A RASTER SCAN CARTOGRAPHIC DIGITIZING SYSTEM

A Thesis

Submitted to the College of Graduate Studies and Research

in Partial Fulfilment of the Requirements

for the Degree of

Master of Science

in the Department of Electrical Engineering

University of Saskatchewan

by

Philip Keith Robertson

Saskatoon, Saskatchewan

July, 1980

ACKNOWLEDGEMENTS

The author would like to express his gratitude to Dr. A. R. Boyle for his advice, guidance and encouragement throughout this work. Thanks are also due to Professor A. E. Krause for his help with the computer system, and to Mrs. J. Miketinac, Mr. J. Lal and Dr. H. Sato for their assistance with various aspects of the work.

Financial support in the form of a scholarship from the University of Saskatchewan is gratefully acknowledged.

UNIVERSITY OF SASKATCHEWAN

Electrical Engineering Abstract 80A207

A RASTER SCAN CARTOGRAPHIC DIGITIZING SYSTEM

Student: P. K. Robertson   Supervisor: A. R. Boyle

M.Sc. Thesis presented to the College of Graduate Studies

July 1980

ABSTRACT

The increasing demand for cartographic data in digital form has resulted in a need for a low cost digitizing system capable of rapidly digitizing large binary image areas. In the past, cartographic data has for the most part been digitally represented in vector form. However, the advent of powerful image display and processing devices, and the generation of large amounts of raster scanned data from sources such as Landsat, have made the digitizing, handling and storing of data in raster form attractive.

This thesis proposes a digitizing system based on a mini-computer controlled flat bed flying spot scanner. Map sheets can be digitized in binary form at a resolution of .004", with additional facility for the detection of lines as narrow as .002". Scanning a 30" x 30" map sheet takes approximately 10 minutes; the raster data is passed to the mini-computer (a PDP 11/60) where it can be further processed if required. A simple run length coding scheme, based on the statistical nature of raster represented map sheets, compacts the data into a form suitable for storage.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

LIST OF FIGURES

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF TABLES

LIST OF ABBREVIATIONS AND PRINCIPAL SYMBOLS

CSR	Control and status register
d	diameter of focused Gaussian beam (of 1/e intensity circle)
dB	decibel(s)
d.c.	direct current
FIFO	first-in first-out (buffer)
Hz	Hertz
"	inch(es)
kHz	kilohertz
MHz	Megahertz
msec	millisecond
nsec	nanosecond
p	position of line centre with respect to spot centre
pixel	picture element
rpm	revolutions per minute
s	sample spacing (as a length)
SOL	start-of-line
SOS	start-of-sweep
t	width of line, or space between lines, on map sheet
TTL	transistor-transistor logic
w	Gaussian beam radius (of 1/e intensity circle)

1. INTRODUCTION

1.1 The Need to Digitize Maps

Large amounts of cartographic data exist in the form of map sheets. Whilst these map sheets are useful for visual examination, newer and more widespread applications require that the data be in some digital form. This makes possible the transmission of cartographic information, calculations of dimensions or other required parameters, and overlaying of additional information such as statistical data or Landsat images. In addition, editing or modification of digitized map sheet information is rapid and avoids re-drafting; this is particularly advantageous when updating is frequent or extensive.

1.2 Representing Cartographic Data in Digital form

Most cartographic information is in the form of lines drawn on map sheets; these lines may represent contours, drainage paths, boundaries, roads or other features. A natural way to describe such information is to specify the path of each line either with a string of points co-ordinated to some chosen reference, or with a chain of successive vectors [3]. Line representation can be convenient for purposes of map modification, or re-drawing using an incremental plotter.

It is, however, becoming increasingly apparent that with recent advances in computational technology, representing maps in raster form can have distinct advantages [17]. In raster form, a map is described by a matrix whose element values correspond to the nature of the map at or around their co-ordinating spatial positions on the map sheet. Usually the co-ordinate axes are mutually orthogonal. With the decreasing cost and increasing speed of high density memory chips, and the resulting advent of powerful image display and processing systems, handling and processing large arrays is feasible, and can be fast. The complex file structure overhead generally characterising vector represented data sets is unnecessary for data in raster form, and real time processing with the aid of image display devices introduces the possibility of performing vector-like operations (such as line removal or shaping) directly on raster data. In fact, a wide range of algorithms has already been developed to process data in raster form [10], and the scope of raster data processing is rapidly expanding, due in part to the generation of large volumes of raster data from such sources as Landsat.

It is thus apparent that in the future, large scale geographic information systems are likely to be raster, rather than vector, based. However, it is unlikely that vector data representation will disappear; certainly most existing digitized cartographic data is in vector or co-ordinate form; and thus a raster based cartographic digitizing system is likely to be of little use if no facility exists for the changing of data format to make it compatible with existing data. In fact, raster to vector (or scan to line) conversion algorithms have recently been widely developed [5,7,8,13,15,18], and direct or modified implementation of these algorithms is feasible using present-day mini-computers.

Given that cartographic data in raster form is required, there remains the problem of digitizing map sheets. For such digitizing to be feasible on a large scale, it must be fast. Reference 2 contains a summary of existing scanning systems, their capabilities and limitations. For the most part, these are designed for grey scale reproduction at high resolution, and few will scan an area the size of a map sheet at high speed. However, cartographic data is two-level, and scanning output need only be binary. In addition, there is little point in scanning at a resolution higher than that after which no further useful information is obtained; this limit is set by original drafting accuracies. The scanning system proposed aims to meet the requirements specific to the digitization of map sheets, achieving high speed while maintaining simplicity of operation and low cost. It is also important that it could be easily integrated into a cartographic data processing environment.

1.3 Outline of Thesis

The preceding sections of this chapter have briefly described the reasons for developing a raster based cartographic digitizing system, and have outlined in general terms the requirements of such a system. In chapter 2, the various methods of digitizing map sheets are described, and the suitability of each for the large scale digitizing of map sheets is discussed. On the basis of conclusions drawn in chapter 2, chapter 3 presents the overall design of the proposed system. Following chapters then treat specific features of this system.

The mechanical and optical designs of the scanner itself are presented in chapter 4. Chapter 5 makes a detailed investigation into the effects of the optical system and electrical hardware signal processing on data resolution and integrity. Design criteria are established, and designs to meet these criteria are presented.

In chapter 6, the hardware and software required to interface the scanner to a PDP 11/60 mini-computer, allowing automatic control of the scanning process, are described. Chapter 7 looks at the problem of handling the large volumes of raw raster data generated by the scanner; the statistics of this data are investigated, and a simple run length coding scheme to compact the data into a form suitable for storage or further handling is developed. Vectorization processes are briefly described, and compaction efficiencies of vector represented and raster coded data are compared.

Chapter 8 makes a critical evaluation of the performance of the digitizing system, presenting the results of digitizing test patterns and map sheets, and isolating areas in which improvements could be made. Finally, in chapter 9, the suitability of the proposed system for the large scale digitizing of map sheets is considered in light of the results presented in chapters 7 and 8.

2. METHODS OF DIGITIZATION

2.1 Introduction.

Since most of the information on map sheets is in the form of lines, conventional methods of digitizing have tended to follow these lines, recording either the absolute position of successive points on the line with respect to a fixed origin point, or the incremental changes in position between successive points given some initial absolute point. This type of digitization is known as line following, and can be done either manually or semi-automatically.

Another possible method is to scan the entire map sheet at some fixed resolution in a raster type process, differentiating between marks on the map, such as lines, and areas of no information. The resulting data can be reconstructed into a matrix which visually represents the map sheet scanned.

In this chapter, both methods of digitization are described, and their relative advantages considered. In addition, the differences in the data format resulting from each process are demonstrated, and methods of changing format are briefly investigated.

2.2 Line Following.

2.2.1 Manual line following.

Manual line following systems range from the straightforward reading of measurements from mutually perpendicular scales to the more recently developed high resolution microprocessor controlled digitizing tablets. For data volumes as large as those from map sheets, some element of computer control or operation is virtually necessary to speed up the digitization process. Most practical systems are based on one of three principles: first, the generation and counting of digital pulses, the number of pulses being proportional to the incremental distance moved in mutually perpendicular directions; second, the time taken for an acoustic signal generated at the point to whose position is to be digitized to reach fixed reference positions (usually orthogonal axes); and third, the variations in electric field caused by a cursor being placed over the point to be digitized, above a grid type network of wires or charge coupled devices.

The first technique can be achieved simply by mounting gear tracks and rotary encoders onto a normal drafting machine; the operator moves the head of the machine along a line to be digitized and a count proportional to the distance moved in each perpendicular direction is generated by the rotary encoder. Spacing between successive points can be time based, or more commonly operator controlled by some mechanical means such as a footswitch or finger button. Counts generated can be entered into a computer on interrupt triggered by the switch. Consequently the data generated is in incremental form with respect to some initial origin point.

Both the second and third techniques generate data in the form of absolute co-ordinates representing position on the map sheet, referenced to the origin of the digitizing device. More complex computer control or built-in processing capability allows effective origin shifting, incremental count calculation and display and other refinements designed for easier usage. Digitizing tablets based on the wire grid principle are in widespread use, and standard interfaces to mini-computers or micro-computers allow continual point logging.

Manual digitizing systems have several drawbacks. Not only is it time consuming to digitize the large amount of information on a map sheet manually, but also the process is highly error prone. Several systems (for example the Hewlett-Packard 9874A) have been developed specifically to minimize operator fatigue, but nevertheless, errors such as jumps between close lines, entering of spurious points or multiple recording frequently occur, making extensive editing often necessary. Typically, a standard sized map sheet of average information content might take between 10 and 30 hours to be digitized by manual methods, with as much time again for editing and error correcting.

2.2.2 Automatic line following.

Automatic, or semi-automatic, line followers have had limited success in cartographic applications [5]. In an automatic line following process, essentially some mobile device locates a line on the map sheet, and tracks it until it reaches a finishing point, recording the position of the mobile device as it follows the line. This can be achieved by several methods; whatever the mobile device used, some form of intelligent control is necessary. Existing systems are mainly mini-computer controlled, often making extensive use of the processor of that computer for long periods of time. This type of system is slow; following speed is limited by the momentum and response of the tracking device, as line jumping or spinning off at sharp corners must be avoided. Severe problems can be encountered at right-angled corners, and junction points must be recorded in some manner. Fairly frequent operator intervention can be required to maintain following integrity, and the problem of repeating lines is by no means trivial. Typically, an automatic line follower might take 10 hours or more to digitize a standard map sheet.

It can be seen then, that the large scale digitization of a map sheet series would be impractical using line following methods. For example, the 1:24000 series covering the USA consists of about 50000 maps, each with up to 10 or so separation sheets; perhaps 500,000 sheets in all. To digitize this series with manual or automatic line following equipment over a 5 year period would take over 500 people working full time with as many machines; clearly not an economically viable proposition. It seems then, that a substantially more rapid method of digitization is required, and scanning systems offer this possibility.

2.3 Scanning.

A map sheet can be scanned in a raster type process to generate data describing the nature of the map at regularly spaced points. In general, the scanning process is designed so that these points form a rectangular or square grid, and the data generated can hence be arranged into a dot matrix visually representing the map sheet scanned.

Scanning systems are either reflective or transmissive. In a reflective system, a light source incident on a point on a map sheet will be reflected if the surface is light, while a dark line will absorb most of the light and little will be reflected. A transmissive system relies on detecting light passing through the map sheet; generally a negative of the sheet would be used, having transparent lines on a dark background.

The output from a scanner of either type is a serial data string which can be re-arranged into a sparse matrix. Each element of the matrix, or pixel, represents the nature of the map sheet at some fixed point on the map co-ordinated by the matrix row or column index. More accurately, it represents a measure of the average density of some neighbourhood around that point, the size of the neighbourhood depending on the physical characteristics of the scanning system such as light spot size and detection technique. This is considered in greater detail in later chapters.

Scanning processes can be extremely fast; the limiting factor in speed of operation becomes the rate at which the large volumes of data generated can be handled and temporarily stored. It is certainly not unreasonable to expect to scan an entire map sheet in a matter of a few minutes. However, the data generated has a high level of redundancy, as a large proportion of a map sheet contains no useful information, and further, the data is in a raster form. As mentioned in chapter 1, it is probable that in the future raster formatted data sets will become more common than vector formatted sets; nevertheless provision must exist for conversion between formats. Consequently it should be recognized that some form of data processing, even if only compaction by efficient coding, must be applied to raster based data in order to render it practical for use and storage. To judge the feasibility of using data from a scanning process, the differences in data format, and the processing required to make data from various digitizing systems compatible, should be more closely examined.

2.4 Input Data Formats.

As described earlier, data from line following types of digitizers consists of either sets of incremental nose-to-tail vectors, or successive co-ordinate pairs (which are easily converted to incremental vectors) describing the track of a line across a map sheet. Lines drawn on a map have a finite thickness or width, and the track described is on average a path along the centre of the line. This is illustrated in figure 2.1; in 2.1(a) a small section of a map is enlarged to show the finite line width. Figure 2.1(b) shows the result after line following; a set of vectors along the centre path of the line. Figure 2.1(c), on the other hand, is a matrix reconstruction of the serial data string produced by scanning the small area; black dots represent samples taken at a fixed resolution spacing where some part of a line was encountered. Clearly if each black dot were joined to its nearest neighbours with a vector, as in the case of absolute co-ordinate pairs resulting from line following, a multitude of spurious vectors would result. If then, vector line representation were required, the matrix of black dots would have to be reduced to a state shown in figure 2.1(d), where effectively a single string of dots remains.

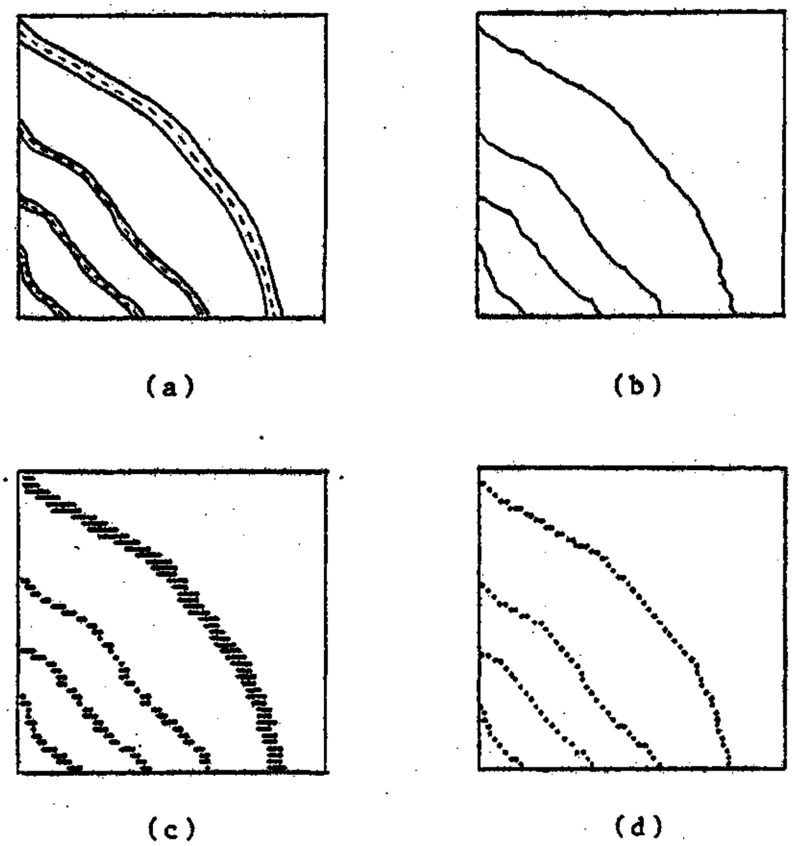


Figure 2.1 Line vector and raster data formats

This process of reducing line widths is known as line thinning, and can be approached in terms of the connectivity of the pixels; by specifying certain rules, precise algorithms can be developed to perform the operation. The resulting data is still in raster form, and further processing can be performed to locate, and convert the apparent lines into vector strings. Software techniques to perform these tasks are described in more detail in chapter 7, but it is clear that if a scanning type of process were used to digitize map sheets, fairly extensive data processing might be necessary to both minimize redundancy and make formats compatible with those of existing data sets.

2.5 Summary.

In fact a substantial amount of work is presently being done in the area of software processing of cartographic data, and results have shown that it is feasible, given certain restictions to be met by the data generating mechanisms, to thin and vectorize the raw raster data in roughly the length of time taken to collect it; that is, a time of the order of minutes rather than of hours. It is thus possible to digitize and process a complete map sheet within a time orders of magnitude smaller than that required for line following types of digitizing.

On this basis, it was decided that a system consisting of a scanner operating on line to the type of computer suitable for data processing should be designed and constructed. The following chapter outlines the factors influencing the design of the scanner and choice of computer, and gives details of the resulting system.

3. DIGITIZING SYSTEM DESIGN

3.1 Introduction.

The type of system most suitable for the digitization of map sheets can be determined from various practical considerations. First, the form of map sheet most frequently required to be digitized, and the required spatial resolution, govern the geometrical and mechanical design. Secondly, the computational system under control of which the scanner is to run, and to which it is to pass the data, places speed and other operational limitations on the design.

This chapter first outlines the characteristics of map sheets to be digitized. It then looks at the two basic types of scanner suitable for scanning large image areas such as map sheets; namely, the drum and flat bed scanners. Possible methods of operating each under the imposed cartographic specifications are discussed, and the requirements of the computer to be used in conjunction with the scanner are examined. Finally, the possibility of adapting the scanner to operate with various computer systems without major modifications is investigated.

3.2 Map Sheet Characteristics.

A complete printed map is a result of overlaying up to as many as perhaps 12 separation sheets. These separation sheets are usually in the form of photographic negatives on a base of mylar, or similar material, which has good dimensional stability over temperature and humidity changes. It is a fairly straightforward task to produce positives from these negatives, if required, though scratches or marks can be more easily opaqued out on a negative. Hence the map sheet can be made available for either transmissive or reflective scanning.

The required resolution of scanning is determined from the visual characteristics of the map, which in turn result from drafting accuracies. In general, it is accepted that a resolution spacing of .004" (or .1 mm) is adequate. One important reason for not using too small a resolution spacing is that the volume of raster generated data increases with the square of the increase in resolution; this can make map data handling before compaction inconvenient or even impracticable. In addition, for most map sheets digitization at a resolution spacing smaller than .004" renders no additional useful information. Lines drawn on a map sheet can vary in width from approximately .004" to .016", though in some cases may be as narrow as .002". Techniques for detecting these narrow lines, while still digitizing at a resolution spacing of .004", are described in chapter 5. It should be noted that various photographic techniques, such as reprinting with sandwiched spacers or diffusers, and subsequent copying onto lithographic film can be used to change the width of map sheet lines if the above specifications are not met.

Detection can be binary in nature; that is, grey levels are not required. It is sufficient to know whether any point on the grid of resolution spacing coincides with a line drawn on the map. If so, a binary '1' is recorded; otherwise a '0'.

3.3 Scanner Types.

Either of two types of scanner can be used for digitizing large image areas. A drum scanner consists of a cylindrical drum around which is wrapped the sheet to be scanned. The drum is rotated about the horizontal cylindrical axis at high speed, and a light source and sensor are slowly moved in a direction parallel to that axis, the beam being directed normal to the map surface. A distance of one resolution element is moved along the axis for each complete revolution of the drum. The map is hence scanned in successive sweeps, the path of each sweep being almost orthogonal to the direction of travel. The angle between the direction of sweep and direction of travel can be determined from the resolution spacing and the length of the map.

In a flat bed scanner, a trolley bearing the map sheet (usually lying in an arced bed rather than a truly flat bed) is moved slowly in one direction while a beam of light is deflected over the surface of the sheet in a perpendicular direction, usually by means of a rotating single or multi-faceted mirror located at the centre of curvature of the bed arc. The directions of sweep and travel can in this case be made truly orthogonal by simply adjusting the axial alignment of the rotating mirror.

Orthogonality of sweep and travel can be important cartographically for overlaying purposes; generally a mis-alignment of more than one resolution element could make rather lengthy software transformation algorithms necessary. This should be avoided if possible.

3.4 Reflective and Transmissive Detection Methods.

Because of the mechanical problems involved with placing either a light source or a detector within a rotating drum, drum scanners lend themselves to reflective rather than transmissive detection methods. Dark markings on a light background can be detected by a lack of reflected light (caused by absorption) at such points, while light will be reflected from light markings on a dark background. A sheet in transparency or negative form can be backed with a sheet of white paper or diffusive mylar to achieve the same effect.

On the other hand, a flat bed scanner is more suited to transmissive type detection (the bed can be made of transparent plastic or perspex), as reflective detection over a wide area (the length of the sweep) is difficult. This is because reflection from a point on a smooth surface gives a uniformly spherical spatial radiation pattern. For detection at a single point the detector can be made to encircle the source, and be placed close to the reflecting surface, resulting in collection of a large proportion of the reflected light. Over a long sweep, however, this is impracticable unless a detector array the length of the sweep is used. It is possible to make the inside of the bed reflective, causing the beam to return along its incident path. However, this requires a half silvered rotating mirror, and in addition care must be taken with the map bed reflecting surface when handling the map sheets.

3.5 Limitations Imposed by Computer Requirements.

If the system is to run in an automatically controlled manner, the characteristics of the computer used for this control must be considered. As mentioned earlier, data from a scanner requires some form of processing to reduce redundancy, and possibly to render it compatible with other data formats (such as vector represented lines). If the machine used for controlling the scanner is also used for processing the data, a minimum computational power is necessary. This suggests the use of a mini-computer. Alternatively, a dedicated micro-computer could be used to control the scanning and pass the data either to a mini-computer for processing, or to some storage medium such as magnetic tape.

The type of machine used also dictates possible data entry methods. Since the total volume of non-compacted data from a map sheet is high, (approximately 60 Megabits from a 30" X 30" sheet scanned at a resolution of .004"), the data entry speed must be high. However, the real problem is not the data entry speed, but rather the rate at which data can be written from the core of the computer to some temporary storage area, such as disk or magnetic tape. This requires the facility of some sort of direct memory transfer, where the processor need only set up the transfer addresses after which it is free to return to handle data entry. Even so, some form of data handling, incorporating a break in the input data stream, is probably going to be required. Consequently it is desirable to have some form of spatial gapping in the generation of data, and hence in the input data stream. A flat bed scanner lends itself well to this by offering the possibility of varying the number of flats on the rotating mirror to allow time gaps between successive sweeps. It should be noted that allowing the computer a break in the input data stream could also be achieved by using a deep input buffer. This buffer would temporarily store the data until the computer became free to accept it, but the cost of a sufficiently deep buffer is prohibitive, and faster than normal input rate requirements on the computer for "catching up" would be imposed. In fact such a buffer (but shallower) is useful for other reasons; this is described more fully in chapter 6 when discussing interface hardware/software configurations.

3.6 Format of Data passed to Computer.

It should be noted that data could be encoded in some way (such as run-length encoded) before entry to the computer. If a suitable coding scheme were chosen smaller data volumes would result, but it was decided not to perform this real time data encoding for several reasons. First, the coding scheme would have to be designed on the basis of the expected statistics of the data set, and it is by no means clear that cartographic data will always be of a predictably consistent statistical nature. Secondly, the hardware to perform such coding would make the interfacing hardware more complex; a sufficiently flexible coding method would involve some form of programmable device, and one aim of the system was simplicity. Thirdly, regularly sampled uncoded data imposes entirely predictable timing and handling requirements on the computer hardware and software, identical for all types of map data. On the other hand, coded data could give rise to unexpected demands on the computer if a particular block of data were unsuitable for the coding scheme being used; that is, the coded data could become more bulky within a given time period than the equivalent uncoded data. To cover this possibility additional hardware buffering and control would be required. Fourthly, since most software available for processing raster data operates on a full data matrix, decoding would have to be performed prior to processing. As processing of some sort would usually be carried out on the raw data, uncoded data is preferable providing sufficient temporary storage space is available for retaining the data until processing. Finally, and perhaps most importantly, since the designed system is essentially a developmental one, and the statistical nature of raster represented cartographic data is not really yet known, it seems more practical to compact the data using software techniques after processing, rather than at an intermediate stage of the digitizing process.

3.7 Choice of Computer.

It was decided that a mini computer (a PDP 11/60 operating under a multi user time-sharing software system, RSX 11-M) should be used to control the scanner, and accept and process the generated data, for the following reasons. First, in a cartographic environment in which automated processes are used, a machine with reliable software support and a powerful software operating system is almost mandatory. Secondly, due to the large data volumes and consequent handling requirements, fast direct memory access to disk and tape is more or less necessary. In addition, since map sheet scanning is unlikely to be a continual process during development, a dedicated processor for running the scanner alone would not be warranted. Thirdly, modifications in operational control and data processing to suit the needs of the user are particularly easy on a mini-computer with reasonable software support. Finally, on practical grounds, most cartographic institutions requiring the use of such a digitizing system already use mini-computers, very often those of the PDP 11 series. Since a PDP 11/60 was available for development of the system, this seemed a logical choice. Further details involving the operating system used and the software design approach are given in chapter 6.

3.8 Adaptability for use with other Co~puter Systems.

A final factor influencing the design of the scanner was the possibility of adapting it to operate with other computers. The previously mentioned facility of the flat bed scanner for inserting breaks in the data stream is particularly advantageous in this respect. For example, if the maximum continuous data entry rate to a computer were half that originally designed for, the average data generation rate over a cycle could be reduced by simply increasing the buffer depth of the interface, and using some of the break time to unload the accumulated data in the buffer, effectively allowing the computer to catch up. This consideration also influenced the interface design as described more fully in chapter 6.

3.9 Choice of Scanner Type.

Because of its relatively simple construction, a flat bed scanner was chosen as being the more suitable type. The mechanical problems of controlling the large revolving mass of a drum scanner are avoided, and the facility for gapping the generated data is particularly attractive. A flat bed scanner more or less determines that transmissive type detection be used; this is accomplished using a transparent bed. Complete details of the scanner construction are given in the following chapter.

3.10 Summary

The approach taken in the design of the system has been outlined in the preceding sections of this chapter. Basically, the intent was that simplicity and flexibility be maintained in so far as was possible while still meeting the cartographic requirements. The final system arrangement, consisting of a mini-computer controlling, and receiving data from, a flying spot flat bed scanner, allows sufficiently for modifications of parts of the system without the need for re-design, while at the same time meeting the cartographic requirements for the digitization of map sheets. The following chapter treats the design of the scanner.

4. SCANNER DESIGN

4.1 Introduction.

The previous chapter establishes the essential features of the overall scanning system required. Details of the design and construction of the scanner itself are presented in this chapter. To begin with, a set of general specifications derived from cartographic requirements and computational restrictions is outlined. As the method of entering data to the computer determines the speed at which the scanner can operate, and hence its geometry and timing relationships, the data entry rate restrictions are first established, following which the mechanical and optical designs are presented. Also described are the methods used for detecting and electrically synchronising the mechanically directed optical signal resulting from the scanning of a map sheet by the flying spot.

4.2 Specifications derived from Cartographic requirements and Computational restrictions.

Based on the considerations of sections 3.1 to 3.7, the following specifications were established. First, the scanner should be capable of digitizing map sheets of up to 30" square, in negative form, at a resolution of .004"

(.1 mm). In addition, some facility for the detection of lines down to .002" in width should be provided. Secondly, map sheets should be digitized to two levels, corresponding to opaque (digital '0') and transparent (digital '1') material on the map sheet. Thirdly, orthogonality of co-ordinate axes should be achievable to within one resolution element error over the entire sheet width. Fourthly, the scanning of a map sheet should take of the order of minutes, and not hours. Finally, the scanner should operate on line to, and be controlled by, a PDP 11/60 mini-computer used under a multi-user operating system, RSX 11-M.

4.3 Data entry to Computer.

In order to keep the system both as simple and as flexible as possible, the use of direct memory access (DMA) to transfer the data from the scanner to the computer was ruled out. (This also on cost considerations.) Instead, a 16 bit I/O port, the DR11-C, normally used as an interprocessor communications buffer on the PDP-11 series, was chosen for data entry. This port consists of a 16 bit input register, a 16 bit output register, and a 16 bit control and status register with facility for interrupt handling.

Using conventional data handshaking methods, the maximum rate of transfer of data from an interface to the computer memory is governed by the time taken to process the instructions making up the input routine, assuming no delays are caused by the interface hardware. On the PDP 11/60 these instruction times cannot be exactly predicted due to the hit or miss nature of the cache memory usage, but a maximum overall cycle time can be estimated. Real time testing by watching flag levels while inputting a continuous stream of data verified that the average time taken to input a 16 bit word was approximately 18 microseconds, and never more than 19 microseconds. Choosing then a data input rate of one word every 32 microseconds leaves a safety margin either for software testing, or possibly for increasing resolution without re-design. This means that a data sampling rate of 500kHz is required, or one sample every 2 microseconds. For this reason, a 1MHz crystal controlled clock was chosen as a real time base for the entire system, and divided by two to generate the sampling rate of 500kHz.

4.4 Mechanical Design.

4.4.1 General structure.

A flat bed scanner was designed to meet the requirements specified in section 4.1. The trolley holding the arced map bed travels along a ground steel edge which defines one axial direction for the generated data matrix.

A 3.5 milliwatt helium-neon laser is used as a light source and is mounted on a short optical bench along an horizontal axis above the trolley. An optical beam collimating and focusing system is also mounted on the optical track, and the resulting beam is deflected vertically downwards by a rotating single faced mirror mounted on a flywheel on the motor shaft. The whole is mounted on a solid levelled structure. Figure 4.1 shows the essential features of the scanner structure.

4.4.2 Sweep timing.

Section 4.3 establishes a sampling period of 2 microseconds for data collection. For reasons described in section 4.6, the sweep and travel motors must be synchronised with the data sampling clock. It is nevertheless desirable that their driving frequency be close to 60 Hz so that testing may be carried out while running from mains or line frequency independently of the clock and control logic. Dividing the 1 MHz clock frequency by successive factors of 2 generates a frequency of 61 Hz, and this was chosen as the motor driving frequency. Using a 900 rpm 4 pole-pair motor to rotate the deflecting mirror, a time of 66 milliseconds per revolution results. In order that map sheets of up to 30" square can be accomodated, and leaving a few inches for edging, an arced bed subtending an angle of 90 degrees at the centre of curvature, the rotating

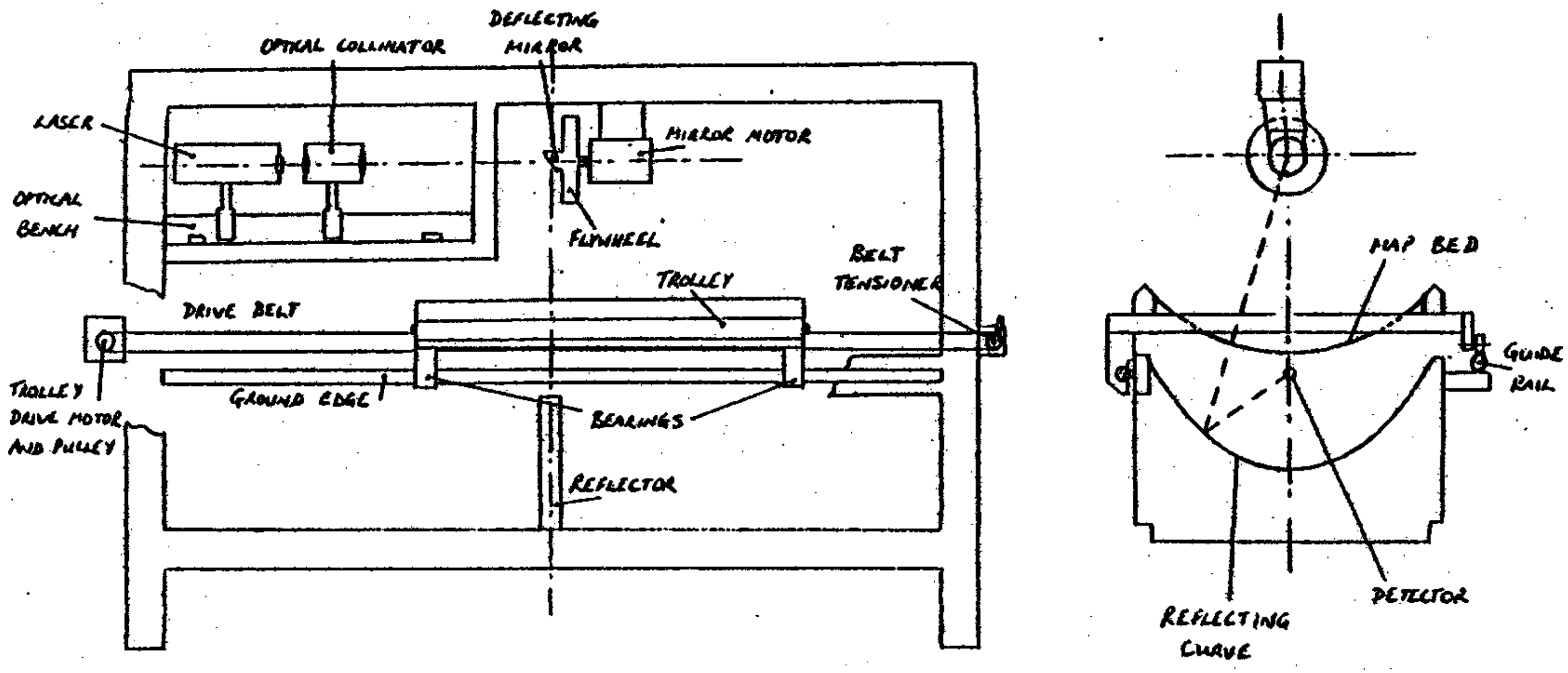


Figure 4.1 Scanner structure

mirror axis, was constructed. This gives a bed which can take map sheets of up to approximately 33" in width, this width being swept in 16.5 milliseconds. To obtain a sweeping speed of .004" per 2 microseconds along the arc, the radius of curvature of the bed must be 21".

4.4.3 Travel timing.

Having established the rotational period of the mirror motor as 66 milliseconds, the linear speed at which the trolley must travel is then determined. Assuming that the computer is able to take data every sweep; that is, that the 49.5 milliseconds gap is sufficient for any data handling between sweeps; the trolley must then move through .004" in 66 milliseconds. This corresponds to a linear trolley speed of .061 inches per second. The trolley is driven by means of a steel belt running around a fixed sized pulley; if a synchronous motor run from a driving frequency of 61 Hz is geared down to 1.02 rpm (1 rpm at 60 Hz) to drive the pulley, a pulley diameter of 1.14" is required.

