The development of a linear diode array detector system for astronomical spectroscopy

A. W. Campbell

THE DEVELOPMENT OF A LINEAR DIODE ARRAY DETECTOR SYSTEM

FOR

ASTRONOMICAL SPECTROSCOPY

-by-

A.W. CAMPBELL, B.Sc.

A thesis submitted to the University of Durham for the Degree of Doctor of Philosophy.

Being an account of work carried out at the University of Durham during the period April 1977 to October 1980.

ABSTRACT

The work reported in this thesis describes the development of an early linear diode array detector system for use in astronomical spectroscopy, outlines factors rendering such a detector suitable for this field, examines some of the more taxing practical problems encountered in early stages of development and, drawing upon these factors and attempted resolutions of the purported initial snags, continues with a look at a recent version and work accomplished with it. In attempting to cover these areas in the development, the description will draw as heavily upon actual laboratory tests carried out as upon the observational fieldwork at both the RGO and the Wise Observatory, partially because of the fact that detailed analysis of the limited data collected is not one of this work's central concerns. As it was necessary to construct a complete working system fairly quickly, the approach adopted and described in detail in what follows may be recognizably far from optimal. Experience gained and various techniques developed could, however, lead to significant improvements in future systems. Therefore, such techniques as they emerge during the early stages are outlined throughout the account as they seem to be one of the more valuable contributions of the project.

CONTENTS

		Page No.
CHAPTER ONE.	DIODE ARRAYS AND THEIR USE IN ASTRONOMICAL SPECTROSCOPY	1
1	Introduction	1
2(a)	The Silicon Diode as a Light Detector	1
2(b)	Quantum Efficiency and Spectral Response	2
2(c)	Practical Detectors	3
2(d)	Dark Current	4
3	Fabrication of Linear Diode Arrays	5
4	Readout Modes	7
5	Noise	8
6(a)	Operation and Noise Performance of a Recharge Sampling Array	8
6(b)	Thermal Shot Noise	10
6(c)	Reset Noise	10
6(d)	Amplifier Noise	11
7(a)	Operation and Noise Performance of a Voltage Sampling Array	12
7(b)	Source Follower	12
7(c)	Machine Readable Dynamic Range	13
7(d)	Random Noise in Voltage Sampling Arrays	13
8(a)	Factors affecting the use of Solid State Imaging Arrays for Astronomical Spectroscopy	16
8(b)	Array Dimensions	16
8(c)	Optoelectronic Performance	17
8(d)	Currently Available Solid State Detectors	18
	References	20

	Page No.
CHAPTER TWO. THE PLESSEY H509 ARRAY	22
1 Introduction	22
2(a) Layout and Scanning	22
2(b) Circuit and Operation of a Single Element	23
2(c) Recharge Operation and Control Pulse Amplitudes	26
2(d) Operation of an Element Pair and NDRO	27
2(e) External Amplifier Configuration	27
2(f) Spatial Noise Considerations	29
3(a) Random Noise	30
3(b) MOSFET Amplifier Noise	30
3(b)(i) Low frequency Noise in MOSFET'S	31
3(b)(ii)Thermal noise	33
3(c) Noise Equivalent Circuit	34
3(d) Switch Capacitance Effects and Charge Pumping Noise	38
3(e) Noise Summary	42
4(a) Driving Electronics	42
4(b) TTL Sequence Generator	43
4(c) TTL to MOS Level Converters	44
4(c)(i) Clock Pulse Drivers	45
4(c)(ii)Charge Pulse Drivers	46
4(c)(iii) Scan Pulse Driver	46
4(d) Power Supplies	46
References	47

		Page No.
CHAPTER THREE.	LABORATORY EVALUATION WITH ATTENDANT PRACTICAL PROBLEMS	48
1	Introduction	48
2(a)	Cooling Systems	48
2(b)	Effects of Cooling	49
3(a)	Measurements of Photoresponse	53
3(b)	Incomplete Recharge	54
3(c)	Amplifier Transistor Lag and V_{REF} Considerations	58
3(d)	Signal Dependence of Lag	59
3(e)	Summary of Non-Ideal Behaviour	62
4	Responsivity at Different Light Intensities	63
5(a)	Effects of Changes in Chip supply voltages and Control Pulse Amplitudes: Introduction	63
5(b)	Clock Pulse Amplitude	63
5(c)	Measurement of Transfer characteristic of T3	65
6(a)	Calibration of Photoresponse, Quantum Efficiency and Noise Measurements: Introduction	66
6(b)	Initial Evaluation of Absolute Responsivity	67
6(c)	Preliminary Noise Experiments	68
6(d)	Combined Noise and Responsivity Measurement	69
6(e)	Experimental Determination of Noise vs. Signal	70
6(f)	Analysis and Results	71
7	Summary of Detector Parameters	73
	References	75

		Page No.
CHAPTER FOUR.	SIGNAL PROCESSING AND DIGITIZATION SYSTEM	76
1	Introduction	76
2(a)	The Present System	77
2(b)	Timing Circuit	78
2(c)	Head Amplifier	78
2(d)	Sample and Hold	80
2(e)	Analogue to Digital Converter	84
2(f)	The Buffer Memory	86
3	Fixed Pattern Removal	86
	References	90
CHAPTER FIVE.	CONTROL AND COMPUTING	91
1	Requirements of a Control System	91
2(a)	Computer Control of Diode Array via CAMAC	92
2(b)	Control Module Operation	93
2(c)	Buffer Memory Operation	94
2(d)	Display System	97
3(a)	Software System	98
3(b)	The Control Program	98
3(c)	Typical Operational Sequence	100
4(a)	The Present Computing System	103
4(b)	The Paper Tape System	104
4(c)	Problems and Further Development	105
	References	107

Page No.

CHAPTER SIX. OPERATION OF THE DETECTOR ON A TELESCOPE SPECTROGRAPH SYSTEM 108

1 Introduction 108

2(a) Detector System for Operation on a Cassegrain Spectrograph 108

2(b) Cooling System Design 109

2(c) Relay Lens System 112

2(d) Registration and Focusing 113

2(e) Organization of the Electronics 114

3 Observing Procedures and Initial Setting Up 115

4(a) Practical Problems Encountered during Field Trials 118

4(b) Odd-Even Effect 118

4(c) Repeatability and Reciprocity under Observing Conditions 120

4(d) Image Retention 122

5 Measurement of Spectral Response 124

References 125

CHAPTER SEVEN. CONCLUSION 126

1 Introduction 126

2(a) Device Considerations 126

2(b) Alternative Operating Procedures 128

2(c) System Improvements 128

References 129

CHAPTER 1

DIODE ARRAYS AND THEIR USE IN ASTRONOMICAL SPECTROSCOPY

1. Introduction

The concern of this thesis is the description of both the development and construction of a system to use a linear diode array for astronomical spectroscopy. The particular array in question was manufactured by the Plessey Co.Ltd. to the requirements of a team working at the Royal Greenwich Observatory. It should be noted that the original intention[1] - that of operating the device in electron bombardment mode (i.e. looking at Astronomical Spectra with an image intensifier) - was never fully realised. Somewhat later, a number of these arrays were made available to the Nuclear Instrumentation Group at Durham University whose main objective at the time was the investigation of array operation under direct illumination. Silicon diode photodetectors had shown themselves as having inherently good response to light, especially in the red part of the spectrum, a region of interest where photographic plates are not at their most useful. Combining this with the fact that, by virtue of their geometrical layout, linear arrays are particularly suited to spectroscopic applications, a decision was taken to attempt to build a system which would allow the use of these chips, under direct illumination, in looking at stellar spectra.

2(a) The Silicon Diode as a Light Detector

The type of detector considered here makes use of the current which flows in a reverse biased p - n junction due to the light incident upon it. Generally the total current flowing will be the sum of this photo-current and that of a thermally generated contribution known as the dark current.

A single photon, if absorbed in the silicon, raises an electron from the valence band to the conduction band thereby creating an electron-hole

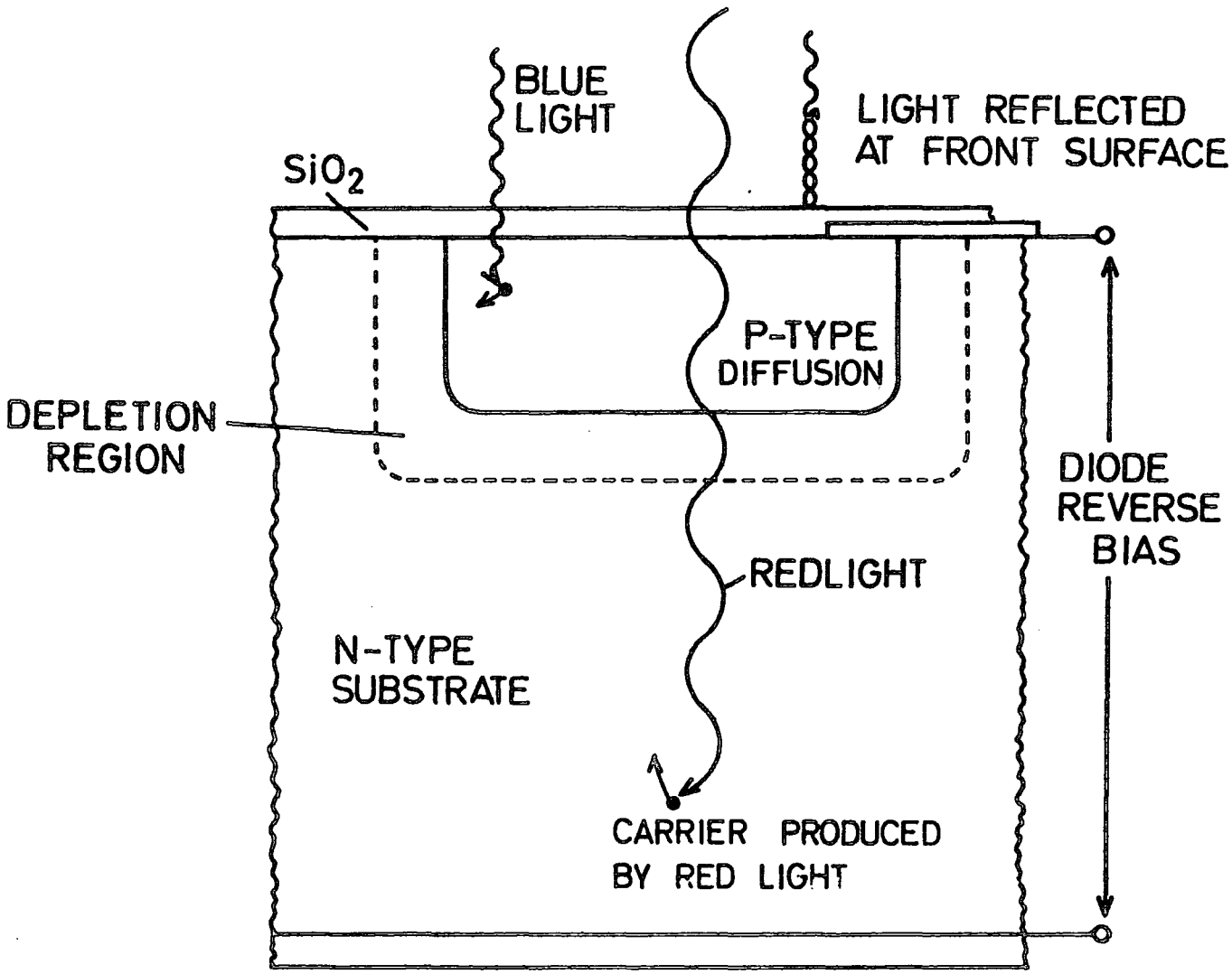


FIGURE 1.1 SECTION THROUGH PHOTODIODE

pair. If the latter reaches the depletion region of the reverse biased junction before it recombines, it will contribute to the photocurrent. The minimum photon energy necessary for absorption in this manner is determined by the band gap energy. So,

For silicon at 42K, the E_g = 1.166 eV.[2]

The above gives

$$\lambda_{max} = hc/E_g = 1.07\ \mu m$$

In silicon, this simple picture is complicated by the fact that, at the absorption edge, the minimum energy transition is not a direct one in crystal momentum and, accordingly, is of low probability; transitions, $\Delta K = 0$ are preferred. Momentum may still be conserved by the creation or annihilation of virtual phonons.

2(b) Quantum Efficiency and Spectral Response

The fraction of incident photons which are detected is called the quantum efficiency RQE of the detector. The losses can be separated into two parts. Firstly, light incident upon the device is attenuated both by absorption in the passivating layers above the silicon and by reflection at the various interfaces. Secondly, e-h pairs created in the silicon may recombine before reaching the junction. This is illustrated in Figure (1.1).

The depth in the silicon at which photons are absorbed depends on their wavelengths. Monochromatic light is atten uated exponentially so that

$$N = N_o e^{-X/L} .$$

Further, the characteristic absorption depth, L increases with wavelength as shown in Figure (1.2).[3] From this it can be seen that most blue photons will be absorbed just inside the surface whereas the red ones have a greater chance of penetrating deeper into the silicon. Considering

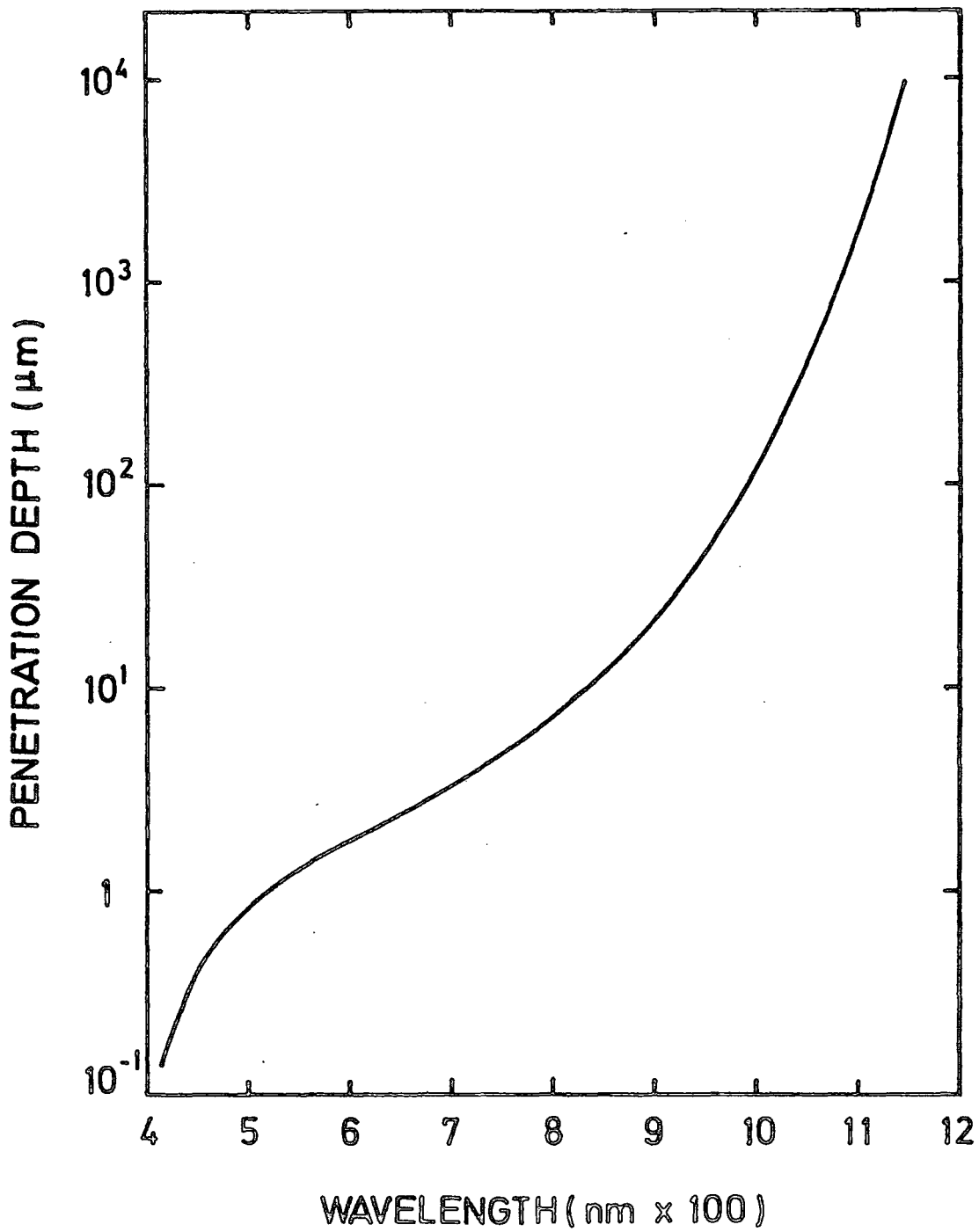


FIGURE 1.2 PENETRATION DEPTH vs WAVELENGTH

a photodiode of typical geometry, the relative spectral response will vary with the junction depth, the width of the space charge layer and with the diffusion length of the carriers. In any practical case the spectral response will differ most from the ideal at very long and at very short wavelengths. The probable lifetime of an e-h pair created near the surface is much shorter because of the existence of surface traps at the Si - SiO_2 interface. Although these can be minimised by good fabrication techniques, in practice, trapping will be much worse here than in the bulk, giving poor blue response.(4) At the red end of the spectrum, those photons with wavelengths approaching the cut-off wavelength are more likely to penetrate deeper into the silicon and to produce carriers which have less chance of reaching the depletion region.(5)

2(c) Practical Detectors

To use a diode as a photodetector it is necessary to measure the photocurrent which, at low levels of illumination, can be very small because one photon can produce a charge flow of only one electron. For situations where the photocurrent is less than any practical system noise, a mode of operation properly titled Photon Flux Integration or, simply, charge storage was developed.(6) In this mode the diode, after being reverse biased, is isolated allowing the charge detected to integrate on the diode's self-capacitance and to discharge it for a set exposure time. A simplified equivalent circuit of this mode of operation is to be found in Figure (1.3). Following this given exposure time the charge required to restore the diode to its initial reverse bias voltage may be measured. On doing so it will be seen that it is proportional to the light detected by the diode integrated over the same exposure time. For constant illumination the signal will increase with the exposure time which may be selected to give an adequate signal to noise ratio. All this ignores the effect of dark current which, in order to be reduced to an acceptable level,

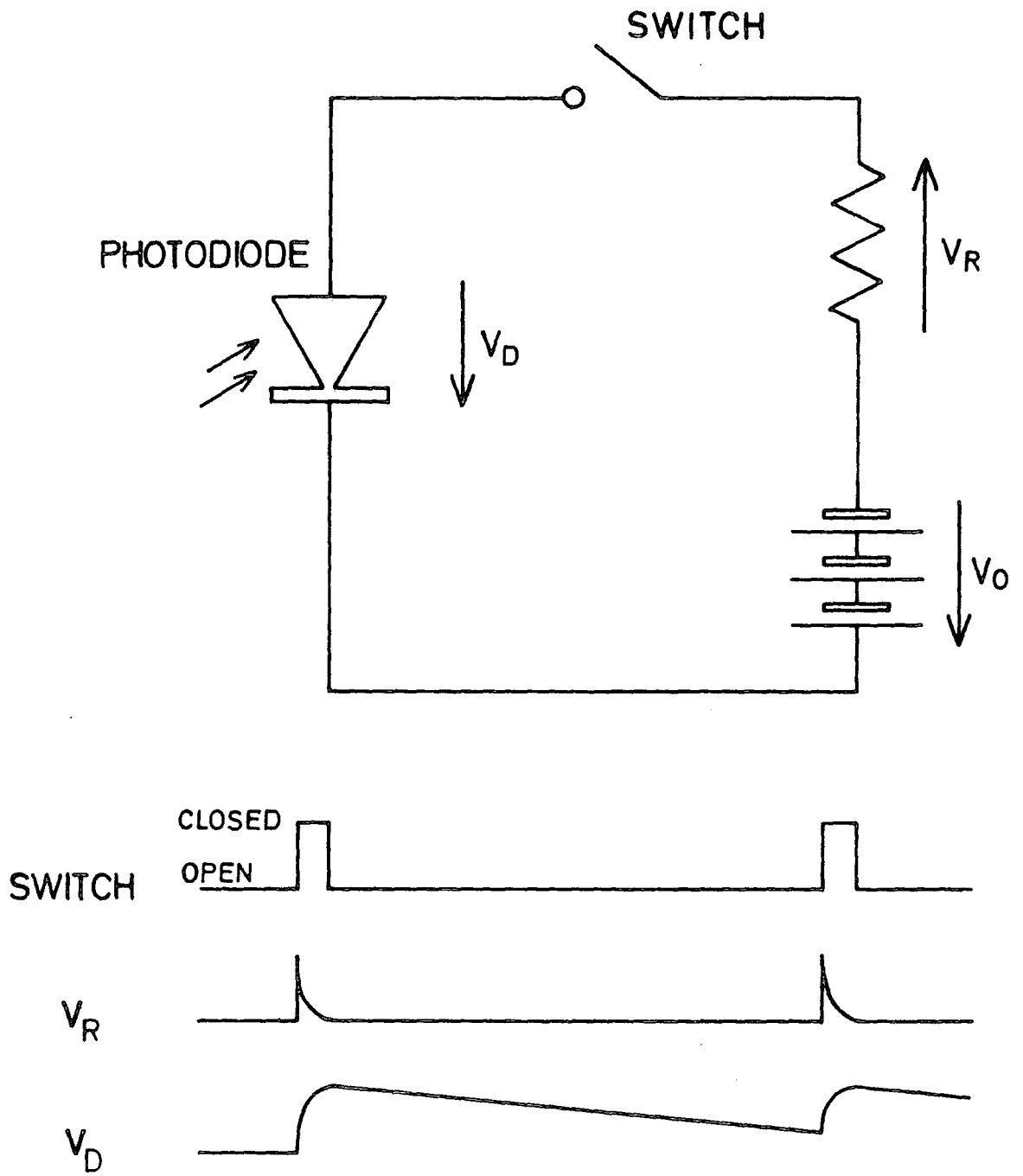


FIGURE 1.3 CHARGE STORAGE OPERATION OF PHOTODIODE

necessitates the cooling of the device in practice. As an alternative to the recharge sampling readout method described above, the decrease in diode voltage may be monitored directly by a buffer amplifier. The relative advantages of each of these systems will be discussed later.

2(d) Dark Current

Dark current consists of e - h pairs which, in theory, are generated thermally throughout the bulk of the device. However, the dominant contribution comes from the depletion region[7] particularly where the latter meets the surface, i.e. around the periphery of a planar device. Hence the dark current will increase with larger reverse bias voltages as these cause the depletion layer to extend. Again, careful fabrication techniques can reduce surface effects, although, for operation in PFI mode with possible exposure times of several minutes, cooling is usually necessary.

The voltage dependence of the dark current means that in PFI mode where the bias voltage changes throughout the exposure, the thermal contribution is non-linear. Therefore in situations where the thermal signal is of comparable size to the light signal, it cannot be easily removed by attempting to take two exposures of identical length (i.e. one illuminated and one dark) and subtracting one from the other. However, a proviso might be added to the effect that if the diode is cooled sufficiently this method can provide a reasonably good approximation.[8] With further cooling the dark contribution can be made negligible for most practical exposure times, but such dramatic cooling has several disadvantages:

(a) the electrical performance of the device may be affected in other ways - all of which are not entirely predictable;

(b) there is a problem with mechanical stresses as the device will have to survive being cycled between hot and cold many times; and

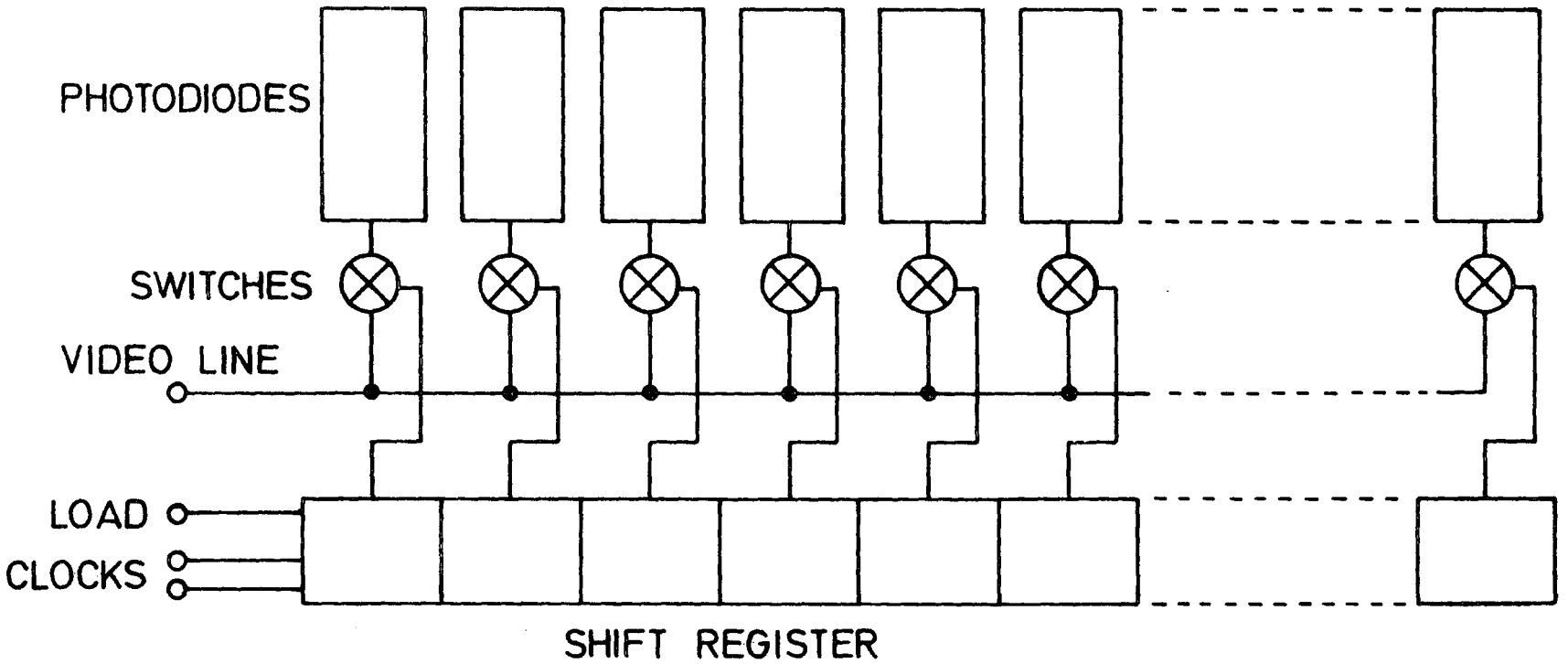


FIGURE 1.4 ORGANIZATION OF AN INTEGRATED DIODE ARRAY

(c) the design of a cryostat may impose restrictions on the use of the device.

There may also be some reduction in the red response of the detector because the band gap energy, E_g is a weak function of temperature.[9(a)] Some typical values here

at 0K, the E_g = 1.17 eV , λ = 1.06 μm, and

at 300K, the E_g = 1.14 eV , λ = 1.09 μm,

Nevertheless, it is possible that the reduction effect is offset by an increase in carrier diffusion lengths[9(b)] at reduced temperatures.

3. Fabrication of Linear Diode Arrays

Most of the linear diode arrays in common use are produced using MOS integrated circuit technology which is a simpler process than, for example, bipolar. The former has the advantage that both photodiodes and switching transistors can be made on the same chip but more important than this is the fact that the process simplicity allows production of large arrays of high yield. In a typical array the video outputs from each diode element are multiplexed onto a common video line; the elements are accessed sequentially by an on chip shift register, parallel to the array. In Figure (1.4) this layout is shown diagrammatically.

The important geometrical parameters of an array, from the user point of view, are: the number of elements, the distance between centres, the dead space between diodes and the array width. All are interrelated and, of necessity, various design trade offs are required to be made. Illustrative of the interrelationship, it is clear that the distance between centres or pitch is determined by both the active area of diode required for a reasonable response and by the complexity of the associated peripheral circuitry which has to be fabricated within the length of one element. In turn, the number of elements is limited by the size of circuit which can be made with reasonable yield, and, ultimately, by the

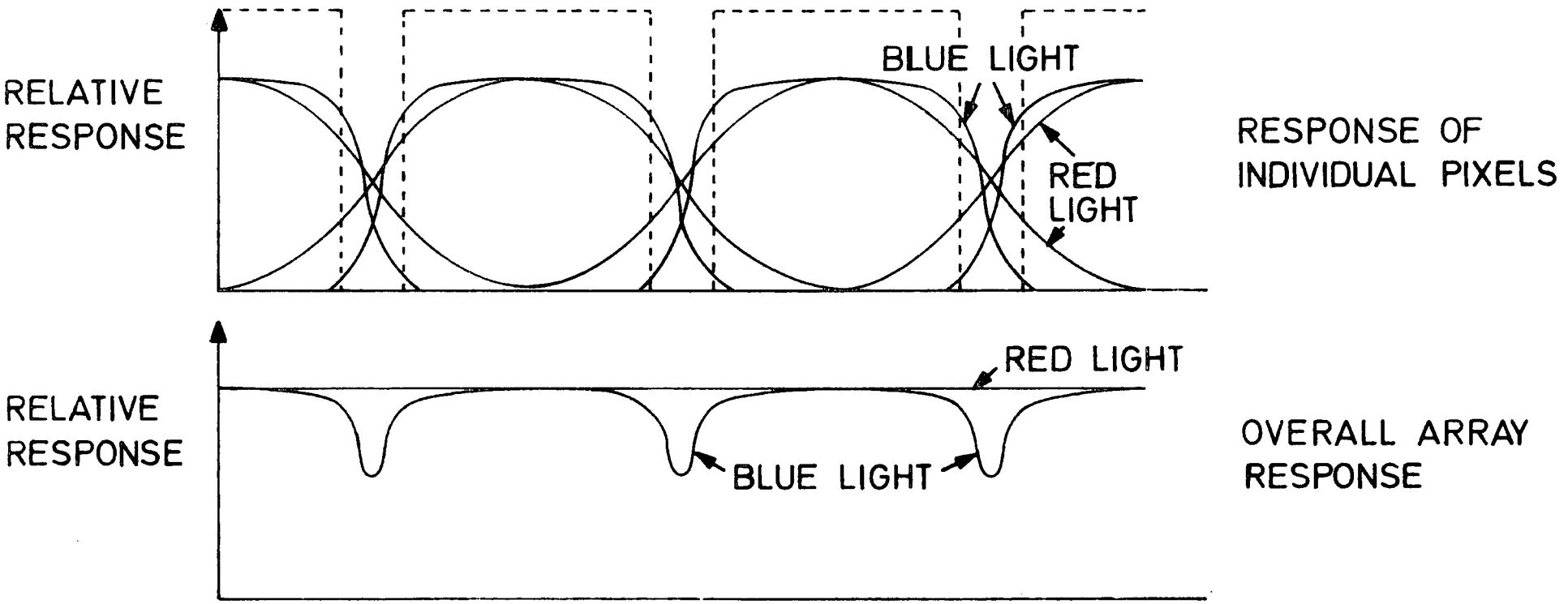


FIGURE 1.5 DETECTOR SPREAD FUNCTION

slice size for a given element pitch. Generally, yield decreases with increasing circuit area.

The dead space between elements is most important when the array's response to bluer wavelengths incident upon it is considered. Most carriers (or electron-hole pairs (e - h pairs) referred to earlier) produced by such light are generated near the surface so that because of the increased recombination probability in this region, their chance of diffusing laterally far enough to reach a junction is considerably lessened. To carriers produced by longer wavelength light deeper in the substrate, this geometrical dead space does not represent any overall loss of quantum efficiency. Instead, it is merely a region of some ambiguity as incident photons may be detected, in the end, by either of the two adjacent diodes. The above process may be summarised by plotting the spread function for different wavelengths as is attempted in Figure (1.5).[10]

In spite of the fact that it is possible to make an array which may be read out in either recharge or voltage sampling mode, the requirements of both simplicity and optimum performance force a choice to be made at the design stage. Assuming that the whole of the diode is to be illuminated, the photocurrent will increase proportionally with the sensitive area of the diode. Accordingly, the signal may be increased by lengthening the diodes, although there is no immediate advantage in making voltage sampling arrays with long diodes because the increase in diode capacitance with area divides out the increase in signal. Nevertheless, an obvious advantage to the user of such long diodes is that they facilitate registration of the device in applications where the light does not fill the array. The aspect ratios of the individual diodes may also have some effect on their signal to dark current performance because, as mentioned earlier, the thermal response increases with the length of the periphery. The number and sharpness of the corners

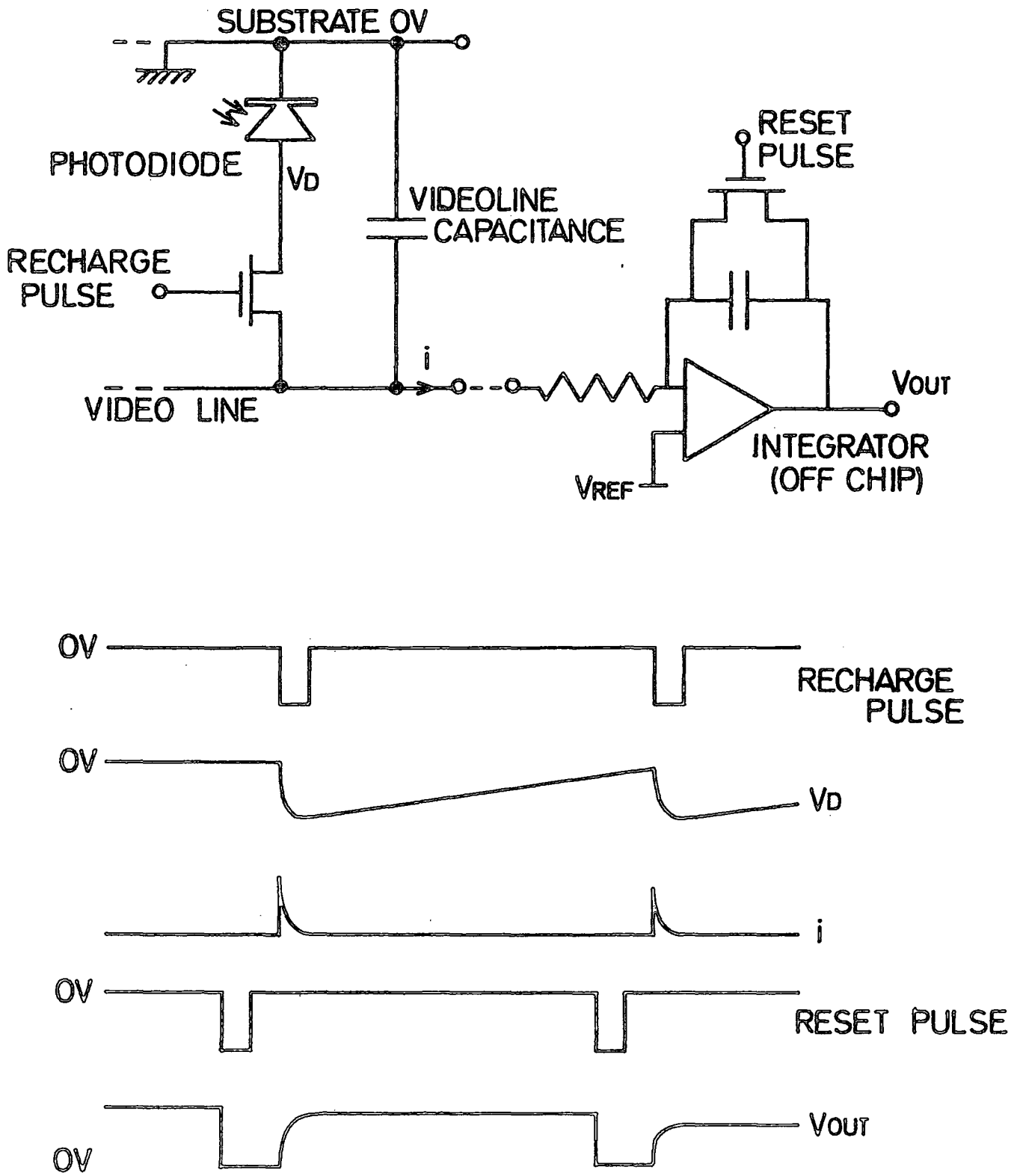


FIGURE 1.6 RECHARGE SAMPLING

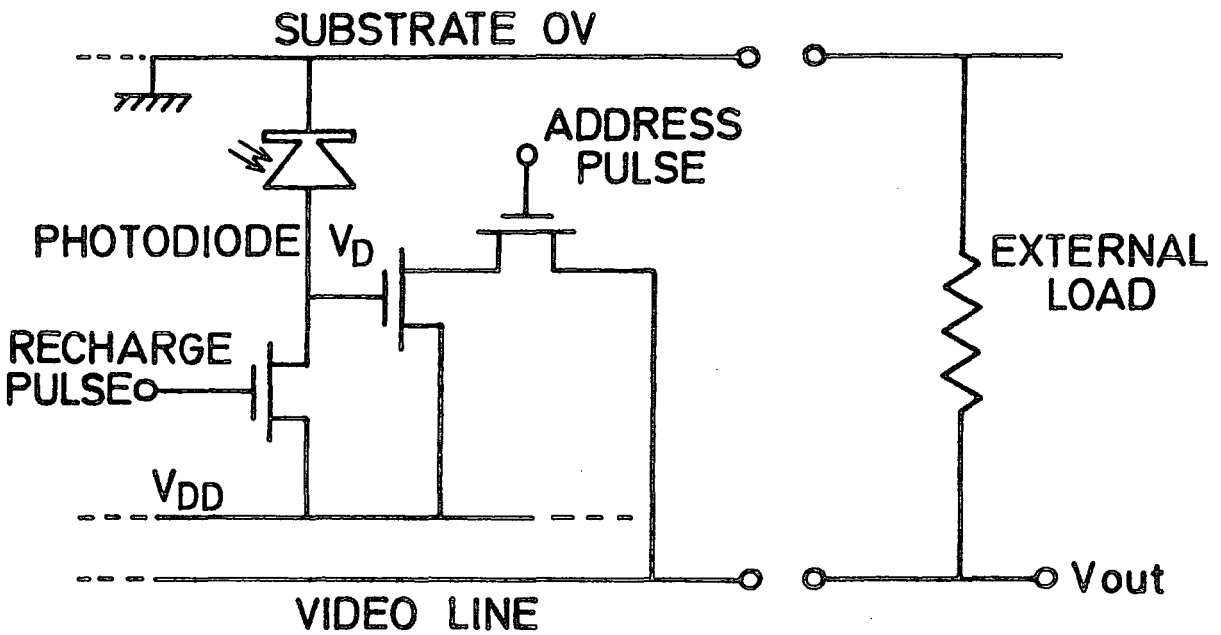
has also been found to be significant.[11]

Table 1.1

Device Reticon 512/24B	Description 512 Diode R.S. Array	Spacing/ Pitch 25 μm	Aperture/ Array width 600 μm	(pF) Cv 50	(pF) C_{RES} 0.6	(els) Q_{MAX} 3×10^7	NDRO X
Fairchild CCD - 101	500 Element Linear CCD	30 μm	30 μm	0.3	NA	1.5×10^6	X
GE - CID 2D Array	100 x 100 Element CID	60 μm	60 μm	27	NA	5×10^6	✓
IPL 4256	Voltage Sampling Diode Array	100 μm	100 μm	0.7	0.7	5×10^7	X
Plessey H 509	256/512 Linear V.S. Diode Array	50 μm	200 μm	0.1	0.1	5×10^6	✓

4. Readout Modes

A photodiode operated in Photon Flux Integration mode may be read out by either measuring the charge which has to be replaced to restore it to its original reverse bias voltage or directly monitoring the decrease of the diode voltage through a high impedance buffer amplifier. Recharge sampling mode and voltage sampling mode are shown in Figures (1.6) and (1.7) respectively. The first method, that of recharge sampling, is more widely used in scientific applications because of its inherent linearity. As it also requires much simpler on-chip circuitry, large arrays of small spacing may be fabricated with high yield. On the other hand, elaborate external processing is needed in order to convert the current pulse on the recharge



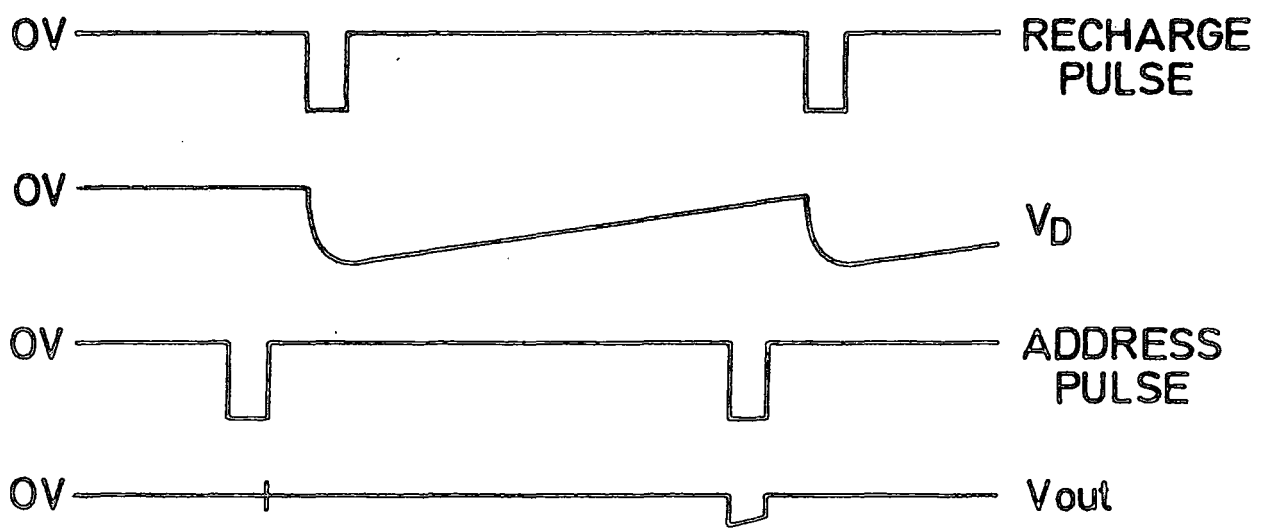


FIGURE 1.7 VOLTAGE SAMPLING

line to an output voltage proportional to the total charge replaced. Conversely, voltage sampling arrays combine more on-chip circuitry with simple external processing and allow for continual monitoring of the signal without necessarily destroying the data since the diodes do not have to be recharged on read-out. With this technique the main drawback is the non-linearity introduced by the variation of the diode capacitance with reverse bias.(12)

5. Noise

Whichever readout system is employed the final form of the video is usually a serial train of voltages or a boxcar waveform. This output is subject to two distinct types of noise:

(1) spatial or fixed pattern noise which is the residual signal seen when an array is read out under zero or flat field illumination; and

(2) random noise which is a combination of the statistical uncertainties in the detection process and of the electronic noise of the system.

The dominant sources of these depend upon the readout mode and will feature in later discussion.

6(a) Operation and Noise Performance of a Recharge Sampling Array

If a reverse biased photodiode is exposed to constant illumination for a set time, t_{ex} after which it is recharged to the original reverse bias voltage, Vo, the signal charge is given by

$$Q_{SIG} = \alpha N t_{ex} = \int_{t}^{t+T} i(t)\, dt, \qquad (1\text{-}1)$$

where α = the responsive quantum efficiency,

N = the number of incident photons per unit time,

t_{ex} = the exposure time,

$i(t)$ = the recharge current, and

T = the time for which the diode is recharged.

This information may be recovered (i.e. ΔQ_{SIG} may be converted to a voltage), either by allowing the current $i(t)$ to be supplied from the video line capacitance C_{VID} and thereby produce a voltage change,

$$\Delta V_{SIG} = \frac{\Delta Q_{SIG}}{C_{VID}}$$

on the video line or, alternatively, the recharge current, $i(t)$ may be amplified and subsequently integrated on an external capacitance, C_{ext}, such that

$$\Delta V_{SIG} = \frac{A}{C_{ext}} \Delta Q_{SIG} \quad ,$$

where A = the current gain of the system.[13]

Considering the operation of an array of diodes, spatial noise will be introduced in two distinct ways:

(i) effects which depend upon how the array is operated, such as switching spikes which may be picked up on the video line and make a contribution to the output of the charge integrator; and

(ii) variations in chip parameters such as interelement spreads in RQE and diode active area. Thus the output voltage of any processing system for an element i of an array is

$$V_{SIG} = Ai + Bi\ Q_{SIG}.$$

In an application where the signal is to be digitized and the data handled by a computer, these two components of the spatial noise may be removed fairly readily. For the d.c. offsets, Ai this is done by subtracting a dark exposure. The responsivity variations then have to be taken out by dividing through with a set of Bi as determined from exposures with a known flat field. In other situations where the video signal is not digitized, such as broadcast applications, it may be desirable to reduce these effects as much as possible by careful processing and, possibly, real-time signal processing. However, if the data may be processed digitally at a later time, such refinements are not essential. It follows from the above that,

even with an array that is cooled sufficiently to produce negligible dark current, the final data are derived from the differences of two readouts:

$$Si = \frac{V_{SIG\ i}(LIGHT) - V_{SIGi}\ (DARK)}{Bi}$$

If the system noise introduced by each readout, σ is independent of signal and if errors in the determination of Bi may be neglected, the total readout noise will be

$$\text{Total Readout Noise} = (2\sigma^2 + \alpha N t_{ex})^{\frac{1}{2}} ,$$

and the signal to noise ratio has the form

$$S/N\ i = \frac{\alpha N t_{ex}}{(\alpha N t_{ex} + 2\sigma^2)^{\frac{1}{2}}} \qquad (1\text{-}2)$$

The readout noise, sets a lower limit to the usefulness of the detector and is usually expressed in equivalent electrons so that it may be compared directly to detected photons. A division may be made of this readout noise into two categories: firstly, factors which are to some extent avoidable, such as poor noise performance of circuit components, pick up of external signals and noise introduced through the power supplies; and secondly, certain "limiting noise sources"[14] which are considered below.

6(b) Thermal Shot Noise

Noise is introduced by the statistical nature of the thermal response generation process. In an operating situation where the device is cooled such that the dark current is negligible, this noise contribution should not be important.

6(c) Reset Noise

Reset noise arises from the thermal noise in the resistance associated with charging a capacitor, in this case, the photodiode. The result of this is an uncertainty in the charge placed on the diode capacitance, C_D at the time of reset. Classical arguments give this noise expressed as

$$\text{Noise (Reset)} = \frac{1}{e} \left(KTC_D\right)^{\frac{1}{2}}. \qquad (1\text{-}3)$$

A more practical expression is

$$\text{Noise (Reset)} = 400 \sqrt{CpF} \text{ electrons}$$

at room temperature. Van der Ziel has pointed out[15] that the above result is only true for lossless dielectrics and in situations where the system is considered over an infinite bandwidth. In any practical situation Equation (1-3) should be regarded as providing an upper bound to the noise introduced by a single reset. It should be remembered, however, that there are in fact two resets associated with every readout of a recharge sampling array, one reset to initialize the diode at the beginning of the integration and another final reset when the signal charge is measured.

6(d) Amplifier Noise

Amplifier noise is the combined effect of all electronic noise processes affecting the video output at the time of readout. This noise is distinct from the limiting noise sources mentioned so far in that it may be minimized by reducing the bandwidth of the signal processing stages which produce the video output. The most important contribution to this amplifier noise usually comes from the first stage connected to the array video line. If this stage is a JFET, the noise introduced, expressed as an equivalent charge on the video line capacitance, is of the form

$$\text{Noise (Amplifier)} = \frac{C_{VID}}{e} \left(4\, KTBR_n\right)^{\frac{1}{2}}, \qquad (1\text{-}4)$$

where C_{VID} = the video line capacitance,

B = the system bandwidth,

Rn = the equivalent noise resistance of the JFET, and

the other symbols have their usual meaning. Equation (1-4) demonstrates that the capacitance at the input of the first amplifier stage is significant

in determining the importance of the amplifier noise contribution. For recharge sampling arrays, this input capacitance is the video line capacitance of the device, which may be as much as 50 pF, and therefore this noise source is often found to be the dominant factor limiting performance in such arrays.

7(a) Operation and Noise Performance of a Voltage Sampling Array

As has been mentioned, the residual voltage on a photodiode operated in charge storage mode may be monitored through an on-chip buffer amplifier, but because of the voltage dependence of the diode's self-capacitance, the change in voltage observed is not a linear measure of the total amount of light detected. However, for some applications where linearity is not necessary, such as shadow edge detection, voltage sampling arrays can be useful since less external processing circuitry is required for their operation.

7(b) Source Follower

The on-chip buffer transistor is usually a MOSFET operating in source follower mode with the video output of the device connected to an external constant current source. Considering the circuit of Figure (1-7), the drain current of the buffer transistor, which is assumed to be operating in saturation, must equal the external current, I const:

$$I_D = \frac{\beta}{2}(V_D - V_T - V_0)^2 = I\ \text{const}, \qquad (1\text{-}5)$$

where V_D = the diode bias at the gate of the transistor,

V_T = the threshold voltage of the transistor,

V_0 = the source voltage,

and β = the gain parameter of the transistor.

Rearranging Equation (1-5) gives

$$V_0 = V_D - \left\{V_T + \sqrt{\frac{2\ I\ \text{const}}{\beta}}\right\},$$

and including the voltage dropped by the resistance of the switching transistor, assuming that this is ohmic, the resulting video output voltage is

$$V_{OUT} = V_D - (V_T + \sqrt{\frac{2\ I_{const}}{\beta}} + I_{const}\ R_{ON}).$$

This suggests that the output voltage will follow the residual diode voltage fairly accurately, although in practice V_T is slightly dependent on V_0 because of a MOST property called body effect. Spreads in the parameters V_T, β and R_{ON} along the array produce a fixed offset pattern and interelement variations in diode capacitance also contribute to the spatial noise. Overall fixed pattern variations are generally worse than those found in recharge sampling arrays.

7(c) Machine Readable Dynamic Range

For a shadow edge incident anywhere on the array to be detected by discrimination of the output boxcar video waveform with a single threshold, the signal voltage output by diodes which are illuminated must differ from that output by those in the dark by an amount greater than the maximum fixed pattern excursion. Furthermore, multilevel discrimination is only possible for

$$n = \frac{V_{VW}}{V_{FP}}$$

levels, where V_{VW} = the maximum signal excursion, called the video window,

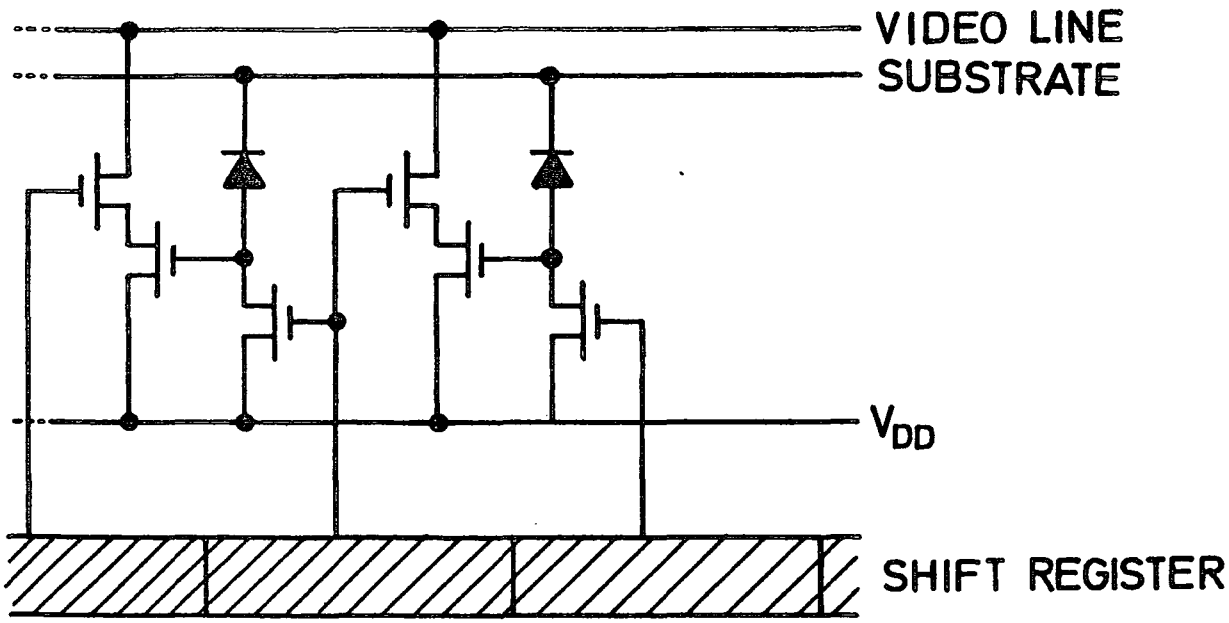
V_{FP} = the maximum fixed pattern excursion,

and n = the machine readable dynamic range.

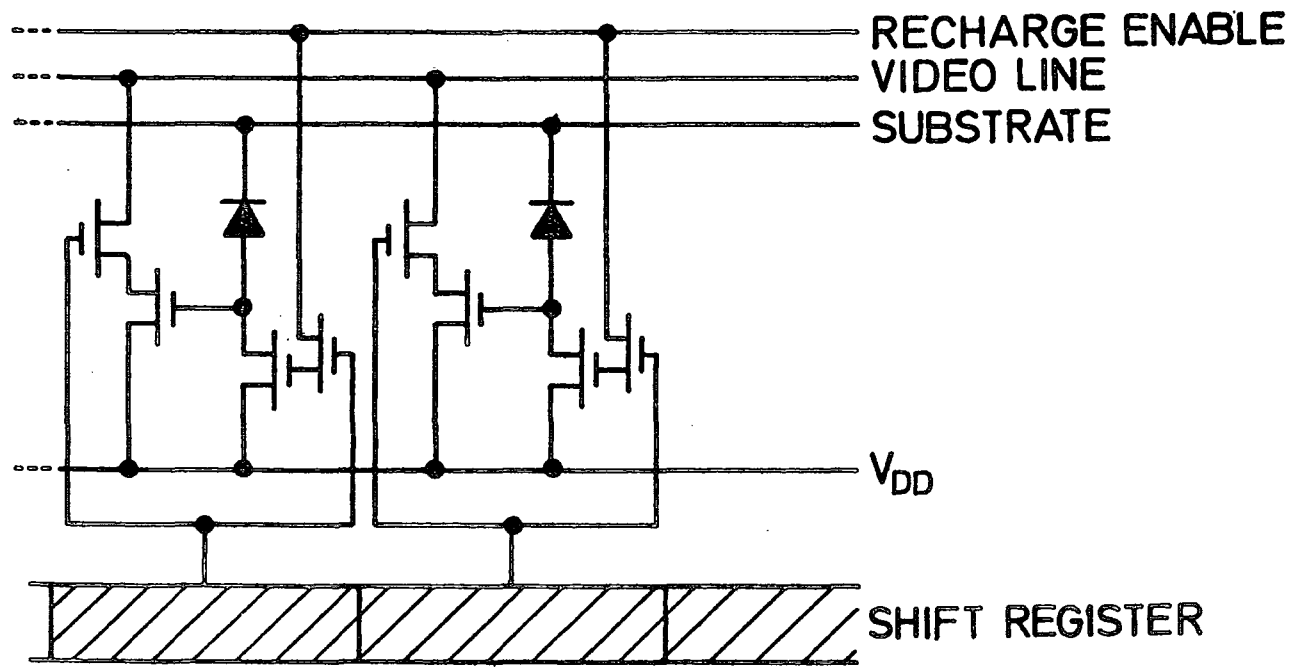
Thus recharge sampling arrays are often preferred in simple real-time applications since they have lower spatial noise.

7(d) Random Noise in Voltage Sampling Arrays

In order to use a voltage sampling array for some "scientific" application it would be necessary to obtain and store a calibration



(a)



(b)

FIGURE 1.8 ORGANIZATION OF VOLTAGE SAMPLING ARRAYS

curve for each element so that the fundamental non-linearity introduced by the voltage dependence of the diode capacitance may be removed from the data. In practice, this probably means approximating the response of each element with some higher order polynomial than that characterised by the two parameters Ai and Bi used in the case of recharge sampling arrays. This calibration would require a number of flat field runs of known relative intensity to be done and would therefore considerably complicate the operating procedure, leading to loss of observing time if the device is being used for astronomical measurements. Nevertheless the above is feasible and, if done, might yield a device with potentially superior random noise performance to that offered by recharge sampling arrays. However, a number of operating procedures are possible for voltage sampling arrays and the limiting noise sources depend upon which approach is employed.

Considering a single voltage sampling element, as shown in Figure (1-7), it will be seen that two control pulses, one to interrogate the element and one to reset it, are required. One possible arrangement, which appears in Figure (1-8a), is to have a common supply line for both recharge and buffer transistor with each element being reset by the interrogating pulse to the next element. Such a device necessarily has destructive readout but this organisation simplifies the external pulses needed to drive the array. Alternatively, control of the reset operation may be brought out independently as with the design of Figure (1-8(b)).

To obtain a value for the change in voltage on the diode after a given exposure, two measurements must be subtracted. Ideally these measurements will be the output voltage immediately following the initial recharge and that at the end of the integration time. Since no reset occurs in the intervening time, the data should not be subject to reset noise as described for recharge sampling arrays. Furthermore, the

capacitance at the input of the buffer transistor is the diode capacitance itself, usually less than 1pF, which is very much smaller than the video line capacitances previously mentioned. In principle the random noise may be reduced even further by performing both the initial and final readouts several times in order to calculate mean values. To do this it would be necessary to define the beginning and end of the exposure with a shutter, doing the multiple readouts in dark times before and after the integration. The problem with the above system is that, in practice, low frequency noise and system drifts, occurring between initial and final readouts (whether single or multiple), are a greater source of error than the reset noise which this technique eliminates.

An alternative voltage sampling method is to recharge the diode again at the end of the integration in order to obtain a zero signal reference level which may be measured much closer in time to the signal readout. With an array of the type shown in Figure 1.8(b) it is possible to recharge the element in the middle of the interrogation time, thereby actually observing the reset operation. In this case, the signal and reset measurements for each diode may both be made in a single scan, the minimum separation time being determined by the time taken for the video output to settle and the speed with which the levels may be accurately sampled. Such operation, designed to minimize the effect of low frequency noise, is termed correlated double sampling. Its disadvantages are that: it makes the readout destructive; the possibility of noise reduction by averaging becomes impractical; and reset noise is once again introduced into the data. In fact, from the point of view of a single element, correlated double sampling is effectively very similar to standard recharge sampling. The potential advantage of lower video amplifier input capacitance is nonetheless retained.

8(a) Factors affecting the use of Solid State Imaging Arrays for Astronomical Spectroscopy

The suitability of a particular imaging array for astronomical spectroscopy depends, firstly, on its geometrical layout, and secondly, upon its optoelectronic properties. In practice, no single device presently available is best in all respects, the optimum choice differing for various spectrographs and specific applications.

8(b) Array dimensions

Any spectrum will have a physical width and a spatial resolution determined by the optical properties of the telescope-spectrograph system used and the apparent size of the object emitting the light to be analysed. If the resolution of the system is not to be wasted, the spatial resolution of the array, as given by the element pitch, should be equally as good. Ideally, according to the sampling theorem, the spatial frequency of the elements should be twice the highest spatial frequency content of the spectrum:

$$2 \times \text{PITCH} = \frac{1}{f_{SPMAX}}$$

Having a device with a pitch very much smaller than necessary is also wasteful because, along with generating a superfluously large number of data points, the readout noise added per unit wavelength interval is correspondingly increased, making it necessary to integrate for longer to obtain an adequate overall signal to noise ratio on a given spectral feature.

In a typical system, when looking at an apparently circular object, the projected width of the spectrum is very similar to the spatial resolution. (This can be seen if the zeroth order spectrum, which is simply an image of the object, is considered). Since the spectrum has to be registered on the array without losses as a result of light spilling

over the edges of the sensitive area, it is helpful if the array width is somewhat larger than the projected spectrum width (i.e. generally it is better if the array width is greater than its pitch). This geometry both assists the positioning of the spectrum on the array in the first place and tolerates some amount of movement of the image which often occurs in practice. From actual observing experience it has been found that changes in the projected position of the spectrum, perpendicular to the array, may result: after changes in the grating angle because the grating axis is slightly out of true; or, during a long exposure, if the telescope guidance system is not able to maintain the first image of the object in exactly the same position on the entrance slit of the spectrograph. Additionally, with a Cassegrain system, where the detector moves about with the telescope, there may be some flexure in the detector mountings. It should be noted that such movements along the array direction, occurring within the exposure time, may degrade the spatial resolution of the system.

Spectrographs designed for use with photographic plates produce, typically, about 50 mm in length of spectrum in the focal plane. To take advantage of the full wavelength coverage available for a given dispersion requires a detector array of this length.

8(c) Optoelectronic Performance

Properties such as large RQE, low noise and linearity are always good to have, but useful work may still be done with lower performance devices. The theoretical dynamic range of a single element of a recharge sampling array, with a typical low noise readout system, can be greater than 10^4 where, for experimental purposes, a signal to noise ratio of 100: 1 is usually more than adequate. However, if very faint objects are to be observed, the limiting factor is the integration time needed to give a reasonable S.N.R.

Neglecting dark signal and sky background, the faintest magnitude

of object that may be detected in any practical exposure time is determined by both the readout noise and the RQE. A figure of merit incorporating both these factors is the Detective Quantum Efficiency:

$$DQE = \frac{(\text{Output SNR})^2}{(\text{Input SNR})^2}$$

$$= \frac{\alpha^2 N}{\alpha N + \sigma_{TOT}^2 + N_B}$$

where α = RQE,

N = number of incident photons, and

σ_{TOT} = total noise added by the readout process expressed as an equivalent number of electrons,

and N_B = number of carriers produced thermally or by sky background etc.

The DQE is useful for comparing detectors because it takes into account the performance of the device and also, through N and N_B, the application envisaged.

8(d) Currently Available Solid State Detectors

To date, linear diode arrays of the recharge sampling type have been employed for spectroscopic work because of their high quantum efficiency, linearity and ideal geometrical layout - that is, in preference to alternative devices such as CCDs which may have lower readout noise. Reticon diode arrays are available with up to 1024 elements on 25 μm centres, an array width of 2.5 mm and a RQE of about 75% at 600 nm. Similar earlier Reticon devices with apertures of 425 μm and 600 μm have been used in systems by several workers. Typical published readout noise figures are listed in Table (1.2) together with two alternative devices.

Table 1.2

SYSTEM	DEVICE	EQUIVALENT READOUT NOISE ELS.
Vogt et al.[16]	Reticon	750
Geary [17]	"	4700
Dravins [18]	"	2000
Livingston et al.[19]	"	950
Aikens et al.[20]	GE CID	800
Meyer [21]	Fairchild CCD 202	30

Since CCDs and CIDs generally have elements with almost square aspect ratios, it is impractical in most cases to use linear arrays of these for spectroscopy. Area arrays overcome this problem to some extent but often with a vast increase in data points and, although the noise per pixel is less than found in diode arrays, the fact that several rows of elements may have to be used reduces this benefit. Table (1.1) gives typical values of dimensions and electrical properties for a variety of devices.

REFERENCES

1. McMullen, D., Wellgate, G.B., Ormerod, J. and Dickson, J. Adv. in Electronics and Electron Physics. Vol.33, pp.873-879.

2. Moss, T.S., Burrell, G.J. and Ellis, B. (1973). Semiconductor Opto-Electronics. (Butterworths), p.302.

3. Snow, E.H. (1976). "Self-scanning photodiode arrays for spectroscopy." Research and Development. April, pp.18-22.

4. Geary, J.C. (1976). "Comments on the use of silicon diode arrays for direct spectroscopy." In proc. IAU colloquium 40. ed. Duchesne, M. and Lelievre, G., Paris-Meudon Obs., pp. 28-1-13.

5. Livingston, W.C. (1976). "Diode Arrays - A Review." In proc. IAU colloquium 40 ed. Duchesne, M. and Lelievre, G., Paris-Meudon Obs., pp. 22-1-13.

6. Weckler, G.P. (1967). "Operation of P-N Junction Photodetectors in a Photon-Flux Integrating Mode." IEEE Journal of Solid-State Circuits. Vol. SC-2, p.65.

7. Van der Ziel, A. (1976). Noise in Measurements. (Wiley). p.124.

8. Percival, J.W. and Nordsieck, K.H. (1979). "Thermal Background Subtraction in Photodiode Detectors." Wisconsin Astrophysics. No.88.

9(a) Kittel, C. (1971). Introduction to Solid State Physics. (Wiley), p.364.

9(b) Geary, J. (1975). Ph.D. Thesis, University of Arizona.

10. White, M.H. (1976). "Design of Solid-State Imaging Arrays." Solid State Imaging. (Noordhoff).

11. Private Feasibility Study, The Plessey Co.Ltd.

12. A good analysis can be found in Jespers, P. (1976) "XY Addressing." Solid State Imaging. (Noordhoff).

13. Buss, R.R., Tanaka, S.C. and Weckler, G.P. (1976). "Principles of Low-Noise Signal Extraction from Photodiode Arrays." Solid State Imaging. (Noordhoff).

14. Livingston, W.C. Op.Cit.

15. Van Der Ziel, A. (1973). "Equivalent circuit and equipartition theorem in lossy dielectric or ferroelectric capacitors." J.Appl.Phys. Vol.44, No.3, pp.1402-3.

16. Vogt, S.S., Tull, R.G. and Kelton, P. (1978). "Self-Scanned Photodiode Array: High Performance Operation in High Dispersion Astronomical Spectroscopy." Appl. Opt. Vol.17, pp.574-592.

17. Geary, J.C. (1976). "Comments on the use of silicon diode arrays for direct spectroscopy." In proc. IAU colloquium 40. ed. Duchesne, M. and Lelievre, G., Paris-Meudon Obs., pp.22-1-13.

18. Dravins, D. "A self scanned silicon diode array for astronomical photometry." (1975). Image Processing Techniques in Astronomy. ed. De Jager, C. and Nieuwenhuijen, H., Astrophys. and Space Sci.Lib. Vol.54, p.97.

19. Livingston, W.C., Harvey, J., Slaughter, C. and Trumbo, D. (1976). "Solar Magnetograph employing integrated diode arrays." Appl. Opt. Vol.15, pp.50-52.

20. Aikens, R.S., Harvey, J.W. and Lynds, C.R. (1976). "The Kitt Peak CID systems." IAU colloquium 40. ed. Duchesne, M. and Lelievre, G., p.25-1.

21. Meyer, S.S. (1980). "Astronomical spectrometer using a Charge Coupled Device Detector." Rev. Sci. Instrum. Vol.51(5), pp.638-641.

CHAPTER 2

THE PLESSEY H509 ARRAY

1. Introduction

The diode array described here was designed and produced by the Plessey Co.Ltd. in collaboration with a group working at the R.G.O.,[1] for use in intensified mode. As it was intended for operation under electron bombardment and built specifically for an astronomical application, it includes some features not found in the commercially available diode arrays more often used for spectroscopy:-

(i) The diodes are accessed in pairs;

(ii) Even and odd numbered diodes have separate recharge lines;

(iii) It is a voltage sampling array and each element incorporates an on-chip voltage amplifier;

(iv) Each chip has four separate arrays.

In this chapter the operation of an array, an external amplifier and basic drive circuitry will be discussed.

2(a) Layout and Scanning

Each chip has four arrays of 256 diodes, arranged in two parallel lines of 512 diodes and is mounted in a 24 pin d.i.l. package as shown in Figure (2.1). Arrays 1 and 2 have common supply voltages and control pulses as do arrays 3 and 4, but each array has its own video output. To date, work at Durham has used only one array of 256 diodes at any one time. However, facilities exist to operate two arrays side by side for comparison purposes (e.g. Star-Sky subtraction), or two arrays in line as a single long array. This has not been done since none of the initial batch of four devices had the necessary combinations of fully operational arrays owing to manufacturing problems. More devices have subsequently been obtained from the manufacturer, some of which have all four arrays

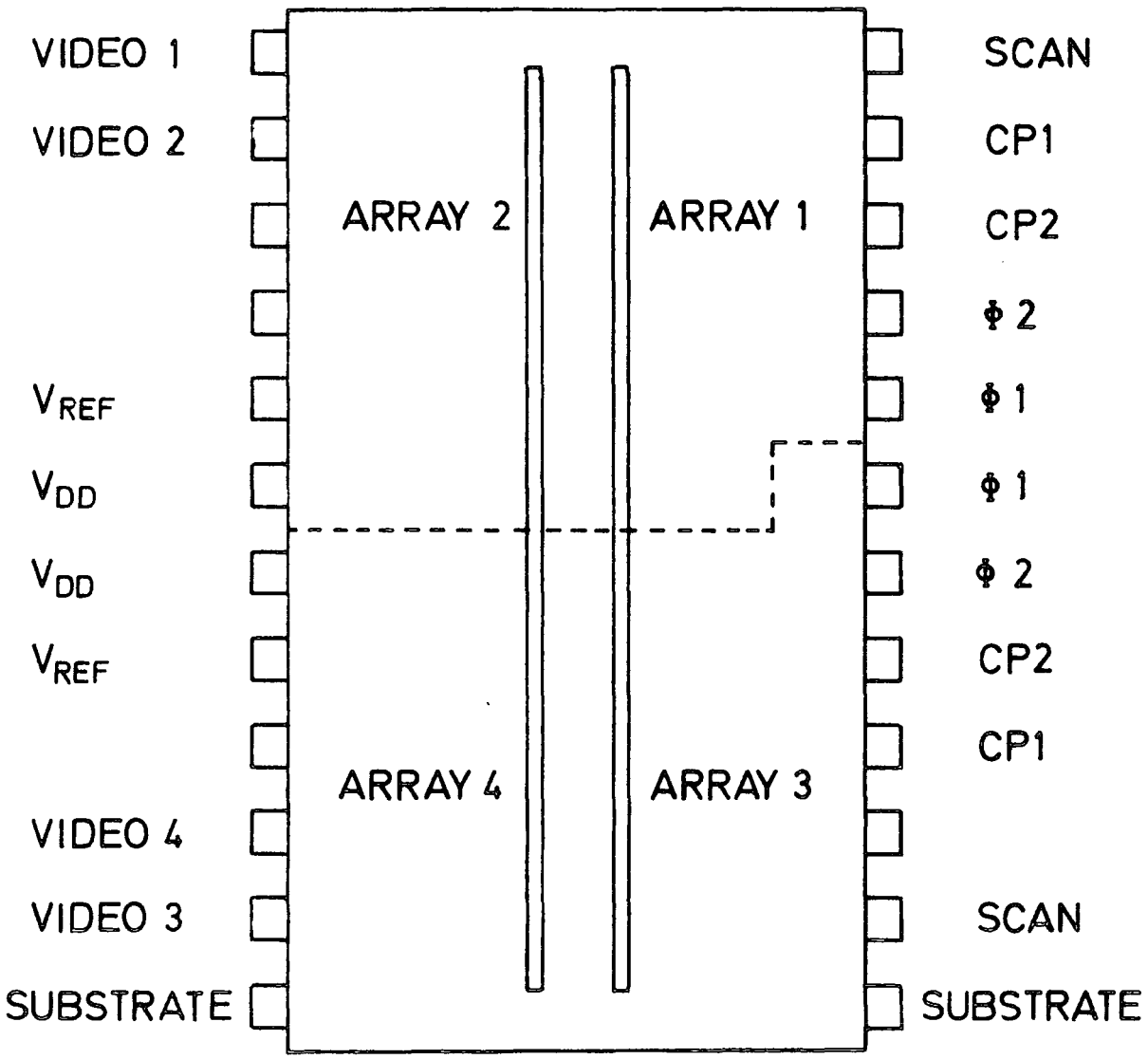


FIGURE 2·1 H509 ARRAY

working. The individual diodes have dimensions of 200 μm x 40 μm and are on 50 μm centres giving an overall array length of approximately 13 mm. The two lines of arrays have a separation between centres of approximately 1.2 mm.

Each array is scanned from the outside towards the middle of the chip by two, 128 element, shift registers; the first addressing odd and the second even elements of the array. These are loaded and clocked in phase which results in alternate pairs of diodes being accessed together. This design, shown in Figure (2.2(a)), doubles the maximum data rate and puts the electronics associated with odd and even diodes on opposite sides of the array, thereby easing the manufacturing constraints on minimum element spacing. The video output, now as a composite of signals from pairs of diodes, must be separated in the external signal processing stage although this leads to greater complexity. The amplifier noise due to on-chip electronics is increased by a factor of $\sqrt{2}$. Figure (2.2(b)) shows part of the composite chip mask.

2(b) Circuit and Operation of a single Element

This array may be thought of as combining some of the features of recharge and voltage sampling. Within each element, shown in Figure (2.3), the photodiode is continuously recharged by the feedback action of transistor T1, such that V_D remains always at some value set by V_{REF}. For simplicity, all voltages and currents will be considered as positive although for a p-channel device these will in fact be negative. The drain current of T1 which is operated in saturation is given by[2]

$$I_D = \frac{\beta}{2} (V_{REF} - V_T - V_D)^2 , \qquad (2\text{-}1)$$

where V_{REF} = the external reference voltage,

V_T = the threshold voltage,

V_D = the diode reverse bias voltage,

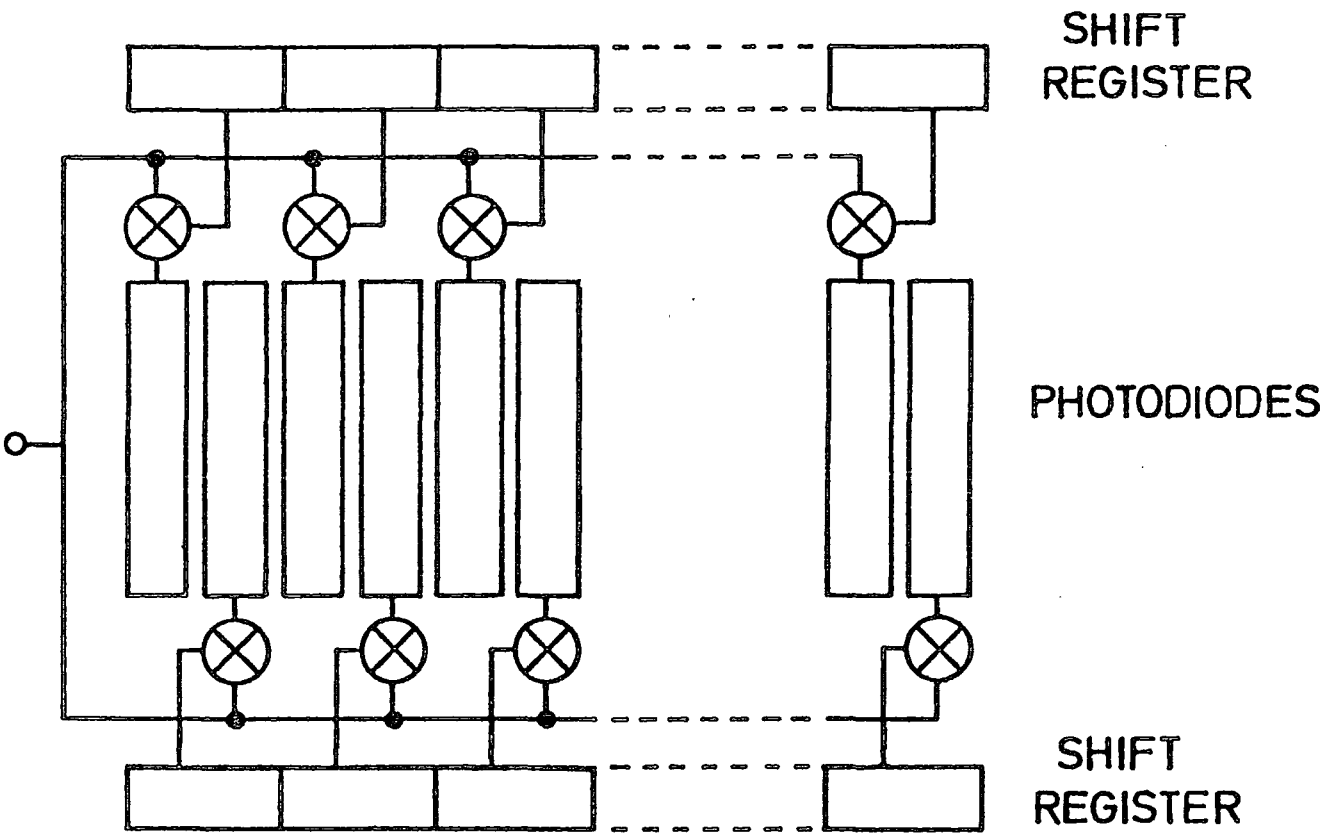


FIGURE 2.2(a) PAIRED ORGANIZATION

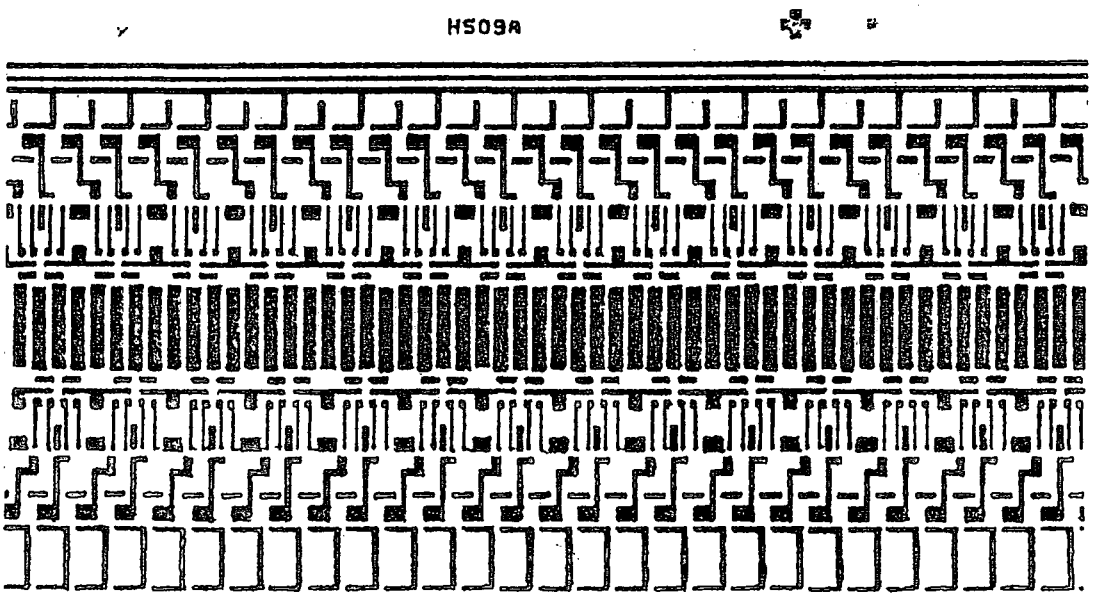


FIGURE 2.2(b) CHIP MASK

β = the gain parameter of the transistor,

which rearranging gives

$$V_D = V_{REF} - V_T - \left\{\frac{2I_D}{\beta}\right\}^{\frac{1}{2}} \qquad (2\text{-}2)$$

In equilibrium, the drain current I_D will exactly equal the photocurrent generated in the diode; the latter is small under normal operating conditions. For a signal resulting in saturation in 10^{-1} seconds, the largest value normally considered, the photocurrent is

$$Ip \simeq 10^{-12} A,$$

and the gain paramater for T1

$$\beta \simeq 4 \times 10^{-6} \; AV^{-2}.$$

Therefore,

$$\left\{\frac{2Ip}{\beta}\right\}^{\frac{1}{2}} \ll V_{REF} - V_T \;,$$

and $$V_D = V_{REF} - V_T.$$

The effective threshold, V_T, is not a constant but varies slightly with the transistor source voltage V_D. An empirical relationship for this is given by the manufacturers:

$$V_T = V_{TO} + 0.4 \sqrt{V_D} \;, \qquad (2\text{-}3)$$

where V_{TO} is a constant for the fabrication process. Figure (2.4) shows the diode bias plotted as a function of V_{REF} using Equation (2-3).

Any current, Ip, generated either optically or thermally in the diode, discharges not its own capacitance, C_D, but the capacitance associated with the gate of the buffer output transistor, T3. This capacitance comprises the diffusion capacitances of T1 and T2, any stray interconnection capacitance and the gate to substrate capacitance of T3. As the latter two contributions which are independent of voltage are expected to dominate, this arrangement should give improved linearity over the response of a standard voltage sampling element. Thus the change in gate voltage at T3 is given by

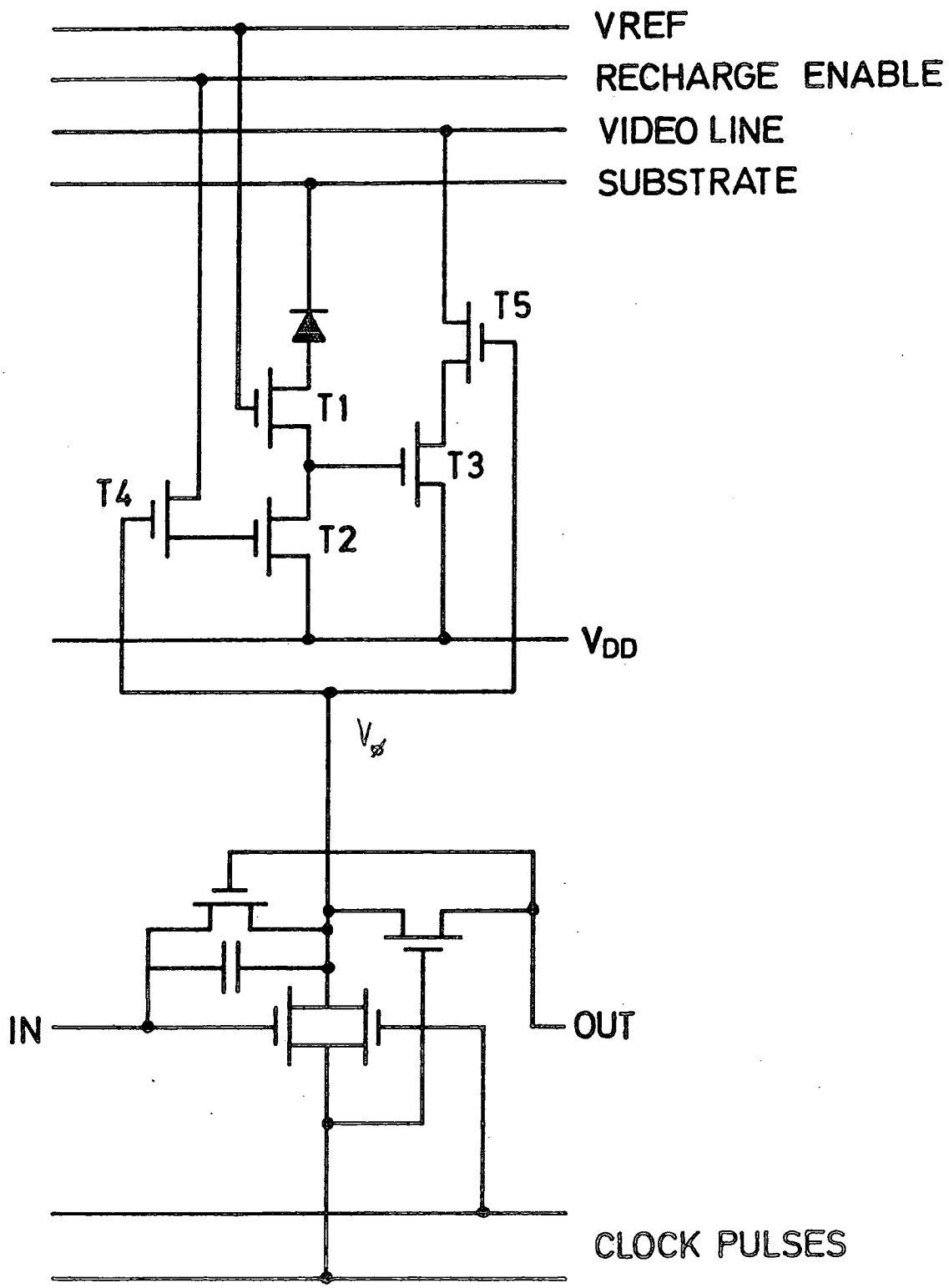


FIGURE 2·3 AMPLIFIED ELEMENT

$$V_G = \frac{1}{C_o}\int_o^t I_p\,dt \quad = \frac{1}{C_o}\int_o^t \alpha NA\,dt \quad , \qquad (2\text{-}4)$$

where Co = total capacitance at gate of T3,

α = R.Q.E.,

N = No of incident photons per unit area per unit time, and

A = the sensitive area of the photodiode.

As the output signal does not depend upon the capacitance of the diode, the sensitive area, A, may be made larger than is usual for voltage sampling arrays. Such a system may be thought of as introducing a notional voltage gain of

$$C_D / C_o \sim 10 \text{ (Manufacturer's estimate)}$$

over the response of an otherwise similar voltage sampling element without transistor T1. This may be found confusing because, as has been said, the output does not depend on C_D. A disadvantage of integrating on Co, which is an order of magnitude smaller than C_D, is that the maximum signal charge, Qmax (see Table 1.1), is decreased. For saturated operation of T1 and proper feedback action, its drain voltage, which is the gate voltage of T3 must satisfy

$$V_G > V_{REF} - V_T = V_D.$$

Thus if V_G is originally charged to the chip supply voltage V_{DD}, the maximum voltage excursion on Co is given by

$$\text{Maximum } \Delta V_G = V_{DD} - V_D, \text{ and}$$

$$Q_{MAX} = \frac{1}{e}(V_{DD} - V_D)\,C_o, \text{ typically } 3 \times 10^6 \text{ electrons.}$$

(In practice the lower observable limit of V_G is set by the operating requirements of T3, the buffer transistor). However, since the diode reverse bias voltage is now constant, this introduces a number of benefits:-

(i) the response at the blue and red ends of the spectrum will not vary throughout the exposure, because the depletion layer width is fixed;

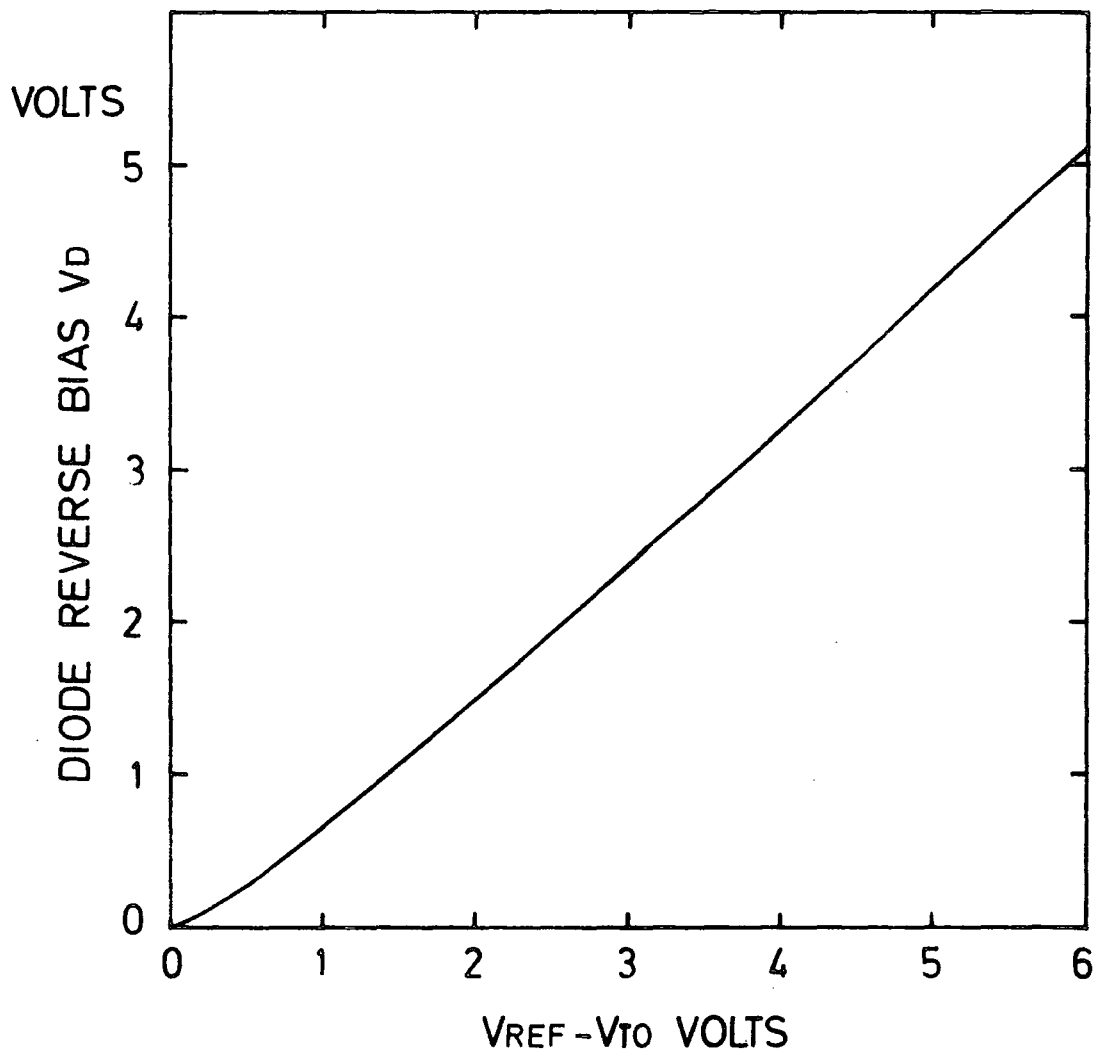


FIGURE 2.4 DIODE BIAS V_D VS V_{REF}

(ii) the dark current which depends on the volume of the depletion layer should be constant throughout the exposure;

(iii) the reverse bias may be chosen to be quite small, thereby reducing the dark current and associated shot noise, although this may impair the potential spectral response.

For the original application of electron counting, at temperatures where dark current is a major problem, the last consideration is most important and T1 can be thought of, with suitably chosen V_D, as enhancing the signal more than the dark current and improving the overall signal to noise ratio.

2(c) Recharge Operation and Control Pulse Amplitudes

A simplified circuit of the detector element is shown in Figure (2.5(a)). When an element is addressed, the enable pulse from the shift register switches on transistor T5 and the signal appears on the video line. The signal's exact relationship to the gate voltage, V_G depends upon the external load employed, as will be discussed later. At the same time T4 is also switched on and, if a recharge pulse is applied on the charge pulse line, it will switch on transistor T2 and the gate capacitance Co will be recharged to V_{DD}. The voltages V_ϕ and V_{DD} are chosen such that when T5 and T2 are switched on, they are in unsaturated operation and have relatively low resistances.

Considering transistor T2, for unsaturated operation

$$V_{CP} > V_{DD} + V_{T2}$$

where V_{T2} = the effective threshold voltage of T2,

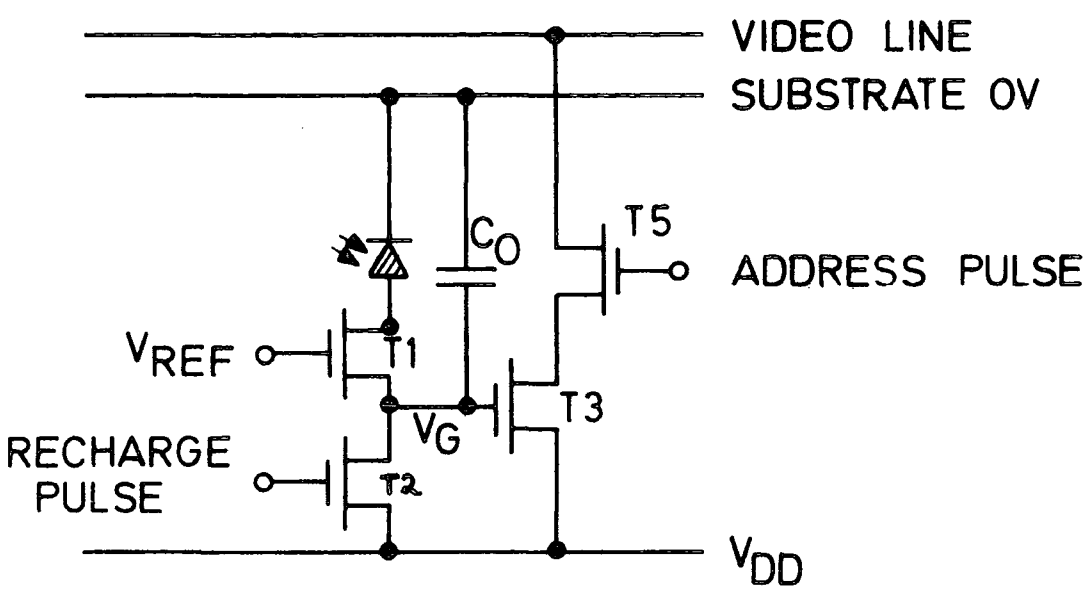
and considering transistor T4

$$V_{CP} = V_\phi - V_{T4}.$$

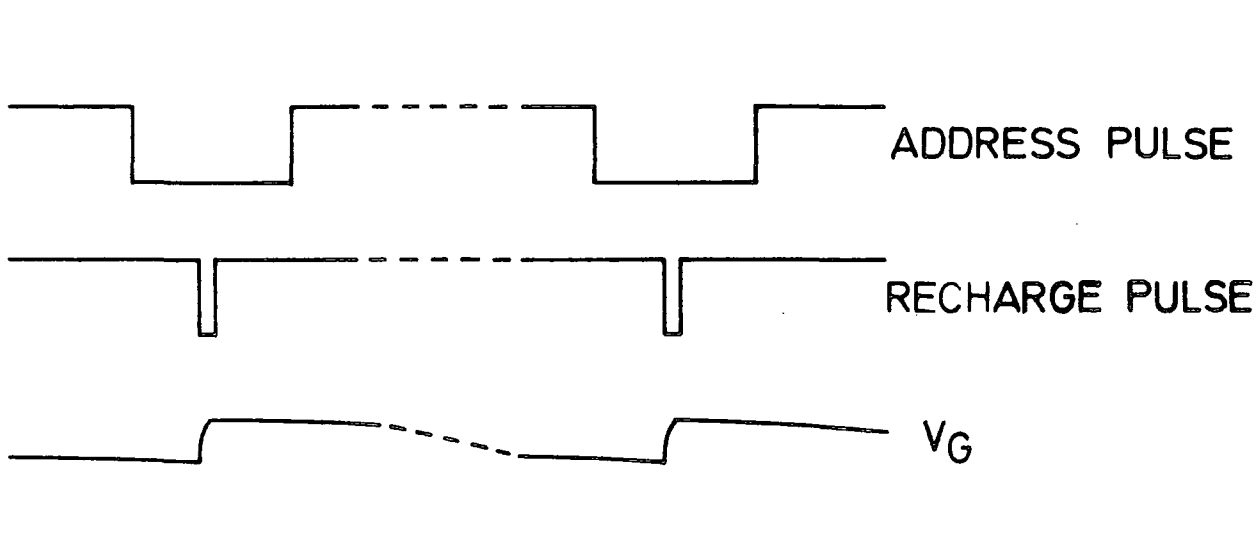
Thus V_ϕ must satisfy the condition

$$V_\phi > V_{DD} + V_{T2} + V_{T4}, \qquad (2\text{-}5)$$

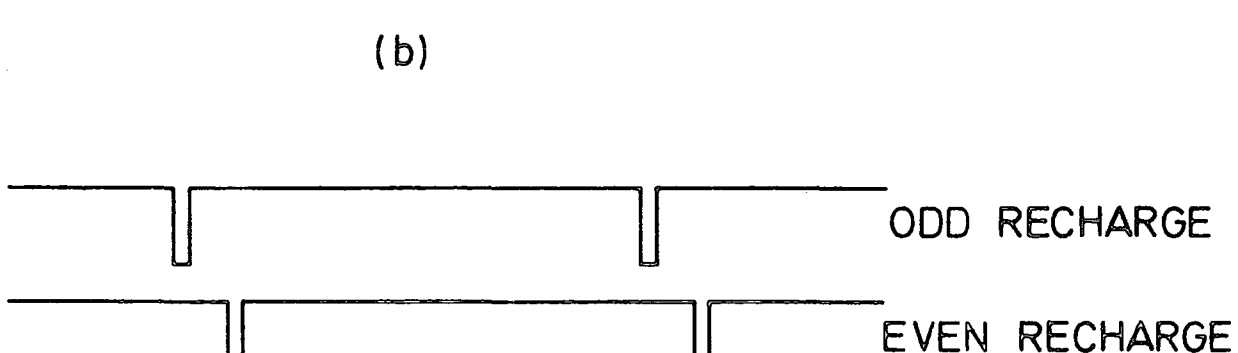
which, neglecting changes in V_T due to body effect, means that the clock



(a)



(b)



(c)

FIGURE 2.5 DETECTOR TIMING

voltage, V_{ϕ} , should be at least two thresholds greater than V_{DD}.

Since the recharge pulse may come at any time while the element is being looked at, the recharge action will be seen as it happens on the video output. This satisfies the requirement for correlated double sampling and, assuming a linear output amplifier configuration, the change in video level seen on recharge will be proportional to the integrated signal. The resulting output signal is shown in Figure (2.5(b)).

2(d) Operation of an Element Pair and N.D.R.O.

As the elements are enabled in pairs whose combined outputs are seen on the video line, it is necessary to recharge odd and even diodes at different times to separate the information from them. This is accomplished by having two independent charge pulse lines and performing two double samples within each access time. The timing for this is shown in Figure (2.5(c)). Non-destructive readout is possible if recharge pulses are not applied, but the output is some combination of the signals from both elements within each pair so that the spatial resolution of the device is halved.

2(e) External Amplifier Configuration

The choice of external amplifier configuration is limited by the need for the output on the video line to be the exact sum of two values which are each a function, preferably linear, of the respective gate voltages of the elements within the pair:

$$S_{VID} = f(V_{GO}) + f(V_{GE}), \qquad (2\text{-}6)$$

where S_{VID} = the signal from an element pair, and

V_{GO}, V_{GE} = the odd and even gate voltages.

If the video line is connected externally to a current source, as shown in Figure (2.6(a)), this is not the case, because the buffer transistors interact through the voltage on the video line. Clearly, for

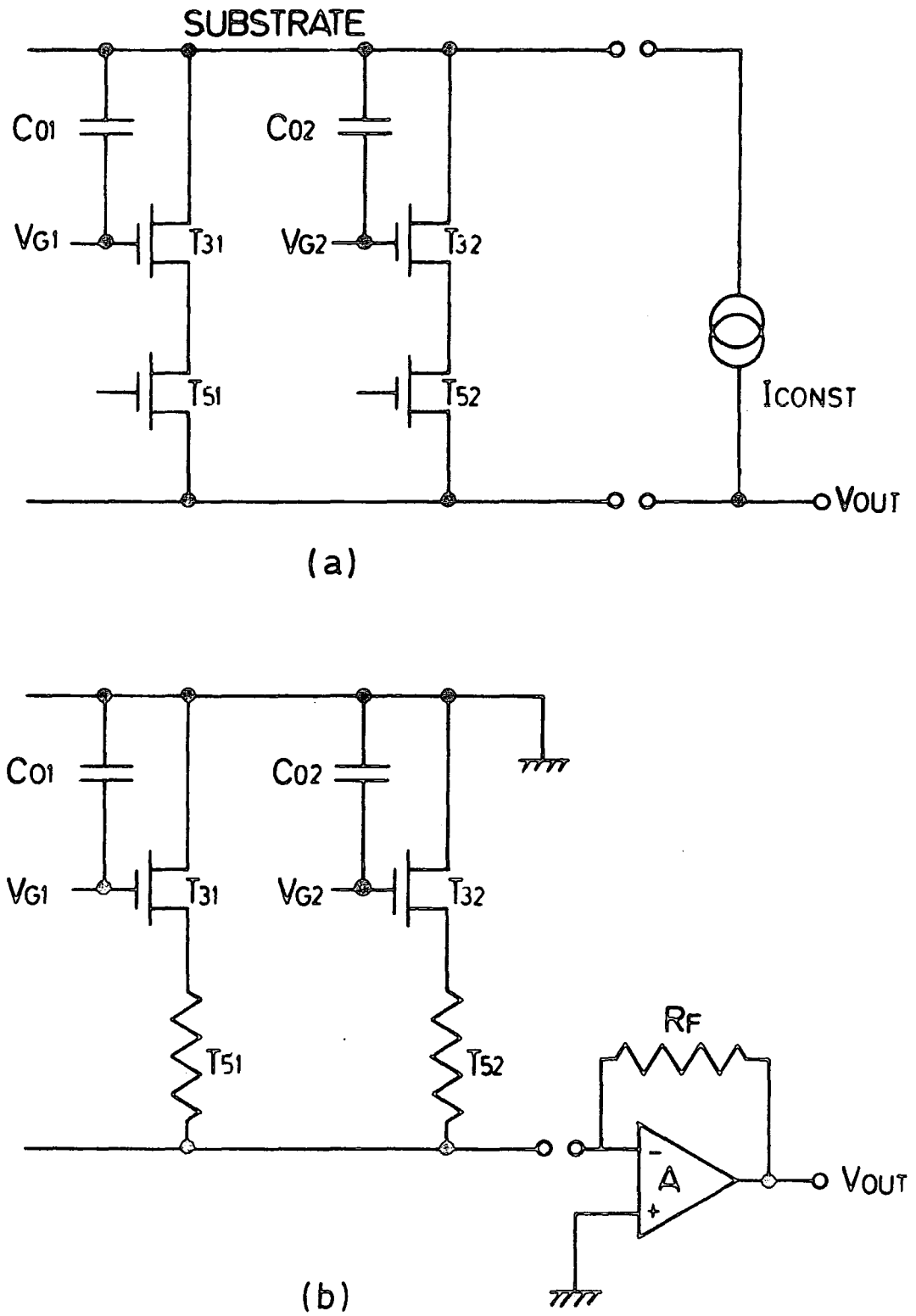


FIGURE 2·6 EXTERNAL LOADS

the two buffer transistors to operate completely independently of one another, the points where the two circuits join, i.e. the video line, must be fixed in potential. The current on the video line which will be the exact sum of the currents from each element can then be measured as follows:

$$I_{tot} = I_o + I_E = f(V_{GO}) + f(V_{GE}) ,$$

taking the form of Equation (2-6). This can be achieved by connecting the video line to a virtual earth amplifier, as shown in Figure (2.6(b)), converting the summed current to a voltage:

$$V_{OUT} = R_f \ (I_O + I_E),$$

where R_f = the feedback resistor employed, and

I_O, I_E = the signal currents from odd and even elements.

Thus the information from the elements may be separated, but the buffer transistors are no longer operating as ideal source followers and their output characteristics are not linear. Considering Figure (2.6(b)), the output current for one element is given by

$$I = \frac{\beta}{2} (V_G - V_T - IR)^2 \ , \text{ for saturated}$$

operation, where R = the 'on' resistance of T5.

Some manipulation yields

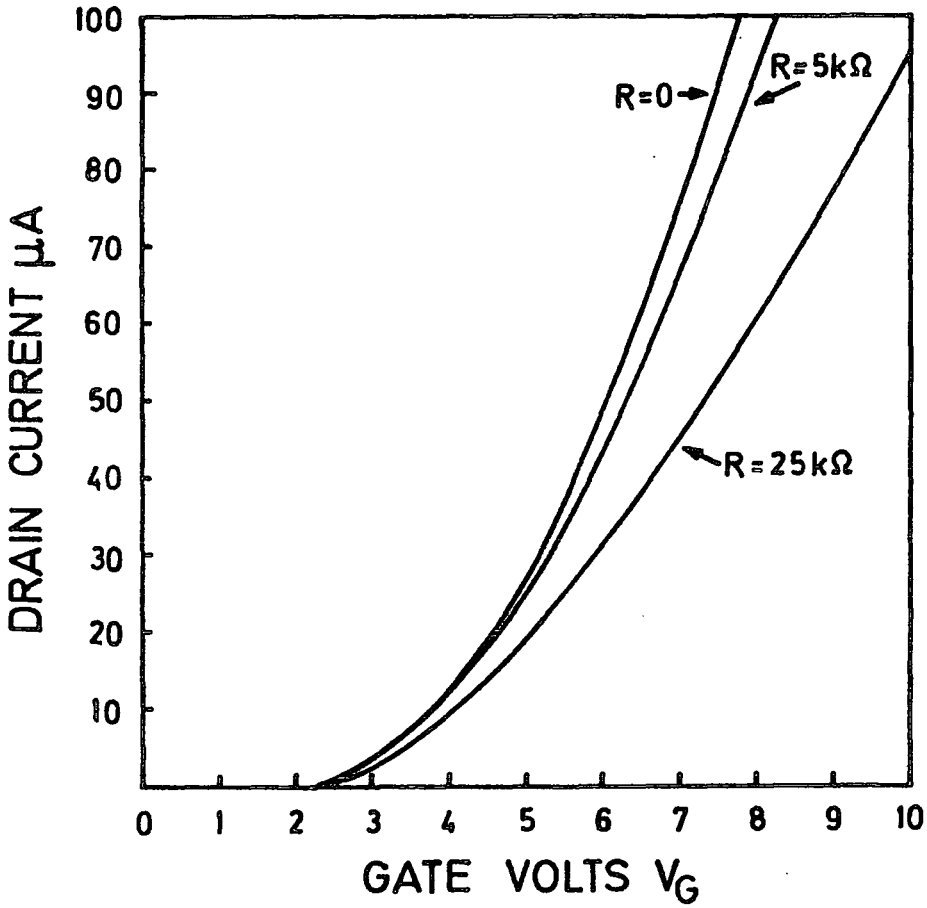
$$I = \frac{V_G - V_T}{R} + \frac{1}{\beta R^2} \left\{1 - (2(V_G - V_T) R\beta + 1)^{\frac{1}{2}}\right\}, \qquad (2-7)$$

which for $2(V_G - V_T) R\beta \ll 1$ may be expanded to give

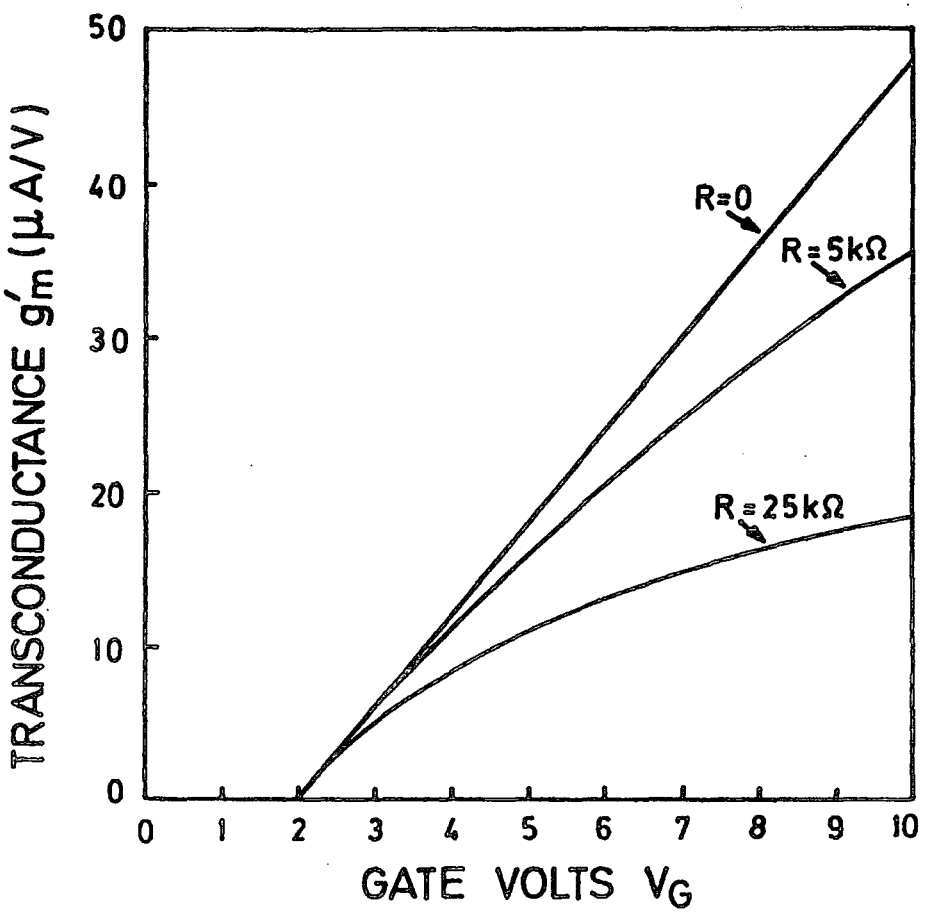
$$I = \frac{\beta}{2} (V_G - V_T)^2 \left\{1 - \beta R(V_G - V_T) + \frac{5}{4}(\beta R(V_G - V_T))^2\right\}, \qquad (2-8)$$

indicating that at low values of source resistance, R, the drain current is related to the gate voltage by the 'square law' and R has only a second order effect. For the switching transistor, T5, in unsaturated operation at low values of drain voltage, conduction is nearly ohmic[3] and given by

$$R_{DS} = \frac{1}{\beta_5(V_\phi - V_T)} \simeq 5K\Omega \ , \qquad (2-9)$$



(a)



(b)

FIGURE 2·7 SOURCE FOLLOWER RESPONSE

where β_5 = the gain parameter of T5: approx. 1.8×10^{-5} AV^{-2},

$V\phi$ = the clock pulse amplitude: typically 13V, and

V_T = the threshold voltage: approx. 2V.

Equation (2-9) may not hold for such a large value of V_ϕ. Similarly, it does not take into account parasitic resistances on-chip, though it should give a reasonable estimate of the source resistance, R. For typical values

$$2(V_G - V_T)\, R \simeq 0.3$$

where $V_G - V_T = 5V$,

$R = 5K\Omega$, and

$= 6 \times 10^{-6}$ AV^{-2},

the power series of Equation (2-8) converges slowly. Figure (2.7(a)) shows a graph of I vs. V_G (as described by Equation (2-7)) for a number of R values. Differentiating to find the small signal transconductance gives

$$g'_m = \frac{1}{R}\left\{1 - \frac{1}{(2(V_G - V_T)\, R\beta + 1)^{\frac{1}{2}}}\right\} . \qquad (2\text{-}10)$$

In the case of the ideal source follower, for large R

$$g'_m \longrightarrow \frac{1}{R} ,$$

but, as has been shown, this is not the case and g'_m varies significantly with V_G, i.e. the response is not linear. Again, this variation is plotted for a number of R values in Figure (2.7(b)). There appears to be no way around this problem apart from, as was suggested for standard voltage sampling arrays, individually calibrating every element. In practice the non-linearity introduced by this is not too great if the device is operated over a restricted range.

2(f) Spatial Noise Considerations

Responsivity variations will arise, as with recharge sampling arrays, owing to both variations in diode area (assuming the diodes are filled with uniform illumination) and to interelement variations in R.O.E. Further

differences will be introduced by spreads in the gate capacitance Co. As the output current is not a linear function of gate voltage, the overall spatial noise cannot be described by two parameters but interelement differences in R, β and V_T will produce a significant offset pattern. This is discussed in more detail by A.R. Hedge.(4)

3(a) Random Noise

The noise performance of this type of array is potentially very good because the reset capacitance and the video amplifier input capacitance, which are important factors in the limiting noise sources found in diode arrays, are now both given by the gate capacitance of the buffer transistor Co, approx. 0.1 pF.

The maximum uncertainty in signal change introduced by one reset operation is now given by(5)

$$\text{Maximum Reset Noise} \simeq 400\sqrt{CpF} \simeq 120 \text{ electrons},$$

at Room Temperature. Taking into account (a) the fact that double sample values have two associated resets and (b) the reduced temperature of operation of the device, a more practical figure is

$$\text{Maximum D.S. Reset Noise} \simeq 460\sqrt{CpF} \simeq 150 \text{ electrons},$$

at 200K. This figure is smaller than the reset noise encountered in most recharge sampling devices where the photodiode capacitance, typically 0.7 pF, must be recharged. However, even greater reductions might be expected in the case of amplifier noise which varies directly as input capacitance.

3(b) MOSFET Amplifier Noise

In general the amplifier noise when referred to the input, may be written

$$NEC_A = \frac{C_V}{e}(4KTBRn)^{\frac{1}{2}}, \qquad (2\text{-}11)$$

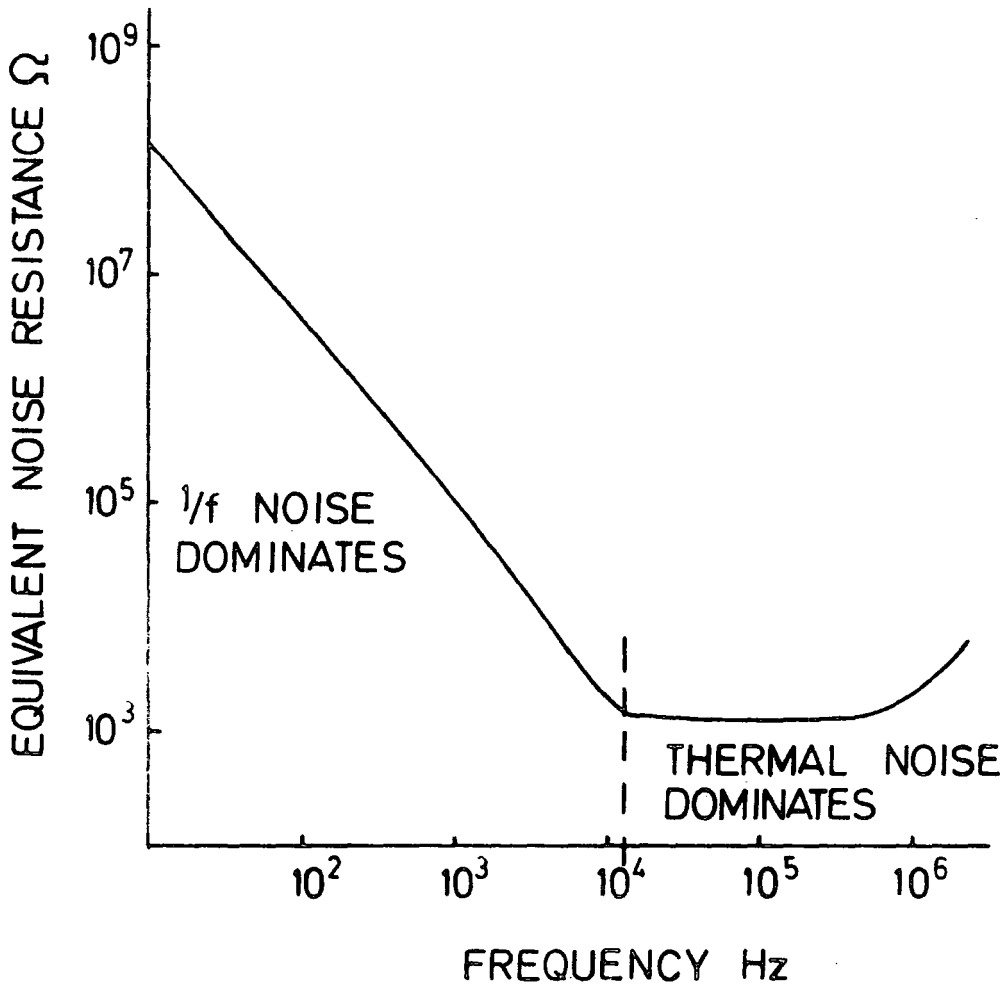


FIGURE 2.8 MOST NOISE SPECTRUM (IDEALIZED)

where Rn = the equivalent input noise resistance of the amplifier,

C_V = the amplifier input capacitance,

K = Boltzmann's constant,

B = the system bandwidth, and

e = the electronic charge.

This suggests an improvement of C_{VID}/C_o over recharge sampling arrays. (The factor may be as much as 500). Also, as the buffer amplifier transistor is on chip, i.e. T3, it is cooled to the device operating temperature, giving a further small improvement. However, MOSFETs on the whole have poorer inherent noise performance than the JFETs commonly used in the front end of low noise signal processing systems. There is also additional noise from T1 and, because two elements are looked at simultaneously, the on-chip amplifier noise of both is imposed on the data from each.

The noise spectrum produced by a typical MOST may be divided into roughly two regions, as has been described by Soares.[6] Below a certain frequency, the noise is approximately 1/f in character. The thermal noise - itself due to the resistance of the channel - dominates with its flat spectrum at the higher frequencies. There is a third region at very high frequencies but this is not of interest in the present application. In general the spectrum has the form shown in Figure (2-8) and so the equivalent noise resistance may be described approximately by

$$Rn = R_{nT} + \frac{K}{f} \quad , \qquad (2\text{-}12)$$

where R_{nT} = the equivalent thermal noise resistance,

f = the frequency,

K = a constant, typically[7] $10^8\ \Omega\ H_z^{-1}$.

3(b)(i) Low Frequency Noise in MOSFETs

A large number of papers have been published on the subject, all of

which attempt to explain the 1/f dependence found over a very wide frequency range. These have been reviewed by Ronen.[8] It is widely agreed that this noise will be:-

(i) inversely proportional to the channel area;

(ii) proportional to the surface state density, Nss; and

(iii) for a MOST in saturation, only weakly dependent on drain voltage but increasing with effective gate voltage.

It is difficult to estimate the frequency range over which 1/f noise will dominate in the transistors T1, T3 and T5, the main contributors to the system amplifier noise, precisely because little is known about the surface state density, Nss. The dimensions of these transistors are known from the manufacturer's data and are given in Table 2.1.

TABLE 2.1

Transistor	T1	T3	T5
Channel length μm	12	8	8
Channel width μm	8	8	8
Area μm^2	96	64	192

Since these are much smaller than transistors which are discussed in the literature, comparatively poor 1/f noise performance might therefore be expected. A theoretical relation for the 1/f noise is developed by Christensson et al.[9] (Their example calculation is done for a device of channel area 4000 μm^2.) Using their relation

$$\frac{V_n^2}{B} = N_t \frac{4 e^2}{Co^2 A} \quad 0.25 \frac{\pi}{2\alpha\omega}$$

the low frequency noise due to T3 may be estimated where V_n^2 = the mean square noise voltage,

B = the bandwidth,

Nt = the number of traps within a few KT of the Fermi level which may contribute to the effect, 3×10^{17} cm^{-3} at 200 K,

Co = the gate capacitance per unit area, 3×10^{-8} F cm^{-2}

$\alpha = 2 \times 10^{8}$ cm^{-1}, and

A = the channel area, 64 μm^2 for T3.

The remaining symbols have their usual meaning. The value of N_T is calculated using a typical value of N_{ss}, 10^{12} cm^{-3} eV^{-1}, but, as has been said, this is not known with any certainty for the Plessey device and may be less. The following is therefore given:

$$\frac{V_n^2}{B} = \frac{1 \times 10^{-7}}{W} \quad V^2 \, H_z^{-1} \quad .$$

Integrating over a typical bandwidth from 500 KHz to 5 KHz and expressing the result as an equivalent noise charge on Co, we have

$$NEC = \frac{V_{nCo}}{e} = Co \left[\frac{10^{-7}}{2} \int_{f1}^{f2} \frac{1}{f} \, df \right]^{\frac{1}{2}} \simeq 170 \text{ electrons.}$$

3(b)(ii) Thermal Noise

The thermal contribution to MOSFET noise has been found[10] to have the form

$$R_{nT} = \frac{\alpha}{gM} \quad , \qquad (2\text{-}13)$$

where R_{nT} = the equivalent thermal noise resistance,

α = a nearly constant term which depends on oxide thickness and substrate doping, and

gm = the device transconductance.

As $\alpha \sim 1$, the importance of thermal noise may be estimated by calculating gm for the transistors T1, T3 and T5. For a MOSFET in saturation

$$I_D = \frac{\beta}{2} (V_G - V_T - V_S)^2 \quad , \text{ and}$$

$$gm = \frac{I_D}{V_G} = (V_G - V_T - V_S) = (2 I_D \beta)^{\frac{1}{2}}.$$

Thus in transistors T1 and T3

$$Rn \propto \frac{1}{VE} \propto \frac{1}{I^{\frac{1}{2}}} \quad , \quad (VE = (V_G - V_T - V_S))$$

so that the thermal noise contributions of these might be expected to vary with existing circuit conditions. Again, we may calculate the noise

of this type generated by T3. For typical values

$$gm_3 = (2\ I_D \beta)^{\frac{1}{2}} \simeq 3 \times 10^{-5}\ \Omega^{-1}\ ,$$

where $I_D \simeq 10^{-4}$A, a typical value corresponding to an output of 5V from the usual external amplifier, and

$$= 6\ \mu A V^{-2}\ ,$$

and the equivalent thermal input noise is

$$NEC = \frac{Co}{e}\left\{4KTB\frac{\alpha}{gm_3}\right\}^{\frac{1}{2}} \simeq 8 \text{ electrons}$$

using $Co = 10^{-13}$ pF,

e = the electronic charge,

K = Boltzmann's constant,

B = the bandwidth, 500 KHz, and

$gm_3 = 3 \times 10^{-5}\ \Omega^{-1}$.

3(c) Noise: Equivalent Circuit

To estimate which transistors make the most significant contributions to the total system amplifier noise, it is necessary to refer all the noises to the same point in the circuit. It is easiest to use the gate capacitance of T3 since here the total noise may readily be evaluated in terms of electrons.

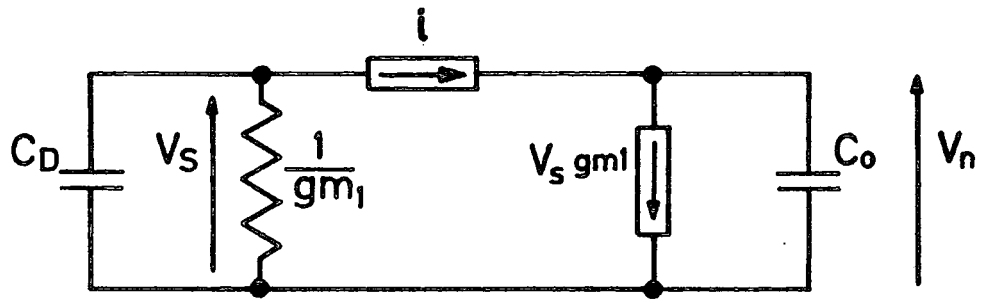
From the equivalent circuit shown in Figure (2.9(a)) it follows that the output noise current of T1 is given by

$$\overline{in^2} = \frac{4KTBRn\ gm^2}{|1 + zgm|^2} \qquad \text{where } z = \frac{1}{j\omega C_D}$$

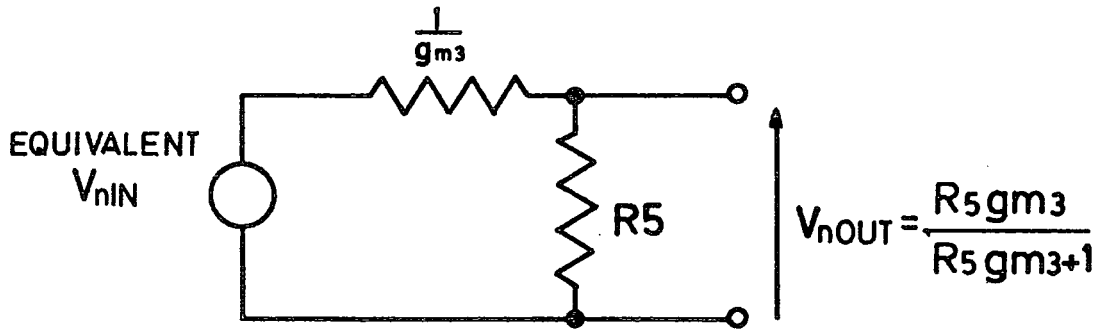
which appears as a voltage Vn across Co such that

$$\overline{Vn^2} = \frac{C_D^2}{Co^2}\left(\frac{4KTBRn}{1 + \frac{\omega^2}{gm^2}C_D^2}\right) .$$

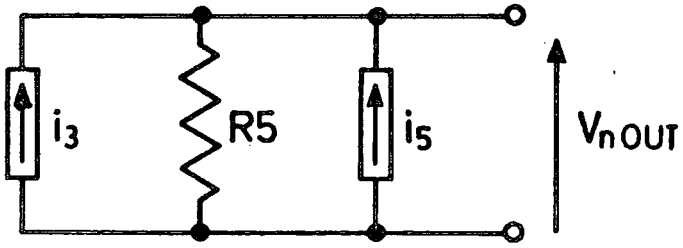
The above represents a low pass filter with a half power frequency,



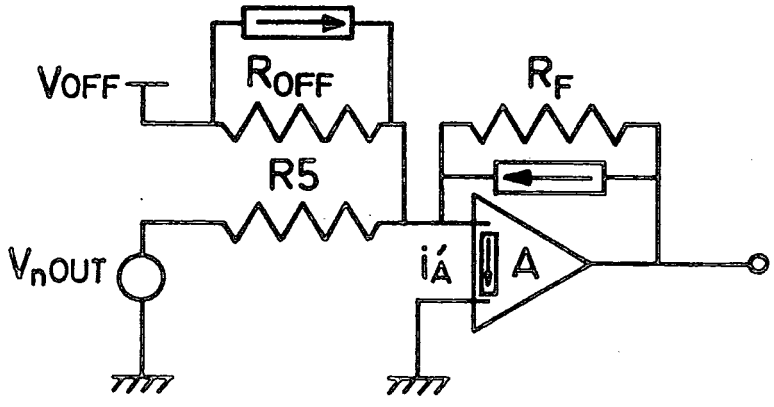
(a) EQUIVALENT CIRCUIT OF T1



(b) SOURCE FOLLOWER EQUIVALENT CIRCUIT



(c) NOISES IN SOURCE FOLLOWER CIRCUIT



(d) NOISES IN EXTERNAL AMPLIFIER CIRCUIT

FIGURE 2.9

$$\omega_o = \frac{gm}{C_D} \quad ,$$

lying below the bandwidth of interest, 5KHz to 500 KHz, for all practical values of gm. In this bandwidth, which is determined by the scanning frequency and the external amplifier bandwidth, the noise voltage is given by

$$\overline{V_n^2} = \frac{4KTBR_n\, gm^2}{\omega^2\, C_o^2} \quad .$$

From the equivalent circuit of Figure (2.9(b)), the effective combined noise of T3 and T5 may be referred to Co. The gain of this circuit is

$$\frac{V_o}{V_1} = \frac{R_5\, gm_3}{R_5\, gm_3 + 1} = G \quad ,$$

where R_5 = the drain resistance of T5, and

gm_3 = the transconductance of T3.

The voltages developed across R_5 by the output noise currents of T3 and T5 may be converted to an equivalent input noise voltage at the gate of T3:

$$\overline{V_{nin}^2} = \frac{(i_3^2 + i_5^2)\, R_5^2}{G^2}$$

$$= \frac{(i_3^2 + i_5^2)\,(R_5\, gm_3 + 1)^2}{gm_3^2}$$

as $gm_3\, R_5 \approx 0.15$, which is small, then the noise voltage is given approximately by

$$\overline{V_{nin}^2} \quad \frac{(i_3^2 + i_5^2)}{gm_3^2} \quad .$$

Similarly, noise generated in the external amplifier circuit may be referred back to Co, as shown in Figure (2.9(c)). If the amplifier has an equivalent input noise current

$$\overline{i_A^2} = i_A'^2 + \frac{4KTB}{R_f} + \frac{4KTB}{R_o} \quad ,$$

where i_A' = the inherent equivalent input noise current of the Operational amplifier used which should be negligible in the bandwith of interest,

R_f = the feedback resistance, usually 47KΩ , and

R_o = the equivalent resistance of any offset circuit.

Combining the terms due to T1, T3, T5 and the external amplifier, the total equivalent noise may be estimated:

$$\text{Noise}_{AMP} = \frac{Co}{e}\left\{4KTB\left(\frac{R_{n1}\, gm_1^2}{w^2\, Co^2} + R_{n3} + R_{n5}\,\frac{gm_5^2}{gm_3^2} + \frac{1}{Rf\, gm_3^2} + \frac{1}{rogm_3^2}\right) + \frac{\dot{i}_A^2}{gm_3^2}\right\}^{\frac{1}{2}}, \quad (2\text{-}14)$$

where R_{n1} = the equivalent noise resistance, thermal and low frequency, of T1,

gm_1 = the transconductance of T1, etc.

Since the magnitude of the low frequency, 1/f type noise is not well known it is best to estimate this contribution independently of thermal noise. In the bandwidth of interest only the 1/f noise from T3 and T5 is significant because noise from T1 is attenuated by the action of Co. Taking into account (a) that $gm_5 \leq gm_3$ and (b) that the channel area of T3 is a third of that of T5, the predominant source of 1/f noise is therefore T3, for which an estimate has already been made using an assumed value of surface state density.

Considering now the thermal noise contributions of the various circuit components, by substituting typical values it is found that all three on-chip transistors are of similar importance - that is, having equivalent noises of a few tens of electrons over the assumed bandwidth (5KHz to 500 KHz). Thermal noise from external circuit resistances may be calculated fairly exactly - being again of the order of a few tens of electrons for typical values of R_f and Ro (eg 47 KΩ).

Altogether it may be said that thermal noise should be small compared to reset noise but no definite conclusion may be made about 1/f noise without further information. Any amplifier noise occurring in processing

stages after the virtual earth amplifier may be referred back to Co and also compared to the signal in a similar manner. It must be remembered, however, that the output characteristic is not perfectly linear and comparison can only be made for a particular value of gm_3. Of interest is the fact that the mean square noise due to the feedback resistor, Rf is inversely proportional to its value. Thus doubling the gain of the system by doubling Rf in fact reduces the noise of this component with respect to the signal, whereas doubling the gain by adding a further amplification stage does not improve this ratio but, on the contrary, may add extra noise sources to the system. Ideally as much amplification as possible should be achieved by the virtual earth amplifier by which is meant Rf should be made as large as possible without driving the amplifier into saturation. In any case, Rf should be a high stability, low noise component as should the resistors of any offset networks which are connected to the summing junction.

Amplifier noise is distinct from other noise sources in that it may be reduced by restricting the bandwidth of the signal processing system. The bandwith quoted above, 5KHz to 500KHz, is roughly that used in the initial development system. As this is far from optimal, significant improvements might be made where thought necessary; such an occasion might be if amplifier low frequency noise were found to be as great as reset noise.

Sources of noise arising from processes not occurring during readout are: reset noise, an upper limit to which has already been estimated; thermal shot noise, which should be negligible if the device is cooled sufficiently; and, possibly, statistical noise due to charge pumping in T2 at the time of recharge.

3(d) Switch Capacitance Effects and Charge Pumping Noise

From simple MOST theory, the amount of charge forming the channel of the recharge transistor in unsaturated operation will be given approximately by

$$Q_{CH} \simeq V_E C_O' A$$

where $V_E = V_{CP} - V_T - V_G$,

C_O' = the capacitance per unit area,

A = the channel area of T2, approx. $100\,\mu m^2$.

V_{CP} = the recharge pulse on-voltage at the gate of T2, and

V_G = the voltage on the gate of T3.

When a recharge pulse is applied to the gate of T2, the carriers necessary to form the conducting channel are drawn in from the diffusions as indicated in Figure (2.10(a)). This being a p-channel device the carriers are holes, and so setting up the channel results in a positive charge being taken from the diffusions of T2. As soon as a conducting channel is formed, the transistor will act like a resistor and if V_G (the gate voltage of T3) is positive with respect to V_{DD}, Co will be recharged until $V_G = V_{DD}$. This is shown in Figure (2.10(b)). On the falling edge of the recharge pulse, when T2 turns off, the charge forming the conducting channel must now be returned to the diffusions as the inversion layer collapses. If any charge is returned to the source diffusion which is connected to the low capacitance environment of Co, it will produce a positive voltage charge and once again a potential difference will exist between source and drain causing a net current to flow restoring equilibrium. Thus, as the gate voltage of T2 starts to fall, the channel charge is initially all returned to the drain diffusion which is connected to the V_{DD} supply. However, when V_{CP} falls below the value necessary for unsaturated operation:

$$/V_{CP}/ < /V_{DD} + V_T/ ,$$

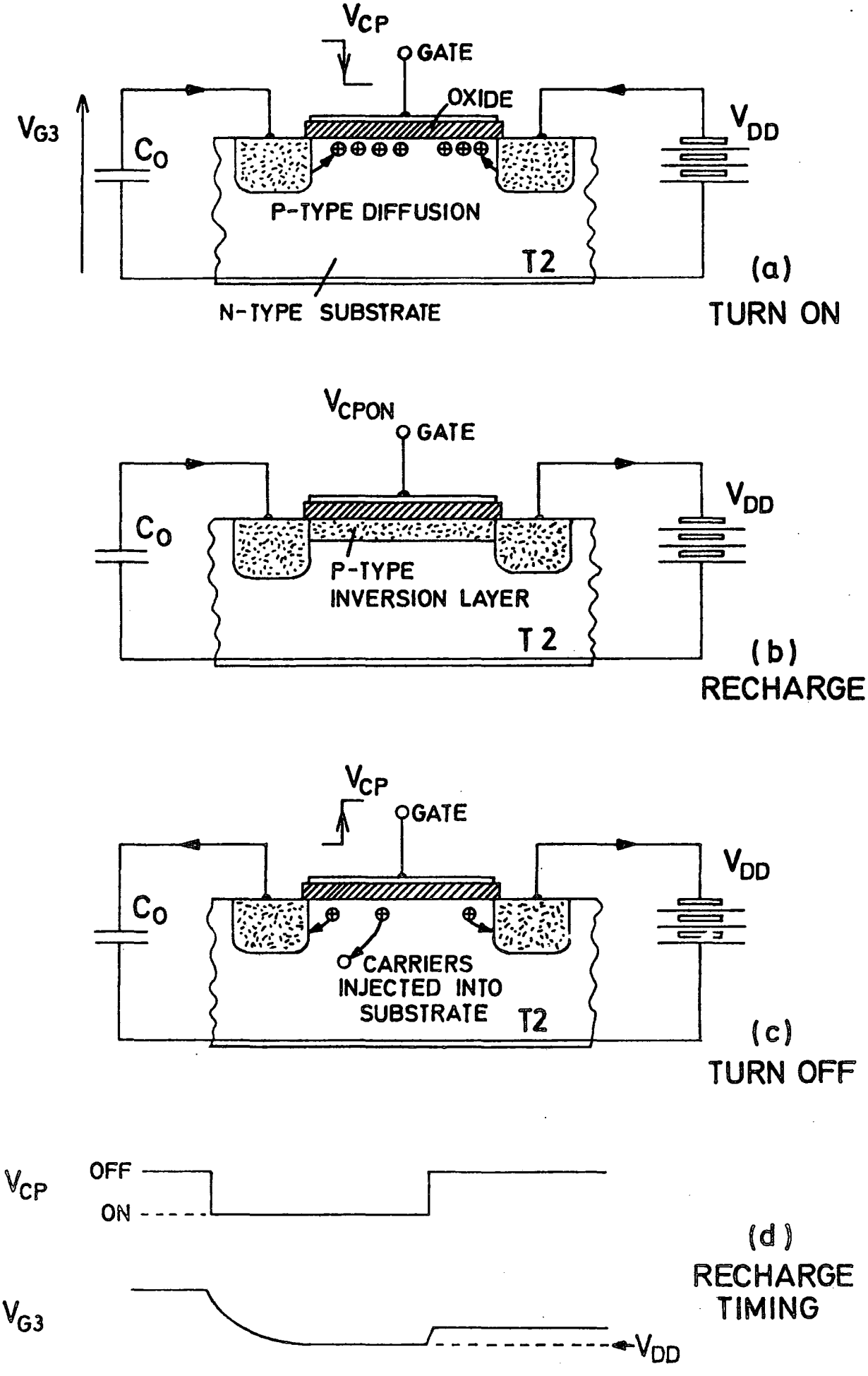


FIGURE 2.10 RECHARGE TRANSISTOR

the channel will pinch off at the ends and conduction will no longer be ohmic. If V_{CP} falls so quickly that some carriers remain in the channel region after the above condition has been met, they will now be returned to both diffusions. Those carriers placed on Co produce a positive offset on the recharge voltage such that

$$V_G = V_{DD} + Q_{OFF}\ Co,$$

where V_G = the gate voltage of T3,

V_{DD} = the chip supply voltage,

Q_{OFF} = the channel charge returned to Co, and

Co = the capacitance associated with the gate of T3.

This effect is clearly seen (if the video output is examined on an oscilloscope during a recharge operation) as a positive signal step occurring at the time of the falling edge of the recharge pulse. Expressed as a voltage on Co this offset is typically < 0.1 V. There are interelement variations in this effect which constitute a further source of fixed pattern offset variation.

It has been found experimentally by various workers[11] that, during the collapse of an inversion layer in a MOST, not all the carriers are able to find their way back to the diffusions. Some recombine in the channel region. The charge injected in this manner causes a small current pulse to flow in the substrate circuit. The total charge returned to both junctions is therefore less than that originally taken from them, implying that, in any on-off switching cycle, a net charge is transferred from the diffusions to the substrate in a direction opposite to the applied bias. This is the effect known as charge pumping.

In the case of T2, the simple situation described above is complicated by an initial voltage difference between source and drain. However, it may be said that this loss of carriers by injection will result in less charge being returned to the source diffusion (connected to the gate of T3)

than would otherwise be the case. The significance of this is that any statistical uncertainty in the number of carriers injected will appear as noise on the value of V_G, placed on Co, following a recharge operation. This noise may be included with the thermodynamic reset noise previously described.

The charge pumped in one cycle is given by Brugler and Jespers as

$$Q_p = A(V_E C_o' \alpha + e N_{ST}), \qquad (2\text{-}15)$$

where A = channel area,

V_E = effective gate voltage,

α = chance of recombination in the bulk,

C_o' = gate capacitance per unit area, and

N_{ST} = the number of fast surface states available to contribute to the effect.

This may be rewritten as

$$Q_p = Q_{CH} + Q_{INT} ,$$

where Q_{CH} = the charge forming the channel at switch off, and

Q_{INT} = the amount of charge trapped at the Si - SiO_2 interface.

The first of these two terms may be subject to some statistical variation. The chance of any given carrier recombining as the gate voltage is reduced depends upon:(a) how quickly this is done; and (b) the distance the carrier must travel to reach a diffusion. For transistors with shorter channel lengths, it has been found that by using gate pulses with slow falling edges the amount of charge lost in the bulk can be made negligible. In practice the transistor T2 is operated with

$$/V_{CP\ OFF}/ > /V_T/ .$$

This still makes a good switch as long as

$$/V_{CP\ OFF} - V_T/ < /V_G/ ,$$

where $V_{CP\ OFF}$ = the off voltage on the gate of T2,

V_T = the effective threshold voltage of T2, and

V_G = the gate voltage of T3.

The above corresponds to the carriers on the diffusions of T2 being separated by a potential barrier. Although the depletion conditions exist in the bulk for the formation of a conducting channel, any carriers are drawn back to the diffusions so that, under steady state conditions, there is no net current flow even if a voltage difference exists between source and drain. In the case where $/V_{CP} - V_T/$ has just changed from a value large enough for a conducting channel to exist to a value just above zero the channel may still contain some carriers. Keeping

$$/V_{CP} - V_T / \; > \; 0$$

probably reduces the charge pumping effect by allowing extra time for the carriers to return to the diffusions. (It may be noted that if V_G becomes small enough during an integration, T2 begins to turn on and stops any further fall in V_G. This is useful because it prevents an element from becoming completely discharged from which state it may take an appreciable time to recover).

The probable magnitude of the uncertainty in the reset voltage which is introduced by this effect is not known, but an estimate may be made if it is assumed that every carrier has a chance γ of being injected. Applying binomial statistics

$$\frac{Q_{OFFSET}}{e} \simeq \frac{Q_0 (1-\gamma)}{e} ,$$

where Q_{OFFSET} = the amount of charge returned to Co,

Q_0 = the amount which would be returned if none were injected, and

γ = the chance of injection.

The error in this will be

$$\text{Noise}_{CP} \simeq (Q_0 (1-\gamma) \gamma)^{\frac{1}{2}}$$

$$= \{Q_{OFFSET} \gamma\}^{\frac{1}{2}}.$$

From observation $Q_{OFFSET} \simeq 10^5$ electrons, typically, which gives,

$$\text{Noise}_{CP} \quad 300(\gamma)^{\frac{1}{2}} \text{electrons.}$$

This should be an overestimate as the part of the injected charge due to trapping at the interface is not apparently subject to the above statistics. In any case, if γ is small this noise source should be small compared to reset noise.

3(e) Noise Summary

The total readout noise in any measurement arrived at by double sampling is given by

$$\text{Noise}^2 = 2x \left\{ \text{Noise}^2_{\text{Reset}} + 2 \int_{f2}^{f1} \text{Noise}^2_{\text{AMPI}} \, df + \int_{f2}^{f1} \text{Noise}^2_{\text{AMPE}} \, df + \text{Noise}^2_{CP} \right\} \qquad (2\text{-}16)$$

where Noise RESET = thermodynamic noise introduced by one reset,

Noise AMPI = internal amplifier noise (multiplied by two because two elements are addressed simultaneously),

Noise AMPE = external amplifier noise, and

Noise CP = charge pumping noise.

The above expression cannot be evaluated very exactly because: (i) Co is not well known; (ii) little is known about the importance of 1/f type noise in this device; and (iii) charge pumping noise is not well understood. However, it is expected that the total readout noise will lie in the range 150 to 300 electrons.

4(a) Driving Electronics

Typical timings of the control pulses employed by the present driving system illustrated in Figure (2.11) are given in Figure (2.12). As noted elsewhere, these arrays were originally supplied to Durham University (together with a driving system developed by both the manufacturers and the R.G.O.) for application in electron counting mode. This early system

Fig 2-11 ARRAY DRIVING ELECTRONICS

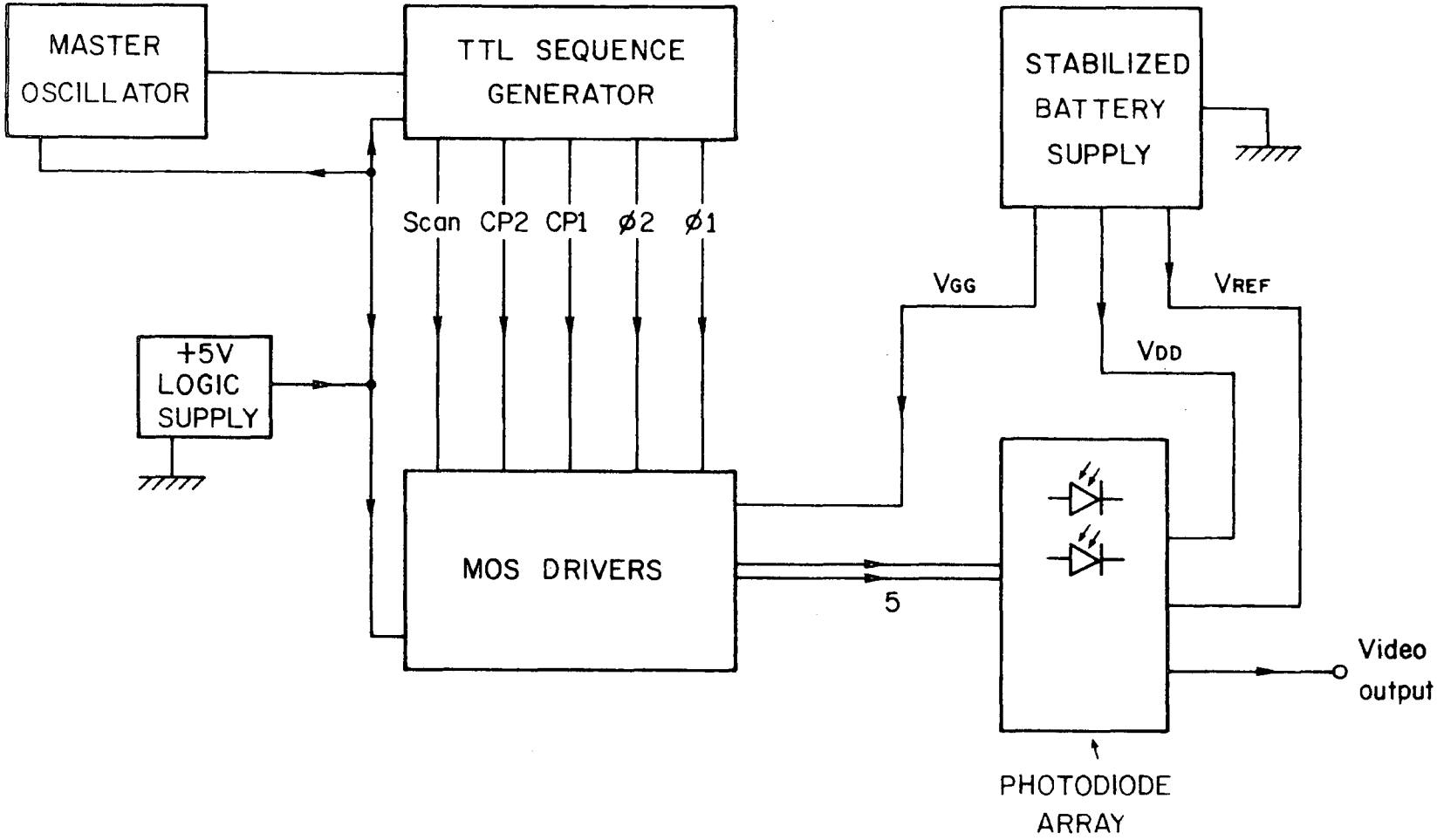
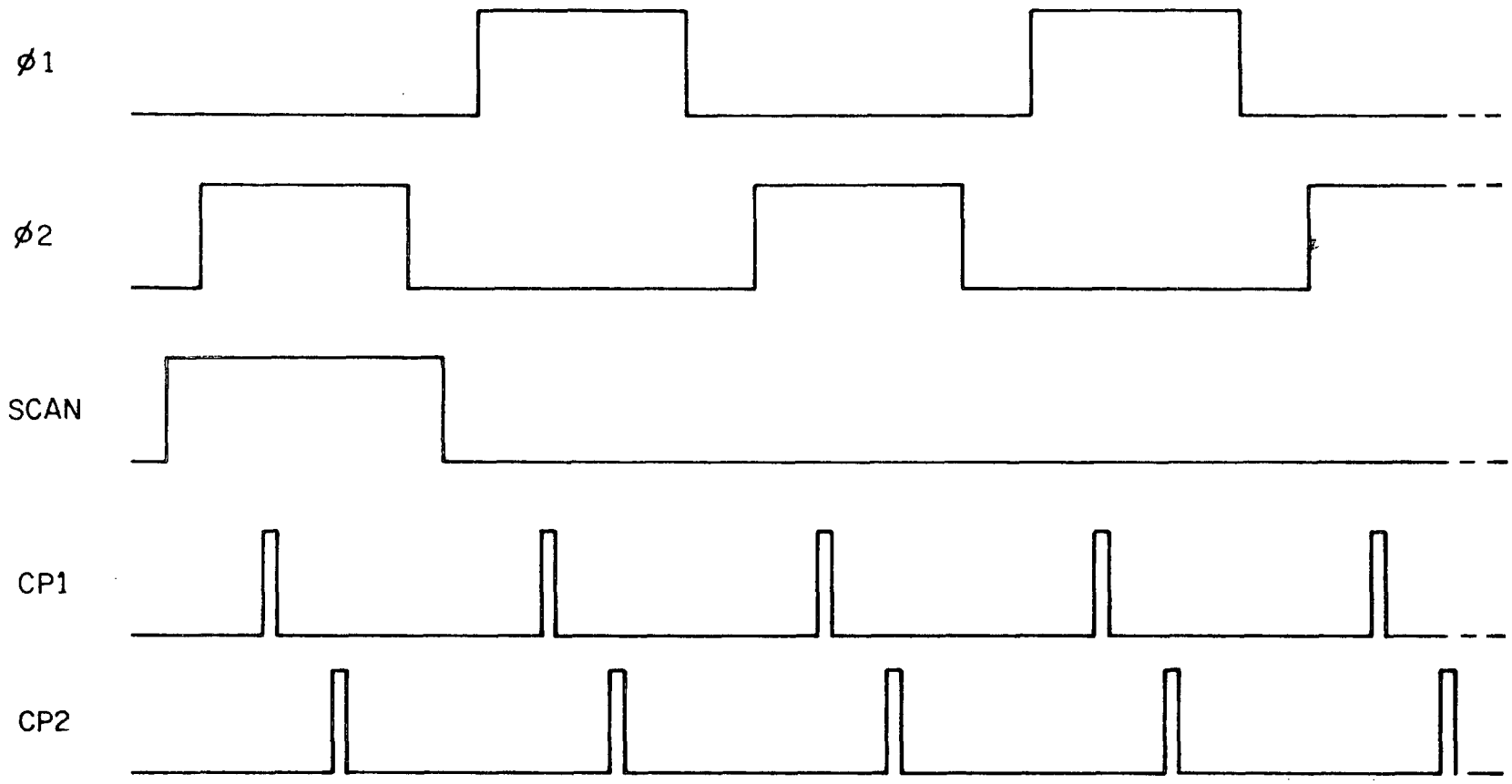


Fig 2·12 ARRAY DRIVE PULSES



generating full period clock pulses has been superseded by systems leaving various amounts of dead time between the two phases of the shift register clock waveform precisely because such full clock pulses were found to cause too much instability for present purposes. However, the original TTL to MOS level converters have been retained despite the fact that alternatives have been tried in laboratory work. All supply voltages which relate directly to device operation in this newer system are obtained from regulated battery units. Other features of the system include a large number of variable parameters and, in particular, those of the scanning frequency and amplitude of the clock pulses - all in need of setting by the individual operator. It might be noted that the optimum values of both parameters mentioned appear to be slightly interdependent and also related to the operating temperature.

4(b) TTL Sequence Generator

It was decided to build a sequence generator producing all the necessary control pulses with the required timing from just one original master clock frequency. Thus the ratios of the lengths of the pulses are hardwired but the overall data rate may be varied by adjusting one frequency. Although perhaps the most flexible approach for initial experimental work, when a particular operating frequency is decided upon, it can nevertheless be very stable - that is, if the master clock is a crystal controlled oscillator. In practice, the recharge pulse widths are set by a TTL monostable but these are not critical so long as they are greater than an experimentally determined minimum. A schema of a typical sequence generator is shown in Figure (2-13) with an internal timing diagram in Figure (2-14).

Experimental set-ups were tried using several pulse generators in delay and one shot mode, all triggered by a master oscillator, to produce

Fig 2·13 SEQUENCE GENERATOR

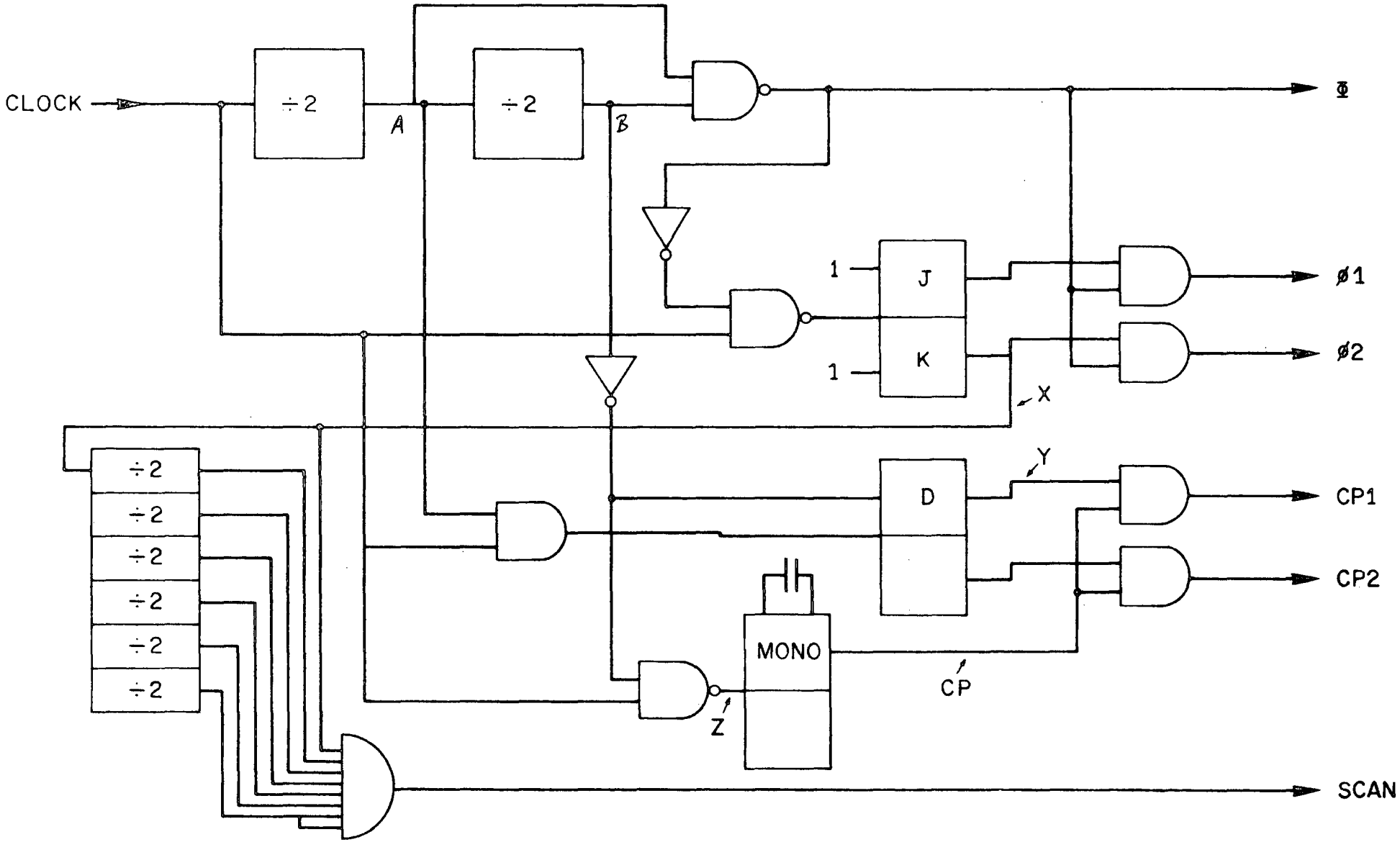
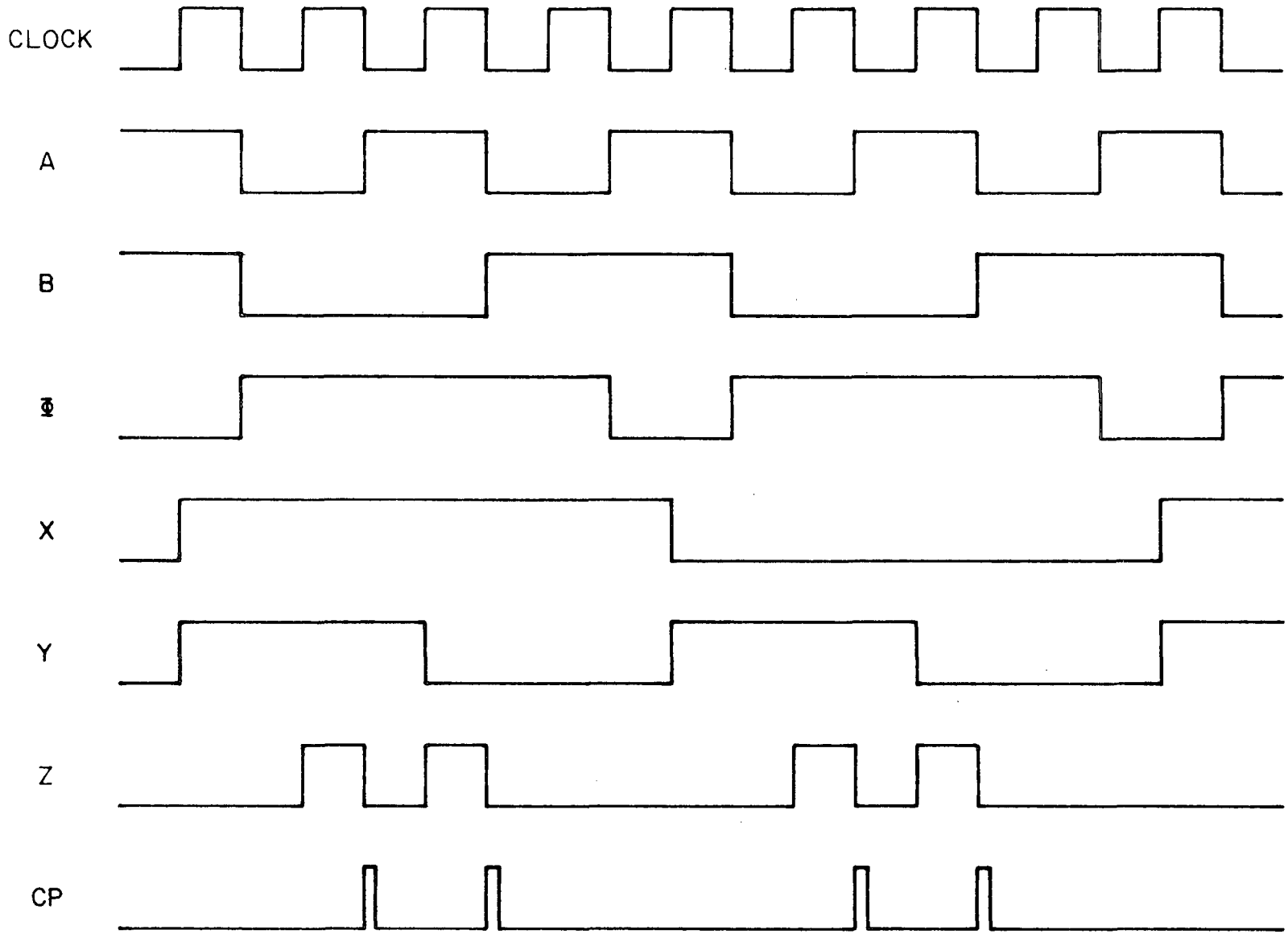


Fig 2·14 INTERNAL TIMING OF SEQUENCE GENERATOR



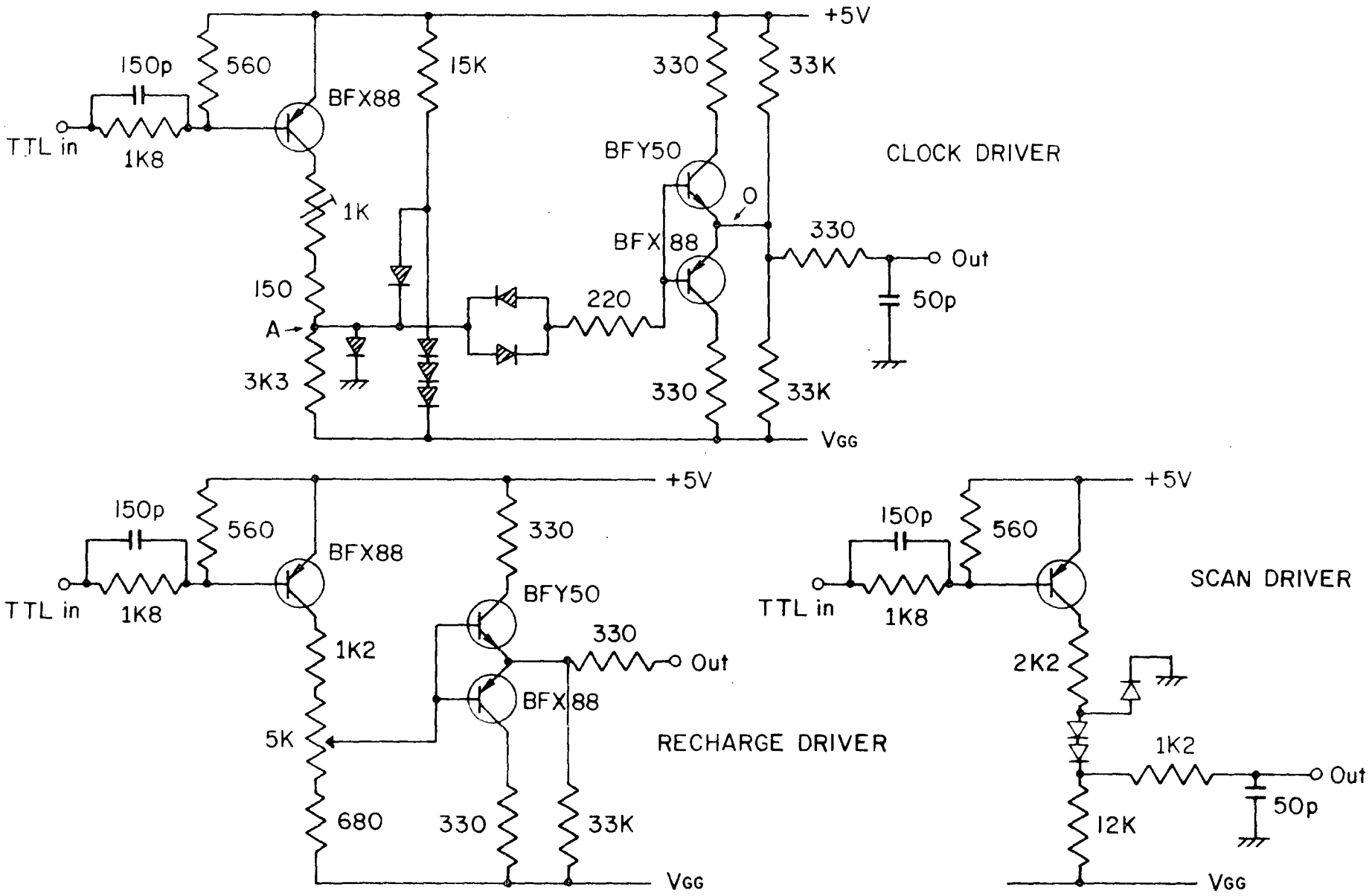
the clock pulses and investigate the operation of the on-chip shift register. It was found that the array operated over a wide range of frequencies and with clock pulses covering any fraction of a period except when the rising and falling edges of the two phases were sufficiently close to each other that they began to overlap because of switching speed limitations. If the latter is the case, the video output shows a wobble which is thought to be due to the output from one pair being enabled before the previous pair fully switches off.

As the final objective was to digitize the array output, it was thought simplest if the signal levels were of equal length and came at a uniform rate. Hence the present system with 25 per cent dead time and three equal video times was adopted. The information from the dead time may provide a useful zero current reference to check against slow electronic drift or, on the other hand, it may be suppressed altogether. The scan pulse must overlap the ϕ 2 clock phase for proper operation of the shift register. To keep the rising and falling edges of the scan pulse in the dead time so that they do not produce spikes on the video output, this time the pulse is produced by dividing a symmetrical pulse of the same phase and frequency as ϕ2. In practice the video output at the time of scan is usually anomalous, but this can only be seen when the array is scanned in consecutive frame times.

4(c) TTL to MOS Level Converters

The original circuits of Figure (2-15) are used to convert the five TTL pulses produced by the sequence generator - namely, ϕ 1, ϕ 2, CP1, CP2, and scan, to suitable MOS levels. The converter circuits are all powered from the same two rails, a variable negative supply, V_{GG} and a +5 volt logic supply. Adjusting V_{GG} varies the heights of all these pulses. For evaluation purposes, the various drivers were all connected to individual negative supplies and adjusted independently. These results

Fig 2·15 MOS DRIVER CIRCUITS



are discussed in greater detail in the next chapter. However, in brief, the amplitudes of the Scan and Charge pulses, as the results show, are not found to be critical provided that they are greater than certain minimum values.

4(c) (i) Clock Pulse Drivers

The function of the clock pulse drivers is to produce an output that switches between an '0' state of zero volts and a '1' state which is an accurately preset negative voltage. This must be done with the same timing as the TTL input from the sequence generator. Any ripple or uncertainty in the amplitude of these pulses may show up in the video output and, consequently, leads to the introduction of undesirable offsets and noise. It is, therefore, crucial that these pulses are fairly clean and flat topped, although absolute speed of the switching edges is not necessarily required unless the array is to be scanned very quickly. Equally, it is useful if these circuits are capable of driving reasonable lengths of cable since it may not always be convenient to place the electronics close to the detector.

Referring to Figure (2-15), a TTL logical '1', applied to the input, switches off transistor T1. The point in the circuit marked 'A' now sits at a voltage determined by the clamping; the voltage will be $V_{GG} + 2V_F$ where V_F is the forward voltage drop of one diode. If the point in the circuit marked '0' is positive with respect to 'A', a current will flow in the base-emitter junction of T2 switching this transistor on and causing the voltage at '0' to swing negative until the condition

$$V_0 = V_{GG} + 3V_F + V_{be}$$

is reached. Thus the negative excursion of the output is produced by the emitter follower stage formed by T2 and R9. When subsequently the input sees a TTL'0', T1 switches on and the point'A'moves positive in voltage

thereby switching off T2 and switching on T3. T3 and R10 form a second emitter follower which handles the positive going edge of the output. Some fine adjustment of the zero level is possible through VR1.

4(c) (ii) Charge Pulse Drivers

Although similar to the above in terms of operation, the charge pulse drivers have a much larger range of adjustment of the off state voltage which may be varied from zero to almost V_{GG}. Another major difference with this circuit lies in the absence of an emitter resistor for T2 - leaving the negative going edge of the output very much slower than is the case with the clock pulse drivers. (positive going in voltage)

4(c) (iii) Scan Pulse Driver

Here, any pulse amplitude > / - 4V/ has been found to work satisfactorily since the scan pulse and this circuit is not very critical.

4(d) Power Supplies

The three stabilized voltages necessary to run the array, V_{GG}, V_{DD} and V_{REF} are all supplied by the one battery unit outlined in Figure (2-16). Each may be independently set within a range of values:

(1) V_{REF} controlling the diode reverse bias voltage is typically - 3 volts;

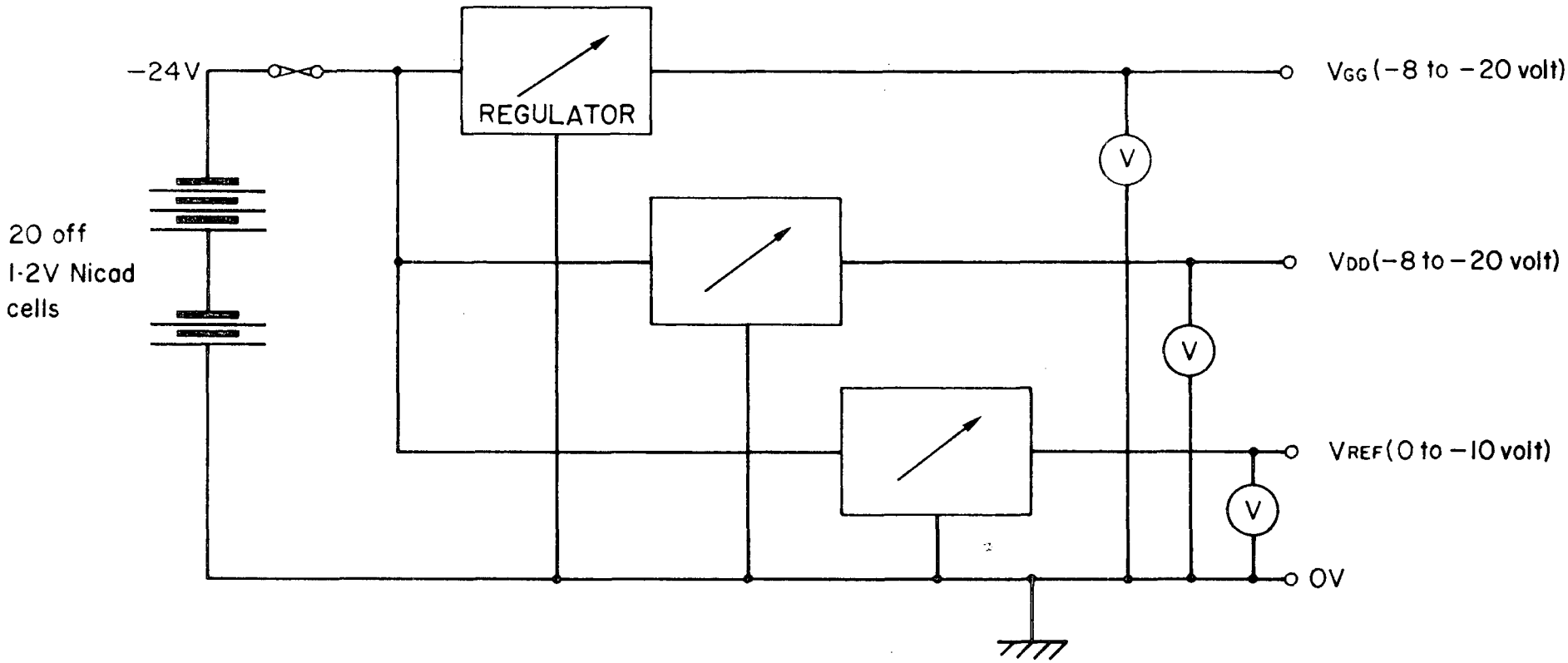
(2) the chip supply voltage, V_{DD} is generally set to about - 9 volts; and

(3) V_{GG} supplying the drivers is usually -14 volts.

A +5 volt rail required by the MOS drivers is usually taken from a mains powered logic supply. Slight variations in this voltage do not directly affect the 'on' values of the MOS driver pulses so that the use of batteries has not been deemed necessary.

In addition to the above power sources, a $\pm$ 15 volt battery powers both the external amplifier and subsequent signal processing stages.

Fig 2·16 MOS SUPPLY VOLTAGES



REFERENCES

1. McMullan, D., Wellgate, G.R., Ormerod, J. and Dickson, J. Adv. in Electronics and Electron Physics. Vol.33, pp.873-879.

2. Richman, P. (1973). MOS Field - Effect Transistors and Integrated Circuits. (Wiley), p.125.

3. Ibid. p.78.

4. Hedge, A.R. Ph.D. Thesis, University of Durham (1981).

5. Van Der Ziel, A. (1976). Noise in Measurements. (Wiley), p.185.

6. Soares, R.A. (1971). "Noise in M.O.S. transistors." Design Electronics. Vol.8, No.7-8, pp.25-30.

7. Van Der Ziel, A. Op.cit. p.53.

8. Ronen, R.S. (1973). "Low-Frequency 1/f Noise in MOSFETs." RCA Review. Vol.34, pp.280-307.

9. Christensson, S., Lundstrom, I. and Svensson, C. (1968). "Low Frequency Noise in MOS Transistors - I." Solid-State Electronics. Vol.11, pp.797-812.

10. Klaasen, F.M. and Prins, J. (1967). "Thermal Noise of MOS Transistors." Philips Res. Repts. Vol.22, pp.505-514.

11. Brugler, J.S. and Jespers, P.G.A. (1969). "Charge Pumping in MOS Devices." IEEE Transactions on Electron Devices. Vol. ED-16, No.3, pp.297-302.

12. and Leunberger, F. (1971). "Charge Pumping and Low-Frequency Noise in MOS Structures." Phys. Stat. Sol. Vol.8, pp.545-550.

CHAPTER 3

LABORATORY EVALUATION WITH ATTENDANT PRACTICAL PROBLEMS

1. Introduction

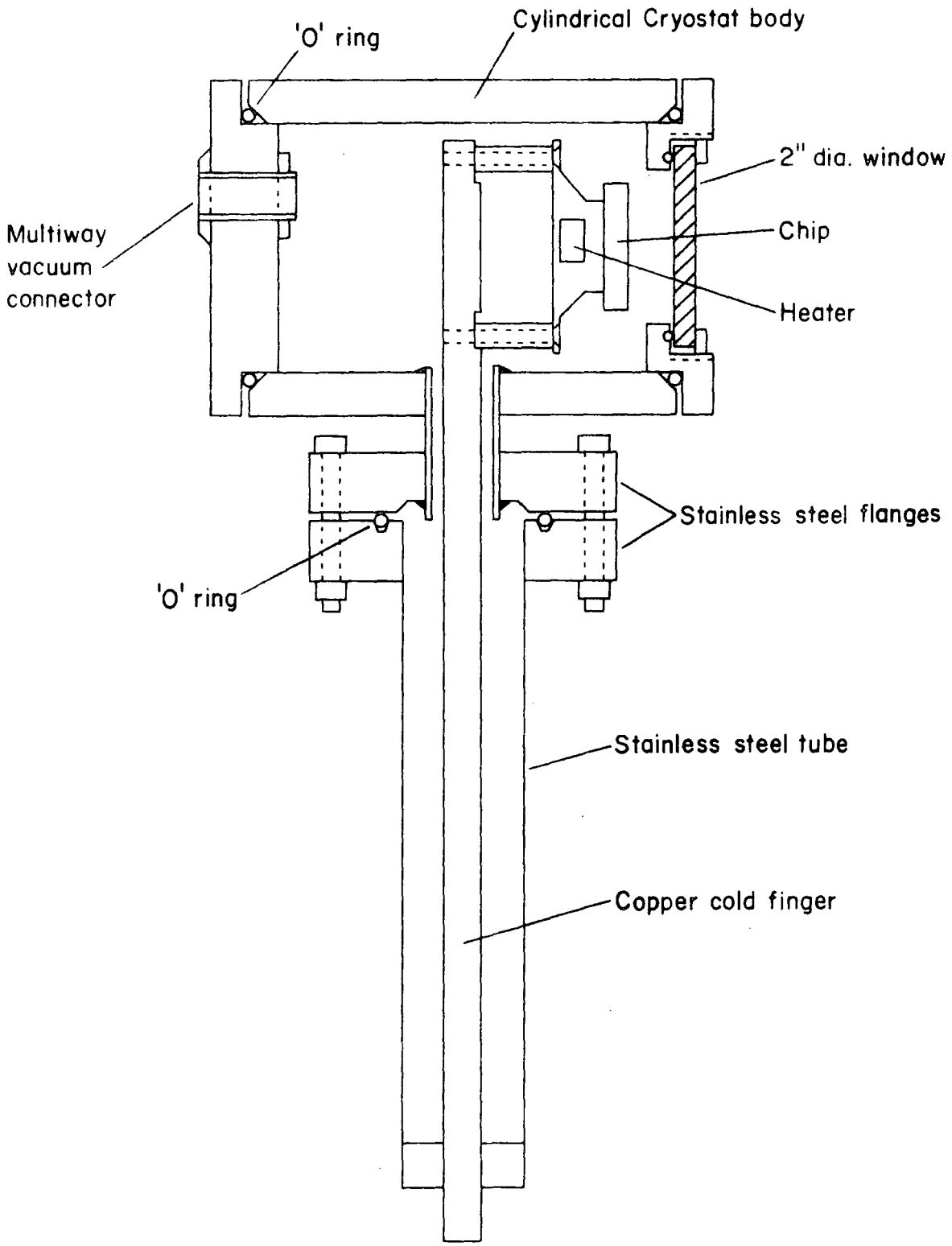
Before undertaking the construction of a complete experimental system for spectroscopic work, a number of laboratory tests were made using simple apparatus to determine how the Plessey array performed and to find approximate values for the photoresponse, quantum efficiency and readout noise. In this chapter, these tests and some of the problems encountered in practical operation will be outlined.

2(a) Cooling Systems

Some initial evaluation work was attempted in measuring the performance of the diode array at room temperature. Large interelement variations in dark current were observed and individual responses were not found to be as linear as anticipated - that is, even allowing for the non-ideal characteristics of the source follower. This was thought to be due to significant thermal leakage at the p - n junctions formed by the diffusions of the two transistors, T1 and T2. The latter are connected to the output capacitance, Co whose bias changes throughout the integration. Measurements of dark current as a function of V_{REF} indicate, by extrapolation, that the dark current at zero depletion layer width is non-negligible. Although the total area of these two diffusions i.e. $\sim$ 100 μm^2 is small compared to the area of the photodiode, $\sim$ 8000 μm^2 where the dominant thermal contribution is from the periphery of the junctions,[1] (i.e. where the depletion region meets the Si - SiO_2 interface), the ratio of the perimeters must be considered. Assuming that the peripheral dark current scales as $V^{\frac{1}{2}}$ and that the junctions have similar doping concentrations, we have for typical values

$$\frac{IDT}{IDD} \simeq \left\{\frac{V_G}{V_D}\right\}^{\frac{1}{2}} \frac{PT}{P_D} = \left\{\frac{10v}{0.5v}\right\}^{\frac{1}{2}} \frac{60\,\mu m}{500\,\mu m} \quad , \qquad (3\text{-}1)$$

Fig 3.1 COLD FINGER CRYOSTAT



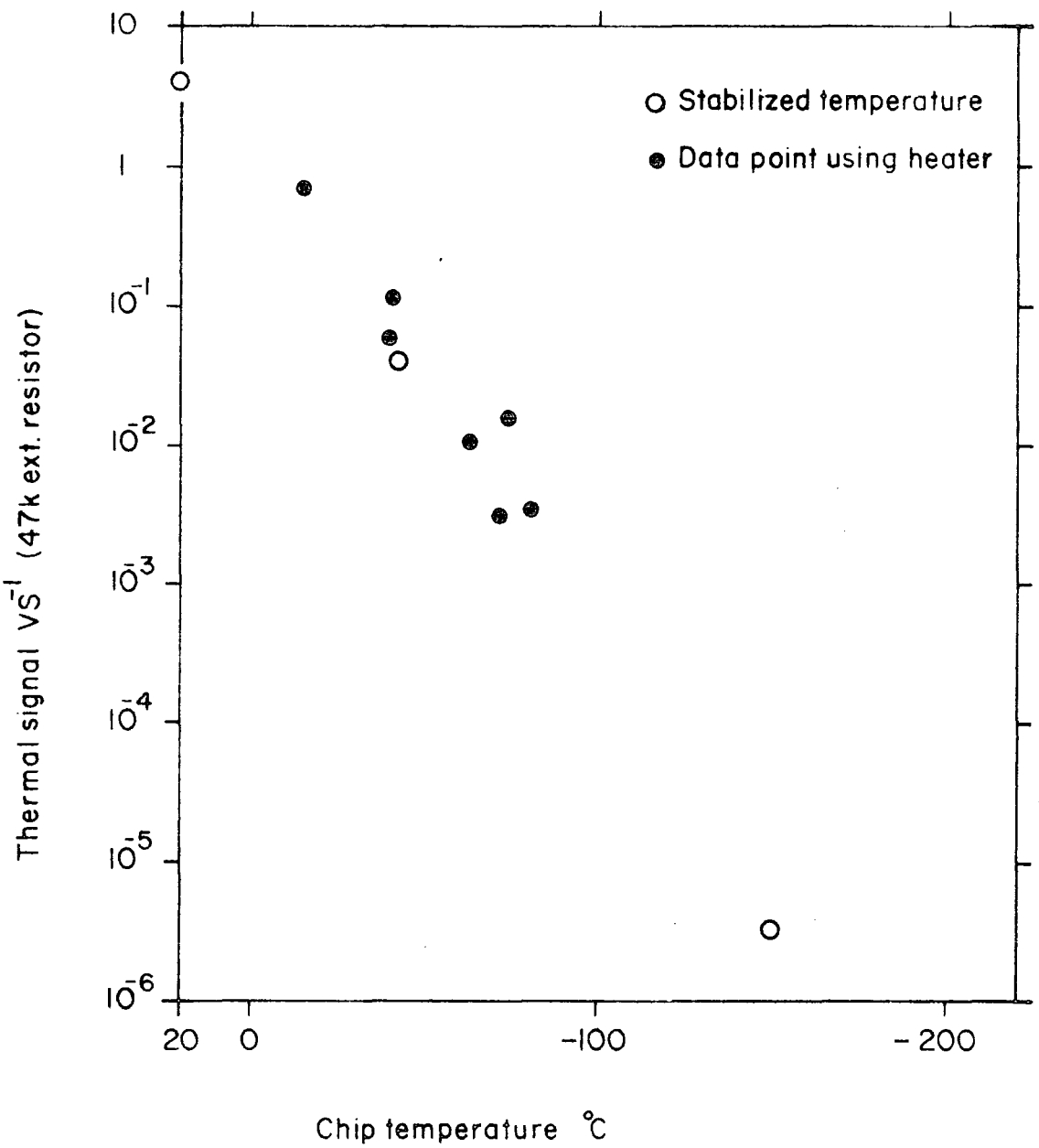
suggesting that the transistors might provide as much as 33% of the dark current at the beginning of an integration.

It was realised that very little useful work could be done with the array at room temperature since the dark current saturated the output in typically less than one second and, even when short integrations were used, this could not be easily removed. Therefore some form of cooling system was immediately required. A cryostat had already been purpose-built for this by the RGO and it, together with the arrays, was loaned to Durham University. This type of cryostat, shown in Figure 3.1 with a cold finger dipping into a glass dewar filled with liquid nitrogen as used in all the early work on the array, proved extremely useful for laboratory tests. A refined version incorporating several practical improvements has since been used for experimental runs on the coudé spectrograph of the 30" telescope at the RGO. Developmental work on this is described fully in a report by A. Humrich.[2] Although the advantages of this particular system are that it is simple to construct and that temperatures of $-120^{o}C$ may be easily attained, it is capable of operation at only one orientation. Accordingly, further work in development was launched in efforts to use the detector on a Cassegrain instrument which would itself require a cooling system unaffected by rotation through large angles. A full discussion of this appears in Chapter 6.

2(b) Effects of Cooling

For present purposes, the desired effect of cooling is the dramatic reduction of the dark current. Such reduction has been measured[3] using the cryostat mentioned above and these results are displayed in Figure (3.2). A drawback of this experiment is that the temperature sensor, in this case a thermocouple, must necessarily be mounted some distance from the chip. This lends itself to difficulty in the determination of whether or not the actual device temperature has stabilized. It is possible to adjust

Fig 3·2 DARK SIGNAL VS. TEMPERATURE



this temperature by means of a small heater in contact with the chip mounting block. At three of the data points - namely, room temperature, $-43^{o}C$ and the minimum temperature, $-120^{o}C$, it has also been possible to allow more time for the system to stabilize so that these particular points are thought to be reasonably accurate. Results show that the dark signal expressed as a rate of change of voltage on C_o is $< 10^{-4} VS^{-1}$ at $-100^{o}C$. This current is equivalent to a signal of approximately one volt being obtained in an integration of three hours. In fact, the dark current has always been found to be negligible at temperatures below $-100^{o}C$ for most practical exposure times.

The appearance of the output signal changes significantly as the device is cooled. Such alteration is due not only to the reduction in dark current but also to changes in the MOS device parameters. Of these, the most noticeable is an increase in β_o which is quoted as being ~ 6 μAV^{-2} at $25^{o}C$ and having a temperature coefficient of $-0.33\%\ {}^{o}C^{-1}$.

From simple theory[4] we have

$$\beta_o = \frac{\varepsilon_{ox}\,\mu}{T_{ox}} \qquad (3\text{-}2)$$

Assuming an approximately linear variation down to $-75^{o}C$, we have for a $100^{o}C$ change

$$\beta_{-75^{o}C} = 6(1 + 0.33) \quad 8\ \mu AV^{-2}.$$

Experimentally, at room temperature and below, the mobility is found to have a T^{-1} dependence which means that this value may be too small. In any case, a substantial increase in β_o is expected on cooling.

From equation (2-8) for $(V_G - V_T)\ R\beta \ll 1$

$$I = \beta_3 V_E^{\,2} - \beta_3^2 \frac{V_E^{\,3} R}{2} + \frac{5}{8}\beta_3^3 V_E^{\,4} R^2 \ldots + C_n \beta_3 V_E^{\,2} (\beta_3 R V_E)^n$$

Also,

$$\beta_3 = \frac{W}{L}\beta_o \propto \beta_o \qquad \text{and} \quad R = \frac{1}{\beta_5 V_E} \propto \frac{1}{\beta_o},$$

giving an expression of the form

$$I = \beta_o \ (aV_E^2 + bV_E^3 + CV_E^4 .....). \qquad (3\text{-}3)$$

Thus as temperature decreases we expect the output signal to increase. Observation supports this expectation although the situation is complicated by an increase in the value of V_{TO} which reduces V_E in Equation (3-3) and diminishes the effect of the increase in β_o as it shifts the whole characteristic along the V_E axis. Manufacturers' data gives V_{TO} = 1 to 2 volts and the temperature coefficient

$$\frac{dV_{TO}}{dT} = -0.2\% \ C^{o-1} \quad \text{at } 25^oC.$$

Again, assuming this holds over a temperature drop of 100^oC, we have

$$V_{TO} \text{ at } -75^oC \ \_\_ \ -1.5 \ (1 + 0.2) = -1.8 \text{ volts.}$$

At larger gate voltages (e.g. $V_G = -10$), the change in V_{TO} has a much smaller effect than the increase in β_o. No attempt has been made to measure large or small signal gain as a function of temperature but, in general, increases in large signal current of about 30% may be observed on cooling from room temperature to around -100^oC.

A much more important consequence of the change in V_{TO} due to temperature is the accompanying change in the diode reverse bias voltage. As noted, this reverse bias is determined by T1. When the array has just been cooled to the operating temperature during which time no light has been allowed onto the photodiodes, it is noticeable that, at the beginning of the first exposure, there is a considerable time delay before any signal is seen at the output. Under constant illumination at temperatures where dark current may be neglected, we have

$$V_D = V_{REF} - V_T(T) - \left\{\frac{2Ip}{\beta}\right\}^{\frac{1}{2}} \simeq V_{REF} - V_T(T). \qquad (3\text{-}4)$$

However, if, without illumination, the device is cooled further the equilibrium value of V_D will decrease but the actual voltage V_D on the diode has no way of discharging so that T1 is left in a region of operation with

$$|V_D| > |V_{REF} - V_T| \quad .$$

So long as this is the case T1 is effectively switched off and, at the beginning of an exposure, the diode itself must discharge an amount

$$\Delta V_D = |V_D| - |V_{REF} - V_T| \quad ,$$

before T1 begins to conduct so that some signal may be seen on C_o, the gate capacitance of T3. This voltage, ΔV_D is the difference between the present equilibrium value of V_D and the value at the temperature at which V_D was last established:

$$\Delta V_D = |V_D(T1)| - |V_D(T2)| = |V_T(T2)| - |V_T(T1)| \quad ,$$

where ΔV_D is a negative signal produced on the diode by cooling from T1 to a lower temperature, T2. This results in a delay at the start of an exposure and a complete loss of signal charge -

$$Q_{LOST} = C_D \Delta V_D \quad , \qquad (3\text{-}5)$$

which is equivalent to a voltage on Co

$$V_S \text{ LOST} = \frac{CD}{Co} \Delta V_D .$$

If we assume the same rate of change of V_T at lower temperatures, a temperature change of 10^oC will produce a negative signal

$$V_S \text{ LOST} \simeq \left\{\frac{C_D}{C_o}\right\} \frac{dV_T}{dT} \Delta T = 0.4 \text{ volt.}$$

Although this effect is most prominent after the device has been allowed to cool in the dark, temperature changes during an exposure will produce similar effects which may not be obvious because of their more gradual nature. Thus any temperature drop between the beginning and end of an integration produces a negative signal offset as described above and, conversely, any increase in temperature will produce a positive offset signal. Again, assuming that the temperature coefficient of V_T is fairly uniform along the array, this effect, particularly on long integrations, will be seen as a DC offset on the signal. In itself this

is not particularly objectionable but an unknown DC offset makes it impossible to accurately divide through the data with a flat field spectrum in order to remove responsivity variations. A consequence of this type of amplified element is that temperature drifts add a fixed pattern of offsets to the data which are not removed by double sampling. Moreover, these offsets are not easy to repeat or estimate. In the case of a positive temperature drift with positive offsets a number of blanked off reference diodes at the ends of the array might be used. However, this would not be of any assistance in dealing with negative signals. If interelement variations in V_T are typically 10%, variations in the temperature coefficient and, thus, the offset might be expected to be similar. Keeping the noise added by these interelement variations less than the system random noise requires

$$\frac{C_D}{C_o} \frac{dV_T}{dV} \Delta T < 10 \ \varepsilon_n$$

if

$$\varepsilon_n \simeq 0.5 \text{ mV} \equiv 250 \text{ electrons if Co} = 0.1 \text{ pF},$$

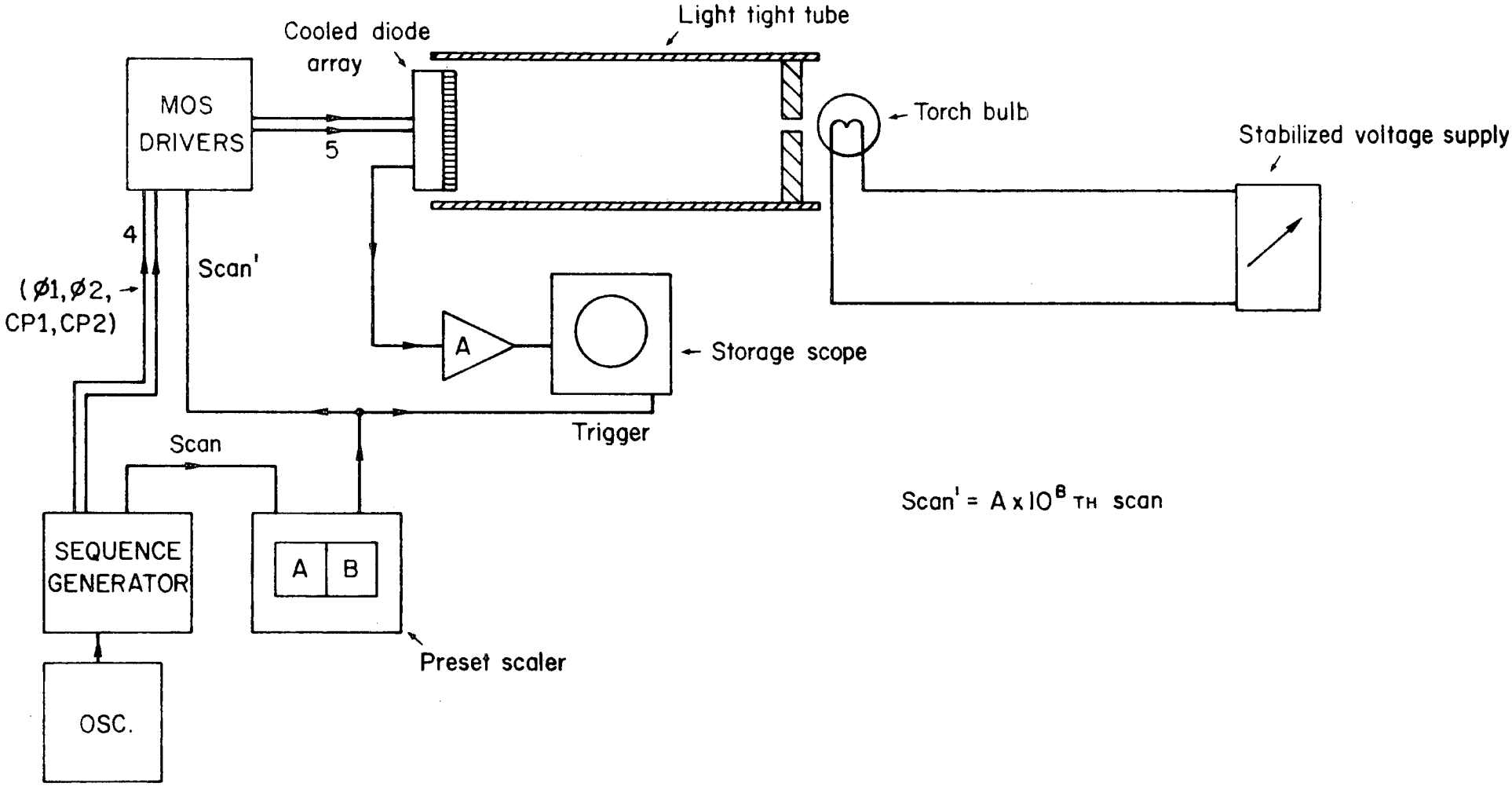
$$\text{then } \Delta T < \frac{5 \times 10^{-4} V}{4 \times 10^{-3} V_T} = 0.125^{\circ}C.$$

Achieving this sort of temperature stability over integration times of one hour or more is technically difficult.

3(a) Measurements of Photoresponse

Using the minimal drive system described earlier (Chapter 2) and a specially built preset scaler to gate the scan pulses, it is possible to make simple responsivity measurements with the cooled detector. The experimental set-up is provided by Figure (3.3). A fixed clock frequency is used and all the control pulses are sent directly to the array except for that of the scan pulse which, alternatively, is routed through the preset scaler. Two thumbwheel switches on the scaler decide how often the array is scanned insofar as they permit a multiplier $\Lambda = 1 \text{ — } 9$, and

Fig 3·3 MEASUREMENT OF RESPONSE vs. EXPOSURE TIME



an exponent B = 0. - - 7, to be set such that only every A x 10^B th scan pulse generated by the driver module actually loads the device shift register. With this apparatus measurements may be taken whose integration lengths vary over several orders of magnitude (that is, if the clock frequency is set such that the basic scan pulse frequency is 100 Hz, the fastest rate used, integration times between 10^{-2} seconds and several hours are possible). When these initial experiments were performed sophisticated data catching equipment was not available - leaving the signal values to be read from the screen of a storage oscilloscope. For simplicity the recharge timing was adjusted so that the diodes in each pair recharged simultaneously and the array was operated as though it had 128 big diodes. The operation of the oscilloscope was in delayed sweep mode with the time-bases adjusted to display the video signal from just one chosen diode pair each time the array was scanned. In this way, not only the difference between signal and reset levels could be seen, but also any changes in these levels between consecutive identical integrations became apparent as the traces registered in differing vertical positions to those previously stored. The zero current level from dead time between the phases of the clock provided a check against overall system drift. Although this system does not provide very good resolution, it is adequate to check approximate linearity with different exposure times under conditions of constant illumination. The array was cooled until the dark current was observed to be negligible and also illuminated by a torch bulb powered from a stabilized supply.

3(b) Incomplete Recharge

In using this system it became apparent that the reset voltage level, after a single recharge, changed with exposure. This had not been expected since, if the recharge pulses are long enough, Co should always recharge to the same voltage. As it was found that this level fell slightly

with increasing integrated signal, it suggested that recharge had been incomplete. The "on" resistance of the recharge transistor, T2, in unsaturated operation, is given approximately by

$$R_{on} = (\beta (V_G - V_T))^{-1} \quad \text{where } V_{DS} \text{ is small}$$
$$= (3 \times 10^{-6} \times 10)^{-1} \simeq 33\ K\Omega$$

If $Co \simeq 10^{-13}$F, this gives a recharge time constant

$$T = RoC_o \simeq 3.3 \text{ ns.}$$

In practice, the rising edge of the charge pulse is much slower than this $\sim 1\ \mu s$ and it is this rise time which determines the minimum necessary charge pulse width. Alterations to the charge pulse width and amplitude were tried but neither provided a cure for the reset voltage problem.

So, in an equilibrium situation where the same light level was looked at repeatedly for identical consecutive integration times, the signal and reset voltages and the difference between them were always the same. When the integration time was altered, the levels in the next few read frames changed gradually to find new equilibrium values. Two possible operational sequences for running the array in the laboratory emerged from this:

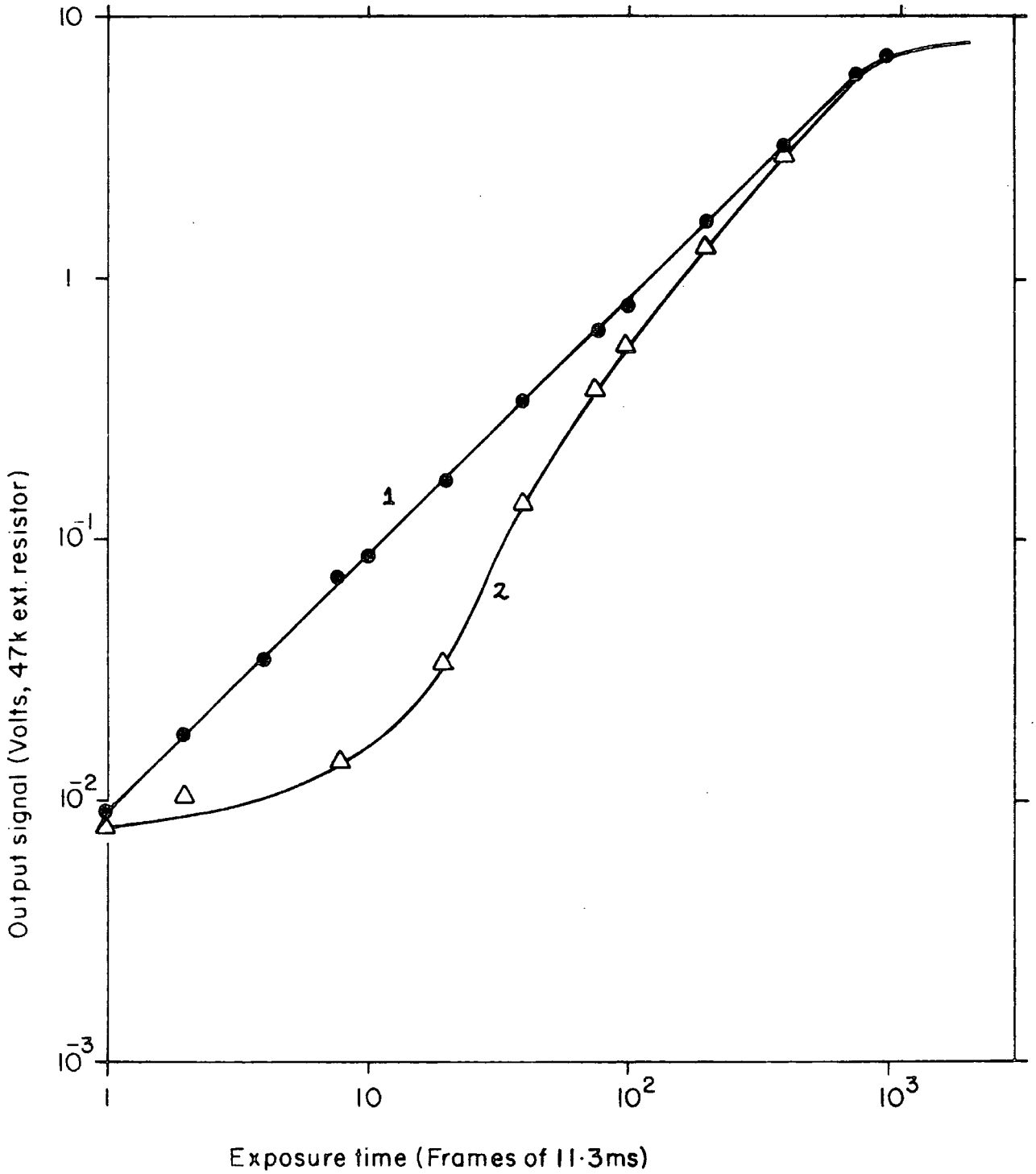
(1) Several consecutive integrations of n frame times are made until the output levels from the array are always the same and, then, the difference between signal and reset is measured; and

(2) Several consecutive integrations of 1 frame time are made (i.e. the array is scanned and recharged continuously) until equilibrium is reached. Then a single integration of n frames is made in order to measure the difference in levels at the end.

Data from these two methods of operation, as a function of integration time under constant illumination, is plotted in Figure (3.4).

The second method for which some additional control circuitry had to be built to allow a smooth change between two scanning rates is

Fig 3·4


10
1
10^{-1}
10^{-2}
10^{-3}
1
10
10^2
10^3
1
2
Output signal (Volts, 47k ext. resistor)
Exposure time (Frames of 11·3ms)

obviously similar to the approach used in diode array systems for astronomical work reported in the literature. Tull[5] mentions a 1% recharge inefficiency and uses a series of about five recharge scans before the beginning of an integration to remove any possibility of image retention. This problem is thought to be due to the external electronics rather than to any inherent property of the Reticon device used. In the case of the Plessey device, the problem seems to be on-chip and may be related to the switch charge storage and offset effect discussed earlier. (See Chapter 2).

As measuring the difference between the signal and reset level is, in any case, an approximation for measuring the difference between the signal level at the end of an integration and the reset level from which it started, it is expected that the first integration method (where the reset level in consecutive reads does not change noticeably) will give better results. In most laboratory tests with a fixed light intensity, this method seems to offer reasonable performance, although it is clearly impractical as a technique for astronomical observations.

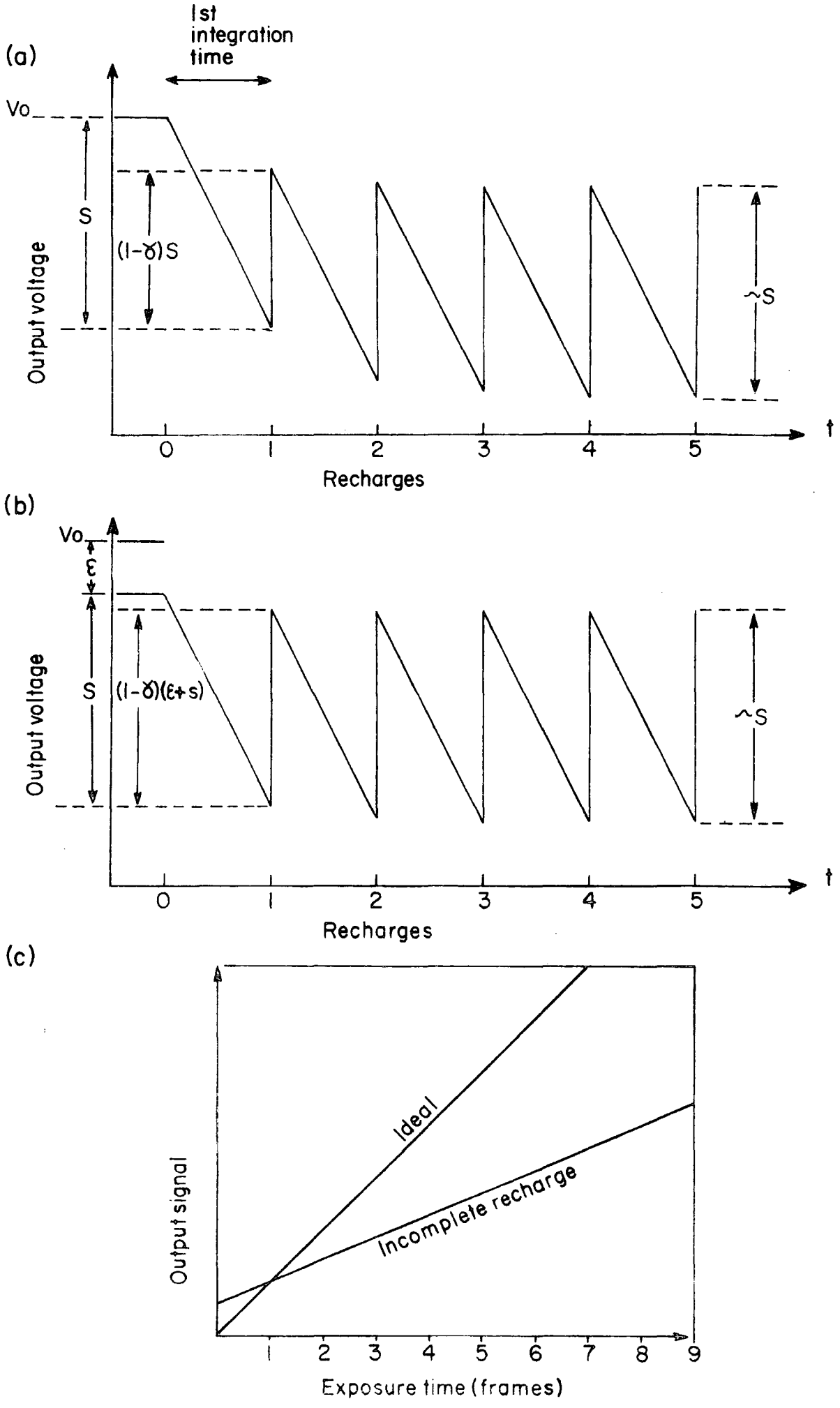
If it is assumed that for any element there is a voltage level, V_o to which the system attempts to recharge C_o, that at every recharge the actual value falls short of V_o by a fraction γ of the initial voltage difference and that the first integration starts from V_o (i.e. S = 0), the double sample signal in the readout frame at the end of one integration will be given by

$$S'_1 = S(1 - \gamma), \text{ where S is the true signal.}$$

The next integration now starts not from the zero signal level, V_o, but from the residual signal γS. This theoretical operation is shown in Figure (3.5(a)). After n consecutive integrations the double sample signal will be given by

$$S'_n = S(1 - \gamma^n) \simeq S \text{ for large } n \qquad (3\text{-}6)$$

Fig 3·5 INCOMPLETE RECHARGE



Again, in practice the first integration does not begin from V_o but from some value $\mathcal{E}$ which is the equilibrium reset value due to an integration of one frame time. (Figure (3.5(b)).

$$\text{Now } S_n' = S(1 - \gamma^n) + \mathcal{E}\gamma^{n-1}(1 - \gamma) \simeq S \text{ for large } n. \quad (3\text{-}7)$$

For the first double sample

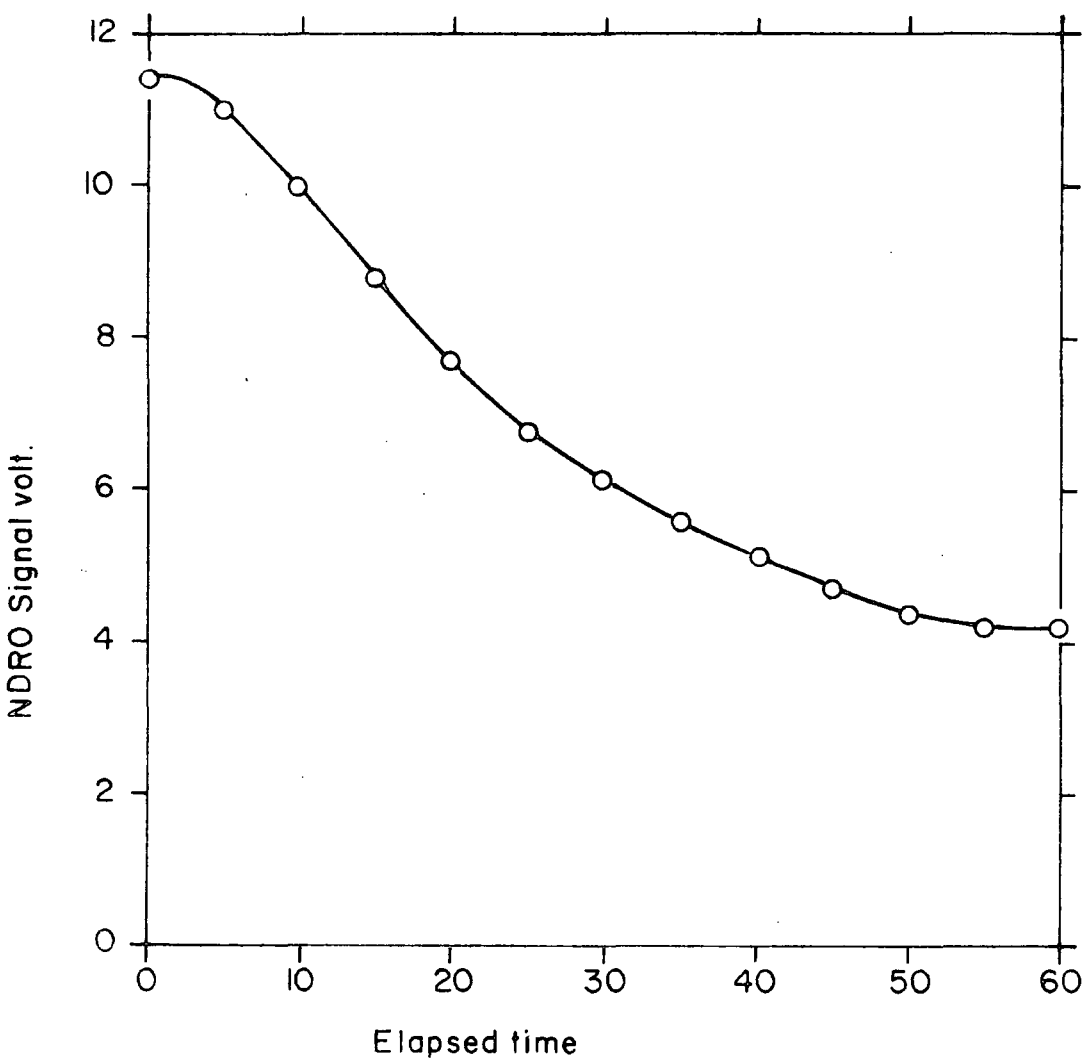
$$S_1' = (S + \mathcal{E})(1 - \gamma) \quad (3\text{-}7(a))$$

In the case where S is the signal in one frame time we know $S_n' = S_{1\ \text{FRAME}}$ for any n. Thus from (3-7(a))

$$\mathcal{E} = \frac{\gamma}{1 - \gamma} S_{1\ \text{FRAME}}.$$

According to this simple model with a value of γ which is independent of signal, the effect of poor recharge on the first double sample value (readout method (2)) is simply to reduce the responsivity by a factor $(1 - \gamma)$ and to add a small offset. (Figure 3.5(c)). The signal on C_o, however, should still be linear with time. Again, in practice this is not found at short integration times nor with more intense illumination as Figure (3-4) shows. As a rule, after a period of more rapid recharging, the output signal from the array tends to hang-up. Figure (3.6) highlights a typical graph of output signal vs. time for the array read out non-destructively - that is, it is scanned continuously but the charge pulses are switched off and the fall in the signal level is plotted as a function of elapsed time. The saturation feature at the bottom of the curve may be explained by the non-ideal source follower characteristic but the slow start effect which is also seen - this time in Figure (3.4(b)) - is unexpected. Since this initial poor responsivity is not as apparent in data taken by the recycling equilibrium method, it is thought to be due to some dynamic effect and is apparently not a fixed distortion in the output characteristic of the array. Noting the manufacturers' realisation that the amplifier transistor, T1 would introduce a lag effect between the signal on the diode and that seen at the buffer transistor, T_3,

Fig 3·6 TYPICAL NDRO RESPONSE



in this experiment, the array is exposed to constant illumination at all times so that the drain current of T1 should reach an equilibrium value equal to that of the photocurrent. Thereafter the time response of T_1 should be unimportant. However, it is difficult to see how the other two transistors involved in the operations of recharge and integration, T2 and T3, might produce the "slow start" effect seen in the data. Both components are common to other voltage sampling arrays and effects of this sort have not been reported. The operation of one such array, an IPL 4064 has been investigated at Durham for comparison purposes.

3(c) Amplifier Transistor Lag and V_{REF} Considerations

Observation of the array being cooled in the dark has shown that T1 can produce a hold up effect at the beginning of an integration, if by some means T1 becomes biased into a region of operation where

$$|V_D| > |V_{REF} - V_T|$$

and T1 is effectively switched off. If such a condition is met an amount of signal charge

$$Q = |V_D - (V_{REF} - V_T)| \; C_D$$

must be integrated on the diode before T1 begins to conduct. Now there will be a switch-on time during which

$$|V_{REF} - V_T| > |V_D| > |V_{REF} - V_T - \left(\frac{2\,Ip}{\beta}\right)^{\frac{1}{2}}|,$$

and finally the equilibrium situation will again be reached where

$$|V_D| = |V_{REF} - V_T - \left(\frac{2\,Ip}{\beta}\right)^{\frac{1}{2}}|.$$

Thus, if a lag is observed as in the experiment described where the photocurrent, Ip remains constant throughout and the temperature is known not to have changed significantly, it seems likely that performing several consecutive recharges is somehow changing V_D and disturbing the equilibrium of T1. That several recharges are required to produce this effect is indicated by the fact that lag is less noticeable in the case of

the first integration method where several integrations are done and the time between recharges is comparatively long. The mechanism by which recharges affect V_D is not known exactly, though it is possible that the rising edge of the charge pulse produces a significant spike on V_{REF} by capacitive coupling through C_{GD} as the gate to drain parasitic capacitance of T1. Any negative spikes picked up on the V_{REF} line will increase the instantaneous value of V_{REF} and distorb the equilibrium of T1. During a long integration this will be unimportant since equilibrium will eventually be restored. As long as the overall value of V_{REF} does not change no net signal will result. However, if a negative spike occurs on V_{REF} during the time of the recharge pulse, T1 acts as a peak value detector given that current may flow only one way through this transistor. At the same time the effect of the photocurrent discharging the diode is not significant over such a short interval. Thus T1 attempts to establish a value of V_D appropriate to the most negative value of V_{REF} seen during this time. The charge necessary for this is supplied by the recharge circuit and not C_o which is recharged simultaneously. Although the slow time response of T1 will reduce this effect, it remains a possible mechanism whereby a small negative signal is placed on the diode in every recharge time. Several repeated recharges at intervals such that the equilibrium of T1 cannot be re-established will enhance this effect. This explanation suggesting a transient effect with a time constant which varies with incident illumination seems to fit much of the observed lag behaviour.

3(d) Signal Dependence of Lag

If at a time t = 0 the transistor, T1 is just switched off,

$$V_D = V_{REF} - V_T,$$

the drain current at a time t later will be given by

$$I_D = \frac{\beta}{2} \left(V_{REF} - V_T - \left(V_{REF} - V_T + \int \frac{Ip-Id}{C_D} \, dt\right)\right)^2,$$

$$= \frac{\beta}{2} \left(Ipt - \int I_D \, dt\right)^2 \frac{1}{C_D^2}$$

where Ip = photocurrent.

Rearranged we get

$$\left\{\frac{2 I_D C_D^2}{\beta}\right\}^{\frac{1}{2}} = Ipt - \int Id \, dt$$

so that in differentiating and solving for $I_D(t)$ we have

$$I_D(t) = Ip \tanh^2 \frac{t}{2\tau} \quad . \qquad (3\text{-}8)$$

where $\tau = \left\{\frac{C_D^2}{2Ip\beta}\right\}^{\frac{1}{2}}$. The form of this is shown in Figure 3.7(a).

Integrating this to find the signal appearing on Co after a time t requires:

$$\int I_D(t) \, dt = Ip \left[t - 2\tau \tanh \frac{t}{2\tau} \right] \qquad (3\text{-}9)$$

as shown in Figure 2.7(b).

For ideal operation, (i.e. under steady state conditions) the signal at a time t should be given by Ipt so that the signal charge discrepancy after a time t is given by

$$\Delta Q = 2 Ip\tau \tanh \frac{t}{2\tau} \qquad (3\text{-}10)$$

which for large t becomes $\Delta Q \rightarrow 2 Ip\tau = \left\{\frac{2IpC_D^2}{\beta}\right\}^{\frac{1}{2}}$ (3-10(a))

$= V_{OFF} C_D$ where V_{OFF} is the difference in V_D between time t = 0 and the equilibrium situation.

Equation (3-10(a)) represents simply the extra charge that must be placed on the diode to raise $|V_D|$ sufficiently so as to turn T1 off.

The fractional error

$$\frac{\Delta Q}{Q} = \frac{2\tau}{t} \tanh \frac{t}{2\tau} \rightarrow \frac{2\tau}{t} \quad \text{for } t \gg \tau \quad .$$

2τ becomes a characteristic time associated with switch-on or the restoration of equilibrium in T1.

For typical values where

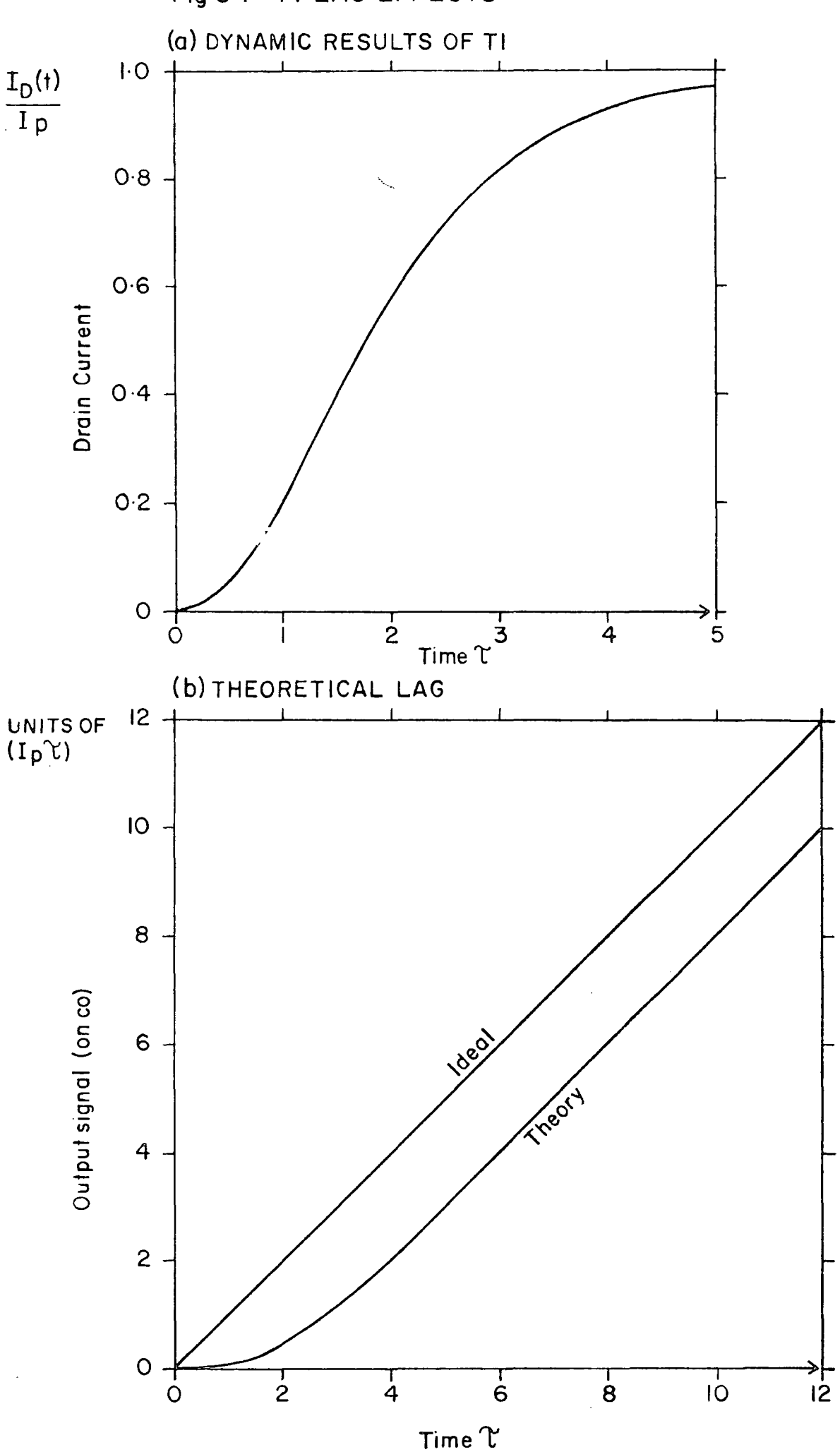
$Ip = 10^{-15}$ A equivalent to saturation in 10^3 S (15 min.),

$C_D = 10^{-12}$ pF, and

$= 4 \times 10^{-6} \, AV^{-2}$,

we have $2\tau = 3 \times 10^{-2}$ S.

Fig 3·7 TI LAG EFFECTS



From Equation (3-10) we see that the time necessary to achieve 1% accuracy is

$$T_{1\%} = 200\tau = 3s.$$

In Table 3.1 these values are compared for a range of incident signals. The fraction of saturation time necessary to achieve 1% accuracy is calculated. It should be noted that saturation time here is arbitrarily defined as the time to obtain a 10V signal excursion on Co ($=10^{-13}$ F) assuming a linear response.

Table 3.1

Photocurrent	Saturation time	2τ	(T1%/Saturation Time) %
10^{-11} A	10^{-1} S	3×10^{-4} S	30%
10^{-13} A	10 S	3×10^{-3} S	3%
10^{-15} A	10^{3} S	3×10^{-2} S	0.3%
10^{-16} A	10^{4} S	10^{-2} S	0.1%

It can be seen from this simple model that although the times necessary for operation to regain linearity become longer for lower signals, these times expressed as a fraction of the saturation time do improve. So, at high light levels, the non-linear part of the output characteristic due to the dynamic properties of Tl extends over a much larger fraction of the total available signal excursion. With an ideal lag free response to obtain a signal to noise ratio of 100: 1 on a diode - that is, neglecting readout noise and considering only shot noise, it would be necessary to detect 10^{4} electrons. This, expressed as a signal voltage on Co is 16 mV (where Co = 0.1 pF) or 0.16% of maximum signal. On the other hand, looking at Table 3.1 it can be seen that, if the effect of lag is included, it may be necessary to integrate for much longer in order to achieve 1% accuracy. This is the case especially at high incident signal levels. Whether or not this presents a serious problem depends somewhat on the application. As lag reduces the signal seen after a given integration time and is

expected to affect similarly all the diodes in the array, it does not reduce resolution (i.e. the ability to see differences in signal on different diodes) in the same way that random noise can. Where intensities of spectral features falling on different parts of the array are compared, lag is important because it degrades the photometric accuracy of the signal. Whereas random noise determines the ability to discriminate between very similar signal levels, lag determines the photometric accuracy with which different signal levels may be compared when large variations of incident signal are seen along the array. (For example, looking at an emission spectrum with strong and weak lines). Having to integrate for longer than would otherwise be necessary is unfortunate because the output characteristic of the buffer transistor, T3 is not ideally linear so that at larger signals the error thereby introduced is greater. However, it might be possible to use a shutter to terminate the exposure and subsequently allow a period of dark integration for the full signal to reach Co before the readout is made. Hence, signal excursions are kept small.

3(e) Summary of Non-Ideal Behaviour

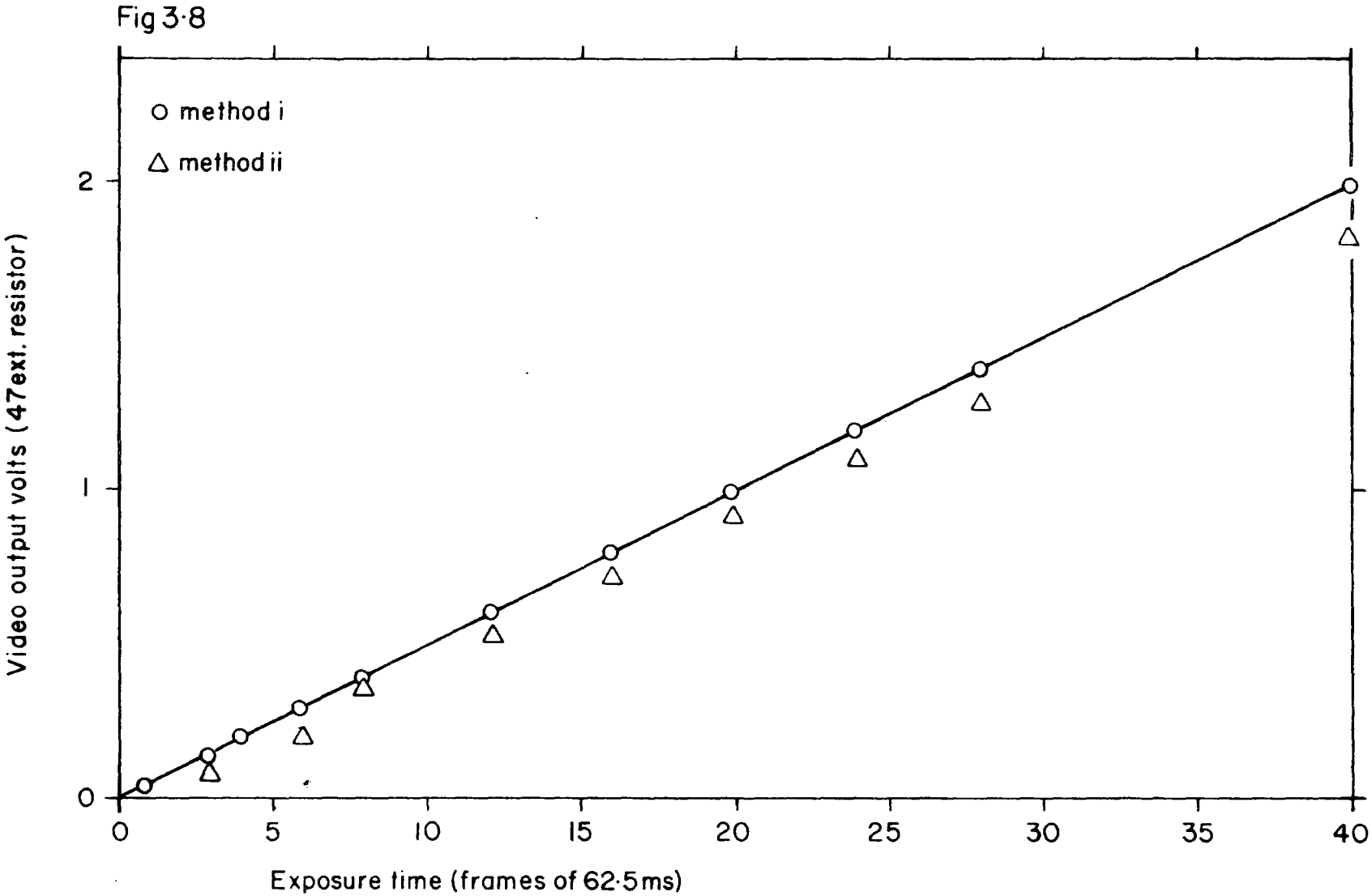
Taking the two effects described, incomplete recharge and lag, it is possible to qualitatively explain the difference between the signal measured by both integration methods. The conclusions are that:

(1) Recharge after just one destructive read is incomplete; and

(2) A period of rapid recharging with very little signal being integrated between recharges causes T1 to effectively switch off by raising V_D above its equilibrium value. This causes a considerable lag effect at the beginning of an integration.

Combining the above with the nonlinear source follower characteristic gives the apparent S shaped response frequently found in laboratory tests which use light sources that are generally more intense than those

Fig 3·8


o method i
Δ method ii
Video output volts (47ext. resistor)
0
1
2
0
5
10
15
20
25
30
35
40
Exposure time (frames of 62·5ms)

encountered in actual astronomical work. On the other hand, we may say that the response is fairly linear if we look at a sufficiently small incident signal for long enough and use a restricted range of output characteristic. Considerable care need be exercised in this. Figure (3.8) shows the output signal values given by the two integration methods, recycling and single shot, for a less intense source. These are plotted on a linear scale.

4. Responsivity at different light intensities

The response of the array using the recycling method of integration has been measured employing different incident intensities. Originally this was done by illuminating the array with diffuse light from apertures of known sizes. Results plotted in Figure (3.9) indicate that, operated in this manner, there is no apparent departure from reciprocity.[6] This experiment has since been repeated with improved apparatus and neutral density filters.[7] This time the resulting graph of signal integrated in a fixed time vs. relative filter transmission shown in Figure (3.10) gives a reasonably straight line.

5. Effects of Changes in Chip Supply Voltage and Control Pulse Amplitudes

(a) Introduction

Direct measurement of the effects of changes in the various control voltages is difficult because these are interdependent and have to be deduced by looking at the output signal - no other access being possible to points of interest within the circuit. Initially work was done on this in efforts to find optimum values for operating the array.

(b) Clock Pulse Amplitude

The variable parameter with the most pronounced effect is that of the amplitude of the shift register clock pulse. In Figure (3.11) the output levels, signal and reset, within a given integration time, are

Fig 3·9 RESPONSE VS. ILLUMINATION

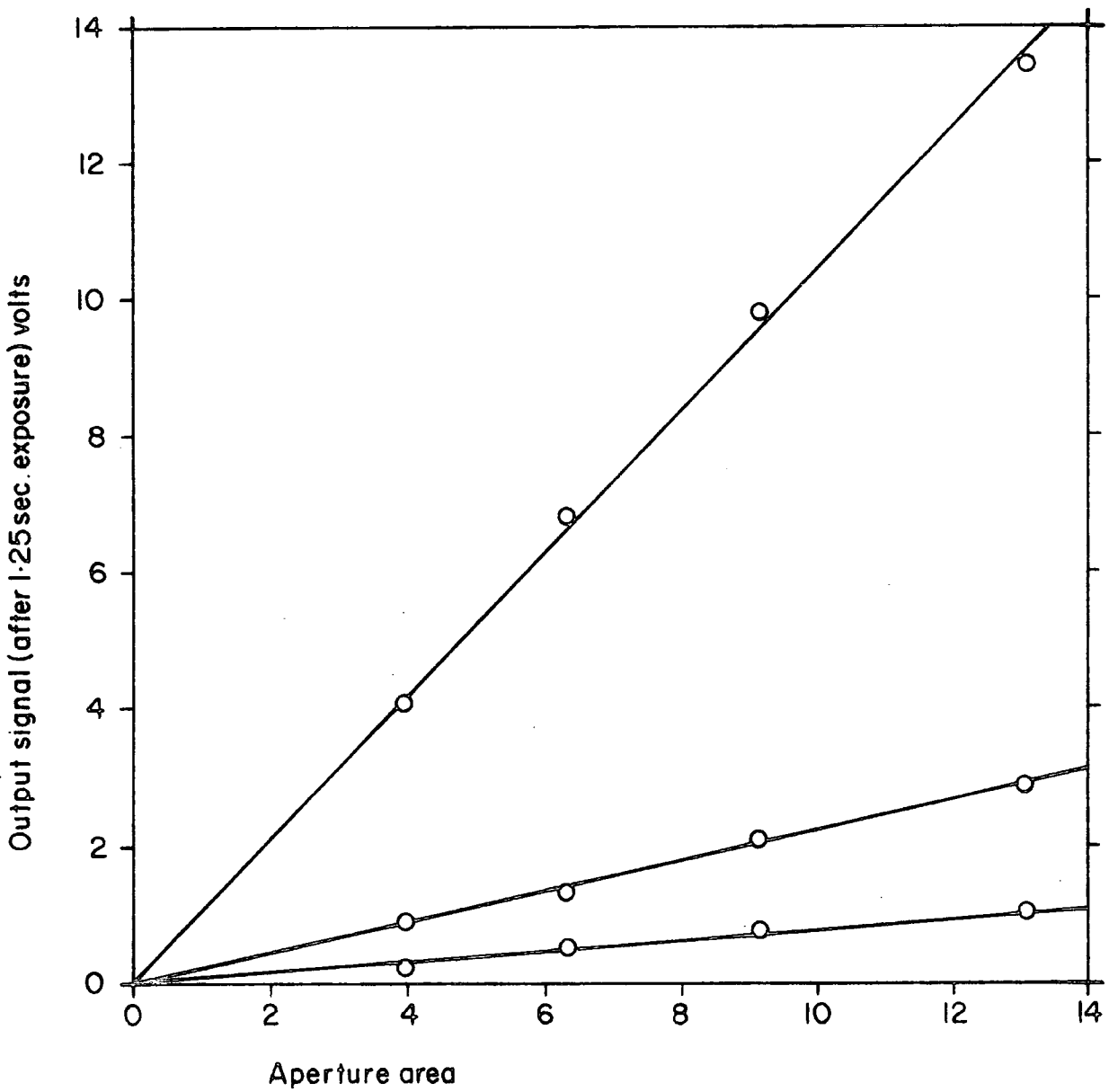
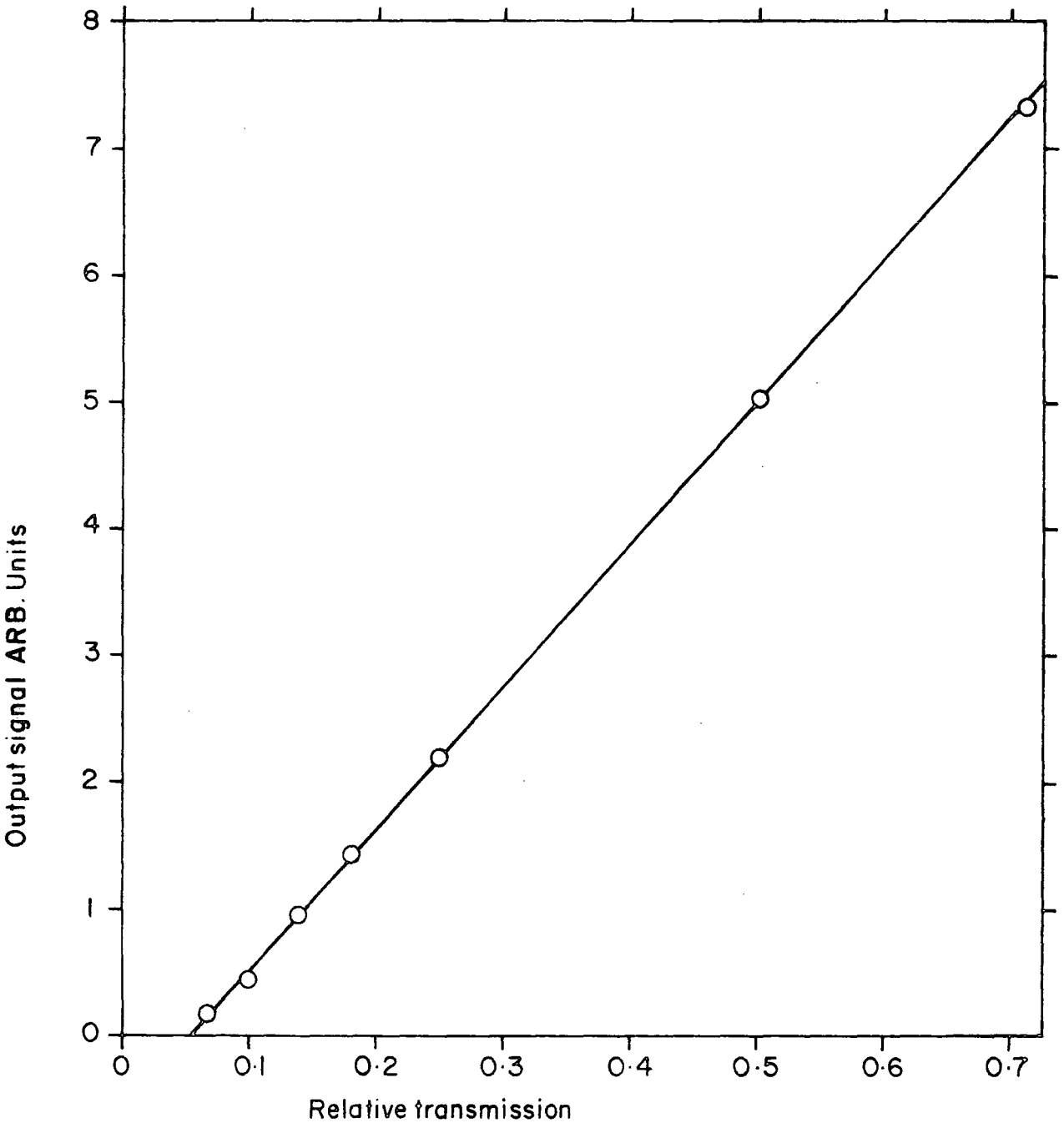


Fig 3·10


8
7
6
5
4
3
2
1
0
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
Output signal ARB. Units
Relative transmission

plotted against the amplitude of just one of the phases of the clock, $V\phi 1$ for various of the values of chip supply voltage, V_{DD}. The amplitude of the other phase, $V\phi 2$ is held constant. Although this was done at room temperature where the signal is mostly thermal, it can be anticipated that the behaviour of the output will be similar, when cooled, if changes in β and V_T are taken into account. The accompanying graph demonstrates the relationship between maximum output signal, $V\phi$ and V_{DD}. The output of an adjacent diode pair which is controlled by the other phase of the clock is also shown. There is a minimum value of $V\phi 1$ necessary for any operation of the shift register. As $V\phi 1$ is increased past this value, the output from the pairs accessed by the other phase rises quickly and saturates. The output signal from the pairs enabled by $V\phi 1$ rises steadily because $V\phi 1$ determines the maximum recharge voltage of the elements. From Equation (2-5)

$$V_{RECH} = V\phi 1 - V_{T4} - V_{T6} \quad , \qquad (3\text{-}0)$$

where $V_{RECH} < V_{DD}$ and saturates as V_{RECH} reaches V_{DD}, the recharge supply voltage. At larger values of V_{DD}, this saturation condition cannot be reached before proper operation of the shift register ceases because $V\phi 1$ has become too large. A slight drop in the outputs of the $\phi 2$ pairs is seen as $V\phi 1$ is increased. This drop may be due to some internal loading effect.

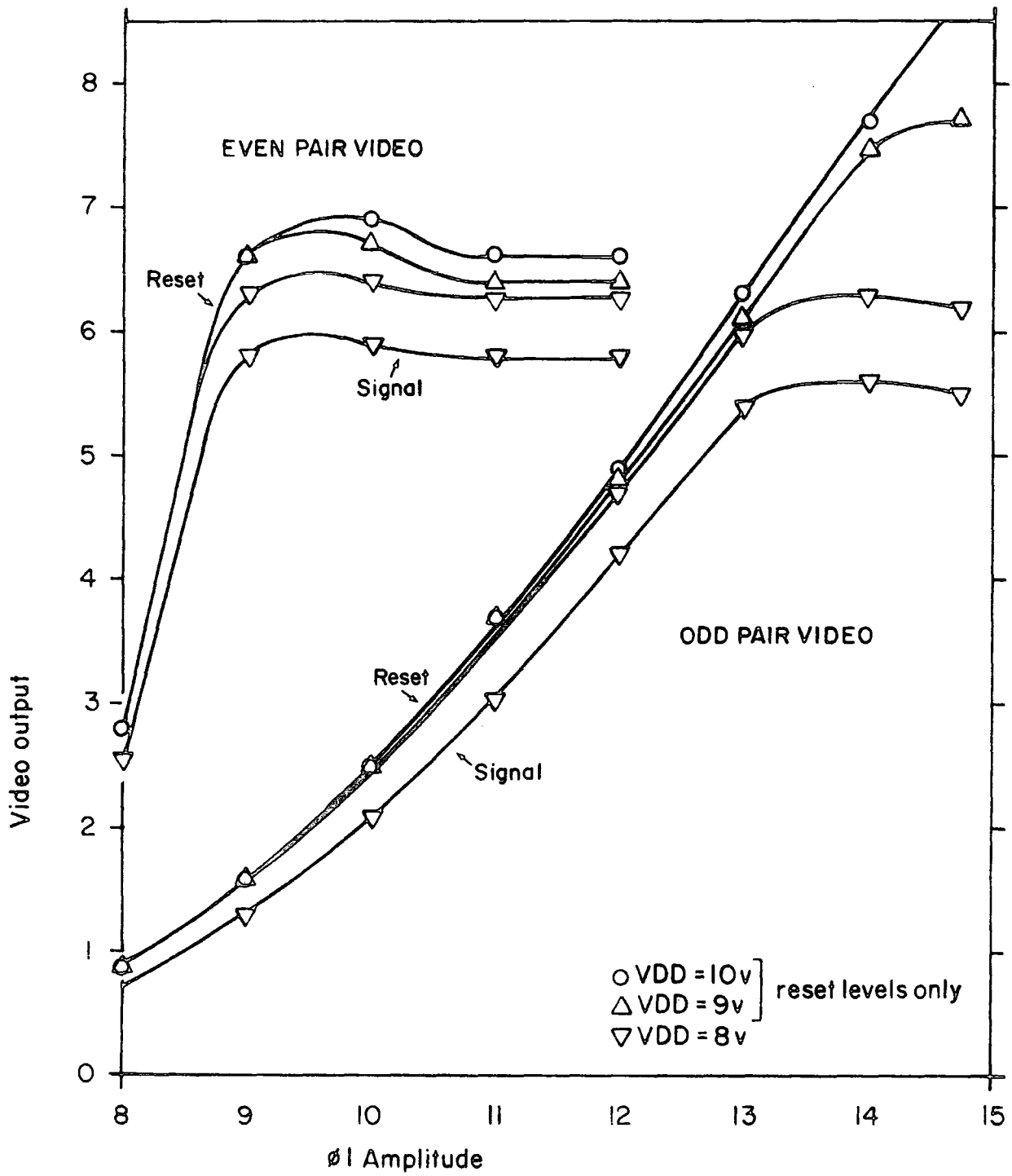
From the graph we see that for $V_{DD} = -8V$ the onset of saturation is at $V\phi 1 \simeq -13$ V which suggests that,

assuming $V_{DD} \simeq V\phi 1 - V_{T2} - V_{T4}$,

$$V_{T2} - V_{T4} \simeq 5V.$$

As both T_2 and T_4 are operated with elevated source voltages, their effective threshold voltages will be greater than V_{T0} for the process. This is consistent with a value of V_{T0} of between 1 and 2 volts. The general implication here is that $V\phi$ should be at least two thresholds greater than V_{DD} if the element is to recharge fully. Recharging to

Fig 3·11 OUTPUT VS. CLOCK AMPLITUDE



less than V_{DD} limits the available signal excursion and may result in greater noise in the recharge operation since the final recharge voltage is determined by $V\phi$ - itself unlikely to be as stable and noise free as V_{DD}.

The reverse experiment of changing V_{DD} with $V\phi$ may be found in Figure 3.12. The output signal from an element first appears when V_{DD}, determining the maximum recharge voltage, passes the threshold voltage of T_3. The latter (i.e. T_3) is always in saturated operation because the maximum voltage which may appear on its gate is V_{DD} - that is, the same voltage which is supplied to its drain. Consequently, the drain current increases steadily as V_{DD} is increased until the T_2 saturation condition as given in Equation (3-0) is reached. The output then flattens off but does not saturate completely suggesting that the drain current of T_3 is not entirely independent of drain voltage in the region of saturated operation. Again, this may be due to modulation of the channel length.[8]

(c) Measurement of Transfer Characteristic of T3

The most convenient way of measuring the transfer characteristic of the source follower stage is by applying a variable d.c. voltage to the charge pulse line in order to measure output as a function of this voltage. Since T_4 is unsaturated as long as

$$V_{CP} < V\phi - V_{T4}$$

it may be regarded as a simple resistance. T_2 will now act in a similar manner to T_1 and so maintain the gate of T_3 at a voltage a threshold lower than its own gate voltage, V_{CP}. Hence plotting output against V_{CP} should give a graph of output versus V_{G3}, shifted by V_{T2}. On estimating the slight change in V_{T2} due to body effect, the small signal transconductance may be measured from the graph. Accordingly in operating the device in this way, it can be seen that both the diode and the amplifier transistor are rendered superfluous. The remaining equivalent circuit is shown in Figure (3.13) and, similarly, a typical graph of output signal for different values of

Fig 3·12 OUTPUT VS. VDD

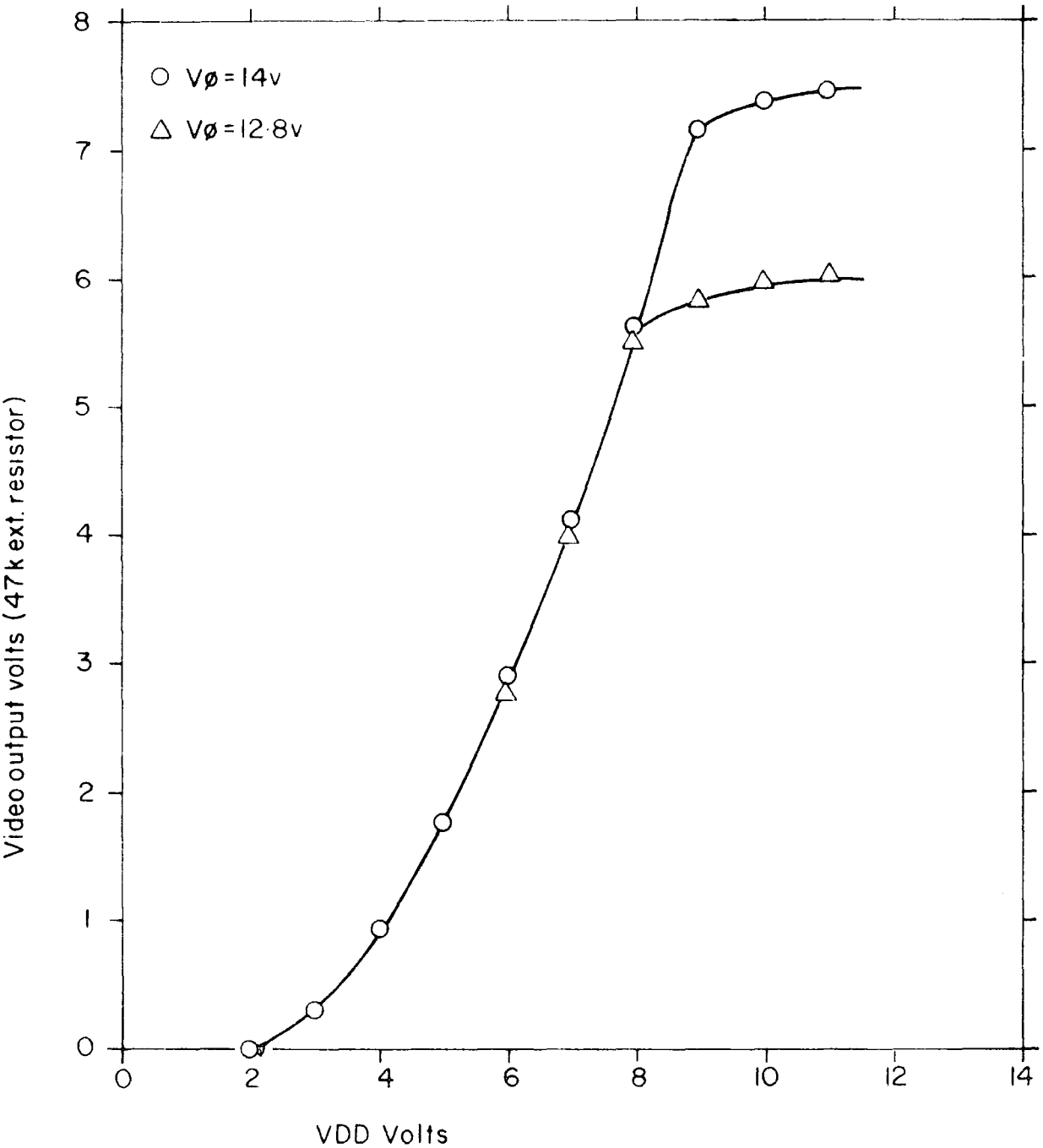
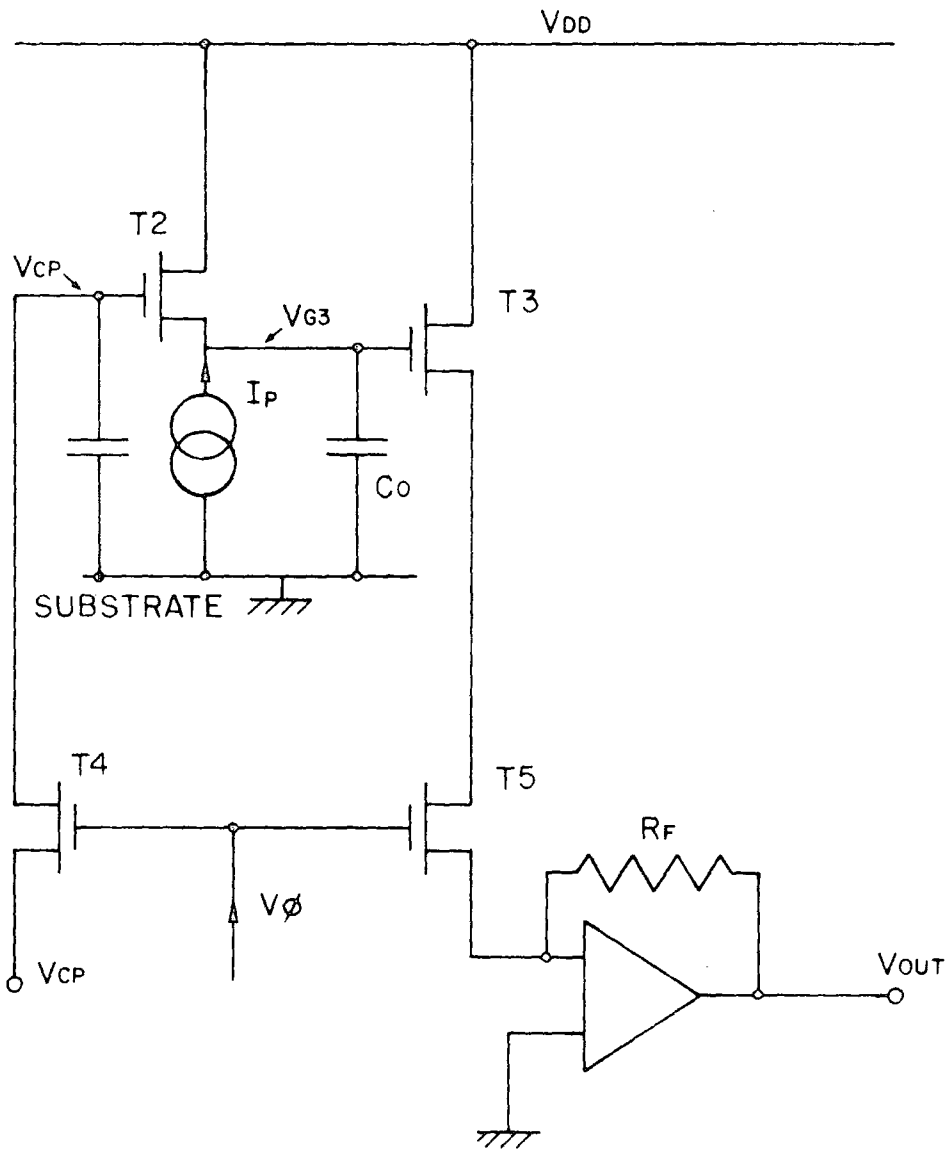


Fig 3·13 MEASUREMENT OF T3 CHARACTERISTIC



applied V_{CP} bias in Figure (3.14). Having allowed for body effect, the slope of the graph gives a small signal gain at the beginning of an integration of approximately 1 for an external amplifier with a 47KΩ feedback resistor. The calculated value for the voltage gain using Equation (2.10) is

$$G_S = \frac{dVout}{d\,V_G} = \frac{Rf}{R}\left[1 - (2\beta R(V_G - V_T) + 1)^{-\frac{1}{2}}\right]$$

putting $Rf = 47K\Omega$,

$R = 5\ K\Omega$,

$(V_G - V_T) = 8V$,

and $\beta = 6 \times 10^{-6}\ AV^{-1}$, gives $\frac{dVout}{dV_G} = 1.67 = G$.

However, as shown in Equation (3.3), the output and small signal gain are proportional to β_o for the process. The manufacturers quote

$$4\ \mu AV^{-1} < \beta_o < 8\mu AV^{-1}.$$

The above discrepancy between theory and experiment may be obtained by a low value of β_o — for example, for $\beta_o = 4\ \mu AV^{-1}$,

$G = 1.11$.

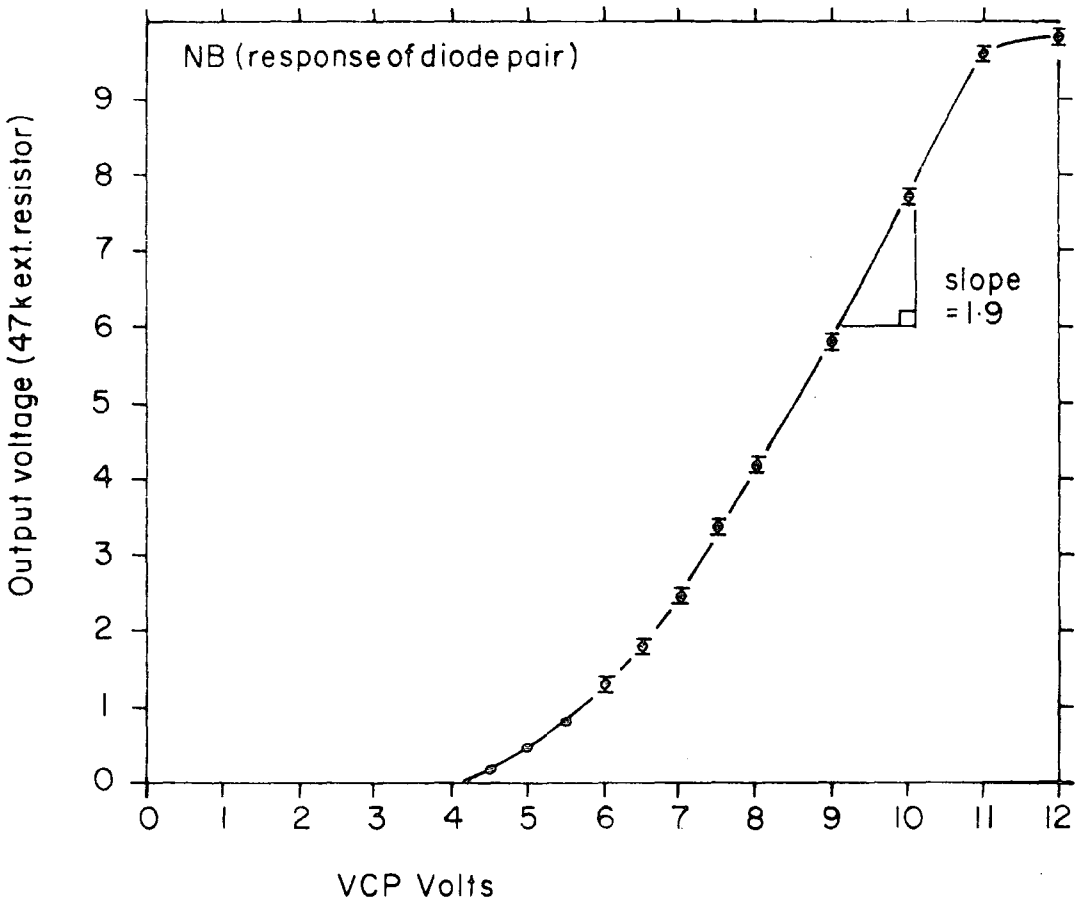
At room temperature similar measurements were made on ten diode pairs of an array for which the values of small signal gain were not found to differ within experimental error.

6. Calibration of Photoresponse, Quantum Efficiency and Noise Measurements

(a) Introduction

To assess the full potential of the array as a detector, having established that it can offer, under certain conditions, reasonable linearity, it is necessary to know: (a) the absolute photoresponse in terms of signal output per incident photon; (b) the responsive quantum efficiency; and (c) the overall system noise. Some idea of the expected output per detected photon has been derived from measurements of the electrical properties of the array. Thus if we can measure the actual

Fig 3·14 MEASUREMENT OF T3 CHARACTERISTIC



output under illumination from a calibrated source, the RQE of the device may be estimated.

$$\text{RQE} = \frac{\text{SIGNAL PER INCIDENT PHOTON}}{\text{SIGNAL PER DETECTED PHOTON}}$$

Where the signal per detected photon is known, the observed output noise at zero illumination may be converted to an equivalent number of electrons. From this figure and the RQE it is possible to calculate the exposure time required to obtain a given signal to noise ratio when looking at a source of known strength. Unfortunately it is difficult to measure with great precision any one of the above quantities, although reasonable estimates are obtainable from the combination of the results of several experiments.

(b) Initial Evaluation of Absolute Responsivity

Output from a light source powered by a stabilized d.c. supply was measured in a particular wavelength band using a narrowband filter centred at 5434 $\overset{o}{A}$ and a Centronics OSI 5K photodetector calibrated at the wavelength of interest. Then output of one element of the array, when placed in the same position as the calibrated photodetector under identical conditions, was measured after a fixed integration time. This method[9] suggested an output of

0.66 μV per incident photon,

with an external amplifier feedback resistor of 47 KΩ .

From simple theory this is given by

$$\text{OUTPUT/INCIDENT PHOTON} = \frac{q_e}{C_o} \quad Gs \times \text{RQE} \qquad (3\text{-}11)$$

where Gs = small signal gain $\sim$ 1,

q_e = electronic charge 1.6×10^{-19}C.,

Co = capacitance associated with the gate of T3 $\sim$ 10^{-13}F.

In using these values we get

$$\text{RQE} = \frac{0.66}{1.6} \simeq 40\%.$$

However, Co is not known from any experimental measurement and the

manufacturers' data suggests that this value may be in error by a factor of 2 so that the above estimate of RQE is subject to the same uncertainty.

(c) Preliminary Noise Experiments

While various attempts have been made to measure the readout noise of the diode array system and to obtain some idea of the shape of its power spectrum, the sampling effect of the scanning system hinders these endeavours. Early tests to find a value for the r.m.s. readout noise simply involved looking at the signal level from a selected element on a storage oscilloscope screen at the highest possible gain. By using the recycling integration method, it is possible to see the spread in vertical position of this level produced by several consecutive similar integrations. This sort of measurement is very sensitive to system drifts which the double sampling ability of the array should be capable of removing. To determine the usefulness of double sampling as a noise reduction technique, it is necessary to examine the difference between a signal and a reset level over several integrations. For small signals, this may be accomplished by using the storage oscilloscope method of data catching. Measured in this manner the noise at zero incident illumination appeared to be $\sim$ 500 μV with a 47KΩ feedback resistor in the virtual earth amplifier circuit. Taking the previously estimated value for output per detected photon of 1.6μV, the derived readout noise of $\sim$ 300 electrons would seem reasonable. The noise seen in a single level (i.e. without using double sampling) was found to be an order magnitude greater, but this was thought to depend greatly upon the length of time over which such measurements were taken. Since the difference between signal and reset levels must be kept small enough for both to register on screen at a sufficiently high gain for the noise to be seen as well, storage oscilloscope readout has its limitations. Continuation of the experiment to large integrated signals would require digitization of the output and, with that, more sophisticated apparatus.

(d)Combined Noise and Responsivity Measurement

If the readout noise due to the effects mentioned in Chapter 2 is independent of signal, the total noise on any signal is given by

$$N_{tot}^2 = \alpha N + \sigma^2 \text{ electrons}^2 \qquad (3\text{-}12)$$

where

α = RQE,

N = total number of incident photons

σ = RMS readout noise in electrons.

Expressed as a noise voltage at the output this will be

$$Vn^2 = G^2 (\alpha N + \sigma^2)$$

where G = the output per detected photon - that is,

given $V_S = G \alpha N$,

$$Vn^2 = GV_S + G^2 \sigma^2.$$

Thus measuring the mean square noise voltage as a function of signal should yield an accurate value of G, the output per detected photon. From this and the noise voltage at zero incident signal, the r.m.s. readout noise in electrons may be found.

In Chapter 2 it was seen that the thermal noise due to the amplifier transistor depends on its drain current, I_D which, in equilibrium, is equal to the photocurrent, Ip. This noise contribution will therefore increase with incident illumination. The thermal noise contribution of the buffer transistor, T_3 is inversely related to the signal voltage on its gate and, accordingly, this noise contribution will increase with integrated signal. Therefore with respect to the Plessey array, it cannot be said that noise may be separated entirely into shot noise and a signal independent readout noise as in Equation (3.12).

In general then

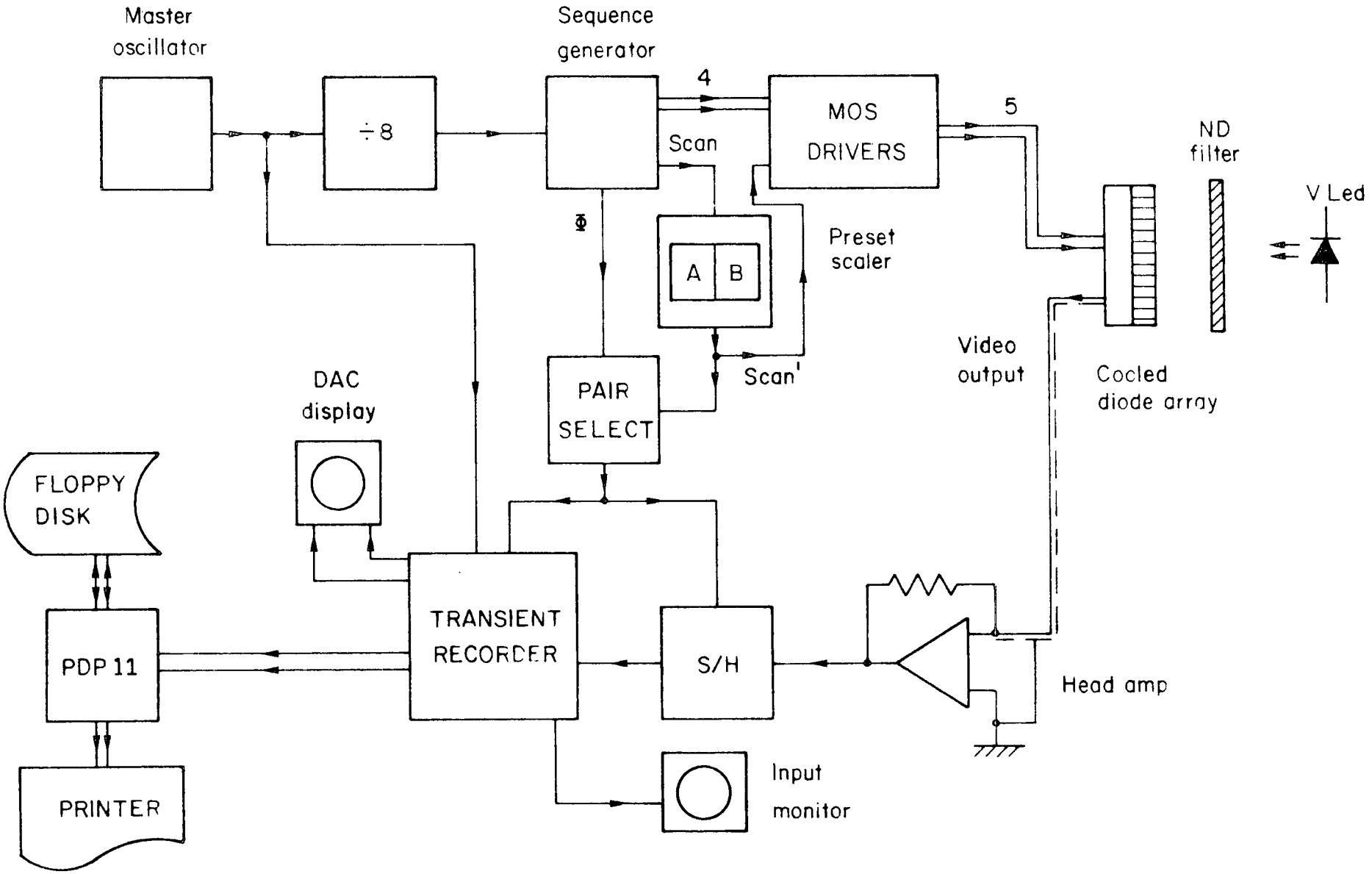
$$\text{Noise}^2\text{tot} = \alpha N + \sigma_o^2 + \sigma(s)^2 + \sigma(I)^2$$

where σ_o = readout noise independent of signal,

$\sigma(s)$ = readout noise dependent on integrated signal, and

$\sigma(I)$ = readout noise dependent on incident illumination.

Fig 3·15 NOISE MEASUREMENT SYSTEM



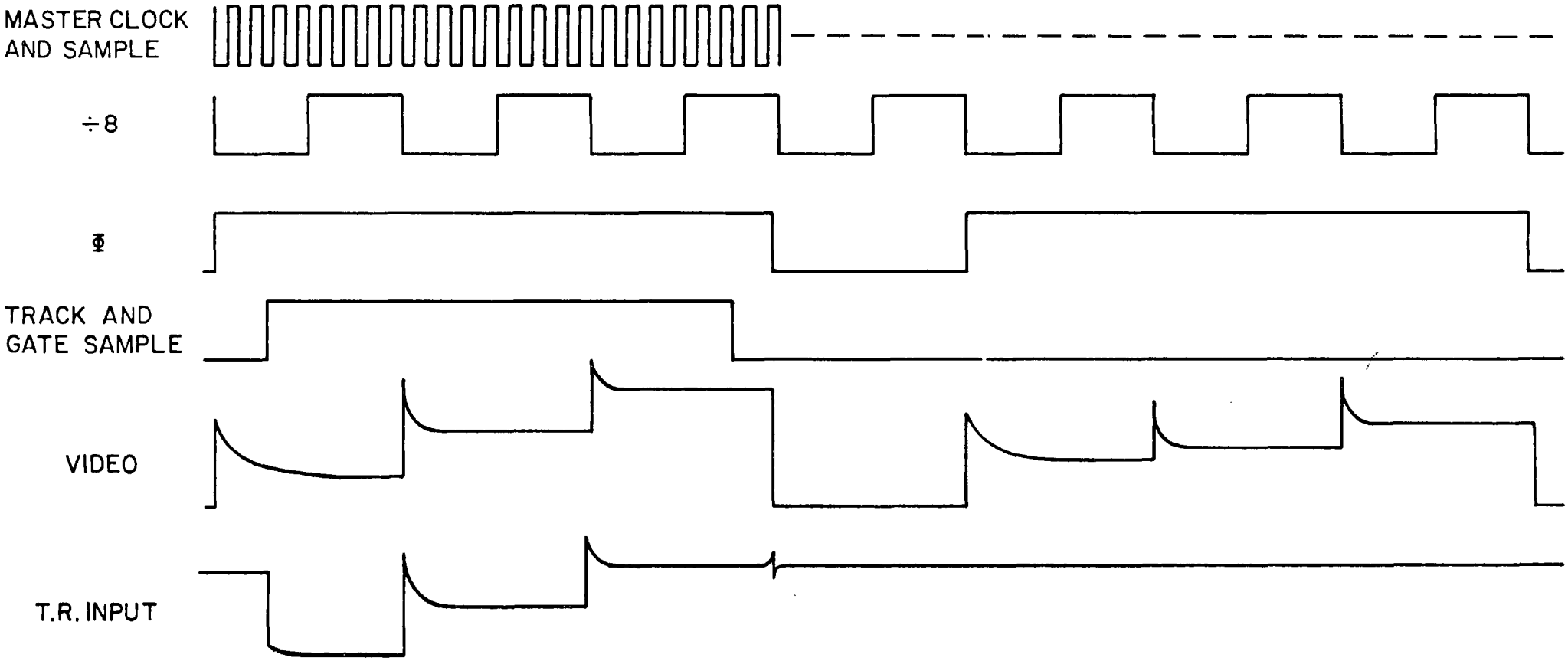
However as G is already approximately known, performing the experiment outlined above provides a useful check on the relative importance of the signal dependent terms of the readout noise. It also gives an upper limit to the value of G.

(e) Experimental Determination of Noise vs. Signal

A measurement has been made of the mean square noise voltage, V_n^2 as a function of the signal by using a special laboratory set up. The system is identical to that used for astronomical work from the array and driving system up to the output of the virtual earth amplifier. Sampling, conversion and storage of the signal are achieved by a DATAL-8 Transient Recorder. So the results obtained from this experiment do not directly apply to the full observing system although they should demonstrate a value obtainable from this detector. The difference between the two systems should lie mainly in the bandwidth - assuming that neither conversion system adds significantly to the noise. If it is true that reset noise dominates the contributions from the factors dependent on external system bandwidth in both cases, then the readout noise measured in this way should be very close to that of the full diode array system.

A system was devised to look at the signal from just one diode pair and to mask out all others so that the input stage of the transient recorder would only have to accommodate the signal excursion of the single pair of diodes. This signal may then be amplified to give the best possible conversion gain. Such a solution obviates the problems arising with the full array system as a result of its fixed pattern offsets. In this way the 10 bit transient recorder system provides a better resolution in terms of digitization step size than does the 12 bit data acquisition system which must look at all of the pairs of diodes. A diagram of the laboratory system is shown in Figure (3.15) and a timing diagram in Figure (3.16).

Fig 3·16 NOISE MEASUREMENT TIMING



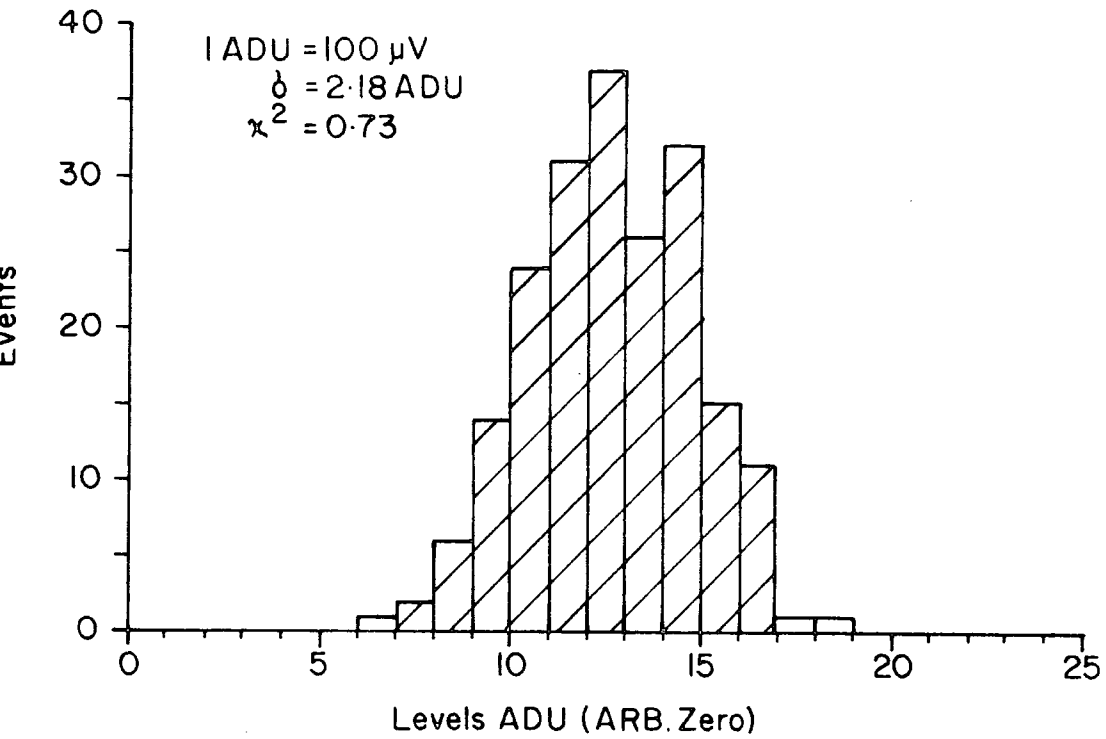
The diode pair to be investigated is switch selectable so that only the signal from this is allowed through the sample and hold module. As the recorder has a 4096 x 10 bit memory, when the signal from the diode pair of interest is sampled in about 20 places, it is possible to do 200 consecutive identical integrations and to store the 20 data points from the ith pair each time the array is scanned and recharged. In doing so we get a digital representation of the three signal levels from the pair in 200 similar experiments. Assuming that the incident light signal is constant during the time necessary to do this (i.e. 200 x integration time), the variation in output signal will be due to both system and signal noise. If it is Gaussian, the r.m.s. noise voltage may be calculated from the distribution of signal noise as plotting the distribution function of the data will verify. A typical distribution appears in Figure (3.17).

(f) Analysis and Results

(i) Several sets of data have been taken at the different intensities provided by using a stabilized LED light source and a set of Ilford neutral density filters. Typical results for square noise voltage as a function of signal are found in Figures (3.18(a)) and (3.18(b)). The analysis was done on a PDP 1103 into which sets of data were read via a CAMAC system - the basis of the experimental diode array system to be described in some detail later. The Fortran analysis programs used to calculate the mean, the variance and to do χ^2 tests in order to check that the data set is Gaussian were the work of G.R. Hopkinson. The operational procedure is described more fully in an internal report.[10] In the preliminary analysis only one data point from each level has been used. The reason for digitizing each level in several places is twofold:

(1) to provide a check that the sampling is taking place on the flat part of the signal rather than near any edges or spikes; and

Fig 3·17 TYPICAL EXPERIMENTAL NOISE DISTRIBUTION



(2) to allow some analysis of high frequency noise.

Despite the fact that this additional information has not yet been used for any quantitative results, it did prove valuable in identifying and removing some ripple problems encountered in the early stages. Having chosen one data point to represent each level, the double sample signal values may be calculated from their differences.

(ii) Results to date seem to indicate a value of G of $\sim$ 4 μV per electron with the usual external amplifier configuration employing a 47KΩ feedback resistor. If Equation (3.12) holds and the signal dependent readout noise is not significant, it suggests that the stray capacitance associated with the gate of T3 is less than the expected value of 0.1 pF. From Equation (3-11)

$$Co = \frac{Gsqe}{G} = 0.04 \text{ pF}$$

for

$G_S \sim 1$ for 47KΩ / feedback resistor,

$G \sim 4$ μV / detected electron, and

qe = electronic charge.

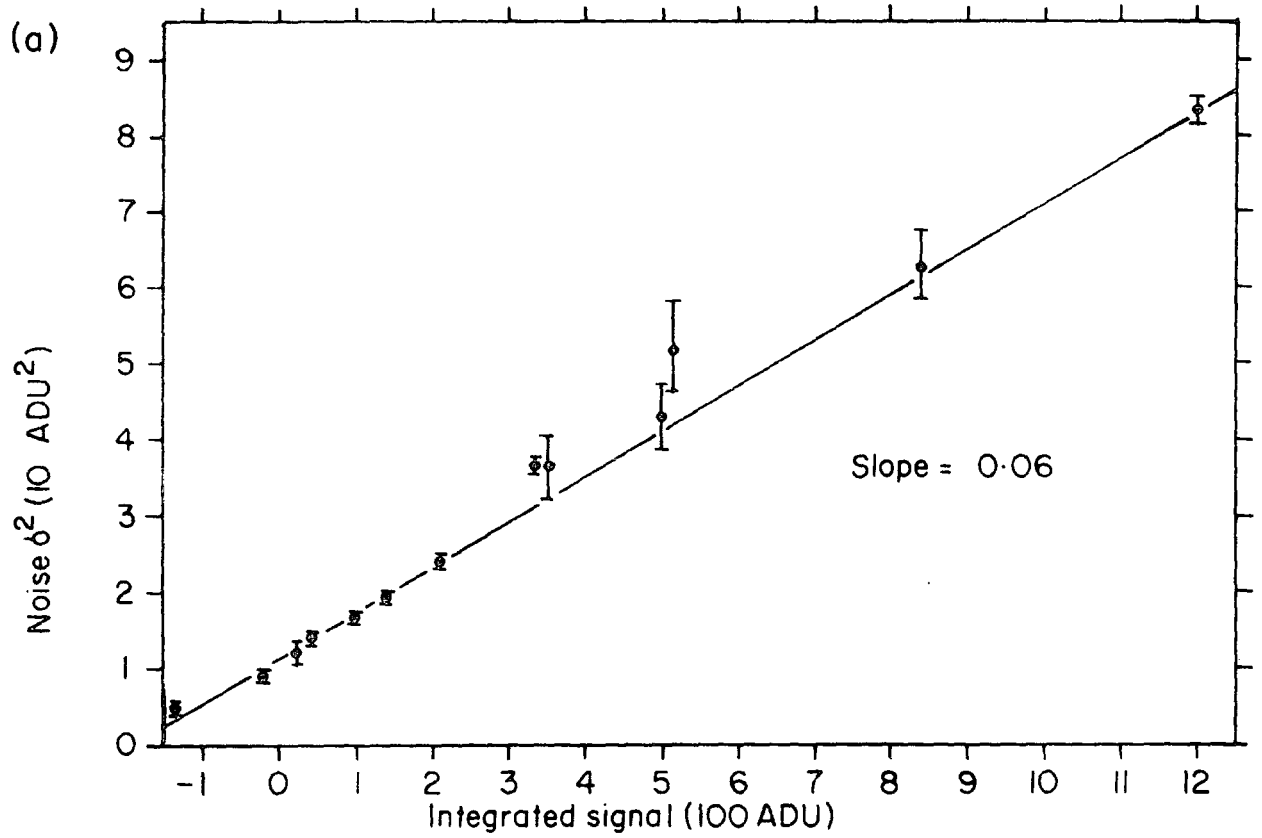
This being the case, the RQE from Equation (3.11) must also be smaller than expected by a similar factor which would mean that it is only $\sim$ 20%. The latter value is feasible seeing that the detector was not built specifically for optical applications. However the expected system readout noise could conceivably be much improved with a maximum value for the reset noise of

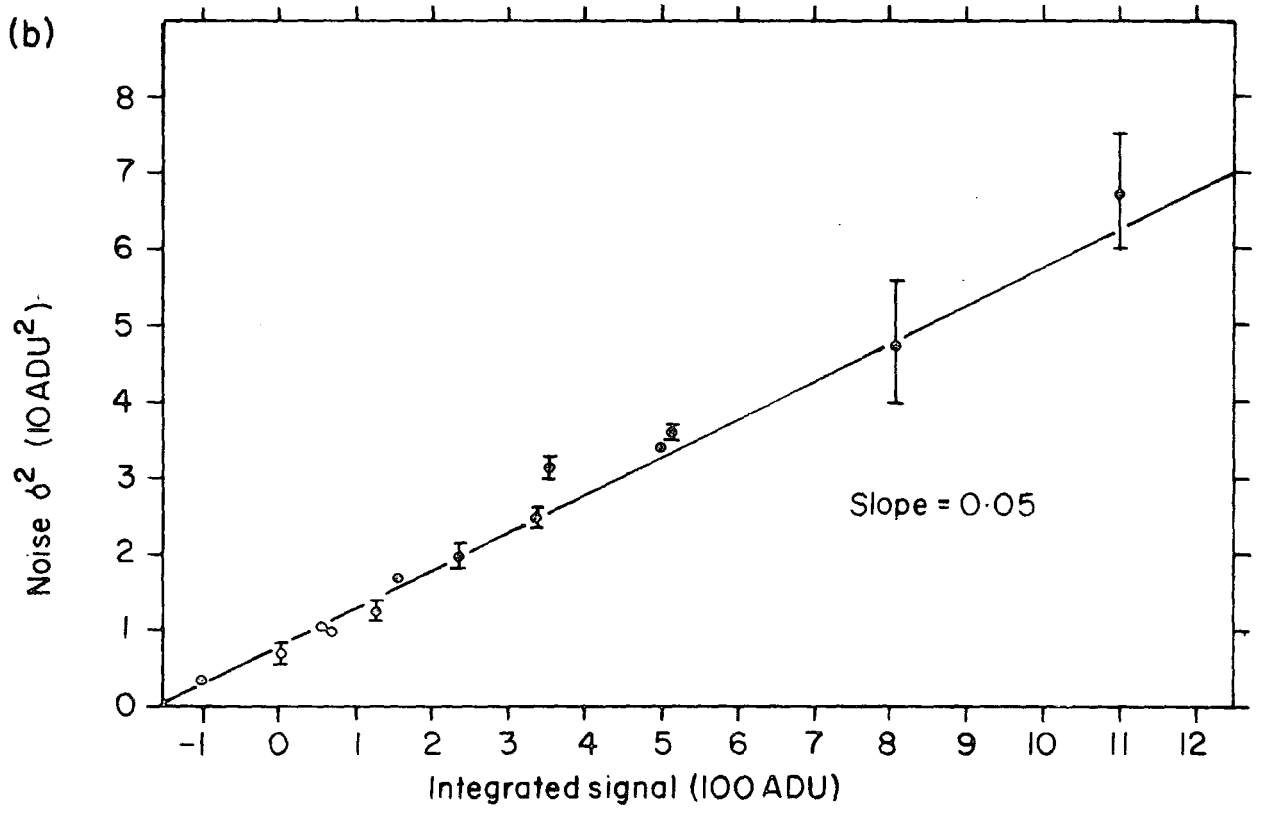
$$\text{MAX RESET NOISE } 327 \sqrt{0.04} = 65 \text{ electrons.}$$

N.B. For a double sampled value this is multiplied by $\sqrt{2}$ to give 92 electrons.

The observed value of root mean square noise voltage (using double sampling) at zero incident signal can be as low as 160 μV which is, if the responsivity is actually 4 μV per electron, equivalent to a noise

Fig 3·18 NOISE VS. INTEGRATED SIGNAL





of only 40 electrons. This experimental finding we note to be lower than that expected theoretically.

(7) Summary of Detector Parameters

The performance of this detector is characterised by its RQE, the noise equivalent charge at zero signal and by the responsivity, G. As yet a consistent set of accurate values has not been found for these parameters because the value of Co, the capacitance associated with the gate of T3 is not well known. Table 3.2 shows these parameters for a range of values of Co. For comparison purposes, double sample reset noise values are also shown.

Table 3.2

Co pF	G Volt/Electron	RQE %	Maximum Theoretical Noise Reset Double Sampling	d.s. electrons σ_o
0.05 pF	3.2	20	104	50
0.1 pF	1.6	40	147	100
0.2 pF	0.8	80	208	200

It should be noted that in the possible solutions outlined in Table 3.2 the number of incident photons required to give a SNR of 1 is given approximately by σ_o/α and, consequently, is almost the same in all three cases. With larger integrated signals where shot noise dominates, the solution providing the largest RQE should give the best performance in spite of having the worst readout noise. This better performance is highlighted when considering the behaviour of the DQE which is a signal dependent figure of merit for a detector. Expressed in directly measurable quantities this is

$$DQE = \frac{Vs\ \alpha}{Vs + V_{no}}^2 \quad G$$

assuming that signal dependent readout noise is negligible

where V_S = signal volts out,

V_{no} = root mean square noise voltage at zero signal,

G = output voltage per detected photon, and

α = RQE.

(a) At small signals

$$DQE \longrightarrow \frac{Vs\ \alpha\ G}{V_{n\,o}^{2}} = \frac{Vs\ G}{V_{no}^{2}}$$

where G = output per incident photon and

Vn^2o is independent of Co.

(b) At large signals

$$DQE \longrightarrow \alpha\ .$$

However, if the figure of 4 μV per detected photon for G (found by measuring noise as a function of signal) is very much too high, we may assume that there is a significant source of signal dependent noise. If located within the device it will degrade performance at large signals so that, again, the theory and the practice are somewhat divergent.

REFERENCES

1. Van Der Ziel, A. (1976). Noise in Measurements (Wiley), p.124.

2. Humrich, A. (1978). Internal Report.

3. Campbell, A.W., Hedge, A.R., Hopkinson, G.R. and Humrich, A. (1978). Internal Report. 4 and 5.

4. Richman, P. (1973). MOS Field - Effect Transistors and Integrated Circuits. (Wiley), p.125.

5. Vogt, S.S., Tull, R.G. and Kelton, P. (1978). "Self-Scanned Photodiode Array: High Performance operation in High Dispersion Astronomical Spectroscopy." Appl. Opt. Vol.17, pp.574-592.

6. Campbell et al. Op.Cit.

7. Campbell, A.W. and Hopkinson, G.R. (1980), Internal Report. 1.

8. Richman, P. Op.Cit. p.90.

9. Campbell, A.W., Hedge, A.R., Hopkinson, G.R. and Humrich, A. (1978). Internal Report. 4 and 5.

10. Campbell, A.W., and Hopkinson, G.R. (1980). Internal Report 1.

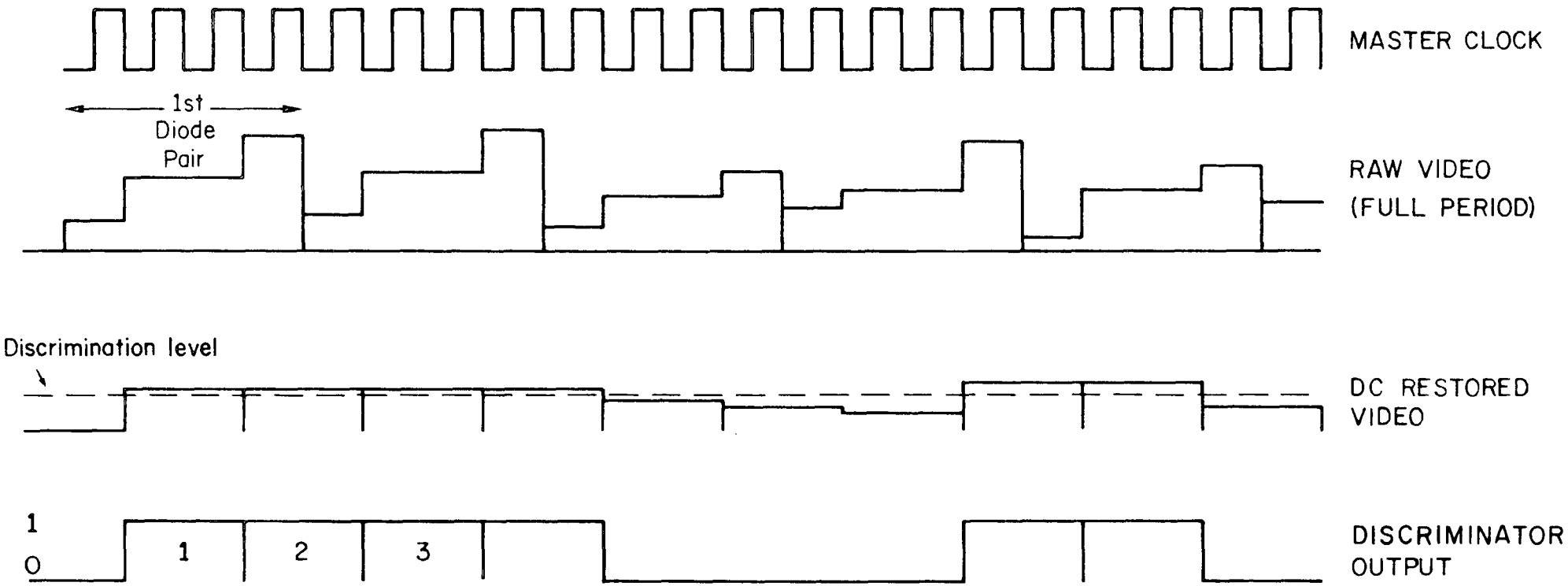
CHAPTER 4

SIGNAL PROCESSING AND DIGITIZATION SYSTEM

1. Introduction

When driven with the system described in Chapter 2, the diode array produces a characteristic boxcar waveform at the output of the external virtual earth amplifier. With the timing chosen, this waveform consists of a serial train of 128 sets of four levels containing the information from consecutive diode pairs. Since the first level of each set is a zero current reference corresponding to the dead time between clock phases, it is usually ignored except when laboratory checks are made on system drift. To recover the video information from the two elements of the pair, the change in output at the transitions level 2 to level 3 and level 3 to level 4 must be found. These effective subtractions may be accomplished either with analogue circuitry or numerically after digitization. Preliminary designs for circuits to be employed in the original application of electron counting make use of the former technique to produce a d.c. restored boxcar video train with amplitudes proportional to the individual integrated signals on consecutive elements. This video signal was then to be discriminated with a single threshold to decide whether or not electrons had hit particular elements. Figure (4.1) illustrates this operation which is similar to that used for discriminating events in nuclear particle detectors. The problem with this type of system is that the data is susceptible to drifts and 1/f noise in any stages following the d.c. restoration, whereas in the conceptually simpler approach of digitizing both the signal and reset levels, there is less electronics between the double sample process and the digitization stage. If a discrimination (or 1 bit conversion) is all that is required the analogue technique is easier to mechanise, but for our application, digitizing to higher accuracy (e.g. 12 bits), the choice is less obvious. D.C.

Fig 4-1 SIGNAL PROCESSING FOR EVENT COUNTING



restoration offers the following apparent advantages:

(i) shorter time between the two samples, because it is not necessary to perform a digitization in this interval, thereby optimizing the noise reduction ability of double sampling;

(ii) removal of the fixed pattern noise[1] and separation of the two signals within each diode pair before digitization, giving more efficient use of the ADC's resolution; and

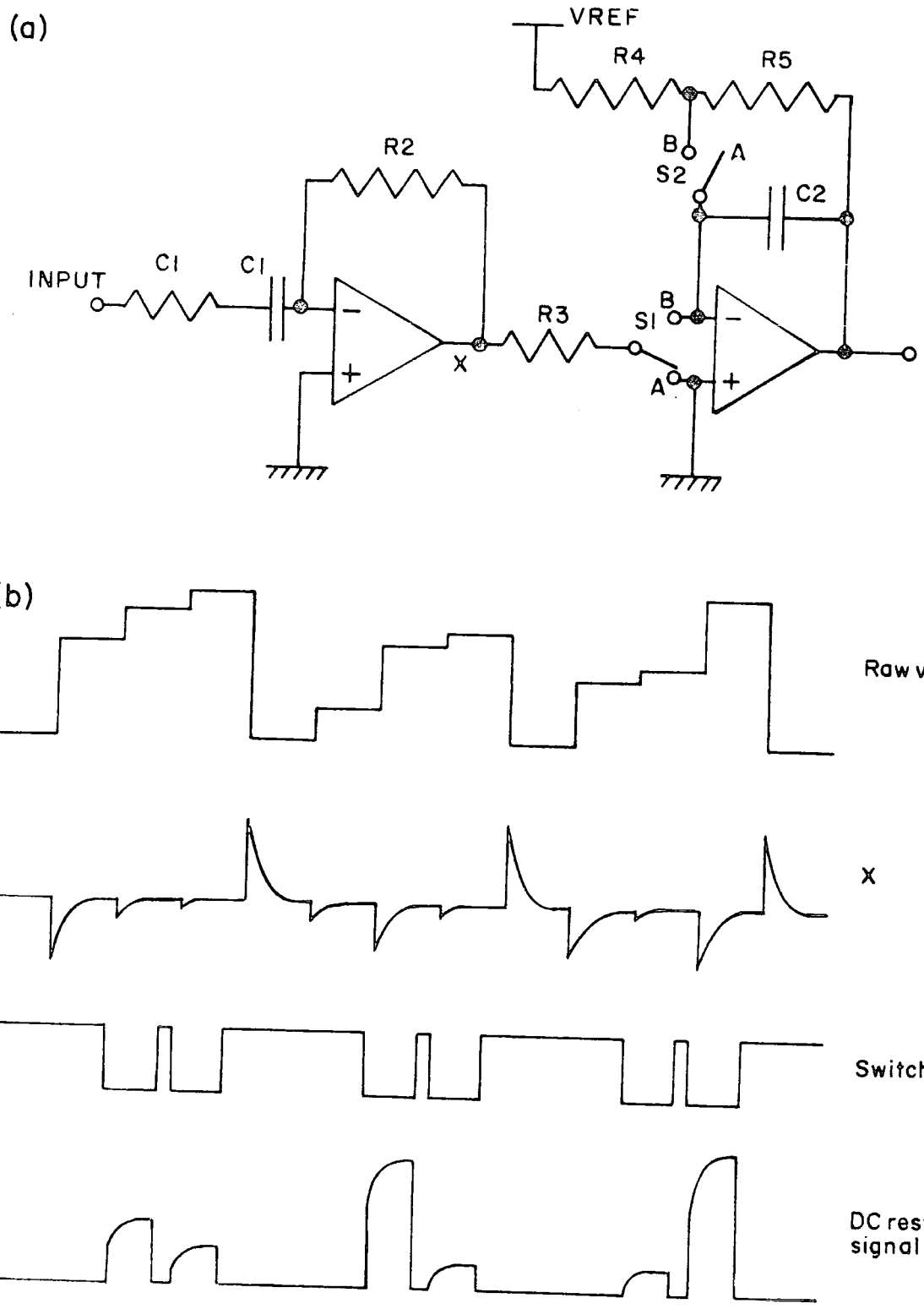
(iii) only one digitization is done to obtain each datum - thus a reduction in noise added by digitization errors ($\pm \frac{1}{2}$ LSB). A d.c. restoration circuit is shown in Figure (4.2a) with operational waveforms in Figure (4.2b). However, this approach is more difficult to implement than that of double digitization chosen for the development system because of its simplicity and the more complete picture it gives of array operation.

As already outlined, correlated double sampling offers the benefit of suppression of low frequency, 1/f type noise, particularly that associated with the on-chip MOSTs which may not have very good inherent performance in this respect. The lower limit to the system bandwidth is determined by the interval between the two samples which, in the present system, is inversely related to the scanning frequency. On the other hand, the effect of system thermal noise generally increases with scanning frequency because the bandwidth of the head amplifier must be made large enough to accommodate the video output rate. Therefore the design cannot be optimized without detailed knowledge of the device power spectrum.

2(a) The Present System

A diagram of the present digitization electronics is provided by Figure (4.3); operational timings follow in Figure (4.4). The necessary control pulses are produced by a circuit which is currently installed as part of the TTL sequence generator described in Chapter 2. Such

Fig 4·2 DC RESTORATION



amalgamation was to facilitate the required synchronization of the operation of these two circuits. Figure (4.5a) is a photograph of the complete timing circuit built as a CAMAC module; Figure (4.5b) shows the digitization system.

2(b) Timing Circuit

Several circuits have been devised to create the four control pulses namely, sample, convert, load memory and increment memory address, while still allowing some flexibility in the exact timings of these. A typical realization is shown in Figure (4.6) where a master clock drives a four bit counter thereby generating a cycle of 16 states from which the control pulses are decoded. On board switches enable the output pulses to be preset to change state at certain points within the cycle. A symmetrical, divide by 16, output supplies the input clock for the array clock pulse sequence generator and this organization automatically maintains synchronism of the conversion cycle with the video output waveform. In this way the ability of the operator to set the data rate by adjusting a single master clock frequency is retained, although the maximum acceptable rate will be limited by the speed of the conversion electronics. Once an optimum data rate and particular digitization hardware are decided upon it is feasible to construct a very much simpler timing circuit, but in the development stages the above approach was found very useful. The only critical timing is that of the sample pulse as variations in the exact moment at which this occurs may introduce noise if there is any slope or ripple in the raw video being sampled. Following the sampling operation, the remaining control pulses are set to come at intervals determined by the characteristics of the ADC and memory.

2(c) Head Amplifier

The external virtual earth amplifier performs several functions: (i) it provides a summing junction for the output currents from the two

Fig 4·3 DIGITIZATION SYSTEM

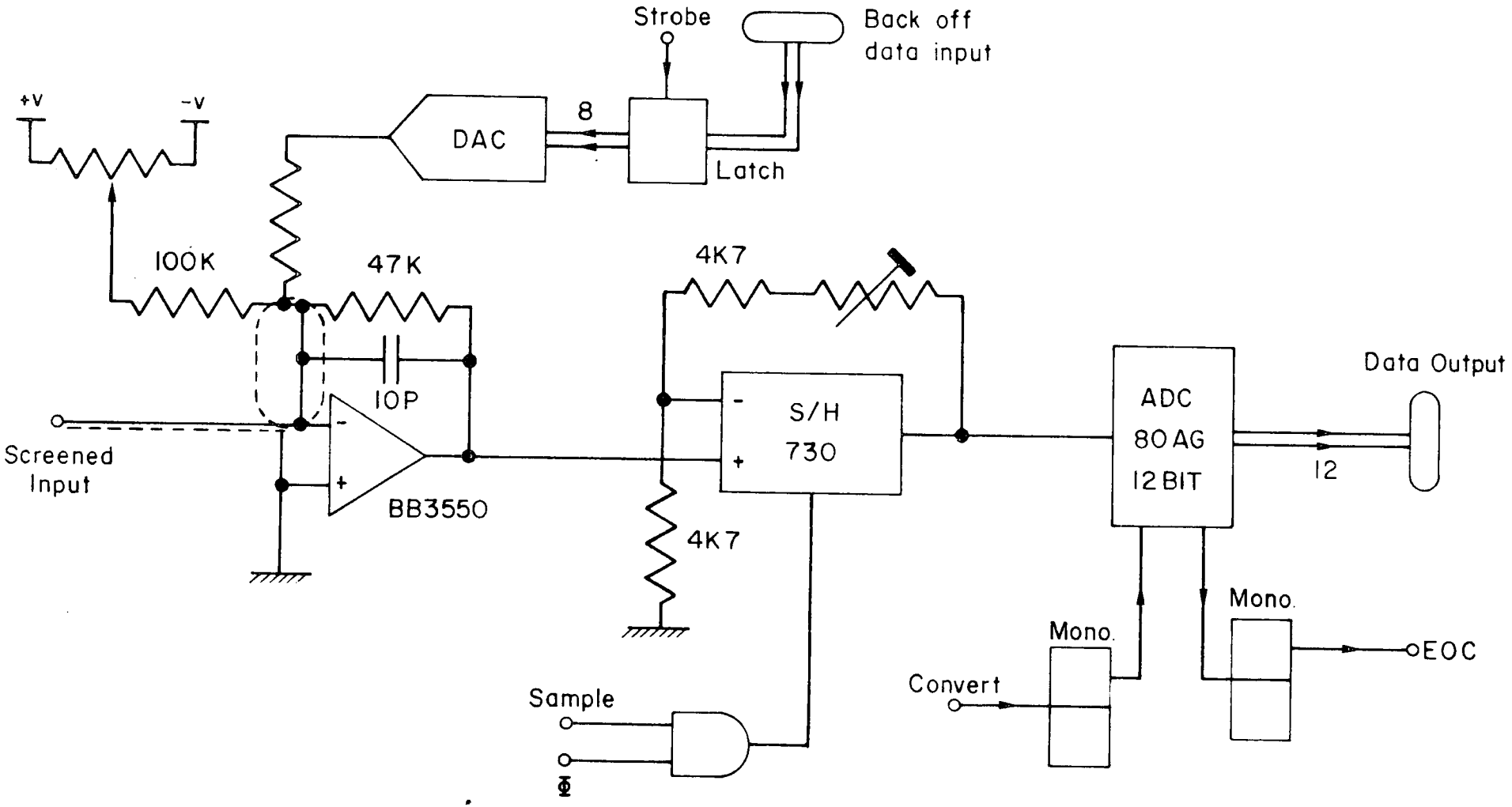
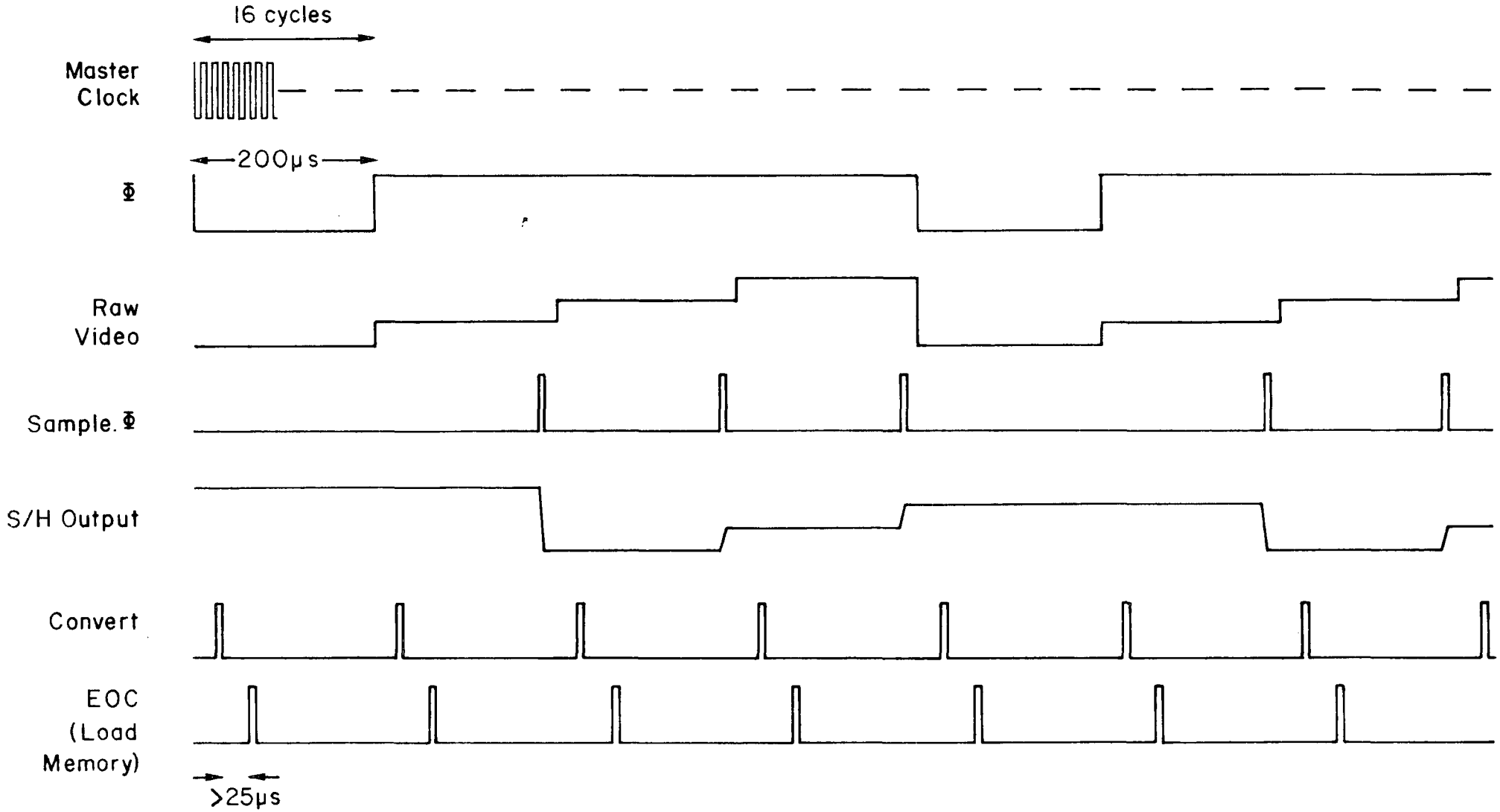


Fig 4·4 DIGITIZATION TIMING



elements of each pair and converts these to a voltage in the range measurable by the ADC; (ii) it allows a d.c. offset to be added to the video signal; and (iii) it enables any desired filtering, such as a high frequency cut-off in order to reduce the noise bandwidth, to be implemented. Simple analysis of the head amplifier (shown in Figure (4.3)) yields that the current to voltage conversion factor is given by the resistance Rf such that

$$\Delta V_{OUT} = -R_f (\Delta Io + \Delta I_E) \quad , \qquad (4.1)$$

where ΔV_{OUT} = the change in output voltage, and

$\Delta I_0, \Delta I_E$ = the change in output current of the odd and even elements.

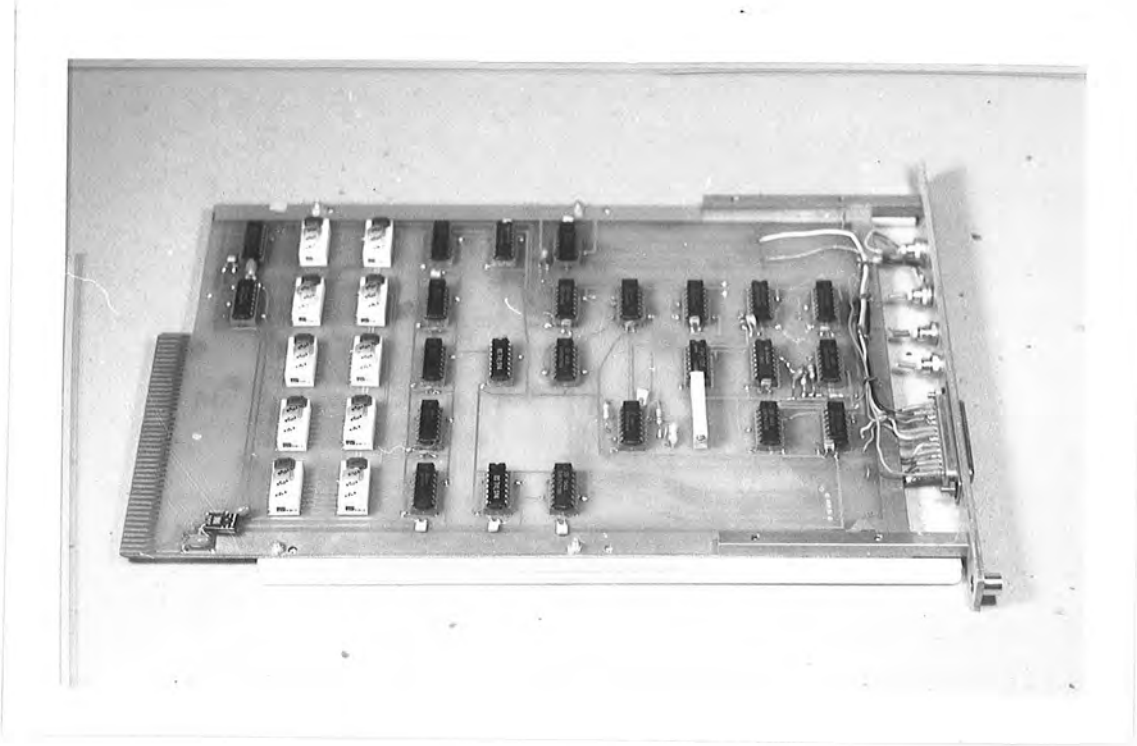
Additionally, the output voltage may be offset by an amount

$$V_{OFF} = V_{VAR} \frac{Rf}{R_{OFF}} \quad , \text{ and the}$$

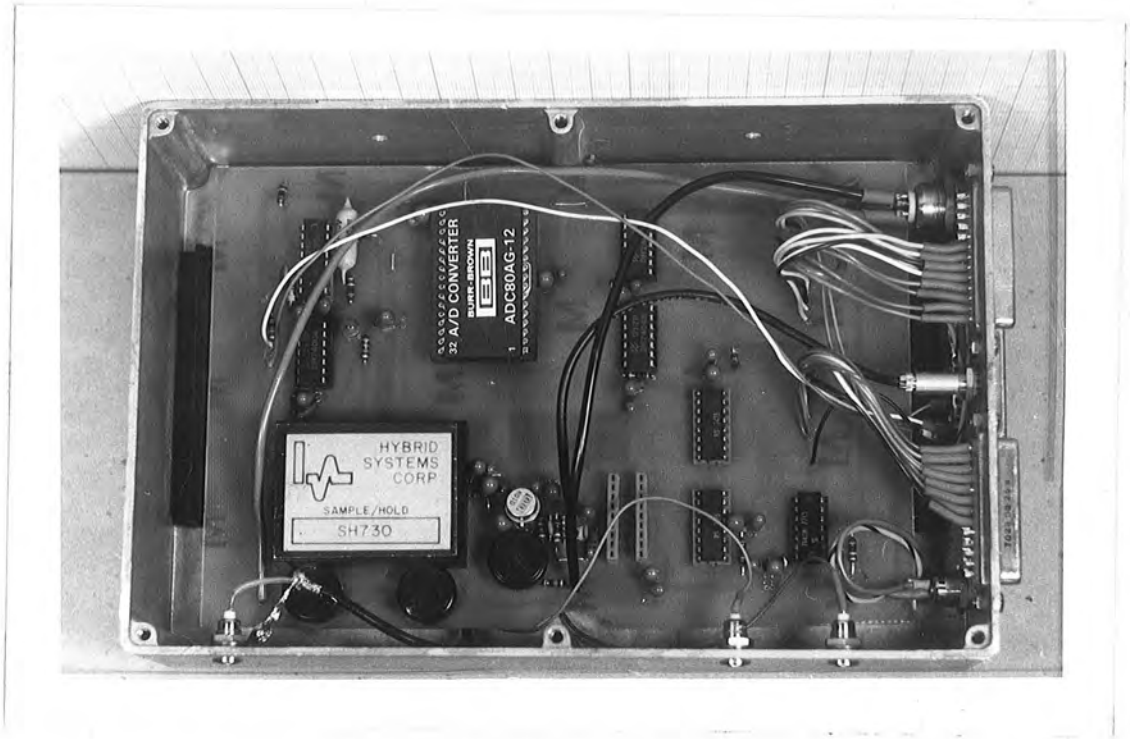
a.c. response of the circuit is characterized by a high frequency cut-off with a half power frequency,[2]

$$f_o = \frac{1}{2\pi R_f C_f} \qquad (4.2)$$

A Burr-Brown 3550J FET operational amplifier was selected for its fast settling low noise performance, although, at the currently used data rate (approximately 5KHz), a more modest unit would probably suffice. This device operates off standard $\pm$ 15 volt. supply rails and is capable of giving an undistorted output swing of $\pm$ 10 volt. As explained in Chapter 2, R_f is chosen to be as large as possible without driving the amplifier output into saturation. Practical values of R_f are usually around 47KΩ, giving a swing of typically 10 volts between the zero current dead spaces and the reset levels, leaving room for the output to accommodate the switching spikes. The offset network has to be adjusted to place the signal and reset levels more or less equally spaced about zero voltage , because the ADC is wired to expect a bipolar signal.



(a) Timing module



(b) Digitization module

Fig 4·5

For simplicity, no further offset adjustments are included in the circuit. Offset networks are made with high stability components and, if possible, decoupled in order to minimize any noise and drifts that might be added here.

2(d) Sample and Hold

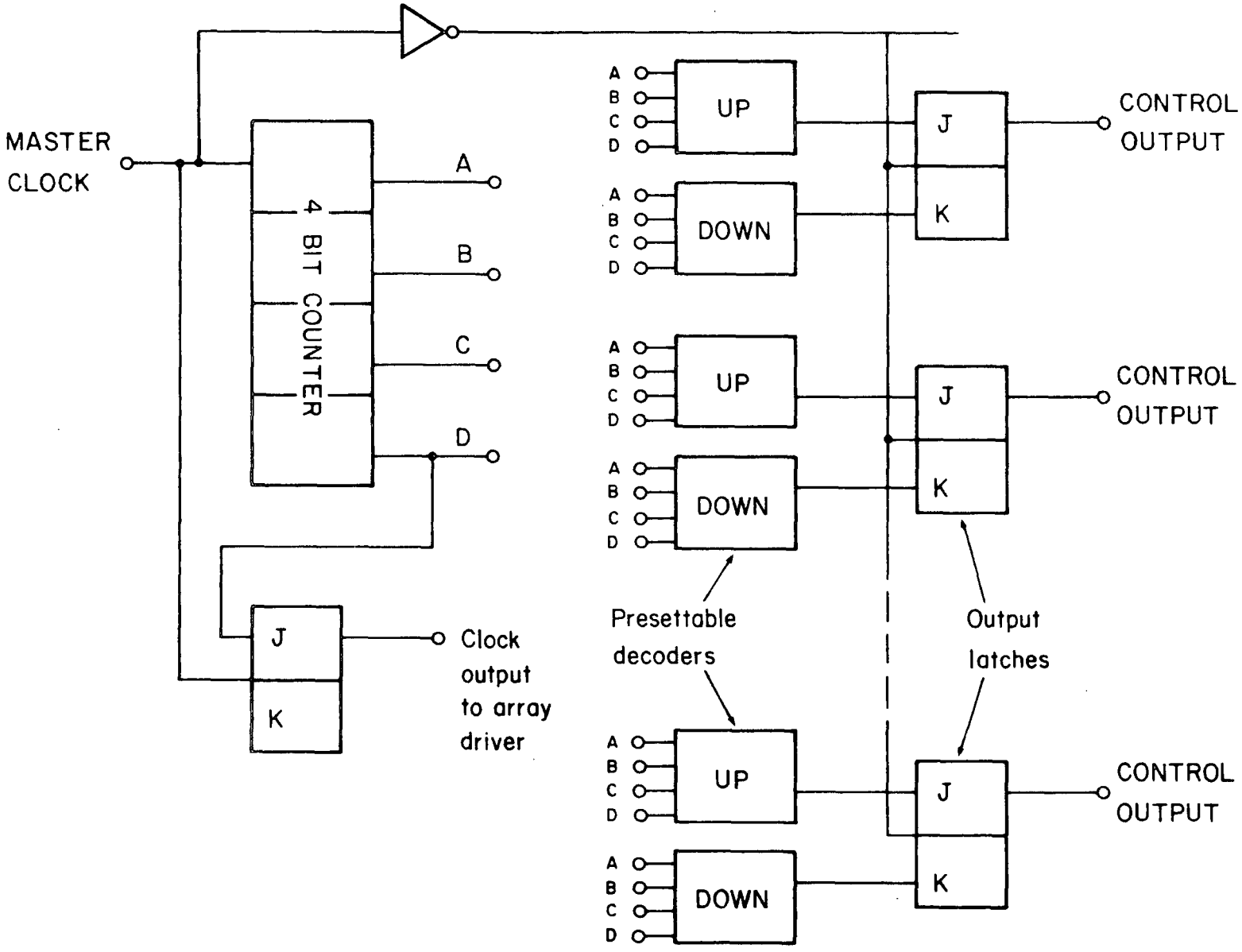
The sample and hold circuit cleans up the raw video by removing both the switching spikes and, more importantly, the pedestals due to the dead spaces, before the video signal reaches the ADC. Better use can then be made of the ADC resolution without saturating the input stage of the converter with large voltage excursions which contain no necessary information. The present sample and hold module, a Hybrid Systems 730, may be connected as an operational amplifier allowing any extra required gain to be introduced. The optimum overall system gain depends on the total system noise. Noise added by uncertainty in the digitization process (i.e. $\pm \frac{1}{2}$ LSB) should be kept small compared to the noise, as was discussed in Chapter 2. Since the data values are calculated from the difference of two ADC values which both have errors with square distribution functions, the total digitization error is $\pm$ 1 LSB, and for this source of error to be negligible the system gain must satisfy the condition

$$1 \text{ ADU or } (1 \text{ LSB}) \ll A \ \sigma_{TOT} \text{ (ADU)}, \qquad (4.3)$$

where A = the overall system responsivity expressed in ADC units per detected photon, and

σ_{TOT} = the total readout noise (assumed to be signal independent), expressed in electrons.

However, adjusting the gain such that the system noise is very much greater than the digitization step limits the range of signals that may be looked at with an ADC of given precision (number of bits). Considering a single element, the usable dynamic range of the system will be, neglecting digitization errors,


MASTER
CLOCK
4 BIT COUNTER
A
B
C
D
J
K
Clock
output
to array
driver
A
B
C
D
UP
DOWN
J
K
CONTROL
OUTPUT
Presettable
decoders
Output
latches

$$\text{Dynamic Range (single element pair)} = \frac{2^n}{A\sigma_{TOT}} \qquad (4.4)$$

where n = the number of bits to which the conversion is done.

Thus if $A\sigma_{TOT}$ = K digitization steps (ADU),

$$\text{Dynamic Range (single diode)} = \frac{2^n}{2K}\ , \qquad (4.4(a))$$

since pairing the diodes halves the output range available to each if the double digitization technique is used. In the present system where n = 12, if A is set to give a readout noise of about 2ADU r.m.s. (i.e. K = 2), the maximum usable dynamic range of a single element is $\sim 10^3$, which is still more than adequate.

When an array of elements is considered, the choice of system gain, A is influenced by the fixed pattern noise. Ignoring the responsivity variation contribution to the fixed pattern noise and considering the phenomenom simply as a spread in the reset levels of the element pairs; it is possible to describe the effect as an equivalent signal charge in much the same way as random noise:

$$\text{Fixed pattern variation (ADU)} = A\,\varepsilon_{FPN}\ ,$$

where ε_{FPN} = peak to peak variation in reset levels expressed as an equivalent signal charge.

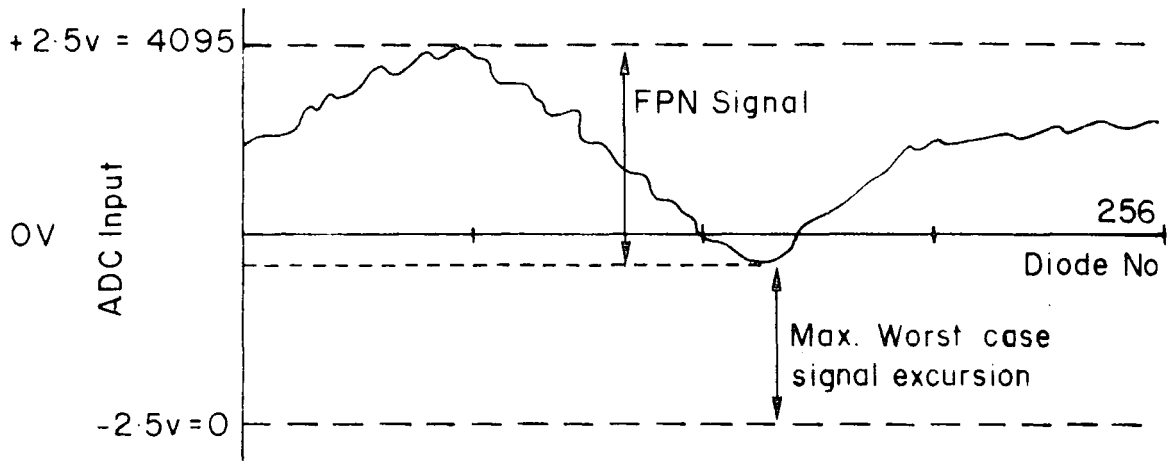
Since the fixed pattern must be accommodated in the ADC input range, much of the available dynamic range as calculated for a single diode pair is effectively used up. Although some elements will still have the performance described by Equation (4.4), the worst-case dynamic range as illustrated by Figure (4.7) is now given by

$$\text{Worst case dynamic range, WCDR} = \frac{2^n - A\varepsilon_{FPN}}{2A\sigma_{TOT}}\ , \qquad (4.5)$$

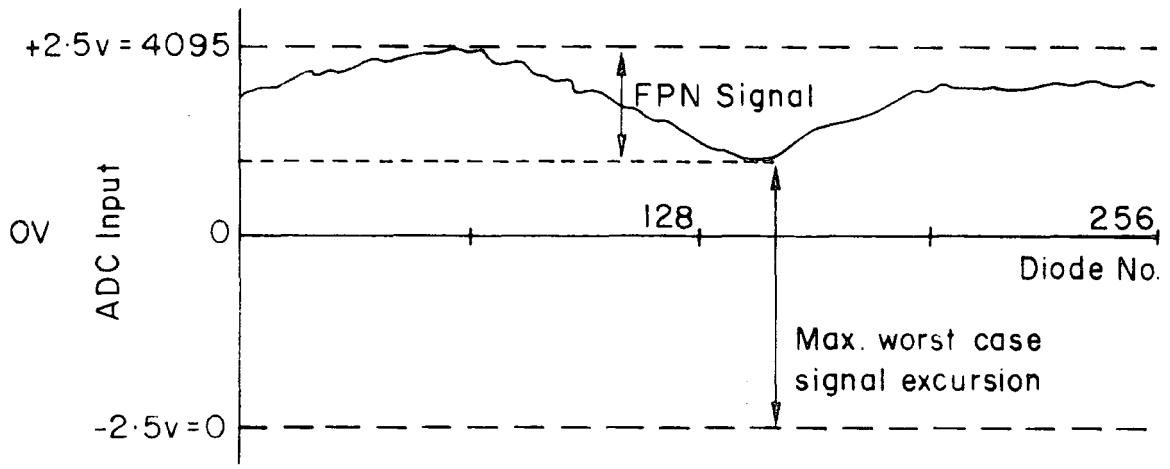
assuming negligible digitization error.

The problem here is that not only does increasing the system gain, A reduce the dynamic range, but, in practice, the equivalent fixed pattern signal

Fig 4·7 DYNAMIC RANGE



(a) Gain = A



(b) Gain = $\frac{A}{2}$

may be so large that the condition

$$A\,\sigma_{TOT} > 1 \text{ ADU}$$

cannot be reached before the WCDR becomes zero:

$$A\,\varepsilon_{FPN} \gg 2^n.$$

For the present system where n = 12 (i.e. a 12 bit ADC is used) this suggests that the equivalent fixed pattern signal must be less than a certain maximum

$$\varepsilon_{FPN} < 2^n\,\sigma_{TOT} = 4000\;\sigma_{TOT},$$

in order for anything approaching the optimum system gain, A to be used without some of the reset levels going out of range. As has been noted, working at less than optimum gain means that the low noise potential of the array is wasted and that the dominant system error is the inherent uncertainty of the digitization process.

To minimize the complications described above it is usual to select arrays for observational use (i.e. actual fieldwork) on the basis of low equivalent fixed pattern noise, although, as only a handful of Plessey arrays exist, the choice is limited. In any case, if the double digitization technique is used, the presence of fixed pattern signal results in either loss of dynamic range or degradation of the overall system random noise performance. This disadvantage is a strong argument in favour of the d.c. restoration approach mentioned earlier. However, the problem may be partially overcome, in the existing system, with the addition of a hybrid digital-analogue back-off unit to be described in a later section.

Returning to the more fundamental properties of the Sample and Hold, the unit chosen (SH 730) has a specified acquisition time, for a 10V change in output, of 1μs, to achieve an accuracy of 1 part in 10^4. Optionally, the acquisition time may be increased by the addition of an external capacitor if a lower droop rate is required. Since the present ADC has a conversion time of 25μs, the fall in output signal from the

S/H module in this time is

$$V = 5mV/mS \times 25\ \mu s = 125\ \mu V.$$

This fall is equivalent to 0.1 ADU (f.s.r. = 5V for a 12 bit system) and is therefore not a significant source of error. Nevertheless, increasing the holding capacitor value may still be worthwhile in order to decrease the system small signal bandwidth, if speed is not essential.

The maximum system throughput rate is determined by the sum of the S/H acquisition time and the conversion time:

$$\text{Max Data Rate, } f_D = (\tau_A + \tau_C)^{-1} \quad ,$$

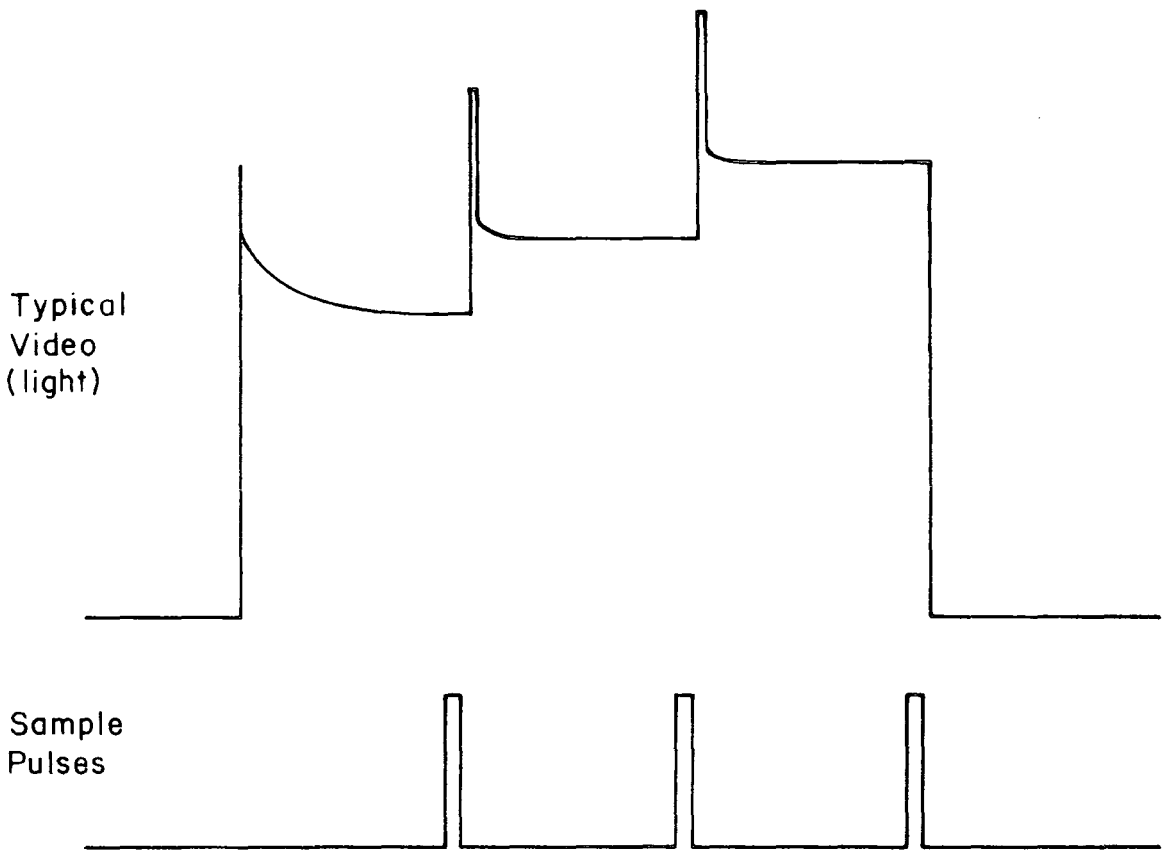
where τ_A = the acquisition time, and

τ_C = the conversion time.

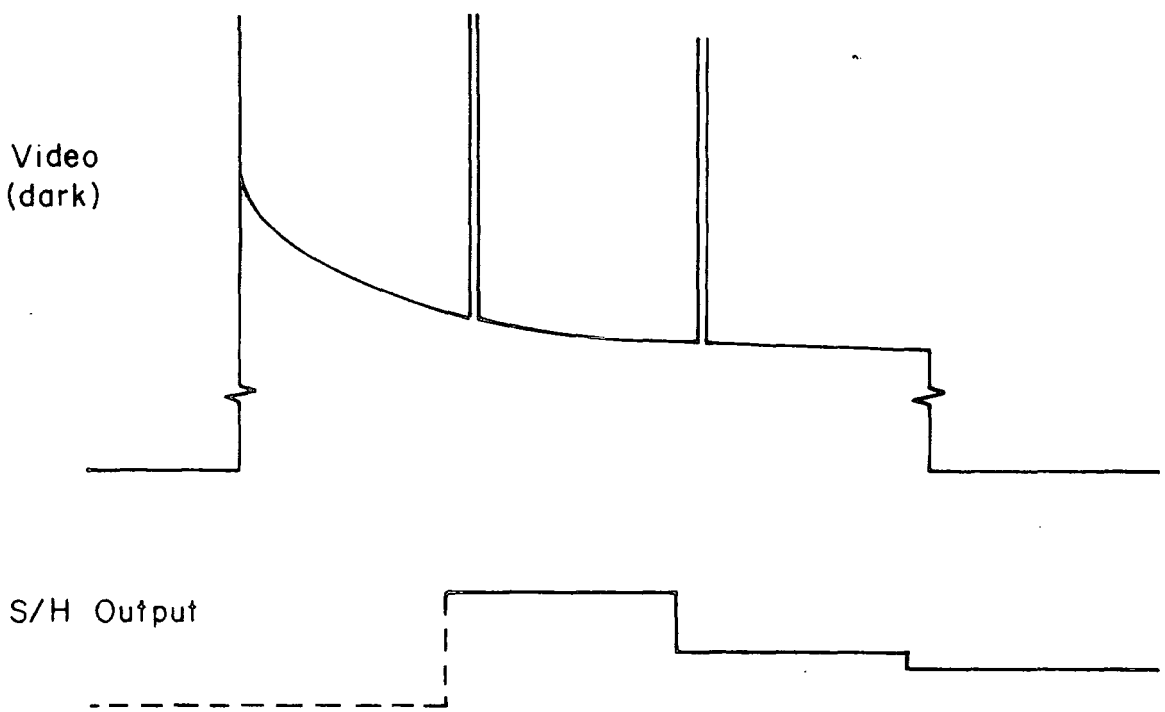
With the double digitization technique, the effectiveness of the double sample at reducing lower frequency noise depends upon f_D, which imposes a lower limit to the system noise bandwidth. From this it can be seen that efforts to reduce the upper limit of the system bandwidth by increasing the hold capacitance will also result in an increased τ_A and a consequent reduction in f_D. In practice, the scanning frequency was chosen more for its convenience of real-time display of the video output on a monitoring oscilloscope than because of any serious attempt to minimize the amplifier noise. The present rate, with a frame time of 10^{-1} sec., gives an acceptably flicker-free trace if the array is scanned continuously. This monitoring is useful for setting up the detector (i.e. adjusting the control voltages, setting up the amplifier offset and possibly focusing of associated optical systems).

As mentioned previously in connection with the generation of the samole pulse, noise may be introduced by any uncertainty or jitter in the effective instant that the sample is taken if the input signal is time varying. Figure (4.8(a)) shows a typical video signal from a pair of diodes, together with the usual positions of the sample pulses. The sampling is done as long as is reasonably possible after the switching

Fig 4·8 TYPICAL VIDEO WAVEFORM



(a)



(b)

transitions, in places where the signal is most flat and free from ringing. Nevertheless the video signal shows an overall fall off from left to right, with the result that digitizing the dark video signal of Figure (4.8b) yields apparent negative signal values when the subtractions are made. Further, the effect is worse on all odd diodes. Thus any sample jitter, particularly in the sample taken on the first video level, will produce a noise effect. Sample and Hold devices have an inherent uncertainty in the delay between the moment at which an external sample command is applied and the effective instant of the sample. There exist many different definitions relating to such Sample and Hold parameters which are discussed in a paper by Tewksbury et al.[3] The uncertainty in the time taken for the device to make the transition from Hold to Sample mode is called the Aperture Uncertainty and is quoted by the manufacturers of the SH 730 as 5nS., making this a negligible source of error since the video level slope is, at worst, only a few ADU per microsecond (working at typical gains).

2(e) Analogue to Digital Converter

From Equation (4.4) it follows that to obtain an accuracy of about 1% on a single element detector, an ADC with a resolution of at least

$$n = \text{Log}_2 (100\ A\ \sigma_{TOT}) \text{ bits}$$

is required (again neglecting quantization errors). Setting the system gain, A to give an r.m.s. output noise of, say, 2 ADU yields a nearest greater integer value of 8 for n. (A discussion of coding errors as a function of noise-to-code width ratio can be found in a paper by Gordon.[4]) When considering the more realistic situation of a detector array, the spatial noise must be taken into account together with the intrascenic dynamic range of the image. For example, to look at a spectrum containing features of interest which differ in relative intensity by a factor of ten, sufficient resolution is required to give the necessary

accuracy of measurement on the weaker features, if comparisons of intensities are to be made from the data of one exposure. Thus, if an accuracy of 1% is needed on signals of one tenth of the maximum worst-case value, an overall dynamic range (worst-case) of 1000 is necessary. Recalling from Chapter 3 that the inherent dynamic range of the detector is much greater than this, the limitation is still the choice of ADC. From Equation (4.5), n is, in the above case given by

$$n = \text{Log}_2 \left[2000 \quad A \, \sigma_{TOT} + A \, \varepsilon_{FPN} \right] ,$$

where A = system gain in ADU per electron,

σ_{TOT} = system readout noise in electron, and

ε_{FPN} = fixed pattern noise in electrons.

Ideally, if $A \, \sigma_{TOT}$ is again set to about 2ADU and if (as is commonly found with the Plessey array) $\varepsilon_{FPN} \simeq 4000 \, \sigma_{TOT}$, the nearest integer value of n required is 13. However, as previously indicated, for observing brighter objects it is possible to decrease the system gain, A, thereby reducing the effect of spatial noise but, at the same time, making the digitization error ($\pm$ 1 LSB) the dominant noise source. Doing this, the number of bits to give a worst-case dynamic range of 1000 is now

$$n = \text{Log}_2 \left[2000 + A \, \varepsilon_{FPN} \right] .$$

Just how much the system gain may be reduced depends upon the brightness of the object and upon the time available to do the integration. Even with negligible spatial noise, the minimum allowable resolution is 11 bits. The foregoing, although somewhat hypothetical, illustrates the effects of the various parameters, readout noise, system gain, fixed pattern noise and intrascenic dynamic range.

In the initial development stages, an 8 bit ADC was used in the data acquisition system. This device was quite adequate for most purposes because high conversion gains were not used and the information from the pairs was not separated (i.e. the detector was treated as 128 big diodes). When the full 256 diode system was built, a 12 bit device was obtained for

improved performance. The conversion time of this 12 bit ADC, a Burr-Brown ADC 80, is 25μs, which is well within the 200μs available to do the conversion at the present 5KHz data rate.

2(f) The Buffer Memory

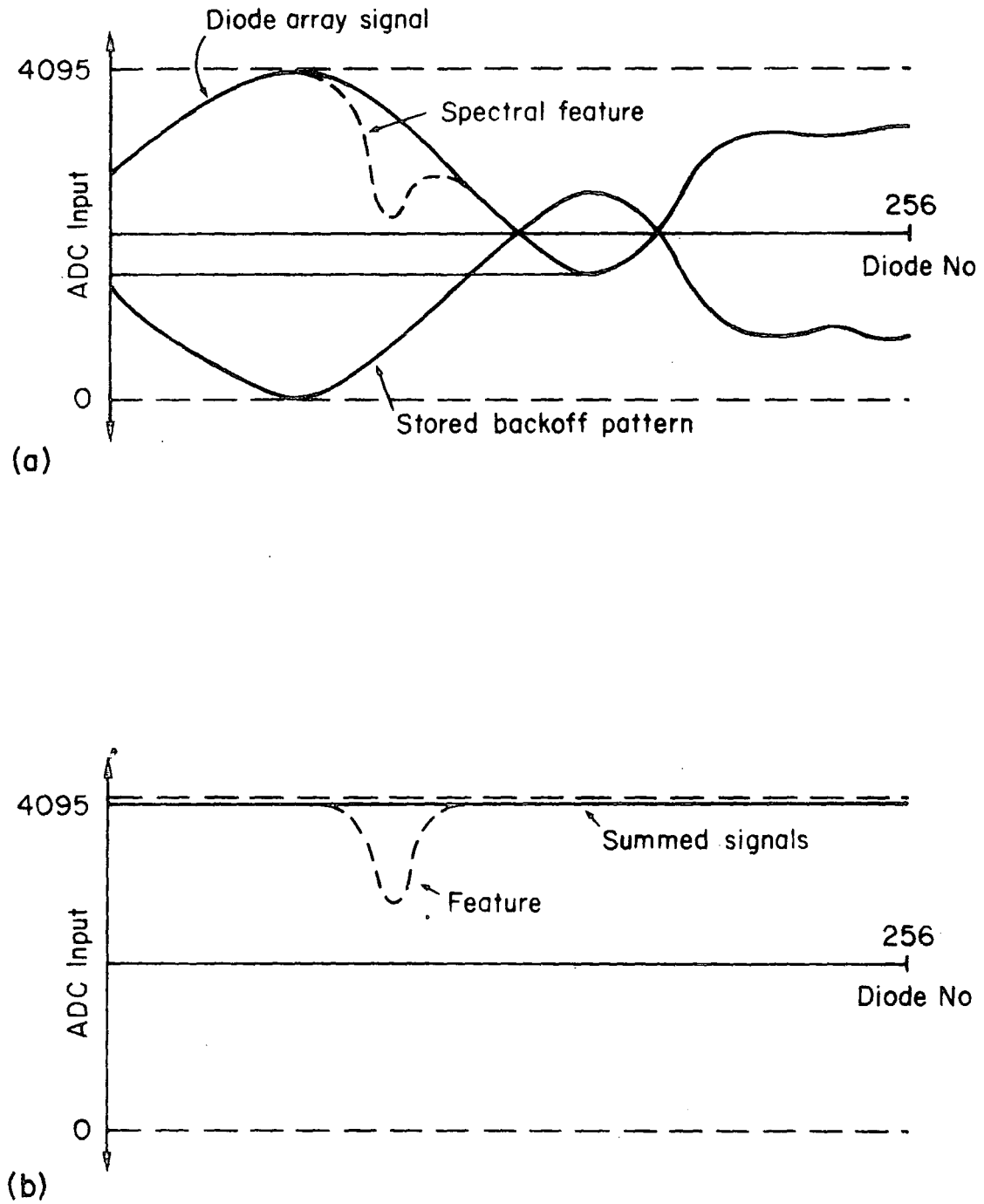
The valid 12 bit data words generated by the ADC are strobed into a buffer memory from which they are recovered at a later time by the computer, as described in Chapter 5. A status flag is output by the ADC module and, in normal use, the falling edge of this signal may be taken to indicate that the conversion is complete. In the circuit of Figure (4.3), a negative edge triggered monostable produces a data valid strobe from the status flag in order to load the data word into the buffer memory. Data is transferred in parallel. Optionally, the cycle controller may be used to generate a load pulse at least 25μs after the pulse which initiates conversion. The present memory also requires an address increment pulse which may occur at any time between the end of the memory load pulse and the next memory load operation. Thus, the sample pulse may additionally perform this increment function, although, in practice, a separate increment pulse is generated by the digitization cycle controller.

3. Fixed Pattern Removal

It has been noted that the overall dynamic range may be improved either by employing an ADC with more bits or by decreasing the system conversion gain. Both of these solutions have disadvantages, the former in the cost of the unit and the latter in the wastage of the optimum system noise performance. From Equation (4.5), it may be seen that there is a third alternative - that of removal of the fixed pattern of offsets which may possibly be more cost effective.

Figure (4.9) illustrates the principle of fixed pattern removal being evaluated in connection with the present system. As mentioned,

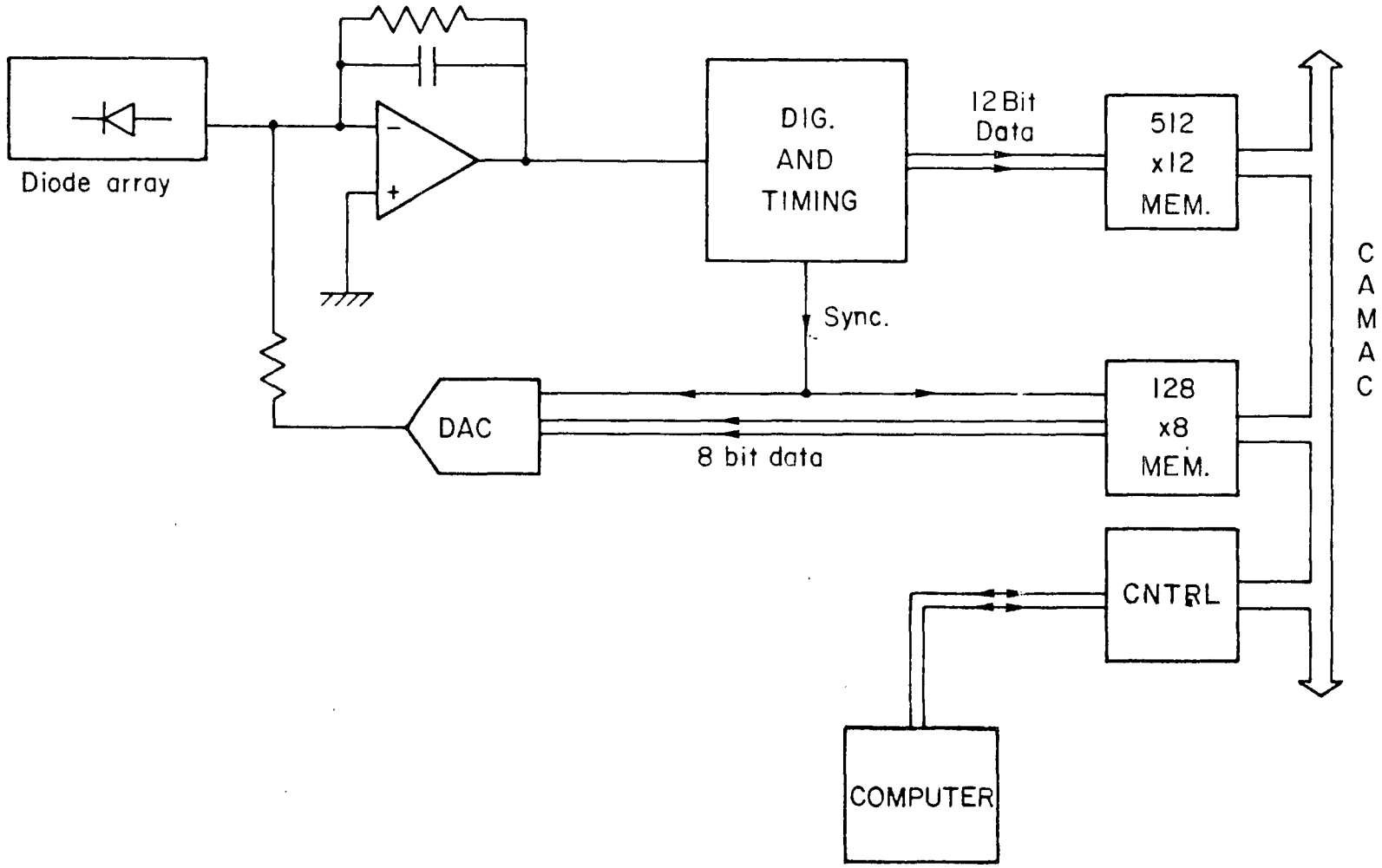
Fig 4·9 FPN REMOVAL



an offset network is, in any case, connected to the summing junction of the head amplifier and this is adjusted in order to make the best use of the input range of the ADC. The idea of fixed pattern removal is to replace this single offset for the whole array with a set of 128 offsets optimizing the use of the ADC by each diode pair. A diagram of a typical mechanization of fixed pattern removal is shown in Figure (4.10). The memory containing the back-off data is coupled to the DAC shown in the circuit of Figure (4.3) in order to implement this programmable back-off facility. As the ADC input still has to accommodate the sum of the signals on two diodes, this digital mechanization does not quite offer the potential of analogue d.c. restoration.

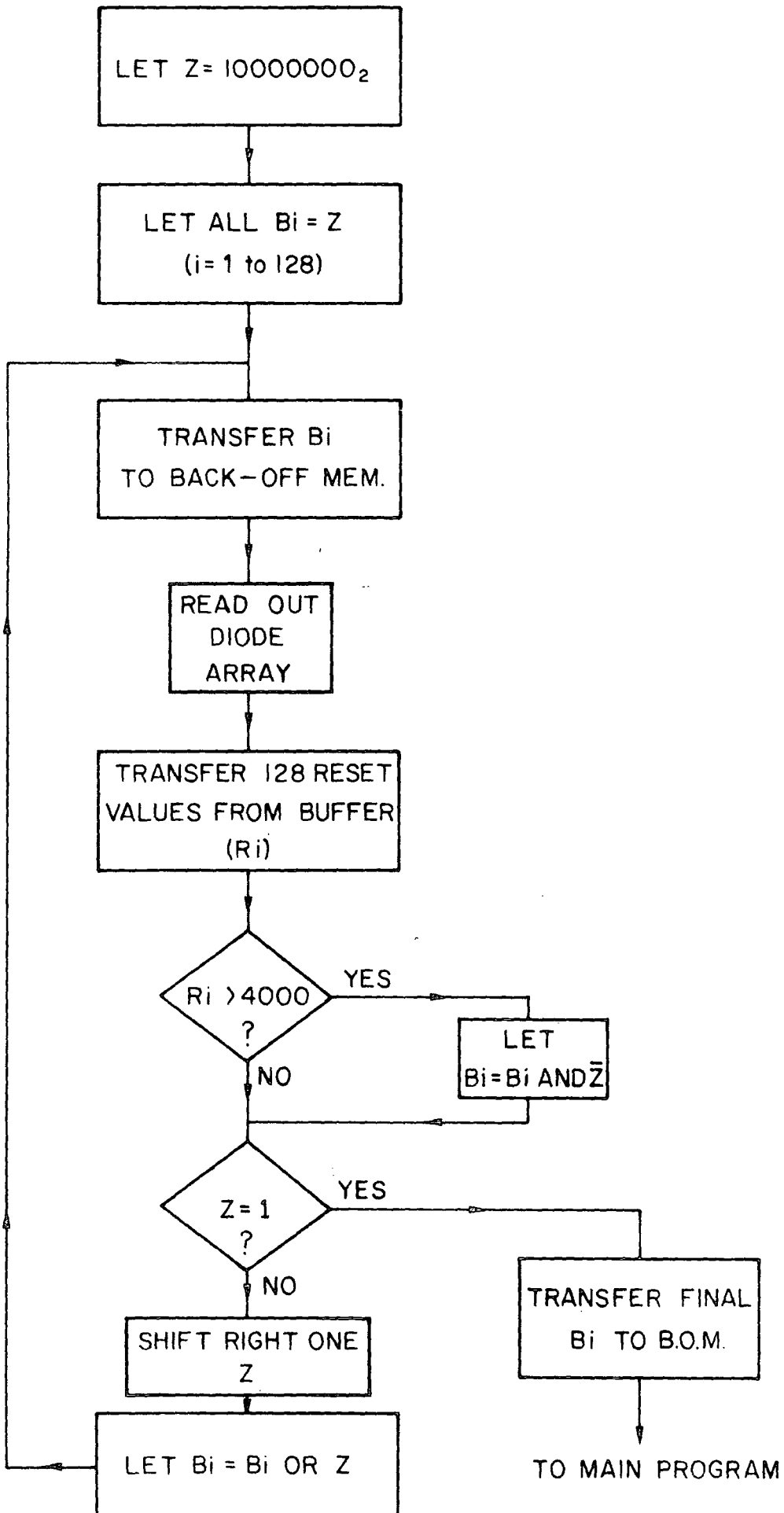
In Chapter 3 it was noted that the system gain changed with the operating temperature of the chip and thus that the fixed pattern of reset levels is also temperature dependent. This dependence and the fact that it may be wished to use several different arrays make it impractical to store the appropriate back-off data very far in advance. A similar system has been described in a paper by Fry[5], using an MDAC to remove responsivity variations in the array output, where the correction terms are stored in a PROM. As the output in the present system is to be digitized and processed by a digital computer, it is fairly simple to arrange for the correction terms to be calculated automatically just before an integration. This setting-up operation can be done (as indicated in the flow diagram of Figure (4.11)) being, in effect, simply a successive approximation 8 bit conversion accomplished in software. The 128 x 8 bit memory is built as a CAMAC module, in a similar fashion to the buffer memory. Initially, all the MSBs of the back-off data words are set to 1, the remaining bits being zeroed. The array is then read out and the digitized values of the reset levels are tested with a given ADC value arbitrarily decided upon as being the zero signal level. To allow a margin of error, this cannot be quite at the extreme of the ADC

Fig 4·10 FFN REMOVAL SYSTEM



range and, in the present system, might conceivably be the code 4000. For those pairs with reset levels having codes less than 4000, the MSB is left set,while in the remaining cases, it is cleared. In the same way, all the other bits are set and tested until, finally after 8 iterations, the set of 128, 8 bit codes which most nearly produce reset levels with the 12 bit code 4000, is found. This technique does not require exact knowledge of the system gain and, indeed, if the approximation routine is carried out before every run alterations to the system gain may be made between exposures if desired. Absolute accuracy of the correction terms is not important since the output is still digitized and reduced in the same way. In addition, the existence of the back-off unit is entirely transperent to the rest of the system operation, features such as NDRO being unaffected in execution. Furthermore,the value of the NDRO facility should be greatly enhanced since most of the fixed pattern is now subtracted in real-time, making any genuine features immediately apparent. The optimum number of bits for the back-off codes may well be less than eight. Setting up the gain of the DAC to give an f.s.r. output equal to the peak to peak fixed pattern variation, this system will reduce the variation in the reset levels by a factor of 2^8 (= 256), which is clearly more than adequate. However, 8 bit DACs are fairly common and cheap, for which reason this back-off system was thought to be a potentially better way of increasing dynamic range, in terms of cost and benefit, than changing to a 14 bit ADC. In practice, some of the output range of the DAC must be reserved to allow the correction terms to change to counteract changes in system gain, without alterations to the offset circuit being necessary. Nothing further can be said here about the actual operation of this back-off unit which has not yet been fully tested. If, in operation, the above system introduces too much additional noise,a base-line restoration circuit, involving differentiation and integration of the signal along the lines of

Fig 4·11 SETTING UP BACK OFF MEMORY



a nuclear pulse amplifier,[6] may provide a better solution, although this would necessitate considerable redesign of such features as the NDRO facility.

REFERENCES

1. A description of d.c. restoration techniques with reference to CCDs is given by White, M.H. (1976). "Design of Solid State Imaging Arrays." Solid State Imaging. (Noordhoff).

2. See for example
Wait, J.V., Huelsman, L.P. and Korn, G.A. (1975).
Introduction to Operational Amplifier Theory and Applications.
(McGraw-Hill).

3. Tewkbury, S.K., Meyer, F.C., Rollenhagen, D.C., Schoenwetter, H.K. and Souders, T.M. (1978). "Terminology Related to the Performance of S/H, A/D, and D/A Circuits." IEEE Transitions on Circuits and Systems. Vol. CAS - 25, No.7, pp.419-426.
See also
"Data Conversion Handbook." Hybrid Systems Corp.

4. Gordon, B.M. (1978). "Linear-Electronic Analog/Digital Conversion Architectures, Their Origins, Parameters, Limitations, and Applications." IEEE Transactions on Circuits and Systems. Vol. CAS-25, No.7, pp.391-417.

5. Fry, P.W. (1979). "Enhancement of The Uniformity of Response of Self-Scanned Arrays." The Radio and Electronic Engineer. Vol.49, No.10, pp.506-508.

6. Herbst, L.J. (1970). Electronics for Nuclear Particle Analysis. (OUP).

CHAPTER 5

CONTROL AND COMPUTING

1. Requirements of a Control System

In Chapter 3 it was noted that there were a number of possible operational sequences that might be employed to run the array, depending on the use to which it was to be put. Only one of these approaches seemed suitable for Astronomical work. In outline, this procedure has three stages: (i) a setting up period during which the elements are charged up to a repeatable initial voltage; (ii) an integration time for which the incident photon flux discharges the elements of the array; and (iii) a final readout scan at which time the double sampled data is taken. This sequence may be accomplished by appropriate gating of the scan pulses, but as it is desirable to be able to use the non-destructive readout (NDRO) facility of this device which itself requires monitoring scans to be made during the integration stage, it is also necessary to gate the recharge pulses. The intervals between the operations in the above sequence may be preset as integer numbers of frame times (the time taken to scan the array once) but, additionally, it is useful if the operator may request NDROs and also terminate the integration at will. Thus an assessment may be made from the rough readout provided by the NDRO as to whether or not to terminate or continue the exposure. However, when the real-time video output from the above system is displayed on an oscilloscope it is difficult to see any signal features (e.g. spectral lines) against the background of fixed pattern offsets. To overcome this, a reference fixed pattern read-out has to be made at the beginning of the integration and the data from this subtracted from subsequent NDROs. It is possible to do this in real-time as discussed in Chapter 4, but otherwise the reduction has to be accomplished digitally as does the subtraction necessary to obtain the final double sampled data.

Although the above control operations together with storage of the data on some permanent medium could be implemented by a hard-wired controller, the data manipulation required is very much easier to perform with a minicomputer system. Also, for a development system where all the control functions that may be required are not necessarily known at the outset, this approach is extremely flexible. If, for example, the wish was to define the exposure length with a mechanical shutter as opposed to the length of the integration time, this facility could be added fairly simply. More minor changes to parameters such as length of integration time and number of initial recharges may be done in software.

2(a) Computer Control of Diode Array via CAMAC

In the present system the sequence of operations necessary to make measurements with the diode array and store the data is controlled by a PDP 1103 microcomputer which is interfaced to a purpose-built CAMAC system by a HYTEC 1100 Crate Controller. The crate contains the integration control module, a buffer memory module and various peripheral interface modules such as a storage display driver for driving a monitor.

Briefly, the control in question may be divided into two parts: (i) an input to the computer from the array driving electronics through an interrupt facility; and (ii) a control output from the computer which may gate various system signals. During an exposure the elapsed time is measured by counting interrupts produced by the scan pulses which are output from the array clock sequence generator. In outline this counting is accomplished as follows: firstly, the scan pulses are made to set a LAM latch within a CAMAC module; secondly, the latch, if enabled, may then interrupt the computer via the crate controller; thirdly, the computer in turn increments a frame counter according to instructions in the control program, tests the value of this counter and prepares the

control module to gate the system control pulses; and finally, the LAM latch is reset.

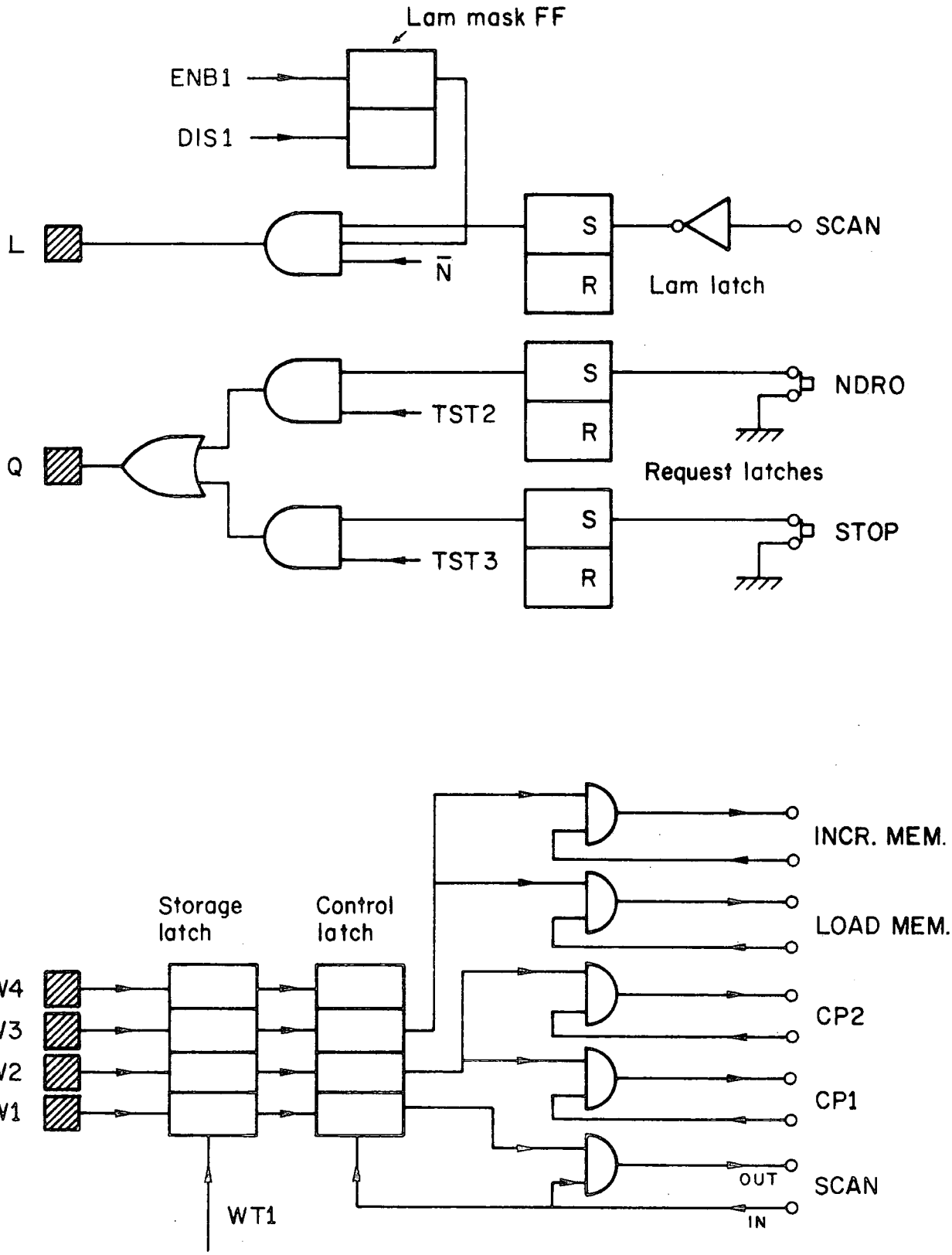
The LAM (Look At Me) Latch referred to above, a feature of the CAMAC specification,[1] allows requests for computer attention to be made by any module within the crate. These requests, at this stage termed interrupts, may then be handled by specific software routines. However, while the interrupts are not being serviced, the computer may perform a number of functions included in which are the transfer of data from the buffer memory into its own store, the simple reduction of data and its display on a monitor.

Once the control program has started running, external operator interaction with the system is possible by means of two switches which allow requests for either a non-destructive read or the termination of the exposure with a destructive read. Again, once the latter request is chosen data may be examined, redisplayed on the monitor or, for purposes of later analysis, transferred to floppy disk or paper tape. With the present system, following termination of the exposure, operations are directed by simple keyboard responses to a dialogue generated by the control program.

2(b) Control Module Operation

In the system described here the input and output control functions are performed within a single control module. This unit contains three LAM latches which may each be enabled, disabled, tested and cleared under program control. Through these latches the three external system signals scan pulse, NDRO request and STOP request may interact with the running control program. A simplified diagram of a typical control module is given in Figure (5.1). After making a decision on the necessary control action, the program may write a 4 bit control word to a register within the control module, where it is stored until updated. At the time of the scan pulse the word in this storage register is loaded into

Fig 5·1 CONTROL MODULE (CAMAC)



the control register. Each bit of the word in the control register gates various control functions in the array system. Bit 1, if set, allows scan pulses to reach the array, whereas bit 2 controls the recharge pulses and bit 3 controls the memory load and increment pulses. The fourth bit is presently spare. Thus, at the beginning of every frame time the current contents of the storage register are loaded into the control register thereby enabling or disabling the system pulses as required by the control program. This double latch system is necessary because the control word must not change in the middle of a frame. To allow the computer plenty of time to make a decision about which pulses to gate, depending on the current value of the frame counter, action is deferred for one frame time by the above system. Allowance can be easily made for the fact that a control word written to the module during the nth frame time, will not be acted upon until the n + 1th frame. At the scan rate presently used, this deferred operation gives the computer 10^{-1} sec. to recognise the interrupt from the scan pulse, increment and test the frame counter, write a control word to the module, set any necessary program flags and finally clear the LAM before the next scan pulse arrives. Not all the above operations are performed each time a scan interrupt occurs but the above represents the most that will have to be done and must be allowed for in the system design. If the frame time were decreased or the interrupt routine made longer such that it could not be fully executed before the arrival of another interrupt, it would be necessary to divide down the interrupt rate by doing some counting of scan pulses within the control module.

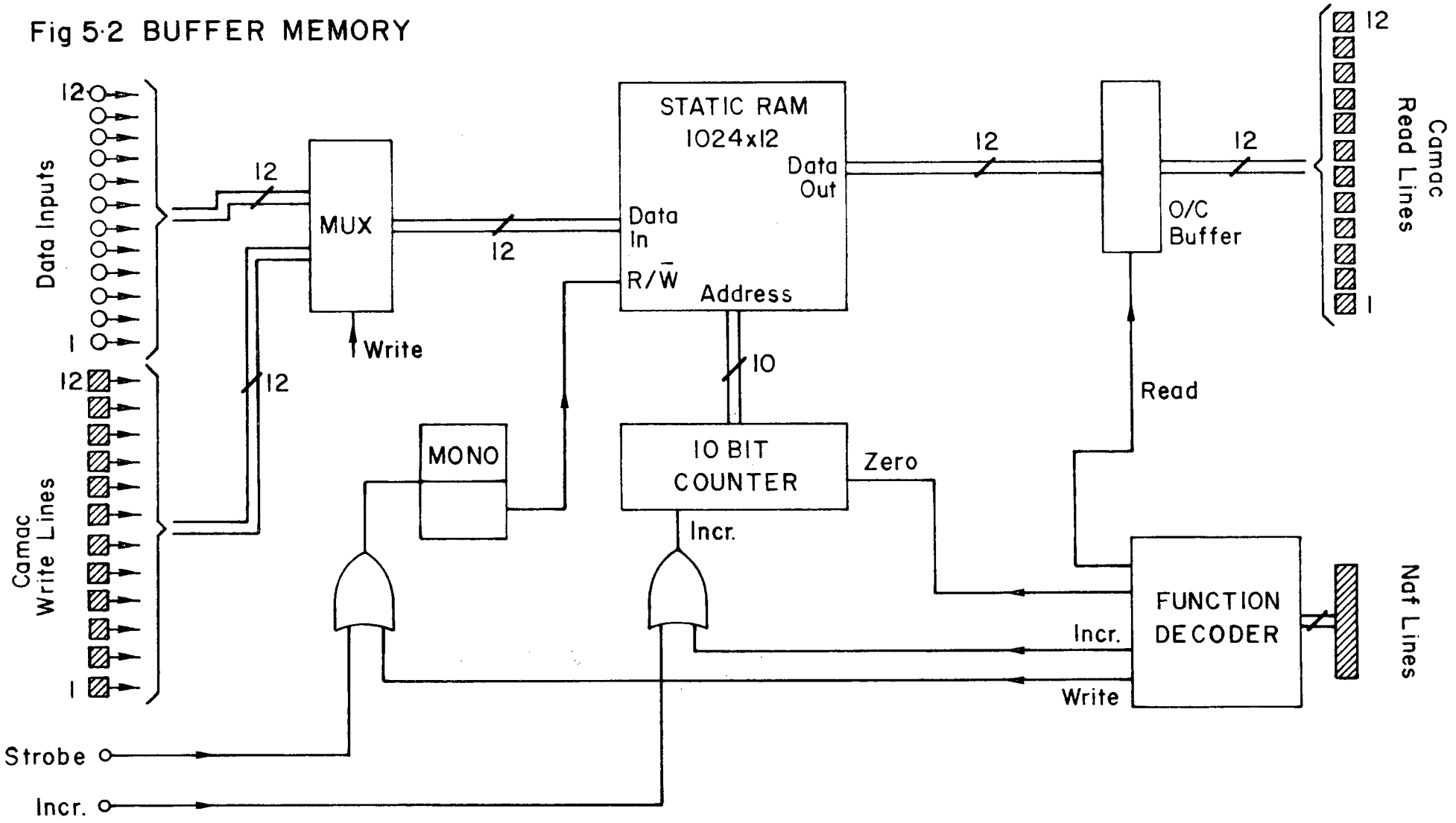
2(c) Buffer Memory Operation

The data values are generated synchronously by the array readout system at a rate determined by the master clock frequency which is set by the operator. At lower data rates input may be made directly to the

computer memory, through CAMAC, under processor control. However, the synchronous nature of the data makes a full handshake procedure impossible with the result that, at higher data input rates, there is an increasing possibility of the computer missing words because it is occupied with internal tasks of higher priority. To avoid this problem it is necessary to read out, firstly, into a buffer store from which the computer may recover the data asynchronously. Data may be transferred directly into the computer memory at much higher rates using DMA but this requires more sophisticated software techniques and has not been considered worthwhile in view of the relatively small amount of data being handled. As one readout scan produces only 512 x 12 bit words of raw data it was thought practical to build a memory large enough to store all of them, thereby allowing the system to be operated at any practical rate. The maximum throughput is now determined by the digitization process which is much slower than typical memory write times. An alternative approach considered was to make a smaller memory of the first-in first-out (FIFO) kind which has the effect of smoothing out the uneven input rate of the computer, but still requires the device data rate to be no greater than the average computer input rate.

The memory currently in use can store up to 1024 x 12 bit words (i.e. two whole readout scans) which might be needed if two arrays on a chip were read out in consecutive frame times or a second readout of an array had to be made before the data from a previous readout had been fully recovered by the computer. This memory is shown schematically in Figure (5.2) and a full realization is given in Figure (5.3). Before a frame of data is accepted by the memory, the address counter is zeroed by a computer command or, alternatively, set to the starting address of the second half of the memory. Loading of this buffer store, via the front panel connector, is controlled by two management pulses, load and increment,

Fig 5·2 BUFFER MEMORY

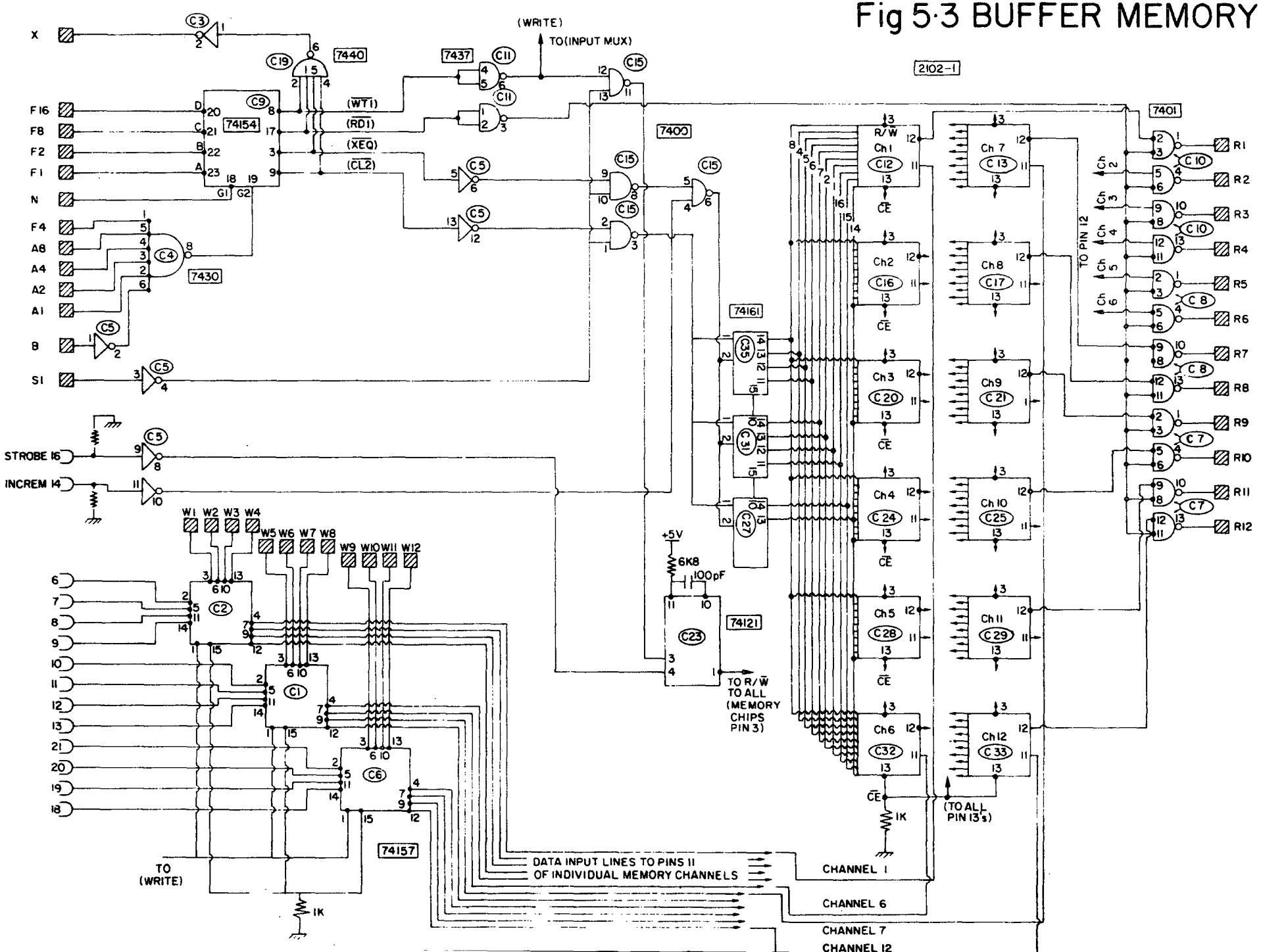


which are generated by the digitization control electronics (as described in Chapter 4).

Synchronization of the above operation is achieved by gating these control pulses with the control module as described in the previous section so that loading begins with the first diode and finishes with the last. As this loading operation takes exactly one frame time there is no need for the memory module to indicate to the computer when all the data have arrived. The data may be recovered from the buffer store, that is, under control of the main program, in any subsequent frame times once the memory address counter has been reinitialized - again by a command from the program. Since the buffer memory is also built as a CAMAC module, the necessary control functions initialize address, read data and increment address are implemented through the crate controller using the standard CAMAC functions. For test purposes data may also be written into the buffer from the computer. Reading back this dummy data and comparing it to the original, possibly by using the display monitor facility, provides a quick check against memory faults.

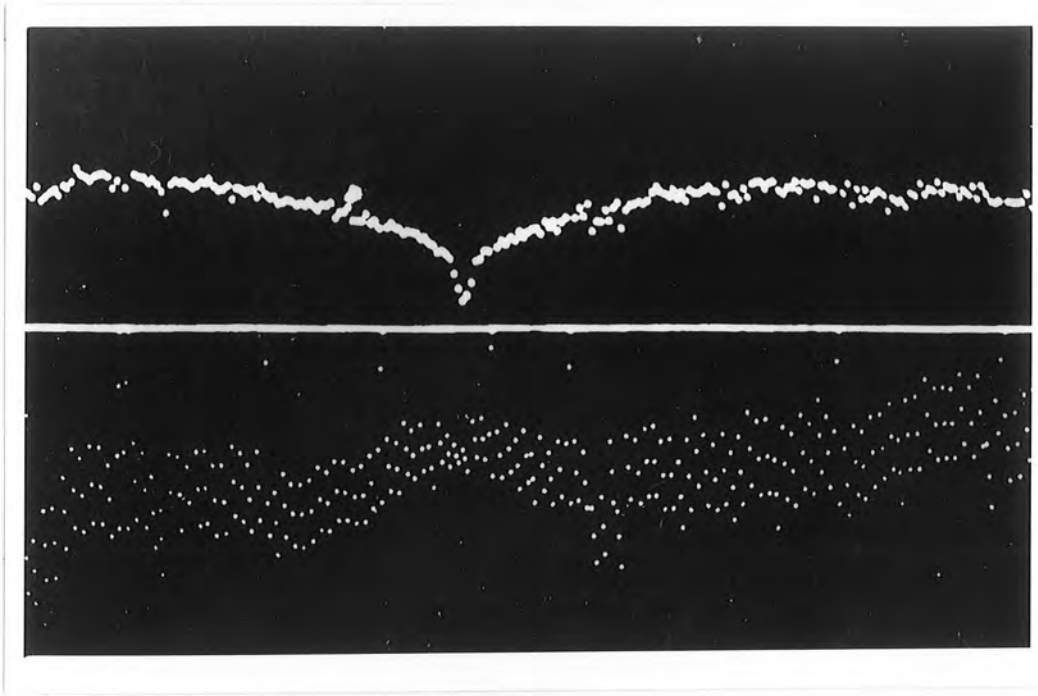
Earlier buffer memories which only stored one frame of data were based on 256 x 4 bit static RAMs and required tri-state buffers at the memory inputs because these RAMs have their inputs and outputs multiplexed onto the same pins. This complication led to some fairly involved wiring which in the prototypes was all done by hand wire-wrapping. Not only was this very time-consuming but it apparently led to some intermittent failures and consequent down time in an experimental situation. For this reason the present greatly simplified design using 1024 x 1 bit RAMs which have separate input and output pins was adopted. This newer unit with its self test ability and reduced number of wire-wrapped interconnections is much easier to service.

Fig 5·3 BUFFER MEMORY

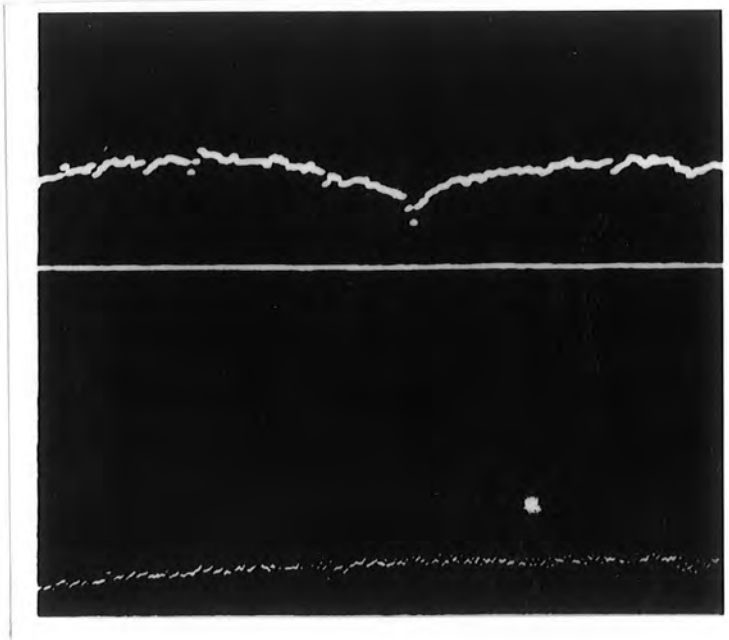


2(d) Display System

For display of the data from NDROs and destructive readouts, a TEKTRONIX 603 bistable storage monitor is used. This display is interfaced to the system through CAMAC by a storage display driver, NE 9028 and a mode generator, NE 7011. The former of these two standard Nuclear Enterprise's modules contains two 10 bit DACs - to control X and Y deflection. The latter module provides facilities such as screen erase together with some adjustment of the bright up period. Conversion gain of the display is set by writing an appropriate control word to the driver module to determine which 10 bits of the data to convert. In the diode array system raw data are produced with 12 bit precision and, for purposes of display, it is usual to convert the 10 most significant bits. However, at small signals the reduced double sample data may not make use of the more significant bits and a gain of either 2 or 4 is possible simply by reselecting the bits to be converted. As was discussed in Chapter 4, a large part of the ADC range is often used up by the fixed pattern offsets and it is known from the outset that the subtracted data will be of effectively lower precision (in terms of bits). For plotting rough spectra, a point plot mode may be selected by setting particular bits in the control word mentioned above. Thereafter it is only necessary to write successive Y coordinates to the module and a point spectrum will be displayed, the X coordinate being incremented automatically between writes. Such displays of reduced NDRO spectra enable the operator to decide whether or not it is worthwhile to continue a run depending on what features are seen to be emerging. At the same time, by looking at the raw data it can be seen whether or not the signal is nearing the saturation part of the output characteristic. To accommodate these two checks, the screen may be split to display two spectra, one above the other. Figure (5.4) shows a typical split display.



(a) Final readout



(b) NDRO

Fig 5·4 MONITOR DISPLAYS

3(a) Software

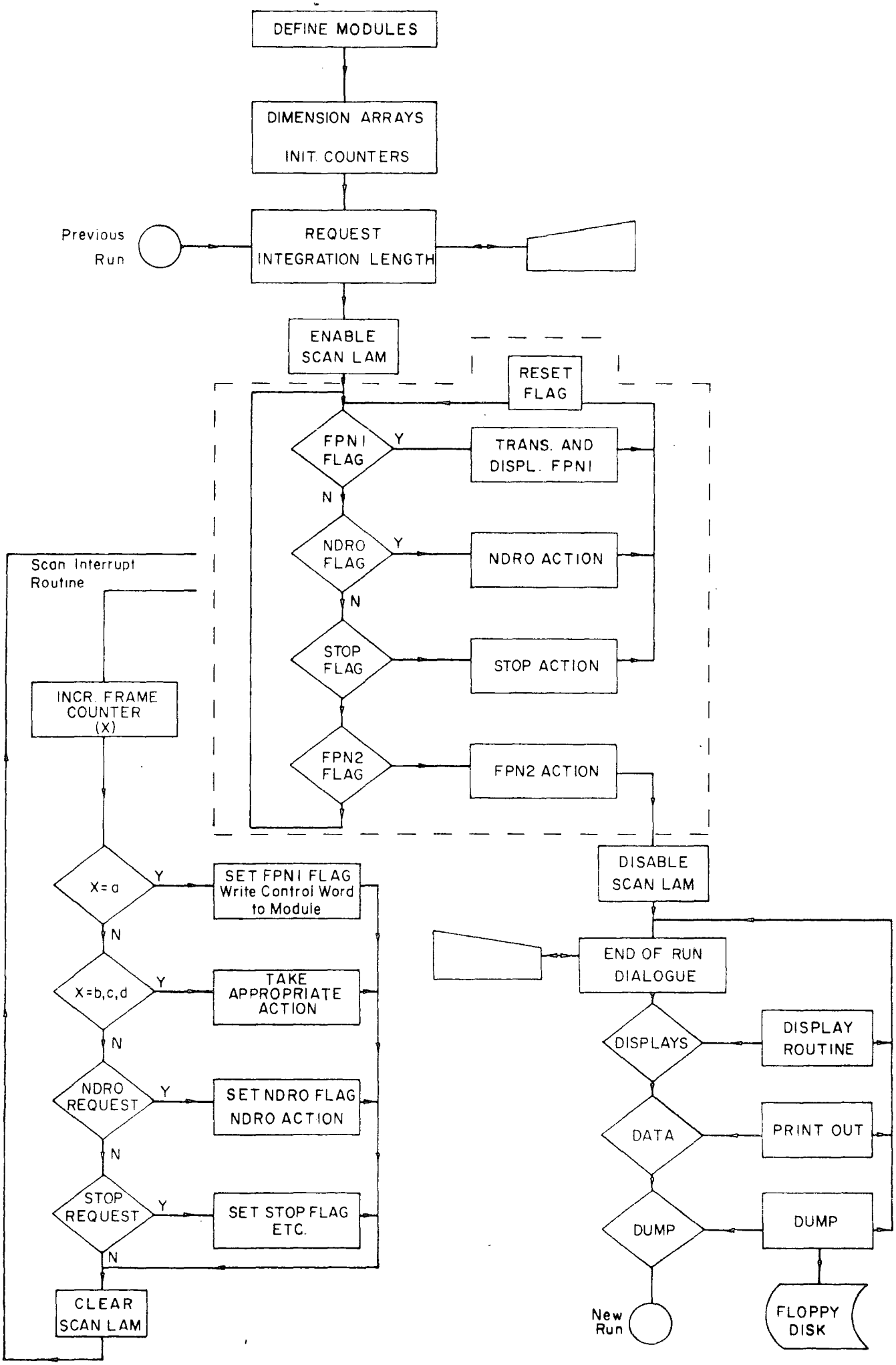
The present system employs a specialised CAMAC control language called CATY which was supplied by Francis Golding Associates. The latest version is for use on a DEC PDP 11 Floppy Disk based system running under RT 11. A CATY program may make use of any of the RT 11 device handlers for input and output which means that data may be dumped into disk files from a running program or output on any other system peripheral. Previous versions of CATY used here were for paper tape operation only and consequently, data had to be dumped to a CAMAC tape punch.

CATY uses statements and commands which are a subset of BASIC and is therefore very easy to learn and use. To handle CAMAC operations it also incorporates some special statements which allow data to be written to and from modules and also the various management functions (e.g. Clear, Test LAM etc.) to be performed. Interrupts generated in any module within the crate may be dealt with by an interrupt routine specified for that module location. The execution speed of CATY is fairly good because the single run command first compiles the entire program and then begins execution, unlike ordinary BASIC which is compiled and executed line by line. However, because the compiler, the source program and the compiled version of the program have to be fitted into resident memory, in our case with the RT 11 resident monitor and any run data, this software approach does not represent very efficient use of memory. Other more obvious disadvantages are that CATY has very limited arithmetic powers, is restricted to the use of integers and may only have 26 variables (vectors and scalars).

3(b) The Control Program

A flow diagram of a typical control and data acquisition program is shown in Figure (5.5). Firstly, the program assigns strings of characters

Fig 5·5 DATA ACQUISITION PROGRAM



to the module locations which are to be used in the crate, enabling the modules in them to be referred to by more meaningful names. Doing this has the advantage that, if the position of a module, called the station number (1, - - 24,) is changed, it is only necessary to modify one line, the assignment statement, and not every line in the program in which that module is referenced.

As CATY only allows the use of 26 variables it is often necessary to define some letters as arrays in order to conserve possible variables. Large arrays are dimensioned for temporary storage of the data because it is inconvenient to dump data immediately to a mass storage device; several small arrays are dimensioned for all variables of a similar type to save letters of the alphabet. For example, all the integration parameters may be placed in one array, say A(n), such that

A(1) = number of initial recharges before the first fixed pattern is measured,

A(2) = the frame number at which integration commences,

A(3) = the operator set integration length,

A(4) = the frame number at which the destructive read is made,

A(5) = the last frame at which an NDRO was requested, and

A(6) = the number of final recharges before the fixed pattern is again measured.

The program may request the operator to input values for some of these parameters, although usually only for the set integration length.

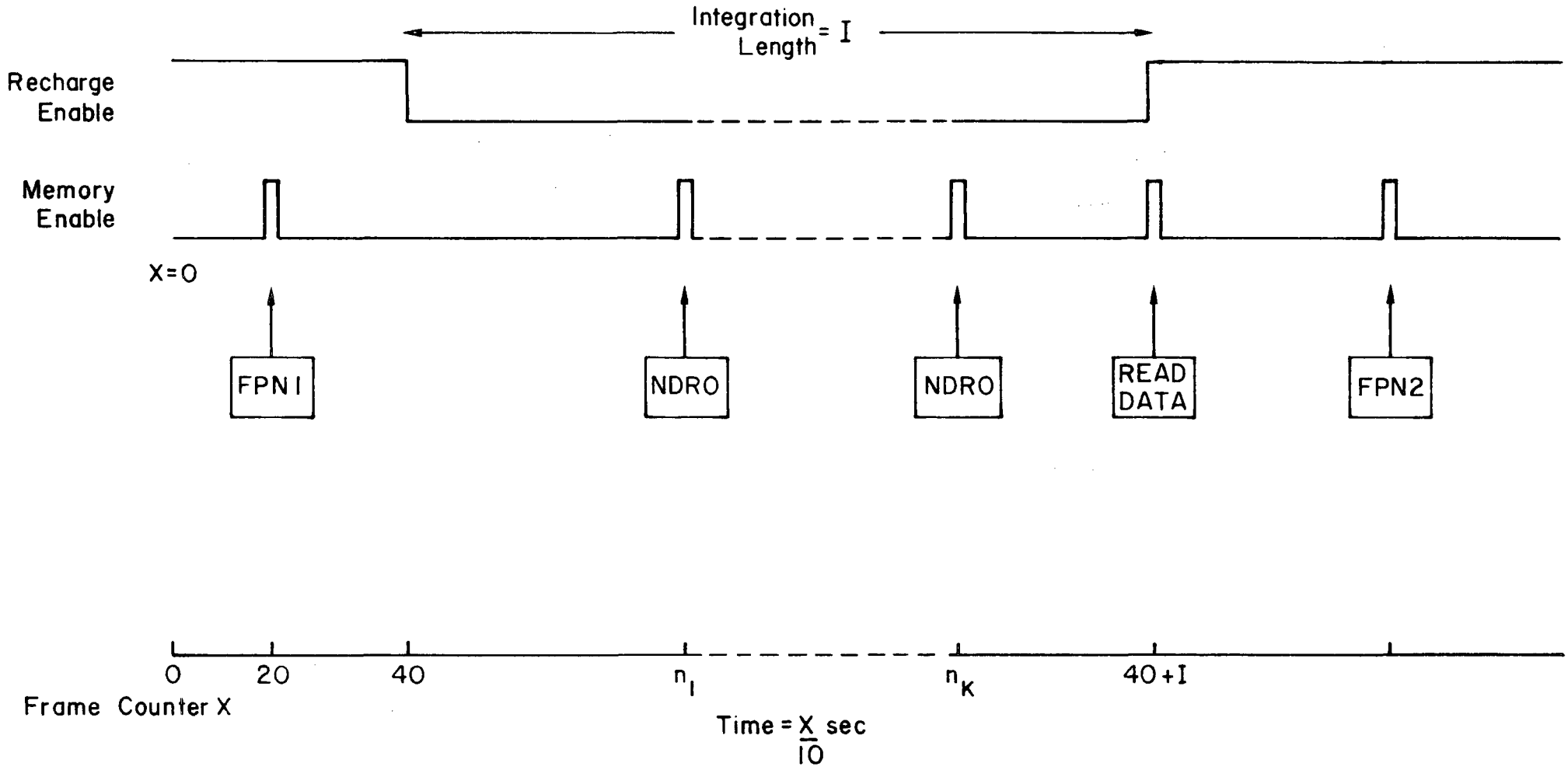
The two most important variables in the program are the frame counter and the status flag which are both zeroed before the operational part of the program begins. As soon as the Scan LAM is enabled the computer will start to count frames, the frame counter being incremented on every occasion that the interrupt routine is invoked. In this routine, the frame counter

is subsequently tested with the values in an integration parameter array such as that described above. Acting on the outcome of this test, any required control word is written to the control module, the status flag is set, if necessary. Finally, the NDRO and STOP LAMS are tested which may result in similar action to that taken upon the occurrence of particular values of the frame counter. In practice, with the present control module the LAM latches are automatically reset when the appropriate control words are written to it. Back in the main program, the status flag is constantly being tested and if it is found to be non-zero, operation of the program will branch to a particular routine depending on the exact value of this variable which may have values 1 to 4. e.g. if, when a Scan interrupt was last being serviced, a NDRO request was identified, the NDRO flag value will be set. This will cause the computer, having waited one further frame time (to allow for the deferred action of the control module), to begin recovery, reduction and display of the data in the buffer memory. When the foregoing has been accomplished the status flag will again be zeroed and the program will resume testing it as before.

3(c) Typical Operational Sequence

The present typical sequence of events, which is shown in Figure (5.6), is summarily outlined below. The array is read out destructively after so many recharge scans, whereupon the frame of data is transferred from the buffer memory and displayed in order to check that the output signal is comfortably within the ADC input range. As soon as this first transfer and display, which typically takes about 2 seconds, has been completed, the integration period is begun in the next possible frame time. Earlier programs did not have a delay between the first fixed pattern read and the start frame, with the result that it was not possible to do integrations of less than about 20 frames (i.e. 2 seconds) and to retain both the

Fig 5·6 TYPICAL OPERATION



information from FPN1 and the final read out. Losing the FPN1 data on such a short integration would not matter because there is not enough time to use the NDRO facility, but having different versions of the control program for different situations (e.g. Star and Standard lamp) would be inconvenient so that the present solution with the delay was adopted. There are other ways of overcoming this problem such as alternately using the first and second halves of the double length memory, but in practice delaying the beginning of the integration works well enough. Performing a fixed pattern read-out, FPN 1 taken several frames before the actual start of the integration in order to have a zero signal reference for any NDRO scans is acceptable here because NDRO is only used as a rough monitoring facility to see what signal, if any, is appearing.

Since the present system does not include a mechanical shutter, the exposure begins at the start of the first frame time in which the recharge pulses are disabled. At any time thereafter the NDRO request button may be pressed, thereby setting a LAM latch within the control module. This LAM, although never enabled (i.e. it does not produce an interrupt) is regularly tested and, if found to be set, produces a CAMAC 'Q' response which causes the control program to execute the non-destructive read sequence. It should be noted that the system usually scans the array and digitizes the video output all the time so, from the array's point of view, it is read non-destructively in every frame time during an integration. The NDRO request facility merely causes the control module to enable the memory load pulses during a particular frame time and, consequently, the non-destructive data may be transferred to the computer resident memory for display on the monitor. Any subsequent NDRO requests update the data stored in the array dimensioned for these, with the result that only the information from the last NDRO is retained.

When the value of the frame counter is such that the set integration length has been reached or the STOP LAM is found to be set, a final destructive read is made from which the desired double sampled signal values are calculated. If the system is executing the NDRO transfer and display routine when the set integration length is reached, the present program terminates the exposure in the first possible frame after the routine has been completed and informs the operator of what has happened. The resulting discrepancy between set and actual integration time, at most 20 frames, is rarely of importance in situations where the NDRO facility is being used, i.e. generally in integrations of 1000 frame times or more. Alternatively, this problem may be overcome by disabling the NDRO facility during the danger period near the end of the integration, which may easily be done in software, or again, as with the similar problem encountered in processing FPN 1, the second half of the long memory may be employed.

As mentioned in Chapter 3 the Plessey Array suffers from an incomplete recharge problem which means that the reset levels measured by the final read scan will be considerably different from those measured at the time of FPN 1 if any significant signal has been integrated. For this reason a second fixed pattern readout, FPN 2, is made a number of frame times later than the destructive readout which terminates the integration. Since the FPN 2 readout is performed after several consecutive recharge operations, as were done before FPN 1, the signal levels measured by FPN 2 should have returned to the repeatable initial condition. Any difference between FPN 1 and FPN 2 is indicative of system drifts and changes in gain which have occurred during the exposure time. As was the case with a long integration, these two fixed pattern readouts may be separated by as much as an hour; at the same time,these measurements provide a necessary check against slow changes in the device

characteristics due to factors such as operating temperature and supply voltage variations.

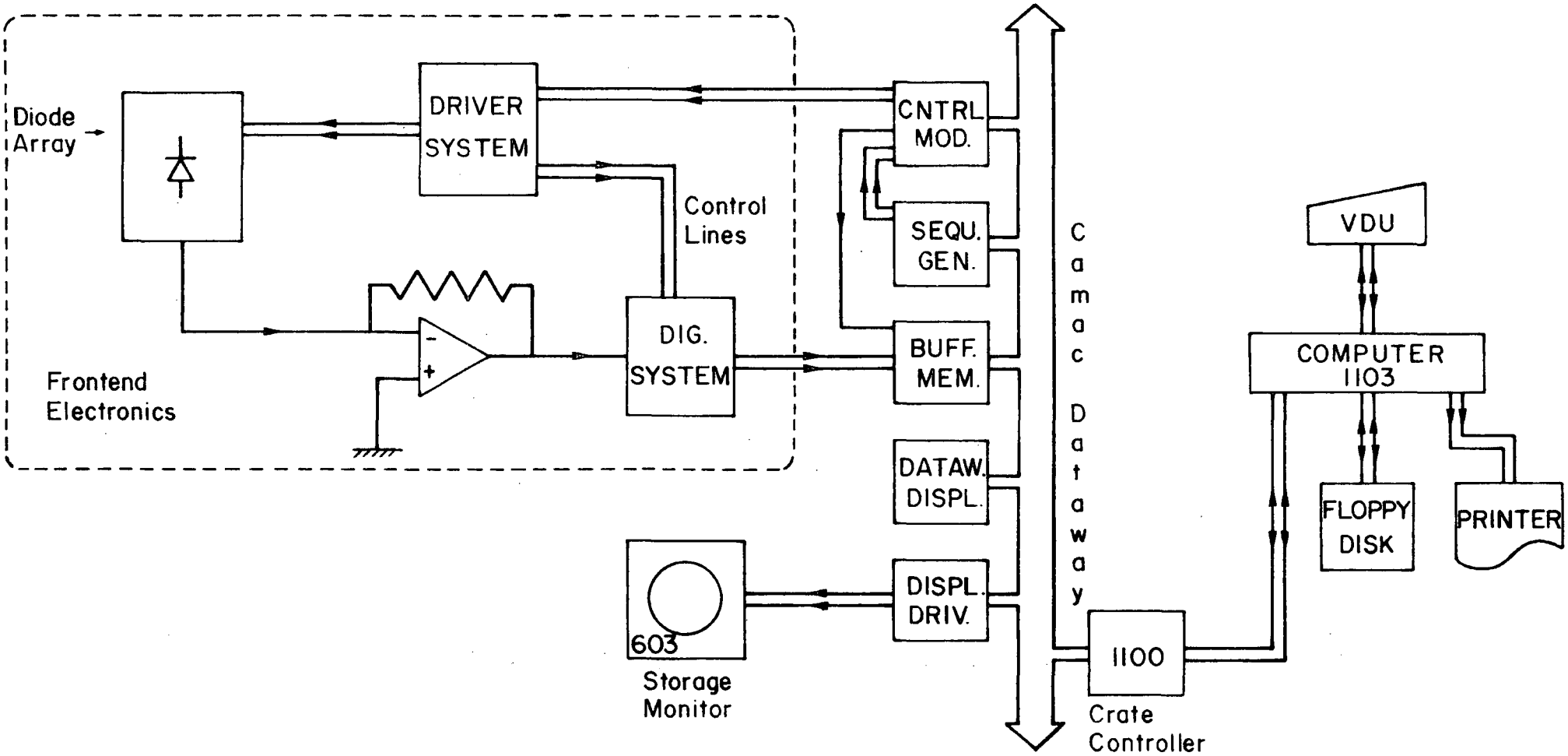
Following the transfer and, if wished, display of the FPN 2 data, the system types out the values of the various integration parameters on the operator's console. The operator is then asked by the system to select from a variety of redisplay and data printout options. This process is directed by single number keyboard responses to multi-choice questions generated by the program. After some re-examination of the data a decision may be made to dump the data to a more permanent storage medium, presently floppy disk, for later analysis. Alternatively, the data may be disregarded and the program will reinitialize all the integration parameters ready for another run.

4(a) The present Computing System

The present full system, which is outlined in Figure (5.7), comprises:- a DEC PDP 1103 microcomputer with a VDU interfaced as the console device; an RX02 dual drive, double density, floppy disk unit for program and data storage; a CAMAC system interfaced through a HYTEC 1104 card which also contains a ROM bootstrap; and a serial printer which may be connected directly to the computer through a serial port or, optionally, may be attached to the VDU as a hard copy dump. This disk system is operated under RT 11 (i.e. DEC's standard software) which supports the use of several high level languages including FORTRAN and BASIC. So far only CATY, which may be obtained in a version compatible with RT 11, has been used for purposes of control and acquisition. Additionally, some simple data analysis with FORTRAN has been tried.

As mentioned earlier, a CATY program may make use of any of the RT 11 device handlers for input and output. The present version of CATY,[2] supplied by FGA, allows up to four i/o streams which may be assigned to any device and file on the system. In the acqusition program it is only

Fig 5·7 Hardware

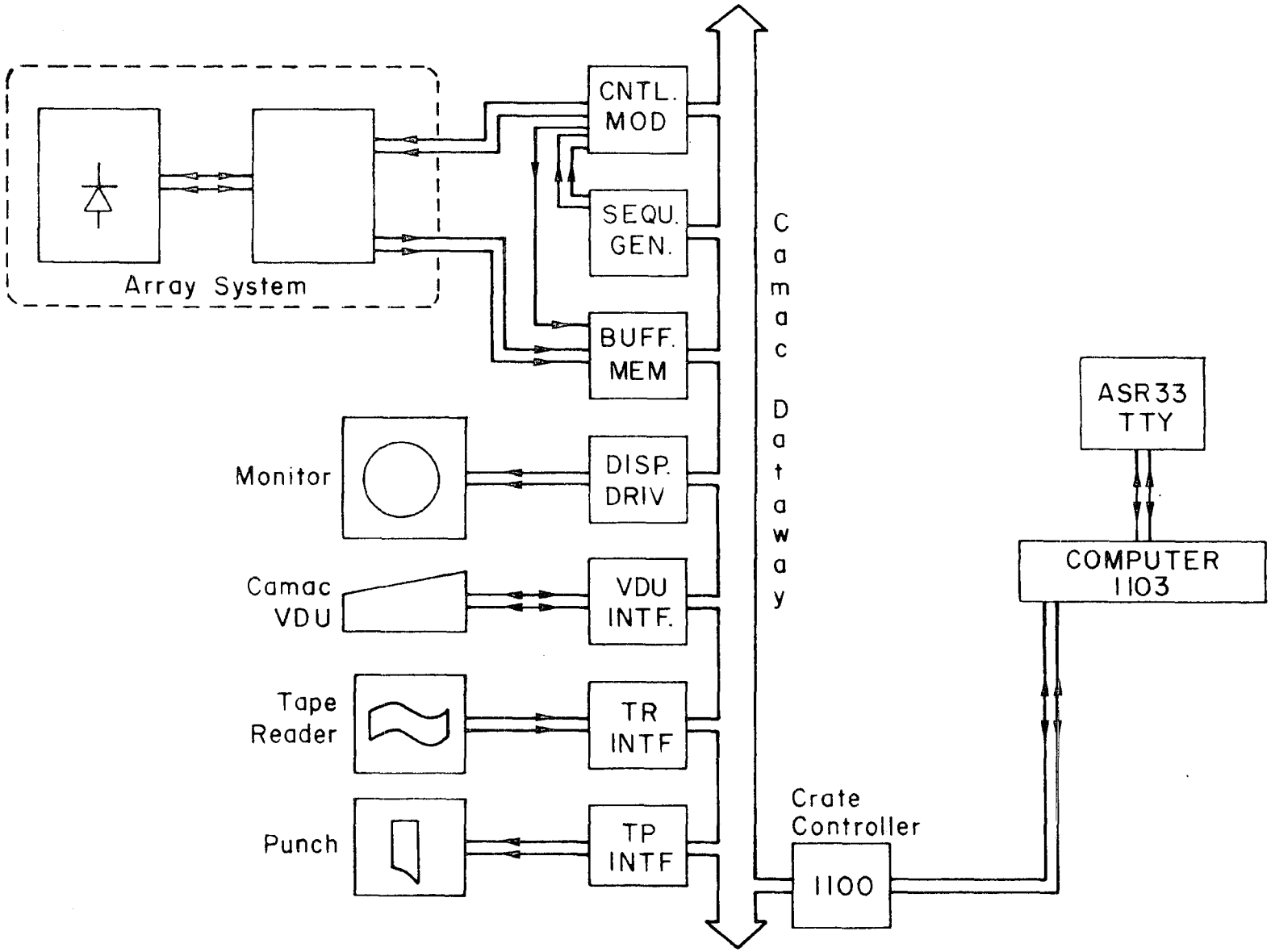


necessary to use one stream assigned to a disk file in order to dump the data from a run. If all the information from an exposure is dumped, Storage space is required for 512 words of FPN 1, 512 words of the most recent NDRO, 512 words of the double sample readout and 512 words of FPN 2. Using CATY this amounts to 2048 x 24 bit words of data. However, these are written into disk files as 8-figure, ASCII coded, decimal numbers (i.e. the maximum 24 bit CAMAC word expressed as a decimal = 16,777,215). Thus, although the system only digitizes the levels to 12 bits (i.e. maximum decimal number = 4095), CATY still treats all data as 24 bit words which, in decimal radix, are written as 4 figure ASCII coded decimal numbers with 4 leading spaces. This is not a very effic ent method of storage because the original 3072 bytes of data are finally written into a disk file as 16,384 bytes of ASCII coded decimal information, half of which is just spaces. (In practice, the numbers are delimited with CR and LF which take up yet more space). One double density diskette may accommodate approximately 512 K[3] bytes of data but, in practice, this can only store about 19 full dumps unless the data is edited and reformatted. However, even with its inherent inefficiencies this system has been found to be far more convenient than the paper tapes which were formerly used for data storage. To maintain compatibility with the old paper tape system, special device handlers have to be obtained for the paper tape punch and reader connected as CAMAC peripherals. By assigning appropriate stream numbers to these devices data may, optionally, be dumped to paper tape or old data may be read from paper tapes for re-display and comparison.

4(b) The Paper Tape System

This system which is illustrated in Figure (5.8) is a minimal configuration dedicated to CAMAC based applications and run by a paper tape version of CATY which is itself a simple operating system. The

Fig 5·8 PAPER TAPE SYSTEM



operator's VDU is interfaced through CAMAC and may be used only for entering and running CATY programs. An ASR 33 teletype is connected directly to the PDP 1103 as its console device and this allows the system to be started up by reading an absolute loader tape with the teletype's integral mechanical reader. A special version of the absolute loader tape has been acquired, thereby allowing further program tapes, in this case the CATY compiler, to be read by the fast optical reader which is interfaced through CAMAC. (Reading the CATY tape on the ASR 33 at 110 baud took about fifteen minutes). Once loaded, the CATY tape supports the use of the CAMAC VDU for further operator interaction with the system. Special CATY commands use the fast CAMAC paper tape punch and reader for the output and input of CATY source programs. However, dumping or reading in of data from these CAMAC peripherals, in run mode, requires special routines of CAMAC commands to be written within the source program. The only in built run-mode facilities are a PRINT statement for typing data on the CAMAC VDU and a PRINTH statement for producing hard copy on the ASR 33 console device. This paper tape system which was used for all the earlier field trials at the RGO is described in greater detail by A.R. Hedge.[4]

4(c) Problems and Further Development

The ease of operation, greater power and flexibility of the disk system has the disadvantages of increased hardware cost, complexity and weight. For field trips the paper tape system was marginally more portable, even with its teletype. A better choice presently available, if portability and ease of use are the requirements, would be a cartridge tape system. (It should be noted that the floppy disk system was successfully transported to Israel, but this is probably not to be recommended).

In processing power the present system is probably most limited by the choice of software. With the paper tape system some crude responsivity correction was attempted using CATY but, as a choice of languages is possible with the disk system, it would probably be best to change to FORTRAN, once the data has been acquired, to do any necessary number crunching. Alternatively, it is possible to do all the necessary functions of control, acquisition and processing with a system such as FORTH.[5] This is fast, makes much better use of memory space than CATY, allows floating point arithmetic and at the same time can handle machine level statements. The apparent disadvantage of FORTH is that it is not DEC supported and requires that the ease and convenience of RT 11 be abandoned.

REFERENCES

1. The basic CAMAC specification is defined in a document EUR 4100 e (1972), "CAMAC A Modular Instrumentation System for Data Handling." Euratom.

2. CATY was originally developed specifically as a CAMAC testing aid at DNPL. It is presently marketed by Francis Golding Associates.

3. Digital, (1978). "Memories and Peripherals". (Digital Equipment Corporation).

4. Hedge, A.R. (1981). Ph.D. Thesis, University of Durham.

5. Moore, C.H. and Rather, E.D. (1977). "The Use of 'Forth' in Process Control." International '77 Mini-Micro Computer Conference, Geneva.

CHAPTER 6

OPERATION OF THE DETECTOR ON A TELESCOPE-SPECTROGRAPH SYSTEM

1. Introduction

In actual operation on an astronomical spectrograph various problems arise both in the connection of the detector to the spectrograph and in devising a workable operating procedure for data collection. To illustrate these points this chapter will begin with an outline of a full purpose-built detector system which has not been described elsewhere. Finally, examples of raw data obtained using different systems will be given with a discussion of further instances of non-ideal detector behaviour emerging from field trials.

2(a) Detector System for Operation on a Cassegrain Spectrograph

Following field trials of the detector on the coude' spectrograph of the 30" telescope at the RGO on three different occasions (the first of these with the equipment used is described by A.R. Hedge[1]), it was decided to build a system to use the Plessey Array at the Wise Observatory in Israel. The project to be undertaken, that of looking at the emission spectra of planetary nebulae, involved the observation of much fainter objects than had previously been looked at with the device. This consideration, together with the likely number of clear nights and suitability of the spectrograph for the work made the 1m. telescope at the Wise Observatory a reasonable choice. However, the fact that this telescope is purely a Cassegrain instrument and that the associated spectrograph camera has a very low f number made adaption of the cooled diode array system which then existed technically difficult. The principal problems encountered here being:-

(i) the construction of a cryostat system operating at a wide range of orientations;

(ii) the design of an optical coupling system to image the spectrum on the detector chip which, because of its surrounding cryostat or cold box, cannot be placed directly at the focus of the spectrograph camera;

(iii) the manipulation of the array to register and focus it accurately in a fast optical system; and

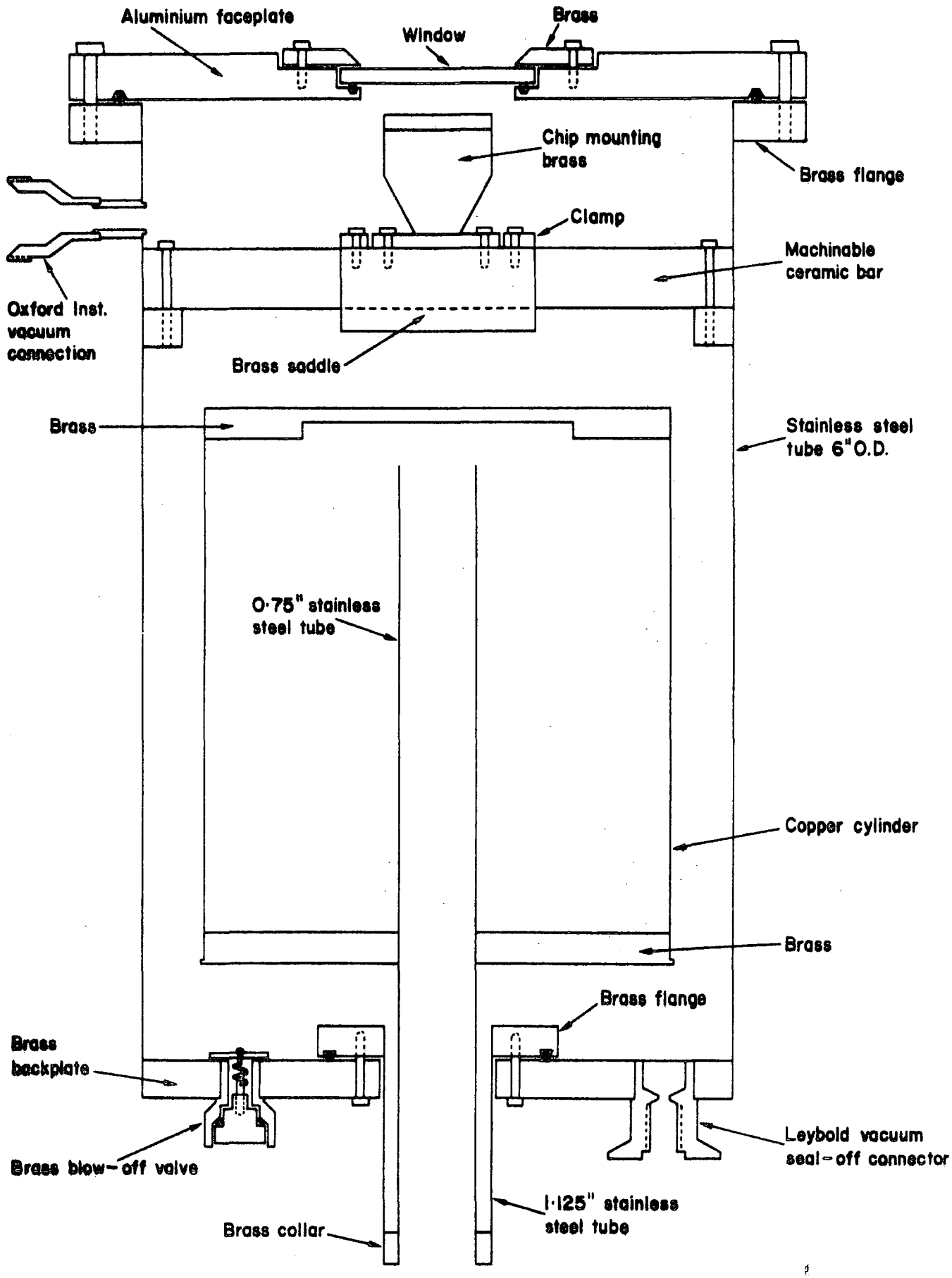
(iv) the provision of long lengths of cable and driver circuits in order that the device may be operated some distance from the control computer and heavier electronics such as power supplies.

2(b) Cooling System Design

As it was necessary to build a working system fairly quickly, the simple approach of the previous cooling system (see Chapter 3), that of keeping the device in contact with a reservoir of the coolant, liquid nitrogen, was retained. Alternative arrangements considered, such as pumping nitrogen boil-off through a heat exchanger connected to the chip mounting block, were dismissed as requiring too much time to develop. A nitrogen reservoir cryostat had already been employed by Geary[2] to mount a Reticon device on a Boller and Chivens spectrograph similar to that at the Wise Observatory. The final design of the cryostat of which Figure (6.1) is a sketch, is an adaptation from a drawing of a cryostat used by the RGO for a CID project.

The detector, mounted on a brass saddle, may be clamped at any position along the ceramic bar behind the front plate of the vacuum container, allowing either the upper or lower pair of arrays on the chip to be centralized in the cryostat window. This is done for two reasons: firstly, having the array in use, central with respect to the cryostat and its mountings, simplifies the lining up procedures, particularly that

Fig 6·1 CRYOSTAT



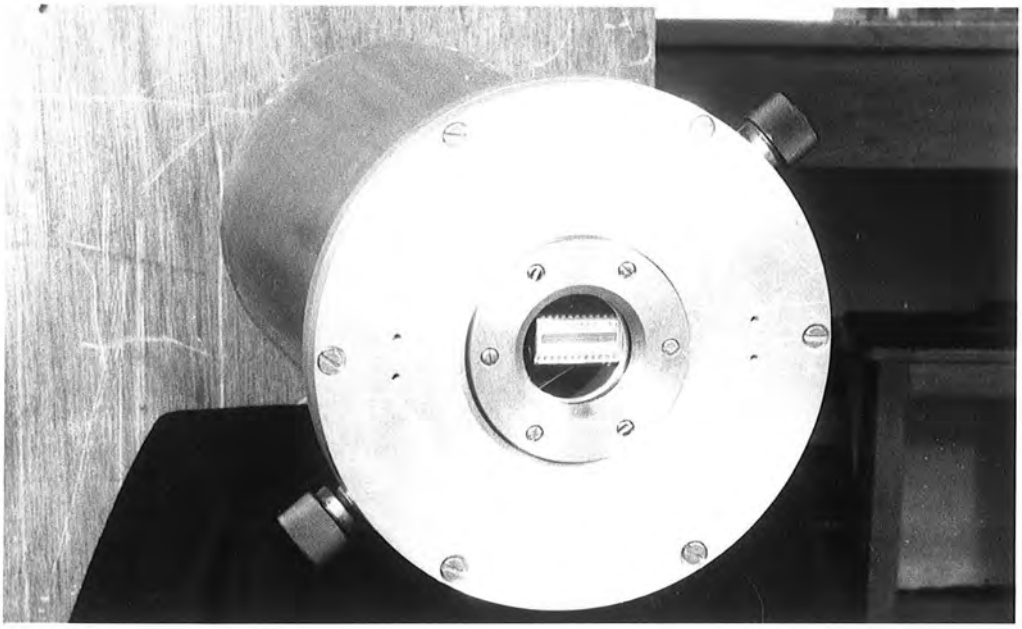
of rotation of the array in a plane perpendicular to the optical axis; and secondly, although the original spectrograph camera produces about 50 mm. of good spectrum, vignetting losses in the coupling system may considerably reduce the intensity of the spectrum away from the camera axis for which reason it is best to start off with everything as nearly lined up as possible, obviating the need for large amounts of adjustment to be incorporated in the mounting design. The machinable ceramic bar, used here as a rigid support for the detector, was first evaluated in the cryostat system developed for use at the RGO. This material: may be worked fairly easily; has been found not to distort significantly when the chip mounting attached to it is cooled; and has an acceptably low thermal conductivity.

Contact between the chip mounting block and the inner nitrogen vessel is made by carefully selected lengths of copper braid, the final stable temperature of the device being to some extent adjustable by variation of the amount of coupling used here. Once a configuration has been found, in a particular ambient temperature, giving the required minimum operating temperature normally used (about $-120^{o}C$), further adjustments may be made with the heater attached to the chip mounting block. In this sort of system the chip sits at some temperature between that of the coolant and the temperature of the surrounding environment, determined by the relative conduction of device's links to the cooling vessel (i.e. the braid) and its links to the outside, principally the lead-in wires and the ceramic bar. Radiation effects are ignored here. Clearly the ambient temperature is quite important in determining the chip temperature. With the cryostat used at RGO it was noticed that the chip temperature fell considerably when the nitrogen dewar was topped up, even if the dewar was still fairly full and the temperature had been stable for a long time. This was thought to be due to boil-off and splashes cooling the outer cryostat body thereby reducing the apparent

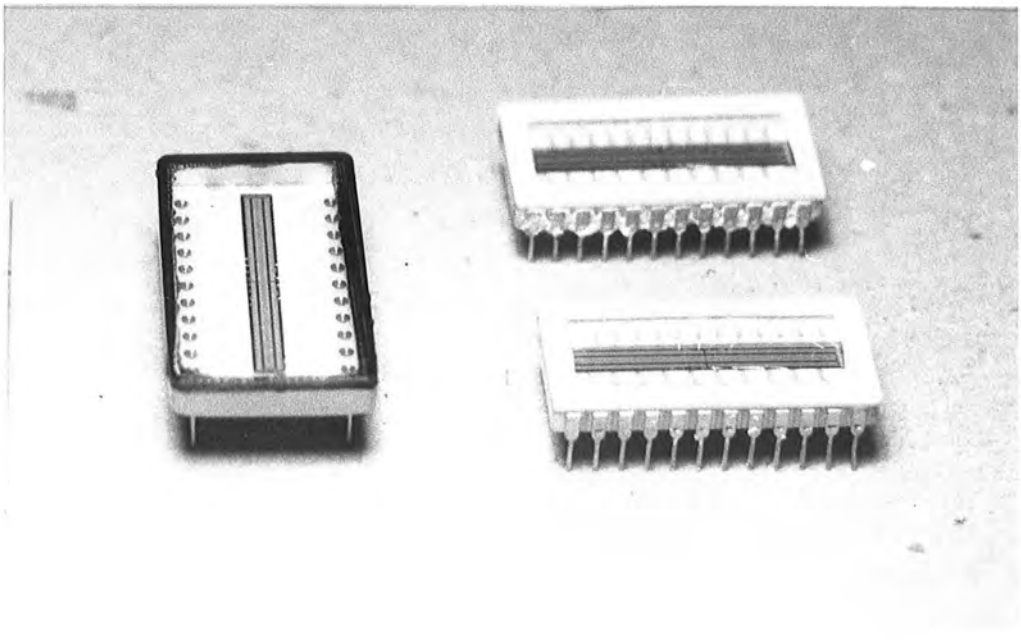
ambient temperature. In this design the boil off is directed away from the apparatus by the re-entrant tail pipe. However, the chip may still be affected by temperature changes in its surroundings and some form of active temperature control would be desirable on long integrations. The device temperature is sensed by a pla tinum resistor mounted below the device package. A proprietary unit is used for this, in preference to the thermocouples incorporated in earlier systems, because a platinum resistance being current driven may be operated some distance from its readout electronics, with a simple two wire connection. All electrical connections to the chip, heater and sensor are made through two M type vacuum connectors mounted on the outside of the cryostat vessel.

The construction of the inner nitrogen reservoir follows the original RGO drawing. So that the temperature of the endplate,(to which the copper braid is attached,) does not change significantly with nitrogen level and vessel orientation, the endplate itself is made of a fairly heavy piece of brass and the cylindrical wall from copper plate for high thermal conductivity. The main system heat loss is through the double tail pipe assembly, which is fabricated from an austenitic grade of stainless steel, chosen for its relatively poor thermal conductivity. Since the inner nitrogen reservoir has a considerable surface area, the effects of radiation losses cannot be ignored. In this design,radiation between the inner and outer vessels is reduced by covering their opposing surfaces with aluminized mylar. To help maintain an adequate vacuum for long enough to carry out a whole night's observing without it being necessary to reconnect the vacuum pump periodically, a getter material, zeolite is used. This material is contained in a sieve fixed to the lower end of the nitrogen reservoir and has the property that,when cooled,it mops up any air entering the cryostat through small leaks.

Initial evacuation of the unit prior to cooling is done by



(a) Cryostat



(b) Chips

Fig 6·2

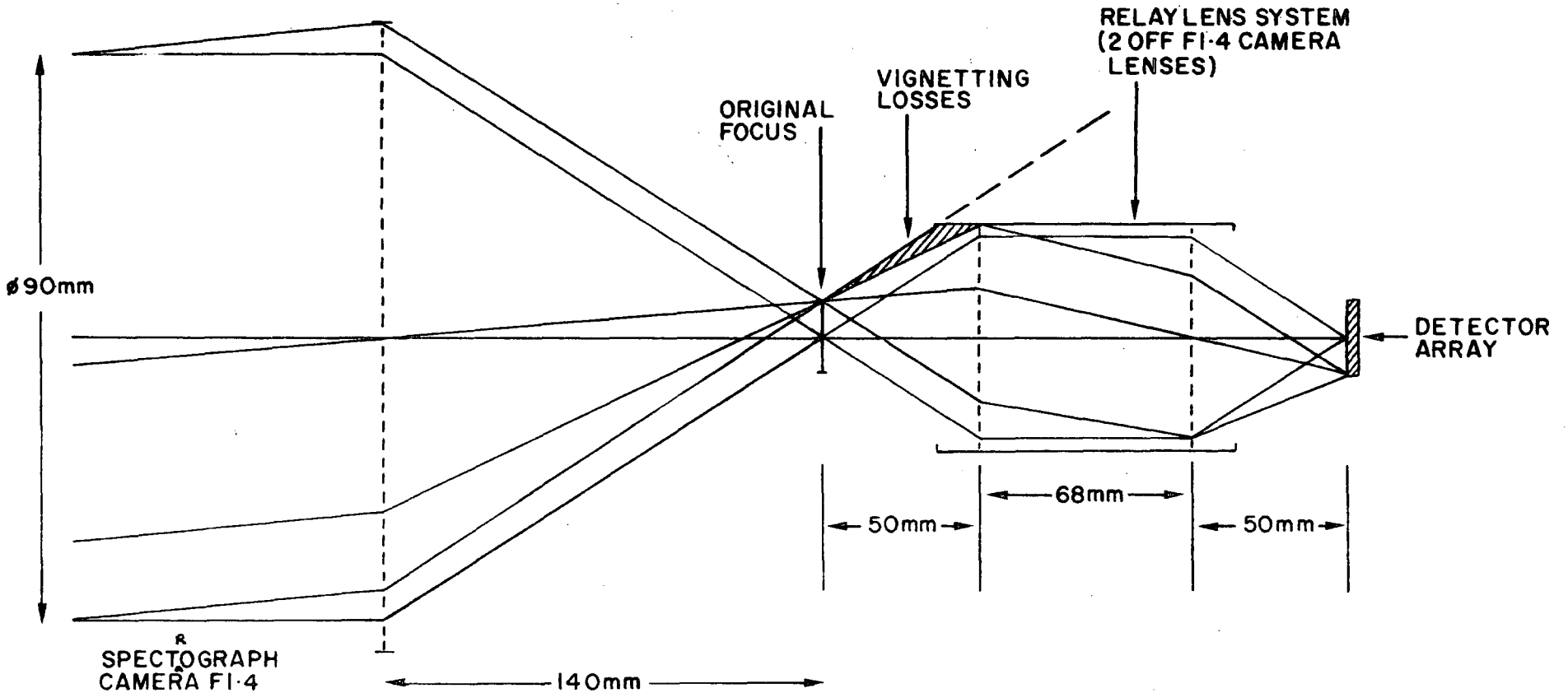
connecting a vacuum pump to the Leybold vacuum seal-off connector mounted on the rear plate of the cryostat. An emergency blow-off valve is also provided in case the vessel is ever left to warm up, having first been allowed to reach atmospheric pressure when cold and then inadvertently resealed. There is also a possibility that the inner nitrogen vessel might rupture in use, due to icing up of the tail pipe, although it has been found in practice that the dry nitrogen boil off, passing steadily out of the reentrant neck, effectively prevents this.

Figure (6.2(a)) is a photograph of the finished cryostat, showing the location of the diode array behind the front plate window. The positions of the two M type vacuum connectors can also be seen. A view of the two types of device package used is given in Figure (6.2(b)).

2(c) Relay Lens System

Since the system magnification and dispersion of the existing spectrograph camera were well suited to the application envisaged, it was attempted to couple this system to the detector with a simple 1:1 relay lens system. A diagram of the optical set-up is shown in Figure (6.3). The dispersion at the first focus, at which in normal use a photographic plate would be placed, is 68Å/mm equivalent to 3.4 Å/diode (i.e. the best possible system resolution). The actual resolution depends upon the projected size of the object observed, if this is greater than the array pitch. Closing down the entrance slit of the spectrograph in order that the wavelength resolution is always determined by the spatial resolution of the array wastes light and,if faint extended objects are to be observed,a trade-off may have to be made between wavelength resolution and signal. For a typical apparent size of planetary nebula of 15 arc sec the projected image size at the focus of the spectrograph camera is given by

Fig 6·3 COUPLING OPTICS



$$\text{Projected Size} = \frac{15 \text{ arcsec}}{648 \times 10^3} \times 1 \text{ m.} \times f_T \times \frac{f_{SP}}{f_{COLL}} = 110 \ \mu\text{m},$$

where f_T = focal ratio of the 1 metre telescope, 7,

f_{COLL} = focal ratio of the spectrograph collimator, 7,

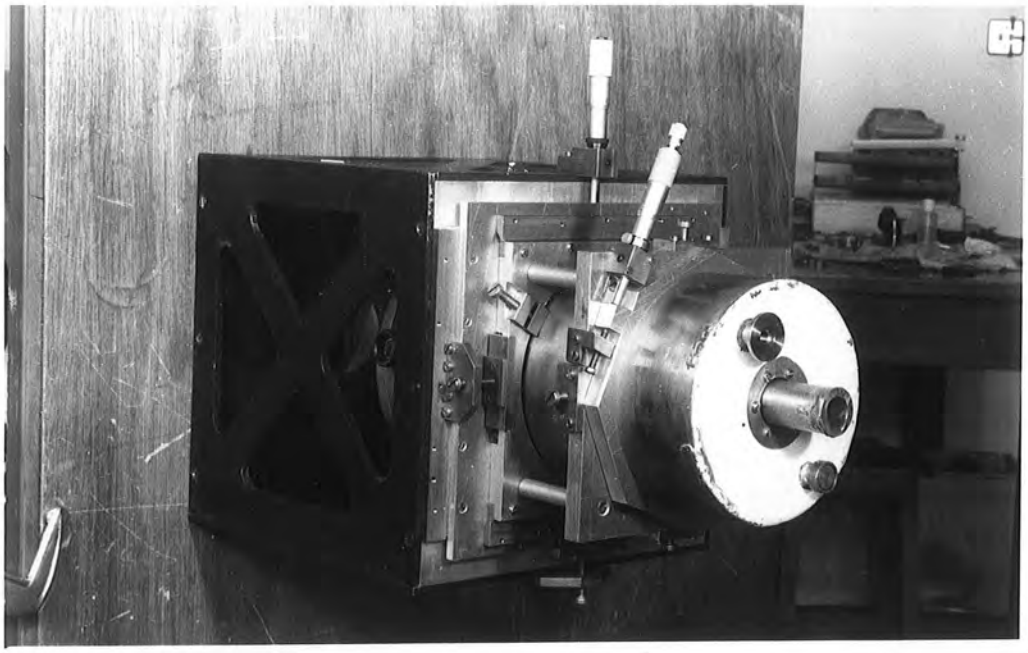
and f_{SP} = focal ratio of the spectrum produced, 1.56.

This projected image size is equivalent to 2.2 diode spacings resulting in an effective wavelength resolution of about 7.5 Å for the system. Recalling that the Plessey array has a 200 μm width, the spectrum produced from such an object should not overfill the sensitive area. As resolution is not limited by array pitch, nothing is to be gained from a relay lens arrangement which increases the size of the projected image and therefore the dispersion. On the other hand, further demagnification would be difficult to achieve with inexpensive camera lenses. The solution adopted was to use two f1.4 camera lenses in combination. These lenses each have a focal length of 50 mm and when mounted together have been measured as having a separation between principal planes of $\sim$ 68 mm. Using this compound lens in a 1:1 geometry produces a second image of the spectrum $\sim$ 168 mm away from the focal plane of the original camera. This second focal plane is situated $\sim$ 13 mm clear of the relay lens, leaving adequate space for the cryostat front plate and window.

2(d) Registration and Focusing

Since the relay lens described above does not change the final focal ratio, focusing will still be fairly critical. The spectrum is produced from a collimated beam of 90 mm diameter with a spectrograph camera lens of focal length 140 mm giving a focal ratio of $\sim$ 1.56. Thus, for defocusing not to degrade the spatial resolution by more than one pixel, it is necessary to place the detector in the focal plane to within an amount

$$\triangle S = 50 \mu\text{m} \times 1.56 \simeq 80 \ \mu\text{m}.$$



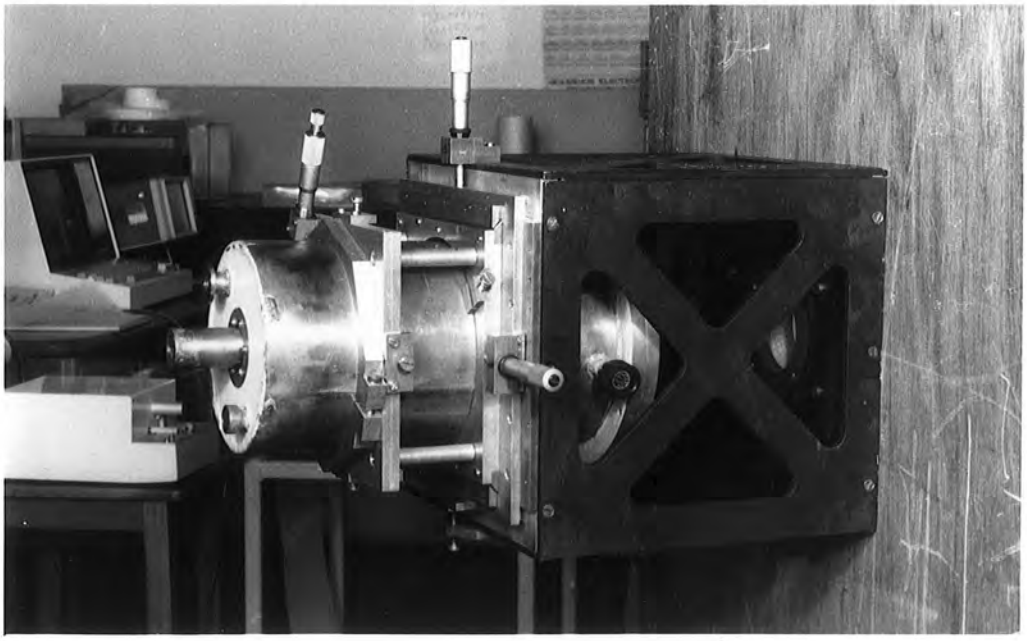


Fig 6·4 CRYOSTAT MOUNTINGS

An equally serious effect of defocusing will be loss of light off the sides of the array. In the example already quoted of a spectrum of width (i.e. the dimension perpendicular to the dispersion) 110 μm, if this is placed centrally in the array aperture of 200 μm, 45 μm of unused sensitive area will be left either side of the image to accommodate movements (as described in Chapter 1) and defocusing. To avoid loss of light in this situation would require that the array be positioned to within 140 μm of the focal plane. It should be noted that for redder wavelengths the device will have an optical thickness, given by the characteristic penetration depth and the refractive index of the medium, of $\sim$ 100 μm, such that perfect focusing is not possible in any case.

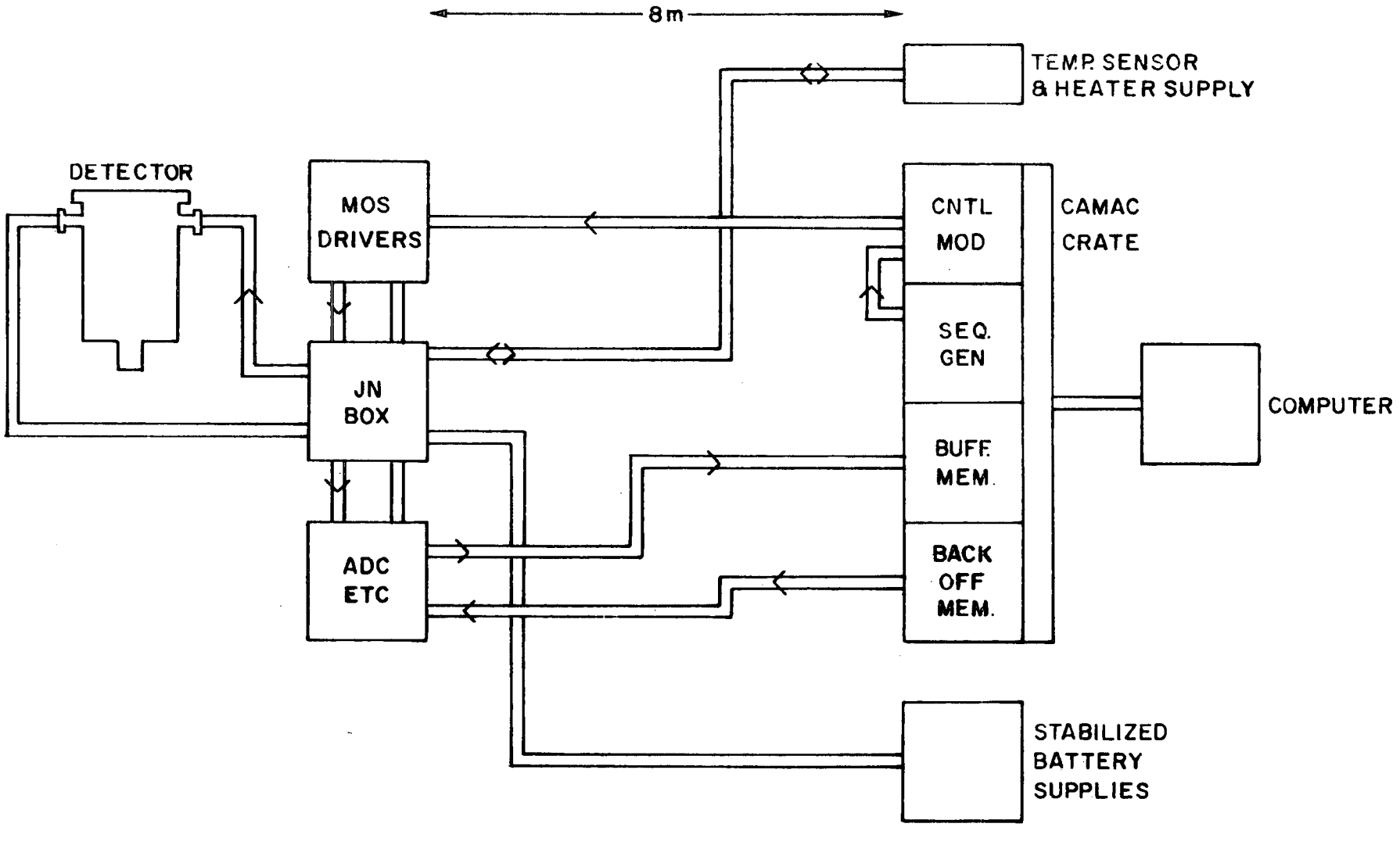
Registration and focusing are accomplished by mounting the whole of the cryostat in slides. Figures (6.4(a)) and (6.4(b)) show the crude mounting arrangement employed, with which the device may be: moved in both the X and Y directions in the plane perpendicular to the optical axis; rotated; and moved in and out for focusing. All these movements may be measured with micrometer gauges, separate clamps being provided to lock each of these adjustments to minimize movement during observations due to free play in the mounting slides.

The above illustrates the mounting problems encountered with a Cassegrain system differing considerably from previous experience on the Coudé spectrograph at RGO. Since the focal ratio of the Coudé system was $\sim$ 47, focusing was not at all critical. The fact that the detector may sit on a horizontal bench simplifies the construction of the cryostat and mountings.

2(e) Organization of the Electronics

From consideration of the telescope and working area inside the dome, it was apparent that the array would have to be operated about 8 metres away from the computer and other rack mounted electronics.

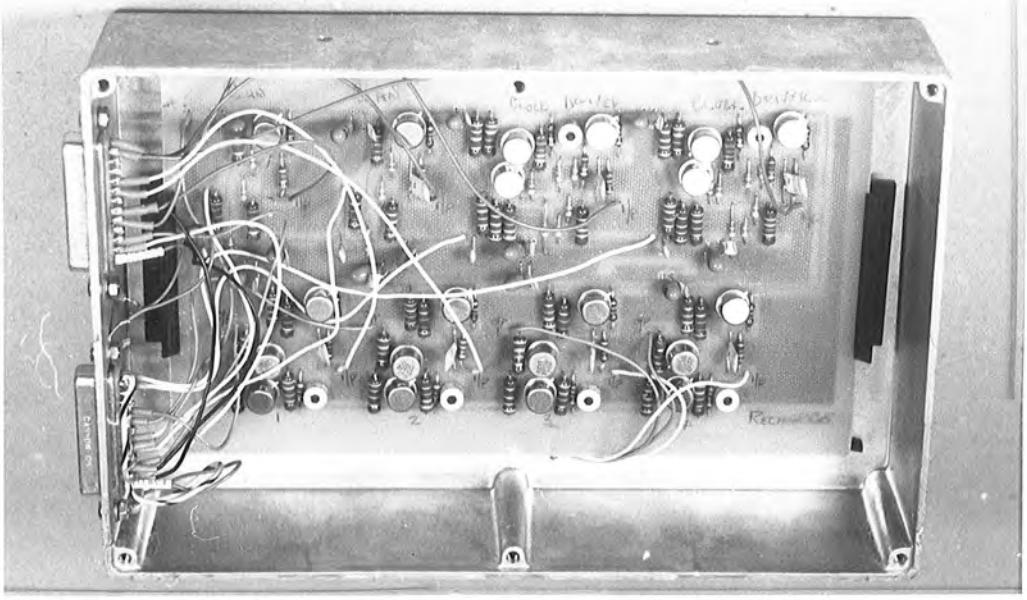
Fig 6-5 ORGANIZATION OF ELECTRONICS



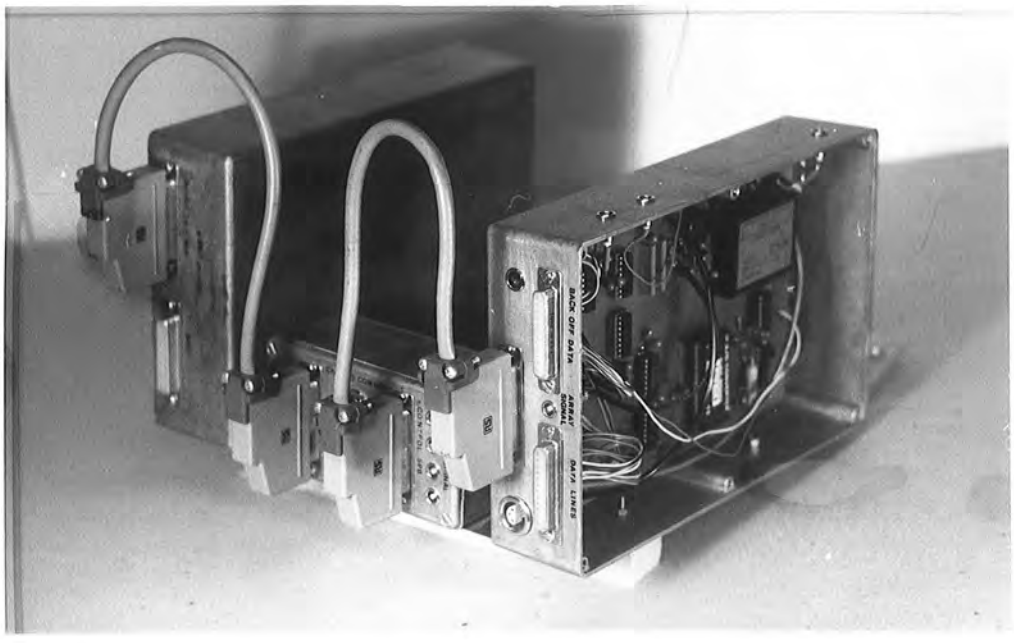
To overcome this problem, the various driving and acquisition functions are separated in such a way that only purely digital signals need be sent along the connecting harness together with the necessary stabilized supply voltages. A schematic diagram of this organization is provided by Figure (6.5), part of which, the acquisition module, was shown in Figure (4.5(b)). Figure (6.6(a)) is a photograph of the TTL to MOS converters and Figure (6.6(b)) gives a view of all the modules which are bolted as a unit to the detector-spectrograph assembly. Those heavier units like the computer, floppy disk drives, CAMAC crate, battery power supplies and battery chargers are mounted in lightweight 19" rack units as illustrated in Figure (6.7). Special TTL buffer modules were used to drive the long multicore control cables. These buffers were found to be necessary because the larger current demand of gates employed to drive cables gave rise to crosstalk within TTL packages. In general, it is best if packages having gates driving into cables perform no other function within a module, (i.e. their other gates are not used). This solution was allowed even to the extent of building simple modules dedicated to line driving, thereby easing decoupling problems. Special line driving chips and receivers were not found to be necessary, the above system performing satisfactorily in the electrically noisy environment of a telescope dome.

3. Observing Procedures, and Initial setting up

The apparatus described in the preceding section is the most sophisticated used to date. The computing system includes the facility of a dual floppy disk drive for program and data storage, previous systems operated at the RGO having relied solely on paper tape. Since the acquisition programs were originally developed at the RGO, they were necessarily designed to dump as little data as was absolutely necessary to avoid accumulating large amounts of punched tape. This problem is to some extent overcome when floppy disks are employed as the storage



(a) TTL to Mos converters



(b) Front end electronics

Fig 6·6

medium, although only about 19 full dumps of all the run data (i.e. 2048 data points) may be placed on one disk; it has been more usual to retain only the 256 reduced data values from the final double sample readout allowing many more runs to be stored without editing the data.

For purposes of later analysis of the data it is necessary to take a flat field or known spectrum, usually provided by a tungsten lamp, in order to be able to divide out responsivity variations due to inter-element variations, instrumental effects and misalignment of the array. To remove all these effects, particularly misalignment, this known spectrum should have exactly the same width and be placed on the array in the same position as the spectrum from the object observed. In practice, it is not possible to do this exactly, but the appearance of the spectrum is generally significantly smoothed by this division. The problems found here are that the apparent size of an object such as a star can be extremely small depending on seeing conditions and that, during a long exposure, the actual position of the object with respect to the entrance slit of the spectrograph may change. Looking at a tungsten source, the entrance slit is filled with light resulting in a spectrum which is larger in the direction perpendicular to the dispersion and probably more uniform than that of the star.

To determine precisely the wavelength interval seen by the detector array, standard lamps giving emission spectra of Neon and Argon are used. Groups of emission lines may be identified from their spacings and relative intensities, enabling the spectrograph grating angle to be preset in any region of interest. This was particularly important for work done on the Coudé spectrograph at the RGO where the dispersion was fairly high (~ 7 Å/mm) with the result that very small grating movements produced large shifts in the position of features on the array. Actual identification, in terms of wavelength, of emission features from standard lamps is not always required for later data analysis, since projects have

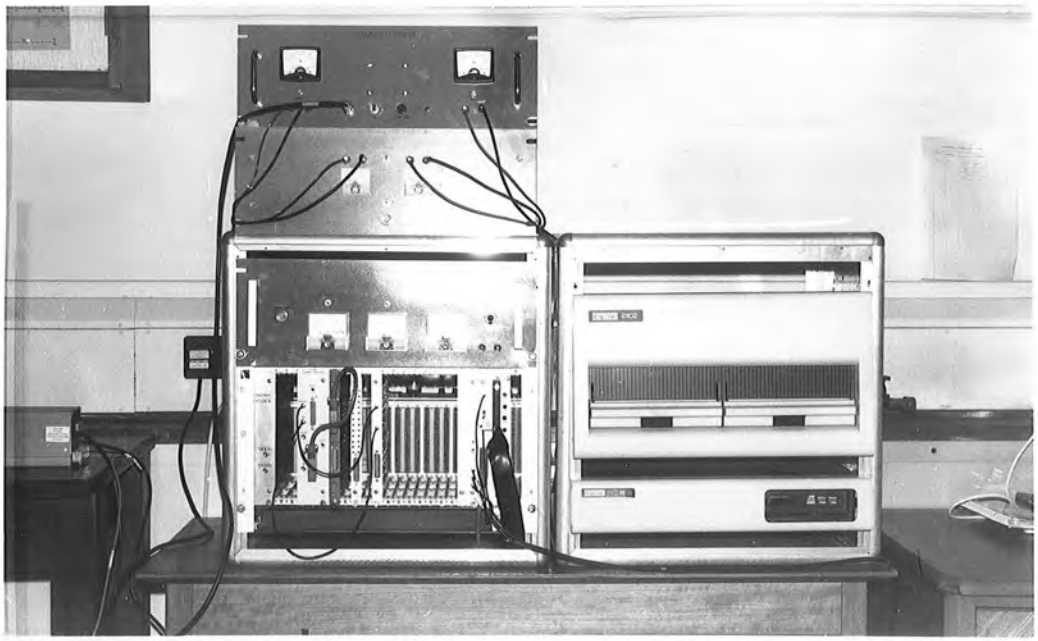


Fig6·7 CAMAC SYSTEM

usually involved the study of stellar features of known wavelength. However, standard lamps assist greatly in finding features in the first place and in performing small grating shifts. In situations where the wavelength resolution is limited by the array pitch, performing several similar integrations at slightly different grating positions (i.e. shifting the spectrum on the array by other than an integral number of diodes thereby changing the phase relationship of the pixels to the spectrum) may enable a spectrum of higher resolution to be obtained from the combined data. In any event, similar spectra taken at different array positions enable genuine features to be distinguished from array responsivity effects which cannot always be completely removed by division with a known flat spectrum.

From the foregoing it can be seen that a single observation typically comprises the taking of several spectra, including a tungsten spectrum at each grating setting, possibly some wavelength calibration spectra and several runs on the actual object.

Initial setting up of focusing and registration is done with the aid of the tungsten and calibration lamps. Typically, a tungsten is first used at the widest possible slit width and length. The array output is monitored in real-time to look for signal, rough lining up being done by trial and error. Having obtained some signal, the slit length is successively decreased and the adjustment procedure repeated. When reasonable registration is achieved, the focusing, which is at first set by rough calculation, may be adjusted. A common approach is to monitor an emission feature from a calibration lamp in real-time, adjusting the focus to obtain a feature which falls on the minimum number of elements. Some further adjustment of registration, particularly rotation of the chip in a plane perpendicular to the optical axis, must then be done. At this stage it is usual to use as small a slit as possible and short integrations, since no signal is detectable by real-time monitoring.

4(a) Practical Problems encountered during Field Trials

Before the first field trial of the detector system at the RGO, the device performance had only been checked using very simple optics. In the responsivity tests in Chapter 3 the array is flooded with light, no imaging system being used. Thus the array had always been overfilled with diffuse light. Operating the detector with a spectrograph image smaller than the array aperture had the expected effect of making the device more sensitive to dirt on its window and the cryostat window, - such blemishes producing marked responsivity variations. Less expected was a pronounced odd-even effect, which varied with the position of the light on the array. Other undesirable effects noticed under observing conditions are: an apparent insensitivity to weaker signals, drifts in the baseline of a spectrum, and, under certain conditions, a kind of image retention.

4(b) Odd-Even effect

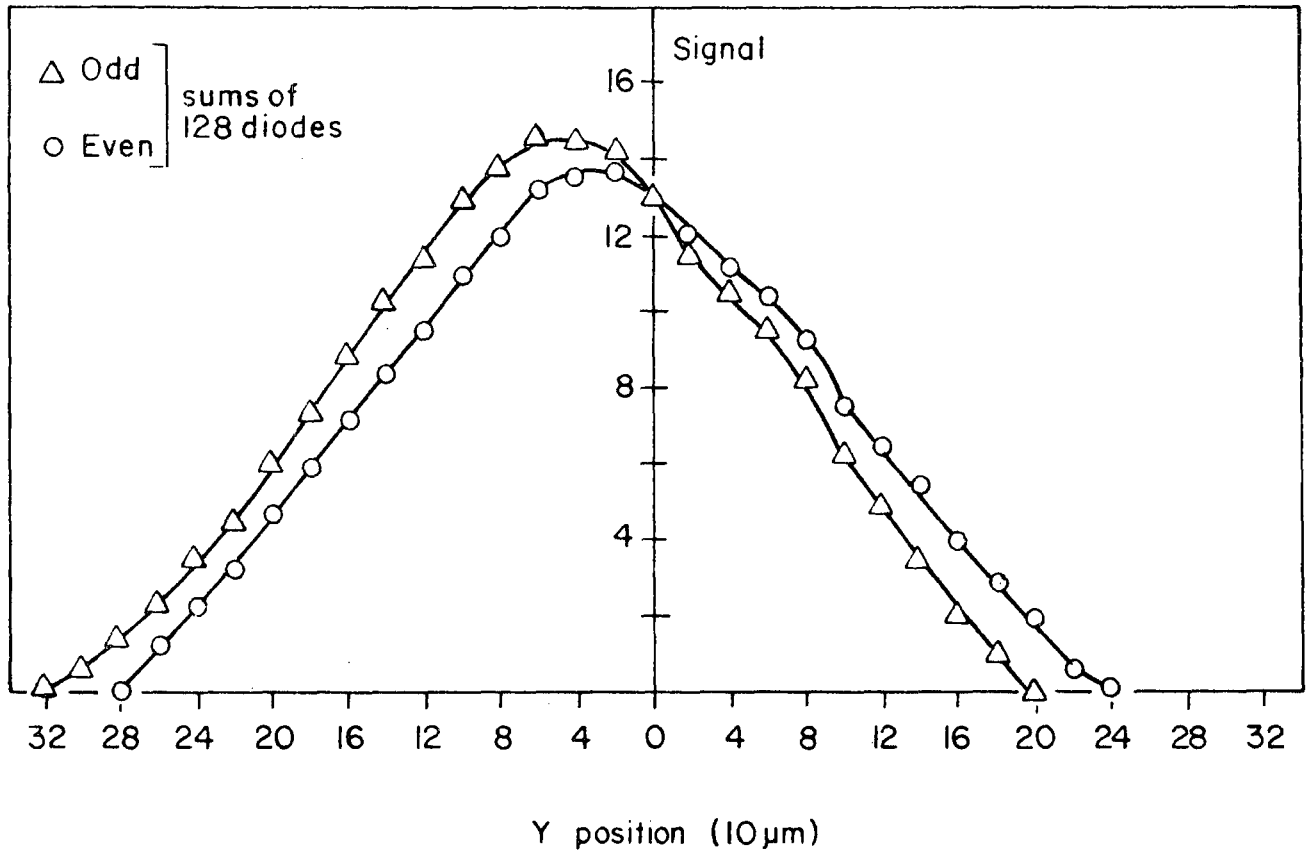
Some degree of odd-even effect is almost inevitable in this type of array, the on-chip electronics for alternate elements being placed on opposite sides of the array and served by different recharge pulses. Thus any effective difference in height or width between CP1 and CP2 may cause a periodic effect. With this particular array where alternate pairs are enabled by different clock pulses, an effect with a periodicity of four diodes is also expected and may be seen in particular devices. Also, in Chapter 4 it was mentioned that the video output has a characteristic shape which introduces a negative signal offset under zero signal conditions, the effect being greater on the odd diode signals.

However, the effect noticed during operation at the RGO changed with the position of the spectrum in the direction perpendicular to the array (usually termed the Y direction). After obtaining a reasonably well aligned tungsten spectrum, it was possible, using the micrometer Y

adjustment, to measure the aggregated output of all odd and all even elements as a function of Y position. The odd-even effect vs. Y is plotted in Figure (6.8). Using a perfectly aligned flat-field image which exactly fills the array, movement in the Y direction would be expected to produce a linear fall off in signal, going to zero at a displacement of $\pm$ 200 μm from the central position. As this array was not designed for optical use, the active area is not sharply defined by an on-chip mask and the rounded shape of Figure (6.8) is quite acceptable especially at longer wavelengths. Slight misalignment and nonuniformity of the tungsten source will also prevent an ideally triangular result being obtained. Recalling that the peripheral electronics associated with odd pairs is placed on just one side of the array and the even circuitry on the other, it appears that light impinging on this electronics either adds to or enhances the signal from the appropriate elements. It is interesting to note that the central position as indicated by maximum signal is not the point of zero average odd-even effect. As it is known that there exists an effect giving odd diodes a greater negative offset, the crossover point might be anticipated to be shifted in favour of even high, but in this experiment the reverse was found. It is not known whether this is the case for all arrays. Pressure of time when using the spectrograph limited the amount of purely instrumental work done, astronomical data-taking being given priority. Further work is required to decide conclusively whether the odd-even effect is signal dependent or merely an offset. Here signal dependence must be further separated into integrated signal and incident signal effects.

Although odd-even differences are in many ways a nuisance requiring some mathematical manipulation for their removal at the data analysis stage, this phenomenon is also quite useful for purposes of alignment,

Fig 6-8 ODD-EVEN EFFECT

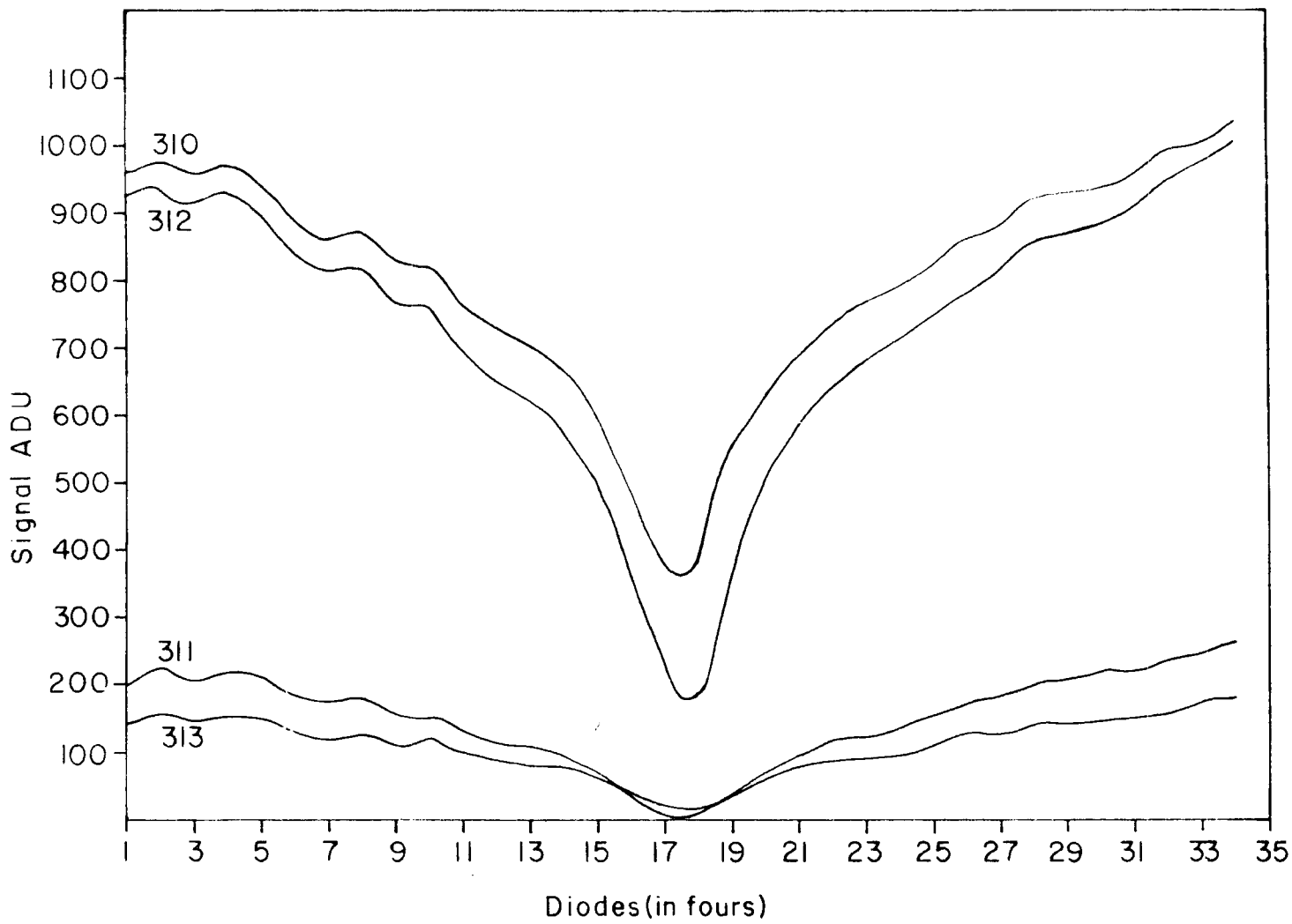


it being immediately obvious when the image has for various reasons moved off the array. If a star or similar object is being observed, an integration resulting in far less signal than expected may be attributed to varying atmospheric conditions. However, if the reason for low signal is misalignment, the odd-even effect will be seen. Thus an alignment check is provided which could only otherwise be done by trial and error movements or placing special registration arrays perpendicular to the observing array.

4(c) Repeatability and Reciprocity under Observing Conditions

One of the first instrumental tests tried at the RGO, using a bright star, Vega under fairly good seeing conditions, was a rough repeatability and reciprocity experiment. Figure (6.9) shows four Vega spectra taken on the same night within an interval of about 15 minutes. The graph shows the part of the spectrum around the Hα feature for four exposures of equal length where, in two of the observations, a 50% neutral density filter has been placed in the light path. (The numbers 310-313 are the experimental run numbers for the observations and indicate the order in which they were made). Attenuating the incident signal by half clearly reduces the total integrated signal by much more than a factor of two. As it has been suggested that the array output may be subject to uncertain drifts in d.c. level, the signal in Figure (6.9) is probably best estimated from the depth of the Hα feature. Total signal will vary slightly between similar integrations as a result of changes in seeing conditions, but alternating between full and half intensity exposures allows any general trend in incident illumination level to be seen. If it is assumed that the incident signal is falling off slightly throughout the experiment, the appearance of the spectra is consistent with the response curve being slightly 'S' shaped. The run done at the time of the largest

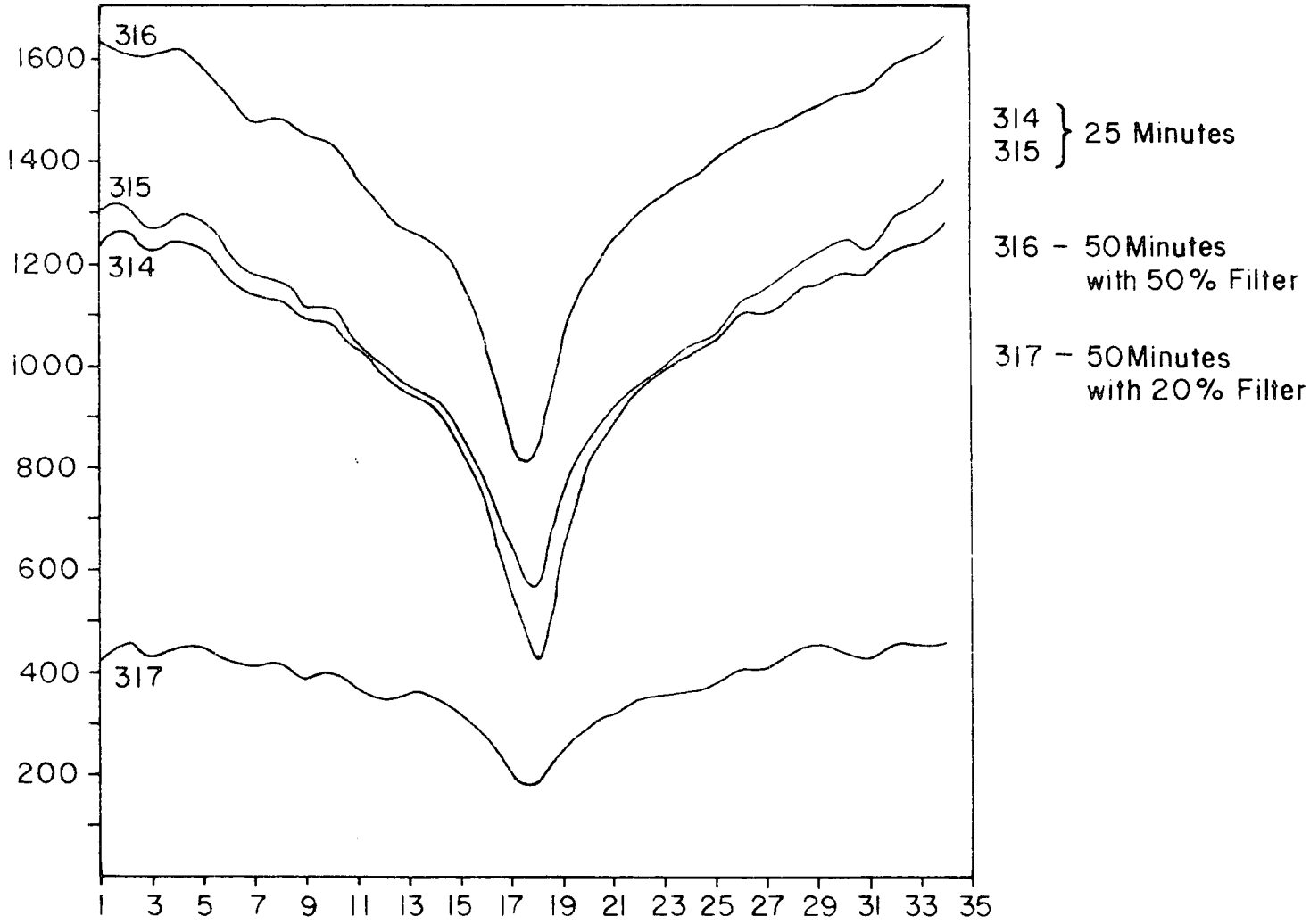
Fig 6·9 H α FEATURE IN VEGA



incident signal,(Run 310), is attenuated because the output stage is beginning to saturate with the result that a similar exposure at a later time,(Run 312), when the incident signal is slightly less, exhibits a deeper absorption feature. At the lower end of the characteristic,the slow start effect (mentioned in Chapter 3) considerably reduces the gain of the detector so that a decrease of incident signal of approximately half (as in Run 311 and 313) yields a very small output signal. This illustrates that,although the response may not be completely linear, this fact does not necessarily matter if small absorption features are to be studied in a bright continuum, as only local linearity of the output characteristic is required. A paper has been written on observations of the OI λ 7773 triplet in intermediate type supergiants,[4] the results of which compare well with those of other workers. However, it might be expected that the array would not perform very well when employed to measure emission features in the presence of very little continuum and experience attempting to look at just such phenomena in the spectra of planetary nebulae supports this.

In Chapter 3 the hope was expressed that linearity would improve with lower incident intensity. The reciprocity experiment was, therefore, tried with much lower intensities and correspondingly longer integration times (i.e. up to 50 minutes). Problems of variation in atmospheric conditions and the ability of the telescope system to track the star become much more pronounced under these conditions. Figure (6.10) shows observations of Hα feature in Vega taken over an interval of about 3 hours. Runs 314 and 315 are consecutive 25 minute exposures taken, once again, to check repeatability. Following these measurements,a filter was placed in the light path to further attenuate the beam by 50% and a 50 minute exposure, Run 316 was taken. This doubled exposure should result in the same total signal being accumulated,but this was not found to be the case,although it may be said that the absorption line has

Fig 6·10 H α IN VEGA



a similar depth. Considering the unpredictable nature of the light source, this sort of experiment is probably best done in the laboratory. A possibly better experiment is to look at particular emission lines from the standard lamps, using neutral density filters to check reciprocity. Figure (6.11) shows typical results obtained by plotting the area under a sharp neon line against filter transmission. These results are more encouraging but, once again, suggest a negative signal offset.

4(d) Image Retention

In attempting to do the measurements on neon lines, it was noticed that some image retention occurred with the result that, if the readings were done in quick succession, the measured signals would depend on the signal in the previous integration as well as the present incident illumination. This meant that, if a high illumination level had just been observed, a following integration on a weak source produced more signal than expected. It was decided to investigate this further using neon lines as these have an easily recognisable position and shape. Thus after an exposure to a sharp neon line and subsequent multiple re-charges (i.e. typically 1 minute of destructive scans or 600 recharge operations), an integration was performed in the dark to see if any further signal emerged. Monitoring the array with non-destructive scans revealed a slow build-up of a signal around the position of the sharp neon line of the previous exposure. Figure (6.12) shows a neon line as obtained with an integration time of 1 frame (1/10th second) together with the signal recovered in an immediately following dark integration of 2500 frames, the retained signal being multiplied by 10 for purposes of display. Although image retention was not unexpected with this device, the spreading out of the retained image was quite surprising. In the example shown, the original sharp line covers at most 6 elements,

Fig 6·11 AREA UNDER EMISSION LINE VS. TRANSMISSION

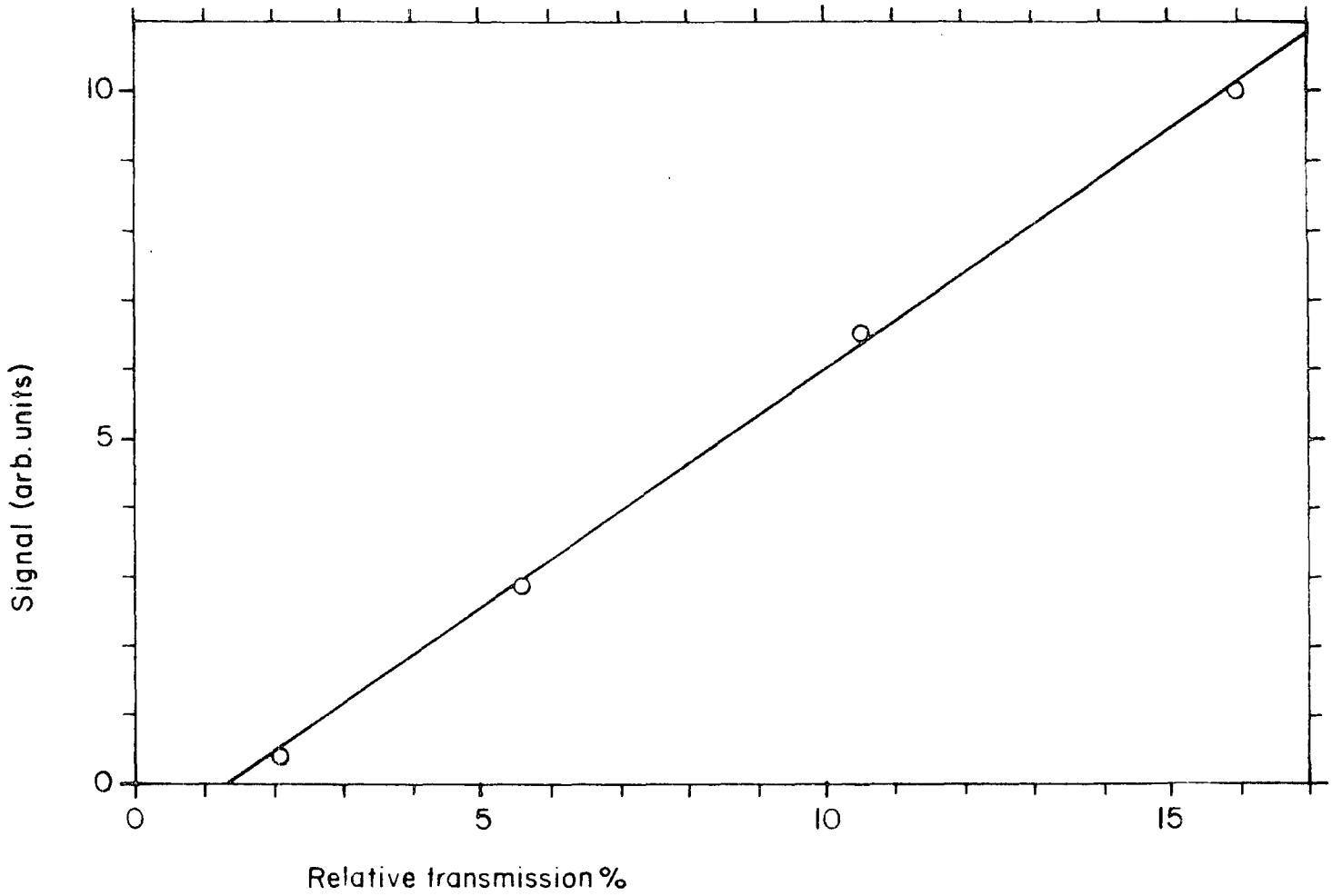
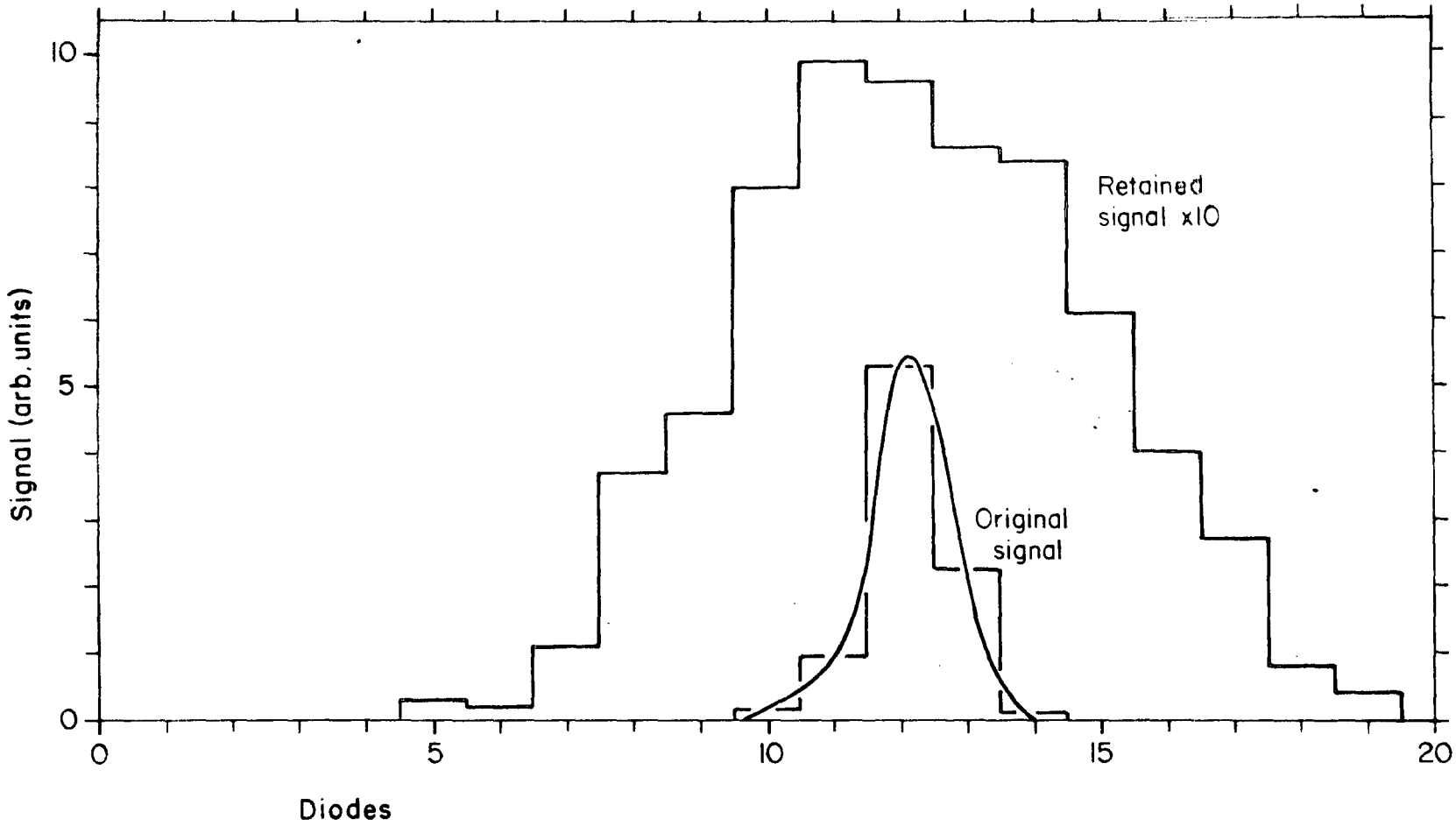
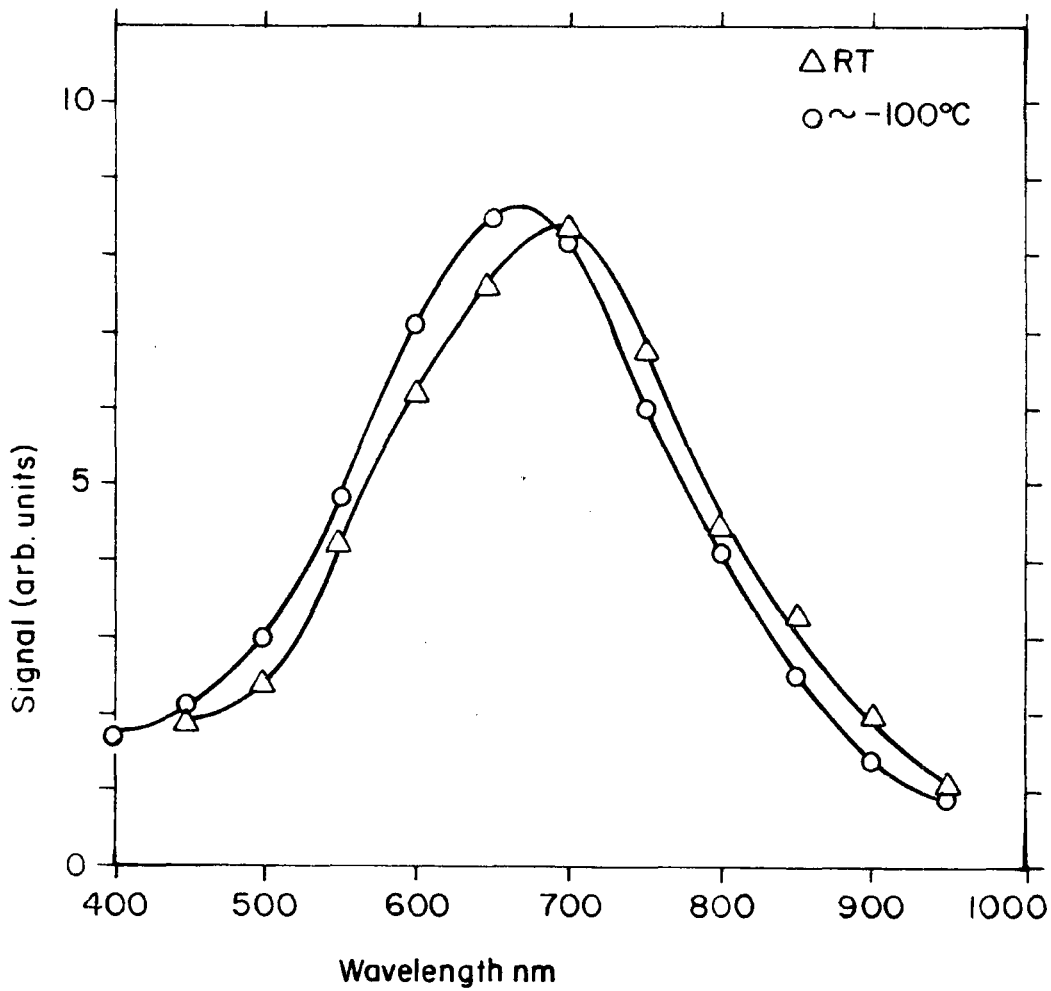


Fig 6·12 IMAGE RETENTION



but the retained image is spread over about 25 elements. To find out when the spreading occurred, a long dark integration was started and the neon lamp was turned on for just 10 seconds near the beginning of the run. In the time after the lamp had been switched off the signal was monitored non-destructively. Further growth of the observed emission lines due to lag effects within the device could be seen, but no spreading occurred at this time. It was concluded that the phenomenon was connected with the recharge operation. Recalling the suggestion of various authors, including Vogt et al[5], that diode arrays might be subject to self heating effects on read-out, it was suggested that in this device local heating might occur at the time of recharge, this heating being dependent on the amount of signal on the particular diode. In this way a sharp emission feature could cause a local hot spot, resulting in significant thermal signal being generated in the area around the original feature for some time afterwards. If this is the case, merely exposing the array to light while it is being recharged should produce an effect which increases with the number of recharges performed under illumination. This was investigated by recharging the array in successive frames while under illumination with a sharp neon line. The neon exposure was controlled with a timer. Immediately the lamp exposure ended, a long dark integration was begun to recover any retained signal. Results so far suggest that the retained signal is not affected by the length of the neon exposure as described above, further investigations indicating that only the accumulated signal in the last integration and the number of recharges since that integration are significant. The implication of this effect for general observing work would seem to be that all exposures should be separated by a few minutes of continuous recharge cycles, possibly with dark integrations, to check that all trace of the last signal has gone.

Fig 6·13 APPROXIMATE SPECTRAL RESPONSE



5. Measurement of Spectral Response

The spectral response of the device has been measured using laboratory facilities available at the RGO. Figure(6.13)shows the relative spectral response measured at room temperature and at the normal operating temperature $\sim -100^{o}C$. This is possibly only a rough experiment as the response of the bolometer, used to measure the output of the monochromator, is assumed flat. However, the wavelength shift in peak responsivity on cooling the Plessey device is clearly seen.

REFERENCES

1. Hedge, A.R. Ph.D. Thesis, University of Durham (1981).

2. Geary, J.C. Ph.D. Thesis, University of Arizona (1975).

3. Hedge, A.R. Op.Cit.

4. Hopkinson, G.L. and Humrich, A. "Observations of the OI λ 7773 triplet in intermediate-type supergiants with a linear photodiode array." Monthly Notices. (to be published).

5. Vogt, S.S., Tull, R.G. and Kelton, P. (1978). "Self-Scanned Photodiode Array: High Performance operation in High Dispersion Astronomical Spectroscopy." Appl. Opt. Vol.17, pp.574-592.

CHAPTER 7

CONCLUSION

1. Introduction

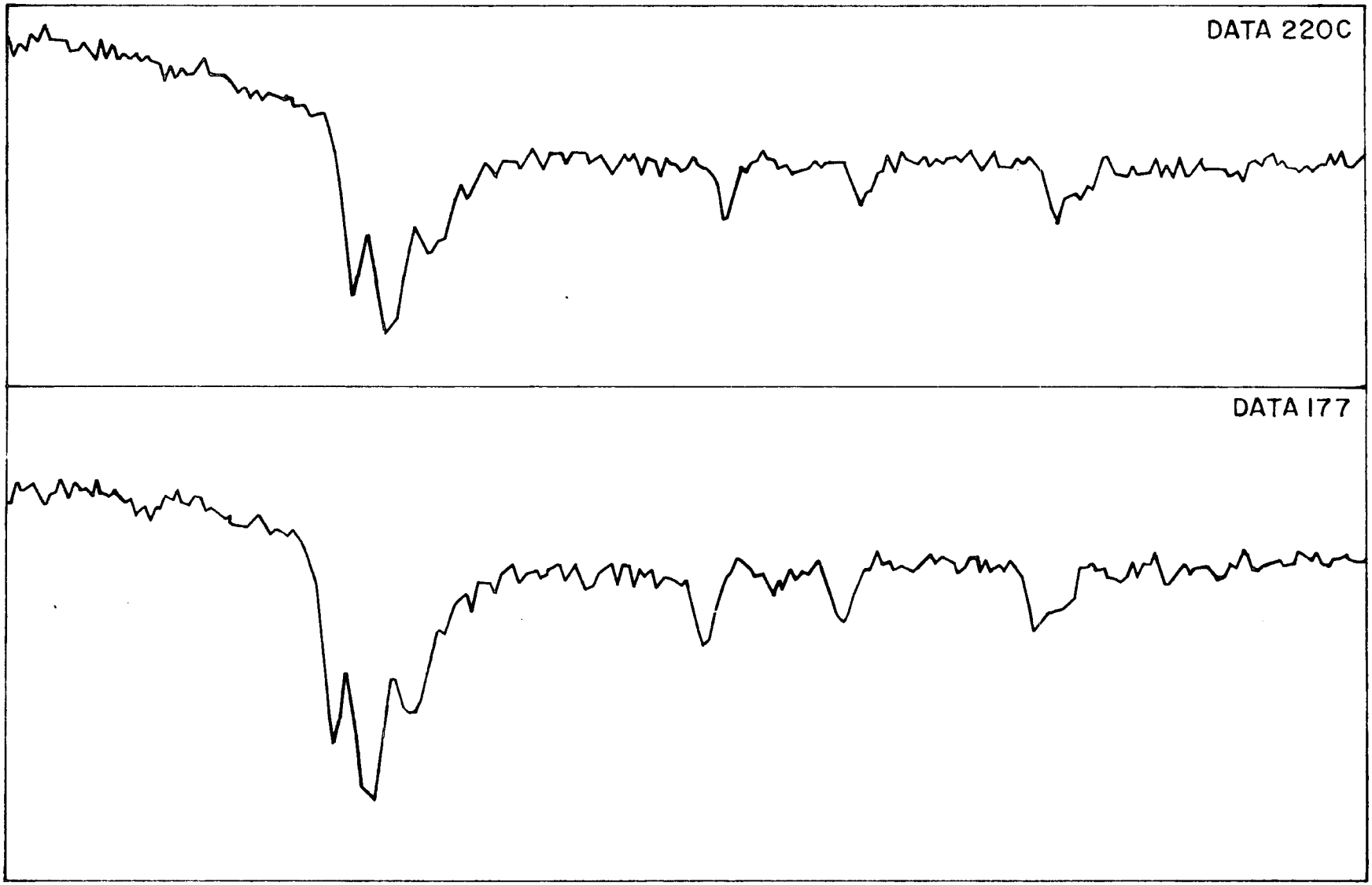
Experimental tests have shown that the existing detector offers potentially low noise performance but has restricted linearity and suffers from some unusual dynamic effects. Under actual observing conditions some useful results have been obtained, usually where the project has involved the analysis of absorption feature in a reasonably strong continuum. As an example of this type of operation Figure(7.1) shows two spectra of α Cygni taken three days apart, indicating that some degree of repeatability is possible. These spectra have been corrected for responsivity variations and odd-even effects by processing of the data on a mainframe computer.(1)

Many improvements to the data acquisition system are possible since the prototype system described was built mainly to evaluate the diode array performance, and the simple signal processing stages employed leave room for significant development. However, from the work done it appears that the main limitations of the detector are inherent in the diode array chip, making further improvements to the rest of the system of doubtful value at the moment. Thus achieving any major improvement to the detector performance as a whole, either requires a new batch of devices to be produced with some modifications to the design, or possibly for the array to be operated in an alternative manner to that originally intended.

2(a) Device Considerations

The most undesirable problem with this device is its apparently poor response to weaker incident signals. Weak objects which, given a long enough integration time, should have been seen, have produced no detectable signal. This failure has led to the belief that the array

Fig 7·1 TWO SPECTRA OF ∝CYG.



suffers from some degree of reciprocity failure. Experimental results have shown that, under certain conditions, the device response is 'S' shaped and that there is a zero error, usually negative. Both these factors help to explain the inability of the detector to see weaker signals but it is not certain whether or not this is the whole story. From theoretical arguments this behaviour is not entirely unexpected because of the inclusion in this array of an on-chip amplifier, although in practice the effects seem worse than indicated by simple calculations. In any case, it seems likely that the device's operation would be much more predictable and less affected by temperature drifts, without the feedback transistor, T1. Other than having more devices made without T1, it may be possible to increase the V_{REF} potential such that this transistor is in unsaturated operation and therefore acts as a simple resistance, thereby eliminating lag effects and baseline drifts due to changes in threshold voltage. All the advantages due to the amplified element design would, however, be lost, resulting in greater readout noise, increased dark signal and possibly poorer overall linearity.

At larger signals, linearity in this device is affected by the characteristic of the source follower transistor which must be attached to a virtual earth to achieve summing of the signals from each diode pair. If the video lines from odd and even elements were brought out separately, as is done with some recharge sampling arrays, the summing requirement would no longer exist and the video lines could be attached to current sources giving a more linear response. Alternatively, individual access to the odd and even shift register clock inputs would enable odd and even pairs to be interrogated at different times. Unfortunately, both these solutions require the production of modified chips.

At present no other voltage sampling arrays are available with a pixel size suited for use in astronomical spectroscopy.

2(b) Alternative Operating Procedures

To overcome the lag effects from which this device suffers the use of a shutter has been suggested (i.e. following an exposure the device is left in darkness for some time prior to read out to allow the signal to transfer through T1). In theory this should eliminate the problem but would slow down the rate at which observations could be made.

The apparent incomplete recharge effect could also be allowed for by storing the signals measured in several consecutive recharge frames after the first destructive read terminating the integration, (e.g. if it takes five recharge frames to recharge an element completely, the sum of the double sample signals measured in these five recharge frames should give the total signal integrated). This complete signal would be subject to $\sqrt{5}\,\sigma$ readout noise and therefore careful evaluation would be necessary to determine whether the improved signal was worth the poorer noise performance.

At the beginning of an integration the setting up procedure employed, of multiple continuous recharge frames, may also be varied. An alternative would be to start off with the whole array saturated with light and use only a fixed number of recharges to initialize the array.

This has been suggested to reduce the "slow start" effect, as has the possibility of using a controlled, pre-exposure light flash. However, it seems unlikely that any of the above will completely rectify the problems that have been observed.

2(c) System Improvements

The greatest improvements here can probably be made in the signal processing stage with some form of front end fixed pattern removal and possibly optimization of the system bandwidth from more detailed knowledge of device noise power spectrum, but, as has been noted, the more fundamental problems of linearity possibly make these improvements of purely academic interest.

REFERENCES

1. Hopkinson, G.R. and Humrich, A. "Observations of the OI λ 7773 triplet in intermediate-type Supergiants with a linear photodiode array." Monthly Notices, (to be published).

ACKNOWLEDGEMENTS

The author would like to thank Professor A.W. Wolfendale, F.R.S., and Professor B.H. Bransden for their support of this work and the use of laboratory facilities. He is indebted to his supervisor, Dr. J.M. Breare, both for his assistance throughout the work and for allowing him time to complete this thesis. Mr. A.R. Hedge, Dr. G.R. Hopkinson and Mr. A. Humrich are thanked for all the contributions they made while working on the same project.

Special thanks are due to Mr. J. Webster who was responsible, in particular, for the production of circuit boards and assembly of finished modules used in this work, provided invaluable technical assistance with most of the mechanical construction and generally supervised the packaging and transportation of equipment on field trips. Mr. R. McDermott and Mr. H. Davison are also thanked for cryostat manufacture.

The author's wife, Karen, is gratefully acknowledged for helpful criticisms of the script and assistance with the photographs. Mr. S. Jobson is thanked for much technical guidance and the loan of photographic equipment. Mrs. Joan McArthur is thanked for her excellent typing.

Finally, acknowledgement is made of the Science Research Council for providing a research studentship and the use of their various facilities throughout the work.