FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 40 No. 5 NOVEMBER 1970

Profile of a Profession

DESPITE its formidable numerical strength there was, at the time of the foundation of the Council of Engineering Institutions, comparatively little known about the 200,000 members making up the fourteen British professional engineering institutions which are federally linked within the Council. Our knowledge still has to be statistical, based on the findings of surveys addressed to approximately 25,000 corporate members and graduate members resident in the United Kingdom—a 15% sample. From two surveys which have been conducted jointly by the C.E.I. and the then Ministry of Technology, in 1966 and 1968,* some facts are emerging and basic information on fields and type of work, salary structure, levels of responsibility, and training and qualifications has been provided.

Certain deductions of interest can be made from a comparison of the 1966 and 1968 surveys. Average incomes across the whole range have increased by 12% in the two years' period while rather more engineers considered they held managerial posts—62·6% in 1968 compared with 59·5% in 1966; those in the sample occupying posts in 'top management' had increased from 11·4% to 12·3%. It is possible that the greater number of managerial posts filled by engineers accounts for the number who state they are 'not in engineering work' more than doubling; apart from this category, the proportions in other types of work (R & D, production, instrumentation and control, consultancy, etc.) have changed only very slightly.

The training of engineers was not enquired into in the 1966 survey but the questions asked in 1968 revealed that almost 40% of the sample undertook at least one training course during the previous 12 months' period. Business studies were taken by a third of those taking courses, followed closely and almost equally by advanced studies in 'own specialism' and in other technical skills. Barely 10% took foreign language courses although the proportion rose with age group, as did the numbers taking business studies. There were about the same number of full-time courses taken as part-time but the former were mostly of less than one week's duration while the latter often extended over long periods which obviously represented further education courses of two or three terms in the evenings at colleges of technology, etc.

The analyses of the survey figures according to membership of Institutions are not yet complete and will be commented on when available. There has however been an interesting breakdown according to industry of the higher degrees (Doctors' and Masters' degrees and post-graduate Diplomas) held by younger engineers (age 25 to 34). Chemical and allied manufacture have the largest number followed by electronic and telecommunications apparatus manufacture; these two accounted for more than half the engineers with higher qualifications in industry (and indeed nearly 80% of all Doctors in industry) whereas between them they employ only about 12½% of all engineers. It may be even more unexpected that holders of Doctor and Master degrees are both divided in roughly equal proportions between industry and the universities and colleges; Diploma holders tend however to be predominantly in industry.

Although many other interesting facts emerge from a study of the 30 tables, graphs and diagrams, this is a 'profile' not a 'portrait' of a profession. The sceptic may question the relevance, and reliability, of surveys of this type even though there was a response of almost 80% of the 25,000 questionnaires in the sample. As the Introduction to the Survey points out, 'National prosperity, in social as well as financial terms, depends upon the efficient use of resources, and technological manpower is as much a resource as money, materials and machinery. Manpower is, however, much less readily measured and much less susceptible to forward planning.' Some of the parameters on which the decisions for the planning of manpower resources must be made will certainly be obtained from this Survey.

F. W. S.

* 'The Survey of Professional Engineers 1968', H.M.S.O. 1970. Price £1. 2s. 6d.

INSTITUTION NOTICES

Clerk Maxwell Memorial Lecture

The Council is pleased to announce that Professor H. E. M. Barlow, D.Sc., F.R.S., C.Eng., F.I.E.E., has accepted an invitation to give the seventh Clerk Maxwell Memorial Lecture. Professor Barlow has taken as the title of his lecture 'Guided Electromagnetic Waves' and it will be given on Tuesday, 9th March, 1971. In view of Professor Barlow's long association with University College London where he is Emeritus Professor of Electrical Engineering, it is very appropriate that the College should have agreed that the lecture may be given in its Engineering Lecture Theatre.

The Inaugural Clerk Maxwell Memorial Lecture was given in 1951 by the late Professor G. W. O. Howe in the old Maxwell Lecture Theatre of the Cavendish Laboratory at Cambridge. Subsequent lectures have been given by Sir John Cockcroft, F.R.S. (1954), Sir Lawrence Bragg, F.R.S. (1957), Dr. Vladimir K. Zworykin (1959), Sir Gordon Radley (1964) and Dr. Maurice J. H. Ponte (1968). The original proposal for this series of lectures arose from the presidential address of Mr. L. H. Bedford in 1948 in which he related Maxwell's work to modern radio engineering practice.

Conference on Noise and Vibration Control for Industrialists

A Conference aimed at increasing awareness of noise and vibration problems and of means of assessment and solution among engineers, architects and consultants and others concerned with these problems in industry will be held from Wednesday, 14th to Friday, 16th April 1971 at the University of Wales Institute of Science and Technology, Cardiff. It is sponsored by Bath University of Technology, The Institution of Electronic and Radio Engineers, The Society of Environmental Engineering and The University of Wales Institute of Science and Technology. The Conference will include tutorial lectures and case history and discussion sessions. Topics covered will include criteria and standards, an introduction to measuring equipment, basic principles of noise control, noise and vibration sources, acoustic materials and hearing conservation programmes.

An exhibition will be held as part of the Conference, and will concentrate on measuring equipment and acoustic systems. A number of representative firms have been invited to take part in the exhibition.

The programme for the Conference is to be published in November. The Secretary of the Organising Committee is Mr. D. R. Hub, UWIST, Cathays Park, Cardiff (Tel. 022-26708) to whom all enquiries should be directed.

Membership Designations

Members are asked to note that, under the Bye-Laws of the Institution, only Corporate Members (Fellows and Members) are entitled to use designations indicating membership, as follows:

F.I.E.R.E. M.I.E.R.E.

Graduates and members in other grades should not use designatory letters after their name.

Corporate Members, who have been duly registered with the Council of Engineering Institutions, are also entitled to add the designation 'C.Eng.', before the letters indicating grade of membership. Honours, awards and academic qualifications should precede 'C.Eng.'; corporate membership of other institutions should follow it.

British National Committee on Ocean Engineering

The Council of Engineering Institutions has set up a Committee with the above title which will be concerned with representation of the United Kingdom on the Committee on Ocean Engineering of the World Federation of Engineering Organizations and on the Engineering Committee on Ocean Resources (ECOR) as well as with the co-ordination of professional engineering interests in ocean engineering in this country.

The Institution's representative on the Committee will be Mr. P. W. Warden (Member). Mr. Warden has served on several Institution Committees including the Programme and Papers Committee and organizing committees for conferences on oceanography and ocean technology, and was Chairman of the Committee for the 1966 Conference on Electronic Engineering and Oceanography. He is manager of the Environmental Sensors Division of the Plessey Company.

The Radio Trades Examination Board

The Council has appointed Mr. W. B. K. Ellis, B.Sc. (Member) as the Institution's Representative on The Radio Trades Examination Board. Mr. Ellis is head of the Electrical Engineering Department at Southgate Technical College. He succeeds Mr. J. C. Martin, M.A. (Member) who has served as I.E.R.E. Representative on the Board since 1961.

Members' Appointments

Mr. Z. Levin (Member 1960) has recently been appointed to the post of Chief Broadcast Engineer of the Engineering Services of the Ministry of Posts, Israel; he has been with the Ministry of Posts since 1949 holding appointments in both communications and broadcasting. Mr. Levin succeeds Mr. H. Langholz (Member) who had held the post since 1948 and has now retired.

U.D.C. 534. 88: 551.462:621. 37/8

Problems in Bathymetric Surveying Presented by Modern Trends in Shipbuilding

By

Rear Admiral G. S. RITCHIE,
C.B., D.S.C., F.R.I.C.S.†

An address given at the Conference on 'Electronic Engineering in Ocean Technology' in Swansea on 23rd September 1970.

A brief history of electronic methods in bathymetric surveying is given. Two developments now required are: (1) a system of sidescan and vertical scanning sonars linked through a computer to give a 400m wide swathe on a chart; (2) a reliable and accurate seabed pressure gauge to record tidal height readings over a period of at least one month.

Until the first World War hydrographic surveying was carried on practically without the use of electricity. The use of theodolite ashore and the sextant and station-pointers at sea had, over the period of one hundred years, been developed into a science in locating the position of a ship or a boat when sounding offshore. Thus whenever visibility was satisfactory there was no problem in relating the position of shoals, and the channels between them, to fixed and visible marks ashore. The coverage of the soundings themselves did not however measure up to a similar accuracy. Sounding with the lead and line in shallow water, or with a lead lowered on wire from a winch on the deck of a stationary ship in deeper water, was not an exact science, and in deep water nothing was known of the errors in depth introduced by the bowing out of the wire by sub-surface currents or drift of ship on the surface. But the greatest deficiency in lead sounding was the paucity of the depth measurements obtained as the boat or ship proceeded, so that fluctuations of the seabed could only be assessed from spot soundings, and dangers could be missed between one sounding and the next.

Development of Echo-sounding

The study of underwater acoustics by Admiralty scientists during World War I, in their attempts to develop the first method of locating enemy submarines, had resulted in a build-up of knowledge in this field; and when the new Admiralty Research Laboratory was opened at Teddington in 1921 the Acoustics Group turned their attention to the use of sound for measuring the depth of the sea. They recorded the time of travel of a sound signal transmitted from the hull of a ship until that sound was received as a returning echo from the seabed.

The earliest echo sounder produced by A.R.L. relied upon a sonic signal developed with an electro-magnetically controlled hammer striking a diaphragm in the ship's hull about every 3 seconds. Successful trials of this set took place in H.M. Surveying Ship *Kellett* in September 1923.

However, this set was limited in its depth range to about 200 metres and was useless beyond the limits of the continental shelf; whilst attempts to use the set in surveying motor boats failed because the screening between the transmitter and the receiver, which was achieved in the ship by placing these on opposite sides of the keel, was not similarly provided by the small wooden keel of a boat.

To make the sound signal more directional A.R.L. developed a magneto-striction sound impulse. And whereas the earlier audio-frequency model required the operator to rotate a depth-calibrated dial until he heard the maximum volume of returning signal before reading off the depth, in the new sets depth was recorded visually by the markings of a timed revolving stylus on iodized paper.

This new set was successfully tested in a surveying boat in the dredged channel approach to Sheerness in 1930, and in deep water beyond the limits of the continental shelf in H.M. Surveying Ship *Flinders* soon afterwards. Kelvin Hughes developed a production model of this set which was subsequently marketed world-wide and offers a fine example of a Defence Department contributing to a national industry.

The surveyor was now able to obtain a continuous line of depths along his pre-planned sounding lines, and when irregularities on the seabed were revealed interlines were run to examine the feature more closely.

Radio Position Fixing Techniques

Sounding, by employing electronics, had now reached a similar state of accuracy to that achieved by visual methods of ship location, but in the latter electronics still played no part, whilst low visibility in any form brought surveying work to a halt. I remember as a young surveying officer in the 1930s that we would bribe the Bosun with gin on days of marginal visibility to keep a leaking hose running past the windows of the cuddy through which our impatient Captain was anxiously peering in hope of better weather. As long as he believed that rain was still falling the sounding boats remained at the davit head and we young surveyors snug in the chartroom.

Surveying productivity could only take another major step forward if electronics were to be introduced

† Hydrographer of the Navy, Ministry of Defence (Navy), Taunton, Somerset.

into the field of ship fixing, and it took the Second World War to bring this about.

In 1942, Mr. Harvey Schwarz of the Decca Record Company placed before the Admiralty his proposals for fixing a ship at sea by electronic means.† One master station and two slave stations, widely placed ashore in positions of fixed location, enabled a vessel with a receiver and recording decometer dials to fix her position as frequently as she wished. The system relies upon phase comparison between the signal transmitted by the master and that sent by the two slaves which are triggered by the master signal. Such phase comparisons, read on the ship's bridge from the two decometers, may be plotted with reference to pre-drawn hyperbolae on the chart. The system was first used on D-Day during the Normandy invasion, and the thing that amazed me most on that longest day was our almost magical ability at that time to plot the position of our survey ship off Arromanches from the readings of two dials on our bridge without the need to take compass bearings of the charted buildings and other landmarks onshore, many of which were in flames.

This system was exhaustively tested in H.M. Surveying Ship *Franklin* in the Bristol Channel in 1945 and led to the first of many Decca Navigator chains which still continue to blossom out all over the world.

However, it was soon clear that the sea surveyor required his own transportable Decca chain, the stations of which he could place as he wished to give the best coverage of the area to be surveyed and thus give the highest accuracy required; and which, above all, would measure the direct distance on the earth's surface from the ship to each of the two slave stations. This latter characteristic enables the surveyor, who has to plot his range patterns onboard after he has fixed the positions of his slave stations, to use circles which are so easy to compute and draw when compared with hyperbolae.

Such a two-range system, the master station being on board ship, was developed by Decca for surveying purposes in the late 1940s, and later the Lambda system was developed with a lane identification facility. This was followed by Hi-Fix, a shorter range but more portable equipment enabling it to be used by smaller coastal surveying vessels working inshore.

So the sea surveyor now has an all-weather fixing system giving accuracies of about 20 metres at ranges of up to 200 miles from the slave stations. Although the standard of echo sounding machines is high, it is towards more efficient methods of searching the seabed from a ship to which the electronic engineer could now well turn his attention.

Future Echo-sounder Requirements

It is the advent of the deep draught tanker and bulk carrier which now throws a spotlight on this need for more knowledge of the seabed in those shallow areas of commercial importance which are largely unstable and insufficiently examined.

The two areas which readily come to mind are the Southern North Sea and the Malacca Strait, both of which are being re-surveyed at the present time and both of which will require further re-survey in the future. Both areas have seabeds of sand or silt under the influence of tidal streams of varying flow rates up to 1½ or 2 knots. In such areas longitudinal sandbanks, some of which break the surface of the sea at low water, exist with the navigable channels running between them, these banks themselves being subject to periodical lateral change. Along the channels transverse sandwaves tend to form, as ripples form on a sandy riverbed, but with much greater amplitudes, sometimes up to 10 metres in height. Little study has as yet been carried out on these sandwaves, which may change their amplitude with the varying speed of tidal streams between spring and neap tides, whilst they may also move along the channel.

In the North Sea area there is the additional danger of thousands of wrecks littering the seabed as a result of two world wars. Most of the larger wrecks have almost certainly been found and are charted, but some smaller obstructions may have escaped the extensive sonar searches, whilst others, previously covered by sand, may in due course become uncovered. Both the transverse sandwaves and small wrecks could be a danger to a deep-draught vessel which, in order to make the maximum financial return on its loaded voyage may be sailing through the shallow areas with but 1 or 2 metres beneath the keel.

If such vessels are to navigate regularly with so little under-keel clearance in these areas of unstable seabed then regular, time consuming, and expensive re-surveys will be required to monitor seabed changes. To facilitate such surveys a broader coverage than that provided by the vertical echo-sounder must be available. It is here that the electronic engineer can come to the surveyor's aid with some form of side-scanning sonar which will not only indicate the presence of a small wreck or the peak of a sandwave, but will measure its horizontal distance either side of track of the sounding vessel and the vertical depth of the shallowest portion of the obstruction beneath the surface of the sea. To achieve this a completely new

† Schwarz, H. F., 'The engineer in state and private enterprise', *The Radio and Electronic Engineer*, 39, No. 1, p. 6, January 1970. O'Brien, W. J., 'Radio navigational aids', *J. Brit. Instn Radio Engrs*, 7, No. 5, pp. 215–46, October 1947.

system based on modern sonar techniques is now essential.

Scanning sonars have revolutionized short range underwater detection and target classification due to their ability to present visually target shape and size. What one now visualizes, and I speak as a layman in the electronic world, is a system of sonars onboard a coastal survey vessel much as follows. For initial detection of wrecks and minor obstructions ahead of the survey ship the need arises for longer operational ranges and a wider scanning sector of about 600 metres and about 60° respectively. This facility would be used in the conventional horizontal scanning mode.

Then on either beam one might have vertical scanning sonars with a very narrow sect or scanning beam enabling range and angle of dip to be taken to the highest part of each obstruction encountered. An onboard computer would be required to convert angle of dip and slant range into horizontal range and vertical depth of the obstruction. Such equipment would be most useful if ranges up to 200 metres either side of the vessel could be covered. Of course the sideways vertical scanners would require to be stabilized. It has also been suggested that the computer would have to be fed with data enabling it to take into account the bend in the sonic path of the slant range due to sea temperature structure, but as this equipment would be designed for use in depths down to 200 metres only I question whether this additional complication would be necessary, but the conference paper by Haslett and Halliday† seems to indicate otherwise.

But certain it is that data processing, storage and visual presentation problems in such a system as I have proposed will doubtless prove no less complex than those involved with the sonar engineering.

It will not be sufficient to store the data on tape, the sea surveyor will require a wide-band visual presentation as the vessel covers the area, so that any doubtful dangers may be investigated on the spot. The surveyor will then be covering swathes of 300 to 400 metres wide, each swathe over-lapping by a small margin the previous swathe. The processing of these data to form a chart will then become more like plotting from air photographs than from single lines of soundings, although the single line of echo sounding beneath the keel will still be required as the most exact depth reference.

If we are successful in providing the complex system I have tried to describe, then the simile I have often used to describe a ship sounding as a ploughman turning a single and lonely furrow will be changed to that of a combine reaping broad swathes of the hydrographic harvest.

One of the great factors discouraging industry from making new and complex hydrographic and oceanographic instruments is the smallness of the potential market for such specialized equipment. In the case of the precise scanning sonar I believe the market could be more extensive than one at first believes. There is no similar equipment available for use in shallow water elsewhere, although it would be strange if the United States and Japan are not thinking along these lines. Many port and harbour authorities throughout the world involved in getting deep draught vessels safely into port through long, comparatively shallow and unstable channels, require such a surveying system if the channels are to be regularly, swiftly and economically monitored.

Led by Sweden, the Danish and German Hydrographers have tackled the problem of swathe sounding in a different way, either by having a number of side boats fitted with echo sounders keeping station either side of the vessel, or by having echo sounding sets in capsules towed from booms or carried in station-keeping floats somewhat on the principle of those used in Oropesa minesweeping.

Such a system appears to me to be unsatisfactory because the echo sounders must be numerous and very close together if no gaps on the seafloor are to be left. In addition any ship with boats in station or towing

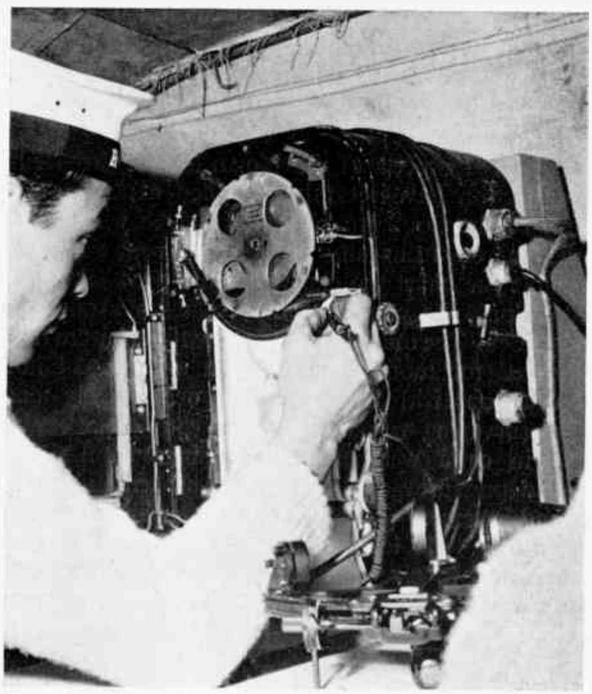


Fig. 1. Early type of echo-sounder recorder.

† Haslett, R. W. J. and Halliday, W., 'Rapid hydrographic surveying using mirror sonars', I.E.R.E. Conference Proceedings No. 9, pp. 129–50, 1970.

floats on the beam becomes a very clumsy and unmanoeuvrable vessel to operate in confined channels through which shipping is continually moving. I believe that we are right, therefore, to go for a more sophisticated system based on the use of sonar, and I am placing my trust in British Industry aided by the Committee of Marine Technology to produce it. Here again as in the case of the echo-sounder, the pioneering work of the Admiralty Research Laboratory in sector scanning sonars can be expected to show us the way. I was much encouraged by five papers and supporting films in the session on sonar systems and applications which showed that a number of scientists and electronic engineers are getting down to the solution of the problem.

Future Tide Gauge Requirements

However there is an associated hydrographic problem which I shall now describe, the first steps towards a solution of which we heard in the paper by Collar and Spencer.† The nautical chart shows the depth of water that a navigator should expect when the lowest astronomical tide situation exists. If he wishes to make use of higher tide to cross a shallow area or harbour bar he computes the height of tide for the required time from the tide tables and adds it to the depths shown on the charts. This should give him a reliable answer if he is near the coast where vertical tidal movement has been regularly recorded throughout at least one lunar cycle making tidal forecasting possible.

However there are shallow areas of strategic importance for deep draught vessels in the Southern North Sea which are over 30 miles from the nearest land and where good tidal forecasting is thus impossible.

The tidal experts of the North Sea Hydrographic Commission have recently compiled a new co-tidal chart of the Southern North Sea which will be published as an Admiralty Chart early next year. To compile the chart the experts used all available tidal data along the coasts of the countries around the North Sea. The only offshore data available were those observed for monthly periods on oil rigs, 24 hour observations with echo sounders occasionally made from surveying vessels at anchor over a flat seabed area in calm weather, and readings from Dutch seabed pressure gauges which have been laid on the seabed in recent years in connection with surveying operations. It is unfortunate that these pressure gauges fail to give readings of the desired accuracy and sometimes have a vertical drift of up to one metre in the period of a month's observations, whilst the stable platform provided by an oil rig is at present restricted to the area of oil discovery off the Norfolk Coast.

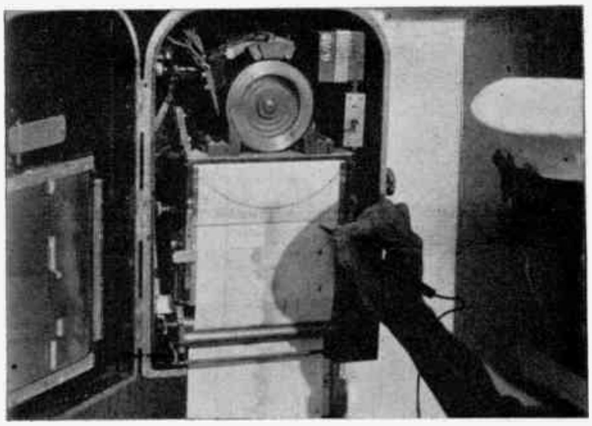


Fig. 2. Modern type echo-sounder recorder.

The co-tidal chart carries curves across the open sea through points having high or low water at the same time for each of twelve half hours before and after the time of high tide at a given standard port, and similar curves for height ratios on the tidal height at the standard port. Such a chart of the Southern North Sea compiled from the slender data available gives us a complex and quickly changing pattern across the area, with two amphidromic points in the open sea where there is no rise and fall of the tide, showing that the water in the North Sea behaves rather like water in a basin that is moved from side to side; the water laps up and down the sides while there is a point of no vertical motion at the centre.

Not only do we require far more month-long tidal observations in the open North Sea so that better co-tidal charts will enable the navigator to forecast more precisely the height of tide in critically shallow areas, but if his surveys are to be accurate then the hydrographic surveyor requires a gauge *in situ* in or near the open sea area of his work so that his soundings may be correctly reduced to that of the lowest astronomical tide.

Many types of seabed pressure gauges have been tried over the last 20 or 30 years, but none have really proved successful. What is needed is a gauge which can be laid on and recovered from the seabed and which will record and log tidal height readings accurately over a period of at least one month.

Here I think there is an assured market for a successful gauge for it is not only the surveyors who require such a gauge; once it is available the operators of deep draught vessels will require to establish such gauges permanently close to areas of critical navigation, and such gauges will be doubly useful for this

† Collar, P. G., and Spencer, R., 'A digitally recording offshore tide gauge', I.E.R.E. Conference Proceedings No. 19, pp. 341–52, 1970.

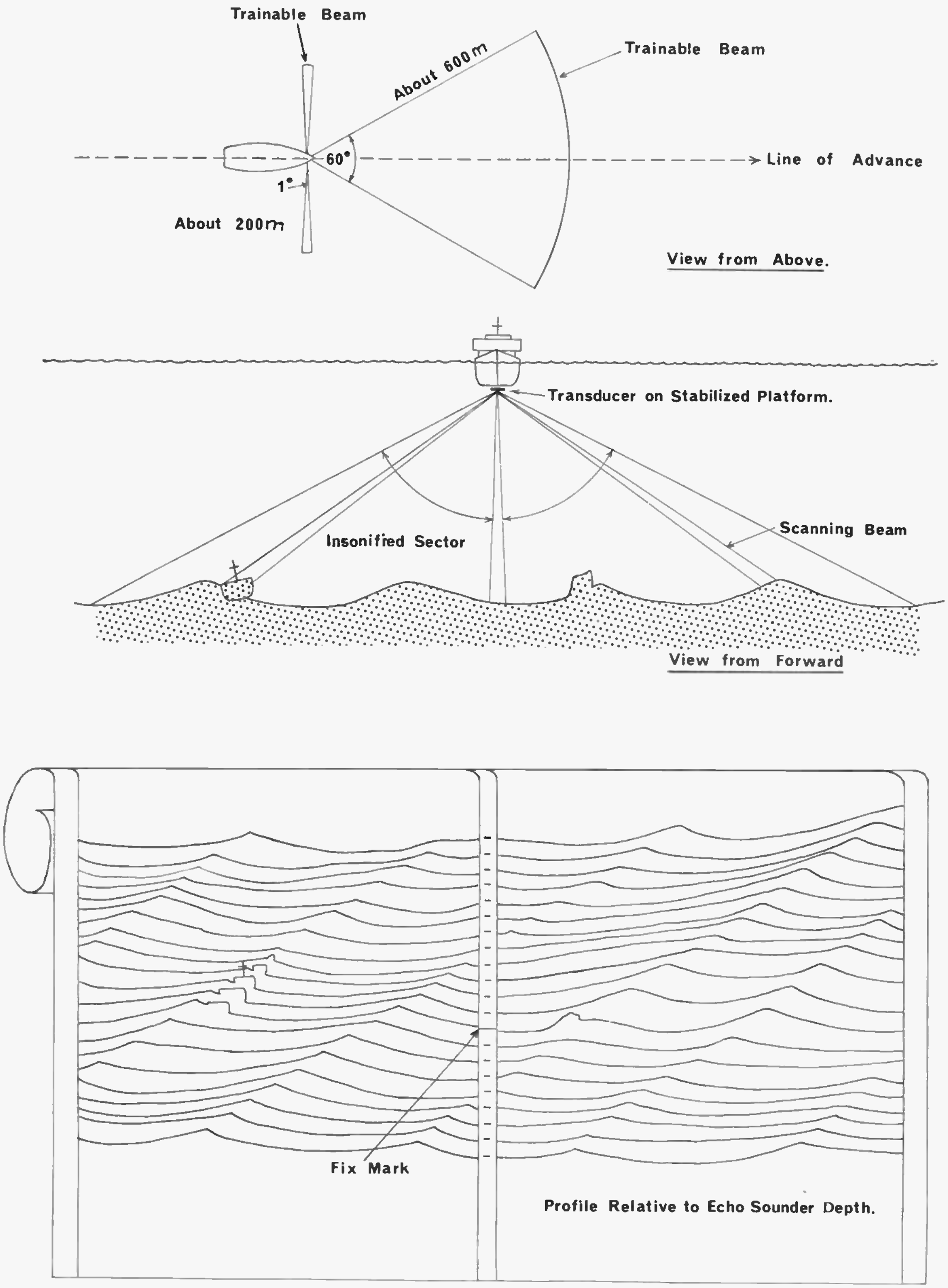


Fig. 3. The proposed surveying sonar system. The lower diagram represents a typical display print-out.

purpose if, by having a watching buoy attached to the gauge the shipmaster will be able to question the gauge by radio as to the present height of tide before he enters the critically shallow area. It is clear that forecasts of tidal heights may be affected materially by meteorological conditions existing at the time and that negative surges of 1 or 2 metres are possible when the tidal height is much less than the mariner could expect unless he has access to a gauge *in situ*.