4.5 Optical design.

4.5.1 Flying spot light source.

A helium-neon laser was chosen as a light source for the optical system. An incoherent light source, such as a filament bulb, has a finite spatial spread, while this does not apply to a coherent source such as a laser. Consequently the distribution of a laser beam focused to a point can be spatially smaller than that from an incoherent source, rendering it suitable for applications where a small spot is required. An optical system is needed to focus the laser beam to a spot. Design of this system requires that the nature of laser beams be more closely examined.

Although the phase front of a laser beam is slightly curved, its transverse intensity distribution is approximately Gaussian in every cross-section of the beam [4], as illustrated in figure 4.2(a). The radius of the beam at which the intensity is $1/e^2$ of the maximum central intensity is know as the beam radius w, and at the beam waist, where the phase front is plane, it has its minimum value, w_o. If z is a measure of distance along the optical axis in the direction of travel of the beam, the beam radius at any point along the optic axis can be expressed as a function of z:

$$w^2(z) = w_o^2 [1+(\lambda z/\pi w_o^2)^2]$$

where λ is the wavelength of the laser light (632.8 nm).

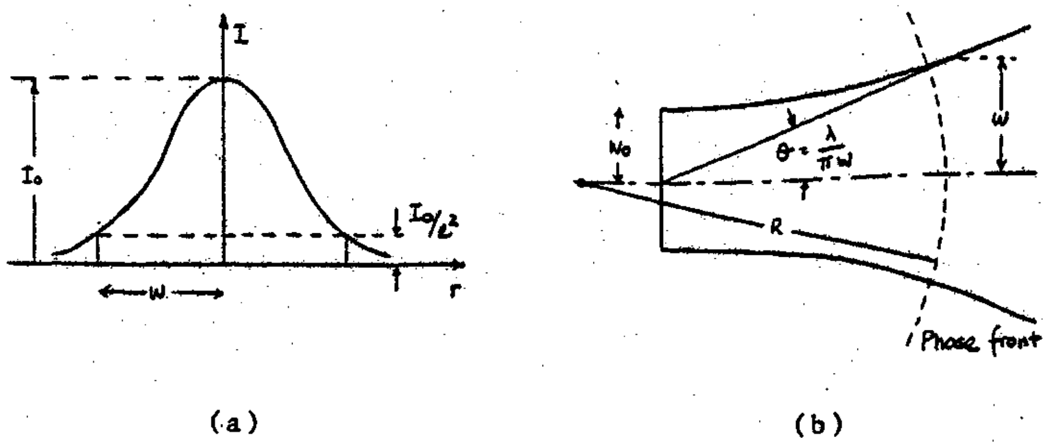


Figure 4.2 Gaussian beam characteristics

If w(z) is plotted for varying z, the beam contour shown in figure 4.2(b) is obtained; this curve is in fact a hyperbola with asymptotes inclined at an angle θ to the axis, where $\theta = \lambda/(\pi w_o)$. This angle is known as the far field diffraction angle, and is a measure of beam divergence. To be able to obtain a sharp focus, beam divergence must be kept small. Since sampling is to take place at spatial intervals of .004", the focused spot should be of this order of size. With the required focal distance, it turns out that focusing the exit beam from the laser (the diameter of this beam is approximately 1 mm.) cannot give such a small spot size. Consequently the initial beam radius w must be increased to reduce beam divergence. Beam radius can be increased by using a collimating system consisting of a pair of lenses such that the focal length of the second is greater than that of the first. This lens arrangement can also be used to focus the effectively expanded beam to a spot; lens positions can be calculated using ray transfer matrices. A complete analysis of the optical system of the scanner, including calculations of the collimating factors and lens specifications, and of the ray transfer matrix of the optical arrangement used, is presented in appendix A. Fine adjustment of focusing was provided for by mounting the secondary lens on a micrometer adjusting track, and setting for minimum pulse rise time. This technique, along with a method for actually measuring spot size, is described in chapter 8.

4.5.2 Light localisation for detection.

The problem of detecting the transmittance or otherwise of the light through the map sheet over the entire width of the sweep was mentioned earlier. One method is to use an array of detectors across the sweep path; however, this can both be costly and involve substantial circuitry. Instead, a reflector as shown in figure 4.1 was designed to such a shape that light deflected from the rotating mirror at all angles throughout the sweep be reflected to a common focal point just below the trolley. In order that a detector placed at this point should not interfere with the incident beam at the mid-sweep position, the reflector was tilted slightly. Due to this tilt, the focal locus becomes a short line, since the path length travelled by light at the edges of the sweep is longer than that travelled by light mid-sweep.

The reflecting curve forms part of an ellipse, with its foci at the rotating mirror and the detector. In order to obtain the true path of this ellipse to use as a template for cutting the reflector shape, its equation can be derived parametrically from the geometry of the scanner. In fact this generates a family of curves, but specification of one point referenced to the scanner frame edge fixes the path exactly. This path was obtained by numerical integration, and a scaled computer plot produced to use as a template for cutting the curve. A piece of 1" thick perspex was cut to the required shape (in fact it was cut to be 1/16" radially deeper than the true required shape) and a strip of 1/16" polished stainless steel was attached to the inside of the curve to act as a reflector. Adjustable mounts provide for alignment. Appendix A contains a parametric derivation of the curve.

4.5.3 Detection of localised light.

The reflector localises the incoming light to a point, or more exactly to a small volume, just below the trolley centre. The angle of incidence of light at this point varies over 180 degrees; since detectors generally have a limited angular field, a method of re-directing the localised light is required. Two methods are possible. The first is to use a wide-angle lens with the detector close to its focal point. However, a wide-angle lens covering 180 degrees is extremely bulky and costly. More practical a method is to place a small cylindrical diffuser at the detection point aligned axially with the focal line resulting from the reflector tilt. The diffused signal can then easily be detected.

A PIN diode was chosen as a suitable optical to electrical transducer for reasons explained in the following chapter. In fact three such diodes were mounted within the diffuser, backed by a ground plane circuit board on which the detector amplifier was built. Three diodes were used both to boost the signal level, and to achieve better signal linearity over the sweep, as small bumps on the reflector surface cause axial deviations of the spot along the diffuser.

4.6 Mechanical/Electrical Synchronization.

4.6.1 Synchronization of motors with data sampling clock.

It was mentioned in section 4.3 that a 1 MHz clock was chosen as a time base for the scanning system as a whole. In addition, the need was pointed out that the motors be run in synchronism with the sampling clock, rather than from mains or line frequency. This is for two reasons. First, fluctuations of up to .001% in line frequency over a short period of time (a second or so) would vary the effective length, in terms of the number of pixels per line, of sweeps by up to 8 pixels; that is, the far edge of the map would be jagged. Fluctuations of up to .001% in line frequency over a longer period of time (several minutes) would vary the effective number of sweeps made in successive scannings of the same physical area; that is, the total length of the map sheet would become indeterminate, as would the exact resolution spacing between sweeps. Such fluctuations are not uncommon, and for slow variations fluctuations of up to .01% are sometimes encountered. The second reason is that physical referencing of the start of each line would become indeterminate to within one resolution spacing. Consequently the motors must be synchronized with the sampling clock. This is achieved using circuitry described by the block diagram of figure 4.3, while circuit schematics can be found in appendix B.

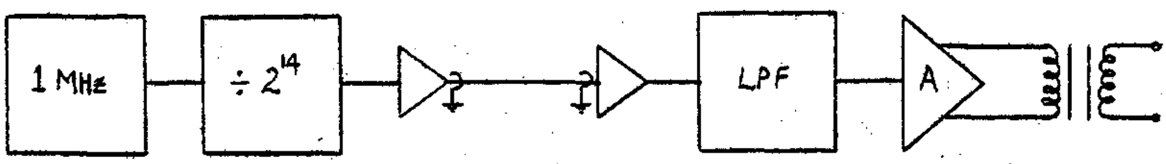


Figure 4.3 Block diagram of motor synchronising circuitry

The 1 MHz square wave from the clock is fed to a string of counters, resulting in a division by 16384 or 2^{14}. This square wave is then driven over a 50 ohm coaxial cable to the scanner, where it is passed through a 2-pole active low-pass filter. The resulting signal is amplified and transformed to give a final waveform of 110v rms at up to several amps. An optic isolator controls the carriage drive, and a safety toggle switch activated when the carriage reaches the end of its track protects the motor drive.

The output signal from the low-pass filter approximates a sinusoid of frequency 61 Hz, but also contains higher order components. A low-pass filter acts to integrate the applied waveform, and components of order 4m-1 (integer m) are phase reversed. Consequently the 3rd harmonic, the only significant component, adds to the fundamental to cause the output to tend to be triangular in nature. The effect of a present 3rd harmonic will be to cause a variation in the rotating field rate, and hence in the motor shaft revolution rate, at a frequency of three times that of the fundamental.

In fact in the sweep motor this variation was difficult to determine, and entirely masked by variations in angular velocity caused by friction in the motor's bearings, which were aperiodic in nature. A large brass flywheel mounted on the motor shaft reduced the variations, and section 8.2 of chapter 8 describes methods used to measure their magnitude. The trolley motor is geared down to such an extent that small periodic variations in driving frequency are of no practical significance.

4.6.2 Physical referencing of generated signals.

Some form of physical referencing to the start of each sweep is required when running the scanner. While this could be achieved by triggering off a transparent line on the edge of the map sheet, this has several disadvantages. First, a dirt or dust fleck obscuring the line would cause a loss of data for that sweep. Secondly, the angle between the sweep and travel axes would depend upon the physical positioning of the map in the bed, and this is unlikely to be achieved to within .004". Consequently a detector was attached to the scanner frame to generate a start-of-line (SOL) signal. This is actually used to gate the data sampling clock, and hence data collection is initiated. In addition, another detector was attached to the scanner frame so that it would be triggered a few milliseconds earlier in the sweep. The signal from this detector signifies the start of each sweep (SOS), and is used to generate the interrupt to the computer which causes it to accept the incoming data for that sweep.

Both detectors have the same electrical requirements; a digital signal only is needed, but in the former case its generated position must be reproducible. A fast rise time is also desirable. Although a phototransistor has a slow response (several microseconds) incorporation within a Schmitt trigger circuit employing positive feedback as shown in figure 4.4 ensures not only a very fast rise time, but also triggering with almost no hysteresis at the point on the transistor's output rise characteristic corresponding to the threshold level of the first gate. A speed-up capacitor reduces transistor turn-off time. Experimental examination of the time of triggering with respect to the start of the rise characteristic showed it to be reproducible certainly to within 50 nanoseconds. Consequently the detectors were determined to be suitable for the required purpose. TTL drivers were used to transmit the signals to the interface board over 50 ohm coaxial cable.

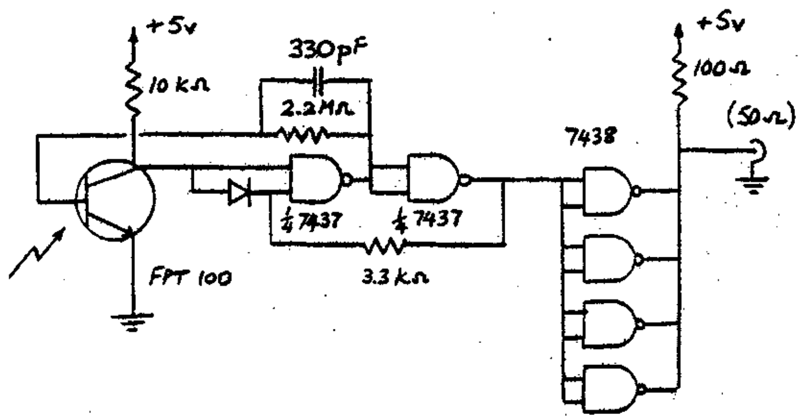


Figure 4.4 Circuit of start-of-sweep and start-of-line detectors

4.7 Summary

This chapter has described the main mechanical and optical features of the scanner, but has treated the electrical design only in so far as is necessary to describe the operation of the scanner itself. The following chapter looks more closely at the nature of the data generated during scanning under the optical and mechanical design constraints, and gives details of the design and construction of the signal detection and processing hardware.

5. SIGNAL DETECTION AND REAL-TIME PROCESSING

5.1 Introduction

Although in concept the detection of light passing through a map sheet and the subsequent generation of a digital signal is simple, in practice the integrity of the generated signal, or the closeness with which it represents the density profile across a prescribed line of the map sheet, depends on a number of factors.

The generation of digital data representing the information on the sheet involves detection of the modulated optical signal, optical to electrical transduction to give an analogue electrical signal, and real-time hardware processing of this signal to produce the required time quantized digital data. Given that the scanner geometry and control are properly designed, the accurate, or more precisely the required, reproduction of the map sheet in digital form is dependent on all of the above processes, each of which may introduce some kind of signal distortion, be it optical or electrical. It is important to be aware of the nature of this distortion so that the hardware processing circuitry may be designed to minimise its undesirable aspects, and possibly utilise particular features of it for specific purposes.

This chapter begins by briefly looking at the overall hardware signal path, after which the effects on resolution and signal integrity of each component in this signal path are examined, in order to determine the detectability or otherwise of lines. This leads to the design of the signal detection, amplification, digitizing and time quantizing circuitry in following sections.

5.2 Overall Signal Path

The transparent lines or areas on a map sheet effectively modulate the light beam to produce an optical signal with an intensity which varies between zero, corresponding to an entirely opaque area, and some maximum value corresponding to an entirely transparent area. The intensity is a measure of power over some defined area, and if the optical to electrical transducer is a device which responds linearly to incident power, the resulting detected signal has an amplitude proportional to the optical power transmitted through the map sheet. This detected signal must be amplified, and then digitized at some chosen threshold level. The digitized signal is sampled at regularly spaced time intervals corresponding to the spatial resolution points. Figure 5.1 shows a block diagram of the overall signal path.

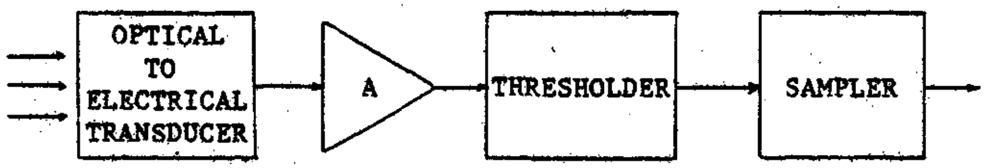


Figure 5.1 Block diagram of harware signal path

5.3 Resolution, Signal Integrity and Line Detection

The geometry and mechanical driving systems of the scanner are designed such that if sampling were to take place every 2 microseconds, a data matrix representing the density of the map at locations .004" apart would result. However, the nature of the matrix entry does not in fact represent the map density at a point location ; rather it specifies whether some measure of the density of a small area of the map centred at the point location exceeds a given value. Further, the relationship between detected signal amplitude and average density of the map over the small area is by no means linear, due to the Gaussian intensity profile of the light spot. It is apparent then, that both the achieved resolution, and the integrity of the data representing the map sheet, are dependent on both the size of the light spot and the chosen threshold value for digitization of the detected analogue signal. In addition the performance, in terms of bandwidth, of the detector and the amplifier through which the signal passes before thresholding, will also affect the data integrity. Finally, due to the inherent properties of a flying spot scanner, the nature and resolution of the signal detection in each of the orthogonal sweep and travel directions is governed by separate mechanisms. Each of the above effects has a bearing which may be more apparent in one of the axial directions, but can also be significant in the other.

It is particularly important to bear these factors in mind when considering the problem of detecting lines or marks, the width of which is less than the specified resolution spacing. If detection of such narrow lines is necessary, it is possible that techniques such as spot widening or mishaping, or using alternative sampling schemes, could be used to achieve detection, even though the detected position would be accurate only to the resolution spacing.

Such methods involve some kind of compromise; for example, spot widening causes a decrease in real resolution and reduces the detectability of spaces between lines. In order to be clear about the consequences of such methods on resolution and signal integrity, and possibly to establish some minimum specifications to be met, more quantitative examinations of the effects of spot size, threshold level, time quantization alternatives and amplifier bandwidth are made in the following sections.

5.3.1 Effects of spot size

It was stated earlier that the intensity profile of any cross-section of a laser beam normal to the axis of propogation is approximately Gaussian. Figure 5.2(a) shows such a profile, together with a view into the axis of propogation of the resulting spot. The dashed circle represents the locus of points at which the radiation intensity is $1/e^2$ or .135 of the maximum central value, and has diameter d. Roughly 86% of the total power in the beam passes through this circle. These $1/e^2$ points correspond to a distance of two standard deviations from the mean of the Gaussian distribution.

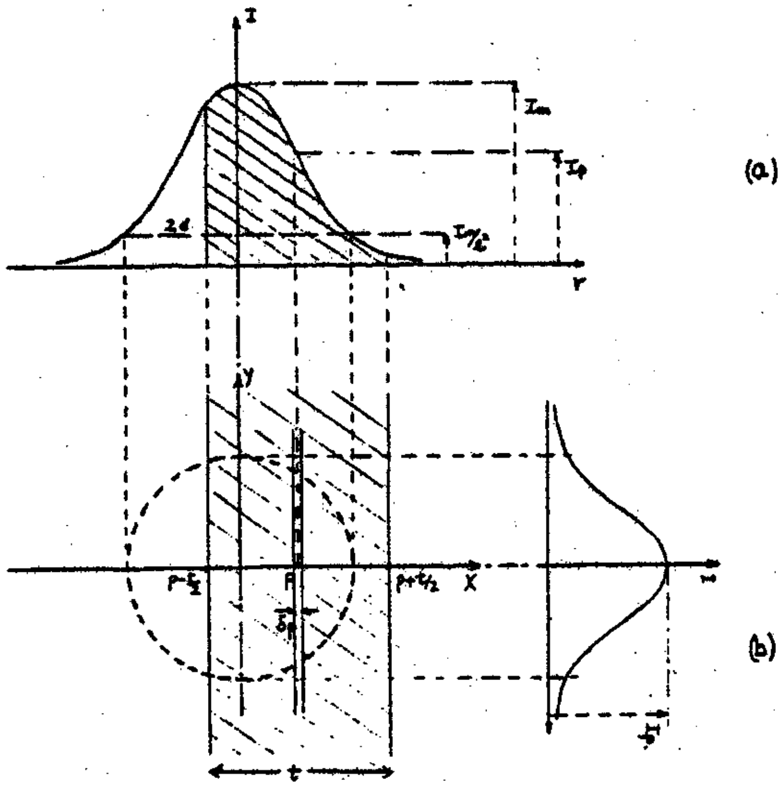


Figure 5.2 Line illuminated by a Gaussian beam

Consider the detected signal amplitude resulting from light transmission through a transparent line on an opaque background, lying in the field of coverage of the light spot, as shown in figure 5.2(b). It can be assumed that the detected signal amplitude depends on the incident power on the detector, or in turn on the power transmitted through the line. In addition, the line can be assumed to be straight over a distance of at least three times the diameter of the spot. Resolving in the plane of the line in directions X normal to the line and Y parallel to it, we can consider the power transmitted through a narrow slit parallel to the Y axis at a position p along the X axis and of width δp. The power transmitted through the slit is then proportional to the area under the Gaussian intensity profile with peak value Ip, and to the slit width δp. As the standard deviations of all parallel intensity profiles across the spot are equal, the area under this profile is proportional to Ip. Thus we have:

$$P(\text{slit}) = Ip \times \delta p \int_{-\infty}^{\infty} \exp(-y^2/2\sigma^2)\, dy \quad , \text{ where } \sigma = d/4$$

Consequently for a line of finite width t, placed with its centre a distance p from the Y axis, the total power transmitted through the line will be proportional to the area under the X axis intensity profile between the limits p-(t/2) and p+(t/2). That is:

$$P(\text{line}) = \int_{p-t/2}^{p+t/2} Ip \int_{-\infty}^{\infty} \exp(-y^2/2\sigma^2)\, dy\, dp$$

It is convenient to consider the transmitted power in terms of the ratio between line width t and spot diameter d, in order to keep the analysis general. Normalising the total cross-sectional beam power to unity, and using tabulated values for areas under the standard Gaussian curve, the power transmitted through a line of width t at a perpendicular distance p from the centre of the spot, can be expressed as a fraction of this maximum transmissible power. Figure 5.3 shows curves representing this power for various line width to spot diameter ratios. The distance p is measured in terms of spot diameters. Thus, for example, for a line of width half the spot diameter (that is, t=d/2), lying with its edge on the spot centre (that is, p=d/4), the detected signal amplitude level will theoretically be 48% of the maximum level. If the same line lies with its centre on the spot edge (that is, p=d/2), the detected signal amplitude level will be 16% of the maximum. When considering space detection, interpretation of these amplitudes is also direct; the signal amplitude is simply reflected about the 50% level. Thus a space of width half of the spot diameter, lying with its edge on the spot centre, will generate a signal amplitude level of 52% of the maximum, while the level from the same space lying with its centre at the spot edge will be 84% of the maximum. Signal amplitude levels for spaces are shown in parentheses on the vertical axis scale.

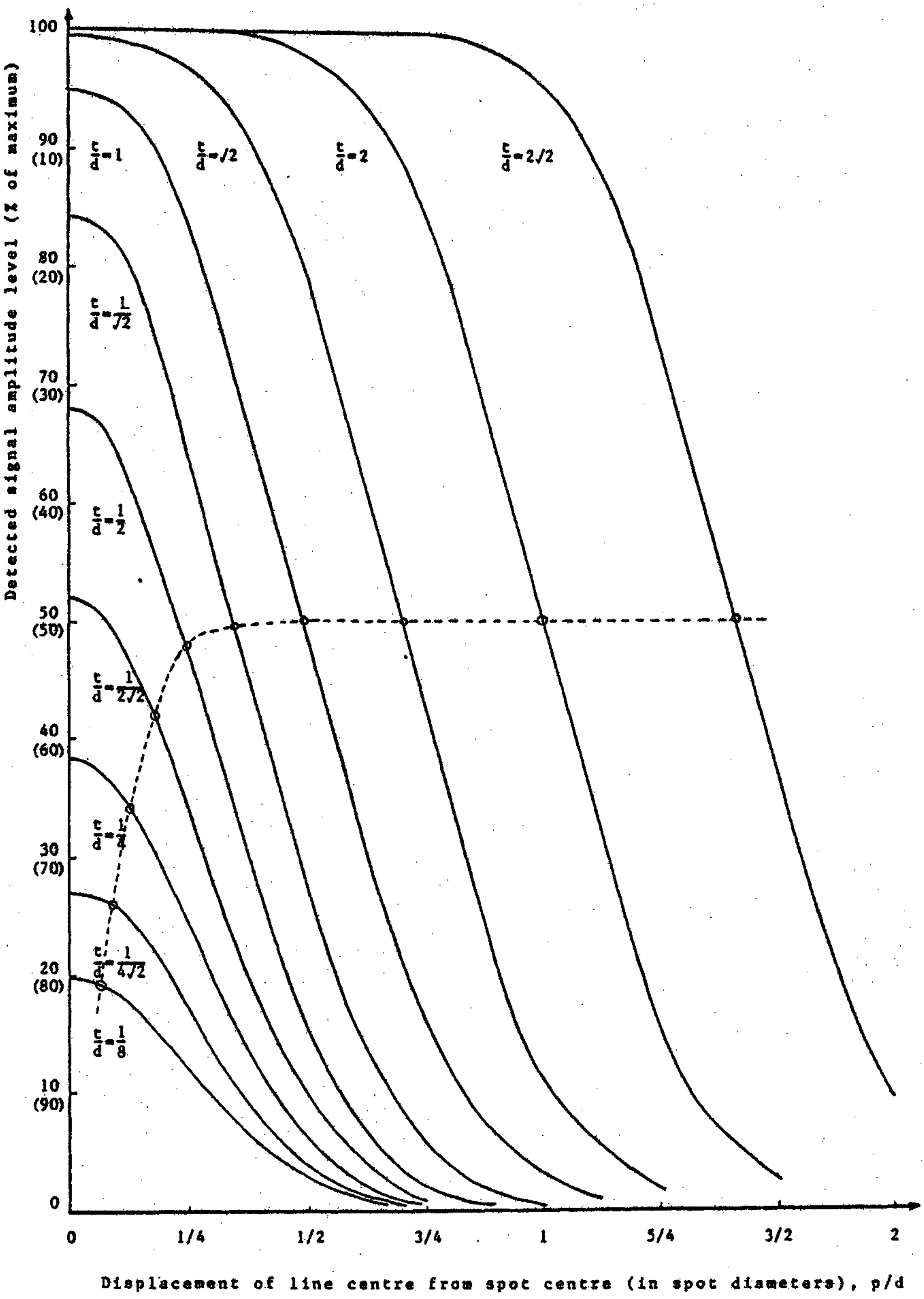


Figure 5.3 Optical power transmission through a line illuminated by a Gaussian beam

The preceding analysis is independent of the angle of the line with respect to the sweep and travel directions; it refers to the signal detected at a specified sample location from any line of width t whose centre lies a distance p from that location. If a sweep runs along a line, the curves of figure 5.3 indicate the steady specific detected signal amplitude obtained, as a fraction of the maximum possible, depending on the position of the line, the line width and the spot size. If the sweep crosses the line normal to its direction, a detected signal pulse, whose shape is given exactly by the appropriate curve of figure 5.3 and its reflection in the amplitude axis, will be obtained. In this case also, the horizontal axis is a measure of distance from the spot centre in terms of spot diameters, or in the real case where the spot is moving, of the distance of the spot from the line centre as it moves across the line.

Having then established a quantitative means of predicting the detected signal amplitude, based on line width, spot diameter and line position, the effects of thresholding and the consequent detectability of lines or spaces can be examined.

5.3.2 Effects of thresholding

The curves of figure 5.3 can be used to determine the threshold value at which digitization should take place if true line widths are to be reproduced. For each plotted curve, the level at which thresholding would return the true line width is shown encircled. As t/d becomes greater than 1, this level approaches 50% of the maximum. Thresholding below the 50% level results in line widening, while thresholding above it results in line narrowing, or space widening. The curves of figure 5.4 show these widening effects over varying line or space width to spot size ratios, for threshold levels between 20% and 80%. As in figure 5.3, percentages in parentheses refer to threshold levels to give space rather than line widening factors. For example, if a line width were half the diameter of the spot, detection at 50% would result in an effective line width of 92% of the true width, if the spot were centrally placed over the line at the instant of sampling. Referring to figure 5.3, if the line were located a distance of half a spot diameter from the spot centre at the instant of sampling, a threshold level of 16% would be required to ensure detection; this would result in substantial widening of lines, and hence narrowing of spaces, as can be seen from figure 5.4. For a space half the spot diameter, thresholding at any value lower than 32% will result in a loss of space detection entirely.

In fact, thresholding at a precise low level is not practicable for several reasons. First, the power output

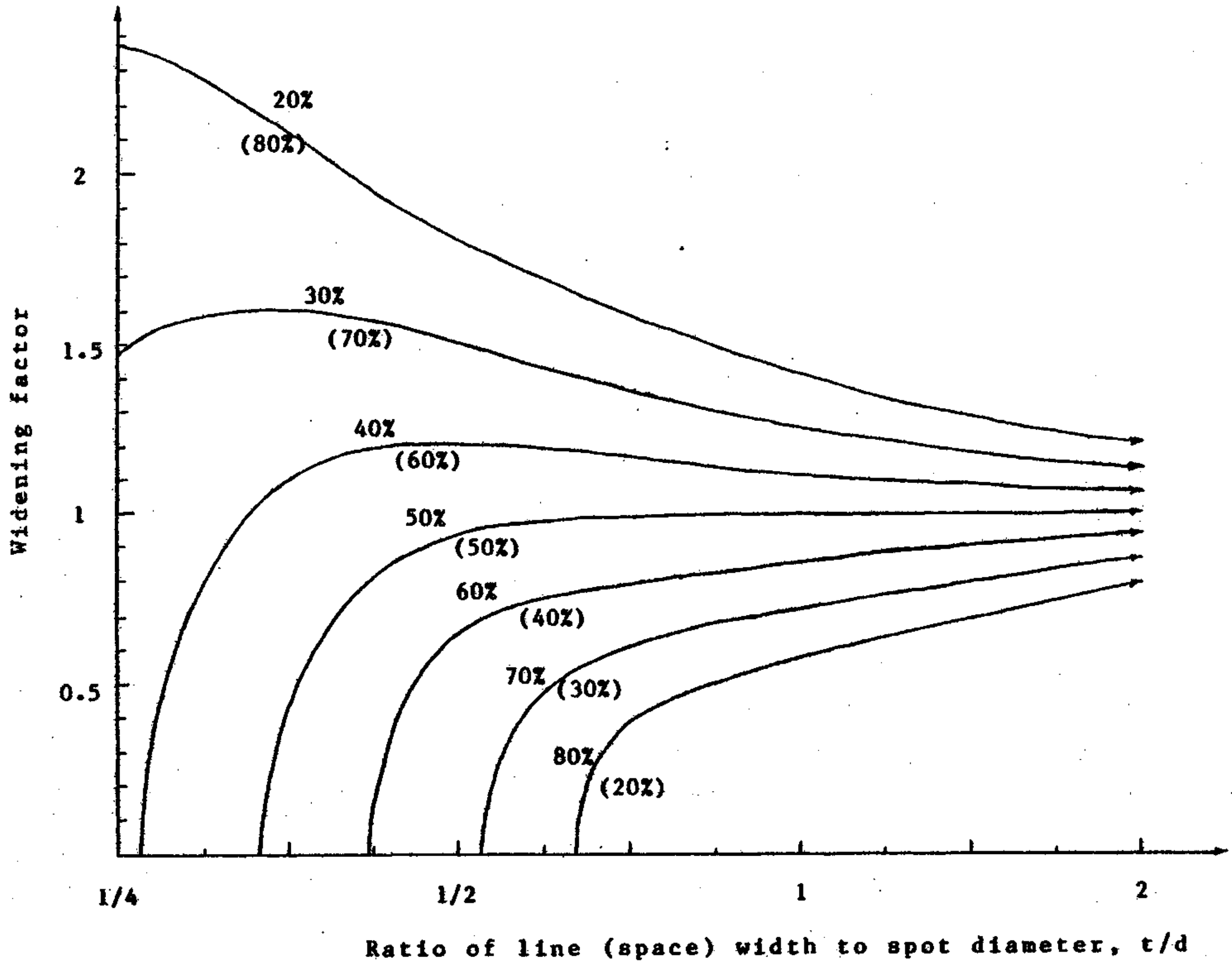


Figure 5.4 Line / Space Widening Effects

from a simple helium-neon laser is not a constant function of time; rather it can vary appreciably due to its random polarisation. Secondly, there exist other unavoidable non-linearities in the optical transmission and detection systems, such as diffuser characteristics, or the fact that the angle of incidence of light on the detector varies over the sweep between 0 and 180 degrees. This suggests that there may be a substantial indeterminacy in correlation between transmitted light power and detected signal amplitude. As a result of this, fine adjustment of the threshold level in terms of signal amplitude is not practicable, and hence absolute prediction of the detection or otherwise of narrow lines is not really possible.

Consequently, while it is feasible to somewhat lower the threshold level from 50% to enable detection of narrow lines which are located fairly close to the spot centre at the instant of sampling, reducing it substantially to detect lines some distance from the spot centre is neither feasible in practice, nor desirable if a reasonable degree of data integrity is to be maintained. Thus if a narrow line lies between two sample points, and is not sufficiently close to either to generate a signal amplitude higher than the chosen threshold level, it will not be detected. It is apparent then, that the choice of a compromising threshold level should not be made without considering the influence of sample spacing on the detection or otherwise of narrow lines. This is investigated further in the following section, and in addition, the effectiveness of a method of detecting narrow lines by sampling at more frequent intervals, as an alternative to low thresholding, is examined.

5.3.3 Effects of sampling

The signal stream generated as the spot sweeps the map sheet represents the density profile across the sweep to a degree of accuracy determined by the spot size, the performance of the detector and amplifier, and the threshold level for digitization. Time quantization is the process by which the signal stream is sampled at fixed intervals to generate a digital synchronous pulse stream. As a result of the quantization noise introduced by this sampling process, some information about the true width and position of lines is lost. For the scanner constructed the sampling interval is 2 microseconds, corresponding to a spacing of .004".

A digitized line in the reconstructed data matrix of scanned data should be connected throughout its length; that is, it must be possible to follow the line along its length by joining set pixels (sample points where the line has been detected) which are either directly adjacent or diagonally adjacent. Section 7.3.2 of chapter 7 looks at adjacency and connectedness more exactly. Consequently, if a line makes some angle θ to one axial direction, assuming the line to be straight over a distance of several resolution spacings, an effective width of $t/\sin\theta$ is presented along that axis, and $t/\cos\theta$ along the orthogonal axis. Thus connectivity is most prone to break down when $\theta = m(\pi/4)$, for integer m; that is, when the line lies parallel to one axis. In this case a minimum effective width is presented in one axial direction, and diagonal adjacency cannot occur; this is then the case for which the connectivity of narrow lines is most likely to be broken. It is this worst case detectability which will be considered in the following discussion.

