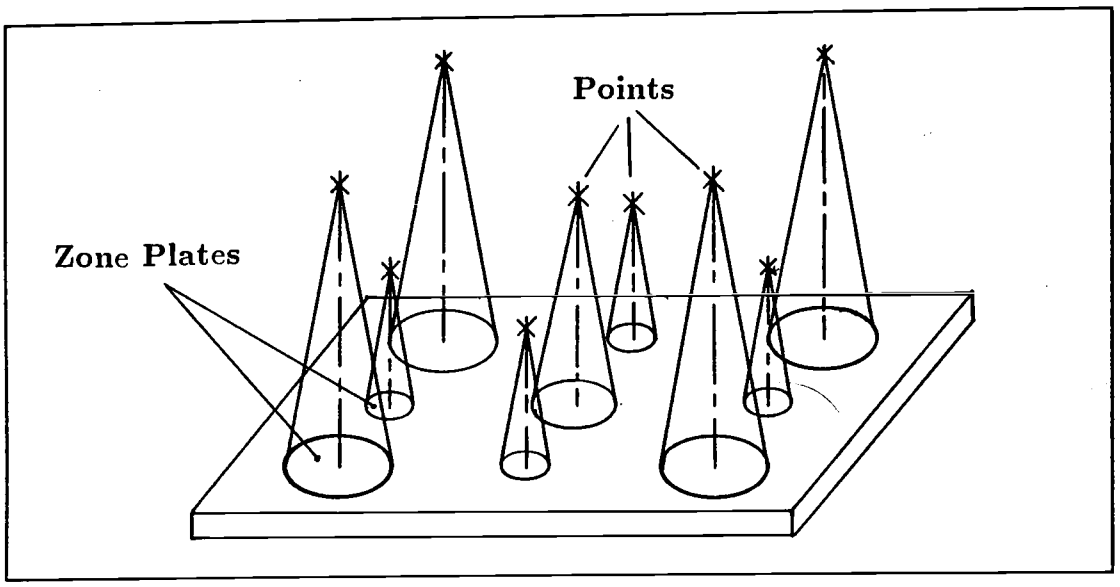


Figure 1.10: Fresnel Zone Plates

of a two dimensional array of Fresnel zone plates (Figure 1.10), which can be considered as the holographic equivalent of concave mirrors, and also an interferometer which is used to locate the unique point in space defined by each of these zone plates (Figure 1.11).

The interferometer is mounted on the moving part of the machine in place of its probe. For each zone plate there is a unique position at which the fringe pattern of the interferometer is a uniform field. Lateral errors from this position introduce parallel fringes and axial errors introduce curved fringes. The system is more sensitive to lateral displacement than to axial displacement, one fringe being introduced for $3\mu m$ laterally and $60\mu m$ axially. Using photoelectric sensors, a sensitivity of one tenth of a fringe, i.e. $0.3\mu m$ laterally and $6\mu m$ axially has been obtained with the system. Later developments have attained an axial sensitivity of $1.5\mu m$, and so the technique has approached its original design aim of $1\mu m$ in all three directions.

The development of this system and another optical system employing mir-

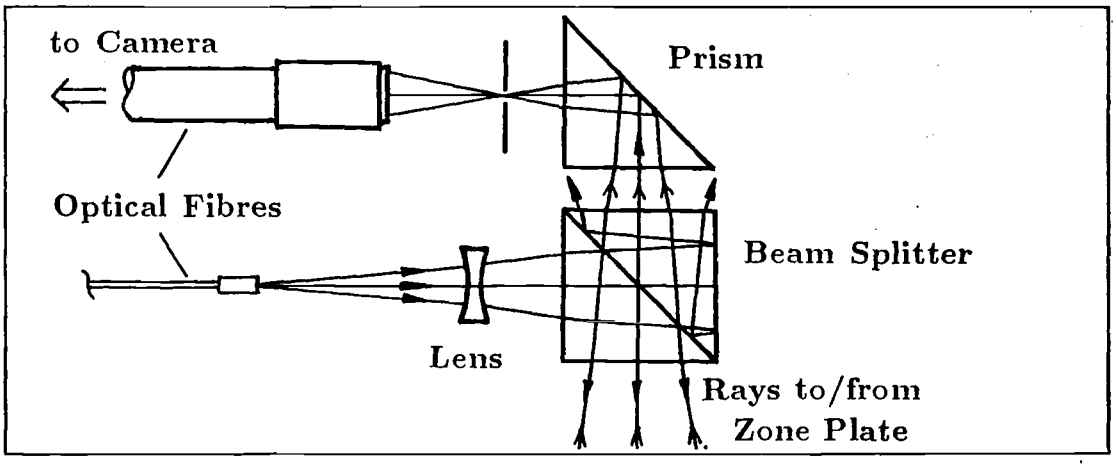


Figure 1.11: Fresnel Zone Plate Interferometer

rors to define the points in space have been described in detail by R.F. Stevens in his thesis [Stevens (1)]. See also the NPL publications [Stevens (2)] and [Stevens (3)]. Apart from their remarkably high precision, these systems have the advantage that they place no material in the working volume being used, and the calibration points can be approached from any direction by the sensing head.

1.4.8 Discussion of Static Systems

In general it can be seen that the requirements for static tests can be met with existing instrumentation, some examples of which are described above. Leaving aside considerations of the cost of the instruments and their associated electronics and computing equipment, different systems have different technical features with their own particular advantages and disadvantages.

A system which requires contact to be made with the robot is inappropriate, as has been shown, but there is no need to make physical contact. Non-contact systems based on a variety of principles and giving the required accuracy are available.

One attribute of these various techniques which can be compared is their adaptability for different circumstances. Some of the systems described require that only measurements at predetermined fixed points in the work envelope can be made. The presence of fixtures of one kind or another is common to the trihedron, Optocator and Fresnel zone plate methods. To make measurements at a different selection of points, these fixtures must be carefully repositioned. This constraint does not apply to the theodolite or photogrammetric methods.

The problem of positioning fixtures gets worse when tests for accuracy (not just repeatability) are being carried out. For instance, the positioning of the reference cubes for the trihedron method is not important for repeatability tests but must be done with great precision prior to tests of robot accuracy. Similarly for the Fresnel Zone Plates, in which the defined points in space are fixed only with respect to the system's base plate. The advantages of this latter system over the trihedron method are that the end effector can approach a single point from any direction, since there is no actual material (corresponding to the reference cubes) present, and that the accuracy of length measurements between well separated pairs of known points can be obtained.

In contrast, movable fixtures are not required in photogrammetry, which needs only a few fixed, known reference points to be in the field of view. With this condition met, the robot can be moved anywhere in the working volume (as long as the targets are visible to the cameras), and the required measurements made.

Another factor which enters into a comparison of different systems is portability. The IPA coordinate measurement system is massive and is clearly not intended as a portable instrument. Some of the other systems could be used in the field, with more or less difficulty, for robot recalibration purposes.

The automatic capture and processing of data is another feature which must be considered. This has received considerable attention with the trihedron, eddy current sensor based system. Theodolite based systems also lend themselves to the automatic collection and processing of measurement data. Photogrammetry, however, is at a disadvantage due to the necessary time lapse for film processing.

It is of some interest to consider the potential of static measurement methods for modification or extension in some way, to meet the more demanding requirements of dynamic measurement. As will be seen in the next section, on dynamic systems, attempts have been made to do this with a number of different instruments.

The natural extension of the eddy current proximity sensor method is to monitor the proximity of the robot end effector to a rail fixed at a known position in the work space. In this way the robot's errors from true straight line motion (for instance) can be observed. Photogrammetry has also been used dynamically, using multiple exposures of the film as the robot moves. To use theodolites dynamically would require a servo mechanism to track the moving target point. Some servoed dynamic systems operating on a similar principle have in fact been built. Developments such as these tend to carry their original disadvantages with them, for example, the film development time in photogrammetry or the inflexibility of the fixtures required for a proximity sensing system. New problems also emerge, usually in connection with moving parts and servo mechanisms in the instrument itself.

Conversely, if a suitable instrument could be developed for dynamic applications, then in meeting the demanding requirements for accuracy, non-contact operation, data acquisition frequency and automatic processing

of the captured data, the instrument would also have met and surpassed the static measurement requirements. Some procedures could be speeded up considerably. The large number of point position measurements required for robot calibration could be obtained by dynamic monitoring of the robot over trajectories through its work envelope. Thus only one type of instrument would be required for position measurement in robot testing.

1.5 Review of Dynamic Systems

In this section an assessment is made of the characteristics of various systems which have been used for dynamic three dimensional position measurement. Some of these techniques have been developed specifically for dynamic measurement in robotics. Others have their origin in quite different fields, such as navigation, inspection, remote sensing and computer graphics. Some were originally static techniques which are being developed or which offer promise for development for dynamic use.

1.5.1 Accelerometers and Inertial Systems

The cost, size, weight and inaccuracy of inertial systems has mitigated against them in industrial applications. A simplified "strapdown" system of accelerometers in which orientation is calculated rather than maintained constant by gyroscopes is half the cost and a quarter the size of a stabilised platform, but weighs approximately $6kg$. The principal use of inertial systems has been for position measurement in aircraft and submarines, over long distances, but in situations where external references of position such as land marks can be used to reset the system. Inertial navigation systems, as used in aircraft, perform a useful function even with an accuracy which locates the aircraft to within 1 kilometer

of its true position.

In mechanical inertial systems the basic measured quantity is force. Inertial platforms consist of three accelerometers aligned normal to each other and supported in a gyroscope stabilised gimbal system. The accelerometers typically have a measurement error of force (and so of acceleration) of 1 part in 10^5. The position data is obtained by integrating first acceleration and then velocity with respect to time. Thus the error in inertial systems increases with time from the last occasion on which it was corrected to some external reference. In circumstances which allow the instrument to be zeroed at shorter time intervals, the cumulative error is reduced.

A Ferranti system used in three dimensional land surveying now offers an error of less than $\pm 25cm$ from the correct location. Another Ferranti inertial platform has been developed for plotting the three dimensional path of the boreholes of oil wells. In this case the inaccuracy is less than $\pm 5cm$. This is achieved by stopping the platform in the borehole at 10 minute intervals to correct the calculated velocity to zero.

The precision which can be maintained even over short time intervals does not meet the criteria for robot performance testing, although it would appear to be accurate enough to be used to monitor the performance of an Autonomous Guided Vehicle or to form part of the vehicle's own sensing system. The requirements of metrology in robotics put more stringent demands on inertial systems than those imposed by military applications. This is true not only for the accuracy of position measurement. The linear accelerations and angular velocities of a robot end-effector can exceed those of a fighter aircraft. Finally, it is doubtful whether the frequency response of accelerometers is good enough to monitor the dynamic behaviour of the robots.

1.5.2 Proximity Sensors

As mentioned in the section on static systems, eddy current proximity sensors can be used to measure local dynamic effects such as corner rounding, overshoot and settling times after arriving at the target position. These sensors can also be used to obtain measurements (to $\pm 10\mu m$) of the departure of an end-effector from longer programmed trajectories, [Warnecke and Brodbeck]. In this type of test the sensors are mounted on a fixture on the robot end-effector, which is then programmed (for instance) to move in a straight line path along a metal bar. The eddy current sensors then give data on the varying distances from, and orientation to the straight bar. The necessity for fixtures to define every tested trajectory is a severe disadvantage.

1.5.3 Tensioned Wires

At Peugeot [Folhanno] and at Unimation, attempts have been made to obtain dynamic performance measurements of robots using a system of tensioned wires. See Figure 1.12. In this system a number of wires from separate dispensers in known positions are attached to "a point" on the robot arm. The lengths of wire paid out are measured at each dispenser and by trigonometry the position (x, y, z) of the point can be calculated. This assumes that the wires are on straight lines and that these lines all pass through one common point at the robot. It is also assumed that each of these lines consistently passes through the same point where the wire emerges from its dispenser.

The resolution of the wire length measurement depends mainly on the resolution of the transducer (e.g. optical shaft encoder) and the length of wire paid out per revolution of the transducer. For example, wire wound on $40mm$ diameter drum with a 4000 pulse per revolution shaft encoder has a tolerance of

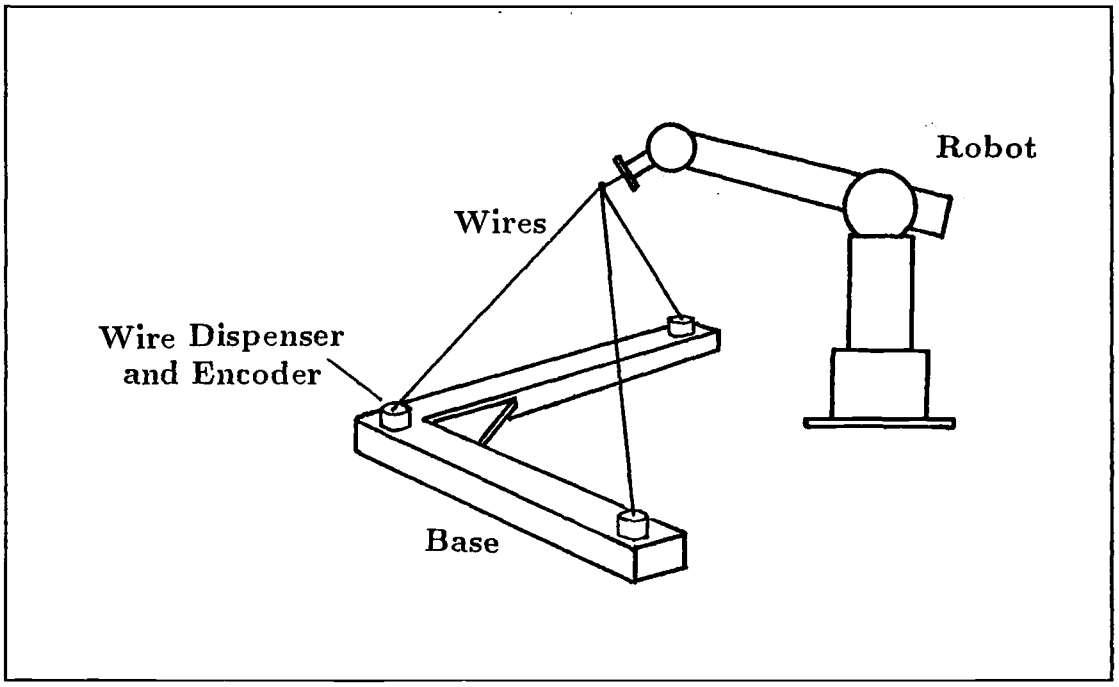


Figure 1.12: Tensioned Wires

$\pm 16\mu m$ on nominal length, ignoring slipping on the drum.

The wire length resolution is not the same as the coordinate resolution however, the latter varying as a function of the coordinates. The transformation from wire lengths to cartesian coordinates can be ill-conditioned. This may be appreciated if one considers a point near the xy plane in which the dispensers are located. A change in z here causes negligible change in wire lengths, so resolution is poor. To guard against this circumstance, either the plane of the measuring points must be well outside the robot working volume, or redundant wires must be added.

Another source of varying resolution of coordinate measurement, and of inaccuracy if this is not taken into account, is the sagging of the wires under gravity. This also varies with (x, y, z). Matters could be improved by varying the wire tension to reduce sagging appropriately in different locations but this introduces

other errors. As tension increases, so the wire stretches, invalidating the length measurement. Tension in the wire also causes some deflection in the robot arm, altering the position it was desired to measure. These errors are likely to be significant with respect to the design aim of $\pm 0.01mm$ accuracy. The system used at Unimation has a constant tension of approximately $15N$ in each wire, and so the force applied to the end-effector by this measurement system can be greater than the maximum load capacity of a machine such as the Puma 560.

In addition, there are other problems. Wire and drum dimensions will vary, not necessarily uniformly, with temperature. Moving parts in the wire dispensers and the wires themselves will be susceptible to vibration, and the wire length readings are subject to random disturbances due to air movement. Finally, the wires restrict the volume in which measurements can be made (without fouling the robot).

It is unlikely that this approach could achieve the desired accuracy of $\pm 10\mu m$.

1.5.4 Multiple Exposure Photogrammetry

At the National Engineering Laboratory the trajectories of a moving robot have been plotted using multiple exposure photogrammetry, [Welsh]. This technique involves illuminating the robot with a number of flashes of light as it moves. Each target appears on the photographs in a series of positions evenly spaced in time. From the two photographs, the three dimensional trajectory can be obtained.

In the dynamic situation a number of practical problems have been overcome. To obtain the requisite close spacing of recorded points on the trajectory, between 150 and 200 flashes are used for each photograph. To see the target steel balls against the overexposed background, they are coated with a highly

reflective paint. To obtain the correct correspondence between targets in each photograph, every tenth flash is made more powerful than the rest.

The accuracy of this approach is similar to the accuracy of photogrammetry used statically. A working volume of $4m \times 4m \times 3m$ deep was viewed, and a precision of location of the targets of less than $0.5mm$ was obtained (i.e. 1 part in more than 8000).

1.5.5 Stereo Vision

Stereo vision in this context is the dynamic analogy to photogrammetry, and it uses two video cameras to survey a scene. The principal driving force behind stereo vision research has been to obtain a machine vision system for robots. Such a system would, ideally, use sensors to acquire a depth map of the scene, construct models of the objects in the scene, and match these models to those of which it already has some knowledge. This information can then be used to plan trajectories for grasping or for navigating past objects. Image analysis, geometric modelling and task planning are, of course, extensive areas of research in their own right. The sensor which provides the information on which all the subsequent modelling is based is however doing the same job as other three dimensional dynamic position measurement systems, in that it is acquiring the spatial coordinates of points in a scene. It is therefore of interest to establish the advantages and limitations of stereo vision, with a view to using the technique for metrological purposes.

The geometry of a stereo vision system is similar to that of photogrammetry as shown in Figure 1.8. In monitoring a robot's trajectory, only particular marked points on the robot need be matched in the two images. Thus the stereo correspondence problem is much alleviated. With a cluster of illuminated targets

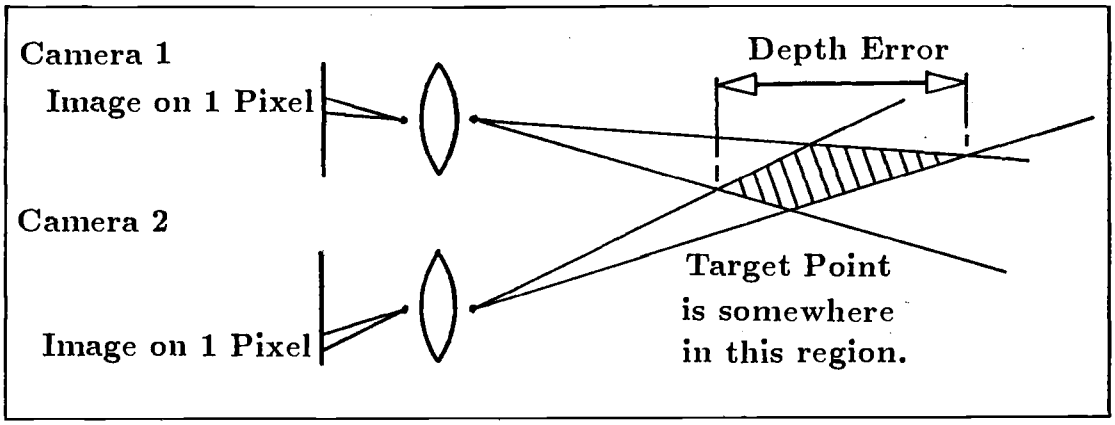


Figure 1.13: Depth Error

(as used in the theodolite method) the corresponding points in each image would be clearly visible and well defined, and so the difficulties of occlusion of matching points could also be minimised.

One difficulty which arises is the poor resolution of video cameras. Each camera image is focused on a two dimensional array of sensing elements. A Charge Coupled Device (CCD) camera might have 512×512 such picture elements or "pixels". Thus in such a camera viewing a scene $2m$ across, it would appear that one pixel represents approximately $4mm$ across the field of view. The consequent error in the calculated depth does of course depend the angle of convergence of the "rays" to each camera. This in turn depends on camera separation and the range to the point, see Figure 1.13.

In experiments to obtain depth maps using two video cameras with 188×244 light sensitive elements, [Yakimovsky *et al.*] obtained a three dimensional resolution of $\pm 5mm$ at a range of $2m$. A factor of ten improvement on this resolution would still leave it a long way from giving the performance required. Even if resolution (and accuracy) could be sufficiently improved, a video rate of $30Hz$ is too slow for dynamic performance measurement of robots.

1.5.6 Selspot

Selspot is a two dimensional imaging system [Selcom] in which the position of images of LEDs (or infrared or laser sources) is detected on a photosensitive solid state device which lies on the image plane of the lens. The position of each point in the image defines a line in space, on which the LED must lie. Two or more of these camera-like instruments are used together and the intersection of two or more lines is used to calculate the cartesian coordinates of the LED. The system is similar in principle to stereo vision, the difference being that the photosensitive detector has a resolution of 1 in 4000. Unfortunately the accuracy of the system is poor, errors being as high as $\pm 0.5\%$ of the measuring range, i.e. 1 part in 200. Such a system, viewing a $3m$ diameter robot working volume would give an error of $\pm 15mm$ on end effector position. The maximum sampling rate for one LED is $10kHz$ which is high, although this decreases as more LEDs are monitored.

1.5.7 Twinkle Box

The Twinkle Box is an optical three dimensional position measurement system [Burton and Sutherland] which was intended as a computer graphics input device. The instrument monitors the position of light emitting diodes (LEDs) which are fixed to the moving object. As with other techniques, an image of the scene is produced with lenses on each instrument. The notable feature in this system is that the image, which is $35mm$ across, is scanned mechanically by narrow slits ($0.3mm$ across), a pulse being detected by a photomultiplier as each slit crosses the image of an LED. The components of a detector are shown in Figure 1.14. The time of each received pulse corresponds to a particular angular position of that slit on the rotating disk, and so locates the LED as lying on

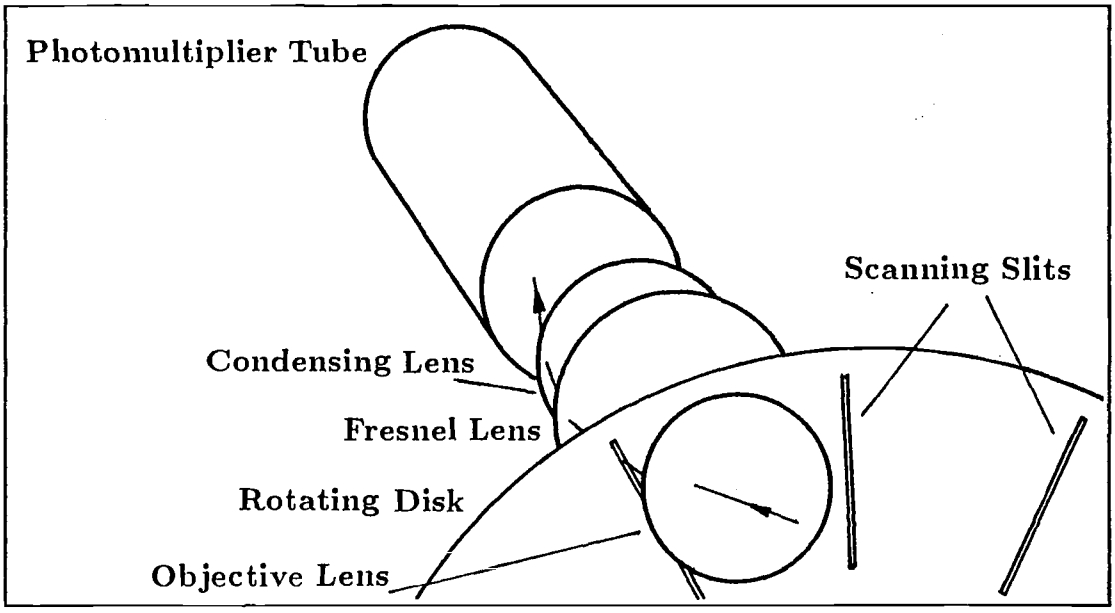


Figure 1.14: The Twinkle Box

a particular plane in space. Other similar detectors determine different planes for the same LED, and the intersection of three such planes gives the cartesian coordinates of the diode.

The position of many diodes can be monitored in this way, by switching the diodes on one at a time. Doing this, there is no ambiguity as to which LED has been sensed during the scan of one slit over the image. With 32 radial slits in the disk and an angular velocity of $3500rpm$, a scan frequency of $1900Hz$ is obtained. This must of course be divided by the number of LEDs being monitored.

The problems incurred with this system include vibration due to the rotation of the disk, errors in the positions of slits on the disk, geometric aberrations in the optics, and errors in the pulse times registered by the photomultipliers. In tests to establish the accuracy of the system, LEDs were moved known distances in the working volume. The standard deviation of the error from zero was found to be $7.3mm$.

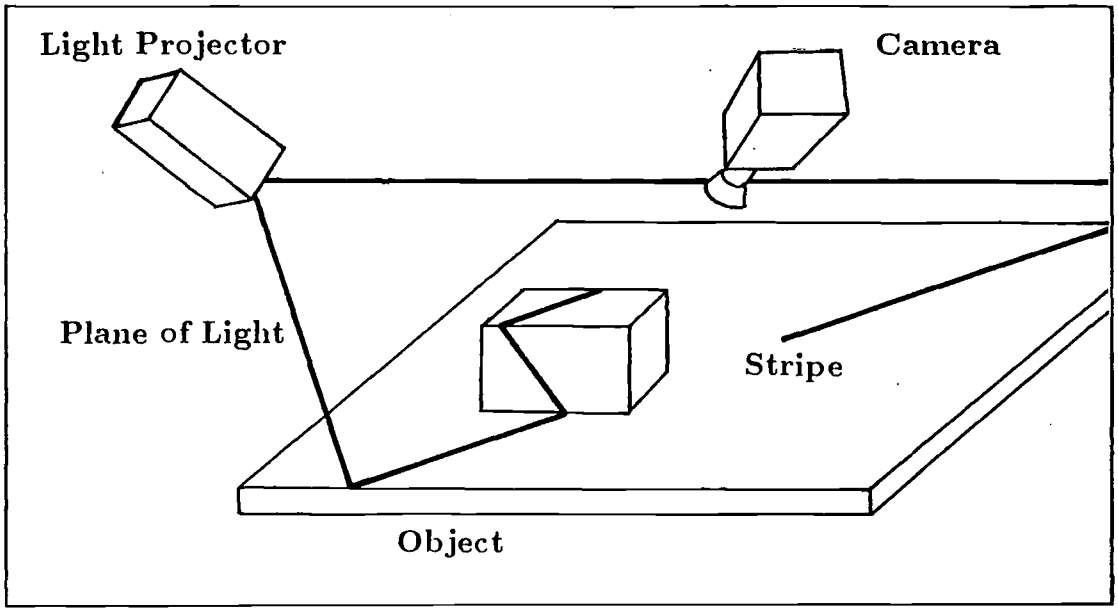


Figure 1.15: Planar Structured Light

1.5.8 Structured Light

In a structured light system, a known pattern of light is projected onto a scene and the consequent illumination of the scene is used to obtain information about the position and orientation of points in the scene. The calculation of the cartesian coordinates of the points of interest in the scene is similar to the calculation carried out for stereo, i.e. by obtaining the intersection of lines and planes in space. One use of structured light has been to obtain geometric data on objects in a scene so that this data can be matched to geometric models of the objects.

Many structured light systems have been built. At the University of Edinburgh [Popplestone *et al.*] the device projected a plane of light and the resulting stripe in the scene was viewed with a video camera. See Figure 1.15. A system which used a point of light [Faugeras] and imaged the scattered light from the object surface using several cameras is shown in Figure 1.16. No performance data is available for either of these systems.

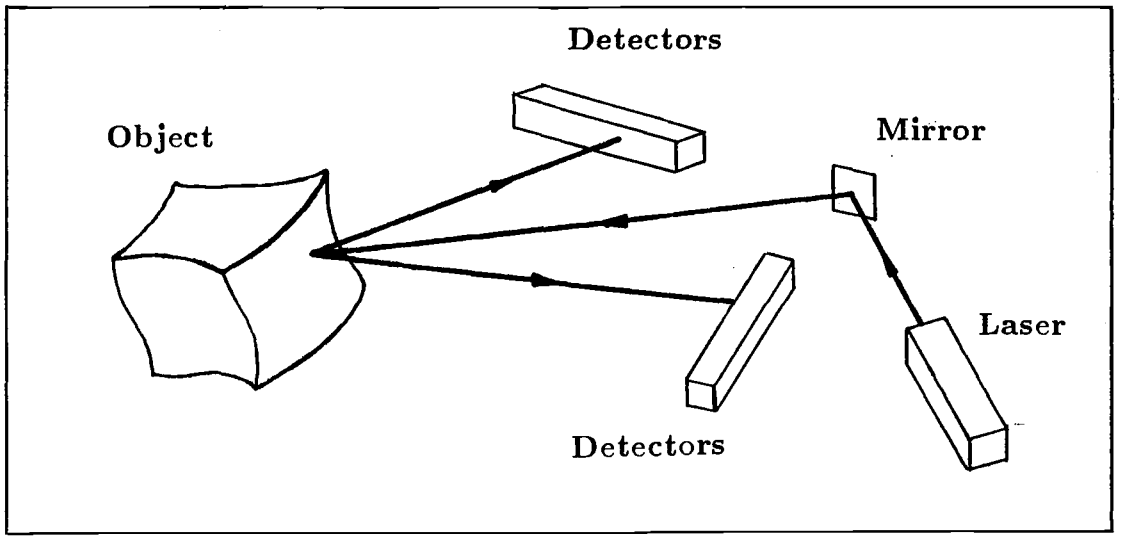


Figure 1.16: Single Point Structured Light System with Imaging Detector

Another system which uses a single laser beam, [Pipitone and Marshall], does not use an imaging system to determine the angular position of the illuminated spot. Instead, a scanning directional detector is used (see Figure 1.17.). In this detector a mirror scans the scene and, in the correct orientation, reflects light from the illuminated spot through a slit to a photomultiplier. The coordinates of the spot can then be found by triangulation. The average data acquisition rate over many scan lines was $62Hz$ and the root mean square (RMS) error of the system in depth measurement is given as $6.4mm$ at $2.438m$

A more complex system [Altschuler *et al.*] uses a rectangular array of laser beams to produce a pattern of dots on the scene. The pattern can be varied using a programmable electro-optic shutter which allows particular columns of beams to be transmitted. By controlling the optics of the projector, the scale of the beam pattern can be varied, giving coverage of small or large objects as required. With this system a large number of points can be determined in each frame scan of the video camera which views the scene. The calculation of point

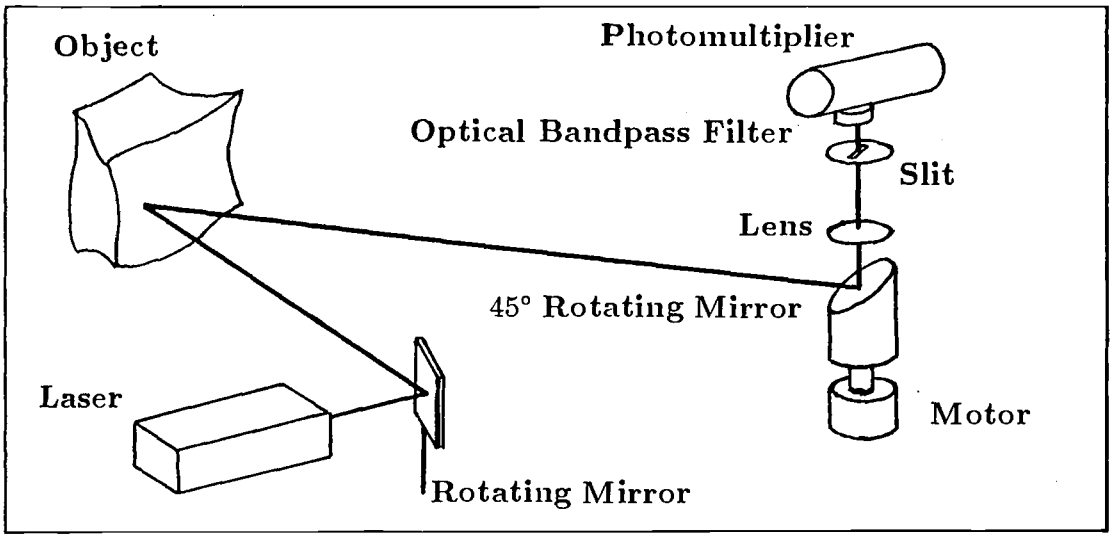


Figure 1.17: Single Point Structured Light System with Scanning Detector

coordinates in this case requires the intersection of two lines for each point.

For an instrument using the above principles to be used for monitoring the position of a *specific* moving point, the spot of light would have to be held continuously on the moving target, steered by a servo system. Potentially, a sheet or point of light can be scanned over a scene with good accuracy. However, if the illuminated scene is viewed with a video camera then, as in the case of stereo vision, the resolution of the camera is a major restriction on accuracy, and the frame rate is a restriction on use in dynamic situations.

Structured light can be used in a different way, which does not involve a video camera or any imaging system. In the CODA-3 system manufactured by [Movement Techniques], planar sheets of light are repeatedly scanned over a scene, using octagonal mirrors rotating at constant velocity. The *time* at which a returning reflected signal (from a retroreflector fitted to the moving object) is received is used to give the angular position of the retroreflector at that moment. It is actually rather more complicated than this however, and is illustrated in

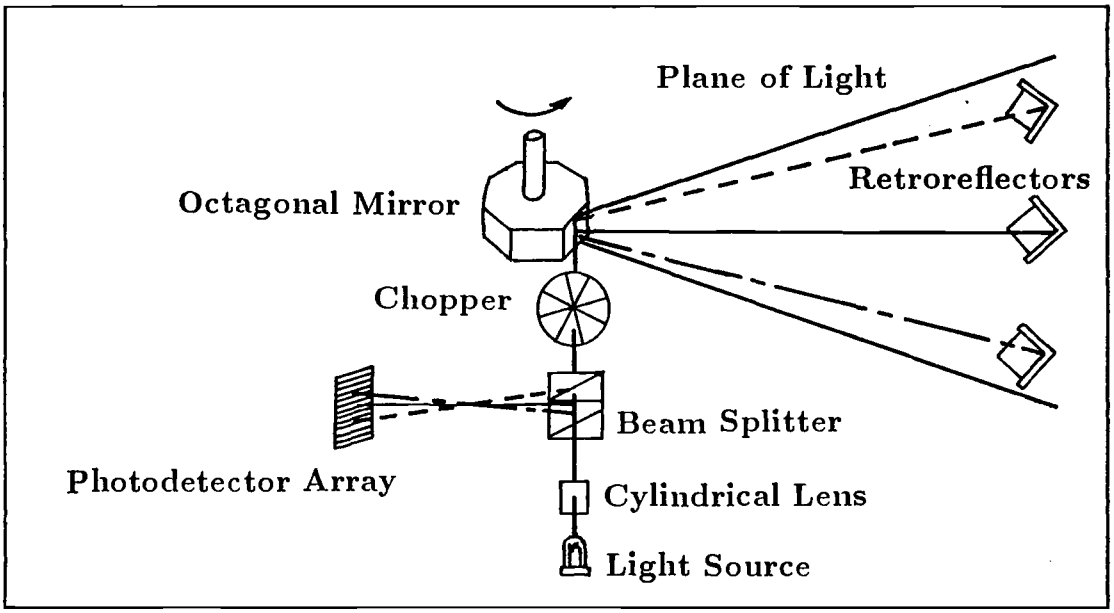


Figure 1.18: "CODA-3"

Figure 1.18. The beam of light returning from a retroreflector does so at an angle (on the plane of light) which is dependent on the position of the retroreflector (on that plane). After being reflected again from a half-silvered mirror inside the instrument the beam strikes a linear array of 128 photodetectors at a position dependent on that returning angle. Thus a vertical sheet of light, scanning about a vertical axis, can return several signals from different retroreflectors simultaneously, and register them on different elements of the array. In addition to this, the output light is "chopped" by a rotating disk made up of different coloured filters. The retroreflectors are also coloured and so the amplitude of a return beam indicates which target it came from.

By intersecting three such planes of light, the position of the target can be calculated. There is, of course, a skew on the times at which each scanner "sees" a particular retroreflector. The makers of this system claim to have overcome this problem in software, and give the performance of the system at $3m$ range

to be $0.3mm$ resolution in the x and y directions and $0.8mm$ resolution in the z (depth) direction. This is at a sampling frequency of $300Hz$.

1.5.9 Moiré Photography

This method is considered here more for its future potential than for an immediate practical application. The projection moiré method is effectively a structured light system. In the conventional method, light from a projector passes through a diffraction grating and a lens, producing a fan of planar sheets of light which illuminate an object, as in Figure 1.19. The object is then viewed by a camera through a similar grating. The difference term of the beating spatial frequencies thus produced is recorded by the camera as fringes, which correspond to contour lines on the object. In special circumstances the contour surfaces which cut the object are equispaced parallel planes normal to the line which bisects the angle (at the object) between the projector and the camera. This is not true in general however, and the contour surfaces are usually second degree surfaces and not equispaced. See [Doty] for an analysis of the geometry of contour fringes.

A non-planar contour geometry requires rather more calculation in the process of analysing the fringe pattern. There are two more serious problems with this method however. Counting fringes in the image allows the relative depth of points on the image to be calculated, but only if peaks can be distinguished from troughs on the object. There is not sufficient information on a single stationary image to do this, using the above method, and various techniques have been developed to overcome the problem. These include imposing known movements on the object or the instrument, or moving one of the gratings to shift its phase. In a dynamic situation where the movement of the object (viewed with a video camera) is one of the things it is desired to find, these modifications are not

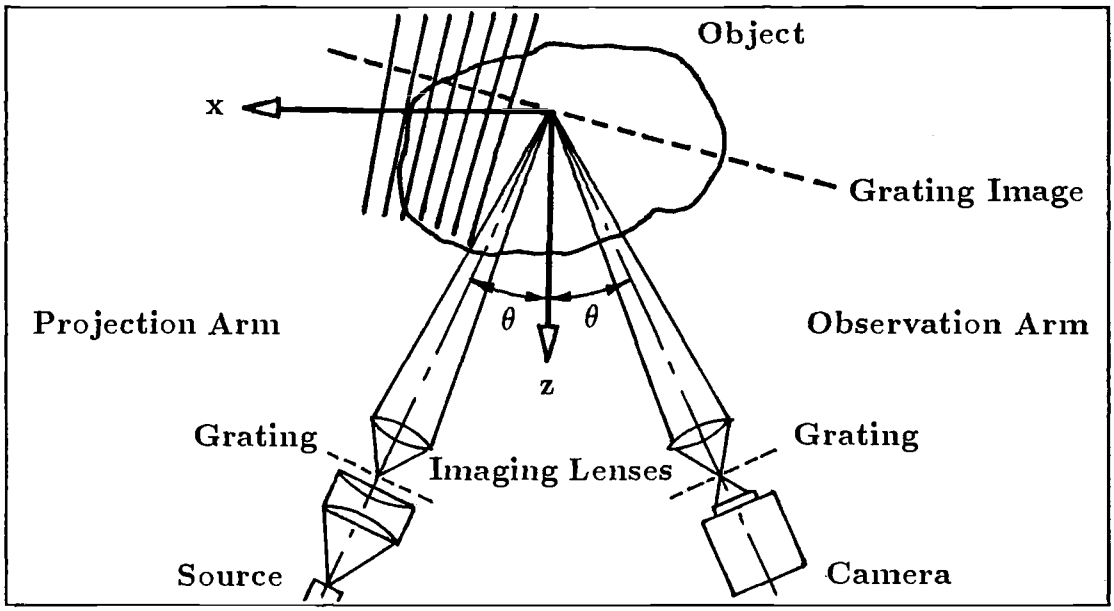


Figure 1.19: Projection Moiré

practical. Another approach, by [Idesawa *et al.*], is the scanning moire method, in which the scan lines of an imaging device are used as a virtual grating in place of the real grating in the viewing system.

The second problem is one of relating the fringe pattern to a coordinate system. Without some means of labelling the contour lines or placing some reference point of known position in the scene, the depth of points in the scene cannot be found. Similarly, breaks in the pattern can occur, making it impossible to tell even the relative depth of one object in the scene with respect to another.

If the above problems can be overcome, and other restrictions such as the depth of focus of the optics are acceptable, then a system can be conceived in which the coordinates of identified points in the scene may be obtained. Each fringe picks out a particular contour in the scene (i.e. a surface intersecting the object), and the coordinates of the point of interest (on that contour) in the camera image define a line in space which cuts the surface. Potentially, with a

sufficiently high resolution camera, the depth of points of interest can be found to the same resolution as the fringes. However, the adaptation of moiré techniques to a dynamic situation can incur similar problems to those of other techniques employing 2-D imaging systems. Once again the resolution and frame rate of the imaging device (e.g. a video camera) are factors limiting the overall resolution and bandwidth of the system.

1.5.10 Ultrasonics

Two main types of system employing ultrasound have been developed for three dimensional position measurement. In one of these, the time of flight of a pulse, travelling between transmitter and receiver is measured. Knowing the velocity of sound in the medium, the distance travelled can be calculated. The second type of system uses a continuous wave over the distance in question and measures phase shift as the distance changes.

Ultrasonics, Time of Flight

An example of ultrasonic position sensing by time of flight is the Lincoln Wand, developed at MIT as a computer graphics input device, [Roberts]. The system had four transmitters in fixed positions at the corners of a rectangle and a receiver (the "wand") whose position was determined as it moved in a working volume $1.8m \times 1.2m \times 1.2m$, see Figure 1.20. The transmitters output in turn at $10ms$ intervals a $20\mu s$ burst of energy with frequencies in the range $20kHz$ to $100kHz$. A complete set of readings from all four transmitters thus required $40ms$, limiting the operating frequency of the system to $25Hz$. The system used a $1MHz$ clock which, with a sound velocity of approximately $300m/s$, would allow a maximum resolution of approximately $0.3mm$ on the distance measurements.

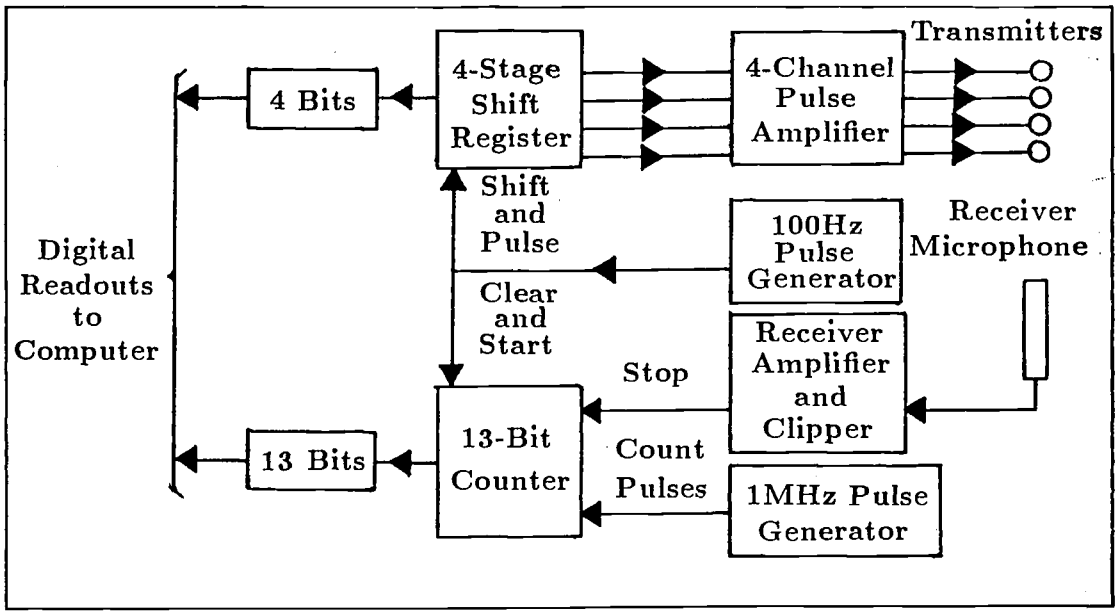


Figure 1.20: The Lincoln Wand

A 13 bit counter allowed distances of up to approximately $2.5m$ to be logged.

Only three distance measurements are needed to calculate the cartesian co-ordinates of the receiver, but the use of four transmitters allows a confidence check to be made on the readings. The problems experienced with this system included reflections at the end of one cycle of transmissions being taken as the first pulse of the next cycle, room noise (particularly from typewriters) being picked up by the receiver, the receiver itself being obscured from transmitters by some other object in the working volume, and air movement or changes in air temperature changing the sound velocity. Taking errors into account, the resolution claimed for this system was $0.51mm$. Accuracy was not verified but was estimated to be $\pm 5mm$ for the experimental set-up being used.

Presumably the resolution and accuracy of each distance measurement could have been improved by using a faster clock and by calibrating the system. However, for dynamic position measurement of robots, this type of instrument (with

four ultrasonic pulse transmitters and one receiver) has serious problems which are rooted in the low velocity of sound. The operating frequency ($25Hz$) is a direct result of this, and is far too low for the monitoring of the trajectory of a robot end effector, even one travelling as slowly as $1m/s$, for which a complete set of coordinates would be obtained at $40mm$ intervals along the trajectory. Worse than this, is the time skew on the signals received from each transmitter, the transmitters operating in turn at $10ms$ intervals within the $40ms$ cycle time. The four distances obtained each refer to different points along the trajectory and cannot be used simply to calculate the cartesian coordinates of one point.

Ultrasonics, Continuous Wave

A device which used continuous wave ultrasound to obtain position and orientation is the Head Mounted Three Dimensional Display developed by I.E. Sutherland. The purpose of the display was to present stereo images separately to each eye of a subject, using two miniature cathode ray tubes fitted to a "helmet" on the subject's head. The subject then experienced the illusion that he was moving around in a graphics world of three dimensional models which had been constructed previously in a computer. To maintain the correct projections of the models in the displays, i.e. consistent with the motion of his head, the position and orientation of the moving subject had to be monitored. After early attempts to obtain this information using a mechanical linkage between the helmet and the laboratory ceiling, a non-contact system employing continuous wave ultrasound was designed, [Sutherland (1)].

The ultrasonic system is shown schematically in Figure 1.21. Mounted on the helmet are three transmitters, operating at $37kHz$, $38.6kHz$ and $40.2kHz$. Four stationary receivers are used, fixed to the ceiling. At each receiver the incoming

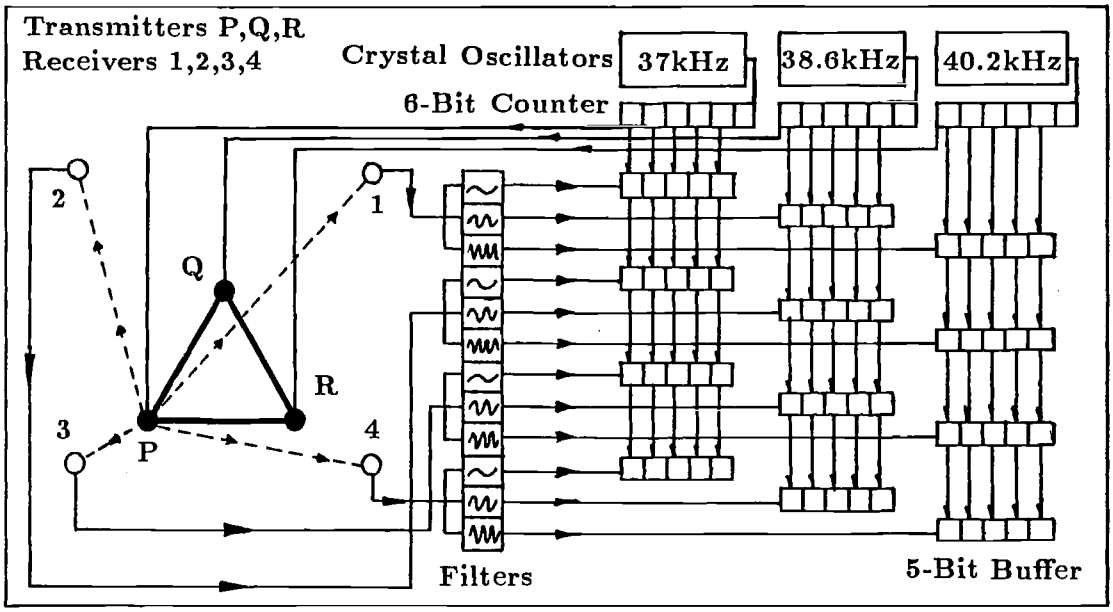


Figure 1.21: Continuous Wave Ultrasonic System

signals at each of the three frequencies is monitored separately, and compared with the reference signal from its crystal oscillator to measure the relative phase shift in that signal. This is recorded with a precision of 5 bits per cycle, i.e. one thirtysecond of a wavelength. The total number of complete wavelengths is also counted. Sound at a frequency of $40kHz$ has a wavelength of approximately $7.5mm$ and so in theory the resolution of each distance measurement is better than $0.25mm$. Each receiver detects all three signals, and so four distances are obtained to each moving point. Calculation of the intersection of three of the spherical surfaces defined by these radii gives the position of one point. Doing this for three points allows the position and orientation of the head to be calculated.

The description of this system in [Sutherland (1)] does not report experimental results, and no mention is made of how continuous wave noise, i.e. reflections from walls etc. might be dealt with. In a later publication [Sutherland (2)], ref-

erence is made only to the mechanical system of head position monitoring, and in [Burton and Sutherland] this ultrasonic method is reported as being unsatisfactory.

1.5.11 Laser Rangefinders

There are two different types of laser rangefinder.[2] One type measures the time of flight of a pulse of light to a target and back. The other type is a continuous wave system in which change in distance is measured by monitoring phase shift.

Laser Rangefinders, Time of Flight

This type of laser rangefinder sends pulses of light to a target. The reflected signal is received back at the instrument and the time lapse between transmission and reception is measured. This is the time required for the double journey to the target and back, and from this the distance is calculated. A rangefinder of this type was built at the Jet Propulsion Laboratory of the California Institute of Technology [Lewis and Johnston] and is shown schematically in Figure 1.22.

The primary purpose of the instrument was to scan a scene and obtain an array of range measurements which were processed to produce a two dimensional "rangepic". Conceivably the instrument could be modified for use in dynamic position measurement of one moving point, however, and so it is worth exploring the limitations of the technique.

From Figure 1.22 it can be seen that the pulse driver for the laser also triggers the "start" of a Time to Pulse Height Converter. The laser pulse is directed to the target by scanning mirrors and the reflected pulse returns via those mirrors to a photomultiplier. The photomultiplier output pulse is received

[2]Excluding structured light systems which sometimes go by this name.

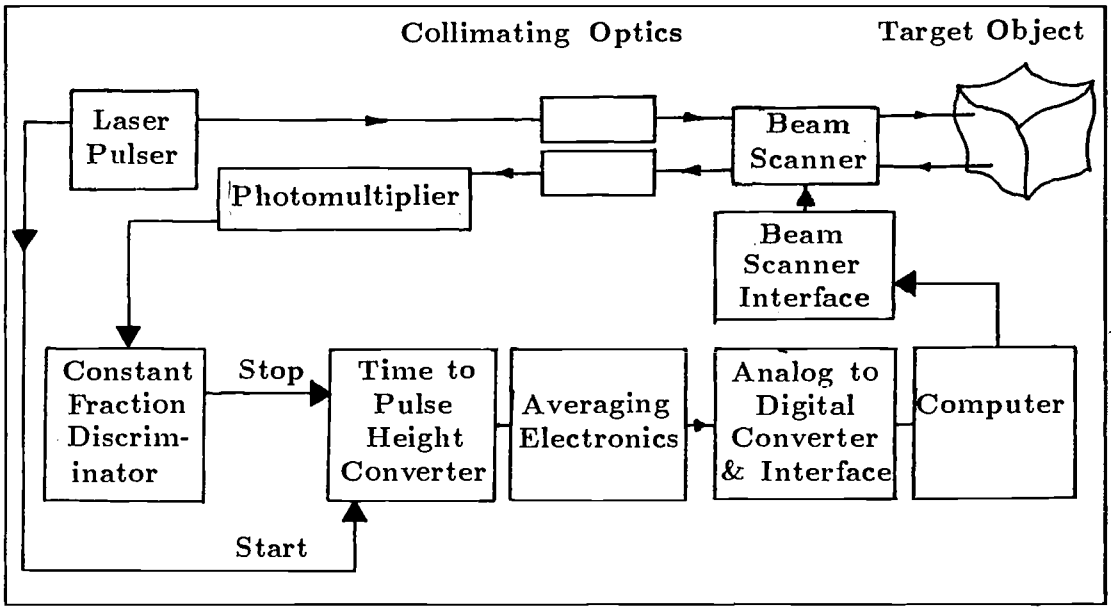


Figure 1.22: Time of Flight Laser Rangefinder

by a Constant Fraction Discriminator which is intended to compensate for the varying strength of returning pulses and which sends a "stop" signal to the Time to Pulse Height Converter. This sequence is repeated at a rate of $10kHz$ and a number of readings are taken for each target point. The output of the Time to Pulse Height Converter is averaged and the result stored in the computer for later analysis. Under computer control, the scanning device then directs the rangefinder at the next target point.

The range measurement problems incurred with this instrument were mainly in connection with the large dynamic range of the return pulse amplitude, which depends on the reflectance of the object, the incident angle and the range. In these circumstances the correct determination of the return pulse time by the Constant Fraction Discriminator was not robust. In favourable conditions, however, an error of less than $\pm 20mm$ on the single distance could be achieved. This corresponds to a time error (over the double path) of $\pm 133ps$.

A similar time of flight system has been built at the Australian National University [Jarvis (1)] for use in conjunction with a structured light and video camera system. The inaccuracy given for this instrument is $\pm 2.5mm$, averaging 100 measurements for each point in a scene and with an acquisition frequency of $100Hz$.

To use a rangefinder system such as these for dynamic position measurement of a robot requires that the rangefinder track a point rather than scan a scene. Conceivably the laser beam steering could be controlled by a similar feedback loop to that used in the University of Surrey and NBS servoed systems described in later sections below. Following the same target material would get over the varying reflectivity problem. Another approach would be to separate the transmitting and receiving parts of the system so that the transmitted pulses come from the robot itself and are "observed" by a stationary receiver.

The resolution attainable with this type of range finder is poor, however. Even if the problems of noise and dynamic range are overcome, the inability to measure time with sufficient accuracy is the fundamental constraint on the resolution of laser time of flight systems. One suggested improvement [Jarvis (2)] is the incorporation of a streak camera into the system. This affords a time resolution of $10ps$, which with no other errors present, would represent a total path error of $\pm 3mm$ or a distance error of $\pm 1.5mm$ to the target point. Sub-millimetric accuracy (let alone $\pm 10\mu m$) is beyond the capabilities of time of flight measurement.

Laser Rangefinders, Continuous Wave

The second type of laser rangefinder is not pulsed, but uses a continuous wave laser beam modulated by a continuous periodic signal, usually in the tens of

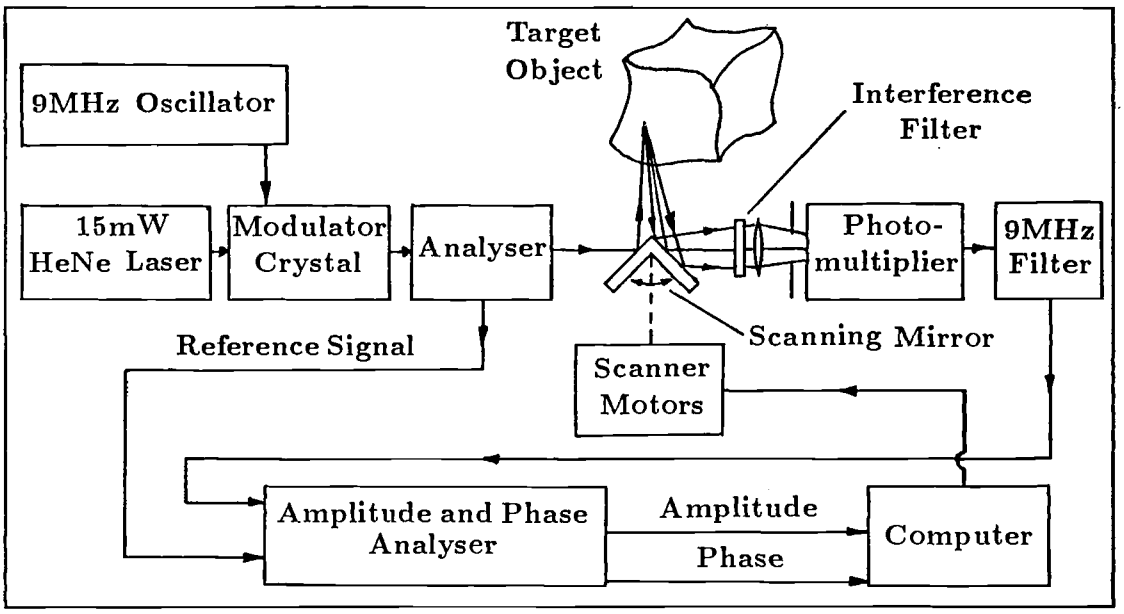


Figure 1.23: Continuous Wave Laser Rangefinder

megahertz region. The reflected signal from the target point is received at the rangefinder and its phase of modulation is compared with the modulation phase of a reference (taken from the outgoing transmitted signal). The rangefinder built at Stanford Research Institute (SRI) was based on this principle, [Nitzan, Brain and Duda], and is shown schematically in Figure 1.23.

In the SRI rangefinder the beam of a He Ne laser passes through a crystal of ammonium dihydrogen phosphate, where it is modulated at $9MHz$, and a sample of the output beam is passed to an amplitude and phase meter. The outgoing laser beam is directed to a target and the reflected signal is received through a narrow band filter by a photomultiplier. The photomultiplier's output is passed through a $9MHz$ filter to the amplitude and phase meter, which has two functions. The first is to compare the phase of the transmitted and received signals and give an output of relative phase shift. The second function is to present the amplitude of the return signal as an output. The two outputs pass

through a multiplexer and A/D converter to a computer, which records both the range and intensity data for measured points.

The rangefinder was designed to operate over a range of between $1m$ and $5m$, and with the $9MHz$ modulation frequency and the ability to sense down to a 0.2 degree phase shift, had a resolution of $10mm$. The low power ($15mW$) of the laser used, and the effects of range and reflectivity, could give rise to weak return signals. Photon noise in the photomultiplier added to the problem, and with the resulting low signal to noise ratio, long integration times ($500ms$) were required for each measurement, to allow the phase shift to be found with the required accuracy. Thus the measurement acquisition rate for this instrument is extremely low, i.e. $2Hz$.

A rangefinder which is similar in principle, [Page and Hassan], has been designed for the inspection of manufactured parts. In this application it was hoped that the high reflectivity of the surfaces involved would allow the signal to noise problems described above to be avoided. The main differences from the SRI design are the use of an infra red laser diode as the transmitter and an avalanche photodiode as the photosensitive receiver, and the modulation of the beam at $50MHz$. This higher frequency would give a theoretical resolution of $1.5mm$ on the measured distances.

As with the time of flight rangefinder, the adaptation to dynamic position measurement of one moving point would require one of two things. Firstly the laser beam might be tracked to follow the point of interest, using some servo system. Alternatively, the instrument could be split into two parts, with a point source transmitter (a laser diode say) mounted on the moving object and the receiver stationary.

1.5.12 Servo Directed Lasers and Interferometers

Retroreflector System

An optical system designed at the University of Surrey [Gilby and Parker] is shown in Figure 1.24. In this system a tracker with a laser, a collimator and two galvanometer driven mirrors directs the laser beam to a retroreflector mounted on the moving robot. The return beam from the retroreflector is also reflected from the two mirrors. It then passes through a beam splitter to a quadrant detector. The two mirror angles (obtained from capacitive transducers on the mirror drivers) and the error registered by the quadrant detector are then used to calculate the equation of a line passing through the centre point of the retroreflector. Two such tracking systems give sufficient information to calculate the cartesian coordinates of the retroreflector centre point. In each tracker the quadrant detector error is also used to drive the mirror galvanometers in such a way as to reduce the error. Using this closed loop servo system, each laser beam follows the retroreflector as it moves with the robot.

In [Gilby and Parker] the inaccuracy of this system is given as less than $\pm 10\mu m$ over a $1m^3$ volume.

2-Axis Gimbal System and Interferometer

Another system which employs tracking laser beams is the Automatic Laser Tracking Interferometer (ALTI) developed at the US National Bureau of Standards, [Industrial Robot]. This device also uses a servoed mirror system to steer a laser beam in azimuth and elevation. For ALTI Version 1 (see Figure 1.25), the target of this beam is another servoed mirror system mounted on the robot being tracked. At the second mirror, a beam splitter and quadrant detector

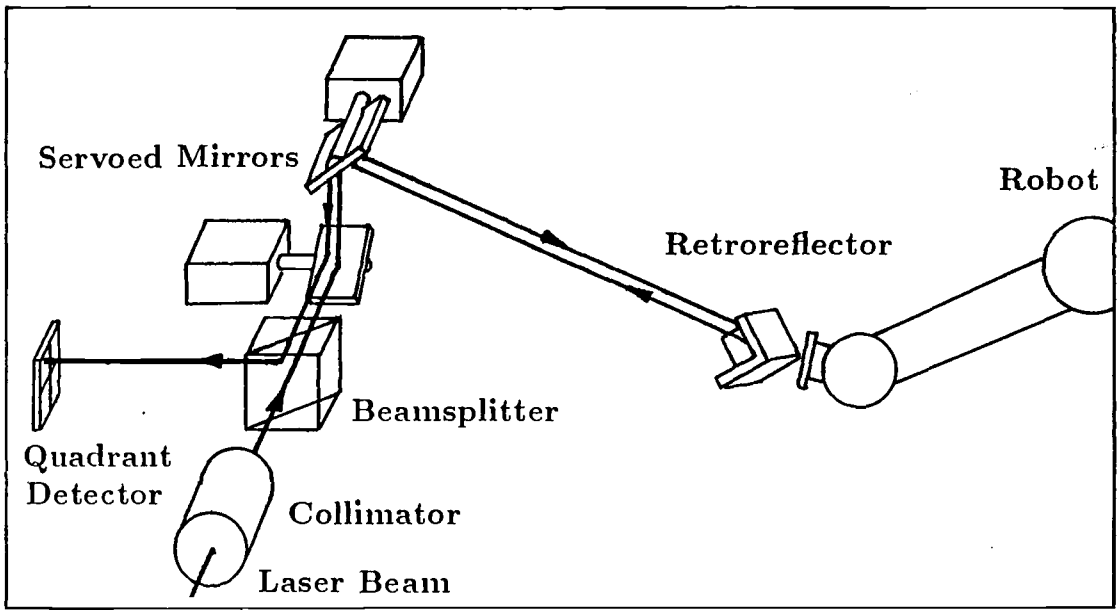


Figure 1.24: Servo Directed Laser with Retroreflector

provide position feedback for the mirror servos. The second mirror reflects the beam back down the same path, onto the centre of the first mirror system. The beam is reflected from there to another beam splitter and on to a quadrant detector which again provides the error signals for the mirror servo systems.

The mirror angles of the "fixed" part of the system give the azimuth and elevation to the centre of the moving mirror. The changes in distance between the mirrors is measured using an interferometer as shown in Figure 1.25. Part of the amplitude of the returning beam is interfered with a component of the output beam, and a count of the fringes obtained is used to calculate changes in distance. This is a relative measurement, and to obtain the actual distance the interferometer must be calibrated using some known distance. Altogether the four angles and one distance which are obtained give five independent measurements, and this can be seen as a five degree of freedom system. In effect this means that if only one such system is used, the three cartesian coordinates of

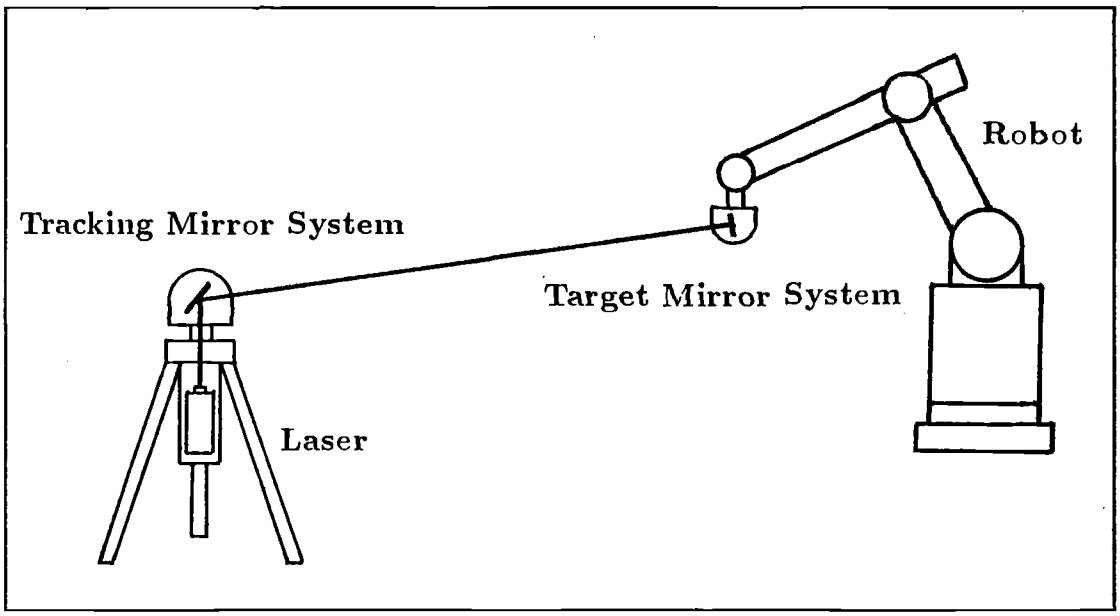


Figure 1.25: ALTI Version 1

the moving mirror centre can be found and, in addition, some identified point on the end effector can be said to lie on a particular circle in space; see Figure 1.26. Two systems like this give sufficient information for the position and orientation of the end effector to be determined.

ALTI Version 2 uses a retroreflector in place of the moving mirror. This gives a system similar to the University of Surrey instrument, except that the interferometer is also present. Thus ALTI Version 2 gives azimuth, elevation and range to the optical centre of the retro reflector, and the cartesian coordinates of that point can be calculated if required. See Figure 1.27.

In calculating the position of the centre of either the moving mirror or the retroreflector, the position errors due angular measurements are dependent on the resolution of the angle transducers used and the accuracy with which they are calibrated (together with errors in the optical components). These errors increase with range. The position errors due to the interferometric measurement

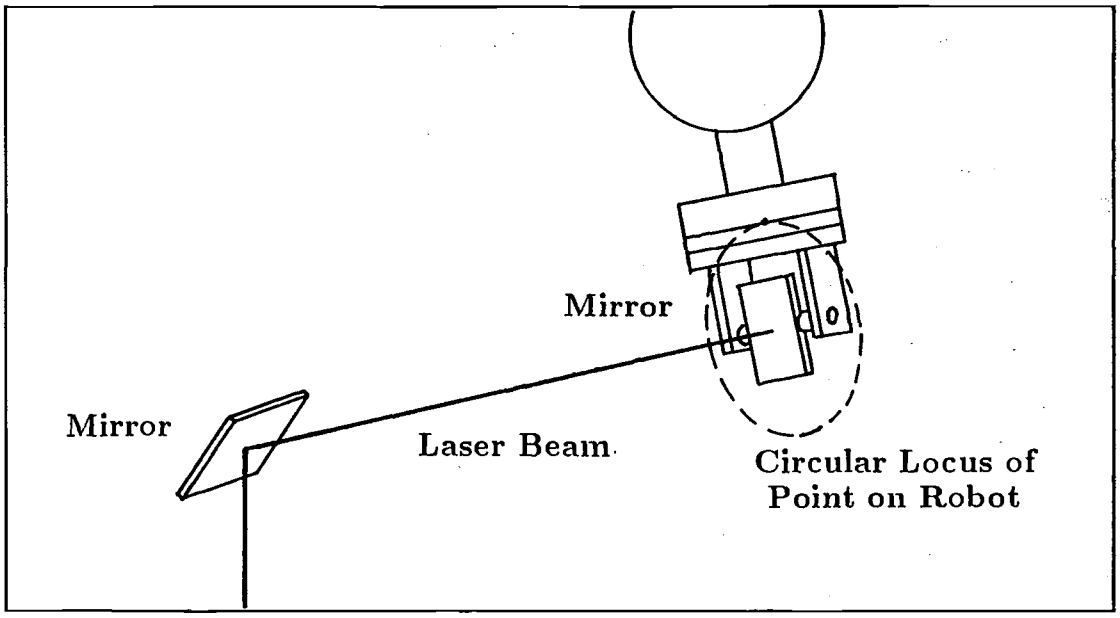


Figure 1.26: Circular Locus of a Point on a Robot

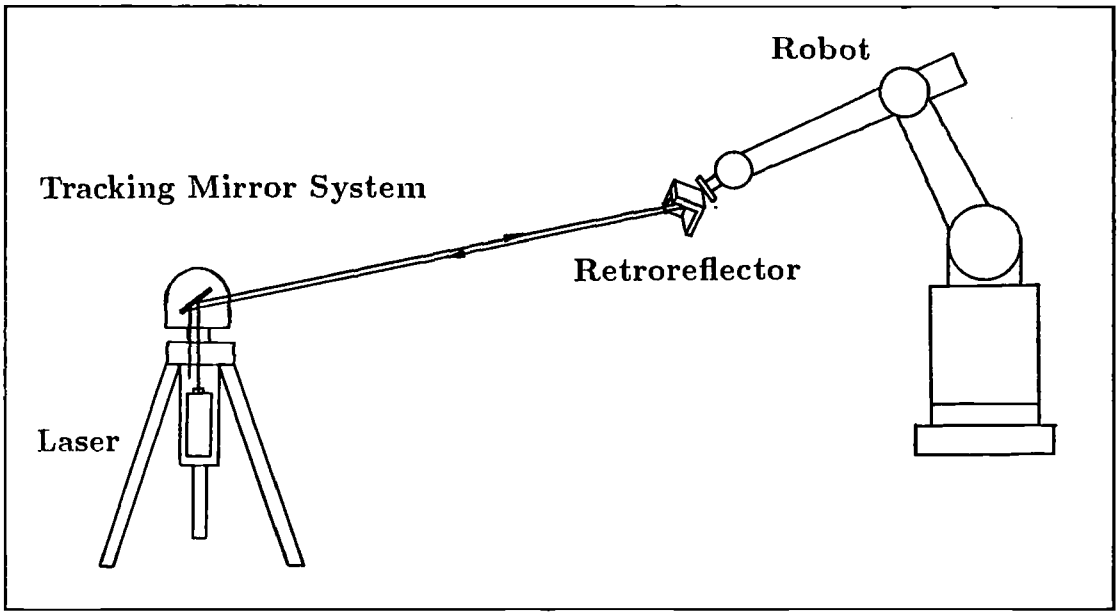


Figure 1.27: ALTI Version 2

of range depend mainly on the calibration accuracy, i.e. the known range from which fringe counting starts. The resolution of the interferometer is, of course, very good, (being less than a wavelength of light) and the error does not increase with range.

Four Linear Interferometers

The previous two systems rely to some extent on angular measurements and so give errors in position which increase with range. The "CMS-1000" system manufactured by Chesapeake Laser Systems Incorporated avoids direct dependence on the accuracy of angle transducers by using only distance measurements in the calculation of point position, and by making these measurements with interferometers [Chesapeake]. As with the two devices described above, a laser beam is directed towards the moving target by a servoed mirror system. The target is a cluster of retroreflectors which returns any beam coming from a direction within a large solid angle. Either three or four beams are aimed at this target simultaneously, and the interferometers in each instrument give, respectively three or four distances to the target from different points outside the robot working volume. As with the tensioned wire method, three instruments are sufficient in ideal circumstances but the use of four allows problems of ill-conditioning to be overcome and so gives more accurate results. This instrument is illustrated in Figure 1.28.

The manufacturers give the performance figures as; resolution $1\mu m$, accuracy $2\text{–}10\mu m$ over a $3m \times 3m \times 3m$ working volume, and a data capture rate of over $200Hz$. As with any interferometer, the accuracy of each instrument depends on the setting up or calibration procedure, and as with any system using measurements from several instruments, the overall accuracy depends on accurate

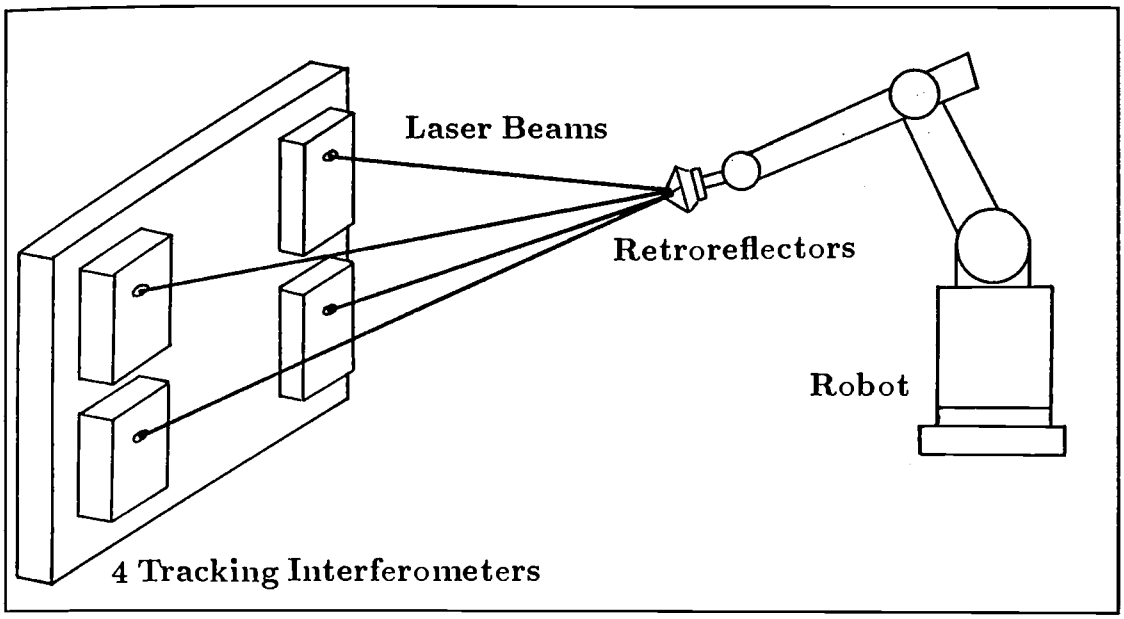


Figure 1.28: Four Linear Interferometers

knowledge of the positions of the trackers themselves.

1.5.13 Discussion of Dynamic Systems

The dynamic systems described above can be categorised in various ways. For instance they might be considered in relation to their original application, e.g. systems used primarily for remote sensing of the environment as against systems designed from the start to be instruments of metrology. They can be considered in terms of the type of measurement they make, e.g. angle, distance, or time, or the type of physical principle they employ, e.g. time of flight of a pulse of radiation, phase shift of continuous wave radiation, inertial systems which use the laws of mechanics, or simply taking the bearings of points. Some devices make contact with the moving object and some rely on radiation from the object to a remote sensor. Some instruments use images of a whole scene and some concentrate on tracking a particular moving point.

Some general remarks can be made about the characteristics of instruments

in particular categories and the problems commonly associated with them.

1. Methods involving contact with the moving object or requiring a lot of instrumentation to be carried on the moving object distort the measurements they are designed to make.

2. Triangulation devices suffer from decreasing accuracy with increasing range.

3. Imaging systems suffer from the problem that they present too much data, from which the important information has to be extracted.

4. Devices which steer beams to track a moving point introduce the complexity and expense of servo systems built into the instrument itself.

5. Ultrasound, at the wavelengths commonly used, suffers from specular reflections from most surfaces. At shorter wavelengths the strength of the signal attenuates rapidly. The low speed of sound restricts the data acquisition frequency.

6. Electro-magnetic radiation, e.g. visible or infrared light, avoids the specular reflection problem in most circumstances but, in the case of time of flight systems, comes up against the problem that given the speed of light, the shortest possible time intervals which can be distinguished still represent distances greater than $1mm$.

Systems which operate, at least in part, remotely from the moving object do all have one thing in common. To obtain all three cartesian coordinates of a point, several instruments (of whatever type) are needed, and the calculation of cartesian coordinates almost invariably requires that the intersection of some combination of lines, planes or spherical surfaces be found in three dimensional space.

1.6 Introduction to the 3-D Interferometer

The purpose of this section is to give a brief outline of the principles of operation of the interferometer which is the main subject of this thesis. The instrument was designed with the dynamic three dimensional position measurement of robots in mind as its application, and it overcomes most of the common problems indicated in the last section.

This interferometer is an optical instrument and, like other interferometers, obtains its measurements by bringing together coherent light from two different optical paths and monitoring the changes in the resulting interference pattern. It is unlike the interferometers used in the existing measurement systems described in subsection 1.5.12, however. The instruments used there are linear interferometers which give measurements of length along the direction of their beam. If the target moves out of the beam, the interference pattern is lost, and so both the instruments described previously require servoed tracking mirrors to keep the beams on target.

The new instrument, on the other hand, does not involve any servo systems or other moving mechanical parts. It works by monitoring the position of a *point source* of laser light attached, say, to the moving robot's end effector. The point source emits spherical wavefronts within the conical volume of the solid angle of transmission. Some of this light is incident on the interferometer itself, positioned outside the robot's working envelope as shown in Figure 1.29. Within the limits imposed by the solid angle of transmission, the interferometer continues to receive the signal as the point source translates and rotates in any direction.

The interferometer itself works by *division of wavefront* in contrast to the

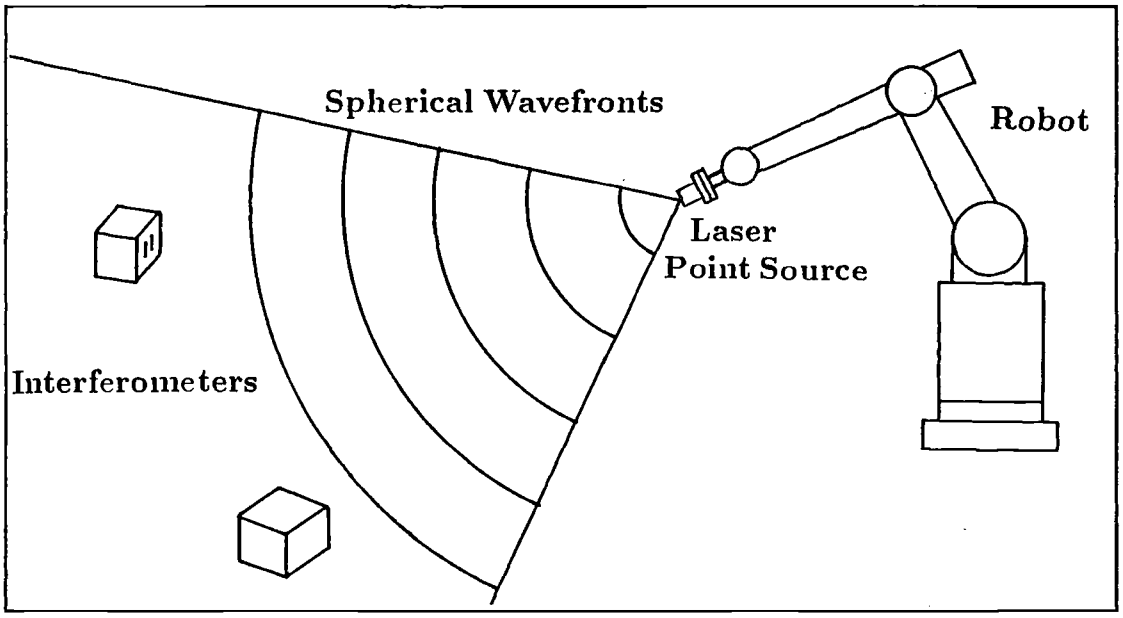


Figure 1.29: Laser Point Source and Receivers

division of amplitude employed by most existing instruments. At the front of the receiver are two parallel slits, and at each of these, incident light is diffracted as it enters the instrument. The two "fans" of diffracted rays continue to diverge and pass into a lens, which is focused at infinity. The diffracted light from each slit produces its own diffraction pattern on the focal plane of the lens, and where the two patterns overlap, a large number of fine, parallel interference fringes are produced. Figure 1.30 is a simplified diagram showing some of the main components of the interferometer.

In the interference pattern, the phase at some particular point on the focal plane is determined by the optical path lengths from the source, through the two slits, to that point. These in turn depend on the distances from the moving source to the two slits. In consequence, as the point source moves, then, in general, the interference pattern moves as a whole on the focal plane. The movement of the interference pattern in a direction normal to its parallel lines

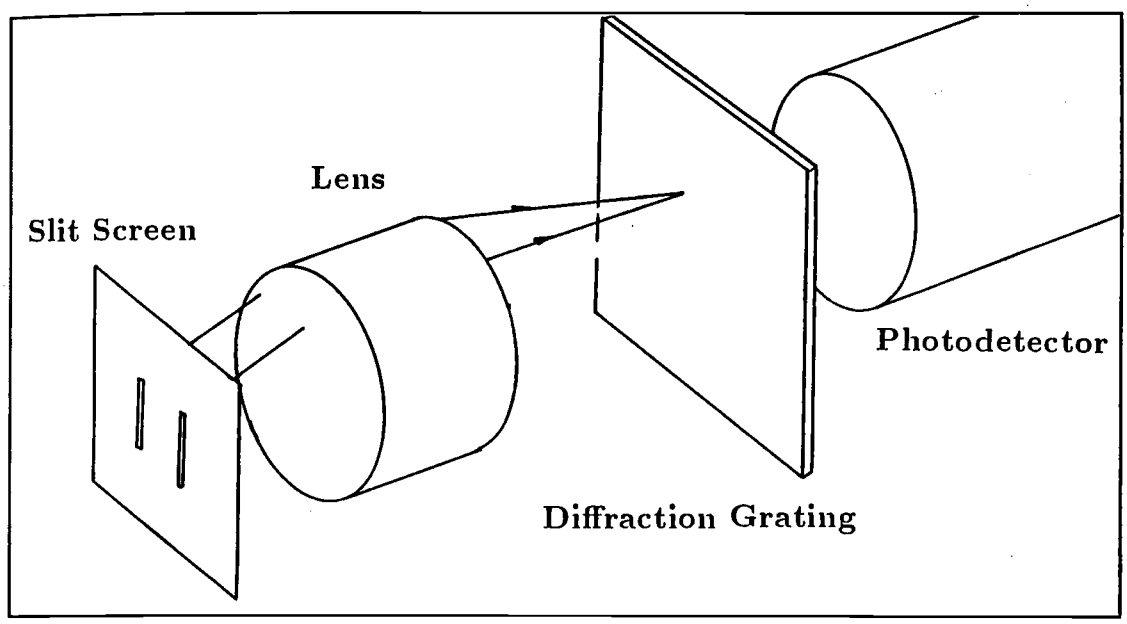


Figure 1.30: Interferometer, Basic Components

can be used to determine one coordinate[3] of the point source in three dimensional space.

Movement of the interference pattern on the focal plane of the lens can be detected using an amplitude transmission linear diffraction grating positioned in this plane. The grating has the same spatial frequency as the fringe pattern, and is oriented with its lines parallel to the fringe lines. As the fringe pattern moves in a direction normal to the grating lines, the amplitude of the light transmitted through the grating varies as a sine wave superimposed on a DC level, with one complete cycle per "line pair" traversed on the grating. The power transmitted through the grating is received by a photosensitive detector which outputs an electronic signal.

By placing a beam splitter behind the lens, two such detectors can be used to give two output signals in quadrature, thus the magnitude and direction of

[3]See Chapter 2.

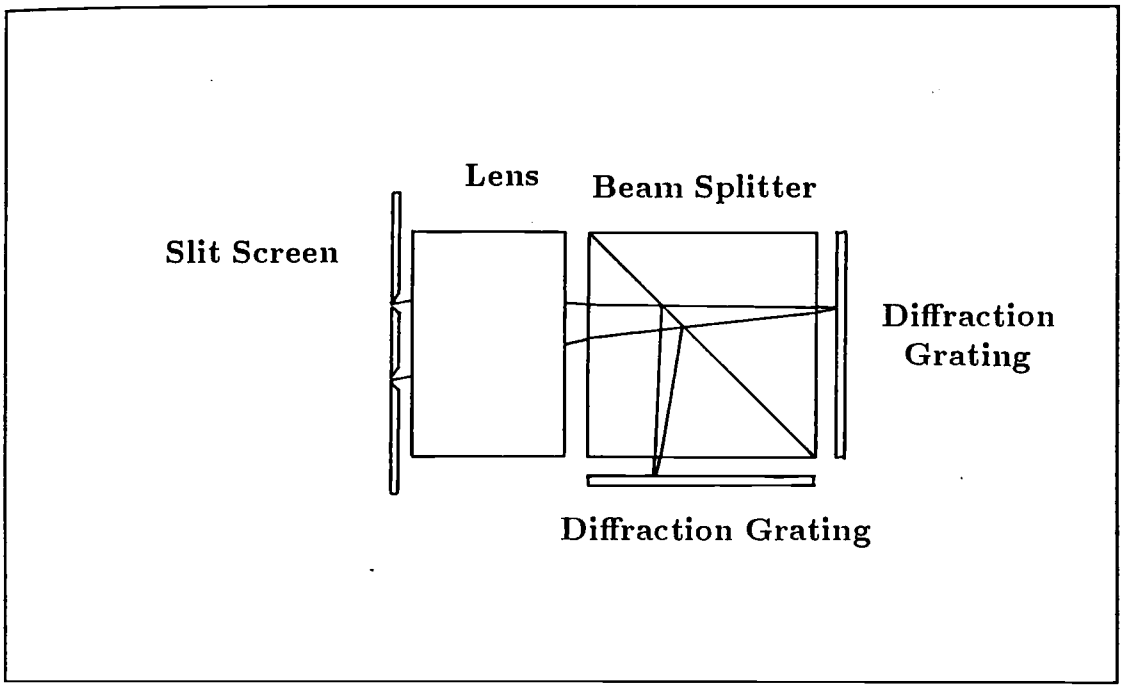


Figure 1.31: Interferometer, Quadrature Detection of Fringe Movement

the fringe movement can be obtained. See Figure 1.31.

The geometric relationship between the source position and the interference pattern position is as follows. For a fixed position of the interference pattern (in the direction normal to its fringe lines), the point source must lie on a particular member of the family of hyperboloids of revolution, of two sheets, with foci at the centres of the two slits. Using three of these instruments, illuminated by the same source, the three cartesian coordinates of the point source can be calculated from one of the intersections of the three hyperboloids. This is illustrated in two dimensions in Figure 1.32.

The optics of the interferometer are described in detail in Chapters 4, 5 and 6. The next two chapters investigate the geometry which allows the position and orientation of an object to be calculated from the interferometer's output.

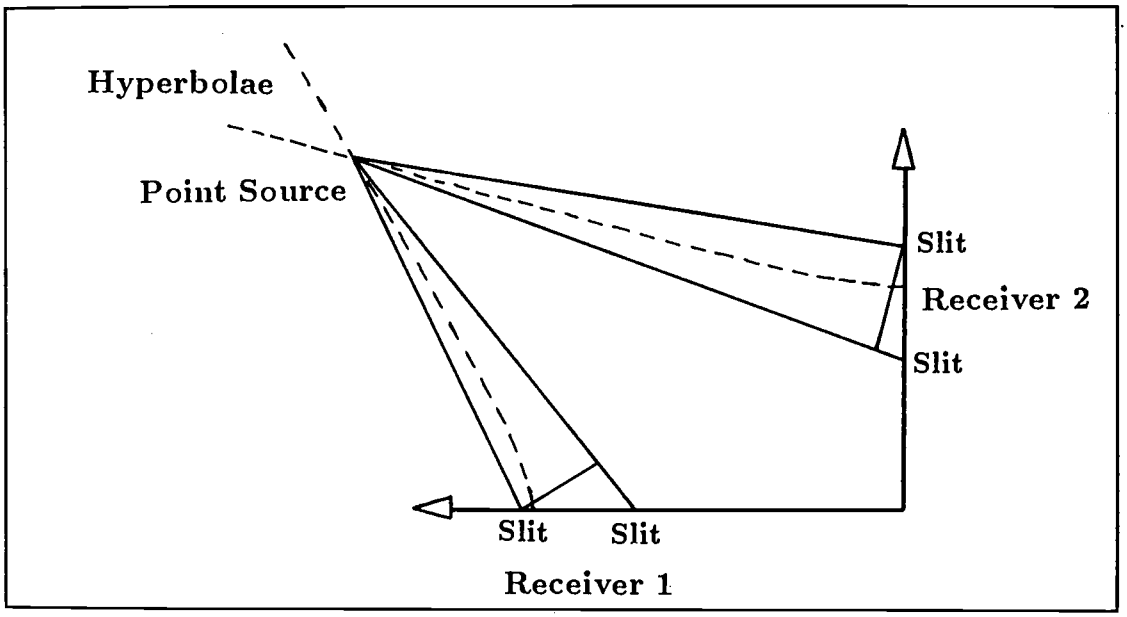


Figure 1.32: Intersection of Hyperbolic Point Loci

Chapter 2

Location of Points

2.1 Position Measurement Systems

The position measurement instruments described in the last chapter can be used in systems which find the location of point in three dimensional space to within some tolerance. They employ a whole range of physical effects, and occasionally some quite ingenious designs to obtain their output. What they do, almost all, have in common is that the coordinate output from each instrument in a system represents a geometrical entity in the three dimensional space. With most instruments this is a line, a plane, or a sphere, and the location of the required point is found by calculating the intersection of some combination of lines, planes and spheres.

In the calculation to identify a particular point, the set of scalar coefficients of components of the position vector must be found. The coordinates output from an instrument might be exactly the scalar coefficients of the position vector, as would be the case in a perfect prismatic coordinate measurement machine. In most systems, a considerable amount of manipulation is required to obtain the position of the point. That is true of the interferometer, (Section 1.6), which outputs coordinates which represent hyperboloids. The calculation to

obtain the intersection of three of these hyperboloids is complex in comparison to calculating, say, the intersection of three planes.

Ultimately all position measuring instruments reach the same result however, and conceptually, most of them go through a similar process to achieve it. To illustrate this, the next section shows that any instrument or combination of instruments as a system can be represented algebraicly as a map.

2.2 Coordinates and Surfaces

2.2.1 Coordinates

Firstly, the idea of a *coordinate* must be separated from the familiar x_1, x_2, x_3 scalar coefficients of a vector in a cartesian reference frame. The term has a more general meaning and it should not be associated with vector spaces. A coordinate is an index to a set. [Do Carmo, pp52,53], [Blyth, Ch 1]. More specifically and restricting the definition to real coordinates, a coordinate is a real number which is a member of a subset of $\Re$. This subset is an index set which, through a map, specifies a subset of $\Re^m$, (m being a positive integer). In a geometrical case, this subset of $\Re^m$ is a point set, i.e. a point, a curve or a surface. A number output from an instrument is a coordinate and so an index. The range of such numbers produced by the instrument between its limits of operation is the domain of the map.

2.2.2 Set of Coordinates

With n a positive integer, a *set of coordinates* is a set of ordered n-tuples of real numbers, each number being a member of a subset of $\Re$ and mapping to a subset of $\Re^m$ as with a coordinate singly. The *intersection* of the image subsets specified through maps by particular coordinates in a member of this n-tuple set

picks out another subset of $\Re^m$, for example, in the geometrical analogy, a point, a curve or a surface. With this formality applied to examples of instruments its simplicity is clear, and it can be seen that a mapping is a compact way of describing the geometric attributes of a system. Initially one linear and one non-linear system are given as examples.

Example: Coordinate Measurement Machine

In a three axis inspection machine the coordinate measured by each axis independently maps to a plane lying normal to that axis direction. This plane is the particular subset of $\Re^3$ picked out by that coordinate, and the remaining information necessary to fix the point is now a member of $\Re^2$. The other two coordinates define intersecting planes which successively remove the remaining degrees of freedom of the point.

In the general case, the axes of the machine are not mutually perpendicular. The plane defined by the coordinate of an axis is given by $\mathbf{x} \cdot \mathbf{n} = d$ where $\mathbf{x}$ is the position vector of a point on the plane, $\mathbf{n}$ is a unit normal to the plane and d is the perpendicular distance from the reference frame origin to the plane, as in Figure 2.1. With no knowledge of any of the coordinates, the point to be found can be anywhere in a volume. Considering only the ith axis of the machine, fixing that single output coordinate constrains the point to lie on a plane. If the characteristics of the machine are known then the direction of the normal to the plane is known, and as the coordinate of that axis varies, the plane sweeps out a volume. In other words, for the ith axis of the machine, the output coordinate $u_i \in U_i$ selects a polynomial p_i from $\{p_i\}$ (the set of all parallel planes in this case) such that an unknown coefficient of the polynomial, d_i is determined by $u_i \mapsto d_i$.

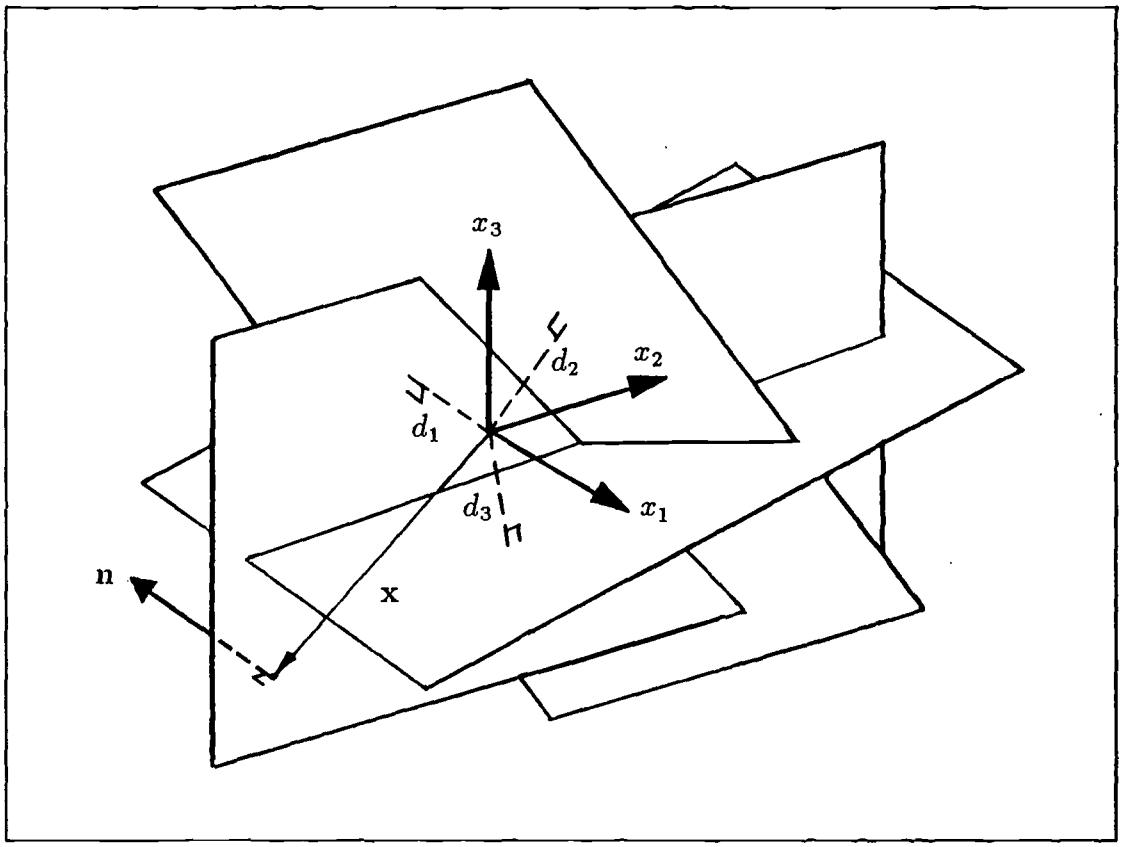


Figure 2.1: Axes of Coordinate Measurement Machine.

The set $V(p_i)$ is the set of triples of coefficients of a point vector to any point on the surface which is defined by putting $p_i(x) = 0$. So each $u_i \in U_i$ defines a plane point set,

$$V(p_i) = \{(x_1, x_2, x_3) \in \Re^3 \mid \quad u_i \mapsto d_i$$
$$p_i = x_1 n_{i1} + x_2 n_{i2} + x_3 n_{i3} + d_i = 0\} \quad (2.1)$$

The volume swept out by all such parallel planes (as u_i varies) is the volume point set,

$$V_i(\{p_i\}) = \{(x_1, x_2, x_3) \in \Re^3 \mid \quad u_i \mapsto d_i$$
$$p_i = x_1 n_{i1} + x_2 n_{i2} + x_3 n_{i3} + d_i = 0\} \quad (2.2)$$

in which d_i is now a variable.

If three coordinates u_1, u_2, u_3 are specified then three planes $V(p_i)$ are picked out and their intersection is the point which satisfies all three equations. The complete map α for the three-axis instrument is,

$$\alpha : U \rightarrow V = \{V(p_1) \cap V(p_2) \cap V(p_3)\} \quad (2.3)$$

or explicitly,

$$\alpha : U \rightarrow V = \{(x_1, x_2, x_3) \in \Re^3 \mid \quad \forall i,\ u_i \mapsto d_i$$
$$\forall i,\ p_i = x_1 n_{i1} + x_2 n_{i2} + x_3 n_{i3} + d_i = 0\} \quad (2.4)$$

This describes the means by which the coordinates obtained by the techniques of this instrument yield a set of simultaneous (in this case linear) polynomials which may be solved for the desired position vector.

Other position measuring instruments can be described in a similar way, the polynomials being different in general, their variety determined by the physics and geometry of that instrument.

Example: Spherical Coordinates

An example rather different from the one above is the ALTI measurement system designed at the US National Bureau of Standards and described in Chapter 1, 1.5.12. The position of the centre point of the moving target can be determined from the system's (spherical) coordinates. The two angles of the tracking mirror allow azimuth and elevation to be determined. The linear interferometer measures changes in range from some calibrated distance and so determines radius.

So in this case with $u_1 = \phi$, $u_2 = \theta$, and $u_3 = r$, the map is,

$$\begin{aligned}
\alpha : U = \{u_1, u_2, u_3\} \rightarrow V = \{(x_1, x_2, x_3) \in \Re^3 \mid \quad & u_1 \mapsto n_{11} = \cos u_1,\ n_{12} = \sin u_1 \\
& p_1(x) = x_1 n_{11} + x_2 n_{12} = 0 \\
& u_2 \mapsto a_{21} = a_{22} = \tan^2 u_2 \\
& p_2(x) = a_{21} x_1^2 + a_{22} x_2^2 - x_3^2 = 0 \\
& u_3 \mapsto r = u_3 \\
& p_3(x) = x_1^2 + x_2^2 + x_3^2 - r^2 = 0\}
\end{aligned} \tag{2.5}$$

The surfaces picked out by these coordinates are a plane, a cone and a sphere,

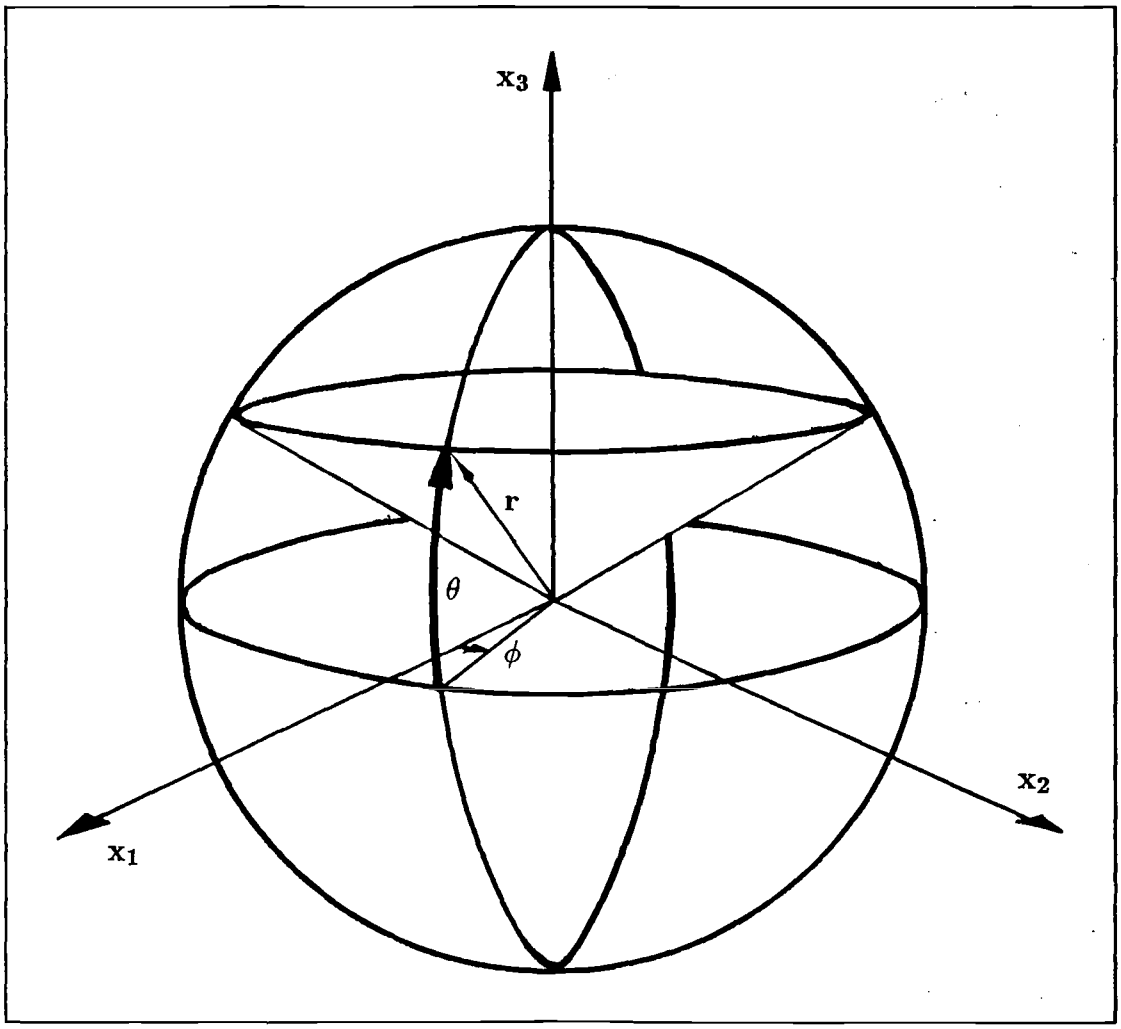


Figure 2.2: Intersecting Plane, Cone and Sphere.

the discriminant matrices of the quadrics both having their canonical form in the same reference frame. See Figure 2.2.

2.3 Reduction of Dimension

In general each of n coordinates maps to an algebraic variety, which is a family of polynomials, each of which is equal to zero, in $\leq n$ indeterminants. In many cases the number of dimensions over which the point is free to "move" in an n-space is reduced by one by each coordinate[1]and thus by each of these equations

applied.

A member of an algebraic variety is a hypersurface in the n-space. In the 3-space which is primarily of interest the hypersurface is usually a surface. Three such surfaces indexed by three coordinates can be used to fix a point in space by taking the intersection of the surfaces, as with the two examples given. Multiple solutions occur with surfaces defined by non-linear polynomial equations. In real situations, other evidence can be brought to bear in determining the correct choice of roots. Examples are the velocity, acceleration and previous sampled position of a moving point.

The method of finding the intersections of the specified surfaces depends on the degree of the polynomials and has to be considered for individual cases. An approach for some three dimensional interferometric measurements is given in the next section, the instrument providing coordinates which index quadric surfaces.

2.4 Geometry of the Interferometer

The geometrical relationship between the position of a point source free to move in three dimensions, and the fringe pattern movement which gives rise to the output signal of the interferometer was introduced in Chapter 1, 1.6. In this section the geometry of the intersection of the hyperboloids determined by the coordinates of three interferometers is considered and an algorithm for finding the position vectors of the intersection points is presented.

In a plane which includes two fixed points, representing the slits of the inter-

[1]Some inconsistencies appear when the varieties are restricted to being real varieties. An equation such as $x_1^2 + x_2^2 = 0$ picks out the x_3 axis alone in $\Re^3$, thus reducing the dimension by two with only one equation. Also, two surfaces may not intersect in $\Re^3$ (e.g. if they are parallel planes). See [Kendig, Ch 1]. In a real measurement system these circumstances can be avoided or accommodated in the design of the instrument and associated algorithms.

ferometer, the movement of the source point which produces no change in the relative phase between the two slits must be movement on the locus of a point with constant difference of distances from these two fixed points. This locus is a hyperbola.

Allowing the point to move in three dimensions, the locus giving constant phase difference between the two fixed points is a hyperboloid (of two sheets) of revolution, rotated about the axis between the two fixed points. See Figure 2.3.

Instrument Map

The coordinate output by the interferometer represents the phase difference between its two fixed measurement points and so, in three dimensions, maps to a member of a family of hyperboloids of revolution with foci the two fixed points. With the x_1 axis passing through the fixed points and the x_3 axis the "optical axis" of the instrument, then in its canonical reference frame the equation of this hyperboloid is,

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} - \frac{x_3^2}{b^2} = 1 \tag{2.6}$$

(see [Sommerville]) where $2a$ is the difference in distances calculated from the phase difference and wavelength, $b^2 = c^2 - a^2$, and $2c$ is the distance between the foci. In a general position and orientation, the equation of a quadric is,

$$\mathbf{x}^T A \mathbf{x} = 0 \tag{2.7}$$

where $\mathbf{x}^T = [x_1\ x_2\ x_3\ 1]$ is a point vector in homogeneous coordinates. A is the symmetric discriminant matrix of the quadric and is determined by the position and orientation of the reference frame and the coordinate u_i measured by the

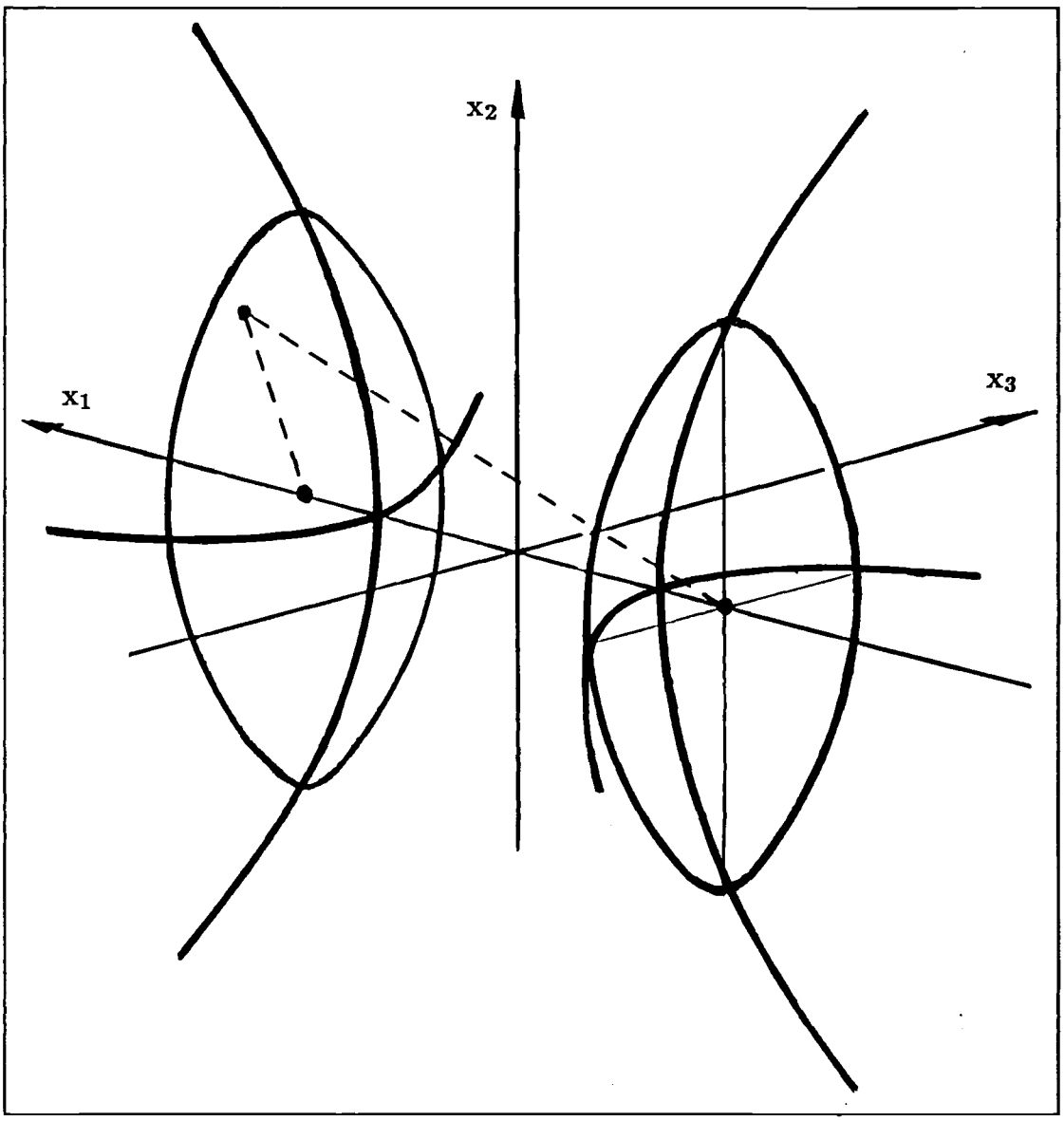


Figure 2.3: Hyperboloid of Two Sheets as a Locus.

instrument. With the position and orientation of the instrument fixed, $V(\{p_i\})$ is the set of surface point sets which may be selected with various values of the coordinate u_i. A particular $u_i \in U_i$ selects the set of vectors $\mathbf{x}$ to points on *one* such surface $V(p_i)$.

Another two interferometers in different positions and orientations give rise to two more surfaces, with discriminant matrices A' and A'' with respect to the same reference frame as A. The position vector of the point source is at one of the intersections of the three surfaces specified by the coordinates, so the map for this instrument is,

$$\begin{aligned} \alpha : U \rightarrow V = \{(x_1, x_2, x_3) \in \Re^3 \mid \quad & p_1 = \mathbf{x}^T A \mathbf{x} = 0,\ u_1 \mapsto A(u_1); \\ & p_2 = \mathbf{x}^T A' \mathbf{x} = 0,\ u_2 \mapsto A'(u_2); \\ & p_3 = \mathbf{x}^T A'' \mathbf{x} = 0,\ u_3 \mapsto A''(u_3)\} \end{aligned} \tag{2.8}$$

With the positions of the instruments known, the elements of A, A', and A'' are easily calculated from given u_1, u_2, u_3. The equations,

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= 0 \\ \mathbf{x}^T A' \mathbf{x} &= 0 \\ \mathbf{x}^T A'' \mathbf{x} &= 0 \end{aligned} \tag{2.9}$$

(i.e. the intersection of three quadrics) must now be solved for $\mathbf{x}$.

Intersection of Two Quadrics

If two quadric surfaces, say with discriminant matrices A and A', intersect then in general they do so in a space curve. This curve is the base curve of a pencil

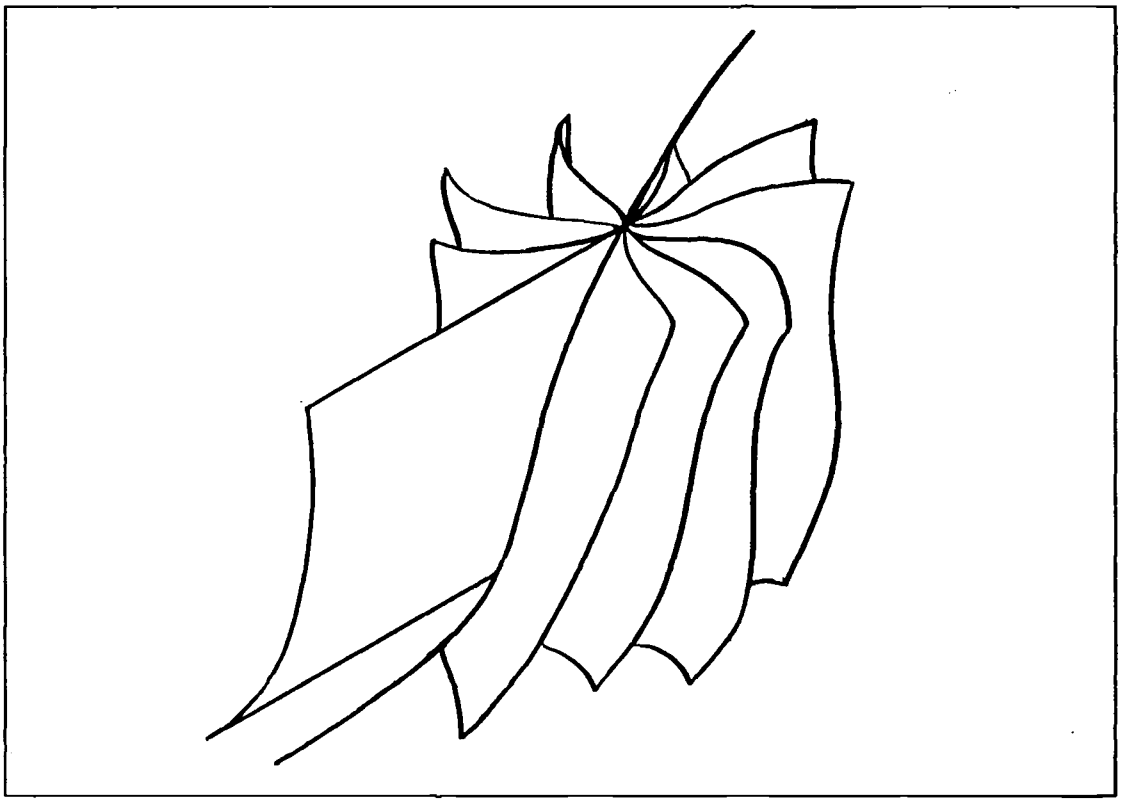


Figure 2.4: Pencil of Quadrics.

of quadrics, (see Figure 2.4),

$$\mathbf{x}^T(A + \lambda A')\mathbf{x} = 0 \tag{2.10}$$

where λ is a scalar which may range from $-\infty$ to ∞. The members of this pencil may be classified according to the ranks of their discriminant matrix and a subdiscriminant matrix, the top left hand 3×3 matrix. [Dresden].

It has been pointed out that the intersection curve of any two quadric surfaces lies on a plane, two planes, a hyperbolic or parabolic cylinder or a hyperbolic paraboloid, all of which are ruled surfaces with singular subdiscriminant matrices $(A_u + \lambda A'_u)$, [Levin]. The equations of these ruled surfaces of the pencil are obtained by putting,

$$\det(A_u + \lambda A'_u) = 0 \tag{2.11}$$

and solving for λ.

In the present case, both A_u and A'_u are the subdiscriminant matrices of hyperboloids of two sheets and are of rank 3 i.e. non-singular, so the equation,

$$|A'_u|\lambda^3 + K_1\lambda^2 + K_2\lambda + |A_u| = 0 \tag{2.12}$$

is a cubic with at least one real root.

Placing each root value of λ in $(A + \lambda A')$, the equations of ruled quadrics passing through the intersection curve are obtained. Selecting one of these, the discriminant matrix is then transformed to its canonical form in another reference frame and expressed in parametric form so that one parameter picks out one ruled line on the surface. With this line equation in parametric form, x_1, x_2, and x_3 can be substituted in the quadric $\mathbf{x}^T A\mathbf{x} = 0$, for instance, to give a quadratic in one indeterminant, the line parameter. Substituting a value of this parameter back into the line equations gives the coefficients x_1, x_2, x_3 at that point on the intersection curve. In [Levin] this approach was taken as part of a process to produce graphical representations of quadric intersections.

Intersection of Three Quadrics

Taking this a stage further, at some point the intersection curve of $\mathbf{x}^T A\mathbf{x}$ and $\mathbf{x}^T A'\mathbf{x} = 0$ cuts the third hyperboloid $\mathbf{x}^T A''\mathbf{x} = 0$. At that point the coefficients x_1, x_2, x_3 found by the above method will also satisfy $\mathbf{x}^T A''\mathbf{x} = 0$. The method of finding *one point* therefore requires an iterative procedure which sets the value of a parameter to pick a line on a parameterisation surface, intersects

this line with another surface on the same pencil and checks to see if this point is on the third (non pencil) surface. Well conditioned solutions can be obtained only if the three interferometers are positioned to give good intersections between their hyperboloids, i.e. if acute angles of intersection are avoided in the working volume of the instrument. This is dealt with in the discussion on the implementation of systems in Chapter 7.

Chapter 3

Location of Objects

3.1 Points and Objects

In applications for metrological instruments it is frequently required that the location of a complete object be described. Information on the location of more than one point on the object is necessary. An example of this type of task is in describing the position and orientation of an industrial robot end effector which can have three translational and three rotational degrees of freedom.

In this chapter the minimum information necessary to locate an object is determined. This is applied in a simple case and an algorithm to give the location of the object is presented. Then the means of employing the hyperbolic geometry of the interferometer to locate an object in three dimensional space is described.

3.2 Determination of Position and Orientation

In locating an object or recording its trajectory through space a complete description requires that the position and orientation of a reference frame fixed in the object can be specified at each sampled point on the trajectory. The position and orientation may be described, with respect to a fixed world reference

frame, by (for instance) three cartesian coordinates and three Euler angles, six independent variables.

Euler angles cannot be measured directly however, and in a dynamic situation measuring cartesian coordinates directly is also a problem. The information obtainable from an instrument in which a measured coordinate constrains a point to lie on some surface can be used on more than one point of the object however. In considering the position of points of a body it can be seen that if the positions of only two points are known (i.e. six coordinates) then the body is still free to rotate about the line through these two points. On the other hand, although knowing the position of three points fixes the body unambiguously nine numbers are required to specify these points.

The number of points and the effect of constraints on a body are now considered. In Figure 3.1, O_0 is the reference frame. The cartesian coordinates and Euler angles of, say, the O_1 frame are known with respect to O_0 or alternatively all three cartesian coordinates of three points are known. In either case the body is fixed.

Now, as an illustration of the effect of linear constraints, fix only the three cartesian coordinates of O_1 and progressively constrain a second point. In Figure 3.2 the two points O_1 and O_2 are used. The position of O_2 is known with respect to O_1 in any body frame but since the orientation of this system in the world frame is unknown, only |w|, the radius to O_1 from O_2 can be used. The locus of possible positions of O_2 is the surface of a sphere, radius |w| and centred on O_1. O_2 is now constrained further by insisting that it must lie on a plane. This intersects the spherical locus in, say, a circle; O_2 still has an infinity of possible positions with respect to O_1 and the body also has an infinity of orientations, rotating about a line through O_1 and O_2.

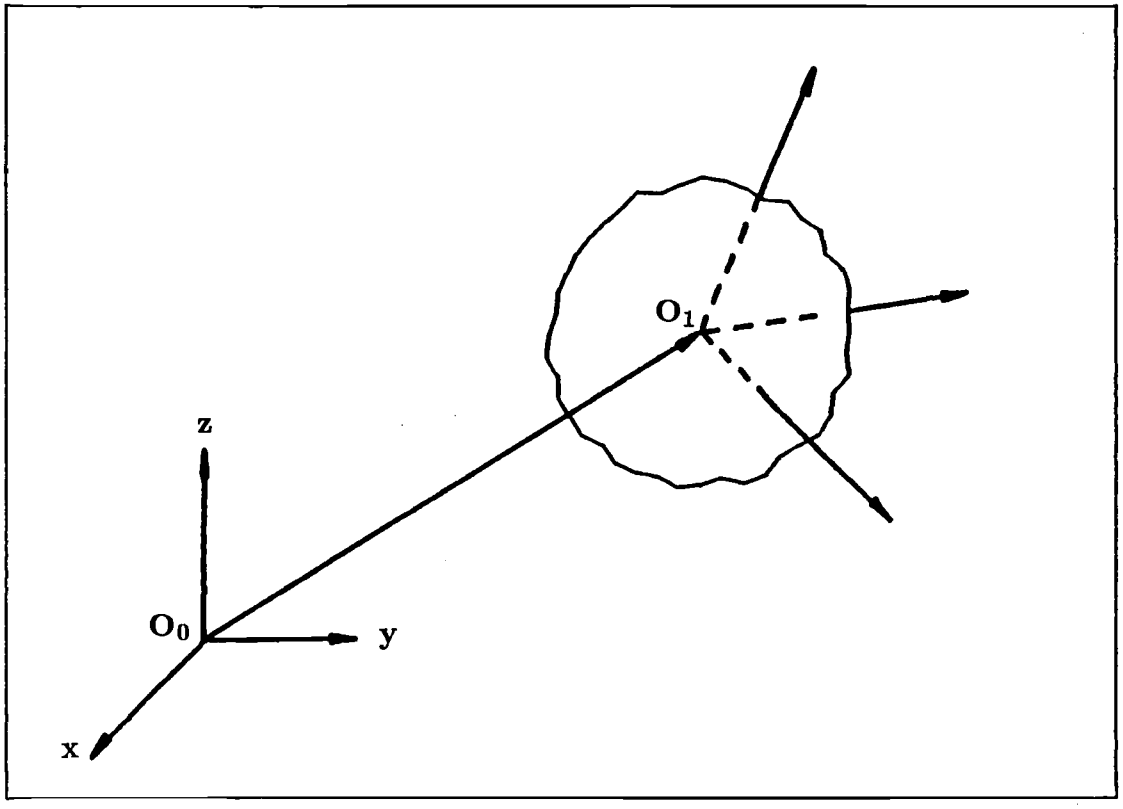


Figure 3.1: Body Constrained in Position and Orientation.

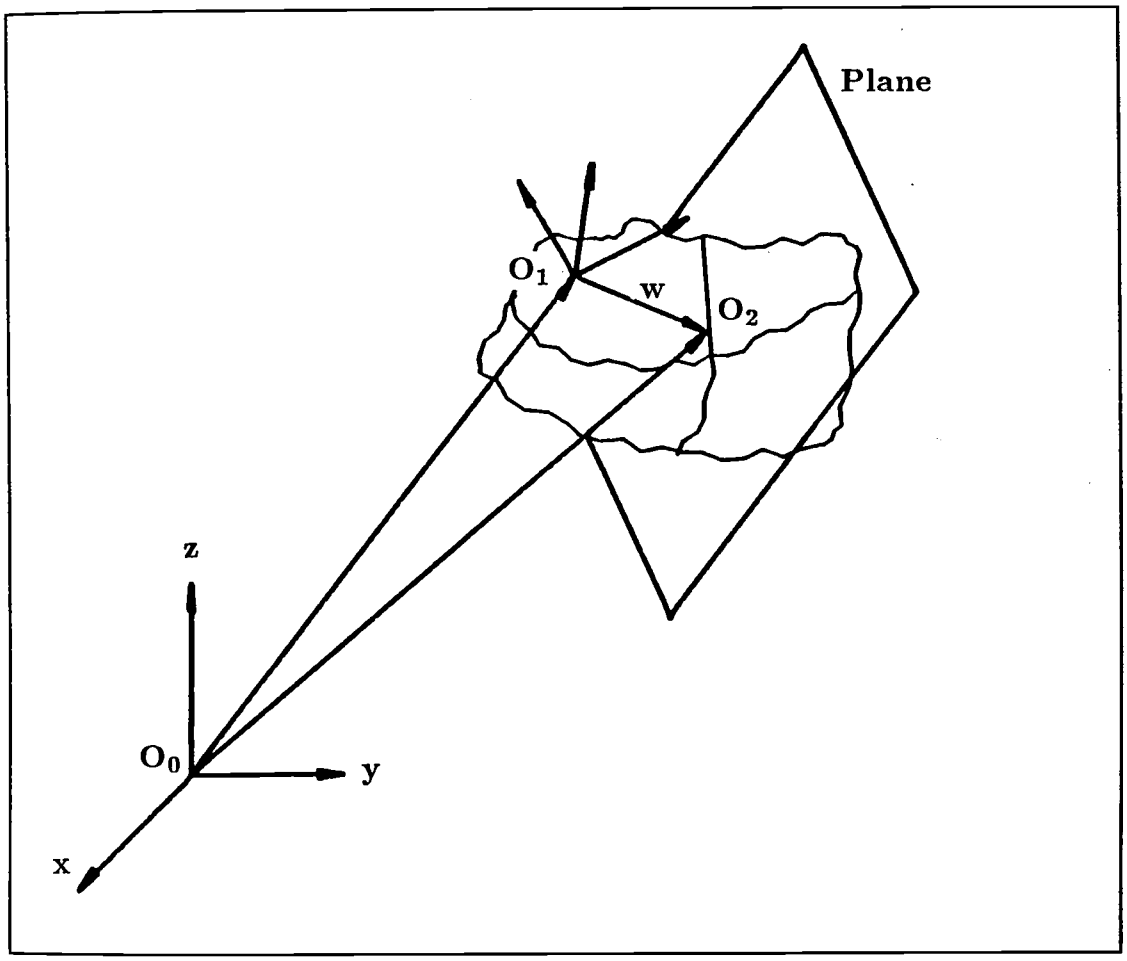


Figure 3.2: Constraints on Two Points: O_1 Fixed, O_2 on a Plane.

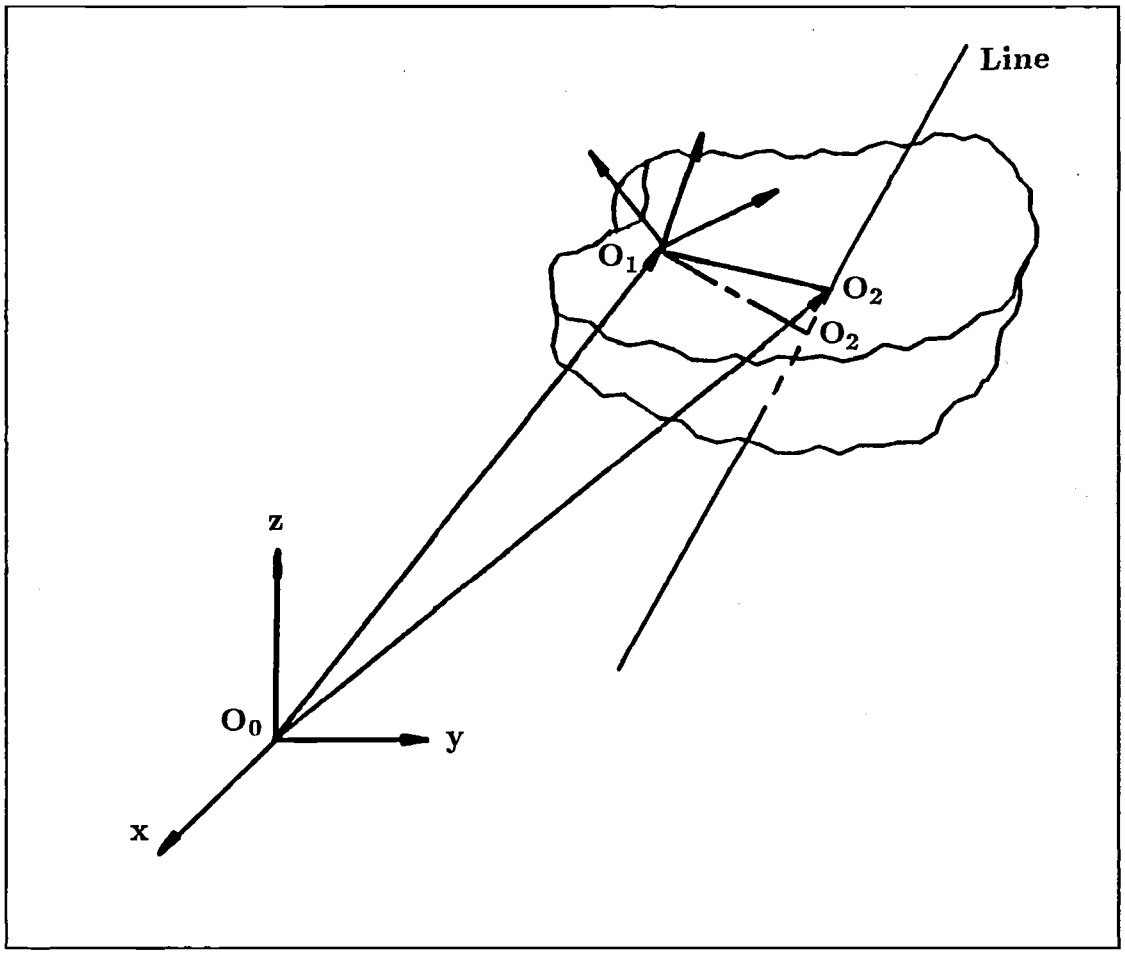


Figure 3.3: Constraints on Two Points: O_1 Fixed, O_2 on a Line.

In Figure 3.3 O_2 must lie on a line fixed in world coordinates. This cuts the spherical locus of O_2 in at most two points. At each of these the body's orientation about O_1O_2 is still undefined.

Even if the positions of O_1, O_2 are given, the body's orientation about O_1O_2 is undefined. The two sets of three cartesian coordinates do not contain as much information as three cartesian coordinates and three Euler angles of a coordinate system at one point. Consideration of cartesian coordinates and relationships between these (e.g. lines and planes) does not specify the position and orientation of the body if only two points are used.

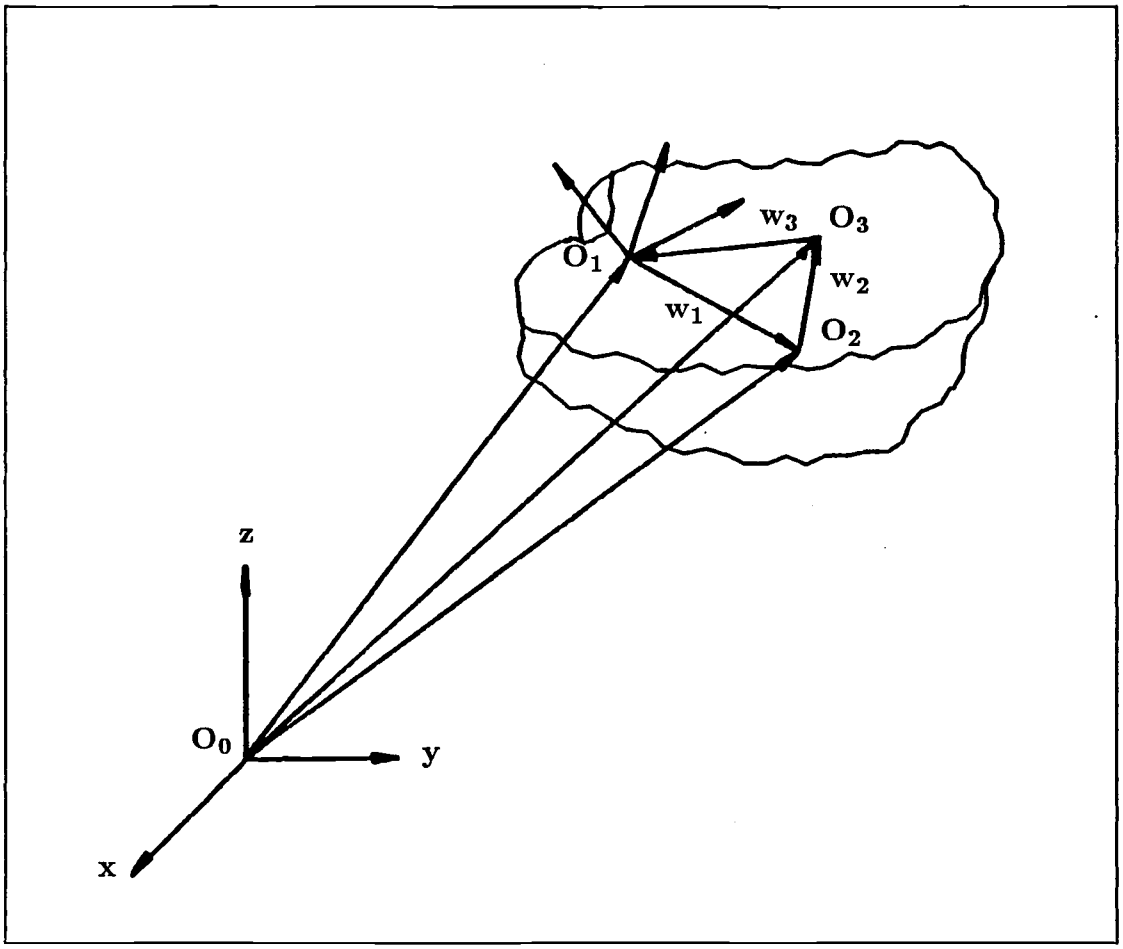


Figure 3.4: Constraints on Three Points.

In Figure 3.4 three points fixed in the body are considered. The three cartesian coordinates of O_1 are known as before. O_2 and O_3 have loci on the spheres of radius $|\mathbf{w}_1|$ and $|\mathbf{w}_3|$ respectively, centred on O_1 . In addition O_3 is on a sphere of radius $|\mathbf{w}_2|$ centred on O_2 .

If O_2 is specified to lie on some plane, its locus is reduced to a circle, but it still has an infinite number of positions. For each one of these the locus of O_3 with respect to O_2 is a sphere and with respect to O_1 another sphere. These intersect in a second circle, again an infinite number of possible positions.

Restricting O_3 as well to lie on some (other) plane does not improve matters

very much. An aid to the imagination here is to think of a set-square in the corner of a box. One vertex, O_1, always touches the same point on the base. O_2 and O_3 must each touch the remaining two sides. This does not fix the position of the set-square.

The starting point of having the three cartesian coordinates of O_1 known is now abandoned. One motive for doing this is that in a real system in which three points are being considered it is likely that similar measured information is available about each point.

If for each point one cartesian coordinate is known or if the three coordinates of the point are related by the equation of a plane in each case then the body may take up an infinity of positions. This may be illustrated by the set-square in the box example. The vertices are now required to touch one surface of the corner each. Obviously it is even less constrained than in the previous example.

- If, however, for each point, two cartesian coordinates are known or if the three coordinates of the point are related by the equation of a line in each case then, in general, there is a solution for the position and orientation of the body.

3.3 Codimension and Constraint on a Body

In Section 2.3, reduction in dimension was considered in respect of a point. Constraint of a point is achieved by the imposition of algebraic varieties. The nature of the variety determines whether, in real space, the dimension is reduced by one, two or three from the original three degrees of freedom of the point. This feature is the *codimension* of the variety [Kendig]. In the case of a point, summing the codimensions of the distinct varieties applied and subtracting this

from three is equivalent to assessing the degrees of freedom left to the point. This can be extended from points to bodies and used to establish whether or not, for instance, the body is completely constrained. This is done as follows.

Initially, with n points under consideration in the body, each point has three degrees of freedom. The total dimension is $3n$. For a rigid body, there are quadratic relationships between the points which are independent of the position and orientation of the body in any particular reference frame. With respect to any point on the body, the other points lie on spherical surfaces at known radii. With $n > 3$, $(3n - 6)$ such equations ensure that the points are fixed relative to each other.

- For a rigid body, six degrees of freedom remain and varieties of total codimension six and different from those within the body are required to fix the body in space. Examples of this are six surfaces, four surfaces and a curve, two surfaces and two curves or three curves, applied to at least three distinct points of the body.

With the knowledge that certain curves or surfaces pass through particular points in a body, the position and orientation of the body may then be calculated. The facility of this calculation depends on the number of indeterminants and the degree of each polynomial equation. Employing all the necessary equations leads quickly to a polynomial of degree greater than four and so in general analytic solutions are not possible. This is true even in the simplest case, in which the measurements from an instrument determine that three points of the body lie on lines. The measured varieties are linear, but the relationships between the points in a rigid body are always quadratic. An algorithm for this case is given in the next section.

3.4 Location of an Object using Three Lines

Three lines, defined in the world reference frame, pass through three points on a body. The relative positions of the three points within the body are known. The problem is to find the position and orientation of the body with respect to the world coordinate system. This is the geometry of a system based on three theodolite measurements, each giving bearings of azimuth and elevation of a different point on an object. Azimuth and elevation give planar and conical surfaces respectively and these intersect in lines, one through each point. This is also the geometry which would be obtained using three of the laser tracking systems developed at the University of Surrey, (described in Chapter 1, section 1.5.12).

One approach, which is similar to that used for solving the kinematic equations of mechanisms, would lead directly to the homogeneous coordinate transformation matrix from the world reference frame to some frame set in the body [Allan]. The transcendental equations obtained in this way are not readily soluble. The method given in the next section does not attempt to find the final transformation directly. Rather it solves for the third coordinate of each of the three points in the body. The transformation from the world frame to the frame fixed in the body may then be found without too much difficulty.

3.4.1 To Find the Third Coordinate of Each Moving Point

In Figure 3.5 O'_n lies on a line in the O_n coordinate system, the azimuth α and elevation β of the line being known. The transformation from spherical to cartesian coordinates is,

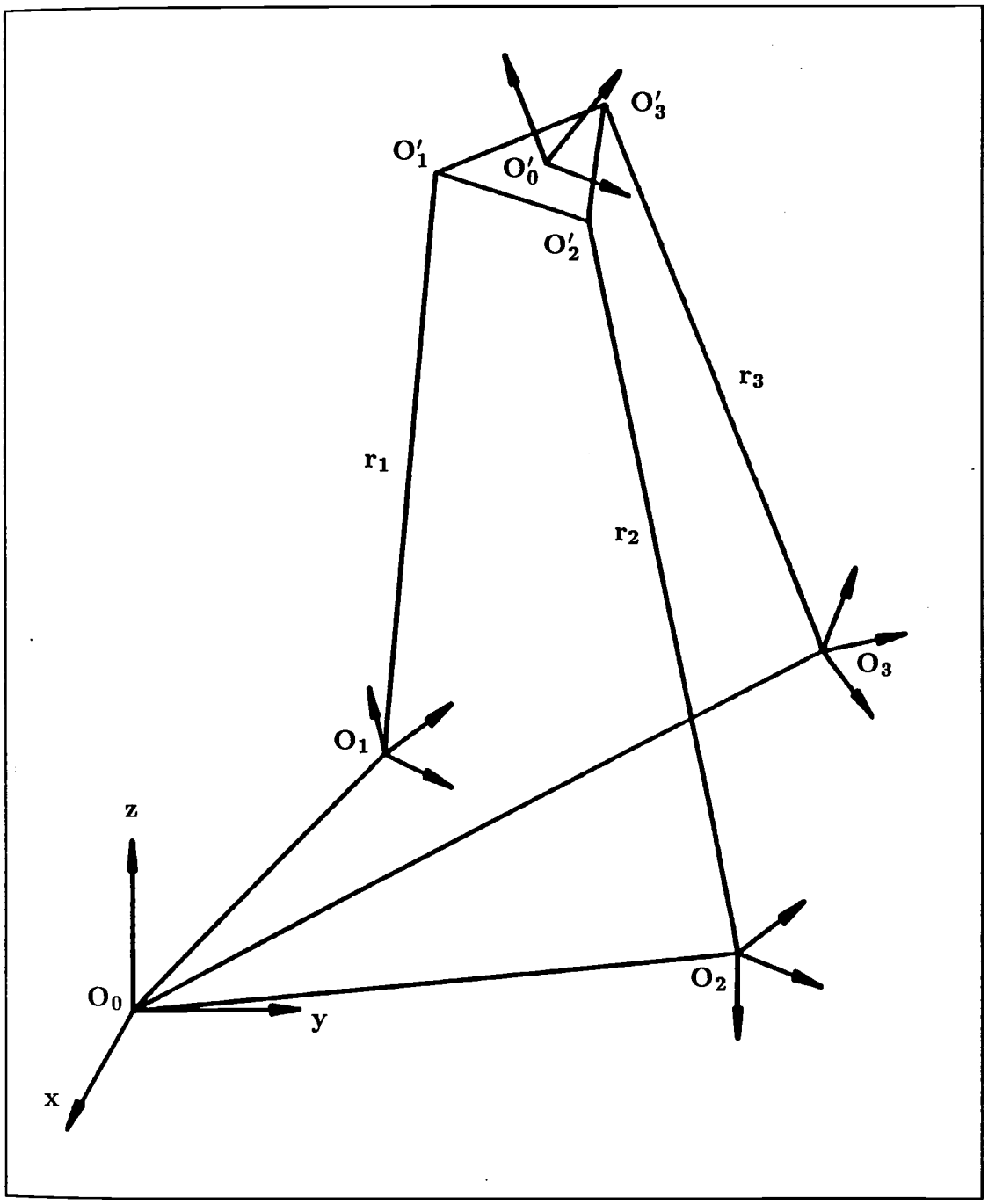


Figure 3.5: Radius Coordinate of each of Three Points.

$$\begin{aligned} x_n &= r_n \cos\alpha_n \sin\beta_n \\ y_n &= r_n \sin\alpha_n \sin\beta_n \\ z_n &= r_n \cos\beta_n \end{aligned} \tag{3.1}$$

Putting,

$$\begin{aligned} \cos\alpha_n \sin\beta_n &= k_{n1} \\ \sin\alpha_n \sin\beta_n &= k_{n2} \\ \cos\beta_n &= k_{n3} \end{aligned} \tag{3.2}$$

the coordinates of O'_n in the O_n system are $r_n k_{n1}, r_n k_{n2}, r_n k_{n3}$ or in homogeneous coordinates O'_n may be represented by the point vector, $\left[k_{n1}\ k_{n2}\ k_{n3}\ \frac{1}{r_n}\right]^T = \mathbf{u_n}$, a column vector in this convention.

The transformation between O_n and O_0 frames is,

$$H_n = \begin{bmatrix} n_{xn} & o_{xn} & a_{xn} & p_{xn} \\ n_{yn} & o_{yn} & a_{yn} & p_{yn} \\ n_{zn} & o_{zn} & a_{zn} & p_{zn} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

and a point vector $\mathbf{u}$ in O_n coordinates is expressed in O_0 coordinates by,

$$\mathbf{v_n} = H_n \mathbf{u_n} \tag{3.4}$$

$$\mathbf{v_n} = \begin{bmatrix} n_{xn} & o_{xn} & a_{xn} & p_{xn} \\ n_{yn} & o_{yn} & a_{yn} & p_{yn} \\ n_{zn} & o_{zn} & a_{zn} & p_{zn} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} k_{n1} \\ k_{n2} \\ k_{n3} \\ \frac{1}{r_n} \end{bmatrix} \tag{3.5}$$

$$\mathbf{v_n} = \begin{bmatrix} k_{n1}n_{xn} + k_{n2}o_{xn} + k_{n3}a_{xn} + \frac{p_{xn}}{r_n} \\ k_{n1}n_{yn} + k_{n2}o_{yn} + k_{n3}a_{yn} + \frac{p_{yn}}{r_n} \\ k_{n1}n_{zn} + k_{n2}o_{zn} + k_{n3}a_{zn} + \frac{p_{zn}}{r_n} \\ \frac{1}{r_n} \end{bmatrix} \tag{3.6}$$

The only unknown in this matrix is the radius r_n, so adding the first three terms in each element to give K_{n1} etc,

$$\mathbf{v}_n = \begin{bmatrix} K_{n1}r_n + p_{xn} \\ K_{n2}r_n + p_{yn} \\ K_{n3}r_n + p_{zn} \\ 1 \end{bmatrix} \tag{3.7}$$

The three vectors $\mathbf{v}_n$ with $n = 1, 2, 3$ are the position vectors in the O_0 coordinate system of the three points in the moving body, in terms of the unknown radius r_n to each point from its O_n origin.

The relationships between points of a rigid body are now used. $\mathbf{w}_n$ are the (free) vectors between pairs of the three moving points.

$$\begin{aligned} \mathbf{w}_1 &= \mathbf{v}_2 - \mathbf{v}_1 \\ \mathbf{w}_2 &= \mathbf{v}_3 - \mathbf{v}_2 \\ \mathbf{w}_3 &= \mathbf{v}_1 - \mathbf{v}_3 \end{aligned} \tag{3.8}$$

Subtracting the component parts of $\mathbf{v}_n$ and putting $p_{x2} - p_{x1} = p_{x21}$ and so on, gives the vectors $\mathbf{w}_n$ in the O_0 world coordinate system in terms of the unknown radii r_1, r_2, r_3.

$$\mathrm{w}_1 = (K_{21}r_2 - K_{11}r_1 + p_{x21})\mathbf{i} + (K_{22}r_2 - K_{12}r_1 + p_{y21})\mathbf{j} + (K_{23}r_2 - K_{13}r_1 + p_{z21})\mathbf{k} \tag{3.9}$$

$$\mathrm{w}_2 = (K_{31}r_3 - K_{21}r_2 + p_{x32})\mathbf{i} + (K_{32}r_3 - K_{22}r_2 + p_{y32})\mathbf{j} + (K_{33}r_3 - K_{23}r_2 + p_{z32})\mathbf{k} \tag{3.10}$$

$$\mathbf{w}_3 = (K_{11}r_1 - K_{31}r_3 + p_{x13})\mathbf{i} + (K_{12}r_1 - K_{32}r_3 + p_{y13})\mathbf{j} + (K_{13}r_1 - K_{33}r_3 + p_{z13})\mathbf{k} \tag{3.11}$$

The lengths $|\mathbf{w}_n|$ between the points O'_n on the body are known, and so using,

$$\mathbf{w_n} \cdot \mathbf{w_n} = w_{nx}^2 + w_{ny}^2 + w_{nz}^2 = |\mathbf{w}_n|^2 \tag{3.12}$$

three quadratic equations are obtained in r_1, r_2, r_3. These are of the form,

$$C_{11}r_2^2 + C_{12}r_2r_1 + C_{13}r_1^2 + C_{14}r_2 + C_{15}r_1 + C_{16} = 0 \tag{3.13}$$

$$C_{21}r_3^2 + C_{22}r_3r_2 + C_{23}r_2^2 + C_{24}r_3 + C_{25}r_2 + C_{26} = 0 \tag{3.14}$$

$$C_{31}r_1^2 + C_{32}r_1r_3 + C_{33}r_3^2 + C_{34}r_1 + C_{35}r_3 + C_{36} = 0 \tag{3.15}$$

all the C_{ij} being known. Two of these equations, e.g. the first and third equation, are used to obtain expressions for r_2, r_2^2, r_3 and r_3^2. These are substituted in the second equation and yield a polynomial of degree 8 in r_1. There is no analytic solution to this and iterative procedures may be employed more efficiently on the three quadratics, as will be shown.

The method which follows is an iterative procedure to find r_1, r_2 and r_3 from the three quadratic equations. The geometrical analogy can be followed through the solution.

In the Equations 3.13, 3.14, and 3.15 the C_{ij} are known and the r_n are the radii which are the third (spherical) coordinates of the three points in the moving body, measured from the O_n origins respectively, (see Figure 3.5). A value is estimated for r_1 and substituted in Equation 3.13. The resulting quadratic in r_2 has, in general, two roots given by,

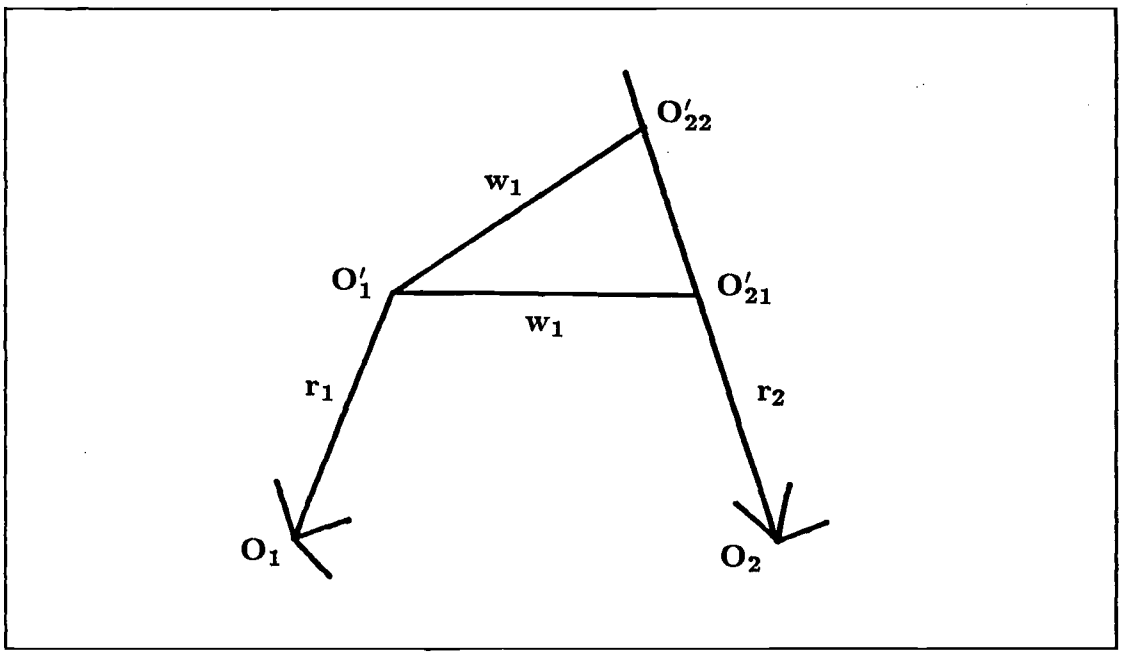


Figure 3.6: Calculation of r_2.

$$r_2 = \frac{-b}{2a} \pm \frac{(b^2 - 4ac)^{\frac{1}{2}}}{2a} \tag{3.16}$$

Referring to Figure 3.6, with r_1 given and the knowledge that the distance between O'_1 and O'_2 is w_1 (known) then O'_2 lies on one of the two points where a sphere of radius w_1 centred on O'_1 is cut by the line on which r_2 lies. Thus there are either one or two real roots r_2 which meet this condition, giving at most two possible solutions for O'_2, i.e. O'_{21} and O'_{22}. Each value of r_2 in turn is substituted in Equation 3.14 to give two quadratics in r_3. Both quadratics are then solved for r_3, deriving two roots from each at most, i.e. four possible values of r_3.

Referring to Figure 3.7, for each calculated r_2 there are two possible lines of length w_2 between O'_2 and O'_3, giving four possible positions of O'_3 which satisfy the equations so far, O'_{31}, O'_{32}, O'_{33} and O'_{34}.

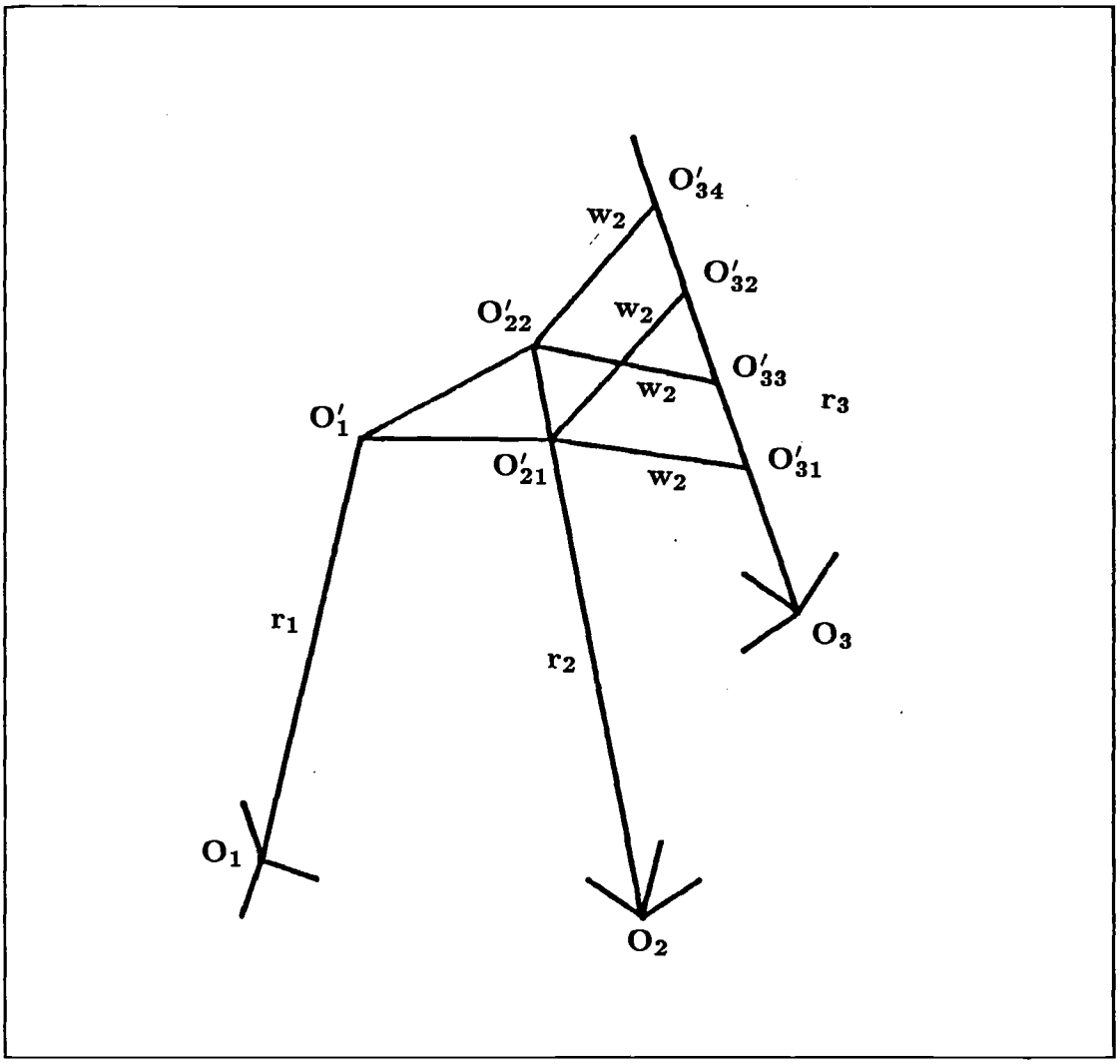


Figure 3.7: Calculation of r_3.

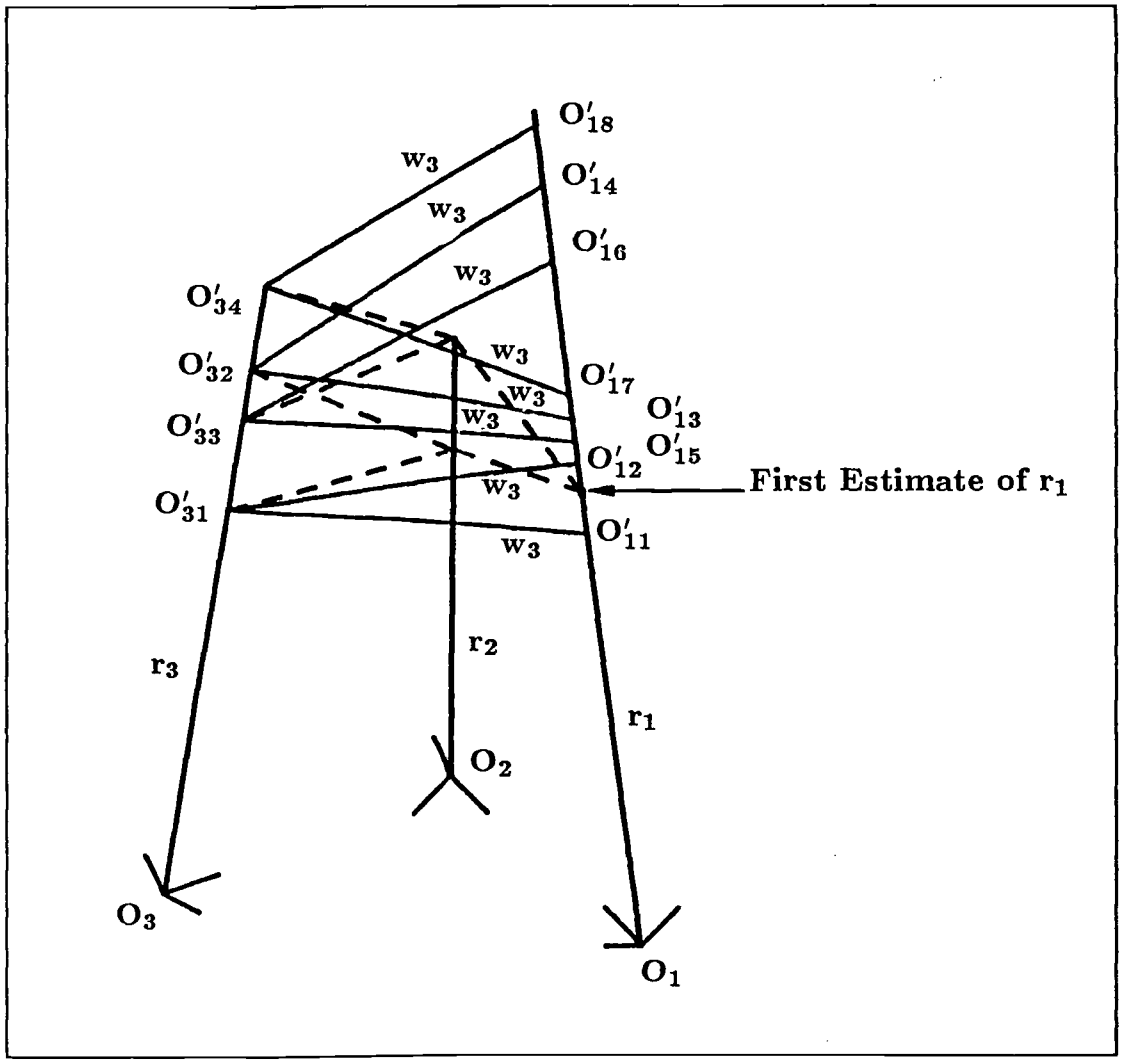


Figure 3.8: Calculation of r_1.

Each value of r_3 in turn is then substituted in Equation 3.15. This gives four quadratic equations in r_1 and therefore eight possible values of r_1. The geometry is shown in Figure 3.8. With the four O'_3 as starting points, there are a total of eight lines of the required length w_3 which meet the line on which r_1 lies. If, as is illustrated in Figure 3.8 none of the eight calculated values of r_1 match the original estimate, then r_1 is adjusted and the procedure repeated until a match to within acceptable tolerance is obtained.

If the correct path were known, that is if the choice between roots of quadratics at each of the three stages had been established, then to improve the estimate of r_1, the mean between the previous value and the resulting calculated value could be taken, converging on the correct r_1 and giving the correct r_2 and r_3 in the process. The first task is therefore to establish the correct series of choices of $+$ or $-$ sign in selecting roots of the quadratics. Since the body moves continuously, a different path may only come about after passing through the situation in which two identical roots are derived from a quadratic. Geometrically, this corresponds to the line through one of the points touching the sphere, centred on one of the other points tangentially. Alternatively it can be seen as the line between two points, O_n and O_{n+1} say, being perpendicular to r_{n+1}.

However, since in reality the values of the coefficients C_{ij} in Equations 3.13, 3.14, and 3.15 are based on measurements of azimuth and elevation at the O_n frames, sampled at some frequency, the data available will not be continuous. Some judgement must then be applied to decide how near equal a pair of roots may become without rechecking that the correct choice of roots is being employed. This judgement will be based on the velocity of the body, the sampling frequency and the rounding errors which occur in computation. Finally, as the computations described above are carried out for each successive set of sampled data, estimates may be obtained by employing previous position and time values to calculate $\dot{r}$ and $\ddot{r}$.

3.5 Location of an Object using Three Quadric Surface Intersection Curves

The coordinate of the interferometer picks out a hyperboloid from a family of hyperboloids of revolution (of two sheets) with the same foci. As was described

in Chapter 2, Section 2.4, two such surfaces intersect in a space curve. With two coordinates available for a point, the point lies on this quadric surface intersection curve. With two coordinates of each of three points in a body the position and orientation of the body can be found with respect to a fixed reference frame by using three quadric surface intersection curves.

The method (although not the geometry) is similar to the case in the last section in which the three curves were lines each passing through one of the points and the third coordinate was found for each point. This method (which has not been implemented) is given below.

For one of the curves, a line on a parameterisation surface of the pencil is selected and intersected with one of the hyperboloids to obtain the coordinates of one point on that curve. If one point of the body is situated at this point on the curve, then the second body point lies on the second curve such that it is a known distance from the first point. Using the quadratic equation which gives the radius from the first to second point would, in combination with the fourth degree equation of the second curve (if that were known) give a polynomial of degree eight.

The intersection of this sphere with the second curve must therefore be found by scanning through the ruled lines of a parameterisation surface of the second curve to obtain candidate points whose radius from Point 1 can easily be checked. The number of points on the second curve which give the correct distance from Point 1 depends on the nature of the curve and the relative positions of Curves 1 and 2 at that instant.

For each successful candidate for Point 2 the procedure is carried out again, this time for possible positions of Point 3 on Curve 3 with respect to Point 2. Then for the third side of the triangle of points the radii from successful Points

3 are taken to possible Points 1 on Curve 1.

As with the case of the three straight lines, if the correct starting point has been chosen on Curve 1, then one of the finishing points is this same point and all three cartesian coordinate for all three points have been found.

Also, as in the case with three lines, multiple solutions are possible, but in a dynamic situation since each sampled set of coordinates is likely to be numerically close to the previous sampled set, once the first multiple choice has been resolved good estimates can be made for subsequent iterations. This speeds up the choice of parametric lines for each curve.

Two possible positions on a curve which satisfy the radius criterion (from the previous point) only approach each other closely on the curve as a tangent plane to the intersecting sphere approaches the point where it includes a tangent to the curve. At this, the two solutions merge into one. When they separate again (as the body keeps moving) the correct choice of solution has been lost. In practice this problem may be overcome by careful positioning of instruments with respect to allowable motions of the object and/or a redundancy of points or instrumentation.

Chapter 4

Diffraction and Interference

4.1 Introduction

In this chapter, the physical background to the interferometer described in Chapter 1, 1.6 is presented. Sections 4.2 to 4.6 describe diffraction and interference due to single and multiple rectangular slits with off-axis incident radiation. These effects form the basis of the interferometer. The consequences of introducing a lens behind the diffracting aperture are considered and, in Sections 4.8 to 4.10, the characteristics of a lens are related to the parameters of the diffraction equations.

4.2 Diffraction at an Aperture

Rigorous solutions of diffraction problems only exist for a few special cases such as the diffraction of monochromatic plane waves by a half plane, (see [Born and Wolf]). There is no exact solution for the diffraction of plane or spherical waves by a rectangular aperture, and so to obtain expressions for the electric field strength due to diffracted radiation beyond such an aperture, various approaches have been tried. Historically[1] these have ranged from the Huygens' construction and Kirchhoff's mathematical formulation of Huygens'

theorem, through the approximate far-field solution of Fraunhofer, the Fresnel integral approximation for near-field diffraction, and Sommerfeld's modification to Kirchhoff's approach, to the more recent return to a geometrical description of diffraction as given by [Keller] or the asymptotic expansion of diffraction integrals (see [Born and Wolf] and [Kline and Kay]).

The Rayleigh-Sommerfeld formula is used in the discussion which follows, in preference to the Kirchhoff[2] formula, which assumes certain mutually incompatible boundary conditions at the aperture (see [Goodman, Ch 3]). In practice the predictions of both formulae are frequently quite similar and Kirchhoff's theory has been much used, (see for instance [Papoulis]). The diffraction problems considered here involve diffraction at an aperture in a plane screen only, and so the Green's function used for Sommerfeld's formula is appropriate, (see [Bouwkamp] for a discussion of modifications to Kirchhoff's scalar diffraction theory). In addition to its self-consistency, the Rayleigh-Sommerfeld formula has the advantage of being slightly simpler.

The Rayleigh-Sommerfeld formula for diffraction of monochromatic radiation at an aperture gives the electric field E at a point P beyond the aperture as,

$$E = \frac{1}{i\lambda} \iint_{\Sigma} E(x,y) \frac{e^{i(kr_1 - \omega t)}}{r_1} \cos\psi \, dx \, dy \tag{4.1}$$

where $E(x,y)$ is the field over the aperture Σ due to incident radiation, ψ is the angle to the z axis and r_1 is the distance to the observation point. See Figure 4.1. Conventionally, three sets of approximations (as described in, for instance, [Goodman]) may then be carried out.

[1] For an introduction to the historical development of diffraction theory see [Goodman] or [Born and Wolf].

[2] See [Wolf and Marchand] for a comparison of the predictions of the two theories.

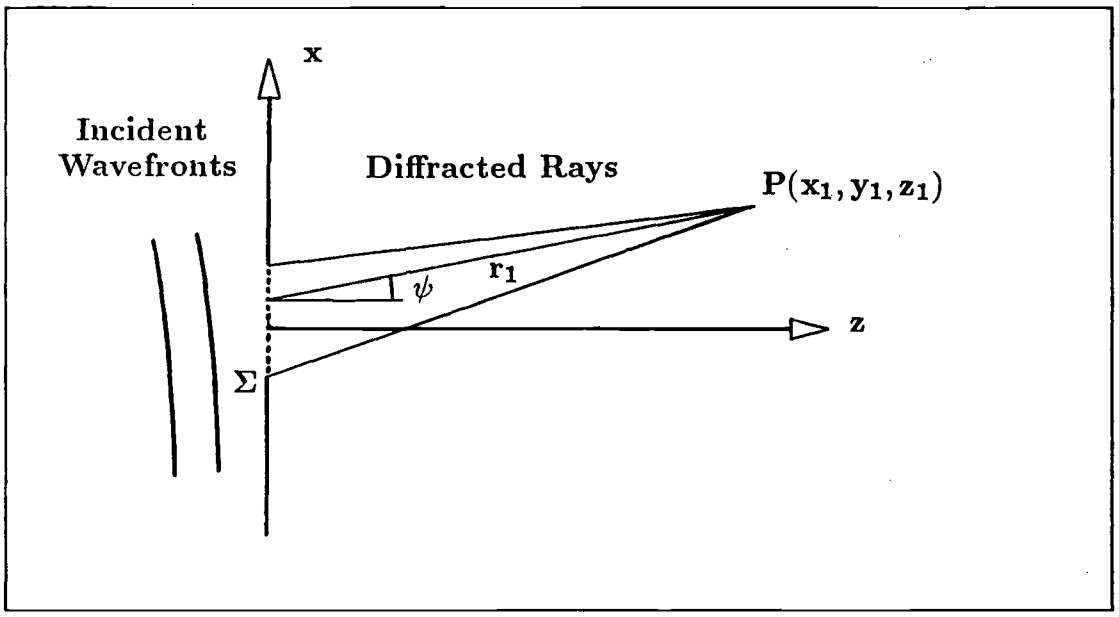


Figure 4.1: Diffraction at an Aperture in the x, y Plane.

4.2.1 The Paraxial Approximation

In Equation 4.1 the amplitude of each diffracted ray from the aperture to the observation point is modified by an obliquity factor $\cos\psi$ for that ray. In the paraxial region $\cos\psi \approx 1$ and so can be taken out of the equation. Further , in this region $r_1 \approx z_1$ assuming an observation plane parallel to the aperture plane, and so the r_1 in the denominator can be taken outside the integral. This leaves, with r_{10} the distance from (0,0) in the aperture to P on the observation plane and ignoring the time dependency,

$$E = \frac{1}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{ikr_1}\, dx\, dy \tag{4.2}$$

4.2.2 The Fresnel Approximation

In the exponent of Equation 4.2 r_1 cannot be put equal to z_1 since the phase at P varies rapidly with r_1. A better approximation to r_1 utilises the binomial

expansion and with $l_1 = \frac{x_1}{r_{10}}$ and $m_1 = \frac{y_1}{r_{10}}$ the direction cosines of the point P, gives,

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{ik\left[\frac{(x^2+y^2)}{2r_{10}} - (xl_1 + ym_1)\right]} \, dx \, dy \tag{4.3}$$

See Appendix A.1.1.

4.2.3 The Fraunhofer Approximation

At long distances from the aperture the Fraunhofer approximation can be used.

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{-ik(xl_1 + ym_1)} \, dx \, dy \tag{4.4}$$

See Appendix A.1.2.

4.3 Diffraction at a Rectangular Aperture: Incident Plane Waves

The diffraction integral, Equation 4.4 is now evaluated for a rectangular aperture with sides at $x = \pm a$, $y = \pm b$, on the plane $z = 0$. The incident radiation is plane wavefronts whose normals have direction cosines l, m, n, to the x, y, z axes.

$$E(x,y) = E_0 e^{ikr} = E_0 e^{-ik(xl+ym)} \tag{4.5}$$

and Equation 4.4 becomes,

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{-a}^{a} \int_{-b}^{b} E_0 e^{-ik(xl+ym)} e^{-ik(xl_1+ym_1)} \, dx \, dy \tag{4.6}$$

which gives,

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 4ab \operatorname{sinc}\left[ka(l + l_1)\right] \operatorname{sinc}\left[kb(m + m_1)\right] \tag{4.7}$$

Irradiance I is found from,

$$I = EE^* \tag{4.8}$$

$$I = \frac{E_0^2}{\lambda^2 r_{10}^2} 16a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)] \tag{4.9}$$

See Appendix A.2.

4.4 Diffraction at 2 Rectangular Apertures: Plane Waves at Normal Incidence

The superimposition of diffraction fields from two rectangular apertures Σ_1 and Σ_2 gives rise to a higher frequency interference term in the irradiance function. Initially with normally incident plane waves and apertures with edges at $(A \pm a)$ and $(-A \pm a)$ in x and $\pm b$ in y, then proceeding from the Fraunhofer approximation as expressed in Equation 4.4,

$$E = E_1 + E_2 \tag{4.10}$$

and with $E_{10}(x,y)$ and $E_{20}(x,y)$ the fields at the two apertures,

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma_1} E_{10}(x,y) e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.11}$$

Here $E_{10} = E_{20} = E_0 e^{-i\omega t}$ and with $e^{-i\omega t}$ understood,

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{-A-a}^{-A+a} \int_{-b}^{b} E_0 e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.12}$$

giving the field at P due to Σ_1,

$$E_1 = \frac{E_0 e^{ikr_{10}} e^{ikl_1 A}}{i\lambda r_{10}} 4ab \operatorname{sinc}[kl_1 a] \operatorname{sinc}[km_1 b] \tag{4.13}$$

Similarly for the second aperture Σ_2,

$$E_2 = \frac{E_0 e^{ikr_{10}} e^{-ikl_1 A}}{i\lambda r_{10}} 4ab \operatorname{sinc}[kl_1 a] \operatorname{sinc}[km_1 b] \tag{4.14}$$

See Appendix A.3 for details.

In this case, the electric fields E_1 and E_2 add to give,

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 8ab \operatorname{sinc}[kl_1 a] \operatorname{sinc}[km_1 b] \cos[kl_1 A] \tag{4.15}$$

This does not happen quite so neatly with non-planar incident radiation and so for later comparison the irradiance is found from,

$$I = EE^* = E_1 E_1^* + E_2 E_2^* + E_2 E_1^* + E_1 E_2^* \tag{4.16}$$

Taking each of the four terms in turn,

$$E_1 E_1^* = \frac{E_0^2 16a^2 b^2 \operatorname{sinc}^2[kl_1 a] \operatorname{sinc}^2[km_1 b]}{\lambda^2 r_{10}^2} \tag{4.17}$$

$$E_2 E_2^* = \frac{E_0^2 16a^2 b^2 \operatorname{sinc}^2[kl_1 a] \operatorname{sinc}^2[km_1 b]}{\lambda^2 r_{10}^2} \tag{4.18}$$

$$E_2 E_1^* = \frac{E_0^2 16a^2 b^2 \operatorname{sinc}^2[kl_1 a] \operatorname{sinc}^2[km_1 b] e^{-2ikl_1 A}}{\lambda^2 r_{10}^2} \tag{4.19}$$

$$E_1 E_2^* = \frac{E_0^2 16a^2 b^2 \operatorname{sinc}^2[kl_1 a] \operatorname{sinc}^2[km_1 b] e^{2ikl_1 A}}{\lambda^2 r_{10}^2} \tag{4.20}$$

Adding Equations 4.17 to 4.20 gives,

$$= \frac{E_0^2 16a^2 b^2 \operatorname{sinc}^2[kl_1 a] \operatorname{sinc}^2[km_1 b]}{\lambda^2 r_{10}^2} \left[2 + e^{-2ikl_1 A} + e^{2ikl_1 A}\right] \tag{4.21}$$

$$I = \frac{E_0^2 32a^2b^2 \operatorname{sinc}^2[kl_1a] \operatorname{sinc}^2[km_1b]}{\lambda^2 r_{10}^2} \left[1 + \frac{e^{2ikl_1A} + e^{-2ikl_1A}}{2}\right] \tag{4.22}$$

$$I = \frac{E_0^2 64a^2b^2}{\lambda^2 r_{10}^2} \operatorname{sinc}^2[kl_1a] \operatorname{sinc}^2[km_1b] \cos^2[kl_1A] \tag{4.23}$$

In Equation 4.22 the high frequency cosine modulation is added to the left hand term i.e. irradiance is always positive. Equation 4.23 gives this in the familiar form of $\cos^2$ of half the argument. This represents the interference fringes within a sinc^2 envelope. The arguments of both cosine and sinc functions depend on the direction cosines of the image point, l_1 and m_1.

4.5 Diffraction at 2 Rectangular Apertures: Incident Plane Waves

In this case, incident plane waves have normals with direction cosines l, m, n. From Equation 4.4 again, and with,

$$E = E_1 + E_2 \tag{4.24}$$

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma_1} E_{10}(x,y) e^{-ik(xl_1 + ym_1)}\, dx\, dy \tag{4.25}$$

remembering that r_{10} is a function of angle to the optical centre line and the position of points on the observation surface.

With non-axial plane waves, the field at the screen is $E_0 e^{-ik(xl+ym)}$, neglecting time dependency.

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{-A-a}^{-A+a} \int_{-b}^{b} E_0 e^{-ik(xl+ym)} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.26}$$

giving the field at P due to Σ_1,

$$E_1 = \frac{E_0 e^{ikr_{10}} e^{ikA(l+l_1)}}{i\lambda r_{10}} 4ab \operatorname{sinc}[ka(l+l_1)] \operatorname{sinc}[kb(m+m_1)] \tag{4.27}$$

Similarly, for the second aperture Σ_2,

$$E_2 = \frac{E_0 e^{ikr_{10}} e^{-ikA(l+l_1)}}{i\lambda r_{10}} 4ab \operatorname{sinc}[ka(l+l_1)] \operatorname{sinc}[kb(m+m_1)] \tag{4.28}$$

See Appendix A.4 for details. The E_1 and E_2 contributions add to give,

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 8ab \operatorname{sinc}[ka(l+l_1)] \operatorname{sinc}[kb(m+m_1)] \cos[kA(l+l_1)] \tag{4.29}$$

The irradiance is now found,

$$I = E_1 E_1^* + E_2 E_2^* + E_2 E_1^* + E_1 E_2^* \tag{4.30}$$

Taking each of the four terms in turn,

$$E_1 E_1^* = \frac{E_0^2 16a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)]}{\lambda^2 r_{10}^2} \tag{4.31}$$

$$E_2 E_2^* = \frac{E_0^2 16a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)]}{\lambda^2 r_{10}^2} \tag{4.32}$$

$$E_2 E_1^* = \frac{E_0^2 16a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)] e^{-2ikA(l+l_1)}}{\lambda^2 r_{10}^2} \tag{4.33}$$

$$E_1 E_2^* = \frac{E_0^2 16a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)] e^{2ikA(l+l_1)}}{\lambda^2 r_{10}^2} \tag{4.34}$$

Adding Equations 4.31 to 4.34 gives,

$$I = \frac{E_0^2 16a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)]}{\lambda^2 r_{10}^2} \left[2 + e^{2ikA(l+l_1)} + e^{-2ikA(l+l_1)}\right] \tag{4.35}$$

$$I = \frac{E_0^2 32a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)]}{\lambda^2 r_{10}^2} [1 + \cos 2kA(l+l_1)] \tag{4.36}$$

or

$$I = \frac{E_0^2 64a^2b^2}{\lambda^2 r_{10}^2} \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)] \cos^2[kA(l+l_1)] \tag{4.37}$$

In Equation 4.37 the envelope of amplitude I is the product of sinc^2 functions of l_1 and m_1 for diffraction patterns due to the long and short sides of the slits respectively. Within this envelope, interference fringes are described by the $\cos^2 kA(l+l_1)$ term, shifted in the same way as the envelope. Thus for a given set of incidence direction cosines, l, m, the phase of the $\cos^2$ term on the observation surface is determined by a constant kAl plus kAl_1, l_1 being a particular direction cosine in the image field.

4.6 Diffraction at 2 Rectangular Apertures: Incident Spherical Waves

A point source S at coordinates x_0, y_0, z_0 emits spherical wavefronts which are incident on the aperture screen. The electric field at the plane of the screen due to this source with strength $\mathcal{E}$ and at radius r from the source is, neglecting time,

$$E(x,y) = \frac{\mathcal{E}e^{ikr}}{r} \tag{4.38}$$

The diffracted field at $P(x_1, y_1, z_1)$ beyond the aperture is,

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.39}$$

employing paraxial, Fresnel and Fraunhofer approximations as per Equation 4.4.

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} \frac{\mathcal{E}e^{ikr}}{r} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.40}$$

Assuming that the distance r from S to the points of the aperture does not vary much and is $\approx r_0$, the distance from S to the origin,

$$E = \frac{\mathcal{E}e^{ikr_{10}}}{i\lambda r_0 r_{10}} \iint_{\Sigma} e^{ikr} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.41}$$

A better approximation for r in the the exponent is,

$$r \approx r_0 + \frac{(x^2+y^2)}{2r_0} - \frac{(xx_0+yy_0)}{r_0} \tag{4.42}$$

At this point, to make the approximation that the second term on the right is negligible is to restrict the source position to be a very long distance from the aperture. This would be in addition to the assumptions already made in which the observation point is far distant from the aperture. If this is done then the spherical waves approximate to plane waves from the l_0, m_0, n_0 direction and,

$$E = \frac{\mathcal{E}e^{ik(r_0+r_{10})}}{i\lambda r_0 r_{10}} \iint_{\Sigma} e^{-ik(xl_0+ym_0)} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.43}$$

The diffracted E-field is thus similar to that due to off-axis plane waves and two rectangular slit apertures as in Equation 4.29. The irradiance pattern will also be similar, i.e. two sinc^2 functions and a $\cos^2$ function as per Equation 4.37.

It may not be desirable to make such approximations however. In Chapter 5 some of the approximations behind the aperture are seen to be unnecessary when a lens is used. The source position is then allowed to be quite close to the aperture and the diffraction and interference patterns are investigated.

4.7 Optical Path Length of Diffracted Rays

The previous two sections have shown that, in the Fraunhofer approximation, the irradiance pattern due to interference includes a $\cos^2$ term within the sinc^2 envelope. When plane wave illumination is used, the approximation requires that the observation point be distant from the aperture. With illumination by spherical wavefronts, both the source and the observation points are required to be distant from the aperture.

The observation distance can be reduced by placing a convex lens behind the aperture. This forms the basis of an instrument in which sensed movement of the $\cos^2$ fringes provides information about changes in the incident field.

Returning to the Rayleigh-Sommerfeld diffraction formula of Equation 4.1 but with the time dependency ignored,

$$E = \frac{1}{i\lambda}\iint_{\Sigma} E(x,y)\frac{e^{ikr_1}}{r_1}\cos\psi\, dx\, dy \tag{4.44}$$

in free space (or air) the contribution to the field at a point downstream of the aperture from a point (x, y) in the aperture is simply the field at the point (x, y) attenuated as $r_1^{-1}\cos\psi$ and with phase determined by e^{ikr_1}.

The optical path length [OPL] from an aperture point to the point P is r_1. Optical path lengths are indicated by square brackets. In free space $[r_1] = r_1$ and can be expressed simply in terms of the (x, y) coordinates (x_1, y_1) of P. With

a multiple element lens in place, however, the [OPL] from (x, y) to $P(x_1, y_1)$ is composed of several, perhaps many straight line segments, some in air, some in glass. To find an expression for $[r_1]$ in terms of x, y, x_1, y_1 would be a complicated task and a solution would be specific to the particular lens geometry used.

- The lens can, however, be considered as a system from which each optical path $[r_1]$ can be determined without reference to the geometry of individual lens elements.

4.8 Lens Model

The two subsections below describe the idealised model of a lens which is used, in geometrical and wave terms respectively.

4.8.1 Geometrical Model

An ideal lens operates on an incident collimated beam of light to focus all rays of that beam to a single point in its image space. As the angles of incidence are varied, the focused point sweeps out a surface in image space, the focal surface. This works from left to right or from right to left, but the surfaces are not assumed to be the same. The left and right focal surfaces are identified as g_1 and g_2 respectively. See Figure 4.2.

Two nodal points N_1 and N_2 lie on the Optical Centre Line (OCL) such that with an incident ray on a line passing through N_1 at ψ to the OCL, there exists a corresponding line through N_2 at ψ to the OCL, which picks out the ultimate destination of the incident ray (and thus of all incident rays parallel to that ray) on the focal surface. The surfaces g_1 and g_2 cut the OCL at F_1 and F_2 respectively.

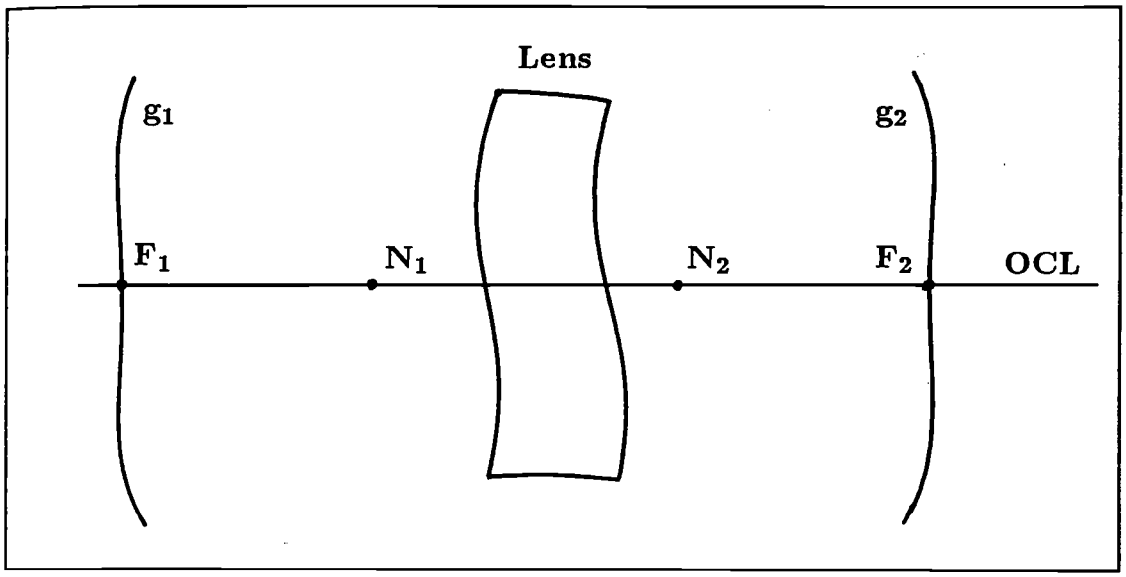


Figure 4.2: Lens and Focal Surfaces.

4.8.2 Phase Model

Phase is taken into account. Incident parallel rays with plane wavefronts are refracted through the lens to produce spherical wavefronts converging to the image point. The OPLs of all rays from one wavefront to a second wavefront are equal. Thus the OPL from any point on a plane, along a ray initially normal to that plane, to their common image point is constant. Even without this order imposed on the incident rays, it is a matter of geometry that the OPL from a plane in object space to its corresponding image point is constant for all points on that plane, whether it is a physical wavefront or not.

4.9 Non-paraxial Diffracted Rays with a Lens

In the paraxial approximations in Section 4.2, the obliquity factor $\cos\psi$ was taken as ≈ 1, making the integrand of Equation 4.1 more manageable. Without a lens, Figure 4.1, each diffracted ray contributing to the field at P (on some

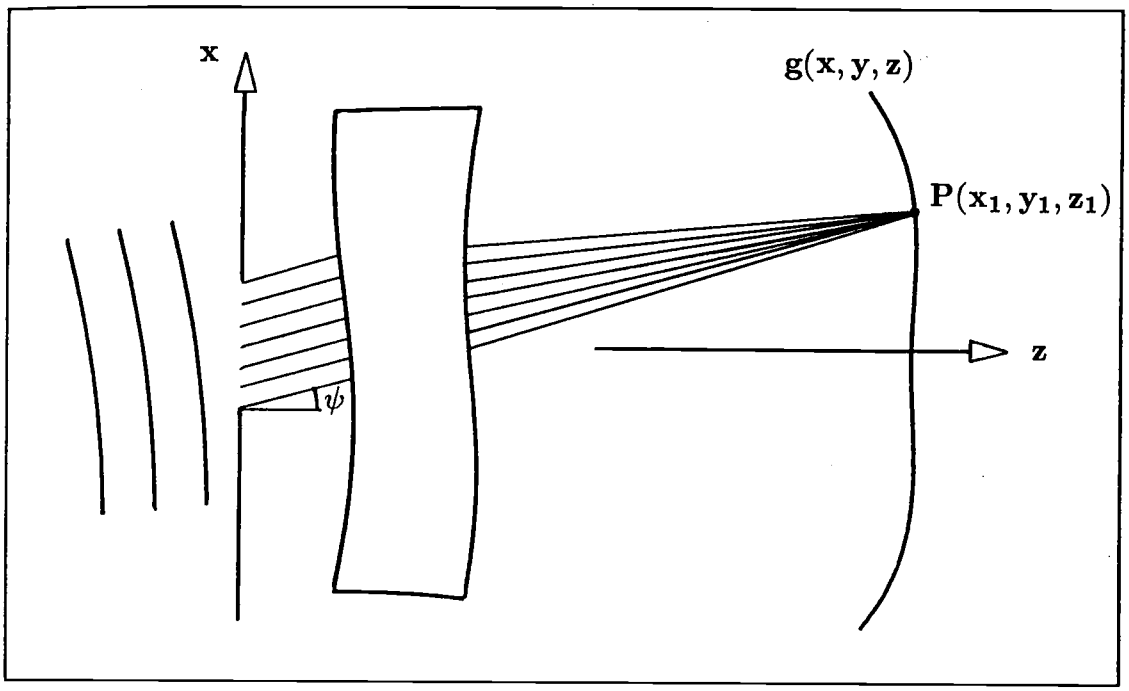


Figure 4.3: Parallel Diffracted Rays through a Lens to Image Point.

meridional plane) has a different value of ψ. With a lens behind the aperture as in Figure 4.3, only incident parallel rays are focused at P and so for some $P(x_1, y_1)$, $\cos\psi$ on the object side of the lens is constant and can be taken out of the integral. Thus the obliquity factor is retained and with a lens the Rayleigh-Sommerfeld formula becomes,

$$E = \frac{\cos\psi}{i\lambda} \int\int_{\Sigma} E(x,y) \frac{e^{ik[r_1]}}{[r_1]}\, dx\, dy \tag{4.45}$$

This can be represented in terms of the free space OPL r_1 as in Figure 4.4. The distance r_1 from any point (x, y) in the aperture to P is a function of the direction of the rays and the focal surface function for the lens. The focal surface need not be a plane.

Each contribution to the summation at the plane Σ_1 representing P is not a segment of the same plane wave but a component of one of many spherical

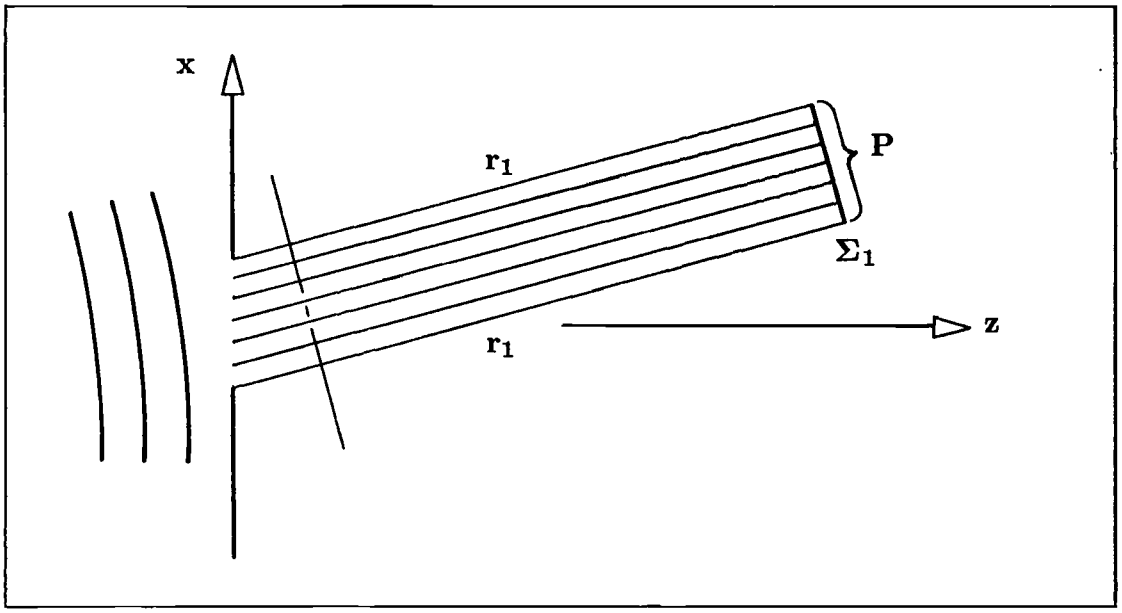


Figure 4.4: Free Space OPL of Parallel Diffracted Rays through a Lens to Image Point P.

waves, and so is attenuated with r_1^{-1}. Also, in general, phase is not constant across any plane normal to the wave direction and the double integral of the electric field contributions over the aperture projection onto that plane will vary as the plane is moved in the "ray" direction. Figure 4.5 shows a low frequency of E across Σ_1, in two dimensions for illustration. As the section x' is moved, the integral of E with respect to x' from x_1' to x_2' varies.

In three dimensions, referring to Figure 4.6, the distance r_1 from each aperture point to the summation plane for a particular direction is obtained as follows. With $\hat{\mathbf{r}}_1$ a unit vector in the $\mathbf{r}_1$ direction, $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$, unit vectors in the coordinate axis directions and l_1, m_1, n_1 the direction cosines of $\mathbf{r}_1$,

$$\hat{\mathbf{r}}_1 = l_1\hat{\mathbf{i}} + m_1\hat{\mathbf{j}} + n_1\hat{\mathbf{k}} \tag{4.46}$$

and with $\mathbf{q}$ the position vector of points in the aperture,

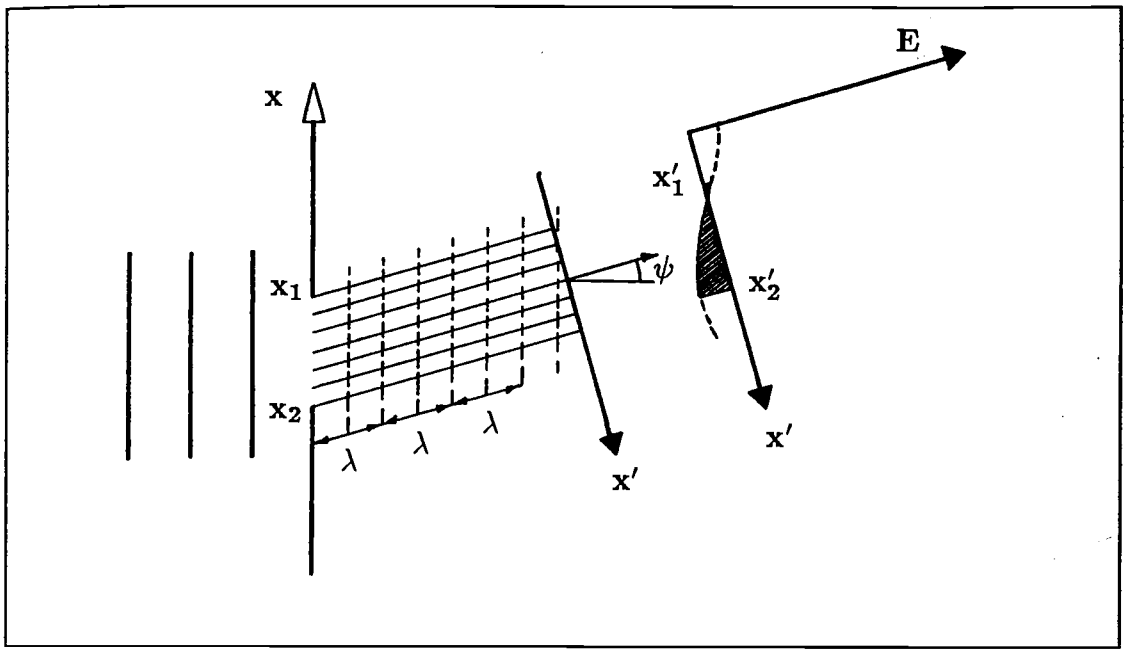


Figure 4.5: Summation of E-field Contributions at a Cross-section x' Normal to Direction ψ.

$$r_1 = r_{10} - \mathbf{q}.\hat{\mathbf{r}}_1 \tag{4.47}$$

$$\mathbf{q} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} \tag{4.48}$$

$$\mathbf{q}.\hat{\mathbf{r}}_1 = xl_1 + ym_1 \tag{4.49}$$

$$r_1 = r_{10} - (xl_1 + ym_1) \tag{4.50}$$

The summation at P is,

$$E = \frac{\cos\psi}{i\lambda}\iint_{\Sigma} E(x,y)\frac{e^{ikr_1}}{r_1}\,dx\,dy \tag{4.51}$$

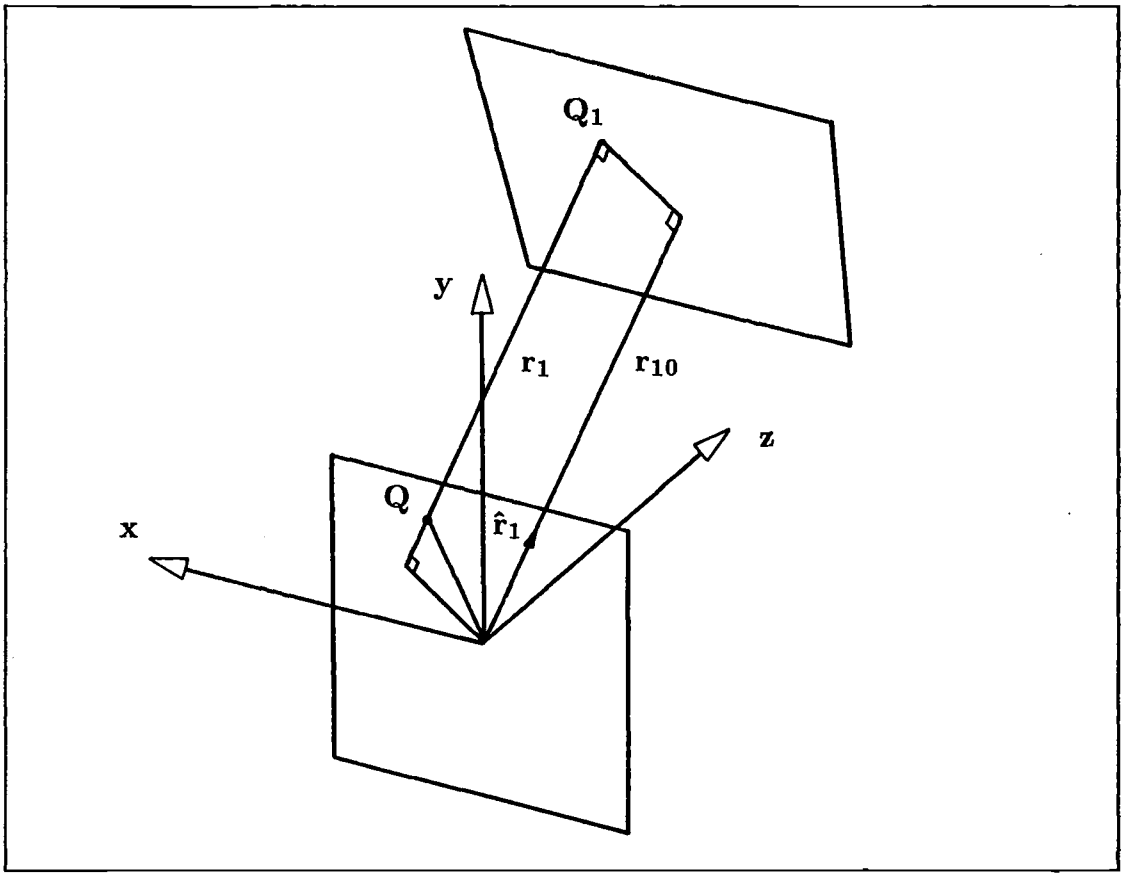


Figure 4.6: Optical Path from Aperture to Image Point.

and since $r_1 = r_{10} - (xl_1 + ym_1)$, where r_{10} is the distance from the aperture origin to the image point,

$$E = \frac{\cos\psi e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{4.52}$$

taking the $r_1 \approx r_{10}$ attenuation in the denominator outside the integral and bearing in mind that r_{10} is a function of angles to the image point. This expression for E, obtained with the lens model, is similar to the Fraunhofer approximation for diffraction observed at a great distance, the difference being that it now includes the obliquity factor, $\cos\psi$, which would have been put equal to unity in a paraxial approximation.

In the calculation of electric field, the determination of r_{10} as a function of the lens characteristics will allow the value of E to be found for all image points (see next section).

4.10 Lens Model, Optical Path Length and Focal Surface Function

The purpose of this section is to show that $[r_{10}]$, the OPL from the aperture origin to a lens image point, can be found from a few simple characteristics of the lens.

With the assumptions of Section 4.8, some properties of a lens can be seen by consideration of optical path lengths alone. Referring to Figure 4.7, a lens has focal points F_1, F_2 and nodal points N_1, N_2 on the OCL. g_1 and g_2 are the focal surfaces. Σ_0 is a plane normal to the OCL and passing through F_1. The Optical Path Length of all rays in a collimated beam parallel to the OCL from the plane Σ_0 to their focal point at F_2 is the same and is equal to $[F_1N_1N_2F_2]$

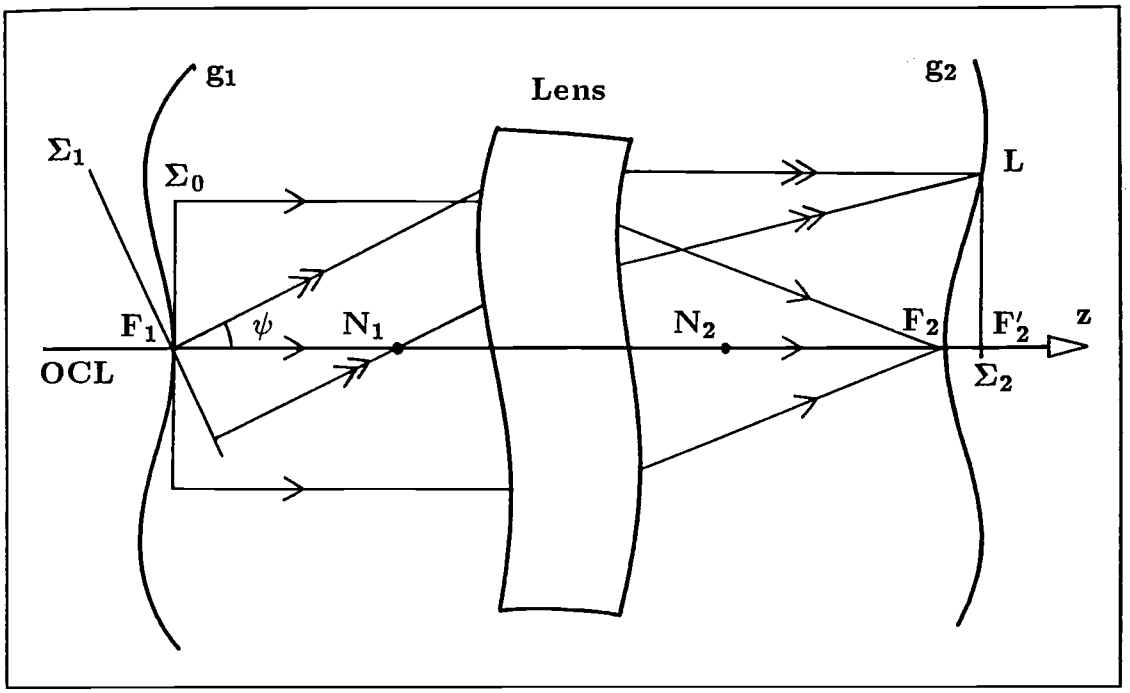


Figure 4.7: Optical Path Lengths between Focal Surfaces.

which in principle is easy to measure.

Rays in a collimated beam at ψ to the OCL from a plane Σ_1 through F_1 all have the same OPL to their image point at L, and this equals $[F_1L]$. A collimated beam from right to left, parallel to the OCL and starting from Σ_2 through L and F_2' is focused at F_1 and $[LF_1] = [F_2'F_1]$ or $[F_1L] = [F_1F_2']$.

- So in tilting the plane through F_1 from Σ_0 to Σ_1 the OPL from that plane to the image point increases from $[F_1F_2]$ to $[F_1F_2 + F_2F_2']$, i.e. *it increases simply by the z-coordinate of L with respect to* F_2.

When the point L on g_2 is at F_2, the distance along a ray through F_1 to F_2 is $[F_1F_2]$, see Figure 4.8. As L moves away from F_2 the distance from F_1 to L is changed by an amount F_2F_2', the z coordinate of L. If F_2F_2' at L decreases the length $[F_1L]$ then to maintain constant distance from L, F_2F_2' must be added

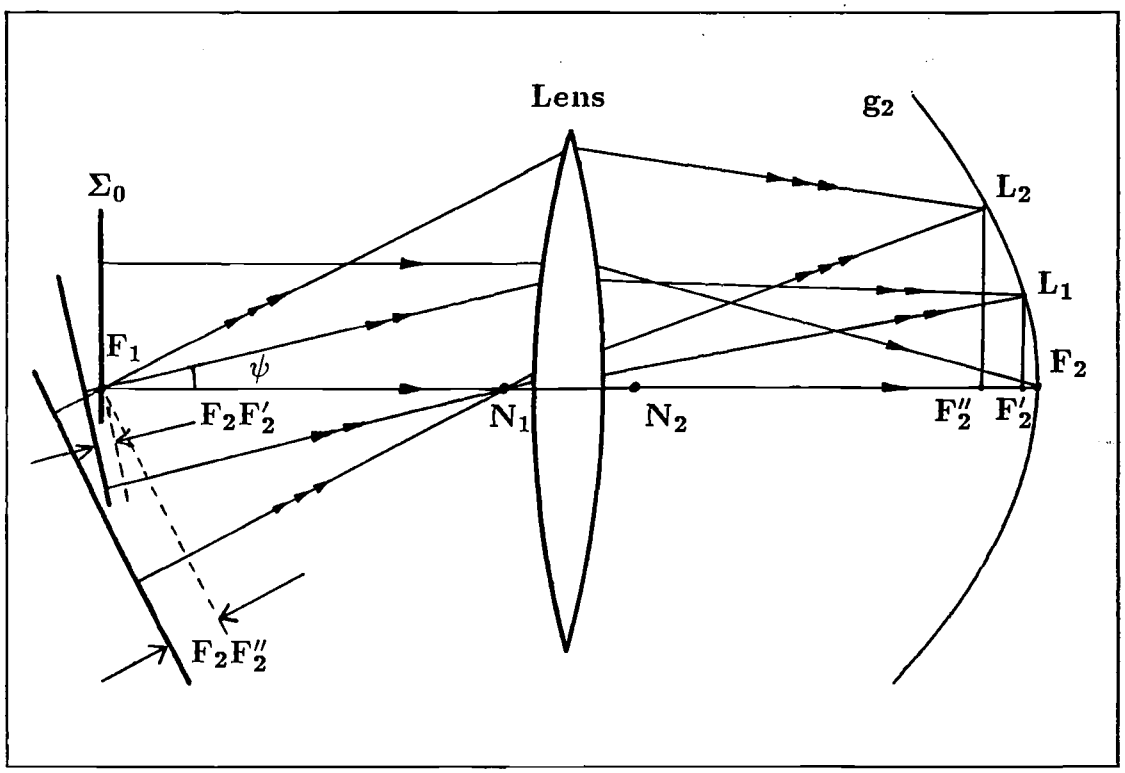


Figure 4.8: Planes of Constant Distance from Focal Surface Points.

to the ray, i.e. the wavefront from which there is constant distance to L on g_2 is perpendicular to a ray at ψ to the OCL through F_1 and F_2F_2'' away from it. Thus the positions of plane wavefronts of some constant distance to points P on the focal surface g_2 can be found for any function g_2, and so for any positive lens.

With the optical path length $[F_1F_2]$ along the optical centre line of a lens known, the distances travelled by diffracted ray beams are now also known and the introduction of a lens into the diffraction calculations makes these no more difficult (nor any easier) than in the earlier cases.

To obtain $[r_{10}]$ through a lens from the aperture origin to the image point is a matter of geometry as illustrated in Figure 4.9 in two dimensions. For an axial incident beam, and with $[F_1F_2] =$

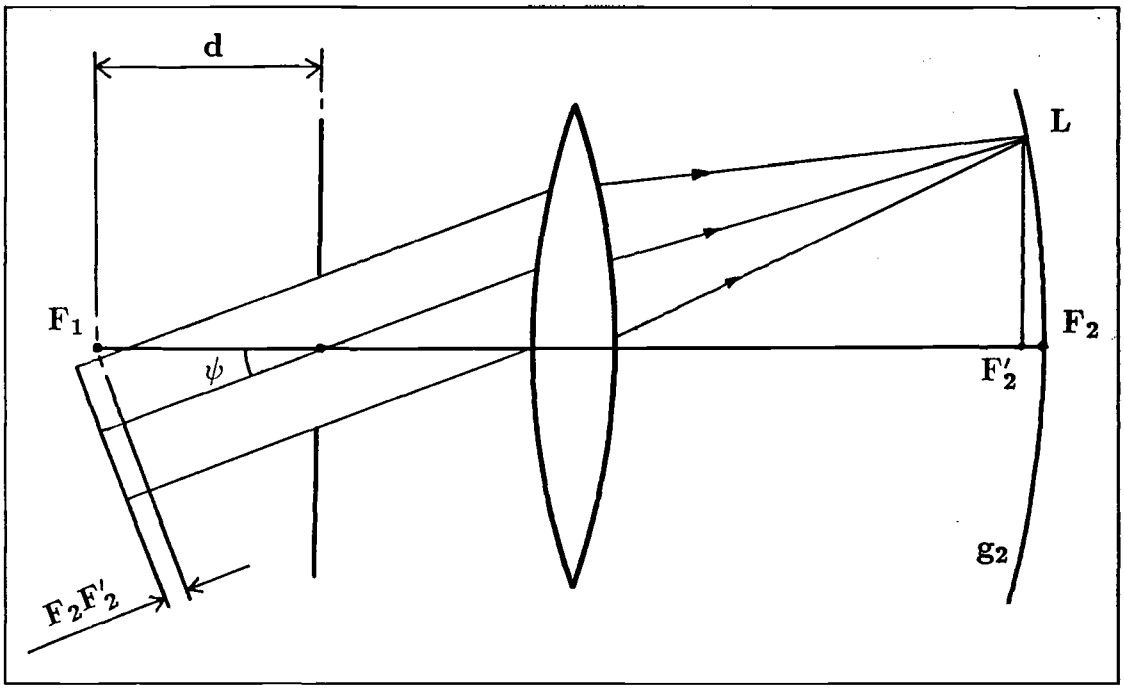


Figure 4.9: Distance from Aperture Origin to Image Point.

$$[r_{10}] = [r_c] - d \tag{4.53}$$

where $[r_c] = [F_1 F_2]$.

For a non-axial beam,

$$[r_{10}] = [r_c] - (d\cos\psi + F_2 F_2') \tag{4.54}$$

Thus in principle, for any focal surface g_2 and incidence angles known, the value of $[r_{10}]$ can be found and used in the diffraction integral equation. The size of $[r_{10}]$ affects both the amplitude and phase of the electric field but only the amplitude of the irradiance pattern.

With small apertures, $[r_{10}]$ can vary with ψ more than $[r_1]$ varies with (x,y) in the aperture for a particular ψ. So it is not unreasonable to use the calculated $[r_{10}]$ in establishing the strength of I at some point on the focal surface. This

is irrespective of the previous approximation that $[r_1]$ does not vary much from $[r_{10}]$ for some incident angle, and its consequent extraction from the integral denominator.

4.11 Conclusions

Without a lens, the irradiance pattern due to diffraction of plane waves by two slits contains a $\cos^2$ term, at least in the Fraunhofer approximation in which the observation screen is far from the aperture. This cosinusoidal pattern is positioned on the observation surface as a function of l, the direction cosine of incident rays with respect to the x axis, perpendicular to the long slit direction. Thus, in this approximation, information about the irradiance pattern position would yield the value of l for the incident plane radiation. Neglecting quadratic terms, incident spherical wavefronts can be treated as in the planar case and once again the direction cosine l is obtained. This latter would require a long distance to the point source as well as to the observation screen.

Introducing a lens behind the diffracting aperture allows the observation surface to be close to the lens, in fact the focal surface of the lens. The Fraunhofer approximation is no longer required behind the aperture. The introduction of this series of refracting elements is not an impediment to the evaluation of the diffraction integrals, even for a lens with a specified, non-planar focal surface.

Plane waves and spherical waves approximated by plane waves have been considered in this chapter. With a moving point source, relatively close to the aperture, supplying the incident monochromatic spherical wavefronts and free to move in three dimensions, once again a fringed irradiance pattern is produced. The measured movement of the fringe pattern gives one "coordinate" of the source position but this coordinate is not the x direction cosine as might have

been suggested by the approximate solution. For a point source at close range the wavefronts cannot be approximated by plane waves. This is dealt with (for a system with a lens) in the next chapter.

Chapter 5

Interferometer

5.1 Introduction

In this chapter some of the results of Chapter 4 are used to obtain expressions for electric field and irradiance at the focal surface of an interferometer. This instrument, which has been built (see Chapter 6), works by division of wavefront and uses a lens to superimpose two diffraction patterns. The relationship between irradiance pattern position and a point source position is determined. A means of detecting fringe pattern movement is presented.

5.2 Diffraction with 2 Slits and a Lens: Incident Plane Waves

When a lens is used the Rayleigh-Sommerfeld formula (Equation 4.44),

$$E = \frac{1}{i\lambda}\iint_{\Sigma} E(x,y)\frac{e^{ikr_1}}{r_1}\cos\psi\, dx\, dy \tag{5.1}$$

becomes (Equation 4.52),

$$E = \frac{\cos\psi e^{ikr_{10}}}{i\lambda r_{10}}\iint_{\Sigma} E(x,y)e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{5.2}$$

For incident plane waves with rays with direction cosines l, m, n, the electric field at the aperture is,

$$E(x,y) = E_0 e^{-ik(xl+ym)} \tag{5.3}$$

and so,

$$E = \frac{E_0 \cos\psi e^{ikr_{10}}}{i\lambda r_{10}} \iint_\Sigma e^{-ik(xl+ym)} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{5.4}$$

which evaluates in a similar way to the Fraunhofer approximation of Section 4.5 to give,

$$E = \frac{E_0 \cos\psi e^{ikr_{10}}}{i\lambda r_{10}} 8ab\, \mathrm{sinc}\,[ka(l+l_1)]\, \mathrm{sinc}\,[kb(m+m_1)] \cos[kA(l+l_1)] \tag{5.5}$$

and,

$$I = \frac{E_0^2 \cos^2\psi 64a^2b^2}{\lambda^2 r_{10}^2} \mathrm{sinc}^2[ka(l+l_1)]\, \mathrm{sinc}^2[kb(m+m_1)] \cos^2[kA(l+l_1)] \tag{5.6}$$

which differs from Equation 4.37 only by the $\cos^2\psi$ term.

5.3 Diffraction with 2 Slits and a Lens: Incident Spherical Waves

Starting from the Rayleigh-Sommerfeld formula, modified for a lens (Equation 4.52),

$$E = \frac{\cos\psi e^{ikr_{10}}}{i\lambda r_{10}} \iint_\Sigma E(x,y) e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{5.7}$$

with a point source of strength $\mathcal{E}$ at (x_0, y_0, z_0) and r the distance from the source to points in the aperture,

$$E(x,y) = \frac{\mathcal{E}e^{ikr}}{r} \tag{5.8}$$

giving,

$$E = \frac{\cos\psi e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} \frac{\mathcal{E}e^{ikr}}{r} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{5.9}$$

With the assumption that r does not vary significantly from r_0 over the aperture, the r in the denominator can be taken out of the integral, and,

$$E = \frac{\mathcal{E}\cos\psi e^{ikr_{10}}}{i\lambda r_0 r_{10}} \iint_{\Sigma} e^{ikr} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{5.10}$$

Now using the first two terms of the binomial expansion to give an approximation to r,

$$r \approx r_0 + \frac{(x^2+y^2)}{2r_0} - \frac{(xx_0+yy_0)}{r_0} \tag{5.11}$$

Substituting for r in the integrand of Equation 5.7,

$$E = \frac{\mathcal{E}\cos\psi e^{ik(r_0+r_{10})}}{i\lambda r_0 r_{10}} \iint_{\Sigma} e^{ik\left[\frac{(x^2+y^2)}{2r_0} - \frac{(xx_0+yy_0)}{r_0}\right]} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{5.12}$$

$$E = \frac{\mathcal{E}\cos\psi e^{ik(r_0+r_{10})}}{i\lambda r_0 r_{10}} \iint_{\Sigma} e^{ik\left[\frac{x^2}{2r_0} - (l+l_1)x\right]} e^{ik\left[\frac{y^2}{2r_0} - (m+m_1)y\right]}\, dx\, dy \tag{5.13}$$

The quadratic terms in x and y can cause problems for integration. With x or y =5mm and r_0 =1m, say, the contribution to phase from the quadratic term (with a wavelength of 632.8nm) is approximately 124 radians and there is no justification here for ignoring this term (if the slits are as long as 10mm). The value of x varies little over each slit by itself, however, in which case that

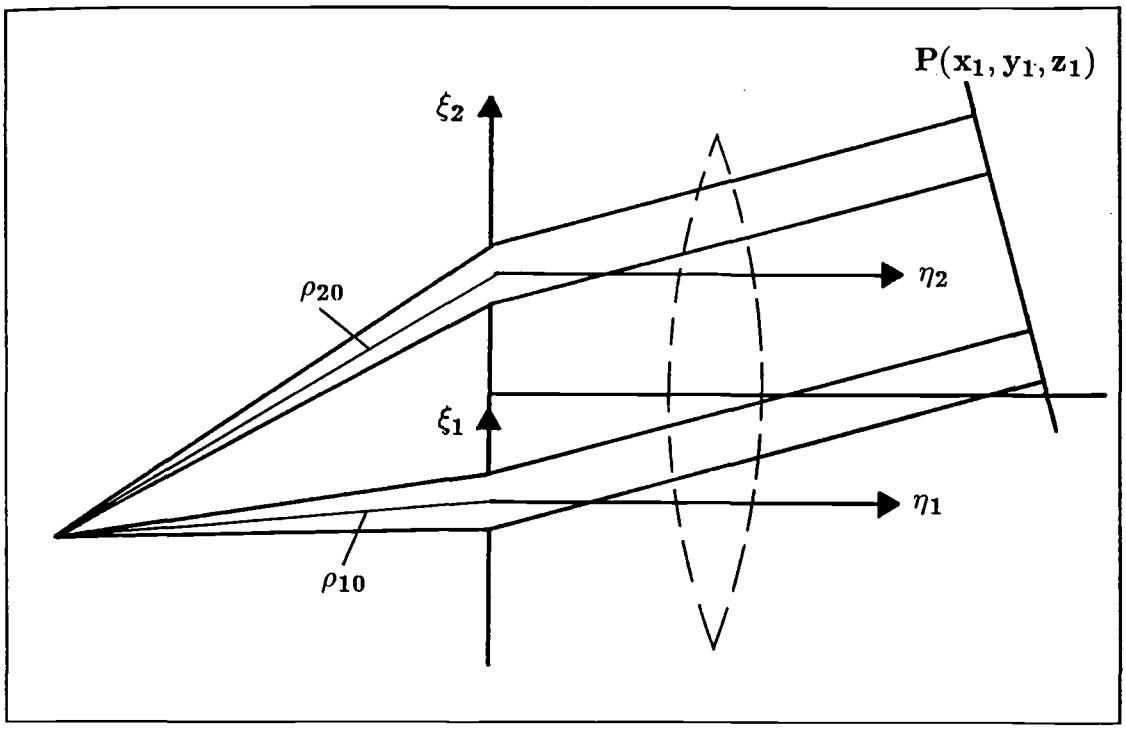


Figure 5.1: Summation of E-field Contributions from Two Slits.

quadratic term can be dropped. The value of E at P is now taken as the sum of E_1 due to one slit and E_2 due to the other, each slit having its own coordinate system (ξ_1, η_1, ζ_1) and (ξ_2, η_2, ζ_2) respectively. See Figure 5.1.

For the first slit the transformation to x, y, z coordinates is,

$$\begin{aligned} x &= \xi_1 + A \\ y &= \eta_1 \\ z &= \zeta_1 \end{aligned} \tag{5.14}$$

$$E_1 = \frac{\mathcal{E}\cos\psi}{i\lambda} \iint_{\Sigma_1} \frac{e^{ik\rho_1}}{\rho_1} \frac{e^{ikr_1}}{r_1} \, d\xi_1 \, d\eta_1 \tag{5.15}$$

where ρ_1 is the distance from the point source to points in the aperture Σ_1.

With ρ_{10} the distance from the point source to the origin of the coordinate system in that aperture,

$$\rho_1 = \left[(\xi_1 - \xi_{10})^2 + (\eta_1 - \eta_{10})^2 + \zeta_{10}^2\right]^{\frac{1}{2}} \tag{5.16}$$

With only the first two terms of the binomial expansion being used (which can be justified for small ξ and η i.e. relatively narrow, short slits),

$$\rho_1 \approx \rho_{10} + \frac{(\xi_1^2 + \eta_1^2)}{2\rho_{10}} - \frac{(\xi_1\xi_{10} + \eta_1\eta_{10})}{\rho_{10}} \tag{5.17}$$

Substituting this in the expression for E_1,

$$E_1 = \frac{\mathcal{E}\cos\psi}{i\lambda}\iint_{\Sigma_1} \frac{e^{ik\left[\rho_{10} + \frac{(\xi_1^2+\eta_1^2)}{2\rho_{10}} - \frac{(\xi_1\xi_{10}+\eta_1\eta_{10})}{\rho_{10}}\right]}}{\rho_1} \frac{e^{ikr_1}}{r_1}\, d\xi_1\, d\eta_1 \tag{5.18}$$

In x, y, z coordinates, $r_1 = r_{10} - (xl_1 + ym_1)$. So in ξ_1, η_1, ζ_1 coordinates,

$$r_1 = r_{10} - [(\xi_1 + A)l_1 + \eta_1 m_1] \tag{5.19}$$

where l_1, m_1 are unchanged by a system translation and r_{10} is a constant in both coordinate systems. Assuming $\rho_1 r_1$ does not change much from $\rho_{10} r_{10}$, the denominator can come outside giving,

$$E_1 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{10}+r_{10})}}{i\lambda\rho_{10}r_{10}}\iint_{\Sigma_1} e^{ik\left[\frac{(\xi_1^2+\eta_1^2)}{2\rho_{10}} - \frac{(\xi_1\xi_{10}+\eta_1\eta_{10})}{\rho_{10}}\right]} e^{-ik(\xi_1 l_1 + Al_1 + \eta_1 m_1)}\, d\xi_1\, d\eta_1 \tag{5.20}$$

The direction cosines of the source point in ξ_1, η_1, ζ_1 coordinates are,

$$\begin{aligned} l_{10} &= \frac{\xi_{10}}{\rho_{10}} \\ m_{10} &= \frac{\eta_{10}}{\rho_{10}} \\ n_{10} &= \frac{\zeta_{10}}{\rho_{10}} \end{aligned} \tag{5.21}$$

$$E_1 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{10}+r_{10})}e^{-ikAl_1}}{i\lambda\rho_{10}r_{10}} \int_{-a}^{a} e^{ik\left[\frac{\xi_1^2}{2\rho_{10}}-(\xi_1 l_{10}+\xi_1 l_1)\right]}\, d\xi_1 \int_{-b}^{b} e^{ik\left[\frac{\eta_1^2}{2\rho_{10}}-(\eta_1 m_{10}+\eta_1 m_1)\right]}\, d\eta_1 \tag{5.22}$$

Considering the contribution of the quadratic terms, with a slit 100μm wide and 10mm long and with a 632.8nm source at 1m range,

$$\frac{k\xi_1^2}{2\rho_{10}} = 1.24 \times 10^{-2} \text{ radians}$$

which can be ignored, but,

$$\frac{k\eta_1^2}{2\rho_{10}} = 124 \text{ radians}$$

which cannot.

The evaluation of the integral in η_1 with a term in $e^{ik\eta_1^2}$ requires graphical or numerical techniques. For present purposes, its solution is not of great interest and can be accepted as being a modification of the diffraction envelope which would be obtained with linear powers of e in the integrand. Of more interest is the summation of terms from the two slit diffraction integrals which are functions of ξ_1, ξ_2.

With the ξ_1 quadratic term,

$$\frac{k\xi_1^2}{2\rho_{10}} \approx 0 \tag{5.23}$$

$$E_1 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{10}+r_{10})}e^{-ikAl_1}}{i\lambda\rho_{10}r_{10}} \int_{-a}^{a} e^{-ik(l_{10}+l_1)\xi_1}\, d\xi_1 \int_{-b}^{b} e^{ik\left[\frac{\eta_1^2}{2\rho_{10}}-(\eta_1 m_{10}+\eta_1 m_1)\right]}\, d\eta_1 \tag{5.24}$$

$$E_1 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{10}+r_{10})}e^{-ikAl_1}}{i\lambda\rho_{10}r_{10}} \left[\frac{e^{-ik(l_{10}+l_1)\xi_1}}{-ik(l_{10}+l_1)}\right]_{-a}^{a} I_{\eta_1} \tag{5.25}$$

The transformation to the (ξ_2, η_2, ζ_2) coordinate system in the second slit is,

$$\begin{aligned} x &= \xi_2 - A \\ y &= \eta_2 \\ z &= \zeta_2 \end{aligned} \tag{5.26}$$

$$E_2 = \frac{\mathcal{E}\cos\psi}{i\lambda}\iint_{\Sigma_2}\frac{e^{ik\rho_2}}{\rho_2}\frac{e^{ikr_1}}{r_1}\,d\xi_2\,d\eta_2 \tag{5.27}$$

where ρ_2 is the distance from the point source to points in the aperture Σ_2.

With ρ_{20} the distance from the point source to the origin of the coordinate system in the aperture,

$$\rho_2 \approx \rho_{20} + \frac{(\xi_2^2+\eta_2^2)}{2\rho_{20}} - \frac{(\xi_2\xi_{20}+\eta_2\eta_{20})}{\rho_{20}} \tag{5.28}$$

Substituting this in the expression for E_2,

$$E_2 = \frac{\mathcal{E}\cos\psi}{i\lambda}\iint_{\Sigma_2}\frac{e^{ik\left[\rho_{20}+\frac{(\xi_2^2+\eta_2^2)}{2\rho_{20}}-\frac{(\xi_2\xi_{20}+\eta_2\eta_{20})}{\rho_{20}}\right]}}{\rho_2}\frac{e^{ikr_1}}{r_1}\,d\xi_2\,d\eta_2 \tag{5.29}$$

In x, y, z coordinates, $r_1 = r_{10} - (xl_1 + ym_1)$. So in ξ_2, η_2, ζ_2 coordinates,

$$r_1 = r_{10} - [(\xi_2 - A)l_1 + \eta_2 m_1] \tag{5.30}$$

Assuming $\rho_2 r_1$ does not change much from $\rho_{20} r_{10}$, the denominator can come outside giving,

$$E_2 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{20}+r_{10})}}{i\lambda\rho_{20}r_{10}}\iint_{\Sigma_2} e^{ik\left[\frac{(\xi_2^2+\eta_2^2)}{2\rho_{20}}-\frac{(\xi_2\xi_{20}+\eta_2\eta_{20})}{\rho_{20}}\right]}e^{-ik(\xi_2 l_1 - Al_1 + \eta_2 m_1)}\,d\xi_2\,d\eta_2 \tag{5.31}$$

The direction cosines of the source point in ξ_2, η_2, ζ_2 coordinates are,

$$\begin{aligned} l_{20} &= \frac{\xi_{20}}{\rho_{20}} \\ m_{20} &= \frac{\eta_{20}}{\rho_{20}} \\ n_{20} &= \frac{\zeta_{20}}{\rho_{20}} \end{aligned} \tag{5.32}$$

giving,

$$E_2 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{20}+r_{10})}e^{ikAl_1}}{i\lambda\rho_{20}r_{10}} \int_{-a}^{a} e^{ik\left[\frac{\xi_2^2}{2\rho_{20}}-(\xi_2 l_{20}+\xi_2 l_1)\right]}\, d\xi_2 \int_{-b}^{b} e^{ik\left[\frac{\eta_2^2}{2\rho_{20}}-(\eta_2 m_{20}+\eta_2 m_1)\right]}\, d\eta_2 \tag{5.33}$$

As with the first aperture, the quadratic term in the first integral is taken as,

$$\frac{k\xi_2^2}{2\rho_{20}} \approx 0 \tag{5.34}$$

$$E_2 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{20}+r_{10})}e^{ikAl_1}}{i\lambda\rho_{20}r_{10}} \int_{-a}^{a} e^{-ik(l_{20}+l_1)\xi_2}\, d\xi_2 \int_{-b}^{b} e^{ik\left[\frac{\eta_2^2}{2\rho_{20}}-(\eta_2 m_{20}+\eta_2 m_1)\right]}\, d\eta_2 \tag{5.35}$$

$$E_2 = \frac{\mathcal{E}\cos\psi e^{ik(\rho_{20}+r_{10})}e^{-ikAl_1}}{i\lambda\rho_{20}r_{10}} \left[\frac{e^{-ik(l_{20}+l_1)\xi_2}}{-ik(l_{20}+l_1)}\right]_{-a}^{a} I_{\eta_2} \tag{5.36}$$

Adding Equations 5.25 and 5.36 gives the sum of the contributions to the electric field at $P(x_1, y_1, z_1)$, $E = E_1 + E_2$,

$$\begin{aligned} E = &\frac{\mathcal{E}\cos\psi e^{ik(\rho_{10}+r_{10})}e^{-ikAl_1}}{i\lambda\rho_{10}r_{10}} 2a \left[\frac{e^{ik(l_{10}+l_1)a} - e^{-ik(l_{10}+l_1)a}}{2ik(l_{10}+l_1)a}\right] I_{\eta_1} \\ &+ \frac{\mathcal{E}\cos\psi e^{ik(\rho_{20}+r_{10})}e^{ikAl_1}}{i\lambda\rho_{20}r_{10}} 2a \left[\frac{e^{ik(l_{20}+l_1)a} - e^{-ik(l_{20}+l_1)a}}{2ik(l_{20}+l_1)a}\right] I_{\eta_2} \end{aligned} \tag{5.37}$$

The irradiance I at $P(x_1, y_1, z_1)$ is,

$$I = E_1 E_1^* + E_2 E_2^* + E_2 E_1^* + E_1 E_2^* \tag{5.38}$$

The first two terms on the right hand side of Equation 5.38 are,

$$E_1 E_1^* = \frac{\mathcal{E}^2 \cos^2\psi 4a^2}{\lambda^2 \rho_{10}^2 r_{10}^2} I_{\eta_1}^2 \operatorname{sinc}^2[ka(l_{10}+l_1)] \tag{5.39}$$

and,

$$E_2 E_2^* = \frac{\mathcal{E}^2 \cos^2\psi 4a^2}{\lambda^2 \rho_{20}^2 r_{10}^2} I_{\eta_2}^2 \operatorname{sinc}^2[ka(l_{20}+l_1)] \tag{5.40}$$

So,

$$E_1 E_1^* + E_2 E_2^* = \frac{\mathcal{E}^2 \cos^2\psi 4a^2}{\lambda^2 r_{10}^2} \left[\frac{I_{\eta_1}^2 \operatorname{sinc}^2[ka(l_{10}+l_1)]}{\rho_{10}^2} + \frac{I_{\eta_2}^2 \operatorname{sinc}^2[ka(l_{20}+l_1)]}{\rho_{20}^2} \right] \tag{5.41}$$

The second two terms on the right of Equation 5.38 are,

$$E_2 E_1^* = \frac{\mathcal{E}^2 \cos^2\psi 4a^2}{\lambda^2 \rho_{10}\rho_{20} r_{10}^2} e^{ik(\rho_{20}-\rho_{10})} e^{2ikAl_1} I_{\eta_1} I_{\eta_2} \operatorname{sinc}[ka(l_{10}+l_1)] \operatorname{sinc}[ka(l_{20}+l_1)] \tag{5.42}$$

and,

$$E_1 E_2^* = \frac{\mathcal{E}^2 \cos^2\psi 4a^2}{\lambda^2 \rho_{10}\rho_{20} r_{10}^2} e^{ik(\rho_{10}-\rho_{20})} e^{-2ikAl_1} I_{\eta_1} I_{\eta_2} \operatorname{sinc}[ka(l_{10}+l_1)] \operatorname{sinc}[ka(l_{20}+l_1)] \tag{5.43}$$

giving,

$$E_1 E_2^* + E_2 E_1^* = \frac{\mathcal{E}^2 \cos^2\psi 8a^2}{\lambda^2 \rho_{10}\rho_{20} r_{10}^2} I_{\eta_1} I_{\eta_2} \operatorname{sinc}[ka(l_{10}+l_1)] \operatorname{sinc}[ka(l_{20}+l_1)]$$

$$\times \left[\frac{e^{ik(\rho_{20}-\rho_{10}+2Al_1)} + e^{-ik(\rho_{20}-\rho_{10}+2Al_1)}}{2} \right] \tag{5.44}$$

$$E_1E_2^* + E_2E_1^* = \frac{\mathcal{E}^2\cos^2\psi 8a^2}{\lambda^2\rho_{10}\rho_{20}r_{10}^2} I_{\eta_1}I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]\,\mathrm{sinc}\,[ka(l_{20}+l_1)]$$

$$\times\cos\,[k(\rho_{20}-\rho_{10})+2kAl_1] \tag{5.45}$$

Adding Equations 5.41 and 5.45 the irradiance I at P is,

$$I = \frac{\mathcal{E}^2\cos^2\psi 4a^2}{\lambda^2 r_{10}^2}\left[\frac{I_{\eta_1}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]}{\rho_{10}} - \frac{I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{20}+l_1)]}{\rho_{20}}\right]^2$$

$$+\frac{\mathcal{E}^2\cos^2\psi 4a^2}{\lambda^2 r_{10}^2}\frac{2I_{\eta_1}I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]\,\mathrm{sinc}\,[ka(l_{20}+l_1)]}{\rho_{10}\rho_{20}}$$

$$+\frac{\mathcal{E}^2\cos^2\psi 4a^2}{\lambda^2 r_{10}^2}\frac{2I_{\eta_1}I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]\,\mathrm{sinc}\,[ka(l_{20}+l_1)]}{\rho_{10}\rho_{20}}$$

$$\times\cos\,[k(\rho_{20}-\rho_{10})+2kAl_1] \tag{5.46}$$

$$I = \frac{\mathcal{E}^2\cos^2\psi 4a^2}{\lambda^2 r_{10}^2}\left[\frac{I_{\eta_1}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]}{\rho_{10}} - \frac{I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{20}+l_1)]}{\rho_{20}}\right]^2$$

$$+\frac{\mathcal{E}^2\cos^2\psi 4a^2}{\lambda^2 r_{10}^2}\frac{2I_{\eta_1}I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]\,\mathrm{sinc}\,[ka(l_{20}+l_1)]}{\rho_{10}\rho_{20}}$$

$$\times\,[1+\cos\,[k(\rho_{20}-\rho_{10})+2kAl_1]] \tag{5.47}$$

$$I = \frac{\mathcal{E}^2\cos^2\psi 4a^2}{\lambda^2 r_{10}^2}\left[\frac{I_{\eta_1}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]}{\rho_{10}} - \frac{I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{20}+l_1)]}{\rho_{20}}\right]^2$$

$$+\frac{\mathcal{E}^2\cos^2\psi 4a^2}{\lambda^2 r_{10}^2}\frac{4I_{\eta_1}I_{\eta_2}\,\mathrm{sinc}\,[ka(l_{10}+l_1)]\,\mathrm{sinc}\,[ka(l_{20}+l_1)]}{\rho_{10}\rho_{20}}$$

$$\times\cos^2\left[\frac{k}{2}(\rho_{20}-\rho_{10})+kAl_1\right] \tag{5.48}$$

5.4 Irradiance Function on a Focal Surface

In Sections 5.2 and 5.3 the irradiance functions due to diffraction of incident plane and spherical waves by two slits contained a cosine (or cosine^2) interference term.

In the case of plane waves, the phase of the cosine^2 term depends on both the incident direction cosine, l, and the diffracted ray direction cosine, l_1. The amplitude is determined by the more slowly varying sinc^2 functions. Equation 5.6 can be written,

$$I = I' \cos^2[kA(l + l_1)] \tag{5.49}$$

or,

$$I = \frac{I'}{2} + \frac{I'}{2} \cos[2kA(l + l_1)] \tag{5.50}$$

which has a maximum $= I'$ when $\cos[2kA(l + l_1)] = 1$. The value of l determines the position of the fringe pattern, and l_1 selects a particular part of that pattern. When l is constant, as l_1 is varied the irradiance is modulated cosinusoidally with l_1. Conversely if only one, fixed l_1 is being observed, then as l is varied, the irradiance in that one, l_1 direction is modulated cosinusoidally as the whole pattern moves.

In the case of incident spherical wavefronts from a point source at distances ρ_{10}, ρ_{20} from the two slit origins, the irradiance function (Equation 5.48) has the form,

$$I = I'' + I' \cos^2\left[\frac{k}{2}(\rho_{20} - \rho_{10}) + kAl_1\right] \tag{5.51}$$

or as in Equation 5.47,

$$I = I'' + \frac{I'}{2} + \frac{I'}{2}\cos\left[k(\rho_{20} - \rho_{10}) + 2kAl_1\right] \tag{5.52}$$

Here, $(\rho_{20} - \rho_{10})$ may be seen as determining the position of the fringe pattern. When l_1 is fixed, the irradiance at that constant value of direction cosine is modulated cosinusoidally with $(\rho_{20} - \rho_{10})$ as the whole pattern moves. If the point source remains on the same hyperboloid with foci at the centres of the two apertures, then $(\rho_{20} - \rho_{10})$ remains constant.

In both cases above, the interference fringes are given as functions of the x direction cosine, l_1, of the diffracted rays in lens object space. With specified lens characteristics the pattern on the lens focal surface may be obtained.

With a planar focal surface normal to the OCL, curves of constant I on the focal surface are hyperbolae, i.e. planar conic sections of the conical surfaces swept out by constant values of l_1. Moving in an increasing x direction from the OCL on the $y = 0$ plane, the irradiance varies periodically and with slowly increasing period.

5.5 Detection of Interference Pattern Movement

5.5.1 Sensing Techniques

Observation of the interference pattern movement on the focal surface can give information on a change in l_1 and for incident plane radiation the new value of l can be found. Similarly for a point source, the new value of $(\rho_{20} - \rho_{10})$ can be found.

The most appropriate means of monitoring interference pattern position in an instrument must depend on the size of object field and the resolution required.

For small object field angles the lens focal length can be long, the slit separation small and the scale of the fringes correspondingly large. In these circumstances the fringes may be detected using the image plane of a video camera, with subsequent analysis of the digitised image.

For larger object field angles and/or higher resolution, a shorter focal length and more widely spaced slits are necessary. In this case the period of the interference fringes can be beyond the resolution of current video sensors. Interference patterns with hundreds of cycles per millimeter can easily be produced. Since only changes of fringe position on the focal surface in the direction of maximum gradient of l_1 are of interest, the movement of high spatial frequency patterns can be detected using a diffraction grating. The grating has the same or almost the same line spacing as the interference pattern produced by the slits and is oriented with its lines essentially parallel to the interference pattern lines. Using an amplitude transmission grating, for instance, the light transmitted through the grating is modulated as a function of the position of the fringes on the grating.

This approach has some similarity to the generation of moire fringes by the relative motion of two diffraction gratings in close proximity, in which one grating produces a large number of closely spaced diffraction patterns and the second grating modulates this output. The differences here are that an interference fringe pattern is generated "at a distance" by the slit screen and lens combination and that pattern movement is caused not by physical movement of a grating but by change in the relative phases at the slits. The similarity is that the grating in the interferometer, like the second grating of a moire pair, has the function of modulating a moving pattern.

5.5.2 Amplitude Diffraction Grating

The principle on which this sensor is based is that a periodic irradiance pattern incident on an amplitude diffraction grating with matching period will transmit (or reflect as the case may be) the convolution of the irradiance function with the grating transmittance (or reflectance) function. Monitoring the grating output power is in effect measuring the convolution variable, in this case the phase shift of the irradiance pattern with respect to the grating pattern.

To illustrate this, examples in which the spatial period of both the irradiance pattern and the grating lines are constant and equal are given. One example is with a square grating comprising alternating, evenly spaced transparent and opaque lines. The other example is with a grating whose transmittance has a sinusoidal variation. Because the period of the fringes and the grating is the same, only one cycle need be considered to obtain a comparison. The power of particular fringes within the sinc^2 envelope is unimportant, only coming into consideration if it is desired to integrate across all fringes to calculate the transmitted power for some particular position.

Referring to Figure 5.2, the irradiance function is,

$$I(\theta) = \frac{1}{2}(\cos\theta + 1) \tag{5.53}$$

$T(\theta)$ is the transmittance of the grating, the ratio of transmitted flux to incident flux and $F(\varepsilon)$ is the power transmitted through the grating, ε being the position of the irradiance pattern on the grating.

$$F = I \star T \tag{5.54}$$

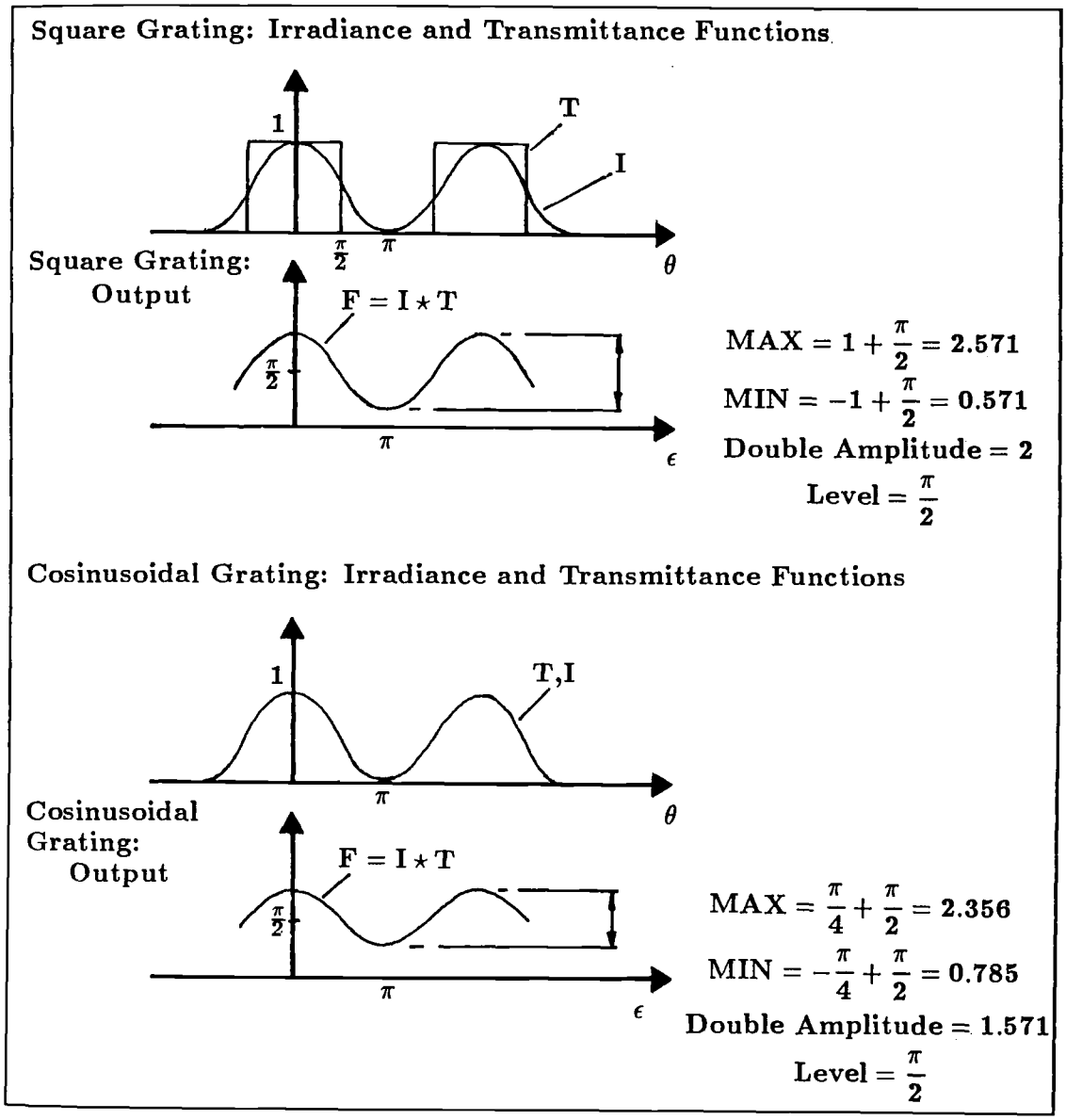


Figure 5.2: Convolution of Irradiance and Transmittance Functions.

$$F(\varepsilon) = \int_{-\infty}^{\infty} I(\varepsilon - \theta)T(\theta)\,d\theta \tag{5.55}$$

Only the interval $-\pi$ to π need be considered.

$$F(\varepsilon) = \int_{-\pi}^{\pi} I(\varepsilon - \theta)T(\theta)\,d\theta \tag{5.56}$$

For the square grating,

$$\begin{aligned} -\pi < \theta < -\frac{\pi}{2},\ \frac{\pi}{2} < \theta < \pi \quad & T = 0 \\ -\frac{\pi}{2} < \theta < \frac{\pi}{2} \quad & T = 1 \end{aligned} \tag{5.57}$$

$$F(\varepsilon) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}[\cos(\varepsilon - \theta) + 1]\,1\,d\theta \tag{5.58}$$

$$F(\varepsilon) = \cos\varepsilon + \frac{\pi}{2} \tag{5.59}$$

For the sinusoidal grating,

$$T = \frac{1}{2}[\cos\theta + 1] \tag{5.60}$$

$$F(\varepsilon) = \int_{-\pi}^{\pi} \frac{1}{2}[\cos(\varepsilon - \theta) + 1]\frac{1}{2}[\cos\theta + 1]\,d\theta \tag{5.61}$$

$$F(\varepsilon) = \frac{\pi}{4}\cos\varepsilon + \frac{\pi}{2} \tag{5.62}$$

The output in both cases is a cosine function plus a constant, the double amplitude being greater with the square grating. Gratings of both types were used experimentally for interference pattern movement detection. The results are reported in the next chapter.

Chapter 6

Experimental Results

6.1 Summary

A series of experiments were carried out to verify the theory given in Chapters 4 and 5 and establish the most important parameters for the design of a developed instrument based on these principles. After the necessary calibration of equipment the stages of the experimental work were as follows. Firstly it was verified that good quality interference patterns could be produced with the equipment and that these patterns moved in the predicted directions with source movement. Next, the amount of fringe movement in response to precise adjustments of the source position or orientation was measured and compared with the theoretical value. Then some of the limits of operation of the components used were explored to determine how the optical design could be improved. The effects of air currents and temperature gradients in the optical paths were investigated. A method of sensing fringe movements using a diffraction grating was tested and the output signal evaluated. Then with the output field split in amplitude, two signals were monitored in quadrature. Finally some observations are made on optical noise and methods of removing it.

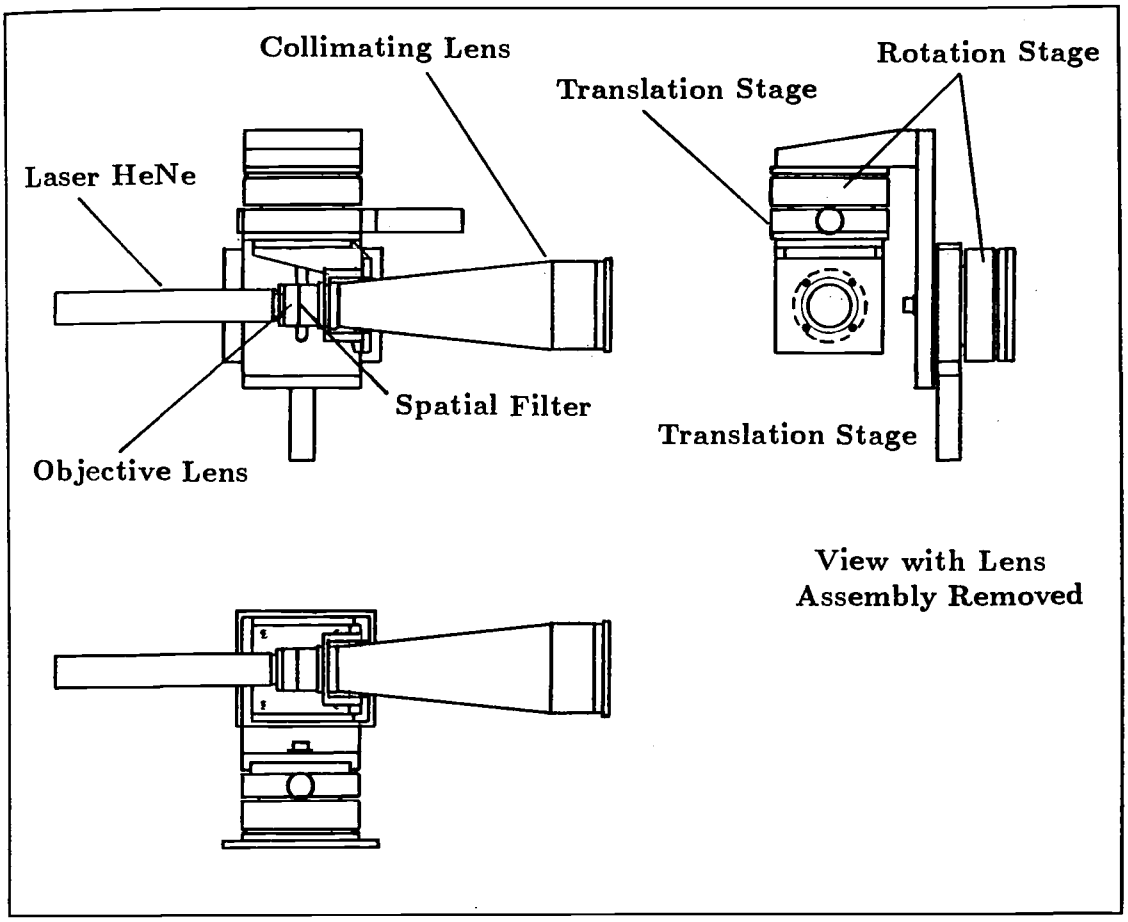


Figure 6.1: Transmitter Assembly.

6.2 Equipment

6.2.1 Transmitter

To produce the plane and spherical wavefronts necessary to test the interferometer a HeNe laser, objective lens, pinhole and (removable) collimating lens were assembled on a four degree of freedom mount. This was built as per the design in Figure 6.1. The assembly was calibrated by adjustment of the positions of the lens and pinhole components of the spatial filter. The correct focus of the collimating lens was obtained by adjusting until the minimum number of fringes were produced on a screen by an etalon positioned in the output beam.

Without the collimating lens in place, spherical wavefronts are produced within a cone. With the output lens in place, a 50mm diameter collimated beam is produced. This beam produced less than half a fringe line pair over the width of the etalon indicating a wavefront flatness of less than 1/4 λ. Only the central 50mm of the 78mm diameter output lens was used. The optics of this arrangement is conventional.

6.2.2 Receiver

The basic arrangement described in Chapter 5 was used and incorporated a single lens reflex camera lens. The receiver components and their development are discussed below.

6.2.3 Slits

Two choices are available for the type of slit screen. The first is to use air slits, made by photo-etching right through very thin sheet copper. The disadvantages of air slits are that they are easily damaged and sensitive to temperature changes. Contamination of the slits with dust particles would adversely affect the interference pattern generated and cleaning is more difficult than with the second type.

The second type consists of a very thin layer of metal (silver or chromium) evaporated onto a glass plate. The metal surface is then photo-etched as before. With this type, much finer slits can be made if required, and with the plane glass surface to the outside, the metal film is protected.

Especially fine slits were not required for these experiments, and so the more readily available air slits were used. See Figure 6.2. The slits produced were 0.100mm wide with 2.5mm separation between the slits. Examination of the

slits with a microscope cathetometer showed the edges to be ragged but the average dimensions were close to these nominal values. The slit screen was held in a standard 52mm camera lens filter holder.

6.2.4 Lens

The first aspect to consider in the choice of lens is the focal length. The size of the interference pattern on the focal plane is a function of focal length and this pattern must match the period of a diffraction grating positioned in this plane. See Section 6.7. The tolerance on the nominal focal length of high quality camera lenses is no better than $\pm 1\%$ however and can be worse. So if the grating line spacing is fixed by the nominal lens focal length, a mismatch can occur.

Since it is quite undesirable to make each grating (in a developed system) especially for its lens, then, should the system prove sensitive to focal length, the most attractive course is to use a lens with adjustable focal length, a zoom lens. Only a small zoom range is required in this application to accommodate tolerances in the slit screen and grating dimensions, but camera zoom lenses are being produced with ever greater ranges of focal length. The lens used initially in these experiments is a Fujica 54 to 270mm, f4.5.

6.2.5 Cameras

Although not part of the instrument, both video and 35mm still cameras have been used to view the interference patterns. The pattern produced by two slits can of course be viewed on a screen some distance away without using a lens. With a lens in place, the pattern on a screen in the lens focal plane is so fine that resolution cannot be achieved by eye.

Two ways round this observation problem have been found. To observe any

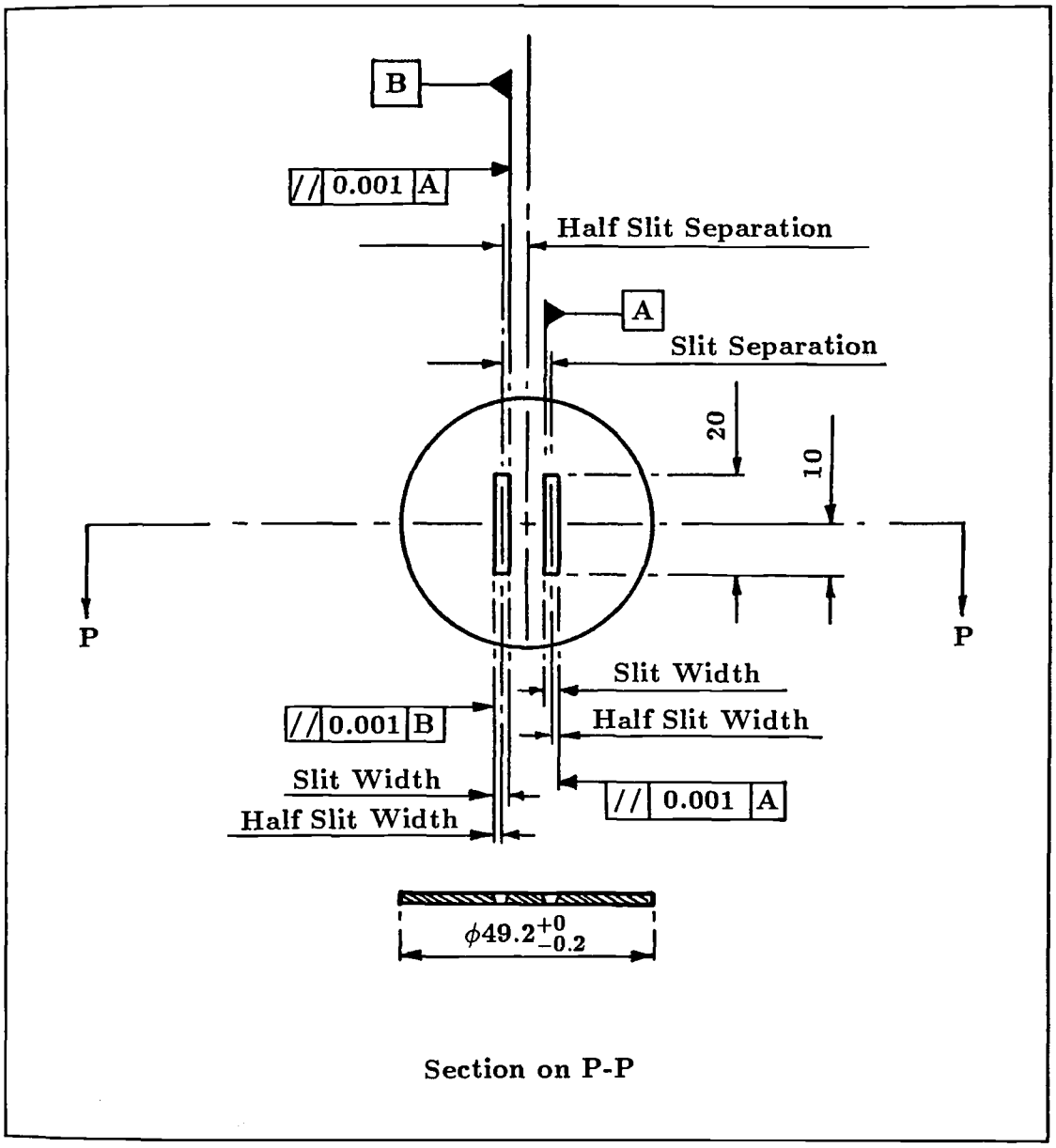


Figure 6.2: Air Slits.

movement of fringes and to assess their visibility, in the first instance a video camera (with its own lens removed) was used, with its image plane in the focal plane of the lens. To obtain permanent records of the patterns a Pentax 35mm SLR camera body was fitted to the lens and a number of still photographs taken on high resolution film (Kodak Technical Pan). Later a Nikon F2 was used, with 55mm Micro-Nikkor and 35-70mm Nikkor zoom lenses. The results are described in Section 6.5.3.

6.2.6 Diffraction Gratings

In general diffraction gratings may be classified as reflection or transmission gratings and within these categories as either phase or amplitude gratings. Amplitude transmission gratings are used in these experiments, since it is required to measure changes in the intensity of light modulated as it passes through the grating. In principal an analogous arrangement may be conceived with reflecting components .

Traditionally gratings are made by cutting a series of parallel grooves in a glass surface with the diamond tipped tool of a ruling engine. Plastic moulded copies are then made using this master grating. The manufacture of the master with the requisite line spacing is however an expensive process.

Another method is to photoetch the grating on an evaporated metal film on which the photoresist has been exposed by a step and repeat process with a single slit. This is a similar process to that described for manufacture of the slit screen and it produces a grating whose transmittance varies as a square wave, that is, alternate, even, translucent and opaque lines.

Two cheaper methods are available. Gratings may be produced photographically using large scale artwork and a series of reductions to give a grating

in photographic negative. This may be on film or glass plate. Again this is a square (or rectangular) wave grating. The second method is to make the grating holographically by interfering two large diameter collimated beams of monochromatic light on a photographic plate. The resulting fringes modulate the transmittance sinusoidally.

Gratings made by the last three techniques were used in the course of the experiments, one of these, a "sinusoidal" grating with 150 cycles per mm, having been made holographically.

6.2.7 Radiometers

The power of light transmitted through the gratings was sensed using first a photodiode radiometer and later an S20 type photomultiplier tube and narrow band interference filter.

In the final stages of the experiments a cube beam splitter was used at the lens output to duplicate the interference pattern and allow two signals in quadrature to be monitored by photomultipliers.

6.3 Interference Patterns

6.3.1 To Verify Geometry of Fringes in Far Field

With the laser and beam expander set up and correctly focused the interference pattern due to two air slits was imposed on a screen one meter from the slits. See Figure 6.3. Using the image plane of a video camera in place of the screen, the interference pattern was displayed on a monitor. See Plate 1. The number of fringes in the central band were calculated to be 50. 45 could be counted on the screen.

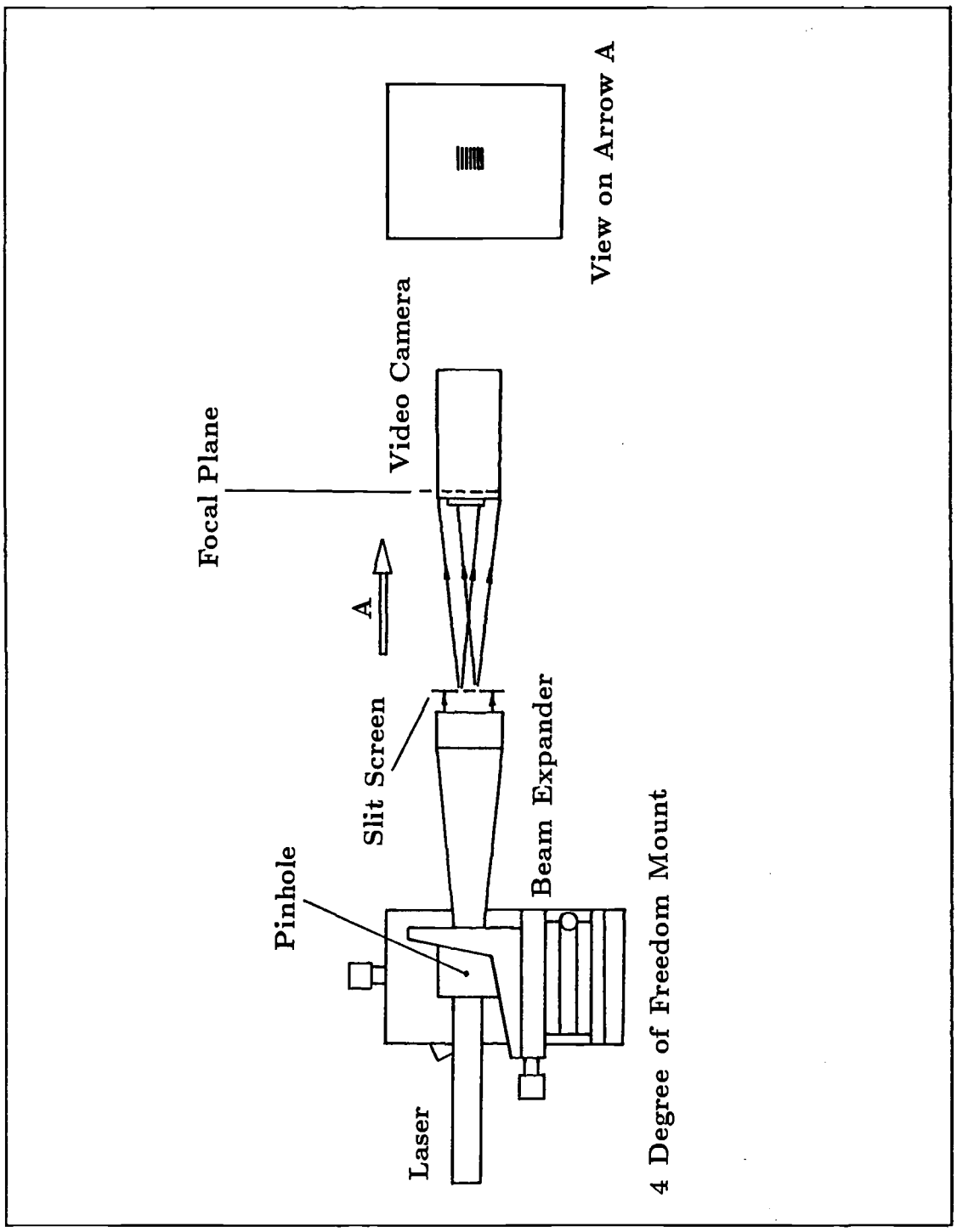


Figure 6.3: Interference with Collimated Beam.

6.3.2 Movement of Fringes with Incident Angle of Plane Waves

Using the collimated beam as before, the angle of incidence to the slit screen was varied both in azimuth and elevation. See Figure 6.4. As expected changes in azimuth caused fringe movement perpendicular to the vertical fringe lines and changes in elevation caused movement parallel to the lines.

The corresponding situation for cylindrical or spherical waves could not be demonstrated in this lensless arrangement as the two central diffraction bands might overlap only at a large distance beyond the slits. To make these bands overlap closer to the slits a lens must be used.

6.3.3 Split Wave Interference using a Lens and Incident Plane Waves

The use of a lens in the system was tried first with the collimated, expanded beam. The reason for this was simply to separate any effects which may have been due to the lens from effects due to spherical wavefronts.

Equipment was arranged as in Figure 6.5 and Plate 2, using the Fujica 54-270mm lens. In the focal plane of this lens was placed the image surface of a vidicon television camera tube. The output of the TV camera could be viewed on a monitor screen. An impression of the path of rays through the instrument is given in Figure 6.6.

The diffraction patterns are spread out mainly in a plane normal to the (long) direction of the slits, horizontally on the focal plane in this case. Vertically the lens focuses incoming parallel rays to a point as usual. The result is a broken horizontal line in the focal plane. Interference should occur within the dashes of light in this line, in effect a row of dots. At the longest focal length of the lens,

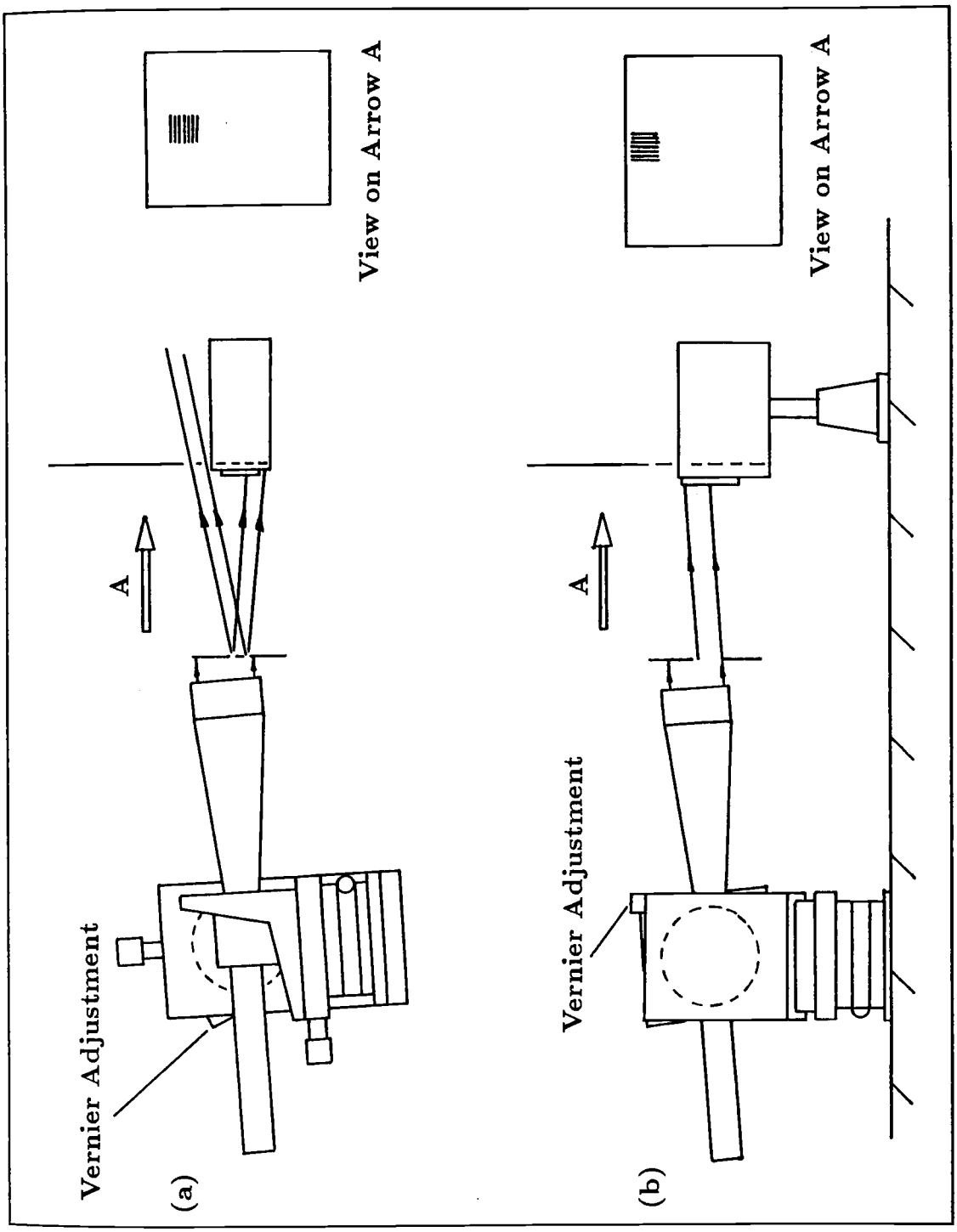


Figure 6.4: Movement of Interference Pattern with Collimated Beam and Changes in (a) Azimuth and (b) Elevation.

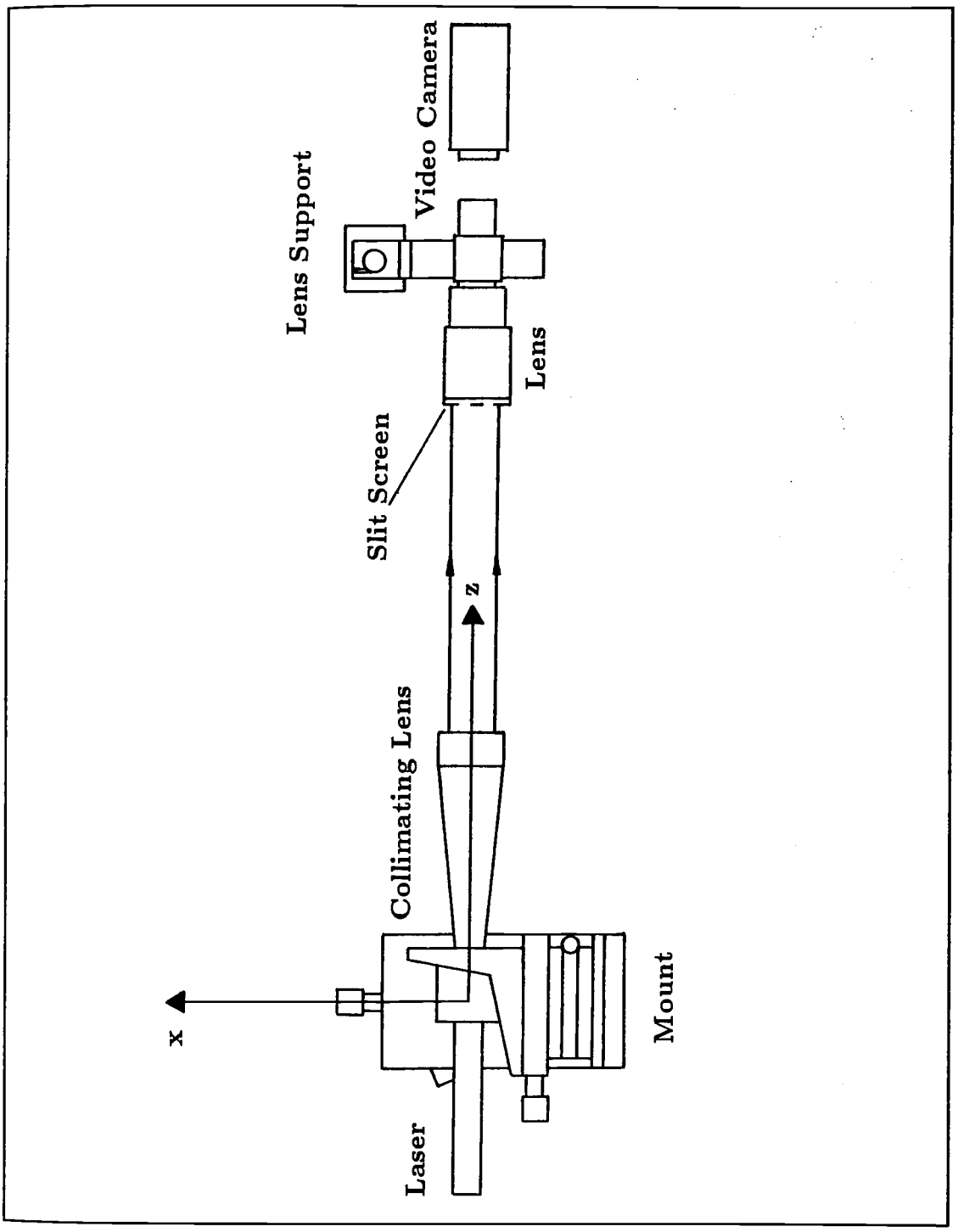


Figure 6.5: Interference with Collimated Beam and Lens.

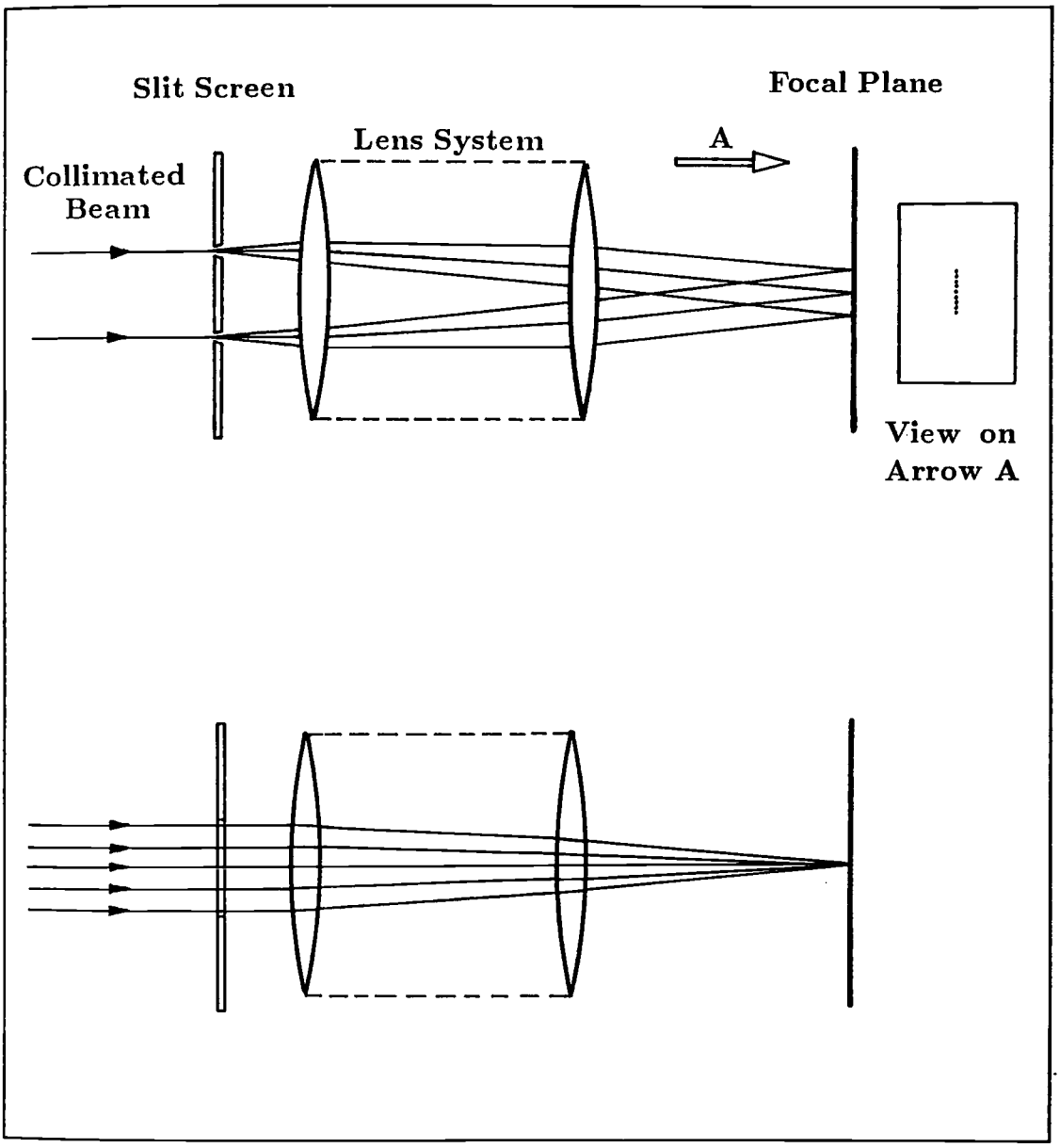


Figure 6.6: Collimated Beam Diffracted and Focused with Lens.

these fringes were just visible on the monitor. See Plate 3.

6.3.4 Split Wave Interference using a Lens and Incident Spherical Waves

With the collimating lens removed from the laser assembly the arrangement of equipment was as shown in Figure 6.7 and Plate 4. The image received by the vidicon was the set of parallel fringes shown in Plate 5. The height of these lines is due to the fact that the rays entering the slits are already diverging and are focused by the lens to vertical lines rather than the points which were obtained with parallel rays. See the ray diagram in Figure 6.8.

The closer the source point is to the slit screen, the greater is the angle of divergence (in vertical planes). At long distances the curvature of the spherical wavefronts at the slit screen becomes small and the interference pattern shrinks in height to approach that caused by plane waves.

It should also be noted that with the point source close enough to the slit screen the diffraction patterns from two slits diverge from each other in a plane normal to the slits to the extent that they will not overlap on the lens focal plane. This effect is less pronounced with fine slits than with wide slits, the diffraction angles being greater with narrow slits.

6.3.5 Movement of the Interference Pattern with Changes in Position of a Point Source

With the origin of the coordinate system centred at the point source, i.e. the pinhole, as shown in Figure 6.7, the complete laser and spatial filter assembly was moved by micrometers successively in the degrees of freedom provided by the laser mount.

As anticipated, rotation about the x axis and rotation about the y axis

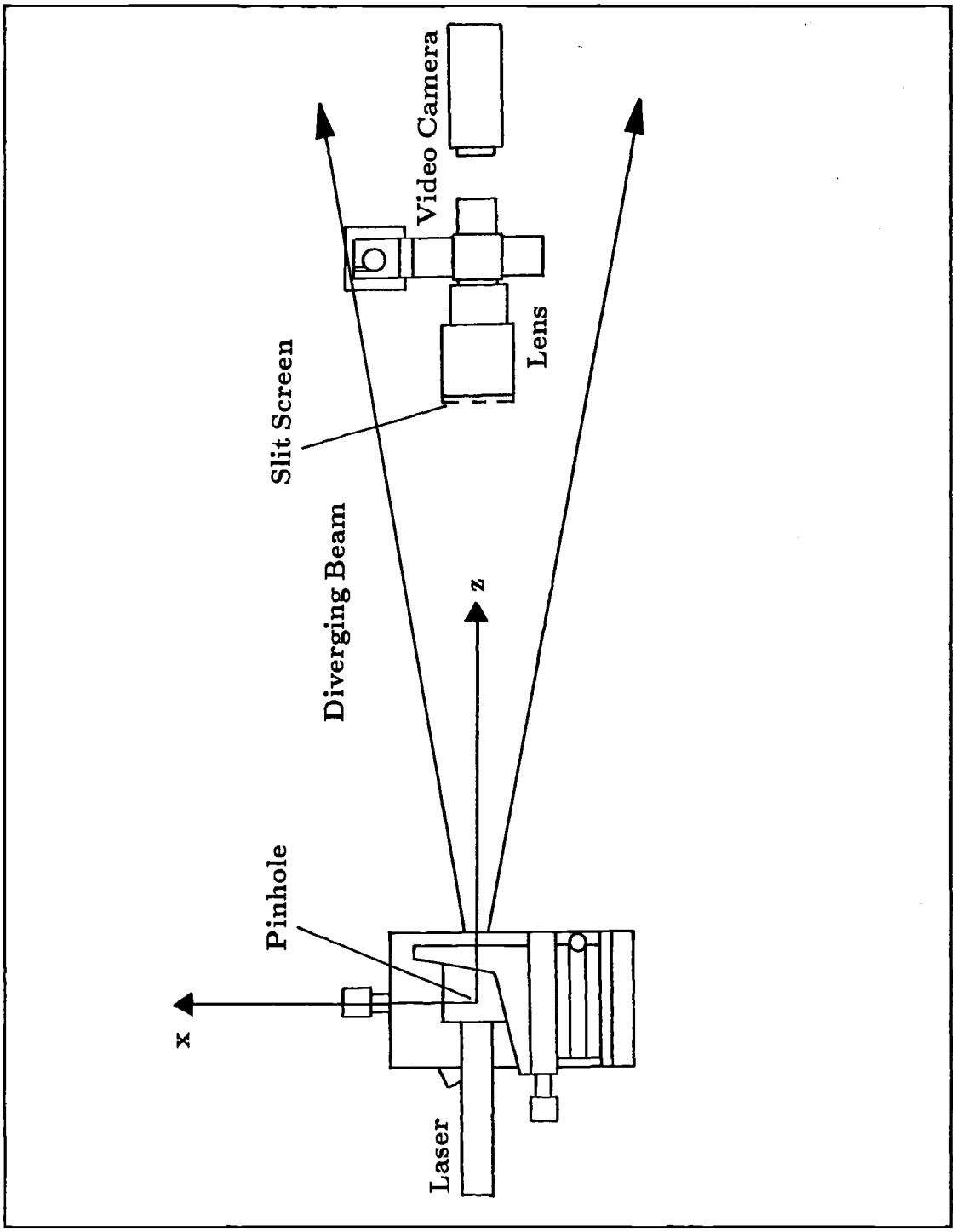


Figure 6.7: Interference with Diverging Beam and Lens.

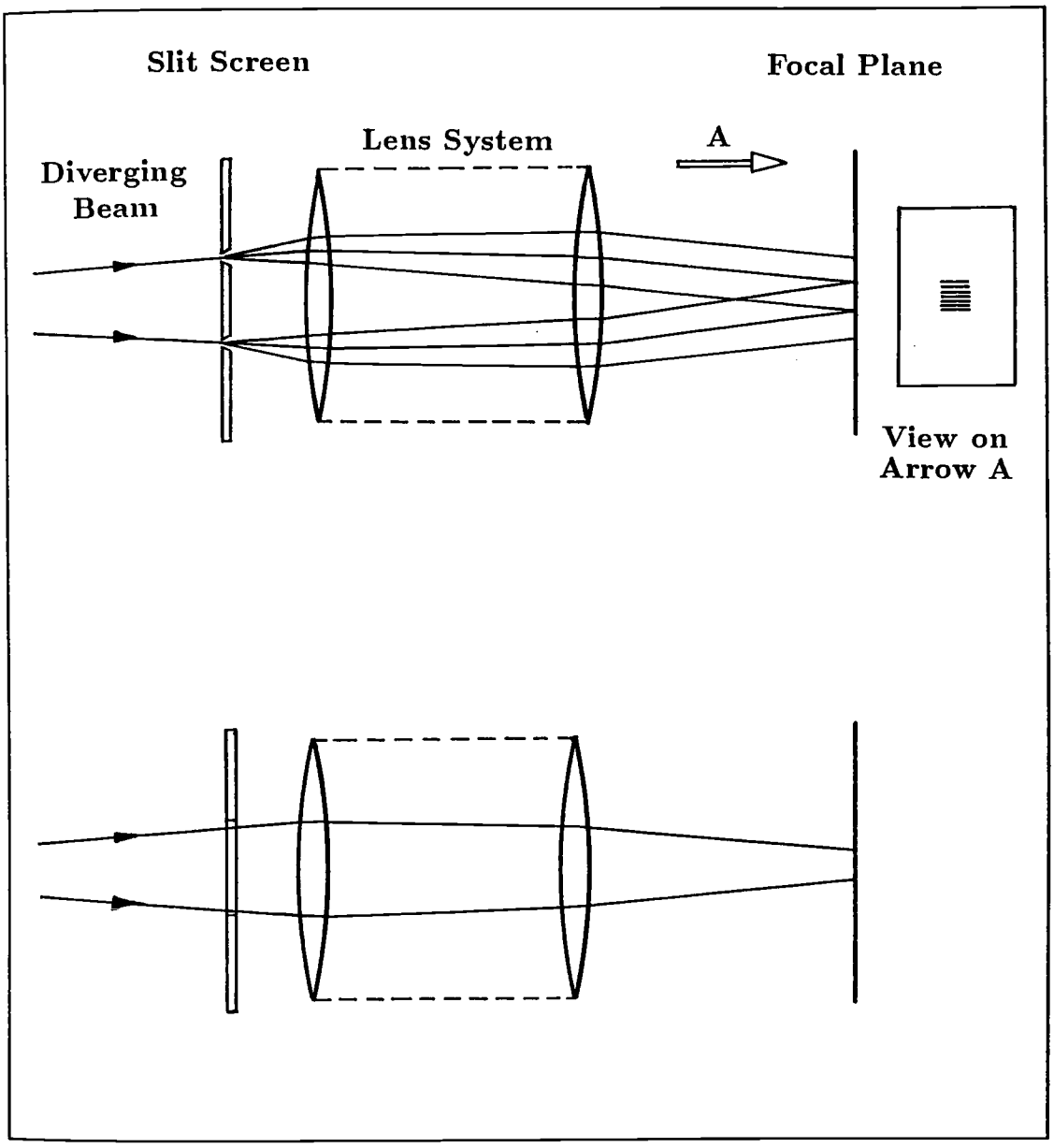


Figure 6.8: Diverging Beam Diffracted and Focussed with Lens.

produced no movement of fringes in the receiver, the pattern (Figure 6.8 and Plate 5) remaining stationary on the monitor. These motions may be thought of as simply rotating the same spherical shells through the receiver slits. No change in phase difference between the slits may be expected.

Also as expected, translation in the z direction caused no fringe movement, the distance from each slit to the source remaining equal in this case since the z axis is on the plane of symmetry of the slit screen and consequently on the central member of the family of hyperbolae with the slits as foci.

Translation in the x direction however may be expected to cause fringe movement because it moves the point source across the hyperbolae. See Figure 6.9. This in fact was the experimental evidence, translations in x causing the fringe lines shown in Plate 5 to move left or right across the monitor screen. See the ray diagram in Figure 6.10.

6.3.6 Comparison of Calculated and Measured Source Displacement

The objective here was to carry out an initial quantitative verification of the model of the optics described in Chapter 5 and demonstrated qualitatively above. Prior to carrying out the tests the slits were measured on a microscope cathetometer to give accurate figures for slit separation ($2c = 2.5166$mm) and slit width ($w = 0.1000$mm).

6.3.7 Verification of Changes in Azimuth of Plane Waves

With the collimating lens fitted to the laser, the incident angle of the collimated beam to the slit screen was adjusted about a vertical axis to cause a 2π shift in fringe position. See Figure 6.11. With the wavefronts initially parallel to the

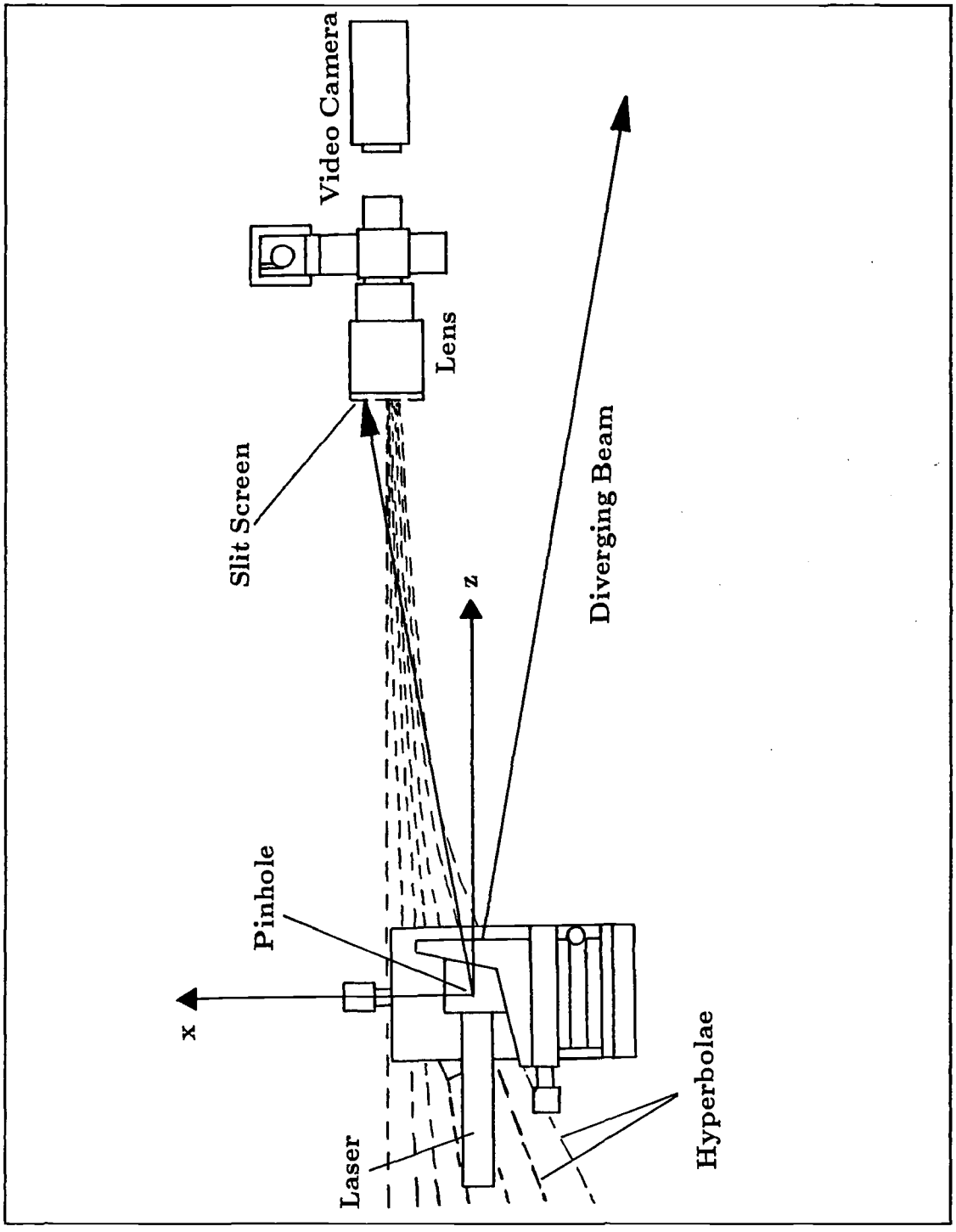


Figure 6.9: Interference with Diverging Beam and Lens with Point Source Displaced in X-Direction.

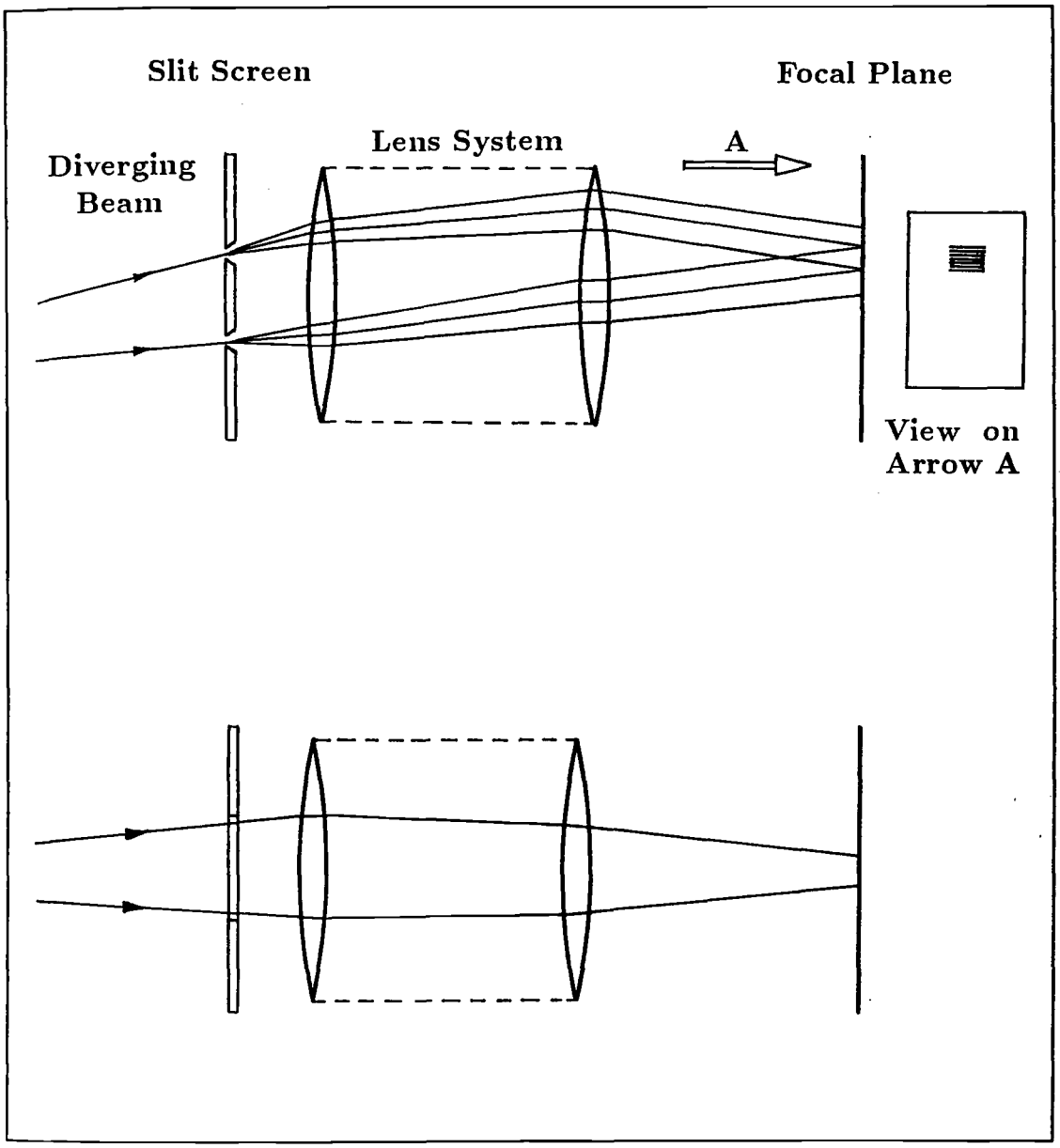


Figure 6.10: Diverging Beam from Point Source Displaced in X-Direction, Diffracted and Focused with Lens.

screen, a change in azimuth α causes fringe movement. $2n\pi$ movement of fringes is caused by $n\lambda$ change in optical path length difference d. In this case, with $n = 1$, $\lambda = 632.8$nm,

$$\sin\alpha = \frac{\lambda}{2c} = 2.5145 \times 10^{-4} \tag{6.1}$$

$$\alpha = 51.865\text{arcseconds} \tag{6.2}$$

Because n is almost linear with α at near normal incidence, the reading of azimuth change from the vernier scale of the azimuth rotation stage was made over $n = 10$ fringe cycles and the average taken for $n = 1$. This gave,

$$\alpha \approx 54\text{arcseconds} \tag{6.3}$$

i.e. a 4% error on the theoretical value. Although the rotation stage used has a sensitivity of 2 arcseconds, its accuracy is only 1 arcminute so some error is to be expected here. In addition there is some subjectivity in judging fringe movements on the monitor screen.

6.3.8 Verification of Changes in X-Coordinate of Point Source

With the collimating lens removed, the laser point source was moved in the x-direction to produce movement of the fringe pattern. See Figure 6.12. With $z = 1$m, $\lambda = 632.8$nm and a 2π shift in fringe position from that resulting from the positioning of the source on the central hyperbola (as near as could be determined), the intersection of the new hyperbola with the x axis was calculated.

For a hyperbola,

$$\frac{x^2}{a^2} + \frac{z^2}{a^2 - c^2} = 1 \tag{6.4}$$

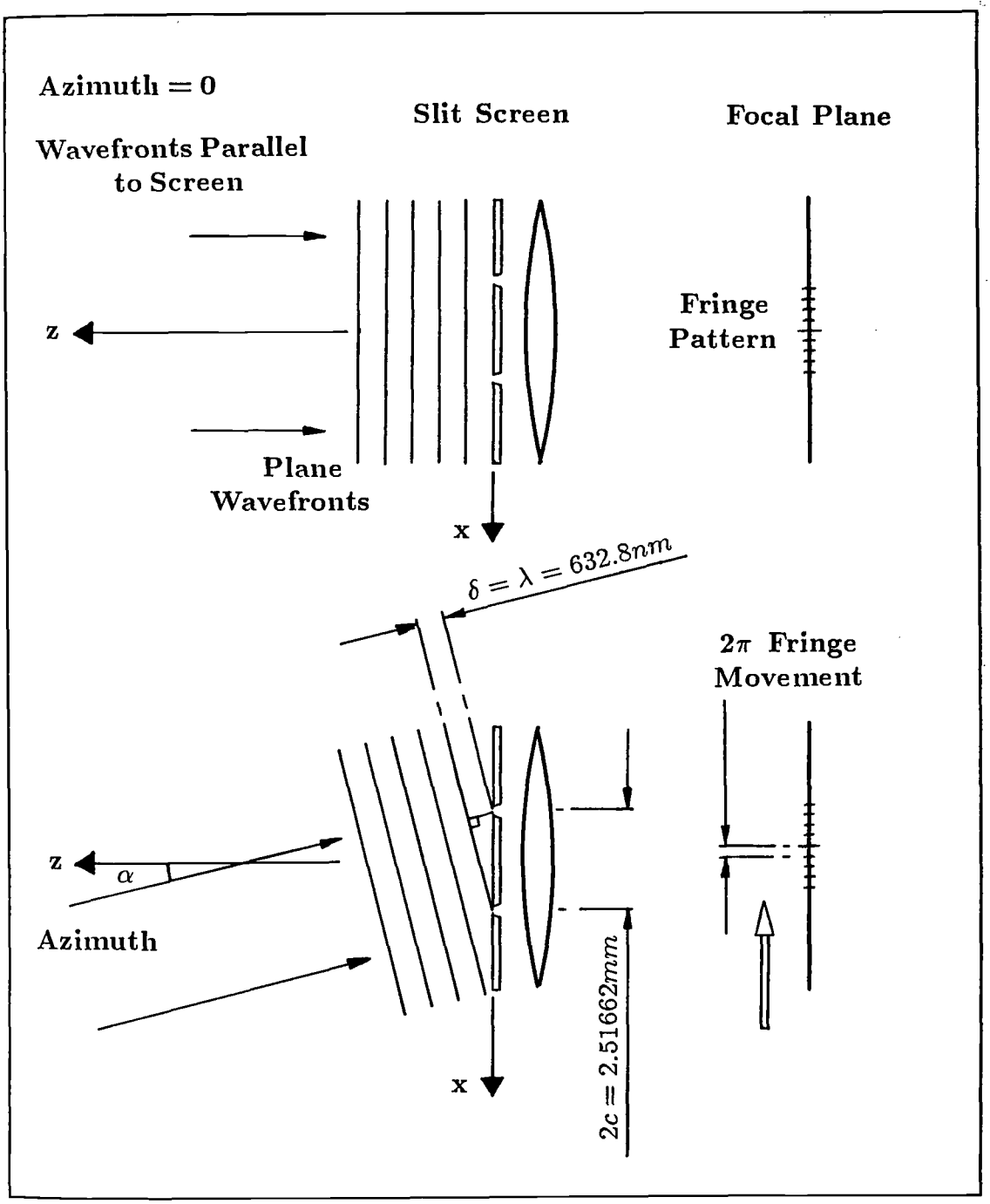


Figure 6.11: Movement of Interference Pattern with Collimated Beam, Lens and Change in Azimuth.

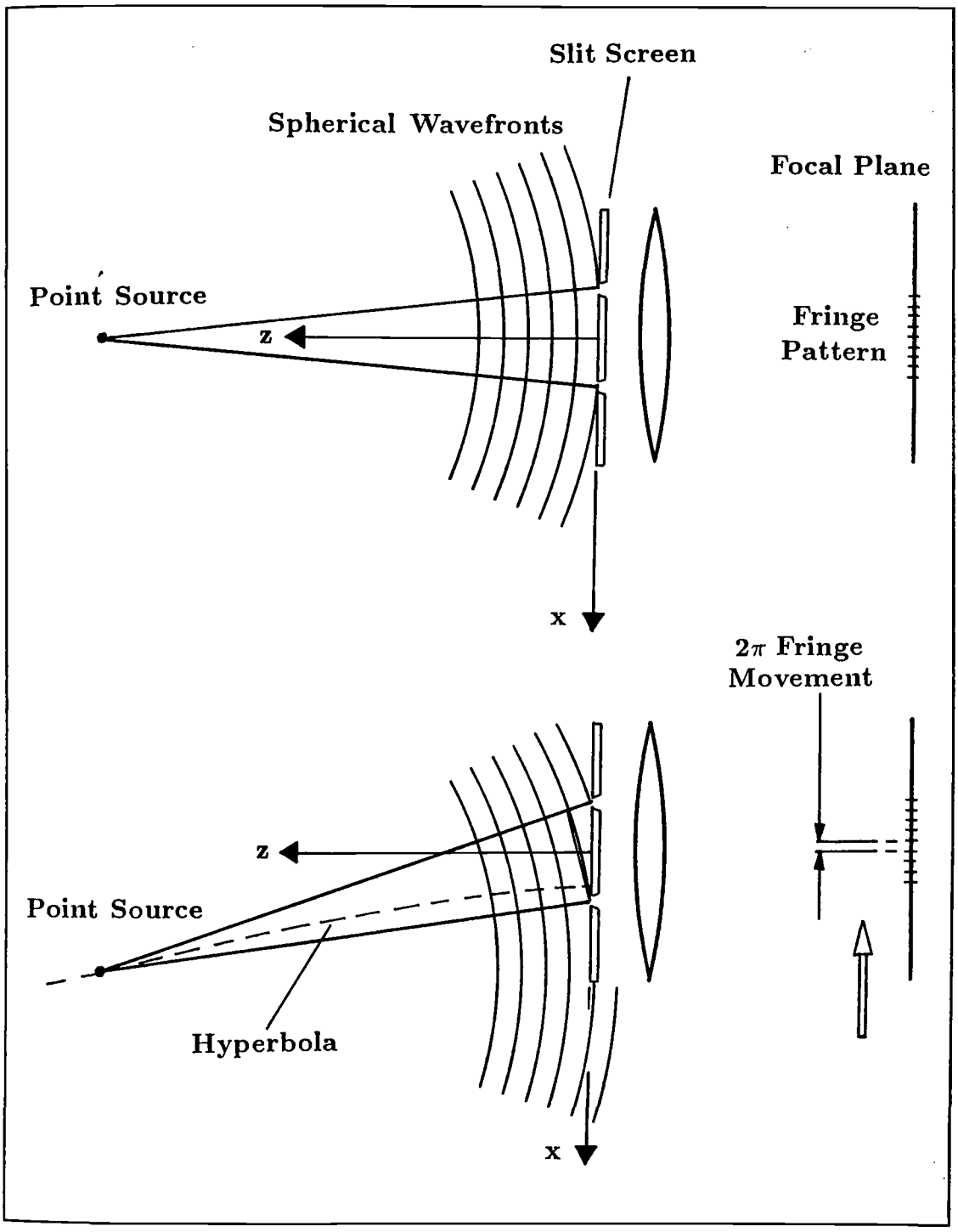


Figure 6.12: Movement of Interference Pattern with Diverging Beam, Lens and Source Displacement in X-Direction.

where $2c$ is the distance between foci (slit separation in this case) and $2a$ is the difference in distances from a point on the hyperbola to each focus ($n\lambda$ in this case). See Figure 6.13. With $2c = 2.5166 \times 10^{-3}$m, and $2a = \lambda = 632.8$nm,

$$x = 251.45\text{microns} \tag{6.5}$$

So, near the centre, this gives approximately 4 fringe cycles per millimeter moved by the source in the x direction. This was in fact what was observed on the monitor when the laser mount was moved with the x micrometer.

6.4 Lens Performance

6.4.1 Focal Length

In general, the size of the image and the field of view of a lens are related by its focal length. Usually the image size is fixed and the focal length is chosen to give the object field required. This is illustrated in Figure 6.14. A 35mm film has a frame size of 36 by 24mm. The diagonal of this rectangle is 43.27mm, suggesting that a 35mm camera lens might be appropriate for producing an image on the central region of a 50mm diameter grating. To obtain a fringe frequency of 150 lines/mm on the focal plane from slits 5mm apart requires a focal length of 52.68mm.

6.4.2 Off-Axis Performance

A bundle of parallel rays filling the input aperture of a lens system may be interrupted in part in its passage through the lens by various stops. Rays parallel to the optical axis are transmitted only in a central diameter. Rays incident at some angle to the axis are, in the main, transmitted only if they pass through

Geometry:
P is any point on the hyperbola,

$$\frac{x^2}{a^2} + \frac{z^2}{a^2 - c^2} = 1$$

A hyperbola is the locus of points at which $PF_1 - PF_2 = \text{constant} = 2a$. The distance between the foci is $2c$.

Apparatus:
F_1 and F_2 represent two slits in the $z = 0$ plane. The difference in optical path length from P to each of these slits is $2a$.

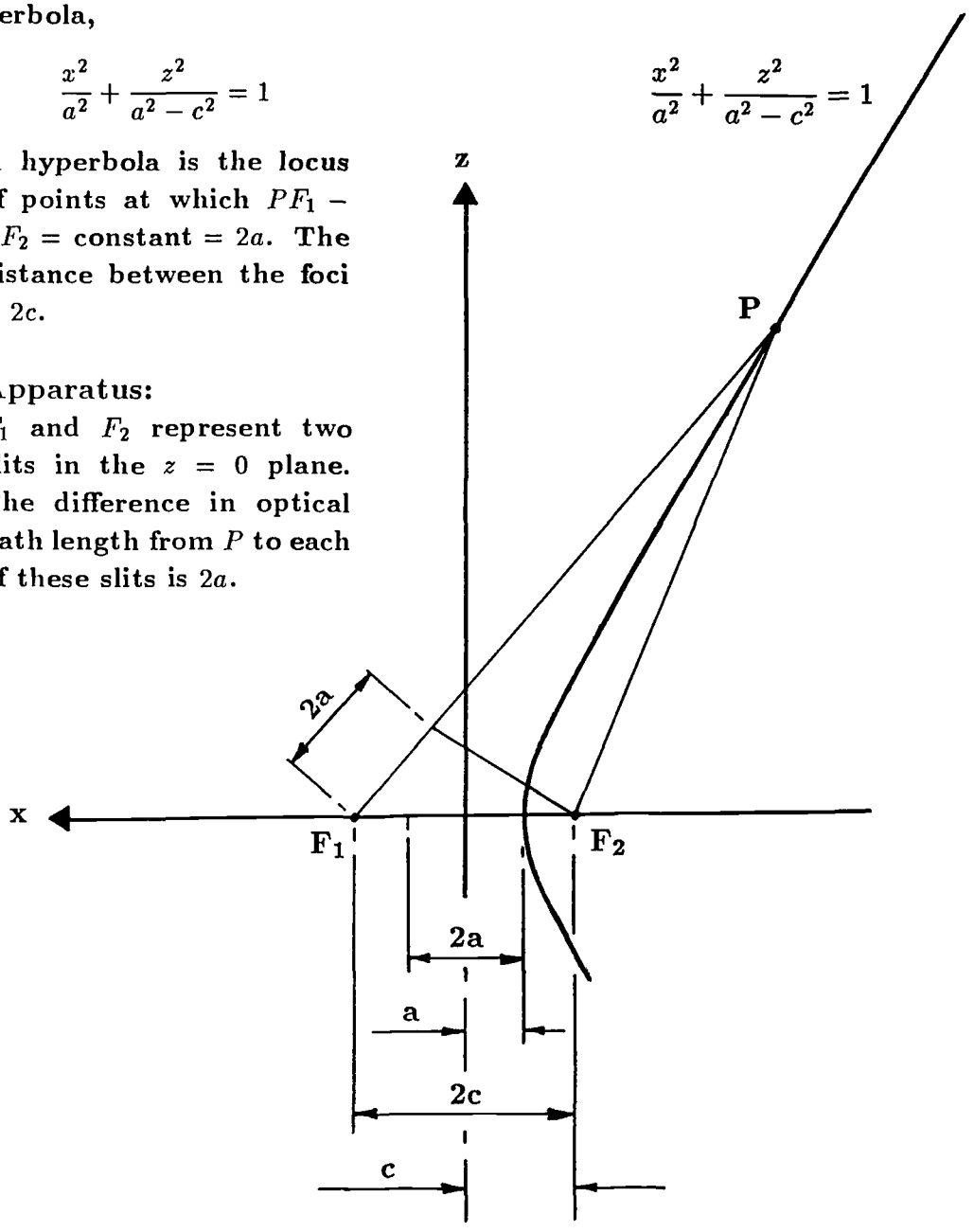


Figure 6.13: Hyperbola Geometry.

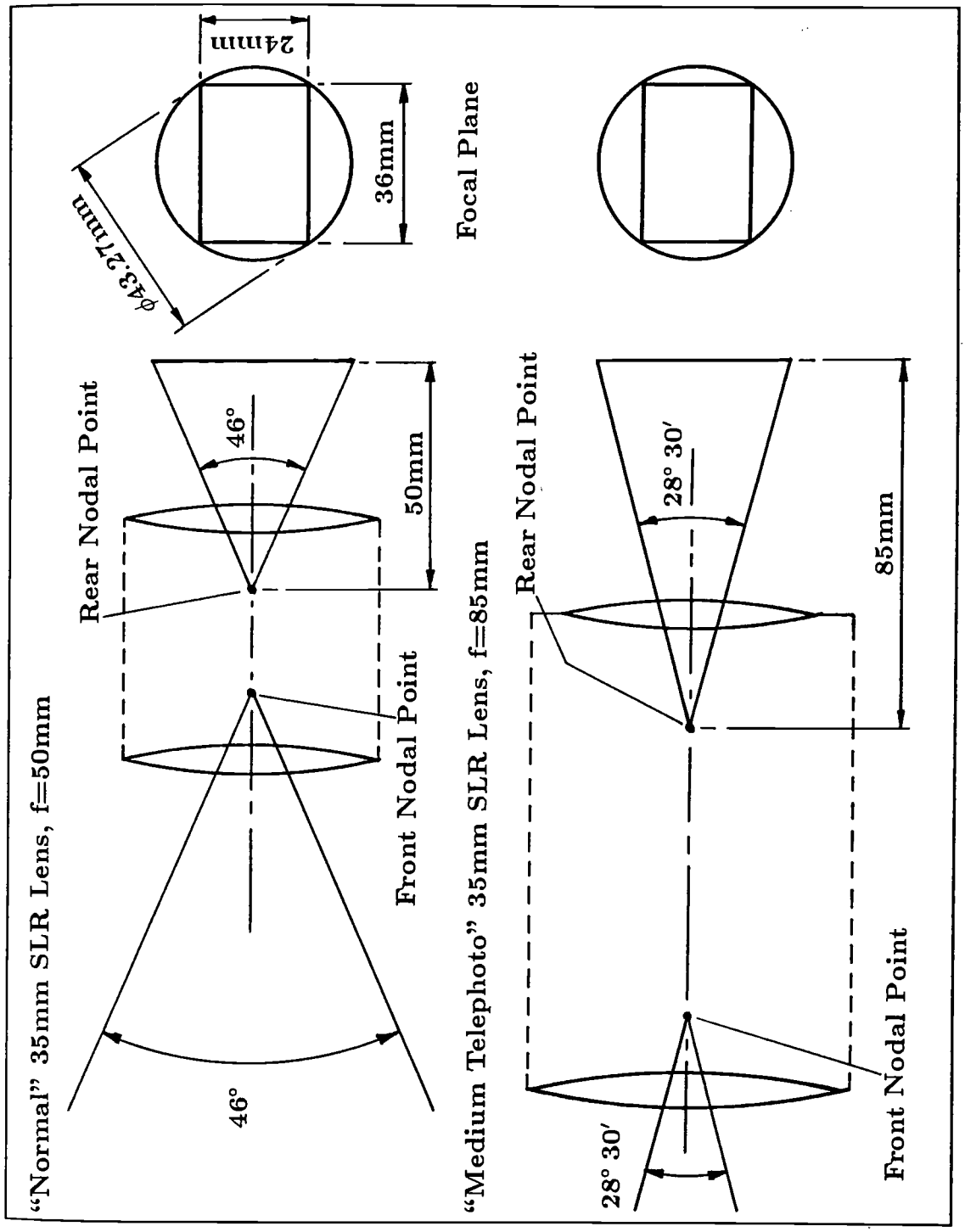


Figure 6.14: Lens Focal Length, Object and Image Fields.

the input lens on the side nearest the source. See Figure 6.15.

This effect occurs to a greater or lesser degree in most camera objectives and is a result of various compromises and optimisations in the design of lenses for that purpose. The consequence of this is a reduced choice of lenses suitable for use as the objective of the interferometer. A lens optimised for a small camera causes vignetting of diffracted rays from the slits as shown in Figure 6.16. The effect can be seen at the output of the lens by placing a screen nearer to the lens than the focal plane i.e. before the diffraction patterns meet. As the source is moved further off-axis, the diffraction pattern from slit S_2 is progressively reduced in height as more of its constituent rays are blocked by stops within the lens. On the image plane, therefore, the height of the interference fringes reduces steadily as the source moves away from the optical axis, reducing in turn the amplitude of the signal at the detector.

This difficulty can be alleviated by using a lens such as the Schneider Kreuznach 240mm. The lens used was similar in construction to the Schneider "Xenar" range for large format cameras, being based on the Cooke triplet design. With this lens considerable angles between source, lens and optical axis were achieved.

What is required, therefore, is a lens with this characteristic of the Schneider large format lens but with a focal length of about 50mm to give a fine interference pattern and high resolution. To this end, some market surveying, borrowing and testing of lenses was undertaken. Obviously it is preferable to use a lens "off the shelf" for this purpose if possible.

6.4.3 Lens Tests

Some tests were carried out on lenses to investigate their suitability for use in the interferometer. The requirements for a useful object field and high resolution

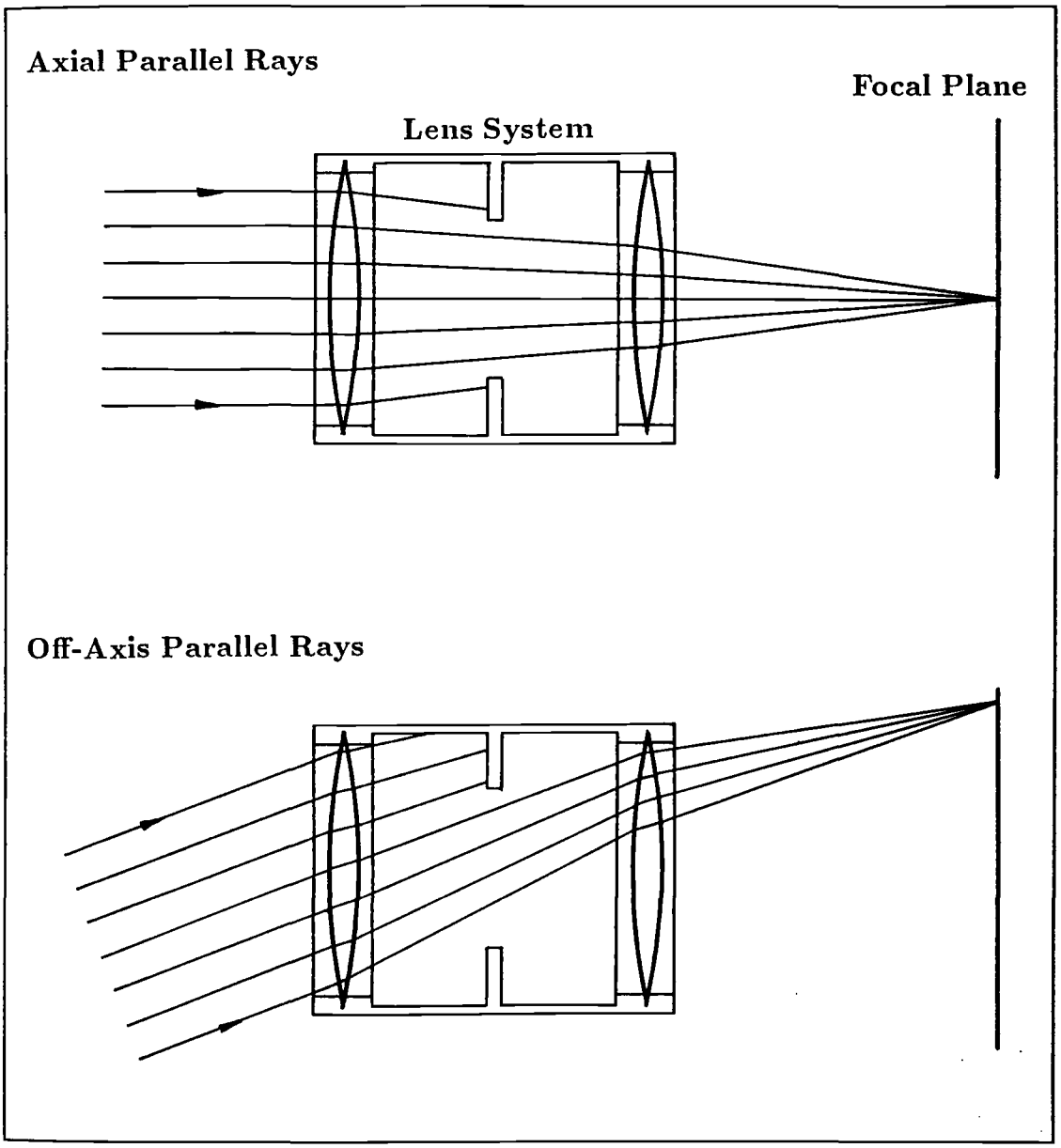


Figure 6.15: Vignetting of Parallel Rays in Lens System.

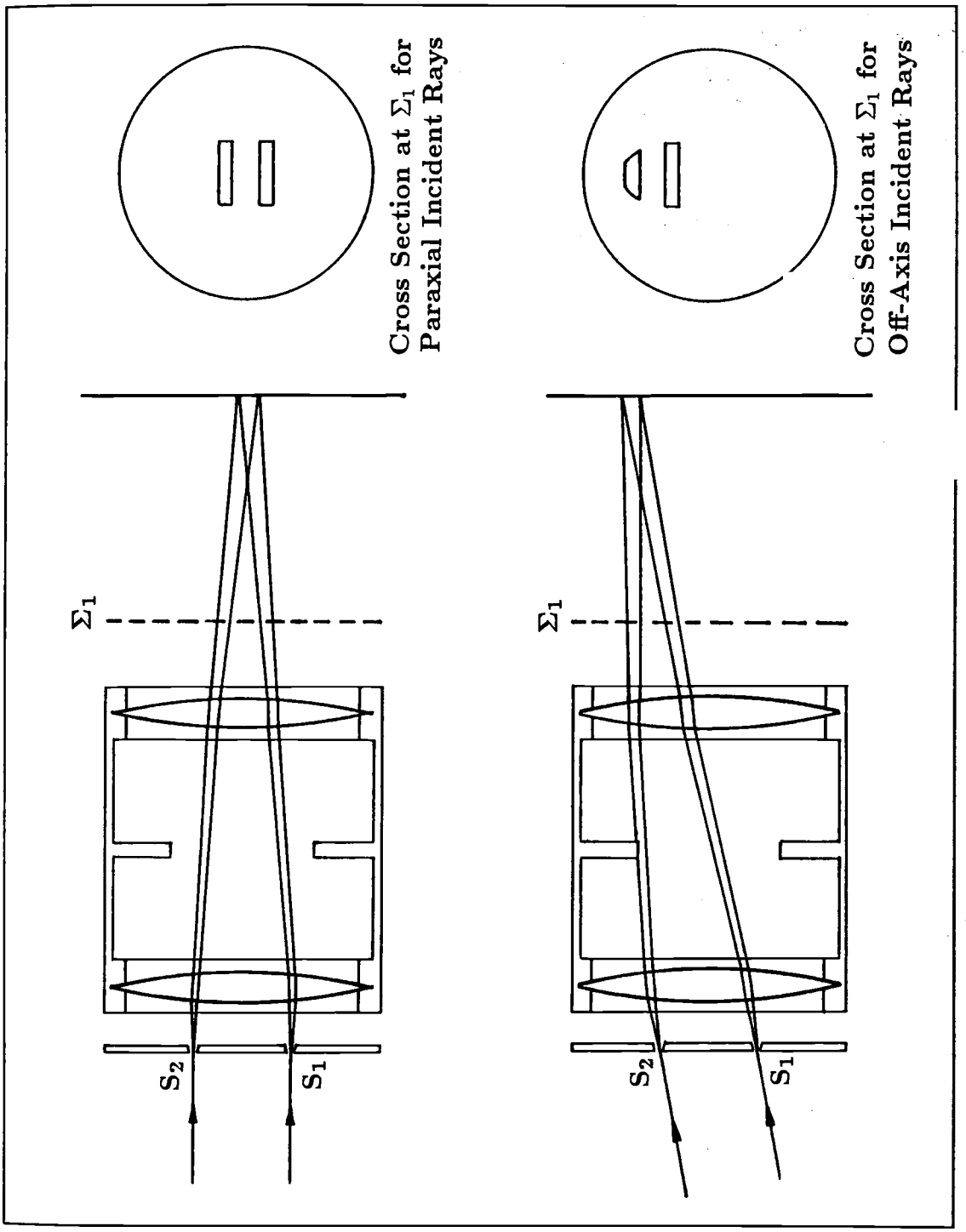


Figure 6.16: Vignetting of Off-Axis Diffracted Rays.

mean that a lens of comparatively short focal length is required. In these circumstances the interference pattern of parallel fringes is small, to the naked eye not much bigger than a pinhead on the image plane. The question arose therefore, as to whether the fringe patterns remain undistorted at this microscopic scale.

Records of the interference patterns obtained with various lenses were made on photographic film in a camera body fitted to the back of the lens under test as in Plate 6. High resolution Kodak Technical Pan 35mm film was used. The developed negatives were then re-photographed through a microscope, scaling to normalise approximately the fringe pattern size on the prints for different tests.

The lenses and focal lengths used were as follows. Firstly, the Fujica 54 to 270mm lens was tested at nominal focal lengths of 270mm, 200mm, 150mm, 100mm, 75mm and 54mm. Another zoom, the Nikkor 35 to 70mm lens was tried, exposures being made nominally at 70mm, 65mm, 50mm, 42.5mm and 35mm focal length. Finally the fixed focus 55mm Micronikkor was tested.

Some examples of the fringe patterns, much magnified, can be seen in Plates 7 and 8. The evidence is that, notwithstanding the problem of vignetting, straight, parallel interference fringes can be produced even with fairly inexpensive optics.

6.5 Effects of Air Movement and Temperature

One potential problem with the use of any interferometer is that the velocity of light and therefore the refractive index in a medium varies with the density of the medium. This relationship is derived from a consideration of the electric dipole moment of the molecules of the medium, which leads to the Lorentz-Lorenz formula,

$$\alpha = \frac{3}{4\pi N}\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{3}{4\pi N}\frac{n^2 - 1}{n^2 + 2} \tag{6.6}$$

where α is mean polarisability, ε is permittivity, N is number of molecules per unit volume and n is refractive index. Proceeding from this formula it can be shown (see [Born and Wolf]) that for a gas,

$$\frac{RT}{p}\frac{n^2 - 1}{3} = \text{constant} \tag{6.7}$$

where R is the gas constant, T is absolute temperature and p is pressure. Using this and the relationship between n and λ for some frequency, it can also be seen that quite small temperature changes and thus small changes in n give rise to large changes in the number of wavelengths of light in the optical path. So with different temperatures, even locally, in the two paths between transmitter and receiver the wavefronts become distorted. Interestingly, in the steady state case this does not matter, since the wavelength and frequency at the receiver remain the same and movement of the source in space causes relative phase change between the slits as before.

In a non steady state situation, that is, while the refractive index of a portion of one of the optical paths is changing, distortion of the interference pattern may be caused. The reasons for this are as follows.

With the source emitting a frequency f_0 and considering a section of length x of the optical path in which n is changing with time, the number of cycles contained within length x changes. For instance if the air inside length x is cooling, the value of n there rises with time and the velocity through x decreases. In that circumstance the number of cycles within x is increasing at some rate f' cycles per second.

The frequency emerging from the section x is therefore,

$$f = f_0 - f' \tag{6.8}$$

$$f = f_0 - \frac{d}{dt}\left(\frac{x}{\lambda_1(t)}\right) \tag{6.9}$$

where $\lambda_1(t)$ is the instantaneous wavelength within x and depends on n, the instantaneous refractive index there.

So the frequency f can be greater or less than f_0 as n is changing, and if this change is different in each optical path, phase shifts between the paths will be registered at the receiver causing spurious pattern movements. With the same change of f at each slit, in theory the pattern would shrink or expand. However, high rates of change of n would be necessary to have an appreciable effect and fast changes of temperature or pressure are required to affect density in this way.

No attempt was made to measure changes in frequency or to evaluate the rates of refractive index change likely to occur in practical situations. A "worst case" type of test was carried out however. This was simply to blow hot air across the optical paths of the interferometer and observe the fringes on a television monitor.

With the hot air stream crossing the optical path near the source, no effect was observed in the parallel fringes. The hot air stream was moved progressively nearer the receiver, surprisingly still with no effect, until at only about 350mm from the receiver the pattern on the monitor began to waver. This was an encouraging result as far as the robustness of this optical arrangement is concerned, particularly when use of the instrument in the open or in industrial environments is considered.

6.6 Detection of Interference Pattern Movement

So far, experimental evidence of the production of fine, straight interference fringes and of the relationship between their position and the position of a point source has been presented. The next stage is concerned with the detection of movement of this pattern in such a way that a signal representing the displacement of the source may be obtained. This was done using an amplitude transmission linear diffraction grating placed in the focal plane of the lens as described in Chapter 5.

6.6.1 Verification of Grating Output Signal Modulation

To check experimentally the behaviour of the interferometer as a system required that a combination of slit separation, lens focal length and grating period be used such that the interference pattern line spacing matched that of the grating.

In the first instance equipment was set up as in Plate 9 and Figure 6.17. An expanded, collimated beam is incident on a photographic grating (5 line pairs/mm) in front of the Fujica lens. The fringes produced by this square grating (or multiple-slit screen) are a series of dots with sharp peaks of intensity separated by several subsidiary maxima and minima, as shown in Figure 6.18.

The focal length of the lens was adjusted until the scale of this pattern in the focal plane matched the period of a second photographic grating (also 5 line pairs/mm). A radiometer based on a photodiode was placed behind this and the output from the radiometer taken to a cathode ray oscilloscope.

By rotating the laser mount about a vertical axis, the angle of incidence to the first grating was varied. This caused movement of the interference pattern as in Figure 6.11.

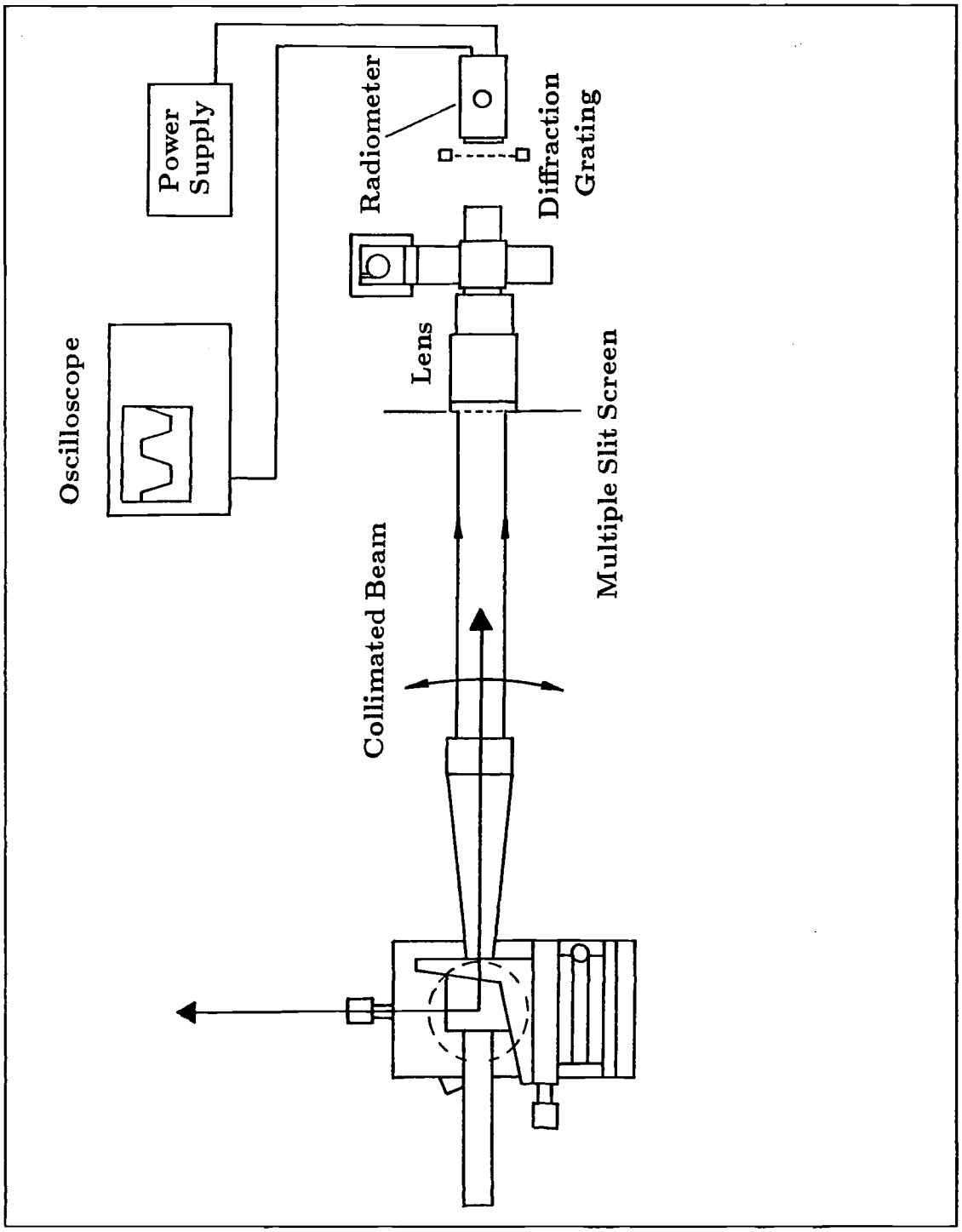


Figure 6.17: Sensing Collimated Beam Angle with Multiple Slits and Diffraction Grating.

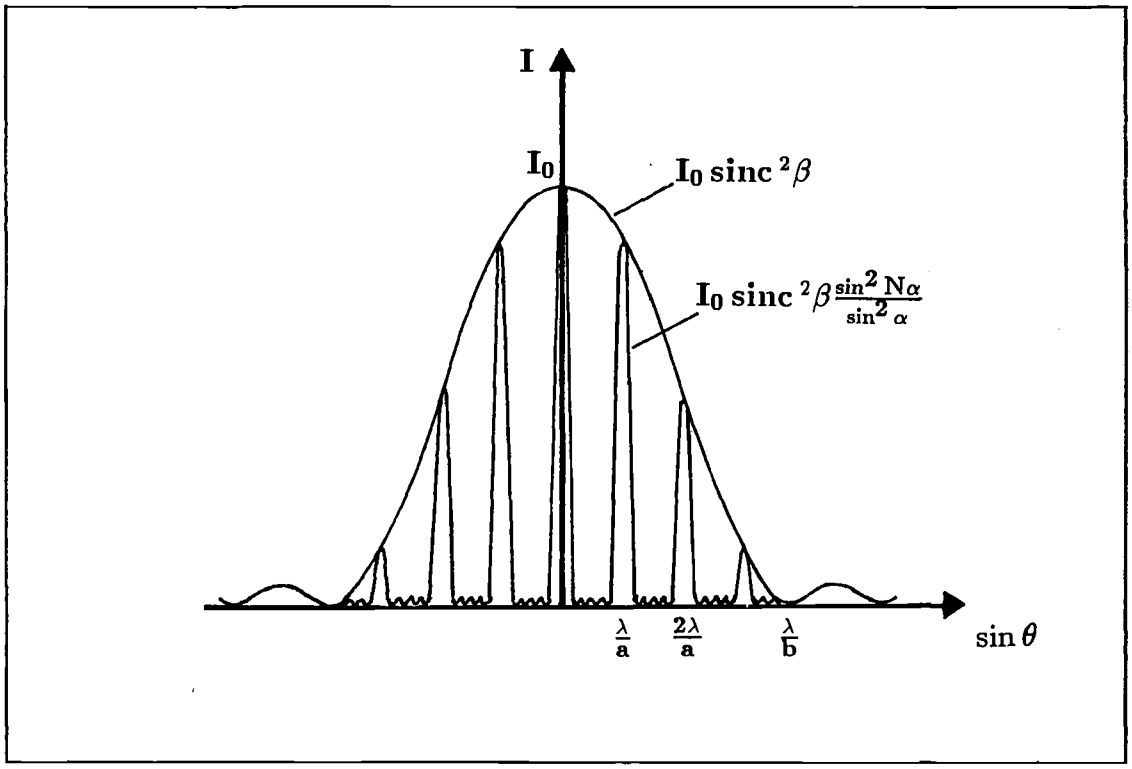


Figure 6.18: Multiple Slit Diffraction.

The signal received by the radiometer was monitored on the oscilloscope and could be seen to vary with incident angle. Rotating the mount as smoothly as possible by hand, the oscilloscope could trigger on the radiometer output and momentarily, the waveforms could be seen. These were not of sinusoidal form, because of the more complex irradiance function due to the first grating, but were regular and periodic. The sensitivity to angle was such that 10 cycles were observed with a rotation of the incident beam of 85 minutes of arc i.e. 8.5 arcminutes per cycle.

Using multiple slits in this way allows much more light into the system, increasing the amplitude of the modulated output signal and therefore the resolution which may be obtained. This only works for incident plane waves, practically, but is useful in that role.

The second experiment uses a two slit screen with 100 micron slits 10 millimeters apart. The lens, set to a focal length of 104mm, scales the interference pattern correctly for a 150 cycle per millimeter diffraction grating. As before, a collimated beam was used and the angle of incidence varied with the micrometer adjustment of a rotation stage on the laser mount assembly. See Figure 6.19.

A photomultiplier was used to receive the grating output and the signal from the anode was viewed on a cathode ray oscilloscope. The waveforms obtained had cosinusoidal form as predicted in Chapter 5. These signals were held in a storage oscilloscope and printed out on an XY plotter. See Figure 6.20. With the optics as described and starting from normal incidence, a 2π phase change between the two slits corresponds in theory to an angular movement of 13.05 arcseconds. Carrying out the measurement with the equipment and averaging over 100 cycles gave 13.8 arcseconds per cycle as well as could be read from the vernier scale of the rotation stage.

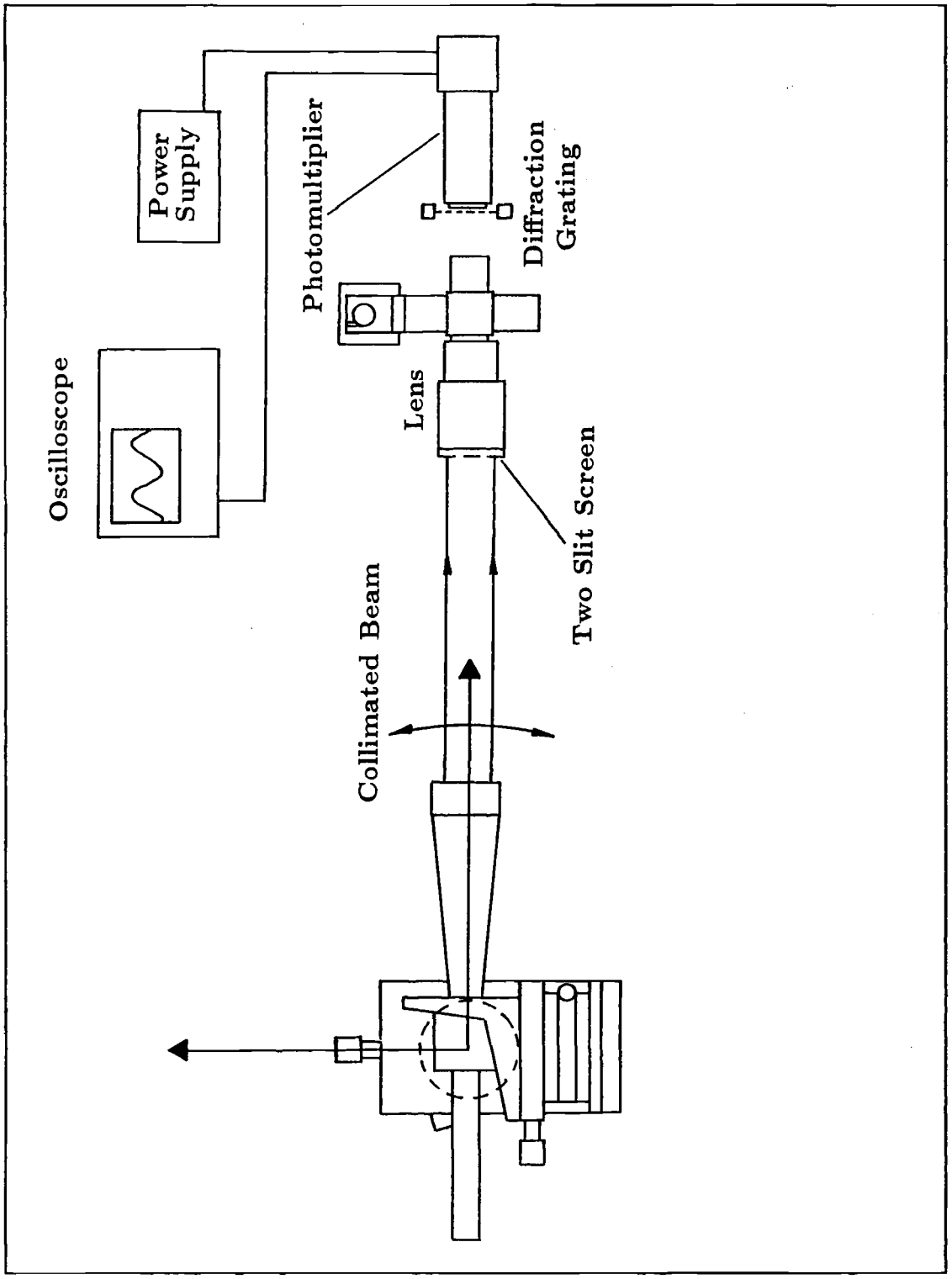


Figure 6.19: Sensing Collimated Beam Angle with Two Slits and Diffraction Grating.

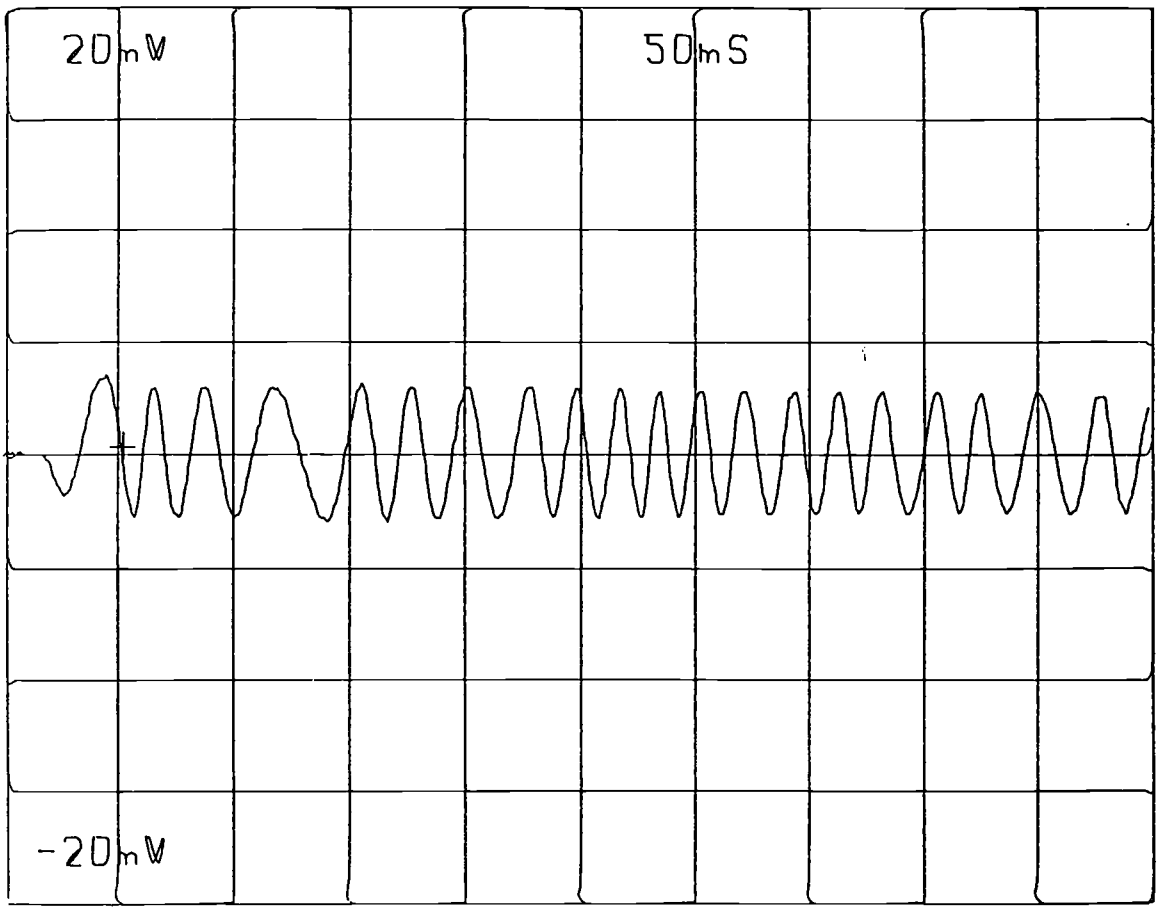


Figure 6.20: Graph of Photomultiplier Output.

6.7 Quadrature

Having obtained the grating output signal, the next step was to use a beamsplitter immediately behind the lens to divide the diffraction patterns and in effect give the lens two focal planes. The equipment was arranged as in Figure 6.21 and the beam splitter cube adjusted in x and y until its two output faces lay in the focal planes of the lens and beamsplitter combination. The presence of this cube of glass in the image field increases the back focal distance of the lens and must have an affect on the shape of the focal surface as well as its position. These changes were not explored in detail.

As the beam angle was changed, two "sinusoidal" signals could be seen on the oscilloscope. The position of one of the gratings was adjusted until these signals were in quadrature. Figure 6.22 is the plot taken from the oscilloscope.

6.8 Optical Noise

In the above experiments, even with interference filters on the photomultipliers the 100 Hz frequency of the room fluorescent lighting was picked up easily on the oscilloscope. Obviously the power of normal artificial lighting in this band of the spectrum (632.8nm) is significant in comparison to the 1mw power of the laser used. This problem may be circumvented in several ways, the more obvious being to switch the room lights off or to use a more powerful laser. To avoid these inconveniences a different type of area lighting might be used, the main criterion in the choice of suitable lighting being that it should have low power in the band of the spectrum used by the instrument.

A better solution is as follows. The signal from the transmitter is amplitude modulated at high frequency. The received signal is then filtered for this

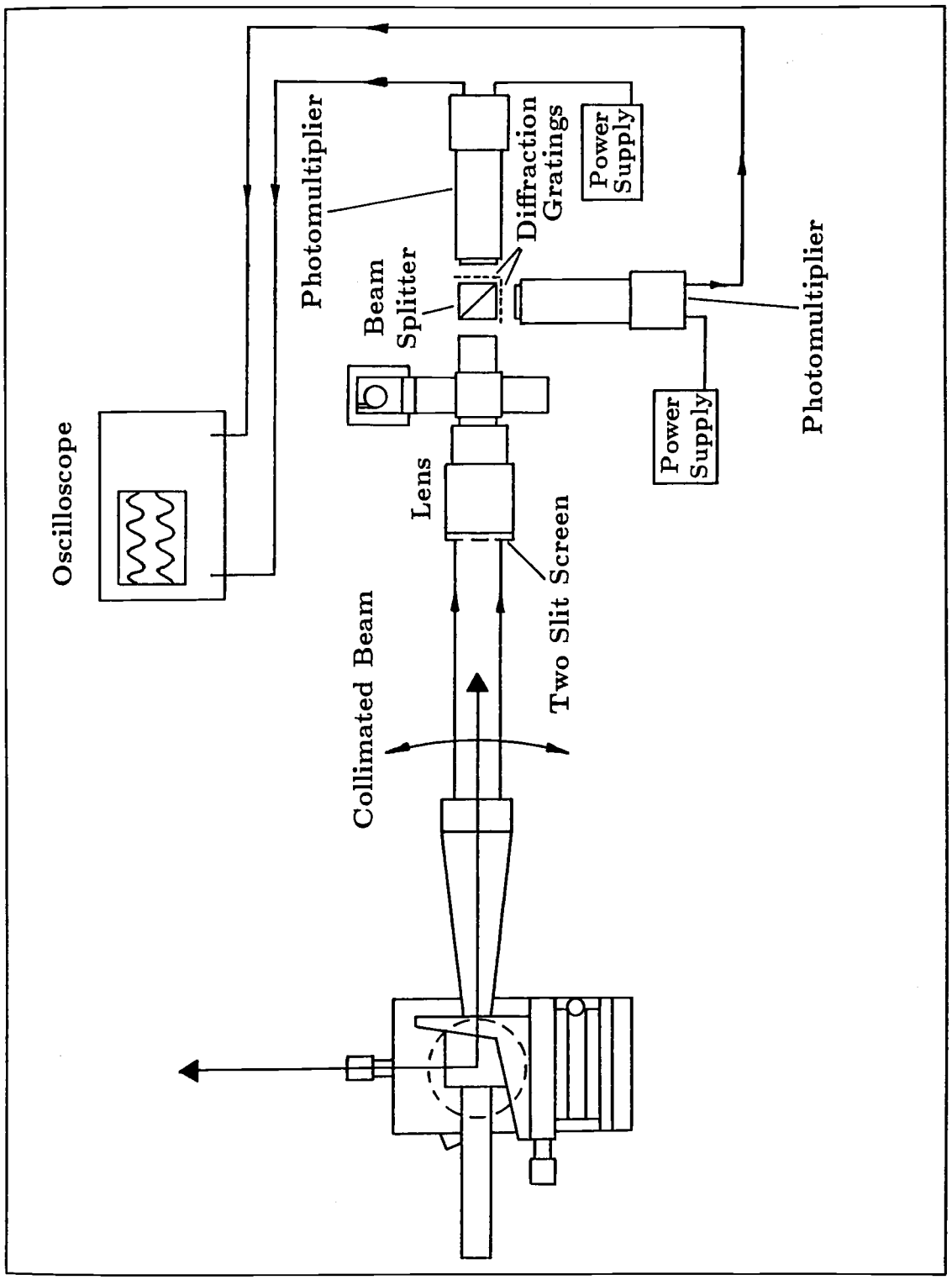


Figure 6.21: Sensing Collimated Beam Angle using Two Signals in Quadrature.

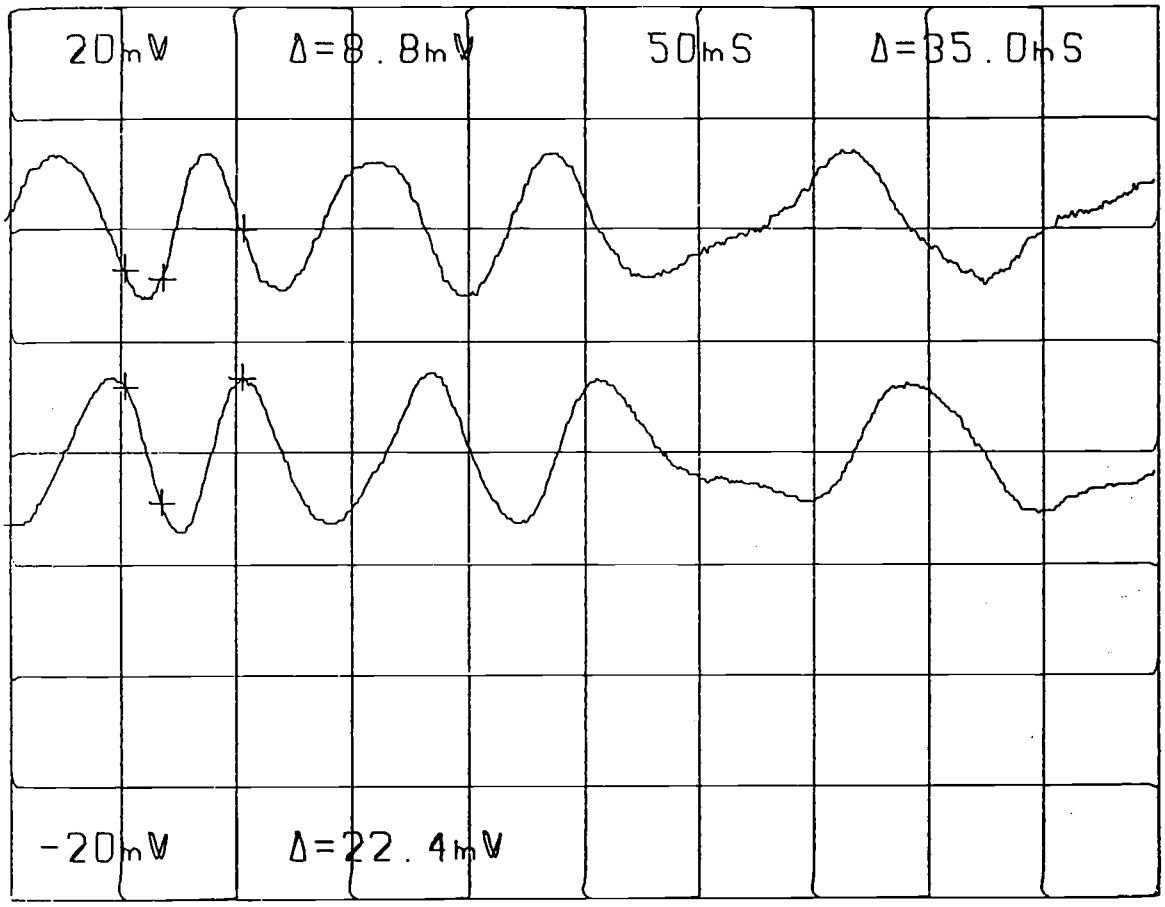


Figure 6.22: Graph of Photomultiplier Output Signals in Quadrature.

frequency and the changes in amplitude due to phase differences are used to reconstruct the output periodic wave caused by source movement.

The effects of noise from ambient natural light have not been explored in these experiments.

Chapter 7

Implementation of Systems

7.1 Introduction

In this chapter the means of implementing a dynamic three dimensional measurement system is considered for the most general case in which it is required that the position and orientation of an object with six degrees of freedom be monitored. The principles of operation of the interferometer have been given in Chapters 4 and 5 and the method whereby the coordinates output from the instrument may be used to locate points and bodies is described in Chapters 2 and 3. The questions which remain as regards implementation concern:

1. The optimum positioning of instruments to cover the required working envelope.

2. The design of the interferometer to give the resolution required for the task.

The example used here to illustrate the process is a requirement to be able to measure the position and orientation of the end effector of a robot arm, moving in a $2m$ diameter spherical working volume. The resolution required of the measuring instrument is $20\mu m$. It is assumed that the moving robot

carries several point sources on its end effector. The positions of these sources are determined using interferometers located at appropriate positions outside the working volume. The instruments should be arranged around the working volume in such a way as to allow the position of points on the robot end effector to be measured anywhere in this volume. The orientation of the end effector can then be calculated from the position data on the several points.

Meeting the requirements of the example is a design process and, as with other engineering design procedures, there is not a single analytical route which leads to a single solution. The procedure used in the next two sections is the nearest approach to such a process which can be achieved.

7.2 Location of Instruments

This section deals with the requirement in Item 1 of the last section, and is largely concerned with geometry.

7.2.1 Implementation Tasks

Discussion

The situation described in the given example consists of several point sources, fixed in position relative to each other, but free to move as a group within the working volume. Each of these point sources transmits light within a conical solid angle. The directions of the cone axes are fixed relative to each other, but as a group they change direction as the end effector changes its orientation.

The interferometers can be considered as *receivers*, positioned around the working volume and outside it. Each of these instruments has an *object field angle*, a conical solid angle within which it can receive radiation from a moving point. The receivers must be sufficiently far away from the working volume, to

be able to "see" the whole volume within their object field angle. On the other hand they must not be so far away from the moving points that the resolution of their measurements becomes too low.

In this case the working volume is a sphere, and the symmetry of the situation suggests that there is no particular reason why any one of the receivers should be further away from the sphere centre than any other. The receivers can therefore be considered to be distributed over the surface of another, larger sphere, concentric with the working volume. In implementing a system, the radius of this sphere must be determined, and the optimum number and distribution of receivers on its surface must be found.

As the end effector rotates, the solid angle illuminated by each point source must always include at least one receiver, if a continuous record of the trajectory of that point is to be maintained. Also, as the object point gets closer to the receiver sphere, the cross-section through its cone of illumination at the sphere becomes smaller, suggesting that receivers be closer together on the sphere surface in this situation. This implies a larger number of receivers and a corresponding increase in the cost of the system. Conversely, a larger number of transmitting components on the robot end effector can reduce the number of receivers required. If one point source moves to a position where it no longer illuminates a receiver, information about several other point sources may still be available. It is important to find the number and distribution of both transmitters and receivers which minimises the number of instruments required.

All the above factors must be taken into consideration when designing a test station such as the one required in the example of this chapter. The solid geometry of the situation is not unique to the interferometer however. Similar problems occur with other three dimensional systems with both active and passive mov-

ing components. For instance the system with a moving two axis servoed mirror and the system employing moving retroreflectors (Chapter 1, 1.5.12) both have their implementation constrained by this geometry. In all such cases, the geometrical description begins with a moving cone of some known solid angle, and receivers with a known object field or acceptance angle. Similar problems arise in establishing how these parameters affect the optimum positioning of multiple receivers (or trackers) around their working volume.

Tasks

In the remainder of this section, the following problems are addressed:

1. How are the optimum number and distribution of receivers determined from a knowledge of the transmission angle at a single moving point?

2. In a realistic case in which several point sources must be used, how is this distribution affected, and how does the receiver object field angle affect the particular design for the robot arm given in the example?

3. What are the maximum and minimum possible ranges from point to receiver when this solution is adopted?

7.2.2 Distribution of Receivers: Single point Source

Restricted Problem

A restricted version of the problem is tackled first. In this case, no account is taken of the number of system components (e.g. point source transmitters) required to be mounted on the robot for calculation of position and orientation. The geometry with only one such component is considered. Furthermore, this component is constrained to make only rotations about one point.

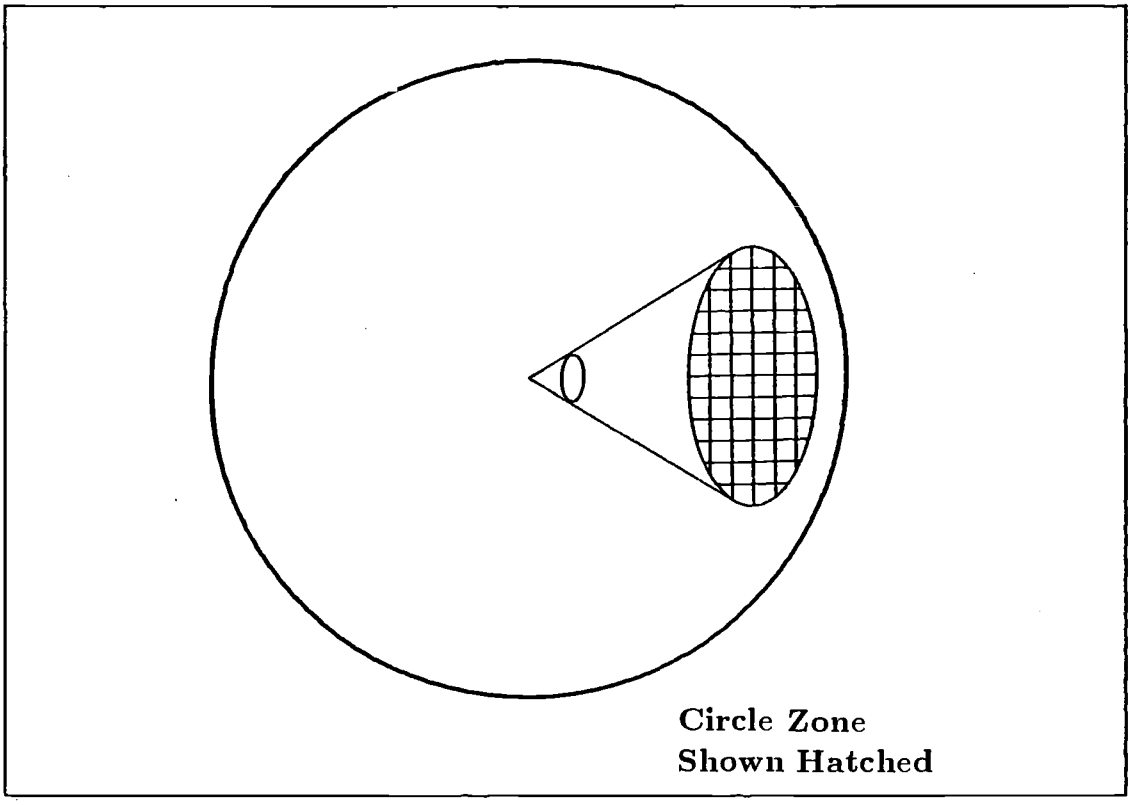


Figure 7.1: Intersection of Cone with Sphere.

Geometrically this system consists of a cone whose vertex is fixed at a point, but which can rotate about any axis through this point. The surface of the cone intersects a sphere, centred on this point, in a circle, see Figure 7.1. The receiving component of the system can only provide information about the moving element if it is within this circle zone on the sphere. The problem in this restricted case is therefore to establish, for a cone with given *solid angle* at its vertex, the minimum number and distribution of receivers required to ensure that at least one of them is "illuminated" by this cone in any orientation. The radius of the sphere is quite immaterial in this case.

Interpretation

As in the above, one might think of the situation as a number of fixed points on the surface of the sphere, representing the positions of the receivers, and a circle moving on the surface enclosing at least one point at all times. Equally one might consider the moving cone's own axis of symmetry as defining a point (moving) on the sphere surface, and each receiver as serving a stationary circular area, see Figure 7.2. This second description allows some headway to be made in determining the number and position of receivers.

The circle zones centred on each receiver must intersect with neighbouring zones to ensure the entire surface is covered. In general, for any particular circle diameter there are an infinite number of ways of covering the sphere. (The amount of overlap can be considered as a measure of the inefficiency of a particular arrangement).

To take as a starting point the solid angle subtended by the circle and to derive the necessary number and position of the circles on the sphere surface is, in general, a difficult problem. Again, in general, the solution would give an irregular spacing of circles and therefore a variation in resolution or utilised range from one tracker to another, see Figure 7.3. Another practical aspect of such a solution may be taken into account. An irregular spacing of receivers requires an irregular (and therefore more complex and expensive) structure to support them.

Regular Solutions

There are a number of regular solutions to the problem, corresponding to particular values of solid angles. The reasoning behind these solutions is as follows.

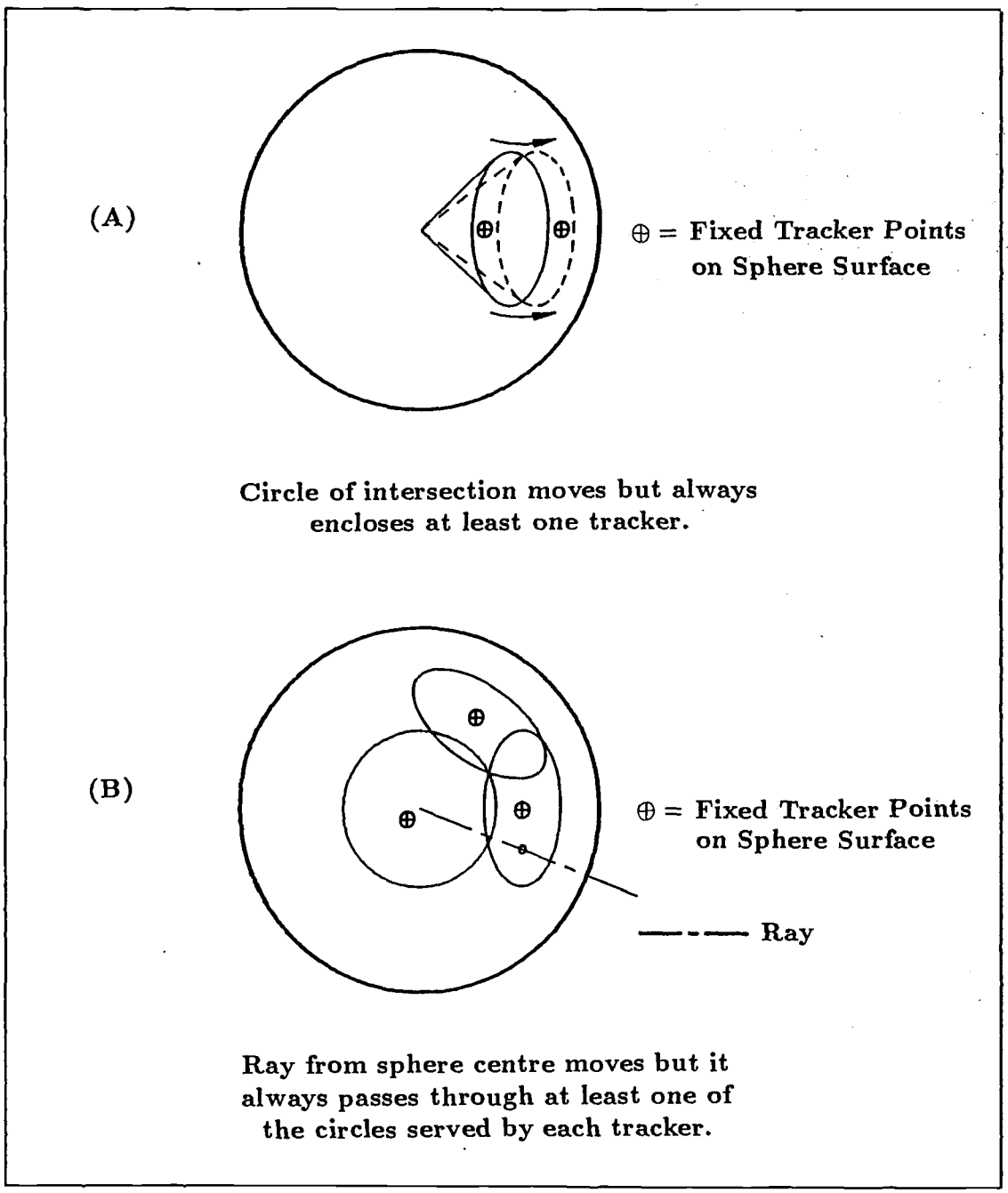
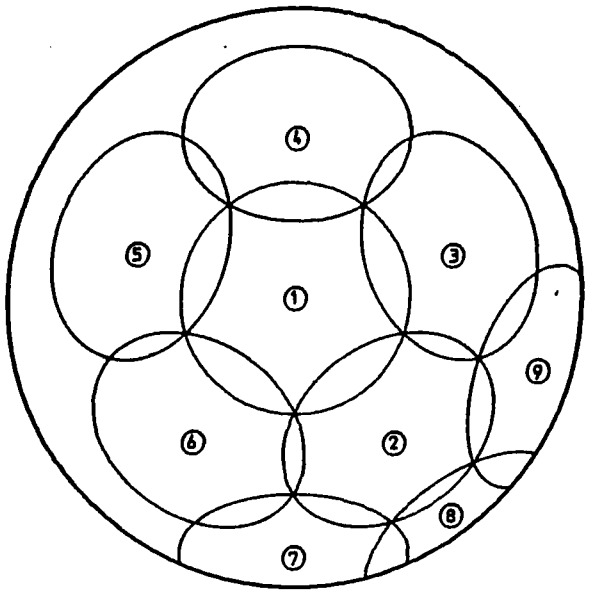


Figure 7.2: Equivalent Descriptions of the Geometry.



**Even when started in a regular manner,
in general the positions of the
circles soon become irregular.**

Figure 7.3: Irregular Spacing of Trackers.

A circle in a plane is surrounded by intersecting circles of the same size with no double overlapping. Joining in turn the points at which the first circle is cut with straight lines produces a regular polygon inscribed in the circle, see Figure 7.4. The analogous operation on the surface of a sphere produces regular spherical polygons. With the correct choice of ratio of polygon side to sphere radius the entire sphere surface can be covered with identical figures inscribed in identical circles. This is, of course, the projection of the edges of a regular polyhedron from the centre to the surface of a sphere.

There are five solutions of this type, corresponding to the five Platonic solids. It can easily be shown that there are no more than five solutions. (See for instance [Coxeter]). If p is the number of sides of a polygon then it is known as $\{p\}$. If a polyhedron has faces of type $\{p\}$, q of which meet at each vertex, it is a $\{p,q\}$. Now the internal angle of a polygon vertex is $[1-(2/p)]\pi$ and q of these are at a polyhedron vertex so,

$$\left(1-\frac{2}{p}\right)\pi q < 2\pi \tag{7.1}$$

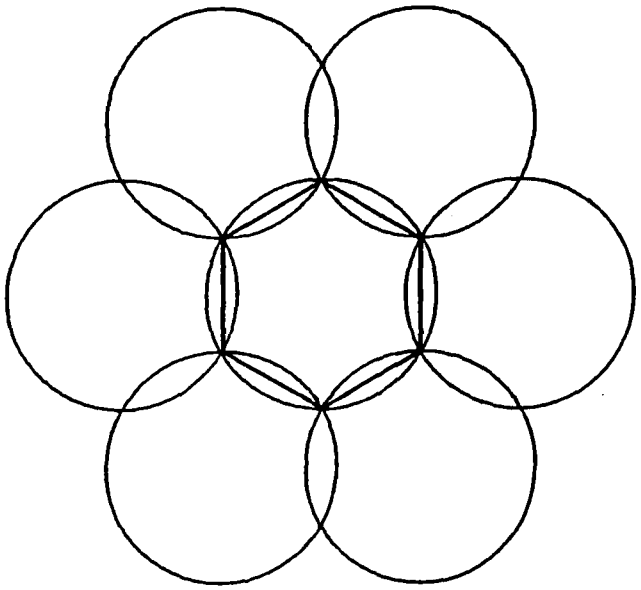
leading to the formula

$$(p-2)(q-2) < 4 \tag{7.2}$$

Since p and q are integers, Equation 7.2 can only be satisfied by {3,3}, {3,4}, {4,3}, {3,5} or {5,3} which are the tetrahedron, octahedron, cube, icosahedron and dodecahedron respectively, see Figure 7.5. (A test rig based on one of these solutions would have receivers sited at positions equivalent to the face centres of one of the above solids).

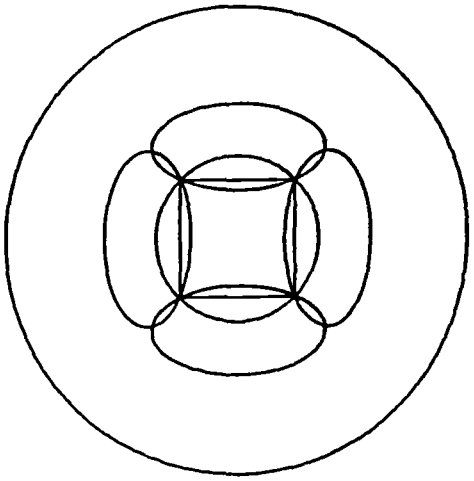
The next step in this process is to determine the solid angles subtended by

(A)



Intersecting Circles on a Plane - One Example

(B)



Intersecting Circles on a Sphere - One Example

Figure 7.4: Intersecting Circles without Double Overlap.

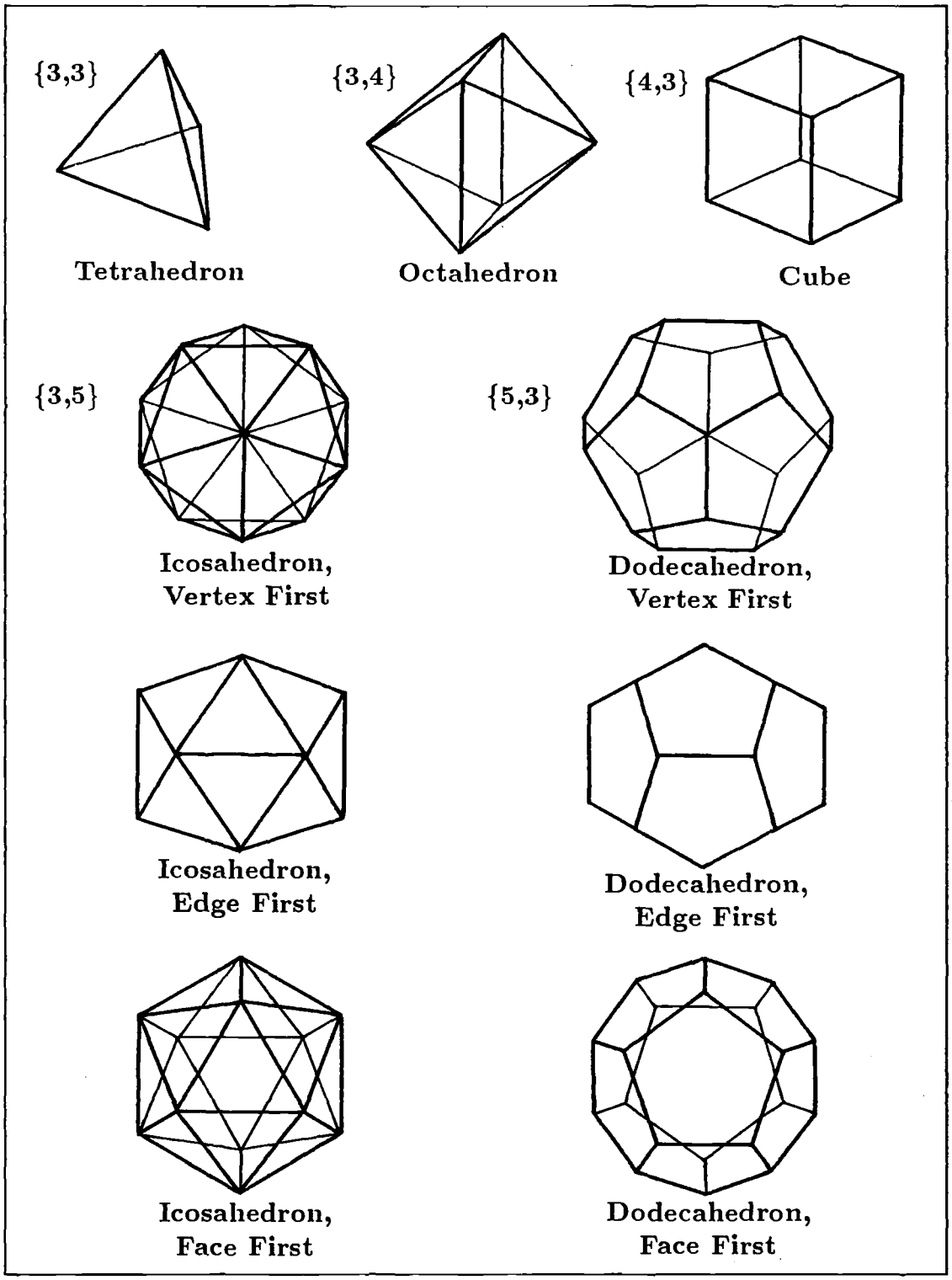


Figure 7.5: The Regular Polyhedra.

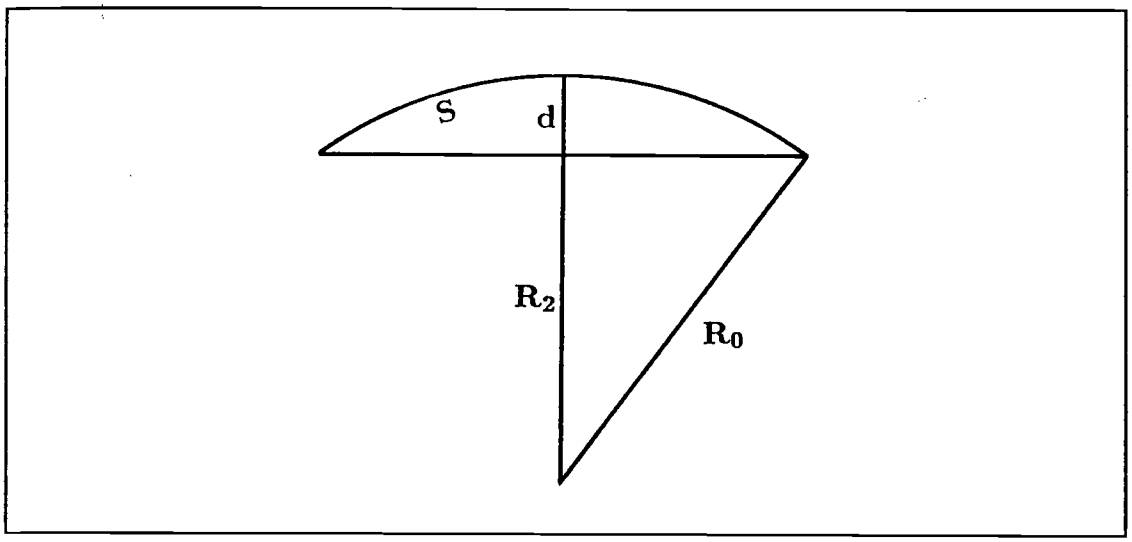


Figure 7.6: Tetrahedron Example.

the circumcircles of the polygonal faces of the regular polyhedra. A sample calculation is summarised below for the regular tetrahedron {3,3}. The results for the rest are given in Table 7.1.

Example 7.2.1 (Tetrahedron) *With an edge length of 2l, sphere circumradius* $R_0 = 1.2247l$ *and sphere inradius* R_2=*0.4082l.*[1] *The area of the circle zone is* $S = 2\pi R_0 d = 6.2830l^2$ *and the solid angle* $= \frac{S}{R_0^2} = 4.189$ *steradians.*

See Figure 7.6.

Non-Regular Solutions

Table 7.1 includes the data for three non-regular solids. Two of these are the quasi-regular polyhedra which have two types of polygonal face each, and are described as,

$$\left\{ \begin{array}{c} p \\ q \end{array} \right\}$$

[1]See [Coxeter].

Polyhedron Projection	No. of Receivers	No. of Zone Sizes	Smallest Zone (sterad.)	Small Cone Angle (deg.)	Largest Zone (sterad.)	Large Cone Angle (deg.)	Cover $> 4\pi$ (sterad.)
Tetrahedron	4	1	4.1890	141.06	4.1890	141.06	4.1890
Cube	6	1	2.6553	109.47	2.6553	109.47	3.3656
Octahedron	8	1	2.6553	109.47	2.6553	109.47	8.6776
Dodecahedron	12	1	1.2902	74.75	1.2902	74.75	2.9140
Icosahedron	20	1	1.2902	74.75	1.2902	74.75	13.2340
Cuboctahedron	14	2	1.1530	70.53	1.8403	90.00	7.6990
Icosidodecahedron	32	2	0.4138	41.81	0.9384	63.43	6.9695
Truncated icosahedron	32	2	0.3923	40.71	0.5268	47.26	2.6765

Table 7.1: Polyhedron Data

A polygon $\{p\}$ is surrounded by p polygons $\{q\}$ and vice versa. The distinct quasi-regular solids are,

$$\left\{ \begin{array}{c} 3 \\ 4 \end{array} \right\} \text{and} \left\{ \begin{array}{c} 3 \\ 5 \end{array} \right\}$$

(the cuboctahedron and icosidodecahedron respectively), see Figure 7.7. Some approximate figures are given in the table for the semi-regular truncated icosahedron (12 pentagons and 20 hexagons).

Configurations for the Restricted Case

If in reality the situation was as described at the beginning of this section, with one moving component which is allowed only rotation then, with the moving solid angle known, the configuration with the minimum number of receivers or trackers of sufficiently close and regular spacing may be selected from Table 7.1. For example, a component with a cone vertex angle of 74.75° or more may be accommodated by 12 trackers positioned on a sphere at the spherical projections of points at the centres of the faces of a dodecahedron. The circumdiameter of each face of this polyhedron subtends 74.75° at the sphere centre. A component with, say, a 70° vertex angle would require, on a complete sphere, 32 trackers

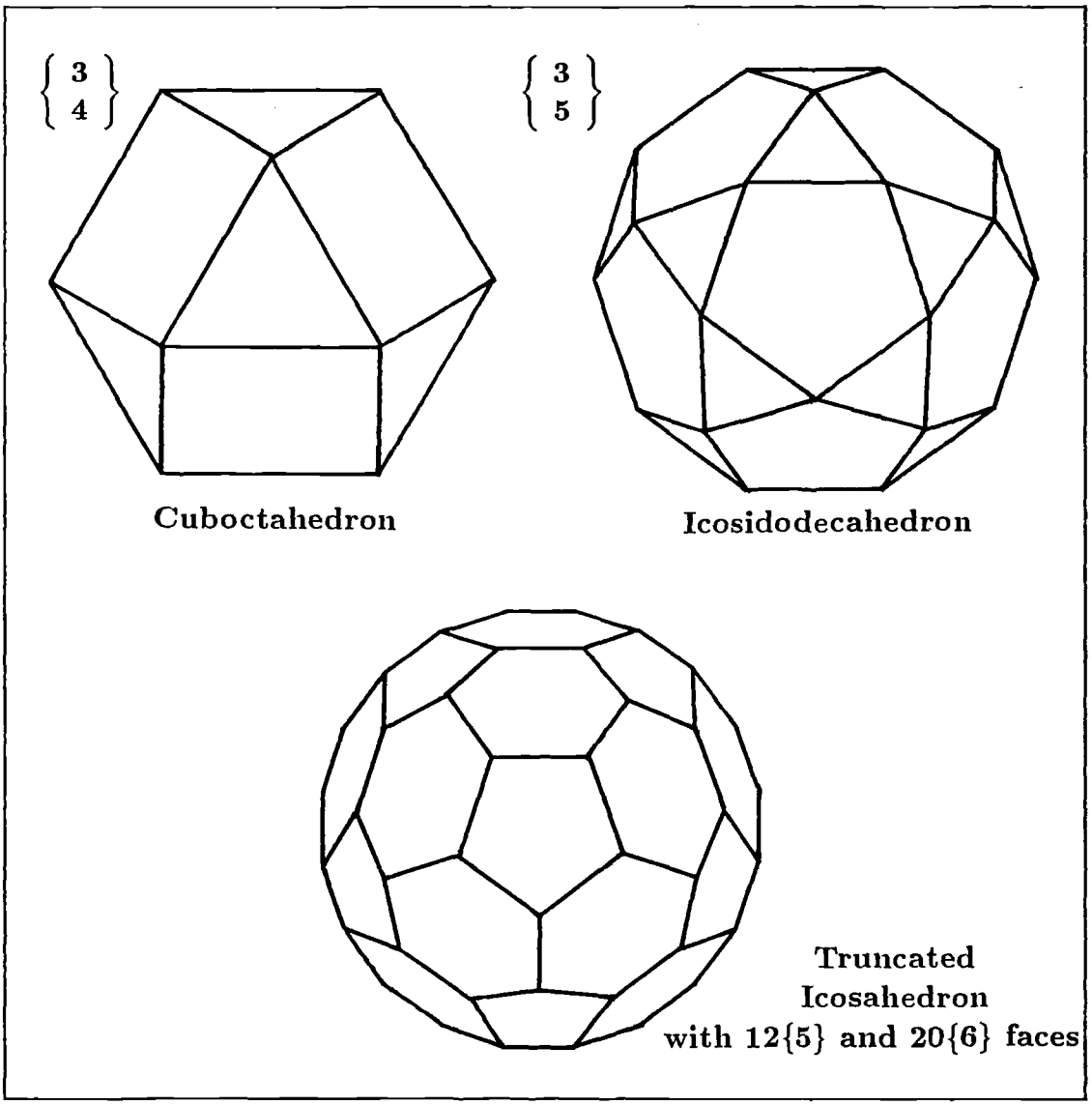


Figure 7.7: Quasi Regular and Semiregular Polyhedra.

spaced on the face centres of an icosidodecahedron or a truncated icosahedron, which have subtended angles of 63.43° and 47.26° respectively from their largest face types.

Relaxation of Constraints

The one moving component is now allowed rotations as before, plus *translations* within some radius of the centre point. The effect of this is shown in the section in Figure 7.8. With, for instance, a point source at the centre point, the radius of the sphere with receivers on its surface is unimportant, because the same solid angle is subtended at the sphere centre by the illuminated area. As the source approaches the spherical surface, however, the illuminated circle on the surface decreases in size, as does the solid angle subtended by this area at the sphere centre.

These two conical surfaces do intersect at a radius which increases with, in effect, the arm radius. For example, with a 90° transmitter at $0.5m$ from the sphere centre and receivers serving 74.75°, the receiver sphere radius has to be $2.665m$ minimum. A $1m$ radius of translation requires a $5.329m$ sphere radius and a $1.5m$ radius of translation requires a $7.994m$ sphere radius. The effect of allowing translation of the single moving point is either to make the test station very large or to increase the number of receivers required.

This problem disappears when multiple point sources are used, provided they are correctly positioned on the moving object.

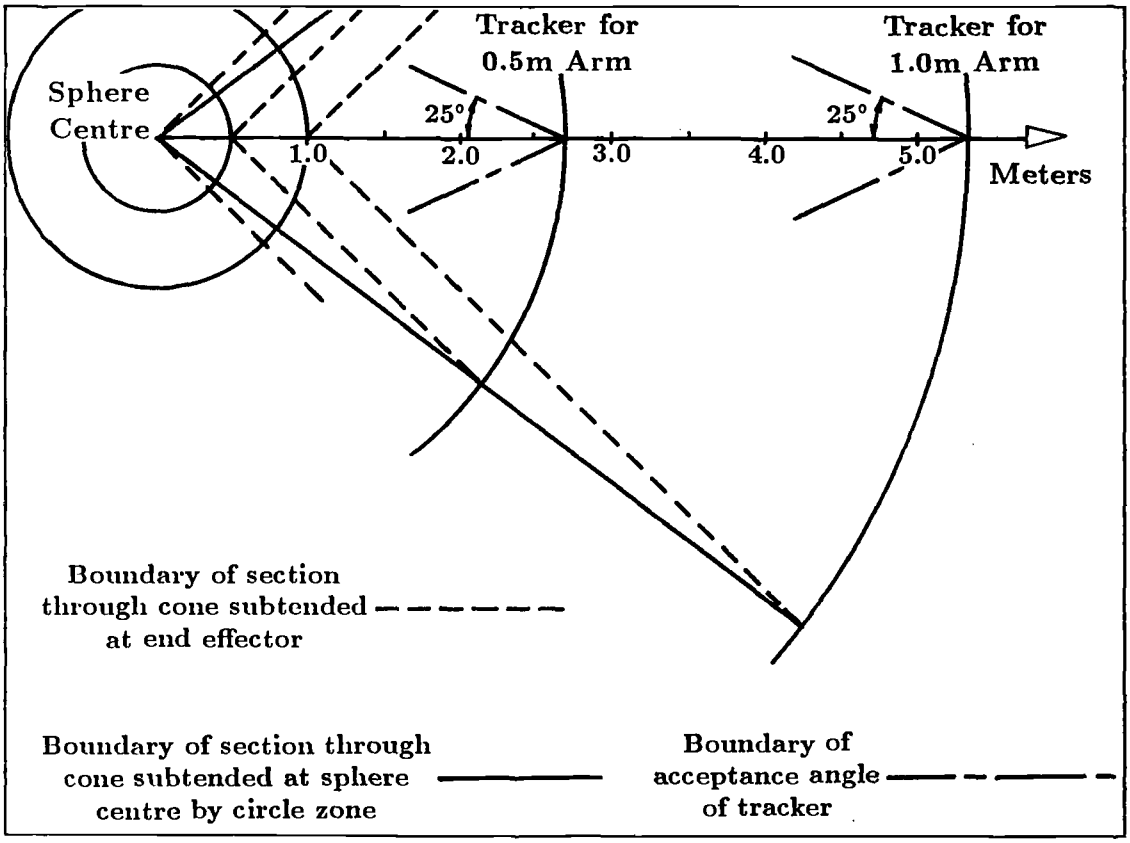


Figure 7.8: Tracker Position for 0.5 and 1.0m Arms.

7.2.3 Distribution of Receivers: Multiple Point Sources

Minimising System Size

Up to this point, a cone of illumination from a single point source has been considered. Initially this cone had its vertex at the sphere centre, and then it was allowed translation towards the sphere surface so that the circle of intersection at that surface reduced in size. This corresponds to the robot arm extended to the edge of its working volume and pointing the source out towards the sphere surface. The circle of intersection then has minimum radius.

In this situation, a second cone on the same axis and in the opposite direction intersects a maximum area on the sphere, also a circle. Between these extremes the intersected area increases from minimum to maximum, as the cone is rotated about its vertex. In intermediate positions, the boundary of this area is a skew, closed figure. This can be seen in the example in Figure 7.9. The important point to note is that at no time in this transition from minimum to maximum does any dimension across the area become smaller than the diameter of the minimum circle. In practice this means that if receivers are sufficiently closely spaced to cope with the size of the minimum circle, then (geometrically) they can track the cone vertex in all positions and orientations inside the defined working volume.

- Furthermore, if for every necessary "cone" there is another, facing the opposite direction, then the receivers may be spaced as though the cones were all centred on the sphere centre.

For every cone whose intersection area becomes too small for the receiver distribution, there is another which has an intersection area larger than necessary

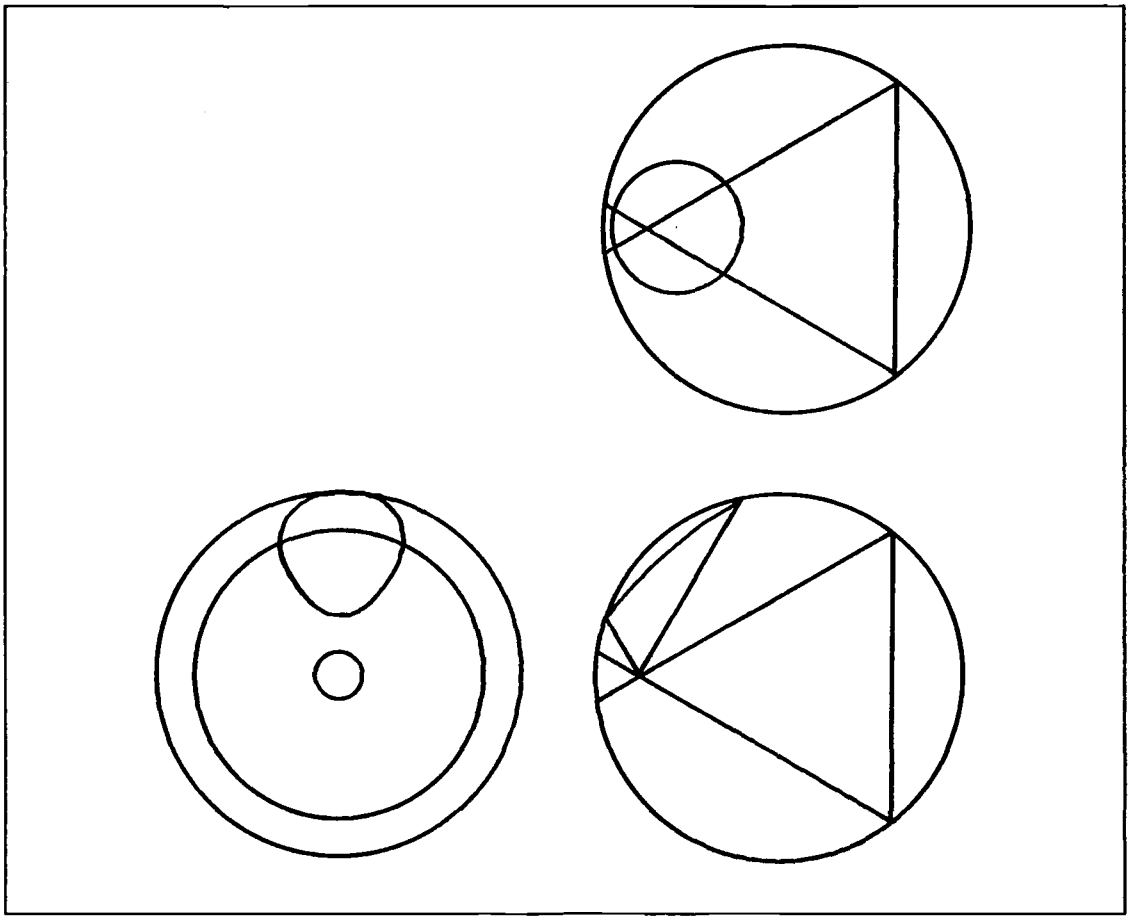


Figure 7.9: Cone-Sphere Intersection, Orthographic Projections.

as the body moves inside the sphere. Positioning the components in this way avoids the need for the very large test station suggested by the results in Figure 7.8. So by using a configuration of components on the moving body made up of pairs of opposite cones, a more compact system can be designed.

In the case of the robot in the example, the calculation of position and orientation requires that at least three separate points be used. If only three points are to be used, then at least two coordinates for each of these points must be found. To be sure that information is available for three points, three pairs of points must be used.

Space Frame Geometry

It has been shown that the number of receivers required in a system, to cover all possible positions and orientations of an object in a working volume, is dependent on the solid angle "illuminated" by each of the moving components. If this is 74.75°, then on a regular structure, 12 receivers are sufficient. But even if the moving component turns out to have a useful angle of rather less than 74.75°, 12 trackers may be acceptable if a small sphere at the centre is acknowledged as being a no-go volume. For some applications, e.g. the performance measurement of industrial robots, this seems reasonable compromise to make to reduce the number of receivers, particularly since this volume is likely to be occupied by the first, stationary link of the robot anyway. It is assumed that this compromise can be made in the given example, and that the working volume is in fact a sphere, but with a smaller sphere at the centre unused.

A design which has been pursued is for a space frame capable of supporting receivers in 12 positions. These positions have the coordinates of the face centres of a dodecahedron. The dodecahedron is not used as a geometrical basis for the

structure however. Its reciprocal figure the icosahedron is used instead. The dodecahedron has 20 vertices, 30 edges and 12 faces. The icosahedron has 12 vertices, 30 edges and 20 faces. The 12 vertices of the icosahedron are distributed in the same way as the face centres of the dodecahedron and in the space frame make material available in the appropriate positions for attaching support for the receivers.

Hyperboloid Coordinates

It was shown in Chapter 3 that the position and orientation of a body can be calculated from knowledge that three or more distinct points of that body lie on particular curves or surfaces in space. In this example, if each receiver supplies only one coordinate of one point, then it would be necessary to have six different receivers illuminated by six different point sources at all times. The information provided by the system in that case, would be that each of the six points lay on a particular hyperboloid. No method of finding the position and orientation of the object from this data is given in this thesis. Alternatively, the interferometers could be placed in pairs at each location on the space frame. In each pair, one interferometer would have its slits oriented at right angles to the slits on the other, and would be positioned immediately beside it. While they are both illuminated from the same point source, they give two coordinates of that point, which can then be shown to lie on the intersection of two hyperboloids, i.e. on a curve in space. The information provided by the system in this case is that each of three points lies on a particular quadric surface intersection curve. A method of calculating the position and orientation of the body from this information is given in Chapter 3.

If only three points on the moving object are to be considered, then, to ensure

continuous monitoring of the object, three opposing pairs of point sources are required. In the application in this example these sources would then be attached to the robot end effector flange.

In practice, a problem can occur when one point source is masked from a receiver by some object in the working volume, for instance, part of the robot itself. This practical point suggests that more than the minimum three pairs of sources should be used, to try to ensure that the necessary information on three points is always available.

7.2.4 Receiver Space Frame

The *shape* of the supporting frame has been determined by the number of receivers required. This in turn was dependent on the working solid angle of the moving components. The effect of radial displacement of the body is countered if the components mounted on it are placed in opposing pairs.

The *size* of the frame must be such that it is outside the swept volume of the object, but another influence on size is the receiver object field. This has now become the limiting factor in minimising the size of the frame.

The receivers should not be positioned too close to the object (see Figure 7.10). In Figure 7.10a the receiver is at the surface of a spherical working volume. Due to the limited object field of the receiver, much of the working volume cannot be seen from this position. In Figure 7.10b the receiver is situated so that the entire volume is within the receiver object field. The minimum radius to the receiver is dependent on this object field angle. An object field of 50° and the $2m$ diameter robot working volume are used to establish the dimensions of the space frame in Figure 7.11. In this case the distance from the centre of the working volume to each vertex of the icosahedron is found to be $R_c = 2.3662m$

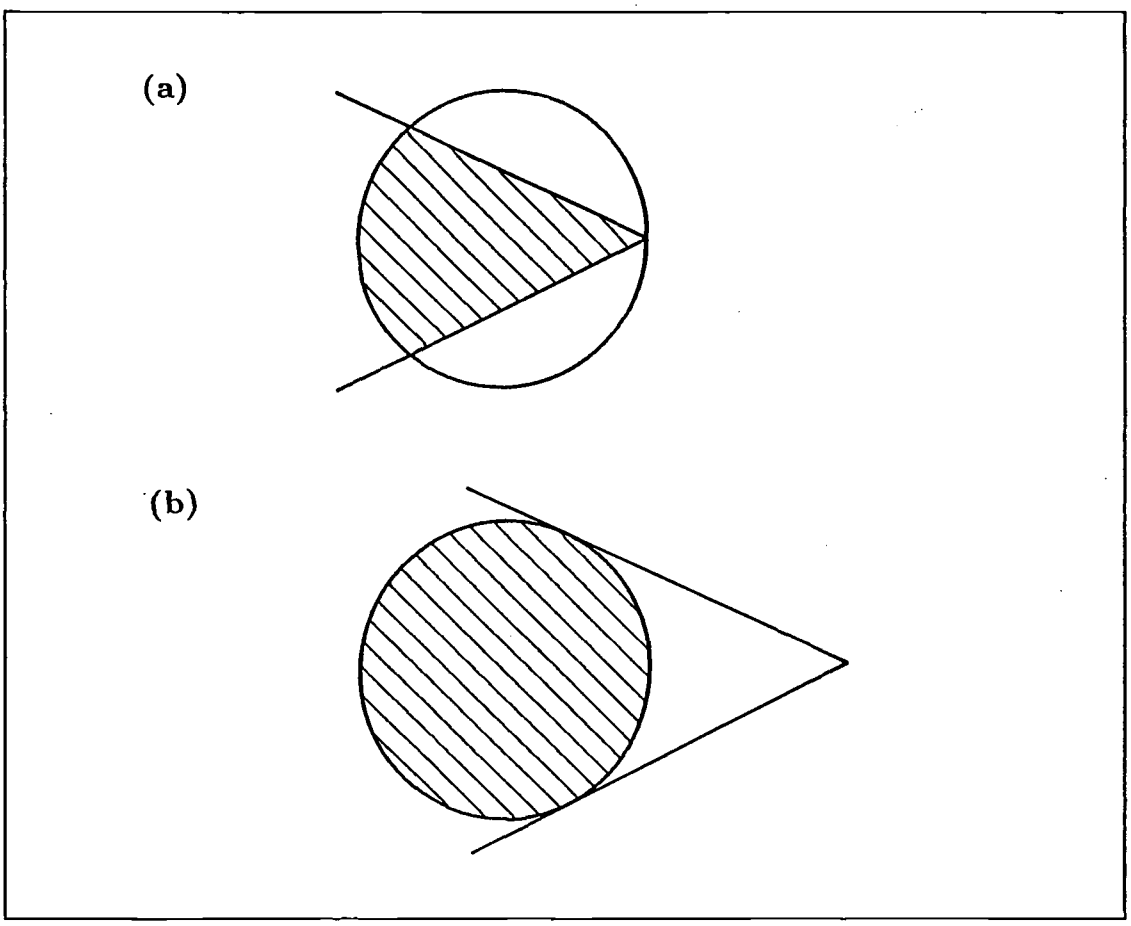


Figure 7.10: Working Volume and Tracker Position.

and since the working volume has a $1m$ radius, the maximum and minimum distances from points inside the working volume to a vertex are $R_m = 3.3662m$ and $R_n = 1.3662m$ respectively. The layout in Figure 7.11 gives the reference dimensions (in meters) for the icosahedral space frame. Installations for other sizes of arm would have dimensions scaled appropriately.

7.3 Interferometer: Design Parameters

This section deals with Item 2 of the requirement in the introduction to this chapter. It is concerned with the resolution of measurement within the work-

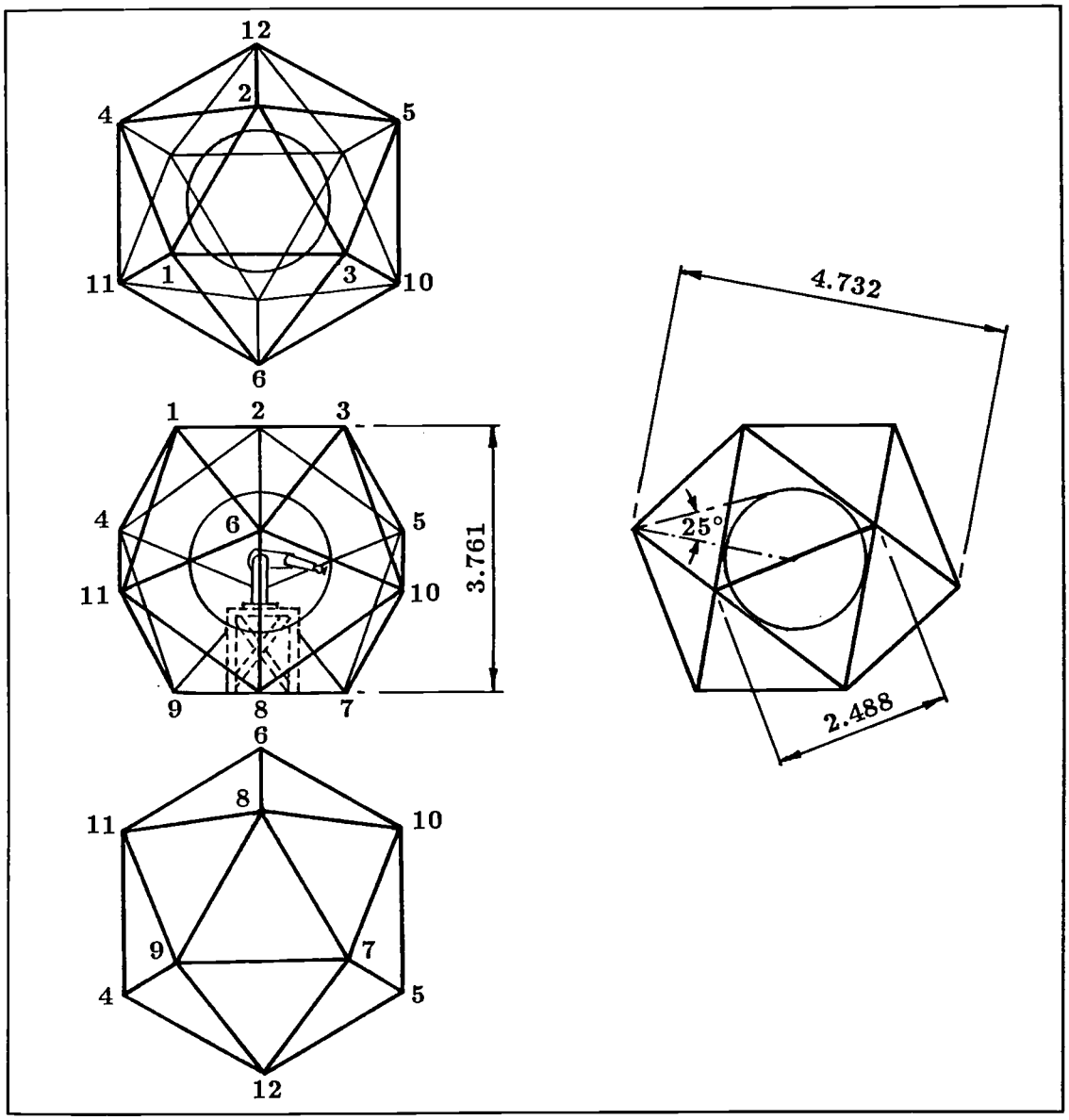


Figure 7.11: Spaceframe Installation.

ing volume and the design of the interferometer itself to meet the performance requirements.

7.3.1 Irradiance in Transmission Solid Angle

The point sources must be small and lightweight so as to provide a negligible addition to the robot end effector inertia. They could be provided by using laser diodes at the end effector or by transmitting the power from a gas laser to the end effector using monomode optical fibres. In either case some ancillary optics may be required, and the nature of these optical components affects the way the available power is spread through the cone of illumination.

The first thing to determine is the irradiance as a function of position within the transmission angle of the laser or laser-lens-spatial filter combination used. As has been seen in Section 7.2, wider transmission angles from each point source allow fewer interferometers to be used in covering all the possible motions of the robot. Alternatively, an increase in transmission angle can be seen as reducing the size of the "no-go" area at the centre of the working volume. It is important, therefore, to ensure a suitable spread of irradiance from the source.

To illustrate the design factors involved, in the following example a collimated beam is focussed to a point using a lens, and the irradiance in the conical beam diverging from this point is found as a function of range R, and angle θ from the cone axis.

In a collimated gaussian beam, the maximum irradiance I_0 is at the centre of the beam, and at a radius r from the centre line of the beam the irradiance is,

$$I(r) = I_0 e^{-\frac{2r^2}{\omega^2}} \tag{7.3}$$

where ω is the radius at which $I(r) = I(\omega) = \frac{I_0}{e^2}$. The power of the beam within

a radius r is,

$$P(r) = \frac{\pi\omega^2 I_0}{2}(1 - e^{-\frac{2r^2}{\omega^2}}) \tag{7.4}$$

The beam is focussed to a point by an ideal lens, of negligible thickness,[2] and with focal length f. The next step is to find expressions for irradiance and power as functions of range R and angle θ, downstream of the focal point.

An very narrow annular section of the collimated beam, with area δA receives approximately constant irradiance $I(r)$. So the power passing through this ring is,

$$\delta P = 2I(r)\pi r\delta r \tag{7.5}$$

As illustrated in Figure 7.12, this power δP is the contribution at θ to the optical axis, to the power passing through the point source. It comes from the projection of δA onto the sphere, radius R', which is centred on the focal point. This is the area $\delta\mathcal{A}$, the difference between the spherical caps at θ_1 and θ_2.

Now using the relationships,

$$\delta r = \frac{R'\delta\theta}{\cos\theta}, \tag{7.6}$$

$$R' = \frac{f}{\cos\theta}, \tag{7.7}$$

and

$$r = f\tan\theta, \tag{7.8}$$

this contribution to the power of the source can be seen to be,

$$\delta P = 2I(r)\pi f^2\frac{\tan\theta}{\cos^2\theta}\delta\theta \tag{7.9}$$

Since the function $I(r)$ is known (Equation 7.3), this gives,

$$\delta P = 2I_0 e^{-\frac{2f^2\tan^2\theta}{\omega^2}}\pi f^2\frac{\tan\theta}{\cos^2\theta}\delta\theta \tag{7.10}$$

[2]For the purpose of these calculations.

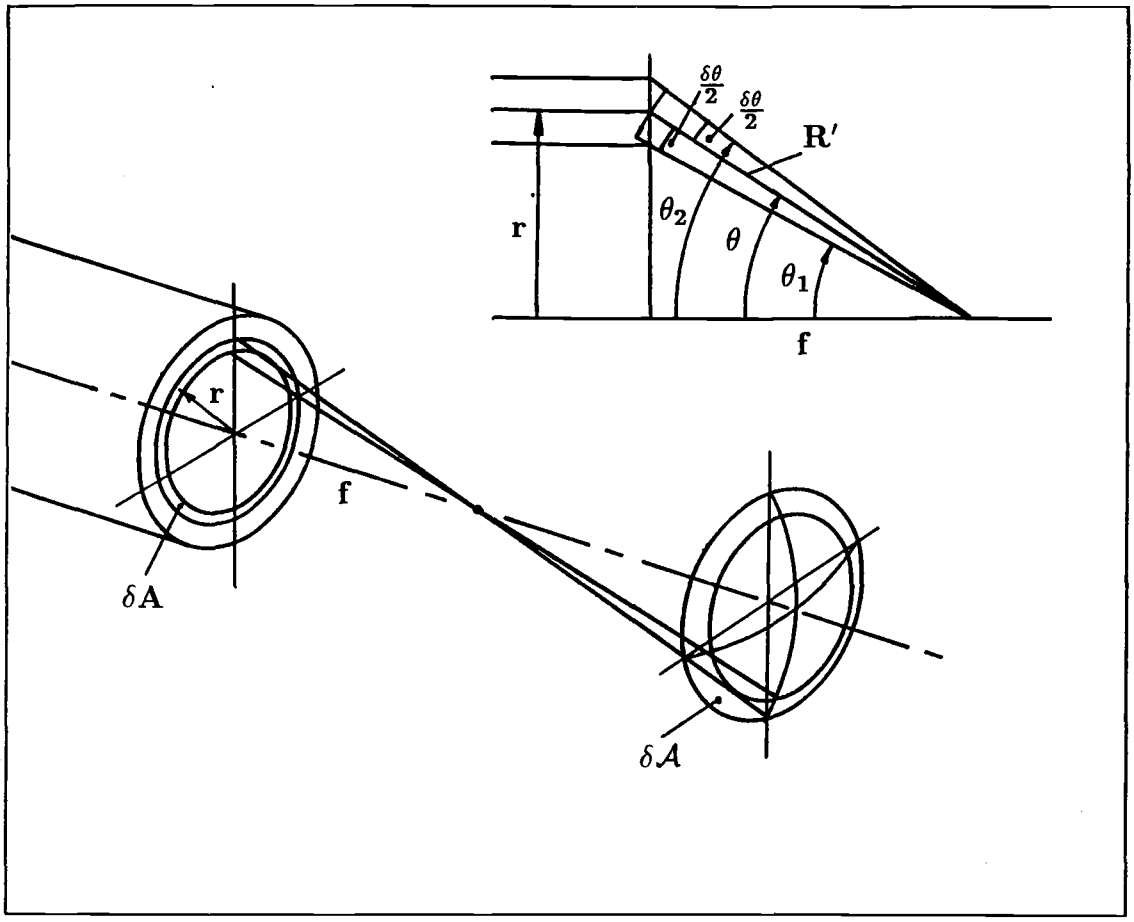


Figure 7.12: Irradiance in Diverging Beam.

Irradiance on a spherical surface at some range R and angle θ is,

$$I(R,\theta) = \frac{\delta P}{\delta \mathcal{A}} \tag{7.11}$$

The area of the spherical caps at θ_1 and θ_2 are $2\pi R h_1$ and $2\pi R h_2$ respectively, where $h = R - R\cos\theta$. So the difference in area is,

$$\delta \mathcal{A} = 2\pi R^2(\cos\theta_1 - \cos\theta_2) \tag{7.12}$$

$$\delta \mathcal{A} = 2\pi R^2(\cos(\theta - \frac{\delta\theta}{2}) - \cos(\theta + \frac{\delta\theta}{2})) \tag{7.13}$$

$$\delta \mathcal{A} = 2\pi R^2 \sin\theta \delta\theta \tag{7.14}$$

- Using Equations 7.10, 7.11 and 7.14 gives the equation for irradiance,

$$I(r,\theta) = \frac{I_0 f^2}{R^2 \cos^3\theta} e^{-\frac{2f^2 \tan^2\theta}{\omega^2}} \tag{7.15}$$

  which can then be used to find the irradiance on the slits of the interferometer, with the source at any point in the working volume.

7.3.2 Variation of Irradiance

As the point source moves inside the sphere, the irradiance at the receiver varies. It can be seen from Equation 7.15 that irradiance depends on both the range and the angular position of the receiver, within the cone of illumination. If it is required to limit this variation of irradiance at the receiver so that the minimum is no less than *one tenth* of the maximum, say, then this restriction can be used to establish some of the characteristics of the optical components of the point source. In Equation 7.15, irradiance at some position (R,θ) depends on R and θ, but also on f, the lens focal length and ω, the collimated beam radius at which $I(r) = \frac{I_0}{e^2}$. By fixing the ratio of irradiance at two points in the diverging beam, the relationship between f and ω can be found.

In this case the maximum range $R_m = 3.3662m$ and the maximum angle from the cone axis is approximately $\theta = 37.5°$ (assuming a cone vertex angle of 75° which is sufficient for this distribution of receivers). The minimum range is $R_n = 1.3662m$ and the minimum angle is $\theta = 0°$. So for a ratio of 1:10,

$$\frac{I(3.3662m, 37.5°)}{I(1.3662m, 0°)} = \frac{1}{10} \tag{7.16}$$

and using Equation 7.15,

$$\frac{R_n^2 \cos^3 0 e^{-\frac{2f^2 \tan^2 37.5°}{\omega^2}}}{R_m^2 \cos^3 37.5° e^{-\frac{2f^2 \tan^2 0°}{\omega^2}}} = \frac{1}{10} \tag{7.17}$$

$$0.3299 e^{-1.1776\frac{f^2}{\omega^2}} = \frac{1}{10} \tag{7.18}$$

$$-1.1776\frac{f^2}{\omega^2} = \log_e 0.3031 = -1.1936 \tag{7.19}$$

$$\frac{f}{\omega} \approx 1.0 \tag{7.20}$$

This determines the relationship between the collimated beam radius and the focal length of the lens. Thus for a lens with $f = 4mm$ say, the collimated beam radius $r = \omega$ at $I(\omega) = \frac{I_0}{e^2}$ is $4mm$ i.e. beam diameter is $8mm$.

A $4mm$ focal length lens which would accept all of the collimated beam up to this diameter would have to have an $f/\#$ of $f/0.5$. It may be more convenient to start with a lens with a particular $f/\#$, such as the Ealing Beck 24-8724 microscope objective, which is an $f/0.6$ with $f = 4mm$, and work through the calculation in the opposite direction to obtain the resulting ratio of minimum to maximum irradiance. This lens has a diameter of $6.67mm$ and so $\omega = 3.33mm$, and the irradiance ratio works out as approximately 1:16 in this case.

7.3.3 Irradiance and Laser Power

In the last section it was shown how the *variation* of irradiance incident on the receivers can be determined in the design process. In this section, the relation-

ship between the irradiance $I(R,\theta)$ at some point, and the *power* of the laser used to generate the point source is found.

In Equation 7.15 the only remaining unknown quantity is I_0, the maximum irradiance in the collimated beam. This can be determined from Equation 7.4. If the power of the collimated beam within the radius $r = \omega$ is $10mW$ say, then with $\omega = 3.33mm$,

$$10mW = \frac{\pi\omega^2 I_0}{2}(1 - e^{-2}) \tag{7.21}$$

$$I_0 = 663.9W/m^2 \tag{7.22}$$

Using this value for I_0 in Equation 7.15, together with the figures for f, ω, R, and θ, allows the maximum and minimum irradiance at the interferometer slits to be calculated.

$$I_{max} = 5.69 \times 10^{-3}W/m^2 \tag{7.23}$$

$$I_{min} = 3.43 \times 10^{-4}W/m^2 \tag{7.24}$$

7.3.4 Range and Power

The minimum irradiance from the point source, as determined in the last section is $3.43 \times 10^{-4}W/m^2$.

On the interferometer, using two slits $150\mu m$ wide by $10mm$ long gives an aperture area of $1.5 \times 10^{-6}m^2$ per slit, and so a minimum power of approximately $10^{-9}W$ entering the instrument from the two slits together. The maximum power, with the point source closest to the slits is $1.6 \times 10^{-8}W$.

Before going on to consider photodetection and other factors in the design of the instrument optics for this application, the relationship between slit separation and the instrument's sensitivity at various points in the working volume is explored.

7.3.5 Variation of Sensitivity and Resolution

The sensitivity of the instrument is not constant throughout the working volume. As discussed earlier, the coordinate output from the instrument identifies a member of the family of hyperboloids which have the two slit centres as their foci. As the point source moves, changing the difference in optical path lengths to the slits, the change in interferometer output indicates that it has moved to another hyperboloid. The instrument has its maximum sensitivity when the source movement is normal to the surface of each hyperboloid the point passes through. This movement is on the surface of an ellipsoid passing through the original point and cutting each member of the family of hyperboloids at right angles. Thus sensitivity must be considered using an orthogonal curvilinear coordinate system consisting of confocal hyperboloids and ellipsoids with foci at the two slit centres. The sensitivity is in fact the derivative of arc length along an ellipse, with respect to changing optical path length difference. The following calculation shows how sensitivity and resolution vary in the application example being considered in this chapter.

In the example, the working volume of the interferometer is a $1m$ radius sphere with centre at $2.3662m$ from the receiver, and lying on its optical axis. Taking the planar cross-section of this volume at the $y = 0$ plane, we wish to find the sensitivity of the instrument at various points on this cross-section. On this plane, the coordinate system consists of confocal hyperbolae and ellipses, as shown in Figure 7.13.

A suitable parameterisation for this system gives,

$$x = c \cosh u \cos v \tag{7.25}$$

$$z = c \sinh u \sin v \tag{7.26}$$

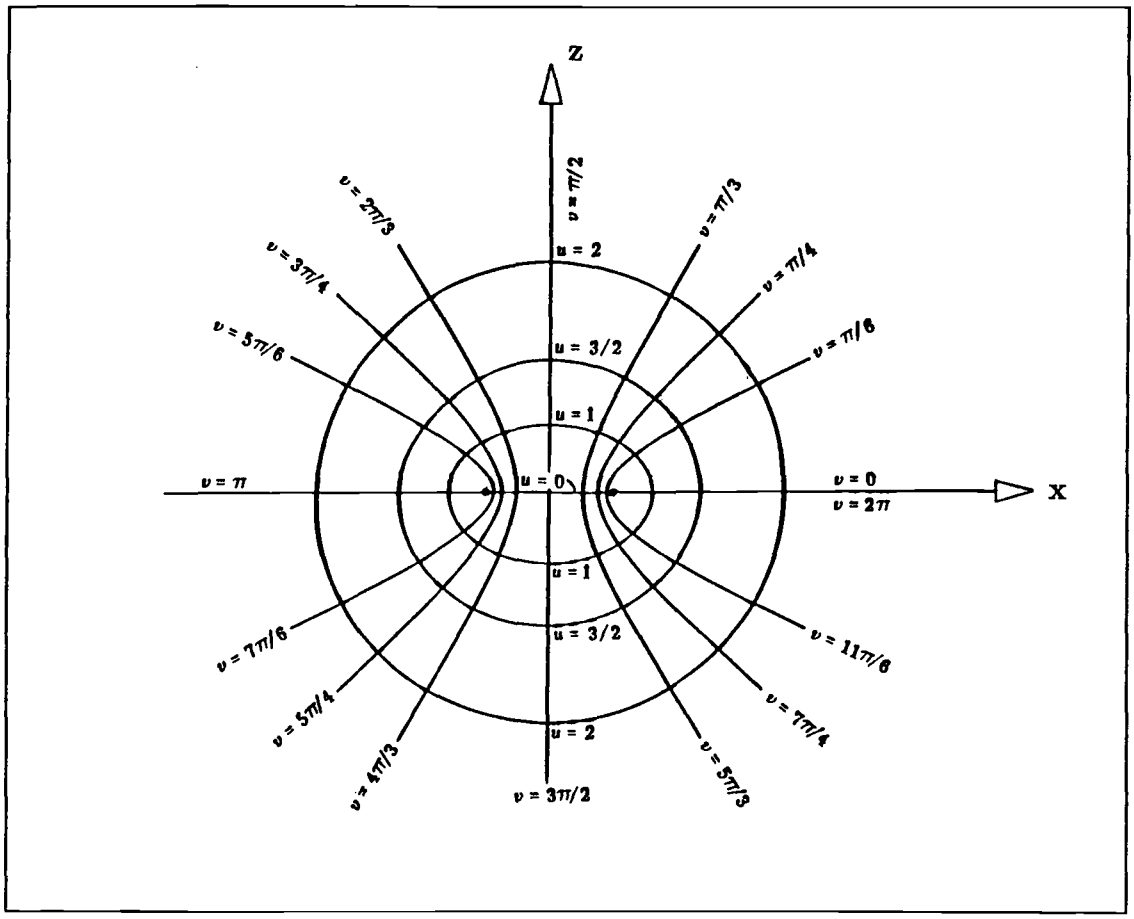


Figure 7.13: Confocal Hyperbolae and Ellipses.

(see [Spiegel]), where u and v are the parameters in the new coordinate system and c is half the distance between the foci.

A differential element of arc length dS of a curve on the plane can be obtained in this coordinate system from,

$$dS^2 = h_u^2 du^2 + h_v^2 dv^2, \tag{7.27}$$

where h_u and h_v are scaling factors (see [Spiegel]) and

$$h_u = h_v = c\sqrt{sinh^2 u + sin^2 v} \tag{7.28}$$

Since we are interested here only in movement *along* an ellipse, $du = 0$ and so,

$$dS = h_v dv \tag{7.29}$$

and,

$$\frac{dS}{dv} = c\sqrt{sinh^2 u + sin^2 v} \tag{7.30}$$

The rate of change of arc length along the ellipse with respect to optical path length difference m is,

$$\frac{dS}{dm} = \frac{dS}{dv}\frac{dv}{dm} \tag{7.31}$$

The second derivative on the right is obtained by expressing the path length difference in cartesian coordinates,

$$m = \sqrt{(x+c)^2 + z^2} - \sqrt{(x-c)^2 + z^2} \tag{7.32}$$

and using Equations 7.25 and 7.26 to convert this to the curvilinear coordinate system,

$$m = c\{\sqrt{(\cosh u \cos v + 1)^2 + \sinh^2 u \sin^2 v} - \sqrt{(\cosh u \cos v - 1)^2 + \sinh^2 u \sin^2 v}\} \tag{7.33}$$

which gives $m = 2c\cos v$. This is differentiated to give,

$$\frac{dm}{dv} = -2c\sin v \tag{7.34}$$

and so Equation 7.31 becomes,

$$\frac{dS}{dm} = \frac{\sqrt{\sinh^2 u + \sin^2 v}}{2\sin v} \tag{7.35}$$

This is the expression for sensitivity in terms of the coordinates u, v.

From Equations 7.25 and 7.26, $\sinh u$ and $\sin v$ can be found for any point (x, z). In fact,

$$\sinh^2 u = \frac{(x^2 + z^2 - c^2) + \sqrt{(x^2 + z^2 - c^2)^2 + 4c^2 z^2}}{2c^2} \tag{7.36}$$

and

$$\sin^2 v = \frac{z^2}{c^2 \sinh^2 u} \tag{7.37}$$

The sensitivity can be converted to resolution by multiplying by the smallest change in optical path length which can be measured at the instrument. The assumption in this example is that one tenth of a fringe cycle is easily discernible in the instrument's output. So the resolution can be found at any point (x, z) by calculating the sensitivity using Equations 7.36, 7.37 and 7.35, and multiplying by 63.28×10^{-9}, i.e. one tenth of the helium neon wavelength in this case. The results for several points in the $1m$ radius working volume are shown in Figure 7.14 in which a $10mm$ slit separation is assumed.

Resolution varies from $8.65\mu m$ to $21.3\mu m$ for the calculated points, and so only exceeds the $20\mu m$ design aim very slightly, at the farthest end of the volume. To improve resolution, either an increased slit separation or the ability to sense movement of a smaller fraction of a fringe cycle is required. For present purposes, it is assumed that the resolution calculated above is sufficient, i.e. that

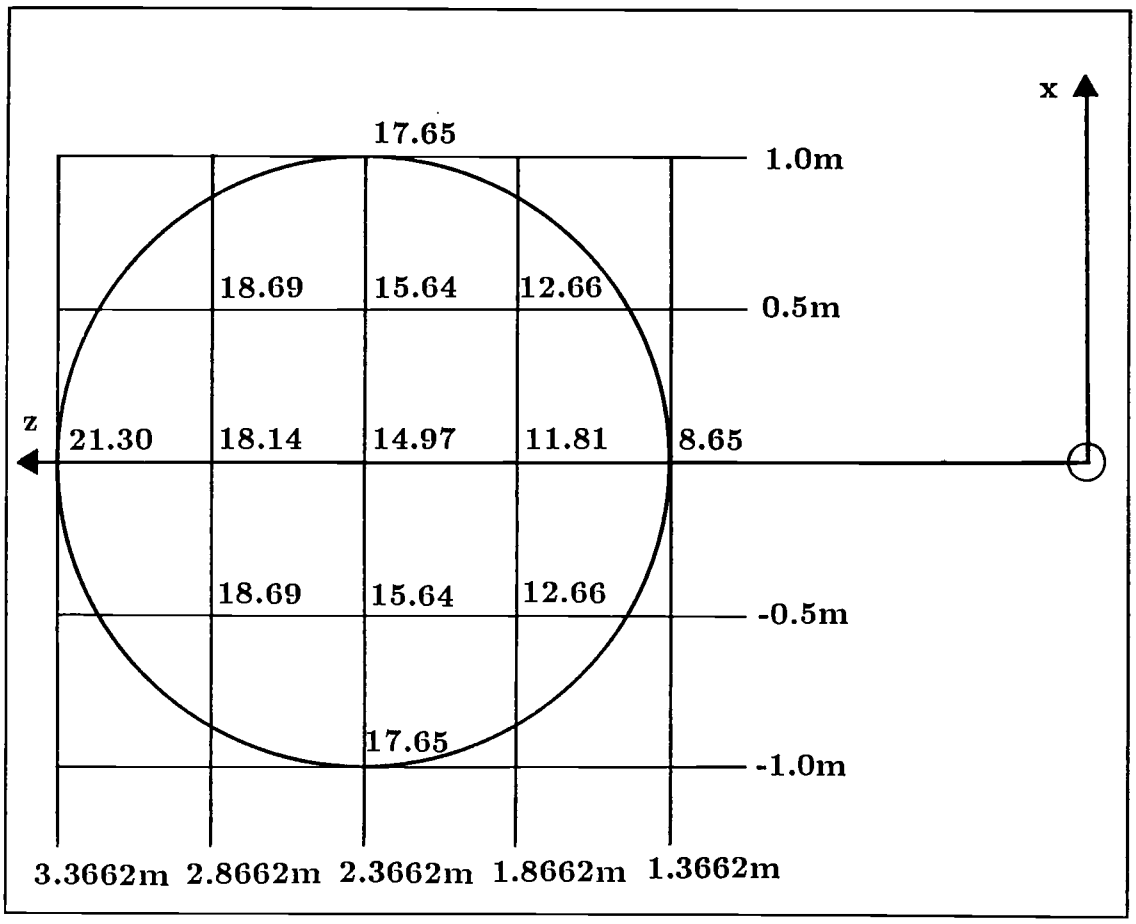


Figure 7.14: Resolution (in microns) in a Cross-Section of Working Volume.

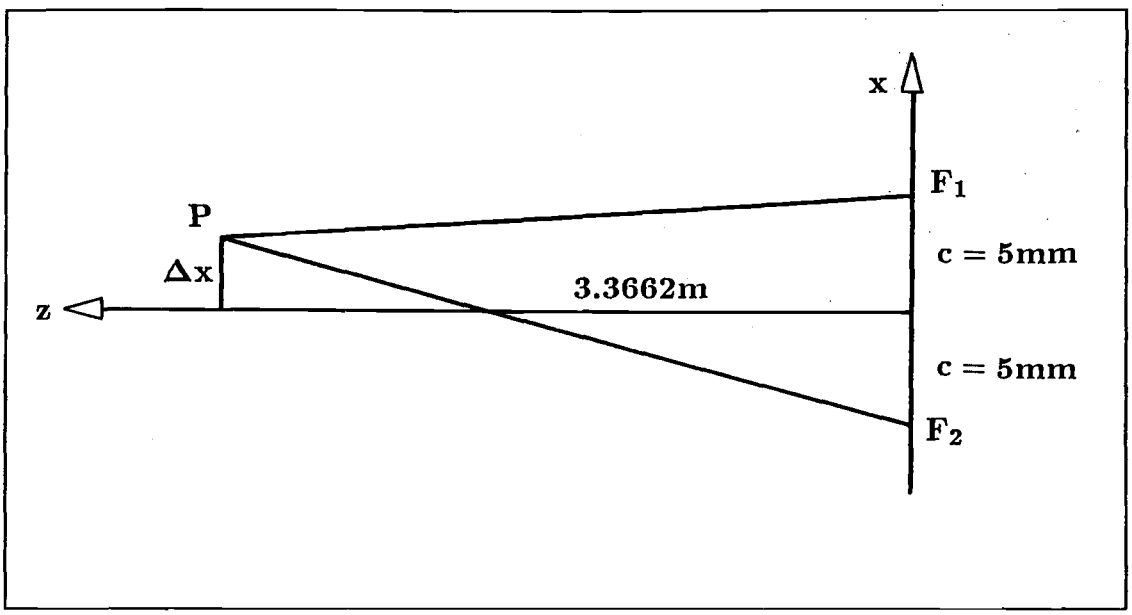


Figure 7.15: Resolution on the Optical Axis.

a slit separation of $10mm$ and the detection of one tenth of a fringe cycle are adequate.

Sensitivity and resolution can be found for any point in the working volume, using the above technique, since the hyperboloids and ellipsoids of the three dimensional coordinate system are solids of revolution about the x-axis. It is only necessary to rotate the plane of cross-section about the x-axis and establish the transformation to cartesian coordinates on the new plane.

For many applications it is useful to be able to check the worst case resolution without going through all the above analysis. This can be done easily for, say, the most distant point of interest on the optical axis (z-axis), as in the following example. In Figure 7.15 the point source at $z = 3.3662m$ moves a distance $\Delta x = 20\mu m$ in the x direction. The slit separation $2c = 10mm$.

From Equation 7.32, the optical path length difference is,

$$PF_1 - PF_2 = \sqrt{(\Delta x + c)^2 + z^2} - \sqrt{(c - \Delta x)^2 + z^2} \qquad (7.38)$$

which gives $59.4 \times 10^{-9}m$. Using light at $632.8nm$, this implies that a resolution of 10.65 parts per fringe cycle is required in the instrument to allow detection of that Δx.

7.3.6 Direction of Maximum Sensitivity

The *direction* of maximum sensitivity at a point can be obtained in several ways. Setting up the 2-D coordinate system described above, in a plane through the point and the x-axis, it can be found using the gradient (with respect to x) of the ellipse passing through the point. Alternatively, and with no reference to the 2-D coordinate system, the direction cosines of the normal to the hyperboloid surface at the point can be found as follows.

In homogeneous coordinates the equation of the hyperboloid is,

$$\mathbf{x}^{\mathrm{T}} A \mathbf{x} = 0 \tag{7.39}$$

where $\mathbf{x}$ is any point on that surface.

The equation of the tangent plane at the point $\mathbf{p}^{\mathrm{T}} = [p_1\, p_2\, p_3\, 1]$ is,

$$\mathbf{p}^{\mathrm{T}} A \mathbf{x} = 0 \tag{7.40}$$

where $\mathbf{p}$ is a point on the hyperboloid surface and $\mathbf{x}$ is any point on the tangent plane at that point. (See [Cohn] for the corresponding equation in 3-vectors). From this equation,

$$\mathbf{p}'^{\mathrm{T}} A \mathbf{x} + [0\,0\,0\,1] A \mathbf{x} = 0 \tag{7.41}$$

where $\mathbf{p}'^{\mathrm{T}} = [p_1\, p_2\, p_3\, 0]$, and since A is the discriminant matrix in its diagonal form,

$$\mathbf{p}'^{\mathrm{T}} A \mathbf{x} + 1 = 0 \tag{7.42}$$

Now putting the normal vector $\mathbf{p'}^{\mathbf{T}} A = \mathbf{q}^{\mathbf{T}}$, and dividing throughout by $\sqrt{\Sigma(q_i^2)}$, the unit normal to the hyperboloid surface at this point is,

$$\hat{\mathbf{n}}^{\mathbf{T}} = \frac{\mathbf{q}^{\mathbf{T}}}{\sqrt{\Sigma(\mathrm{q_i^2})}} \tag{7.43}$$

The coefficients of $\hat{\mathbf{n}}$ are the direction cosines of the normal vector and indicate the direction of maximum sensitivity at that point.

Information about the sensitivity of the instrument and the direction of maximum sensitivity is useful in verifying that the calculation of cartesian coordinates and Euler angles, (using the data from several interferometers in different locations), is well-conditioned.

7.3.7 Focal Length and Diffraction Grating Frequency

In the last section it was shown how the required resolution and the sensitivity of fringe movement detection determine the slit separation in the receiver. This information is now used in the next stage of the design process, which is to determine the focal length of the lens and the spatial frequency of the linear diffraction grating used in the detection of fringe movement.

The irradiance function which describes the interference fringe pattern on the focal plane is given by Equation 5.52, which is,

$$I = I'' + \frac{I'}{2} + \frac{I'}{2}\cos[k(\rho_{20} - \rho_{10}) + 2kAl_1] \tag{7.44}$$

In the argument of the modulating cosine function, $k = \frac{2\pi}{\lambda}$ is the propagation number, $(\rho_{20} - \rho_{10})$ is the difference in optical path lengths, A is half the slit separation ($A = 5mm$ in this example), and l_1 is the direction cosine of diffracted rays in the instrument. Referring to Figure 7.16, for a fixed path length difference, and considering the centre of the pattern, with $\theta_1 = 90°$, and $l_1 = 0$ the

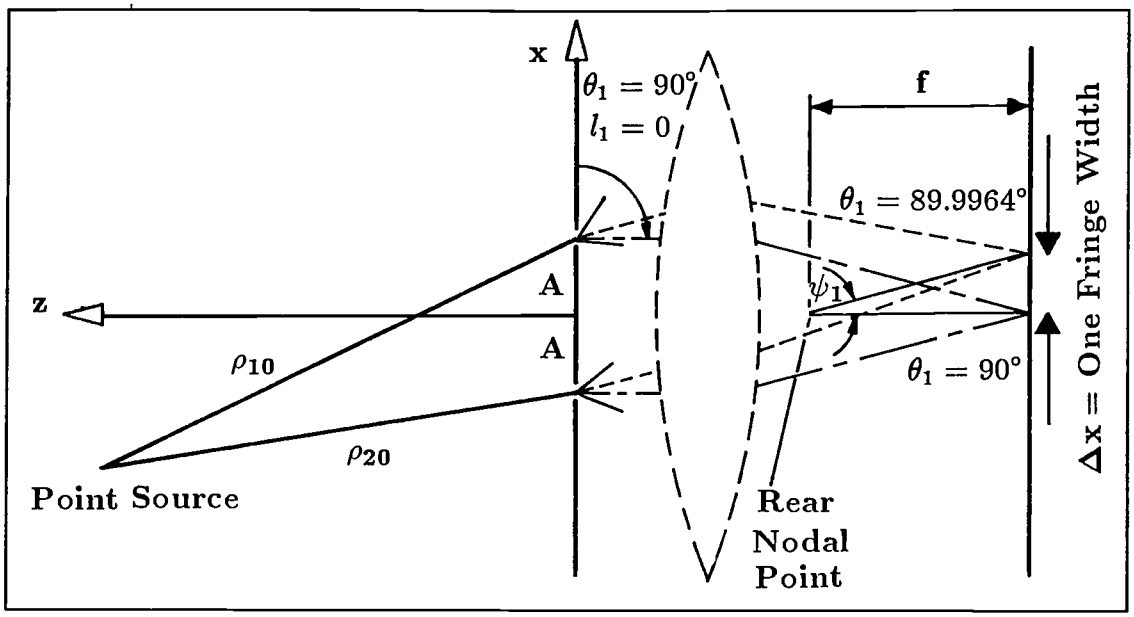


Figure 7.16: Focal Length and Diffraction Grating Frequency

irradiance has a value which depends only on say $(\rho_{20} - \rho_{10})$. It regains this same value when the second term in the argument is increased to 2π. So,

$$2kAl_1 = 2\pi \tag{7.45}$$

or,

$$\cos\theta_1 = l_1 = \frac{\lambda}{2A} \tag{7.46}$$

which gives $\theta_1 = 89.9964°$ and one complete fringe subtends an angle $\psi_1 = 90° - \theta_1$ at the rear nodal point of the lens. Thus with the fringe width Δx and lens effective focal length f,

$$\tan\psi_1 = \frac{\Delta x}{f} \tag{7.47}$$

In this example, $\tan\psi_1 = 633.0 \times 10^{-7}$. Either the lens focal length or the frequency of the diffraction grating can be specified at this stage and the other quantity calculated. The focal length of the lens will affect the overall size of the instrument, and so long focal length lenses are undesirable. A focal

length of $50mm$ is used in this example and, from Equation 7.47, results in $\Delta x = 3.164 \times 10^{-3} mm$, i.e. a grating frequency of approximately $316 cycles/mm$. Alternatively, taking $\Delta x = \frac{1}{250} mm$ as the starting point, the required lens focal length is $63.21mm$.

7.3.8 Slit Width and Minimum Range

To complete the geometric aspects of the design, it must be ascertained that the overlap of the diffraction patterns from each slit is sufficient to produce an adequate number of interference fringes. With incident light from a point source, the diffraction patterns from the two slits do not entirely overlap. As the source approaches close to the slits, the overlap reduces and ultimately leaves just the two separate diffraction patterns on the focal plane. The minimum range depends on the width of the diffraction patterns and on the minimum acceptable pattern overlap, i.e. the minimum number of fringes whose motion can be detected at this close range.

Greater overlap can be obtained with diffraction patterns which are spread over a larger angle, i.e. which have a larger central band. Diffraction pattern central band width increases as the slits become narrower. The diffraction pattern overlap varies with source position, as can be seen in Equation 5.48, but a more simple check on the overlap can be made using Equation 4.9, i.e. assuming plane waves at different angles are incident separately on each aperture.

With $m = m_1 = 0$, (rays parallel to the $y = 0$ plane) then from Equation 4.9,

$$I = I_0 \operatorname{sinc}^2[ka(l + l_1)] \tag{7.48}$$

where a = half slit width, k = propagation number, l is the x direction cosine of the incident rays, and l_1 is the x direction cosine of the observation point on the focal plane. See graph in Figure 7.17.

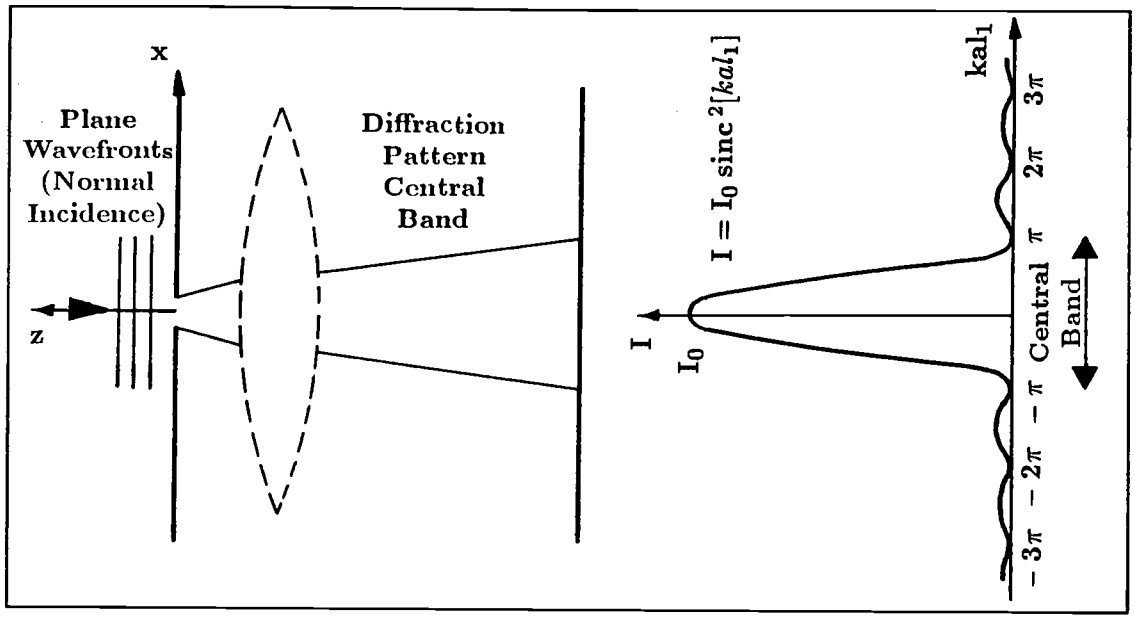


Figure 7.17: Diffraction Pattern Central Band Width

With normal incidence, $l = \cos 90° = 0$ and Equation 7.48 becomes,

$$I = I_0 \operatorname{sinc}^2[ka(l_1)] \tag{7.49}$$

which is zero when $ka(l_1) = \pm\pi, \pm 2\pi...$ The central band of the diffraction pattern lies between $ka(l_1) = \pm\pi$, and so the outside edge of the central band is at an angle θ_1 to the x-axis, where,

$$\frac{2\pi a}{\lambda} \cos\theta_1 = \pi \tag{7.50}$$

So with $\lambda = 632.8 \times 10^{-9} m$ and $2a = 150 \mu m$,

$$\theta_1 = 89.7583° \tag{7.51}$$

and half the central band subtends an angle ψ_b at the rear nodal point, where, $\psi_b = 90° - \theta_1 = 0.2417°$ in this case.

The width of the diffraction patterns can be increased by using narrower slits, but pursuing the above calculation and referring to Figure 7.18 it can be seen that the range for any particular amount of overlap can be found.

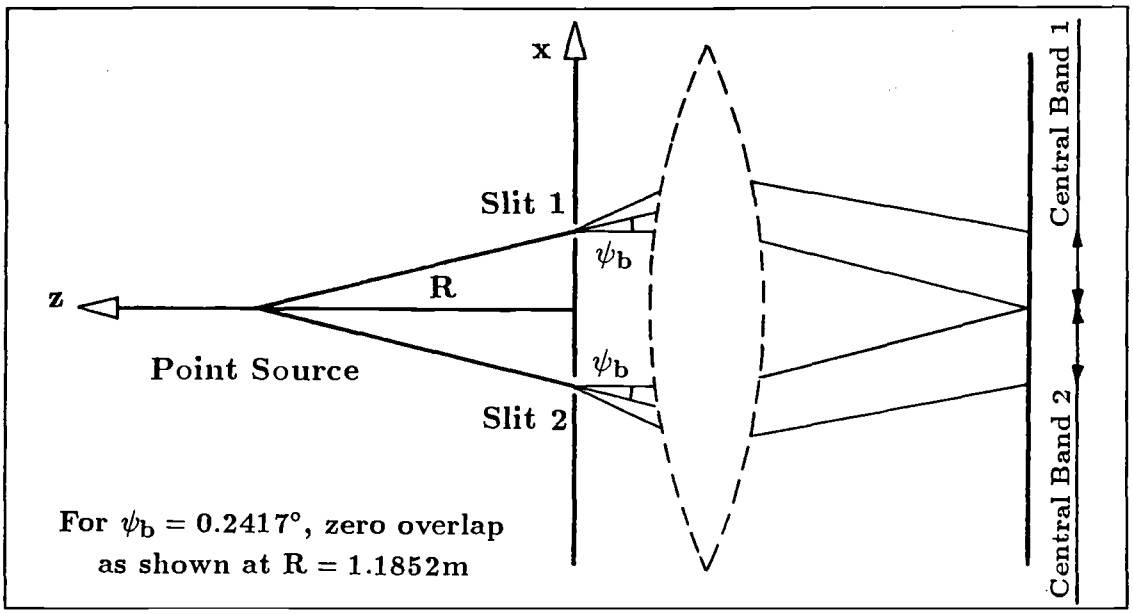


Figure 7.18: Zero Diffraction Pattern Overlap

With the central bands of the two patterns immediately side by side, i.e. zero overlap and with a slit separation of $10mm$ the range $R = 1.1852m$. In the example the minimum working range required is $R_n = 1.3662m$ and clearly at this range there will only be a small diffraction pattern overlap. To remedy this, the implementor may reduce the slit width (with correspondingly less power reaching the photodetector) or he may enforce his required overlap at this range by using some ancillary optical components, for example a wedge prism at each slit, to change the angle of the incident or diffracted rays. It is preferable simply to reduce the slit width, since this avoids the introduction of more components in the system, but in a design process compromises are often required and in this case the choice must be balanced against another factor, the sensitivity of the photodetector.

7.3.9 Photodetector

Photon Flux at Photodetector

In section 7.3.4 it was calculated that in this example, with slits $10mm$ long by $150\mu m$ wide, the power P entering the interferometer ranges from $10^{-9}W$ minimum to $1.6 \times 10^{-8}W$ maximum depending on source position in the working volume. We wish to determine whether a particular type of photodetector is suitable for monitoring this signal. As part of the example, this is now done for a photomultiplier.

From the incident power P in Watts, the photon flux at the diffraction grating can be calculated. The energy (in Joules) per photon is $E_p = h\nu$, where $h = 6.625 \times 10^{-34} Joule\ seconds$, is Planck's constant and ν is frequency in *Herz*. With a wavelength of $\lambda = 6.328 \times 10^{-7}m$ and the speed of light $c = 2.998 \times 10^8 m/s$, then the frequency $\nu = \frac{c}{\lambda} = 4.7 \times 10^{14} Hz$, and,

$$E_p = 3.1 \times 10^{-19} Joules \tag{7.52}$$

So in the case of minimum power entering the instrument, the photon flux F_p incident on the diffraction grating is,

$$F_p = \frac{P}{E_p} = 3.2 \times 10^9 photons/s \tag{7.53}$$

The flux transmitted through the grating depends on the transmissivity of the grating and the relative position of the fringe pattern on the grating lines. In this calculation the assumption is made that when the fringe pattern is in a position which gives maximum transmission through the grating, no light is reflected from or absorbed by the grating, and 100% of the flux is transmitted. Further it assumed that when the fringe pattern and grating pattern are 180° *out* of phase, there is zero flux transmitted. These would not be reasonable

assumptions if we wished to be very precise in predicting the power reaching the photomultiplier, but for the design purpose of establishing that this kind of sensor is broadly suitable, the assumptions can be justified.[3] Thus in this example the photon flux reaching the photomultiplier varies between zero and $3.2 \times 10^9 photons/s$, periodically with fringe position on the grating.

Quantum Efficiency

The quantum efficiency of the photomultiplier is now found for the wavelength of light used in this example. The output can then be calculated and compared with the dark current of the device. The particular photomultiplier considered here is the Mullard XP2233B which was used in the experiments described in Chapter 6. This photomultiplier has a Cathode Radiant Sensitivity $N_{kr} \approx 26mA/Watt$ at $632.8nm$.[4]

Now quantum efficiency QE is given by,

$$QE = \frac{Photoelectron\ Flux}{Photon\ Flux} \tag{7.54}$$

in which,

$$Photoelectron\ Flux = \frac{Photoelectron\ Current}{Electron\ Charge} = \frac{N_{kr}P}{C_e} \tag{7.55}$$

where the Electron Charge $C_e = 1.6 \times 10^{-19} Coulombs$. We have already that,

$$Photon\ Flux = \frac{P}{E_p} \tag{7.56}$$

and so,

$$)E = \frac{N_{kr}E_p}{C_e} \tag{7.57}$$

[3]Losses in lenses and other optical components are also ignored.
[4]From Mullard Technical Handbook, Book 2, Part 3, 1982.

Now since,

$$E_p = \frac{ch}{\lambda} \tag{7.58}$$

where h =Planck's constant and c = speed of light,

$$QE = \frac{N_{kr}ch}{C_e\lambda} = \frac{1.24\,N_{kr} \times 10^{-6}}{\lambda} \tag{7.59}$$

With N_{kr} in mA/W and λ' in nm,

$$QE = \frac{1.24\,N_{kr}}{\lambda'} \tag{7.60}$$

which in this case gives,

$$QE = 5.1\% \tag{7.61}$$

Photomultiplier Output

At the photomultiplier cathode,

$$Photoelectron\ Flux = QE \times Photon\ Flux \tag{7.62}$$

which in this case, with QE= 5.1% and $Photon\ Flux = 3.2 \times 10^9 photons/s$, gives,

$$Photoelectron\ Flux = 1.63 \times 10^8 photoelectrons/s \tag{7.63}$$

Now the cathode current is the product of the photoelectron flux with the charge per electron, and so,

$$Cathode\ Current = 2.6 \times 10^{-8} mA \tag{7.64}$$

This photomultiplier has a gain of 3×10^7, and since anode current is cathode current multiplied by the gain, we have,

$$Anode\ Current = 0.78mA \text{ maximum} \tag{7.65}$$

This is the anode current at the peak of each cycle, as sensed at the photodetector of the interferometer when the source position in the working volume produces minimum irradiance on the slits. It compares well with the dark current for this photomultiplier, which is $6 \times 10^{-5} mA$.[5] Thus noise from the dark current should not hinder the pulse count.

Velocity of Source and Fringe Pattern

The final aspect of this implementation is to consider the effects of source velocity on the signal at the photodetector. The worst case is when the power from the source is at its lowest and the source is moving at maximum speed in the direction of maximum sensitivity, i.e. normal to the hyperboloid surfaces. With the source crossing the optical axis at right angles and at maximum range from the interferometer, one complete cycle of output at the detector represents approximately $210 \mu m$ displacement of the source. If the source velocity is $1 m/s$ say, then one cycle of output occurs in $210 \mu s$, i.e. an output signal frequency of $4762 Hz$. In this worst condition, the photon flux incident on the grating is $3.2 \times 10^9 photons/s$ peak, and so, integrating over one cycle, at $1 m/s$ there are over 10^5 *photons/cycle*. Even with a source velocity as high as $10 m/s$ the photon flux per cycle is over $10^4 photons/cycle$, which suggests that there should be little difficulty in partitioning the output signal into tenths (or perhaps smaller fractions) of a cycle, thus obtaining the required resolution of $20 \mu m$ even at the longest range.

[5] From Mullard Technical Handbook, Book 2, Part 3, 1982.

7.4 Summary

This chapter has shown how the interferometer can be used in practice, to obtain dynamic position measurements. The starting point was the description of the task. The example used was an industrial robot with an end effector which can change position and orientation within a spherical working volume. Important aspects of the requirement were,

1. The need for position and orientation data.

2. The dimensions of the work envelope.

3. The resolution of position measurement.

Given this information, it is shown how a three dimensional dynamic measurement system, based on the interferometer, is designed to meet the requirement. The design process breaks down into two main sections. The first section determines the number and location of interferometers required to give the necessary coverage of the working volume. The second section is concerned with the design of the instrument itself.

In the first section, the solid angle of transmission of the moving point source is used to determine number and angular location of instruments necessary to give the required all-round view of the working volume. The dimensions of the working volume (a sphere in the example), and the object field angle of the interferometers are used to determine the range at which the instruments should operate. Altogether, this information is used to outline the design of a spaceframe suitable for supporting the interferometers at the required positions.

In the second section the amount of power reaching the instrument is established for various ranges within the working volume. The variation of sensitivity

of the instrument at different points in the working volume is examined for this example, and the slit separation required to give the necessary sensitivity (and so resolution) is determined. The method of finding the direction of maximum resolution at each point is also investigated. Characteristics of the optical components of the interferometer, for instance the lens focal length and the diffraction grating spatial frequency are found. Finally it is shown how to establish that the photodetector (in the example a photomultiplier) is suitable for the task.

One important aspect of any implementation of a system using lasers is safety. The diverging beams of laser light used with this interferometer are inherently safer than the collimated beams required by other systems such as the Surrey University, NBS, or Chesapeake instruments, or the various types of laser rangefinder discussed in Chapter 1. With a diverging beam, the proportion of the laser output power which can be incident on the eye decreases rapidly with increasing distance of the eye from the laser. This reduces the possibility of catastrophic damage to the retina. The safety of any implementation of the system does depend on the power of the lasers used and the restriction of access to the test area.

Chapter 8

Development and Applications

8.1 Development

The optical bench tests described in Chapter 6 have verified the theory given in Chapters 4 and 5 as a basis for an instrument. Useful information has been obtained concerning the design of particular components, for example the lens, and on how various parameters of the system may be modified to suit particular applications.

The object of further research is to establish the bounds of operation of the device, to optimise the performance of parts of the system and to investigate ways in which other improvements such as the use of solid state components might be incorporated. Another important aspect of a developed system is the signal processing of the output from the optics and the digitising of sampled data. The requirements for storage or further processing of this data do, of course, depend on the application in which the instrument is being used.

8.2 Optics

Starting from the transmitting part of the system and working through the receiver optics to the detection of light output, the following sub-sections describe

the further developments which seem appropriate.

8.2.1 Primary Source

In the experiments a 1 milliwatt helium-neon continuous wave gas laser was used as the coherent light source. Although fairly small gas lasers can be obtained, it would be desirable to use a laser diode instead if a suitable device was available. This is for two reasons. The primary source forms part of a source mounted on some object whose movements are being measured and in some applications the small size of a semiconductor laser may prove particularly convenient. Secondly, to overcome the problems of optical noise the source power might be amplitude modulated and the signal at the detector filtered electronically for the modulation frequency. A solid state laser is suitable for this purpose.

Although continuous wave laser diodes can be obtained, until recently they have not been suitable for interferometry because of their short coherence length and wide spread of wavelength. Two new developments could overcome this deficiency however, being essentially monochromatic and with long coherence length. These are the cleaved coupled cavity diode and, perhaps more promising, the distributed feedback diode. The stability of the wavelength of these diodes should be investigated. With the appropriate diode the complete source assembly could be made very small as shown in Figure 8.1.

8.2.2 Slit Screen

The air slits used for experiment were well made but required a through etching process which is quite critical. New large scale (and expensive) artwork is required for every design. In the development phase of this instrument, it would be economical to make a mask (as used in chip manufacture) containing a whole

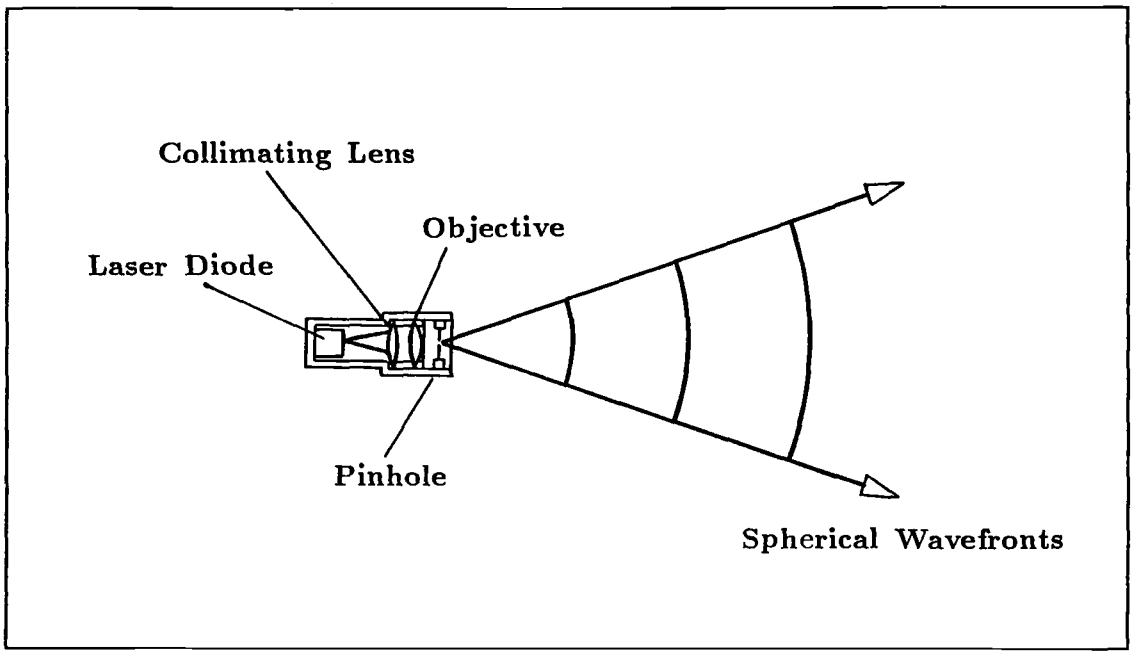


Figure 8.1: Design of Point Source Assembly.

range of slit designs for different applications. The photographic process then becomes simpler, there being no need for multiple reductions. The etching of a thin metal film on a substrate is also a more consistent process. Furthermore, at the time of writing, masks can be made to an accuracy of less than one micron.

Before taking this step however, it would be wise to test one example of "glass slits" to ensure that any problems of reflection from the other side of the thin glass sheet can be overcome. With the slits protected by a glass substrate, the front end of the receiver is far less susceptible to damage and is easily cleaned.

8.2.3 Lens Design

Large scale testing of lenses was not possible in the time available, but some representative camera objectives were used and their shortcomings noted. In any further tests of off-the-shelf lenses, one type which may be of interest in overcoming the vignetting problem is the retrofocus or inverted telephoto lens.

Examples of this type of lens can be seen in [Cox].

Another, and perhaps ultimately the most satisfactory approach is to have a lens designed and manufactured specifically for this job. Such a lens would incorporate a $\pm 5\%$ adjustment on its focal length to accommodate tolerances.

8.2.4 Diffraction Gratings and the Focal Surface

Once an appropriate lens is obtained, the limits of off-axis angle may be explored in more detail, as may the form of fringes on the focal surface as the pattern approaches the edges of this surface. Should geometric distortion adversely affect the integrity of the fringe pattern at the extremes of the focal surface, full use could not be made of the available object field of the lens. In that case, wider object field angles may be obtained by introducing some curvature to the image surface.

8.2.5 Photodetectors

Photomultipliers and a fairly cheap photodiode were used in the experiments and both were adequate for that purpose. The progression to a system with an amplitude modulated signal puts a further requirement on the frequency response of the photodetector, of whatever type. The choice of photodetector is therefore one of the first development tasks. The use of small components such as photodiodes allows the receiver to become quite compact. See the design given in Figure 8.2.

8.3 Electronics

The output of each photodetector in the optical system used in the experiments described in Chapter 6 is an analogue electrical signal. The signal is a sinusoidal

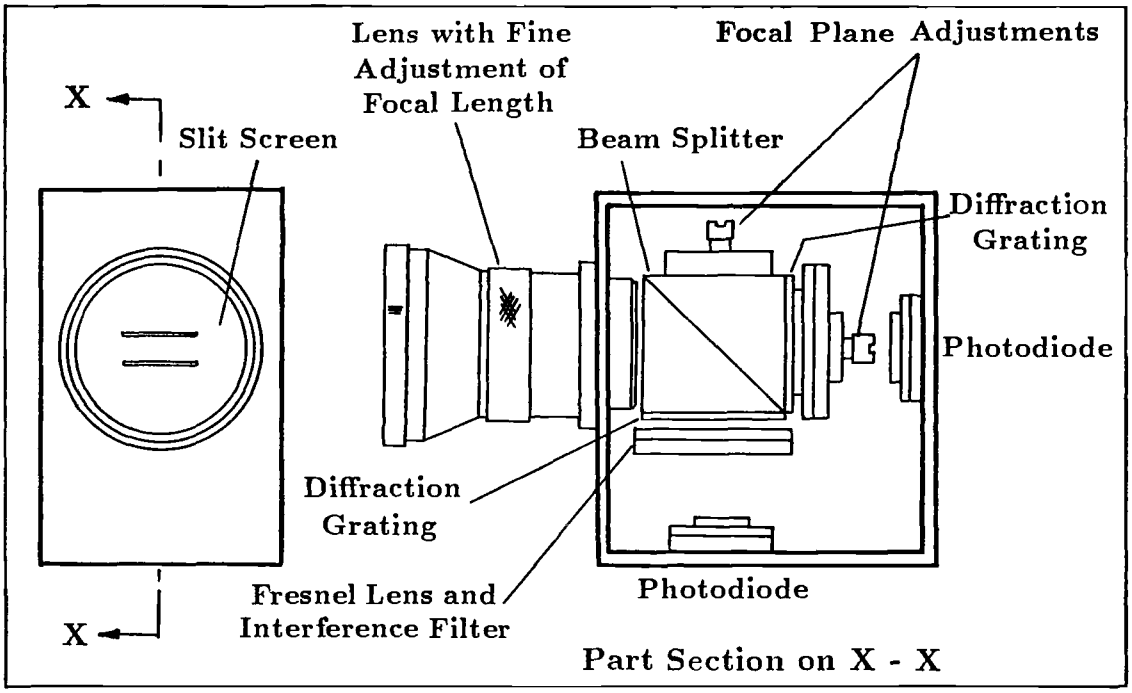


Figure 8.2: Design of Solid State Receiver.

modulation which varies in frequency quickly and in amplitude slowly and is superimposed on a slowly varying level.

The frequency of this basic signal depends on the rate of change of phase difference between the slits, which in turn depends on the position and velocity of the point source, and the distance between the slits. The rate of change of this frequency depends on the position and acceleration of the point source. These factors are different with different applications of the instrument.

For a given application there is a maximum frequency around which the electronics of the instrument are designed. The amplitude modulation frequency is at least an order of magnitude higher than this phase change frequency.

The signal conditioning electronics to be developed filters the signal, removes the varying DC level and produces a normalised, reconstructed sine wave which is then subdivided to give the highest resolution possible at any given frequency.

For examples of the resolution which may be obtained in this way, see Section 8.4.

8.4 Applications

The original motivation for the design of this interferometer came from a requirement to be able to make accurate dynamic measurements of the trajectories of industrial robots. Test results from such an instrument could be used to make comparisons between the performance of different machines or as part of the design loop in improving robot dynamics and control. Errors in particular trajectories could be measured and used to draw conclusions about vibration in the structure, for instance, or inadequacies in the machine's kinematics.

As part of a calibration procedure, accurate static measurements of the coordinates of points on the robot could be obtained and together with information from the joint position transducers on the machine be used to solve for the true kinematic parameters of the robot. This new model could then be placed in the controller and used to improve the machine's accuracy.

A second application is in inspection of manufactured parts. Coordinate measuring inspection machines are massive, very expensive and sensitive to distortion due to temperature variation and their own weight. At least two approaches to the replacement of this type of machine with the interferometer described are possible. In both cases, a striking benefit is that cost does not increase dramatically with working volume as it does with steel and granite coordinate inspection machines.

The first approach is simply to have a mobile point source which can be presented to the workpiece and touched against it at the points of interest. The source and its accompanying probe may be supported from a relatively cheap and compliant mobile structure whose function is merely to support the source

against gravity. The receivers would be mounted in a cluster of three at some convenient position outside the working volume.

Calibration by placing the probe in known positions would be sufficient to establish the true positions of the receivers (if they were not already fixed) and the starting position and zero position of the source. A continuous record of the coordinates of the point moving in the working volume is then available. At the required positions the transformation to cartesian coordinates may be carried out automatically. With appropriate allowance for the probe, measurements may be processed to give the dimensions of the object or checked against the geometrical tolerances of the object's surfaces, for instance.

The second approach to parts inspection introduces a further degree of automation. In this case the means of supporting the point source is servo controlled. This might be a simple servoed cartesian gantry or an industrial robot arm. The inaccuracy inherent in such inexpensive devices is of no consequence here because the point source would be mounted on a compliant fixture at, for instance, the robot's end effector. Thus even if the robot can only move in steps as big as 0.1mm, the error is taken up in the compliant mount and the point source accurately gives the position of the probe on the workpiece, to a few microns if necessary.

One basic motive for requiring a dynamic measurement system to be a non-contact one is to avoid contamination of the results by deflections caused by the forces applied to the workpiece. Another motive is the existence of long distances to the measuring point.

An example of the latter is in the measurement of the dynamic behaviour of large structures such as civil engineering structures under wind loading. This could be done simply by mounting a coherent source at the point of interest

and monitoring its movement with respect to receivers sited on the ground some distance away.

All the above are applications in which a record is being made of the co-ordinates of points either in a trajectory or in a few particular positions. As a position measuring device, however, the interferometer itself may be used as part of the feedback loop in a computer controlled servo system. One example of this type of application is in the fine control of angle between two components, perhaps in a machine or structure such as a telescope or an antenna. Another example of interest is in the control of dynamic machinery such as industrial robots, in which the accuracy of positioning of the end effector under load is increased by using the 3-D interferometer as an *external position transducer.* This overcomes position errors caused by the elastic deformation of the loaded arm. It also allows the tolerances on the machined parts of the robot to be relaxed and admits the possibility of using cheap materials such as plastics in the construction. Errors in the robot's own kinematics and in the model of the kinematics held in its controller would no longer have such importance and the wear of parts such as bearings would not affect accuracy.

For a particular application, the most appropriate values may be chosen for various parameters in the design of the instrument. The aspects of the application to be considered are the range of operation, the working volume, the accuracy required and the environment.

In the design of the instrument these requirements form the basis for decisions on wavelength, power, type of slit screen, slit separation and width, lens focal length and grating frequency. Examples of some design values are now given, with figures for the maximum resolution obtained by various subdivisions of the output sine wave. It can be seen that in the paraxial region in which these values

were calculated, the resolution is practically linear with both slit separation and range to the source. Off- axis, of course, this is not the case.

With 2.5mm slit separation and 1 meter range, 1 fringe cycle corresponds to 251.45 microns movement in the x direction, i.e. across the field of view. This was in fact the first example tested.

With 10mm slit separation and 2 meters range, 1 fringe cycle corresponds to 126.56 microns movement in the x direction. Subdivision of the output sine wave signal by 10 or 100 gives resolutions of 12.6 and 1.26 microns respectively.

The partitioning of sine waves is a technique used in the processing of the moire fringe output of the gratings of coordinate measurement machines. Subdivision to 1000 parts of a cycle has been used. That is preferable in that type of instrument to the alternative of using higher frequency gratings and counting only zero crossings, say. In any case, breaking up the fringe cycles into 10, 100 or even 1000 parts is a conventional way of obtaining good resolution in the sampling of a periodic analogue signal.

The examples of resolution given above are with laboratory work in mind. The theoretical maximum resolution with 20mm slit separation and a range of 1 kilometer is 31.64 millimeters per sine wave. Partitioning the sine waves by 100 gives a resolution of 0.3164mm and by 1000 gives a resolution of 31.64 microns. This begs questions about the effects of atmospheric disturbances over long ranges and would require a high power laser. However, any problems with air currents etc. which may be found would not affect the instrument's use at long ranges outside the Earth's atmosphere.

For this reason, space may well be a good environment for interferometry. The instrument can be used there in several applications which require kinematic data. One example is the provision of information about the deflection of

flexible structures under load, as in, for instance, the control of the orientation of satellites which carry large banks of photocells. This information is required for use in the control of motors if potentially destructive oscillations are not to be set up in the structure. In this case the laser sources would be mounted on the flexing parts of the structure and the receivers in the body of the satellite provide feedback to the attitude control system. The provision of this type of control allows the satellite to operate with a greater degree of autonomy. Similar considerations would apply to space stations, in which large masses are joined together by long, compliant links (should it be desired to manoeuvre such a structure).

A second example of an application in space is in precision navigation or positioning, again potentially with a high degree of automation. In this case the transmitting and receiving parts if the instrument are mounted on different vehicles whose relative position and/or orientation it is desired to control. It appears likely that there is a considerable degree of overlap in the design of control algorithms for satellites and for industrial robots employing externally obtained position data.

Appendix A

Diffraction

A.1 Diffraction at an Aperture

A.1.1 The Fresnel Approximation

For a general point (x, y) in the aperture the distance to the point P is r_1 and,

$$r_1 = \left[(x - x_1)^2 + (y - y_1)^2 + z_1^2\right]^{\frac{1}{2}} \tag{A.1}$$

and with r_{10} the distance from the origin in the aperture to the point P, i.e.,

$$r_{10}^2 = x_1^2 + y_1^2 + z_1^2 \tag{A.2}$$

$$r_1 = \left[(x^2 + y^2) - 2(xx_1 + yy_1) + r_{10}^2\right]^{\frac{1}{2}} \tag{A.3}$$

$$r_1 = r_{10}\left[\frac{(x^2 + y^2)}{r_{10}^2} - \frac{2(xx_1 + yy_1)}{r_{10}^2} + 1\right]^{\frac{1}{2}} \tag{A.4}$$

The binomial expansion gives,

$$\sqrt{1+p} = 1 + \frac{p}{2} - \frac{p^2}{8} + \cdots \tag{A.5}$$

and using only the first two terms of the series,

$$r_1 \approx r_{10} + \frac{(x^2+y^2)}{2r_{10}} - \frac{(xx_1+yy_1)}{r_{10}} \tag{A.6}$$

Substituting this expression for r_1 in Equation A.1 and omitting the $e^{-i\omega t}$ term since the time variation of the field is understood gives,

$$E = \frac{1}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{ik\left[r_{10} + \frac{(x^2+y^2)}{2r_{10}} - \frac{(xx_1+yy_1)}{r_{10}}\right]} dx\, dy \tag{A.7}$$

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{ik\left[\frac{(x^2+y^2)}{2r_{10}} - \frac{(xx_1+yy_1)}{r_{10}}\right]} dx\, dy \tag{A.8}$$

and with $l_1 = \frac{x_1}{r_{10}}$ and $m_1 = \frac{y_1}{r_{10}}$ the direction cosines of the point P,

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{ik\left[\frac{(x^2+y^2)}{2r_{10}} - (xl_1+ym_1)\right]} dx\, dy \tag{A.9}$$

A.1.2 The Fraunhofer Approximation

With the further assumption that $r_{10} \gg \frac{k(x^2+y^2)}{2}$, the Fraunhofer approximation is obtained.

$$e^{ik\frac{(x^2+y^2)}{2r_{10}}} \rightarrow e^0 = 1 \tag{A.10}$$

and,

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma} E(x,y) e^{-ik(xl_1+ym_1)} dx\, dy \tag{A.11}$$

A.2 Diffraction at a Rectangular Aperture: Incident Plane Waves

In this section the Fraunhofer diffraction integral, Equation A.11, is evaluated for a rectangular aperture with sides at $x = \pm a$, $y = \pm b$, on the plane $z = 0$. The

incident plane wavefronts are from a non-axial direction, the wavefront normals having direction cosines l, m, n, to the x, y, z axes. The field at the aperture is,

$$E(x,y) = E_0 e^{ikr} = E_0 e^{-ik(xl+ym)} \tag{A.12}$$

and Equation A.11 becomes,

$$E = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{-a}^{a}\int_{-b}^{b} E_0 e^{-ik(xl+ym)} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{A.13}$$

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \int_{-a}^{a} e^{-ik(xl+xl_1)}\, dx \int_{-b}^{b} e^{-ik(ym+ym_1)}\, dy \tag{A.14}$$

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \int_{-a}^{a} e^{-ik(l+l_1)x}\, dx \int_{-b}^{b} e^{-ik(m+m_1)y}\, dy \tag{A.15}$$

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \left[\frac{e^{-ik(l+l_1)x}}{-ik(l+l_1)}\right]_{-a}^{a} \left[\frac{e^{-ik(m+m_1)y}}{-ik(m+m_1)}\right]_{-b}^{b} \tag{A.16}$$

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 2a \left[\frac{e^{-ik(l+l_1)a} - e^{ik(l+l_1)a}}{-2ik(l+l_1)a}\right] 2b \left[\frac{e^{-ik(m+m_1)b} - e^{ik(m+m_1)b}}{-2ik(m+m_1)b}\right] \tag{A.17}$$

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 4ab \left[\frac{\sin k(l+l_1)a}{k(l+l_1)a}\right] \left[\frac{\sin k(m+m_1)b}{k(m+m_1)b}\right] \tag{A.18}$$

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 4ab\, \mathrm{sinc}\,[ka(l+l_1)]\, \mathrm{sinc}\,[kb(m+m_1)] \tag{A.19}$$

The irradiance I is the product of E and its complex conjugate E^*

$$I = EE^* \tag{A.20}$$

$$I = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 4ab\, \mathrm{sinc}\,[ka(l+l_1)]\, \mathrm{sinc}\,[kb(m+m_1)]$$

$$\times \frac{E_0 e^{-ikr_{10}}}{-i\lambda r_{10}} 4ab \operatorname{sinc}[ka(l+l_1)] \operatorname{sinc}[kb(m+m_1)] \tag{A.21}$$

$$I = \frac{E_0^2}{\lambda^2 r_{10}^2} 16a^2b^2 \operatorname{sinc}^2[ka(l+l_1)] \operatorname{sinc}^2[kb(m+m_1)] \tag{A.22}$$

A.3 Diffraction at 2 Rectangular Apertures: Plane Waves at Normal Incidence

This section gives in detail the derivation of the contributions E_1 and E_2 to the electric field E at a point in the far field. E_1 and E_2 are due to diffraction at apertures Σ_1 and Σ_2 respectively.

$$E = E_1 + E_2 \tag{A.23}$$

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma_1} E_{10}(x,y) e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{A.24}$$

$E_{10} = E_{20} = E_0 e^{-i\omega t}$, and ignoring time dependency,

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{-A-a}^{-A+a} \int_{-b}^{b} E_0 e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{A.25}$$

$$E_1 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \int_{-A-a}^{-A+a} e^{-ikxl_1}\, dx \int_{-b}^{b} e^{-ikym_1}\, dy \tag{A.26}$$

$$E_1 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \left[\frac{e^{-ikl_1 x}}{-ikl_1}\right]_{-A-a}^{-A+a} \left[\frac{e^{-ikm_1 y}}{-ikm_1}\right]_{-b}^{b} \tag{A.27}$$

$$E_1 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 2a \left[\frac{e^{-ikl_1(-A+a)} - e^{-ikl_1(-A-a)}}{-2ikl_1 a}\right] 2b \left[\frac{e^{-ikm_1 b} - e^{ikm_1 b}}{-2ikm_1 b}\right] \tag{A.28}$$

$$E_1 = \frac{E_0 e^{ikr_{10}} e^{ikl_1 A}}{i\lambda r_{10}} 2a \left[\frac{e^{ikl_1 a} - e^{-ikl_1 a}}{2ikl_1 a}\right] 2b \left[\frac{e^{ikm_1 b} - e^{-ikm_1 b}}{2ikm_1 b}\right] \tag{A.29}$$

$$E_1 = \frac{E_0 e^{ikr_{10}} e^{ikl_1 A}}{i\lambda r_{10}} 4ab \operatorname{sinc}[kl_1 a] \operatorname{sinc}[km_1 b] \tag{A.30}$$

Also,

$$E_2 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{A-a}^{A+a} \int_{-b}^{b} E_0 e^{-ik(xl_1 + ym_1)}\, dx\, dy \tag{A.31}$$

$$E_2 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \int_{A-a}^{A+a} e^{-ikxl_1}\, dx \int_{-b}^{b} e^{-ikym_1}\, dy \tag{A.32}$$

$$E_2 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \left[\frac{e^{-ikl_1 x}}{-ikl_1}\right]_{A-a}^{A+a} \left[\frac{e^{-ikm_1 y}}{-ikm_1}\right]_{-b}^{b} \tag{A.33}$$

$$E_2 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 2a \left[\frac{e^{-ikl_1(A+a)} - e^{-ikl_1(A-a)}}{-2ikl_1 a}\right] 2b \left[\frac{e^{-ikm_1 b} - e^{ikm_1 b}}{-2ikm_1 b}\right] \tag{A.34}$$

$$E_2 = \frac{E_0 e^{ikr_{10}} e^{-ikl_1 A}}{i\lambda r_{10}} 2a \left[\frac{e^{ikl_1 a} - e^{-ikl_1 a}}{2ikl_1 a}\right] 2b \left[\frac{e^{ikm_1 b} - e^{-ikm_1 b}}{2ikm_1 b}\right] \tag{A.35}$$

$$E_2 = \frac{E_0 e^{ikr_{10}} e^{-ikl_1 A}}{i\lambda r_{10}} 4ab \operatorname{sinc}[kl_1 a] \operatorname{sinc}[km_1 b] \tag{A.36}$$

Adding the electric fields E_1 and E_2, the $e^{ikl_1 A}$ and $e^{-ikl_1 A}$ terms form a cosine function, and the total disturbance is,

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 8ab \operatorname{sinc}[kl_1 a] \operatorname{sinc}[km_1 b] \cos[kl_1 A] \tag{A.37}$$

A.4 Diffraction at 2 Rectangular Apertures: Incident Plane Waves

This section derives the electric field due to diffraction of off-axis incident plane waves at two parallel rectangular slits. The wavefronts have normals with direction cosines l, m, n.

$$E = E_1 + E_2 \tag{A.38}$$

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \iint_{\Sigma_1} E_{10}(x,y) e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{A.39}$$

r_{10} being the distance from the aperture screen origin to the observation point.

At the aperture screen, the electric field is $E_0 e^{-ik(xl+ym)}$ neglecting time dependency, and so,

$$E_1 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{-A-a}^{-A+a} \int_{-b}^{b} E_0 e^{-ik(xl+ym)} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{A.40}$$

$$E_1 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \int_{-A-a}^{-A+a} e^{-ik(l+l_1)x}\, dx \int_{-b}^{b} e^{-ik(m+m_1)y}\, dy \tag{A.41}$$

$$E_1 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \left[\frac{e^{-ik(l+l_1)x}}{-ik(l+l_1)}\right]_{-A-a}^{-A+a} \left[\frac{e^{-ik(m+m_1)y}}{-ik(m+m_1)}\right]_{-b}^{b} \tag{A.42}$$

$$E_1 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 2a \left[\frac{e^{-ik(l+l_1)(-A+a)} - e^{-ik(l+l_1)(-A-a)}}{-2ik(l+l_1)a}\right] 2b \left[\frac{e^{-ik(m+m_1)b} - e^{ik(m+m_1)b}}{-2ik(m+m_1)b}\right] \tag{A.43}$$

$$E_1 = \frac{E_0 e^{ikr_{10}} e^{ik(l+l_1)A}}{i\lambda r_{10}} 2a \left[\frac{e^{ik(l+l_1)a} - e^{-ik(l+l_1)a}}{2ik(l+l_1)a}\right] 2b \left[\frac{e^{ik(m+m_1)b} - e^{-ik(m+m_1)b}}{2ik(m+m_1)b}\right] \tag{A.44}$$

$$E_1 = \frac{E_0 e^{ikr_{10}} e^{ikA(l+l_1)}}{i\lambda r_{10}} 4ab \operatorname{sinc}[ka(l+l_1)] \operatorname{sinc}[kb(m+m_1)] \tag{A.45}$$

For the second aperture,

$$E_2 = \frac{e^{ikr_{10}}}{i\lambda r_{10}} \int_{A-a}^{A+a} \int_{-b}^{b} E_0 e^{-ik(xl+ym)} e^{-ik(xl_1+ym_1)}\, dx\, dy \tag{A.46}$$

$$E_2 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \int_{A-a}^{A+a} e^{-ik(l+l_1)x}\, dx \int_{-b}^{b} e^{-ik(m+m_1)y}\, dy \tag{A.47}$$

$$E_2 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} \left[\frac{e^{-ik(l+l_1)x}}{-ik(l+l_1)}\right]_{A-a}^{A+a} \left[\frac{e^{-ik(m+m_1)y}}{-ik(m+m_1)}\right]_{-b}^{b} \tag{A.48}$$

$$E_2 = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 2a \left[\frac{e^{-ik(l+l_1)(A+a)} - e^{-ik(l+l_1)(A-a)}}{-2ik(l+l_1)a}\right] 2b \left[\frac{e^{-ik(m+m_1)b} - e^{ik(m+m_1)b}}{-2ik(m+m_1)b}\right] \tag{A.49}$$

$$E_2 = \frac{E_0 e^{ikr_{10}} e^{-ik(l+l_1)A}}{i\lambda r_{10}} 2a \left[\frac{e^{ik(l+l_1)a} - e^{-ik(l+l_1)a}}{2ik(l+l_1)a}\right] 2b \left[\frac{e^{ik(m+m_1)b} - e^{-ik(m+m_1)b}}{2ik(m+m_1)b}\right] \tag{A.50}$$

$$E_2 = \frac{E_0 e^{ikr_{10}} e^{-ikA(l+l_1)}}{i\lambda r_{10}} 4ab \operatorname{sinc}[ka(l+l_1)] \operatorname{sinc}[kb(m+m_1)] \tag{A.51}$$

As with the case of normal incidence the E_1 and E_2 contributions add to give E which has a cosine modulation.

$$E = \frac{E_0 e^{ikr_{10}}}{i\lambda r_{10}} 8ab \operatorname{sinc}[ka(l+l_1)] \operatorname{sinc}[kb(m+m_1)] \cos[kA(l+l_1)] \tag{A.52}$$

Bibliography

[Allan] J.D.C. Allan. *Robot Calibration and Dynamic Measurement System.* NEL report DI/20 (September 1984). National Engineering Laboratory, East Kilbride, Glasgow.

[Altschuler *et al.*] M. D. Altschuler, B. R. Altschuler and J. Taboada. *Laser Electro-Optic System for Rapid Three-Dimensional Topographic Mapping of Surfaces.* Optical Engineering, **20** 6, (November/December 1981).

[Blyth] T.S. Blyth. *Set Theory and Abstract Algebra.* Longman 1975.

[Born and Wolf] M. Born and E. Wolf. *Principles of Optics.* 6th Ed. Pergammon Press 1980.

[Bouwkamp] C.J. Bouwkamp. *Diffraction Theory.* Rep. Progr. Phys. (London Physical Society) **17** (1954), 35-100.

[Brodbeck and Schiele] Bernd Brodbeck and Günter Schiele. *Messungen und Analysen Geometrischer Größen an Industrierobotern.* Proc. 8th International Symposium on In-

dustrial Robots, 4th CIRT, Stuttgart, West Germany. (May/June 1978), 255-268.

[Brunk] W. Brunk. *Geometric Control by Industrial Robots.* Proc. 2nd International Conference on Robot Vision and Sensory Controls, (2-4th November 1982), Stuttgart, West Germany.

[Burch and Forno (1)] J.M. Burch and C. Forno. *The NPL Centrax Lens - A New Lens for Photogrammetry.* Society of Photo-Optical Instrumentation Engineers **399** (1983).

[Burch and Forno (2)] J.M. Burch and C. Forno. *Progress with the NPL Centrax Camera System.* Proc. International Society for Photogrammetry and Remote Sensing. Rio de Janeiro, Brazil, (June 1984).

[Burton and Sutherland] Robert P. Burton and Ivan E. Sutherland. *Twinkle Box - A Three-Dimensional Computer Input Device.* Proc. AFIPS **43**, National Computer Conference, Chicago, Illinois, USA, (6-10th May 1974).

[Chesapeake] Chesapeake Laser Systems, Inc. P.O. Box 4392, Annapolis, MD 21403, USA.

[Cohn] P.M. Cohn. *Solid Geometry.* Routledge & Kegan Paul 1961.

[Cox] Arthur Cox. *Photographic Optics.* Focal Press 1971.

[Coxeter] H.S.M. Coxeter. *Regular Polytopes.* Macmillan 1963 (Dover 1973).

[Davey] Peter G. Davey. *Minutes of Meeting of Tripartite Technical Group, on Performance Testing for Robots.* Science and Engineering Research Council Rutherford Appleton Laboratory. SERC Robotics/159, 15th February 1982.

[Desmaret] J.P. Desmaret. *Method des Mesures Tridimensionelles à l'aide de Deux Theodolites.* Direction des Techniques Avancée en Automatisme, Groupe Mesure, Renault, 92109 Boulogne Billancourt, Paris, October 1981.

[Do Carmo] Manfredo P. Do Carmo. *Differential Geometry of Curves and Surfaces.* Prentice-Hall 1976.

[Doty] J.L. Doty. *Projection Moiré for Remote Contour Analysis.* J. Opt. Soc. Am. **73**, 366-372, (March 1983).

[Dresden] A Dresden. *Solid Analytical Geometry and Determinants.* Dover 1964.

[Faugeras] O.D. Faugeras *et al. Towards a Flexible Vision System.* Robot Vision, ed. J.M. Brady.

[Featherstone] Roy Featherstone. *A Program for Simulating Robot*

Dynamics. Department of Artificial Intelligence, University of Edinburgh, Working Paper 116, 1982.

[Fohanno] Tanguy Fohanno. *Assessment of the Mechanical Performance of Industrial Robots*. General Methods Staff, Automobiles Citroën, France.

[Ford] *Robot Assessment Program: Test Procedures, Test Data, Data Collection*. Ford Motor Company Robotics and Automation Applications Consulting Centre, 15100 Mercantile Drive, Dearborn, Michigan 48120, USA. January 1986.

[Gates *et al.*] J.W.C. Gates, M.C. Hutley and R.F. Stevens. *An Optical Device for the Verification of 3 Axis Measuring Machines*. Proc. NELEX 80, National Engineering Laboratory, East Kilbride, Glasgow. (7-9 October 1980).

[Gilby and Parker] J.H. Gilby and G.A. Parker. *Laser Tracking System to Measure Robot Arm Performance*. Sensor Review, October 1982, 180-184.

[Goodman] Joseph Goodman. *Introduction to Fourier Optics*. McGraw-Hill 1968.

[Idesawa *et al.*] Masanori Idesawa, Toyohiko Yatagai and Takashi Soma. *Scanning Moiré Method and Automatic Measurement of 3-D Shapes*. Applied Optics **16** 8, (August 1977), 2152-2162.

[Industrial Robot] *Enhancing Robot Performance Measurement.* The Industrial Robot, March 1986, 52-54.

[ISO] *Manipulating Industrial Robots - Performance Criteria and Related Testing Methods.* International Organisation for Standardization (working document), July 1987.

[Jarvis (1)] R.A. Jarvis. *A Laser Time of Flight Range Scanner for Robotic Vision.* IEEE Trans. Pattern Analysis and Machine Intelligence **PAMI-5** 5, (September 1983), 505-512.

[Jarvis (2)] R.A. Jarvis. *A Perspective on Range Finding Techniques for Computer Vision.* IEEE Trans. Pattern Analysis and Machine Intelligence **PAMI-5** 2, (March 1983), 122-139.

[Keller] Joseph B. Keller. *Geometrical Theory of Diffraction.* J. Opt Soc. Am., **52**, 2 (February 1962),116-130.

[Kendig] Keith Kendig. *Elementary Algebraic Geometry.* Springer-Verlag 1977.

[Kline and Kay] M.Kline and I.W. Kay. *Electromagnetic Theory and Geometrical Optics.* Interscience, New York, 1965.

[Levin] Joshua Levin. *A Parametric Algorithm for Drawing Pictures of Solid Objects.* Comm. ACM, **19**, 10 (October 1976), 555-563.

[Lewis and Johnston] R.A. Lewis and A.R. Johnston. *A Scanning Laser Rangefinder for a Robotic Vehicle.* Proc. 5th Joint Conference on Artificial Intelligence, (1977). 762-768.

[Machinery] Machinery and Production Engineering, 18th November 1981. 47-49.

[Movement Techniques] Movement Techniques Ltd, 12-14, The Technology Centre, Epinal Way, Loughborough, Leics. LE11 0QE.

[Nitzan, Brain and Duda] David Nitzan, Alfred E. Brain and Richard O. Duda. *The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis.* Proc. IEEE **65** 2 (February 1977), 206-220.

[Ozaki, Ito and Inagaki] Shotaro Ozaki, Kohei Ito and Souji Inagaki. *Standardization of Characterization Relating to Industrial Robots.* Proc. 7th International Symposium on Industrial Robots, Tokyo, Japan, (October 1977), 453-459.

[Page and Hassan] C.J. Page and H. Hassan. *Non-contact Inspection of Complex Components Using a Rangefinder Vision System.* 1st International Conference on Robot Vision and Sensory Controls, Stratford upon Avon, (1981), 245-254.

[Papoulis] Athanasios Papoulis. *System Transforms with Applications in Optics.* Krieger 1981 (McGraw-Hill 1968).

[Paul] Richard P. Paul. *Robot Manipulators: Mathematics, Programming, and Control.* MIT Press, 1981.

[Pipitone and Marshall] Frank J. Pipitone and Thomas G. Marshall. *A Wide-field Scanning Triangulation Rangefinder for Machine Vision.* The International Journal of Robotics Research, **2**, 1, (Spring 1983).

[Popplestone *et al.*] R.J. Popplestone, C.M. Brown, A.P. Ambler and G.F. Crawford. *Forming models of Plane-and-Cylinder Faceted Bodies from Light Stripes.* Proc. 4th Int. Joint Conf. Artificial Intelligence, 1975, 664-668.

[Roberts] Lawrence G. Roberts. *The Lincoln Wand.* Proc. AFIPS **29** Fall Joint Computer Conference, (7-10 November 1966). San Francisco, California.

[Selcom] Selcom Selective Electronic Co AB, Box 250, S-433 01 Partille 1, Sweden.

[Sommerville] D.M.Y. Sommerville. *Analytic Geometry of Three Dimensions.* Cambridge University Press, 1934.

[Spiegel] Murray R. Spiegel. *Vector Analysis and an Introduction to Tensor Analysis.* McGraw-Hill 1959.

[Stevens (1)] R.F. Stevens. *Zone Plate Interferometers for Verifying Three-Coordinate Measuring Machines.* PhD Thesis 1985, Loughborough University of Technology, Loughborough, England.

[Stevens (2)] R.F. Stevens. *The Design of Zone Plate Interferometers for use in Verifying the Positioning Performance of Multi-axis Measuring Machines.* MOM83 March 1987. National Physical Laboratory, Teddington, Middlesex TW11 0LW.

[Stevens (3)] R.F. Stevens. *A Zone Plate Interferometer for Pointing.* MOM84 March 1987. National Physical Laboratory, Teddington, Middlesex TW11 0LW.

[Sutherland (1)] Ivan E. Sutherland. *A Head Mounted Three Dimensional Display.* Proc. AFIPS **33** Part 1, Fall Joint Computer Conference, (1968).

[Sutherland (2)] Ivan E. Sutherland. *Computer Displays.* Scientific American **222** June 1970, 56-81.

[Warnecke and Brodbeck] H.J. Warnecke and B. Brodbeck. *Analysis of Industrial Robots on a Test Stand.* The Industrial Robot, December 1977.

[Welsh] N. Welsh. *Photogrammetry in Engineering* Photogrammetric Record **12**, 67, (April 1986), 25-44.

[Whitney *et al.*] Daniel E. Whitney, Christopher A. Lozinski, and Jonathan M. Rourke. *Industrial Robot Calibration Method and Results.* Charles Stark Draper Laboratory, Inc. CSDL-P-1879.

[Wolf and Marchand] E.Wolf and E.W. Marchand. *Comparison of the*

Kirchhoff and the Rayleigh-Sommerfeld Theories of Diffraction at an Aperture. J. Opt. Soc. Am. **54**, 5, (1964) 587-594.

[Yakimovsky *et al.*] Y. Yakimovsky and R. Cunningham. *A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras.* Computer Graphics and Image Processing **7** (1978), 195-210.

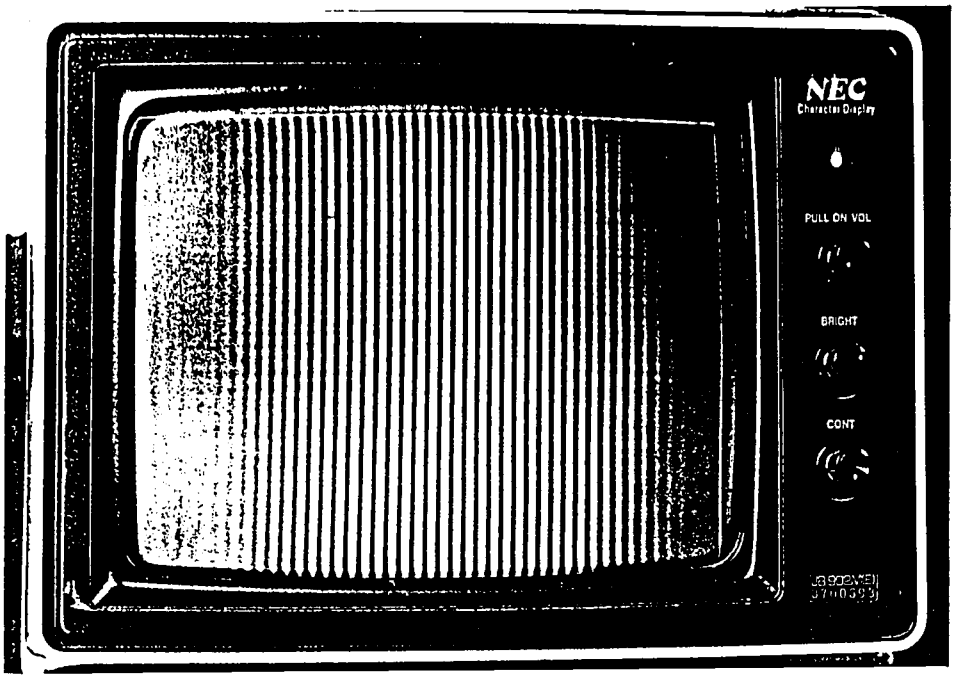


Plate 1. Interference Fringes due to Diffraction of Plane Waves by Two Slits.

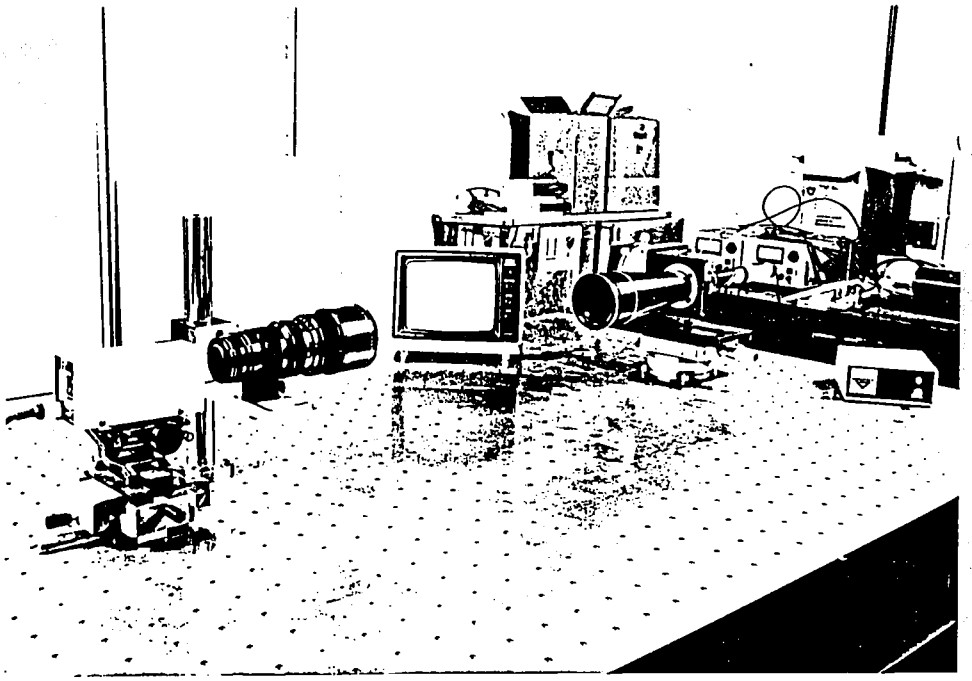


Plate 2. Apparatus to show on a Video Monitor the Interference Fringes focused by a Lens.

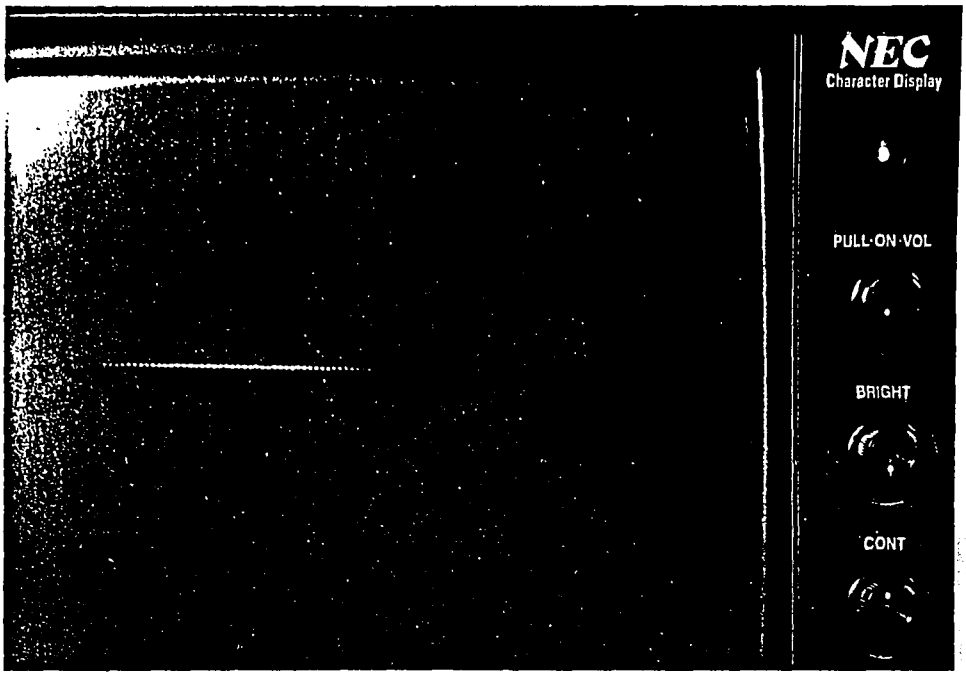


Plate 3. Interference Pattern obtained with Collimated Beam, Slits, and Lens as in Plate 2.

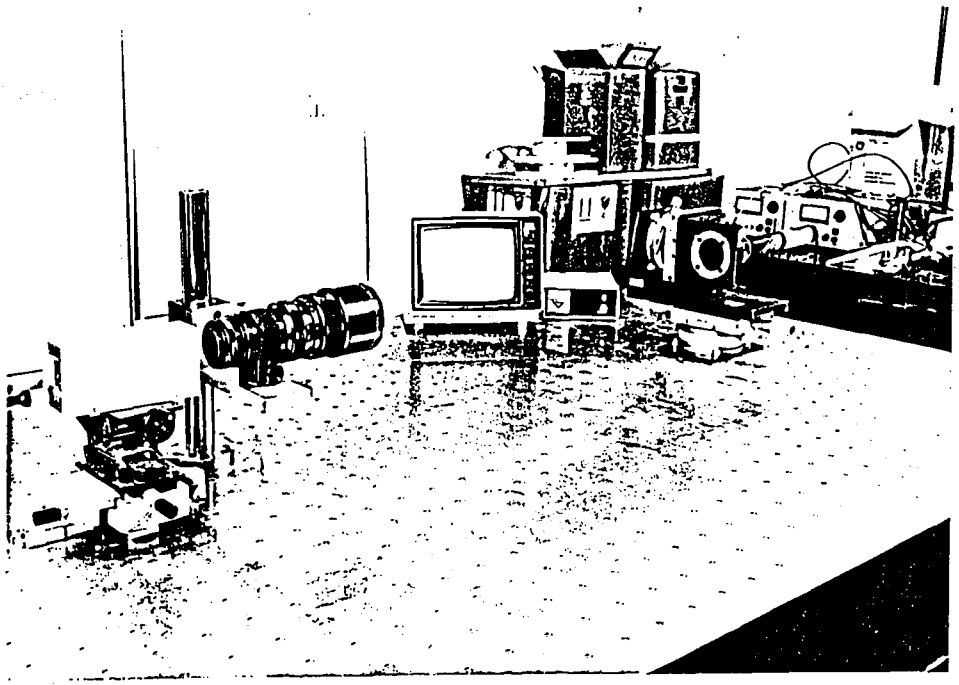


Plate 4. Apparatus with Collimating Lens removed to give Point Source.

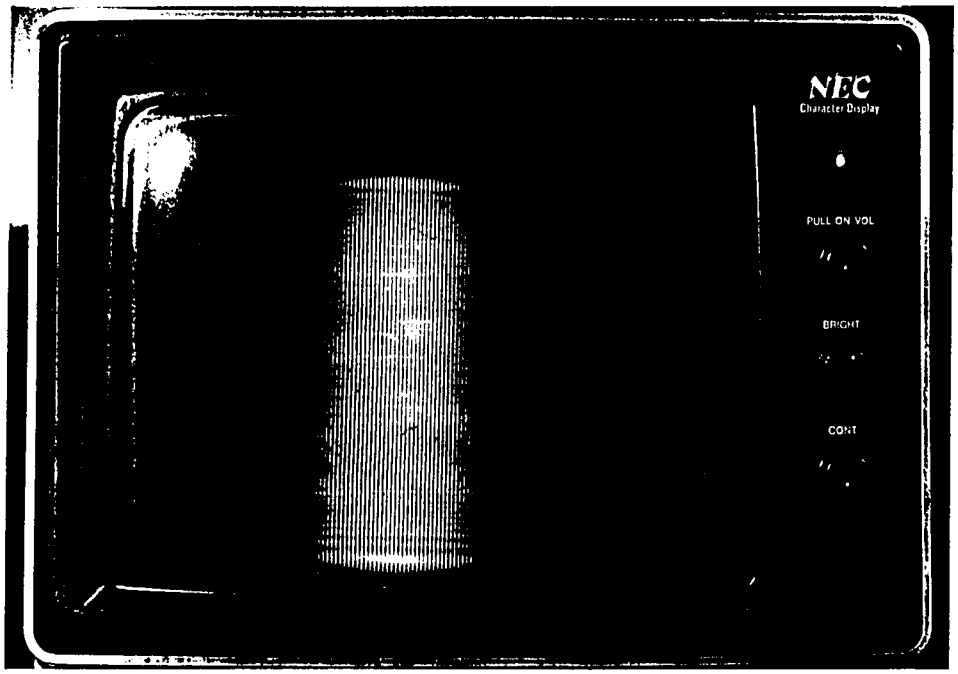


Plate 5. Interference Pattern obtained with Point Source, Slits, and Lens as in Plate 4.

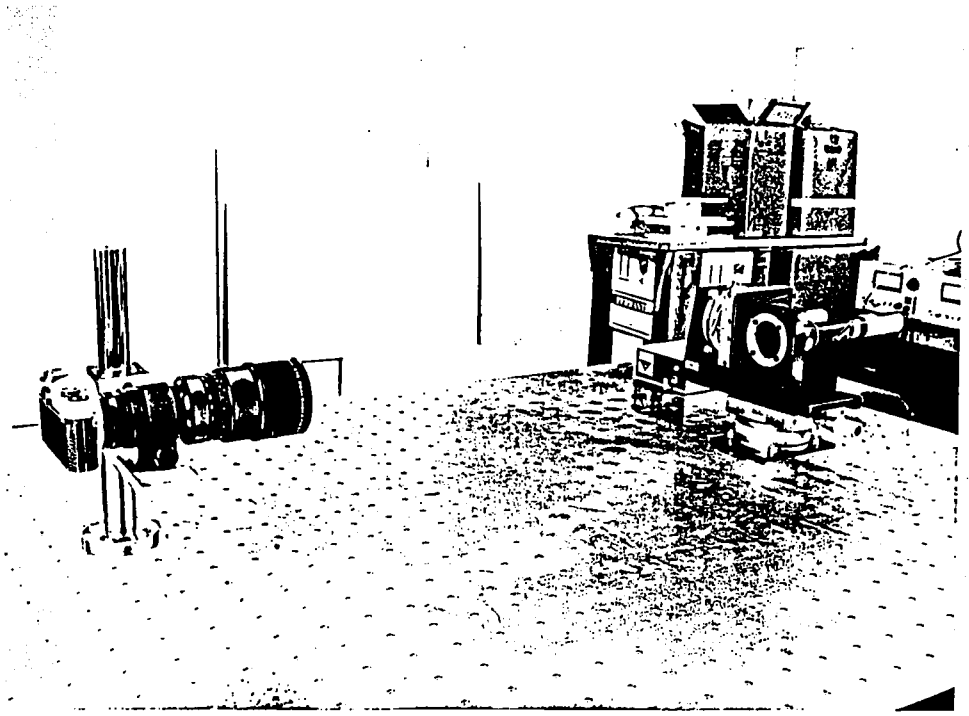


Plate 6. Camera fitted to obtain records of Interference Patterns.

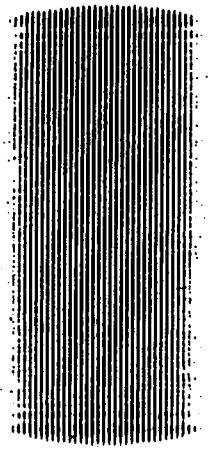


Plate 7. Print of Interference Pattern obtained with 270mm focal length Lens (after rephotographing negative).

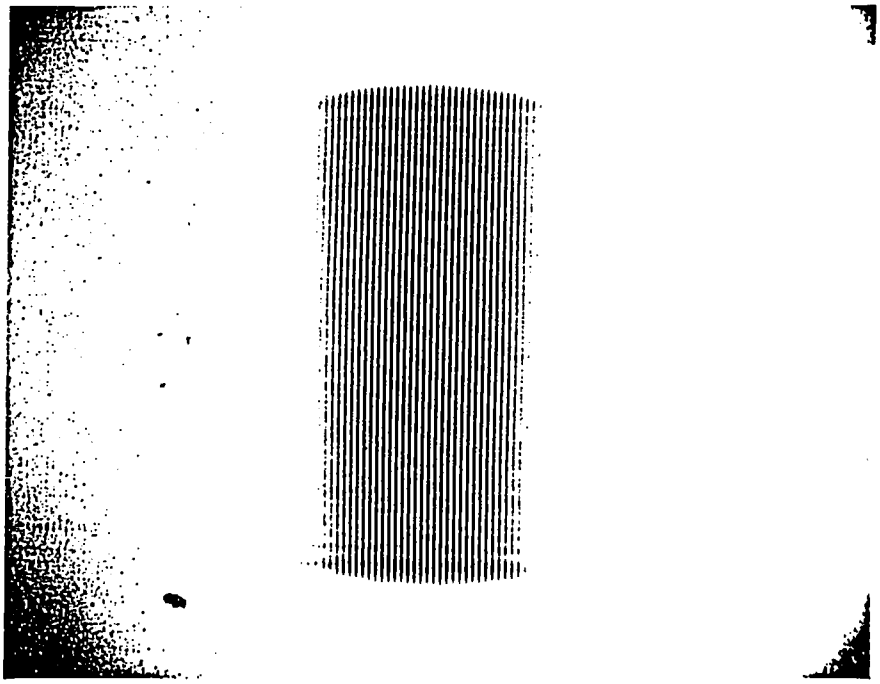


Plate 8. Print of Interference Pattern obtained with 100mm focal length Lens (after rephotographing negative).

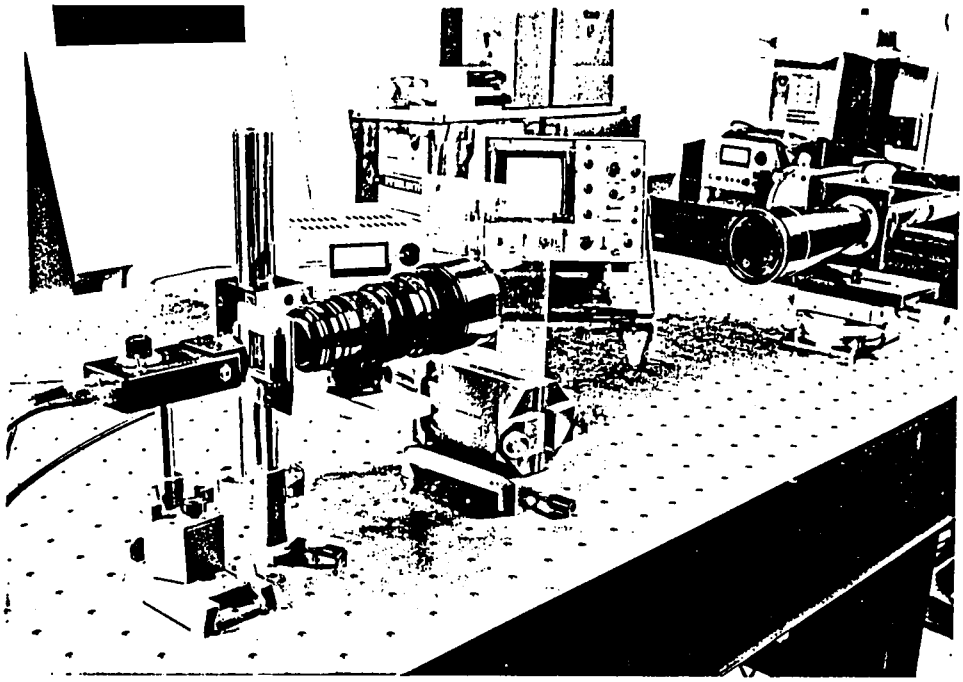


Plate 9. Apparatus to measure Diffraction Grating output with a Grating in front of the Lens.

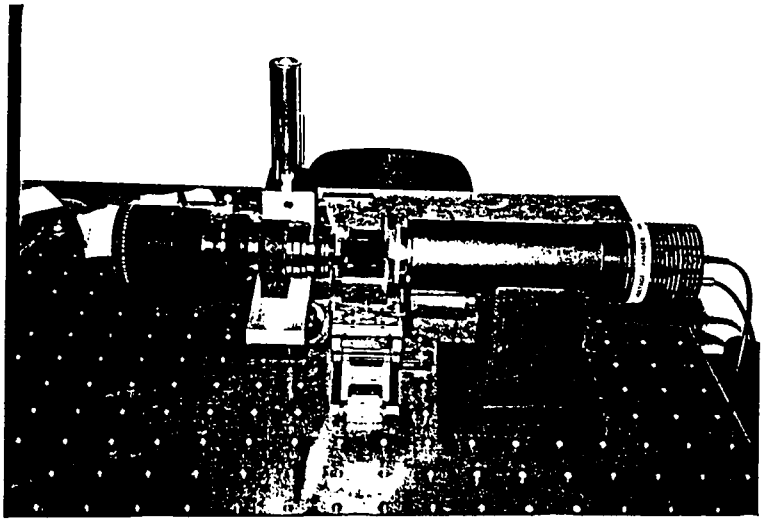


Plate 10. Apparatus to measure Diffraction Grating output with Two Slits in front of the Lens.