Although there are many ways in which the electronic engineer is helping the sea surveyor in such fields as data logging, automated plotting, and the digitization of echo sounding, most of which have received adequate attention during the conference. I have dealt tonight with two major problems which have been presented to us by the new build of deep draught vessels which are now becoming commonplace.

It is difficult here to-night to describe the feeling of awe the sea surveyor feels, when occupied in sounding the seabed in the Southern North Sea with his inadequate tools, he sees the heavily laden giants majestically steaming towards the ports of the east coast of the United Kingdom or North West Europe, knowing as he does how comparatively ill-equipped his surveying vessel is to guarantee the small underkeel clearance upon which the tanker navigators are relying.

British hydrography has led the world for 175 years both in practice and in technological developments. It is accepted philosophy to-day that the United Kingdom cannot tackle all areas of scientific investigation, but only a selected few in which she can really succeed.

Surely our past record shows that the field of shallow water hydrographic surveying could be one such area; whilst the whole of the wider field of oceanography, including all forms of exploitation of the resources of the sea, must also be considered the rightful and proper field of development for the electronic engineers of this great maritime country surrounded by vast areas of continental shelf and almost boundless seas.

Manuscript received by the Institution on 24th September 1970. (*Address No.* 43).

OCEAN TECHNOLOGY CONFERENCE REPORT—II

The address by Rear Admiral Ritchie, printed above, was given at the Conference Dinner in College House at the University College of Swansea. It proved to be a successful and thought-provoking culmination to the Conference sessions in which sonar applications and the measurement of water movements had featured prominently, giving indications of possible ways of meeting the needs stated by the Hydrographer.

The I.E.R.E. President, Mr. Harvey F. Schwarz, was in the chair at the Dinner and introduced Admiral Ritchie in the toast to 'Authors of Papers'. In the course of his remarks Mr. Schwarz read the following message:

'Much regret that medical advisers prevented my joining you tonight as intended. I congratulate Institution on focusing attention on a subject of increasing importance to the world. Also congratulate you on your being joined by Admiral Ritchie who has made such an outstanding contribution to the subject of your convention, and to whom I send warm regards. Especially delighted that the Royal Navy has provided delegates with an opportunity of visiting H.M.S. *Hecate*. (*Signed*) MOUNTBATTEN OF BURMA'

The mention by Lord Mountbatten of H.M.S. *Hecate*, one of the Royal Navy's most modern survey ship's, refers to a notable feature of this Conference—the opportunity to inspect surveying and navigational equipment *in situ*. Similar facilities were provided by the R.R.S. *John Murray*, operated by the Natural Environment Research Council, and the M.Y. *Navigator*, owned by the Decca Navigator Company. A small exhibition of equipment described in some of the papers was arranged near the Conference lecture theatre, and here telemetry equipment and wave current meters were among the items shown.

The Conference was attended by 240 persons from a dozen countries and a list of the 39 papers presented was given in the September issue of *The Radio and Electronic Engineer*. Copies of the volume of papers may be obtained from the I.E.R.E. Publications Department, price £6 10s. per volume.

Designing for Automatic Testing: The Concept and General Approach

By

W. R. OGDEN,
C.Eng., M.I.E.R.E.,†

and

C. R. THOMAS†

Reprinted from the Proceedings of the Conference on 'Automatic Test Systems' held in Birmingham from 14th to 17th April 1970.

This paper deals with conceptual aspects of design which influence, and are influenced by, decisions to use automatic testing techniques in the factory and/or in the field. The early phasing of the influencing process is emphasized, and the importance of 'design-for-autotest' as a design parameter is discussed in relation to other parameters, and the case is made for its co-ordination into the varied aspects of the design requirements. These imply rationalizing design techniques, signal characteristics, equipment constructional forms and factory test organizations. Optimization of design must place a minimum brake on innovation. The applicability of automatic testing to the various stages of factory assembly, and the need to differentiate between the basic testing task and fault analysis by sub-routine, are established. Optimization of the differing requirements of factory and field call for design compromise.

1. Introduction

If we make the assumption that the main reason, and there are others, for using automatic test equipment is cost reduction, then it seems appropriate to examine all associated activities, with a view to maximizing this reduction to the benefit of all.

The decision to use A.T.E. cannot be made lightly nor in any arbitrary fashion. The capital investment, generally large, required by manufacturer and user alike makes it imperative that A.T.E. pays its way. Not everyone, however, even those most concerned with its application, are agreed on the best method of achieving this objective. Some argue that maximum results are obtained when manufacturer and customer use identical A.T.E. Others contend that, since the requirements of both are different, and they invariably are, it follows that the A.T.E. must be different. These arguments are undoubtedly extremes and cannot be assessed in isolation, and no arbitrary decision on the part of either manufacturer or user will satisfactorily achieve maximum cost reduction. The requirements of both must be considered, and a careful evaluation of the advantages and disadvantages of its use by both made at a time sufficiently early to optimize the benefit to both, so that their contribution to the prime equipment design specification gives adequate guidance to a project already burdened with feasibility, reliability, producibility, transportability, testability and maintainability.

To enable a worthwhile gain to be made, therefore, from a decision to use automatic testing techniques (and the potential gains can be considerable), a number of basic questions must be asked, the answers to which will determine the emphasis placed in the guidance given to the design engineer. The following are logical first questions:

1. Is there any economic gain to the manufacturer and/or user by the adoption of automatic testing for the equipment being considered?
2. Can a design approach be made which will allow either or both to benefit from the application of automatic testing techniques?
3. Can the equipment be designed to be used with available A.T.E. (central and peripheral), and available labour grades?

Each of these questions cannot of course be answered immediately by a straightforward Yes or No. Each generates additional questions which in turn produce compromises, all of which contribute to an optimum test and maintenance philosophy which makes demands on the main equipment design. Since the design of any complex equipment is in itself the result of compromise, the testing and maintenance philosophy evolved must undoubtedly influence the design, thus making a contribution to the overall approach, so that the final design optimizes the requirements of all, each recognizing the compromises and accepting the reasons for them.

It is in this area that project management is exercised to its fullest capacity, for not all decisions are easily arrived at, and not all participants will accept such decisions without argument. If optimum design is to be achieved, however, each parameter must be looked at in relation to the others and every decision reached should reflect the objective nature of its assessment.

Any user and any manufacturer embarking on an A.T.E. capital investment programme must be assured that there is adequate gain to be made from such a

† Test Equipment Group, Ferranti Ltd., Edinburgh.

programme in the way of investment return. The acquisition of expensive test equipment requires effective and efficient use and this is best achieved when its capability is used to the full. More efficient use of A.T.E. can be achieved by rationalizing testing techniques, minimizing the use and cost of interface equipment, standardizing in software, etc., etc. There is no doubt, however, that the greatest contribution to the proper cost effective use of A.T.E. is to assess its usefulness at the beginning of a project and to include automatic testing as a design parameter.

Many questions can be asked concerning the relationship between the use of auto-testing techniques and the design of equipment. Why should its use influence design? How should it be influenced? By whom and with what objective: How important is it? These are some of the questions and if the answers are to have a useful and acceptable influence on the design they must be answered convincingly, logically and honestly.

2. Why should Equipment Design be Influenced by the Requirements of Automatic Testing?

The days are not long past when the only restraints placed on the design engineer, particularly in the avionics field, were size, weight and operational requirement. It came as some measure of surprise to him therefore that when production, testing and maintenance of his design were attempted, those responsible for these activities found it necessary to recommend changes, in some cases major ones, which indicated a somewhat less than complete approach to the overall concept than had hitherto been believed necessary. The more objective engineers recognized this deficiency, as also did the customer, and over a period evolved the need for appraisal at various stages in design, this being intended to allow the production, testing and maintenance organisations, manufacturers and customers alike, the opportunity of constructively criticizing the design in order to achieve a more acceptable product.

The days are now past when the design engineer's plea, 'If we were doing it again we wouldn't do it this way' was an acceptable argument for design deficiencies. It is now recognized that a successful design embraces more than just the satisfying of the functional parameters. The design engineer, however, cannot be expected to be an omniscient unity on whom should be placed the sole responsibility for producing an all embracing complex equipment, the entire product of his own uninfluenced ideas. Peculiar and unfamiliar requirements have entered his sphere of influence. Producibility, standardization of components, rationalization of test equipment, testability, reliability, maintainability and a later entrant, more difficult to recognize and achieve, more all embracing, more frightening—cost of ownership.

To many engineers and particularly the young engineer newly graduated and keen to make his mark in the innovative field, all these hitherto unknown factors constitute a bewildering and sometimes frustrating problem, which if not satisfactorily resolved will cause a severe inhibiting load on his innovative ability.

It is now recognized that specialists in associated fields must make a contribution to individual projects in order to achieve a design compatible with all the various requirements. It must be stated here that, too often, ruthless and efficient activity in some areas has weighted the priorities heavily towards a design where the maximum benefit was gained in those areas but at considerable expense in others. An excellent example is the familiar phrase 'We can't redesign it now—it would cost too much'. But did anyone assess the cost in manufacture, test and maintenance resulting from this generally unevaluated wholly biased decision? Here again the management team must evaluate the individual contenders for recognition and make decisions accordingly.

Manufacturers will vary, as will their products, in the relative importance given to the many aspects of the design requirement, but those who by proper evaluation achieve a balanced design will undoubtedly find their products more acceptable to the customer, who invariably pays for errors or deficiencies.

Referring again to the question heading this section let us visualize what will happen if design for automatic testing is considered an irrelevant requirement. Systems will continue to be designed where complex and expensive interface and peripheral equipment will be needed if automatic testing is to be achieved, others will be designed where no gain can be achieved at all. In some cases either by enlightened engineering or fortuitously, the requirements for automatic testing will be designed in. If design for automatic testing is assessed as a design parameter and considered at the conception stage there is no doubt that the requirement must influence the design.

The alternatives to the acceptance of the need to design for automatic testing are:

1. the denial of its relevance to the project;
2. the acceptance of the cost implications of some other test process;
3. the arbitrary acceptance of A.T.E. costs when proved desirable;
4. the introduction of the changes necessary to allow it to take place;
5. the dissatisfaction both of production test and the customer and their decrease in confidence in future design.

The curve shown in Fig. 1 is representative of an area of production activity. Over the years, despite a continued upward trend in labour rates, assembly costs have not increased alarmingly, thanks to improved production techniques, productivity agreements, etc. Some improvement is also attributable to designing for producibility.

For many reasons, however, testing costs have failed to show a similar pattern. Increased requirements to meet higher quality and environmental standards, higher accuracies, requiring costlier test equipment, increased complexity causing increased diagnostic and re-test times and requiring higher labour skills, have increased the cost of testing rather than decreased it. The need to reduce the cost of testing is emphasized by extrapolation of the trends shown in Fig. 1 and the disproportionate share of total production costs have made detailed examination of the reasons imperative. The reasons can be many and varied and it is not the purpose of this paper to discuss them, interesting and in some cases surprising though they are. There is no doubt, however, that an effective method of cost reduction is to design for efficient testing and in many cases designing with automatic testing in mind can compress the test time by at least 10/1. With cost gains of this magnitude achievable, designing-in the necessary compatibility with automatic testing principles and with, preferably, available hardware is logical and economical, always bearing in mind the impact in other areas of interest to ensure the maximum effectivity. Associated manufacturing problems in the use of A.T.E. will be discussed later.

We have already seen that the needs of a customer are not necessarily coincident with those of the factory. The reliability and maintainability parameters of an equipment in the environment of its use and upkeep, may give quite different priorities to the considerations on which the requirements for automatic testing are based, including the availability of A.T.E., his desire to acquire A.T.E. and his ability to use it effectively.

These two issues of field and factory must both have their cases clearly stated at the equipment conception stage and only a balanced view of both will achieve an optimized compromise in the design. If the design compromise is not balanced, the result can be higher than necessary test costs in the factory, or unreasonable maintenance costs in the field.

There seems no doubt therefore, that with relevance to both conditions of use, the most efficient utilization of A.T.E. with its potential gains to manufacturer and customer alike, will be achieved only if the initial equipment design is conceived with its use in mind.

3. How should Equipment Design be Influenced?

The implementation of any requirement in a design can be achieved by its specific and mandatory inclusion in the design specification or by the application of a company-based design discipline. The former can be as detailed as necessary to ensure compliance; the latter can be instructive or advisory depending on the design control organization and the personalities of the design team.

Obviously design engineers vary considerably in their capabilities, their attitudes, their demands for individual consideration or team integration. The design organization must therefore be flexible enough to allow full scope to the innovative talents of the

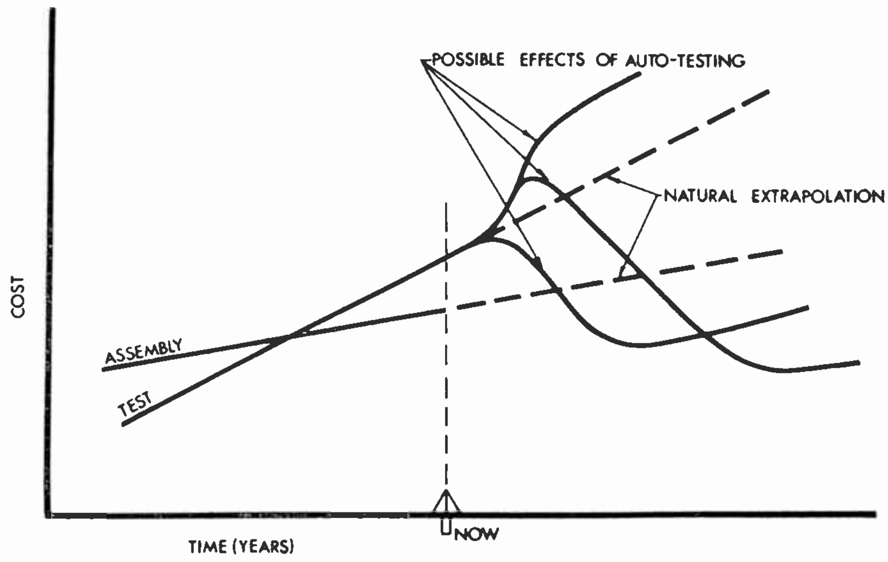


Fig. 1. Pattern of production costs.

individual and yet be specific enough to give proper guidance and control where this has a beneficial effect to the project.

The inclusion of detailed requirements in a design specification presents few problems from the designer's point of view. If the requirements are achievable, then their attainment is the designer's goal and will no doubt be met.

Under these circumstances, however, the customer now finds himself in a position where he has to be considerably more explicit than he would sometimes like to be, or indeed than is often possible. On the other hand he may be, and often is, willing to consider specialist guidance from the manufacturer and is more likely to accept this guidance if he is aware of the manufacturer's clearly defined design policy. Under such circumstances he may not feel it quite so essential to define his requirement in specific terms.

From the company's point of view it is generally possible to make some statement of design intention, although engineers are in many cases justifiably loth to be too definitive too early.

Where automatic testing needs are being assessed there is considerable merit in posing, at an early stage in the project, guideline questions which the designer should be able to answer, with help if necessary, to his own satisfaction.

The fact that all questions cannot be answered satisfactorily at the first time of asking should not preclude the necessity for an answer as soon as possible. They perform a useful reminding function, assist in maintaining an awareness of the detailed elements of the problem, and a re-appraisal of the questions and answers will often reflect a desirable change in priority or emphasis that might otherwise be overlooked.

The establishing of objective design appraisals will increase the probability that an optimum solution will be produced. If a particular guideline is disregarded then there should be a reasoned, presentable case to justify that rejection. If there is an inadequate or no reason, then it is the function of project management to take the appropriate remedial action.

Such an arrangement allows the design team, advised or guided by specialists where necessary, to make the decisions, to take the responsibility for having made them, and indeed to justify them.

Some typical questions posed to design engineers in a recent project are shown in Appendix 1. Obviously these are slanted towards a particular project with a particular application but they serve to illustrate the type of questions which hopefully lead the design engineer into thinking along the right lines. A subsequent appraisal of the answers against achievement can also be revealing.

A similar questionnaire relating to the test requirements on the same programme was prepared and is shown in Appendix 2.

It should be pointed out that in both cases the questions illustrated were only some in a more comprehensive questionnaire, and are shown only to suggest one method of ensuring design awareness of the test requirement.

It should be pointed out here that the number of times a design engineer will say 'You know, I hadn't thought of that', is ample justification for asking the questions.

The effective use of automatic test equipment is one of standardization wherever possible to a basic and efficient technique. Standardization of organization, of test philosophies, of design and manufacturing techniques of modular construction, of inspection procedures and test specifications take on added importance.

The need for stability of design cannot be over emphasized, in order to minimize the cost due to the re-programming necessary to cater for changed or additional parameters. Design instability can add considerably to the test cost in the need for diagnostic and re-test times, and the whole case for automatic testing can hinge on whether or not an item has a reasonable probability of being right first time at the test stage, for on this probability rests the necessity for detailed subroutine diagnostic testing or the provision of alternative back-up test equipment.

Right-first-time testing invariably involves a stable design, designed for producibility, and effective prior inspection.

So design engineers are influenced by many other considerations apart from auto testing, and the way in which they are influenced is largely dependent on company products, priorities, commitments and attitudes.

But irrespective of what the influencing factors are or who recommends them, the design engineer should make the decision; and if properly advised, he will ensure that the drawings produced will not only meet the requirements of efficient production and provide an adequate inspection record, but that they will be instrumental in the compilation of satisfactory testing programmes. Gentle persuasion, perhaps even a little pressure, at the right time is all that is usually necessary to ensure that the minimum conversion work is required for programme compilation in a usable form. But however the influence is applied, the problems must be fully considered and the gains and penalties assessed.

Automatic testing must always be an intended, assessed and planned activity.

4. By Whom should it be Influenced?

We have seen that, where knowledgeable in his requirements, a customer can have a considerable influence on the design by making his maintenance requirements known. In general, however, particularly where projects are complex and innovative, he may be unable to do this in sufficient detail to ensure adequate compliance. Under these circumstances he will expect guidance from the design team in the implementation of a general, stated maintenance philosophy. In such cases the design engineer has the responsibility for interpreting the broad wishes of the customer. In addition, the requirements of production testing in the factory must also be catered for and if automatic test equipment is in use or contemplated, then careful thought must be given to ensuring that the design allows maximum benefit to both. With the design engineer's efforts in the main concentrated on satisfying the functional requirements, and rightly so, with all its problems, it is unlikely that even with the best of intentions he can give sufficient thought to the requirements of test and maintenance. He must rely on specialist services in many areas, like component standardization, reliability, new materials and production techniques, etc.

The utilization of these specialist services, the dovetailing of the advisory and the executive tasks, is largely a matter of company policy reflected in project management. It is becoming increasingly obvious, where companies and their customers are embarking on an automatic test programme, with its high capital investment, or with the declared objective of test cost reduction, that the implementation of this policy cannot be contained within a single project, nor be treated as a hoped-for achievement.

Projects tend to be inward looking, dealing justifiably with the particular problems contained within the project. The solution to a functional requirement may come as a welcome relief to the design engineer but it may present a greater problem in manufacture, production test or maintenance.

A specialist group whose major task is ensuring that a company's test philosophy is implemented, and whose representative is seconded into a project at its conception, will make a considerable contribution to satisfying both production test requirements and those of the customer.

Such an engineer must have a thorough knowledge of test equipment and testing techniques, be capable of making recommendations which will benefit the project, and act among other things as an adviser to production test. To accomplish this he must gain the confidence and respect of the project and have a clearly defined place in the company structure. Appendix 3 gives some of the terms of reference which seem to be appropriate to this task. We have no doubt there are others.

Such a Test Engineering group can integrate the total testing requirements of a company, however diversified its products, so that maximum testing is achieved at minimum cost. This can only be done, however, where Test Engineering influences the design of the equipment, for maximum benefit can never be achieved unless the decision to design for automatic testing is made consciously and not achieved by fortuitous chance or unfortunate default. Such a group, with its broad-based project involvement can, with success, evaluate how new techniques evolved for one project can, with benefit, influence another so that automatic test equipment can be used more effectively and more efficiently.

5. The Relevance to Different Levels of Testing

In the factory the manufacturer starts, broadly speaking, with components, and finishes with black boxes connected together into systems. At any stage in manufacture and for many reasons, a component, module, sub-assembly black box or system may be defective. The manufacturer must, of necessity, inspect and test at various stages in the production cycle in order to ensure the maintenance of quality standards and fault free assembly. He must also carry out verification testing of the complete system to satisfy himself and his customer that the product achieves the correct performance and quality. If spares have to be provided the manufacturer must also ensure interchangeability, both mechanical and electrical. It is axiomatic that the later a defect is discovered in the production cycle, the more costly is its rectification. Defects must therefore be discovered as early as possible, rectified as quickly as possible and re-tested as cheaply as possible. A policy intended to implement these requirements is, of course, aiming at right-first-time assembly, and it is certainly not the intention to discuss its achievement in this paper. But a major cost in any defect investigation is undoubtedly diagnosis and re-test. The use of A.T.E. can substantially cut these costs provided the unit under test is designed with this purpose in view. This necessitates the provisioning of adequate verification and diagnostic test points in such a way as to assist logical fault deduction. In addition equipment designed with logical functional blocks can be performance tested much more readily, and troubleshooting is considerably reduced; and the smaller the number of functional blocks the easier is the task. Implementing this concept on a purely test basis would almost invariably reduce the equipment to its most easily tested constituents, and might well in some cases bring the task within the viable throw-away cost. The penalties of such a reduction, however, may be unacceptable in the areas of reliability, interchangeability, and increases in cost elsewhere.

There is obviously a point where the overall gain is maximized and the penalties acceptable and careful evaluation should help to approach this point.

The problems of interface between main equipment and automatic test equipment must not be overlooked at the design stage. If automatic test equipment is to be used effectively, and with flexibility it must not result in complex and costly interface equipment. Test Engineering should be able to recommend and possibly design, interfaces which will result in the most cost-effective approach. In this area there is considerable room for standardization not only within a company but within the U.K., if not wider.

Where equipment, for any reason, requires modification the need to minimize the impact on the automatic test equipment must be clearly understood. In many cases with proper and careful consideration by the design team, costly and sometimes unnecessary changes in automatic test equipment or software can be avoided. In this area again the guidance of Test Engineering can play a useful role.

If we now look at the user side of the problem we see it somewhat in reverse. Hopefully he starts with a system satisfactory in performance and quality, and his aim is to achieve maximum operating time with minimum maintenance cost and an acceptable performance.

Where lack of reliability in performance, as in avionics, can be costly in many ways, it is desirable to be able to monitor the performance either continuously or at the latest point in time before actual use. Built-in test equipment was evolved mainly as a means of increasing user confidence in successful performance. It can however with considerable advantage be effectively integrated into the automatic testing techniques necessary to test performance at lower assembly servicing levels as well as to provide diagnostic capability. This capability must always be viewed, however, in the light of the user's maintenance philosophy and the facility afforded should be no more than is necessary to achieve the desired result at the level of servicing intended. This additional capability should only be provided if there is no unacceptable penalty in reliability or other parameter.

Design engineers must always be wary of providing a system which is incompatible with the user's philosophy, his existing automatic test equipment, or his access to the necessary supporting software.

These then are some of the problems where the use of automatic test equipment at different levels of test and servicing should influence the design. There are of course others, such as skill levels, provision of print-out, repeatability-independence, etc., which must be considered.

6. Co-ordination with other Design Requirements

We have seen that design decisions are essentially compromises arrived at after assessment of the relative importance of all parameters bearing on a requirement. This compromise will not be correct if the balance is wholly or partly biased in any direction. This conference is concerned with one specialized area in which many compromises have been shown to be necessary within that area alone. There are many similar areas of specialized interest and we make no apology for introducing here other parameters which reinforce, oppose or modify the idealized case for automatic testing.

Briefly, a designer ought at some stage to be able to answer the following questions, sometimes to his own and/or his company's satisfaction, always to the customer's:

(a) Is it effective in meeting its operational requirement?

(b) Is it reliable?

(c) Is it producible?

(d) Is it testable?

(e) Has interchangeability been achieved?

(f) Is it maintainable?

If he can truthfully answer 'Yes' to all these questions and in addition assure himself and others that it could not have been done better or cheaper, then he is indeed entitled to the credit that such a performance merits.

It is unlikely, however, that all designers will give a fully balanced assessment of the importance of the above parameters. Additionally, a specialist concerned with emphasizing his own particular aspect, will almost certainly tend to over-emphasize his case.

The logical approach to improving the probability of achieving a proper balance of the invariably conflicting requirements listed above, is to consider them as early parameters for objective assessment. System design is the phase when the priorities should be properly allocated, and the targets set internally and externally for the broadly defined elements of the intended design. Future evaluations of achievement against intention will determine whether or not the balance was correctly assessed.

7. Conclusions

To sum up there is no doubt that, in order to achieve maximum benefit from the use of A.T.E., the design team must evaluate its usefulness in both factory and field. Any decision to use automatic test equipment must be accompanied by the conscious inclusion, in the design, of the requirements for its use, and the proper emphasis given to the influence that such a decision has on the design. It should not be over emphasized, but evaluated as part of the design specification, and as a sensible, relevant parameter.

If this paper has tended to be aggressive in its affirmation that designing for automatic testing is a necessity, it is because the authors are convinced that unless a firm rational approach is made to the problem of its many uses right from the initial project assessment, then automatic test equipment will not provide the benefits, either to company or customer, that are undoubtedly possible.

There is little doubt that automatic testing techniques will continue to improve, and the apprehension which presently exists about their cost effective use will be dispelled. The time scale in which this will be achieved can be considerably shortened, if and when project teams can convince the users, internal and external, that the needs of testing and maintenance have received the design attention which their cost demands.

8. Appendix 1: Suggested Design Checklist for Field Maintainability and Factory and Field Testability

1. Has the field servicing philosphy applicable to the equipment been determined in consultation with the customer?
2. Has the equipment design been conceived in a way which gives due consideration to the customer's servicing policy?
3. Where accessibility is necessary for any servicing operation, has it been reasonably provided?
4. Has the requirement for tools (and especially uncommon tools) for maintaining the equipment been minimized?
5. Has built-in test been conceived to assist in the matter of field maintainability?
6. Has there been any sacrifice of maintainability by the introduction of advanced or uncommon techniques? If so, is it justified or improvable?
7. Has there been any sacrifice of maintainability due to cost or producibility considerations? If so, is it acceptable or justifiable?
8. Have test point facilities been provided on a rational basis for the equipment as a whole, and for its constituent parts? (The method of incorporation of monitor/test points will obviously depend on the level of servicing at which they are intended to operate. They must be readily accessible at that level.)
9. Has an assessment been made of probable average repair times, with differentiation between diagnostic, repair and confirmatory testing aspects? Are they reasonable or improvable, as a whole or in particular areas?
10. Has consideration been given to the economics of servicing by throw-away modules (based on theoretical reliability, scrap cost, etc.)? (Throw-away-ability is only reasonable if the isolation and analysis aspects of unserviceability cost little in time and equipment. Self-evident unserviceability or designed-in diagnostic assistance is necessary to get full value from the throw-away concept.)
11. Has there been a signal rationalization task carried out, to limit the signal types, ranges and scale-factors, where relevant and advantageous?
12. Have the areas of divergence between 'factory' and 'field' testing been defined? Has any attempt been made to decrease this divergence in the interests of factory or customer?
13. Has the equipment been designed to require the minimum of specially-designed test equipment?
14. Does the equipment lend itself to programmed testing? Has it then been conceived to be so tested in the factory or field? Is automatic programmed testing advantageous in factory or field?
15. Do the drawings for cable-forms and cable assemblies include provision for easy compilation of a test program to allow automatic cable testing to be carried out?
16. Are signals which are transferred between units and between modules, monitorable in a meaningful way?
17. Does the relevant documentation contribute effectively to the user's capability of performing any necessary servicing in the field? Does it differentiate effectively between proof-testing and analytical testing?
18. Has module size been optimized and standardized (or rationalized) throughout the equipment?
19. Does the equipment design cater for, or require staged testing at factory pre-completion stages?
20. Has the need for 'select-on-test' components been eliminated? If not, have the implications in the factory test area and in the field been fully considered, and the difficulties minimized?
21. Where electrical series-monitoring is required, are facilities provided for the easy breaking-in and re-making of the circuit?

etc.

Final Question

Are you satisfied that the approach made in the design of the equipment gives full and proper regard for the good name of the Company and for the interests of the customer in the areas of testability and maintainability?

9. Appendix 2: Suggested Design Checklist for Test Equipment

1. Has the main equipment design been studied thoroughly?
2. Has the customer's testing philosophy been declared?
3. Has the factory testing philosophy been discussed and agreed?
4. Have the areas of divergence between factory test requirements and field test requirements been studied and narrowed where possible?
5. In recommending the test equipment considered necessary for field or factory testing, was it accepted that the main equipment design could be influenced by testability considerations?
6. Have the special test gear requirements been kept to the minimum by selecting available instruments wherever possible and economic?
7. Has a servicing and calibration policy been laid down for the test equipment in the field?
8. Does all the recommended test equipment have a capability which meets or exceeds the requirements of the equipment it must test?
9. Has a component rationalization policy for the project been adhered to in the test equipment?
10. Has accessibility been designed in, to give reasonable servicing facilities for the test equipment?
11. If programmed sequential testing is desirable, has consideration been given to the importance of time and skill in the test operations in factory and in the field, and, therefore, also to automatic test procedures? Is there clear differentiation between test modes and analytical modes of operation?
12. Have the test equipment size, weight, portability and cost (development and manufacture) been optimized to the advantage of the customer and ourselves?
13. Do the drawings for cableforms and cable assemblies include provision for ready compilation of a test program to allow automatic cable testing to be carried out?

etc.

10. Appendix 3: The Test Engineering Task

1. Formulate the company's testing philosophy with the object of cost-effective testing.
2. Provide specialist guidance to the project team in order to influence the design towards efficient testing.
3. Analyse the cost and effectiveness of current factory test equipment and procedures with a view to their improvement.
4. Carry out the evaluation and implementation of improved testing techniques in the interest of the company and the customer.
5. Assist in the preparation of production test procedures in co-operation with the Design Engineers.
6. Investigate and ensure corrective action from production test failures.
7. Estimate the test work content in each project including necessary skills, cost and test equipment.
8. Maintain an effective interface between Production Test and Design Engineers.
9. Discuss, advise and ensure the achievement of, the customer's maintenance policy.

Manuscript received by the Institution on 19th January 1970.
(Paper No. 1349/IC32)

U.D.C. 535.2:621.382

A Laser Machining System for Making Integrated Circuit Masks

By

M. S. QURESHI,
B.Sc., Ph.D.†

and

K. G. NICHOLS,
M.Sc., C.Eng., F.I.E.R.E.‡

An automatically controlled laser machining system for cutting integrated circuit masks in thin films deposited as fused silica substrates is described. A high-pressure helium-neon pulsed laser is used for cutting precision edges and a continuously operated carbon-dioxide laser is used for fast machining of larger areas in the mask. Details of the machining speeds and edge precisions obtained with each laser are given. The requirements for machining masks with the aid of both lasers are discussed. Results obtained in cutting test masks are presented and it is shown that considerable saving in machining times can be effected using the system of two lasers. The need for a vector generator, to control the movements of the work tables, as a peripheral to a computer is explained. Details of an experimental high-level mask-making computer language used with the above system are given.

1. Introduction

The use of lasers for machining integrated circuit masks has been described by several authors.[1,2,3] In principle, the method uses the high intensity focused output beam of the laser to machine an opaque thin film deposited on a transparent supporting substrate. The substrate is mounted on a work table which is moved in its own plane beneath the focused beam in order to machine the required pattern into the mask. The minimum width of a machined line is of the order of 5 to 10 μm with a sharpness of edge of the order of 1 μm. While this may be an adequate precision for a final mask§ for some applications, it is more usual to use the laser machined mask as the master for reduction and replication in a step and repeat camera. A reduction of linear dimensions by a factor of between 10 and 30 of the laser machined mask and a replication of its pattern into an array to give the final mask occurs in such a camera. The minimum line width and edge precision in the final mask is then comparable with that obtained by the conventional mask making process of repeated photo-reduction of a large scale master. In fact, optimally, the reduction factor used in the step and repeat camera is determined by the required minimum line width in the final mask. Quite apart from the question of minimum line width, the use of the laser machined mask as the final mask would require the replication of the pattern to be carried out on the laser machine. In general, this is very time-consuming and also requires a more sophisticated computer program to control the laser work table. However, much here depends on the step and repeat camera which is available. If, as is the case for the authors, the available step and repeat camera is manually operated, there is an incentive to carry out replication of the mask pattern automatically by the laser machine. This is to avoid the extreme tedium and possibility of human error inherent in the operation of a manual step and repeat camera. The mask making program (LASMASK), described later in the paper, has a facility for replication of the pattern at the laser machining stage. A solution to the problem of the manual step and repeat camera is to split the replication function into two stages. Thus replication of a pattern into a five by five array on the laser machine followed by a further replication of five by five in the step and repeat camera is certainly much less tedious than a twenty-five by twenty-five replication in the latter. This also has the advantage that the minimum line width and edge precision in the final mask is acceptable for all but the most precise work. Figure 1 is a photograph of a test pattern which has been replicated on the laser machine|| and Fig. 2 is the same pattern after replication in the step and repeat camera. Where a sophisticated automatic step and repeat camera is available, the virtue of replication on the laser machine is not so obvious. However, many small experimental integrated circuit fabrication laboratories do not have such cameras.

The need for narrow line widths and high precision of edge in integrated circuit masks has greatly influenced the choice of laser type used in mask making machines. This requirement for precision means that the lasers used in this application have low area machining rates with the result that it often takes a day or more to produce a set of masks for an integrated circuit.

† Department of Electronics, University of Southampton; now returned to the West Pakistan University of Engineering and Technology, Lahore.

‡ Department of Electronics, University of Southampton, Southampton, SO9 5NH.

§ A final mask is the one which is used directly to define areas on the photo-resist coated silicon slice.

|| This was done on the carbon-dioxide laser which is described later in the paper.

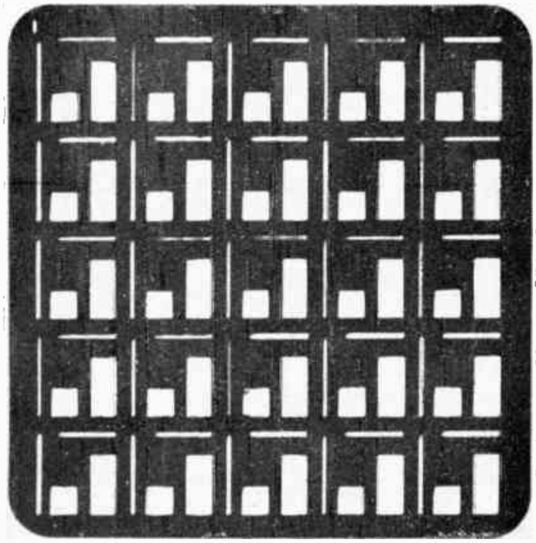


Fig. 1. Replicated test pattern machined by laser. (Overall size 4 cm × 4 cm).

2. Brief Description of System and Performance

In the sections which follow, a system which uses both a high pressure, pulsed, helium-neon laser and a continuous output carbon-dioxide laser is described. The helium-neon laser is used to cut the high precision edges required in the mask, and the carbon-dioxide laser, which has a much higher area machining rate but lower precision that the helium-neon laser, is used to *rough-out* the larger transparent areas of the mask.

The work tables of both machines are controlled by a digital vector generator. The input data to this generator are produced with the aid of a high-level mnemonic language (LASMASK) which is programmed for use on a small computer. This language is designed primarly for use in conjunction with experimental work of the kind carried out in a university integrated circuit fabrication facility, such as that at Southampton, and its structure and versatility reflect this requirement. The vector generator and the program language are discussed in later sections of the paper.

The system is used to machine masks in thin metal films on 5 cm × 5 cm (2 in × 2 in) fused silica substrates. The edge definition achieved in machined areas is about 1 µm and the minimum machined line width has been standardized at 20 µm. The time to machine a mask is very much a function of the pattern to be cut but is typically 15 minutes (see Sect. 5) for a mask of not too great a complexity. The computing time used by the LASMASK program is negligible but the time required to prepare input data manually for the program depends very much on the complexity of the mask. For example, the mask of Fig. 1 requires only eleven lines of simple data (see Fig. 11).

3. The Helium-Neon Laser

The high pressure, pulsed, helium-neon laser was developed in England by Boot and Clunie.[4] Several authors[5,6,7] have investigated its properties and mechanism of operation. Its application to the machining of thin films on glass substrates and to the production of integrated circuit masks in particular has also received attention.[3,6,8] While for these applications the precision of machining is inferior to that which can be achieved with electron beams, the ability to machine in normal atmospheric conditions and the transparency of the substrate to the machining beam are marked advantages in favour of the laser.

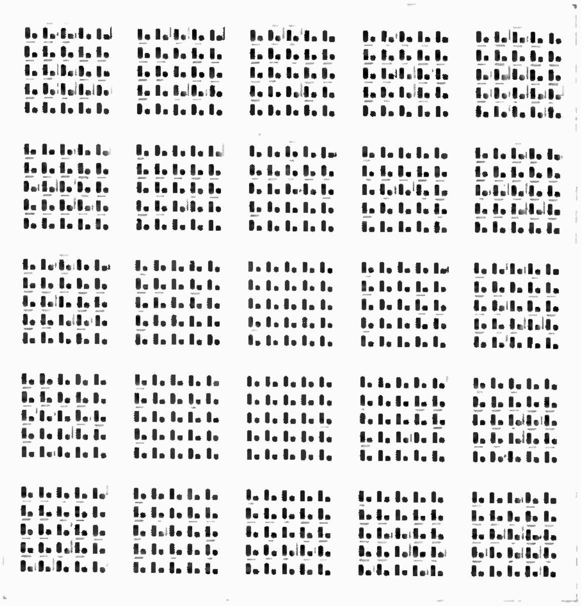


Fig. 2. Pattern of Fig. 1 replicated by 5 by 5 in step and repeat camera. (Overall size 4 cm × 4 cm).

The output of this laser occurs in pulses of about 1 µs duration. The wavelength of the output is 1·15 µm (near-infra-red) approximately and glass is still transparent at this wavelength. The peak power in the pulses can be made to exceed 200 W and the pulse repetition rate to exceed 2 kp/s. The power conversion efficiency is low,† about 0·1%, and a rather massive pulsed power supply is required. Figure 3 shows a schematic diagram of the laser tube and its power source. Pulses of about 50 kV are applied to external electrodes on the tube and discharge currents of 50 A or so flow in the tube during

† It is for this reason primarily that there is much interest currently in the Q-switched YAG laser[1] which has an effiicency greater than 10%.

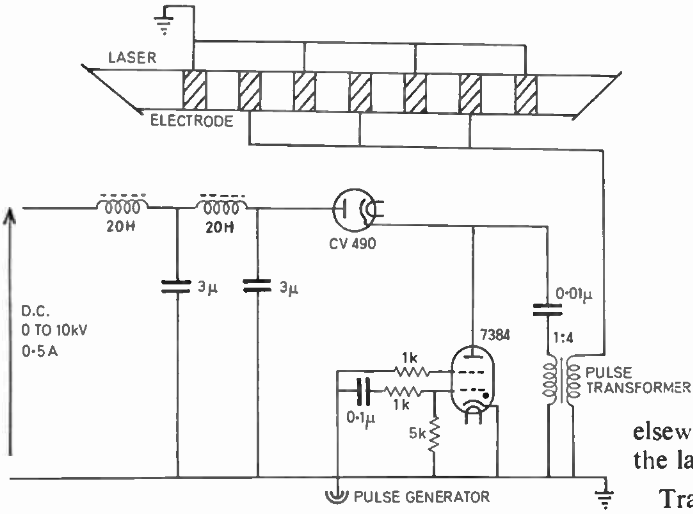


Fig. 3. Schematic of He–Ne laser tube and pulse power supply.

the pulses. The tube itself contains helium and neon to pressures of about 250 torr and 8 torr respectively.

In order to obtain a high output power with an acceptable reflector alignment tolerance, a near hemispherical optical cavity, external to the tube, is used. This is provided by multi-layer dielectric reflectors at each end of the tube, one reflector being concave and the other plane. The power is coupled out of the cavity through the plane reflector, which has a reduced reflectivity to allow for this. The output beam is focused on the film to be machined by a multiplet lens of focal length 2·5 cm and numerical aperture 0·95. The focused spot diameter can be as small as 3 μm but, for reliable and repeatable results, is more commonly arranged to be 10 μm. The peak power in the spot can be as high as 1 GW/cm^2 which is adequate for almost any machining purpose. This laser has a high precision, crossed roller bearing, work table with a rectilinearity of movement better than ± 2 μm. It is moved in its co-ordinate directions by micrometer lead screws which are in turn driven by electrical stepping motors. The movement of the table per step of a motor is about 3 μm (3·19 μm precisely). Details of this table and an associated interferometric measuring system have been described elsewhere.[9] Figure 4 shows a schematic diagram of the laser and work table.

Tracks are machined as a series of over-lapping spots. With a table movement step of 3 μm, a machined track of width 12 μm can be expected to have a sharpness of edge much better than 1 μm. In practice, an edge definition of about 1 μm is achieved. This can be seen from the single track of the photograph in Fig. 5. The maximum pulse repetition rate of the laser and/or the maximum stepping rate of the table limit the machining speed. In our case the limit is imposed by the latter at about 300 steps/second in order to achieve absolute reliability of table movement. This gives an area machining rate of about 1×10^{-4} cm^2/s. A rough estimate of the time required to machine a mask, using this laser only, can be obtained from this figure and by assuming that the area of the mask is 4 cm^2 and that 25% of this area is to be machined. This gives 10^4 s or approximately three hours. Often as many as seven masks are needed in the processing of a circuit and so it would take at least a day to machine such a set of masks. It was the need to reduce this time which led us to introduce the concept of a second, less precise but fast, laser for *roughing-out* much of the required area of a mask.

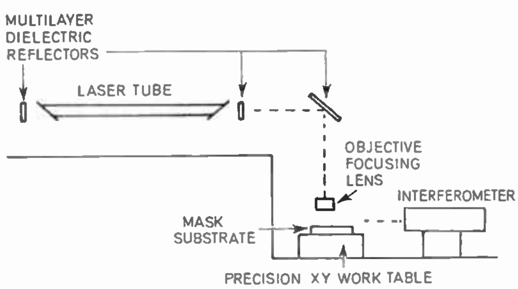


Fig. 4. Schematic of laser and work table.

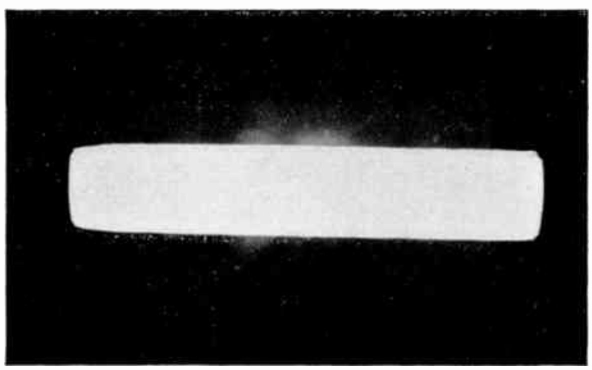


Fig. 5. Single 20 μm wide track machined in 40 nm thick antimony film with He–Ne laser.

4. The Carbon Dioxide Laser

The continuous output carbon-dioxide laser is more widely known and used than the pulsed helium-neon laser. It is capable of producing, according to size, gas mixture and flow rate, continuous output powers, at a wavelength of 10·6 μm, from a few watts to tens of kilowatts. A schematic diagram of the laser used in this work is shown in Fig. 6. The tube is 1·25 m long and is supplied with premixed gas containing 82% helium, 12% nitrogen and 6% carbon dioxide by pressure. The gas is flowed at a slow rate through the tube, the pressure within the tube being about 7 torr. The tube itself is cooled by water flowing in a surrounding jacket, as is also the metal cathode (grounded) which acts as an internal electrode within the gas. Additional cooling of the cathode is provided by a fan. The metal anode end of the tube, which again acts as an internal electrode, does not require cooling. A 12 kV d.c. supply, in conjunction with a ballast resistor, maintains a discharge current in the tube of up to 45 mA.

The optical cavity is semi-confocal and is formed by two concave gold-plated total reflectors within the gas tube. The output power is coupled out of the cavity through a small sodium chloride window in the centre of the plane reflector. A maximum output power at 10·6 μm of 18 W is available although all machining work is carried out with the power in the range from 5 to 10 W.

The output beam is focused on the film to be machined by a single-element, 5 cm focal length, potassium bromide lens. This produces a focused spot with a diameter in the range 100 to 300 μm. The arrangement of the laser and the machining table is similar to that of the helium-neon laser machine (Fig. 4) except that a less precise machining table is used. This table is also driven by lead screws and stepping motors with a basic step movement of about 60 μm (56·4 μm precisely). Machining rates of thin films under the continuous beam vary from 1 to 4 cm/s according to the laser power and the thin film material but most work is carried out at about 1 cm/s. Track widths also vary from 150 to 400 μm according to the laser power, film material and machining speed. The precision of edge of the tracks is of the order of 25 μm as can be seen from the photograph of a single track in Fig. 7. Area machining may be carried out by overlapping adjacent tracks with their centre lines separated by a distance of four table steps (assuming a track width of about 250 μm). At a machining speed of 1 cm/s, this gives an area machining rate of $2{\cdot}5 \times 10^{-2}$ cm^2/s which is approximately three hundred times greater than that obtained with the helium-neon laser.

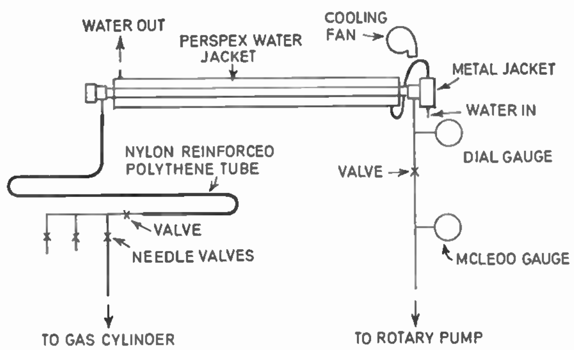


Fig. 6. Schematic of CO_2 laser.

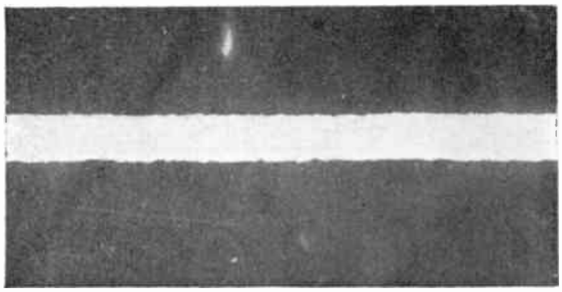


Fig. 7. Single 275 μm wide track machined in 40 nm thick antimony film with CO_2 laser.

It is necessary to remove the continuous beam when the work table is moving the substrate between areas to be machined and, of course, to recover it in order to start machining the new area. The removal or replacement of the beam energy must be achieved within the time the table takes to move one step, that it, within 6 ms at a table speed of 1 cm/s. In fact, a little longer time is available for removing the beam owing to a delay which occurs while the vector generator reads new data. It is not practical to switch the d.c. supply to the laser in the times available and the alternative of using a shutter in the path of the beam has been adopted. The shutter itself is a low inertia reflector which is actuated by a solenoid under control from the vector generator. It is placed just in front of the focus of the beam which it reflects harmlessly to an absorber. The shutter withdrawal time is less than 6 ms and its insertion time is about twice this figure.

5. Thin Film and Substrate Materials

The normal substrate material, glass, is opaque at the carbon-dioxide laser output wavelength of 10·6 μm. Materials which are transparent at this wavelength, e.g. sodium chloride, are generally unable to withstand the stresses set up by the temperature gradients resulting from machining and tend to fracture. At some power levels this occurs only in the presence of a metal film and not if the film is absent, thus indicating that conduction of heat from the film to the substrate is a significant process. Fused silica, although opaque at 10·6 μm, has been found suitable as a substrate material provided the laser output is carefully controlled. At one extreme, the power must be adequate to machine the film, and at the other it must not be so high as to cause undue damage to the substrate.

The metal thin film need only be thick enough to be substantially opaque to the visible light used in the step and repeat camera or to the ultra-violet light used to expose photo-resist. This usually implies a film thickness of about 100 nm but this does depend on the particular film. Vacuum-deposited chromium films are particularly attractive for this application because of their strong adherence to the substrate and have been used in this work. However, antimony and bismuth films, while not so strongly adherent to the substrate, give a wider margin between the two extremes of the carbon-dioxide output power, mentioned earlier, and are now used fairly extensively by us for making masks. Preliminary work with these metals suggests also that it may be possible to use them satisfactorily on borosilicate glass substrates. It is also possible to use polymer films[3] spun on the substrates as in the spinning of photo-resist films, but these films are thicker and we have found their machining with the carbon-dioxide laser less satisfactory.

6. Mask Machining with Both Lasers

Experience has shown that it is necessary to border the large areas which are going to be *roughed-out* with the carbon-dioxide laser with about eight four-step spaced 12 μm tracks machined with the helium-neon laser. That is, a total helium-neon laser machined border of about 100 μm. This figure is determined by the cumulative uncertainties of the precision of edges of the laser machined tracks and of the re-registration of the substrate from one work table to the other. Normally, the coarse machining is carried out first and, at the same time, a registration mark is machined on a corner of the film. This mark is used subsequently as an origin for the precision machining on the helium-neon laser. Relative parallelism of registration between the two work tables is achieved to within the required accuracy (25 μm in 2 cm) by the precision of manufacture of the two substrate holders and the substrate fixing arrangement on these holders.

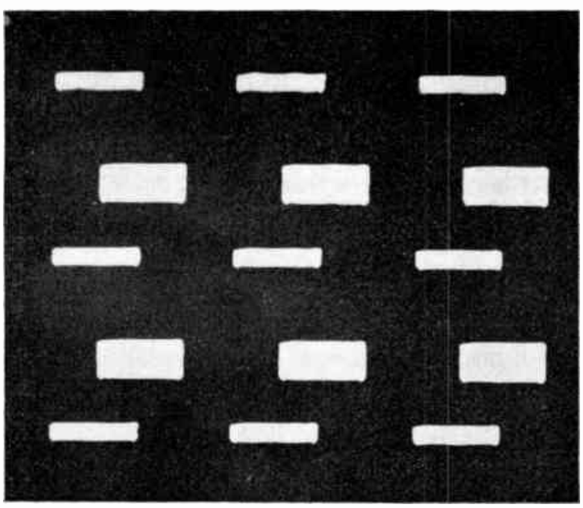


Fig. 8. Test pattern machined with CO_2 laser.

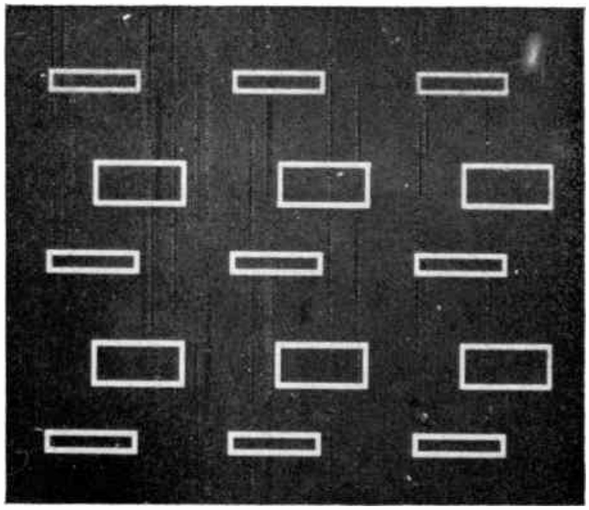


Fig. 9. Frame for pattern of Fig. 8 machined with He–Ne laser.

Figure 8 is a photograph of a replicated test pattern machined with the carbon-dioxide laser. The dimensions of the larger rectangles in this pattern are approximately 2 mm by 1 mm. Figure 9 shows the bordering frame only machined with the helium-neon laser. Each frame has an edge of eight 20 μm tracks spaced four steps apart, that is each edge is a rectangle of approximately 100 μm wide. Figure 10 shows the result of bordering the pattern of Fig. 8 by machining the frame of Fig. 9 into the same film to produce the final test mask. It is useful to compare the total machining time for this mask with the time it would take to machine the complete mask with the helium-neon laser alone. On the basis of the machining rates for the two lasers given earlier, these two figures are approximately (800+40) seconds and 2700 seconds.†

† Unfortunately, we are unable to verify this experimentally at the present time owing to the limited reading rate of the paper tape reader input to the vector generator. Because of the latter, reading data for the carbon-dioxide laser takes very much longer than the actual machining with the laser. This situation is shortly to be corrected with the acquisition of a 300 ch/s reader.

The combined laser system therefore offers a time improvement factor of approximately three, a figure which rapidly increases as the areas to be machined increase in size.

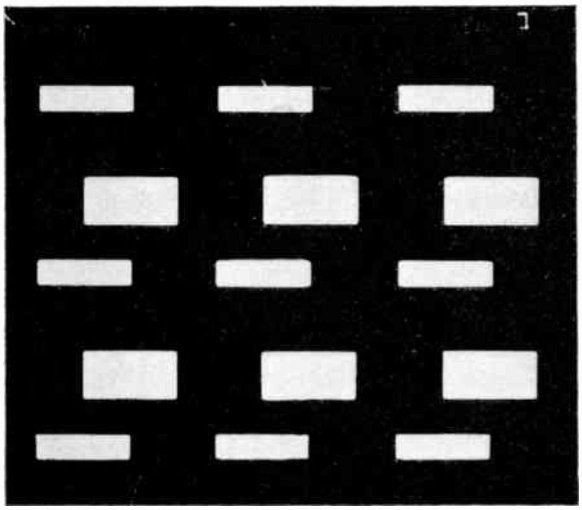


Fig. 10. Final pattern obtained with aid of both lasers.

In a subsequent 10 : 1 reduction of linear dimensions by a step and repeat camera, the 2 mm by 1 mm rectangles would be reduced in size to 200 μm by 100 μm. Extrapolation of this estimation to a rectangle of area (final mask) 50 μm by 50 μm shows that the combined laser system still performs the task in approximately half the time that would be taken in using the helium-neon laser alone. Some integrated circuit masks, particularly those for base diffusion and for interconnexion patterns, have about 80% of their total area in rectangles at least as large as this and it is these masks which have previously taken so much time to machine. It is therefore expected that the time to machine a set of masks with the combined laser system will be at least halved. Further, it is not unrealistic to expect a reduction in the width of the frame needed to be machined by the helium-neon laser with improvements of the carbon-dioxide system. An improved arrangement for re-inserting the shutter into the beam of the carbon-dioxide laser is currently being investigated and this is expected to improve the edge definition of the laser machining near the corners of rectangles. This alone might allow the width of the frame to be halved with the consequence of almost doubling the time reduction factor. Further improvements can be expected when higher power focusing systems for 10·6 μm wavelength become available and possibly also with developments of optical cavities for carbon-dioxide lasers.

7. The Digital Vector Generator

The stepping motors which drive the tables require to be supplied with one pulse for each step of the table. A polarity signal level is also required to indicate the direction of movement. If the tables were to be controlled more or less directly from a computer in an *off-line* mode, an input medium such as paper or magnetic tape would be required. Even for the simplest mask patterns, the very bulk of input data needed would make the quantity of paper or magnetic tape prohibitively large. Also, particularly with paper tape, the time used to read input data would approach that used in machining. *On-line* control directly from the computer would be feasible but the interface lines would be very *busy* dealing with the demands of the tables for data. This would reduce the availability of the computer for other functions.

The alternative is to provide pattern generation equipment, as an *on-* or *off-line* peripheral to the computer, to drive the tables. If this equipment were made too sophisticated in its functions, it would become a complex computer in its own right and its dependence on the parent computer would become very slight. The cost of a solution in these terms is not acceptable. What is required is a compromise between these two extremes with some pattern generation being performed by software within the parent computer and some by relatively simple hardware in the peripheral equipment. The digital vector generator used to control the laser tables is such a compromise.

The vector generator accepts as input, at present from paper tape, the end-point co-ordinates of a straight line which the laser is to machine. The generator is incremental in the sense that these co-ordinates represent changes from the present position of the table. The generator produces the co-ordinates of all intermediate points along (or near) the straight line and it supplies pulses and polarity signals to the stepping motors accordingly. It also provides beam pulses for the helium-neon laser and a beam shutter signal for the carbon-dioxide laser. Because of its limited functions and also because of its rigid input format, the logic system of the generator is relatively simple and inexpensive. However, even with its limited functions, it reduces the bulk of data required of the computer to manageable proportions and, if and when it is connected *on-line* to the latter, the demands for data will be so limited that the computer will be able to service these demands in interrupt mode. This means that other tasks being performed by the computer will only be minimally disturbed.

The format of the input data has been described in detail elsewhere.[9] It suffices to remark here that, although it is possible, the format is not suitable for direct key-board punching of input data tapes. Quite apart from the absence of any diagnostics in the vector generator to pick up operator punching errors, there is the tedium of punching data for, say, a rectangle as a large number of adjacent lines. The input for the vector generator is therefore prepared by programming the parent computer in a high level, user-oriented, mnemonic language. This language is discussed in the next section.

8. LASMASK: A Laser Mask Making Program

The computer used is a Honeywell DDP–516. This is a small data processor having a 16-bit word and, in our case, 12 k words of core store. The LASMASK program is coded in DAP–16, a symbolic form of the machine code of the machine and an assembly language. It occupies 3 k words of the core store including 1 k word for the storage of diagnostic message text. Additional store is required for generated data the amount of which depends on the complexity of the mask to be machined. This can, however, be provided from any suitable backing store. The input to the program is in terms of four-letter mnemonics as listed in Table 1.

A simple example of the use of the program is given in Fig. 11. This figure, which is largely self-explanatory with reference to Table 1, gives the data which was used to produce the mask of Fig. 1.

```
STRT
NLNE Ø 15Ø
STRT
BLNE 5Ø 5Ø 15Ø 5Ø
RECT 12Ø 7Ø 15Ø 1ØØ
BLNE 5Ø 75 5Ø 15Ø
RECT 8Ø 12Ø 15Ø 15Ø
AGIN 5
NLNE Ø 15Ø
AGIN 5
HALT
```

REPEAT 5 TIMES

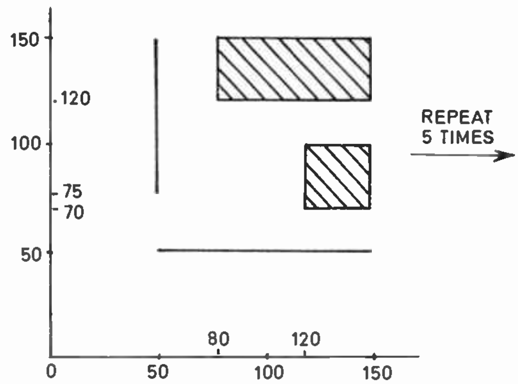


Fig. 11.
Example of use of LASMASK.

Table 1

Mnemonic	Meaning	Followed by data
NLNE	NO LiNE to be cut. (Move table to new position)	One pair of co-ordinates giving the point to which the table is to be moved.
BLNE	Begin to cut LiNE	Two pairs of co-ordinates giving the initial and final point of the line. A NLNE to the initial point is automatically generated.
CLNE	continue to cut LiNE	One pair of co-ordinates giving the end point of the continued line.
RECT	RECTangle	Two pairs of co-ordinates giving lower left-hand and upper right-hand corners of the rectangle.
(Rectangles are always scanned parallel to the long edge with a lateral separation of scan equal to four steps. Optionally this separation can be changed. The final scan always has the correct separation to complete the rectangle. A NLNE is automatically generated to get to the bottom left-hand corner. If necessary, a NLNE is generated to get to the top right-hand corner on completion of the rectangle.)		
STRT	StaRT	No data. It is a marker to indicate the point from which data is to be repeated.
AGIN	AGaIN	The number of times the preceding data is to be repeated from the appropriate STRT.
*XYZ	XYZ is the name of a user defined pattern in terms of NLNE, BLNE, CLNE and RECT instructions.	No data. The program finds the data for the pattern either from a library or from a special data tape.
LBRY	LiBRarY pattern	A list of *XYZ's being names of library patterns which will be required subsequently.
DEFN	DEFiNition of a pattern	*XYZ, name of pattern followed by instruction lines defining pattern.
FNSH	FiNiSH, end of library list, library tape or definitions	No data.
HALT	End of data	No data.

The input data may optionally be expressed in units of table steps or in microns for either machine. The format of the data is fairly free, erases and nuls being ignored everywhere and spaces ignored where non-significant. A repertoire of error messages largely protects the user against punching and format errors and against other invalid actions. The STRT, AGIN recursive facility can be nested to almost any depth and defined or library patterns can be nested within each other to a large degree. Each line of the basic input data requires three 16-bit words of storage, if required in backing store as mentioned previously.

9. Conclusions

The laser machining system described is likely to effect a reduction in mask machining time by a factor of at least two, and probably significantly more with further developments of the system, compared with other systems. As mask machining time is the principal limiting factor in present systems, a reduction of this time by a factor of even two is a significant development.

10. Acknowledgments

The authors are indebted to R. M. Langdon, D. A. Rolt, A.C. Burley and E. A. Adams, all of whom are no longer with the Department, for their pioneering work on the system. They are also indebted to R. L. Bassett for much detailed construction work. They wish to acknowledge the facilities provided by the University and the support provided by the Ministry of Technology, early in the project, under research contract PD/40/029/AT. Finally, they wish to thank Professor G. D. Sims for his constant encouragement.

11. References

1. Cohen, M. I., *et al.*, 'Laser machining of thin films and integrated circuits', *Bell Syst. Tech. J.*, **47**, pp. 385–405, 1968.
2. Jackson, T. M., *et al.*, 'Automated interconnection processes for semiconductor integrated circuit slices', 'Integrated Circuits', I.E.E. Conf. Publ. No. 30, pp. 181–191, 1967.
3. Jackson, T. M. and Rowe, T. J., 'Laser cuts masks to size, eliminating most errors', *Electronics*, **42**, pp. 81–83, January 20th, 1969.
4. Boot, H. A. H., *et al.*, 'Pulsed laser operation in a high-pressure helium neon mixture', *Nature*, **198**, pp. 773–774, 1963.
5. Boot, H. A. H., *et al.*, 'Micromachining with a pulsed gas laser', *Electronic Letters*, **2**, p. 1, 1966.
6. Rowe, T. J. and Moule, D. J., 'Laser micro-machining of thin metallic films', 'Electrical Methods of Machining and Forming', I.E.E. Conf. Publ. No. 38, pp. 19–26, 1968.
7. Langdon, R. M., *et al.*, 'The density of He(2^3S) atoms in a high-pressure pulsed helium-neon laser', *J. Phys. A.*, **2** (Ser. 2), pp. 611–619, 1969.
8. Rolt, D. A., 'A Laser Machining System', Ph.D. Thesis, University of Southampton, 1968.
9. Rolt, D. A. and Nichols, K. G., 'An interferometer controlled micropositioning table', *Proc. Internepcon*, Brighton, England, pp. 576–586, October 1968.

Manuscript first received by the Institution on 22nd June 1970 and in final form on 14th September 1970. (Paper No. 1350/IC33)

U.D.C. 621.396.7

The Planning and Commissioning of a Communications Satellite Earth Station and its Integration with Existing Telecommunications Systems

By

N. WHEATLEY,
C.Eng., M.I.E.R.E.†

The growth of the INTELSAT system of satellite communications is recalled and the considerations for on-site planning are discussed in the context of the construction of the earth station at Bahrain, Arabian Gulf in 1969. The relationship of other wideband systems in the Gulf is described, along with the resultant growth of traffic in an area where the demand had been suppressed due to the limitations of narrow-band high-frequency outlets.

1. Introduction

The history of satellite communications has been brief, fast-moving and technically exciting. The concept of artificial geostationary satellites was first proposed by A. C. Clarke in 1946 but had to await the arrival of precision rocketry before it could be tested. It has added a new dimension to telecommunications since the first active satellite, *Telstar I*, was put into orbit in July 1962. This was of an experimental nature and three large-aerial earth stations were built at Andover, U.S.A., Goonhilly Down, Cornwall, and Pleumeur-Bodou, France, to participate in tests which successfully demonstrated the feasibility of tracking medium altitude satellites in elliptical orbit. Later satellites, *Syncom II* and *Syncom III*, in 1963 proved the feasibility of such stations manoeuvring satellites placed into near-synchronous orbits in a near-equatorial plane to give a quasi-stationary position relative to a point on the earth's surface.

The question of whether the one-way transmission time of 270 milliseconds between two points on the earth via a satellite at 36 100 km (22 500 miles) distance would be acceptable for telephone conversation still had to be resolved. As a trial, Atlantic telephone cable circuits were rerouted via satellite working, without any prior public announcement, to test public reaction to long delay times in conversation. When this proved to be acceptable, the way was clear to develop a world-wide communications system via geostationary satellites.

In August 1964 the International Telecommunications Satellite Consortium—INTELSAT—was formed and has been responsible for the construction and placing in orbit of the *Intelsat* series of satellites. *Early Bird*, now known as *Intelsat I*, was the first of these in April 1965 since when the third generation has been put into global service, the latest being the eighth in the series launched during the last week of July 1970. This unfortunately failed to reach its planned orbit.

Orders for eight of the fourth generation have been placed and the first of this *Intelsat IV* series is due to be put into orbit during the first quarter of 1971.

2. Brief Description of Intelsat III

The satellite is essentially a communications repeater which employs two highly linear transponders. It is designed for 1200 two-way voice channels using a total 450 MHz bandwidth (225 MHz per transponder). The frequency arrangement of the transponders are shown in Fig. 1.

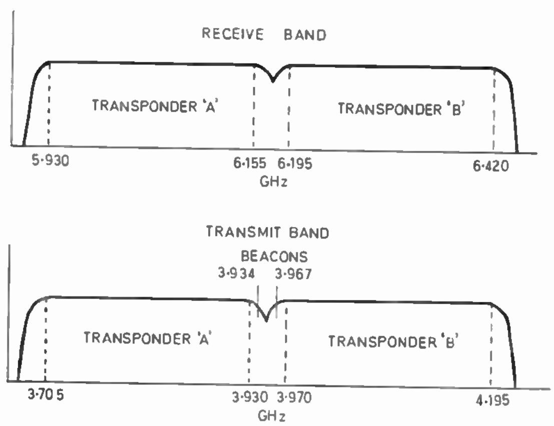


Fig. 1. *Intelsat III* satellite receive and transmit frequency bands.

Each transponder relays communications signals, receives commands and transmits telemetry information. Communications signals in the 6 GHz common carrier band are received and amplified without altering their modulation. Signal amplification occurs entirely at radio frequency through a two-stage tunnel diode amplifier and two travelling-

† Cable and Wireless Ltd., Bahrain, Arabian Gulf.

wave tubes. After a single frequency conversion, these signals are retransmitted in the 4 GHz common carrier band to earth stations. Command and telemetry signals are carried in the 6 and 4 GHz bands respectively. Each transponder has a command decoder and telemetry encoder. Commands are amplitude modulated tones received by the omnidirectional antenna. Telemetry information phase modulates each of two 4 GHz beacon signals which are used for tracking and identification by the earth stations.

3. International Telecommunications Satellite Consortium

INTELSAT is a consortium, comprising 76 countries at August 1970, each of which has a financial quota of varying degree based on its expected usage of the system, and is governed by two Agreements.

An Interim Agreement between Governments set out the general objectives and conditions to apply in the field of satellite communications. It established the Interim Communications Satellite Committee (ICSC) as the policy-making body and has appointed the United States firm, Communications Satellite Corporation, (COMSAT), as Manager for the space segment, i.e. for the procurement, launching and control of the satellite under instructions from the ICSC.

The second agreement is the Special Agreement which is between telecommunications entities and again relates to the space segment; it does not affect the ownership of earth stations which remain national responsibilities of the countries illuminating the satellite. Signatories, or groups of signatories, whose total quota is at least 1·5% of the cost of running the system may have one representative on the policy-making ICSC. This committee meets regularly at approximately two-monthly intervals and has standing sub-committees for technical, finance and contracts matters to assist it.

Both these agreements are essentially interim and the ICSC was required to render a report to each party to the Interim Agreement by 1st January 1969 containing the Committee's recommendation on definitive arrangements for the future, reflecting fully all shades of opinion of the members of the Consortium.

The Committee's Report to Governments was issued to all INTELSAT members, who considered it at a Plenipotentiary Conference held in February and March 1969. This Conference appointed a Preparatory Committee to resolve in an objective manner differences of views presented during the Conference and to prepare draft texts for consideration by the Conference when it re-convened.

The Resumed Plenipotentiary Conference, attended by 67 INTELSAT representatives and ITU and UN observers, re-convened in February 1970 to discuss the Preparatory Committee's Report. Differing views on the role of the United States as the major quota holder (52·82%) in INTELSAT, and COMSAT as Manager, emerged and an Intersessional Working Group (IWG) was established to meet during May and June 1970. The IWG is to prepare a single set of recommended texts based on a proposal submitted by Australia and Japan relating to the major issues of management arrangements, voting in the Governing Board, determination of investment shares and structure. The IWG was due to report its findings to the re-convened Conference on 8th September 1970 or, if it was unable to complete its work by then, not later than 90 days from that date.

Meanwhile the Interim Arrangements continue in force until superseded by the Definitive Arrangements.

4. Setting up of Earth Stations

It was against this earlier background that national telecommunications bodies—Government administrations or recognized private operating agencies authorized by those Governments—were required to decide on the timing and location of earth stations to utilize the potential of such satellites. There were two main considerations:

(a) To provide relief for the rapidly congesting submarine telephone cables across the Atlantic and Pacific Oceans and the South China Sea.

(b) To use the satellite system to provide wideband communications of high quality in areas hitherto served either by narrow-band submarine telegraph cables or by h.f. radio links, which are subject to the vagaries of the ionosphere.

It is the situation under (b) which is discussed here and in particular the Earth Station opened in Bahrain in the Arabian Gulf by Cable & Wireless Ltd. in July 1969; this was the first in the Middle East or Africa to operate to the INTELSAT system. The coverage of the INTELSAT Indian Ocean Region satellite is shown in Fig. 2 along with operational and planned earth stations of the whole system.

4.1. *Preliminary Studies*

Obviously before financial approval can be obtained for a project costing between £1·5 and £2 million, the economic viability must be gauged. In the case where the earth station is to augment telephone cable systems on high-density routes it is relatively simple to extrapolate normal traffic growth over wideband links.

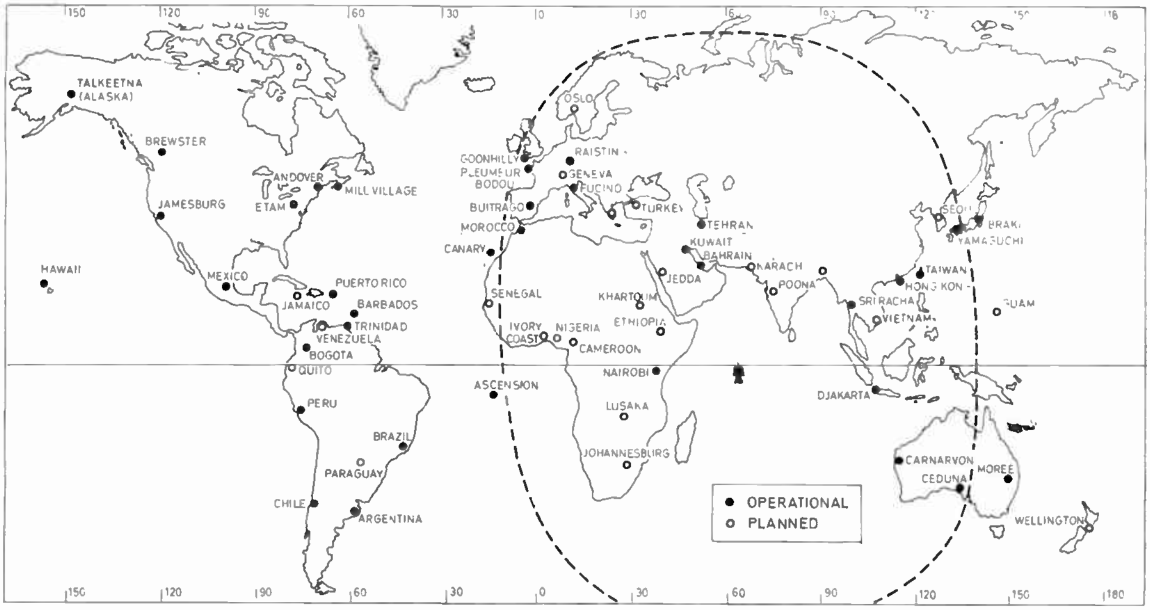


Fig. 2. INTELSAT system showing (a) coverage area of Indian Ocean Region satellite and (b) operational and planned earth stations for all Regions.

But where no such links exist it is necessary to make a close study of (i) business potential, (ii) political stability (where the communications carrier and prospective earth station owner is not a Government administration), (iii) estimation of suppressed demand on existing h.f. systems and market research into likely public reaction to an improved service—although it has been proved many times that a good communications link stimulates use beyond the planner's expectations—and (iv) the best physical location in the traffic catchment area. Also the implications of the chosen location on existing concessional agreements with adjoining states has to be considered, since the ownership of earth stations is rapidly becoming a status symbol in the telecommunications world.

The logical choice on all the above counts for the Arabian Gulf area was Bahrain, which has been a major commerce centre for over 100 years.

4.2. *Choice of Site*

The main factors for consideration at Bahrain were:

(a) The optimum location on Bahrain island to suit the angles of elevation (57·15°) and azimuth (153·53° East of North) to be subtended by the Indian Ocean Region satellite located over the equatorial point 62·5° East.

(b) The method of connexion to the inland telephone network. This was chosen to be by a 2 GHz microwave link.

(c) Possible interference from other terrestrial radio sources. Ideally, the site should be in electrically-quiet surroundings bounded by low hills to give protection from noise at angles below a few degrees.

Most of Bahrain is relatively flat desert and, as no such ideal site existed, extensive tests were necessary in early 1967 to determine the level of interference at received r.f. (3·7–4·2 GHz) and at i.f. (70 MHz). Extracted results are shown in Table 1. The main concern was the possibility of interference from the radars of ships plying the Arabian Gulf and also from a known planned 1·3 GHz surveillance radar, the 3rd harmonic of which would fall into the earth station receive band. A solution was reached with the radar operating authority, which provided 60 dB harmonic suppression filters, and the proposed site for the earth station was moved from the west to the east side of the island giving better shadow protection from the radar.

The result has been successful and no interference has been encountered to date.

Table 1

Bahrain Earth Station Site—Ras Abu Jarjur
Radio Noise Measurements 4th–12th July 1967
Corrected signal levels into wideband dipoles in dBW

Frequency spot MHz	Corrected signal level dBW	Remarks
71·0	−112	Arabic broadcast. Unidentified
75·0	−108	Multiplex harmonic
76·4	−110	Doha (Qatar) private r.t. link
79·2	−103	Television video, Dharhan, S. Arabia
88·5	−104	F.m. broadcast, Dharhan
95·5	−99	F.m. broadcast, Dharhan
3021	−83	Bahrain Airport radar type AR.1. Tx 'B'
3025	−114	Radar due east. 1 day 0050–0158LT only
3040	−116	As 3025. 7·5 rev/min
3065	−104	Radar due east. 5 rev/min
3075	−104	Radar due east. 7·5 rev/min
3090	−101	Radar bearing north. 7·5 rev/min

Secondary considerations at the chosen site were:

(d) The availability and reliability of electric power. The economics of on-site generation versus the public utility supply were examined closely over the 15 years amortized life-span of the station. The result was marginally in favour of the public supply, with provision of standby diesel generating sets to cover temporary public supply failure.

The normal method of mains distribution in remote areas of Bahrain is by overhead feeder, but in this case underground cable was stipulated within half a mile of the station perimeter to reduce the risk of electrical noise.

At the same time opportunity was taken, when planning the capital costs of this supply, to provide for power at another remote site at Jebel Dukhan, beyond the earth station, which was to house a tropospheric scatter station forming part of the area wideband network.

(e) Geological soil survey to determine load bearing capacity and electrical conductivity had to be done. Although the ground was known to be hard limestone-type rock beneath about 2 ft of sand, test bores were made to ensure no cavities existed beneath the dish antenna pedestal building. This was vital since the dish itself weighs some 250 tons and the pedestal roughly the same.

(f) Meteorological conditions relating to wind speed, rain, temperature, humidity and sand storms had to be studied and data statistically analysed. Temperature of exposed dish surfaces and sand storms merited much consideration before a final dish design was chosen. The possible corrosive effect of fumes from the Bahrain oil refinery several miles away had to be considered under certain wind directions.

(g) Finally, site accessibility for staff and supplies during and after installation plus a good water supply for building and domestic use are most important. Since it is rarely possible to expect staff to live on site under desert conditions, transportation for staff and heavy or wide loads in relation to road works, heights of power lines, etc., all needed detailed checking before the chosen site could be finally accepted.

The outcome was that a plot of land sufficient to accommodate technical buildings and two antenna structures (in case a second dish is required) at Ras Abu Jarjur, some 13 miles from the main town area, was leased from the Bahrain Government; and a contract for the provision of the station was signed with the Marconi Company.

5. Site Layout and Station Description

Figure 3 shows the site layout of the above station which is divided mainly into (i) technical and power buildings and (ii) dish antenna and its support structure.

5.1. *Technical and Power Buildings*

The main Technical Block houses the radio equipment and control console room, test equipment maintenance room, stores, electrical distribution and airconditioning plant rooms, offices and staff facilities. The Power Block contains the supply authority's distribution sub-station to convert the 11 kV incoming line to 440 V, three 300 kVA standby diesel alternators, two of which are capable of carrying the station load, and heavy stores and workshops.

A small building at the north edge of the site contains the 2 GHz duplicated radio equipment for the microwave link connecting the station via a 3 metre dish mounted 45 m (150 ft) high to the International Telephone Exchange in Manama town at the north of Bahrain island. The associated multiplex equipment is in the equipment room of the Technical Block. Cross-site trenches interconnect the

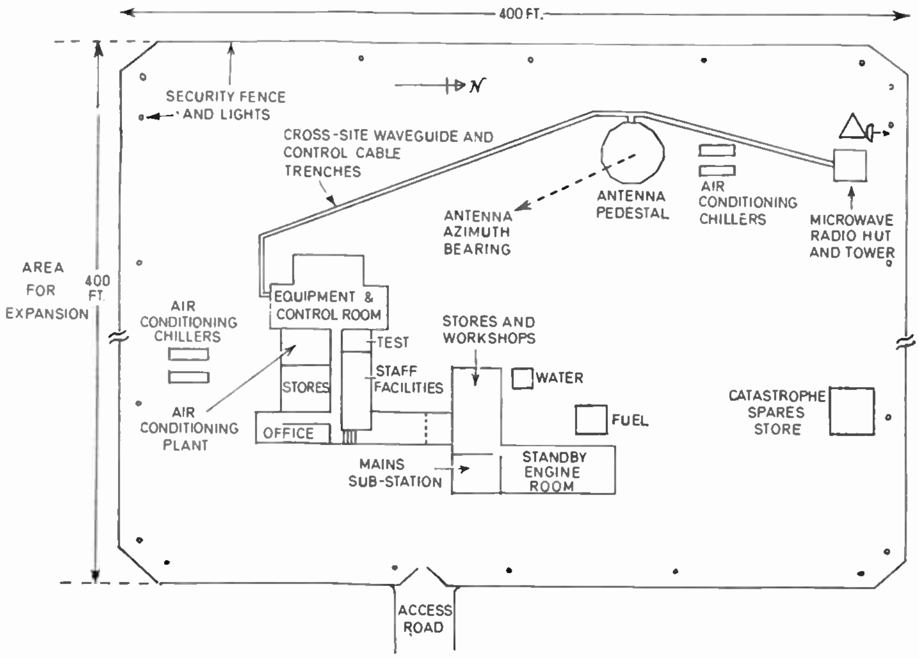


Fig. 3. Bahrain earth station—site layout.

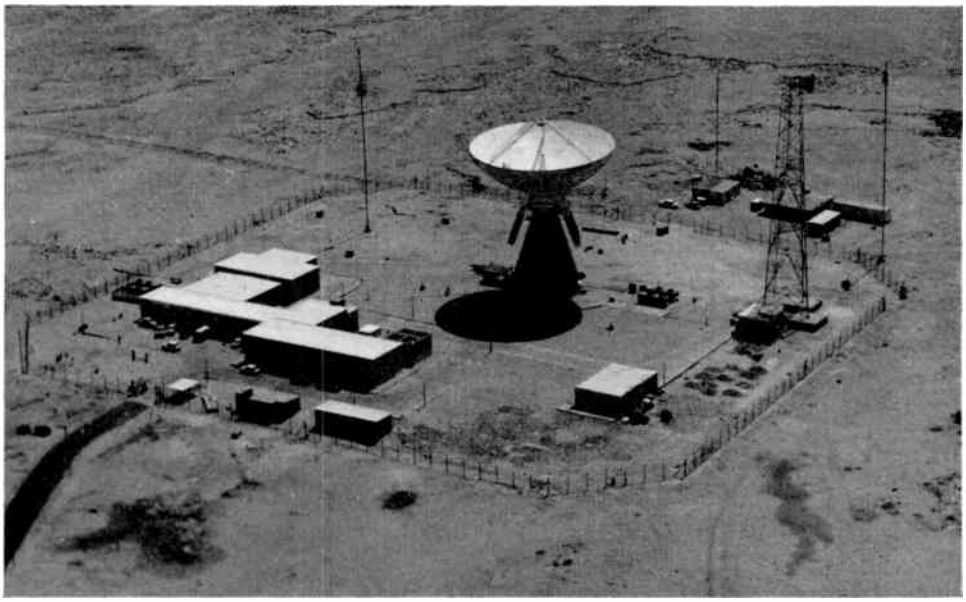


Fig. 4. Aerial view of Bahrain earth station towards the end of the installation phase.

separate buildings and are sealed from the elements to protect the waveguide runs and control cables.

Space has been provided to allow for up to double the equipment for future expansion in the event that a second antenna is required to illuminate another satellite simultaneously, e.g. the Atlantic Ocean 6°W *Intelsat III* or one of the *Intelsat IV* series as they come into service.

Building work commenced in April 1968 using a local contractor and a design prepared by the Chief Architect's Department of Cable and Wireless. The Technical Block is formed from prefabricated steel frame and blockwork structure and the essential equipment room was ready for occupation in September 1968. The dish pedestal was finished in November 1968 and there was a steady build-up of staff to about 70 in March and April 1969 when virtually all the essential equipment was on site. Installation work and testing finished towards the end of June and the station became operational on 14th July 1969—just over a year from the start of building.

Figure 4 is an aerial photograph of the station at the end of the installation phase.

5.2. *Dish Structure and Support Pedestal*

Figure 5 illustrates a cross-section of the dish and its support pedestal which is 12-sided in plan and conical in elevation with three floors. The first two storeys house main power supplies and servo control racks and the twin 5 kW transmitters, the output of which is fed to the feed horn at the centre of the dish.

The antenna is supported in an elevation-over-azimuth mode by a main central kingpost which has a lower oil-immersed plastic bearing to take the main weight. This bearing housing is built into the fabric of the building and is supported by 8 of the 24 macalloy rods also incorporated in the pedestal and pre-stressed during construction so that the whole of the weight of the dish is distributed throughout the pedestal from the top ring. An upper azimuth bearing, comprising six large plastic pads resting against a steel ring set into the building, takes the sideways thrust. Immediately behind the feed horn

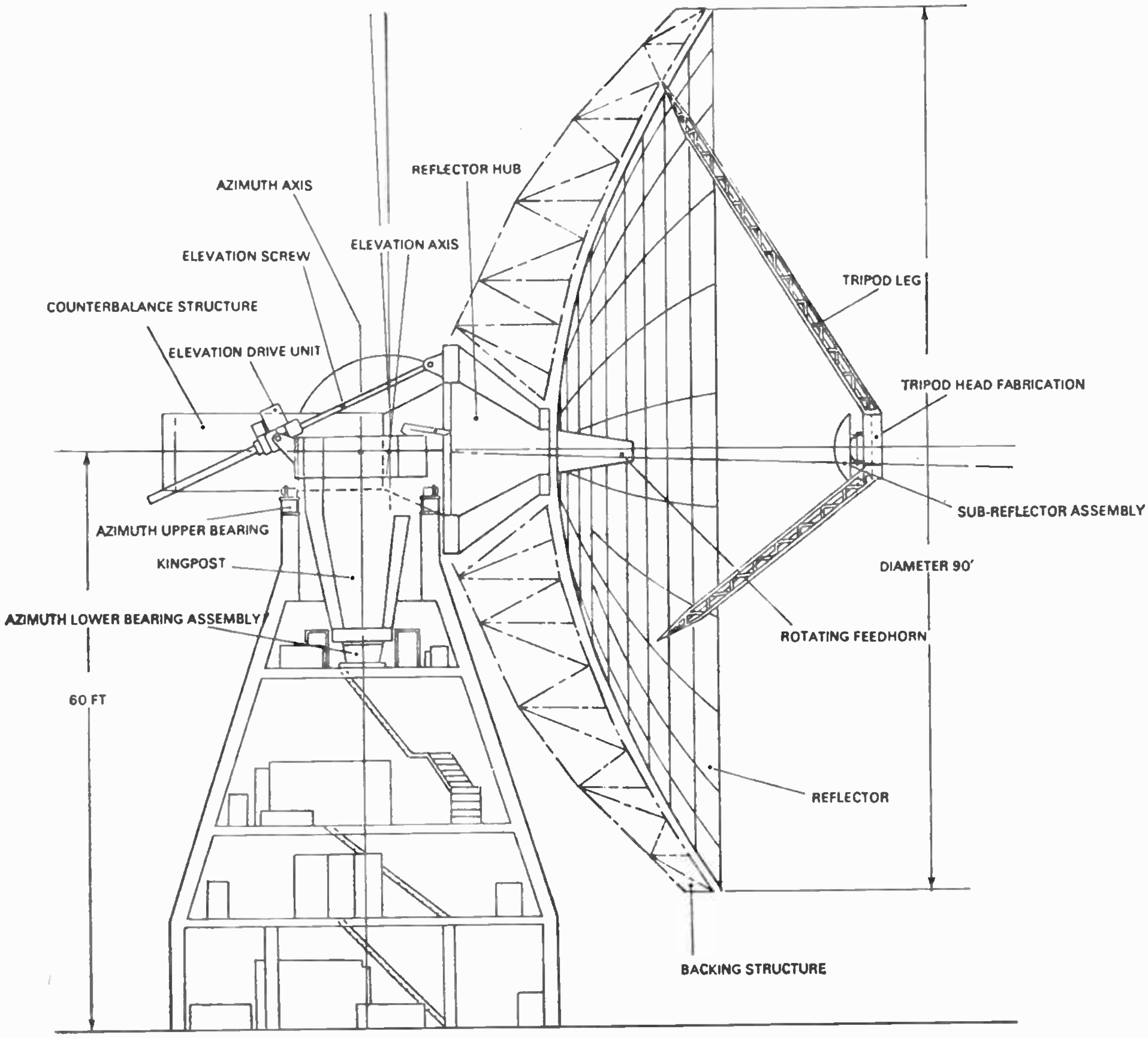


Fig. 5. Cross-section of pedestal and dish.

is a room which houses the parametric amplifier prime receivers and cryogenic system, both of which are duplicated. This room has a self-levelling floor for ease of maintenance at all elevation angles.

The dish itself is 27 m (90 ft) diameter quasi-parabolic employing a Cassegrain hyperboloid sub-reflector with a focal length of 9·08 m (29·8 ft). It is fully steerable and is composed of 24 radial steel trusses which were assembled on jigs at ground level before hoisting for attachment to the hub-cone. Aluminium panels, 2·1 mm (1/12 in) thick, were then fitted individually on to the trusses and their joining intercostals, each panel being capable of individual adjustment in relation to the next. The net result gave a surface accuracy of 0·1 mm (0·040 in) r.m.s. checked by a Pentag optical device from the centre. The backing structure detail of the dish is illustrated in Fig. 6.

Electrically controlled, large twin recirculating ball screw jacks move the dish in elevation. Similarly, two 10 hp d.c. motors driving pinions against a toothed bull-ring fixed to the tower control activate the azimuth movement. Both systems are in constant but infinitesimal tracking motion controlled by the auto-track receiver tuned to one of the satellite tracking beacons. Programmed tracking facilities are also available. A.c. motors can be brought into use for speedy movement to a stow position in an emergency.

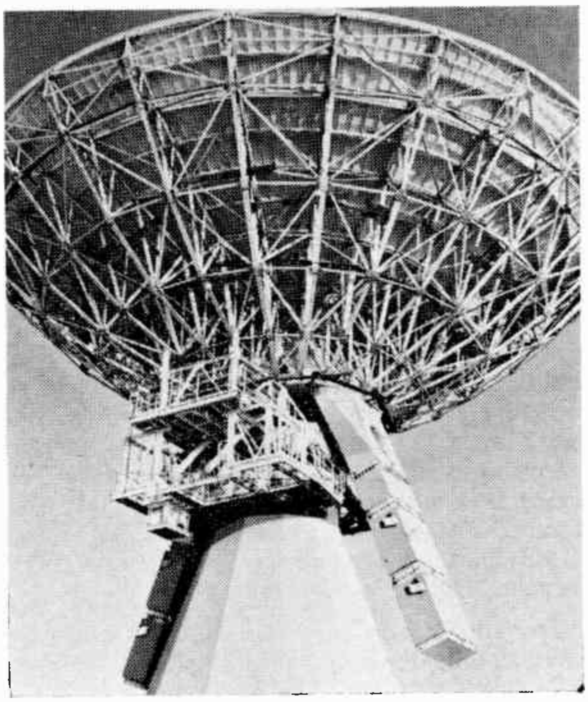


Fig. 6. Close-up detail of the dish backing structure.

5.3. *Dish Performance*

The design objective was to meet or better the ICSC requirement of a standard earth station to have an antenna gain, G, greater than 57 dB at 4 GHz, and a figure of merit, G/T, not less than 40·7 dB before permission is given to use the satellite.

The figure of merit is defined as the ratio of

$$\frac{\text{gain of antenna}}{\text{system noise temperature (in degK)}} = \frac{G}{T}.$$

The dish diameter chosen was 27 m for which it can be shown that, at 4 GHz, the gain $G_o = 61{\cdot}21$ dB. Losses due to profile errors, mechanical errors and blockage by the Cassegrain sub-reflector and tripod supports total approximately 1·83 dB.

Thus, effective gain,

$$G = 61{\cdot}21 - 1{\cdot}83 = 59{\cdot}38 \text{ dB.} \quad \ldots.(1)$$

The system noise temperature is made up of three major portions, namely,

T_1, low noise receiver system = 22°K

T_2, transmission network noise = 10·1°K

T_3, aerial noise due to the Earth's surface and the atmosphere = 29·3°K

Total, T = 61·4°K

or 17·88 dB. ..(2)

Therefore

$$\frac{G}{T}\text{attainable} = (1)-(2) = 59{\cdot}38 - 17{\cdot}88 = 41{\cdot}50 \text{ dB.}$$

This objective was very closely attained on completion of the dish structure by relating gain from the radio star Casseopeia A.

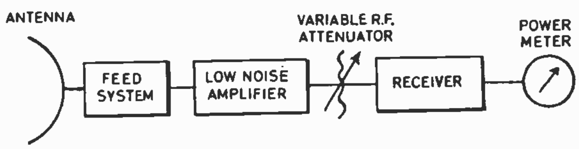


Fig. 7(a). Simplified circuit for G/T measurement.

Figure 7(a) shows a simple circuit for measurement of G/T. It relies on the known flux densities of certain radio stars shown in Table 2, on the Y-factor technique using hot and cold reference loads, and application of a correction factor for angular extent from the radio star point source, i.e., the antenna beam width.

Table 2

Star	Frequency GHz	Flux density W/m²/Hz	Most probable error in flux %
Cassiopeia A	4·080	$1047{\cdot}0 \times 10^{-26}$	2
Taurus A	3·950	$716{\cdot}9 \times 10^{-26}$	3
Cygnus A	4·161	$465{\cdot}1 \times 10^{-26}$	3

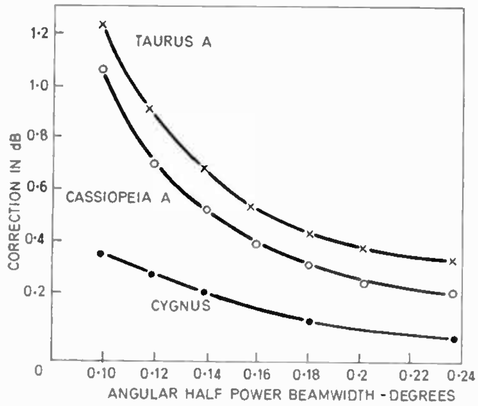


Fig. 7(b). Angular correction factor, from three radio stars.

The basic equation used is

$$\frac{G}{T} = \frac{8\pi k(Y-1)}{S.\lambda^2}.K_1 K_2 \text{ dB}$$

where k = Boltzmann's constant.

Y = ratio of noise power available when antenna is directed towards cosmic source to that available when antenna pointed to background clear sky at same elevation angle.

K_1 = correction factor for atmospheric attenuation and approximates to $10^{(0.004/\sin\theta)}$ where θ is measurement elevation angle.

K_2 = angular correction factor plotted in Fig. 7(b).

S = known flux density from Table 2.

6. Description of Circuitry

The simplified block diagram in Fig. 8 shows the signal progression between the earth station antenna and the baseband equipment of the microwave link. Except for the polarizer, diplexer and rotating waveguide joint (of annular cross-section to cater for both transmit and receive frequencies), there is full redundancy of all equipment.

6.1. *Transmit Path*

Identical signals are fed from the baseband multiplex rack to the baseband amplifier and modulator cabinet where they are converted to 70 MHz intermediate frequency before being fed via cross-site cable to the transmit i.f. amplifier and transmit drive rack.

The amplified signal next modulates the upconverter before it is passed at radiated frequency (6230 MHz for Bahrain) to two travelling wave tubes in cascade, which form the main transmitter, to produce 5 kW maximum into the antenna. Highly purified water is employed for cooling the final t.w.t. Normal power used with the present channel loading is approximately 95 W which, allowing for feeder losses gives an estimated isotropic radiated power of 80·3 dBW. Full provision of non-reflective loads at various points in the transmit chain is made for test purposes.

6.2. *Receive Path*

Main and standby parametric amplifiers, cooled by liquid helium to a temperature of 17°K (−255°C), are located immediately behind the dish centre. They provide a gain of 30 dB to the −130 dBW signal received from the satellite, under clear sky conditions, before passing to a tunnel diode amplifier (10 dB gain) fitted with threshold extension and thence to a low-noise t.w.t. (28 dB gain). A flexible waveguide allows for elevation movement and a cable banding and stowage basket allows ±270° azimuth movement for control cables.

The amplified signal is fed via the single rotating joint and cross-site waveguide at r.f. to the Technical Block, where passive stripline dividers feed to downconverters each controlled by phaselock loop local oscillators. These oscillators are highly stable devices and during installation the maximum drift over 28 days was recorded at 558 Hz from the nominal 100 MHz, whilst the average was less than 100 Hz. Each local oscillator is set to allow the downconverter to extract and filter the allocated carrier of the station required and to convert to i.f. Currently, two carriers are being received—Goonhilly 3900 MHz, and Kuwait 3995 MHz. Lebanon 3790 MHz was scheduled to open to Bahrain in October 1970 and during 1971 earth stations in India 3845 MHz, Hong Kong (2) 4015 MHz and West Pakistan 3780 MHz are all scheduled to come on stream and will operate with Bahrain. East Africa will connect later.

Each destination is equipped at Bahrain with separate, fully duplicated downconverters, the signals from which, after i.f. amplication, are demodulated and fed at baseband frequency via an automatic selection switch to the receive multiplex equipment

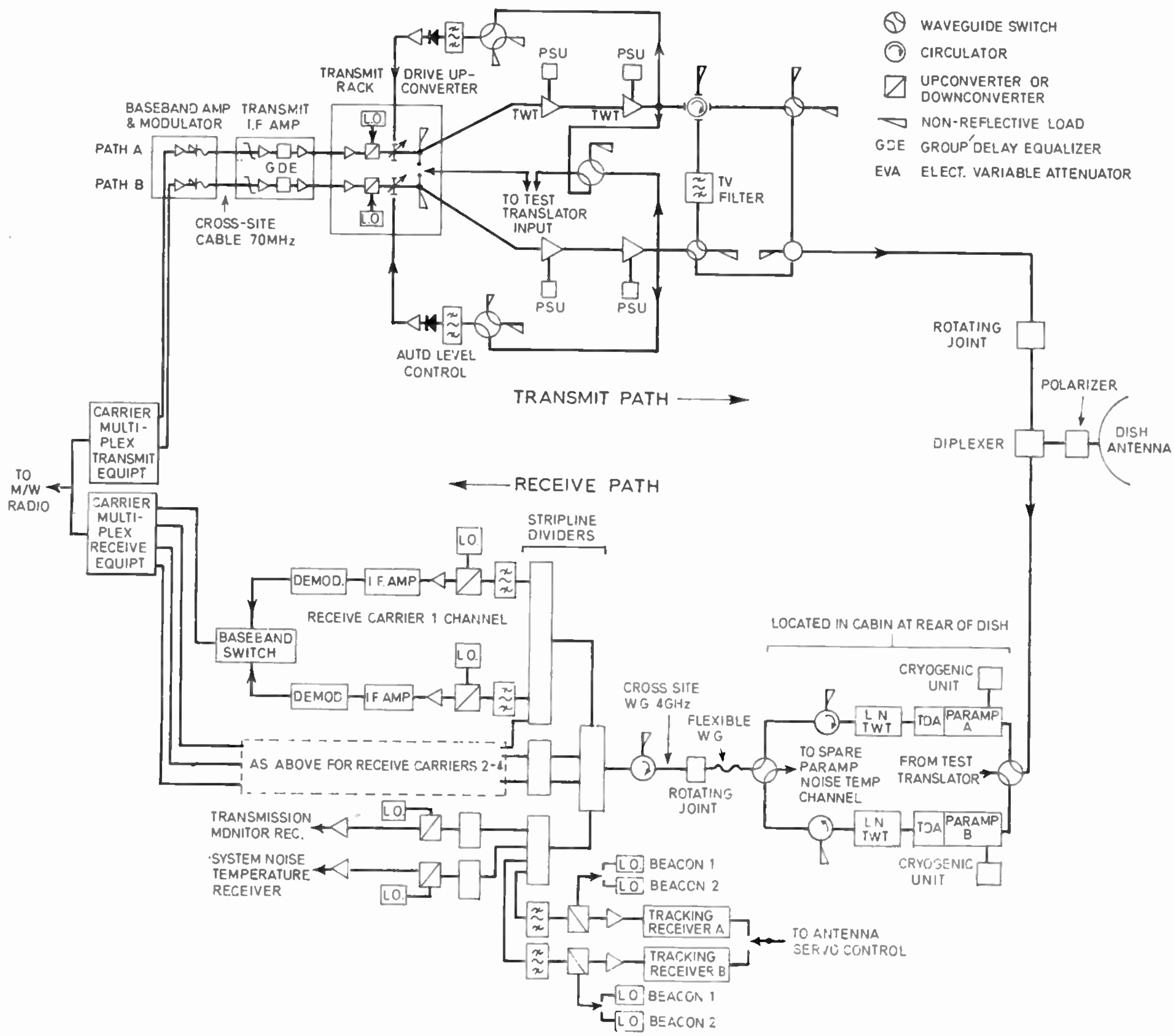


Fig. 8. Bahrain earth station—simplified schematic.

for onward transmission at super-group level over the microwave end-link.

Bahrain's own transmission and the system noise temperature can be similarly monitored. Also either of the two satellite beacons—3934 MHz for *Intelsat III*, Flight 3 identification and 3967 MHz common to all *Intelsat III* series—are demodulated and used to control the antenna servo mechanism to provide automatic tracking of the satellite. Data supplied from COMSAT are available to feed to the digital encoders for program track in the event of auto-track failure.

The status of each major unit in both transmit and receive paths is continuously monitored and fed back to the control console. Alarms indicate any parameter variation and a mimic diagram at the rear of the console gives visual indication of any such changes.

7. Baseband Configuration

The baseband configuration to be transmitted and received by each earth station is agreed internationally at regular meetings of operations representatives of earth station owners and the system manager. Table 3 shows the arrangements in force for Bahrain and projected to the end of 1973.

It is interesting to note the multi-user facility, for example on Group 5 to Hong Kong and UK, which is best explained on the receive side. Although Bahrain will receive separate carriers from these two

Table 3

Baseband for Bahrain Transmit: period to year end 1973
Carrier: Bahrain—Multi-user 6001 Supergroup

Destination	Channels 4 kHz	Carrier MHz	0	12	60	108	156	200	252 kHz
6230/4005 MHz	(60)	10		Gp. A	Gp. 5	Gp. 4	Gp. 3	Gp. 2	
Hong Kong	2				2				
India	4			4					
Kuwait	12						12		
Lebanon	11							11	
West Pakistan	5			5					
United Kingdom	20				8	12			
Total	54								

earth stations fed by separate down-chains to common baseband equipment, each country will be modulating discrete pre-arranged channels in the same group which will be extracted at channel level by the Bahrain channel multiplex equipment located at the remote end of the microwave link in the International Transmission Maintenance Centre (ITMC) before being passed to the International Telephone Exchange switching equipment.

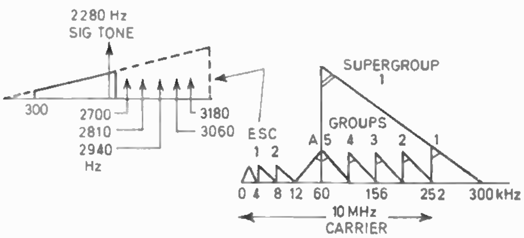


Fig. 9. Group arrangement and engineering service channels.

The group and supergroup displacement of the 10 MHz carrier is shown in Fig. 9. The 0–4 kHz portion of the spectrum is utilized for energy dispersal and 4–8 kHz and 8–12 kHz are used for engineering service circuits. Each of the latter two 4 kHz bands are capable of division into one voice and five telegraph channels for use between earth stations and system control points.

8. Connecting Communications Network

As mentioned in Section 4, the prime objective of the Bahrain earth station is to provide high quality international voice channels from the Arabian Gulf. In order to exploit this outlet fully and to meet the needs of neighbouring states, equally good communications to CCITT standards were required between Bahrain and the other important locations in the area. Therefore, concurrently with the earth station planning, the following wideband projects were instigated and have been or are about to be brought into service to form a comprehensive network.

8.1. *Bahrain-Doha* (*Qatar*) *Tropospheric Scatter Link*

This portion of the network was the first to open, in August 1967. It operates over approximately 90 km of sea and flat desert on 2 GHz, using dual 100 W output fed into twin parabolic 9·2 m (30 ft) diameter dishes, thus providing quadruple diversity (space and frequency) at the receiving end. The capacity of 24 channels is almost fully occupied and is about to be expanded to provide for 120 channels.

8.2. *Bahrain-Dubai* (*Trucial States*) *Tropospheric Scatter Link*

This link operates over a 390 km mainly sea path between the two highest points in each country, Jebel Dukhan (Bahrain) and Jebel Ali (Dubai), and is connected by microwave 7 GHz links at each end to the main town areas.

The tropo-scatter link uses two 10 kW transmitters in the 1·7–2·3 GHz band feeding into twin 27 m

(90 ft) billboard reflectors of parabolic curvature via offset feedhorns. The four receivers also operate in quadruple diversity with tunnel diode amplifiers as the prime receivers.

The capacity is 72 channels but is presently lightly loaded with adequate provision for expansion for several years. It commenced operation in July 1969 simultaneously with the earth station.

The Jebel Dukhan station and terrain is illustrated n Fig. 10.

8.3. *Bahrain-Dammam (Saudi Arabia) Microwave Link*

This is the final portion of the expansion programme and is to be implemented in mid-1971. It will replace a 6-channel v.h.f. link over a 61 km sea path and will be a 300-channel capability system, sub-equipped initially for 36 channels and operating in the 2 GHz band.

Figure 11(a) shows schematically the pattern of communications which has resulted from these various links. It should be read in conjunction with Figs. 11(b) and 11(c) which together indicate the full range of services and destinations available as outlets via the new earth station.

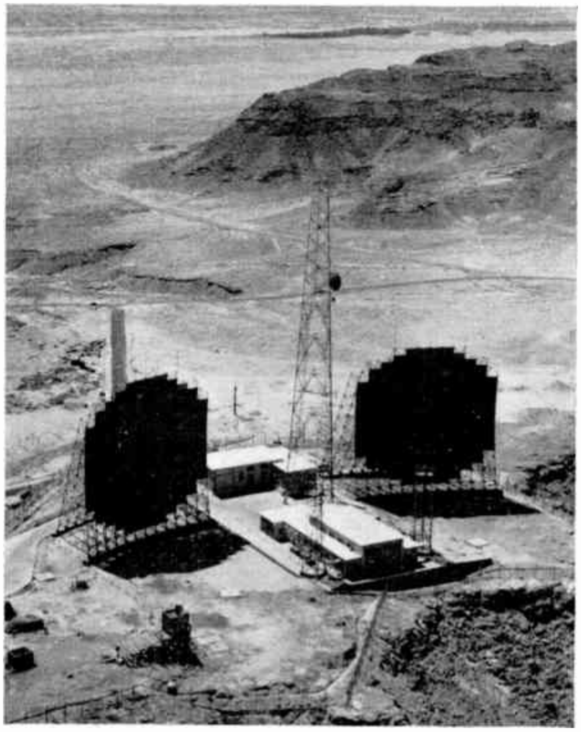


Fig. 10. Tropospheric scatter station at Jebel Dukhan, Bahrain, showing surrounding terrain; the station is 300 ft above sea level.

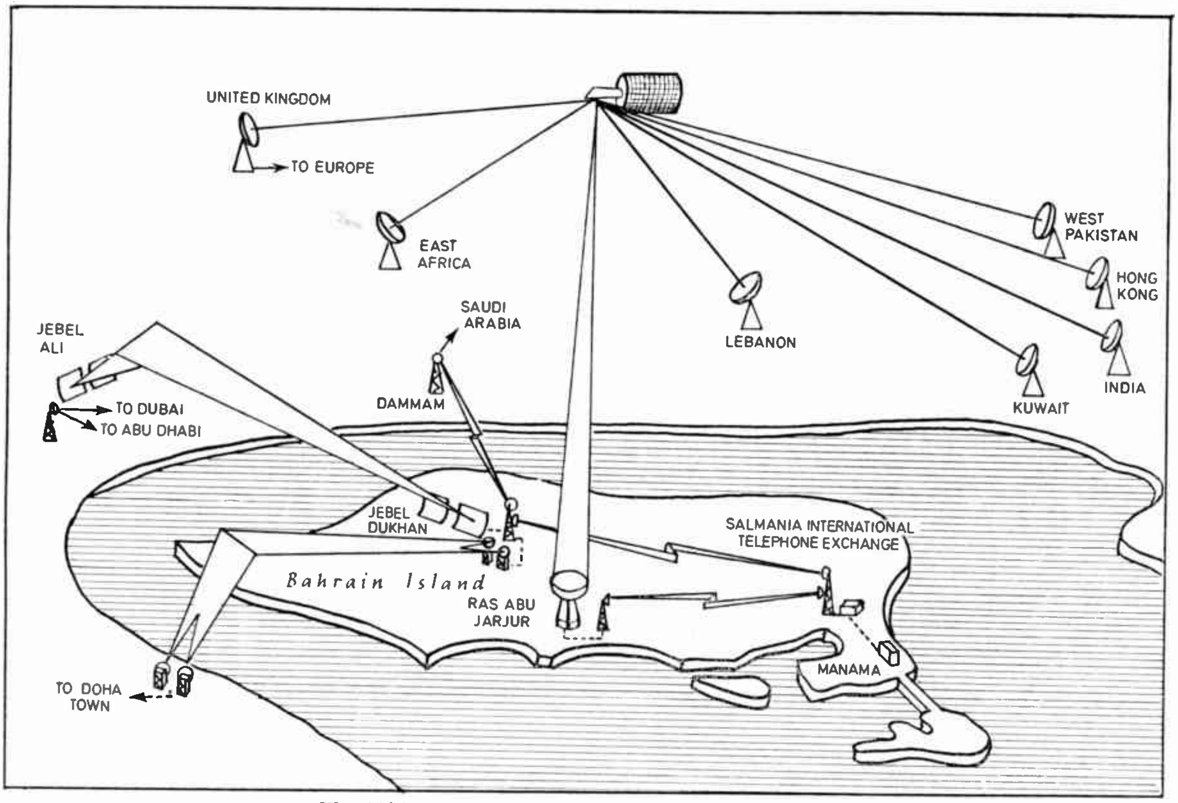


Fig. 11(a). Pattern of Bahrain communications links.

9. International Telephone Exchange

Obviously, an expansion of such order could not be handled by existing exchange facilities geared essentially to the previous h.f. telephone circuit capacities for international calls. The final criterion upon which any telecommunications enterprise will succeed or fail in the estimation of the public is the performance of the telephone instrument or record equipment in the customer's home or office, and the availability of the required connexion on demand.

Therefore, new I.T.Es, with associated International Transmission Maintenance Centres, at Salmania, (Bahrain), Dubai and Doha were commissioned to be in service to meet the increased channels of the new links. The new exchanges are 4-wire cross-bar type equipped for CCITT No. 5 signalling. A full description of all functions of the I.T.E. is beyond the scope of this paper.

10. Traffic Growth

The graphs in Fig. 12 show the traffic growth in paid minutes of international telephone calls from four months prior to the opening of the Bahrain earth station to the United Kingdom and tropo-scatter

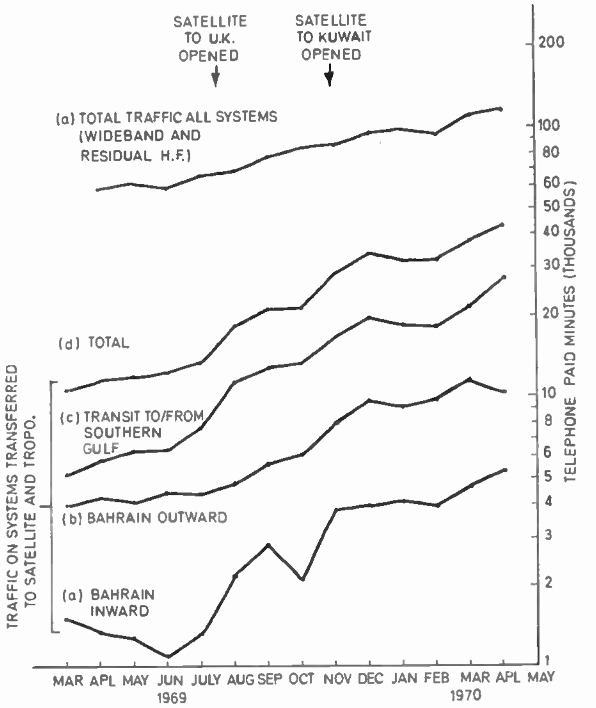


Fig. 12. Bahrain International traffic growth March 1969 to April 1970.

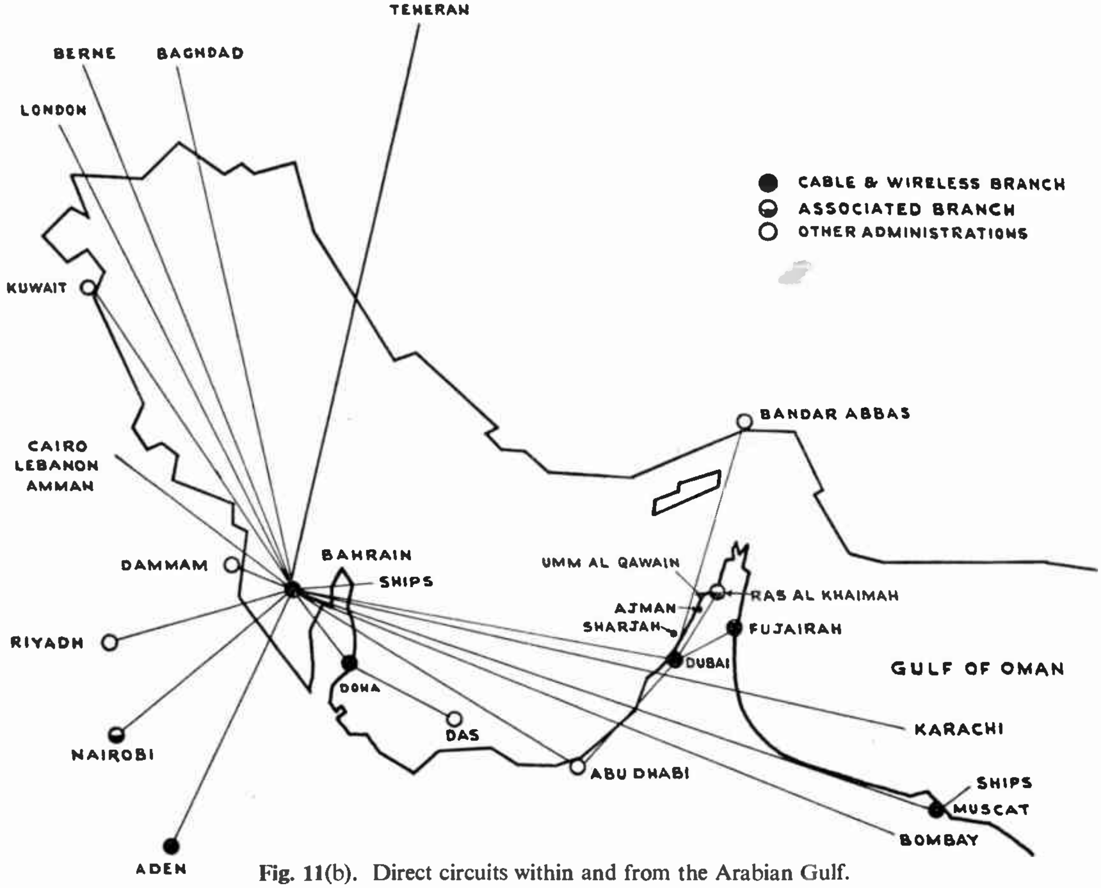


Fig. 11(b). Direct circuits within and from the Arabian Gulf.

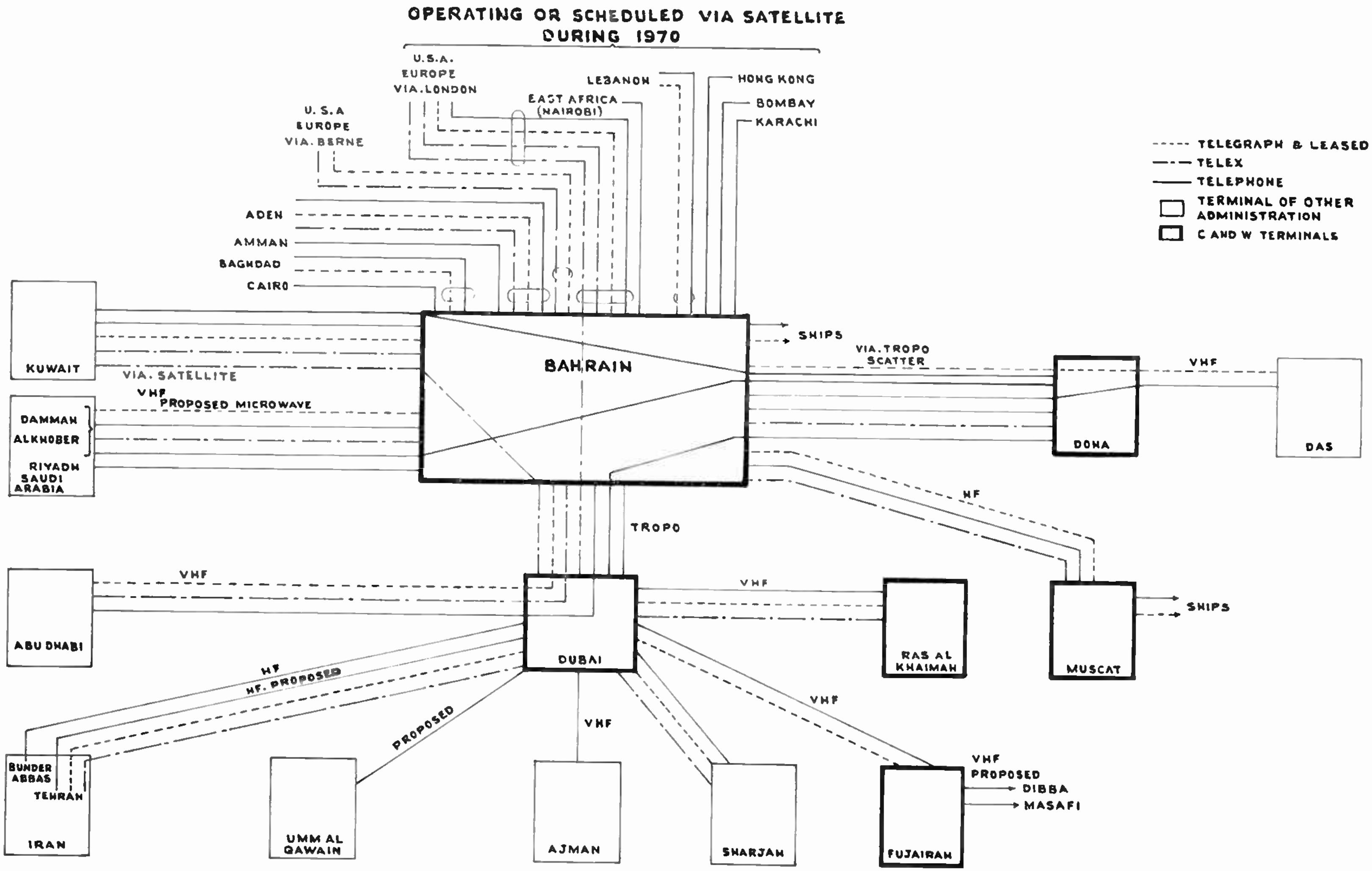


Fig. 11(c) Gulf telecommunications links.

system to Dubai until April 1970, the latest month for which statistics are available at the time of writing.

Curves (a) and (b) show inward and outward traffic terminating in Bahrain, whilst (c) indicates transit traffic in both directions to and from the Trucial States in the southern part of the Arabian Gulf, on links which transferred from h.f. to wideband systems in July 1969.

Although taken over a relatively short statistical period, it is interesting to note that the total of this traffic (d), which increased at a rate of 13·5% in the four months before the opening of the wideband systems, has shown approximately 200% increase in the nine months after the opening. In the case of Bahrain inward traffic (a), a falling trend has been reversed to give about 250% increase in the same period.

The curve of total traffic over all systems (e) shows a less steep growth—100% over a 12-month period—since it includes circuits still carried on residual h.f. It is expected that this rate of growth will be maintained as the Lebanon and Indian sub-continent earth stations commence operations to Bahrain, after which it is likely that growth will be regulated by factors of commerce rather than channel capacity. But at the time of writing it is certain that the capital outlay and technical effort expended on wideband systems is being justified by a very satisfactory traffic increase.

11. Conclusion

The outstanding success of the *Intelsat III* global system has ensured that satellites, backed up by submarine cables, will provide the main means of communications for at least the next decade. The larger and higher-powered *Intelsat IV* series, and its successors, with more sophisticated transmission patterns will probably result in smaller earth station dishes with a consequent reduction in the cost of earth stations from the present £1·5–£2 million.

But, eventually, one satellite over any region will become saturated, requiring a second satellite to meet the traffic demands. Unless the cost of an earth station does in fact fall drastically to make a two-dish station economically attractive, those stations with one dish will only be able to communicate via one satellite with other countries also using the same one. An increase of all earth stations to two-dish mode could also lead to congestion of the space communications frequency spectrum.

A number of solutions have been put forward, such as dividing user countries into regions to work to a given satellite or using one satellite for high density and the other for low density routes. These ideas would either reduce international commercial contact in telecommunications or produce unacceptable delay by requiring two-hop paths.

Possibly the most likely answer will be some form of satellite-to-satellite relay before bringing the signal back to earth, provided it does not cause the signal delay to exceed the generally accepted maximum of 450 to 500 ms.

Satellite systems using microelectronic technology have produced an excellent new mode of communications; but there is no room for complacency if the continuing demand for more capacity is to be met. For the moment, however, the communications needs of the Arabian Gulf area will certainly have been met when the remaining microwave link to Saudi Arabia is completed shortly.

12. Acknowledgments

The author is grateful to Cable and Wireless Ltd. for the use of technical literature relating to Bahrain Earth Station and for traffic statistical data of the Arabian Gulf area. He wishes to express also his appreciation of the ready assistance given by the project leader and engineers of the installation team and other colleagues in Bahrain and by the design engineers in London during the preparation of this paper. The help provided by engineers of the Marconi Company during the building of the station is also gratefully acknowledged.

13. Bibliography

1. United Kingdom Seminar on Communication Satellite Earth Station Planning and Operation, May 1968, Vols. 1, 2 and 3.
2. INTELSAT Satellite Systems Operations Guide, May 1969, Vols. 1 and 2.
3. Kreutal, R. W. and Pacholder, A. O., 'The measurement of gain and noise temperature of a satellite communications earth station', *Microwave J.*, **12**, No. 10, pp. 61–6, October 1969.
4. *Intelsat III* Commercial Globe Communications Satellite System, undated, TRW Inc. Systems Group, Redondo Beach, Calif., USA.
5. 'Developments in satellite communications', *Lenkurt Demodulator*, August 1969.
6. 'INTELSAT Report, Meeting of Preparatory Conference on Definitive Arrangements', *Telecommunications J.*, **3**, No. 10, October 1969.
7. Traffic statistics of Cable and Wireless Ltd., Bahrain for Arabian Gulf area, March 1969–April 1969.
8. Communications Satellite Corporation, Reports to The President and The Congress, April 1970.

Originally submitted as a Technical Paper to gain exemption from the Institution's Graduateship Examination. Manuscript first received by the Institution on 5th March 1970 and in final form on 1st September 1970. (Paper No. 1351/Com. 34)

A Simple Data Recording System with Computer Analysis

By
G. R. WHITFIELD, Ph.D.,†

This system uses a new technique of f.m. recording to record accurately a small number of slowly varying parameters, and speech, on one track of a cheap commercial tape recorder. The record can readily be played back into a computer through a simple interface unit. The paper also describes transducers for converting pressure and temperature into frequency.

1. Introduction

This system was designed to fill a gap in the present range of data logging systems. It records accurately a small number of slowly varying parameters and speech on a single channel of magnetic tape, using a cheap commercial recorder and relatively simple electronics. The whole system is small, light and cheap. The record can readily be played back into a computer.

The design is based on a number of ideas which are individually almost self-evident:

(1) The only suitable recording medium is magnetic tape; if possible a cheap domestic portable recorder should be used.

(2) The easiest way of recording accurately is to use f.m. recording, and the best way to do this is to transduce directly into frequency. This gives the added advantage that there are no errors from contact resistance, contact potential or variation in amplifier gain between the transducer and the recorder.

(3) Accurate f.m. recording requires constant tape speed. But in practice, the output frequency will be measured by a frequency counter—a timed gate and counter—for input to the computer. If the timed gate is put *before* the recorder, then the record consists of a sample of the signal containing a number of cycles proportional to the frequency, and the tape speed need not be constant.

(4) When designing the transducers, stability and low drift are essential, but linearity is not, since the computer can be used to apply a calibration curve.

The prototype system was developed to measure the performance of gliders. It had to record height, airspeed and elapsed time, within $\pm 0{\cdot}1\%$ or better, for computer analysis. The complete airborne apparatus had to occupy less than 0·03 m^3, weigh less than 15 kg with power supplies and cost less than £100 for materials. This specification was met comfortably. Systems are now being developed for other applications, some with more parameters. In this paper, the discussion has been kept as general as possible, but the prototype system (being small) is used as an example.

† Department of Applied Physical Sciences, University of Reading, Whiteknights, Reading, RG6 2AL.

2. The System

A block diagram of the system is shown in Fig. 1, and typical waveforms in Fig. 2. The transducers convert the input quantities into frequency, variable

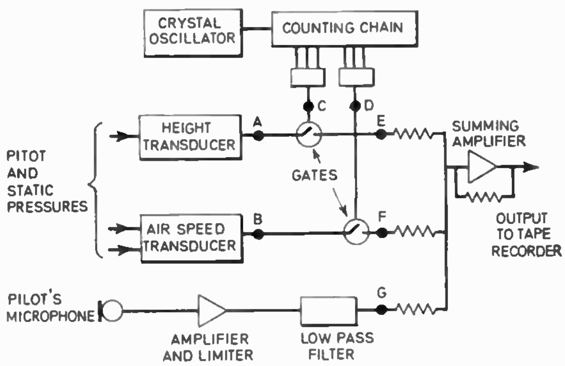


Fig. 1. Block diagram of recording system.

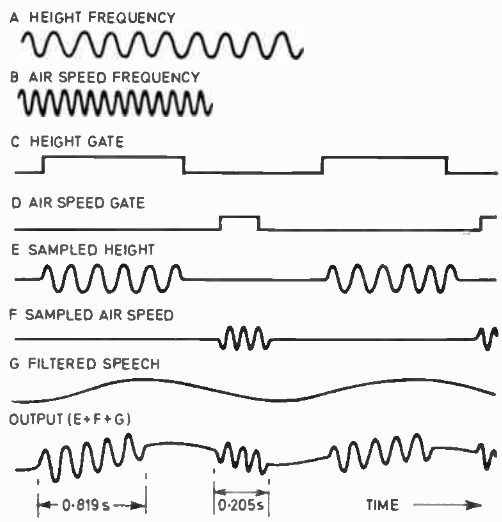


Fig. 2. Waveforms in recording system.

from 4–10 kHz. A 5 kHz crystal oscillator drives a counting chain to generate gating pulses, which are used to pass samples from each transducer in turn, with gaps between them, to the summing amplifier and recorder. The number of cycles in each sample is proportional to the frequency coming from the transducer, and so is a measure of the input quantity. In addition to the data, a filtered speech channel (up to 3·5 kHz) is also fed to the summing amplifier. The combined signal is fed to the recorder, which runs continuously. The speech channel is limited to avoid overloading the tape, because this can spread speech signals into the 4–10 kHz band which carries the data.

On replay, the speech is filtered out, and the sinusoidal data signals are made square by a Schmitt trigger circuit and differentiated to give pulses which are used to interrupt a PDP-8 computer. The computer is programmed to count the number of interrupts (and hence cycles) in each sample; it recognizes the gaps by the absence of interrupts for several milliseconds. Subsequent analysis depends on the application.

Most of the circuits are conventional, using r.t.l. microcircuits for counting and generating the gating waveforms, and discrete components elsewhere. The summing amplifier (Fig. 3) is a virtual-earth amplifier, and the gates are field effect transistors in series with the input resistors. This arrangement gives a switching ratio of 300 (in volts) with little shift of zero when each switch operates. No trouble has been experienced with multiple inputs (up to seven so far).

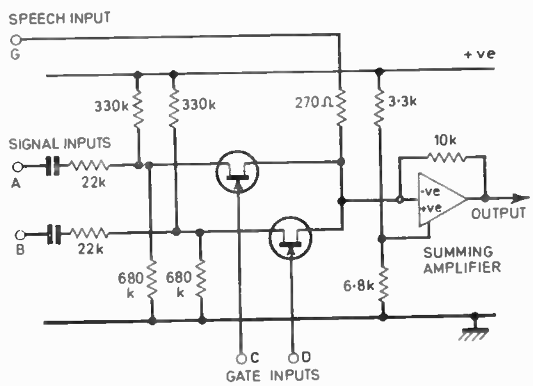


Fig. 3. Gates and summing amplifier.

3. Transducers

The transducers developed for this system have mostly been R–C oscillators of the general form of Fig. 4. The loop contains a phase splitter and R–C pair which gives a phase shift that varies with frequency, a further 270° phase shift independent of frequency, and an a.g.c. stage. The circuit oscillates at the frequency $f = 1/2\pi RC$ at which the R–C circuit gives a 90° phase shift.

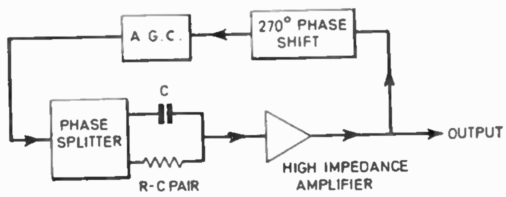


Fig. 4. Block diagram of transducer oscillator.

The height transducer (Fig. 5) uses a fixed resistor and a variable parallel plate capacitor, whose plates are mounted on two standard aircraft altimeter capsules (obtained from Smiths Instruments). As the height increases, the pressure falls, the capsules expand and the gap between the plates decreases, varying linearly with height. To a first order, ignoring edge effects and strays, the capacitance is inversely proportional to the separation of the plates and so the frequency also varies linearly with height. In practice the law is not quite linear, the slope changing by 10% over the height range, but it is easy to correct for this in the computer. There is no sliding friction, so sticking and hysteresis are minimized, and the use of two opposed capsules compensates for acceleration.

The principal difficulty in the design of the circuit is the low capacitance of C, typically 30 pF, which gives the phase shifter a very high impedance at a few kilohertz. The following stage uses a field effect transistor and two cascaded emitter followers. The a.g.c. stage is a virtual-earth amplifier with a capacitor as the input impedance and a vacuum-mounted thermistor as the feedback resistance. This gives the required 270° phase shift independent of frequency, while the thermistor keeps the amplitude constant.† To minimize stray capacitance, the lower plate of the capacitor is screened by a plate driven by the second emitter follower, and the whole transducer is mounted in an earthed metal case.

The airspeed transducer is similar, but uses two opposed airspeed indicator capsules. The pipework is more complex as pitot pressure has to be applied to the inside of each capsule and static pressure to the outside. Since a.s.i. capsules are too weak to carry plates, the capsules themselves are used as the plates of the capacitor.

† Baxendall, P. J. 'Transistor sine-wave oscillators', *Royal Radar Estab. J.*, No. 45, p. 5, October 1960.

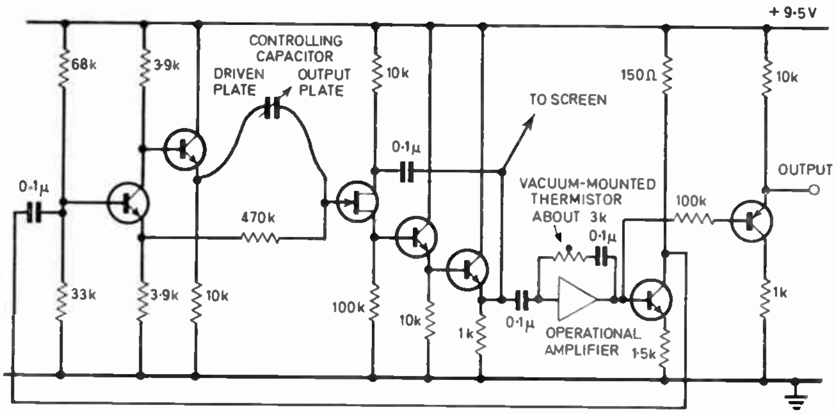


Fig. 5. The circuit of the height transducer.

Obvious developments of these transducers can be made to measure pressure, displacement or acceleration (with a heavy plate on a spring). All use the same circuit. Angular displacement can be measured using a vane type variable capacitor in a similar oscillator circuit.

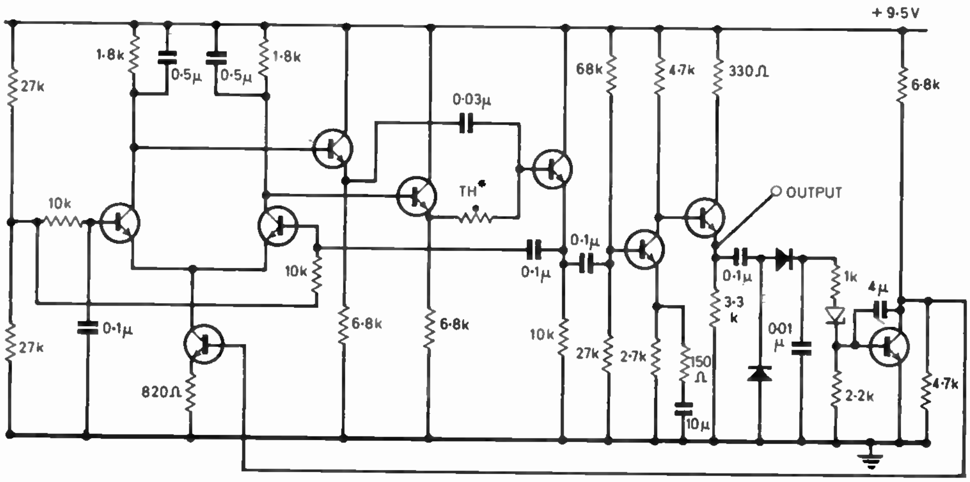
For thermometers, it is easier to use a thermistor and a fixed capacitor as in the circuit of Fig. 6. This circuit is simpler because the impedances in the phase shifter are much lower. The use of an a.g.c. system in place of the thermistor of Fig. 5 widens the temperature range of the oscillator, and combines neatly with the long-tailed-pair phase-splitter. This type of circuit was not used for the height transducer as it was rather more temperature-sensitive than the earlier version—but this is negligible in a temperature transducer.

Transducers that give analogue outputs can be used with the system by adding an analogue to frequency converter. Voltage controlled multivibrators have been used successfully; the square wave output causes no trouble.

4. System Performance and Limitations

The system records samples (or strictly, averages over defined short periods) of each parameter at defined intervals of time. Since all the timing is derived from a crystal oscillator, it is easy to make timing errors negligible—less than 1 in 10^5 or 10^6.

The reading accuracy of each sample depends on the bandwidth available and the length of the sample, for there is an inherent uncertainty of ± 1 in the count. If the bandwidth is B and the sample time T, this uncertainty is $\pm 1/BT$ of the full-scale range of the



ALL SILICON PLANAR TRANSISTORS
2N3708 OR SIMILAR

* THERMISTOR MULLARD VF 3100 TO 3116 } ABOUT 1kΩ AT CENTRE TEMPERATURE

Fig. 6. The circuit of the temperature transducer.

parameter. For example, the prototype with a bandwidth of 6 kHz (from 4–10 kHz) and a sampling time of 0·8192 s has a reading accuracy of $\pm 1/4900$ of full scale in height, or about $\pm 0{\cdot}6$ m in a range of 3000 m. Thus the system can record data of low accuracy at a high data rate, or high accuracy at a low data rate. It is often convenient to take long samples of one parameter only, so that on replay its count always exceeds that of any other. The computer can then distinguish between the different parameters without explicit labelling.

Apart from the increased cycle time, there is no fundamental limit to the number of parameters that may be sampled. The gates of Fig. 3 have been used successfully up to seven parameters, with no indication of nearing a limit.

Absolute accuracy is determined by the stability of the transducers and the care taken in their calibration. The height and speed transducers appear to have long-term drift of less than 0·3% over a period of several months, but this is difficult to establish as the transducers are much more sensitive than the manometers used to calibrate them. Temperature coefficients of frequency of about 1 Hz degC^{-1} (in 10 kHz) have been achieved without undue difficulty.

5. Conclusions

The system described fills a gap between the small, cheap but rather inconvenient data recording systems and the large, expensive systems. The prototype has been successfully used to measure the performance of gliders† and for meteorological recording. A system is now being built for meteorological research that will record 14 parameters and speech on a two-channel magnetic tape recorder.

The advantages of the system are simplicity, and small size, weight and cost, and the fact that data and speech can be recorded simultaneously on one channel. The system is very resistant to recorder imperfections, since neither amplitude nor frequency carries any information. Even an occasional 'drop out', which might cause the loss of a very significant digit in a digital system, can only cause an error of 1 in the least significant place.

The disadvantages of the system compared with a digital one are the large amount of recording tape used for recording a relatively small amount of data, and the large amount of computer time required for playback; but the second could be avoided by using off-line equipment or a small computer for replay, and passing only the counts to the main computer.

6. Acknowledgments

The author is indebted to the Paul Instrument Fund of the Royal Society for a grant to help develop this system, and to the technicians and students who have built and used it.

† Whitfield, G. R., 'Automatic Recording and Analysis for Glider Performance Testing', 12th OSTIV Congress, Marfa, Texas, June 1970.

Manuscript first received by the Institution on 10th December 1969 and in final form on 23rd September 1970. (Short Contribution No. 140/IC34).

TRAVELLING SCHOLARSHIP

An I.E.E.E. travelling scholarship of £300 is offered for visit(s) to foreign electrical or electronic research or manufacturing establishments by a postgraduate student. The purpose is to promote an exchange of research and technological ideas and to foster a closer relationship between young engineers in different countries. Candidates must submit a programme for their visit(s), and the award will be made to the candidate whose programme is judged most likely to promote the objects of the scholarship.

The scholarship is financed by the U.K. and Republic of Ireland Section of the Institute of Electrical and Electronic Engineers, which is acting in collaboration with the Institution of Electrical Engineers and the Institution of Electronic and Radio Engineers. Entrants must be Student or Graduate Members of one of these three institutions.

Further information and entry forms from Dr. A. J. Ellison, Queen Mary College, Mile End Road, London E1 4NS.

The closing date for applications is 31st December 1970.

U.D.C. 621.319.4

Plastic Film Capacitors

By

P. D. HABERMEL†

Presented at a Components and Circuits Group Symposium on Capacitors held in London on 28th April 1970.

A review of the various materials used in plastic film capacitors shows the advantages of polystyrene, polyethylene terephthalate and polycarbonate. The technique of construction of foil and film and of metallized types is described and the limitation of various electrical parameters pointed out. Life and reliability factors are explained as well as the mechanism of failure for both a.c. and d.c. operation.

1. Introduction

Plastic films for the manufacture of capacitors were first introduced in Germany some 25 years ago but the general acceptance by the British electronics industry did not occur until 1950. It should be recognized that the basis of producing inexpensive, good quality, uniform film has originated from the use of those films in the general packaging industry. Without a mass market, from which a dielectric quality can be produced or selected, it is unlikely that the cost of plastic film capacitors could have reached their present levels.

2. Properties of Plastic Films

The basic requirements of a dielectric film are that it should have:

(1) good dielectric strength;
(2) uniform thickness;
(3) freedom from pin-holes;
(4) wide temperature range;
(5) resistance to moisture;
(6) acceptable loss angle in the frequency range used.

The most common of the dielectric films in use today are shown in Table 1 with the various trade names used by manufacturers of the basic film.

Each material thus has specific features and its popularity is largely dictated by the cost per picofarad. The most commonly used films today are polystyrene, polyethylene terephthalate and polycarbonate. These are abbreviated to Styrene, P.E.T. and P.C. and established pattern of their use with respect to capacitance is as follows:

0   10 pF   100 pF   1000 pF   0·01 μF   0·1 μF   1·0 μF   10 μF

←——— POLYSTYRENE (STYRENE) ———→

←POLYETHYLENE TEREPHTHALATE (P.E.T.)→

←——— POLYCARBONATE (P.C.) ———→

A further trend is seen with the application of styrene, P.E.T. and P.C. components in electronic circuits with styrene being used in LC tuned circuits and filters where low losses, close tolerance and linear temperature coefficient are the important parameters, and P.E.T. and P.C. in coupling and decoupling where small size, good insulation resistance and low cost are the relevant factors.

The working voltage stress that may be applied to these films is a function of the material and type of component. It is usually between 25 V/μm and 40 V/μm of dielectric thickness. The actual figure will be related to the reliability factor of the final product. Close liaison with the supplier of films is producing an optimum thickness for the established B.S. and I.E.C. R5 series of voltage ratings, i.e. 63, 100, 160, 250, 400 and 630 V.

3. Production Techniques

The different production techniques may be divided into two groups, foil and film where an aluminium or lead/tin foil is used with plastic film, and metallized film. Most plastic films can be metallized and this is done by evaporating aluminium on to the film under vacuum. Metallizing is carried out on wide rolls of materials with a masking plate to give a margin between adjacent strips of metallizing. Alternatively the complete surface area of the film is metallized and a spark erosion technique used to give the required margin.

Figure 1 shows a roll of metallized film which has been prepared for manufacturing into 25 mm wide capacitor cells. It is subsequently slit and wound into reels for processing into capacitors.

The thickness of the metallized layer is fairly critical as too thin a layer will have too high resistivity and

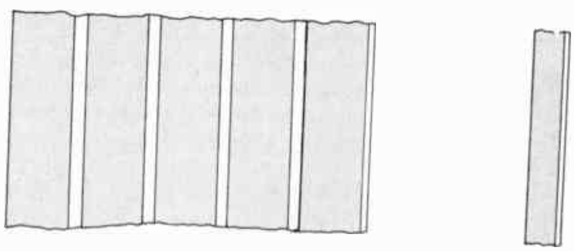


Fig. 1. Roll of metallized film.

† Mullard Ltd., Consumer Electronics Division, London, WC1E 7HD.

Table 1
Characteristics of plastic films

	Polyethylene terephthalate	Polycarbonate	Polypropylene	Polystyrene	Polytetra-fluorethylene	Lacquered foil
Trade Names	Mylar Melinex Hostaphan	Makrofol	Propafilm Bexphane Trespaphan	Styrofol	Teflon Fluon	
Dielectric K 1 kHz	3·2	2·8	2·2	2·4	2·1	10
or ε 100 kHz	3·1	2·75	2·2	2·4	2·1	—
Dielectric losses, 1 kHz	50×10^{-4}	9×10^{-4}	4×10^{-4}	2×10^{-4}	1×10^{-4}	200×10^{-4}
tan δ 100 kHz	160×10^{-4}	15×10^{-4}	6×10^{-4}	4×10^{-4}	2×10^{-4}	1000×10^{-4}
Insulation resistance at 25°C (MΩ/μF)	5×10^4	7×10^4	8×10^4	25×10^4	50×10^4	2×10^4
Temperature range	+120°C	+125°C	+100°C	+85°C	+200°C	+85°C
Temperature coefficient (parts in 10^6 deg^{-1} C)	+350	−75	−250	−150	−50	+1000
Long term stability	±3%	±2%	+2%	+1·5%	+1·0%	±5%
Minimum thickness available	3·5 μm	2 μm	10 μm	8 μm	8 μm	1 μm
Main feature	Good tensile and dielectric strength; inexpensive	Low losses	Low losses; inexpensive	High insulation resistance; linear temperature coefficient; low losses; high stability	High temperature; low losses; high insulation resistance; linear temperature coefficient	High capacitance volume

(1 μm = 10^{-6} m, i.e. 25·4 μm = 0·001 in.)

Polystyrene: polymer of the aromatic styrene monomer, non-polar, extruded.
P.E.T.: reaction of terephthalic acid and ethylene glycol }
P.C.: polyester—reaction of bisphenol A and phosgene } polar, solvent cast on drum.

consequently the a.c. rating will be restricted and the loss angle will be excessive. Too thick a layer will result in problems in the subsequent production processes. Typical figures for the resistivity of the metallizing are between 1 Ω and 4 Ω per square.

After winding the two metallized layers together so that the films are slightly staggered the cells are often flattened before being sprayed with a molten mixture of copper, tin or zinc.

The process is fairly critical and the molten particles penetrate into the capacitor cell to a depth of the stagger of the winding. It is the velocity and temperature of the spray which gives bonding or adhesion to the electrodes and provides the base for attachment of the terminations.

At this stage electrical tests are carried out and a voltage is applied from a controlled source impedance. This is known as 'healing' and operates by burning away the metallizing from the area around any pin-holes which might be present in the film. Thus any potential short circuit is 'healed' and the capacitor restored to its specification.

The other basic production technique is foil and film where the foil may be aluminium or lead/tin and connexions could be inserted electrode or extended foil. The extended foil construction is similar to the metallized except that separate foils and films are used (Fig. 2). These are wound together so that the metal foils extend beyond the film and are then spun over and connecting wires attached. The inherent

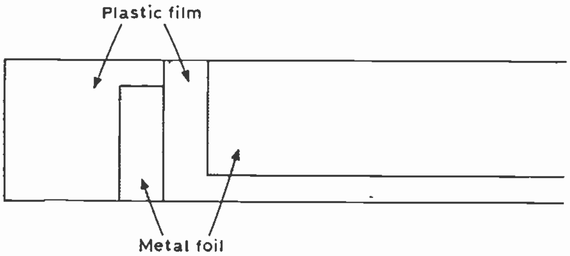


Fig. 2. Layers of plastic film and metal foil.

self-inductance of this type of construction is extremely low as well as the effective series resistance and enables high peak a.c. currents to be handled.

The inserted electrode construction is usually a tinned copper wire which is prepared by flattening its end into a 'spade' shape and welding directly to the aluminium foil of the capacitor. In theory the connexions are made at a point directly opposite each other to cancel out the inductance of the winding.

The most commonly inserted electrode capacitor is polystyrene which can be manufactured at relatively low cost due to the unique properties of the film. The terminations are usually of the welded type giving low d.c. resistance. The basic film is bi-axially stretched during its manufacture and thus the molecular structure is stressed at a temperature above the working point. After being slit and wound with metal foils the whole unit undergoes a heat shrink cycle at a temperature in the region of 110°C and thus the film is reduced to its original dimensions. This effectively seals the ends and virtually eliminates voids in the winding. The film therefore forms the encapsulation as well as the dielectric and the resulting capacitor is extremely stable and fairly inexpensive. The winding process is usually carried out with a capacitance bridge attached to the machine in order to ensure only small spreads in tolerance.

Extended foil construction can also be used but due to the low softening point of styrene which is approximately 90°C, the limit of temperature and time are extremely critical when the terminations are attached.

To summarize, the properties of polystyrene film are:

(a) high stability
(b) low dielectric loss
(c) linear temperature coefficient
(d) high insulation resistance.

On the debit side the upper operating temperature limit is only 85°C and the size and cost of large values, i.e. above 10 000 pF, is disproportionately high. A slight disadvantage with un-encapsulated or shrunk types is a humidity absorption coefficient which is +60 to +200 parts in 10^6 per degC over the range 50% to 95% relative humidity.

4. Capacitor Properties

Polyethylene-terephthalate capacitors are manufactured with foil/film as well as metallized film. They are probably the most widely used capacitors today and have largely replaced paper capacitors for d.c. applications up to 1000 V. The foil/film types invariably employ extended foil technique with the terminations connected directly to the extended aluminium foils.

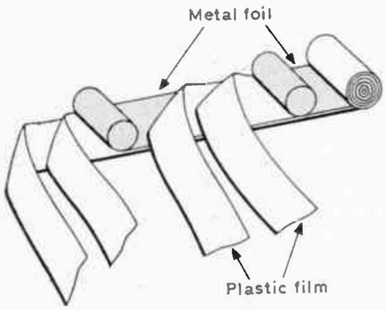
Over the past few years a demand has appeared for smaller components to allow greater packing densities in printed circuits. The metallized component has thus tended to replace the foil/film type without adversely affecting the cost. Although the basic cost of metallized film is high, the fact that only two films are used instead of two films and two foils results in greater winding efficiencies and thus the costs per microfarad tend to equate. For an equivalent capacitance a reduction in volume of over 50% is possible.

It must however be recognized that in pulse circuits, e.g. television time-bases, deflexion and switching circuits, the peak current may be several amperes and a metallized construction cannot be used. This is due to the effective series resistance which is comprised of the metallized layer on the dielectric and the sprayed end-sections. A particular capacitor thus has a power rating in watts which is derived from I^2R. This is generally greater than 60 mW per square cm of surface area and is a function of the film width which in turn corresponds to the thermal transmission across the metallizing to the terminations.

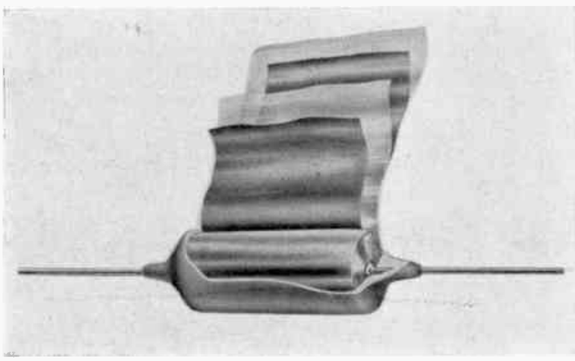
With a complex waveform or if the electrical circuit constants are unknown the maximum power dissipation can be found empirically by measuring the temperature rise on the body of the component. It is generally accepted that a maximum temperature rise of 10 degC at an ambient of less than 40°C measured after one hour's operation is permissible. Alternatively the r.m.s. current can be measured with a thermocouple instrument and should not exceed 400 mA.

The finite figure is a function of the geometry of the capacitor cell, i.e. the length/diameter ratio. For example, a given capacitance could be wound with a 1:1 ratio and have a lower series resistance than a 2 or 3:1 ratio. The relative production costs however invariably dictate a 2:1 length/diameter ratio.

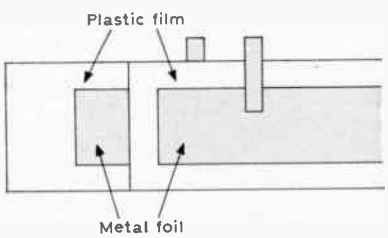
A further limitation exists where the average power is not exceeded but the instantaneous power is and for this reason a limit of dv/dt is given for the waveform. This can be either positive or negative, i.e. rise time or discharge. A convenient limit is 10 V/μs



(a)



(b)



(c)

Fig. 3. Typical examples of types of construction of film capacitors.

which can be realized without special attention to the sprayed end-section.

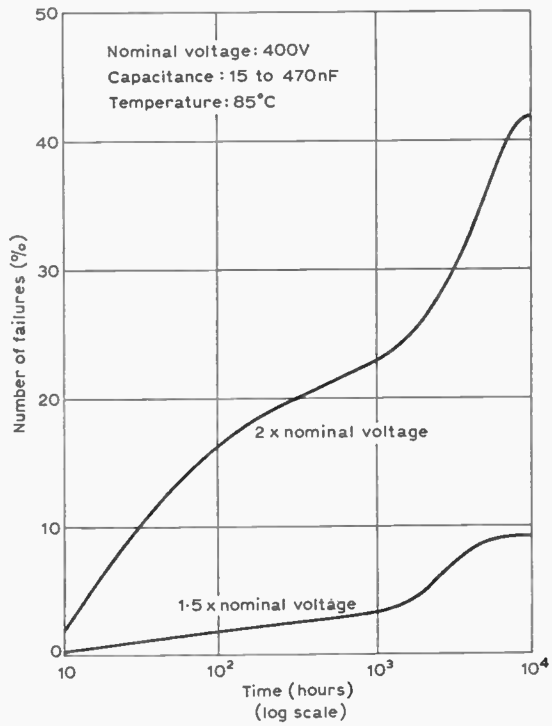
The accepted tolerance for P.E.T. capacitors is ±10% or ±20% and consequently the use of a bridge during winding is not necessary. Usually the winding is done automatically on a multi-head machine with a pre-determined length of materials.

Some four years ago metallized polycarbonate capacitors appeared and their construction techniques are similar to P.E.T. The main advantage from the application view point is lower dielectric losses which become significant at frequencies higher than 10 kHz.

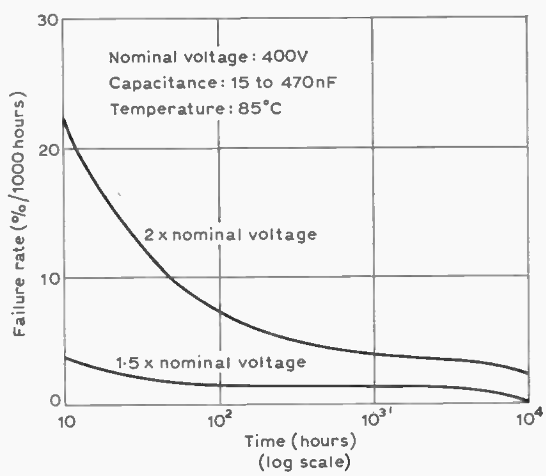
To summarize, the properties of P.E.T. and P.E.C. are:

(a) Good stability.
(b) High insulation resistance.
(c) Wide temperature range.
(d) Self healing in metallized version.
(e) Low cost.

On the debit side there are few serious disadvantages for d.c. rated components. The temperature coefficient of P.E.T. is non-linear but this could only be important in compensation circuits, and where capacitance stability over a wide temperature range is paramount.



(a) Percentage of failures as a function of time.



(b) Failure rate expressed percentage failure per 1000 hours.

Fig. 4.

5. Failure Mechanisms

Having examined the production processes involved we can consider the mechanism of failure in life. This can be divided into three main groups:

(1) Total failure—the inability of an item to perform its function.

(2) Partial failure—resulting in deviation beyond the specified limit.

(3) Degradation failure—which is both gradual and partial during life (affecting tan δ, IR and ΔC).

The mechanism of failure is fundamentally a function of construction and for foil/film it is invariably a perforation on the dielectric. Provided the component is used within its ratings the failure rate will be less than 0·1% per 1000 hours. Failure due to other factors are insignificant provided the initial quality is good.

The evaluation of the life and reliability is assessed at a maximum continuous rated temperature and, to accelerate the results, at 1·5 and 2·0 times the nominal voltage. Under these conditions a failure pattern for low and high voltage types with respect to capacitance is shown on Figs. 4(a) and (b).

These show a maximum rate during the first 100 hours and then a significant drop.

In the case of a metallized construction, failure due to a perforation of the dielectric is unlikely to occur due to the self healing features and the failure rate for all categories is less than 0·01% per 1000 hours at the rated voltage. On circuits where the

Table 2

Life test requirements at $1{\cdot}5E$ rating and 85°C for polyester capacitors

	1000 hours	10 000 hours	
Foil/film 15 nF–470 nF 400 V	3·3%	9·4%	Catastrophic failure (punctured dielectric only)
Metallized All Types	0·4% 0·15%	1·3% 0·8%	Catastrophic failure Degradation failure

Catastrophic failure: short circuit (less than 1 MΩ) or open circuit.

Degradation failure: ΔC greater than 5% after 1000 hours or 15% after 10 000 hours.
tan δ greater than 2× requirement after 1000 hours.
Insulation resistance less than 0·1 × requirement after 1000 hours.

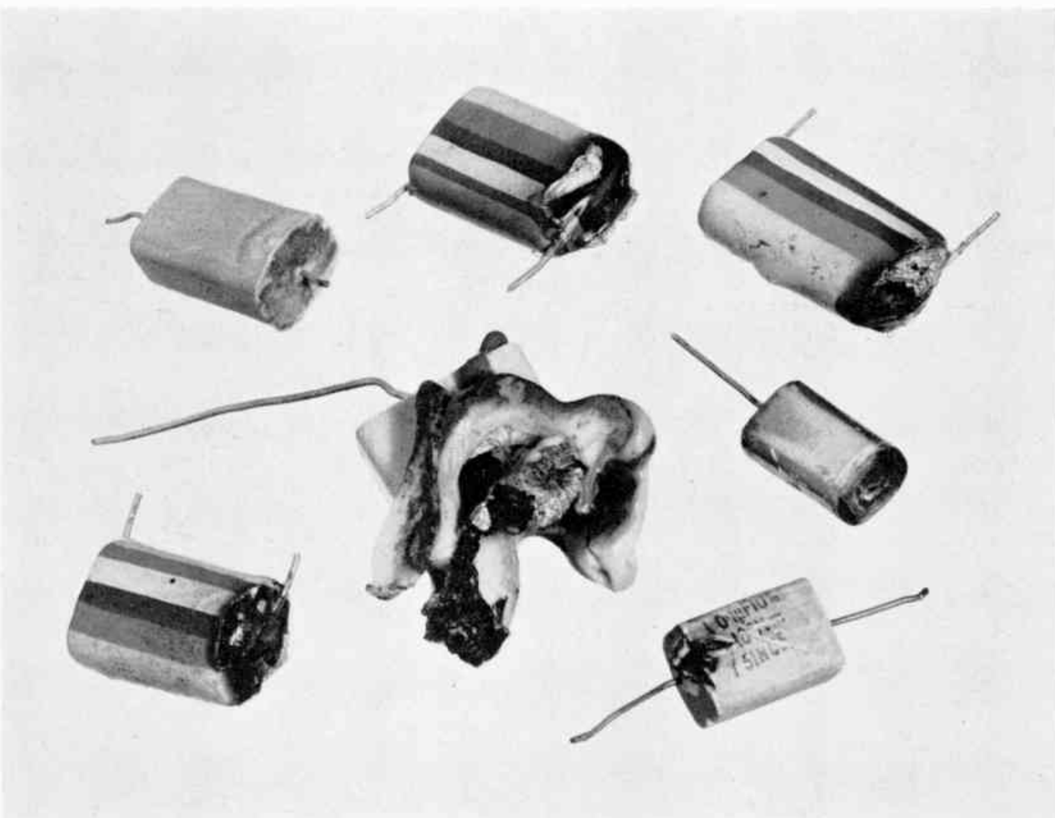


Fig. 5. Examples of types of failures.

capacitor is operated at low d.c. potentials, the power available to burn out a weak point is insufficient and a partial failure in insulation resistance could appear.

The most usual cause of failure with plastic film capacitors is due to exceeding their ratings or using them incorrectly. A frequent cause is using a d.c. rated component on a.c. mains as in interference suppression of small motors, and thyristor-power control circuits. In this case 700 V peak-to-peak is applied to the component and in a foil/film construction the occluded air in the windings will ionize and the resultant hot spots will eventually puncture the dielectric. Vacuum impregnation of the windings is not possible due to the sealing between the films. For this reason a.c. rated capacitors have an additional paper tissue wound with the plastic film and this acts as a carrier to remove the air and subsequently fill all the voids with a suitable impregnant.

Failure of a metallized construction used on low source impedance a.c. supplies is usually due to continual healing occurring or excessive current flow beyond the rating of the metallized layer and connexions. Continual healing can occur whenever the peak voltage from a low source impedance, i.e. less than 100 Ω, delivers sufficient peak power to start a chain breakdown of the weak points in the film. Excessive current flow usually occurs at higher frequencies such as television time-base and deflexion circuits and inverters operating at frequencies above 10 kHz. In this case the peak current could be several amperes which flows through the path of lowest resistance. The current density will therefore be extremely high at these points and the sprayed end sections will be burnt away from the metallizing. The process continues until the capacitor either ignites the dielectric and encapsulation or goes open circuit.

Examples of various types of failures are shown in Fig. 5.

The most recent film material available to capacitor manufacturers is polypropylene which is electrically between polycarbonate and polystyrene and the price structure of the raw material indicates that it will be less expensive than any of the existing materials. At the present time it is only available in limited quantities in minimum thickness of 10 μm which would be suitable for a 400 V d.c. rating. The dielectric constant is lower than P.E.T. or P.C. and consequently until much thinner films can be developed suitable for the mass market which requires 100 V–250 V rating, it is unlikely that it will be exploited for d.c. applications.

Manuscript received by the Institution on 6th May 1970. (Paper No. 1352/CC 90).

The Author

Mr. P. D. Habermel obtained his technical training during Military Service with R.E.M.E. and subsequently at Hendon Technical College where he obtained a Higher National Certificate in radio and electronics subjects. From 1950 to 1953 he worked on the development of electronic switching of X-ray diagnostic equipment with General Radiological Limited. He then joined Suflex Limited, initially as a sales engineer and subsequently became sales manager for plastic film capacitors. Since 1961 he has been with Mullard Limited as technical manager for capacitors within the consumer electronics division.

U.D.C. 534.88

Monte-Carlo Simulation of an Active Sonar

By

J. E. HUDSON, B.Sc., Ph.D.†

This paper describes the analysis of a non-linear sonar processing system by simple Monte-Carlo methods. The target detection capability with isotropic background noise and spatially coherent reverberation is found with attention being paid to the effects of frequency filters in the input channels from the transducers.

The resolution and detection of multiple and diffuse targets is assessed under noise free conditions and with the assumption of target scintillation due to forward scattering or medium inhomogeneity.

List of Symbols

- D mean absolute deviation of phases
- h impulse function
- i time variable
- j space variable (channel numbering)
- k space separation or time delay
- $\mathbf{x}, \mathbf{y}$ vector components of a signal phasor
- ϕ signal phase at an array element
- θ phase difference between two adjacent array elements
- σ^2 variance

1. Introduction

A sonar system which used hard limiting amplifiers and a digital processing system was recently constructed in this Department.[1,2] A theoretical prediction of the performance of the system was desirable for two reasons: firstly, to provide target detection curves against which the experimental curves could be checked to prove correct operation, secondly, to optimize the design for future systems.

As with many non-linear systems a purely analytic approach was found to be almost impossible and was abandoned in favour of a Monte-Carlo simulation on a digital computer. The system operates largely with signals having random properties such as noise, reverberation, and fluctuating targets and for these situations the Monte-Carlo method is admirably suited.

2. System Description

In concept the system is extremely simple even though the detailed circuitry is quite complex. The signals from a seven-element line receiving array are hard-limited after amplification to give seven square waves carrying only phase information. These seven signals are the processor inputs. The processor computes the phase differences between the six possible pairs of signals from adjacent elements. Two functions of this set of six phase differences are then computed. The first is the mean phase difference. This is proportional to the phase gradient across the receiving array and measures the bearing of the predominant target. The second is the dispersion of the phase differences about their mean and measures the signal/noise ratio of the predominant target.

The acoustic frequency is 48 kHz and the sampling rate is 16 kHz, i.e. every third period is sampled. This enables good range resolution to be obtained when a short pulse is transmitted from a nearby projector.

The display is arranged on a conventional oscilloscope. Range is measured horizontally along the time base which is synchronized to the transmitter pulses. The vertical deflection is continuously connected to the bearing (mean phase difference) output of the processing unit and therefore constantly fluctuates as the bearing of the predominant target in successive range increments varies. The display, in this form, is of little use since there is, as yet, no arrangement for eliminating the spurious bearings caused by random noise at ranges where there is no target. Noise is eliminated from the display by backing off the beam current of the cathode-ray tube and only applying a bright-up pulse when a target with a high signal/noise ratio is detected. This pulse is supplied by the processor which compares the phase difference dispersion to a pre-set threshold at each sampling and only if the dispersion is less than the threshold, proving that there exists a good plane wave at the array, is the pulse produced.

This display differs from a normal range/bearing or B scan display in that only one target can be displayed at each range increment, whereas a linear system could theoretically display seven, and in that the spots are all of uniform brightness. The processing system has the advantage that clipped signals are of uniform height irrespective of signal strength so that no time-varying gain is needed as in a linear system. Target strength is always measured relative to locally originating reverberation, or to the ambient noise level, whichever predominates.

3. Simulation Procedure

The aim was to simulate within a computer program the signals appearing at the array elements due to noise, reverberation, and plane waves from targets and

† Department of Electronic and Electrical Engineering, The University of Birmingham.

to sample and process these signals as realistically as possible so that accurate target detection curves could be drawn. The characteristics of the three types of signal present are quite different. Thermal noise is uncorrelated between the receiver channels. With good transducers and amplifiers acoustic noise in the medium may predominate over thermal noise but this may also be assumed to be uncorrelated between channels, if it is isotropic, so there is no need for distinction. Reverberation on the other hand is caused by backscattering from mud, fish etc., within the transmitter beam. If this is directional the scattering is localized and there is high correlation between the reverberation signals at the different array elements, especially adjacent ones.

If the bandwidth of the amplifiers is limited a further correlation is produced in the time domain. This affects the performance of the system in certain modes of operation.

Both noise and reverberation may be assumed to be Gaussian processes and therefore are quite easy to generate in a computer. The detailed aspects of this process are described in the following sections.

The signal from a target is a plane wave of constant amplitude but with a phase gradient across the channels depending on target bearing. For simplicity the target was assumed to be directly ahead, producing the same signal in each channel.

3.1 *Vector Representation of Uncorrelated Noise*

Because the reverberation and noise in the present case are narrow band they may be adequately represented in the form,

$$f(t) = x(t)\cos(\omega t) + y(t)\sin(\omega t) \qquad \ldots\ldots(1)$$

where x and y are functions of the time varying slowly in comparison to (ωt) and with a random Gaussian amplitude distribution and ω is 2π times the centre frequency. The phase of any channel signal at time t is given by $\tan\theta = y(t)/x(t)$. Thus, since the sonar works entirely in terms of relative phases, it is possible to simulate the system by working entirely with randomly generated values of the functions $x(t)$ and $y(t)$.

The first simulation runs were done under the assumption that the noise and reverberation had no time coherence between successive time samples of the phases. Thus successive values for $x(t)$ and $y(t)$ were uncorrelated.

Each sample value for $x(t)$ was formed by adding twelve independent samples from a rectangular distribution (the only distribution available in the computer) to form a truncated Gaussian variate. Each $y(t)$ sample was formed in the same way using a further twelve rectangular samples. Each rectangular sample has a variance of $\frac{1}{3}$ being drawn from a distribution uniform from -1 to $+1$ so the variance of each x and y sample is 4. Thus the variance of the noise process of equation (1) is $\frac{1}{2}(4+4)$ or 4.

Seven independent noise samples were generated, one for each channel. Together with any signal, the seven random samples were operated upon as described in Section 2 forming a trial value of mean phase and phase dispersion. These two values were stored. A second set of seven phases was then computed for another trial and so on. In each run of the program, 1000 trials were made forming 1000 samples each of mean phase and phase dispersion for statistical evaluation.

Because the length of each run is finite the noise power in each channel may not be the same even though derived from the same basic process. The expected error in the variance of the noise in each channel can be calculated quite easily and Aitken[3] shows that for 1000 trials the fractional error is about $(2/1000)^{\frac{1}{2}}$ or 0·2 dB. This is well within the accuracy needed in the simulation result.

Different runs of the program, with different signal conditions, all used the same set of vectors as noise so the set was carefully examined to ensure true randomness. Tests were carried out to measure (i) autocorrelation of successive complex vectors, (ii) cross-correlation between real and imaginary parts of each vector, (iii) cross-correlation in different channels. These tests showed that the generated noise had small correlations but not outside the limits which are permissible and expected for the sample size. Some difficulty was experienced with early programs using the *Mercury* Autocode system so testing of the noise is essential.

3.2 *Generating Noise with Time Coherence*

In one mode of operation the experimental system used a simple form of integrator which required that four consecutive time samples of the signal phase differences satisfy the threshold before displaying a picture point. Clearly the time coherence of the noise will affect the joint probability of all four samples passing the threshold.

The output time function of the tuned amplifiers is found by convolving the input time function with the impulse responses of the tuned circuits.[4] In the experimental system the phase sampling occurred at exactly one-third of the centre frequency and therefore the impulse response, $\exp(-t/\tau)$. $\cos(\omega t)$ was only used at the peaks of the cosine, i.e. at $\omega t = 0, 6\pi, 12\pi, \ldots$ etc. and consequently is just a sampling of $\exp(-t/\tau)$.

The complex noise vector samples are written in the form,

$$x_i, y_i, \quad i = 1, 2, 3, \ldots, 1000. \qquad \ldots\ldots(2)$$

and the impulse function in the form,

$$h_k, \quad k = 0, 1, \ldots, n. \qquad \ldots\ldots(3)$$

The impulse function must be truncated at the nth value h_n when it becomes insignificantly small, perhaps 1% of h_0, otherwise computing time is wasted.

The time coherent noise sequence, x_i', y_i' is formed by the convolution,

$$x_i' = \sum_{k=0}^{n} x_{i-k} \cdot h_k \qquad \ldots\ldots(4)$$

$$y_i' = \sum_{k=0}^{n} y_{i-k} \cdot h_k \qquad \ldots\ldots(5)$$

This convolution increases the variance of the noise in the program (though not in the real channel) in the ratio,

$$\sum_{k=0}^{n} (h_k^2) \qquad \ldots\ldots(6)$$

To complete the convolutions, n additional incoherent vectors must be supplied so that x_i, y_i run from $i = 1-n$ to 1000.

For the single tuned circuit it is not necessary to perform a complete convolution and the equivalent sequence is produced by the process, if $h_0 = 1$,

$$x_i' = h_1 x_{i-1}' + (1-h_1^2)^{\frac{1}{2}} x_i \qquad \ldots\ldots(7)$$

The convolutions above are valid only for narrow band filters with symmetrical frequency responses. Middleton[5] has discussed this question in detail.

3.3 *Introducing Inter-channel Correlation*

The noise inputs to the channels may be correlated between channels by virtue, in sonar systems, of mutual coupling between the array elements[6] or because the noise is in fact reverberation produced by a directional transmitter.[7,8] Coherence between the noise inputs at the different channels of the processor may be introduced into the simulation by a similar process to that used for time coherence. The inter-channel correlation function ψ_k is assumed to be known. This is the correlation between channel j and channel $j+k$.

An impulse function h_k must be found whose auto-correlation function is equal to ψ_k. A suitable function is easy to find if ψ_k has a recognized mathematical form but otherwise a suitable function may be found by trial and error.

The correlated noise is produced by the convolution

$$x_{ij}'' = \sum_{k=0}^{n} x_{i,\,j-k} \cdot h_k \qquad \ldots\ldots(8)$$

$$y_{ij}'' = \sum_{k=0}^{n} y_{i,\,j-k} \cdot h_k \qquad \ldots\ldots(9)$$

where i is the time variable, and j is the channel number. In the present case the channels were connected to a line array so that j may be also referred to as a space variable.

The variance of the noise is increased by the convolution in the ratio

$$\sum_{k=0}^{n} (h_k^2) \qquad \ldots\ldots(10)$$

3.4 *Addition of Signal Vectors*

A signal may be added to the noise by adding a constant vector x_s (or y_s) to each noise sample and the signal-power/noise-power ratio is given by the ratio

$$\frac{\langle x_s^2 \rangle}{\langle x'^2_i \rangle + \langle y'^2_i \rangle}$$

4. Results with the Digital Sonar System

The methods outlined above were used to simulate the operation of the experimental system. Signals of the following types were simulated:

(i) Plane waves in uncorrelated noise.

(ii) Plane waves in time-coherent noise, spatially uncorrelated.

(iii) Plane waves in spatially coherent noise but incoherent in the time.

Noise with both spatial and temporal coherence was not used because it was thought that the effects of time and spatial coherence would act independently.

After the generation of the signals additional processing was performed. From the generated vectors, the signal phases were obtained by the function, available in the computer,

$$\phi_i = \arctan(y_i/x_i), \qquad i = \text{channel number}.$$

The phase differences were obtained by subtraction and truncation as in the equipment.

$$\theta_i' = \phi_{i+1} - \phi_i \qquad \ldots\ldots(11)$$

if

$$-\pi < \theta_i' < +\pi, \qquad \theta_i = \theta_i' \qquad \ldots\ldots(12)$$

if

$$\theta_i' > \pi, \qquad \theta_i = \theta_i' - 2\pi \qquad \ldots\ldots(13)$$

if

$$\theta_i' < -\pi, \qquad \theta_i = \theta_i' + 2\pi \qquad \ldots\ldots(14)$$

The mean phase difference was the function

$$\bar{\theta} = \sum_{i=1}^{6} \left(\frac{\theta_i}{6}\right) \qquad \ldots\ldots(15)$$

while the phase deviation was equal to

$$D = \sum_{i=1}^{6} \frac{|\theta_i - \bar{\theta}|}{6} \qquad \ldots\ldots(16)$$

4.1 *Plane Waves in Uncorrelated Noise*

By generating seven concurrent uncorrelated series of noise, superposing signal vectors, and processing in the fashion outlined above it was possible to compute a cumulative probability function for the mean phase deviation. The same set of noise vectors were used for all values of signal/noise ratio to provide an accurate comparison of the results. The characteristics of systems with various numbers of channels were subsequently computed.

Figure 1(a) shows a typical detection characteristic for a fixed dispersion limit of $\pi/4$ and various numbers of channels. The false alarm rate is given by the detection probability for zero signal. In Fig. 1(b) the false alarm rate has been fixed at 0·01 corresponding to perhaps 40 spurious bright-ups in each range scan. The threshold must be varied as the number of channels is changed to keep this constant. Otherwise the variables are as before.

It is seen that the sensitivity of the system increases at roughly 2–4 dB per doubling of the number of channels. This is the sort of improvement—ideally 3 dB—which would occur with a linear processing system, for example, a fast scanning unit. If the width of each element were doubled keeping the number of channels the same the equivalent improvement in sensitivity would be had but in doing so the width of the scanned sector would be halved.

4.2 *Noise with Time Coherence*

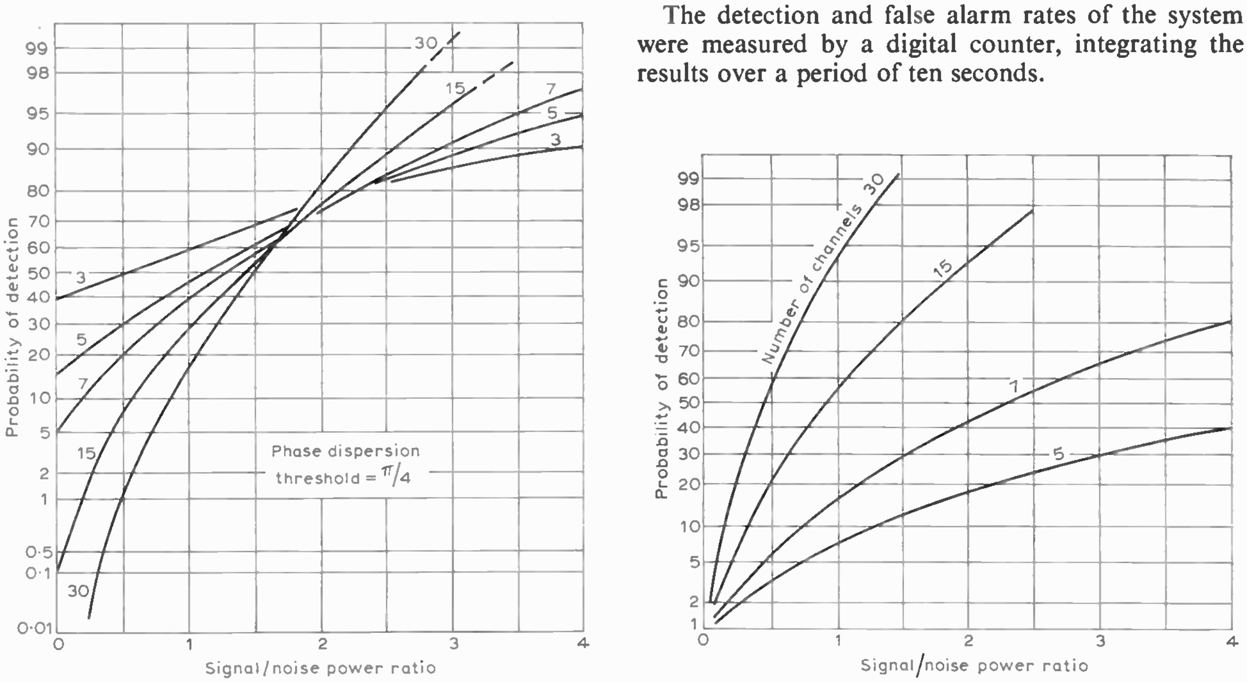
The experimental system used a multiple decision detector in which four time samples were taken consecutively and each had to satisfy the threshold before the bearing was displayed. A single tuned circuit was used in each channel with an impulse response given by (as in Sect. 3.2).

$$h_k = e^{-0 \cdot 48k} \qquad \ldots\ldots(17)$$

Time coherent noise was generated using this impulse function and after adding signal vectors the whole was processed and a detection curve was produced. This is shown in Fig. 2 together with some experimental measurements which show good agreement.

The experimental results were obtained by transmitting a c.w. signal from a submerged transducer at a range of about 20 metres so that a plane wave was produced at the receiving array. The outputs of seven independent narrow band noise sources were added into each channel and, by varying the intensity of these sources, the signal/noise ratio was artificially adjustable. The addition of synthetic noise was preferred to using the natural noise of the preamplifiers. It was thought that there was coherence of the self-noise in the channels because of coupling in the leads to the transducers, acoustic coupling from element to element, and also because the noise in the preamplifiers was partially set up by reception of local radio broadcasting transmitters.

The detection and false alarm rates of the system were measured by a digital counter, integrating the results over a period of ten seconds.



(a) Detection characteristics for fixed dispersion limit.

(b) Similar characteristic for fixed false alarm rate.

Fig. 1.

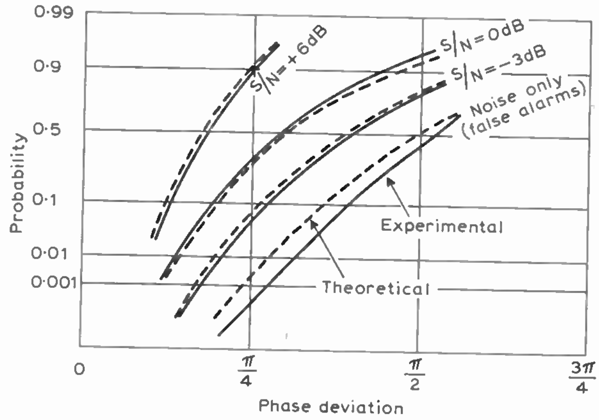


Fig. 2. Cumulative probability of the phase deviation function in a seven-channel system using four time samples.

4.3 *Reverberation with Inter-element Correlation*

This arises with a directional projector. The sonar system was used with either of two transmitting projectors. The first was 0·75 wavelengths wide by 5 wavelengths long vertically and had a horizontal directivity pattern 70° wide between the 3 dB points. The second was a binomial tapered array with eight sections horizontally, each section again being about five wavelengths long vertically. The shading function was in the amplitude ratio 1, 7, 21, 35, 35, 21, 7, 1, and the amplitude directivity pattern was of the form $(\cos\theta)^7$ with a 20° width between the 3 dB points.

For the first transmitter the impulse function giving the required correlation of the reverberation between the channels was $h_0 = 0{\cdot}98$, $h_1 = 0{\cdot}2$. For the binomial transmitter the correct function was found to be $h_0 = 1{\cdot}0$; $h_1 = 0{\cdot}86$; $h_2 = 0{\cdot}545$; $h_3 = 0{\cdot}08$.

A set of noise sequences was produced with this impulse function across the channels. Eight sequences for the $\frac{3}{4}$ wavelength and ten sequences with the binomial transmitter were necessary to form seven mutually coherent sequences for processing which afterwards proceeded just as before.

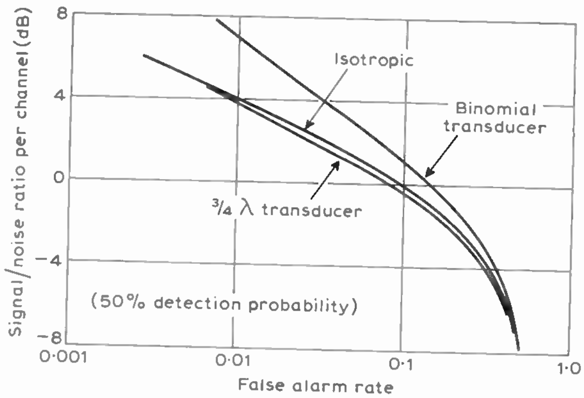


Fig. 3. Comparative system performance with three types of transmitter under reverberation limited conditions.

Figure 3 shows the comparative sensitivity of the system using the two projectors and thirdly with an isotropic projector. The binomial array shows a 3 dB loss in processing sensitivity which must be subtracted from its 10 dB directivity. The $\frac{3}{4}$ wavelength transducer shows very little loss and has a 3 dB directivity. Taking into account both system sensitivity and projector directivity the system is 3 dB more sensitive using the $\frac{3}{4}$ wavelength transducer and 7 dB more sensitive using the binomial array as compared to an isotropic projector under reverberation limited conditions. Under noise limited conditions the binomial array would show its full 10 dB gain. There is as yet no experimental confirmation of these figures.

5. Multiple Targets

The performance of the system under multiple target conditions was of great interest but is difficult to measure experimentally. To simulate multiple target conditions, targets with selected amplitudes and phases each radiating continuously were assigned positions at regular intervals of bearing and the signals which they would produce at a receiving array similar to the experimental array were calculated by vector summation. In this way the effect which the targets would produce in a pulsed system if they were all situated at the same annulus was reproduced. Many calculations of the mean bearing indicated by the system were made while the targets fluctuated.

The target amplitudes and phases were caused to fluctuate by sampling independently from Gaussian distributions as described in Section 3.1. Fluctuation of target strength is highly probable in underwater work arising from relative motion of target and sonar or disturbances of the acoustic field due to turbulence or surface scattering.

Figure 4 shows the display bearing probability density function $p(\bar{\theta})$ derived from 1000 trials with

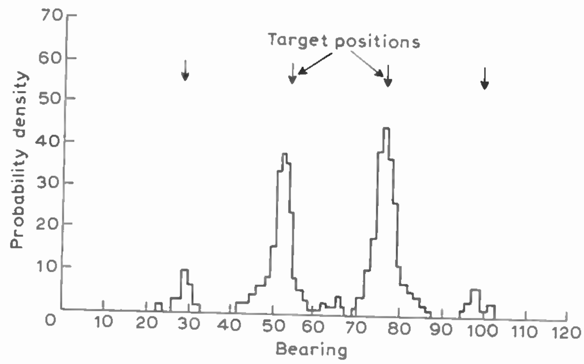


Fig. 4. The probability density function of the displayed bearing with four random targets.

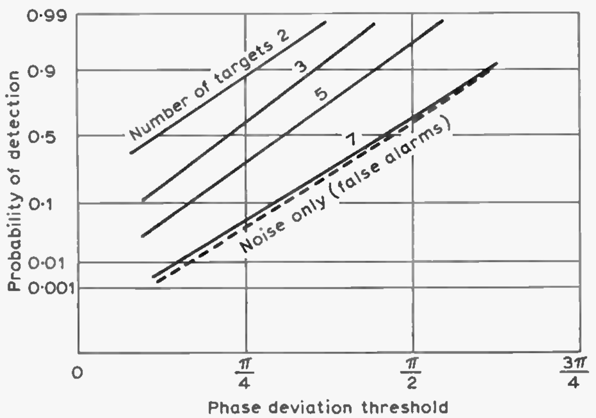


Fig. 5. Detection curves for various numbers of random targets.

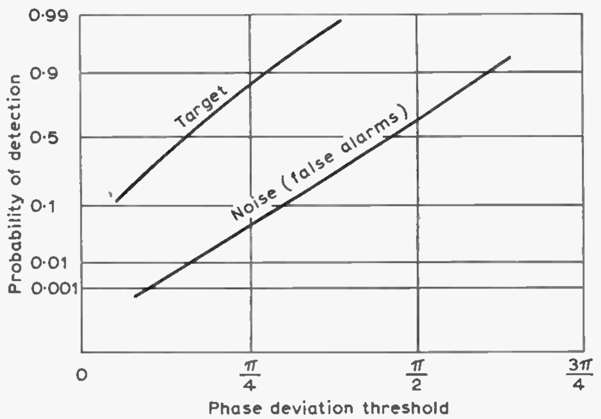


Fig. 7. The detection curve for the rectangular diffuse target.

four targets while Fig. 5 shows the detection characteristics for various numbers of targets spaced at intervals of about 20°.

This resolution of multiple targets using statistical methods was not expected nor has it been explained satisfactorily. In these tests there was always a tapering of probability of display for targets near the edges of the sector. This effect could be caused by a higher probability of phase truncation (Sect. 4) for these marginal targets.

Another type of target which occurs is the diffuse target such as a fish shoal or a rocky shore. These were simulated by placing 40 incoherent targets in a small sector of bearing. Figure 6 shows the histogram of bearings displayed for a rectangular target distribution of 50° width, while Fig. 7 shows the detection characteristics. Clearly the system gives a good representation of the target distribution but a large number of trials are needed to form the histogram and the detection performance is not good. The rectangular target of 50° width has a detection characteristic similar to that of a plane wave signal of +4 dB signal/noise ratio.

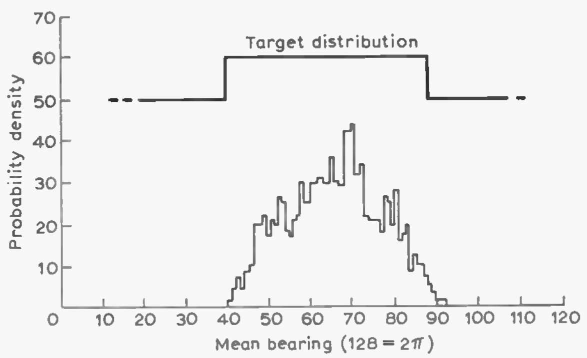


Fig. 6. The probability density function of the displayed bearing with a rectangular diffuse target.

6. Conclusions

Some representative results of a simple Monte-Carlo simulation of the sonar have been described. The most important conclusions which were drawn are as follows.

The detection sensitivity of the system improves by about 3 dB for each doubling of the number of elements in the array. This is the same relationship as would be found if a linear additive processing system were used with the same array and it is concluded that the prime factor governing sensitivity is the area of the array as with the linear system. The size of the scanned sector is governed by the width of the sections into which the array is divided. There is a requirement with the digital system of having at least five sections in the array or else the performance will be relatively poor due to sampling error.

The performance of the system in multiple and diffuse target environments was found to be good in that, given a large number of trials, the target configuration could be accurately imaged on a long persistence display screen. It was found that fluctuation of the target echoes was essential for this to occur. Unfortunately the rejection of background noise and reverberation is very poor because the threshold must be relaxed in order that multiple targets are accepted by the processing unit. If the average intensities of the echoes from targets in a cluster varies widely the strongest target will be displayed and the weaker ones suppressed.

These multiple target resolution results remain unconfirmed except for the case of two targets at the same range. Here the displayed bearing alternated between the positions of the targets as their relative phase varied. The imaging of distributed targets such as fish shoals and shore lines is satisfactory.[9] In this case there are many independent samples of the targets at successive range increments which satisfies the many trial requirement for just one transmitted pulse.

Other simulation programs, not described here, were used to find the effect of pulse length and receiver bandwidth under noise and reverberation limited conditions. The advantage of using a short pulse and wideband receiver amplifiers under reverberation limited conditions, an early design postulate, was verified.

7. Acknowledgments

The author would like to thank his colleagues for their help. Particular gratitude is due to Professor J. W. R. Griffiths, who directed the work, and to Dr. D. J. Creasey, who conducted field trials. The work was performed at Birmingham University and was supported by Professor D. G. Tucker.

8. References

1. Nairn, D., 'Clipped digital technique for the sequential processing of sonar signals', *J. Acoust. Soc. Amer.* 44, No. 5, p. 1267, November 1968.
2. Creasey, D. J. and Braithwaite, H. B., 'Experimental results of a sonar system with a digital processing unit', *Applied Acoustics*, 2, p. 39, 1969.
3. Aitken, A. C., 'Statistical Mathematics', Sect. 70 (Oliver and Boyd, London, 1962).
4. Lee, Y. N., 'Statistical Theory of Communication' (Wiley, New York 1963).
5. Middleton, D., 'Introduction to Statistical Communication Theory', (McGraw-Hill, New York, 1960).
6. Tucker, D. G., 'Correlation due to inter-element coupling in line arrays'. *The Radio and Electronic Engineer*, 36, No. 2, p. 95, August 1968.
7. Hudson, J. E., 'The Correlation of Reverberation in Sonar Systems', University of Birmingham, Department of Electronic and Electrical Engineering, Internal Memorandum No. 307.
8. Cron, B. F., Hassell, B. C. and Keltami, F. J., 'Comparison of theoretical and experimental values of spatial correlation'. *J. Acoust. Soc. Amer.* 37, No. 3, p. 523, May 1965.
9. Creasey, D. J. and Brauthwaite, H. B., *loc. cit.*, Figs. 8, 14 and 21–23.

Manuscript first received by the Institution on 8th January 1969, in revised form on 23rd February 1970 and in final form on 8th June 1970. (Paper No. 1353/AMMS34).

STANDARD FREQUENCY TRANSMISSIONS—October 1970

(Communication from the National Physical Laboratory)

Oct. 1970	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		Oct. 1970	Deviation from nominal frequency in parts in 10^{10} (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 60kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	− 299·9	+ 0·1	+ 0·1	702	742·1	17	− 299·9	+ 0·1	+ 0·1	690	632·4
2	− 300·0	+ 0·1	0	702	641·5	18	− 299·9	+ 0·1	+ 0·1	689	631·7
3	− 300·0	+ 0·1	+ 0·1	702	640·5	19	− 300·0	+ 0·1	0	689	630·5
4	− 299·9	0	+ 0·1	701	640·2	20	− 300·0	+ 0·1	0	689	630·3
5	− 299·9	+ 0·1	+ 0·1	700	639·6	21	− 299·9	0	+ 0·1	688	629·9
6	− 300·0	0	0	700	639·5	22	− 299·9	+ 0·1	+ 0·1	687	629·3
7	− 299·9	+ 0·1	0	699	638·9	23	− 299·9	+ 0·1	+ 0·1	686	628·5
8	− 299·9	0	0	698	638·5	24	− 300·0	—	+ 0·1	686	—
9	− 300·0	+ 0·1	0	698	637·7	25	− 299·9	—	+ 0·1	685	—
10	− 299·8	+ 0·1	0	696	636·8	26	− 299·9	—	+ 0·1	684	628·5
11	− 299·8	0	0	694	636·5	27	− 299·9	+ 0·1	+ 0·1	683	627·2
12	− 300·0	+ 0·1	+ 0·1	694	635·7	28	− 300·0	+ 0·1	+ 0·1	683	626·2
13	− 299·8	+ 0·1	+ 0·1	692	635.7	29	− 299·9	+ 0·1	+ 0·1	682	624·7
14	− 299·9	+ 0·1	+ 0·1	691	634·8	30	− 300·0	+ 0·1	+ 0·1	682	624·0
15	− 300·0	+ 0·1	+ 0·1	691	633·7	31	− 299·9	+ 0·1	+ 0·1	681	623·3
16	− 300·0	+ 0·1	+ 0·1	691	633·2						

Note: The frequency offset for 1971 will be -300×10^{-10}.

All measurements in terms of H.P. Caesium Standard No. 334, which agrees with the N.P.L. Caesium Standard to 1 part in 10^{11}.

* 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.

Contributors to this issue

Mr. W. R. Ogden (M.1961; G.1959) obtained his technical education at Paisley Technical College and the Royal College of Science and Technology Glasgow (now Strathclyde University). After civilian appointments with the Royal Electrical and Mechanical Engineers and the Admiralty Signal and Radar Establishment, he joined Ferranti Limited, Edinburgh, in 1958 and worked on the development of test techniques for monopulse radar systems. During this period he spent eighteen months in the Aerospace Division of Westinghouse Electric Corporation in Baltimore, Maryland. Mr. Ogden was appointed manager of the Test Equipment Group in 1968 and is responsible for the design, development and manufacture of special purpose test equipment for civil and military avionic projects. He is a past chairman of the Scottish Section of the Institution.

Mr. C. R. Thomas joined the R.A.F. as an apprentice in 1938, finally retiring after 15 years service to join the Guided Weapons Department of Vickers-Armstrongs. Prior to his present position, he was for some six years a group leader of the electronic instrumentation group of Brush Electrical Engineering Company's research division at Loughborough. He has been with Ferranti Limited at Edinburgh since 1961, and he is currently support systems evaluation engineer in their electronic systems department. Mr. Thomas was a contributor to 'Electronic Engineer's Reference Book' on the subject 'Electrical/mechanical analogies applied to the vibratory/oscillatory condition'.

Dr. J. E. Hudson graduated in electrical engineering at Birmingham University in 1964. He then continued in the Department to do research work on underwater propagation and clipped signal processing toward the degree of Ph.D. which he received in 1968. Dr. Hudson then worked for about a year at the G.E.C.-A.E.I. Stanmore Laboratory on sonar processing systems. At present he holds a research fellowship at Birmingham which commenced in 1968; his work is concerned with target classification and mutual coupling effects in sonar arrays.

Dr. G. R. Whitfield read natural science (physics) at Trinity College, Cambridge, and subsequently obtained his Ph.D. in radio astronomy at Cambridge for a radio star survey at 38 MHz, work on radio star spectra, and observations of *Sputniks 1* and *2*. From 1958 to 1966 he was at the Royal Radar Establishment, Malvern, where he worked mostly on c.w. radar research. Since 1966 he has held an appointment as lecturer in electronics in the Department of Applied Physical Sciences, University of Reading. Dr. Whitfield is particularly interested in flying and gliding and his present paper is associated with this interest.

Mr. Norman Wheatley (M. 1970) served during the war with the Royal Navy on radio and radar maintenance. He joined Cable and Wireless Limited in 1947 and has served at the Company's branches in Aden, Mombasa, Nairobi, Singapore, Accra, Hong Kong and Bahrain. He is currently the Cable and Wireless Regional Engineer (Gulf) and Manager (Engineering), Bahrain, where he has been closely associated with the early on-site planning and subsequent operation of the Bahrain Earth Station.

Dr. M. S. Qureshi received his B.Sc. degree in electrical engineering from the West Pakistan University of Engineering and Technology, Lahore, in 1962. After serving as an assistant engineer with the Water and Power Development Authority, Lahore, for about 15 months, he returned to the University as a lecturer in the Department of Electrical Engineering. Under the Overseas Training Scheme of that University, he came to the Electronics Department of the University of Southampton in October 1967 to work on semiconductor devices for a doctorate. On completing his Ph.D. studies in August 1970, Dr. Qureshi returned to Lahore.

A note on the career of **Mr. K. G. Nichols** (F. 1966; M. 1960) was published in the July 1970 issue of *The Radio and Electronic Engineer;* he is now a reader in the Electronics Department of the University of Southampton.