The curves of figure 5.4 show the effects of line widening or reducing as the ratio of line thickness to spot size varies, for specified threshold level values. For any line thickness to spot diameter ratio, provided that the spot diameter to sample spacing ratio is known, the probability of worst case line detection (as defined above) can be obtained from the effective line width (true width X widening factor) and the sample spacing. The curves of figure 5.5 show detection probability curves for lines lying parallel to one resolution direction for threshold levels between 20% and 80% in 10% steps. The horizontal axis gives line thickness to spot diameter ratios, while the vertical axis is scaled in terms of the spot diameter to sample spacing ratio d/s. For a given ratio of d/s, the axial

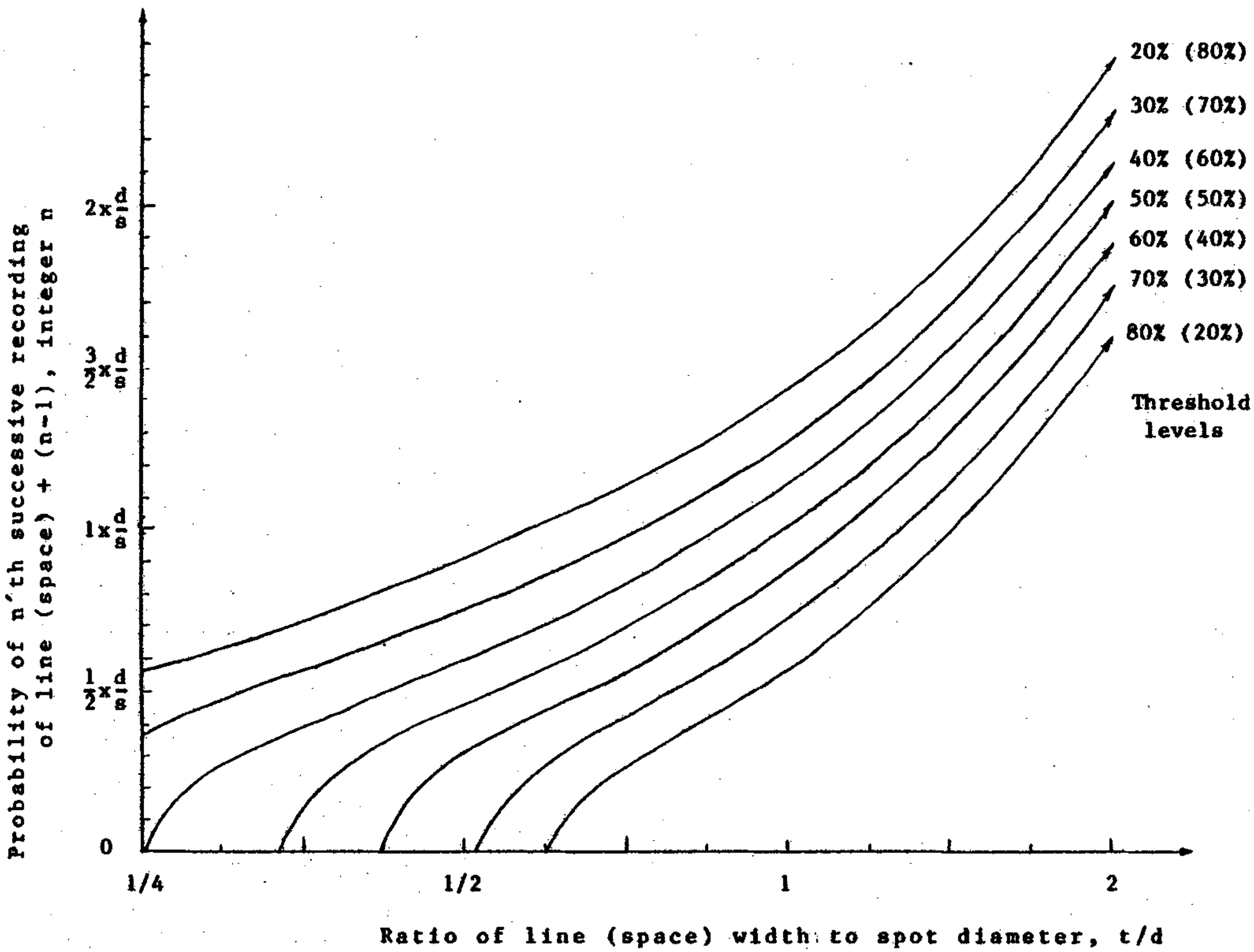


Figure 5.5 Worst Case Line / Space Detection Probability

scale is defined, and between any pair of integers (n-1) and n for n=1,2..., if a value of 0 is assigned to (n-1), and a value of 1 assigned to n, then the intermediate values between 0 and 1 give the probability of the n'th successive recording of a line or space in the sampled data stream in a direction perpendicular to the line or space orientation. Thus if the spot size were equal to the sample spacing (d/s=1) using a 50% threshold level, a line of thickness .002" (t/d=1/2) orientated parallel to one resolution direction, would have a detection probability of 0.46. If d/s=2 for the same line width (t/d=1/2 again), detection probability is .92, while if d/s=4, probability of detection once is 1, and of detection in two successive samples taken in a direction perpendicular to line orientation is .84. It should be noted that for lines lying truly parallel to one resolution direction, entire lines will either be detected in full, or not at all. However, this is unlikely to occur often in practice; narrow lines are generally contours, and hence not straight but rather have on average random orientation at any point. Thus such lines of this thickness will appear broken at irregular intervals, depending on the chosen threshold level. If the threshold level is chosen so that the worst case probability of detection is 1 for a given line thickness, all lines of this thickness will be connected.

One possible way to improve the probability of detection of narrow lines would be to decrease the sample spacing. Since a data matrix is required whose elements represent information only at the originally specified resolution spacing of .004", if sample spacing is reduced, some aspect of the information contained in the matrix representing the data at the higher resolution would have to be contracted into the samples of the required matrix. For example, if sampling were to be performed at twice the specified resolution, four samples would have to be contracted into one before entering data to the computer. If a logical 'OR' of the four samples were taken, narrow lines would have a higher detection probability. Figure 5.6 shows how a line detected at any of a group of four points would be recorded at a single sample, which would be effectively placed at the centre of the four points. This in fact corresponds to a situation in which sampling takes place at only this point, but the spot size and intensity profile are modified to those shown in figure 5.6; the profile is no longer a single Gaussian, but rather a non-additive combination of four Gaussians. For the previous worst case lines of width .002" (which under straightforward sampling at .004" would have a detection probability of .46 when thresholding at 50%), decreasing s by a factor of two while still thresholding at 50% now results in a detection probability of .92. Worst case space detection probability will be zero unless the effective space width is greater than 2 sample spacings.

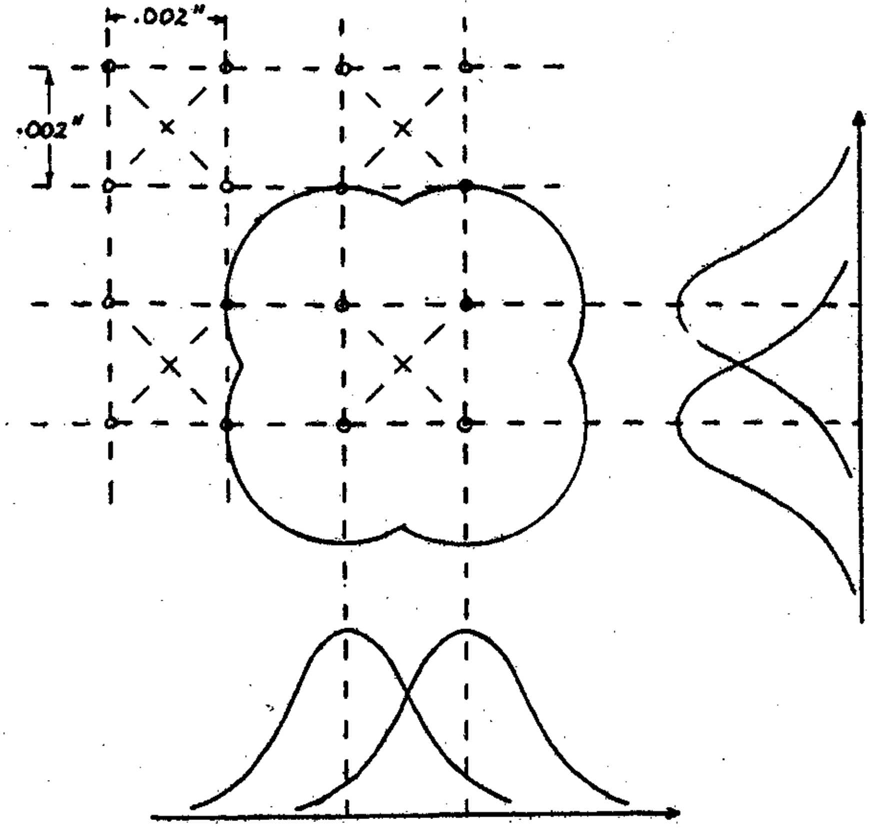


Figure 5.6 Effective spot shape and profile under alternative sampling scheme

Although such sampling might be desirable for some types of data, with the constructed scanner geometry and motor speeds, sampling between lines is not directly possible. Sampling between successive serial data points is, however, quite possible. By sampling every 1 microsecond, or .002", and performing a logical 'OR' on successive pairs of samples, narrow line detection will be improved. However, resolution processes in the sweep and travel direction will now no longer be identical. A line running parallel to the sweep axis will have a worst case detection probability given by the curves of figure 5.5, while one lying perpendicular to it will have a higher worst case detection probability. This suggests that it might be desirable to perform the same sampling operation in the other resolution direction. Implementation of such a scheme would mean sweeping at half resolution spacing and temporarily storing, say, each odd indexed line of data. For each incoming even indexed line of data, a logical 'OR' operation could be performed between identically time co-ordinated positions in stored and incoming lines, and the resulting single stream would give the required information. This sampling scheme could be implemented in the following manner.

1. Sweep at .002" intervals, and perform a logical 'OR' operation on successive serial pairs of samples.

2. Temporarily store each odd indexed line of data.

3. Perform logical 'OR' on corresponding samples from incoming even indexed line (in real time), and stored odd indexed line.

The result would be a sampling process which would detect information at .002" spacing, and would re-position it onto a .004" spaced grid. It should be noted that since the final sample value derived from the values at each of the four points is obtained by logically 'OR'ing these values, the order in which pairs of the four are 'OR'ed together does not affect the result. If 'OR'ing is performed first on successive pairs of samples from each sweep, the size of the temporary storage buffer required is 8 kbits rather than the 16 kbits that would be required if pairs of corresponding samples from successive sweeps were first 'OR'ed.

Step 1 could be achieved in either of two ways. The first would be to use the scanner and timing arrangements as they stand, but have a dual faced rotating mirror instead of a single faced one. The generated data stream would simply be sampled at 1 microsecond intervals, or at 1 MHz, and the logical 'OR' performed between successive pairs directly on clocking of the second sample of each pair. Storage of alternate lines of data would require a synchronously clockable buffer of size 8 kbits, or 1024 bytes, together with some fairly simple control logic. The third step could be achieved by connecting the temporary storage buffer as a ring buffer (or alternatively using a first-in first-out stack), and clocking it in synchronism with the sampling of the incoming data stream. The required logical 'OR' could then be taken on the next clock cycle, and the incoming even-indexed line of data discarded. The signal would then represent the required samples at .004" spacing. The input data rate to the computer would not be affected. The second way requires more additional buffer space than does the first, but does not require a dual faced mirror. The mirror rotation rate would be doubled, and lines time quantized at 2 MHz to generate the samples at .002" spacing along the sweep. The stream at 1 MHz resulting from 'OR'ing of successive serial pairs of samples would be stored temporarily, and stored and incoming lines similarly 'OR'ed. This has the disadvantage that the final data stream would be produced at 1 MHz for 8 milliseconds, rather than the designed 500 kHz for 16 milliseconds. Continual input at this rate is beyond the capabilities of the computer, so an additional buffer of roughly half a line length (that is, 4 kbits, or 256 16-bit words) would be required.

It was decided that in addition to straightforward sampling at .004" spacing, partial provision for narrow line detection would be made by implementing as an alternative sampling scheme the logical 'OR'ing of successive pairs of samples at .002" spacing in the sweep direction. Section 5.5 of this chapter descibes the hardware constructed for this purpose.

5.3.4 Bandwidth of detector and amplifier

If either the detector, or the detector amplifier, has a limited bandwidth, a shaping effect will take place on the detected pulses. For an input step function, a limited bandwidth will impose finite rise and fall times. Depending on the rise and fall time relationships, and the threshold level chosen for detection, the output signal will be delayed with respect to the input, and may be either shorter or longer in duration. Although not necessarily the case, in general the rise and fall times of the detector and amplifier will be roughly equal. Consequently, if the threshold level is chosen at about 50% of the signal amplitude, only a delaying distortion occurs. However, it has already been noted that depending on spot size, line width and other less quantifiable non-linearities in the system, pulses may vary substantially in amplitude. Maintaining a constant low threshold level means that a pulse of smaller amplitude will be foreshortened due to the rise time tail.

Since it seems desirable to avoid introducing yet another non-linearity into the detection process, the bandwidth of the detector and amplifier should be sufficiently large that the pulse shaping effect of rise and fall times be minimal compared with the effects due to spot size. As it is the difference between delays caused by the rise and fall times which introduces the undesirable distortion, it seems reasonable to ensure that this difference is of an order of magnitude at least less than the sampling interval of 2 microseconds. An overall detector and amplifier bandwidth of about 1 MHz would satisfy these conditions, and hence seems a reasonable aim.

5.3.5 Conclusions

The preceding four sections suggest that in the interests of maintaining both a fixed degree of data integrity, and consistency of resolution, the distorting effects of any techniques used to obtain information at a higher spatial resolution than that specified be carefully examined.

The detection probability curves of figures 5.5 can be used to determine the effects of spot widening on worst case line and space detection. However, interpretation is not particularly easy if spot diameter variations are to be investigated. If the sample spacing is fixed (at .004"), the detection probability of worst case lines of a specific width can be plotted for varying spot diameter. This is shown in figure 5.7, where a line width of .002" was chosen, this being the minimum width for which detection is desired. It can be seen that depending upon the threshold level, there is a spot diameter which gives maximum widening of .002" lines. At a 40% threshold level, this maximum occurs for a spot diameter of approximately .004". For a spot diameter much greater than .004", worst case space detection is very poor. If the spot diameter is less than .004", line widening is not substantial, giving low detection probability for worst case lines of this width. In fact, with the optical system used, .004" was about the minimum spot diameter obtainable.

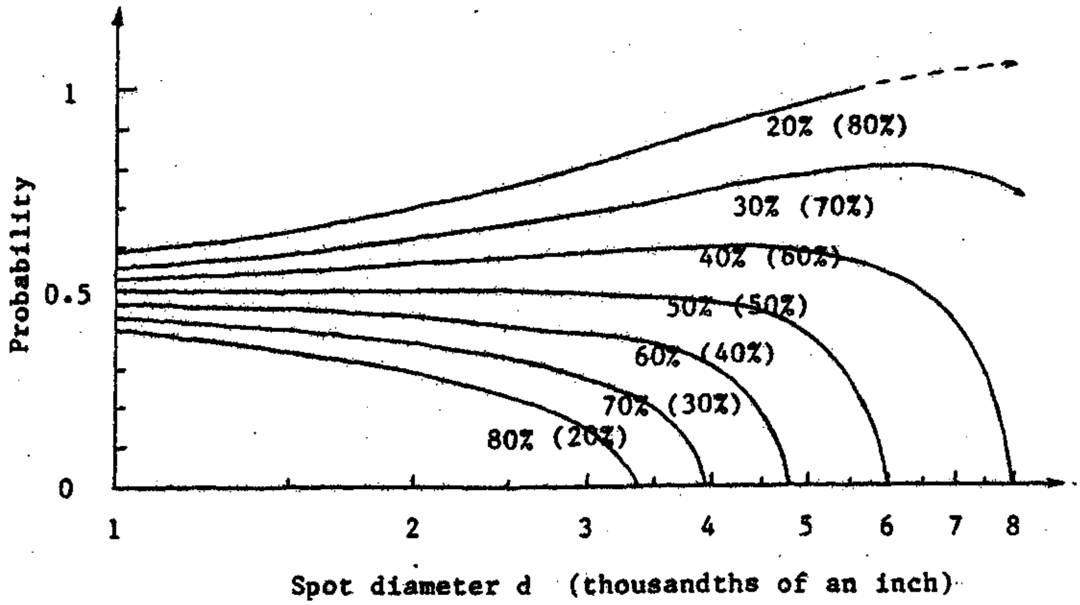


Figure 5.7 Worst case detection probability for lines/spaces of width .002"

It seems that a reasonable compromise is to use a spot diameter of .004", and to threshold the signal at approximately 30% of the maximum level. This gives a .75 detection probability for worst case lines of width .002", and, referring back to figure 5.5 with d/s=1, a detection probability of .72 for worst case spaces of .004". Worst case lines of width greater than .003" will always be detected. This ensures that randomly orientated lines of width .002" on the map will seldom have connectivity broken.

The possibility of sampling at an increased resolution, and thresholding nearer to the 50% level should not be overlooked. Sampling at half the final resolution grid spacing in the sweep direction does not pose problems, as described in section 5.3.3, but to do so in the travel direction requires additional hardware. To avoid this, it might be possible to ellipticise the spot, with minor axis in the sweep direction so that resolution in this direction is maintained, but the same arguments about thresholding still apply, as the intensity distribution along the major axis of the ellipse would still be approximately Gaussian. It seems that some technique, such as spot wobbling along an axis perpendicular to the sweep direction, would be required. This has the advantage that a small spot can be used, and an intensity profile which is very much squarer in nature than a Gaussian is obtained along the axis of wobble. Spot wobbling could be used in the sweep direction also; this would avoid the peaky spot intensity profile resulting from intermediate sampling. However, providing that the connected detection of lines of width .002" is the only criterion not met by digitization at .004", if this can be achieved by electrical rather than mechanical means, this seems more attractive.

Based on these considerations, and bearing in mind that the system is a developmental one and hence avoiding unnecessary mechanical complexity, it was decided that sampling at half the original resolution spacing in the sweep direction, as an option to straightforward sampling, should be implemented. This technique generates digital samples at twice the specified resolution, and any particular logical extraction of information may be obtained, depending on the characteristics of the map sheet, to generate the final samples placed on a .004" grid. Implementation of this scheme in the sweep direction gives a degree of ensurance of detection of worst case narrow lines without requiring a very low threshold level, while under normal operation the intermediate samples would be ignored.

Having then established the requirements of each section of the signal processing hardware, the following sections describe their design and construction in detail.

5.4 Signal Detection

A PIN photodiode was chosen as a suitable optical to electrical transducer. Section 5.3 of this chapter determined that the bandwidth of the detector/amplifier combination be greater than 1 MHz ; this corresponds to a detector rise time of better than about 200-300 nanoseconds. The required sensitivity is less easy to determine, as the absorptivity of the plastic map bed and the transparent areas of the map sheet, and the effect of the diffuser are difficult to quantify. A proportion of the power in the laser beam will be reflected from each surface, and in particular the collection of the incident power on the diffuser will depend upon the the detector positions within the diffuser, and the nature of the diffusive material. The large dynamic range of a PIN diode, coupled with its low noise threshold level detection performance, make it electrically suitable for the purpose. Although a photomultiplier tube would also be electrically suitable, the small physical size of a PIN diode makes it very much more attractive in the chosen geometry of map trolley, reflected signal path and detector position.

In fact a combination of three diodes within the diffuser was used. In order to obtain better linearity of response over the entire width of the sweep, in addition to a centrally placed diode, diodes were mounted at each end of the diffuser cylinder. Small bumps in the reflecting surface, and the tilt of the reflector to keep the detector out of the path of the incoming beam, cause the reflected beam to strike the diffuser non-centrally; multiple diodes positioned as described improve the linearity to within acceptable limits.

In order to obtain a high speed of response from a PIN diode, the diode must have a small active surface area.

Unfortunately such a diode is not the most ideal for detection of diffused light, and consequently signal amplitude levels will be low. The Hewlett-Packard HP 5082-4203 diode was chosen for its wide directional sensitivity characteristics, and also for its high speed, high sensitivity and low noise properties.

A PIN diode can be operated either under zero bias, or under reverse bias. Under zero bias it behaves as a voltage source and operates in photovoltaic mode, while under reverse bias it operates in photoconductive mode and acts as a current source. The characteristics of the HP 5083-4203 PIN diode are given in appendix B. Fast response requires that the diode be operated under reverse bias at the sacrifice of noise performance. Specifically, a load resistance of less than about 100 kohms with low capacitance is necessary if the rise time is to be less than 200 nanoseconds. Consequently the photoconductive mode of operation was chosen. A transconducting amplifier is required to render the generated signal suitable for voltage thresholding, and this is treated in the following section.

5.5 Signal Amplification

The detector amplifier must have high gain, fast response and low noise performance. In addition, since steady d.c. signal levels may be encountered when the sweep runs along a line on the map, a wideband amplifier is necessary. The combination of these requirements suggests that the input stage of the amplifier be discrete. Although hybrid operational amplifiers with input offset and bias currents of the order of nanoamps are available, to achieve a wide bandwidth and low noise performance at high gain is not feasible. In the constructed detector, generated currents were of the order of tens to hundreds of nanoamps, and for a voltage output of several volts, a transconductance of the order of 10^7 is required. In this section, the design, construction and performance of the first stage of the amplifier, which converts the low level input current into an output voltage of reasonable level, are described. The following section describes thresholding, and the amplification and d.c. level removal necessary to make the threshold level immune both to thermal drift in the first stage, and to the density of the base material of the map sheet.

The circuit of the first stage of the amplifier is shown in figure 5.8. A low noise NPN transistor (2N 2484) was chosen for its ability to amplify very low level currents. Following this is a voltage gain stage; a PNP transistor (2N 3251) is used so that thermal effects will cause opposite d.c. drifts in the first two stages, tending to reduce overall drift. Finally an emitter follower buffers the amplified signal.

The wideband noise vs. source resistance and noise figure vs. frequency curves for the input transistor are shown in appendix B. In order to keep the noise figure below 1 dB, a source resistance of 10-12 kohm was chosen, and a collector current of 100 microamps compromises between bandwidth and noise, giving a value for R3 of 100 kohm. R3 is the source resistance for the second stage transistor; the noise figure vs. collector current characteristics for this transistor are also shown in appendix B; with R3 = Rs = 100 kohm, the noise figure is less than 2 dB if Ic < 70 microamps and less than 1 dB if 20 microamps < Ic < 30 microamps. Again in order to keep the gain bandwidth product high, the higher collector current was chosen at the slight expense of noise performance, since for this stage noise is less critical. A collector current of appoximately 70 microamps requires a collector resistor of about 100 kohm. This value was also chosen for the collector resistor of the emitter follower, giving a collector current of approximately 100 microamps. Adjustment of R2 and the feedback resistor Rf allows the damping of overshoot and provides for fine offset balancing. In fact R3 was trimmed to 94 kohm to bring the d.c. output offset to zero for no input current.

The amplifier was built on a ground plane to minimise noise pickup, and careful power supply bypassing was employed. The entire first stage was also shielded. Figure

5.9 shows the performance of the amplifier for a square wave test signal input through a 10 Mohm resistor, giving an input current of 50 nA. Overshoot damping is dependent on the resistance and capacitance of the input source. Overdamping increases the rise and fall times of the amplifier, while underdamping causes longer pulse settling times. The PIN diodes have a combined reverse bias resistance of greater than 10 Gohm, and a combined capacitance of about 6 pF; these conditions should be met as closely as possible when adjusting for minimal overshoot. Ideally, final trimming should be performed using a modulated light source to create a pulsed input for the amplifier under its actual operating conditions, but in practice overshoot damping was set to give minimum signal rise and fall times without visible distortion, using the smallest achievable laser spot diameter crossing opaque to transparent boundaries.

The constructed amplifier has the following measured characteristics.

Transconductance = 2×10^7

Bandwidth = 2.3 MHz

Noise at output (grounded input) = 0.6 mV rms

Signal to noise ratio (for 25 nA rms input) = 30 dB

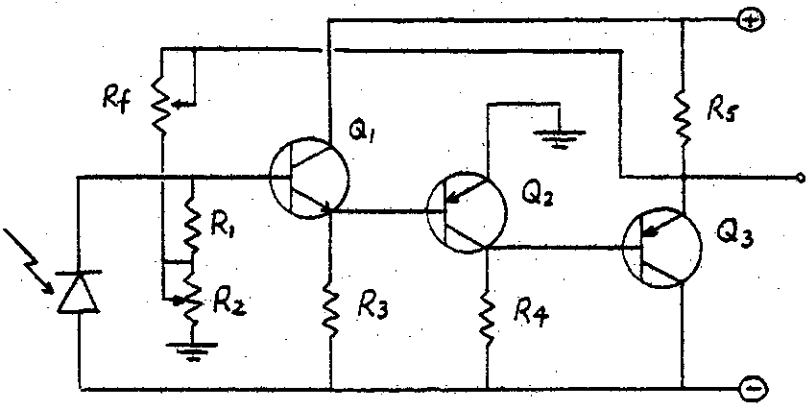
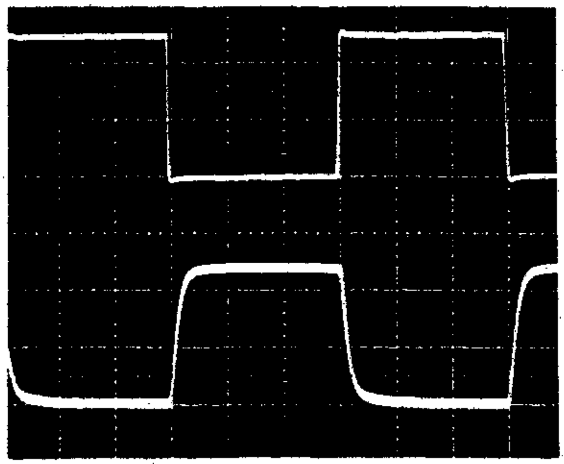
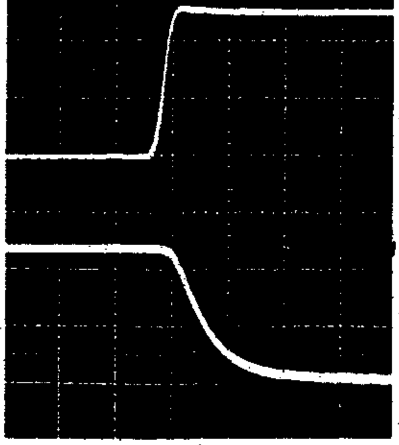


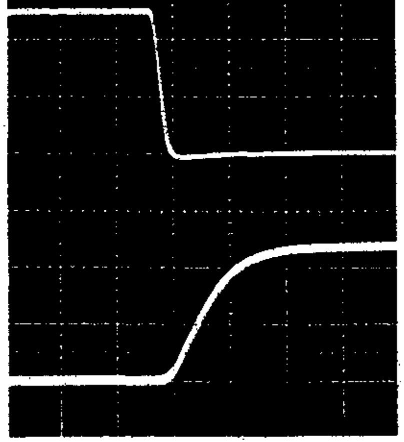
Figure 5.8 1st stage of detector amplifier



(a)



(b)



(c)

Figure 5.9 Detector amplifier test performance

5.6 Threshold Detection

It was decided that the threshold level for digitization of the analogue signal should be independent of the steady d.c. voltage signal level for two reasons. First, one problem encountered when a discrete first stage is used for signal amplification is that the d.c. output level varies with temperature. Because the transconductance gain is so high, this d.c. shift can be appreciable compared with the signal level, and this makes maintaining an exact threshold level difficult if the signal is d.c. coupled. Secondly, the opaque density of map sheets can vary depending on the base material thickness and type. Since the diodes are capable of detecting extremely low light levels, a less dense sheet introduces a small d.c. offset; the threshold level should be independent of this.

As noted earlier, it is convenient to have a d.c. signal available for static adjustments, but under normal operating conditions this is not necessary. Therefore the following method was used to make the threshold level independent of the d.c. component of the signal. An intermediate gain stage is used to bring the signal level to about 1v maximum, following which a buffer amplifier isolates a passive low-pass filter from the intermediate stage. This filter tracks the d.c. component of the signal, provided that no appreciable energy is contained in the signal pulses representing light transmission through the map sheet. As the pulses occur on average infrequently over the sweep, and not at all for 3/4 of the sweep cycle time, this is in fact the case. This d.c. voltage can then be subtracted from the d.c. signal to generate the required signal which can then be amplified to a level suitable for thresholding. Keeping the signal amplitude after the intermediate stage to about 1v maximum ensures that even for d.c. drifts of several times the signal amplitude, no clipping will occur. This allows the possibility of scanning map sheets of variable base densities without hardware adjustment. LM 318 operational amplifiers, with a unity gain-bandwidth product of 15 MHz were used, and trimming capacitors control overshoot.

Actual threshold detection can now be made with respect to a fixed reference level. This level is significant as far as signal integrity is concerned, and is hence adjustable. The amplified pulses represent a measure not just of the width of the line on the map, but also of the laser spot characteristics as described in earlier sections of this chapter. Hence the threshold level represents some non-simple measure of the instantaneous signal level at which the system is to specify a recorded mark on the map. Due to the large number of influencing factors involved, adjustment of this level must be made according to the characteristics of the data being scanned.

An LM 311 comparator was used for threshold detection. Hysteresis was employed to improve noise immunity, as the comparator is capable of changing output state for very small input voltage changes. The resulting digital signal stream must be transmitted over coaxial cable to the time quantization circuitry mounted at the computer interface. TTL drivers were used for this purpose, and a buffer inverting gate isolates the comparator and low level signal circuitry from the relatively high switching currents used by the drivers. In fact, it was found that even with stringent power supply bypassing, the switching transients due to the drivers caused power supply spikes, and unnacceptable degradation of the analog signal, and it was necessary to use a separate power supply for the drivers. The driver chip was also mounted off the ground plane.

Figure 5.10 shows the complete signal detection, amplification and digitization circuitry, while figures 5.11(a) and (b) show its ground plane construction, complete with detector diodes and diffuser. Screening was removed for the photographs. Performance of each stage of this circuitry in actual scanning is shown in the final section of this chapter, and examined more carefully in chapter 8.

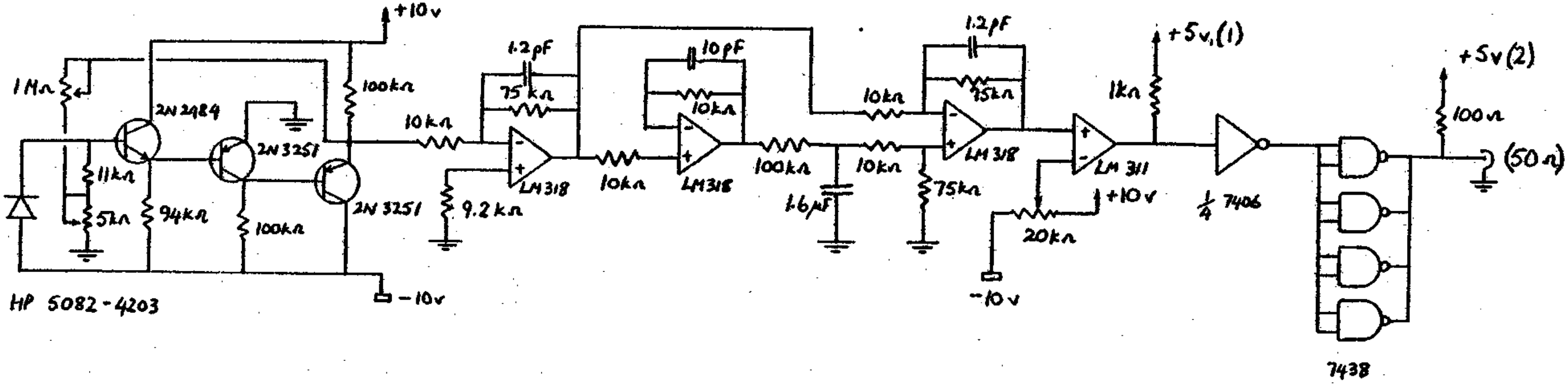
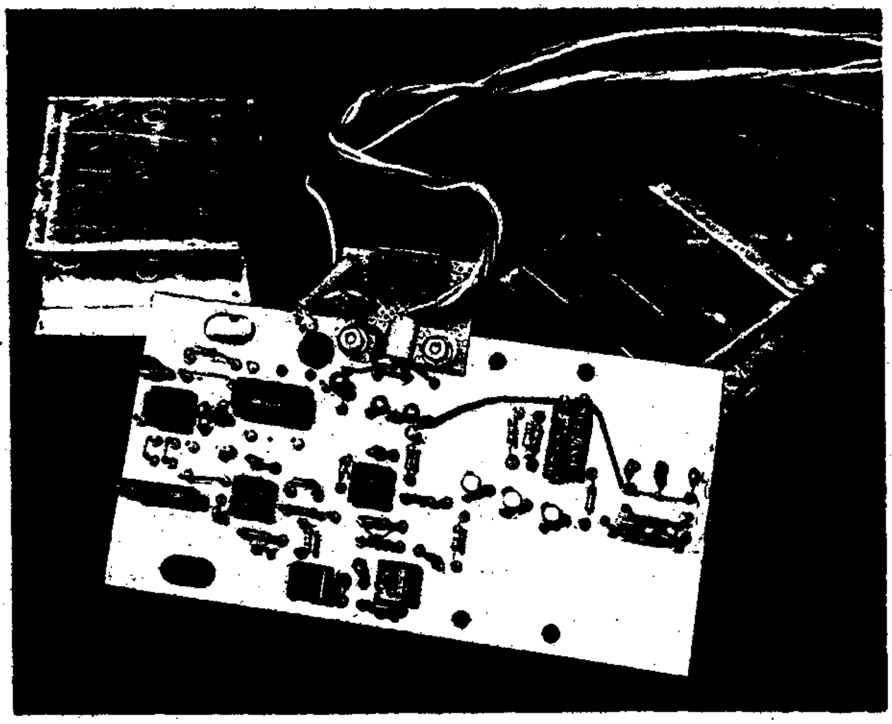
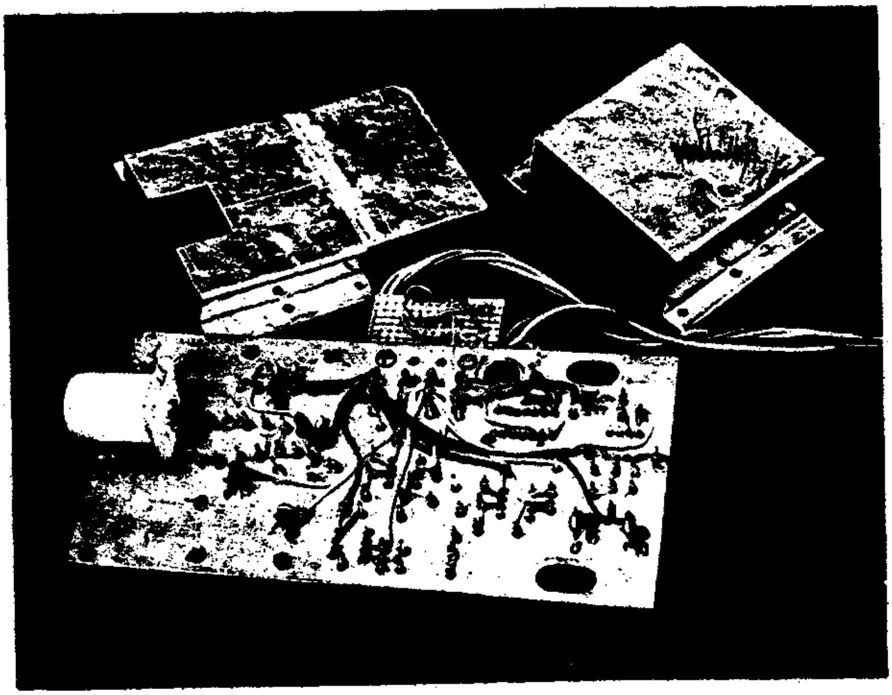


Figure 5.10 Signal detection, amplification and digitization circuitry



(a)



(b)

Figure 5.11 Signal detector/amplifier board

5.7 Time Quantization

The base 500 kHz sample rate is derived from the 1 MHz crystal controlled clock which generates the time base for the entire system. The samples are taken by clocking a synchronous flip/flop on the positive going edge of the 500 kHz clock signal, and the output is accepted into the serial-to-parallel converter and buffer (described in the following chapter) on the negative going edge of the same signal to ensure a determinate state at the output.

Included as a switchable option is the time quantization of successive serial pairs of samples taken at 1 microsecond intervals. Again in order to avoid any indeterminate states occurring due to sampling the output of a flip/flop too close in time to a state change, the inverted 1 MHz clock signal is used to generate two sampling frequencies of 500 kHz which are then 90 degrees and 270 degrees respectively out of phase with the base sampling frequency of 500 kHz described above. The logical 'OR' of the result of these sampling operations is directed to the input of the base 500 kHz sampling flip/flop. This then means that the resulting single sample effectively represents the central point between successive serial pairs.

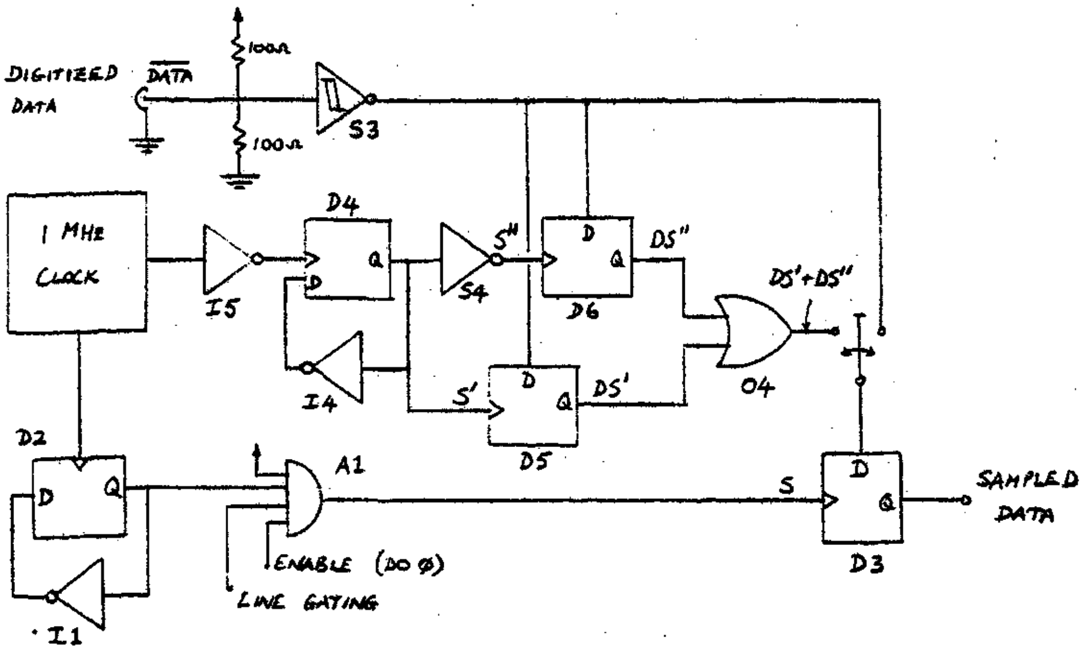


Figure 5.12 Logic to perform time quantization

Figure 5.12 shows the logic circuitry constructed to perform time quantization. Figure 5.13(a) demonstrates straightforward sampling on a data stream, while figure 5.13(b) shows the timing sequence for the sampling logic to implement successive serial pairs sample combining, and the result of this sampling on the same data stream. In the following section, results of both types of sampling on actual scanned data are shown. The time quantization logic was constructed on the interface board, the layout of which is given in appendix B.

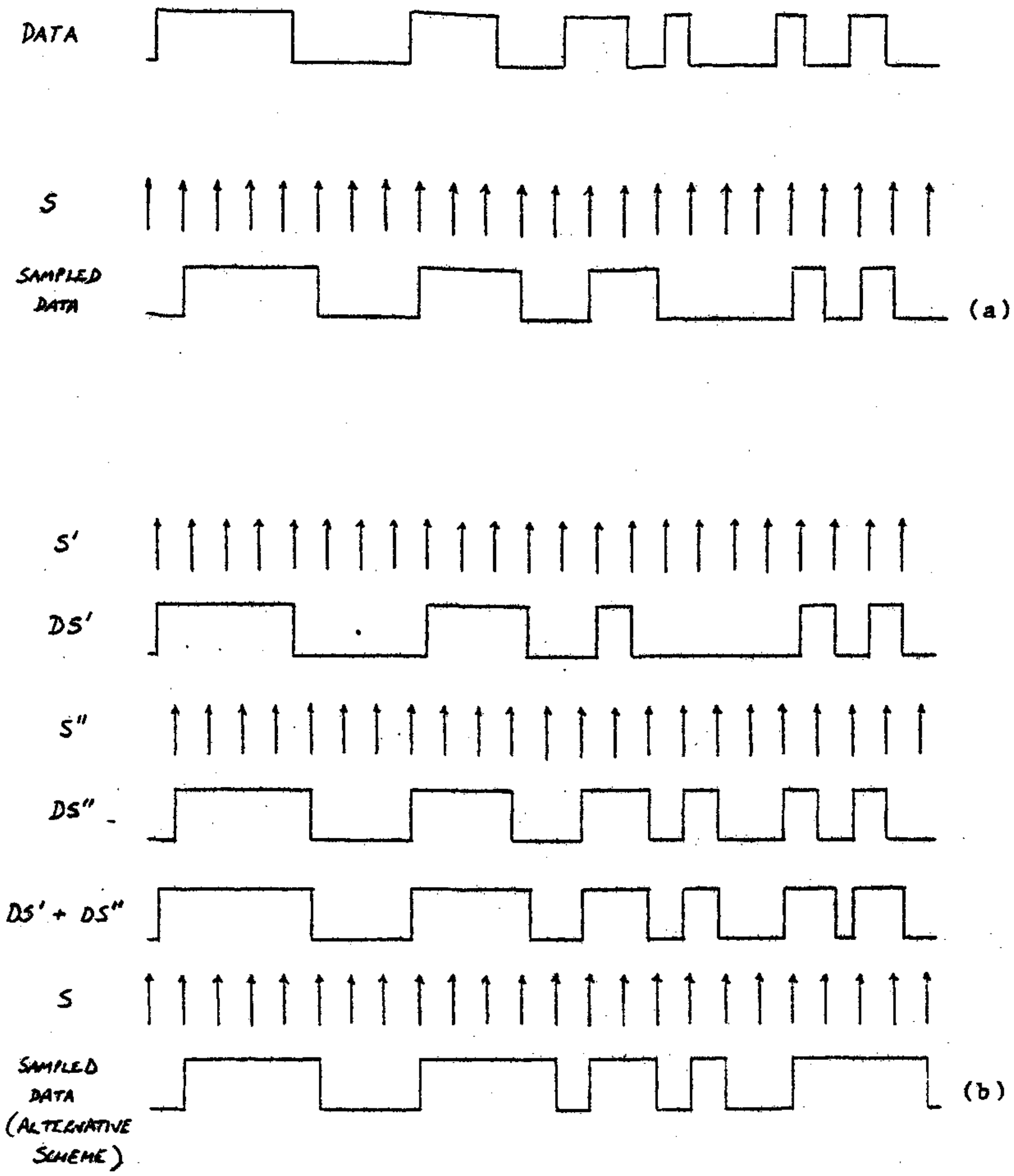


Figure 5.13 Data stream sampling under direct and alternative sampling schemes

5.8 Summary

This chapter has described the hardware signal processing required for the generation of a synchronous digital pulse stream that can be reconstructed into a matrix to represent the scanned map sheet in a non-exact manner. This non-exactness has been examined, and the signal processing method and circuitry designed to minimise its undesirable aspects, while making use of some of its features, have been described. Having obtained a serial synchronous data stream, the next step is to pass this data to the computer through the 16 bit parallel I/O port described in chapter 3. The hardware and software necessary to do this are described in the following chapter.

6. INTERFACE BETWEEN SCANNER AND COMPUTER

6.1 Introduction

The interface hardware accepts the time quantized serial pulse stream from the scanner and converts it into 16 bit parallel form suitable for entry to the computer. It also handles all necessary control and synchronising signals both to operate the scanner, and to pass the data to the computer. The interface software receives the incoming data upon hardware generated request, and transfers it through memory to a chosen storage medium (disk or magnetic tape). In addition, further software supervises and controls the scanning process. The software structure is such that other use can be made of the computer while scanning is in progress.

This chapter is divided into three main sections. The first examines the factors which govern the design of the interface hardware and software, while the second gives details of the interface hardware; namely, the buffering and control circuitry. The third section describes the software organisation developed to control the scanner and handle the incoming data.

6.2 Factors Governing Interface Design

The interface hardware must be designed to operate in conjunction with the software such that data throughput is continuous during the scanning process. Facility must be provided, either in hardware or software, to ensure that delays in computer response do not result in a loss of data. These delays can be of two types; first, the short delay in response from an input routine currently with control of the processor; and secondly, longer delays caused by interference from other processes taking place in the computer.

Facility for accomodating the first type of delay in response time must be provided in the interface hardware; specifically, by the provision of some small buffering capacity, together with suitable control signals to handle data transfer between the interface and memory in a handshake manner. However, so accomodating the longer response delays would require an excessively deep buffering capacity. This would be unsatisfactory not only on a cost basis, but also in that delays in response caused by operating system overheads can often be of the order of several milliseconds, rather than microseconds, and it would be impractical to attempt to temporarily store the data generated during this time. Consequently it seems more attractive to eliminate such delays by software means if possible, using both the routine running priority structure of the operating system, and any other necessary blocking of interfering processes during the data collection cycle. The following two sub-sections look more specifically at the hardware and software requirements for accomodating or preventing each of the two types of delays described above.

6.2.1 Accomodation of short computer response delays

Data samples are generated serially at the rate of one every 2 microseconds. The most straightforward approach to performing a serial-to-parallel conversion would be to use a set of shift registers, and simply enter data to the computer every 16 samples. However, this requires that the computer must be able to respond to an input request within two microseconds; it is not realistic to expect this from a program controlled input mechanism. Clearly some form of temporary storage is necessary, at least to store one word while the next is being generated. This would give the computer 32 microseconds to respond, which is quite feasible. (In fact tests showed that with a minimum of software overhead, the input routine could handle one word in approximately 19 microseconds.) Although a single word of storage (such as a set of 16 registers) would be sufficient to deal with short response delays, it turns out from requirements described in the next section that a buffer at least several words deep is necessary. Consequently it was decided that a 16 word deep first-in first-out (FIFO) stack should be used for the combined purposes of serial-to-parallel conversion and interface buffering. This is described in detail in section 6.3.1, and the necessary control signal generating logic to perform data transfer in section 6.3.2.

6.2.2 Accomodating and preventing operating system delays

As with most multi-user single processor operating systems, RSX-11M uses a defined hardware and software priority structure to determine where control of the processor is held at any time. Hardware interrupts have a priority higher than any software priority, and these interrupts are processed by transferring control to an interrupt service routine which executes at a high software priority. Consequently a running task can at any time be briefly suspended for the processing of a hardware interrupt. By convention, service routine processing time is usually limited to less than 100 microseconds; further required processing is performed at a lower priority by a subsidiary task set up from the service routine. Providing the guidelines for I/O driver operation are adhered to, delays during the running of a high priority task should not be longer than about 100 microseconds, and usually would be shorter. Since it is desirable to operate within these guidelines, the interface buffer should be capable of accomodating delays of up to 100 microseconds. This means temporary storage of 3 or 4 words (as a word is generated every 32 microseconds), with one additional word to accomodate the short software response delays described in the previous sub-section. A 16 word deep FIFO stack meets this requirement with a comfortable safety margin.

However, there is another type of system delay which can cause appreciable breaks in the running of a task. This is caused by the running of other tasks at a priority above that of the data input task. Care can be taken that the priority of the input task is set higher than that of any other user tasks, but there is still a problem with the operating system task which keeps track of task priorities and determines where control of the processor resides. This scheduling task comes in once a second, as determined by an interrupt derived from the 60 Hz line frequency. It scans all issued task requests and priorities, and establishes a scheduling order. If a large number of tasks are waiting to run, this scheduling can take up to several milliseconds; this would be an unacceptably long delay in the data input process. Since data input occurs for a period of 16 milliseconds over the 64 millisecond cycle time, on average one such task scheduling would take place during each data input sweep. It was decided that the easiest way to avoid loss of data was to prevent the scheduling from taking place during the data collection period. This can be achieved by simply disabling the line clock interrupt at the start of a sweep, and re-enabling it after collection is completed. Since only one scheduling operation in four is prevented, disturbance to other computer users is not serious. This procedure, and the necessary support software to make it possible, are described in section 6.4 of this chapter.

6.3 Interface Hardware

The interface hardware handles data throughput by providing a serial data sink to the scanner, and a parallel data source to the computer. Control logic looks after the timing of this operation, and in addition allows the process of scanning a map sheet to be automatically performed under program control. The following two sub-sections describe the buffer and control logic in detail.

6.3.1 Interface buffer

A first-in first-out (FIFO) buffer allows asynchronous data input and output operations. A Fairchild TTL macrologic 9403 16 words by four bits chip was chosen as being suitable for the purpose; four such chips are required for a 16 bit word. Data can also be removed serially from this chip, leaving open the possibility of using a high speed serial computer input line with only minor modifications. One other feature of the buffer structure is that it can easily be changed to make the output 8 bits in parallel. This would make it particularly suitable for use with a fast dedicated microprocessor controlled magnetic tape drive.

Upon receiving 16 bits of data, the FIFO generates a data ready signal which can be used directly as a request to the computer. Transfer of data to the output register of the stack and output enabling are performed directly by the return acknowledge signal from the computer, and data is then available on the DR11-C input register lines. Additional control signals required for FIFO operation are generated by the control logic described in the following section.

6.3.2 Control logic

Control signals of two types are required. First, the process of data transfer to the DR11-C I/O port requires standard flagging signals, and secondly, if the overall scanner operation is to be automatic, facility for starting and stopping the motors and the data collection process must be provided. It was mentioned in chapter 4 that the DR11-C interface consists of three 16 bit registers; an input register, an output register, and a control and status register (CSR). The CSR has two outgoing flag lines and two incoming request lines. These request lines can be used as interrupts if required, but need not be. The flag and request lines are suitable for data transfer handshaking, while the output register can be used for the essentially static overall control signals. Figure 6.1 shows the schematic structure of the DR11-C I/O port, with request and flag lines shown.

Figure 6.2 shows the logic constructed to perform the control functions described below. Gate identification is consistent with board layouts given in appendix B. In chapter 4 the start-of-sweep (SOS) and start-of-line (SOL) detectors were described. The SOS signal is used to generate an interrupt request to the computer, via bit 15 (REQ B) of the CSR. The interrupt request flip/flop (D1) is reset by the interrupt acknowledgement sent under program control via bit 0 of the CSR. The SOS signal is also used to reset the FIFO stack (MR). The SOL signal gates the system clock to start the data quantization (on the positive going edge) and synchronous filling of the stack (on the negative going edge). Clock gating is achieved by means of a monostable (M2) with a time constant slightly longer than the time taken to input a line of data. This obviates the need for an end-of-line detector; line length is determined by the number of words accepted by the software, which can be an entered variable. (This means that maps of variable size can be scanned without re-positioning and aligning an end-of-line detector, as would be necessary unless large

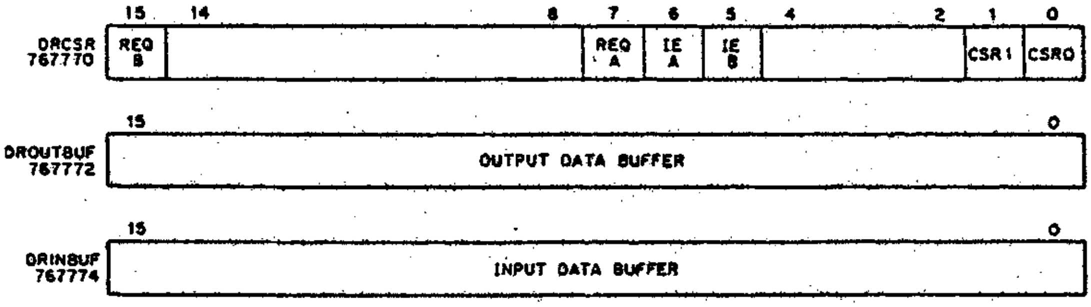


Figure 6.1 Structure of DR11-C I/O port

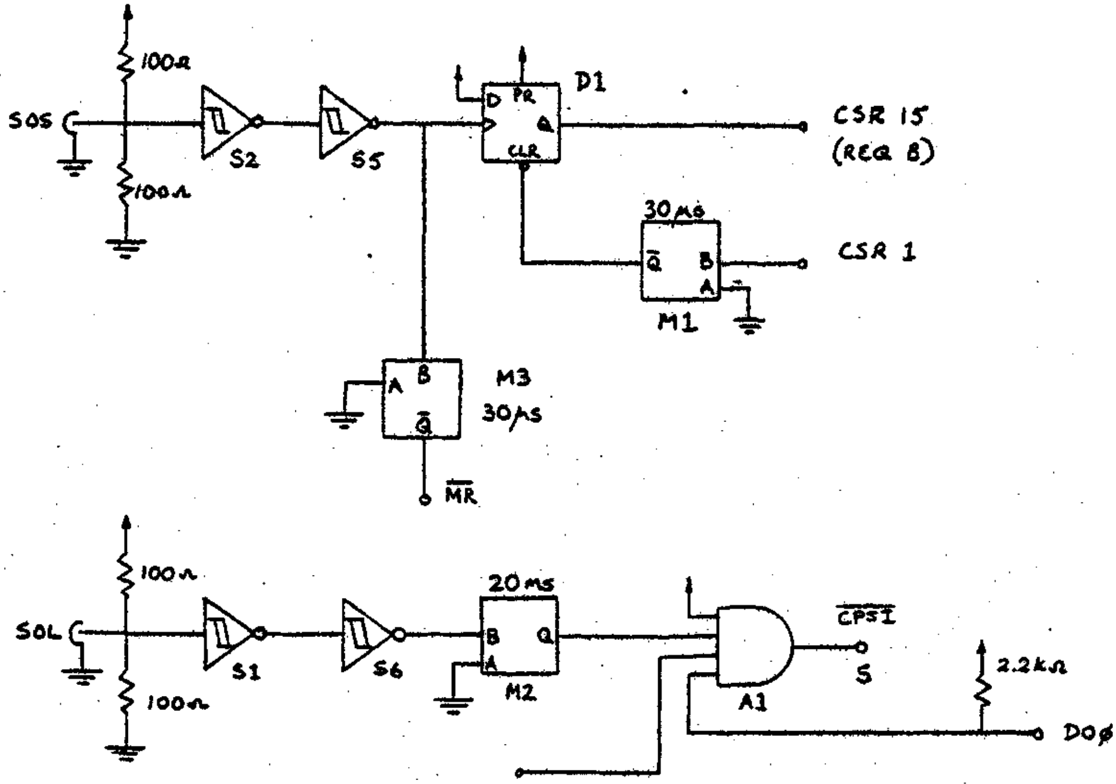


Figure 6.2 Interface control logic

quantities of off-map data were collected.) As successive words of 16 bits fall through the stack, the data ready (REQ A) and acknowledging output clocking (CSR1) signals control data transfer to the computer. At the end of a line (as determined by the line length specified when initiating the scanning), the data ready signal is simply ignored by the input routine, and at the end of the monostable timing period, clocking is gated out. Where necessary, buffers or monostables are used to avoid possible uncertain states due to spikes on the DR11-C lines. (Even with careful grounding, cross-talk between lines was encountered.) The logic is designed to prevent latch-up in any particular state caused by unsynchronised timing conditions such as program initiation at some critical phase of the collection sweep.

The operation of the input and the output sections of the interface board can be tested separately, either by running coupled to the scanner generated signals but disconnected from the computer, or by running with a test input program. Appropriate test points are provided on the interface board. In fact the time quantization circuitry, the frequency dividing circuitry and TTL drivers to send signals to the scanner over coaxial cable were built onto the interface boards. Appendix B contains details of board layout and wiring.

6.4 Software Design

The control and operational software is designed to minimise the effect of the entire scanning process on other users of the computer. To this end, three separate programs are used. The scan input program SCAN is responsible for accepting data from the DR11-C interface and writing it out to disk or magnetic tape. As noted earlier, data collection only takes place during one quarter of each sweep period, and the start of each data collection period is marked by an interrupt. The interrupt service program (INTRUP) detects this interrupt, and allows SCAN to take control of the processor for the data collection interval. The control program (CONTRL) co-ordinates the scanning operation. It is written in Fortran, and can be user modified as required. CONTRL initiates the other programs; necessary information is passed to SCAN, and return information back to CONTRL.

Support software to allow SCAN and INTRUP to communicate with the DR11-C port is also required, but it can be installed within the computer's operating system and hence be transparent to the user. In the following sections, the schematic logical flows of each program are given, along with more detailed descriptions of their operations.

6.4.1 Scan input program SCAN

SCAN performs two basic operations; first to transfer data from the DR11-C interface to memory; and secondly, to write this data from memory to some temporary storage medium, such as disk or tape.

The process of receiving data from the DR11-C is initiated by an event flag set by INTRUP. As described earlier, the first action of SCAN is to disable the line clock interrupt to keep control of the processor. The program then waits until it receives a data ready signal from the interface control logic, and upon doing so, acknowledges it and reads a word of data from the input register into memory. This continues until it has accepted the required number of lines of data (passed to it initially from CONTRL), after which it becomes dormant until the next setting of the event flag by INTRUP. During this input sequence, it is possible that a fault might occur (for example, a connection break or hardware failure) that would leave the program continually waiting for a data ready request. As the line clock is at this stage disabled, recovery of control of the processor from SCAN would not be possible. To avoid this, some form of check must be made within the wait loop, and an exit provided, incorporating re-enabling of the line clock interrupt. This check could be time based, or a counter could be incremented and exit performed if a specified count were exceeded. To minimise software overhead, a simple check of the computer's switch register is made; if a value of 1 is loaded into the register by the operator, exit and hence recovery is achieved.

The second operation performed by SCAN is to transfer data from memory to some temporary storage location on disk or magnetic tape. It is not possible to write out each line of data individually to disk or tape as it comes into memory, as disk access times and tape start up times are longer than the 48 milliseconds available between sweeps. Consequently larger amounts of data must be written out in a single continuous operation, and a dual buffering system used when inputting data to memory. A buffer size of 8 kwords was chosen; such a buffer can be written to either tape or disk in less than half the time taken to fill up the second buffer, leaving a substantial safety margin. Writing takes place under control of the device driver about once a second, and the processor need only transfer to the driver the address in memory of the buffer to be written out.

Finally, SCAN also keeps a record of data transfer operations. This is in the form of I/O status codes generated by the system software after an I/O operation has taken place, describing the success or otherwise of the transfer, and the size of the data block transfered. If a data transfer error occurs, its position and details can thus be determined. After writing this information out to disk, SCAN activates CONTRL and sends the disk address at which the I/O status buffer is written, so that it can be checked from CONTRL. Figure 6.3 gives the logical flow of SCAN, while the complete program listing is given in appendix C.

6.4.2 Interrupt process program INTRUP

The purpose of INTRUP is to recognise the start of a sweep cycle, and indirectly allow control of the processor to be taken by SCAN. This is achieved by setting a system event flag and declaring a significant system event; the effect of such a declaration is that the scheduler determines which tasks, if any, are waiting to become active pending the setting of the event flag. SCAN leaves itself in such a state after a 16 millisecond data collection interval, and can thus proceed to take control of the processor for the next 16 millisecond period. The minimal size, and non-consequential function of INTRUP, mean that it can quite safely be permananatly installed in the operating software if required. It can also be used as a manual or automatic interrupt detector for stepping through hardware or software development, testing or debugging routines. Figure 6.4 shows the logical flow of INTRUP, while the program listing is given in appendix C.

SCAN

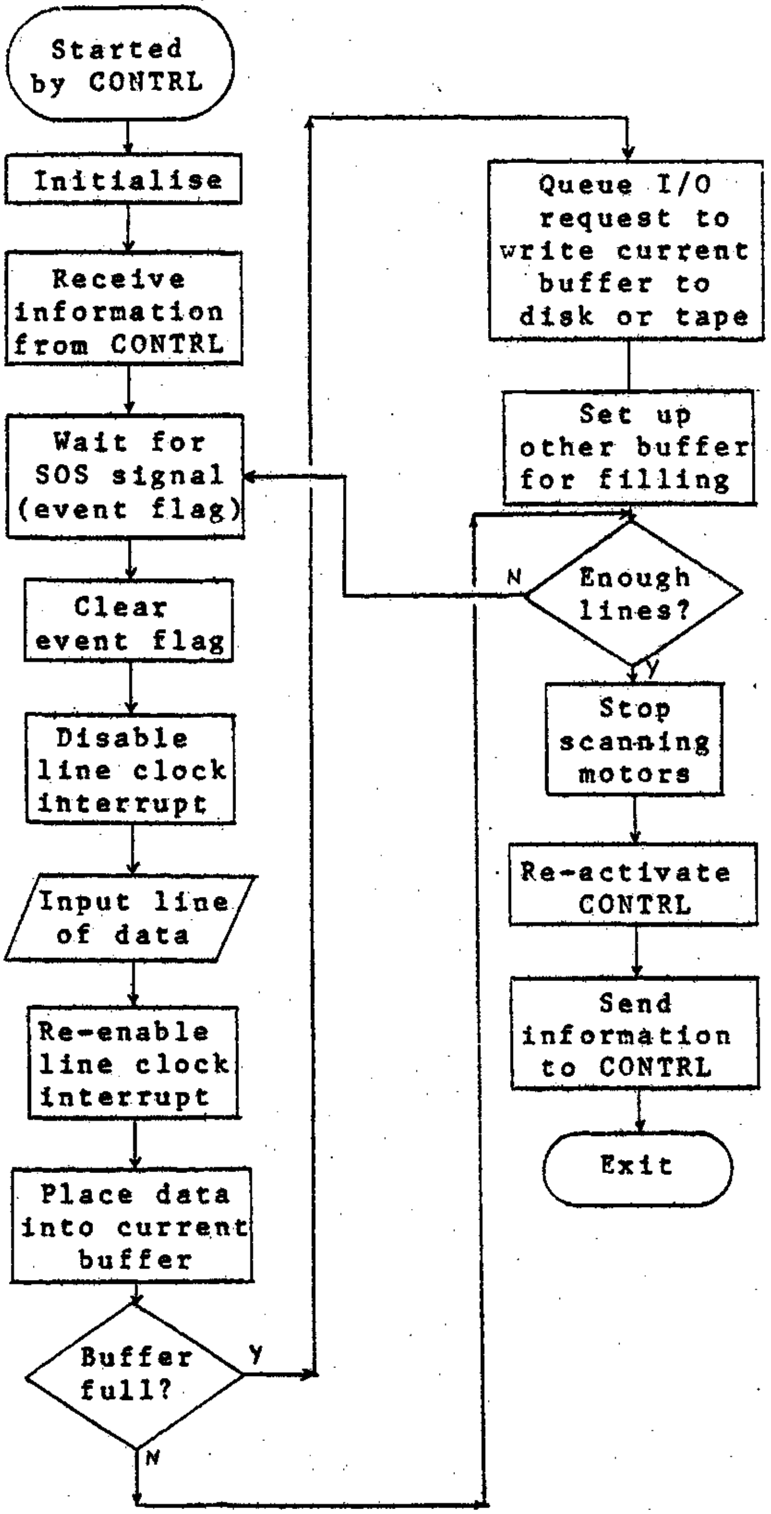


Figure 6.3 Logical flow of SCAN

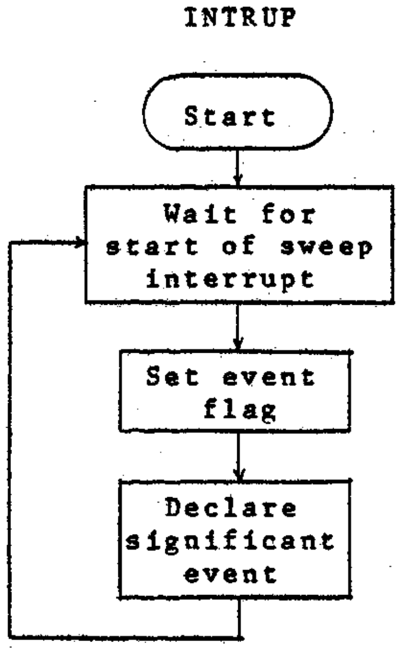


Figure 6.4
Logical flow of INTRUP

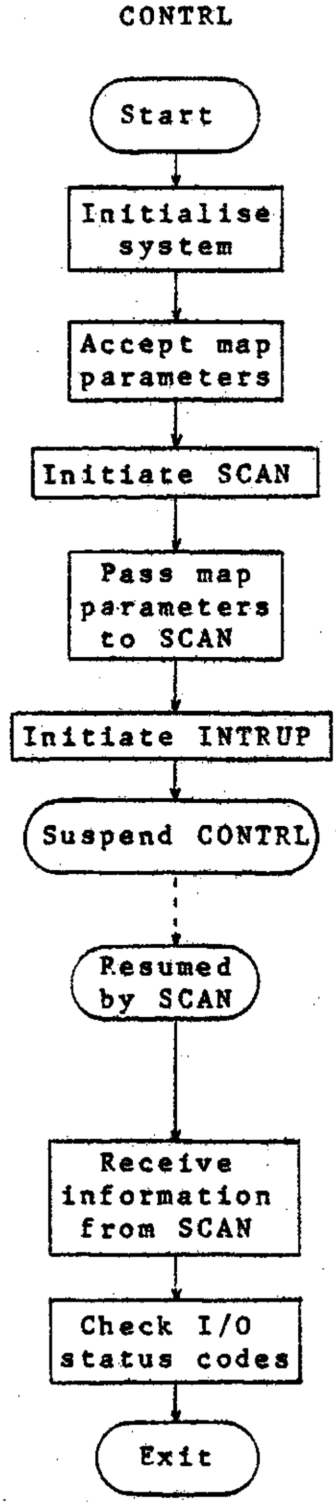


Figure 6.5
Logical flow of CONTRL

6.4.3 Control program CONTRL

The scanning operation is initiated by CONTRL. Information which must be transfered to SCAN, such as the size of the map to be scanned (line length and number of lines), and the storage medium (disk or tape), is input through CONTRL and passed to SCAN. While scanning is taking place, CONTRL is dormant, and can be checked out of memory onto disk to allow fuller use of memory by other users. When scanning is completed, CONTRL is brought back by SCAN and information concerning the running of the operation is returned to it, as described in section 6.4.1. Figure 6.5 shows the logical flow of CONTRL, and the program listing is given in appendix C.

6.4.4 Support software

The extent to which the design of the system is dependent on the computer's software operating system was pointed out in earlier sections. Because of this, it is worth briefly describing the support software necessary to run the process. In order that INTRUP and SCAN can communicate directly with the DR11-C interface, the address of the DR11-C registers must be defined. This can be done within a program segment, but this imposes size and structure limitations on the program. Better is to create what is known as a device common area at the required location in memory; named registers with their physical addresses defined in this area are then accessable to any program which is linked to the area. The common area is built by the program DCOM. Another program, SCOM, builds a similar area to allow access to the line clock interrupt register and the computer's switch register (used for a software trap in case of an error condition occurring while the line clock is disabled).

Further support software includes programs to display scanned data on various types of display systems, and other data handling programs. Because of the large volumes of data involved, an image processor display system was found particularly useful for visually examining scanned maps. Data display routines, along with appropriate subroutines to pack and unpack data, and other programs to transfer data between disk and tape, are not included in this thesis.

6.5 Summary

The preceding sections of this chapter describe the hardware and software structure used to transfer data generated by the scanner onto a temporary storage medium (disk or tape). This data is in raw raster form; a bit value of '1' means that a line or mark on the map was encountered at this point; otherwise a '0' is recorded. Successive words represent serial groups of 16 bits. Figure

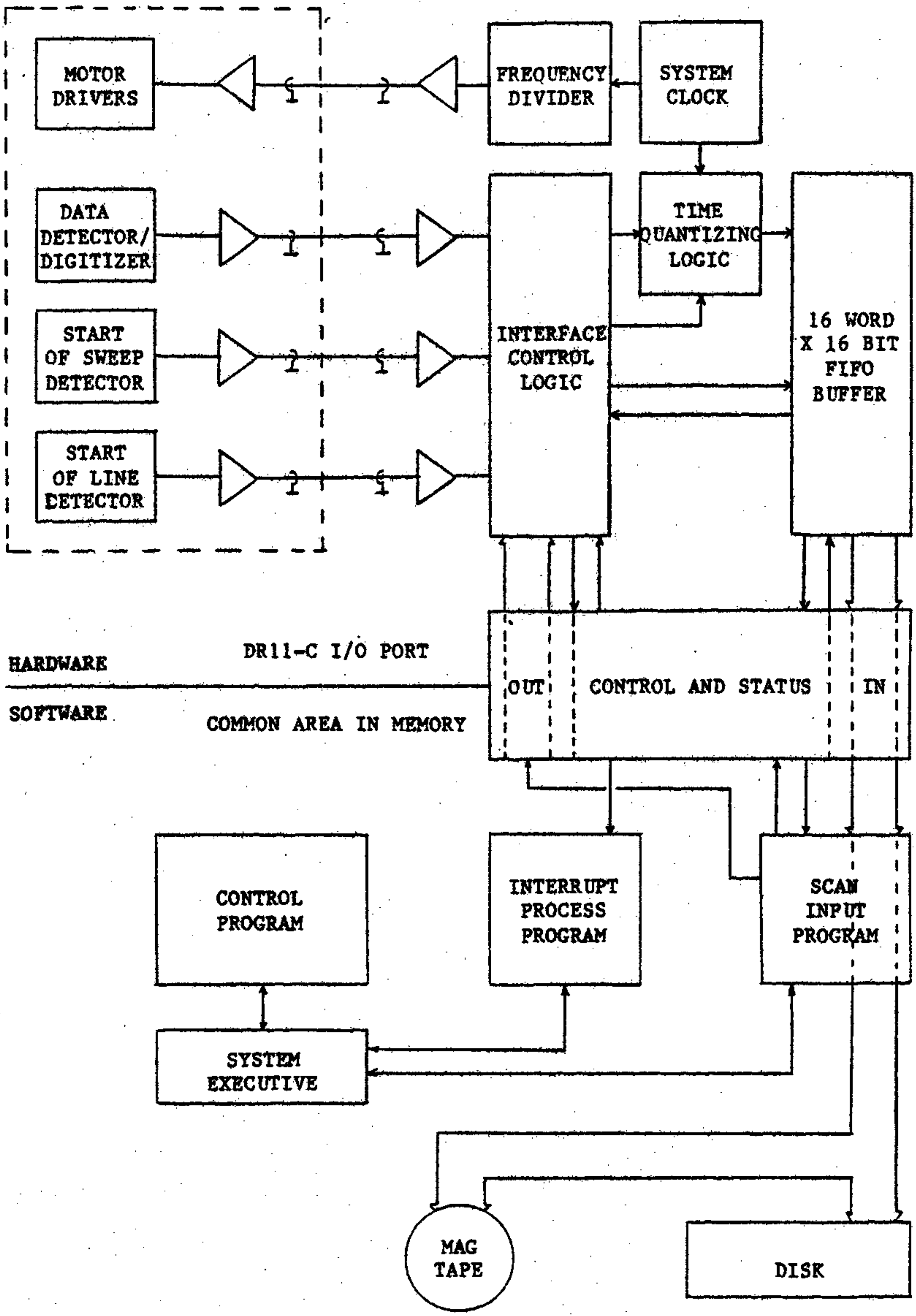


Figure 6.6 Scanning System Interface Structure

6.6 summarises in block form the overall interface structure. The scanner hardware is also included for completeness.

While the data could be stored directly in its generated form, it turns out that for most types of map sheet data, a direct raster format has a substantial degree of redundancy. The following chapter looks at software methods of reducing the total data volume to render it more suitable for permanant storage.

7. SOFTWARE DATA PROCESSING

7.1 Introduction

The scanning process described in previous chapters results in a digital description of a map sheet such that every stored bit represents directly information obtained about small areas of the map centred on resolution grid points. Software processing is performed on this data for one or both of two reasons. First, storing the data in this form is not particularly efficient; if a long run of zeros occurs, essentially no information takes up considerable storage space. Some method of reducing data redundancy is required if raster data is to be stored directly. Secondly, as pointed out in chapter 1, the data may be required in vector form. For most types of cartographic data, representation in vector form results in a more compact data set than when in raw raster form. In addition, it is possible that the first step of the vectorization process, that of line thinning, might be performed on raster data if line widths are not to be retained.

In this chapter, the first of these processes, that of data coding, is looked at in some detail, after which vectorization processes are briefly outlined.

7.2 Coding of Raster Data

7.2.1 Requirements of a coding scheme

Although data compaction is the purpose for which coding is performed, a suitable coding scheme need not necessarily optimally encode the data. More important is that it be a fast process, and that the reverse process of decoding also be fast. In addition, it is desirable that it be possible to perform coding and decoding in a high level language (specifically Fortran), to make the data universally usable.

Scanned data consists of a serial string of bits, arranged in words of 16 bits in size, which can be segmented at appropriate points to reconstitute individual sweeps across a map. Data is generally characterised by fairly long strings of '0' bits, representing spaces between lines, followed by shorter strings of '1' bits, representing the crossing of lines. Run length encoding is particularly suited to this type of data. Provided the average length of a run of like bits is greater than the number of bits required to describe that number, compaction is achieved.

This suggests that the statistics of map data in raster form be investigated in order first that a suitable run length encoding scheme can be chosen, and secondly that the compaction achieved by applying such a scheme to a variety of map sheet data sets can be estimated. The following section makes such an investigation, and in section 7.2.3, the limitations imposed by the requirement that coding be performable in a high level language are evaluated.

7.2.2 Statistical basis for coding

If a suitable coding scheme is to be chosen, some information about the likely zero-run and one-run lengths is required. These lengths will depend upon map sheet characteristics, but to some extent have a measure of predictability. More specifically, the length of zero-runs will depend on the type of map (contours, culture, drainage etc.), the nature of the lines on the map (for example, contours are usually aligned to some extent and spaced apart in a non-random manner), and the angle at which a sweep cuts the lines. Similarly, the lengths of one-runs will depend on the above factors, but rather on the line width than spacing. If the map sheet area is sufficiently large, the angle at which lines are cut can be assumed to be on average random, suggesting that, for example, the probability distribution for the occurrence of lengths of one-runs could to some degree be predicted. The fact that lines are often nearly parallel for reasonable proportions of their lengths suggests that if the zero-run lengths and the one-run lengths are represented by discrete variables, these variables will not be entirely independent.

A thorough approach to choosing a suitable coding scheme would involve using map sheets representative of various types of cartographic information to generate histograms of the relative frequencies of occurrence of specified pairs of successive runs of zeros and ones. If a distribution of known characteristics could then be found to well approximate the frequency histograms for each type of map sheet by varying only the parameters of the distribution, coding efficiencies under any specified coding scheme could be calculated for encountered ranges of these parametric variations. To get an indication of the feasibility of such an approach, data sets from contour, culture and drainage map sheets were examined. These data sets were obtained by scanning the three map sheets shown in chapter 8, figures 8.8.1, 8.9.1 and 8.10.1. A two dimensional set of bins was constructed, and run length pairs (a run length pair consists of a zero-run length and a one-run length) were binned according to their lengths. Since the length of a run that can be represented by n binary bits is given by 2^n-1, bin boundaries were set at these values for n=1,15. Figure 7.1 shows the resulting two dimensional relative frequency histograms in the form of profiles along the zero-run length axis, for each one-run length bin number. Bin numbers correspond to the exponent of the upper limit of the bin, given by 2^n-1, and hence the

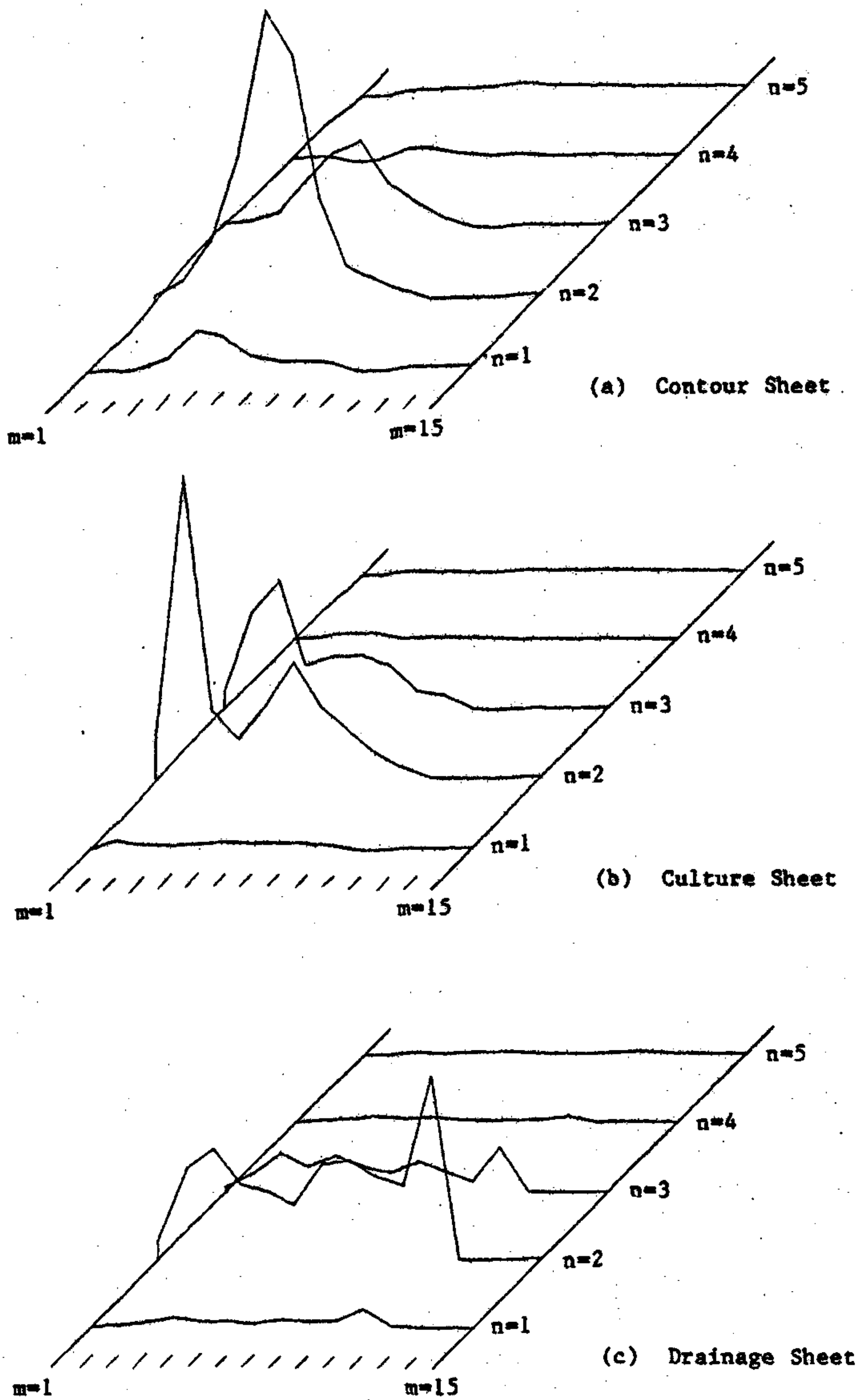


Figure 7.1 Raster Run Length Statistics

run length measure is logarithmic. The histogram profiles of figure 7.1(a) are for the contour map sheet, those of 7.1(b) for the cultural sheet, and those of 7.1(c) for the drainage sheet. Individual parameters for zero-run lengths and one-run lengths were also calculated, as were correlations between the two. These parameters are summarised in table 7.1.

Map sheet		Contour	Culture	Drainage
Area of sheet (sq. inches)		50	50	50
Raw data volume (kwords)		16	16	16
Total no. of run pairs		50231	55213	7152
Zero-runs :	Mean	59.3	53.2	436.5
	St. Dev.	235	300	634
One-runs :	Mean	3.3	3.8	3.3
	St. Dev.	2.9	4.1	3.1
Correlation coefficient		0.02	-0.02	-0.08

Table 7.1 Summary of raster map data statistics

It can be seen that the data sets have far from uniform characteristics. The double peak in the culture sheet profiles is due to the fact that roads are represented by parallel boundary lines, the space between these lines giving rise to the low zero-run length peak. As these lengths are small, the average is not greatly affected, but the standard deviation is high. The drainage sheet is fairly sparse, and the total data volume is hence small. The pronounced peak for m=12 reflects the fact that over much of the sheet, a single group of lines is the only data encountered.

It was decided that to perform a comprehensive analysis of the statistics of raster represented map data was beyond the scope of this work. Rather a more practical approach to coding the data was taken. A simple method largely determined by convenience for use with a high level language, and with a minimum of calculative overhead was chosen, and its compacting performance for each of the data sets illustrated in this section was calculated. The following section describes this coding scheme, and the method of applying it to data from a full map sheet. In addition, the storage penalties incurred by making it usable entirely with a high level language are investigated.

7.2.3 Coding method

If coding and decoding are to be performed in a high level language at high speed, run lengths must be represented by directly addressable volumes; that is, bytes or words. Each coded entry consists of a pair of numbers, one representing the zero-run length, and one representing the one-run length. Since run lengths longer than 255 occur infrequently, it was decided that a 16 bit word should be used to represent each run pair. If an 8 bit byte is used to represent each run, runs are then directly addressable. An equally simple run length encoding scheme records a count of the number of successive zeros occurring before each one; this results in a total number of entries equal to the total number of ones. Preliminary investigation showed that this method yields an expansion rather than a compression of the data representing map sheets, due to the frequent occurrence of multiple ones. Consequently it was not investigated further.

To prevent the possibility that a single error (for example, as might be encountered when reading data from tape) should be cumulative, it was decided that data should be encoded in chunks. A '0' word, which would never be generated under normal coding, was used to signify the start of a new chunk of data, and the following word was set aside as an information word for specific user applications. (For example, a specific value could indicate that the data for that line was coded using a 16 bit word for each run length if that chunk had a predominance of long runs.) A chunk could be simply a line of data, or if preferred, an area. Although retrieval of data from disk is faster if done sequentially, as it is written onto disk line by line, coding by areas has the attraction that if sections of the map are to be edited or displayed, retrieval of that section is simpler and faster. Areas 512 elements square are

suitable since this is the size of most display screens. The volume occupied by uncoded data representing an area 512 elements square is 16 kwords; this is as large a volume as can comfortably be handled in memory if some form of processing is to be performed on the data. The co-ordinated position of each area on the map sheet can be registered in the word following the zero word. If a particular chunk size and shape were decided upon, data structures on disk could be modified to optimise data transfer times.
Line orientation has very definite trends over localised areas. This suggests that compactions will be highly dependent on scanning direction. This was found to be the case, but the added programming complexity and time penalty incurred in first determining the direction for which most efficient compaction occurs, and then possibly re-manipulating large amounts of data, would make it unattractive to make use of this feature, unless storage volume were at a premium.

The coding scheme is not particularly efficient at describing runs of longer than 255 elements. Since most such runs will be zero-runs, it could be that using a larger bit field to represent the zero-run than that for the one-run would result in a more efficient coding method. This might require an assembly language subroutine to code and decode, though a simple magnitude test and subtraction could be implemented in a high level language to add one extra bit to the zero-run field. To examine the possible gain in compaction efficiency using such a scheme, storage volumes and hence coding compactions obtained when using a p-bit zero-run field, and a q-bit one-run field (where p+q=16) were calculated for each of the three data sets of the previous section. Figure 7.2(a), (b) and (c) shows curves of compaction factors obtained for p=1,15 for contour, culture and drainage map sheets respectively. It can be seen that the possible gain in compaction by increasing the zero-run bit field is small for each of the contour and culture sheets, though for the drainage sheet which has larger areas of no information, there is a significant increase. However, overall compaction for this sheet is fairly high, and its coded data volume low; overall the increase in storage volume is small. Based on these compactions, it was decided that the simple coding scheme using one byte to represent each run length should be implemented. One additional advantage of having equal zero-run and one-run bit fields is that maps scanned in positive form can be equally well coded without changing the scheme, and further, sheets with large shaded areas can also be fairly efficiently coded.

It was not intended that this coding scheme should in any way optimally encode map sheet data; rather the intent was to compact the original raster data in such a way that decoding is simple and rapid, while still achieving a

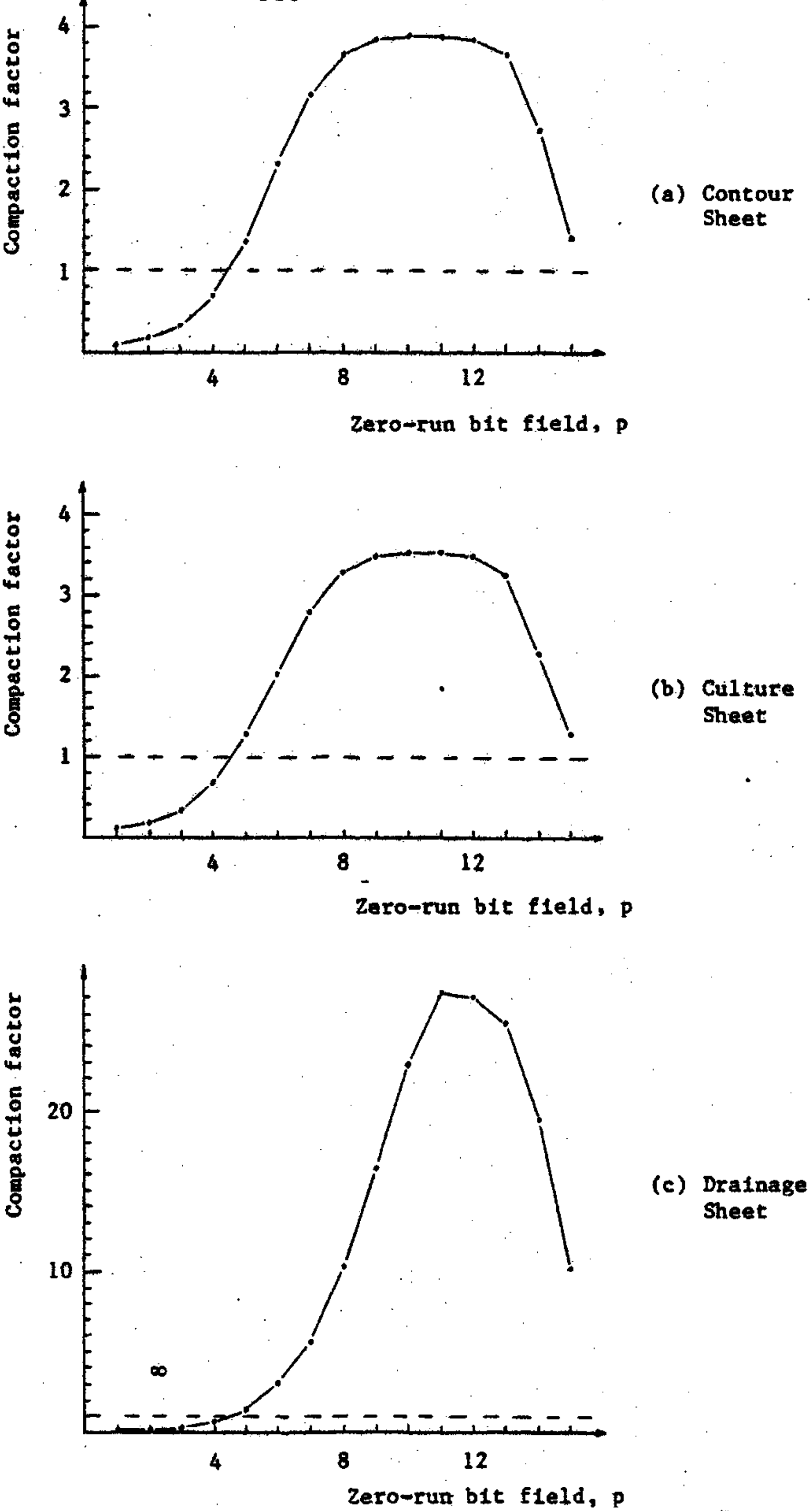


Figure 7.2 Coding Compaction Factors

reasonable degree of compaction. The coding scheme was implemented in Fortran, and data was transferred from uncoded form on disk to coded form on magnetic tape. Coding integrity was also checked by decoding and displaying the resulting data.

7.3 Raster to Vector Conversion

7.3.1 Line thinning

Line thinning is a process by which the directionality and length of lines is maintained, while their width is reduced to a minimum. Thinning processes, and their operation on the elements of a map or picture (called pixels) can be described in terms of pixel adjacency and line connectivity [11]. Figure 7.3(a) shows a pixel (i,j) of a picture, with its 4-neighbours (i-1,j), (i,j-1), (i,j+1), (i+1,j). Each of these points is 4-adjacent to (i,j). If diagonal elements are included, as shown in figure 7.3(b), the points are 8-adjacent to (i,j). A sequence of distinct points $(i_0,j_0),(i_1,j_1)\ldots(i_n,j_n)$ is called a path from (i_0,j_0) to (i_n,j_n) if (i_m,j_m) is adjacent to (i_{m-1},j_{m-1}) for $1<m\leq n$. If there is a path between two points in a picture, then these points are said to be connected.

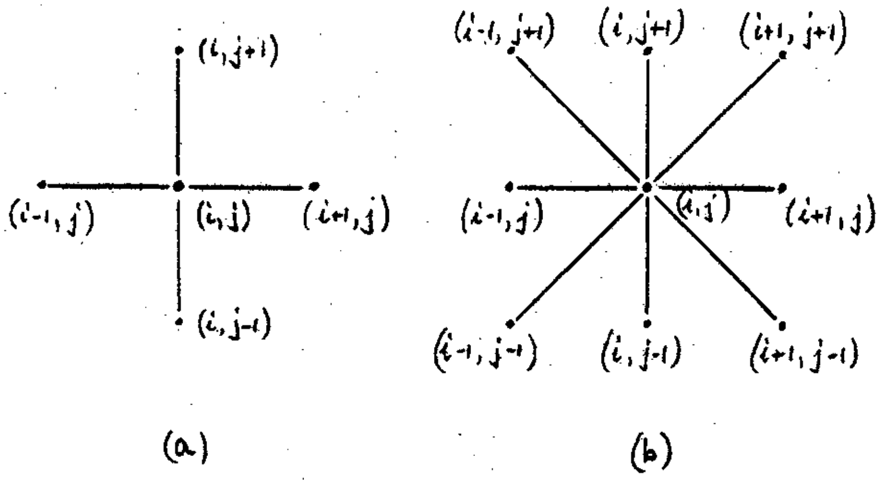


Figure 7.3 Pixel adjacency

Thinning should not destroy the connectivity of the end points of a line; that is, it should be possible to trace the path of a line by following from one set pixel to the next in steps of either one resolution element, or one diagonal element; and at the same time, set pixels should be adjacent to only two other pixels, which themselves are not adjacent, except at singular points such as line junctions or line ends. For describing lines on a map sheet, 8-adjacency is normally used to give a smoother curve representation (and also a smaller data volume). When thinning lines from a map sheet, generally the central track of the line is the required thinned path. Given the above conditions then, it is possible to construct an algorithm which may be applied to each pixel of a scanned map sheet to produce thinned lines. However, this cannot necessarily be achieved with a single pass through the data.

Thinning processes can be either parallel or sequential in nature. In a parallel process, an algorithm applied to each pixel determines if that pixel is required to maintain path connectivity on the basis of adjacent pixel values in the original data. That is, the previous state values of adjacent pixels are used to determine the next state value of each pixel. Sequential thinning, on the other hand, is a process which is applied to pixels sequentially, and determines if the pixel is required to maintain path connectivity from the previous state values of as yet unprocessed pixels, and the current state values of processed pixels. Sequential thinning is thus directional. Because of the large volumes of map data, and the consequent long processing times, a single pass thinning process is attractive, even if some form of data pre-processing (such as nibbling) is necessary so that thinning can be satisfactorily performed.

7.3.2 Line vectorization

If raster data representing lines satisfies the conditions for connectivity described in the previous section (that is, it is thinned) it can be converted into strings of vectors describing the lines. This vectorization consists essentially of two processes; the first is that of following the path of a line, representing it as a series of successive vectors from one point to its next adjacent point; and the second is that of determining if a group of these successive vectors represents a "straight" line segment, for some given definition of straight, and if so, of replacing the group with a single vector. These two processes are known as line following and weeding respectively.

7.3.3 Summary of vectorization processes

The implementation of raster to vector conversion processes incorporating thinning, line following and weeding, were the subject of work being done in conjunction with this project [7]. An asymmetric sequential operator [8] was chosen to perform thinning on the map data in a one-pass process. This operator imposes the requirement that no line be wider than 4 elements; exceeding this limit may result in prominences along the thinned line. For normal map data this is the case, provided that the scanning process does not artificially widen lines. Line following and weeding were also performed on the thinned data in order to compare vector storage volumes with those obtained using raster coding techniques.

7.4 Raster/Vector Format Compaction Comparisons

The three data sets described earlier were used to compare the compaction achieved by raster coding with those achieved by line vectorization [7]. Table 7.2 shows the resulting storage volumes for each type of data. It can be seen that on the basis of storage volumes alone, vectorization is not warranted.

Map sheet	Compaction factors		
	Line vector	Coded raster	
		8-bit zero-field 8-bit one-field	12-bit zero-field 4-bit one-field
Contour	3.9	3.7	3.9
Culture	1.4	3.3	3.5
Drainage	9.5	10.3	27.0
Overall	2.8	4.5	5.2

Table 7.2 Data compaction factors

7.5 Summary

This chapter has looked at the problems of data reduction and conversion from a practical viewpoint. In its raw state, the large amount of raster data generated by a scanning process could well pose serious storage and handling problems; perhaps to the point that dealing with such data would be impracticable. The intent of this chapter, and the work done on data coding, was to demonstrate that it is quite feasible to reduce the raw data volume to an amount comparable to that which would be generated by a line following type of digitization process, using only a simple, rapid coding scheme which can be implemented in a high level language. This is considered in chapter 9, along with the overall performance evaluation made in chapter 8, in judging the suitability of the scanning system for the large scale digitizing of cartographic information.

8. DIGITIZING SYSTEM PERFORMANCE EVALUATION

8.1 Introduction

The performance of the digitizing system can be judged to some extent subjectively by visually examining scanned map sheets, but a more specific evaluation can be made by scanning a test pattern. Since the mechanisms governing the digitizing process in the sweep direction are not entirely the same as those governing it in the travel direction, a test pattern which allows two dimensional uniformity of resolution to be gauged is desirable. For this purpose a test pattern consisting of a circle of alternate black and white sectors, each subtending an angle of 3 degrees at the circle centre, was drawn up, reduced and repeatedly photographed onto a strip of 35 mm. film. Figure 8.1(a) shows a true size section of this strip. This enabled digitization along any strip across the sweep or travel to be examined. A straightforward test pattern with precise line and space widths, as shown in figure 8.1(b) was also found useful while setting up the system.

In the course of previous chapters, specific components of the digitizing system were cited for further testing; section 8.2 describes procedures used to test the performance of each of these components. Section 8.3 looks at the system performance in scanning various types of actual map sheets, while in section 8.4 a summary of the aspects of the system which could benefit from improvement, and the suggested nature of these improvements, is made.

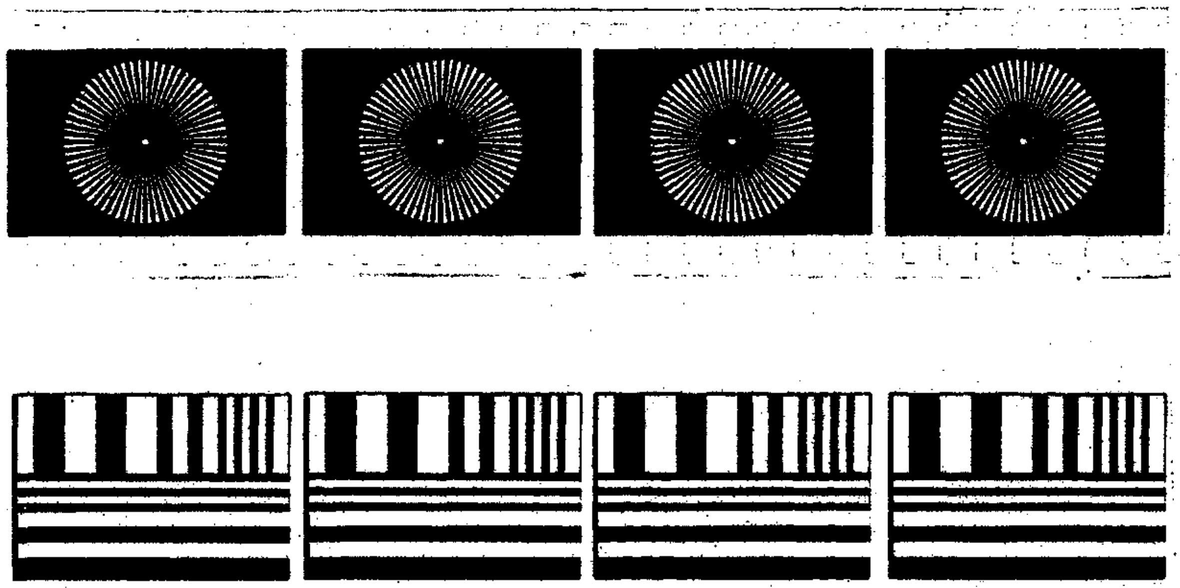


Figure 8.1 Test pattern strips

8.2 Testing of Specific Aspects of System Performance

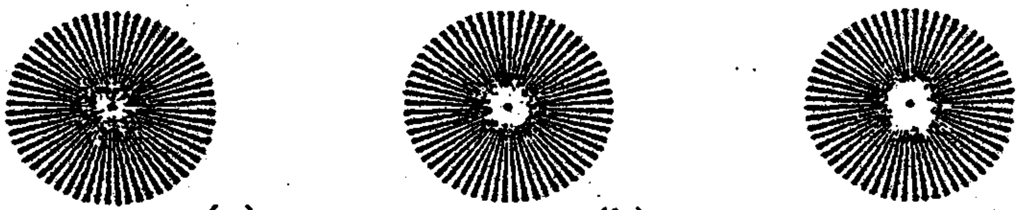
8.2.1 Consistency of rotating mirror motor speed

Section 4.6 of chapter 4 pointed out the need for testing the consistency of the mirror motor rotational speed. Small variations in motor speed will create edges to lines or areas on the reconstructed map sheet. It was found that over the sweep, fluctuations in speed occurred in a non-reproducible manner. Very close to the triggering detector, little or no fluctuations were apparent (that is, lines were barely jagged), but within several inches on the map (one inch corresponds to about 3 degrees of angular displacement of the motor armature) of this detector, variations of line position of one bit either side of a mean were encountered. These variations occurred apparently at random throughout the sweep; bit offset for any particular line was only consistent over several inches. More than one bit offset from the average position was seldom encountered. A close examination of the 61 Hz driving frequency for the motor showed minimal, if any, inconsistency in delay distortion, rising edges being reproducible to within the measurement accuracy, and certainly to within 50 nanoseconds or so. It was concluded that friction in the motor bearings was responsible for the fluctations. Previously, the motor had been run without a flywheel, and fluctuations were of an order of magnitude greater. In fact flywheel size was fairly arbitrary; the motor shaft protruded only at one end of the motor, and hence to prevent undue unbalancing, flywheel mass was kept to approximately that of the motor itself. It is not insignificant that in commercial scanning systems, motors are of high quality and often have air bearings for reduced frictional drag. Section 8.4 looks further at the limitations of the motor used when discussing improvements.

The effect of a present third harmonic in the motor driving frequency was also investigated using the test pattern strip of figure 8.1(b). If noticeable, these effects should be apparent as an expansion or compression of linear distance in the reconstituted data, with expansion and compression maxima 60 degrees apart. No discernible periodic fluctuations in linear spot speed were discovered, though it is possible that they might have been masked by the fluctuations due to frictional drag. However, it seems that the combined effects of the inductive properties of the motor and the inertia of the flyweel were sufficient to smooth out any variations due to this residual third harmonic at least to within the resolution accuracy of the digitizing process.

8.2.2 Uniformity of digitizing processes in sweep and travel directions

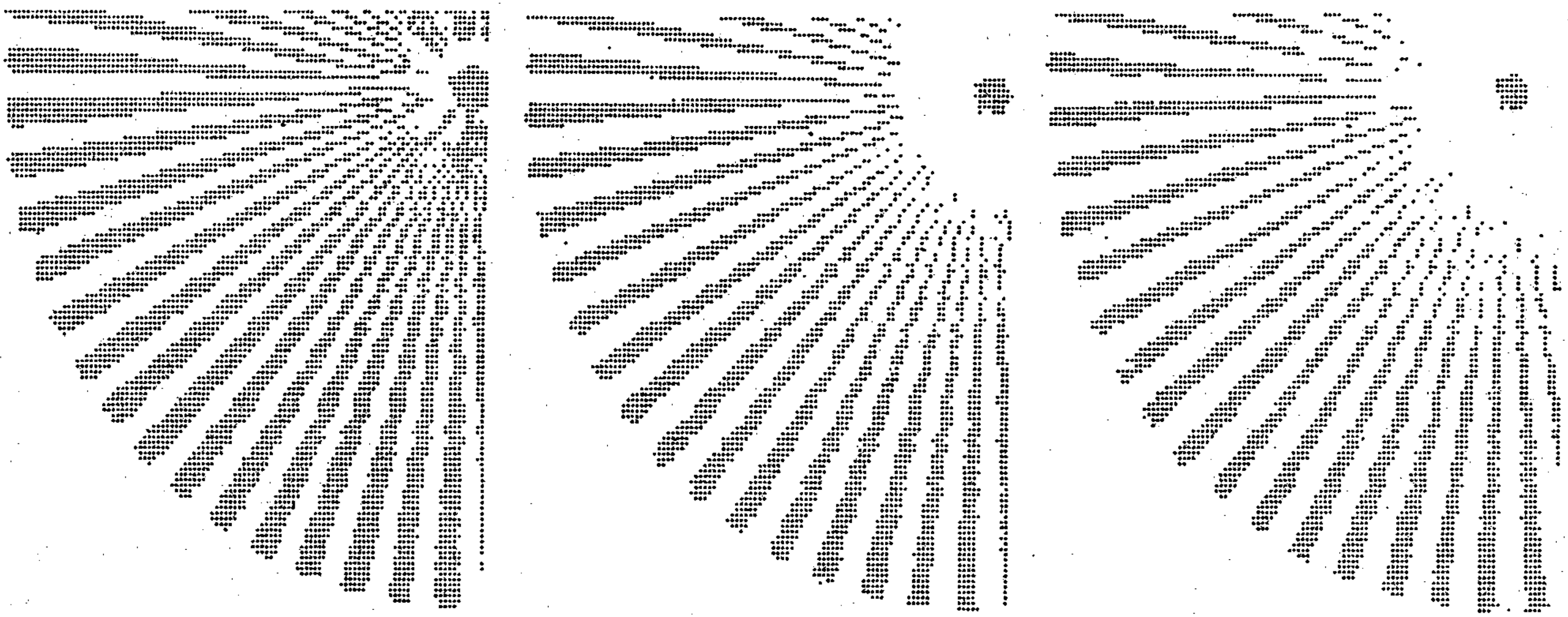
Figure 8.2 shows the result of scanning the test pattern of figure 8.1. The patterns were replotted at a resolution of 200 points per inch, rather than the 250 points per inch at which they were scanned. The oval shape is due to the fact that the pulley wheel driving the trolley had a diameter of 1.25", rather than the 1.14" specified in chapter 4. This results in a ratio of major to minor axis lengths equal to that measured from the patterns. All scanning was performed with this oversized pulley, though replacing it with a correctly sized one would be a simple task. The resulting resolution spacing in the travel direction is then .0044", meaning that narrow line detection probability along a cross-section in this direction is reduced (referring to figure 5.5). This is most apparent when the threshold level is above 50%, and can be seen from the oval shaped loss area in the centre of figures 8.2(b) and (c), for which thresholding took place at about 60% and 70% of the maximum level repectively.



(a) (b) (c)

Figure 8.2 Plots of scanned test pattern

Figures 8.3(a) and (b) and (c) show enlarged plots of sections of figures 8.2(a), (b) and (c) respectively. A horizontal to vertical scale factor of 4/5 is introduced in figure 8.3 due to the geometry of the output plotting device. These test patterns indicate that neither the detector/amplifier bandwidth, nor the transmission of the signal from the scanner to the interface, causes appreciable degradation of resolution in the sweep direction.



(a) (b) (c)

Figure 8.3 Plots of scanned test patterns showing individual pixels

8.2.3 Spot diameter measurement

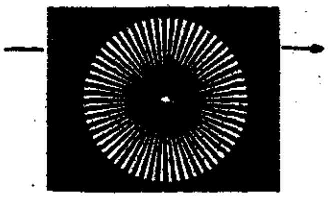
If the focused laser spot moves across an opaque/transparent boundary, a detected signal will be generated with a rise time dependent on the spot diameter, providing that the detector/amplifier bandwidth is sufficiently high. From the curves of figure 5.3, it can be seen that for t>d, a spot moves through approximately .63 of its own diameter between the 10% and 90% detected signal amplitude levels. Thus measurement of the 10% to 90% rise time of the generated signal will enable calculation of the spot diameter. This measurement was made for the smallest obtainable spot size for secondary lenses of various focal lengths. Spot size was minimized by adjusting focus for minimum rise time in each case, and the smallest aperture which did not cause beam truncation was used. With a 50 mm. secondary lens, a spot diameter of approximately .007" was obtained, and with a 100 mm. lens, the diameter was approximately .004". Lenses were of equivalent quality (Nikkor PC lenses). It can be seen that the theoretical limitations derived in appendix A are substantiated, the smallest spot size being obtained with the lens of longest focal length (and hence maximum collimation factor).

8.2.4 Effects of finite spot size and threshold level

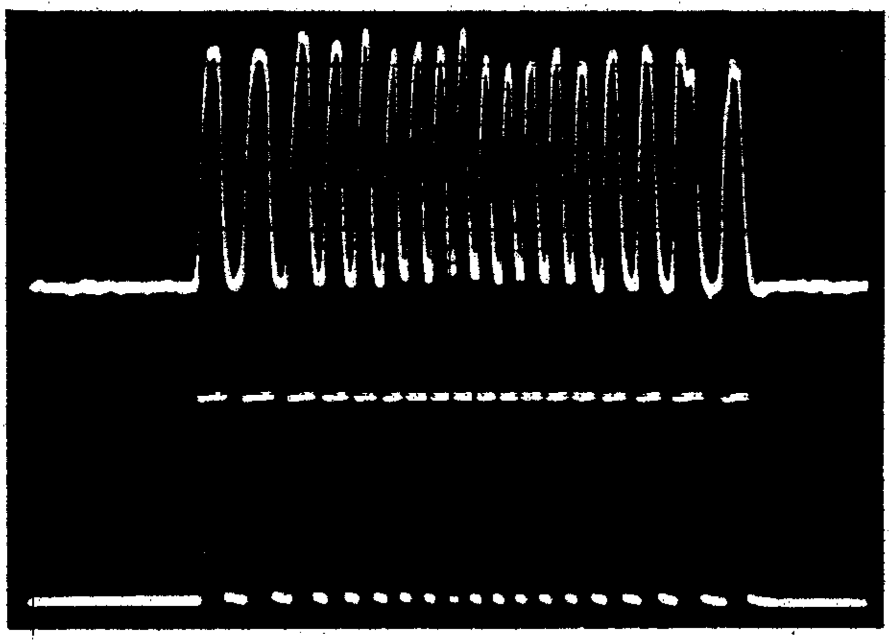
As described in the previous section, a finite spot size results in finite detected signal rise and fall times; if a line is narrower than the spot diameter, a signal amplitude less than 95% of the maximum will be obtained. Figure 8.4 shows a small portion of a sweep across a contour map sheet. Lines producing pulses (c) and (h) have a width of approximately .007", those producing pulses (e), (f) and (g) have a width of between .002" and .003" along the line of sweep, and those producing pulses (a), (b), (d) and (i) have a width of .002". Using figure 5.3, and extrapolating between curves where necessary, the pulse heights correspond to those predicted for a spot diameter of .007" to within 10% (the map was scanned using the 50mm secondary lens). Thus the effects of a finite spot size can clearly be seen, and close agreement with the spot size measured in section 8.2.2 is obtained.

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

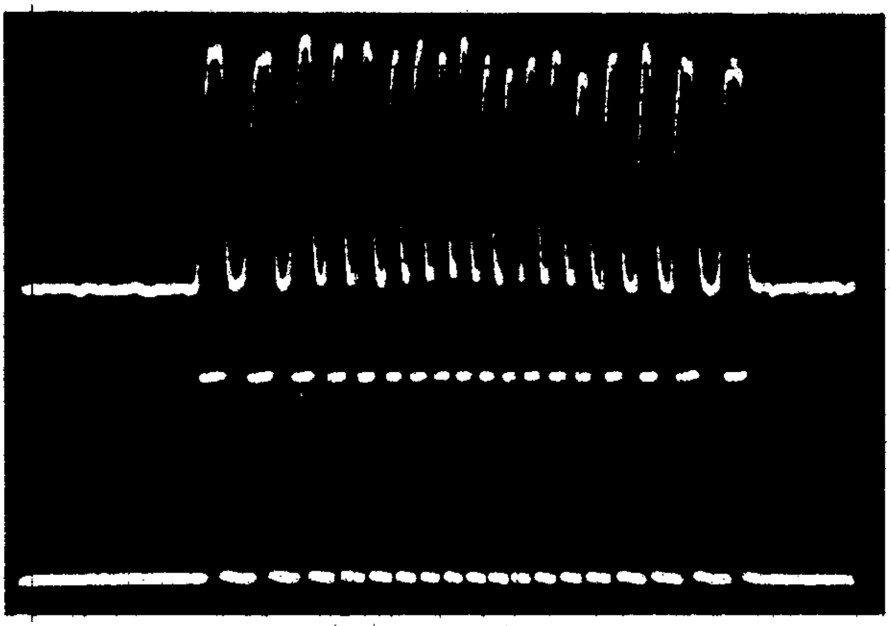
Figure 8.4 Analogue signal detected along a map sheet sweep



(a)



(b)



(c)

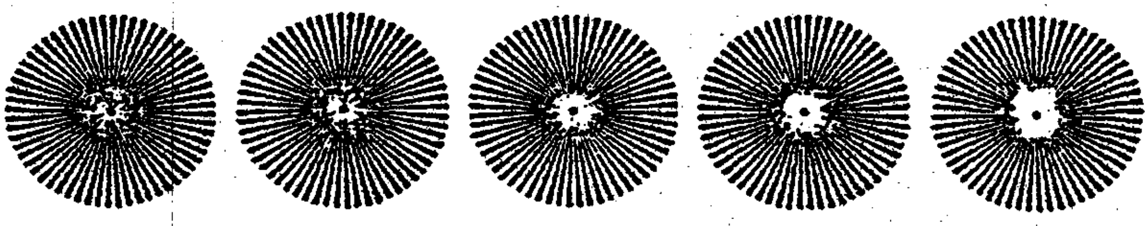
Figure 8.5 Effect of threshold level on line/space widths

The effects of changing the threshold level on line and space width can be seen from figure 8.5, in which a section of the test pattern is scanned (shown in figure 8.5(a)). A threshold level of approximately 30% (figure 8.5(b)) widens lines and improves line detection probability, while thresholding at approximately 70% (figure 8.5(c)) widens spaces and hence improves space detection. In each case the upper trace shows the amplified detected signal, and the lower trace shows the result after thresholding. (These traces also show localised non-linearities in response over the sweep as described in the following sub-section.)

8.2.5 Non-linearities in response over sweep

Non-linearities in response along the sweep can be separated into two types; first, the variation in signal level of long time constant, averaged over a distance of an inch or so, over the length of the sweep; and secondly, the localized variations of shorter time constant over localized areas. In this section, the first of these types is considered, while the following section deals with the latter.

With a diode placed at each end of the diffuser, average linearity over the sweep was within approximately 10%, and was dependent largely on the exact alignment of the reflector surface; small irregularities and blemishes caused areas of slightly lower detected signal amplitude. With a third diode centrally placed in the detector for increased signal amplitude, response became peaked towards the centre of the sweep. Although this was employed for testing and scanning small map sheets, a more precisely shaped reflector would allow a smaller detector diffuser, and the two end-placed diodes would be sufficient. Alternatively, the use of a wide-angle lens to direct the incoming light to a small area would greatly increase detected signal amplitude levels. Modification in reflector position and size would be required if such a lens were used, as with the present system geometry, the lens would have to cover an angle of 180 degrees. Figure 8.6 shows the result of scanning a strip of patterns lying across the sweep from the centre (8.6(a)) progressively outwards to the edge (8.6(e)).



(a) (b) (c) (d) (e)

Figure 8.6 Response fall-off towards sheet edge

It became apparent in the course of testing that linearity over the sweep was quite highly dependent on the positions of the optical components of the scanner, and in particular, of the reflector and the detector. For optimum linearity, some form of micrometer adjustment of detector position would be necessary. One method of linearising the response over the sweep would be to use an automatic gain control stage in the detector amplifier and set the time constant to be of the order of the time taken to sweep an inch or so on the map; this could be implemented quite easily in the intermediate stage of amplification.

8.2.6 Localized non-linearities in response

The second type of non-linearity, that over localized areas, was more responsible for signal degradation on the pixel scale. Several factors were responsible for this degradation. First, the clear plastic from which the map tray was constructed was found to affect the passage of the laser beam in a non-uniform manner; driving the trolley with no map in the tray gave a signal amplitude level which varied by about 10% consistently, and by a substantial amount more occasionally. This seems to be caused by interference with the transmitted beam, due partly to fine score marks on the surface of the plastic, and partly to the nature of the structure of the material itself, and stresses within it from the bending. As a test, the trolley was removed and a sheet of clear perspex was positioned in its place; moving the perspex caused no variation in steady signal level (with the exception of fluctuations in laser power output described below). However, the perspex was not stressed by bending. Consequently, if improved localized linearity in response is required, further investigation into a suitable material from which to construct the map tray should be made.

The second factor isolated as partly responsible for localized signal degradation is the power output fluctuations of the laser. With no part of the scanner moving, and taking a measurement from the d.c. coupled (1st) stage of the amplifier, variations in detected signal level of around 5%, and up to 10%, were encountered. These variations were not periodic in any way; rather they occurred at random, with time constants of the order of seconds, and in a less determinable way, with shorter time constants. This amount of variation is greater than that which would be expected due to reflection of a randomly polarised beam, the polarisation ellipse orientation of which is varying.

Finally, general localized signal degradation was caused by mechanical vibrations from external sources. The scanner was located in a non-ideal position, on an upper building level close to an air conditioning unit which exhibited resonances with a period of the order of a second or so. In addition, treading in the vicinity of the scanner produced small vibrations, which when transmitted through the frame to the optical components, and in particular to the laser, caused visible jumps in the location of the line produced by the flying spot. These vibrations also affected the start-of-line detector pulse; although it could be made less sensitive to variation in spot alignment by using a detector with a larger surface area, this would reduce the precision with which the start of each line was registered. It became clear, as various signal degradations were isolated as being caused by mechanical vibrations, that more careful mounting of the entire structure would be required to obtain consistent performance. It was also suspected that a perceptible play in the bearings of the mirror motor was responsible for slight variations in spot positioning.

It should be noted that although localised signal level fluctuations caused by any of the factors mentioned above could be quite substantial, this does not necessarily imply that the data collected at these points will be affected seriously. If a fluctuation occurs while the signal level is high, there will be a noise immunity equal to the difference between the high level of the signal and the threshold level. If it occurs on the rising or falling edge of the pulse, some immunity is provided by the hysteresis of the comparator used for thresholding. In both cases, the sampled signal will only be affected if a change in the digitized level occurs at the sampling instant, and this in turn depends on the time extent of the irregularity. Thus the extent of the effects of these localised non-linearities in response on the final signal integrity is dependent on both their magnitude and time extent, and is consequently difficult to predict.

8.2.7 Alternative sampling scheme

Due to other non-linearities in the system, a careful evaluation of the effects of the alternative sampling could not be made. Figure 8.7 shows a series of sweeps made repeatedly along a single line across a map sheet. The alternative sampling scheme was switched in at the instant indicated by the arrow. The effect of the scheme should be to increase the width of a line by on average half a pixel, or .002".

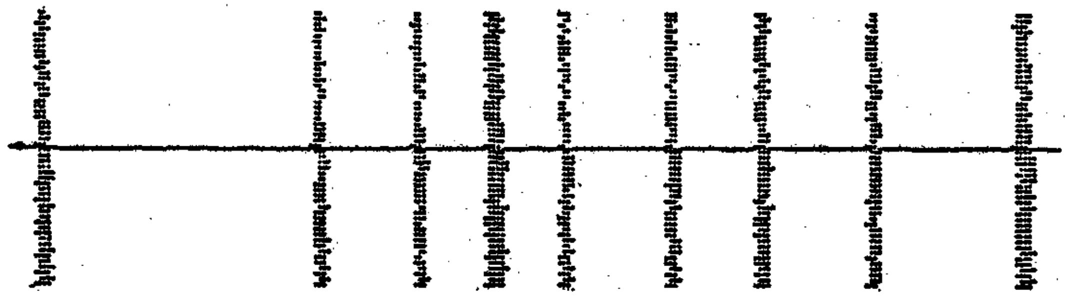


Figure 8.7 Effects of alternative sampling on line widths

Counting the total number of set pixels in the 25 sweeps before switching it in gives an average of 25.68 set pixels per sweep; and for the 25 lines after switching it in, of 30.08 set pixels per sweep. This is a difference of 4.4 pixels, and since 9 lines were swept, this corresponds to an average increase in width of 0.49 pixels, which is very close to the predicted 0.5. Hence it can be deduced that the logic of the alternative sampling scheme was functioning as designed.

8.3 Performance Scanning Map Sheets

This section looks at the performance of the system while actually scanning map sheets. Three sheets were chosen for scanning; a contour sheet with minor contour lines of thickness .002" (a majority of the lines on the sheet), and major contour lines of thickness .007"; a culture sheet with various line widths, and in addition narrow space widths between road boundaries; and a drainage sheet. Actual size positives of the scanned negatives are shown in figures 8.8.1, 8.9.1 and 8.10.1 respectively. Figures 8.8.2, 8.9.2 and 8.10.2 show full size plots of the scanned data; plotting was performed on a raster plotter (Versatec 311-8222) at 200 points per inch, giving a magnification from the original of 1.25. It should be noted that due to the mis-sized trolley drive pulley, there is also a compression of data in the travel direction by a factor of 1.1. Figures 8.8.3, 8.9.3 and 8.10.3 show enlargements of selected sections of each sheet such that individual pixels can be seen. Each of these enlargements represents an area of 128 x 128 pixels, or approximately 1/2" x 1/2", on the original sheet.

Following the analysis and discussion in chapter 5, thresholding at 30% of the maximum signal level was decided upon for the contour sheet, in order to pick up the lines of width .002" as connected lines. Judging from the results, this level appears to be about the best compromise. It cannot be further lowered if contour separation is to be maintained, while isolated breaks in narrow line continuity do still occur occasionally. However, these breaks only occur at the edges of the sheet (figure 8.8.3 is at the edge of the sheet and was selected for enlargement for this reason) where signal amplitudes are lower. In fact, the edge mean threshold level was just over 40% of the original. Breaks in continuity occur largely in sections of lines parallel to the sweep. This is in keeping with the fact that resolution spacing is .0044" in the travel axial direction, rather than the designed .004", due to the over-sized trolley drive pulley. Otherwise, positionally the breaks occur at random, suggesting that a combination of laser power fluctuations and irregular map tray transmission characteristics as described in section 8.2 are causing the signal levels to occasionally dip below the threshold level.

Thresholding was also performed at 30% when scanning the culture sheet. However, this sheet does not contain narrow lines, and looking at figure 8.9.3 (chosen from the centre of the sheet where detected signal levels are highest), it is apparent that lines have been overwidened; there is a small number of points where connectivity exists across a space. measurement of the line and space widths on the original showed space widths to be marginally larger; since space detection is equally as important in this case, thresholding should have been made nearer, if not at, the 50% level.

On the drainage sheet, thresholding was made approximately at the 50% level, as again no narrow lines are encountered. Results were satisfactory, as shown in figure 8.10.3; even at the edge only one break in continuity occurred, which could well have been caused by a dirt fleck or signal level fluctuation of longer than average duration.

Unfortunately, the scanner had to be dismantled due to building construction, and large scale plotting and examination of the scanned data was not performed until after dismantling. Hence sheets could not be re-scanned at different threshold levels on the basis of closer examination. However, it does seem apparent that thresholding at around 50% of the maximum signal level for sheets not containing lines narrower than .004", and at around 30% of the maximum level for sheets containing narrow lines (of down to .002") generates results which substantiate the theoretical expectations derived in chapter 5, and appear satisfactory in a visual sense. Further general performance conclusions are drawn in chapter 9.

8.4 System Shortcomings and Areas for Improvement

In the course of evaluating the performance of the scanning system, various shortcomings in design and construction were uncovered, and in addition, some possible improvements became apparent. This section looks at the mechanical, optical, electronic and computational aspects of the system, isolating problems in each and discussing methods to reduce or eradicate them.

8.4.1 Mechanical

It became clear during testing that most of the problems encountered while operating the system were of mechanical cause. These problems are summarised as follows:

1. Vibrations transmitted through the frame of the scanner caused minor deflections of the beam from its intended path, resulting in localised signal degradation, and causing occasional mis-triggering of the start-of-sweep and start-of-line signals.

2. Frictional drag in the bearings of the mirror motor caused minor localised variations in motor speed, and hence localised signal degradation.

3. Perceptible play in the bearings of the mirror motor caused a slight variation in beam deflection angle, and hence again localised signal degradation. This was markedly more noticeable after the motor had been running for some time; increasing tolerances due to thermal expansion are probably responsible for the increased slop.

4. The three detectors were physically positioned by bolting them to portions of the scanner frame. This made fine positional adjustment difficult, and methodical adjustment almost impossible. It became apparent that the detectors require a precise micrometer positioning device; in the case of the start-of-line and start-of-sweep detectors, adjustment in a single dimension is sufficient, but the data detector must be positioned accurately in three dimensions.

The vibrational problems indicate that the scanner must be located on some sort of vibration free platform if improved localised signal integrity is required. Alternatively, a compromise might involve the use of vibration damping feet, and more careful mounting of the moving components of the scanner on the frame. The flywheel on the mirror motor should also be more precisely balanced.

It might be possible to reduce problems caused directly by the mirror motor (2 and 3 listed above) by adjusting the bearings to reduce play, and by using a larger flywheel. In fact flywheel size was fairly arbitrary; since the motor shaft extended only on one side of the motor, a balanced flywheel arrangement could not be used, and a flywheel of approximately the same mass as the motor, but of a large diameter, was used. However, a more satisfactory solution would be to replace the motor with one of higher quality, perhaps with air bearings, designed for use in a scanning system. Such motors are commercially available.

8.4.2 Optical

Limitations in the optical system were encountered as follows:

1. The minimum spot size that could be obtained with the optical system used was between .004" and .005"; it was found desirable that a spot size of approximately .002" should be at least obtainable, even if not often used.

2. Optical alignment was performed visually, using the laser beam as a trace. Consequently, it is unlikely that the resulting spot size was actually the smallest that could be produced with the components used. In addition, it was difficult to determine if the spot shape was distorted due to lens mis-alignment.

3. Variations in laser power output were encountered, adding to uncertainty in the level at which the detected signal is thresholded.

4. Imperfections in the light localising reflector surface in the form of bumps, depressions and a slight overall skew meant that although over most of the length of the sweep the beam was focused to a small volume around the intended point of detection, deviations were caused by these irregularities. In order to prevent data collection "holes", a relatively large cylindrical diffuser was necessary (approximately 1" in length and 3/4" in diameter). This meant that detected signal amplitude was low, making a high amplifier gain necessary, and hence reducing signal to noise ratio. As the beam occasionally hit the edges of the diffuser linearity over the sweep was not particularly good.

5. The map tray was found to affect the passage of light through it in a non-uniform manner. Interference (due to fine scouring and the nature of the structure of the plastic) caused localised signal degradation.

Although using a microscope objective as a primary lens and a camera lens as a secondary lens achieves collimation and beam focusing, such lenses are designed for compromising compensation for the various major optical aberrations. In particular, in a camera lens, effort is made to correct for chromatic aberration. However, a laser is a monochromatic source, and hence lenses may be designed for use at a single wavelength, and can be corrected primarily for spherical aberration which is largely responsible for image degradation in laser optics. Hence an optical collimator designed for use with a laser light source is likely to have an improved performance over that of the lens system used, bringing the spot size closer to that theoretically obtainable with an ideal optical system with a specified degree of collimation.

The light localising reflector surface was made from a strip of polished stainless steel, attached to the curve cut from perspex. problems were caused by imperfections in the steel strip, rather than in the curve shape. It would be possible to have the curve cut from solid steel, and have an exact surface ground and polished directly; since the curve need not be very wide this would be quite feasible. However, the fact that performance was satisfactory apart from the small imperfections suggests that the designed reflector would be quite adequate if more care were taken to ensure that the strip had no such blemishes. Slight problems were encountered with the steel rising slightly off the curve; a thinner strip would have less tendancy to do so.

Signal degradation due to the map sheet tray should be entirely eliminated by choosing a material which neither scuffs easily, nor noticeably affects the passage of the laser beam. Glass would be ideal, but due to its weight, and the comparative difficulty of forming it into a curve, it is less attractive than a plastic. Preliminary investigations showed perspex to be suitable, though the effects on transmission of stress due to bending should be investigated.

8.4.3 Electronic

Performance of the electronic components of the system was as designed, with the following provisos:

1. It is possible that the low-pass filter which picks out the fundamental component from the 61 Hz square

wave to drive the motors should have a steeper roll-off than 12 dB per octave. However, as mentioned earlier, periodic mirror speed variations were entirely masked, if at all present, by speed fluctuations due to frictional forces. If a motor of higher quality, possibly with air bearings, were installed, the effects of the residual third harmonic should be investigated more carefully; this could be done by scanning a fine grid at various points over the sweep and checking for consistency in spacing. If necessary, the roll-off of the filter could be increased by cascading an identical stage. (Amplifier gain would have to be marginally increased.)

2. It became apparent that the use of an automatic gain control (with a several millisecond time constant) to make detected signal response over the sweep more linear would have been a simple way to compensate for these non-linearities. Although this would result in a variable signal to noise ratio, maintaining a consistent maximum signal level, and hence threshold level, would result in a more consistent degree of data integrity over the area of the map sheet. This would not be difficult to insert in the intermediate stage of the amplifier (an integrated circuit A.G.C. such as

the LM 370 would be suitable) as a switchable option; time constant adjustability would also be desirable.

8.4.4 Software

The interface software performed data input as designed, though several points should be mentioned.

First, the software using the interrupt to initiate data collection for each line was operative under RSX-11M version 3.1. With the installation of version 3.2, and some new hardware in the computer system, the interrupt vector address assigned to the DR11-C was moved out of the range recognized by the operating system. Recognition of this vector address required that a software system generation be performed on the computer to re-define the addressing area; this was not done before the dismantling of the scanner. Consequently most data input was performed with a slightly modified version of SCAN which simply waited for the interrupt request line to be set after each sweep. This has the disadvantage that other use of the CPU of the computer is blocked throughout scanning.

The second point is that initially, unsynchronised lines appeared occasionally in the scanned data. This loss of synchronism was found to be caused when the line clock interrupt bringing in the operating system scheduler occurred immediately before the line clock interrupt was disabled by SCAN; the interface buffer was then insufficient to hold the data generated while the scheduler held control of the processor, and data collection commenced after the start of the line. There are a number of possible solutions to this problem, but rather than increase the depth of the hardware interface buffer, or interfere too deeply with the operation of the system, it was decided that the distance between the start-of-sweep and start-of-line detectors should be increased so that sufficient wait time (between disabling the line clock interrupt and receiving the first word of data) be provided that the scheduler can complete its task without interfering with data collection. Although this increases slightly the time for which the scanning software blocks other users from the processor, it has the advantage that it solves the problem for data input under either the interrupt mode, or that described above.

Thirdly, a situation might exist in which the start-of-sweep detector generates a signal, but the start-of-line detector does not, due perhaps to physical vibration or intermittent electrical fault. In this case, a line of data would be lost, and the map compressed in the travel direction. The simplest way of pinpointing if this should occur would be to increment a counter at each start of sweep signal detection, and place the count into a common area of memory accessible to all the programs. SCAN could then check this count against its line count, and if a discrepancy should occur, either exit with a message to that effect, or continue but record the line number, passing it to CONTRL at the end of scanning.

The points mentioned above demonstrate that although the software was not developed to the extent that would be desirable for a continuously operating system (due partly to the dismantling of the scanner and the unavailability of the interrupt vector address), the flexibility maintained was advantageous in the developmental system. Having various possible modes of operation was found useful not only for testing, but also in dealing with the loss of interrupt facility. There is another reason why software complexity should be kept to a minimum: in a computer operating system such as the one used, task priorities, modes of running and general system procedures must be tailored to ensure smooth system usage; making any one task or set of tasks highly dependent on certain system conditions limits the capabilities of the system. Hence keeping modes of operation as straightforward as possible, and keeping software complexity to a minimum, is highly desirable. The priorities at which the data input programs were run, and the event flags used for communication, might well have to be changed for satisfactory data input under different operating system conditions.

8.5 Summary

This chapter has evaluated as objectively as possible the performance of the scanning system. Specific tests have been performed, and the results of scanning test patterns and map sheets have been presented. Departures in performance from original design, and drawbacks in design discovered as a result of operation, have been described; their causes have been isolated and suitable remedies have been discussed.

The following chapter uses this evaluation, together with the conclusions reached in chapter 7, to consider the extent to which the scanning system as designed and constructed meets the originally stated requirements for a rapid automatic cartographic digitizing system.

Figure 8.8.1 True size positive of contour sheet

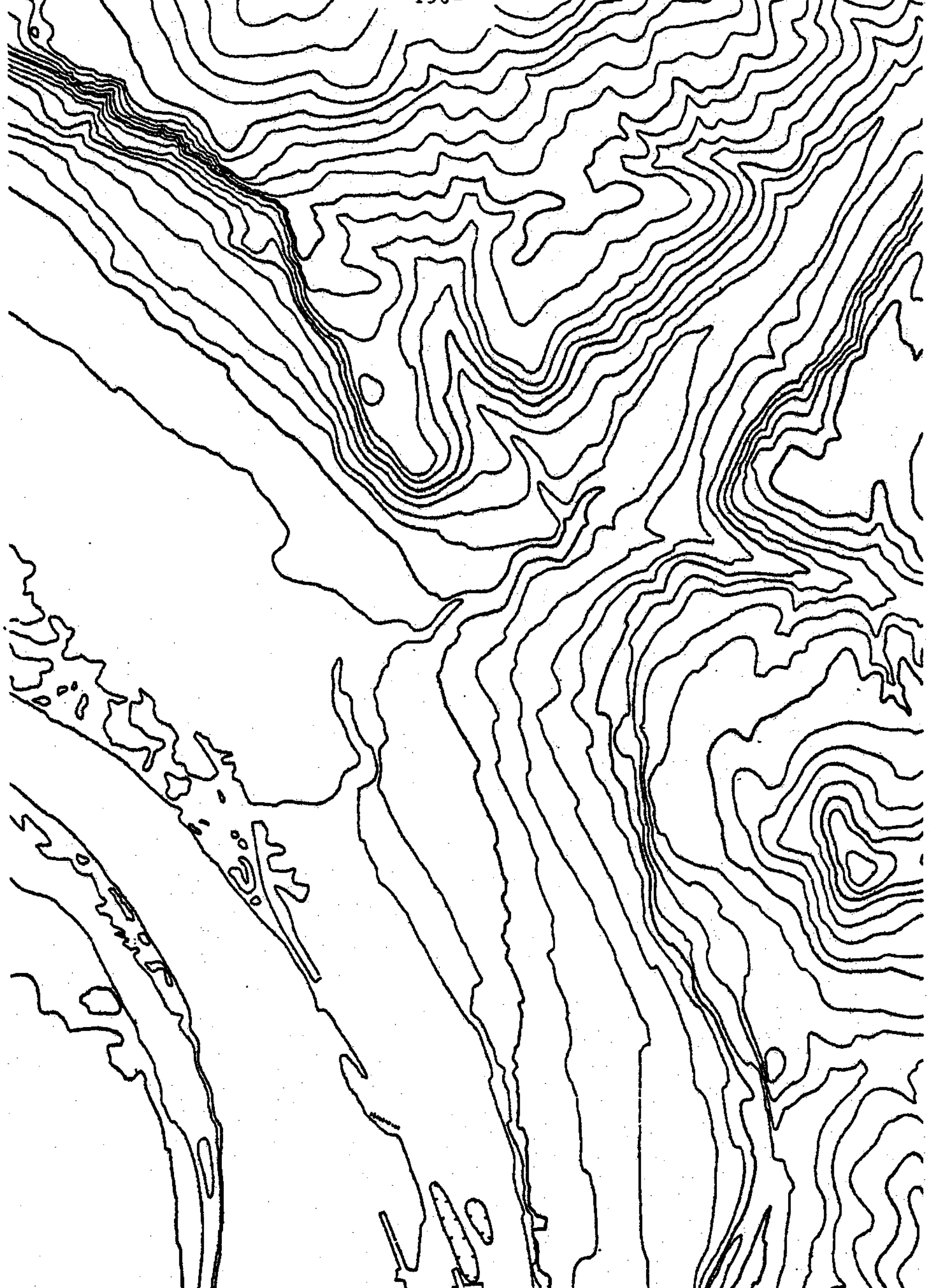


Figure 8.8.2 Matrix plot of scanned contour sheet

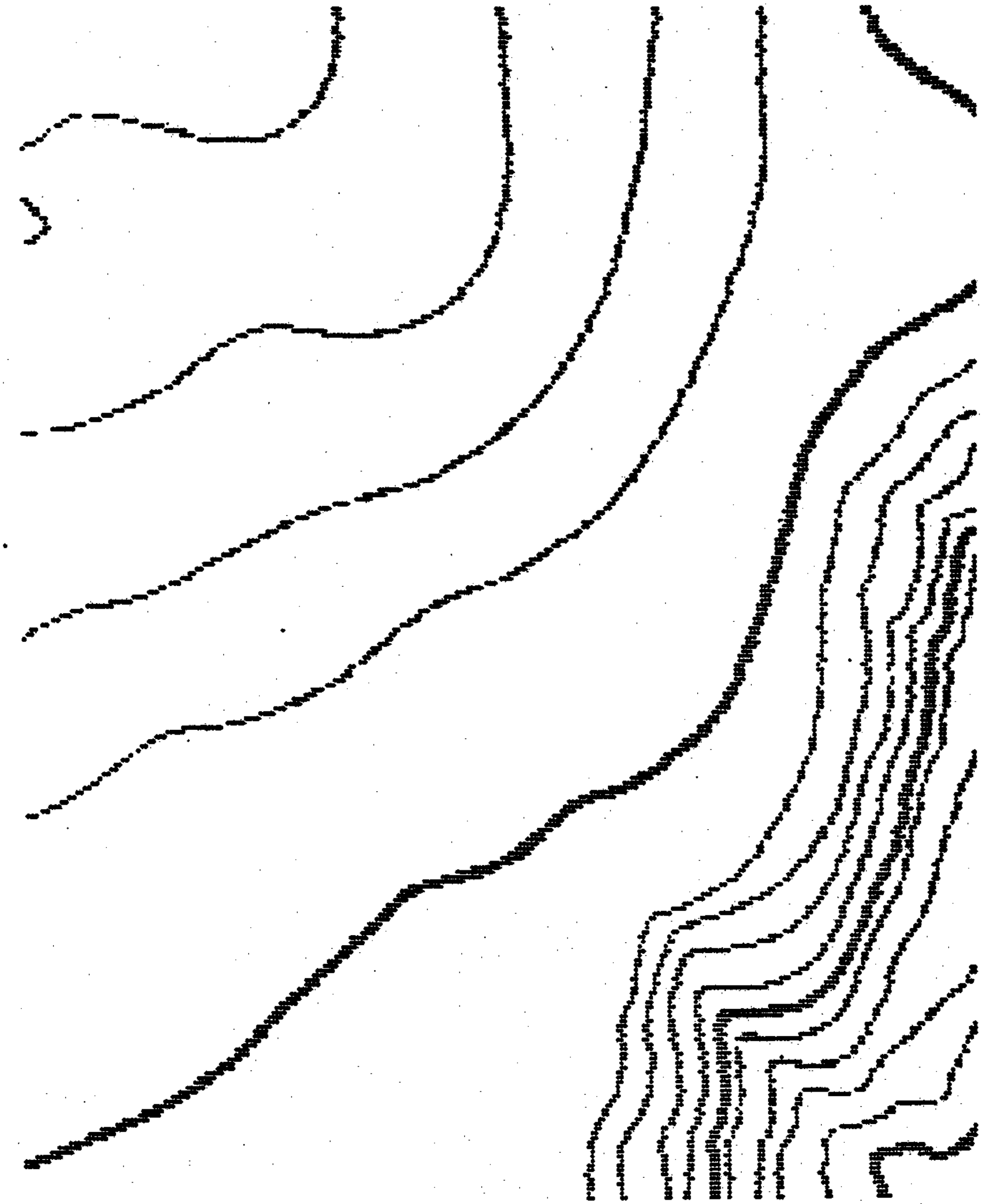


Figure 8.8.3 Enlarged section of scanned contour sheet

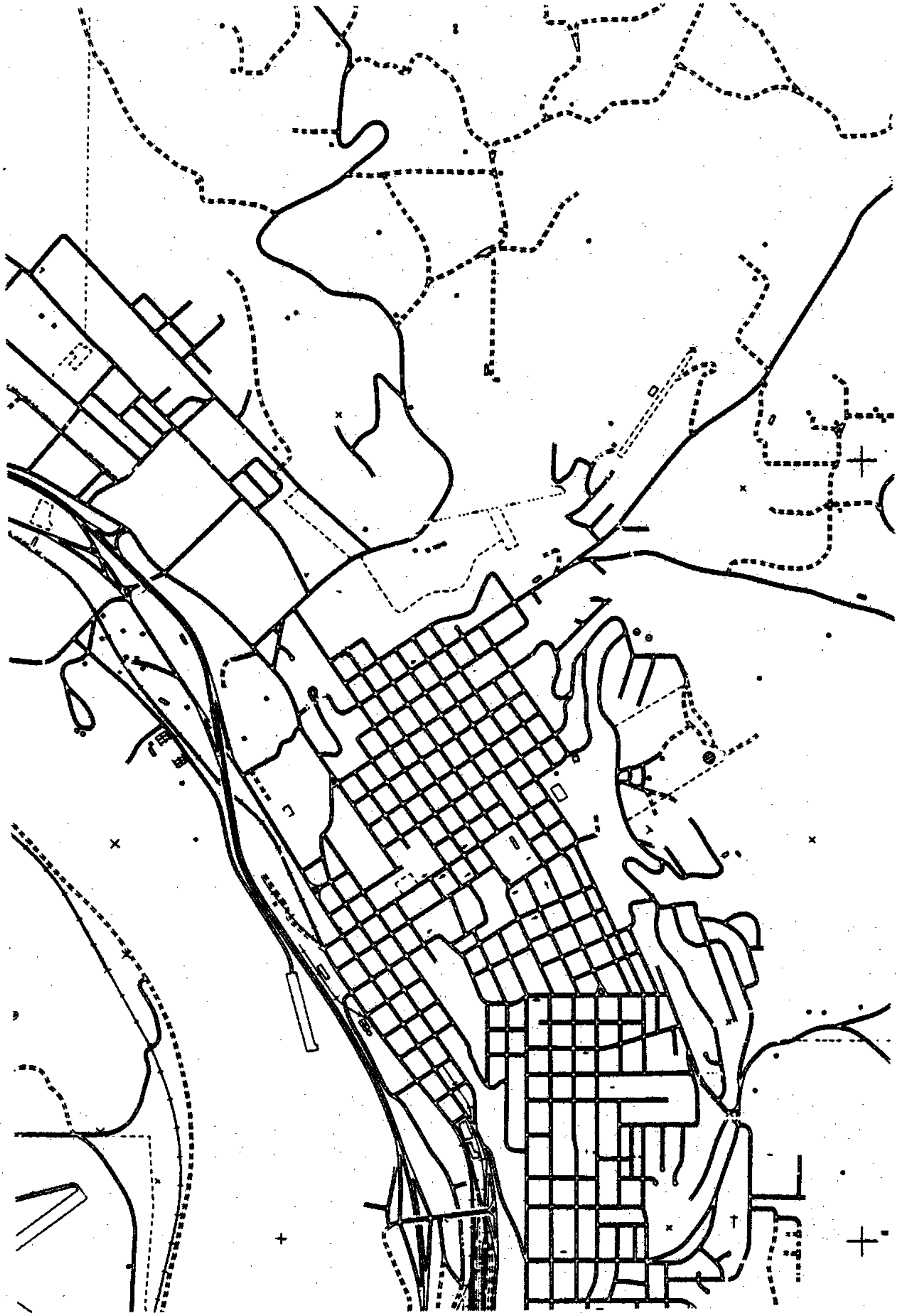


Figure 8.9.1 True size positive of culture sheet

Figure 8.9.2 Matrix plot of scanned culture sheet

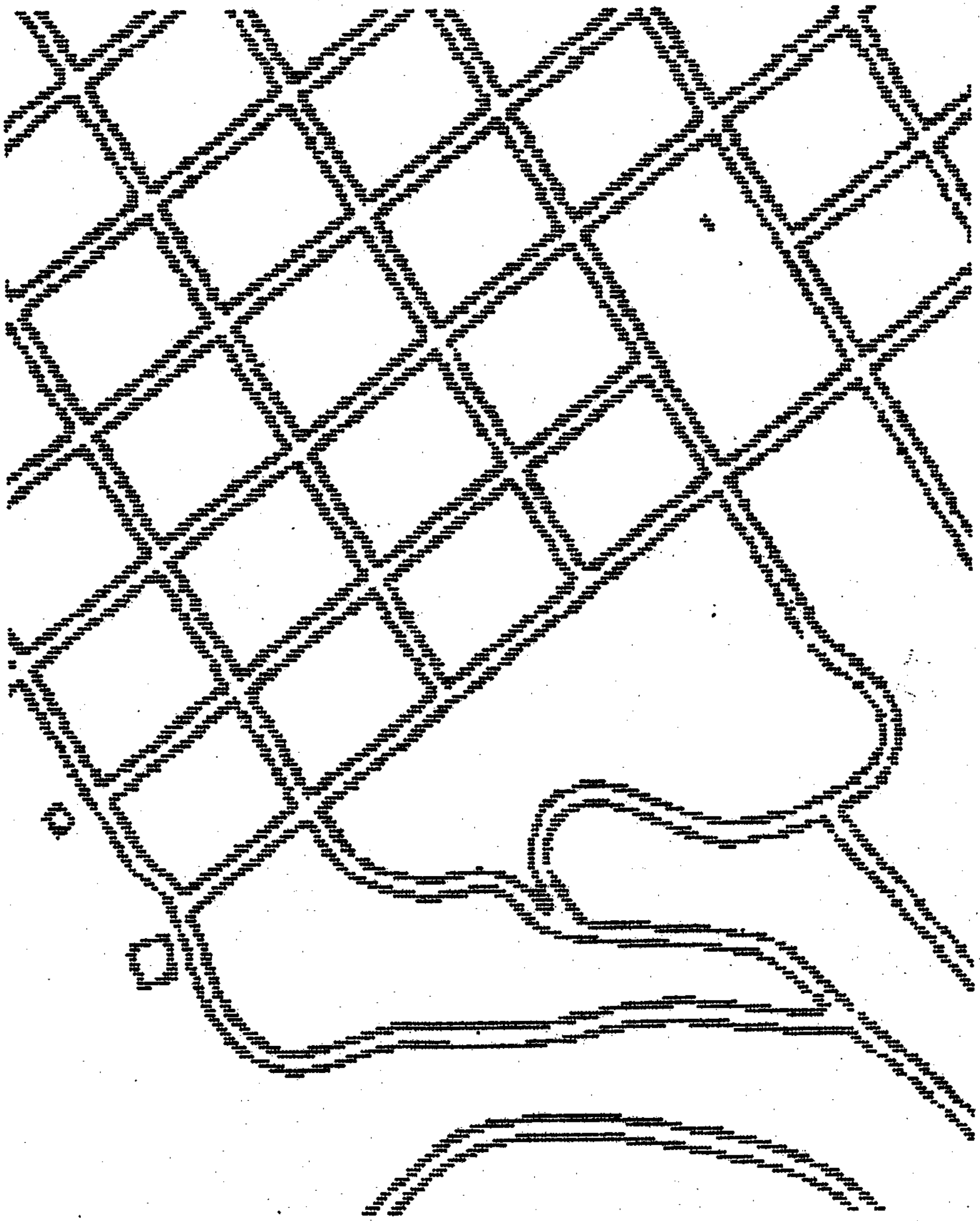


Figure 8.9.3 Enlarged section of scanned culture sheet

Figure 8.10.1 True size positive of drainage sheet

Figure 8.10.2 Matrix plot of scanned drainage sheet

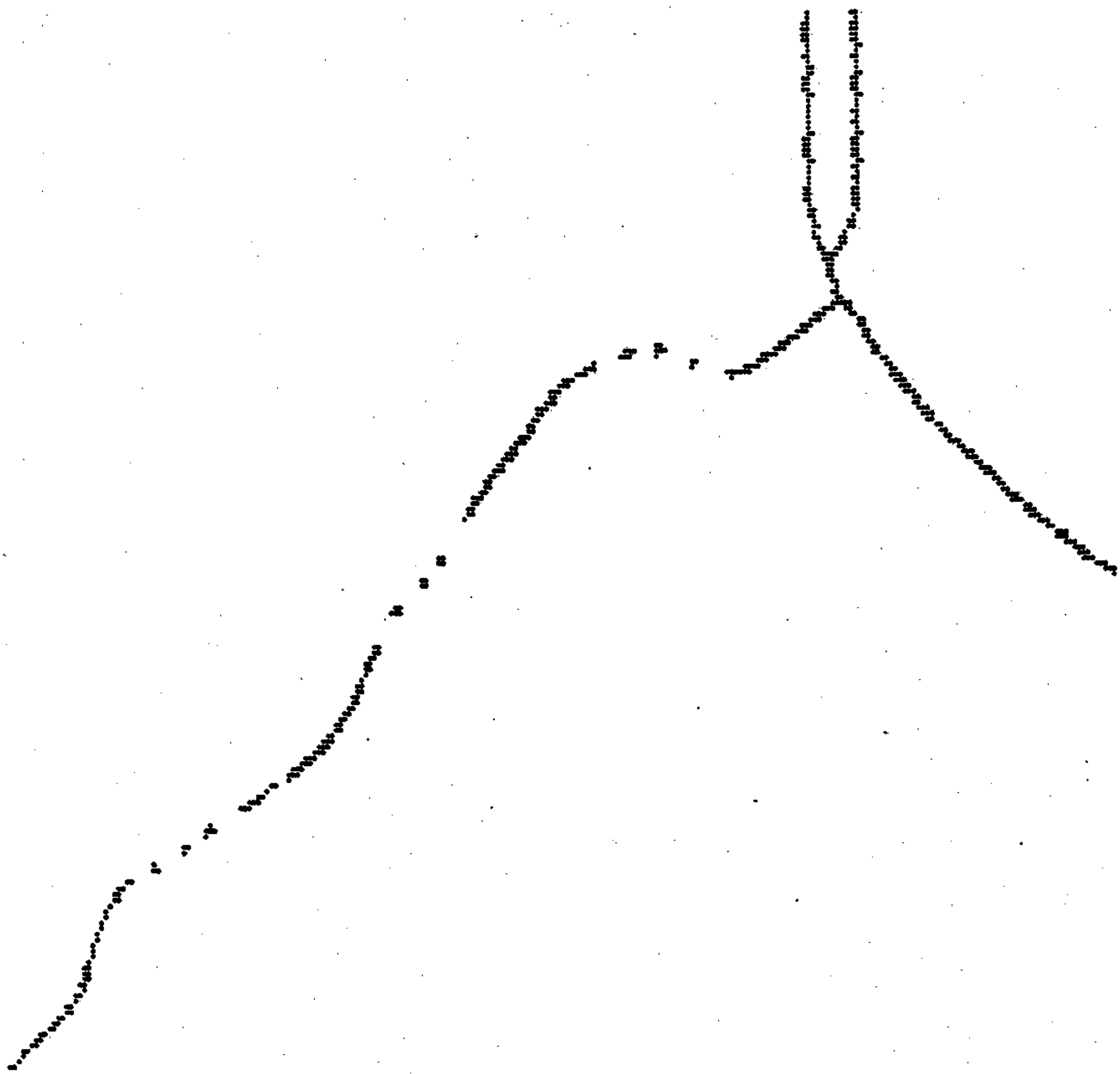


Figure 8.10.3 Enlarged section of scanned drainage sheet

9. CONCLUSIONS

The demand for cartographic data in digital form has highlighted the lack of systems capable of digitizing an area the size of a map sheet rapidly, and at low cost. This thesis has proposed such a system, consisting of a flat bed flying spot laser scanner interfaced to, and controlled by, a mini-computer. Map sheets can be digitized in a raster scanning process at a resolution of .004" (≈.1 mm.), with added facility for ensuring detection of lines as narrow as .002". Scanning a full sized map sheet takes approximately 10 minutes, and the data is temporarily stored on disk or magnetic tape prior to processing.

From the results presented in chapter 8 it can be seen that the system performs largely as designed. Some features, mainly mechanical, were isolated as being the cause of unsatisfactory or inconsistent performance, and design or constructional modifications to overcome these problems were discussed. None of these modifications would involve any major changes or cost a great deal. For a fully operational system, further software development might be necessary or desirable. Any specific operating environment will have its own characteristics within which software must operate, but the flexible structure of the design should allow room for any required alterations without affecting the overall operation of the system. Consequently it can be concluded that with the minor improvements cited, the developed system should more than adequately meet the stated requirements for digitizing cartographic data.

As was discussed in chapters 1 and 7, such a system would only be of use if the generated data can be comfortably handled and stored. The system was designed around a mini-computer, as this is the type of machine most likely to be found in an automated cartographic processing environment, and consequently it should be feasible for such a computer to handle and process the large volumes of data generated. This was found to be so; chapter 7 shows that simple run length encoding can reduce the data volumes to a level comparable with that required for vector representation, and vectorization processes can render the data compatible with existing, or manually digitized, line vector data sets.

Consequently the concept of a raster based geographic information system appears attractive. Given that map sheets can be rapidly digitized to generate raster data, large rapid access magnetic disks and powerful image display and processing systems make handling and processing the large volumes of data generated a viable proposition. With data in raster form, re-plotting either on matrix plotters, or with a modulated scanning device, then becomes a feasible means of reproducing or re-drawing updated map sheets.

Finally, it should be pointed out that application of the proposed system is not limited to cartographic digitizing; rather it would be suitable for any type of two level large area digitizing, such as of engineering drawings.

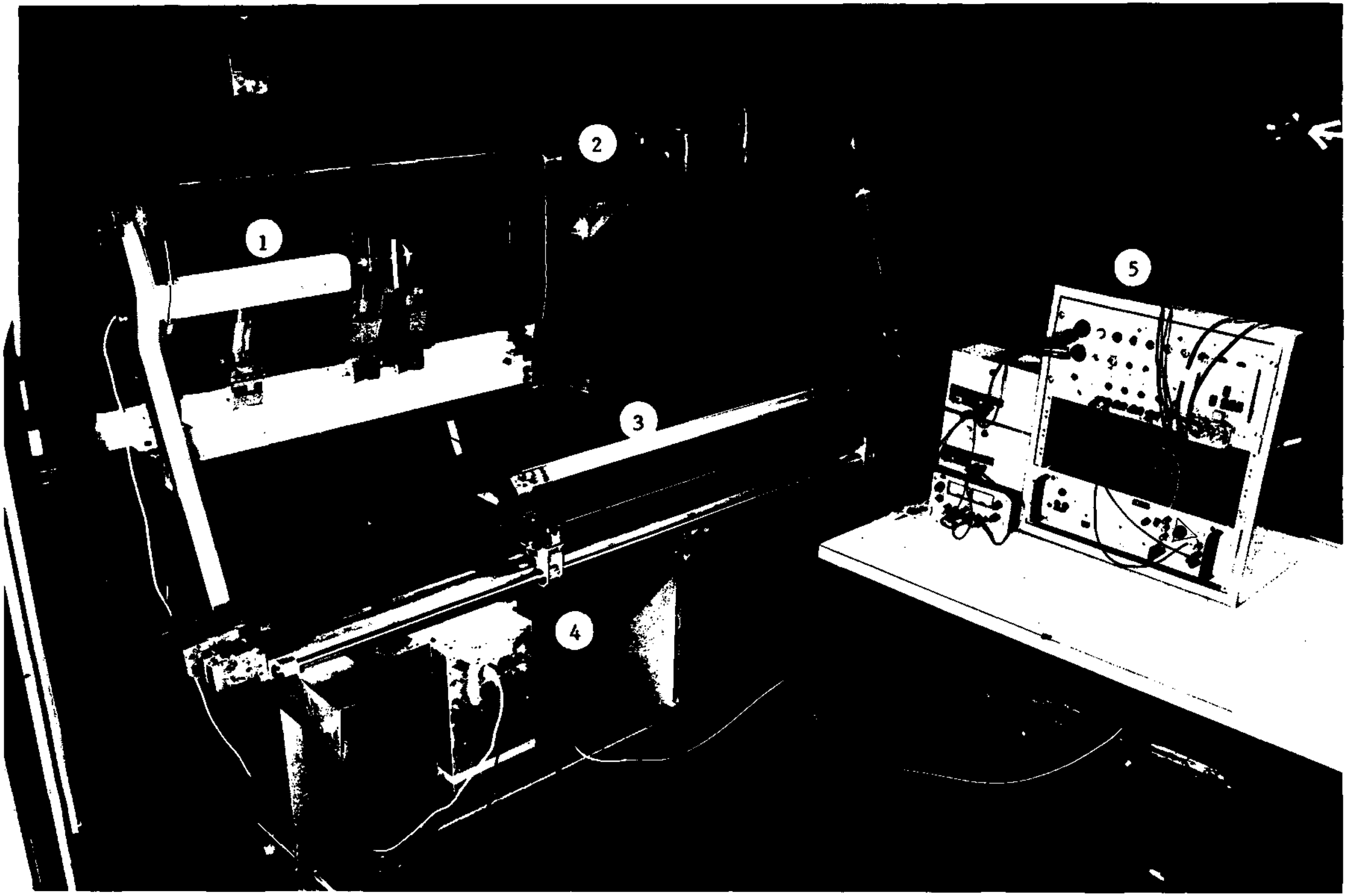


Figure 9.1 Scanner assembly and control hardware

1. Optical system
2. Sweep motor
3. Map trolley
4. Reflector
5. Motor control hardware

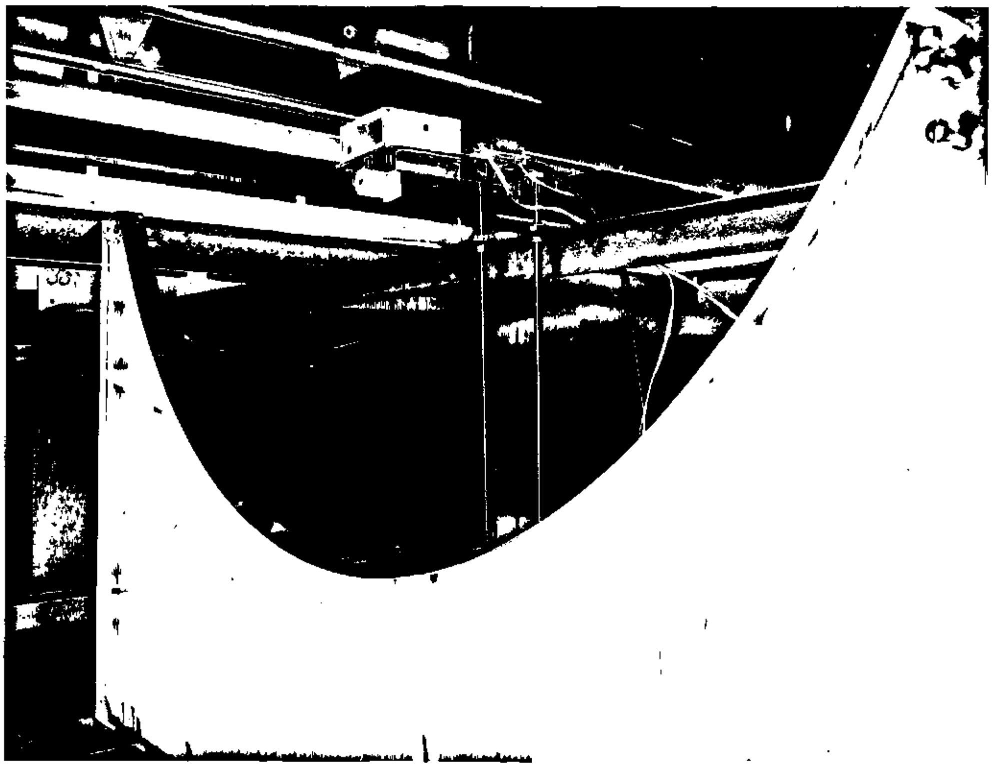


Figure 9.2 Reflector and detector arrangement

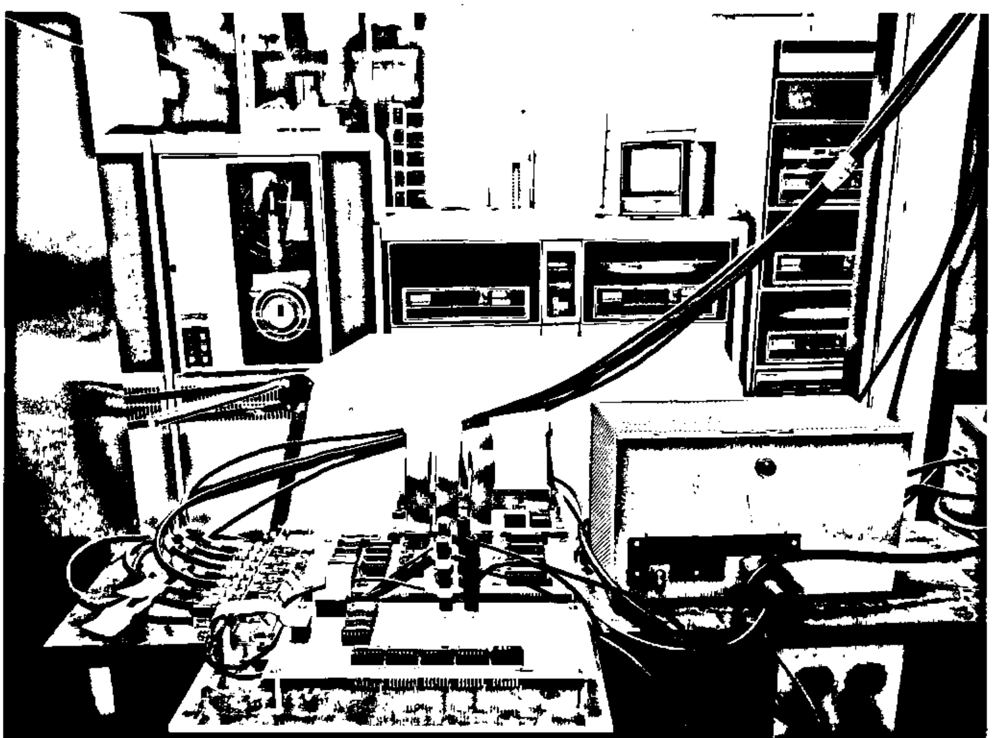


Figure 9.3 Interface hardware

10. REFERENCES

1. Amdon, E.L. and Aiken, G.S., "Algorithmic Selection of Best Method for Compressing Map Data Strings", Comm. ACM 14, 1971, p 769-774.

2. Davey, P.G.,Harris, J.F.,Preston, G., "Automatic Boundary Digitising for the Department of the Environment", Oxford University, Nuclear Physics Laboratory Ref:91/78.

3. Freeman, H., "On the encoding of arbitrary geometric configurations", IRE Trans. EC-10,2, June 1961.

4. Kogelnik, H. and Li, T., "Laser Beams and Resonators", Proc. IEEE, Vol.54 No.10, Oct 1966.

5. Kreifelts, T., "Skeletonising and Line Following in Raster-Digitised Line Structures", Presented at the GI/NTG Congress on Digital Image Processing, Munich, March 1977.

6. Levialdi, S., "On Shrinking Binary Picture Patterns", Comm. ACM, Jan 1972, Vol.15, No.1.

7. Miketinac, J., "Raster to Vector Conversion of Map Data", University of Saskatchewan, Dept. of Electrical Engineering Internal Document, 1980.

8. Mueller, C.S.M. and Christensen, A.H.J., "SCANLINE: a computer program for converting raster maps into line maps", NRIMS of the CSIR, Pretoria.

9. <u>Optoelectronics Applications Manual</u>, Prepared by The Applications Engineering Staff of the Hewlett-Packard Optoelectronics Division, McGraw Hill, 1977.

10. Peuquet, D.J., "Raster Processing of Spatial Data in Geographic Information Systems", Ph.D. dissertation, State University of New York at Buffalo, 1978.

11. Rosenfeld, A. and Kak, A.C., <u>Digital Picture Processing</u>, Academic Press, New York, 1976.

12. Rosenfeld, A., "Connectivity in Digital Pictures", J.ACM 17, Jan 1970.

13. Rosenfeld, A., "Extraction of Topological Infromation from Digital Images", Paper presented at Harvard Symposium on Topological Data Structures for Geographic Information Systems, Oct 1977.

14. Saraga, P., "Thinning Operators", Mullard Research Laboratories, Technical Note No.1294, Feb 1974.

15. Seuffert, P., "An Application of Line and Character Recognition in Cartography", IEEE Congress on Pattern Recognition, New York, July 1977.

16. Spectra Physics, Inc., "Design of Optical Systems for Use with Laser Beams", Laser Technical Bullletin Number 5, Mountain View, California, 1966.

17. Turke, K., "Digitizing of Geometric Data for Thematic Mapping: State-of-the-Art and Future Developments", Translated version from Reihe II: Ubersetzungen-Heft Nr 35, 1978.

18. Woetzel, G., "A Fast and Economical Scan-to-Line Conversion Algorithm", Gessellschaft fur Mathematik und Datenverarbeitung MHB, Bonn, Dec 1977.

APPENDIX A

A.1 Requirements and design of beam focusing system

When a parallel beam of light is focused to a point, the intensity distribution in the plane perpendicular to the optic axis at the focal point (the focal plane) can be obtained, and hence a measure of spot size (defined as the diameter of a circle within which some fraction of the total energy in the beam lies) can be determined. In simple optical systems, where rays are paraxial, spot size in the focal plane is directly proportional to the f number of the optical system, and depth of focus is proportional to the square of the f number.

A system can be analysed in terms of normalised dimensionless co-ordinates u and v, in the direction of the optic axis, and in the plane perpendicular to it, respectively. Analysis is then independent of f number. If z and r are the actual physical dimensions parallel and perpendicular to the optic axis respectively, then u and v can be expressed in terms of z and r:

$$u = kz/4f \qquad v = kr/2f \qquad \text{where} \quad k = 2\pi/\lambda$$

If we consider the focusing of an ideal Gaussian beam to a point O on the optic axis, the intensity in the vicinity of

O can be given approximately by [16]:

$$I(u,v) = \frac{Io}{\pi(1+u^2)} \exp\left[\frac{-v^2}{(1+u^2)}\right]$$

where Io is the total energy of the beam.

In the focal plane u=0, and hence $I(v) = (Io/\pi)\exp(-v^2)$.

Along the optic axis v=0, and hence $I(u) = Io/[\pi(1+u^2)]$.

In the focal plane, a circle of radius v=1 contains 63% of the total energy, while one of radius v=1.5 contains 78% of the total energy. This forms a useful approximation to spot diameter (as defined by the $1/e^2$ points on the intensity distribution, where a circle of this radius contains 86% of the total energy).

Hence if we require a spot diameter < .004", or a radius of < .002" (or < 50 microns);

$$v = kr/2f, \quad \text{or} \quad f = kr/2v$$

$$\text{hence} \quad f < 166 \quad \text{for } r < 50 \text{ microns}, \; v \approx 1.5$$

For the given scanner geometry, $z \approx 100$ cm, and from the specifications of the laser used, beam radius is approximately 0.5 mm. Thus, $f \approx 2000$, and the beam must be widened by a factor of $2000/166 \approx 12$ to achieve a spot diameter of approximately .004". Beam widening, or collimation, can be achieved by using a pair of lenses, the ratio of their focal lengths being the beam widening factor.

A high quality microscope objective lens of nominal focal length 4 mm (and effective focal length 4.4 mm) was used as a primary lens. A secondary lens of focal length greater than 50 mm is then required to be able theoretically to obtain a 78% spot diameter of .004". Using a 100 mm secondary lens, a minimum 78% spot diameter of .002" is theoretically obtainable. These correspond to $1/e^2$ spot diameters (86%) of approximately .005" and .0025" for the 50 mm amd 100 mm lenses respectively. However, the presence of spherical and other aberrations, and possible non-alignment of the optical system components, mean that actual spot diameters will be somewhat larger.

In the presence of spherical aberration the proportion of the total energy which lies within a circle of given radius in the focal plane decreases as beam truncation increases. Increasing the f number of a photographic lens reduces spherical aberration, but may also cause, or if already present increase, beam truncation. Ideally, design should compromise betweem the spot expanding effects of beam truncation and the spot contracting effects of reducing spherical aberration. In practice, the aperture of the secondary lens was reduced until a reduction in detected signal amplitude was obtained, indicating that discernible truncation was taking place.

A.2 Derivation of ray transfer matrix and lens positions

Following Kogelnik and Li [4], the propagation of paraxial rays through simple optical structures can be described by ray transfer matrices. These matrices relate the displacement x with respect to the optic axis of the incoming ray, and its slope x′ at the point of entry, to the corresponding displacement and slope of the outgoing ray, y and y′, respectively, by:

$$\begin{bmatrix} y \\ y' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ x' \end{bmatrix}$$

where A, B, C and D can be related to the focal length of the elements and principal, input and output plane separations of the optical system.

The matrices for ray transfer over a distance d, and through a thin lens of focal length f, are given in table A.1. The matrix for a combination of a lens of focal length f, followed by a distance d, can be obtained by matrix multiplication:

$$\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} = \begin{bmatrix} (1-d/f) & d \\ -1/f & 1 \end{bmatrix}$$

Hence ray passage through two successive lens/distance combinations of focal lengths and distances f1 and d1, and f2 and d2 respectively, can be obtained as follows:

$$\begin{bmatrix} 1-\frac{d2}{f2} & d2 \\ \frac{-1}{f2} & 1 \end{bmatrix} \begin{bmatrix} 1-\frac{d1}{f1} & d1 \\ \frac{-1}{f1} & 1 \end{bmatrix} = \begin{bmatrix} 1-\frac{d2}{f2}-\frac{d1}{f1}+\frac{d1d2}{f1f2}-\frac{d2}{f1} & d1-\frac{d1d2}{f2}+d2 \\ \frac{-1}{f2}\ \frac{+d1}{f1f2}\ \frac{-1}{f1} & \frac{-d1}{f2}+1 \end{bmatrix}$$

For a parallel beam entering the primary lens, $x'=0$ for all x. A condition for the lens system to focus the beam to a spot is that for the composite system, $y=0$ for all x. Hence $y = 0 = Ax$ must be satisfied. Equating the A term of the matrix for the composite lens system to zero results in:

$$1-\frac{d2}{f2}-\frac{d1}{f1}+\frac{d1d2}{f1f2}-\frac{d2}{f1} = 0.$$

For the chosen scanner geometry, $d1+d2=D$ is a fixed value. Substituting D-d1 for d2, we get a quadratic equation for d1 in terms of f1, f2 and D:

$$d1^2 - d1(D+f1) + (Df1+Df2-f1f2) = 0$$

Substituting for D = 112 cm, f1 = .44 cm, f2 = 5 cm, and solving for d1, we obtain d1 = 5.7 cm,

and so d2 = 106.3 cm.

Figure A.1 shows the resulting optical system, and table A.2 summarises its specifications.

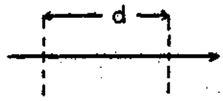
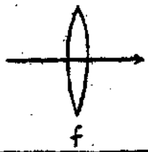
NO	OPTICAL SYSTEM	RAY TRANSFER MATRIX
1		$\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$
2		$\begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}$

Table A.1 Ray transfer matrices for simple optical structures

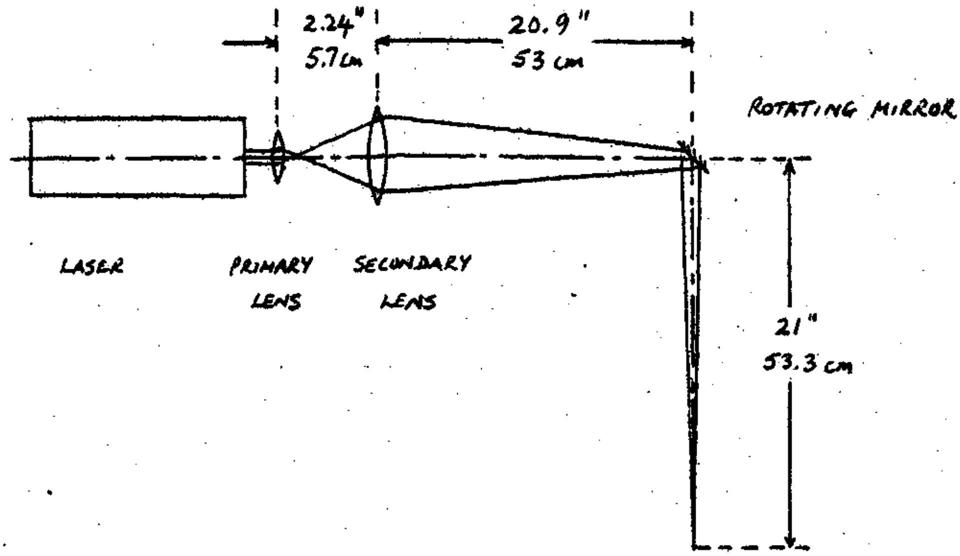


Figure A.1 Beam focusing optics

Lens	Primary	Secondary(1)	Secondary(2)
Type	Ealing microscope objective	Nikkor photographic	Nikkor photographic
focal length	4.4 mm	50 mm	100 mm
aperture	N.A.=0.65	Max: f 1.4 Used: f 4	Max: f 2.5 Used: f 4

Table A.2 Optical system lens specifications

A.3 Derivation of reflecting curve for light localization

Consider the curve given by the locus of points such that an incident light beam from the rotating mirror, point A, will be reflected to a single detection point D below the map bed, as shown by figure A.2.

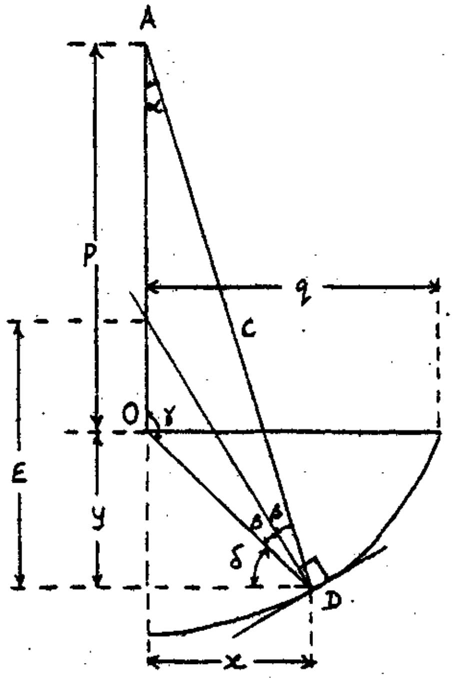


Figure A.2 Reflecting curve for light localization

The gradient of the normal at any point on the curve is given by:

$$\text{gradient normal} = \tan(\delta+\beta) = \frac{\tan\delta + \tan\beta}{1-\tan\delta\tan\beta}$$

and hence the gradient of the tangent at any point on the curve is given by:

$$g(x,y) = \frac{-1}{\tan(\delta+\beta)}$$

As $\tan\delta = y/x$, $\quad g(x,y) = \dfrac{x - y\tan\beta}{y + x\tan\beta}$ -(1)

and thus β is required to evaluate g(x,y). Using the sine rule in triangle ABD:

$$\frac{\sin 2\beta}{p} = \frac{\sin\alpha}{(x^2+y^2)^{1/2}} = \frac{x}{(x^2+y^2)^{1/2}[x^2+(y+p)^2]^{1/2}}$$

whence $\beta = (1/2)\sin^{-1}\left[\dfrac{px}{(x^2+y^2)^{1/2}[x^2+(y+p)^2]^{1/2}}\right]$ -(2)

Equations (1) and (2) can be used to obtain a set of curves satisfying the required conditions for reflection. Inserting the boundary condition that x=q for y=0 determines a single curve, whose locus may be obtained by numerical integration as follows:

$$x=q,\ y=0 \Rightarrow \beta = (1/2)\sin^{-1}\left[\frac{p}{(p^2+q^2)^{1/2}}\right]$$

and $\quad g_0(x_0,y_0) = -1/\tan\beta$

whence given p and q, g_0 can be calculated.

Then $x_1 = x_0+\Delta x$, and $y_1 = y_0+\Delta y$, where $\Delta y = g_0\Delta x$.

Hence g can be evaluated at (x_1,y_1), and thereafter,

$g_n = \dfrac{y_n\tan\beta - x_n}{y_n + x_n\tan\beta}$ where $\beta = (1/2)\sin^{-1}\left[\dfrac{px}{(x_n^2+y_n^2)^{1/2}[x_n^2+(y_n+p)^2]^{1/2}}\right]$

and $y_{n+1} = y_n+g_n\Delta x$

A Fortran program was written to perform the numerical integration, using values of p and q measured from the scanner. (p=23.875", q=16.125".) A true scale plot was produced, and used as a template for cutting the curve.

APPENDIX B

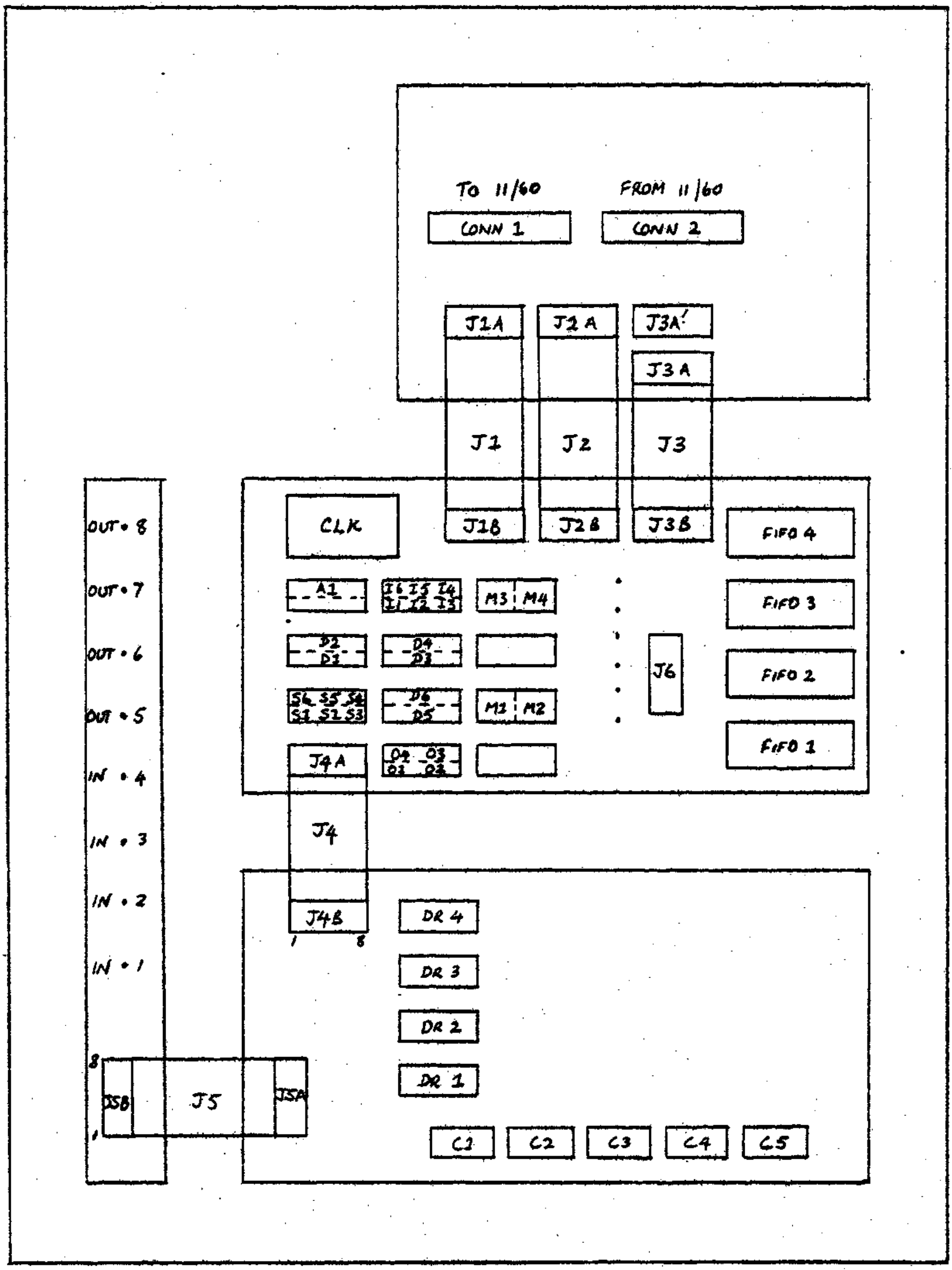


Figure B.1 Interface board layout

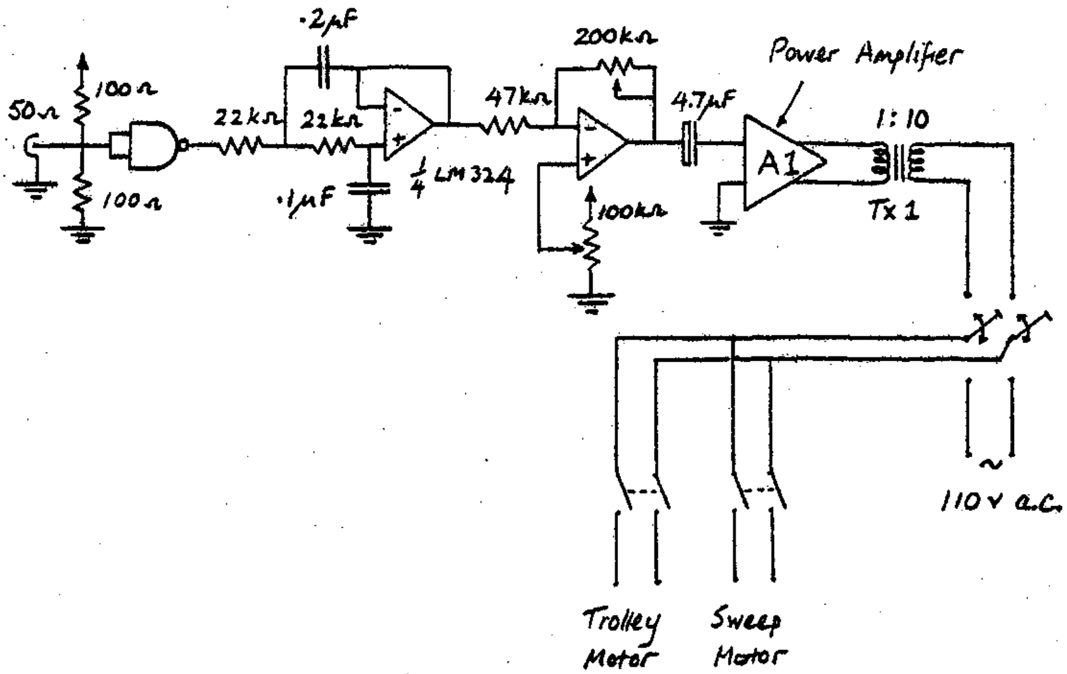


Figure B.2 Motor synchronising circuitry

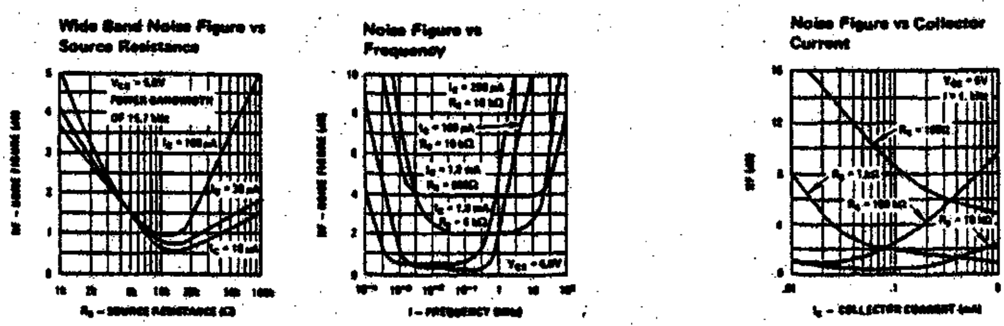


Figure B.3 1st stage detector amplifier transistor characteristics

Pin diode characteristics

HEWLETT PACKARD
COMPONENTS

PIN PHOTODIODES

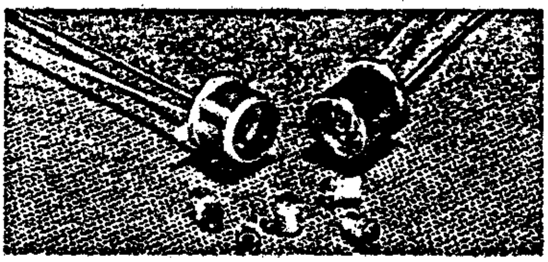
5082-4200 SERIES

TECHNICAL DATA APRIL 1978

Features

- HIGH SENSITIVITY (NEP <-108 dBm)
- WIDE DYNAMIC RANGE (1% LINEARITY OVER 100 dB)
- BROAD SPECTRAL RESPONSE
- HIGH SPEED (T_r, T_f,<1ns)
- STABILITY SUITABLE FOR PHOTOMETRY/ RADIOMETRY
- HIGH RELIABILITY
- FLOATING, SHIELDED CONSTRUCTION
- LOW CAPACITANCE
- LOW NOISE

Active area: 1mm Diam	5082-4207	TALL SIZE (TO-18)
0.5mm Diam	5082-4203	TALL SIZE (TO-18)
	5082-4204	TALL SIZE (TO-18)
	5082-4220	— Short (TO-46)
0.25mm Magnified 2.5x	5082-4205	— Subminiature



Description

The HP silicon planar PIN photodiodes are ultra-fast light detectors for visible and near infrared radiation. Their response to blue and violet is unusually good for low dark current silicon photodiodes.

These devices are suitable for applications such as high speed tachometry, optical distance measurement, star tracking, densitometry, radiometry, and fiber-optic termination.

The speed of response of these detectors is less than one nanosecond. Laser pulses shorter than 0.1 nanosecond may be observed. The frequency response extends from dc to 1 GHz.

The low dark current of these planar diodes enables detection of very low light levels. The quantum detection efficiency is constant over ten decades of light intensity, providing a wide dynamic range.

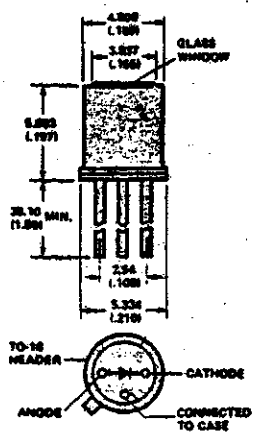
The 5082-4203, -4204, and -4207 are packaged on a standard TO-18 header with a flat glass window cap. For versatility of circuit connection, they are electrically insulated from the header. The light sensitive area of the 5082-4203 and -4204 is 0.508mm (0.020 inch) in diameter and is located 1.905mm (0.075 inch) behind the window. The light sensitive area of the 5082-4207 is 1.016mm (0.040 inch) in diameter and is also located 1.905mm (0.075 inch) behind the window.

The 5082-4205 is in a low capacitance Kovar and ceramic package of very small dimensions, with a hemispherical glass lens.

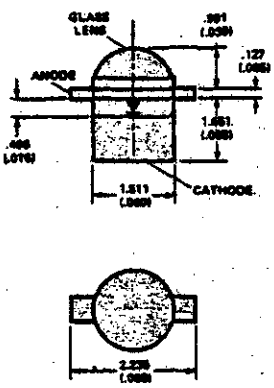
The 5082-4220 is packaged on a TO-46 header with the 0.508mm (0.020 inch) diameter sensitive area located 2.540mm (0.100 inch) behind a flat glass window.

Package Dimensions

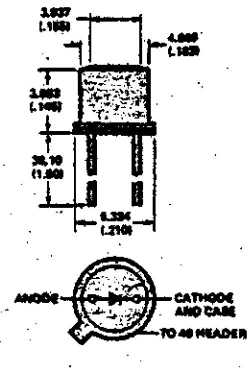
DIMENSIONS IN MILLIMETERS (INCHES).



-4203, -4204, -4207



-4205



-4220

Pin diode characteristics (cont.)

Absolute Maximum Ratings

Parameter	-4203	-4204	-4205	-4207	-4220	Units
P_{MAX} Power Dissipation[1]	100	100	50	100	100	mW
Peak Reverse Voltage[2]	200	200	200	200	200	volts
Steady Reverse Voltage[3]	50	20	50	20	50	volts

Electrical/Optical Characteristics at T_A=25°C

Symbol	Description	-4203 Min.	-4203 Typ.	-4203 Max.	-4204 Min.	-4204 Typ.	-4204 Max.	-4205 Min.	-4205 Typ.	-4205 Max.	-4207 Min.	-4207 Typ.	-4207 Max.	-4220 Min.	-4220 Typ.	-4220 Max.	Units
R_e, D^{**}, $R_\phi \cdot A$	Axial Incidence Response at 770nm(4)		1.0			1.0			1.5*			4.0			1.0		$\mu A / (mW/cm^2)$
A	Active Area[4]		2 x 10^{-3}			2 x 10^{-3}			3 x 10^{-3}*			8 x 10^{-3}			2 x 10^{-3}		cm^2
R_ϕ	Flux Responsivity 770 nm[5] (Fig. 1, 3)		.5			.5			.5			.5			.5		μA/μW
I_D	Dark Current[6] (Fig. 4)			2.0			0.6			15			2.5			5.0	nA
NEP	Noise Equivalent Power[7] (Fig. 8)			5.1 x 10^{-14}			2.8 x 10^{-14}			1.4 x 10^{-14}			5.7 x 10^{-14}			8.1 x 10^{-14}	$\frac{W}{\sqrt{Hz}}$
D^*	Detectivity[8]	8.7 x 10^{11}			1.6 x 10^{12}			4.0 x 10^{12}*			1.6 x 10^{12}			5.5 x 10^{11}			$\frac{cm\sqrt{Hz}}{W}$
C_J	Junction Capacitance[9] (Fig. 5)		1.5			2.0			0.7			5.5			2.0		pF
C_P	Package Capacitance[10]		2			2						2					pF
t_r, t_f	Zero Bias Speed (Rise, Fall Time) 11		300			300			300			300			300		ns
t_r, t_f	Reverse Bias Speed (Rise, Fall Time) 12			1			[illegible]			1			[illegible]			1	ns
R_S	Series Resistance			50			50			50			50			50	Ω

*see Note 4.

NOTES:

1. Peak-Pulse Power

   When exposing the diode to high level incidence the following photocurrent limits must be observed:

   $$I_p \text{ (avg MAX.)} < \frac{P_{MAX} - P_\phi}{E_c}; \text{ and in addition:}$$

   $$I_p(PEAK) < \frac{1000\ A}{t\ (\mu sec)} \text{ or } < 500mA \text{ or } < \frac{I_p \text{ (avg MAX.)}}{f \times t}$$

   whichever of the above three conditions is least.

   I_p - photocurrent (A) f - pulse repetion rate (MHz)
   E_c - supply voltage (V) P_ϕ - power input via photon flux
   t - pulse duration (μs) P_{MAX} - max dissipation (W)

   Power dissipation limits apply to the sum of both the optical power input to the device and the electrical power input from flow of photocurrent when reverse voltage is applied.
2. Exceeding the *Peak Reverse Voltage* will cause permanent damage to the diode. Forward current is harmless to the diode, within the power dissipation limit. For optimum performance, the diode should be reversed biased with E_c between 5 and 20 volts.
3. Exceeding the *Steady Reverse Voltage* may impair the low-noise properties of the photodiodes, an effect which is noticeable only if operation is diode-noise limited (see *Figure 8*).
4. The 5082-4205 has a lens with approximately 2.5x magnification; the actual junction area is 0.5 x 10^{-3} cm^2, corresponding to a diameter of 0.25mm (.010"). Specification includes lens effect.
5. At any particular wavelength and for the flux in a small spot falling entirely within the active area, responsivity is the ratio of incremental photodiode current to the incremental flux producing it. It is related to quantum efficiency, η_q in electrons per photon by:

   $$R_\phi = \eta_q \left(\frac{\lambda}{1240}\right)$$

   where λ is the wavelength in nanometers. Thus, at 770nm, a responsivity of 0.5 A/W corresponds to a quantum efficiency of 0.81 (or 81%) electrons per photon.
6. At −10V for the 5082-4204, −4205, and −4207; at −25V for the 5082-4203 and −4220.
7. For (λ, f, Δf) = (770nm, 100Hz, 6Hz) where f is the frequency for a spot noise measurement and Δf is the noise bandwidth, NEP is the optical flux required for unity signal/noise ratio normalized for bandwidth. Thus:

   $$NEP = \frac{I_N/\sqrt{\Delta f}}{R_\phi}$$

   where $I_N/\sqrt{\Delta f}$ is the bandwidth – normalized noise current computed from the shot noise formula:

   $$I_N/\sqrt{\Delta f} = \sqrt{2qI_D} = 17.9 \times 10^{-15} \sqrt{I_D}\ (A/\sqrt{Hz}) \text{ where } I_D \text{ is in nA.}$$
8. Detectivity, D* is the active-area-normalized signal to noise ratio. It is computed: for (λ, f, Δf) = (770nm, 100Hz, 6Hz).

   $$D^* = \frac{\sqrt{A}}{NEP} \left(\frac{cm\sqrt{Hz}}{W}\right) \text{ for A in } cm^2.$$
9. At −10V for 5082-4204, −4205, −4207, −4220; at −25V for 5082-4203.
10. Between diode cathode lead and case — does not apply to 5082-4205, -4220.
11. With 50Ω load.
12. With 50Ω load and −20V bias.

Pin diode characteristics (cont.)

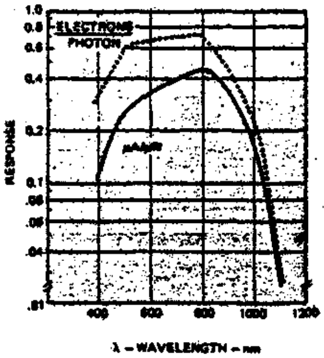


Figure 1. Spectral Response.

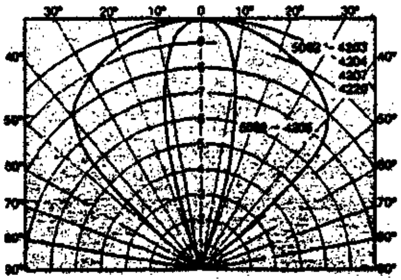


Figure 2. Relative Directional Sensitivity of the PIN Photodiodes.

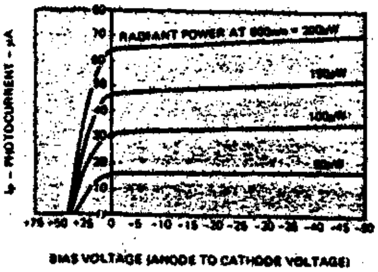


Figure 3. Typical Output Characteristics at λ = 900nm.

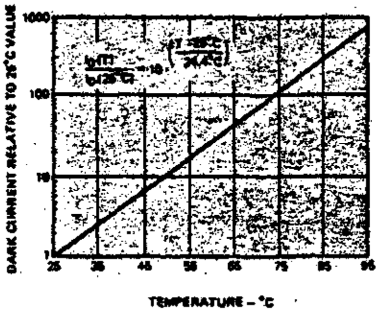


Figure 4. Dark Current at −10V Bias vs. Temperature.

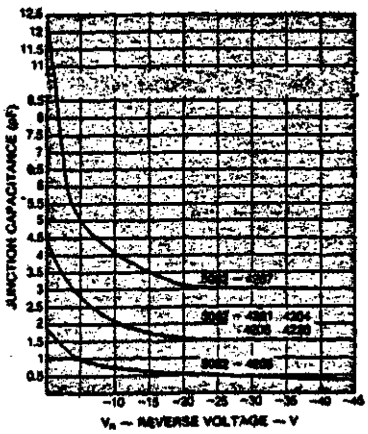


Figure 5. Typical Capacitance Variation With Applied Voltage.

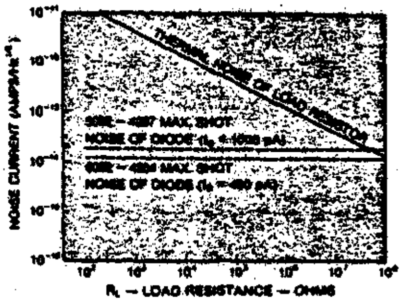


Figure 6. Noise vs. Load Resistance.

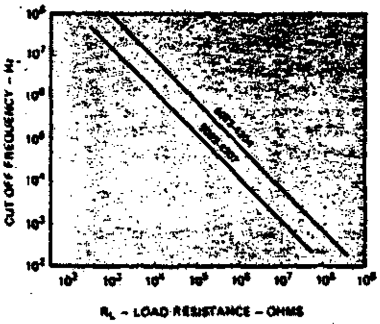


Figure 7. Photodiode Cut-Off Frequency vs. Load Resistance (C = 2pF).

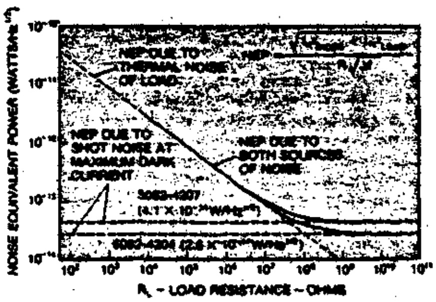


Figure 8. Noise Equivalent Power vs. Load Resistance.

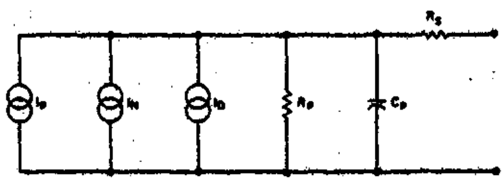


Figure 9. Photodiode Equivalent Circuit.

I_P = Signal current ≈ 0.5μA/μW x flux input at 770 nm

I_N = Shot noise current

<1.2 x 10^{-14} amps/$Hz^{1/2}$(5082-4204)

<4 x 10^{-14} amps/$Hz^{1/2}$(5082-4207)

I_D = Dark current

<600 x 10^{-12} amps at -10 V dc (5082-4204)

<2500 x 10^{-12} amps at -10 V dc (5082-4207)

R_P = 10^{11}Ω

R_S = <50Ω

APPENDIX C

C.1 Listing of CONTRL

```
C***********************************************************************
C
C   CONTRL.FTN   -   PROGRAM TO CONTROL DATA SCANNING SYSTEM OPERATION
C                    INITIATES SCAN.MAC AND INTRUP.MAC
C
C                    P.K.ROBERTSON 1980
C
C***********************************************************************
C
C   INITIALISE
C
        INTEGER BUF(13),RBUF(15),IOUT(3),PRL(6),IOST(512),ISB(2)
        DATA A/6RINTRUP/
        DATA B/6RSCAN  /
        CALL CLREF(52)
        WRITE (5,400)
400     FORMAT(' ON WHAT DISK IS DATA TO BE WRITTEN?,/,
     + ' NOTE : IF NOT DISK 2, DISK # MUST BE CHANGED IN SCAN')
        ACCEPT *,IDK
        CALL ASNLUN (1,'DK',IDK)
        WRITE (5,500)
500     FORMAT(' HOW MANY CHUNKS ? (1 CHUNK = 32 SWEEPS)')
        ACCEPT *,IC
        WRITE (5,501)
501     FORMAT(' HOW MANY WORDS/LINE?')
        ACCEPT *,IW
        BUF(1)=IC
        BUF(2)=IW
1       WRITE (5,510)
510     FORMAT(' READY TO START - TYPE 0 IF SO, -1 TO EXIT')
        ACCEPT *,IR
        IF (IR) 99,2,1
C
C DECLARE SIGNIFICANT EVENT AND QUEUE BUF FOR SCAN
C
2       CALL SEND(B,BUF,,IDS)
        IF (IDS) 77,3,3
C
C ACTIVATE SCAN
C
3       CALL REQUES(B,,IDS)
        IFL=1
        IF (IDS) 77,4,4
C
C ACTIVATE INTRUP
C
        CALL RUN (A,,1,1,,,IDS)
        IFL=2
        IF (IDS) 77,5,5
C
C SUSPEND
C
5       CALL SUSPND
C
```

Listing of CONTRL (cont.)

```
C
C RESUMED BY SCAN
C
        IFL=3
C
C ABORT INTRUP
C
        CALL ABORT (A,IDS)
        IFL=4
        IF (IDS) 77,6,6
C
C RECEIVE DATA FROM SCAN
C
6       CALL RECEIV (B,RBUF,,IDS)
        IFL=4
        IF (IDS) 77,7,7
C
C UNPACK SENDING TASK NAME AND OUTPUT
C
7       CALL R50ASC(6,RBUF,IOUT)
        WRITE (5,520) IOUT, (RBUF(K),K=3,15)
520     FORMAT(' RECEIVED FROM ',3A2,/,1X,13I4)
        WRITE (5,530)
530     FORMAT(' DO YOU WISH TO SEE I/O STATUS CODES?, 1 FOR YES')
        ACCEPT *,IOC
        IF (IOC .NE. 1) GO TO 99
C
        PRL(2)=1024
        PRL(5)=RBUF(4)
        CALL GETADR (PRL(1),IOST)
        CALL WTQIO("001000,1,2,,ISB,PRL,IDS)
        DO 100 I=1,64
100     WRITE (5,540) (IOST((I-1)*8+J),J=1,8)
540     FORMAT(' ',8(I5,3X))
C
        GO TO 99
C
77      WRITE (5,590) IFL,IDS
590     FORMAT(' STAGE ',I3,' DIR ERROR IDS = ',I5)
99      STOP
        END
C
C***********************************************************************
```

C.2 Listing of INTRUP

```
;*********************************************************
;       INTRUP.MAC :                    P.K.ROBERTSON 1980
;       INTERRUPT DETECT AND SERVICE
;       BUILD WITH  TKB
;                   TKB>INTRUP/PR:5=INTRUP
;                   TKB>/
;                   ENTER OPTIONS
;                   TKB>RESCOM=DCOM/RW
;                   TKB>//
;*********************************************************
;
        .TITLE  INTRUP
        .MCALL CINT$S,SETF$S,DECL$S,EXIT$S
;
INVEC:  ISR
        414
;
INTRUP: BIS     #40,$DCSR
        CINT$S  #INVEC,#120000,#ISR,#FIN,PR5
WAIT:   WAIT
        BR      WAIT
;
;
ISR:    SETF$S  #52.
        DECL$S
        RTS     PC
;
FIN:    BIC     #40,$DCSR
        EXIT$S
;
FLAGS:  .BLKW   4.
;
        .END    INTRUP
;*********************************************************
```

C.3 Listing of SCAN

```
;*********************************************************************
;       SCAN.MAC  BUILD WITH                          P.K.ROBERTSON 1980
;                      SCAN/PR:0=SCAN,IOP
;                      /
;                      RESCOM=DCOM/RW
;                      RESCOM=SCOM/RW
;                      //
;
;       SCAN INPUTS DATA FROM DR11-C, WRITES TO DISK OR TAPE
;       INITIATED BY CONTRL.FTN
;       OPERATES WITH INTRUP OR DIRECTLY USING REQUEST LINE
;       USES IOP.MAC, A SET OF I/O ROUTINES
;*********************************************************************
        .MACRO  ARRAY,ADDR,SIZE,ROWS
        MOV     R0,-(SP)
        MOV     R1,-(SP)
        MOV     R2,-(SP)
        MOV     #ADDR,R0
        MOV     #SIZE,R1
        MOV     #ROWS,R2
        JSR     PC,PRTARR
        MOV     (SP)+,R2
        MOV     (SP)+,R1
        MOV     (SP)+,R0
        .ENDM   ARRAY
;
        .MACRO  PRINT,BUFFER
        MOV     R0,-(SP)
        MOV     #BUFFER,R0
        JSR     PC,PRTSUB
        MOV     (SP)+,R0
        .ENDM   PRINT
;*****************************************************************
;
        .TITLE SCAN
MESS1:  .ASCIZ  <15><12>/STAGE ONE/
MESS2:  .ASCIZ  <15><12>/STAGE TWO/
MESS3:  .ASCIZ  <15><12>/STAGE THREE/
        .EVEN
        .PSECT
        .MCALL  ALUN$C,QIO$S,EXIT$S,MRKT$S,WTSE$S,QIOW$S
        .MCALL  RCVX$S,RSUM$C,SDAT$C
;
;   INITIALIZE SHEET
;
SCAN:   ALUN$C  1.,MM,0                 ; LUN 1 = MM0
        ALUN$C  4.,DK,2                 ; LUN 4 = DK2
        RCVX$S  CON,#RDAT               ; RECEIVE DATA FROM CONTRL
        MOV     #RDAT,R0
        ADD     #4,R0
        MOV     (R0)+,NUM               ; # LINES = RDAT(3)
        MOV     (R0)+,WORDS             ; # WORDS/LINE = RDAT(4)
        MRKT$S  #3.,#5.,#2.
        WTSE$S  #3.
```

Listing of SCAN (cont.)

```
        CMP     #0,NUM                          ; CHECK FOR ZERO LINES
        BNE     COM1
        MOV     #16.,NUM                        ; DEFAULT OF 16 LINES
COM1:   CMP     #0,WORDS                        ; CHECK FOR 0 WORDS
        BNE     COM2
        MOV     #200.,WORDS                     ; DEFAULT OF 200. WORDS/LINE
COM2:   MOV     #IOST,R4                        ; R4 POINTS TO I/O STATUS BUFFER
        MOV     #37777,$DOUT                    ; DROUTBUF -> 00111111,11111111
        PRINT   MESS1
        MOV     #140377,$DOUT                   ; DROUTBUF -> 11000000,11111111
        BIS     #2,$DCSR                        ; INITIALISE LINE F/F
        PRINT   MESS2
        BIC     #2,$DCSR
;
; INITIIALIZE LINE
;
        MOV     #D1,R3                          ; R3 POINTS TO BUFFER D1
        MOV     #0,R1                           ; SET DISK LOCATION POINTER
INIT1:  MOV     #0,FLAG
INIT2:  MOV     #BUF,R0                         ; R0 POINTS TO INPUT BUFFER
        BIC     #1,$DCSR                        ; ENABLE STACK TRANSFER
;
SWP:    MOV     WORDS,COUNT                     ; RESET WORD COUNT
;*******************
;  MODIFICATION TO OPERATE DIRECTLY OR WITH INTRUP
;*******************
;       WTSE$S  52.                             ; WAIT FOR INTERRUPT
;*******************
        BIT     #100000,$DCSR                   ; IS SWEEP REQUEST B SET
        BEQ     SWP
;*******************
        CLR     $LKSR                           ; DISABLE LINE CLOCK INTERRUPT
        BIS     #2,$DCSR                        ; CLEAR SWEEP REQUEST F/F
;
; COLLECTION  LOOP
;
LOOP:   CMP     #1,$SWRG                        ; CHECK SWITCH REG FOR 1
        BEQ     7$                              ; IF SO, EXIT
        BIT     #200,$DCSR                      ; IS REQ A SET?
        BEQ     LOOP                            ; NO
        BIS     #1,$DCSR                        ; YES, SET CSR0 TO 1
        MOV     $DIN,(R0)+                      ; GET DATA
        BIC     #1,$DCSR                        ; CLEAR CSR0
        DEC     COUNT                           ; ENOUGH WORDS ?
        BNE     LOOP                            ; NO, INPUT NEXT
;
;
7$:     MOV     #300,$LKSR                      ; RE-ENABLE LINE CLOCK INTERRUP
        BIC     #2,$DCSR                        ; RESET CSR1
        INC     FLAG                            ; INCREMENT BUFFER FLAG
;
```

Listing of SCAN (cont.)

```
        MOV     #BUF,R0                         ; RESET R0
        MOV     #512.,R2
LINE:   MOV     (R0)+,(R3)+                     ; LINE TO TEMP BUFFER
        SOB     R2,LINE
        CMP     #16.,FLAG                       ; TEMP BUFFER FULL ?
        BNE     T32                             ; NO, CHECK OTHER ONE
        QIO$S   #IO.WLB,#4.,#1.,,R4,,<#D1,#16384.,,,R1>
        ADD     #32.,R1                         ; ADVANCE DISK POINTER
        ADD     #4.,R4                          ; ADVANCE I/O STATUS BUF POINTER
        MOV     #C1,R3                          ; R3 NOW POINTS TO BUF C1
        JMP     INIT2                           ; RETURN TO FILL C1
;
T32:    CMP     #32.,FLAG                       ; IS C1 FULL ?
        BNE     INIT2                           ; NO, RETURN FOR MORE DATA
        QIO$S   #IO.WLB,#4.,#1.,,R4,,<#C1,#16384.,,,R1>
        ADD     #32.,R1                         ; ADVANCE DISK POINTER
        ADD     #4.,R4                          ; ADVANCE I/O STATUS BUF POINTER
        MOV     #D1,R3                          ; R3 NOW POINTS TO BUFFER D1
FIN:    DEC     NUM                             ; ENOUGH LINES ?
        BEQ     EXIT                            ; YES, THEN EXIT
        JMP     INIT1                           ; NO, BACK FOR MORE
;
EXIT:   PRINT   MESS3
        MRKT$S  #3.,#2.,#2
        WTSE$S  #3.
        MOV     #SDAT,R0                        ; ADDRESS OF SEND BUFFER
        MOV     #1.,(R0)+                       ; 1 -> SDAT(1)
        MOV     (R1),(R0)+                      ; IOST ADDRESS ON DISK -> SDAT(2
        QIO$S   #IO.WLB,#4.,#1.,,R4,,<#IOST,#1024.,,,R1>
;       ARRAY   SDAT,13.,8.
;       MRKT$S  #3.,#5.,#2.
        SDAT$C  CON,SDAT,52.                    ; SEND DATA TO CONTRL
;       ARRAY   IOST,200.,8.
;       MRKT$S  #3.,#10.,#2                     ; DEBUGGING AIDS
;       WTSE$S  #3.
        RSUM$C  CON                             ; RESUME CONTRL
        EXIT$S
;
COUNT:  .WORD   0
FLAG:   .WORD   0
NUM:    .WORD   0
BUF:    .BLKW   512.
IOST:   .BLKW   512.
D1:     .BLKW   8192.
C1:     .BLKW   8192.
RDAT:   .BLKW   15.
SDAT:   .BLKW   13.
CON:    .RAD50 /CONTRL/
        .END SCAN
;*************************************************************************
```