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."

VOLUME 41 No. 5

MAY 1971

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

An Independent Communications Authority

NEW electronic techniques of communications, many of them postulated and shown to be practicable and 'cost-effective' years ago, are gradually becoming realities. Some of the prospects and the problems were foreseen in the first of the Institution's Post-war Development Reports:* in addition, anticipated growth of data transmission networks, area mobile radio telephones, and electronic telephone exchanges is now taking place. Certain of these new facilities are natural extensions to the operations of the Post Office which has been rationalized in recent years by administrative separation of its telecommunications side from the postal services, giro and savings bank. The desirability of such a change was the subject of an editorial in the Institution's Journal nearly six years ago† but one of the other proposals made in that article has *not* been progressed.

Briefly, the case was argued for the establishment of a 'National Communications Authority', which was envisaged as an independent body on which would be represented *all* classes of users of communications systems, including defence services, broadcasting, navigational services for ships and aircraft, commercial, medical, industrial and amateur users and, of course, the telecommunications side of the Post Office. Some of the functions of such an authority were outlined in the Institution's recommendations to the Committee on Broadcasting (the 'Pilkington' Committee)‡ as long ago as 1961 and an analogy was made with the Federal Communications Commission in the United States. In 1961 the question of television broadcasting standards (number of lines, type of colour system) was a particularly thorny issue which might well have been resolved sooner had there been a controlling body.

The concept of a 'Ministry of Communications' appeared to be accepted by the appointment in 1970 of a Minister of Posts and Telecommunications. The Post Office Corporation is, however, directly under the aegis of this Ministry; thus the responsibility of the Minister for matters of, for instance, frequency allocation sets the Post Office Telecommunications in an anomalous position *vis-à-vis* the other users.

An assessment of the present position of communications in Great Britain with reference to the future was recently made by Professor H. M. Barlow in a paper prepared for the National Electronics Council§ in which the difficulty of an interested party acting as arbitrator was stressed. The planning of a truly comprehensive network of telephones and data links as well as the extension of cable distribution and, of course, the linking of overseas communications by satellites represent complex tasks which must necessarily sometimes conflict with the claims of other services.

The proper utilization of resources is necessary whether these be water or oil, or radio frequencies. Ill-founded frequency allocations can be considered to pollute the electromagnetic environment and one important function of some form of independent communications authority would be to provide the disinterested control necessary to achieve the greatest possible utilization with the minimum of r.f. interference.

F. W. S.

* 'Post-war Development Report—Part 1', *J.Brit.I.R.E.*, 4, pp. 134–50, October/December 1944.

† 'A ministry of communications', *The Radio and Electronic Engineer*, **30**, No. 1, p. 3, July 1965.

‡ 'Radio and television broadcasting in Great Britain', *J.Brit.I.R.E.*, **21**, No. 5, pp. 379–87, May 1961.

§ Barlow, H. M., 'Telecommunication services in the U.K.—future development and overall policy', *National Electronics Review*, **7**, No. 2, pp. 30–31, March/April 1971.

Contributors to this issue

Mr. G. H. Kyte served an engineering apprenticeship with B.T.-H., Rugby, from 1929 to 1935, and obtained the degree of B.Sc.(Eng.) of London University by part-time study. Subsequently he worked for J. Stone & Co., Deptford, on the development of small voltage regulators and with the Sperry Gyroscope Co. on aircraft navigational equipment. During the past ten years Mr. Kyte has been manager in charge of development of n.d.t. equipment and techniques with T.I. Steel Tube Division. He serves on several B.S.I. Committees concerned with n.d.t. specifications and standards.

Mr. K. R. Whittington read for the Mechanical Sciences Tripos at Christ's College, Cambridge, obtaining his degree in 1953. After National Service as an Engineer Officer in the Royal Navy, he joined Sperry Gyroscope in 1955 where he worked on inertial guidance systems. In 1959 he took up employment with Racal Communications and from then until 1961 worked on s.s.b. mobile communications equipment. Since 1961 Mr. Whittington has been the head of electronics at the Research Laboratories of Tube Investments where his work mainly concerns quality control and n.d.t. instrumentation.

Mr. R. K. Sharma (G. 1971) was awarded the degree of B.E.(Hons.) in Electrical Engineering from the Government Engineering College, Jabalpur, in 1965. From 1966 until 1967 he was a Pilot Officer in the Technical Electrical Branch of the Indian Air Force. Since 1968 Mr. Sharma has been with Instrumentation Limited, Kota, initially as an engineer in the Research Division, and he is at present in the Company's Development Division.

Mr. K. R. Haigh (Member 1950, Associate 1947), after receiving his technical education at Hull Technical College, served as an Electrical Officer in the R.N.V.R. from 1943-46. After a few years in industry, he joined the Royal Naval Scientific Service in 1952 and is at present a Principal Scientific Officer at the Admiralty Underwater Weapons Establishment at Portland. From 1968-70 he was an exchange scientist with the U.S. Naval Oceanographic Office and was a crew member of the submersible *Ben Franklin* during its thirty-days' Gulf Stream Drift Mission. Mr. Haigh has also carried out oceanographic research in a number of other submersibles and currently holds the diving record for a U.K. citizen of a dive to 6180 ft. In addition to these underwater activities, Mr. Haigh is also a qualified diver. He is the author of several papers and articles; arising out of his interest in industrial archaeology he has published a book on the submarine cable industry.

Mr. S. J. Lent (M. 1957; G. 1956) joined the B.B.C. during the war and started his technical training on transmitters at Daventry. He later served with the R.A.F. where he trained as a navigator and after the war rejoined the B.B.C. in the Engineering Research Department. He spent some time in the sound and recording sections and later transferred to the television group of the Department. Mr. Lent has since been engaged in a wide range of subjects in the television field including the earlier colour television experiments, telerecording, standards conversion and colour television systems. At present he is an engineer in the image scanning section and is mainly concerned with colour television cameras and displays.

Mr. S. G. Allen (M. 1967; G. 1958) received his technical education at the South East Essex Technical College and subsequently gained the degree of M.Sc. in telecommunications systems from Essex University in 1970. After serving a four-years' apprenticeship with the Marconi Company, he joined the G.E.C. Applied Electronics Laboratories in 1959 and worked on the development of microwave components for guided weapons. Since 1964 he has been a senior research engineer at Standard Telecommunication Laboratories where he is currently concerned with digital microwave radio systems.

Biographical notes on Messrs. F. H. Wise and D. R. Brian follow their paper on page 212.

U.D.C. 621.396.65:621.376.3:621.376.56

A Comparison of P.C.M. and F.D.M.–F.M. Microwave Radio Systems

By

S. G. ALLEN,

M.Sc., C.Eng., M.I.E.E., M.I.E.R.E.†

The present and future position of p.c.m. microwave line-of-sight systems is reviewed in relation to the more conventional f.d.m.–f.m. microwave systems. A comparison is made of the relative noise performance, transmitter power and system bandwidth between the two systems. The effect of rainfall attenuation is discussed.

1. Introduction

With a few exceptions, the whole of the present microwave communication network of the U.K. is composed of f.d.m.–f.m. systems. At the present time the use of p.c.m. transmission is largely confined to 24/32 channel systems over existing junction cables, where it has proved to be an attractive economic proposition. Wider acceptance of p.c.m. transmission is inhibited by interface problems with the existing f.d.m. analogue network and by the large bandwidth required for the transmission of p.c.m. telephone channels. Apart from these problems, p.c.m. offers numerous advantages such as transmission over long distances with small impairment, the ability to multiplex speech, high-speed data and television in a common digital stream, and low-cost switching of the digital stream. For microwave transmission in particular, the tolerance to interference enables the spectrum utilization to be comparable to f.d.m.–f.m.

In addition to these basic advantages, there are new forms of transmission media such as glass fibres using laser sources and long-haul waveguide carrying 40 GHz underground, both of which will almost certainly use digital transmission. The enormous capacity of these media could eventually reduce the cost of bandwidth and enable the transmission of digitally-encoded television and viewphone to be realized economically. These concepts lead one to visualize and predict an all-digital communication network to exist at some time in the distant future.

The present interest in p.c.m. microwave systems is related to two of the above developments:

(i) The large number of 24/32 channel systems, which must eventually be multiplexed to a higher capacity for communication between concentrations of these systems. This could lead logically to 96 channel groupings, or 6·3 Mb/s digital streams, which could carry multiples of 24/32 channels or viewphone. In general, it is uneconomic to use microwave transmission for less than 24 channels.

(ii) The very large capacity systems. Microwave radio may be capable of carrying up to 500 Mb/s digital streams, which could be used to feed long haul waveguide systems, or be used as an independent transmission system. 500 Mb/s is equivalent to 7500 telephone channels or four digital colour television channels. Because the microwave spectrum required for the transmission of 500 Mb/s is wide, it precludes the use of the existing microwave communications bands below 10 GHz. Above 10 GHz, rainfall attenuation becomes severe and reduces the spacing between repeaters, to the extent where the spacing is about 4 km at 40 GHz. Such spacings are unacceptable for f.d.m.-f.m. line-of-sight microwave systems if the C.C.I.R. recommended noise performance is to be realized, so that the higher microwave frequency line-of-sight systems will probably be used only for the transmission of p.c.m. or other digital signals. A problem for the design engineer is to make such routes competitive with alternative communication systems.

P.c.m. microwave repeaters have been designed for intermediate numbers of channels (e.g. 240 channels) but the problem here is to find suitable sources of traffic. P.c.m. multiplex for more than 32 channels is uncommon, so for the present time and for the near future the market for microwave systems having intermediate numbers of channels must be dominated by f.d.m.–f.m.

However, the frequency bands currently being used by f.d.m.–f.m. microwave line-of-sight systems (i.e. 2, 4 and 6 GHz) are becoming saturated in the developed countries, which forces the systems planners to take a hard look at the higher microwave frequencies and consequent digital transmission.

An evaluation of digital microwave radio system performance follows below. This comprises a comparison of the noise performance, transmitter power, bandwidth, modulation and effect of rainfall attenuation relative to f.d.m.–f.m. systems.

2. Noise Performance

The major technical difference between p.c.m. and f.d.m. microwave radio systems is the noise performance. P.c.m. signals may be regenerated and therefore do not accumulate noise in the same way as f.d.m.–f.m. analogue systems. This permits close spacing between p.c.m. microwave repeaters and consequently allows the possibility of operating at frequencies above 10 GHz where rainfall attenuation reduces the repeater spacing.

2.1 *F.D.M.–F.M. Noise Performance*

The noise performance of f.d.m.-f.m. microwave systems has been studied extensively and is tabulated as a detailed list of recommendations by the C.C.I.R.[1]

† Standard Telecommunication Laboratories Limited, London Road, Harlow, Essex.

Briefly, the recommendations are that the noise power at a point of zero relative level in any telephone channel on a 2500 km hypothetical reference circuit for f.d.m.–f.m. radio relay systems should not exceed:

(a) 7500 pWp† mean power in any hour.

(b) 7500 pWp one minute mean power for more than 20 % of any month.

(c) 47 500 pWp one minute mean power for more than 0·1 % of any month.

(d) 1 000 000 pW for more than 0·01 % of any month.

For systems carrying a given number of telephone channels the C.C.I.R. also recommend the r.m.s. frequency deviation, usually 100 or 200 kHz.

Figure 1, curve (b), shows the effect of these recommendations on the signal/thermal noise performance of a typical 120 channel f.d.m.–f.m. microwave radio system. Calculations for this and other systems are given in Appendix 10.1.

The restriction of 7500 pWp is an important one when comparing with p.c.m. microwave radio systems. For a multi-hop microwave route each hop contributes to the total noise so that as the repeater spacing is reduced the permitted noise per hop becomes less. For an f.d.m–f.m. route using reduced repeater spacing the thermal noise contribution per hop could be reduced by increasing the transmitter power but the intermodulation noise contribution per hop could not similarly be reduced. One therefore rapidly approaches a repeater spacing below which it is not possible to make an f.d.m.–f.m. route that would meet the C.C.I.R. recommended noise performance. For microwave systems requiring small repeater spacing it is therefore necessary to use digital transmission with regeneration to achieve high quality performance.

2.2 *P.C.M. Noise Performance*

The C.C.I.R. are studying p.c.m. microwave systems but a list of recommendations does not exist at the present time. Because of the different nature of noise associated with p.c.m. systems there is some debate as to what should and should not be considered acceptable.

The essential differences between the two systems are illustrated in Fig. 1 which shows the signal/noise ratio in a telephone channel as a function of the received carrier to noise ratio. For the condition of zero fade, digit or bit errors due to thermal noise are negligible, curve (a). There is, however, another source of noise termed quantization noise which is generated by the terminals and not by the transmission medium, curve (c).

The magnitude of this noise is determined by the choice of the number of quantization levels, but although of apparently large magnitude, this form of noise is quite acceptable, because the quantization noise is directly proportional to the signal level (i.e. the signal/quantization noise is constant), over the major part of the signal dynamic range which is encountered in practical systems.

For a p.c.m. microwave system therefore, one cannot sensibly specify a mean noise performance of 7500 pW0p as is possible with f.d.m.–f.m. systems.

† pWp refers to psophometric power in picowatts.

The way in which the noise performance of p.c.m. microwave systems will almost certainly be specified[2] is that a certain maximum error rate will be permitted for a certain percentage of any month.

The most likely figures will be either

(a) a maximum error rate (of possibly 10^{-3}) for not more than 0·1 % of any month,

or

(b) a maximum error rate (of possibly 10^{-3}) for not more than 0·01 % of any month.

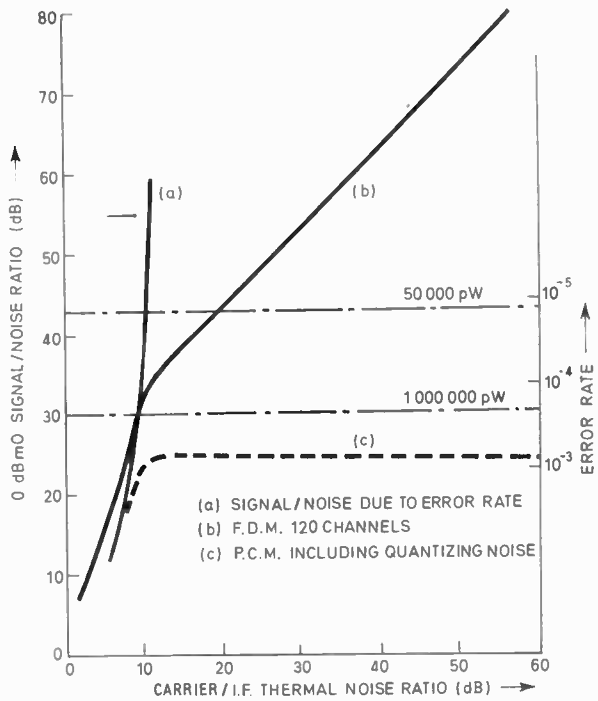


Fig. 1. Comparison of p.c.m. and f.d.m. microwave systems (single hop).

Assuming Rayleigh fading, the difference in the transmitter power required for these two cases is 10 dB. On the other hand, differences in the choice of maximum error rate have relatively little effect on the transmitter power as indicated in Fig. 1.

By relating the above performances to those of comparable f.d.m.–f.m. microwave systems it is observed that the f.d.m.–f.m. system performance degrades gradually past the 47 500 pWp noise level, whereas the p.c.m. system performance degrades abruptly beyond the 10^{-3} error rate level, as the received carrier signal fades.

The choice of case (a) would therefore tend to give a worse system performance than equivalent f.d.m.–f.m. systems, whereas the choice of case (b) would give a better system performance, but unfortunately uneconomically because higher transmitter powers would be required than for equivalent f.d.m.–f.m. systems.

For the present calculations a compromise solution assuming a 40 dB fading margin for a 50 km hop is assumed.

3. Rainfall Attenuation

The main effect of rainfall is to increase the path attenuation and this effect becomes rapidly more noticeable as the microwave frequency is increased. For microwave frequencies above 10 GHz rainfall attenuation eventually predominates over multi-path attenuation, the exact frequency depending upon the geographical location. Fading due to rainfall attenuation creates a number of serious problems.

(a) Simple diversity techniques do not help, although it may be possible to use another frequency band in the lower microwave frequencies or alternatively use route or geographical diversity.

(b) Increasing the transmitter power by 10 dB has relatively small effect on the repeater spacing in areas of heavy rainfall as seen from Fig. 2.

(c) For very small percentages of the time, multi-path fading tends towards Rayleigh fading, for which the fading margin has to be increased by 10 dB for a reduction of 10 times in outage time; whereas for rainfall attenuation the statistics of rainfall are such that the fading margin in decibels has to be approximately doubled for 10 times reduction of outage time.

(d) A large proportion of deep fades due to multi-path propagation are of milliseconds duration, which can be tolerated by the telephone network. With rainfall attenuation a large proportion of deep fades are of several minutes duration, which can cause the loss of telephone calls which have already been set up. (This loss of calls is made slightly worse for multi-path fading when p.c.m. transmission is used because loss of signal for 0·5 ms will cause loss of synchronization of the digital stream. Synchronization is normally re-established within 5 ms on recovering the signals, so that the effective duration of deep fades is increased by up to 5 ms.)

The British Post Office have developed a model for rainfall attenuation,[3] which takes into account the fact that high density rainfall tends to occur in localized cells of generally less than 2 km diameter. A set of curves based on the model has been published and these account for most of the experimentally observed rainfall attenuation on microwave links. For outage times less than or equal to 0·1 % the curves can be described by the following equation:

$$A = B \times 1{\cdot}88^{(\log_{10} 1/10z)}\ 1{\cdot}63^{(\log_{10} y/10)}$$

where A is the rainfall attenuation margin in dB
z is the percentage outage time
y is the path length in miles
B is a constant which is a function of frequency and of geographical location.

The following table gives some values of the constant B at a frequency of 11 GHz:

S.E. England	5·6
Western region U.S.A.	6·9
Gulf Coast U.S.A.	24

The way in which constant B varies as a function of frequency may be determined from graphs in the literature and from more recent Post Office measurements,[4,5] but the way in which constant B varies as a function of geographical location is something that must ultimately be determined experimentally, although it may be possible to relate the constant B to the total annual rainfall at a given location together with the number of hours per year that a certain rainfall rate is exceeded at that location.[6,7]

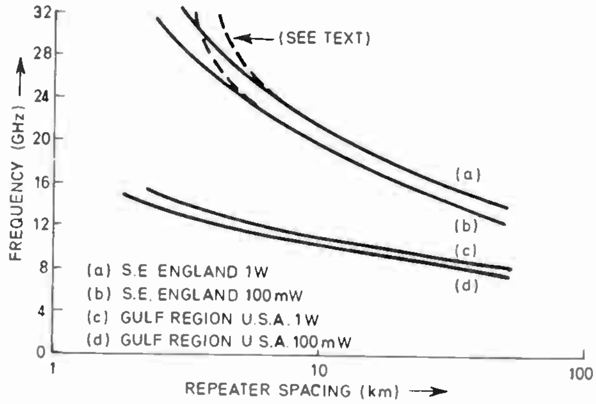


Fig. 2. Repeater spacing 250 km route (32 channels 0·01 % outage).

Using the above equation to calculate the fading margin, and the path loss formulae of Appendix 10.3, the repeater spacings as a function of frequency for a 250-km route have been plotted for a total outage time of 0·01 % in Fig. 2. A 32-channel system having transmitter powers of 100 mW and 1 W has been assumed, together with aerial beamwidths of 1°.

The dotted curves appearing on Fig. 2 are an intuitive allowance for the fact that for small repeater spacing the rainfall attenuations of each hop are no longer uncorrelated and therefore the solid curves in these regions give a pessimistic answer.

Route or geographical diversity has been shown to be technically feasible to the extent that a second parallel route placed at 2 km distance from the first route could reduce the total outage time due to rainfall by a factor of ten,[8] when the output of the two routes was combined. However, this particular solution may prove to be an expensive one. It could be less expensive if there were three routes of which any two could be selected, similar to the way in which telephone traffic is switched to an alternative route if one route is blocked.

4. Transmitter Power

Unlike f.d.m.-f.m., transmitter powers for p.c.m. microwave radio are determined by calculating the carrier/noise ratio corresponding to the maximum error rate, then adding a suitable fading margin. For example, if phase reversal modulation with differential detection is used then the theoretical carrier/thermal noise ratio required to achieve an error rate of 10^{-3} is approximately 8 dB, and is independent of the number of telephone channels. Since the thermal noise power at the receiver is directly proportional to the receiver equivalent noise bandwidth, and the receiver noise bandwidth is directly proportional to the number of telephone channels, it follows that for a given route the transmitter power required is directly proportional to the number of telephone channels. The transmitter power, when measured against the number of telephone channels on a logarithmic scale therefore has a 3 dB per octave slope as drawn in Fig. 3.

This situation does not apply to f.d.m.–f.m. transmitter powers. The transmitter powers for f.d.m.–f.m. are determined by the modulation index of a test tone. The r.m.s. frequency deviation of the test tone is recommended by the C.C.I.R., and is typically 200 kHz. So long as this frequency deviation remains the same for different numbers of telephone channels, as for example for 300, 600 and 960 channels, then the test tone modulation index is inversely proportional to the number of telephone channels, and the transmitter powers for f.d.m.–f.m. therefore increase at 6 dB per octave when measured against the number of telephone channels on a logarithmic scale. This difference is illustrated in Fig. 3, which shows the different test tone frequency deviations, such as 100 kHz, 140 kHz, and 200 kHz recommended by the C.C.I.R. for different numbers of channels, which cause small departures from the basic trend of 6 dB per octave. Nevertheless, the outcome is that for large numbers of telephone channels, p.c.m. microwave requires less transmitter power than f.d.m.–f.m. whereas the reverse is true for small numbers of channels. This could be a significant conclusion, because it implies that solid-state transmitters can be used for very large capacity systems that would otherwise require the use of travelling wave tubes if f.d.m.–f.m. techniques were used.

Pre-emphasis has been taken into account in the above reasoning. For a system with a given number of channels, say 120 channels, the C.C.I.R. recommend a pre-emphasis network which has the effect of giving approximately constant phase deviation over the whole baseband in order to equalize thermal noise between the top and bottom channels. Similarly for a 960-channel system there is a pre-emphasis network which achieves the same objective. In comparing the p.c.m. and f.d.m.–f.m. transmitter powers above, however, it is the average phase deviation for the 120-channel system which is compared with the average phase deviation for say the 960-channel system which gives rise to the basic 6 dB per octave slope.

Calculations of the f.d.m. and p.c.m. transmitter powers are given in Appendix 10.1.

For small capacity systems, Fig. 3 indicates that there is not much to choose between the transmitter powers required for p.c.m. and those required for f.d.m.–f.m. (For 24-channel systems a fade margin of 35 dB provides a better comparison with f.d.m.–f.m.) In fact, the transmitter powers are likely to be higher for small capacity p.c.m., because Fig. 3 assumes four-phase modulation with coherent detection whereas for small capacity systems binary f.m. is a more likely form of modulation.

An interesting possibility exists if one considers the use of diversity for p.c.m. Assuming that the fading is due to multi-path then the use of diversity results in a considerable reduction in the required fading margin and consequently a considerable reduction in the required transmitter power for a p.c.m. microwave

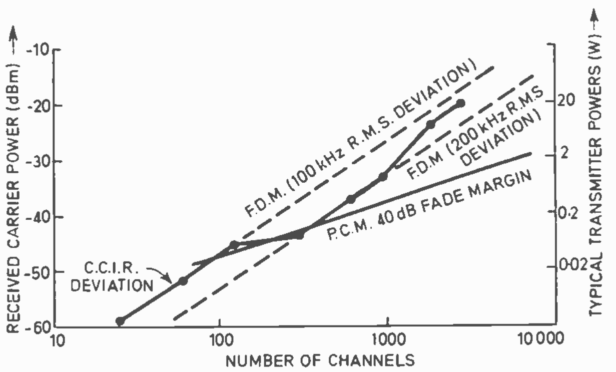


Fig. 3. P.c.m. and f.d.m. transmitter powers.

system. This is not possible for an f.d.m.–f.m. system because there must always be sufficient received carrier power to meet the C.C.I.R. mean noise recommendation.

5. Bandwidth

P.c.m. transmitted over cable normally requires excessive bandwidth when compared to f.d.m. cable systems, but because f.d.m.–f.m. microwave is extravagant of bandwidth and p.c.m. is tolerant to adjacent channel interference it is possible, for small to medium numbers of channels, to transmit the same number of telephone channels over the same microwave frequency allocation as is used by an f.d.m.–f.m. system.

This may be seen simply as follows; if one wishes to make a 120-channel f.d.m.–f.m. system then the frequency allocation in the 7 GHz band is 7 MHz spacing between bearers. By straightforward division this amounts to a spectrum allocation of 58·3 kHz per telephone channel.

For p.c.m. using 8 digits and 8 kHz sampling the bit rate per telephone channel is 64 kbits/s which has a filtered spectrum of 64 kHz, and as a double-sideband signal becomes 128 kHz per telephone channel. For transmission over a microwave link, four-phase or quaternary modulation can be used, together with cross-polarization (assuming that sufficient cross-polarization isolation can be maintained), which amounts to sending the same bearer frequency on the opposite plane of polarization. These two facilities enable four times the traffic capacity to be carried in the same spectrum, which is equivalent to only 32 kHz per telephone channel.

In more detail, the spectrum required for the transmission of an f.d.m.–f.m. signal may be derived from Carson's rule and the Holbrook and Dixon channel loading factors. This has been done in Appendix 10.2 and is compared with the spectrum required for the transmission of p.c.m. using both full-raised-cosine and half-raised-cosine filters. The results are shown in Fig. 4 and illustrate that for small numbers of channels the spectrum required for p.c.m. is less than for f.d.m.–f.m., but that for large numbers of channels the spectrum required for p.c.m. is approximately twice that required for f.d.m.–f.m. Four-phase modulation, together with the efficient use of spectrum possible with p.c.m., which is described later in this Section, was assumed for Fig. 4.

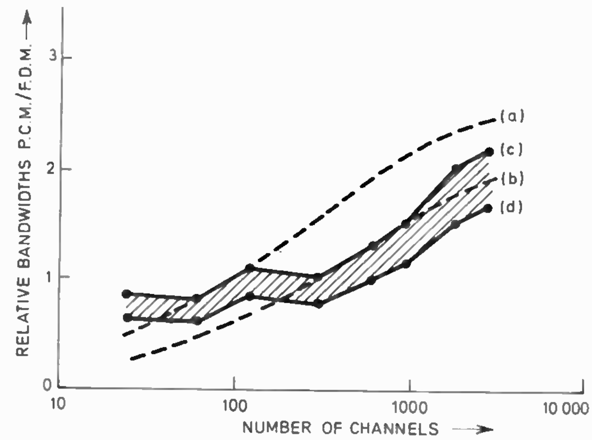


Fig. 4. Relative bandwidths for p.c.m. and f.d.m.-f.m.

	F.D.M.-F.M.	P.C.M.
(a)	100 kHz r.m.s. deviation.	Full raised cosine.
(b)	200 kHz r.m.s. deviation.	Full raised cosine.
(c)	C.C.I.R. deviation	Full raised cosine.
(d)	C.C.I.R. deviation.	½ raised cosine.

It is to be noted from Fig. 4 that the general trend is for the p.c.m. system performance to be worse than f.d.m.–f.m. for large numbers of channels, whereas the graph for transmitter powers (Fig. 3) showed the opposite trend. This leads one to conclude that for the system parameters chosen, if there was freedom to choose any frequency deviation for f.d.m.–f.m., then there would not be much difference between the system performance of p.c.m. and f.d.m. when compared on a transmitter power-bandwidth basis.

This is in fact confirmed by Fig. 5 which shows the power bandwidth relation between the two microwave systems for different numbers of channels when a 40 dB fade margin is selected for the p.c.m. system and varying deviation for the f.d.m. system. If some of the receiver threshold margins[9] are allowed, then the p.c.m. system performance becomes worse than that for f.d.m., but if on the other hand the necessary fade margin is less than 40 dB, then the p.c.m. system performance becomes better than f.d.m. However the point is somewhat academic since one cannot sensibly use f.d.m.–f.m. at the high microwave frequencies.

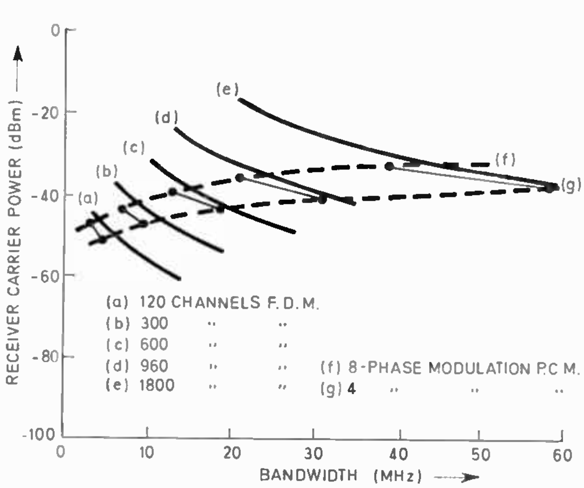


Fig. 5. Power bandwidth relation between p.c.m. and f.d.m.

The more efficient use of the spectrum possible with p.c.m. arises because the Carson's rule bandwidth is normally the bandwidth over which a linear amplitude and phase response is required from the channel dropping filter. These linearity requirements necessitate a filter with a 3 dB bandwidth which is greater than the Carson's rule bandwidth and with a sharp cut-off beyond this point to provide high attenuation at the adjacent channel frequencies.

For the transmission of p.c.m. in a communications band having a fixed channel spacing, the spectrum can be used far more efficiently by increasing the bit rate until the adjacent spectra just begin to overlap, thus enabling the p.c.m. spectrum to occupy the full space between adjacent channels as shown in Fig. 6(b).

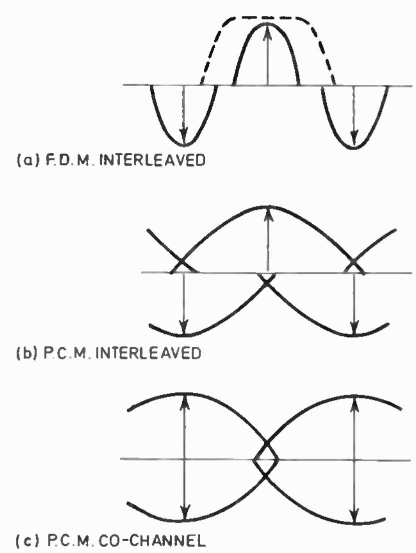


Fig. 6. Channel arrangements for p.c.m. and f.d.m.

Such a method of utilizing the spectrum is to be preferred to the simple cross-polarization referred to earlier because it reduces the numbers of bearers, but cross-polarization can be used efficiently as indicated in Fig. 6. This figure also shows the conventional channel arrangements for an f.d.m.–f.m. system for comparison.

The channel dropping filters can become an integral part of the overall transmission filter characteristic. This overall filter characteristic is designed to give minimum intersymbol interference at the receiver, but must also be allocated in a compromise way between the receiver i.f. filter, channel filters and transmitter band limiting filter, all of which require either minimum bandwidth or maximum selectivity.

For optimum reception of p.c.m.–p.m. in the presence of Gaussian noise, a cosine filter has been assumed for the receiver with a similar cosine filter for the transmitter (suitably modified for rectangular pulses) giving a resultant of a raised cosine spectrum at the receiver detector.

However, for microwave systems having a high carrier frequency or a small number of channels, the carrier frequency tolerance is of concern and the raised-cosine response is less ideal. The 3 dB bandwidth of the raised-cosine response is fixed by the bit rate and any attempt to increase the bandwidth to allow for the carrier frequency tolerance simply increases the intersymbol interference. A Gaussian type of filter response is then to be preferred because it permits the bandwidth to be increased, to allow for carrier frequency tolerance, without increasing the intersymbol interference. The carrier frequency spacing between adjacent channels is nevertheless similar to that used for the raised-cosine calculations and so the bandwidth calculations remain representative.

Further parameters for p.c.m. microwave filters and for adjacent channel interference are given in Appendix 10.4.

6. Modulation

6.1 *Modulation Systems*

Studies of the various digital modulation systems have been made recently by the C.C.I.R. The results of the C.C.I.R. studies have been tabulated as the steady state signal power divided by the noise power density and bit rate for the different methods of modulation. The figures are for an error rate of 10^{-6} and represent the relative transmitter powers required for the different methods of modulation.

Angle modulation, which gives better noise immunity, is most suitable for microwave transmission. The C.C.I.R. figures[2] for signal power divided by the noise power density and bit rate for angle modulation are as follows:

Phase modulation with coherent detection	Binary	10·5 dB
	4-level	10·5 dB
	8-level	13·8 dB
Phase modulation with differentially coherent detection	Binary	11·4 dB
	4-level	12·8 dB
	8-level	16·8 dB
Frequency modulation discriminator detection	Binary	13·4 dB
	Duobinary	15·9 dB
	4-level	20·1 dB
	8-level	25·5 dB

The relative bandwidths of the different levels of modulation are given by

$$\text{bandwidth} = \log_2 (\text{number of levels})$$

From the above figures it is evident that phase modulation with coherent detection gives the lowest transmitter power for a given error rate and that four-phase modulation can be used to double the information rate without any increase in transmitter power relative to binary modulation if coherent detection is used. It is therefore not surprising that this modulation technique is proposed for digital satellite systems where the maximum performance must be obtained from the system. Nevertheless, there are other considerations which lead to the choice of alternative modulation methods.

For tropospheric scatter systems, for example, differential detection is preferred because it is many times less sensitive to unwanted frequency or phase modulation effects which may arise due to multi-path propagation or differential time delays in the arrival of digital signals. This problem will be less severe with microwave line-of-sight systems, but could cause difficulty in the transmission of 500 Mb/s where each pulse length is only 2 ns.

Recovering the carrier is a necessary part of the detection of a phase-modulated signal and the instrumentation to achieve this is simpler for differentially coherent detection than for coherent detection. For small capacity systems, however, the change in the phase shift of the delay line as a function of temperature and of the carrier frequency tolerance can cause problems in the use of differential detection.

A typical method of recovering the carrier in a binary coherent detection system is to multiply the incoming signal by two, which removes the phase modulation, then divide by two and lock the phase of the resulting signal to a voltage-controlled oscillator which is then used as the reference carrier for detection. The reference carrier so established may have either of two phases separated by 180°, thereby introducing an ambiguity in the detected signal. This ambiguity is avoided by differentially encoding the digital stream at the transmitter, so that the signal is then modulated in terms of transitions rather than absolute values. Additional logic is required in the receiver to decode the signal.

A similar differential encoding technique is used for differential detection systems but for this case the detection system automatically decodes the signal.

If one is prepared to trade a few decibels of transmitter power for a simpler and consequently lower cost modulation system, as would normally be required for small capacity systems, then the use of other modulation methods becomes attractive.

For example, double sideband amplitude modulation with envelope detection, which requires 6·5 dB more transmitter power than phase modulation with coherent detection to achieve the same error rate of 10^{-6}, is being used in some current small capacity p.c.m. microwave systems because of the simplicity of instrumentation and consequent low cost.

A better compromise could be binary f.m. with discriminator detection which requires only 2·9 dB more transmitter power than coherent phase detection. If binary f.m. is used then a peak-to-peak frequency deviation of about 0·65 times the bit rate is required to achieve optimum spectrum occupancy. Multi-level f.m. systems are simple, but are not a sensible choice because of the rapid increase in the required transmitter power as the number of levels is increased; although a possible exception to this is the duobinary f.m. technique which achieves three-level modulation by passing a binary stream (suitably coded) through a narrow bandwidth filter.

6.2 *Codes*

Apart from the question of modulating the microwave carrier, there is also the question of choosing a suitable code for the digital stream. Most junction cable p.c.m. systems use alternate mark inversion (a.m.i.) to remove the d.c. component and to provide adequate timing in an economical manner. Such a code is wasteful of spectrum for microwave applications. The requirements for a microwave system are also different. Removal of the d.c. component is not absolutely essential as it is for line systems. It is nevertheless convenient to remove it because the baseband amplifiers can then use a.c. coupling and the possibility exists of inserting supervisory and speaker circuits at the low frequency end of the spectrum. The other requirement of the code is to improve the timing content and here the situation is also different. A line system comprises large numbers of regenerators, all of which are extremely limited as to power consumption and all of which must be of minimum cost. A microwave system on the other hand has repeaters spaced at 40 km compared to a line system spacing of 2 km, does not necessarily require regenerators at every repeater, and has comparatively unlimited power to drive the regenerator. Consequently, a phase-locked regenerator is possible which requires a smaller density of timing information. A digital code suitable for a microwave system would therefore achieve the necessary timing content and removal of the d.c. component with the minimum redundancy possible.

A small percentage redundancy on a high-speed digital stream should easily provide sufficient numbers of digits to enable a delta modulation omnibus speaker circuit plus all the supervisory and alarm signals to be inserted into the main digital stream. Thus drop and insert of these added facilities at intermediate repeater stations becomes extremely simple and does not involve the use of filters. The added digits can also be used to provide a continuous monitor of error rate, which in turn can be used to initiate switching of the spare channel.

7. Conclusions

The commonly used frequency bands for line-of-sight microwave systems are becoming increasingly saturated in the developed countries, so that systems planners are being forced to explore the possibility of using the higher microwave frequencies above 10 GHz for line-of-sight applications.

The basic disadvantage associated with these higher frequency bands is that rainfall attenuation reduces the spacing between repeaters so that more repeaters are required for a given route length. Apart from economic considerations, the noise allocation per repeater becomes proportionately smaller, so that a situation is soon reached where the use of f.d.m.–f.m. techniques is not tenable if the C.C.I.R. noise recommendations are to be met.

The obvious technical solution to this problem is to use digital modulation techniques or more specifically p.c.m. Digital signals can be regenerated at each repeater and therefore performance is sensibly independent of the number of repeaters. Such a technical solution fits in well with the future development of digital telecommunications, where the use of 24-channel p.c.m. is growing rapidly throughout the world.

With regard to the problem of rainfall, the path attenuation caused by rainfall increases very rapidly as the rainfall rate is increased and as the microwave frequency is increased. In the common carrier band at 6 GHz, the microwave link outage time caused by rainfall is so small that these links can be used almost anywhere in the world. The reliable operation of microwave links at 18 GHz is not possible in areas of tropical rainfall, and even in the south east of England it is necessary to reduce the repeater spacing to 15 km to achieve an outage time of 0·01 % over a 250 km route at this frequency. For frequencies above 18 GHz, the geographical limitations are severe, although even at 36 GHz serious consideration is being given to digital microwave line-of-sight repeaters spaced by only a few kilometres. It is to be noted that for these applications outage time can be reduced considerably by using route diversity, since high density rain is usually concentrated in cells of less than 2 km diameter.

The overall conclusion emerging is that the use of digital microwave transmission will become increasingly preferred as the use of frequencies extends above 10 GHz and that the use of f.d.m.–f.m. will become increasingly more difficult as the use of frequencies extends above 10 GHz. At frequencies above 40 GHz the use of line-of-sight microwave systems will become untenable and systems using such frequencies will be forced to go underground into enclosed waveguides. One might therefore expect to see the emergence of digital microwave line-of-sight systems in the frequency range 10 GHz to 30 GHz carrying high capacity digital traffic such as digitally-encoded viewphone, television or feeds to long-haul waveguide or satellite terminals. It is interesting to consider that the available spectrum between 10 GHz and 30 GHz is twice as wide as the available spectrum below 10 GHz.

One of the most serious problems to be faced in introducing these types of digital microwave links into international communications networks is the problem of outage time. There is, as yet, no general agreement on whether one should design for a hypothetical route to have an outage time of 0·1 % or 0·01 %, even though the difference in the transmitter powers required for these two cases is large, particularly if the outage time is dictated by rainfall. Secondly, and perhaps more important, is that outage time due to multi-path fading is generally made up of thousands of fades, many of short duration, so that individually they do not seriously affect the communications network. Outage time due to rainfall on the other hand is generally made up of a smaller number of fades each of several minutes duration, so that these fades would cause the loss of telephone calls which had already been set up, unless of course an alternative communication route was provided during these occasions. For the transmission of data the pattern of longer outage times but longer periods of low error-rate transmission might be preferred.

The question of the large bandwidth required for p.c.m., which is a frequent argument against its use, is largely solved for microwave systems by the tolerance to adjacent channel interference. This tolerance permits the same carrier or bearer frequency to be used on both planes of polarization simultaneously (assuming that sufficient cross-polarization isolation can be maintained), and also permits the spectrum of adjacent channels on the same plane of polarization partially to overlap. These combined effects enable p.c.m. microwave systems to occupy less spectrum than f.d.m. microwave systems for small capacity systems (less than 200 channels). For large capacity systems (2000 channels) p.c.m. occupies approximately twice the spectrum of an equivalent f.d.m system.

On the other hand, p.c.m. requires less transmitter power than f.d.m. for large capacity systems, so that when a power-bandwidth system performance comparison is made, there is very little to choose between p.c.m. and f.d.m. microwave.

Preferred methods of modulation for digital microwave systems are phase modulation with differential detection for small capacity systems and phase modulation with coherent detection for large capacity systems. For cases where one is prepared to pay a transmitter power penalty of a few decibels, such as for small capacity systems, there are other modulation systems, such as frequency modulation with discriminator detection or amplitude modulation with envelope detection. These other systems are attractive because of the simple instrumentation required.

It has been an implicit assumption in the conclusions so far that digital microwave systems are associated with reduced repeater spacing. For small capacity systems carrying a few multiplexed 24/32 channel p.c.m. inputs, the use of the microwave spectrum can be more efficient than for f.d.m.-f.m., so that one may well see these systems in the frequency bands below 10 GHz where there are no rainfall problems. A further incentive is that p.c.m. multiplex terminal equipment is cheaper than f.d.m. terminal multiplex equipment.

For large capacity systems it is more reasonable to associate digital microwave with the higher microwave frequencies and hence with the problems of reduced repeater spacing, low-cost repeaters, high equipment reliability and rainfall attenuation.

Solutions to these problems are to be found in route diversity, mast mounted repeaters having small aerials, small power consumption, small Fresnel zone clearance, reduced civil works and installation costs, integrated circuit design and high traffic capacity reducing the cost per channel.

It is a challenge to engineers to solve these problems, but it is for these reasons that the final usage of these systems is speculative at the present time.

In the meantime, one might expect to see high-capacity digital systems in the frequency range 10 to 18 GHz, where relatively large repeater spacings can be used.

8. References

1. C.C.I.R. Vol. IV, 'Radio Relay Systems', Part 1. Oslo 1966.
2. Conclusions of the Interim Meeting of Study Group IX, Geneva, p. 98, September 1968.
3. Turner, D. and Easterbrook, B. J., 'Prediction of attenuation by rainfall in the 10·7–11·7 GHz communication band', *Proc. Instn Elect. Engrs*, **114**, No. 5, pp. 557–65, May 1967.
4. Baker, A. E. and Brice, P. J., 'Radio wave propagation as a factor in the design of microwave relay links', *P.O. Elect. Engrs J.*, **62**, Pt. 4, pp. 239–46, January 1970.
5. Turner, D. J. W. and Turner, D., 'Attenuation due to rainfall on a 24 km microwave link working at 11, 18 and 36 GHz', *Electronics Letters*, **6**, No. 10, pp. 297–8, 14th May 1970.
6. Russak, S. and Easley, J., 'A' practical method of estimating rainfall rate frequencies directly from climatic data', *Bull. Amer. Met. Soc.*, **39**, No. 9, 1958.
7. Gameson, A., 'The yearly distribution of rainfall intensities', *Met. Mag.*, June 1965, p. 173.
8. Hogg, D. C., 'Millimeter wave communication through the atmosphere', *Science*, **159**, pp. 39–46, January 1968.
9. Yoshida, K. *et al.*, '2 GHz microwave p.c.m. system', *Japan Telecomms. Rev.*, **11**, No. 1, pp. 18–29, 1969.
10. Plotkin, S. C. and Medhurst, R. G., 'F.m. bandwidth as a function of distortion and modulation index', *I.E.E.E. Trans. on Communication Technology*, **COM-16**, No. 3, p. 500, June 1968.
11. C.C.I.T.T. Blue Book, Vol. 3, p. 84.

9. Selected Bibliography

9.1 *Digital Transmission*

Neu, W. and Kündig, A., 'Project for a digital telephone network', *I.E.E.E. Trans. on Communication Technology*, **COM-16**, No. 5, pp. 633–47, October 1968.

Ito, S. *et al.*, 'Microwave communication system using p.c.m.', *NEC Res. Dev.*, No. 9, pp. 98–108, April 1967.

Conference on 'Digital Satellite Communication', London November 1969 (I.E.E. Conference Publication No. 59).

Yoshida, K. *et al.*, '2 GHz microwave p.c.m. system', *Japan Telecomms. Rev.*, **11**, No. 1, pp. 18–29, 1969.

Richards, D., 'Transmission performance of telephone networks containing p.c.m. links', *Proc. Instn Elect. Engrs*, **115**, No. 9, pp. 1245–58, September 1968.

Endacott, K., 'P.c.m. radio relay systems', *Telecommunication J. Australia*, pp. 257–260, October 1968.

9.2 *Modulation Systems*

Whelan, J. W., 'Analogue f.m. versus digital p.s.k. transmission', *I.E.E.E. Trans. on Communication Technology*, **COM-14**, No. 3, pp. 275–82, June 1966.

Cahn, C., 'Performance of digital phase modulation communications systems', *I.R.E. Trans. on Communication Systems*, **CS-70**, pp. 3–6, May 1959.

Tjhung, T. T., 'Band occupancy of digital f.m. signals', *I.E.E.E. Trans. on Communication Technology*, **COM-12**, No. 4, pp. 211–16, December 1964.

Lender, A., 'Correlative level coding for binary data transmission', *I.E.E.E. Spectrum*, **3**, pp. 104–15, February 1966.

Shaft, P., 'Error rate of p.c.m.-f.m. using discriminator detection', *I.E.E.E. Trans. on Space Electronics and Telemetry*, **SET-9**, No. 4, pp. 131–137, December 1963.

Miyauchi, K., ' A design method of filtering system for radio p.c.m., *J. Instn Elect. Commun. Engrs, Japan*, **51**, No. 8, pp. 15-17, August 1968.

9.3 *Economic Considerations*

Cosgrove, T. and Chipp, R. D., 'Economic considerations for communication systems', *I.E.E.E. Trans. on Communication Technology*, **COM-16**, No. 4, pp. 513–25, August 1968.

Dolan, H. and Lockley, N. A., 'Power supply requirements for low consumption communication equipment', Joint U.K.A.E.A.-E.N.E.A. International Symposium on Industrial Applications for Isotopic Power Generators, September 1966, pp. 77–83.

9.4 *Propagation*

Pearson, K., 'Method for the prediction of the fading performance of a multisection microwave link', *Proc. Instn Elect. Engrs*, **12**, No. 7, pp. 1291–99, July 1965.

Turner, D. and Easterbrook, B. J., 'Prediction of attenuation by rainfall in the 10·7–11·7 GHz communication band', *Proc. Instn Elect. Engrs*, **114**, No. 5, pp. 557–65, May 1967.

Turner, D., Easterbrook, B. J. and Golding, J. E., 'Experimental investigation into radio propagation at 11·0–11·5 Gc/s', *Proc. Instn Elect. Engrs*, **113**, No. 9, pp. 1477–89, September 1966.

Weibel, G. E. and Dressel, H. O., 'Propagation studies in millimeter-wave link systems', *Proc. Inst. Elect. Electronics Engrs*, **55**, No. 4, pp. 497–513, April 1967.

Hickin, E. M., 'Atmospheric attenuation of radio signals above 10 GHz', *GEC-AEI Journal*, **35**, No. 3, pp. 133–36, 1968.

Forbes, E. J., 'Planning 13 GHz t.v. relay systems', *I.E.E.E. Trans. on Broadcasting*, **BC-14**, No. 1, pp. 19–24, March 1968.

Hogg, D. C., 'Millimeter communication through the atmosphere', *Science*, **159**, pp. 39–46, January 1968.

9.5 *Relevant Textbooks*

Bennett, W. R. and Davey, J. R., 'Data Transmission'. (McGraw-Hill, New York, 1965)

Carlson, A. B., 'Communication Systems: An introduction to signals and noise in electrical communication'. (McGraw-Hill, New York, 1968)

Schwartz, M., Bennett, R. and Stein, S., 'Communication Systems and Techniques'. (McGraw-Hill, New York, 1966)

Yonezawa, S., *et al.*, 'Microwave Communication', p. 346. (Maruzen, Tokyo, 1963)

Panter, P. F., 'Modulation, Noise and Spectral Analysis'. (McGraw-Hill, New York, 1965)

Goldman, S., 'Frequency Analysis, Modulation and Noise'. (McGraw-Hill, New York, 1948)

10. Appendices

10.1 *Received Carrier Powers*

(a) F.D.M.–F.M.

A calculation is shown for 24 channels. Other channel capacities are extrapolated from this calculation.

Assumed parameters: route 2500 km, 50 hops at 50 km per hop, pre-emphasis 8 dB, baseband 12–108 kHz, deviation 100 kHz r.m.s., receiver noise figure 8 dB.

To meet the 7500 pWp noise requirement, 2500 pWp is assumed to be allocated to thermal noise, whilst the remainder is allocated to intermodulation noise, local oscillator, modem, co-channel, etc.

The permitted total mean thermal noise power for the route is therefore 2500 pW0p or −56 dBm0p. For each repeater the permitted mean thermal noise is 50 pW0p or −73 dBm0p.

Calculations to find the thermal noise in the top telephone channel are as follows; the f.m. improvement in the top channel is given by

$$S = 20\log\left(\frac{F}{f_m}\right)+P_1+P_2\ \text{dB}$$

where F is the r.m.s. frequency deviation

f_m is the maximum modulating frequency

P_1 is the pre-emphasis in the top channel

P_2 is the psophometric weighting

Therefore

$$S = 20\log\left(\frac{100}{108}\right)+4+2{\cdot}5$$
$$= 5{\cdot}8\ \text{dB}$$

The thermal noise in a 3·1 kHz telephone channel is −139 dBm. Therefore the thermal noise (T) in the top telephone channel referred to the receiver input, with a receiver noise figure N_f,

$$T = -139-S+N_f$$
$$= -139-5{\cdot}8+8$$
$$= -136{\cdot}8\ \text{dBm}$$

The relationship between the carrier/i.f.-noise ratio (C/N_1) and the test tone signal/noise ratio in a telephone channel (S/N_2) is given by

$$\left(\frac{C}{N_1}\right)\text{dB} = \left(\frac{S}{N_2}\right)\text{dB}-20\log m-10\log\frac{B}{2f_2}$$

where B is the i.f. bandwidth

m is the modulation index

f_2 is the telephone channel bandwidth

Including the other f.m. advantages of pre-emphasis and psophometric weighting, the equation can be usefully modified to give the received carrier power as

$$C = \left(\frac{S}{N_2}\right)\text{dB}+T$$
$$= 73-136{\cdot}8 = -63{\cdot}8\ \text{dBm}$$

−63·8 dBm is the carrier power which must be received under the mean conditions, and since one can expect an average or simultaneous fade of approximately 5 dB on all hops, the received carrier power under 'no fade' conditions.

$$C_0 = -58{\cdot}8\ \text{dBm}$$

Under 'no fade' conditions the thermal noise power per repeater will be −78 dBm0p.

Similar calculations have been made for other channel capacities and are as follows:

Channels	24	60	120	300	600	960	1800
Deviation	100	100	100	200	200	200	140
C_0	−58·8	−51·5	−44·7	−42·8	−36·6	−32·9	−24·1

With these received carrier powers the fading margin available is usually sufficient to meet the noise requirement that the one minute mean noise power should not exceed 47500 pW0p for more than 0·1% of any month.

(b) P.C.M.

For p.c.m.–p.m. using binary coherent detection the probability of error (P_e) is related to the carrier to noise ratio by

$$P_e = \tfrac{1}{2}\left(1-\text{erf}\sqrt{\frac{C}{N}}\right)$$

For an error rate of 10^{-3}, which exists on only one link at any one time, the carrier/noise ratio is 6·8 dB.

From Appendix 10.4, the equivalent noise bandwidth of the receiver filter is 1·2 times the bit rate for a Gaussian characteristic. Continuing the example for 24 channels, the receiver noise bandwidth becomes 1·2 × 1·54 = 1·85 MHz. To achieve an error rate of 10^{-3} the received carrier power must be

$$-174+8+62{\cdot}7+6{\cdot}8 = -96{\cdot}5\ \text{dBm}$$

which also applies to 4-phase modulation.

However, there are other considerations that will finally determine the minimum received carrier power for a p.c.m. microwave system. The figure of -96.5 dBm assumes a perfect receiver.

More realistically, in a practical system the overall transmission bandwidth is restricted to reduce spectrum occupancy and this results in a reduction of the detected pulse amplitude. Jitter, drift of slice level, deviation of the phase shift of the delay line in a differential detector, temperature, ageing, intersymbol interference, adjacent channel interference and carrier tolerance can all contribute to increase the required figure for the minimum received carrier power by up to 6 dB.[9]

For the present calculations a margin of 4 dB is used which accounts for intersymbol interference, temperature, ageing, etc, but does not allow for external interference.

The minimum received carrier power for 24 channels is therefore -92.5 dBm.

For the purpose of the graph shown in Fig. 1, the signal to error rate noise (S/N_e) in a decoded p.c.m. telephone channel can be approximated by

$$\frac{S}{N_e} = \frac{1}{4P_e} \text{ for } P_e < 10^{-2}$$

10.2 *Bandwidth Comparison*

The bandwidth occupied by the f.d.m. spectrum is given by

$$B_1 = 2f_1(mk+1) \qquad \text{......(1)}$$

where m is the modulation index

f_1 is the maximum modulation frequency

and k is the multi-channel loading factor.

Equation (1) is Carson's law bandwidth for retention of the significant sidebands and negligible distortion of the demodulated signal. Correspondence on f.m. bandwidth as a function of distortion and modulation index[10] has shown that Carson's law gives only -90 dB distortion due to band limiting. It can therefore be assumed that any energy outside the bandwidth given by equation (1) can be removed by filtering without distorting the signal.

To obtain a similar bandwidth for p.c.m. a radio frequency bandwidth equal to twice the bit rate is required if a full raised cosine transmission characteristic is used, or a bandwidth equal to 1·5 times the bit rate if a ½ raised cosine (or 50% roll off) transmission characteristic is used. The bandwidth is therefore

$$B_2 = 2R \text{ or } 1{\cdot}5R \qquad \text{......(2)}$$

where R is the bit rate

For the f.d.m. case, values of K have been published for various numbers of channels with and without peak limiting.[11] The occurrence of peaks for which limiting occurs is 10^{-5}, and since this case corresponds more closely with the bandwidths of existing radio links, the published figures for peak limiting are used for calculating bandwidths.

Substituting into equation (1) and (2) gives Table 1.

The bandwidths give the spectrum occupied by f.d.m. and p.c.m., but take no account of the relative tolerance to adjacent channel interference, of multi-level p.c.m. modulation, or of carrier frequency tolerance.

For high capacity p.c.m., four-phase modulation with coherent detection will most probably be used, which enables the bandwidth to be halved whilst using the same transmitter power. The bandwidth figures in Table 1 can therefore be reduced by ½ for p.c.m. systems using four-phase modulation. Further increase of modulation levels tends to be uneconomic because of the rapid increase of transmitter power and modem complexity.

The tolerance to adjacent channel interference enables p.c.m. to make more efficient use of a given channel spacing as is indicated in Fig. 6. Figure 6(a) shows how the conventional f.d.m. spectra are located on vertical and horizontal polarizations. The dotted line represents the channel dropping filter characteristic which has a linear amplitude and phase characteristic over the frequency band occupied by the f.d.m. spectrum and has high attenuation at adjacent channel frequencies. The spectra of the adjacent channels do not overlap even though they are located on alternate planes of polarization.

Table 1

Number of channels	F.d.m. r.m.s. deviation (kHz)	Bandwidth (MHz) F.d.m.	P.c.m. raised cosine	P.c.m. ½ raised cosine
24	35	0·71	3·08	2·28
	50	0·92		
	100	1·63		
60	50	1·44	7·68	5·75
	100	2·37		
	200	4·23		
120	50	2·30	15·36	11·5
	100	3·49		
	200	5·88		
300	200	9·55	38·4	28·8
600	200	14·45	76·8	57·5
960	200	19·9	122·8	92·1
1800	140	27·9	230	172·5
2700	140	38·6	346	260

With p.c.m. the more efficient use of the channel spacing is indicated in Fig. 6(b). A comparison with f.d.m. could be made with the knowledge that the adjacent channel spectra can be permitted to overlap to an extent where the separation between channels is about 1·7 times the bit rate.

A simpler comparison (slightly pessimistic for the p.c.m. case) is made by observing from Fig. 6(a) that by duplicating the p.c.m. spectra on the opposite planes of polarization the traffic capacity is at least doubled relative to f.d.m. and so the p.c.m. bandwidths in Table 1 can be reduced again by ½. This gives the relative bandwidths shown in Table 2.

The relative bandwidths given in Table 2 are plotted in Fig. 4 (together with dotted curves showing the effect

Table 2

Number of channels	F.d.m. r.m.s. deviation (kHz)	Relative bandwidth p.c.m./f.d.m. Raised cosine	½ raised cosine
24	50	0·84	0·62
60	100	0·81	0·61
120	100	1·1	0·83
300	200	1·0	0·75
600	200	1·33	1·0
960	200	1·54	1·15
1800	140	2·06	1·55
2700	140	2·24	1·68

of keeping the f.d.m. r.m.s. frequency deviation at a constant value), for the full raised cosine response. It is then more obvious that the increase in the relative bandwidth required for p.c.m. for large numbers of channels is due to the decreasing modulation index of the f.d.m. system.

The comparison in Table 2 is between f.d.m. microwave systems using normal frequency deviation and p.c.m. microwave systems using four-phase modulation and cross-polarization.

10.3 *Path Loss Formulae*

(a) Free space path loss

$$= 20 \log_{10}\left(\frac{4\pi L}{\lambda}\right) \text{dB}$$

$$= 32{\cdot}5 + 20 \log_{10} L + 20 \log_{10} F \text{ dB}$$

where L is the path length in km
F is the frequency in MHz.

(b) The gain of two parabolic aerials assuming 0·57 illumination efficiency

$$= -84 + 40 \log_{10} D + 40 \log_{10} F \text{ dB}$$

where D is the aerial diameter in metres.

10.4 *P.C.M. Filter Parameters*

The optimum 3 dB bandwidth of a Gaussian filter is a compromise between intersymbol interference which occurs if the bandwidth is too small and thermal noise which is directly proportional to the bandwidth. If both of these effects are considered as a deterioration of system performance in decibels then the optimum 3 dB bandwidth for the filter occurs at about 0·8 times the clock rate.

The raised cosine filter is a narrower bandwidth filter in which the pulse zero crossings are arranged to be coincident with the centres of adjacent pulses.

For optimum noise performance the Gaussian filter characteristic is distributed between the transmitter and receiver such that the square root of the filter characteristic is located in the receiver and the square root in the transmitter. This gives rise to the following parameters:

GAUSSIAN FILTER

1. Optimum overall 3 dB bandwidth = 0·8 times the clock rate.
2. 3 dB bandwidth of receiver and transmitter filters = 1·13 times the clock rate.
3. Receiver equivalent thermal noise bandwidth = 1·2 times the clock rate.

RAISED COSINE FILTER

1. Optimum overall 3 dB bandwidth = 0·5 times the clock rate.
2. 3 dB bandwidth of receiver and transmitter filters = 1·0 times the clock rate.
3. Receiver equivalent thermal noise bandwidth = 1·0 times the clock rate.

These bandwidths become larger when the filters are modified by $x/\sin x$ for rectangular pulses.

Manuscript first received by the Institution on 29th September 1970 and in final form on 12th February 1971. (Paper No. 1381/Com. 42.)

U.D.C. 621.317.72 : 621.391.822 : 621.397

A New Equipment for the Measurement of Video Noise

By

F. H. WISE, B.Sc., Dip.El.†

and

D. R. BRIAN (Graduate)†

Reprinted from the Proceedings of the Conference on Television Measuring Techniques held in London from 11th to 14th May 1970.

Describes a video noise measuring equipment which has been developed to meet the needs of a television broadcasting organization. The various measuring techniques available for this purpose are reviewed and the chosen method, which employs a sampling technique, is fully described. Details are then given of the practical design together with some remarks on its use.

1. Introduction

Measurement of noise represents an important part of the testing and maintenance of television transmission equipment but a practical difficulty arises in that many items of equipment will only operate correctly when line sync pulses are present. These signals are needed, for example, in receivers for the operation of automatic gain control and in video circuits for d.c. stabilization. In consequence, it is necessary that any instrument designed for noise measurement be unaffected by the presence of sync pulses.

A noise measuring equipment is described which satisfies the above requirement and which is suitable for use in studios, transmitting stations and in connexion with vision links. The operation depends upon a sampling process which has not been previously used in this application and which is shown to give satisfactory results.

2. Methods of Measuring Noise in the Presence of Sync Pulses

The equipment to be described is needed for the measurement of noise in 75 ohm circuits at a standard level of 1 V peak–peak (p–p), that is, with 0·3 V sync pulses. It is customary to express noise level in dB with respect to the picture amplitude (0·7 V) and this is normally measured as a p–p quantity for low frequency (power supply) noise and as an r.m.s. quantity for wide bandwidth (random) noise. The minimum sensitivity required is about −65 dB for r.m.s. and −55 dB for p–p measurements, so that it is clear that some specific provision must be made to discriminate against sync signals present which are some 40/50 dB greater than the above levels.

A number of methods have been used to achieve this result:

(a) The most basic method is by direct assessment of the p–p value from a c.r.o. display. If required, the r.m.s. value may be estimated from the p–p value by use of a correction factor.

(b) By application of a pulse train equal and opposite to the sync pulses so that cancellation occurs.

(c) By time gating the signal such that the sync pulse information is excluded leaving about 50–80% of the active line time for measurement.

(d) By frequency separation, making use of the property that the spectral components of a line sync pulse train are located at integral multiples of line frequency. If measurements are made in the narrow frequency bands between them no sync information is present and only the circuit noise will be detected.

(e) An extension of the sampling principle in which a single sample is taken of the voltage appearing at one point along each line.

Although in principle extremely simple, method (a) has a number of disadvantages. First, it is not possible with some types of oscilloscope to employ sufficient gain to observe the noise present without at the same time affecting the display because of overload due to sync pulses. Second, for the measurement of wideband noise, the correction factor used to convert from p–p to r.m.s. is arbitrary and depends upon the spectral density of the noise and the definition of p–p. In practice, a factor of about 17 dB is normally used, which, for 'white' noise assumes that p–p is defined as that value exceeded for not more than about 0·05% of time. Experience has shown that different observers can 'measure' the same r.m.s. noise level when using this method and obtain results differing by as much as 6 dB. For most purposes this degree of uncertainty is too great and a more accurate technique is needed. Methods (b), (c) and (d) have been described in the literature.[1,2,3]

For many purposes, method (c) is satisfactory and several instruments have been put into production which depend upon this principle. The latest application of the method would be to measure noise in the so-called 'quiet line' of the field blanking interval by gating out all information except that occurring in one line.

A practical difficulty occurs with this type of instrument because unless low frequencies are excluded, any line tilt on the measured waveform shows up as noise. In the case of a v.h.f. transmitter operating at white level, a 2% sag across the line due to power supply regulation is not unusual so that during the gated interval, about 1% of tilt will be observed. This sets the minimum measurable level of p–p noise to around −35 dB which is some 20 dB short of requirements.

The frequency separation technique described under (d) has been used with success but it gives a result which is applicable to one part of the spectrum only. For certain purposes this can be an advantage but is not convenient for operational use.

† Experimental and Development Department, Independent Television Authority, 70 Brompton Road, London S.W.3.

Primarily for the above reasons a new method of measurement (e) has been proposed which avoids this difficulty. This is an extension of the well-known sampling principle. A sample is taken in each line but whereas in method (c) a large amount of noise data is gathered in each gated-on interval, in this method, only a single 'instantaneous' value or sample of the noise is obtained. The sampling technique is conventionally applied only to the case where the highest frequency to be processed does not exceed one-half of the sampling frequency. In this case we use the sampling technique to process noise in a bandwidth of several hundred times the sampling frequency. The method is believed to be new and is described in detail in Section 3.

3. The Method Selected

The sampling method is described and its inherent characteristics outlined. By taking a single sample in each line, noise is effectively measured along a narrow vertical stripe down the television picture.

3.1 *Explanation of Operation*

It is well known that any waveform which contains components up to a frequency f may be described exactly by a train of samples taken at a rate of greater than $2f$ and that signals of frequency identical to that of the sampling frequency will not be discerned. For measurements on the television waveform where sync pulse information has to be excluded, the sampling rate must, therefore, be at an integral multiple of the line repetition frequency and if the sampling rate is set at once per line, then the effects of line tilt or any other variations along the line are eliminated. This does mean that only frequencies up to 7·8 kHz† may be retrieved undistorted from the output waveform. If we consider the application of higher frequency signals to the sampling circuits, we see that an output is obtained having the same amplitude but changed in frequency. This is illustrated in Figs 1(a) and 1(b).

Consideration of the problem shows that the output frequency f_o is related to the input frequency f_i and the sampling frequency f_s by the formula $f_o = |f_i - N.f_s|$

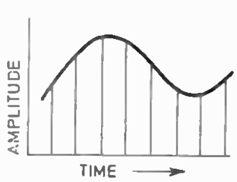


Fig. 1(a). Sample signal of frequency less than ½ sampling rate.

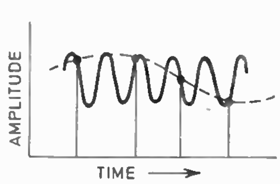


Fig. 1(b). Sample signal of frequency greater than ½ sampling rate.

† For 625-line case.

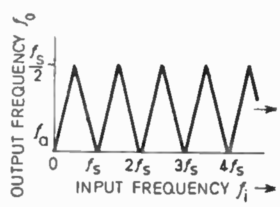


Fig. 2. Relationship between input and output frequency.

where N is integer and $f_o < f_s/2$. This is shown graphically in Fig. 2.

Thus the sampling process compresses high frequency signals into a low frequency ($f_s/2$) bandwidth. For the case where the output from the sampling gate is applied directly to a low-pass filter as in Fig. 4, the limit of this process occurs when the sample duration becomes comparable with the period of the input waveform. The transfer function of the process is readily evaluated with reference to Fig. 3 which shows the waveform B cos ωt

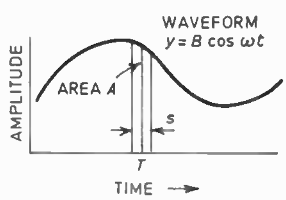


Fig. 3. Effect of sample pulse duration.

sampled at time T over a period S. The mean value M of the waveform during this period S is given by:

$$M = A/S = \frac{B}{S}\int_{T-S/2}^{T+S/2} \cos \omega t \, dT = B \cos \omega t \, \frac{\sin \omega S/2}{\omega S/2}$$

Thus the original amplitude is preserved provided

$$\frac{\sin \omega S/2}{\omega S/2}$$

is close to unity, i.e. ωS is small. In fact the error becomes about 10% when the sample period is one-quarter of the period of the input frequency. In order that the response is substantially constant up to 10 MHz, the sample pulse duration should not be greater than 25 ns. The problem has been considered in the time domain but it is also of interest to consider the problem in the frequency domain where the sampling system may be regarded as a 'multi-heterodyne' receiver. The receiver analogy is in fact very close as illustrated by the circuit in Fig. 4.

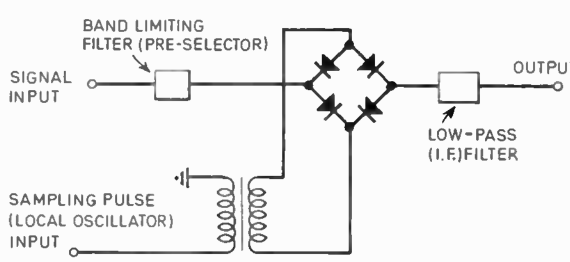


Fig. 4. Circuit of sampling gate.

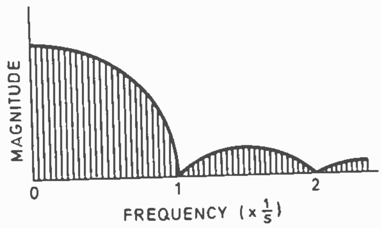


Fig. 5. Spectrum of sampling pulse train.

The sampling pulse input is in effect a series of local oscillators reproducing the pulse spectrum. If the sampling waveform consists of a train of pulses of duration S and repetition frequency f_s, then Fourier analysis yields the result that the spectrum is as shown in Fig. 5. The local oscillator signal, therefore, consists of a series of sine waves spaced in frequency by the sampling rate f_s and of substantially constant amplitude up to the frequency given by $\frac{1}{4}S$. If now the filter following the sampling gate is of the low-pass type, having a cut-off frequency of $f_s/2$, then the signal appearing in the output circuits will be due to the beat between the input signal and only one of the local oscillator components. As the input frequency is changed, so the output frequency changes as shown in Fig. 2. Each discontinuity in the characteristic at the output frequency $f_s/2$ occurs at the point where a new local oscillator takes over.

The frequency response of the 'multiheterodyne' receiver becomes as shown in Fig. 6.

The arrangement adopted in the practical equipment differs from that described above in that output from the sampling gate is applied to a hold circuit before passing through the low-pass filter. This causes the input to the filter to be of histogram form rather than a series of narrow pulses and results in a large increase in system gain in the ratio of 20 ns to 64 μs, or about 70 dB.

When this sample and hold technique is used, the sample duration becomes unimportant provided that it remains small compared with the interval between samples and that the system frequency response, from input to the hold capacitor with the gate conducting, is adequate at the highest frequency of interest. It is necessary, however, that the sample gate turn-off time is very short and the criteria for pulse width given earlier, now apply to gate switch-off time. With this modification, the mathematical relationships remain substantially unchanged and the system still measures along a narrow vertical stripe, defined in this case by the turn-off of the sampling gate.

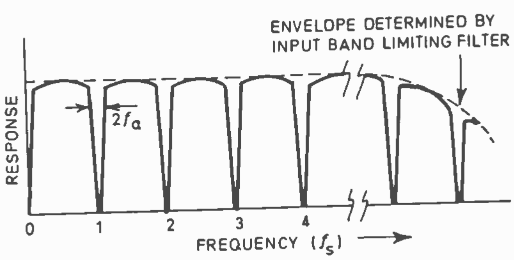


Fig. 6. Frequency response of 'multiheterodyne' receiver.

The use of a narrow sampling pulse is of no disadvantage and in fact a 20 ns pulse is used in the practical equipment described in later sections.

3.2 *Sensitivity*

At first sight it may be thought that by taking measurements during extremely narrow samples of the television waveform, the inherent sensitivity of the equipment will be low. In fact this is not so and is illustrated by the following simplified analysis.

Referring to Fig. 7:

S is the effective sampling pulse width, say 20 ns. (The turn-off time for the sample and hold system.)

B is the effective noise bandwidth defined by effective sampling pulse width S and is approx $1/2S$, i.e. 25 MHz.

E_s is the noise voltage to be measured in bandwidth B.

E_g is the noise voltage due to the sampling gate in bandwidth B.

b is the bandwidth of the audio low-pass filter.

E_a is the equivalent input noise voltage of the audio amplifier.

R_s is the source resistance.

R_g is the 'on' resistance of the gate.

Operating conditions are chosen such that the frequency response at capacitor C, measured whilst the gate is conducting, is substantially constant up to an input frequency of 10 MHz and the input impedance of the following amplifier is chosen such that the potential across C remains substantially constant during the interval between sampling pulses. Under these conditions, the transfer function of the sampling gate is unity, i.e. an input of 1 V in the bandwidth B appears at the sampling gate output as a signal of 1 V but in the bandwidth b.

We see that the inherent noise voltage is determined by the root of the sum of squares of E_g and E_a. The value of E_a is about 10 μV r.m.s. or 0·7 V r.m.s. −97 dB for a m.o.s. f.e.t. input stage when driven from a capacitive source of 500 pF, while the value of E_g is theoretically about 1 μV r.m.s., assuming an ideal diode gate 'on' resistance of 10 ohms and a noise bandwidth of 25 MHz. Practical semiconductor devices exhibit excess (flicker or $1/f$) noise at low frequencies (within the bandwidth b) which makes the contribution due to E_g some 30/40 dB larger. There is also some contribution to the gate noise E_g from the sample pulse source (which is equivalent to local oscillator noise occurring in a superheterodyne receiver). This is minimized by the use of a sampling gate which is balanced for the sample pulse input. Practical measurements show that E_g is about 30 μV or 0·7 V r.m.s. −87 dB.

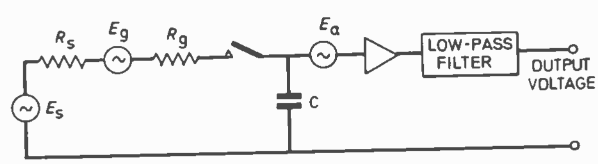


Fig. 7. Equivalent circuit of sampling gate.

It is seen that E_g is the dominant source of noise limiting the equipment sensitivity. It could be somewhat reduced by using a greater value for S and hence reducing the effective bandwidth B, but this would result in a less good high frequency performance and a less than *pro-rata* improvement in residual noise because the excess noise tends to be concentrated towards the low frequency end of the spectrum. The response is substantially uniform across the whole video band but there are 'holes' exhibited at each multiple of the line frequency. The presence of these 'holes' ensures that no response is exhibited to any signal which is repetitive at line rate. If the amplifier following the sampling gate has a response which cuts off at some low frequency f_a, then the 'holes' referred to above, each have a width of $2f_a$.

The system provides an output in the audio band which has statistical properties very similar to those of the original signal, with the same r.m.s. and p–p value but without any sync pulse information. Because the input bandwidth is not continuous but is of a form shown in Fig. 6, some difference exists between the c.w. sensitivity and the distributed (random noise) sensitivity. This difference amounts to rather less than 1 dB except in the case of c.w. signals close to an integral multiple of the line frequency.

After conversion into a sync pulse-free signal in the audio band, the noise may be amplified and measured using any conventional technique. The concept of r.m.s. and p–p is well understood in the case of periodic signals. Where random noise is considered some additional remarks are in order.

If the noise is truly random (white) noise, then the probability P of the peak value having an amplitude of at least E is given by:

$$P = 1 - \mathrm{erf}\,(E/V\sqrt{2})$$

where V is the r.m.s value and

$$\mathrm{erf}\,(x) = \frac{1}{\sqrt{\pi}}\int_0^x e^{-x^2}\,dx$$

This is rather more meaningful if shown graphically as in Fig. 8.

Oscilloscope brightness is normally set such that p–p measurements correspond to those values exceeded for between 0·1% and 0·001% of time. This results in a factor p–p/r.m.s. of between 16 dB and 19 dB.

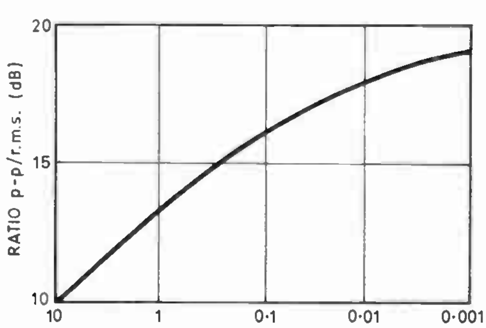


Fig. 8. Probability of a given ratio p–p/r.m.s. being exceeded for random noise.

The design of the p–p detector in this equipment is such that it responds to values exceeded for about 2% of the time and the resulting factor p–p/r.m.s. for random noise is about 13 dB.

Because the r.m.s. measurements are made using a thermal device, the readings are not waveform dependent.

3.3 *Special Requirements*

A system has been outlined which will measure noise using a sampling method. So far it has only been considered in relation to the measurement of noise in the presence of line sync pulses. It is also required that measurements may be made in the absence of sync pulses (see 5.1) or in the presence of both line and field sync pulses (see 4.2.2).

Provision is made for separating the noise into various bands by use of input filters. For the 625-line standard, wideband measurements may be made over the frequency range 40 Hz to 5·5 MHz. For measurements of power supply components a low frequency band 40 Hz to 7·5 kHz may be selected whilst for high frequency measurements, the l.f. components may be rejected by selection of input filters to give a passband 7·5 kHz to 5·5 MHz. Thus parts of the spectrum may be separated and different sources of noise identified. A different division of the spectrum is useful if it is desired to separate power supply components which may be modulated onto, rather than added to, the television waveform and this requires that components close to all line frequency harmonics are separated. This is achieved by adding a high-pass filter after the sampling gate having a cut-off frequency of about 1 kHz. Figures 2 and 6 show how this achieves the desired result.

It has been shown that the equipment measures noise appearing in a narrow stripe down the television picture. This provides an adequate statistical sample of the noise provided that the noise is not correlated line to line. Normally this condition is satisfied except in the case of line store standards converters where so-called curtain noise appears as a series of vertical stripes. In this case the equipment will not detect the curtain noise but will, of course, measure any other non-correlated noise. To overcome this problem the time delay between sync pulse and sample is made to vary so that measurements are made along diagonal stripes thus cutting across each of the curtain elements in turn.

The simplified system block diagram needed to meet these requirements is shown in Fig. 9.

4. Practical Circuits

4.1 *Filters*

The filters and weighting networks used for noise bandwidth selection and weighting have conventional passive components and the designs were based on existing information.[4] However, due to availability of space and the necessity of shielding, some mechanical development was found to be desirable in order to reduce the physical size and minimize pick-up. The former problem was overcome by using a recently available high frequency ferrite for all except the two low-frequency splitter filters, thus reducing the coil size to approximately

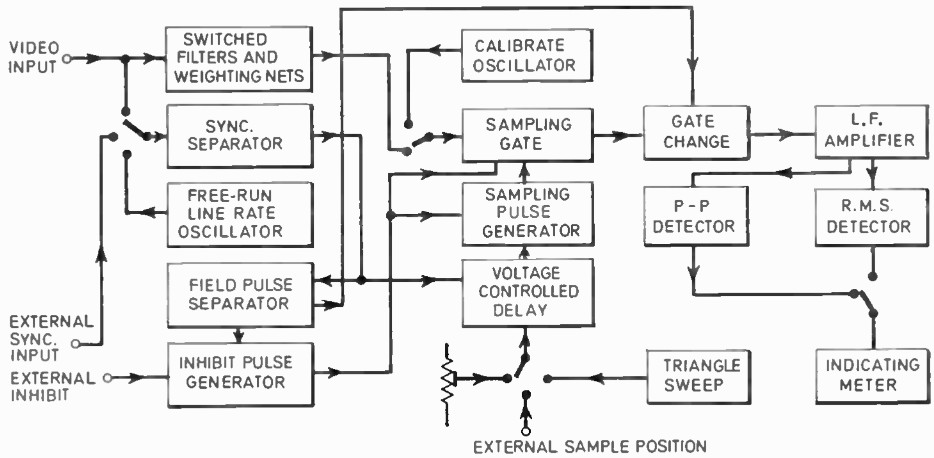


Fig. 9. Block diagram of video noise measuring equipment.

1 cm³. The latter problem arises due to the necessity of operating the equipment in strong r.f. fields, e.g. transmitting stations. Enclosure in a light gauge tin-plate box formed the r.f. shielding for each filter and in the case of the 5 kHz and 7·5 kHz splitter filters, mu-metal was used to overcome the additional mains pick-up problem (see 4.4.1).

The filters, together with a stable sine wave calibration source, are enclosed in a 4 in wide ISEP† module which includes the bandwidth and weighting network selection switches and the equipment function switch.

4.2 *Timing and Gating Waveforms*

4.2.1 The sampling pulse

The 20 ns sampling pulse is generated in the sampling gate module and the 0·5 μs trigger pulses are generated in the timing unit and timed from the line synchronizing pulses. A line sync separator is used to provide a clean sync waveform, being driven either from the incoming signal or a free running generator of either 10 or 15 kHz nominal frequency, depending upon standard selection. This waveform is used to drive a sawtooth generator, the output of which is applied to one input of a fast differential amplifier. A d.c. level applied to the other input controls the trigger pulse delay with respect to the sync pulse by an amount proportional to the d.c. level. Variation of this level, therefore, moves the sampling position across the line, i.e. varies the position of the vertical stripe along which the noise is measured. If a slow (4 Hz) triangular waveform is used in place of the d.c. level the sample is effectively swept diagonally back and forth across the line enabling 'curtain noise' from a standards converter to be measured.

4.2.2 The field inhibit pulse

If the incoming signal contains field as well as line sync information, the sampling process must be interrupted during the field and equalizing pulse period or large errors may be introduced. This is achieved by using a field sync separator followed by an inhibit pulse generator which gates off the sampling pulses during the unwanted period. This pulse is also used to operate the second gate in the sampling unit.

† International Standard Equipment Practice.

4.3 *The Sampling Gate*

The basic sampling gate employs a diode bridge, followed by a 20 dB amplifier. The 20 ns sampling pulses are generated from the trigger pulses by a differential amplifier switching an h.f. transistor, which drives the primary of a 3-winding toroidal transformer. The pulses produced in the other two windings are applied in balanced mode to the diode bridge and sample the incoming signal, the level being held on a storage capacitor. The signal is then amplified, using an f.e.t. differential amplifier and fed via the second gate to the l.f. amplifier.

4.3.1 The second gate

The sampling pulses are stopped during the field inhibit pulse period hence interrupting the sampling process. It was found however that due to the relatively long period for which the gate was off, the sampled level tended to drift causing large errors. The problem was overcome by using a further gate and clamping the level to an average value during the interval that this gate is switched off. The second gate consists of an f.e.t. series switch which is held on during normal operation, allowing the sampled signal to pass through unaffected. If field pulses are present on the input waveform, the f.e.t. gate is switched off during the field inhibit pulse period thus interrupting any output from the first gate. The level at the output of this second gate is then held at the value stored on an RC time constant, this being approximately the average value over the last field.

Originally, the same inhibit pulse was used to control both the sampling pulses and the second gate. Due to switching transients it was subsequently found to be desirable to restart the sampling process a few lines before re-opening the second gate and the trailing edge of the second gate inhibit pulse was delayed slightly to achieve this. Figure 10 shows a block diagram of the two gates and their switching circuits.

4.4 *The Low-frequency Amplifier*

By the sampling theory, the maximum output frequency of a sampling system is one-half the sampling rate. Hence the noise contained in the incoming signal is bandwidth compressed to 5 kHz or 7·5 kHz for 405 and 625 operation respectively. The effective r.m.s. voltage and p–p excursion however remains the same. It is

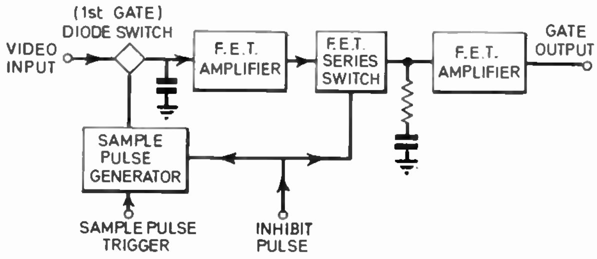


Fig. 10. Block diagram of sampling gates.

desirable therefore to filter the signal, providing a cut-off at half the sampling rate and a high attenuation at the sampling, or line frequency to ensure that components due to the sampling pulse are eliminated. A gain of 40 dB, switchable in 10 dB steps is also required.

4.4.1 Passive system

A conventional passive filter was initially employed, preceded by a switched attenuator and followed by a 40 dB amplifier comprising a single linear integrated circuit operational amplifier. This was found quite satisfactory, until operated in close proximity to a mains transformer when pick-up became considerable. Because the module was intended for use in an ISEP rack, together with a self-contained power supply, some form of magnetic shielding was essential. The mains power unit was therefore shielded in mu-metal and the filter replaced by an active system using selective feedback.

4.4.2 Active system

A five-pole Chebyshev filter with 0·1 dB ripple was used, the RC selective feedback components being switchable for dual-standard operation. The system has an overall gain of 40 dB and line rate attenuation of 30 dB.

4.5 *The Detectors*

For ease of operation a single meter was employed to display both r.m.s. and p–p measurements. It was therefore desirable to use an identical decibel scale for both detectors and the law had to be made the same for each. The simplest way to achieve this was to make the response of both detectors linear.

4.5.1 The r.m.s. detector

A thermal device is a good choice for r.m.s. measurements as true values may be measured without reliance upon some non-linear characteristic. The original arrangement proposed, employed two indirectly-heated thermistors run at constant power dissipation.[5] Unfortunately, the relationship between input voltage and

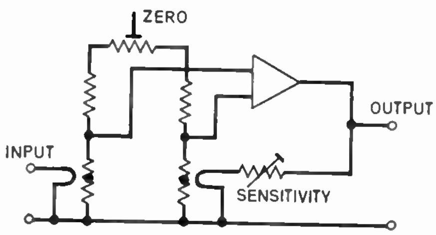


Fig. 11. Diagram of r.m.s. detector.

d.c. feedback although accurately defined, is not linear, so that the alternative arrangement shown in Fig. 11 was adopted. This operates in a similar manner but the noise voltage applied to one thermistor is balanced by a d.c. voltage fed back to the other.

The thermistors in the latter arrangement are not operated at constant power and in order to ensure adequate stability, the thermistors are enclosed in an environment which is temperature stabilized within about 5 deg C.

In the original design it was ensured that application of excessive input signals could not burn out the thermistors. Experience has shown however that this precaution was insufficient and that additional protection is necessary to limit the maximum input power to well within manufacturers' ratings. In the absence of this additional protection it was found that some drift in the thermistor characteristic occurred making subsequent zeroing of the instrument impossible.

4.5.2 The p–p detector

Operational amplifiers are again employed to reduce any distortion introduced by the peak detecting diodes. Two similar circuits are used, one operating on the positive peaks and the other on the negative peaks, the resulting balanced output being converted to unbalanced by another operational amplifier and fed to the same meter as for the r.m.s. detector via a relay. Figure 12 shows the basic circuit of this detector.

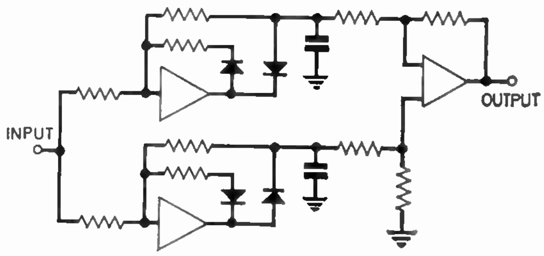


Fig. 12. Diagram of p–p detector.

5. Performance

5.1 *Normal Operation*

The inherent noise level of the instrument is such that accurate r.m.s. measurements can be made down to about −70 dB (with reference to 0·7 V) and p–p measurements down to about −60 dB. Measurements are normally made on a line repetitive waveform, i.e. one containing line synchronizing pulses and an identical line to line signal. This need not be black level or a white bar but may be, for example, a full test line. Care must be exercised in positioning the sampling pulse if the signal has sharp transients. If the sample pulse were positioned on a sharp rise, any timing jitter in the sample pulse generator would show up as noise. Close monitoring of the sample pulse position is therefore required when dealing with such signals. The advantage, however, is that the noise at any point along a line may be accurately measured and the noise on separate steps of a staircase may be determined. If the incoming signal has no line synchronizing pulses the sample pulses may be triggered from an external sync source or from the internal free-running pulse generator.

5.2 *Special Features*

The measurement of line-to-line correlated 'curtain' noise from a standards converter has been mentioned (see 4.2.1) but the line information must be constant, i.e. black level, and no low-frequency band limiting filters can be used as these would introduce tilt which would show up as noise in the swept mode.

The presence of a composite signal containing field as well as line sync pulses causes interruption of the sampling process (see 4.2.2) hence causing a drop in the r.m.s. sensitivity. This is compensated by the addition of approximately 1 dB of gain, controlled by a logic circuit which senses the presence of field syncs and the selection of 'r.m.s.' on the function switch. Peak–peak measurements are unaffected by the inhibit process and no gain change is required. The sampling process may be gated-off by an external signal enabling the noise contribution of individual heads of a video tape recording machine to be measured. The sensitivity reduction must be taken into account and the results weighted accordingly. A switchable active low-pass filter is incorporated after the sampling gate producing a 'comb' filter effect and removing information lying ± 1 kHz from harmonics of line frequency (see 3.3).

A calibration and zero check is provided together with external sync, gating and delay inputs and a remote output from the detector circuits.

6. Acknowledgments

The authors wish to thank the Director of Engineering of the I.T.A. for giving permission to publish this paper.

7. References

1. Weaver, L. E., 'The measurement of random noise in the presence of a television signal', BBC Monograph No. 24, March 1959.
2. Weaver, L. E., 'Some new video measurement techniques and apparatus', *EBU Review*, Part A Technical, No. 64, December 1960.
3. Edwardson, S. M., 'An instrument for measuring television signal/noise ratio', BBC Monograph No. 37, June 1961.
4. CCIR Recommendation 451.
5. Bozic, S. and Whitbread, J., 'Thermistor powermeter', *Electronic Engineering*, 39, No. 476, pp. 637–40, October 1967.

Manuscript first received by the Institution 16th March 1970 and in final form on 6th October 1970. (Paper No. 1382/Com. 43.)

The Authors

Mr. F. H. Wise obtained his B.Sc. degree in physics from the University of Southampton in 1951 and the Postgraduate Diploma in Electronics from the University in 1952. He then joined Vickers-Armstrongs (Aircraft) Ltd. and worked on u.h.f. receiving systems and flight simulators. In 1958 he took up an appointment with the Independent Television Authority and is currently the Head of the Radio Frequency Section where the range of duties have included propagation prediction, various theoretical studies and development of specialized broadcasting equipment. From 1964 to 1969 Mr. Wise was a member of the CCIR Study Group on Propagation through Non-ionized Regions of the Atmosphere and since 1970 he has been a member of the CCIR Joint Study Group on Application of Satellites in Broadcasting. He has been author or co-author of nine papers on the subjects of radio propagation, computer techniques in broadcasting and equipment design.

Mr. D. R. Brian (G. 1969) received his technical education at the North West Kent College of Technology and at Woolwich Polytechnic where he obtained the Higher National Certificate with Endorsements. His first appointment was with the British Scientific Instrument Research Association as a Research Assistant in the Industrial Measurement and Control Group where he worked on automation, electronics and computer programming. In 1967 he joined the Experimental and Development Department of the Independent Television Authority and since then has worked on the development of specialized broadcasting measurement equipment.

U.D.C. 534.22 : 620.179

A High-speed Ultrasonic Testing Machine for Tubes

By

G. H. KYTE,
B.Sc.(Eng.), C.Eng., M.I.E.E.,†

and

K. R. WHITTINGTON,
M.A., C.Eng., M.I.E.E.‡

Reprinted from the Proceedings of the Conference on Industrial Ultrasonics held at Loughborough from 23rd to 25th September 1969.

An ultrasonic flaw detection equipment has been developed for testing steel tubes at speeds compatible with modern rates of production. The system uses a series of identical probes arranged in a ring encircling the tube. Fast sequential pulsing of the probes, together with slow rotation of the tubes gives effective circumferential scanning speeds several times faster than is obtainable by purely mechanical methods. The probe rings are immersed in a small tank of water through which the tube is passed via specially designed water retaining valves.

1. Introduction

The increasingly stringent demands of present day engineering construction on the quality of tube for critical applications make it more and more necessary to reduce the incidence of defects which can occur from time to time in the process of manufacture.

Discontinuities in the form of cracks and 'laps' (overlapping material on the surface of the tube) are often initiated by inclusions and voids in the raw material. Since the manufacture of a tube usually involves considerable elongation of the material such defects tend to be longitudinal in nature.

All tubes contain discontinuities many of which are so small as to have no deleterious effects, and it is therefore a necessary feature of the inspection process that it should have an ability to discriminate between these and unacceptable defects. As a result of increasing working stresses smaller discontinuities are becoming more significant. The consequence of any failure to detect harmful discontinuities may be disastrous in such fields as steam generation plants, nuclear and aircraft applications. In nuclear fuel can element applications, testing specifications call for the detection of discontinuities of 0·05 mm (0·002 in) depth, and even in tubes for conventional steam plant the corresponding specification figure is of the order of 0·3 mm (0·012 in).

The requirements of an inspection equipment are that it shall be capable of detecting discontinuities down to the level specified by the customer and of operating at speeds consistent with the rate of production.

Modern non-destructive testing methods make extensive use of both eddy current and ultrasonic techniques. Flaw detection by eddy current testing depends on the change in current in a coil inductively coupled to the part under test, due to discontinuities in the latter. Small diameter tubing can be tested in typical eddy current machines at about 100 m/min (300 ft/min), which is considered a suitable speed to match present production rates.

There is, however, a limitation to the ability of eddy current techniques to detect small defects and, in particular, defects on the inside wall of the tube. Furthermore, eddy current techniques tend to be unduly sensitive to variations in wall thickness of the tube, which produce signals difficult to distinguish from those arising from harmful defects.

Ultrasonic flaw detection depends on the reflexion of high-frequency acoustic radiation from discontinuities in the material. Probes containing acousto-electric transducers of lead zirconate-titanate or similar material are used, with resonant frequencies up to 10 MHz, depending on the application. The transducers are energized momentarily with a pulse of electrical energy, producing a damped mechanical oscillation which decays, typically, in about 1 μs. This pulse of mechanical vibration travels outwards through the object tested and is reflected from any discontinuities which it meets. The return echoes are then detected by the same or a separate similar transducer acting as a receiver, are amplified and displayed on a cathode-ray tube or made to operate alarm circuits. The time delay between the transmitted pulse and the reflected pulse is used to distinguish true flaw echoes from energy reflected from the inside or outside walls of the tube. Water or other liquid is generally used as a coupling medium between the transducer and the tube.

In the more conventional ultrasonic machines, small diameter tubing (up to 3·7 cm), can be tested at about 6 m/min. This speed limitation of ultrasonic testers is mainly because the small diameter probes used must be scanned helically over the surface of the tube in order to achieve complete coverage. This is normally done by rotating the tube as it passes through the machine.

Higher scanning speeds can be achieved by rotating the probe assembly, 3000 rev/min or more being possible in machines of advanced design, when testing tubes of about 5 cm o.d. Attempts to exceed these speeds rapidly run into increasing mechanical difficulties.

The scanning rate of the machine to be described is not limited by mechanical considerations and in its present form is between two and three times as fast as the most advanced rotating probe system. This has been achieved by surrounding the tube being tested by a ring of probes, whereby the tube can be irradiated by acoustic energy around the whole circumference without the need for rapid rotation. Since the probes cannot all be energized simultaneously because of problems of interaction and interference, circuitry has been devised for switching them sequentially, thus moving the point of inspection around the tube at a speed considerably in excess of that achievable by mechanical means.

† Tube Investments, Steel Tube Division, Development Engineering Department, Walsall, Staffordshire.

‡ T.I. Research Laboratories, Hinxton Hall, near Saffron Walden, Essex.

In the design of the equipment priority has been given to reliability and ease of operation and maintenance. Modern constructional materials and solid-state switching techniques have been used. A prototype equipment was operated on a trial basis in a factory for 12 months and a second equipment has been in satisfactory operation in production since mid-1967.

2. Basic Ultrasonic Requirements

It was decided to design the prototype equipment to cover the range of diameter 1·25–5 cm ($\frac{1}{2}$ in–2 in) and to inspect for longitudinal defects only, since a preliminary survey had indicated that these were the most important considerations where quality testing of tubes would be required.

The fundamental system of testing used is the conventional one for this type of application, namely a pulse echo immersion system using a frequency of 5 MHz generating transverse waves in the material under test.

Preliminary investigation established that the effective beam width of conventional transducers was such that in order to inspect the whole circumference of the tube it was necessary to use a minimum of 72 transducers spaced at 5° intervals around the tube periphery (see Section 5). Although this is just feasible, the difficulties of mechanical disposition of transducers and of the number involved (all of which would have to be matched and individually set up) made this undesirable.

By accepting the necessity for slowly rotating the tube as it passed through the test head the same degree of inspection could be achieved while using a smaller number of probes.

In the present equipment a ring of 24 probes is used, i.e. a probe spacing of 15°, and these are pulsed sequentially at 8000 pulses/s. The tube is rotated at such a speed that during the period of 24 successive pulses the tube rotates through 5°; thus during the period of three complete switching cycles of the 24 probes the complete periphery of the tube is inspected.

2.1 *Alternative Probe Arrangements*

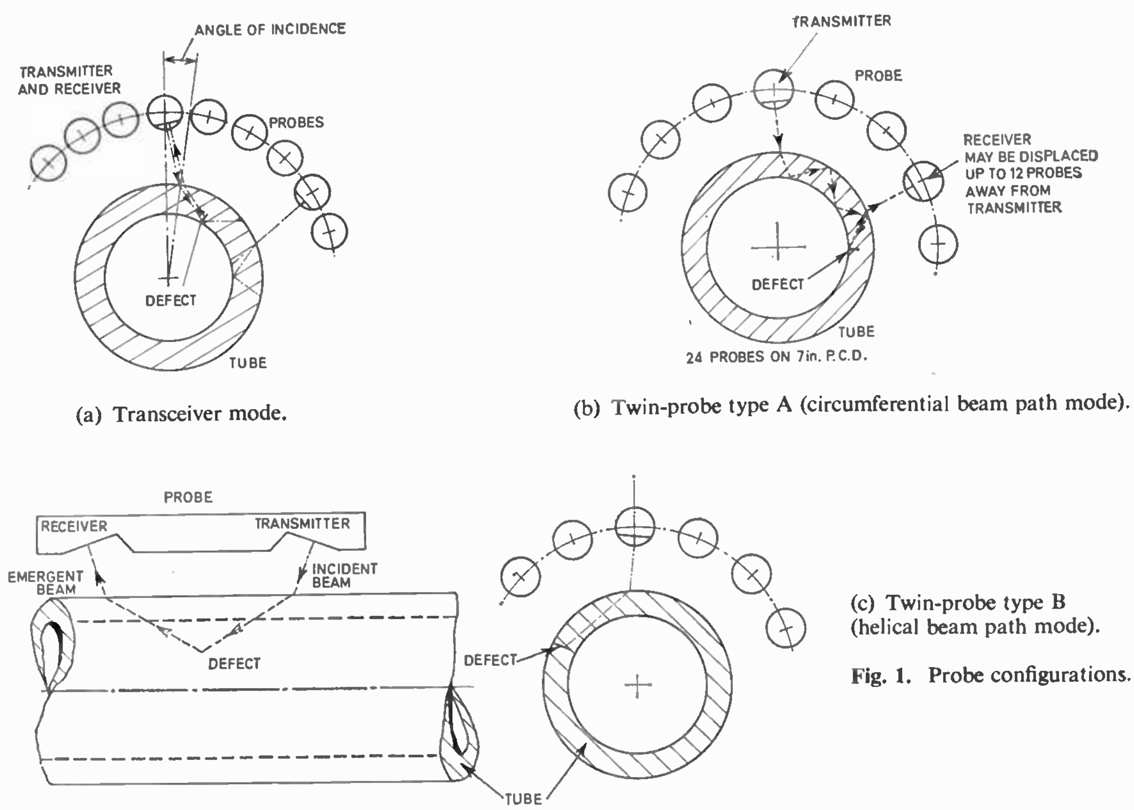
The test head is designed to allow the use of several alternative probe configurations. These are illustrated in Fig. 1.

2.1.1 Transceiver mode (Fig. 1 (a))

This is the most common system in general use; each probe in turn transmits a pulse of energy and receives any resulting echo. The probes are orientated to transmit and receive in a plane perpendicular to the longitudinal axis of the tube. The beam strikes the tube surface at a slight angle and produces a shear wave that travels around the tube wall in a circumferential plane. Any shear wave reflected by a defect becomes re-converted to a longitudinal wave at the tube–water interface and returns to the probe.

2.1.2 Twin type A (Circumferential beam path) (Fig. 1(b))

In this system the probes are disposed in the same manner as for the transceiver system, but reflected energy from a pulse transmitted by one probe is received by another probe further around the ring. The advantage



(a) Transceiver mode.

(b) Twin-probe type A (circumferential beam path mode).

(c) Twin-probe type B (helical beam path mode).

Fig. 1. Probe configurations.

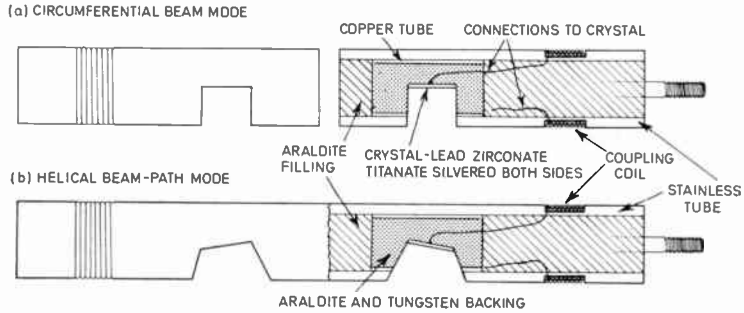


Fig. 3. Sections through transducer assemblies.

of this system is that possible interference from the large echo reflected by the tube/water interface is avoided; however, the longer path length in the material can result in excessive attenuation and loss of signal in certain circumstances.

2.1.3 Twin probe type B (Helical beam path) (Fig. 1(c))

This differs from the type A system in that the transmitting and receiving probes are displaced longitudinally and so orientated that the ultrasonic beam follows a helical path within the wall of the tube.

This mode enables the interface echo to be eliminated by interposing an acoustic screen between the probes while retaining the advantage of short path length in the tube. The disadvantage is that an additional set of transducers is necessary.

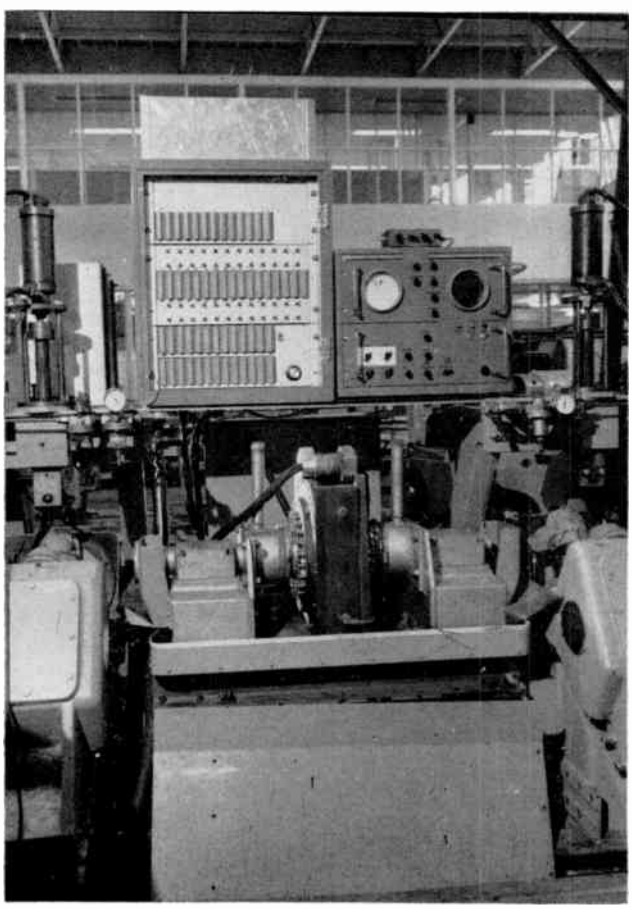


Fig. 2. Inspection head assembly.

3. Inspection Head

The probes are mounted in the probe head, which itself forms part of the complete inspection head assembly (Fig. 2).

3.1 *Probes*

The problem of mounting as many as 24 transducers in close proximity to the tube presented some difficulty. In the common type of immersion probe the transducer crystal is mounted at the end of a cylindrical body often 5 cm to 7 cm in length, with a trailing co-axial cable permanently sealed into the other end. The ultrasonic beam is projected axially. While this is convenient for many applications using only a small number of probes, the difficulties of mounting a large number of such probes, and of providing sufficient space to allow each to be adjusted over a range of angles, made it quite unsuitable for use in the high speed tester and a different construction was therefore adopted.

As will be seen from Fig. 3(a) the transducer crystal is mounted in the side of a cylindrical body so as to project the ultrasonic beam radially. With this construction the direction of beam travel relative to the tube surface can be easily varied by rotating the body about its axis. To further facilitate this, and also to simplify the problem of inserting and removing probes, the energy is transmitted to (and received from) each probe by means of a pair of inductively coupled windings, one on the probe body and the other fixed in the main body of the probe head assembly.

The basic probe assembly (Fig. 4) thus consists of a simple and robust rod with no trailing leads or slip rings.

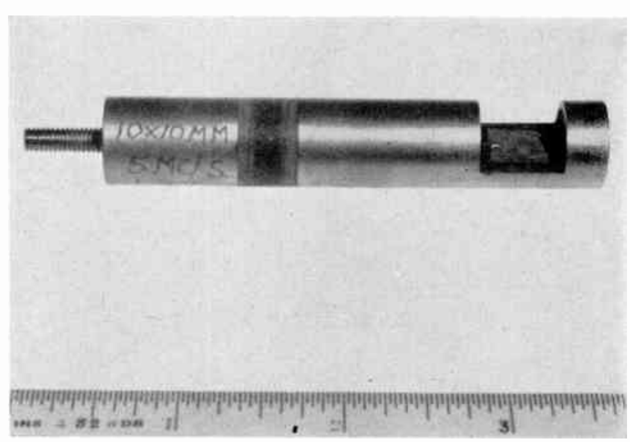


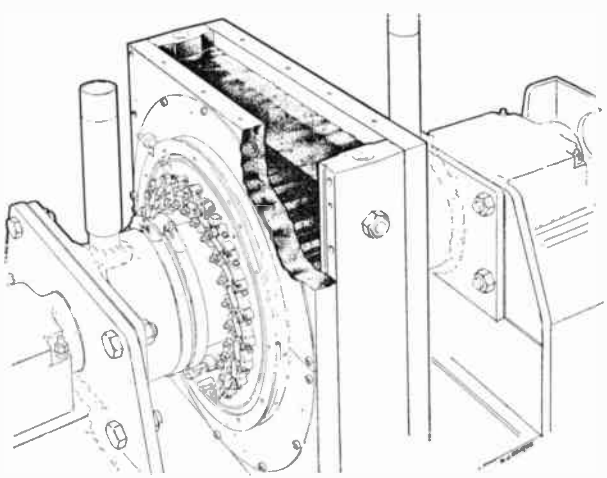
Fig. 4. Transceiver probe.

In the case of probes for the helical beam path mode, the two crystals are mounted in the same body and set by a jig during manufacture so that the beam paths converge at a predetermined angle (Fig. 3(b)).

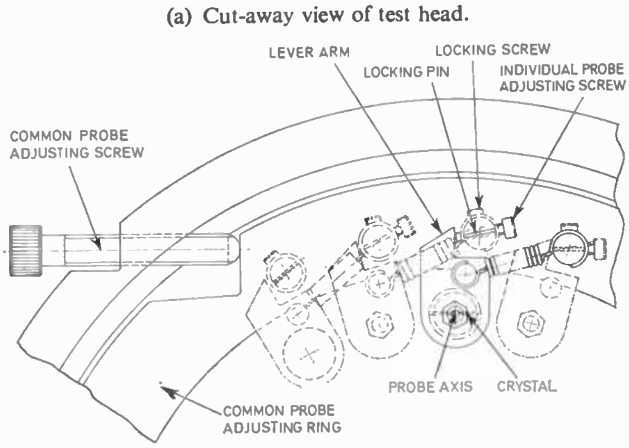
3.2 *Probe Head*

The construction of the probe head is shown in Fig. 5(a). It is essentially a narrow tank with two end plates each drilled to receive 24 probe bodies equally spaced on a 17·8 cm (7 in) diameter pitch circle. In the cases of probes for the transceiver and circumferential beam path modes, two rings of 24 probes can be fitted, one in each end plate, thereby permitting testing to take place simultaneously with ultrasonic beams travelling clockwise and counter-clockwise around the tube. Probes for the helical beam path mode, which are necessarily longer, bridge across from one end plate to the other.

The probes are mounted so that their angles of incidence can be adjusted (by rotating the probes on their axes) both individually and collectively, as required. The system is shown in Fig. 5(b). Small lever arms on the ends of the probes are spring-loaded on to locating pins fitted on the common adjusting ring. The latter can be rotated through a small angle by means of the common probe adjusting screw, thereby moving all probes simultaneously, while fine individual adjustments are made by means of adjusting screws abutting on the probe lever arms.



(a) Cut-away view of test head.



(b) Detail of probe adjustment.

Fig. 5. Construction of probe head.

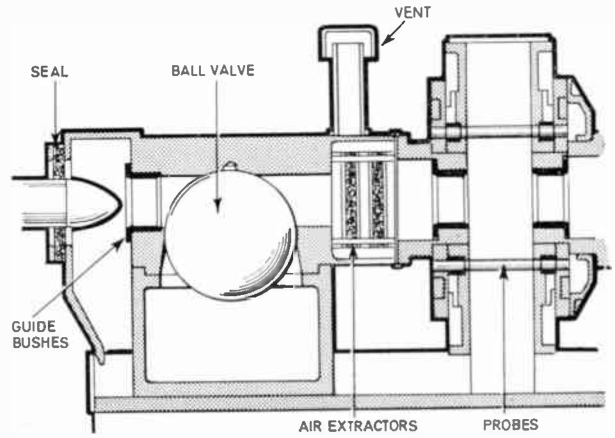


Fig. 6. Part section through inspection head assembly.

3.3 *Inspection Head Assembly*

Figure 6 shows in diagrammatic form a longitudinal section through this assembly. The head is maintained full of water up to the necessary level from a large external tank. At entry and exit the water is sealed off by a special seal consisting of a buoyant ball which is pushed down by the tube as it enters and floats back to its sealing position as the tube leaves. A small amount of water is lost with each passage of the tube, but this is easily replaced by gravity feed from the reservoir. Any air brought in by the passage of the tube is trapped in small chambers, between the seals and the testing zone, which are vented to the atmosphere. Soft rubber disks in these chambers serve to wipe the tube clear of air bubbles which would otherwise cling to the surface and be carried into the testing zone.

4. Electronics

A block diagram of the electronic circuitry, arranged for the helical beam path mode of testing, is shown in Fig. 7. The system may conveniently be considered as being divided into four main sub-systems namely master pulse generator, transmitter switching units, receiver switching units and the display and alarm units.

The master pulse generator controls the probe switching sequence and the time-base circuits.

The transmitter switching units comprise a transmitter power output stage associated with each transmitting probe, a system of gates to allow pulses from the master pulse generator to be fed to one output stage at a time, and a transmitter ring counter controlled by timing pulses from the master pulse generator, to define the sequence in which the transmitter gates are opened.

The receiver switching units comprise: a system of receiver gates to allow one receiving probe at a time to be connected to the main receiver amplifier; a set of switched gain units, each associated with one receiver probe; and a receiver ring counter, timed from the master pulse generator and controlling the receiver gates. The switched gain units are brought into circuit one at a time, each with its associated probe, under the control of the ring counter. Each gain unit is provided with a manual adjustment to enable the gain of that particular probe channel to be made equal to all the others.

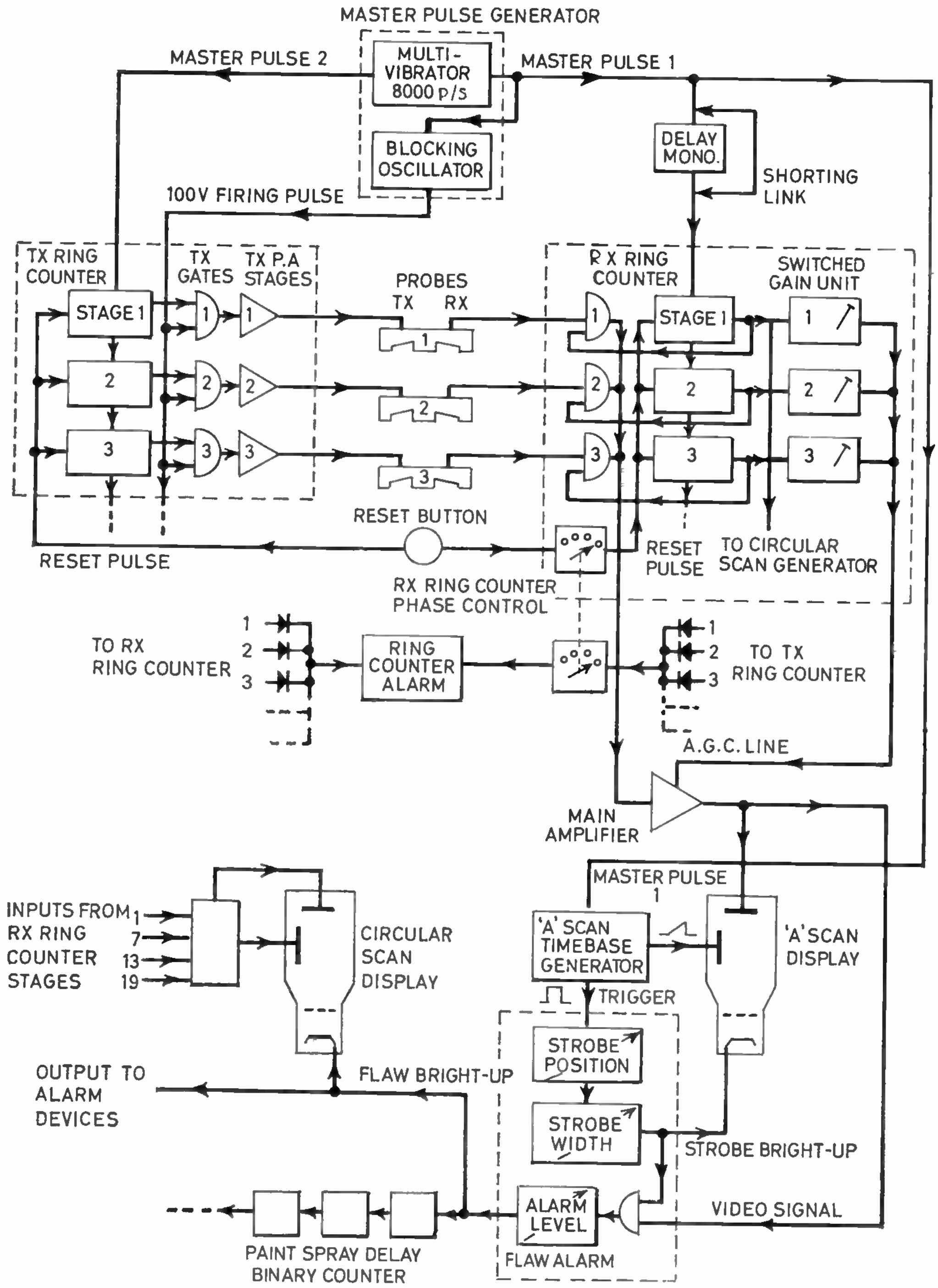


Fig. 7. Block diagram of electronic circuits (helical beam path mode).

Two c.r.t. displays are provided, one a normal A-scan display of signal amplitude against time, the other a circular presentation of probe position, on which the firing of the alarm circuits by an echo from a particular probe is represented as a bright-up pulse. The flaw alarm circuits themselves are arranged to give a visual or audible warning when return echoes exceed a present amplitude. In order to avoid triggering of the flaw warning units from signals other than true echoes, 'strobe' signals are provided, synchronized to the timebase, but variable in position and width. These strobes are used to gate the receiver output into the flaw alarm units.

A solenoid-operated paint spray is provided at the outlet from the probe head for marking the defective area of the tube. This is operated from the flaw alarm units, via a delay circuit to allow for the time taken for the tube to pass through the head.

The various units of the electronic system are described in more detail in the following sections. Figure 8 shows the relationship between the waveforms in various parts of the circuit.

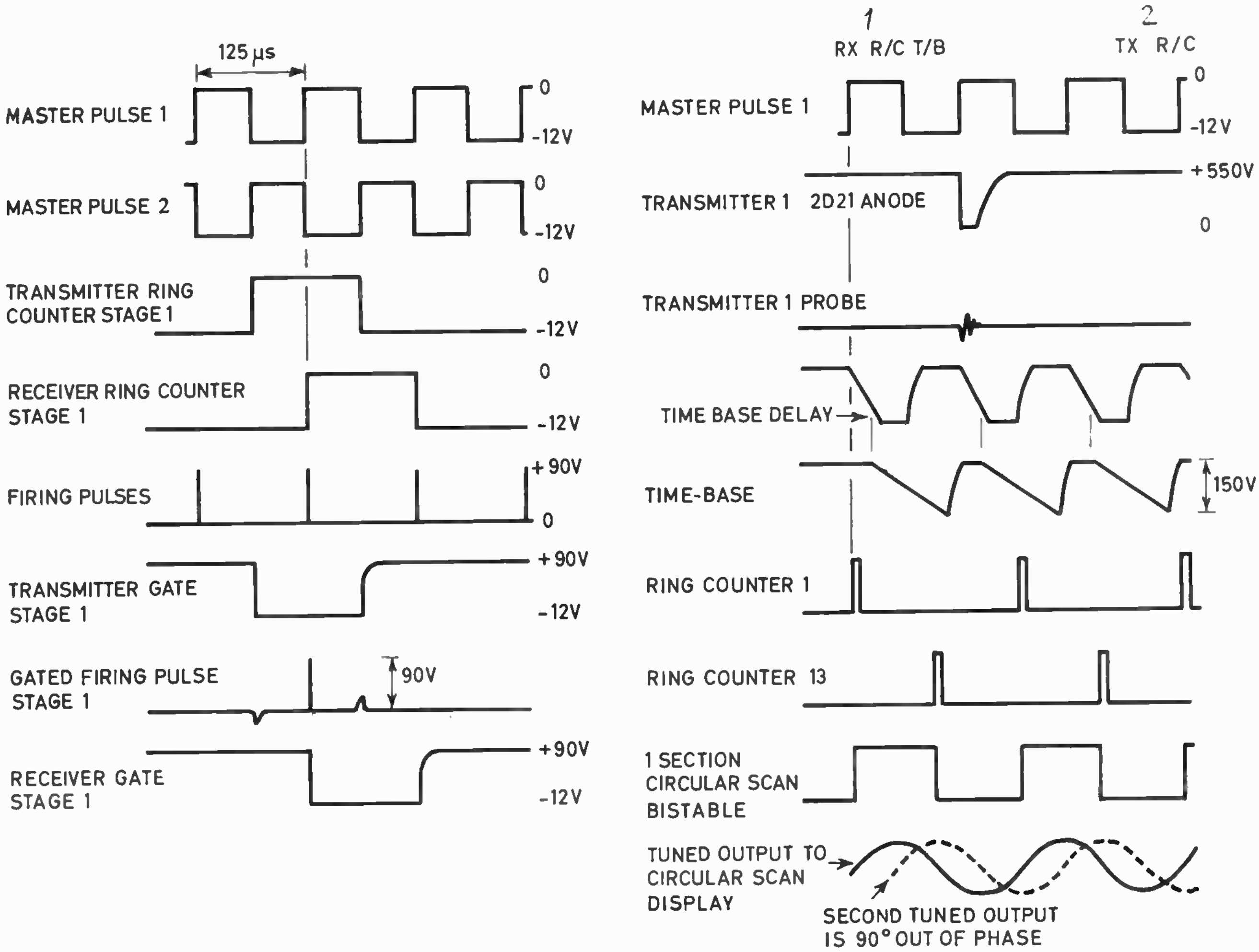


Fig. 8. High-speed ultrasonic flaw detector waveforms.

4.1 *Master Pulse Generator*

The master pulse generator controls the timing and sequencing of all electronic operations. It comprises a transistor multivibrator, free-running at 8000 pulses per second, from which two symmetrical square waves 180° out of phase with each other are available. These two outputs are designated master pulse 1 and master pulse 2. Master pulse 1, besides being fed to other circuits, triggers a transistor blocking oscillator which provides a rectangular 'firing' pulse of 1 µs duration and +100 V amplitude to fire the transmitters.

4.2 *Transmitter Switching Circuits*

Each of the 24 transmitter output stages consists of a small thyratron fed from a 550 V d.c. rail. As each thyratron in turn receives a gate firing from the master pulse generator, it causes a capacitor to be discharged into the appropriate transmitting probe.

The transmitter gating circuits which route the firing pulse to the appropriate power output stage consist of transistor/diode circuits supplied from a 90 V rail When a transistor is turned on by a pulse from the transmitter ring counter, the associated diode becomes forward biased and allows the firing pulse from the master pulse generator to reach the appropriate thyratron. In order to ensure that the firing pulse falls centrally within the transmitter gating waveform, the former is triggered from master pulse 1 and the latter from master pulse 2. (See Fig. 8.)

The transmitter ring counter is in effect a 24-position electronic switch which controls the transmitter gates. It consists of 24 transistor bistable stages appropriately connected, and obtains its timing pulses from master pulse 2, as described above.

4.3 *Receiving Switching Circuits*

The receiver ring counter is identical to the transmitter ring counter in circuitry and construction, but is triggered from master pulse 1. It thus runs synchronously with the transmitter ring counter but out of phase with it (Fig. 8). The phase difference is necessary to ensure that the receiver gates are open at the correct time to receive any flaw echoes, allowing for the time taken for the transmitted acoustic pulse to travel the requisite distance through the coupling medium. Also, since the transmitter firing pulse and the receiver switching pulse coincide (both being triggered from master pulse 1), any unwanted signals from these two sources occur at the same moment and are more easily dealt with. In some testing conditions the water path may be so long that the return echo does not occur within the period for which the receiver gate is open, despite the phase delay. For such conditions an additional delay monostable circuit is provided, which may be plugged into the timing pulse line to the receiver ring counter. This is replaced by a shorting link when not required.

In the circumferential beam path mode of operation, the receiving probes and transmitting probes are displaced circumferentially from each other by a selected number of probe positions (see Fig. 1(b)) This requirement dictates that when initially zeroing or resetting the ring counters, the receiver ring counter must be reset to a different condition from the transmitter ring counter, so that the circulating pulses in the two counters are displaced by an appropriate number of stages. This is done by means of a phasing switch (ring counter phase control, Fig. 7), a displacement of up to 12 positions being possible. An alarm is provided in the ring counter unit which will operate a warning light should the transmitter and receiver ring counters get out of step and require setting. Resetting, whether initially or in the event of a fault, is carried out by means of a push-button on the front panel.

The receiver ring counter controls conventional transistor gates which allow the outputs from the individual receiving probes to be fed sequentially into the main receiver amplifier.

In addition to controlling the receiver gates, the receiver ring counter also switches sequentially into circuit the series of switched gain units having preset controls (Fig. 7), which allow the gain of the main amplifier to be programmed to suit the characteristics of the probe pair (transmitter and receiver) operative at any one time. In this way the overall loop gain of the system is kept constant regardless of any variations in the individual probes.

In the circumferential beam path mode of testing, two independent arrays of 24 probes are used, one looking clockwise and the other counter-clockwise around the tube. In this case, two separate receiver amplifiers, two sets of receiver gates and two sets of transmitter output stages, are required. The extra units are provided in the equipment, but have been omitted from the block diagram in the interests of clarity. The additional receiver gates are switched by the existing receiver ring counter.

4.4 *Display and Alarm Units*

Amplification of the return echoes, after gating by the receiver gates as described above, is provided by a conventional 5 MHz tuned amplifier, using thermionic valves. A nominal gain of 110 dB is available, the actual gain being determined by the preset potentiometers of the receiver switched gain units, which are switched by the receiver ring counter, as explained previously, into what would normally be the automatic gain control line of the receiver.

The amplified return echoes are displayed on a conventional A-scan c.r.t., the time-base for which is triggered from master pulse 1.

A delay unit is incorporated to enable the flaw echoes to be positioned as required on the c.r.t. screen, and a bright-up pulse, controllable in position and width, is also derived from the time-base circuitry. This bright-up pulse is used to control a gate which assures that only those echoes returning during a selected period can operate the alarm circuits. It is referred to subsequently as a 'strobe' signal, by analogy with radar practice, although commonly termed a 'gate' signal in ultrasonic flaw detector usage. The term 'strobe' is used to avoid confusion with the transmitter and receiver gate signals previously discussed.

The alarm circuit is triggered by echoes of suitable amplitude which fall within the strobe waveform. A sensitivity control is provided, and the 'alarm' signal closes relay contacts which light a front-panel lamp and operate external circuits as required. There are two alarm units, A and B associated with the two receiver channels. They differ in certain respects, mainly in that unit A contains the circuit supplying the strobe pulse to the A-scan display, while unit B provides bright-up pulses to the circular c.r.t. display described below. Unit A also contains pulse lengthening circuits to allow signals from flaw echoes to be recorded on an ordinary pen recorder. For simplicity only one flaw alarm unit, combining the features of both, is shown in Fig. 7.

The second c.r.t. displays two concentric circular traces which are produced by time-base circuits using sine and cosine waveforms derived from the receiver ring counter by appropriate shaping techniques. This display is therefore synchronized to the 'rotation' of the electronic switch energizing the probes, and bright-up signals can be displayed on it which give a positional indication of the particular probe on which an echo large enough to trigger the alarm is being received. The two concentric displays relate to the two receiver channels. Channel A only is used in the helical beam path mode of testing, but both channels are used in the circumferential beam path mode.

One use of the alarm system is to operate a paint spray to mark the position of detected defects on the tube being tested. Since the tube cannot be marked until it has emerged from the test head, a delay circuit is needed between the alarm unit and the paint spray. This is provided by a 20-stage shift register each stage representing $7\frac{1}{2}$ cm of the tube length at the normal speed of testing. A flaw signal from alarm unit B injects a pulse into the first stage of the counter, and an internally generated timing signal passes the information down the shift register to the paint spray. The exact delay can be adjusted within limits by varying the timing pulse repetition frequency.

5. Mechanical Handling

The design of feed equipment capable of moving the tubes through the probe head with the desired helical motion presents a number of problems.

The tube must be accurately centred relative to the head, even though tubes in diameters ranging from 1·25 cm to 5 cm have to be handled; a water seal must be maintained even when the tubes cannot be fed nose-to-tail (see Section 3). and the peripheral speed of the tube and the helical angle should be maintained constant regardless of tube diameter changes. The complete mechanical handling system is capable of passing tubes, one at a time, with a helical motion through the test head, marking the position of a defect, and directing the

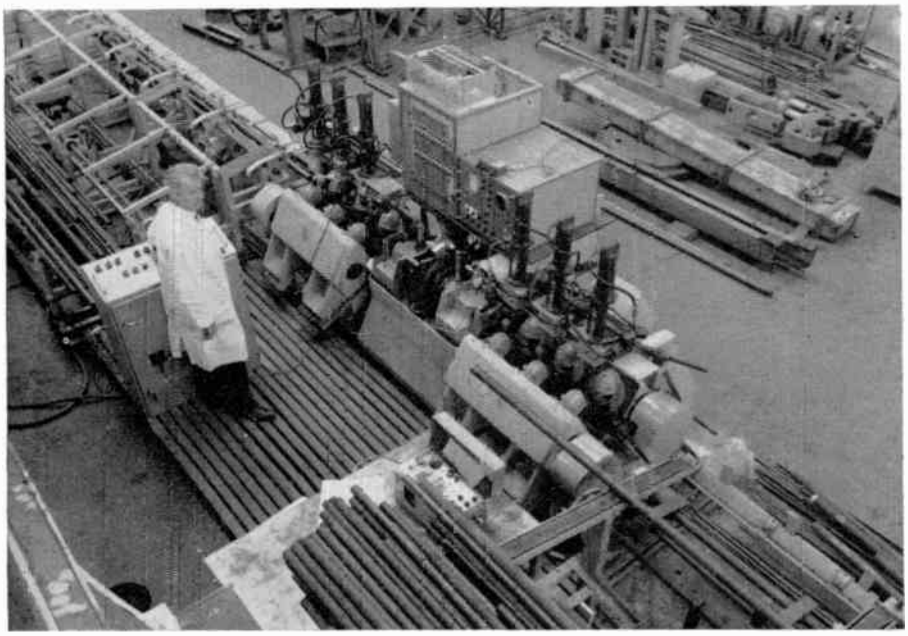


Fig. 9. General view of installation.

tube to an accept or reject rack; Fig. 9 is a general view of the installation.

A diagram of the helical feeder is shown in Fig. 10. It consists of two feed roller assemblies, one on the input and one on the output side of the test head, and a variable speed drive unit. The feed roller housings can move in slide-ways in order to centralize the tube under test in the test head. The tube is held firmly in contact with the feed rollers by air-operated top rollers. The helical feed angle can be adjusted so that any tube motion can be produced from pure rotation through the complete range of helix angles to straight feed without rotation.

The peripheral surface speed of the tube being tested can be varied by means of an electronic speed control that will also maintain the speed selected with close tolerance.

The handling system is designed to feed the tubes in succession automatically from the inlet side, shown on the right hand side of Fig. 9, through the head and to deliver them into either 'accept' or 'reject' racks depending on the result of the test. Should a fault occur, or the operator desire to inspect a tube more closely, the automatic operation can be over-ridden.

In order that the cause of defects can be ascertained, and where possible rectification work carried out, a means of marking the position of defects in the reject tubes has been incorporated. This consists of a pneumatic paint spray gun, situated on the outlet feed roll assembly and operated by a delayed signal from the ultrasonic tester. This is described in more detail in Section 4.

The usual electrical and pneumatic services are required for the machine, but it has been found necessary for some applications to include a water re-circulation system which can by the addition of inhibiting chemicals reduce the corrosion of tubes after testing.

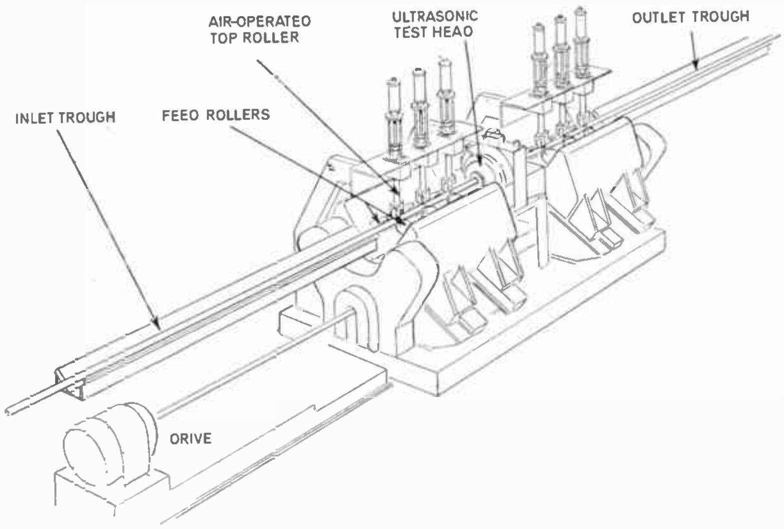


Fig. 10. Arrangement of helical feeder.

6. System Performance

6.1 *Relationship between Probe Coverage, Flaw Size Rotational Speed and Linear Speed*

In order to detect flaws regardless of their positions on the tube, it is necessary that:

(1) For every consecutive axial position of the tube, every part of the circumference must be examined.

(2) For every circumferential scan as implied in (1) above, a defined axial length of tube must have been inspected dependent on the relationship between beam width and depth of flaw to be detected. This axial tube length examined per circumferential scan defines the effective helical scanning pitch (which is, of course, not the same as the mechanical feed pitch).

As previously explained it is not possible to dispose enough probes around the circumference to enable a complete circumferential scan of the tube to be achieved with only one complete switching cycle of the ring of probes. It is therefore necessary to perform more than one switching cycle for each complete circumferential scan of the tube, and since each probe has its own particular arc to cover it will be appreciated that the number of cycles must be an integral one.

Knowing the circumferential arc coverage of each probe, and the total number of probes around the circumference, it is possible to calculate:

(a) The minimum rotational tube speed necessary to allow the whole circumference to be examined in the fewest possible switching cycles.

(b) The minimum number of complete switching cycles necessary to complete one circumferential scan.

Knowing (b) above, and knowing the length of tube examined per circumferential scan, it is possible to calculate the maximum axial velocity (ft/min of tube) permissible for complete examination of the tube.

6.2 *Circumferential Scan*

It is first necessary to determine the circumferential arc coverage of each probe. For this purpose tubes with spark-eroded artificial defects, of equal size on the bore and outside surfaces are used.

It should be explained that for each probe, as a flaw is rotated through the ultrasonic beam, the signal received from it rises to a maximum and then falls away. The probes are first adjusted mechanically so that for each probe this maximum signal is the same for both the internal and external defects. The 'loop gain' of all the probe channels is then equalized (i.e. the gains of the channels are individually adjusted so that the signal amplitudes are all equal).

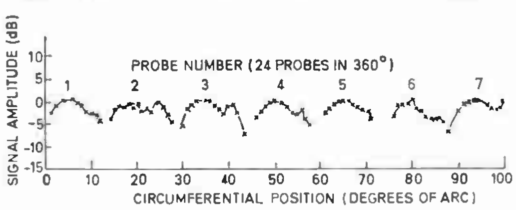


Fig. 11. Typical graphs of signal amplitude from probes plotted against angular position of tube.

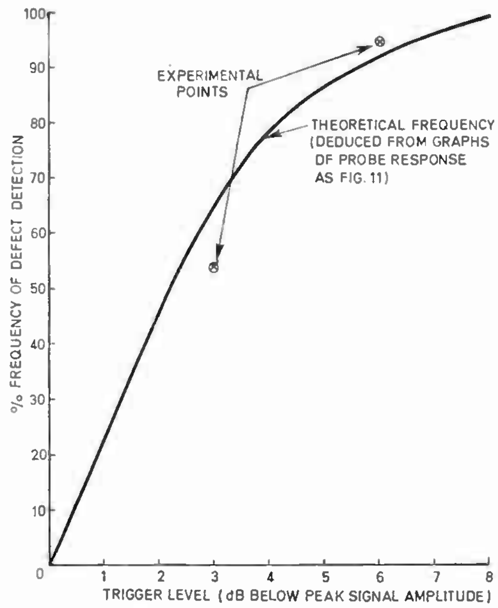


Fig. 12. Theoretical and experimental frequencies of defect detection.

Plots of signal amplitude against angular position of tube are obtained for each of the 24 probes, an example being shown in Fig. 11. From this diagram it will be clear that if the flaw warning trigger level is set so that a warning is received only when the peak amplitude is reached, the effective arc of coverage of each probe will be zero. An increasing arc of coverage will be obtained as the trigger level is reduced until at 6–10 dB below peak amplitude, warnings will be obtained from defects at virtually any angular position of the tube. However, as described in the introduction, it is usually required that level of discrimination between acceptable orientations in the tube and non-acceptable discontinuities be fixed at as low a value as is practicable so that with trigger setting 6–10 dB below the maximum it is found in practice that very small surface irregularities, at the optimum angular position, would also cause warning signals to appear. Thus, a compromise is necessary, and in routine operation of the high speed tester a trigger level 3 dB down on the maximum signal amplitude is used. This gives an arc of coverage, per probe, of the order of $7\frac{1}{2}°$–$10°$ or somewhat more than 50% of full 360° coverage of the tube.

On this basis, one would expect that if a tube with a single artificial defect were sent through the tester without rotation, a large number of times, with random defect orientation, the defect would be detected on at least 50% of the passes. Further, it is possible to plot theoretical frequency of detection against trigger level and to compare this with the achieved performance. This was done and the curves obtained are shown in Fig. 12 from which it may be seen that the predicted performance is experimentally confirmed.

6.3 *Rotating Tube*

Once the percentage arc coverage has been determined, it is possible to calculate the minimum rotational speed necessary to allow the whole circumference to be examined in the fewest possible switching cycles. For example, with 5° arc coverage per probe, and 24 probes spaced 15° apart, 33% of the tube surface, i.e. 5° in each sector, can be covered during each complete probe switching cycle; it is therefore necessary to carry out three such cycles to ensure complete coverage, the tube being rotated 5° in the period of each switching cycle Since each switching cycle takes 1/333 seconds (for 8000 pulses per second switching speed) the tube should complete one revolution in

$$\frac{365}{5} \times \frac{1}{333} \text{ seconds,}$$

equivalent to 277 rev/min. If the trigger level were such that the coverage was $12\frac{1}{2}$% in each sector of 15° it would be necessary to rotate the tube $2\frac{1}{2}$° after each switching cycle in order to cover the remaining $2\frac{1}{2}$° in each sector. This is equivalent to 138 rev/min. Figure 13 shows rotational speed plotted against percentage arc coverage, for a pulse repetition frequency of 8000 pulses per second.

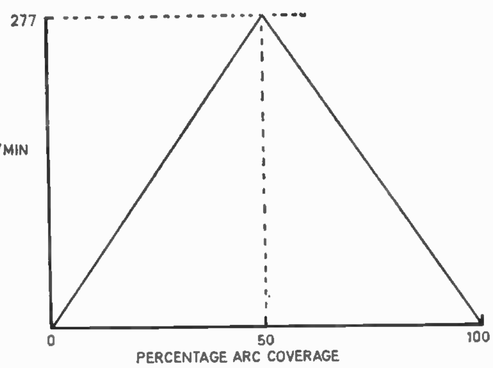


Fig. 13. Rotational speed versus % arc covage.

At any rotational speed the minimum of switching cycles to complete one circumferential scan must be an integral (otherwise part of the circumference will not be completely examined) and this integral value changes from infinity (at zero arc coverage) to unity (at 100% arc coverage) in a hyperbolic manner, as shown in Fig. 14. For example, at 50% arc coverage the number of scans necessary is two, at $33\frac{1}{3}$%, three, at 25%, four, and so on.

6.4 *Longitudinal coverage*

The longitudinal coverage of each probe will vary with trigger level setting in a similar way to the circumferential coverage; however, in the case of longitudinal defects, the magnitude of the effect will be smaller. Where the defect length is considerably greater than the beam width, the signal amplitude can be considered to be reasonably constant over the length of the flaw, and in practice even with shorter flaws there is always a finite

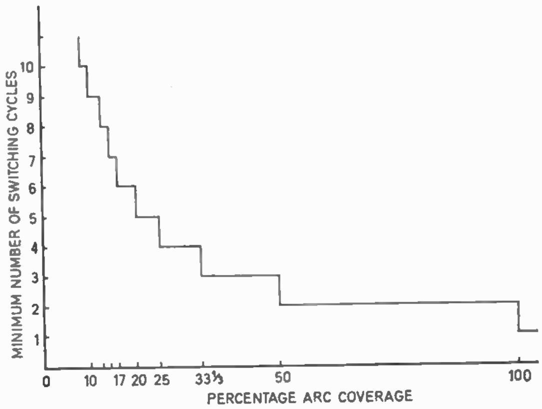


Fig. 14. Minimum number of switching cycles versus % arc coverage.

length in the axial direction within which the circumferential response does not change significantly, or in which the fall-off in response can be accommodated by allowing a small additional margin in the trigger level over and above what is necessary for the circumferential coverage.

Thus the length of tube which is inspected during each completed circumferential scan is a function of the relationship between the flaw length and the effective beam width, and is usually determined in practice by experiment.

Having established the tube length, Z, inspected per scan, the permissible axial speed is then given by $V = ZF/nP$ where n is the number of switching cycles per circumferential scan, as determined above, P is the number of probes, and F is the pulse repetition rate.

In the present case, for a $\frac{1}{2}$ in (1·25 cm) long flaw it is found in practice that $Z = 0{\cdot}24$ in (6·2 mm). Thus since $r = 8000$ pulses, $n = 3$, $P = 24$, the permissible axial speed is

$$\frac{0{\cdot}24 \times 8000}{3 \times 24} \times \frac{60}{12} = 133 \text{ ft/min (40 m/s)}$$

and this can be correspondingly increased for longer defects.

7. Acknowledgments

It will be appreciated that the development of an equipment of this type is the result of the efforts in various degrees of many engineers. The authors acknowledge in particular the contributions of Mr. D. Chattway and of Messrs. D. Terry and F. Bowen of Accles & Pollock, Mr. D. Wilson of T.I.R.L. and Messrs. R. Ashford, J. Beamand and T. Forks of D.E.D.

The authors wish to thank the Directors of Tube Investments Ltd., for permission to publish this paper.

Manuscript received by the Institution on 30th June 1969. (*Paper No. 1383/IC44.*)

U.D.C. 621.374.33: 621.387.322.3: 681.327.12

A Low-cost Seven-segmented Display System

By

R. K. SHARMA,

B.E.(Hons.) (Graduate) †

An economical seven-segment display system is described which employs 21 diodes and 7 driving transistors as against 49 diodes and 14 driving transistors in a conventional display. The single driving transistor employed for each segment has to be biased so that it is normally saturated and the lamp is illuminated. To extinguish the lamp, the transistor is held below cut off by providing a suitable excitation.

1. Introduction

The most popular system at present for digital read-out is by means of numerical indicator tubes. If for some reasons these tubes are not used, the next most popular system is the seven-segment system for displaying the digits. A seven-segment display system may be made of miniature neon or filament lamps. In a conventional seven-segment display system, seven lamps are arranged in the shape of the numeral eight (8) (Fig. 1). These lamps are driven by 14 transistors, two transistors for each lamp. A diode matrix of 49 diodes is used for illuminating a particular set of lamps which form a particular digit. In this system, all the lamps remain extinguished until excitation is applied.

In the system to be described all the lamps remain illuminated until excitation is applied. When excited, the lamps unwanted for a particular digit are extinguished. The circuit employs only 21 diodes and seven driving transistors against 49 diodes and 14 driving transistors in a conventional seven-segmented display system.

2. Operation

The arrangement of lamps is as shown in Fig. 1, and Table 1 gives the display code tabulating the lamps required to be extinguished for displaying a particular digit. The full circuit diagram of the system is shown in Fig. 2.

For automatic operation, the input terminals 1 to 9 are connected to a decoder circuit of diode AND gates employing positive logic. (In positive logic, state '1' is more positive than state '0'. Thus logic state '1' is a positive voltage and logic state '0' is a negative voltage.) When voltage V_B (positive voltage) is not applied, there is no input voltage available and hence, the biasing provided by resistors R1 and R2 is sufficient enough to saturate the transistors and thus all the seven lamps are illuminated.

Let the voltage V_B be applied. Depending upon the decoder circuit, only one input terminal is at logic stage '1', i.e. a positive voltage, and the rest are at logic state '0', which is the threshold voltage at the input terminals. In the case of an input terminal at logic state '1', V_B will send a current through Rb to the driving transistors connected to it via the diodes. This current is sufficient to neutralize the biasing of the transistors through resistors R1 and R2 by V_{BB}, and holds them below cut-off. Thus the lamps connected to these transistors are extinguished.

Only one input terminal is energized at a time and the rest eight input terminals are at the threshold voltage, by virtue of their respective AND gates in the decoder circuit being at logic state '0'. Because of this, all the current from V_B through Rb is bypassed via the AND gates and hence the transistors connected to these input

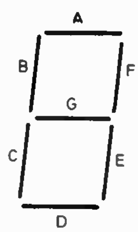


Fig. 1

Table 1

LAMPS	LAMPS TO BE EXTINGUISHED FOR DIGITS									
	0	1	2	3	4	5	6	7	8	9
A		X			X					
B		X	X	X				X		
C		X		X	X	X		X		X
D		X			X			X		
E			X							
F						X	X			
G	X	X						X		

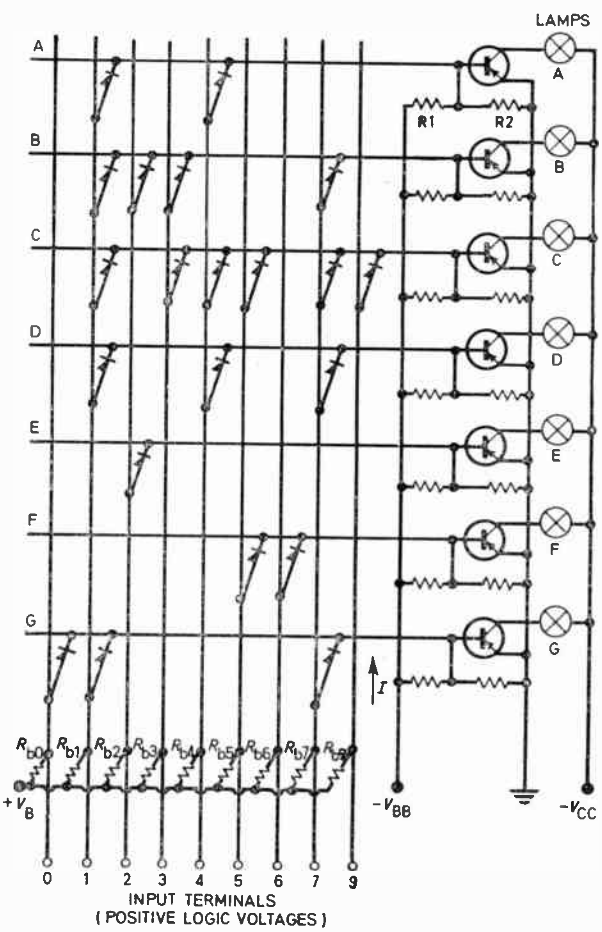


Fig. 2. Full circuit diagram of the display system.

† Instrumentation Limited, Kota-5, Rajasthan, India.

terminals via the diodes are not affected. Thus, only those lamps which are connected to the input terminal which is at logic state '1' are extinguished, displaying the digit represented by that input terminal.

3. Discussion

If this technique of having all lamps illuminated without excitation is employed, there is no need to have one input line for digit 8. Thus, no AND gate is required in the decoder circuit for digit 8.

A minimum value of voltage V_B is required for a particular set of resistors Rb so that current fed to the driving transistors is sufficient enough to cut them off. