VOLUME 42 No. 4   APRIL 1972

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

---

A Common Heritage

THE oceans and seas of the world can truly be regarded as a common heritage which can prove either a source of food and wealth generally or, if neglected and abused, turn into sterile wastes. These perhaps oversimplified and let us hope overstated views of current concern with marine pollution came repeatedly to mind during the recent 'Oceanology International 72' Conference and Exhibition at Brighton. This was the second such event and, in comparison with 1969, impressed by its evidence of the realizations, three years later, of what were then in many cases feasibility studies or, at best, research prototypes. Thus, while many items shown and discussed were familiar to those who had been at the first exhibition, or were aware of the proceedings of the I.E.R.E. Conference on Electronics in Ocean Technology in 1970, it is apparent that in research and manufacture a very worthwhile area of scientific and engineering co-operation has been firmly established.

The usual definitions of the terms 'ocean engineering', 'ocean technology' and 'oceanology' overlap to a certain extent and the whole theme of activity within the oceans is essentially an interdisciplinary one. Exhibits ranged from underwater houses and diving chambers, through hydrographic survey equipment to sensors for physical, chemical and biological properties of the sea. This is not the time, nor is there adequate space, to describe in any detail either the exhibits or the subjects of papers in the Conference but the ubiquity of electronics was apparent. The exploration and exploitation of 'hydrospace' may not be dependent on electronic techniques to quite the extent that 'aerospace' is, but many important ventures would be extremely difficult to envisage if the information and control functions of electronic 'black boxes' did not exist.

Apart from calling for the co-operation of practitioners of many sciences and technologies, work in the oceans is the concern of many nations, acting either collaboratively or individually. That there is world-wide preoccupation with oceanology was to be seen in the range of exhibits and papers, and, of course, attendance. Five countries—Canada, France, Germany, Japan and the Netherlands—took national stands, grouping both government and industrial exhibits, and organizations from several other countries, notably the United States and Italy, were among the individual stands.

The British approach to the future was well expressed by Mr. David Price, Parliamentary Under Secretary to the Department of Trade and Industry, who when opening the Conference, said:

'Industry is faced with the difficulty of developing a technology where the time-scale between the identification of a market and the realization of profits is often large. In most of our countries the commercial exploitation of the resources of the sea and the seabed require a partnership between Government and industry. At this stage of development the Government should initiate and finance feasibility studies and other preliminary investigations in the closest co-operation with industry to help to identify where the opportunities may lie and what is necessary to exploit such opportunities.

'In the years immediately ahead we would do well to concentrate our attention largely upon the Continental Shelf and the problems of exploiting the potentials of the ocean at greater depths should wait. After all, the continental shelves of our oceans cover some ten million square miles, the equivalent of one-fifth of the land surface of our planet. We are gradually entering into a new phase in our endeavours, where normal terrestrial market considerations will play an increasing part. In saying this I do not detract one iota from the importance of continuing scientific endeavour, especially in the deeper oceans. But here the objective will remain for the time being at least the acquisition of more knowledge about the nature and working of basic forces in our planet without an early opportunity for commercial exploitation.'

While the oceans represent a common heritage, it is one which calls for care and forethought if it is to contribute to man's well being.

Contributors
to this issue

Professor E. A. Faulkner (Fellow 1966, Member 1964) graduated in physics from University College, Oxford. He worked at G.E.C. Research Laboratories, Wembley and subsequently at C.S.I.R.O. Division of Tribophysics, Melbourne, specializing in radio frequency measurements in solids. In 1960 he was appointed to the J. J. Thomson Physical Laboratory, University of Reading to take charge of teaching and research work in electronics. He was appointed to the Chair of Solid State Electronics in the same department in 1970. A strong advocate of university-industry collaboration, Professor Faulkner works as a consultant in linear circuit design for a number of leading electronics firms and is a director of Brookdeal Electronics Ltd. He has published about 60 papers in scientific and engineering journals and is the author of two books. He is a member of the Institution's Papers Committee and of the Thames Valley Section Committee.

Dr. M. J. Buckingham graduated in physics from the University of Reading in 1967. Since then he has been engaged in research into the properties and applications of semiconductor devices at the J. J. Thomson Physical Laboratory, University of Reading. More recently, at the same university, he has pursued an interest in the problems associated with the detection of gravitational radiation.

Mr. C. H. Jones was educated at Bournemouth Grammar School, after which he completed his National Service in the R.A.F. as an Air Wireless Mechanic. In 1955 he entered the Scientific Civil Service to work on night vision and digital communication. He graduated in 1964 in electrical engineering, receiving the external B.Sc. honours degree of London University.

His present position is Senior Scientific Officer at the Signals Research and Development Establishment where he is responsible for investigation into circuit design, fabrication and failure analysis of all forms of microcircuitry to be used in military equipment.

Dr. J. B. Knowles received his B.Sc. Tech., M.Sc. Tech., and Ph.D. degrees in electrical engineering from the University of Manchester Institute of Science and Technology in 1957, 1958 and 1962 respectively. Doctoral research work in collaboration with Ferranti Ltd., Wythenshawe, stimulated his interest in direct digital and adaptive control systems, and provided the practical basis for his subsequent papers on the design of these systems. As a lecturer at U.M.I.S.T. in control and communication systems from 1958 to 1968, he was an initiator of the post-graduate control school, and during this time also published contributions to medical electronics and spectral analysis. Currently, he is a Principal Research Fellow in the Control and Instrumentation Division at U.K.A.E.A. Winfrith, where he is engaged in problems on fluid dynamics, heat transfer and discrete data systems.

Mr. D. W. Leggett obtained his B.Sc. degree in electrical engineering in 1967 at Battersea College of Advanced Technology and received an M.Sc. in electronics from Southampton University in 1968. From 1960-1968 he was both a craft and student apprentice with U.K.A.E.A., Winfrith, and in 1968 he joined the Control and Instrumentation Division at Winfrith as an engineer. Since graduating, he has been engaged on the design of control and safety circuits for nuclear reactors and he is currently with a small team investigating interference problems and applying new techniques to reactor instrumentation.

Mr. A. Brian Keats (Member 1966, Graduate 1961) obtained his professional qualifications through part-time study whilst working at the Post Office Research Station, Dollis Hill. During that time (1950–1960) he was associated with the development of telegraph transmission systems and telegraph distortion measuring instruments. On joining the U.K.E.A. in 1960 his experience in logical design and digital electronic circuit techniques was directed towards the development of solid-state safety circuits for nuclear reactors. This activity extended to the application of digital techniques to other areas within the nuclear reactor control and instrumentation field including on-line computer applications.

Biographies of other contributors to this issue will be found on page 162.

Infra-Red Atmospheric Temperature Sounding from Satellites

P. J. ELLIS, Ph.D., C.Eng., M.I.E.R.E. *

G. PECKHAM, Ph.D. *

R. SANDWELL, B.Sc. *

Professor S. D. SMITH, D.Sc., F.Inst.P. *

J. T. HOUGHTON, D.Phil., F.R.S. †

C. D. RODGERS, Ph.D. †

and

E. J. WILLIAMSON, D.Phil. †

SUMMARY

On 8th April 1970 the Selective Chopper Radiometer (S.C.R.) was launched into a polar orbit on the *Nimbus 4* meteorological spacecraft. Since that date the instrument has produced nearly four million temperature profiles of the Earth's atmosphere, from the ground up to 50 km, in most cases to an accuracy of 1 or 2K. This has therefore made it possible for the first time to study the temperature parameters of the atmosphere daily on a global basis. Results from the S.C.R. and future experiments will therefore have a marked effect on meteorology, which in the past has had to rely on rather sparsely distributed radio and rocket sondes.

The principle of operation of the experiment is the measurement of infra-red energy in the 15 μm band originating from atmospheric carbon dioxide, but within the band, precise wavelength selection is obtained by combination of interference filters and selective chopping by carbon dioxide cells. This enables the six channels to be tuned to regions of different absorption coefficient in the 15 μm wavelength band, and hence to receive radiation from different heights within the atmosphere.

* *Heriot-Watt University, Riccarton, Currie, Midlothian*
† *Clarendon Laboratory, Oxford.*

1. Introduction

In order to obtain sufficient data for progressing numerical weather forecasting, it is desirable to achieve complete global monitoring of atmospheric parameters up to high levels in the atmosphere on a daily basis. The six parameters which are required are:

vertical distribution of temperature
surface pressure
density field
and the three components of wind velocity.

Of these the most easily measured is the vertical temperature profile, and fortunately this is probably also the most important since pressure is related to temperature via the hydrostatic equation

$$p(z) = p_0 \exp(-z/H)$$

where $p(z)$ is the pressure at height z

p_0 is the surface pressure

and H is the scale height, which is a function of temperature:

$$H = \frac{kT}{Mg}$$

where M is the mean molecular mass.

Hence knowledge of $T(z)$ and p_0 gives information also on $p(z)$.

In the past, data have been gathered by radio sonde balloons and a lesser number of rocket sondes, but the coverage given by these methods is limited and is mostly (75%) concentrated over the land masses of the Northern Hemisphere. This situation is rapidly being changed by the recent success of remote temperature-sounding instruments mounted on polar orbiting satellites which can give the required degree of global coverage.

2. Principle of Remote Temperature Sounding

Various researchers[1-3] in the early 1960s suggested instruments for the remote temperature sounding of the atmosphere by observation of the emission of infrared energy from the 15 μm absorption band of carbondioxide. The advantages of working in this spectral region are:

(i) carbon dioxide is uniformly mixed at all layers in the atmosphere up to ~ 75 km, to within 1 or 2%;

(ii) the temperature of the atmosphere is normally in the range 200 to 300 K, for which the Planck black-body emission curve peaks close to 15 μm;

(iii) there exist suitable optical materials and detectors for this wavelength.

The spectrum of carbon dioxide in 14–15 μm region is shown in Fig. 1 (top trace). It can be seen to consist of many separate lines, and this implies that the absorption coefficient is a rapidly varying function of wavelength. Consider now that the atmosphere is viewed from above by a multi-channel radiometer. By Kirchhoff's law, radiation is emitted from regions which absorb, and at the wavelengths which correspond to a

high CO_2 absorption coefficient (low transmission), the emission received at the satellite originates from levels high in the atmosphere. Any radiation at these wavelengths, which propagates upwards from lower regions in the atmosphere, will be re-absorbed at the higher levels. Conversely, radiation at wavelengths with a low CO_2 absorption coefficient is free to escape from the lower levels. Therefore an instrument which can selectively respond to spectral regions with different absorption coefficients, is able to identify emission originating from different heights within the atmosphere.
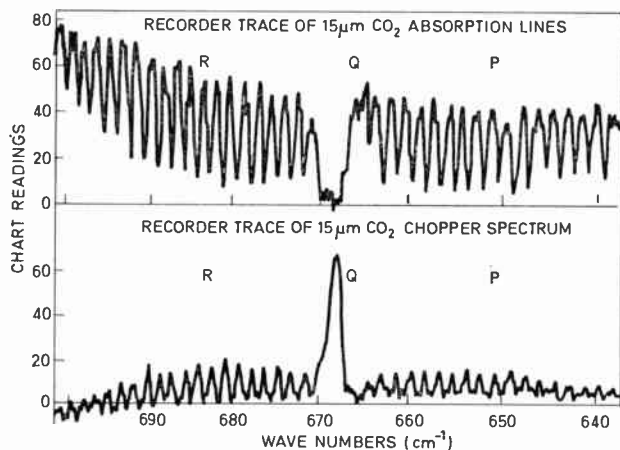

Fig. 1. Spectral form of CO_2 chopping.

The intensity of radiation reaching such a radiometer in a spectral range Δv will thus be a weighted mean of the black-body intensity, $B(z)$ as a function of height, z:

$$I = \frac{\int_0^\infty B(z)K(z)\mathrm{d}z}{\int_0^\infty K(z)\mathrm{d}z}$$

The term $K(z)$ is known as the *weighting function* and it can be shown that

$$K(z) = \int_{\Delta v} \frac{\mathrm{d}Tv(z)}{\mathrm{d}z}\mathrm{d}v$$

when we integrate over frequency and where $Tv(z)$ is the transmission between the height z and the radiometer.

The line widths of the molecular spectrum of CO_2 in the 15 μm region vary between 0.1 cm^{-1} and 0.001 cm^{-1} for the pressure ranges concerned. Therefore the absorption coefficient and the viewing depth in the atmosphere changes by several orders of magnitude within a narrow spectral range (over 0.5 cm^{-1} for example).

For an ideal instrument having a monochromatic bandwidth (less than 0.0001 cm^{-1}), the weighting functions would be approximately 10 km deep. Conventional instruments of the grating spectrometer or interferometer types require a bandwidth of around 3–5 cm^{-1} in order to obtain the required accuracy consistent with a reasonable physical size, and this degrades the vertical

resolution by a factor of approximately 2, and limits the maximum height of temperature sounding to 30 km. This latter fact is particularly significant when it is remembered that pressure is an exponential function of height, and that an increase in height of temperature sounding from 30 to 50 km corresponds to a pressure change from 10 to 1 mbar (i.e. 1000 mbar at the earth's surface).

The aim of the development programme carried out jointly between Oxford and Heriot-Watt Universities was to develop a simple and robust instrument capable of temperature sounding to a height of 50 km and to have sufficient signal to noise ratio to give an accuracy of 1 K or better.

This has now been achieved by means of *selective chopping*. Wavelength selection of the instrument is firstly defined by high resolution interference filters with bandwidths of 3 to 10 cm^{-1}, and these filters themselves were the subject of a considerable development programme extending over several years, since at the start of the project no suitable devices existed.

Within the bandwidth of the filters, however, further wavelength selection is performed by alternately switching the incoming radiation first through a cell containing carbon dioxide, and then through an evacuated cell. The CO_2 will absorb radiation at wavelengths where the absorption coefficient is high, and transmit at wavelengths where it is low. The vacuum cell will transmit equally at all wavelengths. Thus the resulting radiation will be a modulated beam whose amplitude will depend upon the intensity of radiation absorbed by the CO_2, that is, the intensity of radiation at wavelengths where the absorption coefficient is high, which as we have seen corresponds to radiation originating from high (above 30 km) levels in the atmosphere. The modulated signal is detected, amplified and integrated via a phase-sensitive detector.

With a cell length of 10^{-2} m and a pressure of 5×10^{-3} N m^{-2}, the weighting function of the instrument peaks at 45 km. Temperature sounding at slightly lower heights (35 km) can be achieved by admitting a small amount (2×10^{-3} N m^{-2}) of CO_2 into the 'vacuum' half of the cell. This has the effect of absorbing out radiation from the central region of each line, i.e. that coming from the very top of the atmosphere.

For sounding at low levels (below 30 km) a slightly different technique is appropriate, since in this case it is required to detect radiation from the wings of the CO_2 lines (i.e. regions of lower absorption) rather than the line centres. A simple on/off (black) chopper is used, and within the spectral response defined by the filter, radiation at the line centres is removed by a single absorbing cell of CO_2, through which all the radiation is passed. The viewing depth into the atmosphere of such an instrument can be adjusted by selection of the filter wavelength and CO_2 path length.

Using two selective chopping and four selective absorbing channels, six approximately equally spaced weighting functions can be defined to cover the entire atmosphere up to 45 km (Fig. 2).
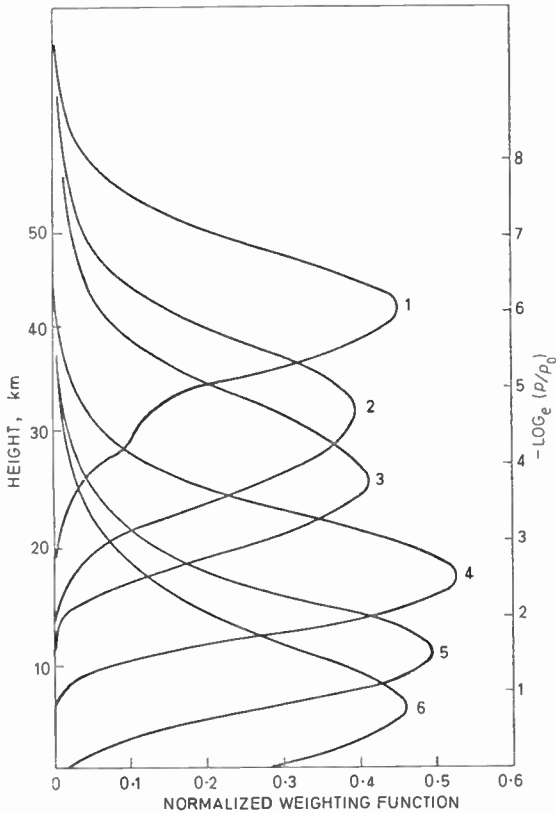
Fig. 2. S.C.R. weighting functions.

The three main advantages of these techniques are as follows:

(i) Since radiation at all wavelengths where the absorption coefficient is the same will be detected, the device has multiplex properties—i.e. it can accept useful radiation over many lines. The equivalent integrated bandwidth of the instrument is approximately 1.5 cm^{-1} and the resulting weighting functions show very little degradation from the ideal monochromatic response.

(ii) The interference filters which are used to define broadly (3 cm^{-1}) the spectral response around the required 15 μm region can accept radiation in a $20°$ cone and are approximately 1 inch in diameter. Hence the energy grasp of the system is large, and unlike most conventional optical systems is limited by the low resolution element (the filter) and not by the high resolution element (the gas cells).

(iii) All the optical components are mechanically stable and the alignment is reasonably uncritical. The engineering problems of the device are therefore simplified.

These three advantages taken together mean that a fairly simple, compact and robust instrument can be built which is capable of temperature sounding to greater heights, with improved resolution and better signal/noise ratio than equivalent grating or interferometer instruments.

3. Instrument Design

The Selective Chopper Radiometer has been described in several previous papers, notably those by Houghton and Smith[4] and Abel et al.[5]

The optical layout is shown in Fig. 3. Lenses within the optics define the field of view of the instrument to be a $10°$ cone, which gives a viewing area on the Earth of 100 miles in diameter.

Radiation from the Earth is reflected through $90°$ by a calibration mirror (not shown) into the CO_2/vacuum cells. In the case of the selective chopping channels, in front of the cells is a 10 Hz vibrating mirror mounted at $45°$ on resonant spring cantilever vanes. The mirror views deep space which for these purposes may be considered to have zero radiance. Therefore on one half-cycle the zero space signal is passed through the vacuum cell, and the Earth's radiation through the CO_2 filled side, a situation which reverses itself on the next half cycle.

In the case of the selective absorption channels a fixed space viewing mirror covers half the entrance aperture and a black blade chopper passes space and Earth radiation on alternative half-cycles. Behind the entrance aperture is a cell of CO_2 common to both halves of the beam.

It is convenient for the instrument to view space as a reference since the Earth temperatures being measured are less than or comparable with the temperature of the chopper. In the system described, the chopper is in the field of view of the instrument at all times, and since the electronics responds only to modulated signals, no radiation from the chopper is detected.

After passage through the CO_2 cells, the $10°$ beam is converged into a $20°$ beam and passed through the interference filter. Final convergence is performed by a conical light pipe at the end of which is mounted the 2.5 mm × 2.5 mm thermistor bolometer detector.

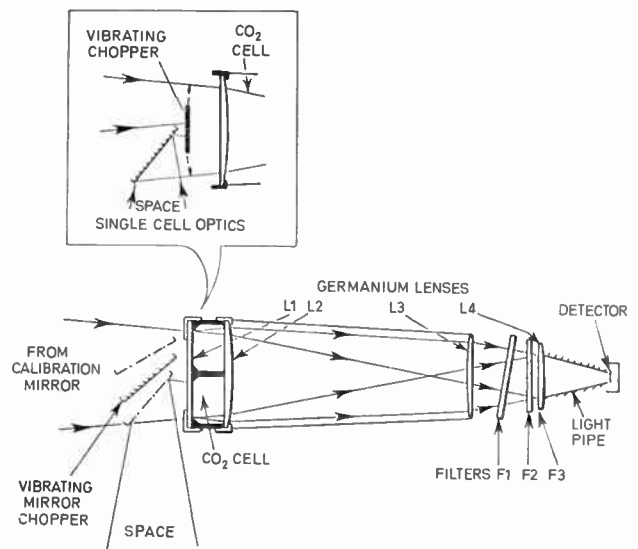In the case of the selective chopping channels, the interference filter is mounted on a rotatable wheel so that
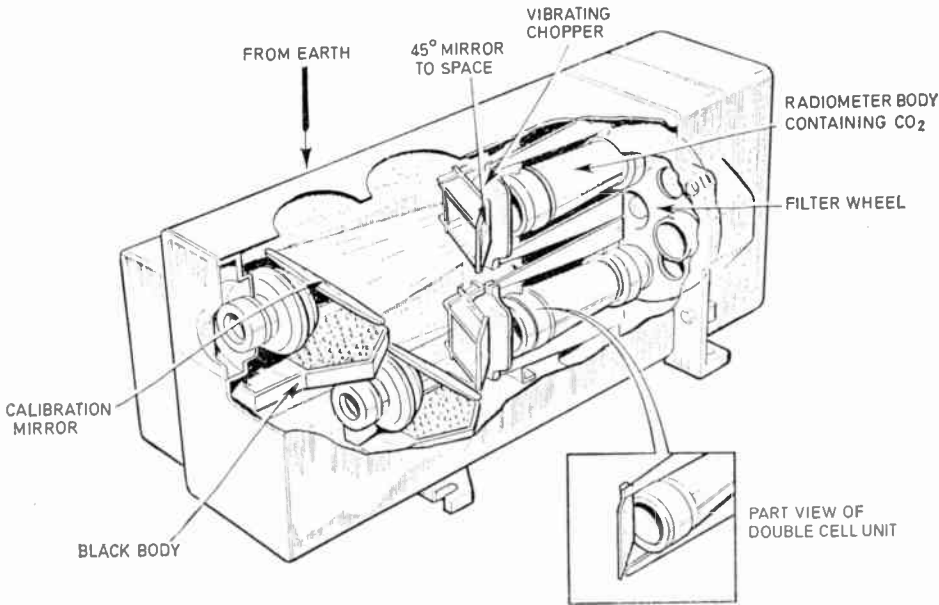

Fig. 3. S.C.R. optical system.

Fig. 4.
Sensor housing of the Selective Chopper Radiometer.

the 15 μm filter may be replaced by one tuned to 12 μm, where CO_2 is transparent. Using this filter it is possible to check the balance of the system, since the chopping effect due to the CO_2 absorption is removed.

A similar mechanical arrangement is used to exchange the filter in channel 3, which normally views a region around 25 km in the atmosphere, for one in an atmospheric window channel at 10·6 μm. At the same time the field of view of the channel is reduced to a slit 7 miles long and 70 miles wide, so that high-resolution spatial observations of ground or cloud top temperature may be obtained. Knowledge of these parameters is of importance for the interpretation of results in the lower atmosphere.

The calibration mirror, which normally views the Earth, may be rotated through 90° to view space in order to provide a zero calibration, and then again by a further 90° to view an internal black body of known temperature to provide a full scale calibration. This cycle normally occurs every 34 minutes in flight, the space and black-body views each being of 32 seconds duration.

The vibrating choppers are an interesting item in themselves. The chopping blade or mirror is mounted at the end of two cantilever beryllium copper springs, to which are bonded two piezo-electric crystals, one of which acts as a pick-off. The signal from this is amplified and fed back to the other (drive) crystal, so forming an electro-mechanical oscillator, vibrating with 1 cm amplitude at its natural frequency of 10 Hz with an input power of less than 1 mW. This should be compared with say 6 W for a conventional motor-driven rotating chopper, which also suffers from the disadvantage that it requires lubrication.

The microvolt level signals from the thermistor bolometer detectors are amplified by factors of up to 10^6, phase-sensitively rectified, integrated and passed to the spacecraft telemetry system for storage and subsequent transmission to ground tracking stations.

The experiment is packaged in three 'sensor housings' each of which contain two of the six optical channels. A common electronics module contains amplifiers, power supplies and the command and control system for operating the calibration mirrors and filter wheels. The layout of one of the sensor housings is shown in Fig. 4.

The spacecraft configuration is a hollow toroid structure known as the 'sensory ring' on which the Earth viewing experiments, including the S.C.R., are mounted. Above this ring is located the attitude control system (a.c.s.) which by means of flywheels and gas jets is capable of stabilizing the spacecraft to better than 1° in all three axes. Large solar paddles are driven by the a.c.s. so that they track the Sun and supply up to 450 W of power during satellite day.

4. Data Handling

Data from the S.C.R. and other experiments are stored in on-board tape recorders which are capable of holding one orbit's data, and of playing back this information in 4 minutes when the spacecraft passes over a tracking station, the main one being in Fairbanks, Alaska. From there, the data are transmitted via a wide-band microwave link to Goddard Space Flight Center, Washington, where the S.C.R. data are extracted and sent on a 4 kbits[-1] trans-Atlantic cable to Oxford and Edinburgh.

The two basic methods of presenting the data are:

(i) as temperature profiles, showing the vertical structure from the ground up to 50 km above one area on the Earth, or

(ii) as maps, showing isothermals over the entire globe for one particular height in the atmosphere.

Much of the data can be analysed and processed into these forms in real time during the two-hour period of reception each day (Fig. 5).
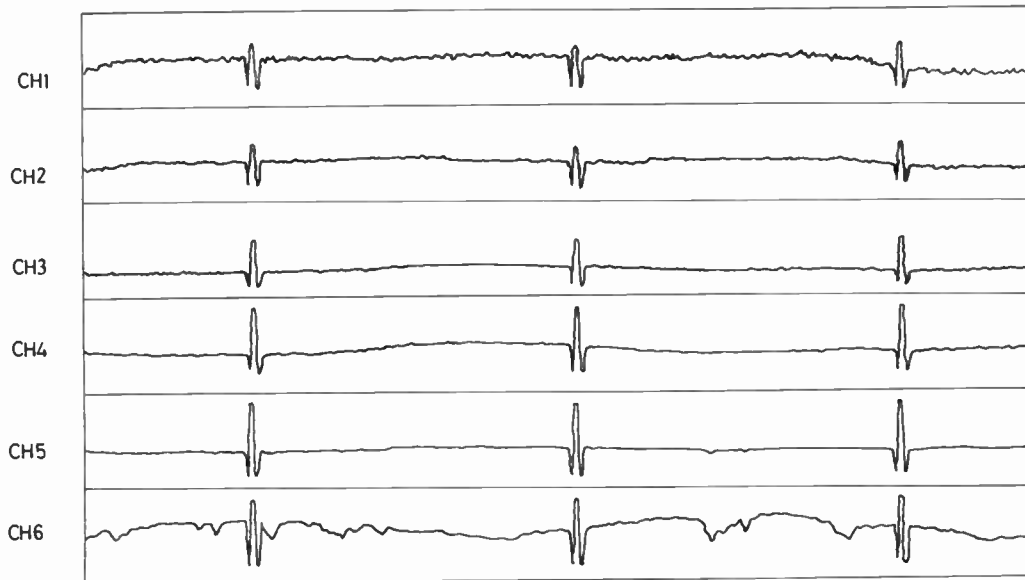
Fig. 5. Plot of typical data received (and averaged over 16 seconds) for one complete orbit.

The following points can be noted:

1. The orbit starts and finishes over Alaska.

2. Three calibration cycles occur during the orbit when the field of view of all channels is switched from Earth to space (having zero radiance) to the internal black-body (at approximately 300 K), to space and finally back to Earth.

3. The effect of clouds is pronounced on the lowest channel (CH6) but is virtually absent from all other channels, hence illustrating the vertical temperature resolution of the instrument.

4. The temperature of the regions of the atmosphere corresponding to channels 3 and 4 is less than two regions above and below this height (see Figs. 6 and 7).

5. Results

The *Nimbus 4* spacecraft was launched from N.A.S.A.'s Western Test Range in California on 8th April 1970, and the S.C.R. experiment was switched on during Orbit 6. Since that time the instrument has provided nearly four million complete and independent temperature profiles of the atmosphere.

The output from each channel is sampled every second, but for the higher channels it is necessary to integrate several samples to obtain the required accuracy. This is permissible, since the spatial rate of change of temperature at these levels in the atmosphere is small. The required integration time to obtain 1 K accuracy in each channel is shown in Table 1.

Atmospheric temperature profiles derived from the S.C.R. have been compared with results obtained by

rocket and radio sondes. A typical result from a series of tests performed over Wallops Island on 11th October 1970 is shown in Fig. 6, and over Berlin on 19th September 1970, in Fig. 7.

The experiment performed satisfactorily until March 1971, at which time the noise level of channels 5 and 6 increased by approximately an order of magnitude, which rendered the results of these channels useless. It is believed that this noise increase was caused by an intermittent fault within the bias supply to these two detectors. However, following a period during which S.C.R. was switched off due to a spacecraft malfunction, the noise level progressively decreased, and useful data were again obtained.

The operating conditions of the equipment itself are extensively monitored in orbit by means of 'housekeeping' telemetry. This can be used to provide correction factors where necessary for the radiometric data, and to evaluate the engineering performance of the experiment. All such telemetry channels, together with the sensory data, were calibrated before launch at the time of spacecraft thermal-vacuum tests, during which the spacecraft and all the experiments were cycled four times between 10°C and 35°C over a two-week period.

All supply voltages generated by d.c.-to-d.c. converters within the experiment are monitored to 8-bit (1 part in 256) resolution. It is necessary that in most cases these voltages remain constant to within a few percent.

Table 1

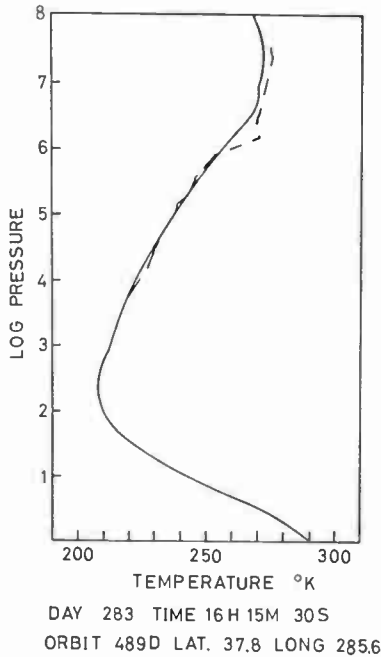| Channel	1	2	3	4	5	6
Viewing height (km)	45	35	25	18	10	5
Integration time to obtain 1K accuracy	60	25	7	3	1	1
No. of fields-of-view covered by integration time	2·3	1·1	0·5	0·2	0·07	0·07

Fig. 6. S.C.R. temperature profile (solid line) comparison with rocket sonde data (dashed line) over Wallops Island. 11th October 1970.

In fact, as shown in Table 2, in the course of nearly a year's operation the stability on all supplies is better than 0·4%.

Table 2. S.C.R. power supply voltages since launch (and comparison figure from thermal-vacuum tests).

	S.C.R. power supply voltages (nominal)				
Orbit No.	+9 V	−9 V	+3·5 V	+8·2 V	−8·2 V
214	+8·97	−8·91	+3·38	+8·22	−8·21
1025	+8·97	−8·91	+3·38	+8·22	−8·21
2258	+8·97	−8·91	+3·39	+8·21	−8·21
Thermal-vacuum tests	+8·96	−8·91	−3·40	+8·23	−8·20

Similarly, the stability of amplitude of oscillation of the vibration choppers is important, especially in the case of the selective chopping channels, Nos. 1 and 2. It can be seen from Table 3 that the amplitudes of the choppers in channels 1, 2, 3 and 4 have remained constant to a figure comparable with the telemetry resolution (0·4%), but that of channel 5 has increased by 4% and that of channel 6 by 1·7%. This can probably be explained by an increase of 2 deg C since launch in the average temperature of the housing containing these two channels. It was noted during thermal-vacuum tests that the amplitude of channel 5 chopper was abnormally sensitive to temperature variations. However, the actual increase in chopper amplitude has a negligible effect on the radiometric performance.

Table 3. S.C.R. chopper amplitudes since launch (and comparison figures from thermal-vacuum tests).

	S.C.R. chopper amplitudes (volts)					
Orbit No.	channel 1	channel 2	channel 3	channel 4	channel 5	channel 6
214	2·52	2·47	1·69	1·90	2·92	2·28
1025	2·53	2·48	1·69	1·91	2·96	2·30
2258	2·53	2·48	1·70	1·90	3·01	2·31
Thermal-vacuum tests	2·51	2·45	1·68	1·90	3·08	2·26

The thermal performance of the instrument has been mostly satisfactory. Due to the sophisticated active temperature control of the spacecraft, day/night variations in the temperature of the S.C.R. components have been generally less than 2 deg C, and the long-term drift since launch is in most cases less than 1 deg C.

6. Future Developments

Two further spacecraft in the *Nimbus* series, *Nimbus E* and *F* are to be built. *Nimbus E*, scheduled for launch in late 1972, will contain an enlarged and improved Selective Chopper Radiometer built by the same British team. This experiment will have 16 channels compared with 6 in the case of the *Nimbus 4* unit, and will employ optical beam splitting and filter switching in order that the optimum packing density can be achieved.[6] The field of view has been reduced from 10° to 1·5°, so that the
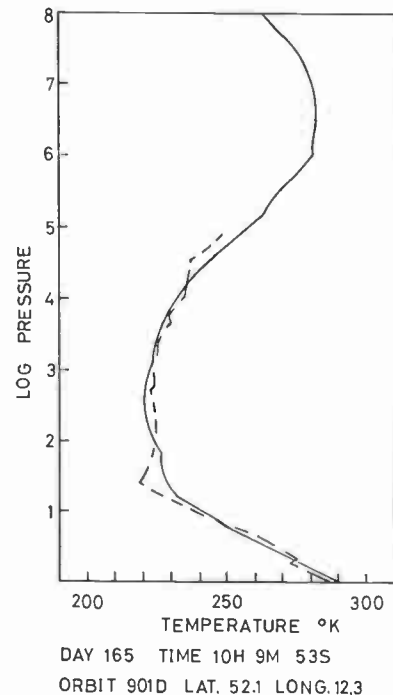

Fig. 7. S.C.R. temperature profile (solid line) comparison with radio sonde data (dashed line) over Berlin. 19th September 1970.

instrument can view small holes in cloud cover and therefore extend the coverage of accurate temperature sounding. The additional channels will observe water vapour content of the atmosphere, ice particles in cirrus cloud, reflected sunlight and near-infra-red emission from clouds.

The *Nimbus E* spacecraft will also have on board an American experiment, the Infra-red Temperature Profile Radiometer (I.T.P.R.), which employs interference filters similar to those used in the S.C.R. Another filter instrument, the Vertical Temperature Profile Radiometer (V.T.P.R.) is being developed for a *Tiros* satellite.

The *Nimbus F* spacecraft will fly the Pressure Modulated Radiometer (P.M.R.) developed by the S.C.R. team at Oxford University, which will perform temperature sounding at extreme heights to 70 km in the atmosphere by replacing the technique of beam switching through CO_2 cells, as used in the *Nimbus 4* radiometer, with a cell in which the pressure of CO_2 is modulated by means of a piston.

Finally, in conjunction with the Meteorological Office, Bracknell, it is hoped that a development of the Selective Chopper Radiometer will be built for the *Tiros N* operational meteorological spacecraft.

7. Acknowledgments

8. References

1. Kaplan, L. D., 'Interference of atmospheric structure from remote radiation measurements', *J. Opt. Soc. Amer.*, 49, pp. 1004–7, October 1959.

2. Wark, D. Q., 'On direct temperature soundings of the stratosphere from satellites', *J. Geophys. Res.*, 66, pp. 77–82, January 1961.

3. Houghton, J. T., 'Meteorological significance of remote control measurement of infra-red emission from atmospheric carbon dioxide', *Quart. J. R. Met. Soc.*, 87, pp. 102–104, January 1961.

4. Houghton, J. T. and Smith, S. D., 'Remote sounding of atmospheric temperature from satellites: Part 1', *Proc. Roy. Soc. A*, 320, pp. 23–33, 24th November 1970.

5. Abel, P. G., Ellis, P. J., Houghton, J. T., Peckham, G., Rodgers, C. D., Smith, S. D. and Williamson, E. J., 'Remote sounding of atmospheric temperature from satellite: selective chopper radiometer for *Nimbus D*: Part 2', *Proc. Roy. Soc. A*, 320, pp. 35–55, 24th November 1970.

6. Ellis, P. J., 'Analysis and control of the permanent-magnet stepper motor', *The Radio and Electronic Engineer*, 41, No. 7, pp. 302–308, July 1971.

Contributors
to this issue

Professor S. D. Smith graduated from the University of Bristol in 1952. He then did research at the University of Reading, receiving his Ph.D. for a thesis on 'infra-red multilayers' in 1956. He was awarded the D.Sc. degree from Bristol University in 1966 for his work on the application of infrared optics to the fundamental study of solids and the development of infra-red techniques. From 1965 to 1958 he was a Senior Scientific Officer at the Royal Aircraft Establishment, Farnborough, and in 1958 he was appointed a Research Physicist in the Meteorology Department at Imperial College. He moved to the University of Reading as a Research Physicist in 1959 and four years later obtained a Lectureship. He was appointed a Reader in 1965.

During 1968 Dr. Smith was Royal Society European Visiting Fellow at the University of Freiburg, West Germany. In 1970 he was appointed to a Chair in Physics at Heriot-Watt University, Edinburgh. A member of the Meteorology Working Group of the British National Committee for Space Research from its formation in 1960 until dissolution in 1965, Professor Smith has been a member of an S.R.C. Working Group on Terrestrial and Planetary Atmospheres since 1968.

Dr. P. J. Ellis (Member 1967, Graduate 1962) received his technical education at Redhill Technical College and Brighton College of Technology, while working at Mullard Research Laboratories as an electronics apprentice and later as an engineer. In 1962 he joined the Space and Weapons Research Laboratory of Elliott Automation. He then received an appointment as senior research assistant at the J. J. Thomson Physical Laboratory, University of Reading, joining a group engaged in the design of instruments for the *Nimbus* weather satellites. The group has since moved to Heriot-Watt University, Edinburgh, He was awarded the Ph.D. degree of Reading University in June 1971.

Dr. G. E. Peckham obtained an Open Scholarship to Trinity College, Cambridge, in 1954 and after reading mathematics and physics, received his B.A. degree in 1960. From 1960 to 1964 he carried out research at the Cavendish Laboratory on phonon dispersion relations in crystals and subsequently received his doctorate. Dr. Peckham was appointed to a research post at the University of Reading in 1964 and in 1966 he became a Lecturer in J. J. Thomson Laboratory at Reading. In 1971 he moved to the Physics Department at Heriot-Watt University, Edinburgh, as a Lecturer.

Dr. J. T. Houghton who is now a Fellow of Jesus College, Oxford, first joined the College in 1948 and after obtaining his B.A. went on to read for his D.Phil, which he was awarded in 1955. After three years at the Royal Aircraft Establishment, Farnborough, he returned to Oxford as Lecturer in Atmospheric Physics. In 1962 he was appointed to a Readership in Atmospheric Physics and he is at present Head of Department. From October to December 1969 Dr. Houghton was a Visiting Professor at the University of California. In March of this year he was elected a Fellow of the Royal Society for his studies of radiative processes in the atmosphere and the use of infra-red spectrometry from satellites in these studies.

Mr. R. Sandwell joined Elliott Automation Ltd. in 1960. He was awarded the B.Sc. in applied physics in 1966 after pursuing a sandwich course at City University, London. From 1966 to 1968 he was Project Engineer at Elliotts for the *Nimbus D* Selective Chopper Radiometer experiment. From 1968 to 1970 he was seconded by Elliotts to the General Electric Company at Valley Forge, Pennsylvania as Subsystem Engineer responsible for the investigation of the incorporation of the S.C.R. on to the *Nimbus D* spacecraft. From 1970 to the present time he has been working at Heriot-Watt University on the *Nimbus E* Selective Chopper Radiometer and future satellite experiments.

Dr. C. D. Rogers graduated from the University of Cambridge, where he was at Caius College, in 1960. In 1963 he was awarded his Ph.D. and was appointed a Research Assistant in the Department of Atmospheric Physics at Oxford. From 1966 to 1967 he was a Visiting Scientist in the Department of Meteorology, Massachusetts Institute of Technology. Since 1967 he has been a University Lecturer in the Department of Atmospheric Physics at Oxford.

Dr. E. J. Williamson was at University College, Oxford, from 1956 to 1959. After obtaining his B.A. he continued at the University and obtained the D.Phil. in 1964. For three years he was research Assistant in the Department of Atmospheric Physics at Oxford and in 1967 to 1968 went to the Goddard Space Flight Center as a Research Associate. Since 1968 he has been Senior Research Officer in the Department of Atmospheric Physics at Oxford.

The Principles of Pulse Signal Recovery from Gravitational Antennas

M. J. BUCKINGHAM, Ph.D. *

and

Professor E. A. FAULKNER,
Ph.D., C.Eng., F.I.E.E., F.I.E.R.E. *

SUMMARY

The general engineering problem of pulse signal recovery from a reactive (i.e. energy-storing) transducer is considered. It is shown that by using active filters (i.e. filters which are not subject to the constraints imposed upon passive systems by the second law of thermodynamics), pulses of energy less than $k\theta/2$ (where θ is the absolute temperature) are detectable.

The arguments are then extended to apply to the situation involving a resonant detector—this is relevant to the gravitational wave detectors which are currently being engineered. In the resonant case the appropriate system function is centred on the resonance frequency $\omega_0/2\pi$ rather than on zero frequency.

Expressions for the signal/noise ratio and the minimum detectable energy are obtained for a detector in terms of the Q values of the bar and the transducer, and β, a parameter representing the degree of electro-mechanical coupling. Under optimum operating conditions the split bar is found to be more sensitive than the single bar by a factor of about 10; if, however, a time resolution of a few milliseconds is required for coincidence measurements the single bar shows a penalty of nearly 3000 compared with the split bar.

* *J. J. Thomson Physical Laboratory, University of Reading, Whiteknights, Reading RG6 2AL.*

1. Introduction

The signal/noise problem in gravitational antenna systems has received considerable discussion from physicists, both in the literature and elsewhere[1-5] but in the opinion of the present authors there has been insufficient reference to the state of knowledge in the field of electronic engineering. In this paper we shall relate the problem of the gravitational antenna to the well-known engineering problem of pulse signal recovery from a non-resonant reactive (i.e. energy-storing) transducer.

2. The Signal/Noise Problem in Reactive Transducers

2.1. *Minimum Detectable Pulse Energy*

The circuit of Fig. 1 is one way of representing a non-resonant reactive transducer with time-constant $T_1 = R_1 C_1$, which we shall assume to be initially (i.e. before the arrival of the signal energy) in thermal equilibrium with its surroundings at 'room temperature' θ. We assume that the incident signal is in the form of a pulse occupying a time much shorter than the time-constant T_1.
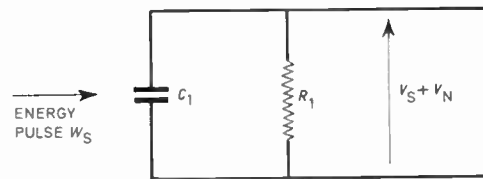

Fig. 1. A reactive transducer.

The energy W_S usefully absorbed by the transducer is converted to a signal voltage V_S whose peak value V_{SP} is given by the relation

$$W_S = C_1 V_{SP}^2/2 \qquad \ldots.(1)$$

and the noise voltage V_N has a mean-square value determined by the equipartition requirement that the mean noise energy stored in the capacitor should be $k\theta/2$:

$$W_N = C_1 \overline{V}_N^2/2 = k\theta/2. \qquad \ldots.(2)$$

The signal/noise ratio is therefore given by the ratio

$$V_{SP}^2/\overline{V}_N^2 = 2W_S/k\theta. \qquad \ldots.(3)$$

It is tempting to deduce from these considerations that the minimum detectable energy is invariably fixed at $k\theta/2$. However, this conclusion is fallacious because it does not take into account the process by which the energy W_S arrived in the capacitor C_1.

In the situation we are considering where the limitations of thermal equilibrium do not necessarily apply, it is necessary to consider the fluctuations in terms of *random forces*† (or couples, voltages, currents etc.) which compete with the corresponding force exerted by the signal. Although the statistical properties of these random forces are derived from equilibrium conditions, they remain valid in non-equilibrium situations.

Accordingly we re-draw Fig. 1 in the form shown in Fig. 2(a), where the coupling of the signal into the

---

† A review of the random force method is given by McCombie.[6]

capacitor C_1 is represented by a signal current generator $x(t)$, and the voltage fluctuations in the capacitor are assumed to be due to a random current generator $i_N(t)$. As is well known, the equilibrium mean energy $k\theta/2$ in the capacitor is obtained by assuming that the mean-square value of i_N in any angular frequency interval $\delta\omega$ is given by the relation

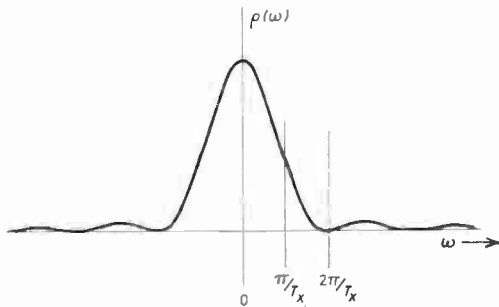$$\delta\bar{i}_N^2 = 2k\theta\delta\omega/\pi R_1. \qquad \ldots.(4)$$

Now we shall suppose for the sake of simplicity that $x(t)$ is a 'square' pulse of height x_P and width T_x, centred at $t = 0$, with $T_x \ll T_1$. This function and its power spectrum are shown in Figs. 2(b) and (c).
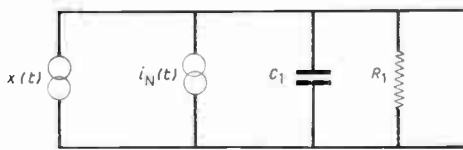
(a) Equivalent circuit of transducer showing signal and noise current generators.
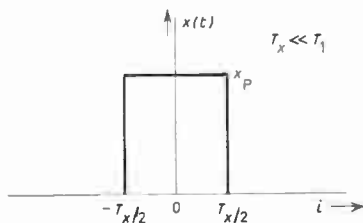
(b) The single pulse input

(c) The power spectrum of the pulse of (b).
Fig. 2.

The resulting peak voltage across the capacitor is

$$V_{SP} = x_P T_x/C_1 \qquad \ldots.(5)$$

and the corresponding signal energy

$$W_S = x_P^2 T_x^2/2C_1. \qquad \ldots.(6)$$

The important fact that emerges from the random-generator treatment is that the inherent signal/noise ratio is not the ratio of W_S to $k\theta/2$ but the ratio of x_P^2 to the mean-square value of the noise current generator i_N in the frequency range occupied by the signal. By inspecting Fig. 2(c) we find that a reasonable estimate of the signal bandwidth $\Delta\omega_x$ is π/T_x and the resulting
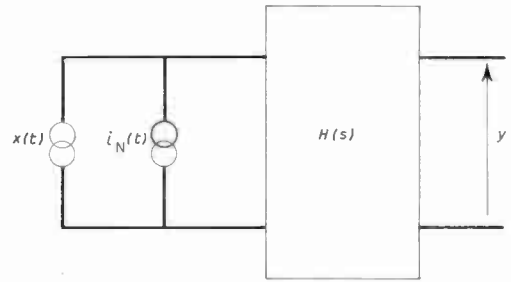

Fig. 3. The system diagram.

expression for the inherent signal/noise ratio is, (making use of equations (4) and (6)),

$$S/N = x_P^2/(2k\theta\Delta\omega_x/\pi R_1)$$
$$= \pi W_S T_1/k\theta T_x^2 \Delta\omega_x$$
$$= W_S T_1/k\theta T_x. \qquad \ldots.(7)$$

From this calculation it emerges that the 'minimum detectable energy' is smaller by a factor of approximately $2T_x/T_1$ than the value of $k\theta/2$ predicted by the first argument.

2.2. The Need to Increase the Bandwidth

In order to establish the methods of realizing in practice the inherent signal/noise ratio, we adopt a 'systems' approach and draw the equivalent circuit shown in Fig. 3.

Here the signal current generator and noise current generator are represented as being connected to a 'black box' which is a noiseless (but not in general passive) linear system with output voltage $y(t)$ and transfer function $H(s)$ having the dimensions of impedance.* With this formulation the circuit of Fig. 2(a) has a transfer function

$$H(s) = R_1/(1+sT_1). \qquad \ldots.(8)$$

This function immediately shows us why the original system gives much less than the optimum signal/noise ratio with the stated condition $T_x \ll R_1 C_1$: the cut-off frequency of the system is so low that the signal is heavily attenuated in relation to the low-frequency components of the noise. Such attenuation always acts to the detriment of the signal/noise ratio because the peak-square value of the signal is reduced at least as the square of the bandwidth while the noise power is reduced approximately in proportion to the bandwidth.

We see now that in order to obtain the optimum signal/noise ratio we must increase the bandwidth of the system $H(s)$. In order to simplify the argument we shall for the present assume that it is to remain a first-order function, although as we shall see later it is usually made more complicated in practice. We write for the new function

$$H(s) = AR_1/(1+sT_2) \qquad \ldots.(9)$$

where A is some appropriately chosen constant, and we

* Physicists who are unfamiliar with the engineering conventions in Laplace-transform terminology are referred to 'Introduction to the Theory of Linear Systems', E. A. Faulkner (Chapman & Hall, London, 1969).

must now discuss the required value for T_2. We have already seen that the signal bandwidth $\Delta\omega_x$ approximates to π/T_x and the optimum value for T_2 will therefore be not less than T_x/π. In fact, it is not generally necessary to accept such a large fraction of the signal spectrum if we are simply interested in detecting the existence of the pulse rather than timing it as accurately as possible. This is illustrated by the fact that the peak response y_P of the system described by equation (9) is given by the relation

$$y_P = Ax_PR_1[1-\exp(-T_x/T_2)] \qquad ....(10)$$

so that the system will give little attenuation of the peak height of the pulse ($y_P \geqslant 0.7Ax_PR_1$) provided that $T_2 \leqslant T_x$. Assuming that the bandwidth can be increased without the introduction of further noise generators, the exact expression for the signal/noise ratio will be the ratio of y_P^2 to the mean-square value of the noise component y_N in the output. Using (4), (9), and (10) we find

$$S/N = \frac{y_P^2}{\overline{y_N^2}} = \frac{\pi R_1 x_P^2[1-\exp(-T_x/T_2)]^2}{2k\theta\int_0^\infty d\omega/(1+\omega^2 T_2^2)}$$

$$= \frac{2W_ST_1}{k\theta T_x}[1-\exp(-T_x/T_2)]^2\frac{T_2}{T_x} \qquad ....(11)$$

When plotted against the ratio T_x/T_2, the function in brackets has a very broad maximum of 0.4 in the region of $T_2 = T_x$, giving the result

$$S/N = 0.8W_ST_1/k\theta T_x. \qquad ....(12)$$

This agrees well with the result obtained in equation (7) by a simplified argument which assumed that the noise spectrum could be regarded as being sharply cut off; we see that for a first-order system the optimum bandwidth is in fact given by $T_2 = T_x$, although almost the same signal/noise ratio is obtained by substituting into equation (11) the condition $T_2 = T_x/\pi$ which was used in deriving equation (7).

2.3. Methods of Increasing the Bandwidth

We must now consider the methods by which the bandwidth of the system can be increased from $1/T_1$ to approximately $1/T_x$ in order to obtain the possible improvement of a factor $2T_x/T_1$ in the minimum detectable energy.

If we simply connect a resistor R_2 across the capacitor C_1 we shall certainly succeed in increasing the bandwidth but at the same time (if R_2 is at room temperature) we will increase the mean-square value of the noise current generator i_N by a factor $(1+R_1/R_2)$ so that the signal/noise ratio is not improved. To look at the matter from another point of view, the mean energy stored in C_1 is still $k\theta/2$ because the system is still in thermal equilibrium at room temperature. It follows that we cannot hope to perform the required function by means of components which are in thermal equilibrium with their surroundings.

At this point in the argument the physicist starts to think in terms of liquid helium—the electronic engineer, somewhat more economically, in terms of active circuits. An active system (one incorporating power supplies and transistors) is not subject to the constraints on (a) power

gain, (b) noise generator values, and (c) reciprocity, imposed by the second law of thermodynamics and the principle of microscopic reversibility on systems which have no energy sources.[7] One consequence of (b) and (c) is that a resistor at a low temperature can be simulated by the circuit shown in Fig. 4, where the amplifier is assumed to have voltage gain A_1 and negligibly small input admittance.

The system input impedance is $R_F/(1-A_1)$ and this approximates to a resistance in the frequency range where A_1 is substantially real. However, the noise generators are those associated with the feedback resistor and with the amplifier itself; these have no necessary relationship with the system input resistance and in practice one can simulate a resistor of very low temperature by this technique, at least in the audio-frequency range (say 10 Hz – 1 MHz).
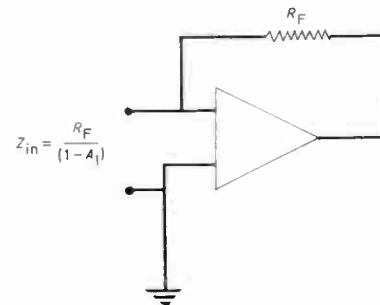

Fig. 4. Simulation of a resistor at low temperature.

Now although the circuit of Fig. 4, suitably designed, can be shunted across the circuit of Fig. 2(a) to improve the pulse signal/noise ratio towards its theoretically 'best possible' value given by equation (7), we would not normally adopt this method in practice. There is an alternative way of making use of the non-equilibrium properties of active circuits, and that is by means of *active filters*.

The term 'active filter' has become associated over the past few years with a multitude of contributions to the literature describing various methods of implementing resonant second-order transfer functions without the use of inductors. In the present context, however, we are using the term to include any filter which, by reason of its active nature, is not subject to the thermodynamic constraints listed above. With suitable design an active filter may be used to provide almost any realizable rational system function without introducing a significant amount of noise.

In applying the 'active filter' concept, we redraw Fig. 3 in the form of Fig. 5.

Here the original system function $H_1(s)$, determined primarily by the parameters of the transducer, is assumed to be unchanged. The output voltage of the original system is connected to the input terminals of the filter which has negligible input admittance and voltage transfer function $H_2(s)$. The overall transfer function $H(s)$ is then given by the product
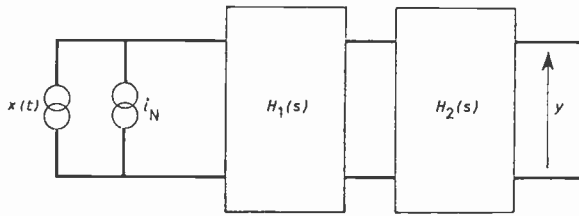
$$H(s) = H_1(s)H_2(s). \qquad ....(13)$$

Fig. 5. Transducer with active filter.

If we assume that the required transfer function is that given in (9), and that the transducer function is that of (8), we obtain for the filter transfer function

$$H_2(s) = A(1+sT_1)/(1+sT_2). \qquad \ldots(14)$$

This function is not in fact realizable, having infinite bandwidth, and in practice a function of higher order must be used. A typical filter function is

$$H_2(s) = A(1+sT_1)/(1+sT_2)^2 \qquad \ldots(15)$$

giving an overall transfer function

$$H(s) = A/(1+sT_2)^2. \qquad \ldots(16)$$

The signal/noise ratio obtainable with such functions can be calculated by setting up an expression analogous to (11) and optimizing it as in (12). As one would expect from power-spectrum considerations, the optimum signal/noise ratios calculated for these higher-order systems differ little from the values obtained from the simple first-order models.

The 'active filter' method has been known for many years to the designers of nuclear pulse amplifiers[8], although it has not usually been discussed in terms of an overall transfer function. A system described by equation (14), usually approximated by

$$AsT_1/(1+sT_2),$$

is known as a *differentiating circuit*, and the time-constant T_2 as the *differentiation time*.

2.4. Double Pulses

In certain systems the signal to be detected is in the form of a double pulse of zero total area. We may approximate this by the function shown in Fig. 6(a), which is a pulse like that of Fig. 1 immediately followed by a similar pulse in the opposite direction. The power spectrum of this function is given in Fig. 6(b).

This pulse can, of course, be handled by the systems previously discussed, and in this case gives approximately the same signal/noise ratio as a single pulse. However, the power spectrum, Fig. 6(b), has a peak at $\omega \simeq \pi/T_x$ and we may therefore expect to optimize the signal/noise ratio by choosing a bandpass function of the form

$$H(s) = A.2as/[s^2+2as+a^2+n^2]. \qquad \ldots(17)$$

In the familiar 'high-Q' formulation this appears as

$$H(s) = A.\omega_0 s/Q \left[s^2 + \frac{s\omega_0}{Q} + \omega_0^2 \right] \qquad \ldots(18)$$

where $\omega_0 \simeq n$, the frequency of free oscillation, and the Q-factor is defined by the relation

$$Q = n/2a. \qquad \ldots(19)$$

We must now consider the question of the optimum Q factor. When we examine the power spectrum of the double pulse, which is

$$\rho(\omega) = \frac{4x_P^2}{\omega^2}(\cos \omega T_x - 1)^2 \qquad \ldots(20)$$

and compare it with the corresponding spectrum for a single step function

$$\rho(\omega) = x_P^2/\omega^2 \qquad \ldots(21)$$

we find that the spectral density of the double pulse at its peak angular frequency is greater than that of the single step by a factor of no less than 16. This corresponds to the fact that the double pulse function can be regarded as being made up of three step functions, of value x_P, $-2x_P$, and x_P respectively. If these are perfectly timed with respect to the oscillatory step response of a highly resonant system, they will reinforce one another and give a resultant amplitude corresponding to a single step input of value $4x_P$.
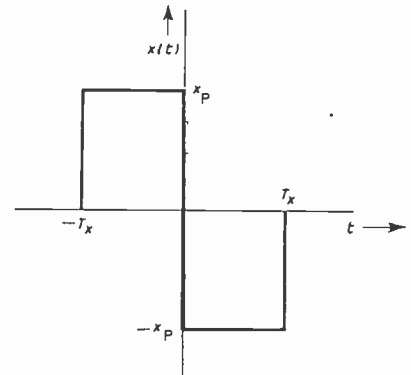
(a) A double pulse input $x(t)$
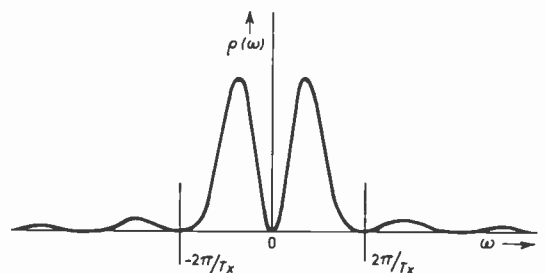

(b) The power spectrum of (a).

Fig. 6.

However, this fact does not imply that a high-Q response will optimize the signal/noise ratio. The step response of the function (17) is the inverse transform of $2aA/[s^2+2as+a^2+n^2]$ which is

$$a(t) = \frac{2Aa}{n}\exp(-at)\sin nt \qquad \ldots(22)$$

and this reduces in the high-Q approximation to

$$a(t) = \frac{A}{Q}\exp(-nt/2Q)\sin nt. \qquad \ldots(23)$$

For $Q \geqslant 1$ this step response is an oscillatory one lasting over many cycles, and the amplitude of the
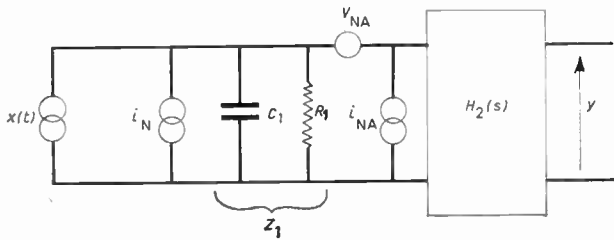
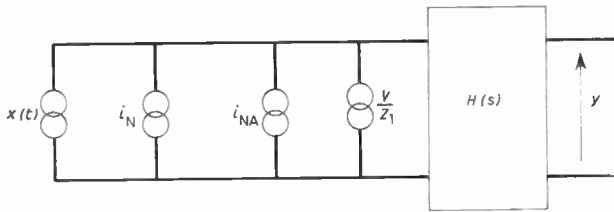(a) System diagram showing noise generators of active part of the system.

Fig. 7. (b) System diagram with all generators referred to the input.

response to a double pulse exactly at the resonance frequency is four times the step response. The effect of the narrow bandwidth on the noise will be to alter the spectrum so that it is sharply peaked around the resonance frequency; the filtered noise will take the form of slowly increasing and decaying sinusoidal oscillations. The fact that the signal/noise ratio has been reduced is evident by the following approximate argument:

(a) No additional signal information is gained by the existence of the decaying 'tail' of the response—a measurement of the signal envelope is of no more value than the first peak, no averaging effect being available because the noise components decay no faster than the signal.

(b) The amplitude of the signal response is proportional to $1/Q$ according to equation (23); on the other hand, the bandwidth (and therefore the mean-square value of the noise) is also proportional to $1/Q$; therefore the mean-square signal/noise ratio is reduced by the same factor $1/Q$.

In fact the double pulse spectrum (20) has the general appearance of a bandpass function with $Q \simeq 1$, and it is in this region that the optimum Q factor is found. As in the case of (14), the required active filter function is not realizable and a function of higher order than (17) will in practice be required.

2.5. Amplifier Noise

So far we have been considering only the effect of thermal noise in the transducer, which is the fundamental and unavoidable limitation. In a practical system there will also be non-equilibrium noise generated by the active elements in the system, and the effect of these can be represented by random voltage and current generators at the terminals of the active system. In Fig. 7(a) these generators have been added to the system of Fig. 5 with the components of the passive part of the system being explicitly shown. In Fig. 7(b) the additional noise

generators have been moved to the input side of the circuit and all noise generators are shown directly in parallel with the signal current generator; the voltage generator V_{NA} has been represented as a current generator by the use of Norton's theorem. This generator has a different frequency dependence from V_{NA} because of the factor Z_1.

This technique of referring all noise generators to the input of the transducer is an extremely important and general one, which not only provides insight into signal recovery problems but also may lead to a major simplification in the calculations because the overall signal/noise ratio can now be evaluated at one step by integrating all the noise generators over the frequency response of the system. The mean-square values of the non-equilibrium noise generators may, of course, be functions of frequency.

We shall not take the question of amplifier noise further than this stage, not only because of the wide variety of noise models that may be appropriate in different circumstances (each of which modifies the earlier conclusions about optimum transfer functions) but also because it is often possible, by competent circuit design, to ensure that the amplifier noise is negligible.

3. Signal Recovery from the Gravitational Antenna

3.1. The Antenna and its Equivalent Circuit

The general principles involved in detectors of gravitational radiation have been outlined by Fellgett and Sciama.[2] The gravitational antenna consists of a large metal bar resonant at a frequency of about 1 kHz, the mechanical vibration of which is coupled into an electrical circuit by means of a piezo-electric crystal. The antenna is used to detect a force pulse whose waveform is unknown but which is likely to be a double pulse of width comparable with the longitudinal resonance period of the metal bar.

The analysis of the detection system is carried out by the use of an electrical equivalent circuit to represent the whole transducer, i.e. the antenna and the piezo-electric device. Fellgett and Sciama[2] give a circuit in which the electro-mechanical coupling is represented by a transformer but we shall find it more convenient to eliminate the transformer and use the circuit of Fig. 8, which was originally given by Weber.[1]
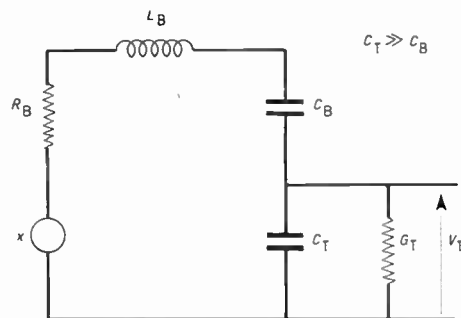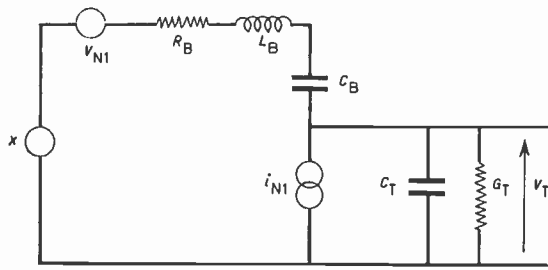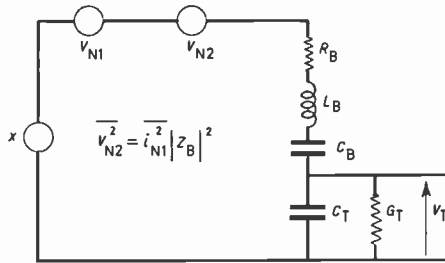

Fig. 8. Equivalent circuit of a gravitational wave detector.

(a) Transducer equivalent circuit showing noise generators.

(b) Circuit with noise generators referred to the input.
Fig. 9.

The piezo-electric device is represented by a capacitance C_T and associated loss G_T, which are the *actual* parameters of the device. The mechanical resonant system is represented by L_B, C_B, R_B with signal input voltage generator x, and the electro-mechanical coupling corresponds to the insertion of the larger capacitance C_T in series with this circuit. The degree of this coupling is represented by the ratio

$$\beta = C_B/C_T \ll 1. \qquad \ldots.(24)$$

From a purely linear-system point of view the ratio C_B/C_T is of little significance because to a good approximation it simply introduces a numerical scaling factor into the system function; however, it is a key factor in signal/noise considerations and we must therefore choose the value of β so that it correctly represents the degree of electro-mechanical coupling. In these terms we may define β by the relationship

$$\beta = \frac{\text{electrical energy stored in } C_T \text{ (open circuit)}}{\text{elastic energy stored in mechanical system}}. \quad ..(25)$$

at any instant. The actual value of β will depend on the properties of the piezo-electric device and on the degree of mechanical coupling between it and the bar, but even with the most effective piezo-electric materials and perfect mechanical coupling it is not possible to achieve values of β greater than about 0·3.

For a given piezo-electric device (assumed to have a low loss angle over the frequency range of interest) with given mechanical coupling to the bar, we are now able to define the equivalent bar capacitance C_B.

If the bar has mechanical resonance angular frequency ω_0 and Q factor Q_B with C_T short-circuited, we have also

$$L_B = 1/\omega_0^2 C_B, \qquad R_B = 1/\omega_0 C_B Q_B. \quad \ldots.(26)$$

The remaining parameter to be determined is the scaling factor between the force F applied to the bar and the equivalent voltage generator x. In order to do this

we must equate the energy supplied to the mechanical system by a gravitational pulse to the energy supplied to the equivalent electrical system by the generator x.

So far as the dominant (longitudinal) mode of oscillation is concerned, the mechanical system can be regarded as a pair of localized masses M coupled by a spring. An impulse of momentum $P = \int F \, dt$ will impart to the mass an energy $W_M = P^2/2M$; similarly in the electrical equivalent circuit an impulse of flux $\phi = \int x \, dt$ will impart to the inductor L_B an energy $W_E = \phi^2/2L_B$. Hence equating W_E and W_M we obtain

$$x = F\sqrt{L_B/M} \ = \ \frac{F}{\omega_0\sqrt{\beta C_T M}}. \qquad \ldots.(27)$$

This approach to the equivalent circuit differs from that of Fellgett and Sciama[2] in that the circuit components are electrical impedances rather than mechanical ones and the generators represent voltages proportional to forces rather than forces themselves.

3.2. Noise Generators

In Fig. 9(a) is shown the transducer equivalent circuit with its thermal noise generators, and in Fig. 9(b) the current generator i_{N1} has been referred, by the use of Thévenin's theorem, to the signal input voltage generator x.

Figure 9(b) is somewhat more complicated than Fig. 2(a), but the principle is the same. The interpretation, however, is a good deal more complicated because of the frequency dependence of the generator V_{N2}. The noise generators in Fig. 9(b) have the mean-square values

$$\delta \overline{V}_{N1}^2 = 2k\theta R_B \delta\omega/\pi \qquad \ldots.(28)$$
$$\delta \overline{V}_{N2}^2 = 2k\theta |Z_B|^2 G_T \delta\omega/\pi \qquad \ldots.(29)$$

where Z_B is the impedance of the series combination of R_B, L_B, and C_B.
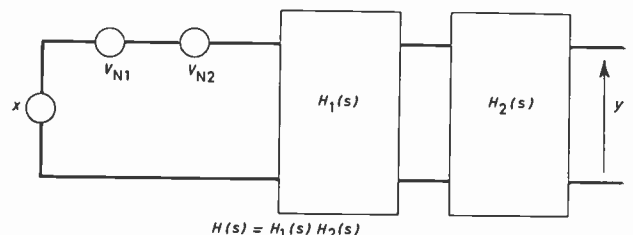

$$H(s) = H_1(s) \, H_2(s)$$

Fig. 10. System diagram used when considering the optimum transfer function.

3.3. The Optimum Transfer Function

When considering the optimum transfer function, we use the system diagram of Fig. 10, which is derived from Fig. 9(b).

This is analogous to Fig. 5, although the overall transfer function $H(s)$ is now a dimensionless voltage ratio. The function $H_1(s)$ is the transducer transfer function, and before writing this down we must decide whether the input impedance of the active system $H_2(s)$ should approximate to an open circuit or to a short circuit. (In principle the input impedance might have any value, but in practice we normally consider the two 'ideal' alternatives of the voltage amplifier with series

feedback and the trans-impedance or charge amplifier with parallel feedback.) The normal choice for a piezo-electric transducer would be a charge amplifier, and furthermore in this particular case, as can be seen from Fig. 9, the short-circuit alternative will avoid the problem of the noise spectrum peaking at a different frequency from the signal spectrum. We shall therefore assume that $H_2(s)$ is a current-to-voltage (impedance) function, and that $H_1(s)$ must therefore be written as an admittance function relating the short-circuit current at the trans-ducer terminals to the input voltage generators. From Fig. 8 we derive the expression

$$H_1(s) = 1/Z_B(s) = 1/[R_B + sL_B + 1/sC_B]$$

$$= \frac{1}{R_B} \cdot \frac{s\omega_0/Q_B}{s^2 + s\omega_0/Q_B + \omega_0^2}. \qquad \ldots(30)$$

If the spectral density of the noise generators were independent of frequency we should now choose for $H_2(s)$ a function whose product with $H_1(s)$ gave a broad-band function like that of equation (16). However, as we see from (29), the spectral density of V_{N2} has a minimum at the bar resonance frequency and increases rapidly on each side of resonance, implying that the optimum signal/noise ratio will be provided by a transfer function which selects a frequency range around ω_0. To a good approxi-mation the impedance Z_B can be expressed in the region of resonance as

$$Z_B = R_B + 2jL_B(\omega - \omega_0)$$

so that equation (29) becomes

$$d\bar{V}_{N2}^2 = 2k\theta[R_B^2 + 4L_B^2(\omega - \omega_0)^2]G_T \, d\omega/\pi. \quad \ldots(31)$$

If we consider a frequency interval of width $\Delta\omega$ centred at ω_0, we see that the integral of the reactance term makes a contribution to the mean-square noise which is proportional to the *cube* of the bandwidth $\Delta\omega$, while (as in the non-resonant case considered in Section 2) the square of the peak signal response will increase as the *square* of $\Delta\omega$. The pulse response will be a slowly decaying 'ring' rather than a single peak, but the entire signal information is conveyed by the initial peak because the subsequent decay is no slower than the decay of noise components so that nothing is gained by averaging over the decay time.

The need for a narrow-band overall response can be seen from another point of view by considering the transfer function $H_2(s)$ that would be required for the active filter in order to counteract the sharply resonant nature of the transducer transfer function and give a broad-band overall response. The gain would require to be much higher in the frequency range outside the resonance region and in this range would give rise to an amount of noise disproportionate to its contribution to the signal.

We now find ourselves in a situation which has many physical analogies with that of Section 2, with the system response now centred around ω_0 rather than zero fre-quency. The decay time of the resonant circuit, which is $2Q_B/\omega_0$, is analogous to the decay time T_1 of the first-order circuit of Fig. 1, and the 'short pulse' condition $T_x \ll T_1$ is equivalent to the condition that the spectrum of the incident gravitational pulse is broad compared

with the bandwidth ω_0/Q_B. Since $Q_B > 100$ this is equivalent to the condition that the gravitational input should not consist of a long sinusoidal function of frequency ω_0.

The 'short pulse' condition being satisfied, we can expect to improve the signal/noise ratio by increasing the bandwidth—provided that this can be done without introducing additional noise generators. Thus a noiseless resistor inserted in series with R_B will reduce the noise current while having no effect on the signal amplitude because the initial value of the step response of the circuit depends on the reactive elements only. The reduction in Q will of course increase the rate of decay of the step response, but this does not correspond to a reduction in the effective amplitude of the signal.

As before, we are using the active filter as the preferred alternative to the 'cold damping' method. The simplest practicable form for the overall transfer function is that of (16), displaced so as to be centred about ω_0 rather than zero frequency. The required function is

$$H(s) = \frac{As^2\omega_0^2}{Q_M^2(s^2 + s\omega_0/Q_M + \omega_0^2)^2} \qquad \ldots(32)$$

with $Q_M = \omega_0 T_2/2$, to a very good approximation if $Q_M \gg 1$.

The parameter Q_M is the overall Q factor of the measurement system ($Q_M < Q_B$) and we can now calcu-late its optimum value. From Fig. 10 in conjunction with (28), (29), and (32) we obtain after integration over the frequency range from 0 to ∞ the following expression for the mean-square value of the noise component y_N in the output:

$$\bar{y}_N^2 = \frac{A^2k\theta}{2\beta C_T Q_M} \left[\frac{1}{Q_B} + \frac{1}{\beta Q_B^2 Q_T} + \frac{3}{\beta Q_T Q_M^2} \right]. \ldots(33)$$

In this expression the parameter Q_T is defined by the relation

$$Q_T = \omega_0 C_T/(G_T + G_{NA}) \qquad \ldots(34)$$

where G_{NA} is the mean effective noise conductance of the active filter; this depends on both the voltage and the current noise generators (see Fig. 7) and is a function of frequency, but we find that in the practical situations discussed here the value of G_{NA} is substantially constant over most of the relevant frequency range and for the purposes of system design it can be taken as constant.

If the input voltage x is a step function of magnitude x_P, we have also for the peak value y_P of the response the relation

$$y_P = Ax_P/Q_M e \qquad \ldots(35)$$

and the corresponding peak energy stored in the mech-anical system is given by

$$W_S = \beta C_T x_P^2/2. \qquad \ldots(36)$$

From (33), (35), and (36), in conjunction with the condition $\beta Q_T Q_B \gg 1$ which applies to all the practical systems that have been proposed, we now obtain the following expression for the signal/noise ratio:

$$S/N = y_P^2/\bar{y}_N^2 = \frac{4W_S}{k\theta e^2} \cdot \frac{1}{\dfrac{Q_M}{Q_B} + \dfrac{3}{\beta Q_M Q_T}} \qquad \ldots(37)$$

The choice of optimum overall transfer function is now made by optimizing (37) with respect to variation of Q_M, the result being

$$\left.\begin{array}{l} \text{optimum value of } Q_M = \sqrt{3Q_B/\beta Q_T} \\ \\ \text{optimum S/N} = \dfrac{2W_s}{k\theta e^2} \cdot \sqrt{\dfrac{\beta Q_T Q_B}{3}} \\ \\ \text{minimum detectable energy} = \dfrac{\sqrt{3}}{2} \cdot \dfrac{k\theta e^2}{\sqrt{\beta Q_T Q_B}} \end{array}\right\} \quad ..(38)$$

On the basis of (38) we may now compare the two main mechanical schemes that have been proposed: the single bar originated by Weber[9] and also used by Tyson,[3] and the split bar which was suggested by Dr. P. S. Aplin of the University of Bristol, has been discussed by Gibbons and Hawking[4], and is currently being operated by Drever.[5] In making the comparison we must assume a value for Q_T, and we shall here take a value of 10^3 which should be attainable with modern piezo-electric materials and properly designed amplifiers. The approximate values of the relevant parameters are given in Table 1.

Table 1. Comparison of optimized detection systems.

	$\omega_0/2\pi$	Q_B	β	Q_T	Minimum detectable energy	Optimum Q_M
single bar	1 kHz	10^5	10^{-5}	10^3	$0.2\,k\theta$	5500
split bar	1 kHz	10^3	10^{-1}	10^3	$0.02\,k\theta$	5.5

In terms of minimum detectable energy, Table 1 indicates that there is a factor of 10 between the single bar and the split bar. On the other hand, we are implicitly assuming that the experimental situation allows the choice of optimum Q_M, and it is by no means certain that this is the case. It has been pointed out that for coincidence measurements a time resolution of a few milliseconds would be desirable, and this implies a Q_M value of less than say 10 at the resonance frequency of about 1 kHz. If the single bar were operated in a system with $Q_M = 10$ it would show a penalty of nearly 3000 compared with the split bar.

In considering the system design, we have been assuming that the input $x(t)$ is in the form of a step function. The response of a resonant system to a double pulse is discussed in Section 2.4, and these considerations will apply equally to the present split-bar and single-bar systems because the resonance frequencies are about the same.

3.4. Rectification

Having decided that the optimum overall transfer function is a narrow-band one, we are faced with the problem that the response to a pulse input will be a decaying sinusoid rather than a pulse. In fact, the output voltage of the system will be expressible in the form

$$y(t) = f_1(t) \cos \omega_0 t + f_2(t) \sin \omega_0 t \quad ....(39)$$

where $f_1(t)$ and $f_2(t)$ are low-frequency functions in the sense that their power spectra are concentrated in a frequency range well below ω_0. An alternative formulation to (39) is

$$y(t) = f_E(t) \cos (\omega_0 t + \alpha_E(t)) \quad ....(40)$$

where $f_E(t)$ is the *envelope function* defined by the relation

$$f_E(t) = [f_1^2(t) + f_2^2(t)]^{1/2} \quad ....(41)$$

In practice, the functions $f_1(t)$ and $f_2(t)$ can be extracted by the use of analogue multipliers in conjunction with a two-phase oscillator, and $f_E(t)$ can be obtained by means of an 'envelope detector' consisting of diode rectifiers connected in the feedback path of an operational amplifier; in each case the output will contain unwanted components of frequency at least $2\omega_0$ which can be removed by means of a suitable filter.

In the type of system we are considering here the function $f_E(t)$ contains all the signal information, the phase angle α_E (which remains constant during the response to a given pulse) being arbitrarily defined.

A question which has received some attention is how the signal/noise ratio would be affected by 'post-detection filtering', i.e. by the use of low-frequency filters following the envelope detector. Although there is no simple answer to this in the general case, a very simple result emerges when we study the response of a resonant system to pulse inputs and random-noise inputs. In this case a post-detection filter with response function $H_D(s)$ (this being assumed of course to be a low-frequency function) has the same effect as a narrow-band filter in the linear (pre-detection) part of the system, with response function $H_3(s)$ given by the relation

$$H_3(s) = H_D(s - j\omega_0) + H_D(s + j\omega_0) \quad ....(42)$$

3.5. Practical Design of the Active Filter and Rectifier

The required filter function is (32) divided by (30), that is

$$H_2(s) = \frac{AR_B Q_B(s^2 + s\omega_0/Q_B + \omega_0^2)s\omega_0}{Q_M^2(s^2 + s\omega_0/Q_M + \omega_0^2)^2}. \quad ....(43)$$

In the same way that the overall resonant transfer function (32) was derived from the low-frequency function (16) by a frequency shift, so the filter function (43) can be derived from the filter function (15) as the sum of the functions obtained by substituting $(s \pm j\omega_0)$ for s:

$$H_2(s) \simeq \frac{AR_B(1 + sT_1 + j\omega_0 T_1)}{(1 + sT_2 + j\omega_0 T_2)^2} + \frac{AR_B(1 + sT_1 - j\omega_0 T_1)}{(1 + sT_2 - j\omega_0 T_2)^2} \quad ....(44)$$

where $Q_M = \omega_0 T_2/2$, $Q_B = \omega_0 T_1/2$.

The multiplier R_B is required because the filter function is now a current-to-voltage transfer rather than the dimensionless voltage gain given by (15). Function (44) approximates closely to (43) in the high-Q limit, the discrepancy in the complex poles and zeros being less than ω_0/Q_M^2 and in the real zero less than ω_0/Q_M.

The frequency-shift relationship between (44) and (15) is of more than academic interest, because it indicates a very convenient method of actually constructing the filter. In the 2-path structure of Fig. 11, the four boxes

Fig. 11. Two-path realization of equation (44). All unmarked resistors are of value R_0 and $R_0C_1 = T_1$, $R_0C_2 = T_2$.
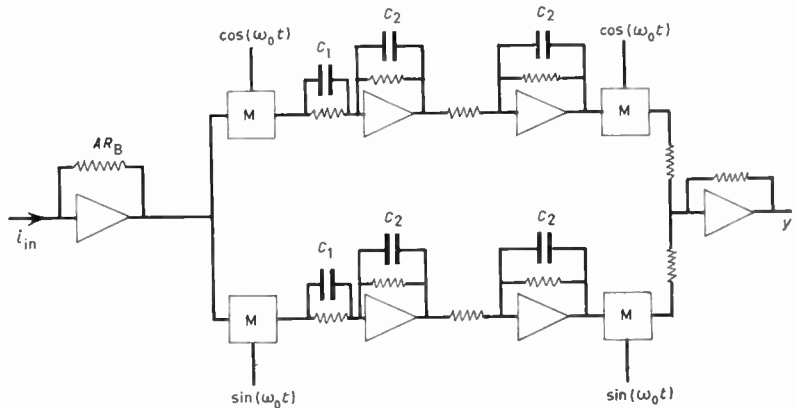
marked M are modulators (analogue multipliers) and the active RC filters between the modulators realize the low-frequency function (15). The modulating function is obtained from a two-phase local oscillator so that the filter can easily be 'tuned in' to the required resonance frequency, and the Q values can readily be changed because they are determined only by the RC products.

Although Fig. 11 shows the operating principle of a suitable filter, the construction must be modified to incorporate an envelope detector. Functions $f_1(t)$ and $f_2(t)$ of equation (39) are available at the outputs of the low-frequency filters, so the second pair of modulators may be replaced by analogue squarers (in practice, the same analogue multipliers connected in a different way) so that the output of the summing amplifier is

$$f_E^2(t) = f_1^2(t) + f_2^2(t). \qquad \dots(45)$$

This is believed to be the principle currently used by Weber.[1] Alternatively an envelope detector can be used at the output, in which case we may remove the second stage (transfer function $1/(1+sT_2)$) from each low-frequency filter and connect one of them at the output; as indicated in Section 3.4, this will remove unwanted high-frequency components while leaving the effective overall system function unchanged.

4. Conclusions

We have endeavoured to show that the essential signal-recovery problem in the case of the gravitational antenna is the design of a suitable linear active filter. We have suggested a suitable fourth-order filter function (43) and have shown how its parameters may be optimized, and one way in which it could be realized. It may be that a filter of higher order would give better results, but in view of the general nature of the discussion it seems unlikely that any substantial improvement could be obtained by increasing the complexity in this way.

5. Acknowledgments

6. References

1. Weber, J., 'Evidence for the discovery of gravitational radiation', *Phys. Rev. Letters*, 22, p. 1320, 1969.
2. Fellgett, P. B. and Sciama, D. W., 'Gravitational wave astronomy: an interim survey', *The Radio and Electronic Engineer*, 41, No. 9, p. 391, September 1971.
3. Tyson, J. A., 'Pulsed tensor radiation experiment at 4 kHz and absolute antenna calibration'. Paper presented at the Sixth International Conference on Gravitation and Relativity, Copenhagen, July 5th–10th, 1971.
4. Gibbons, G. W. and Hawking, S. W., 'The detection of short bursts of gravitational radiation', *Physical Review* D, 4, No. 8, p. 2191, October 1971.
5. Drever, R. W. P., 'Observations on pulse response of a wide-band gravitational wave detector'. Paper presented at the Sixth International Conference on Gravitation and Relativity, Copenhagen, July 5th–10th, 1971.
6. McCombie, C. W., 'Fluctuation theory in physical measurements', *Rep. Prog. Phys.*, 16, 266, 1953.
7. Faulkner, E. A., 'Reciprocity and passivity', *Electronics Letters*, 6, No. 10, p. 304, 1970.
8. Gillespie, A. B., 'Signal, Noise and Resolution in Nuclear Counter Amplifiers (Pergamon, Oxford, 1953).
9. Weber, J., 'Observation of the thermal fluctuations of a gravitational wave detector', *Phys. Rev. Letters*, 17, p. 1228, 1966.

An Analysis of Transport Delay Simulation Methods

J. B. KNOWLES,
M.Sc.Tech., Ph.D., C.Eng. M.I.E.E. *

and

D. W. LEGGETT, M.Sc. *

SUMMARY

The inherent limitations of existing lumped parameter networks and electro-mechanical systems for simulating a continuously variable transport delay are shown. Electronic sampled-data storage schemes overcome these difficulties and result in a much higher performance transport delay unit. Cost effectiveness is shown to be markedly enhanced by linear interpolation and m.o.s. shift-register storage of the sampled data.

* Atomic Energy Establishment, U.K.A.E.A., Winfrith, Dorchester, Dorset.

List of Principal Symbols

P_N Nth order Padé approximation
ω angular frequency (rad/s)
τ generic symbol for a time constant
ζ generic symbol for a damping factor
s Laplace transform variable
S_N Nth-order Single's approximation
Φ spectral density function
R_m mth Fourier series coefficient for a rectangular pulse train

Other symbols are defined in text.

1. Introduction

Large-scale mathematical models of a nuclear reactor plant are developed prior to the finalization of a design for the following reasons:

(i) They indicate the dynamic performance of the system particularly with respect to load-following and fault conditions.

(ii) They give insight into the 'controllability' of the the system.

(iii) They provide operators with simulated plant experience even for hazardous conditions.

The practical utility of such models is therefore unquestionable.

As a result of pipework or any form of flow between units, transport delays or distance/velocity lags occur in many types of plant. Some of these delays have a marked effect on the stability and dynamic performance of the plant, so that they must be included in an analogue simulation of the mathematical model. Further, it is often necessary to simulate a continuously variable transport delay, because the velocities of the fluid are a function of the power demanded from the plant. Hence the problem is that of synthesizing the transfer function $\exp(-s\tau)$, where the delay τ is controlled by another simulated variable. After many years' experience in the simulation of nuclear plant, the Atomic Energy Establishment, Winfrith, produced the following extracted specification for a continuously variable transport delay unit:

Step-function response
   time to 1% $\leqslant 10$ ms } with a constant
Step-function response   } delay
   overshoot $\leqslant 1\%$

Delay band 1–10 s over 0–1 Hz   } with 0·1%
Delay band 10–100 s over 0–0·1 Hz } uncertainty
Delay band 0·1–1 s over 0–10 Hz   }

There are three basic methods of realizing a transport delay simulation unit:

(i) Lumped parameter networks involving operational amplifiers.

(ii) Electromechanical devices.

(iii) Electronic sampled-data storage schemes.

An examination of the inherent limitations of these schemes with respect to the above specification will be shown to lead quite naturally to the design finally

adopted. The application of a delay unit is not restricted to the nuclear engineering industry. Such a computational element finds application in the simulation of steel rolling mills, long-range flight navigational systems, motor-car suspension units, etc.

2. Lumped Parameter Networks

It can be seen from Appendix 1 that the gain frequency characteristic of an Nth order Padé approximation[1, 2] is the ideal value,

$$|P_N(j\omega)| = 1 \cdot 0$$

Further, at low frequencies

$$\arg P_N(j\omega) = -\omega\tau \quad \text{for } \omega\tau \ll 1$$

but at high frequencies the phase frequency characteristic is non-ideal, which leads to bandwidth limitations. The error quantity

$$-(\arg P_N(j\omega) + \omega\tau)/\omega\tau = \text{phase error}$$

is shown as a function of normalized frequency in Fig. 1. An important aspect of the networks is that an
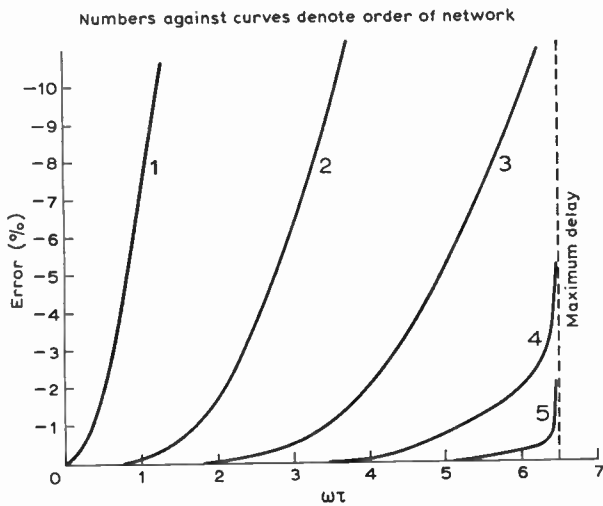
Numbers against curves denote order of network

Fig. 1. Error in phase characteristic of Padé networks. Numbers against curves denote order of network.

arbitrary increase in the order of the approximation fails to reduce the error beyond a delay-bandwidth product of 6·5. This value is below specification. In addition, the non-minimum phase characteristic of the networks results in quite intolerable step-function responses, as illustrated for a second-order network in Fig. 2.

Improvements in step-function response and circuit noise level have been claimed using the so-called cut-product approximation[3] or an empirical choice[4] of damping factors in the expression:

$$S_N(s) = \frac{1}{1 + s\tau_0} \prod_{k=1}^{N} \left[\frac{(\tau_k s)^2 - 2\zeta_k(\tau_k s) + 1}{(\tau_k s)^2 + 2\zeta_k(\tau_k s) + 1} \right] \quad \ldots\ldots(1)$$

The time lag τ_0 is small relative to the actual delay required, but sufficiently large to attenuate spurious noise signals.

$$P_2(s) = \frac{1 - \frac{1}{2}(s\tau) + \frac{1}{12}(s\tau)^2}{1 + \frac{1}{2}(s\tau) + \frac{1}{12}(s\tau)^2}$$

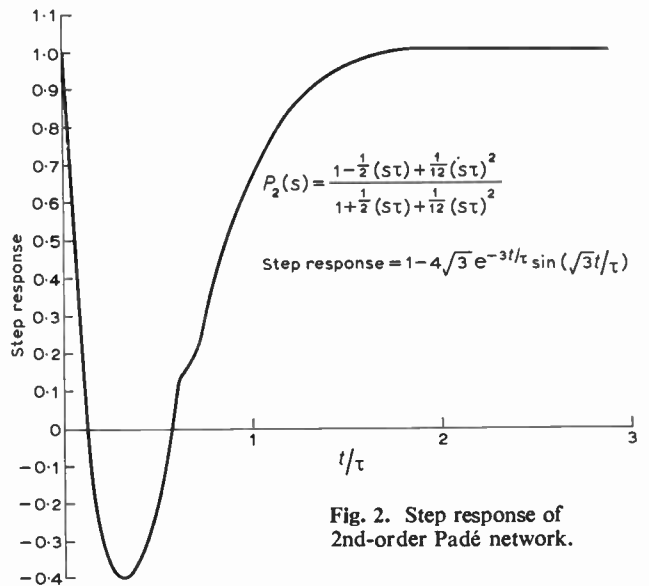$$\text{Step response} = 1 - 4\sqrt{3} \, e^{-3t/\tau} \sin(\sqrt{3}t/\tau)$$

Fig. 2. Step response of 2nd-order Padé network.

For a fourth-order empirical approximation[4] the transport delay generated at low frequencies is shown in Appendix 2 to be:

$$\tau = \tau_1(9 \cdot 83 + \tau_0/\tau_1) \qquad \ldots\ldots(2)$$

The expression for the actual phase lag of the network can be normalized with respect to the delay defined in equation (2), and the error in this normalized phase frequency characteristic is shown in Fig. 3. It is observed that the frequency response of a Single approximation is very inferior to the Padé network of the same order. Thus the mentioned improvement in step-function response is achieved at the expense of steady-state accuracy, which is a not uncommon exchange in systems engineering.
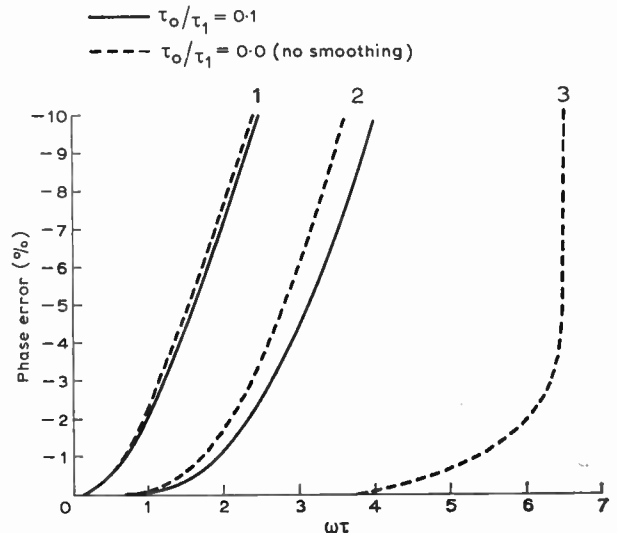
Fig. 3. Error in phase characteristic of different lumped approximations.

For the Single approximation $\tau = (9 \cdot 83 + \tau_0/\tau_1)\tau_1$.
For the smoothed Padé approximation $\tau = (1 \cdot 0 + \tau_0/\tau_1)\tau_1$.
(1) The Single approximation.
(2) 2nd-order smoothed Padé approximation.
(3) 4th-order Padé approximation.

More recent work[14] has produced lumped parameter active networks for which the overshoot and initial undershoot in the step-function response can be controlled to as low as 0·5%. Unfortunately, the ratio of delay-time to rise-time for the step response (equivalent to delay-bandwidth product) deteriorates with decreasing overshoot. Even allowing 10% overshoots with a sixth-degree denominator polynomial, this ratio is only 5·046 which is below specification.

So far most attention has been directed at lumped parameter networks which have been used previously to generate a fixed transport delay, but several investigators[5, 6] have proposed schemes for synthesizing a continuously-variable transport delay using the same basic components. It is pertinent to question if such techniques are capable of meeting the previous specification with a variable delay, when their constant delay performance is so inadequate. Further, a transport delay is inherently associated with a distributed parameter system, whose defining relationships are partial differential equations. Hence a lumped parameter network, with its associated set of ordinary differential equations, is believed to be fundamentally the wrong technique for constructing a continuously variable transport delay unit. This conclusion is confirmed to some extent by the very limited delay-bandwidth product, over which the convergence of the lumped network techniques is guaranteed.[6] Finally, the generation of variable transport-delays by analogue networks evidently requires the adjustment of the circuit's parameters. In common with other situations, this adjustment is far less easily implemented than for digital systems, such as that described later.

3. Electromechanical Systems

Magnetic recording systems may be used to generate a transport delay by displacing the write and read heads in correct relation to the tape motion. The transport delay obtained is evidently

$$\tau(t) = \int_{x=0}^{D} \frac{dx}{v(t)}$$

where D is the physical separation of the write and read heads, and v is the tape velocity. A variable transport delay is achieved by using a velocity servo to control the speed of the tape winding motor. If the tape velocity changes relatively little over the transit time of a data element, then the transport delay produced is

$$\tau = D/v(t)$$

Direct amplitude or frequency modulation of analogue data onto the magnetic tape is not suitable for the present purpose, because variations in tape velocity induce simultaneous amplitude variations in the signal recovered by the read head. This difficulty has been circumvented[7] by converting the analogue data into p.w.m. format. However, this and all other magnetic recording schemes possess other serious disadvantages.

If an endless tape transport is employed, the splicing of the tape destroys its homogeneous structure. As a result, spurious short duration transients of up to 20% peak output have been observed by G. B. Collins of
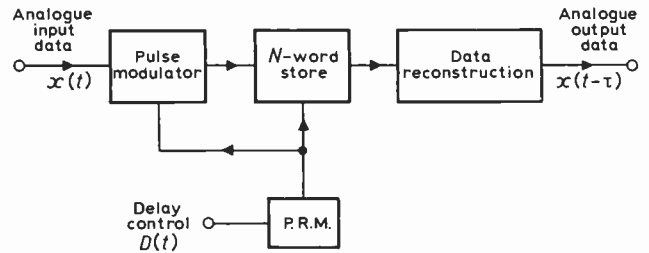

Fig. 4. Electronic sampled-data storage scheme.

A.E.E.W.† when the splice passes under the read head. In addition, the inertia of the tape drive mechanism produces a sluggish step-function response, whose duration is several seconds. It is therefore impossible to simulate step changes in fluid velocity with reasonable accuracy, unless the corresponding transport delay is of the order of several minutes. The same performance limitation clearly applies to any proposed design which utilizes a mechanical drive. Thus, electromagnetic tape, paper tape and stepped magnetic drum systems[8] are all unsuitable for the present application. These considerations indicate the necessity of using the completely non-mechanical techniques described in the next section.

4. Electronic Sampled-data Schemes

An electronic sampled-data storage scheme incorporates the basic elements shown in Fig. 4. The response time of the electronic pulse rate modulator (p.r.m.), which consists of a reset integrator circuit, is such as virtually to eliminate the previously described performance limitations of electromechanical systems. For the present application, there are three possibly viable data storage techniques.

(i) Analogue storage.

(ii) Core storage.

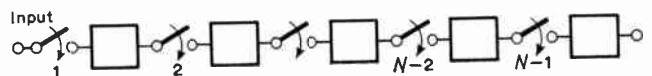(iii) M.o.s. shift register storage as m.s.i. circuits.

Fig. 5. Sequential analogue store.

Digital integrated circuits have been available as relatively inexpensive elements for some time, and the same is now true for linear circuits. There is therefore the possibility of using so-called 'sample and hold circuits' for the sequential store in the manner indicated in Fig. 5. The m.o.s.f.e.t. switches included in each element are normally open, and they are designed to close then open sequentially in the order

$$N-1, N-2, N-3, \ldots 4, 3, 2, 1$$

Currently available units have a sampling aperture of 0·1 μs which is small enough to be useful in a practical design. A core store is a familiar digital storage system, and no explanation of this intricate device is considered

---

† Unpublished report.

necessary. Metal oxide silicon technology has enabled a very high component density to be achieved at a low per unit cost. One such product is a 200-bit shift register, which possesses only six terminal connexions for input, output, clock and power supply. This limited number of external leads results in a very inexpensive device, that is particularly attractive as regards the construction of the required sequential store. The relative economics of these different data storage units will be considered in more detail later.

As the delay unit is to be used with an analogue computer system, the sampled-data output of the sequential store must be reconstructed into analogue form. There are many possible methods for achieving this, but from the viewpoint of limiting hardware complexity only clamping and linear interpolation are considered here. If the most recent past output sample from the store occurred at t_k, then the output of the clamp interpolator is defined by

$$z(t) = y(t_k) \quad \text{for } t_k \leqslant t \leqslant t_{k+1} \qquad \ldots\ldots(3)$$

and the output of the linear interpolator is defined by

$$z(t) = y(t_k) + \frac{t}{t_{k+1} - t_k}\left[y(t_{k+1}) - y(t_k)\right] \quad \text{for } t_k \leqslant t < t_{k+1} \\ \ldots\ldots(4)$$

It will be observed that equation (4) would be physically unrealizable, if it were not for the time-delay generated by the store. Schematic procedures for the practical implementation of equations (3) and (4) are given in Fig. 6.

Reference to Fig. 4 establishes that the transport delay (τ) imposed on the input sample at time t_n is given implicitly by the equation:

$$\text{number of store words } (N) = \lambda \int_{t_n}^{t_n+\tau} D(t)\, \mathrm{d}t$$

where λ is a constant associated with the pulse rate modulator. If the delay control signal $D(t)$ changes relatively little over the transition period for this sample through the store, then

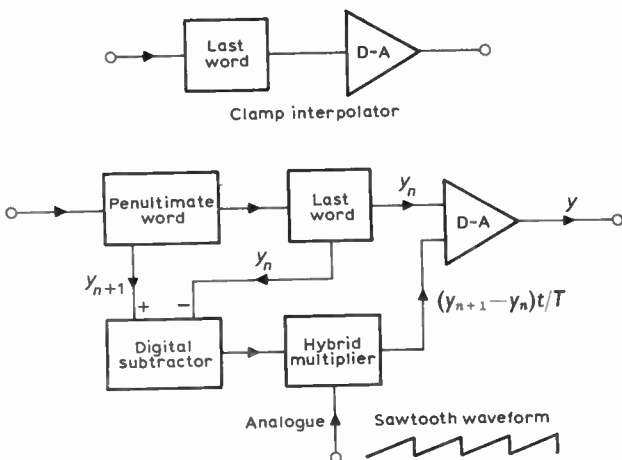$$\tau = N/f_s \qquad \ldots\ldots (5)$$

Clamp interpolator

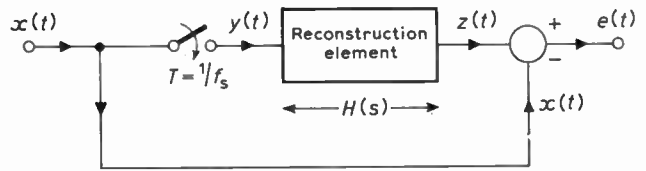**Fig. 6.** Clamp interpolation and linear interpolation with digital data.

Fig. 7 Model p.a.m. carrier system for assessing reconstruction errors.

where f_s is the corresponding clock frequency. The clock frequency is also the sampling frequency for the input data, so that it cannot be reduced below a certain specific value. This value must enable the specified accuracy to be achieved with the adopted reconstruction technique. It is proposed therefore to analyse the reconstruction errors appertaining to clamping and linear interpolation of sampled-data at a fixed sampling frequency. By this means, it is possible to select the reconstruction element that provides the required transient† and steady-state performance for the minimum sampling-frequency and therefore the minimum number of store words.

5. Analysis of Reconstruction Errors

Neglecting data quantization errors[9, 10] the performance of the reconstruction element may be assessed using the model p.a.m. carrier system shown in Fig. 7. By virtue of the narrow input sampling window and the rapid data transfers within the delay unit, the width (γ) of the pulses forming the signal $y(t)$ may be regarded as significantly smaller than the constant sampling period. The filter $H(s)$ must differentially attenuate the side-bands produced by the sampling operation so as to produce a close approximation to the signal $x(t)$. This process is illustrated in Fig. 8.
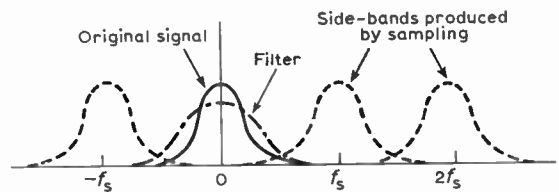

Fig 8. Spectra associated with the data reconstruction process.

From Appendix 3 it can be seen that the required error quantities are as follows:

$$\overline{e^2}\Big|_{\text{clamp}} = \frac{(2\pi f_0/f_s - \sin 2\pi f_0/f_s)^2 + (1 - \cos 2\pi f_0/f_s)^2}{(2\pi f_0/f_s)^2} + \\ + \sin^2(\pi f_0/f_s)\sum_{m=1}^{\infty}\left[\frac{1}{\pi^2(m+f_0/f_s)^2} + \right. \\ \left. + \frac{1}{\pi^2(m-f_0/f_s)^2}\right] \quad (6)$$

† With a sampled-data scheme, the design specification of the unit relates only to the transient and steady-state reponses of the reconstruction element.

$$\overline{e^2}\Big|_{\substack{\text{linear}\\\text{interpolator}}} = \frac{(\pi f_0/f_s + \sin \pi f_0/f_s)^2 (\pi f_0/f_s - \sin \pi f_0/f_s)^2}{(\pi f_0/f_s)^4} +$$

$$+ \sin^4(\pi f_0/f_s) \sum_{m=1}^{\infty} \left[\frac{1}{\pi^4(m + f_0/f_s)^4} + \frac{1}{\pi^4(m - f_0/f_s)^4} \right] \quad (7)$$
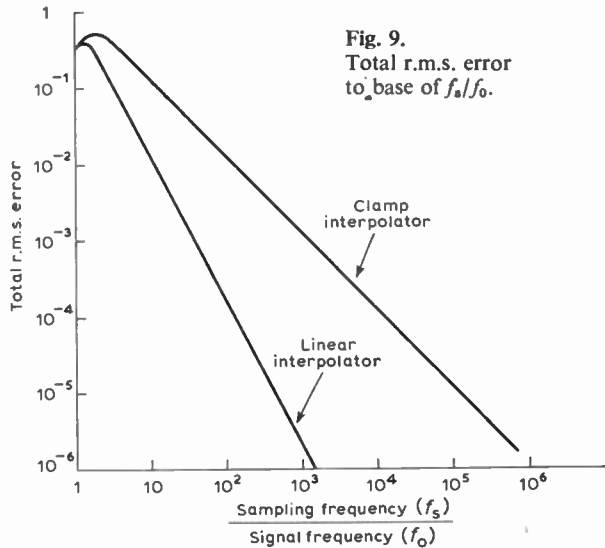
Fig. 9.
Total r.m.s. error to base of f_s/f_0.

Table 1. Minimum sampling frequencies

	Input frequency ranges		
	0–1 Hz	0–0·1 Hz	0–10 Hz
Minimum sampling frequency for clamp interpolator	4·4 kHz	440 Hz	44 kHz
Minimum sampling frequency for linear interpolator	65 Hz	6·5 Hz	650 Hz

and the results are plotted in Fig. 11. If the sampling frequency for the linear interpolator is now increased to 100 Hz in the 0–0·1 Hz and 0–1 Hz frequency ranges and assumes the minimum value of 650 Hz in the 0–10 Hz range, then the step function response and steady-state design specifications are now completely satisfied. However, the cost saving obtained by the reduction in storage elements must be weighed against the increased cost of the linear interpolation circuit itself. The total cost incurred in the realization of a typical 5-second delay in the 0–1 Hz frequency range is now used to assess the cost-effectiveness of the two reconstruction elements.

The appropriate store lengths (L) required for a 5-second delay in the 0–1 Hz range are evaluated from Fig. 9 and equation (5) as:†

$$\left.\begin{array}{l}\text{clamp interpolation } L = 2\cdot2 \times 10^4 \text{ words}\\ \text{linear interpolation } L = 0\cdot5 \times 10^3 \text{ words}\end{array}\right\} \quad ...(9)$$

Even with digital stores as large as 20k words, m.o.s. gate technology is claimed to provide a cheaper sequential memory than a magnetic core system.[13] Thus the choice remains only between analogue and m.o.s. shift register storage. At current prices, the amount of storage space demanded by equation (9) is such as to make the digital technique the economic choice, despite the extra expenditure on a–d and d–a convertors. Although a
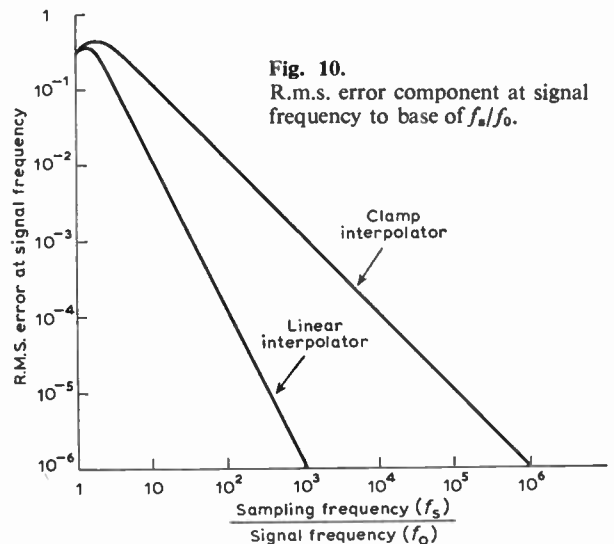
6. Choice of Basic System Elements

Using equations (6) and (7), the total r.m.s. steady-state errors in the amplitudes of the two reconstruction elements are plotted against the normalized frequency f_s/f_0 in Fig. 9 for the unit power sinusoidal input. The steady-state r.m.s. error components at the input signal frequency are obtained using the first terms only in equations (6) and (7). These results are shown in Fig. 10 again for the unit power sinusoidal input. It is observed by comparing Fig. 9 and Fig. 10 that the reconstruction element's error is completely dominated in both cases by the input signal frequency component. Although a computing amplifier may not respond significantly to frequency components around multiples of the sampling frequency, it will certainly respond to the input signal frequency component. Hence the relative superiority of the linear interpolation element as suggested by Fig. 9 is quite real, especially at the higher sampling frequencies.

Examination of Fig. 9 shows that the largest steady-state error occurs at the highest input frequency. For most probability distributions, three times the standard deviation contains most, if not all, probable values of the variate. Hence for a reconstruction accuracy of $0\cdot1\%$ it is appropriate to put

$$\sqrt{\overline{e^2}} = 3\cdot33 \times 10^{-4} \, (= \tfrac{1}{3} \times 10^{-3}) \qquad ......(8)$$

in Fig. 9, and the minimum sampling rates which are consistent with the previous design specification are obtained as shown in Table 1.

The step-function responses of the reconstruction elements are easily derived from equations (3) and (4),

Fig. 10.
R.m.s. error component at signal frequency to base of f_s/f_0.
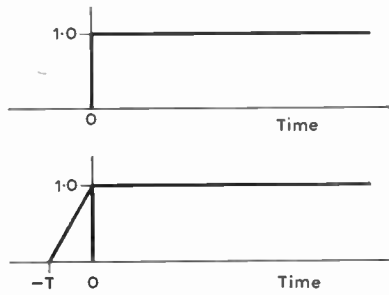
† For a 100 Hz sampling rate.

Fig. 11. Step function responses of clamp and linear interpolators.

linear interpolation element involves relatively more hardware, substantial overall cost economies of the order of 40 : 1 may be achieved at the moment by virtue of the significantly smaller m.o.s. storage requirement. Hence a linear interpolator element with m.o.s. shift register storage and an electronic pulse rate modulator is judged to represent currently the best design for the continuously variable transport delay simulator unit.

7. Conclusions

Lumped parameter networks and electromechanical systems have been used previously for simulating a continuously variable transport delay. It is established that the inherent limitations of these methods are:

   (i) Limited delay-bandwidth product.

   (ii) Poor step function response.

   (iii) Relatively sluggish response of the mechanical drive of tape or drum.

   (iv) Tape splicing problems.

Completely electronic sampled-data storage schemes overcome these problems, and result in a much higher performance unit. For this form of realization, significant economic advantages are shown to accrue from linearly interpolating the sampled-data as compared to simple clamping. Cost effectiveness is still further enhanced by the use of m.o.s. shift-register storage. The proposed system achieved a performance which is markedly higher than previous designs, whilst maintaining an economic price.

The hardware realization and design considerations of the proposed transport delay unit form the basis of the companion paper.[15]

8. References

1. Johnson, C. L., 'Analog Computer Techniques' (McGraw-Hill, New York, 1956).

2. Smith, G. W. and Wodd, R. C., 'Principles of Analog Computation' (McGraw-Hill, New York, 1959).

3. Warfield, J. N. and Weaver, L. E., 'Cut-product approximations in electronic analogue computers', I.R.E. Regional Conference Paper, Dallas 1958.

4. Single, C. H., 'An analogue for process lags,' *Control Engineering*, 3, p. 113, October 1956.

5. Chernyshev, M. K., 'Analysis of the limiting possibilities of devices for the reproduction of a variable delay using operational amplifiers', *Avtomat. i Telemekh.* (USSR) No. 12, p. 2249, December 1965. In Russian. English translation in: *Automat. Remote Control* (USA), **26**, No. 12, p. 2172, December 1965.

6. Norkin, K. B. and Shoya, L. D., 'Certain methods for simulating an element with inertialess variable time delay', *Avtomat. i Telemekh.* (USSR), No. 4, p. 154, April 1969.

7. Kohr, R. H., 'Simulation of automobile suspension and steering systems', Mid-West Simulation Council Meeting, September 1958.

8. Simulog Group Model 1.011, Electronics Associates Inc., Long Branch, N.J.

9. Knowles, J. B. and Edwards, R., 'Effects of a finite-word length computer in a sampled-data feedback system', *Proc. Instn Elect. Engrs*, 112, p. 1192, 1965.

10. Knowles, J. B. and Edwards, R., 'Finite-word length effects in multirate direct digital control systems', *Proc. Instn Elect. Engrs*, **112**, p. 2376, 1965.

11. Laning, J. H. and Battin, R. H., 'Random Processes in Automatic Control' (McGraw-Hill, New York, 1956).

12. Knowles, J. B. and Leggett, D. W., 'A Time Series Analysis and Evaluation of Digital Transport Delay Simulation Units', A.E.E.W. M914, 1969.

13. 'INTEL' advertisement, *Electronics*, October 1970.

14. Rakovitch, B. D. and Djurich, B., 'Method of synthesis of non-minimum phase transfer functions for time-delay simulation', *The Radio and Electronic Engineer*, 41, p. 273, June 1971.

15. Keats, A. B. and Leggett, D. W., 'A transport delay simulator using digital techniques', *The Radio and Electronic Engineer*, 41, No. 4, pp. 179–84, April 1972.

9. Appendix 1 : Padé Approximation

The familiar Maclaurin Series for the required transfer function is

$$e^{-s\tau} = \lim_{n \to \infty} \sum_{k=0}^{n} (-s\tau)^k / k!$$

but its convergence is significantly slower than the Padé series:[1, 2]

$$e^{-s\tau} = \lim_{u+v \to \infty} F(u, v, -s\tau)/G(u, v, -s\tau) \quad ......(10)$$

where

$$\left.\begin{array}{l} F(u, v, -s\tau) = 1 + \sum_{n=1}^{\infty} \left[\dfrac{\prod_{k=0}^{n-1}(v-k)}{\prod_{k=0}^{n-1}(u+v-k)} \right] \dfrac{(-s\tau)^n}{n!} \\[2em] G(u, v, -s\tau) = 1 + \sum_{n=1}^{\infty} (-1)^n \left[\dfrac{\prod_{k=0}^{n-1}(u-k)}{\prod_{k=0}^{n-1}(u+v-k)} \right] \dfrac{(-s\tau)^n}{n!} \end{array}\right\}(11)$$

A transfer function for a so-called Nth-order Padé approximation is obtained by terminating the above series at $n = N$ and setting

$$u = v = N$$

In this manner, the following 1st, 2nd and 3rd order Padé approximations are derived

$$P_1(s) = \frac{1-(s\tau/2)}{1+(s\tau/2)}$$

$$P_2(s) = \frac{1-\frac{1}{2}(s\tau)+\frac{1}{12}(s\tau)^2}{1+\frac{1}{2}(s\tau)+\frac{1}{12}(s\tau)^2}$$

$$P_3(s) = \frac{1-\frac{1}{2}(s\tau)+\frac{1}{10}(s\tau)^2-\frac{1}{120}(s\tau)^3}{1+\frac{1}{2}(s\tau)+\frac{1}{10}(s\tau)^2+\frac{1}{120}(s\tau)^3}$$

Replacing s by $j\omega$ in (10) and (11) immediately establishes that

$$F(u, v, -j\omega\tau) = \overline{G(u, v, -j\omega\tau)}$$

so that

$$|P_N(j\omega)| = 1\cdot 0$$

Thus, the gain-frequency characteristic of a Nth order Padé approximation is unity.

10. Appendix 2: A Single's Approximation

A substitution into the equation (1) for a fourth-order approximation[4] yields

$$S_4(s) = \frac{1}{1 + s\tau_0} \left[\frac{(\tau_1 s)^2 - 2\zeta_1(\tau_1 s) + 1}{(\tau_1 s)^2 + 2\zeta_1(\tau_1 s) + 1} \right] \times$$
$$\times \left[\frac{(\tau_2 s)^2 - 2\zeta_2(\tau_2 s) + 1}{(\tau_2 s)^2 + 2\zeta_2(\tau_2 s) + 1} \right] \quad (12)$$

with

$$\zeta_1 = 0\cdot4, \qquad \zeta_2 = \sqrt{1\cdot5}, \qquad \tau_2/\tau_1 = 1\cdot68 \quad ...(13)$$

For this network, the gain-frequency characteristic is

$$|S_4(j\omega)| = \frac{1}{\sqrt{1 + (\omega\tau_0)^2}}$$

and at low frequencies the transport delay generated is

$$\tau = \tau_1(9\cdot83 + \tau_0/\tau_1)$$

11. Appendix 3: Reconstruction Errors

Using the nomenclature defined in Fig. 7, it follows that $e(t) = z(t) - x(t)$.

If the input process is assumed to be wide-sense stationary,[11] then the spectral density function of the reconstruction unit's steady-state error is

$$\Phi_{ee}(\omega) = \Phi_{xx}(\omega) - \Phi_{xz}(\omega) - \Phi_{xz}(-\omega) + |H(j\omega)|^2\Phi_{yy}(\omega) \quad ......(14)$$

As shown in reference 12, the spectral density functions for the p.a.m. signals are given by:

$$\left. \begin{array}{l} \Phi_{yy}(\omega) = \sum\limits_{-\infty}^{\infty} |R_m|^2 \Phi_{xx}(\omega - m\omega_s) \\ \Phi_{xz}(\omega) = R_0 H(j\omega)\Phi_{xx}(\omega) \end{array} \right\} \quad ......(15)$$

where

$$\left. \begin{array}{l} R_m = \frac{\gamma\omega_s}{2\pi} \left[\frac{\sin(m\omega_s\gamma/2)}{m\omega_s\gamma/2} \right] \\ \simeq \frac{\gamma\omega_s}{2\pi} \quad \text{for } m\omega_s\gamma/2 \ll 1 \end{array} \right\} \quad ......(16)$$

Substituting equations (15) and (16) into equation (14) yields

$$\Phi_{ee}(\omega) = \Phi_{xx}(\omega)|1 - R_0 H(j\omega)|^2 +$$
$$+ |H(j\omega)|^2 \sum\limits_{\substack{m=-\infty \\ \neq 0}}^{\infty} |R_m|^2 \Phi_{xx}(\omega - m\omega_s)$$

and using the result[11]

$$\overline{e^2} = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \Phi_{ee}(\omega) \, d\omega$$

the mean square value of the reconstruction unit's steady-state error is obtained as:

$$\overline{e^2} = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \Phi_{xx}(\omega)|1 - R_0 H(j\omega)|^2 \, d\omega +$$
$$+ \frac{1}{2\pi} \sum\limits_{\substack{m=-\infty \\ \neq 0}}^{\infty} |R_m|^2 \int\limits_{-\infty}^{\infty} |H(j\omega)|^2 \Phi_{xx}(\omega - m\omega_s) \, d\omega \quad (17)$$

A useful physical interpretation of equation (17) is as follows. The first term represents the low-frequency distortion produced by virtue of the non-uniformity of the reconstruction filter's gain characteristic over the base-band of the input signal $x(t)$. The second term represents the high-frequency noise produced by the incomplete attenuation of the side-bands generated by the sampling process.

A measure of the effectiveness of the linear processing units can be obtained by evaluating equation (17) for the unity power, sinusoidal input:

$$x(t) = \sqrt{2} \sin(\omega_0 t + \phi) \qquad ......(18)$$

where ϕ is uniformly distributed between 0 and 2π. It is readily shown that the power spectrum corresponding to equation (18) is

$$\Phi_{xx}(\omega) = \pi[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]$$

and substituting this result into equation (17) yields:

$$\overline{e^2} = |1 - R_0 H(j\omega_0)|^2 + \sum\limits_{\substack{m=-\infty \\ \neq 0}}^{\infty} |R_m|^2 \times$$
$$\times [|H(jm\omega_s - j\omega_0)|^2 + |H(jm\omega_s + j\omega_0)|^2] \quad (19)$$

The frequency response function for a clamp and linear interpolator are readily derived as

$$H_{\text{clamp}} = [1 - e^{-j\omega/f_s}]/j\omega\gamma$$
$$H_{\substack{\text{linear} \\ \text{interpolator}}} = 4 \sin(\omega/2f_s)/(\omega^2\gamma/f_s)$$

Substituting the above equations into equation (19) yields the required error quantities

$$\overline{e^2}\Big|_{\text{clamp}} = \frac{(2\pi f_0/f_s - \sin 2\pi f_0/f_s)^2 + (1 - \cos 2\pi f_0/f_s)^2}{(2\pi f_0/f_s)^2} +$$
$$+ \sin^2(\pi f_0/f_s) \sum\limits_{m=1}^{\infty} \left[\frac{1}{\pi^2(m + f_0/f_s)^2} + \right.$$
$$\left. + \frac{1}{\pi^2(m - f_0/f_s)^2} \right] \quad (20)$$

$$\overline{e^2}\Big|_{\substack{\text{linear} \\ \text{interpolator}}} = \frac{(\pi f_0/f_s + \sin \pi f_0/f_s)^2(\pi f_0/f_s - \sin \pi f_0/f_s)^2}{(\pi f_0/f_s)^4} +$$
$$+ \sin^4(\pi f_0/f_s) \sum\limits_{m=1}^{\infty} \left[\frac{1}{\pi^4(m + f_0/f_s)^4} + \right.$$
$$\left. + \frac{1}{\pi^4(m - f_0/f_s)^4} \right] \quad (21)$$

A Transport Delay Simulator using Digital Techniques

A. B. KEATS, C.Eng., M.I.E.E., M.I.E.R.E. *

and

D. W. LEGGETT, M.Sc. *

SUMMARY

This transport delay simualtor was designed to extend the facilities of an analogue computer when used, in particular, for plant simulation purposes. An abridged performance specification is given in the Appendix which shows that this delay simulator has compatible performance with a wide range of modern analogue simulator equipment.

Its principle of operation, chosen after consideration of several alternatives, is based on sampling and digital storage. The store comprises m.o.s. integrated-circuit shift registers. Linear interpolation is used to minimize the sampling rate and storage capacity required to achieve the specified accuracy.

An alternative mode of operation is as a function generator, the ordinates of the given function being initially written into the store from punched paper tape or an analogue signal source.

*Atomic Energy Establishment, U.K.A.E.A., Winfrith. Dorchester, Dorset.

1. Introduction

The transport delay simulator (t.d.s.) was developed at the Atomic Energy Establishment, Winfrith, to provide greater accuracy and flexibility in plant simulation studies on the analogue computer. It replaces a rotating capacitor drum store type of simulator. After consideration of a number of methods including Padé approximations, 'endless loop' magnetic tape storage, curve writer/followers, analogue sample store, etc., a digital sampled data storage system was chosen as being the most suitable.[1] A block diagram is shown in Fig. 1.

The analogue input is sampled by an analogue-to-digital converter (a.d.c.) and the sampled value written into the first stage (or location) of a digital store. Each time a new sample is taken, the earlier values are advanced to the next store location. Having passed through the store, the samples are presented sequentially to a digital-to-analogue converter (d.a.c.) which reconstructs the original analogue signal waveform.

The accuracy with which an input function is reproduced at the output of the delay simulator is proportional to the number of samples per cycle of the highest frequency contained in that signal. Therefore, for a specific accuracy, the number of words of storage required to yield a given delay is determined by the bandwidth and the delay time. The sampling rate for a given accuracy and consequently the digital storage capacity are drastically reduced by employing 'linear interpolation' which 'smooths' the output waveform between successive samples.[1] The resolution (word length) of the a.d.c. and d.a.c. are dictated independently of sampling rate by the required accuracy of reproduction of a d.c. signal.

Transport delay in most physical systems is influenced by changes in other parameters. This requirement is met in the t.d.s. by a pulse-rate-modulator through which the sampling rate and hence time-delay, is under the control of an analogue 'delay control' input voltage having a dynamic range of 100 : 1.

The magnitude of the time delay T_D achieved by the simulator is proportional to the number of samples stored (N) and inversely proportional to the sampling rate (f_s), i.e.

$$T_D = \frac{N}{f_s}$$

The equipment can alternatively be used as a function generator by loading ordinates of a given amplitude/time function into the store and then recirculating them at the required rate. During recirculation, the ordinates are presented sequentially to the d.a.c. (and linear interpolator) to reproduce the stored function in analogue form. The original function may be written into the store either from an analogue signal source or from a punched paper tape.

2. Principal Components of the System

The simulator can be divided conveniently into five basic blocks which interact through the control logic (Fig. 1). The analogue-to-digital and digital-to-analogue converters are self-contained commercially-available
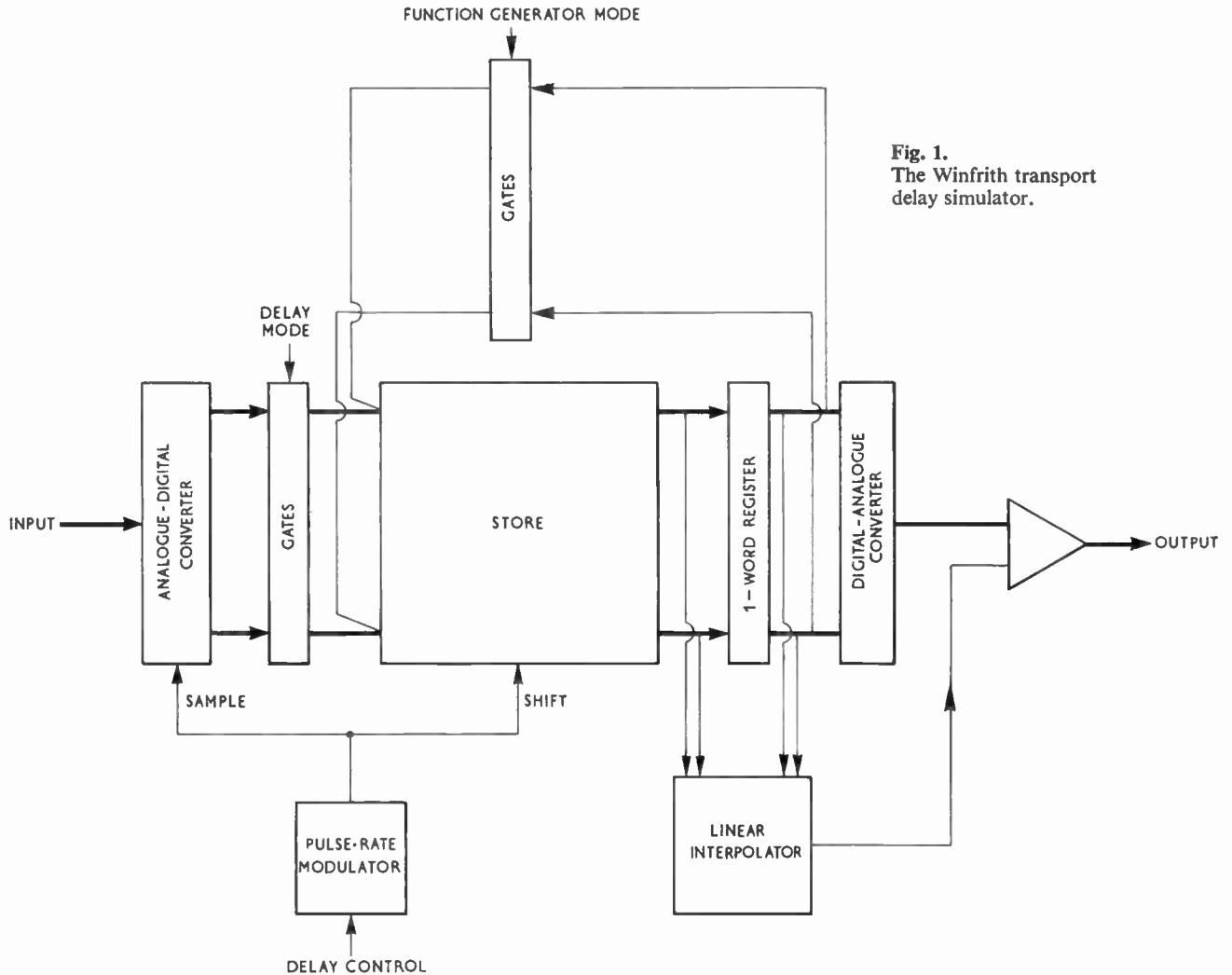
FUNCTION GENERATOR MODE

Fig. 1.
The Winfrith transport
delay simulator.

modules and will therefore be described only briefly. The store, linear interpolator and pulse-rate-modulator are specially built, mainly from standard integrated-circuit elements, and will be considered in more detail.

2.1. Analogue-to-Digital Converter

The a.d.c. is of the successive approximation type and converts an analogue input voltage within the range $+10$ to -10 V into 12-bit 'offset binary' form in 80 µs. The a.d.c. module is controlled by the external logic via its 'reset' input. When 'reset' is applied, the previous reading is cleared from the output buffer register. When 'reset' is removed the next digitization commences. A 'complete' signal from the a.d.c. indicates to the external logic that the conversion process has finished and the new measurement is available in the output buffer register. All the digital signals are at standard TTL levels.

The 12-bit (11 bits + sign) resolution chosen for the a.d.c. is based on the specified overall error tolerance for the simulator of $\pm 0.1\%$. The maximum error introduced by the a.d.c. is $\pm$ half the least-significant-bit, i.e. $\pm 0.025\%$.

2.2. Digital Store

The sequential digital store comprises m.o.s. integrated

circuit static shift registers assembled into 400-word plug-in modules. The word length is 13 bits which includes the 12-bit word length of the a.d.c. and a 'marker-bit' which is used to locate data in the store; the store could be regarded as 13 parallel shift registers each of 400 bits length.

This particular size of store module was based on the currently available m.o.s. shift register integrated circuits. The rapid development of large-scale integration (l.s.i.) will enable an increasingly larger size of store module to be practicable.

The digitized information is shifted through the store by a single phase clock, the information being written in when the clock is in a low state and shifted, and subsequently read out, on the clock pulse transition from a low to high state.

Each t.d.s. has capacity for five 400-word modules, i.e. 2000 words of storage.

2.3. Digital-to-Analogue Converter

The d.a.c. module comprises a 12-bit digital input buffer followed by a digital-to-analogue decoding resistor network. The digital input (at standard TTL levels) is strobed into the buffer register under the

control of the external logic. The analogue output is a voltage in the range $+10$ to -10 V.

The 12-bit resolution is again based on the overall error tolerance ($\pm 0.1\%$) of the simulator. The maximum d.a.c. error accounts for a further 0.025% (i.e. $\pm\frac{1}{2}$ least-significant-bit).

2.4. Linear Interpolation

Figures 2(a) and 2(b) depict a section of a typical signal at the t.d.s. output showing how linear interpolation 'smooths' the curve between consecutive samples to give a more accurate representation of the function
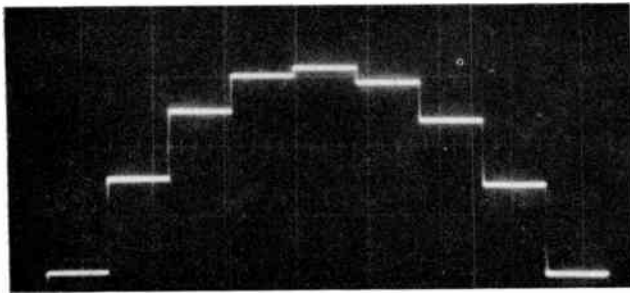
(a) Without interpolator.
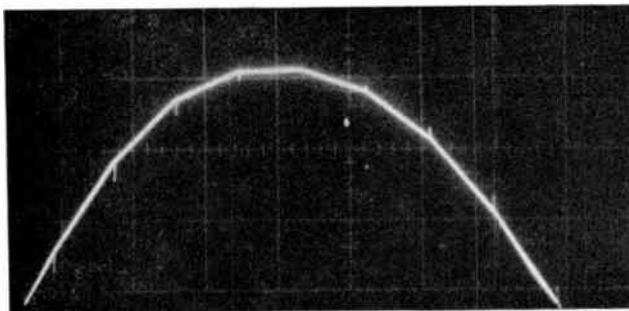
(b) With interpolator.

Fig. 2. Output of transport delay simulator.

(as discussed in reference 1). Linear interpolation is achieved (Fig. 3) by adding a linear ramp to each ordinate y_n, the amplitude of the ramp being the difference between the two consecutive ordinates y_n and y_{n+1}. The period of the ramp is equal to the sampling period T so that the sum of y_n and the ramp just reaches y_{n+1} in the sampling interval. The difference ($y_{n+1} - y_n$), obtained by digital subtraction, is applied to the digital input of a 'hybrid multiplier' (or multiplying d.a.c.). The corresponding analogue input is a ramp of unit (10 V) amplitude and period T. The output of the hybrid multiplier (Fig. 4) is the product of the digital and analogue inputs, i.e. a linear ramp of amplitude ($y_{n+1} - y_n$) and period T. Access to the two adjacent samples y_n and y_{n+1} is provided by the single-word storage register between the main store output and the d.a.c. input. An analogue summing amplifier is used to combine the outputs of the hybrid multiplier and the d.a.c.
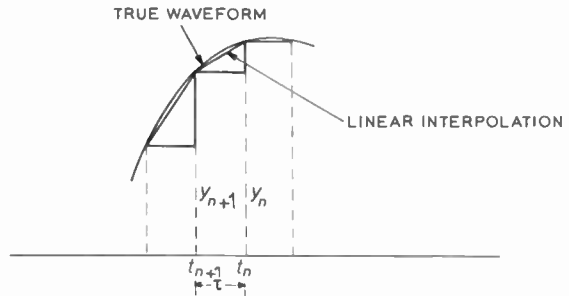

Fig. 3. Reconstruction of waveform from sampled data.

The circuitry required to perform linear interpolation is depicted in Fig. 5. The digital difference between the last and penultimate samples is obtained from the full subtractor. If the output from the subtractor is a positive quantity, this is fed directly to the digital input of the hybrid multiplier; if the output is negative, then 2's complement is performed on the difference quantity to convert it into an equivalent positive number.

The output amplifier in the interpolator is a switched invert/non-invert amplifier which gives the difference signal its correct sign. The dual purpose amplifier is conditioned by the 'off' and 'on' state of the f.e.t. TR1.

The maximum change likely to occur between two consecutive samples in the output signal is ± 1 V and since the interpolator operates only on the difference, the 'difference' word length is restricted to the 8 least significant bits of the full word length.

2.5. Pulse-rate-modulator

The magnitude of the time delay introduced by the simulator depends upon the number of words (samples) stored and on the sampling rate. The sampling rate (and consequently the rate at which sampled values are shifted through the store) is controlled by a pulse-rate-modulator (p.r.m.) from the 'delay control' input. An additional requirement of the pulse-rate-modulator is to produce the sawtooth (ramp) input required by the hybrid multiplier for linear interpolation. This requirement led to the choice of a linear integrator as the basis of the pulse-rate-modulator circuit (Fig. 6). The circuit integrates alternately in opposite directions to avoid the problem of fast resetting which would be required in a unidirectional integrator.
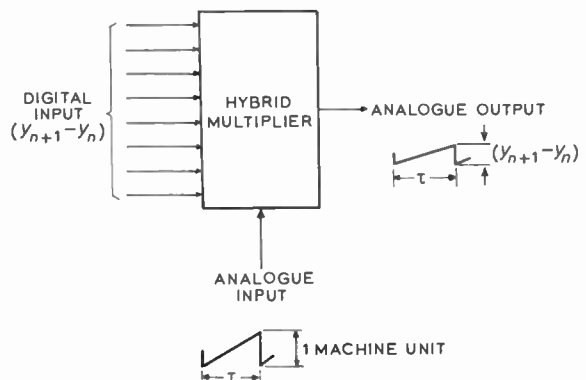

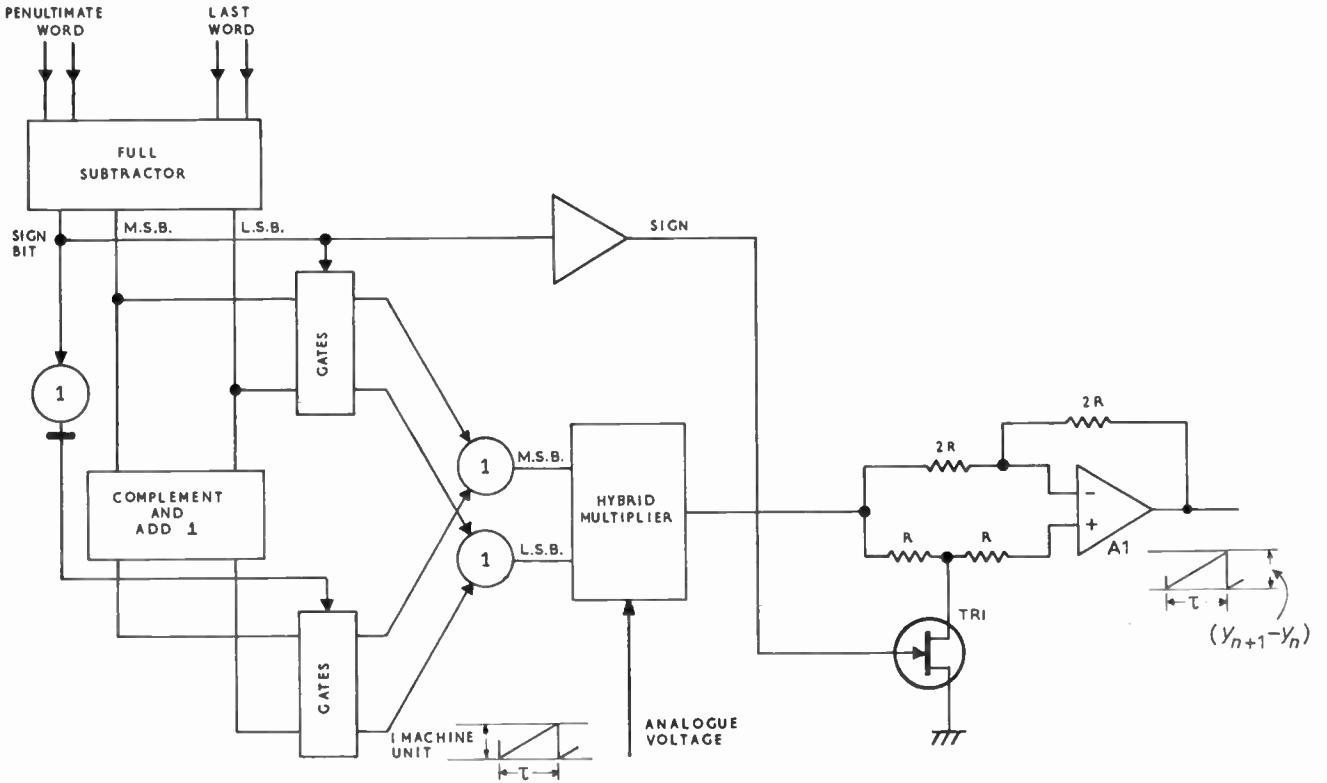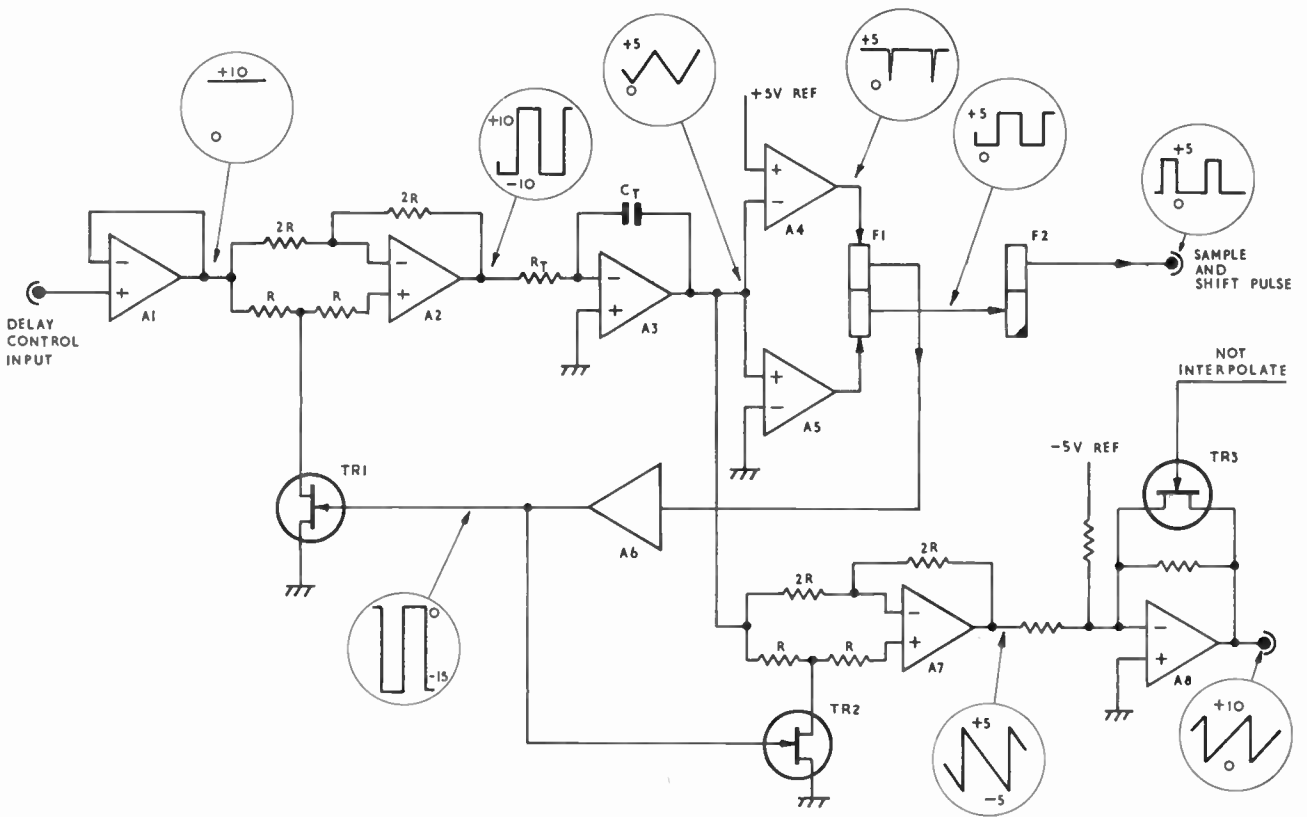Fig. 4. Hybrid multiplier.

Fig. 5. Linear interpolator.

Fig. 6. Pulse rate modulator.

The delay control voltage is applied to the input of the integrator via a high impedance buffer stage A1 and a switched invert/non-invert amplifier A2. The integrator output voltage changes linearly, at a rate proportional to the delay control voltage, until it reaches the threshold of one of the comparators A4 or A5. The resultant output of the appropriate comparator reverses the state of bistable F1 which in turn reverses the condition of the invert /non-invert amplifier A2 via f.e.t. switch TR1. The polarity of the input voltage to the integrator is therefore reversed and the output commences to change linearly in the opposite direction until the opposite threshold is reached. The integrator output is thus a symmetrical triangular waveform. The sawtooth waveform required by the hybrid multiplier is derived from the output of the integrator by passing it through a second invert/non-invert amplifier A7 controlled by bistable F1 via f.e.t. switch TR2. The sense of this control is such that A7 inverts during the positive-going excursion of the integrator and is non-inverting during the negative-going excursion. The result is that during one complete cycle of the integrator, the output of A7 changes from $+5$ V to -5 V, then reverts rapidly to $+5$ V at the end of the cycle. Amplifier A8 is an inverter and level shifter. It is also switched to zero gain by the f.e.t. switch TR3 when linear interpolation is not required for certain test purposes. The main timing pulse for control of the a.d.c. and store is generated by monostable F2 once per cycle of the bistable F1.

3. The Control Logic

The functions of the delay simulator are supervised mainly by the 'mode' controls of the associated computer. In addition to these however, there are manual controls on the simulator itself which predetermine the functions of the mode controls. The alternative use of the simulator as a 'function generator' accounts for most of these controls but there is a four-position range switch for sampling rate and an 'interpolate' switch which are common to both modes of operation. Other controls, used only in the function generator mode, select the method of loading a function (automatically from punched paper tape or manually from a variable d.c. voltage source) and methods of transmitting the function after loading (single shot or repetitive). Starting and stopping the function generator after loading is under the control of the computer mode switches.

A large proportion of the control logic is associated with setting up, detecting and recirculating the 'marker bit'. This is used to locate data in the store, particularly in the function generator mode to indicate completion of loading.

4. Error Analysis

The errors associated with the simulation of transport delay may be considered in two parts.

Firstly, there is the amplitude error between corresponding points on the input and output waveforms (amplitude distortion). This error may be further divided into contributions due to the finite sampling rate and the finite resolution (12 bits) of the a.d.c. and d.a.c. (The error due to drift and offset in the analogue output circuits can be neglected.) If it is assumed that these errors are independent,[2] their mean-square values, when summed, yield a total r.m.s. amplitude error of 0.039%.

For practical purposes the errors will be less than $3 \times$ r.m.s. error, i.e. 0.12%. This is a pessimistic calculation and all measurements made have shown that the error is less than 0.1%.

The second form of error is in the delay time due to errors in the frequency of the pulse-rate modulator and to the limited accuracy to which the delay control voltage can be defined by the computer. For long delay times, when the delay control voltage input to the p.r.m. is a minimum (0.01 machine unit), the timing accuracy is limited by the $\pm 1\%$ absolute tolerance on the delay control voltage. For short delays, where the delay control voltage lies within a tolerance of $\pm 0.01\%$, the timing accuracy is limited by the tolerance on the frequency of the p.r.m. itself to $\pm 0.1\%$. This limit has been confirmed by measurement over the full range of conditions covered by the specification. (See Appendix.)

5. Operational Experience

A prototype version of the transport delay simulator has been in service on the analogue computer installation at Winfrith since February 1971. It has yielded the expected improvement of at least two orders of magnitude in accuracy over the existing rotating capacitor drum delay simulator. The wider range of operation, coupled with the improved reliability inherent in solid-state systems, has greatly increased the computer capability to solve a wide range of problems in the plant simulation field.

An example of its use is shown in Fig. 7. Figure 7(a) is a simplified representation of a typical section of plant comprising a heat source, heat exchanger, circulating pump and interconnecting pipework. The transport delay simulator is used here to represent delay in the length of pipework between the heat source and the heat exchanger. Its value is proportional to (amongst other things) the reciprocal of pump speed. Pump speed is therefore fed into the delay simulator as a variable which controls the delay, i.e. the 'delay control' voltage. A temperature disturbance in the circulating fluid is represented by the waveforms in Fig. 7(b). Whilst the disturbance is passing between points A and B (i.e. through the delay simulator) the pump speed is dropped. This results in an increase in transport delay which in turn causes a lengthening of the disturbance as observed at B.

Two new developments, which will be incorporated into future models, have been evolved from user trials of the prototype. The first is the provision of a more direct method of loading functions from a punched-paper tape reader. Originally, this facility was implemented through some existing equipment which assembled 5-bit tape characters into digital words and performed digital-to-analogue conversion. The ordinates of the function to be loaded were then fed into the delay simulator in analogue form. The intention now is to feed the tape characters directly into the simulator,
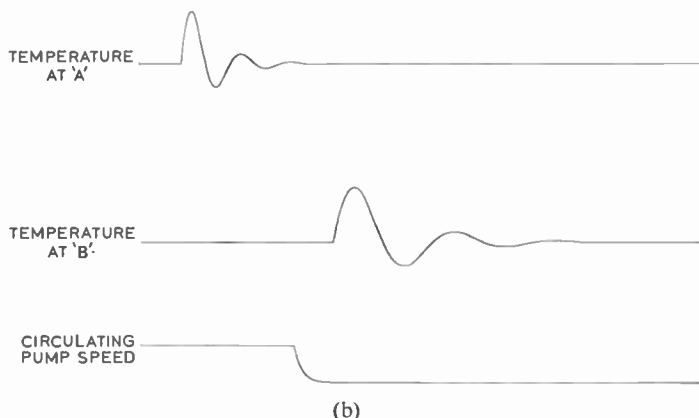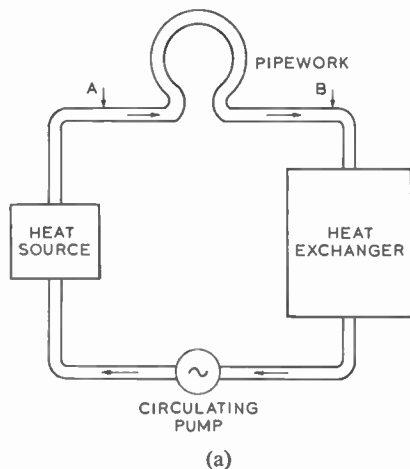
Fig. 7. (a) Typical plant model involving 'transport delay'.
(b) The effect of a variation in transport delay on a temperature transient.

assemble them into 12-bit words and then load them directly into the digital store.

The second development is a new facility to enable the results of one computation to be stored and then used as the input function for a further computation. This change from 'store' to 'function generator' mode is required to be controlled from the computer mode controls.

6. Conclusions

This transport delay simulator was designed to improve upon the performance of earlier simulation methods in respect of resolution, range, stability and reliability. The improvement is made possible by the availability, at low cost, of complex integrated electronic circuits whose high reliability has been well established in other fields. The adoption of a sampling technique coupled with digital storage of the sampled data enables the resolution to be precisely defined and obviates most of the sources of drift and offset usually associated with analogue techniques. The use of digital electronic principles relaxes the bandwidth constraints inherent in electromechanical storage and lumped parameter networks, thus permitting a much wider range of bandwidth and delay times to be achieved. During the first six months of operation there has been only one failure in the 52 m.o.s. integrated circuits used and none in the 200 bipolar circuits. This inspires confidence in the reliability of the equipment and is consistent with the high reliability record of integrated circuits generally.

7. Acknowledgments

The authors take pleasure in acknowledging the contribution made by Dr J. B. Knowles on the theoretical analysis of the benefits to be gained by using linear interpolation; also the co-operation given by Messrs. F. C. Kingdon and J. R. Taylor of A.E.E. Winfrith in commissioning the prototype and pursuing further developments.

8. References

1. Knowles, J. B. and Leggett, D. W., 'An analysis of transport delay simulation methods', *The Radio and Electronic Engineer*, 41, No. 4, pp. 172–8, April 1972.

2. Knowles, J. B. and Edwards, R., 'Effects of finite word length computer in a sampled-data feedback system', *Proc. Instn Elect. Engrs*, 112, p. 1192, 1965.

9. Appendix: Abridged Performance Specification for the Transport Delay Simulator

Analogue input and output signal voltage range: ± 10 V ($= \pm 1.0$ machine unit).

Delay control input voltage range: $+0.1$ to $+10.0$ V ($= 0.01$ to 1.0 machine units).

Delay ranges:

	Range 1	Range 2	Range 3	Range 4
Delay control input (machine units)	1.0 to 0.01	1.0 to 0.01	1.0 to 0.01	1.0 to 0.01
Delay time (ms per word of storage)	0.1 to 10	1 to 100	10 to 1000	100 to 10 000
Sampling rate (Hz)	10k to 100	1k to 10	100 to 1	10 to 0.1
Signal bandwidth corresponding to sampling rate above (Hz)	100 to 1	10 to 0.1	1.0 to 0.01	0.1 to 0.001

Waveform accuracy
   (i.e. instantaneous difference between output and input amplitudes): $\pm 0.1\%$.

Delay time accuracy:
   $\pm 0.1\%$ at minimum end of each range. (Determined by the accuracy of the pulse-rate modulator.)
   $\pm 1\%$ at maximum end of each range. (Determined by the accuracy with which the associated equipment can define the required delay control voltage, i.e. 0.01 machine unit. A typical computer has an accuracy of 0.01% machine unit, therefore an output of 0.01 machine unit can be guaranteed to an accuracy of only 1%.)

Store capacity:
   2000 words in 400-word blocks.

Total delay time:

$$T_D = N/f_s \text{ seconds}$$

where N = number of words of storage,

f_s = sampling frequency (Hz).

Fault Investigation of some Silicon Integrated Circuits

C. H. JONES,
B.Sc.(Hons.), C.Eng., M.I.E.E.*

SUMMARY

This paper reviews work in the field of design and fault investigation of special integrated circuits intended for use in military radio equipment.
The circuits considered were designed, developed and master layouts prepared by the author and his colleagues. Fabrication was carried out by industry. The fault investigation was mainly centred around the spread of design parameters and the faults associated with device fabrication. Initial investigation was by means of 'black box' measurements; more detailed measurements were carried out within the circuit chip using optical and electron microscope and mechanical probes.

* Signals Research and Development Establishment, Ministry of Defence, Christchurch, Hants.

1. Introduction

The advent of microcircuitry, and in particular the silicon integrated circuit (s.i.c.), in military equipment has led to problems that were not present in previous methods of equipment construction. Many of these are associated with the implementation of this new technology, but equally important are the problems resulting from the high standard of quality and reliability required in this field of application.

It is well known that the s.i.c. can only be produced cheaply when large numbers of units are involved to a limited number of standard designs. The levels of quality and testing are thus dictated almost entirely by their commercial application for fields such as computers, automation, etc.

In the military field the same cost-effective criteria cannot be rigidly applied. This market is relatively small and uncertain, while the quality and reliability are required to be of a higher order than for commercial applications. It is also evident that many required circuit functions cannot be engineered using the standard designs available. This has led to the special integrated circuit designed for a specific application.

The s.i.c. requirements for the latest generation of military communication equipment have been satisfied by a mixture of standard and special devices, all procured through approved specifications to ensure that the appropriate quality and reliability will be maintained. The standard devices are obtained from more than one approved manufacturing source, while the special circuits are vested in single manufacturers.

As the Qualification Approval Authority for much of this equipment, it was considered essential that the Signals Research and Development Establishment must have up-to-date and detailed knowledge on all aspects of microcircuitry from basic semiconductor physics and fabrication methods to modes and results of failures. The work of the author and his colleagues has been to create a nucleus of such information and skills. The following is a review of some of these activities.

To date S.R.D.E. has designed and fully evaluated six special integrated circuits. These cover the fields of radio- and video-frequency analogue circuits as well as low- and medium-speed logic circuits. The devices have all been fabricated by industry using standard bipolar technology. The circuits were designed for possible use in army radio equipment and this implied reliable operation over extremes of environment and a low failure rate in use. Several of these circuits have been accepted and will go into production in due course.

The design of a limited number of special circuits was undertaken mainly to give suitable vehicles for the 'in house' work. A satisfactory design gave indications of the order of yield that could be expected at first hand. The resulting satisfactory devices were then available for use in prototype equipment. More important in the context of this paper, there was a ready supply of typical faulty devices, whose design history was known, for use in the failure investigation exercise.

The number of faulty devices that were available or that could be examined in a reasonable time was obviously limited, hence no reliable statistical deductions could be made, but from the data that have been derived to date, sensitive areas in circuit design, diffusion, and encapsulation have been indicated. This information has been fed back to the industry and has resulted in improved yields of satisfactory devices.

2. The Integrated Circuit Laboratory

The S.R.D.E. laboratory for integrated circuit work has evolved to obtain knowledge of aspects of micro-circuitry, especially in the field of special function devices. Several pieces of complex equipment have been obtained and in some cases modified and adapted to suit their role. This was especially the case with mechanical probe equipment, most of which was expensive and too complex for the requirements visualized in this laboratory.

For circuit design work, standard electronic laboratory test equipment was obtained. In the construction of prototype circuits prior to their conversion into monolithic form, the components were chosen to simulate their integrated counterpart as closely as possible. In the case of active components these were obtained from industry as encapsulated multiple transistors made to the same geometry and diffusion schedules as those in the final circuit. The resistors used were 20% Grade 2 carbon to simulate the wide tolerance and temperature coefficient experienced in the diffused component. Experience had shown that the diffused capacitor was a difficult component to simulate due to its wide tolerance, low breakdown voltage and its capacitance/voltage characteristic. In practice a foil or polycarbonate type was used and the effects of voltage extremes theoretically assessed. Wherever appropriate a theoretical analysis of the design was made and a computer simulation created to give correlation between theory and practice.

A satisfactory design was converted into the monolithic form in a series of steps. Scale layouts with inter-connexion variations and test patterns were prepared within the laboratory, the diffusion and encapsulation were performed under contract with industry. The testing and evaluation of the first diffusion samples were undertaken again in the laboratory.

3. Circuit Design and Fabrication

The conversion of a schematic circuit diagram into a scale layout has proved to be very time consuming and requires an intimate knowledge of the circuit in question and the 'layout rules', as supplied by the industry. Our experience has shown that the person best fitted to do this work was the circuit designer since the drawing skills required were not excessive. The drawings were prepared on temperature and humidity stable plastic sheet divided into a millimetre grid, this being the finest resolution required. The grid absolute accuracy was such that it did not deviate by more than ± 0.5 mm in 70 cm. The drawing scale was 250 times final size, hence each millimetre drawn represented 4 μm on the final device. The final 'clean'

drawings were ink-lined using a 0·1 mm tracing pen and the various diffusion areas hatched or coloured to a standard code.

On acceptance of a scale layout as fulfilling all the related design rules, a first diffusion run was made and a small number of visually inspected circuit chips were encapsulated and sent to the designer for evaluation. These were intended to check for any latent layout and design errors; especially as in the latter case some of the effects of converting from discrete to integrated components cannot be estimated or anticipated. Subject to satisfactory results at this stage, further diffusions were ordered and large numbers of devices were tested against a tentative specification provided by the designer. The results of this exercise enabled the specification to be finalized, the production yield estimated and a price fixed for production. Throughout these diffusion exercises a close watch was kept by the designer on the test results, any fault trends or abnormalities being investigated by practical examination as detailed later. The results of these examinations were fed back to industry where action was taken as appropriate. The objective was to have consistently well-understood and reliable devices that could be reproduced with high yields.

4. Fault Investigation

Fault investigation in monolithic microcircuits required some equipment and skills that were not normally found in the electronic laboratory. The faults usually handled came under the headings of diffusion errors, bonding and encapsulation, design and tolerancing margins and malfunctions due to introduction of 'phantom' components inherent in the monolithic process.

Initial comparison measurements to the design specification were made treating the device as a 'black box'. Other more complex external measurements were made as required until some theory could be proposed for the reason for malfunction. The next stage consisted of opening the encapsulation. For plastic devices, the leads were soldered to a suitable base to maintain their spacing and the plastic removed by means of a solvent. The hermetically-sealed welded packages were opened by grinding away one face of the package using fine emery paper on a flat surface. When it was almost punctured, the remainder was peeled away using a sharp knife. The packages sealed with soft solder were opened by application of a hot soldering iron of a large thermal capacity for a short time to the lid, whereupon the expanding gas within the package forced the lid off. In all cases the previous electrical tests were repeated to ensure that no thermal or mechanical damage had been done. With open devices every precaution was taken to ensure no further damage or contamination occurred. To this end the devices were handled under clean conditions in a laminar flow cabinet and stored in covered dishes.

On satisfactorily opening a device—this could not always be guaranteed—a visual examination was made at a magnification of up to 500 times. Inspection was made of the interconnect pattern for incomplete etching

or breaks, especially in the areas indicated by the fault mode. The bonds and bond leads were also inspected for presealing damage and poor welds. At this stage any suspect areas were photographed on colour Polaroid film. The colour record was found to be a great help in distinguishing the various diffusion areas. The colour variations were due to the build-up of oxide after each diffusion step, the thickness being such as to cause interference colours in the reflected light. Closer examination of these coloured areas in conjunction with the original layout drawings identified the different diffusions. Breaks and discontinuities in these areas could easily be seen, and this was especially the case with isolation diffusions.
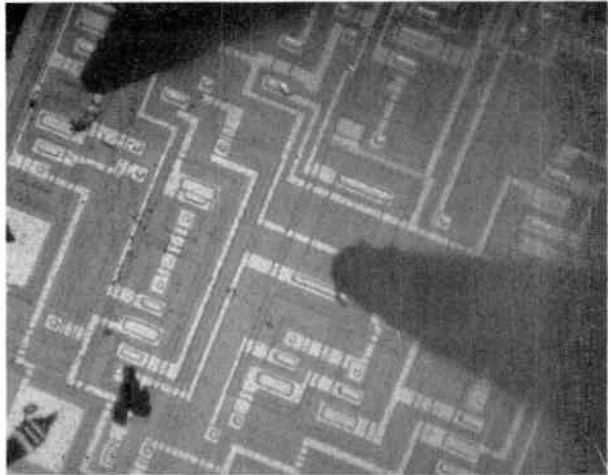

Fig. 2. Probing of chips.

screen profile projector, (Fig. 1) and the chip surface examined at a magnification of 100 times. Normal electrical supplies were then applied to the device. The potentials, waveforms etc., were then monitored at various parts of the circuit by means of probes. The contact area and resolution of these were such that any metallized area within the circuit could have a connexion made to it. This required a probe point of less than 8 μm diameter and combined resolution and backlash of less than 4 μm (Fig. 2). The satisfactory use of a mechanical probe to these accuracies required considerable practice if a satisfactory contact was to be made with minimum damage. In the case of passivated devices, i.e. chips that were protected with an overall oxide layer, connexion within the circuit was made by increasing the downward and tangential probe pressure until the oxide was punctured.

To enable individual suspect components to be isolated from the shunting effect of other components, the aluminium interconnect pattern was severed where appropriate. One method used a probe to scribe across the tracks. This was not always successful, especially when the track passed over oxide steps, since there was also the risk of damaging the oxide layer below. A second more satisfactory method was to vaporize the track over a short length. Providing the track was narrow and well clear of other tracks there was little risk of any molten aluminium causing bridging in the vicinity. The break was also easily seen as the heat caused a local discolouration of the chip surface (Fig. 3). To break the interconnexion, two probes were put down as close together as possible on the conductor path without touching each other. A current-limited power pack was then discharged between the probes. A supply of about 6 V at 100 mA was found to be satisfactory. Too low a power caused the aluminium to incandesce rather than vaporize, giving prolonged heating and hence damage to the circuit diffused areas below. By these means the faulty component or area was easily traced; the characteristics of individual diodes, transistors and resistors were displayed on a standard transistor curve tracer.
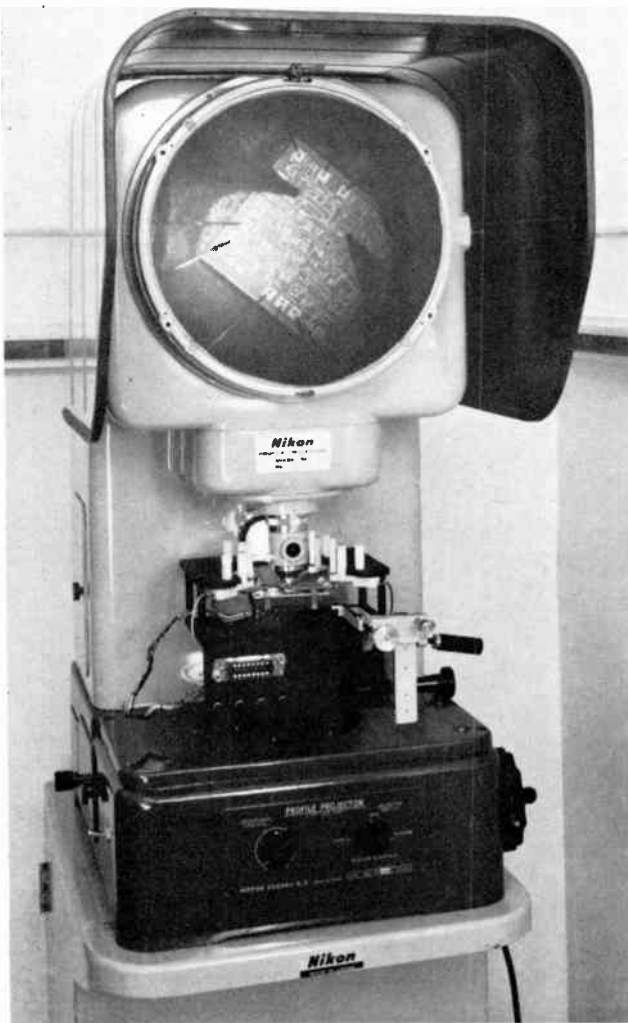

Fig. 1. Profile projector showing probe attachment.

Where faults could not be related to visual imperfections, electrical measurements were made within the silicon chip area. The opened device was first mounted on a special printed circuit board carrier to ease handling and so that electrical connexions could be made to the package leads. The carrier was then fitted into a large-

Fig. 3. Component isolation by aluminium vaporization.

The light emitting property of a reversed bias p–n junction was also used in the case of devices whose malfunctions were voltage dependent. When this biased device was examined at 100 times magnification in the dark, areas of breakdown were indicated by spots or lines of light. In some cases the supply voltage was elevated to increase the light intensity and help break-down identification. Breakdowns of this sort were usually attributed to incorrect impurity concentration or oxide pin-holes which allowed 'rogue' diffusion areas to be scattered throughout the chip.

The visual inspection of devices using optical methods had a major disadvantage in that the depth of focus was very short at high magnification. This made examination of bonds and surface foreign matter difficult at high magnifications. In these cases the scanning electron microscope (s.e.m.) was used.[2, 3] This was capable of much higher magnification with a very large depth of field and produced what appeared as a 'three-dimensional' picture. On scanning an s.i.c. chip, biased into its operating condition, with the s.e.m. electron beam, secondary electrons were produced at the area of impact. The intensity of these was dependent on the relative potential of the chip at that point, since recapture of these electrons could occur. The secondary electrons were used to modulate the brightness of a c.r.t. display. The s.e.m. was adjusted so that the most positive chip potential caused high recapture. The display showed black for these areas and bright areas for those at zero potential. Light intensities between these extremes were produced for other potentials. By this means the gradient across resistors, conducting and non-conducting junctions and biased epitaxial and substrate areas were displayed. Examples of these are shown in Fig. 13(a) where saturated transistors have light collector areas and the forward biased base/emitter junction indicated by a slight darkening of the base connexion relative to that of the emitter. Non-conducting transistors have dark collector areas with the base and emitter areas of similar brightness. Current flow along resistors can also be distinguished. With the s.e.m. switched to its 'television mode', where the display

was 625 lines per frame, and 50 frames per second, it was possible to examine low-speed dynamic phenomena within the devices. Switching of transistors, change in current levels, clock line levels, etc., could easily be checked. On comparing these results with those obtained from a fault-free device, the fault area and often the reason for the fault became obvious.

5. Detailed Fault Analysis

Each of the designs produced had their own peculiar fault patterns as well as some that were common to more than one design. It was found that the experience gained from each circuit fault examination was invaluable in the design of the following new devices. This was especially noticeable in the packing density and circuit complexity. To highlight the diffusion faults, circuit layout and design errors found, they are listed as follows:

5.1 Wideband Amplitude Modulator

This was the first design that was attempted at S.R.D.E.[1] At that time the design rules provided by the diffusion house were not in their final form and there was some doubt whether this type of work could be done outside the industry. To overcome this doubt and to encourage further work, the author made every effort to ensure that this device gave some yield. This was done by undertaking a relatively simple design, using a low packing density, wide track resistors (up to 16 μm) and large geometry transistors. This returned a dividend in that a working yield from untested devices of 80% in a batch of 25 samples was achieved. It was found that the spread of supply currents was small, the audio- and radio-frequency responses were consistent with those estimated and the range of values for an external preset resistor was well within that indicated by the design equations. This showed well the repeatable nature of the integrated circuit.

5.2 Current-sensitive Control Unit

This was produced as two separate designs, the second being a further development of the first. Owing to changes in specification outside the control of the author, it was decided to develop a second design to incorporate all the proved features of the first with some hindsight modifications. The initial diffusion of the first design was found to be outside the schedule tolerances. With this knowledge, the author ordered a small number of these for initial checking of the layout. It was found that they suffered from low voltage break-down and thin base diffusion. This resulted in excessively high current gains in the transistors and high values of diffused resistors. From this batch it was found that three devices fulfilled the initial specification, proving the design and the layout. Subsequent diffusions showed that the low voltage breakdown and thin base diffusion had been cured. This resulted in an acceptable yield of circuits that were within specification. Measurements on the out-of-specification devices showed a trend toward low currents in a constant current load. This was designed around an absolute value of diffused resistance, the current spread to be not greater than

±20% of the design figure. It was found that this resistor, in each case, was outside the accepted diffusion spread of ±20% and in some cases was as high as +50% above the design value. Chip examination indicated that there was a narrowing of the resistor tracks over that indicated by the diffusion masks. This was attributed to a swelling of the photoresist coating after development. The resistor track widths on subsequent diffused slices were measured to demonstrate that the quality control had been tightened.

Work involved in monitoring the value of resistors in these early batches was extended to examining the temperature coefficient of the diffused resistor and as a means for measuring the chip operating temperature. This was used in turn as an indication of package dissipation at various ambient temperatures. The curves of Fig. 4 were typical of the temperature/resistance behaviour found in several batches of devices and were close to that quoted for the diffusion schedule. In every case there was a point of minimum resistance at about 0°C.

Fig. 5. Bridge method of measuring diffused resistors.

temperature after each operation. The sand ensured that the temperature change was slow, allowing measurements of resistance to be made at equal steps in temperature.

The actual chip surface temperature and hence the junction temperature (T_j °C) was further obtained by placing the 'calibrated' device in various ambient temperatures and varying its dissipation by means of its external supplies. The voltage across the test resistor was maintained at its calibration value for each measurement and these readings were converted to temperature from the calibration curve. The curves of Fig. 6 are plots of junction temperature against chip dissipation at ambient temperatures of 20°C, 80°C and 125°C. This was for a chip mounted in a 14-lead 'flat pack' free standing in still air. Hence for a maximum junction temperature of 175°C it was shown that this type of package could dissipate 500 mW at an ambient of 80°C, but only 280 mW at an ambient of 125°C. This technique could obviously be extended to the measurement of other packages and the effectiveness of heat sinks.

The redesigned circuit attempted to incorporate design improvements as well as satisfying a much tighter
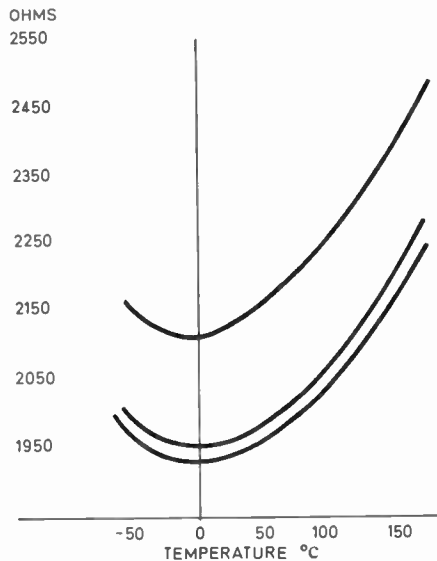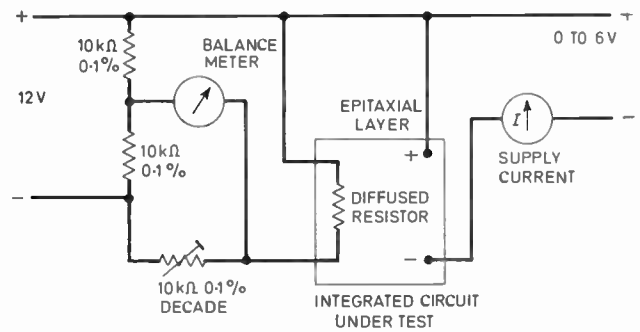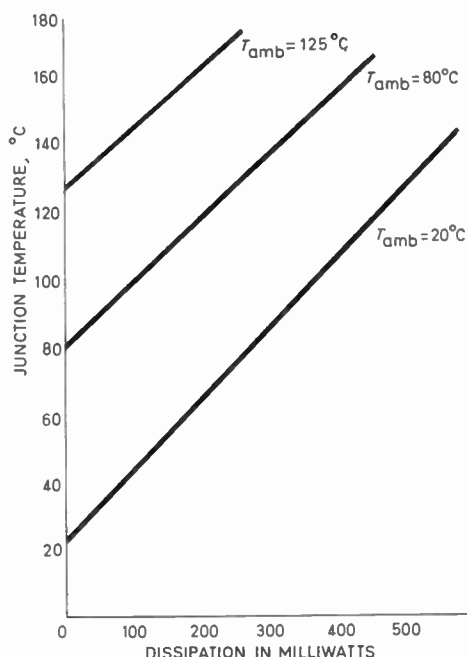
Fig. 4. Variation of resistance with temperature for diffused resistors.

When measuring the diffused resistor, isolation from the epitaxial layer was maintained by appropriate bias between the chip substrate and the epitaxial layer with one end of the test resistor connected to the most positive potential as shown in Fig. 5. It was fortuitous in this circuit layout that the chosen resistor was connected in the design between the positive supply line and an external pad connexion. This enabled a Wheatstone bridge method of measurement to be used with the bridge bias negative with respect to the chip epitaxial layer voltage. The resistor measured was the highest value in the circuit (nominally 2·2 kΩ) so that the local heating due to its own dissipation was small.

This integrated circuit with a thermocouple in close contact was buried in a beaker of sand and heated to a temperature of about 170°C and then cooled in a refrigerator to −60°C, being allowed to return to room

Fig. 6. Junction temperature/dissipation for ¼-inch square flat pack.
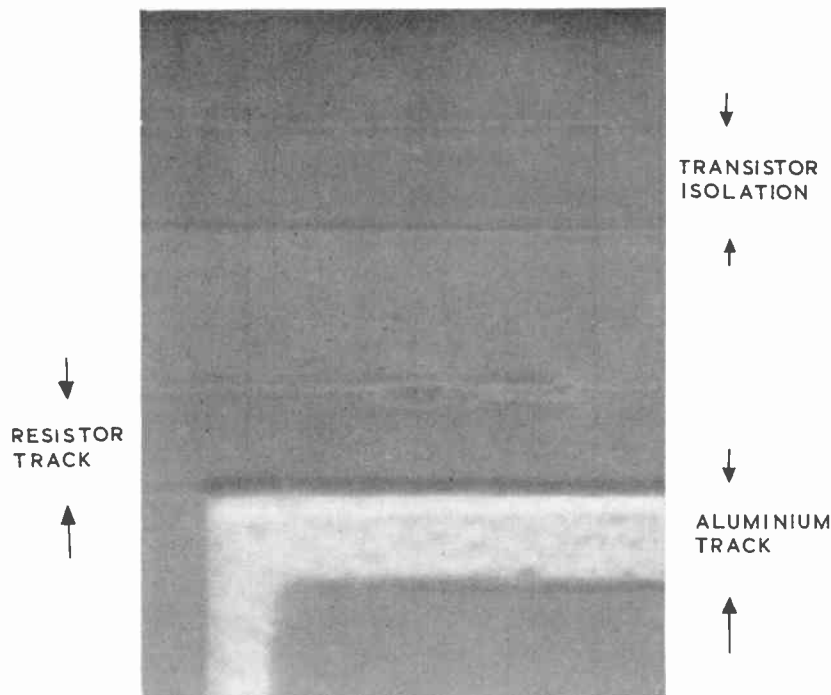
RESISTOR
TRACK

TRANSISTOR
ISOLATION

ALUMINIUM
TRACK

Fig. 7. Pinholes in oxide layers
(magnification × 1000).

operating specification with some success. This was borne out in the first diffused samples of this design where a yield of 20% was obtained from visually inspected and encapsulated chips. Examination of the faulty devices indicated a high incidence of malfunctions which appeared to be voltage dependent. Detailed examination attributed this to resistors connected to the most positive supply line (18 V). Certain of these displayed a 'Zener' breakdown characteristic which latched the lower potential end of the resistor to approximately 5 V below this supply rail. In isolation these resistors showed that they possessed a Zener characteristic, i.e. a sharp breakdown 'knee'. This indicated the presence of emitter diffusion in contact with the resistor diffusion areas, which could only occur through oxide or mask faults (Fig. 7). Operating the devices in the dark showed in many cases the light-emitting behaviour associated with reverse bias breakdown. These facts confirmed that the devices had been assembled from a diffused slice with an unacceptably high incidence of oxide pinholes. These were traced to a malfunction of the 'whirler' used in the photoresist coating of slices prior to exposure to the emitter diffusion mask. This caused excessive thinning and in some cases puncture of the photoresist over previously etched steps. The malfunction was corrected and no further problems were anticipated in this area.

5.3 Low Off-set Unity Gain Gating Amplifier

A previous proposed design,[4] had proved to be difficult to produce with a satisfactory yield and the author was requested to start a parallel exercise as this device was expected to be a holding item in the project.

A severe restriction was placed on the preparation of the master masks for this circuit in that the pad connexions had already, in effect, been defined by the

associated equipment designers. This led to considerable difficulty in producing a compact layout as detailed later.

On evaluating the initial batch of diffused devices a yield of 20% was achieved which fulfilled the requirements in all except one item. This shortcoming was attributed to a design error that was subsequently corrected. Examination of the 'out-of-specification' devices in this and later batches showed that the design was very sensitive to 'phantom' diffused components. This phenomena was further examined in some depth.

The design error resulted from an attempt to improve the switching speed of the circuit so as to increase its area of applications. The basic circuit is shown in Fig. 8. The amplifier was inhibited by switching the point B to the negative supply rail and biasing the point A below 0 V. The fastest inhibiting time was 200 ns and was dictated by the switching speed of the
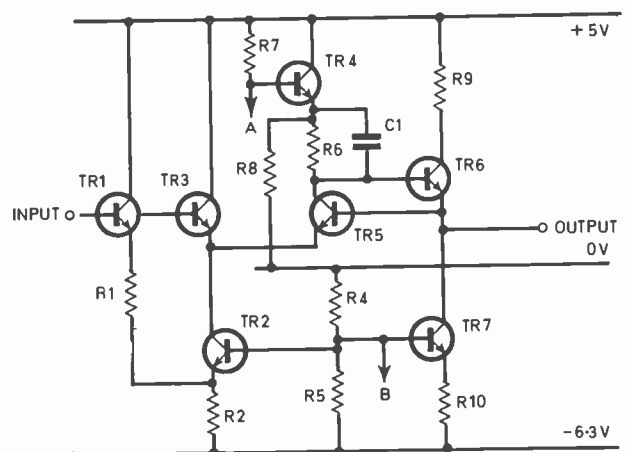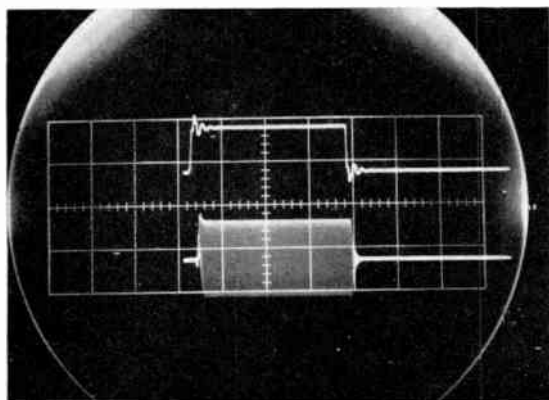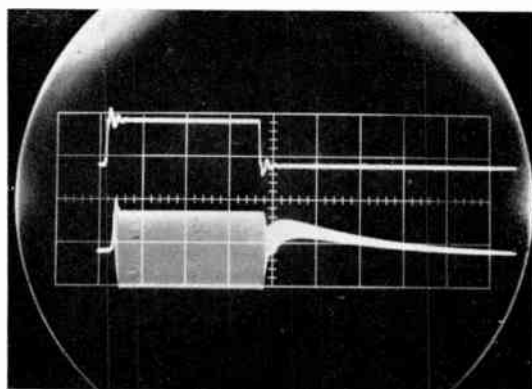

Fig. 8. Simplified circuit diagram of gating amplifier.

(a) With R8 included.

(b) With R8 removed.

Fig. 9. Switching behaviour of gate amplifier.
Horizontal scale—0·5 µs per square.

circuit associated with TR4 and TR6. During the falling switch transient at A, TR4 acted as an emitter follower with R8 as its d.c. load. Hence the base of TR6 was forced down due to the charge on C1 in series with the falling emitter of TR4. Similar reasoning applied with a rising transient when the base of TR6 was forced up. With the circuit fully gated off the output resistance was very high since each of the transistors TR5, 6 and 7 were non-conducting.

In certain applications more than one of these circuits were connected to a common output, but with only one circuit in a conducting state. By this means multiplexing or signal routing was achieved. Under these conditions the output connexion of the non-conducting circuit was modulated by the voltage from the conducting circuit. In order that there should be negligible loading of the output by the former, its output resistance was required to remain high with a leakage current not exceeding ± 5 µA for a modulation swing of ± 1 V. In practice it was found that this leakage was exceeded and was attributed to the two leakage paths:

(1) R8, R6 and V_{be} of TR6.

(2) R8, R6 and V_{cb} of TR5.

These became effective only when the modulation voltage exceeded the transistor p–n forward bias voltage.

A simple correction was made by the removal of resistor R8. This was taped over in the original 'cut and strip' layout and a new diffusion mask prepared at small cost. The removal of this resistor reverted the circuit to a slower 'switch-off' behaviour of about 1 µs. This was still within the specification limit required. The histograms of the two switching behaviours are recorded in Fig. 9. It will be noted that only the 'switch-off' transient was affected; the transistor TR4 still acted as an emitter follower in the 'pull-up' mode.

The elimination of the above design fault uncovered a percentage of devices in subsequent diffusions that still displayed excessive leakage in the gated-off mode due to a diffusion fault. Isolation of components within these chips proved that collector leakage in either of the transistor TR2 or TR7 was the prime cause. It was found that the specified value of 10 nA for collector leakage was in some cases grossly exceeded. This is shown in Fig. 10 with base returned to the negative rail. Here an ohmic leakage was evident and its value was of the same order as the circuit leakage. Samples of TR2 were also examined and similar results were recorded. In some devices both transistors were at fault. It was interesting to note that the faulty transistor could be identified without resorting to circuit probing by monitoring the output leakage while varying the input to the amplifier. In the case of leakage through TR2 via the base/emitter junction of TR5, this had a component of the input signal, since leakage in TR2 allowed TR3 to act as an emitter follower. In the case of leakage in TR7 this was not affected by the input signal. The factors in the diffusion process that contributed to this leakage phenomenon were corrected and subsequent devices have been free from this fault.

It has been mentioned above that the pad connexions to this device were specified prior to commencing the layout. The best configuration that could be achieved required the routing of the negative supply rail outside four of the connecting pads as shown in the bottom
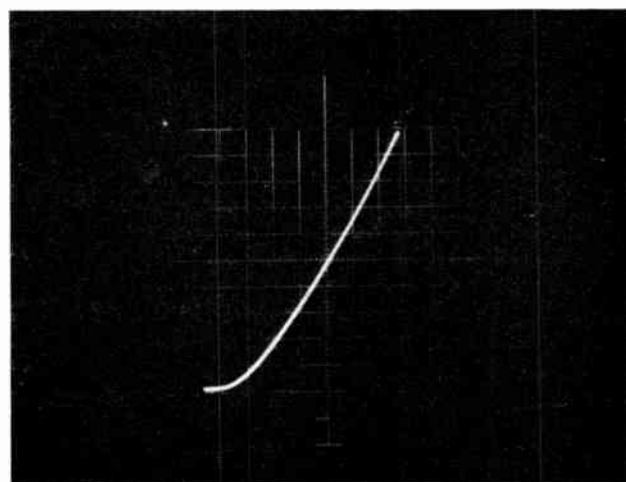
Fig. 10. Characteristics of leaking transistors.
Scales: horizontal 1 V/div; vertical 10 µA/div.

left-hand corner of the chip photographs of Fig. 13 (a and b). In production devices that have a surface passivation coating this was not considered a detriment even though the bond leads were longer than normal. In pre-production devices this coating was not added so that this layout presented a limited risk since shorts between these pads and the supply line could occur due to over-laying of the bond lead.

This was the case in one diffusion batch where the lead was drawn tightly over the chip surface. This is shown in detail in the s.e.m. micrograph of Fig. 11. Careful manipulation of a microprobe enabled the lead to be lifted from the chip surface and the fault cleared. This small incidence of faults was reported to the manufacturer and appropriate action was taken by their Quality Assurance Department.
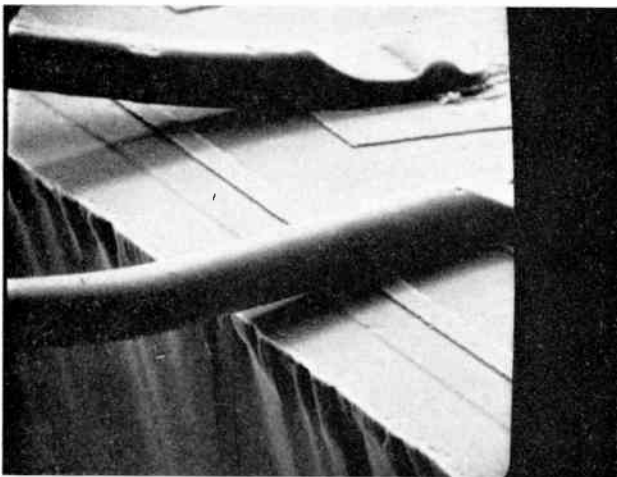

Fig. 11. S.E.M. micrograph of bond lead short.

A further and more serious malfunction of this device was its sensitivity to 'phantom' diffused components. Anywhere within an integrated circuit where three diffused areas come into contact there is the possibility of the creation of a phantom transistor, either n–p–n or p–n–p, though under normal conditions the current gain is less than unity. In the case of the normal epitaxial n–p–n transistor with its emitter connected to the most negative voltage in the circuit, i.e. the substrate potential, a phantom p–n–p transistor is formed using the base, collector and substrate of the former to form the emitter, base and collector of the phantom. In practice this arrangement has a very low current gain due to its thick 'base' diffusion which is partly made up of the highly doped buried n layer. This absorbs most of the current flowing from the lightly doped 'emitter' giving a poor base transport factor. When the normal n–p–n transistor is driven into saturation as in RTL applications, its collector voltage is at a lower potential than the base. This results in the phantom p–n–p also being made to conduct, but in this case the current flows into the substrate. In complex RTL circuits this can result in appreciable substrate currents which, unless a low resistance substrate connexion is made, can create potential elevation of the chip relative to the negative supply rail.

The sensitivity of the circuit design in question to the various configurations of phantom transistors appeared as devices which took excessively high line current in some cases sufficient to destroy the device. In some this was a permanent phenomenon and in others it was precipitated by a switching transient. Detailed investigation indicated a phantom four-layer diode (s.c.r.) whose characteristic is shown in Fig. 12. This was obtained by means of a transistor curve tracer which was used to sweep the positive supply rail of the device relative to its negative supply rail. Most of the devices showed this characteristic, the rejects having a breakdown voltage close to or below the normal supply voltage. Circuits that required breakdown voltages just in excess of normal supplies were sensitive to transients on the lines that would trigger the s.c.r. In other devices the breakdown voltage was very high and was masked by the normal breakdown behaviour of the diffusion process. It was considered that the four layers of the p–n–p–n diode were made up of:

(i) A resistor with one end connected to the positive supply rail.
(ii) The epitaxial layer.
(iii) The substrate.
(iv) A collector region in a transistor which was in a saturated state and whose emitter was connected to the negative supply rail.

The s.c.r. 'trigger' electrode was (iii) above, i.e. the substrate. The bias on this depended on the substrate currents flowing from the phantom p–n–p devices detailed earlier. It was further shown that if the substrate bias was reduced by making a low ohmic contact between the negative supply rail and the substrate, the s.c.r. behaviour was inhibited. This was effected by means of an extra bond lead between the chip negative supply pad and the inside of the metal encapsulation. This created a good contact to the substrate due to the large area of the eutectic bond between the back of the chip and the metal encapsulation.
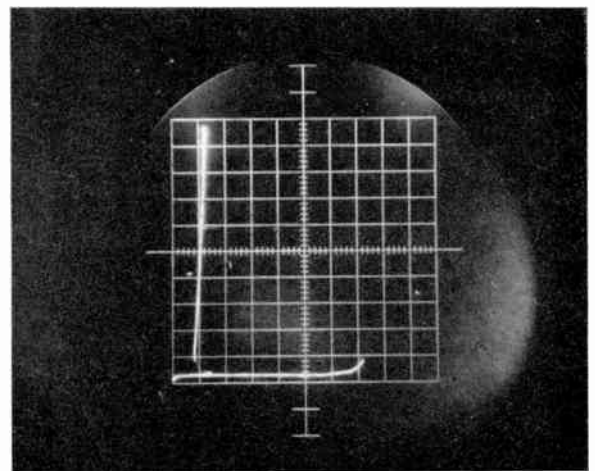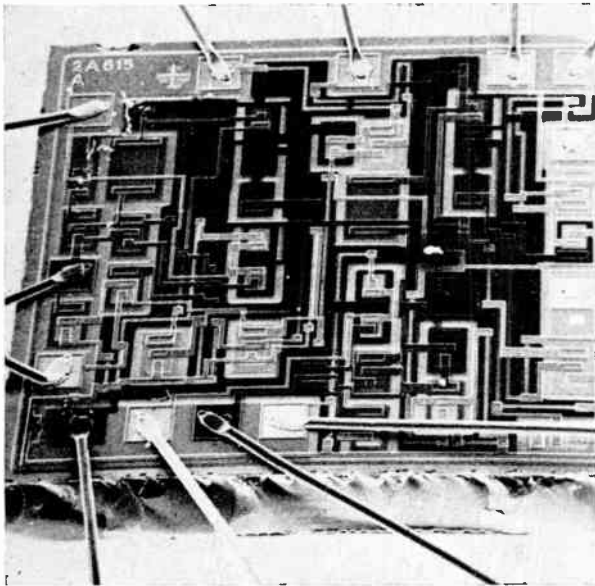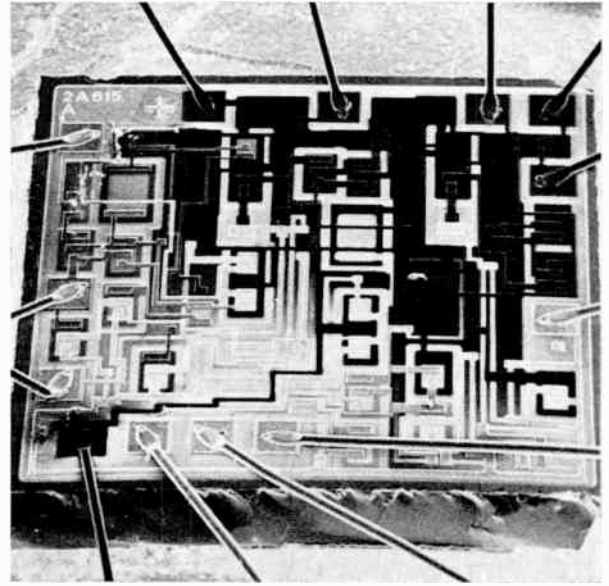

Fig. 12. Characteristic of phantom p–n–p–n device.
Scales: horizontal 1 V/div; vertical 1 mA/div.

(a)   (b)

Fig. 13. S.E.M. micrographs of biased device.

The s.e.m. micrographs of current paths in a circuit exhibiting the s.c.r. characteristic compared with one operating in its correct mode are shown in Fig. 13(a) and (b). Figure 13(a) shows correct current distribution with supplies connected only to the positive and negative supply pads. Here the epitaxial layer is at the positive potential over its whole area and the various transistors can be seen to be either in their non-conducting or saturation condition. Figure 13(b) shows a similar circuit where the phantom s.c.r. has triggered. The lower part of the epitaxial layer can be seen to be biased below its normal potential corresponding to (ii) above. The transistors either in the bottom right-hand corner or above bottom centre, both of which are in saturation, correspond to (iv).

6. Conclusions

During the last three years the author and his colleagues have investigated and gained experience in the following areas associated with integrated circuits: circuit design, integrated circuit simulation, layout drawing and preparation, diffusion faults, bonding and encapsulation faults, the use of mechanical probes in electrical measurement and the isolation of diffused components, and the use of the scanning electron microscope for the examination of the chip surface and its behaviour under static and dynamic biased conditions. This has ensured an appreciation of the advantages and disadvantages of this technology in military communications equipment. First-hand experience has been obtained in the new skills required and the difficulties that have to be overcome.

In circuit design, a considerable expertise has been built up in the creation of complex circuit functions in the restricting framework imposed by the s.i.c. medium. This has relied on a close adherence to the manufacturer's diffusion rules. Reliable circuit simulation was found to be important in this learning process and resulted

in the use of the conventional 'bread-board' technique as well as computer simulation.

One area of weakness in the producing of a special s.i.c. has been highlighted as the creation of the diffusion mask layouts. These can, in the case of complex bipolar devices, form the major part of the circuit producing cycle. This is due entirely to the requirement for a high component packing density within the s.i.c. chip, which can only be achieved by means of human judgement and dexterity.

Failure analysis was divided into two areas. Firstly, the examination of standard devices that have failed in use showed that these faults were usually of the 'misuse' or 'poor workmanship' category and were routine. It was considered that the second area covering faults and weakness in new designs was more fruitful and relevant. As well as covering all aspects of quality of workmanship essential in determining the levels of inspection to be specified in production, the following were studied in detail. A technique was developed whereby the heat dissipation of various s.i.c. packages could be assessed with and without heat sinks. By this means a conservative chip surface temperature could be defined at the highest equipment operating temperature, contributing greatly to increased reliability. Some of the unspecified secondary diffusion parameters were also considered of great importance due to their effect on device yield.

Tightening of the associated specification limits and extra attention to process detail by the manufacturer resulted in an elimination of high leakages in reverse-biased transistor junctions in sensitive circuit designs. The effects of 'phantom' components were controlled by a thorough understanding of their effect on the circuit operation. This was assessed where possible before the monolithic stage was reached. In practice this was not always achieved and corrections to the layout and design had to be done in hindsight.

7. Acknowledgments

The author would like to acknowledge the cooperation of his colleagues K. Head and P. Aitken who were closely associated with the work of this paper. Acknowledgment is made to the Plessey Company who processed the designs, for their helpful cooperation and interest in the results of the failure analysis.

The paper is reproduced with the permission of the Controller, Her Majesty's Stationery Office and is Crown copyright.

8. References

1. Stewart, A. and Jones, C. H., 'A wide band amplitude modulator as a special silicon integrated circuit', *The Radio and Electronic Engineer*, 38, No. 3, pp. 137–45, September 1969.

2. Head, K., 'Scanning Electron Microscopy', S.R.D.E. Report No. 70023.

3. Devaney, J. R., 'Application of scanning electron microscopy to integrated circuit failure', *Solid State Technology (U.S.A.)*, 13, pp. 73–7, March 1970.

4. Jones, C. H., 'Development and Engineering of a Low-offset Unity Gain Gating Amplifier as Special Integrated Circuit', S.R.D.E. Report No. 68025.

5. Adams, J. and Workman, W., 'Semiconductor network reliability assessment', *Proc. Inst. Elect. Electronics Engrs*, 52, pp. 1624–35, December 1964.

STANDARD FREQUENCY TRANSMISSIONS—March 1972

(Communication from the National Physical Laboratory)

March 1972	Deviation from nominal frequency in parts in 10¹⁰ (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		March 1972	Deviation from nominal frequency in parts in 10¹⁰ (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	+0·1	0	0	632	660·8	17	0	+0·1	+0·1	632	654·0
2	0	+0·1	0	632	659·6	18	0	0	0	632	653·8
3	0	0	−0·1	632	659·9	19	+0·1	0	0	631	653·6
4	0	0	−0·1	632	660·0	20	+0·1	+0·1	0	630	653·0
5	−0·1	−0·1	−0·1	633	660·7	21	0	0	0	630	653·0
6	−0·1	0	−0·1	634	661·1	22	0	−0·1	+0·1	630	654·4
7	0	0	−0·1	634	655·9	23	+0·1	+0·1	+0·1	629	653·5
8	0	0	−0·1	634	656·0	24	+0·1	+0·1	+0·1	628	652·3
9	0	0	−0·1	634	656·0	25	+0·1	0	+0·1	627	652·5
10	0	0	0	634	655·8	26	0	+0·1	0	627	651·3
11	+0·1	+0·1	0	633	656·0	27	0	0	0	627	649·9
12	+0·1	+0·1	+0·1	632	655·5	28	0	0	−0·1	627	649·7
13	0	0	0	632	655·7	29	0	+0·1	0	627	648·7
14	0	0	0	632	654·4	30	+0·1	0	0	626	648·9
15	0	0	0	632	654·4	31	0	0	0	626	648·6
16	0	0	0	632	654·6						

All measurements in terms of H.P. Caesium Standard No. 334, which agrees with the N.P.L. Caesium Standard to 1 part in 10¹¹.

* Relative to UTC Scale; (UTC$_{NPL}$ − Station) = + 500 at 1500 UT 31st December 1968.

† Relative to AT Scale; (AT$_{NPL}$ − Station) = + 468·6 at 1500 UT 31st December 1968.

Terrestrial Microwave Radio Relay System Development at Frequencies above 10 GHz

Joint Propagation Studies by the Radio and Space Research Station and the Post Office Research Department

Fig. 1. Standard lattice steel microwave tower in the Post Office radio-relay network near Plymouth; the lowest dish antenna (1·5 m diameter) on the left-hand leg, with a protective radome, is installed for the new 11 GHz link to Caradon Hill.

To assist in meeting the explosive growth expected in telecommunications in the next twenty years the Post Office and the Science Research Council have begun what is probably the largest microwave-propagation study yet to be mounted anywhere. This project could have a major influence in opening up the 10-100 GHz range of the radio spectrum (wavelength down to 3 mm) to new microwave telecommunication system use. As frequencies rise about 10 GHz, transmissions are increasingly affected by the weather and the object of the study is to find out just how radio transmissions at these millimetric wavelengths are affected and what can be done to avoid disruption of service. This article describes the broad details of the project.

Large increases in the telecommunications network of the British Post Office will be required in the next two decades to cope with the forecast growth in demand for existing services and with the introduction of new services such as Viewphone. There is a trend towards digital transmission, with the goal of a versatile all-purpose network, while data services in particular are growing rapidly, with increasing interest in high speed data. In these circumstances the Post Office must study the technical and economic feasibility of meeting the demands with various methods of cable and radio transmission so that the network can quickly respond to customer needs.

Microwave radio-relay systems operating in frequency bands centred on 2, 4 and 6 GHz have played a major part in the trunk network in the past two decades but growth of this traffic has been so great that these frequency bands are becoming congested and it is necessary to extend upwards the range of frequencies to be used. Radio propagation, even at the frequencies now in use, is affected to some extent by the weather because variations of refractive index of the atmosphere cause fading of the received signals. At the higher frequencies now being considered, extending up to 40 GHz and possibly as high as 100 GHz, the effect of the weather and especially rainfall becomes increasingly severe.

System Development at Frequencies above 10 GHz

Recognizing the possibilities for exploiting frequencies above 10 GHz to meet its expanding communication requirements, the Post Office is pursuing a research and advanced development programme centred at present on two main types of digital microwave radio-relay system:

An 11 GHz system transmitting digital information at bit rates of about 100 Mbit/s on each of a number of microwave carriers and suitable for multiple 'hop' lengths of some 30 km.

A 20 GHz system transmitting digital information at bit rates of up to around 500 Mbit/s on each of a number of microwave carriers and suitable for multiple 'hop' lengths of some 5 to 10 km.

Both systems use pulse code modulation of the microwave carriers instead of the frequency division multiplex/frequency modulation methods characteristic of present 2, 4 and 6 GHz systems. Digital techniques offer major advantages, notably:

better quality and stability of transmission,

the use of regeneration at each relay station which permits a virtually unlimited number of stations to be connected in tandem without loss of signal quality, and

suitability for transmitting different types of signal, e.g. telephone, data, facsimile and Viewphone, on the same carrier without mutual interference.

The 11 GHz system is intended for use at the stations of the existing extensive Post Office microwave radio-relay network in the United Kingdom, and on spur routes to that network (Fig. 1). By using existing sites it will be capable of early exploitation, e.g. to provide high-speed data transmission over inter-city links.

The 20 GHz 'short-hop' system will use compact solid-state low-power microwave equipment mounted on poles up to 24 metres in height by the road-side at intervals of up to 5 km or 10 km, the poles being similar to lighting columns used on motorways and at large intersections. The poles—much less obtrusive than lattice steel towers—would stand by the roadside at appropriate intervals, particular attention being paid to the environmental amenities aspect of the design (Fig. 2). The shorter hops are necessitated by the increased effect of attenuation by rain at the higher frequency; space-diversity, discussed later, may also be required in some applications to mitigate this effect. This system is in an earlier stage of development than the 11 GHz system. Its main field of application will be over junction routes of moderate capacity; other potential applications include lower capacity, e.g. 10 Mbit/s, digital 'tails' to customer premises.

Both systems are being developed in co-operation with the United Kingdom telecommunications industry.

A longer-term possibility, at present in an early stage of study, is the use of frequencies in the 50 to 100 GHz range for local distribution, e.g. over distances of up to 1 km, covering, say, a residential area and providing a wide range of telecommunication services direct to individual premises.

All these applications of microwave radio-relay systems require comprehensive statistical information on the characteristics of microwave propagation extending over at least a year so as to cover representative weather conditions, and in a form applicable to the various areas of the country. This information relates not only to variation of attenuation due to rainfall and other causes, but also to the 'multi-path' characteristics of typical routes because these condition the maximum digital bit rates that can be transmitted.

In short, detailed and comprehensive microwave propagation information will be essential for the economic and effective design and utilization of these microwave systems in the future.

Fig. 2. Artist's impression of the slender pole that will carry Post Office microwave equipment of the future. The 2 m × 1 m mast-head canopy carries all electronic equipment and two 60 cm dish antennas; it will be possible to lower it to ground level for maintenance using a cable and winch inside the pole.

Planning of Terrestrial Microwave Links

Although variation of refractive index may cause fading, as at the lower frequencies, rain falling in the radio path can cause significant attenuation of the signal: at the higher frequencies, raindrops both scatter and absorb the energy of the radio wave. Snow melting as it falls may produce substantial attenuation. The diameters of droplets found in clouds and fog are relatively small so that their effect is much less marked. Examples of the effect of rain and snow are given in Fig. 3. Because the amount of radio energy scattered and absorbed depends on the diameter of the raindrops and because heavy rainfall has a larger mean raindrop diameter than lighter rainfall the attenuation depends on rainfall intensity as well as frequency. This is shown in Fig. 4.

To plan radio-relay systems which may be affected in this way by rainfall it is necessary to know how frequently rainfall of a given intensity is likely to occur in a particular area, how far it extends and how long it will last. The cell-like structure of areas of rainfall has been known for many years and has become more generally appreciated with the increasing use of weather radar displays, such as the one shown in Fig. 5. This cell-like structure offers the possibility that if one radio path should be severely affected by a rainstorm an alternative path may be found which avoids the rain cell causing the loss of signal. The improvement in reliability to be obtained by providing an alternative path is demonstrated by the curves of Fig. 6.

(a) Snowfall fade on a 31 km path.

(b) Rainfall fade on a 24 km path.

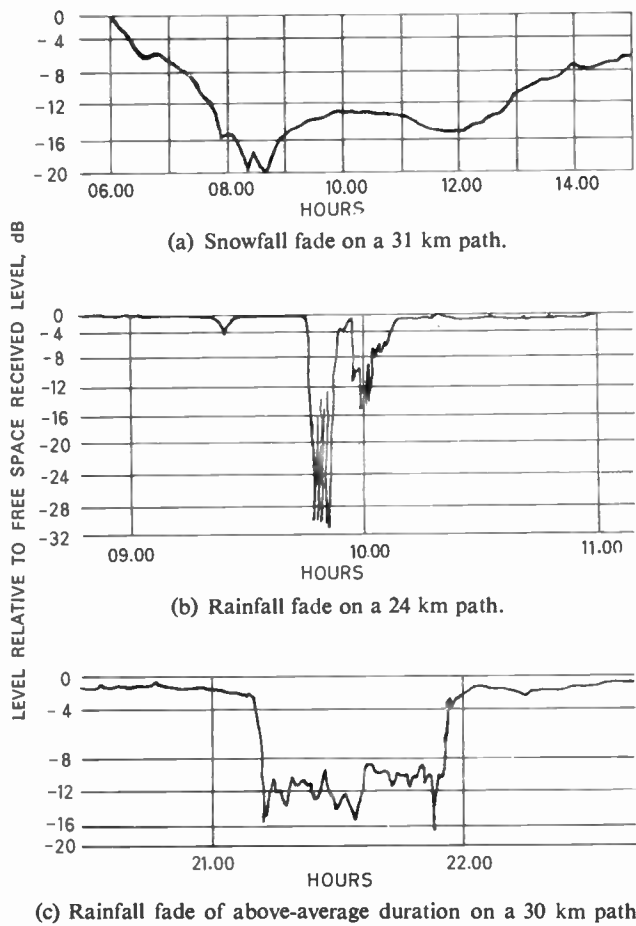(c) Rainfall fade of above-average duration on a 30 km path.

Fig. 3. Examples of fading on parts of various lengths of a 11 GHz radio link due to precipitation.

It is probable that the full development of microwave radio-relay systems will ultimately require the use of orthogonal polarizations of the same carrier frequency over the same link. For satisfactory system performance the cross-polar discrimination will therefore need to meet certain requirements. Rainfall, in addition to producing attenuation of the signal, may also affect the cross-polar discrimination. This arises because raindrops are rarely spherical and usually take forms approximating to oblate spheroids. The effect on a radio wave is to produce a signal component with orthogonal polarization and this appears as interference in the other channel. Variations of refractive index, by tilting the received wavefront, may also produce the same effect.

Significant fading is sometimes observed in clear air under stable weather conditions, particularly during nights with little cloud; when this occurs, the atmosphere may become stratified so that at a particular altitude a rapid change of refractive index may occur. Some of the energy from the transmitting aerial travelling at an angle from the direct path to the receiving aerial may arrive by reflexion from regions of rapidly changing refractive index so that two or more signals may be received. The reflected signals will have travelled by longer paths and so will be delayed with respect to the direct signal and may be in opposition to it, causing a fade, and digital systems may suffer inter-symbol

interference. The frequency with which this multi-path propagation is likely to happen and its severity and persistence when it occurs are largely unknown, particularly at the higher frequencies. An example is given in Fig. 7.

The efficient planning of a network of microwave radio links for frequencies higher than those in current use which will meet internationally agreed standards of performance clearly requires determination of the extent of the effects described.

Joint Post Office-R.S.R.S. Studies

The Post Office Research Department, in conjunction with the Radio and Space Research Station of the Science Research Council, has begun an investigation designed to find out just how much influence the troposphere is likely to have on the performance of U.K. terrestrial radio-relay systems of the future. The major part of the investigation is being carried out in East Anglia in the area between Martlesham, the site of the Research Department's new establishment, and Mendlesham Radio Station, an existing radio-relay installation 24 km distant.

At first sight, East Anglia, with annual rainfall lower than average, seems an injudicious choice for an investigation concerned with the effects of rainfall. However, as already described, radio-wave attenuation by rainfall is determined by instantaneous intensity and not by the annual fall, and the frequency of occurrence of high intensity rainfall in East Anglia is not so much less than average that the results would be unrepresentative. In addition, East Anglia is known to experience relatively frequently the tropospheric conditions which are conducive to multi-path transmission. Finally, the cost of installing and maintaining an extensive field experiment has been minimized by making use of spare
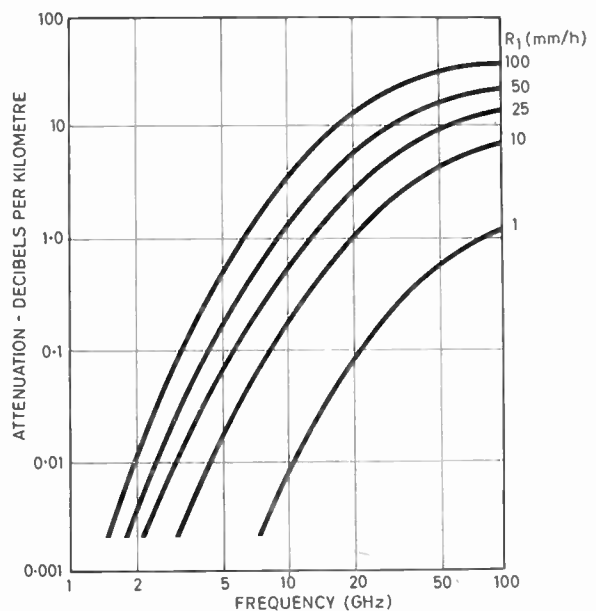

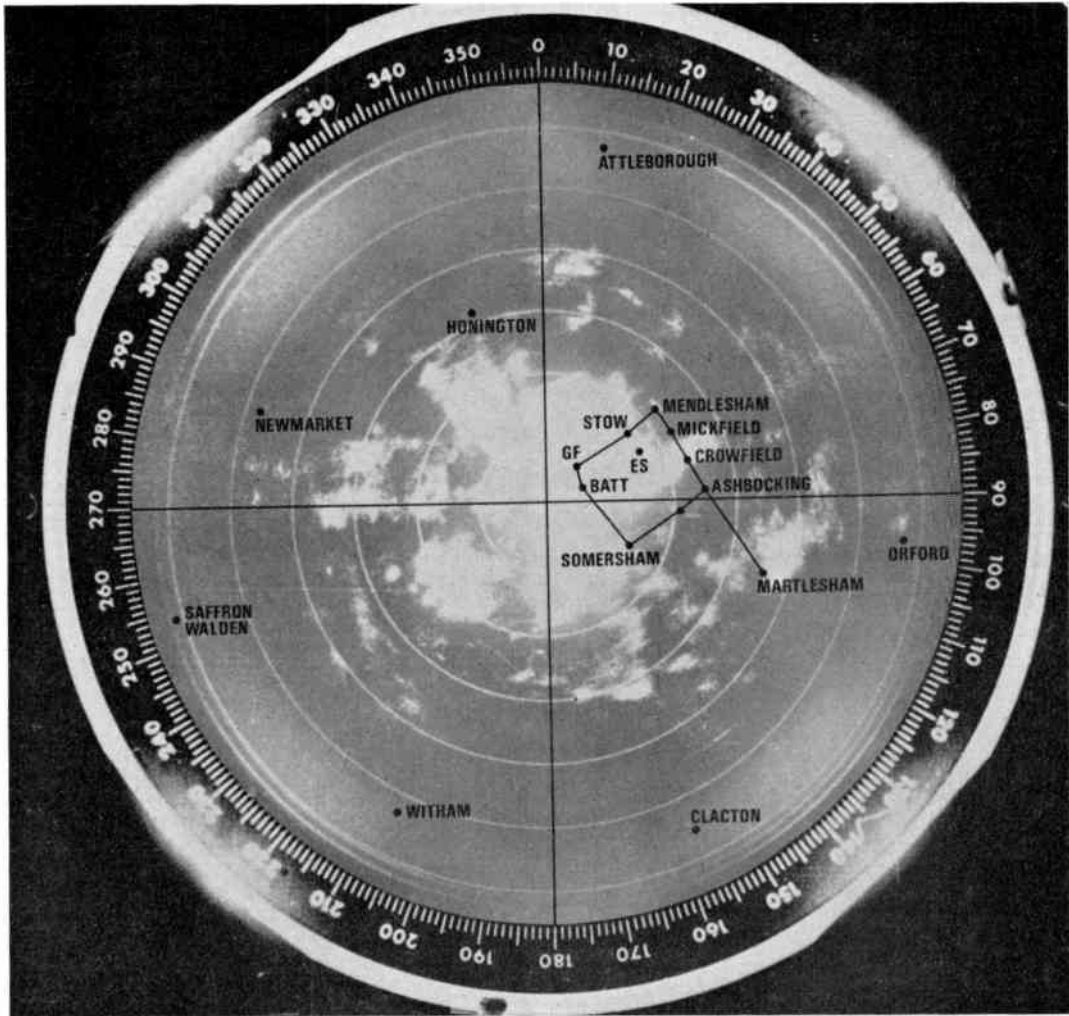Fig. 4. Variation of attenuation with frequency due to rainfall of different intensities.

Fig. 5. Weather radar display of rain showers. An outline of the project is superimposed with the positions of nearby towns to give an indication of scale.
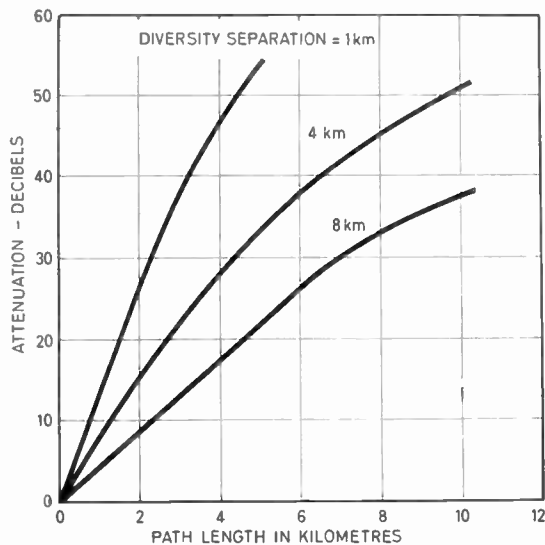

Fig. 6. 18 GHz attenuation appropriate to 0·001 % probability against path length for various diversity separations (after D. C. Hogg).

aerial mounting capacity on the mast at Mendlesham radio station and all installations in the experiment are within easy travelling distance of Martlesham.

The investigation has four main objectives:

(1) To determine the spacing required between radio paths so that should one path be severely affected by rainfall. an alternative path could be used.

(2) To determine the characteristics of multi-path transmission.

(3) To obtain further data relating rainfall intensity and radio signal attenuation.

(4) To determine the effect of rainfall and refractive index variations on the cross-polar characteristics of the radio signal.

Diversity Studies

To determine the best separation between radio paths, two sets of approximately parallel paths giving spacings of 4 km, 8 km and 12 km have been arranged with one set perpendicular to the other as shown in Fig. 8. This has been done to include the effect, if any, of persistent asymmetry in the distribution and shape in an area of
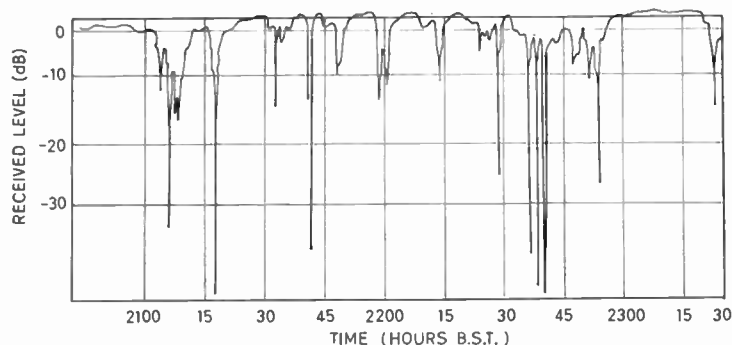
Fig. 7. Section of recorder chart to show an example of multi-path fading at a single frequency.

T – TRANSMITTERS
R – RECEIVERS

Fig. 8. Path diversity experimental network. Transmitted frequencies are 22 GHz and 37 GHz where stated.

rainfall. One set of paths lies approximately in a SW–NE direction, the direction of the prevailing wind. Single-frequency transmissions at 22 GHz and 37 GHz are used because rainfall is not expected to produce any significant change in signal loss with frequency over a typical r.f. channel bandwidth.

Data collected at each receiving station are transmitted to Mendlesham Radio Station over a radio telemetry system working at 460 MHz. These data will be analysed to determine the coefficient of correlation for simultaneous fading on the paths of various lengths and spacing, of different orientation and at the frequencies 22 and 37 GHz.

Multi-path Studies

To keep costs down, the multi-path transmission experiment makes use as far as possible of the same stations and only those combinations of frequencies and path lengths likely to yield the most significant information. These are shown in Fig. 9. Because multi-path transmission is sensitive to small changes of frequency, the transmitter frequency is stepped through a 500 MHz bandwidth in each case. The stepped-frequency responses showing multi-path activity are recorded at Mendlesham on magnetic tape for subsequent analysis by a combined analogue simulator and computer technique.

Attenuation and Rainfall Studies

Study of the relationship between attenuation and rainfall rate is fundamental to an understanding of the effect of rainfall on radio propagation. The first detailed theoretical work was carried out by J. W. Ryde during the 1939–1945 War. Since then there have been numerous attempts to verify Ryde's work.

Ideally, such studies require a knowledge of the instantaneous profile of rainfall intensity between the transmitter and receiver and the signal attenuation at the same instant. The latter is readily measured but the former requires a large number of rapid-response rain gauges distributed along the path. The measurements done so far have usually had neither sufficient rain gauges nor rain gauges of adequate response time.

The Radio and Space Research Station has developed a gauge from which rainfall rates averaged over suc-
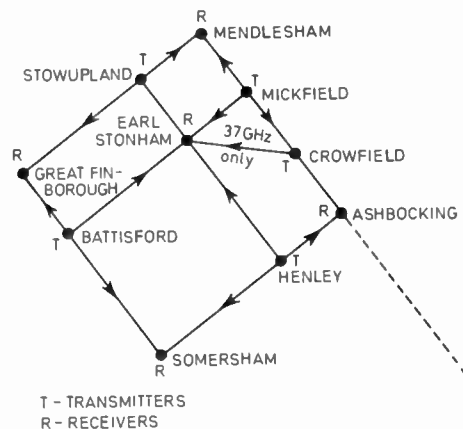
cessive 10-second intervals give a good approximation to instantaneous rates, a substantial improvement of rain gauges used previously. A photograph of the gauge is shown in Fig. 10. The gauge is of the standard-drop type, i.e. the rain collected by the funnel is formed into drops of constant size, the rainfall rate being determined by counting the number of drops falling through a light beam in a given period.

Twenty-one of these gauges have been installed at spacings of 0·8 km on the Somersham–Mendlesham path, and a further 19 between Crowfield and Mendlesham at 0·4 km spacings. The attenuation due to rainfall on these paths will be compared with data obtained from the rain gauges to (a) evaluate the significance of spacing in determining the near-instantaneous profile of rainfall-rate, (b) determine any persistent anisotropy in the rainfall distribution, and (c) study the relationship between characteristics and the design of diversity networks. Similar rain gauges are also installed at each transmitter and receiver site.

The rain gauges have been designed to function unattended for several months, the data collected by each gauge being transmitted to Mendlesham on command over a radio telemetry system. Each gauge contains a test and calibration system which can be
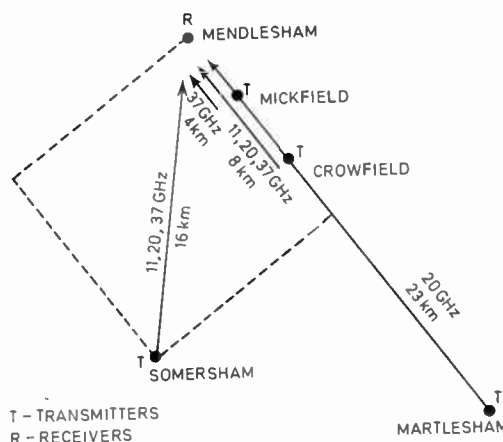

T – TRANSMITTERS
R – RECEIVERS

Fig. 9. Multi-path transmission test links.

operated from Mendlesham for periodic checks. It was decided that it would be an advantage to mount the rain gauges, logic circuitry and radio telemetry equipment at the tops of 20-ft high telegraph poles. This arrangement reduces the chance of vandalism, occupies minimum ground area because no fencing is needed, and, because the location is often restricted to the edges of fields, it keeps the gauge clear of debris and minimizes the possibility of spurious readings due to the presence of a hedge.
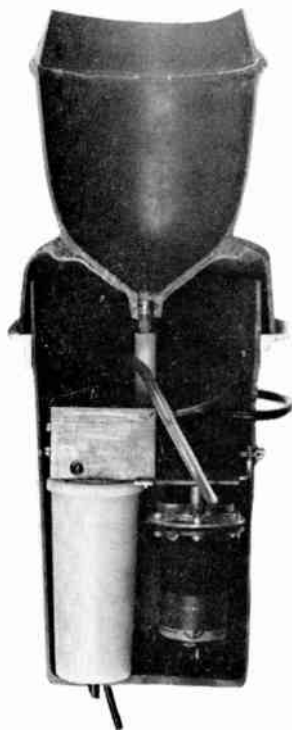

Fig. 10. Cut-away view of rapid-response rain gauge, developed by R.S.R.S. The light bulb and phototransistor and associated electronics are in the small metal box on top of the reservoir on the left; on the right is the calibration pump.

Supporting investigations are in progress at R.S.R.S., Slough. These include studies of rainfall data obtained since January 1970 with rapid-response rain gauges, simultaneous measurements of rainfall rate and of attenuation at a frequency of 36 GHz on a 500 m path and the development of a microwave scattering technique for probing rainfall structure using a spaced transmitter–receiver arrangement with narrow antenna-beams. The rainfall data illustrate the importance of short bursts of intense rainfall in the occasional heavy thunderstorm in the design of microwave relay links.

In addition, a link operating at a frequency of 110 GHz (2·7 mm wavelength) has been set up between Windsor and R.S.R.S., a distance of 2·7 km. This link has been in continuous operation since August 1970 and measurements have been made of the attenuation caused by precipitation, with the objective of determining the potential application in short-hop relay systems of frequencies much higher than those being studied in the experiments in Suffolk. The results show that, even

at 110 GHz, serious fading would only occur on a link about 3 km long for about 1 hour per year. Consequently, there may be useful applications in communications, especially in urban areas, for millimetric frequencies much higher than those hitherto exploited.

Cross-Polarization Studies

Equipment is also being installed on the Crowfield to Mendlesham and Mickfield to Mendlesham paths which will determine at 22 GHz and 37 GHz the effect on cross-polar discrimination of rainfall and refractive index variations in the path between the transmitting and receiving antennas.

Work on cross-polarization already carried out by the Post Office Research Department at Martlesham has shown that the presence of rain in an antenna feed aperture alone can affect significantly the cross-polar discrimination. The antennas for this study will therefore be equipped with radomes. Changes in cross-polar discrimination arising from propagation between antennas will then be measured by switching the polarization from vertical to horizontal every half second and measuring the co-polar and cross-polar components at the receiver.

Data Control, Logging and Analysis

Mendlesham Radio Station is the data control, logging and analysis centre for the four parts of the investigation. Data are received from the rain gauge network over a u.h.f. telemetry system and directly from the microwave receiving equipments at Mendlesham. These are recorded on magnetic tape for subsequent processing. In addition atmospheric temperatures and wind velocities will be recorded. Daily weather reports will be separately logged.

Extrapolation to Other Areas

In the joint experiments in Suffolk, the Radio and Space Research Station has installed radars, operating at wavelengths of 3 cm and 10 cm, to scan the region of the microwave links and rain gauges. The photographs of rainfall structure thus obtained will be used to relate the data from the links and rain gauges to the general weather pattern. R.S.R.S. is also proposing to obtain data from rapid-response rain gauges in other parts of the United Kingdom, to compare with the results being acquired in Suffolk and at Slough. A number of these gauges will be located on 11 GHz links being installed in the main microwave network. The P.O.R.D. is arranging to measure all significant fading which occurs on these links for correlation with the measured rainfall rate.

In parallel with these measurements, the theoretical treatment of attenuation caused by rain, and the extent to which theory and experiment agree, are being studied at R.S.R.S. As a result of this composite study it is hoped to facilitate an extrapolation to other areas of the United Kingdom of the conclusions, derived from the Suffolk network, on the characteristics of microwave attenuation and the design of diversity networks.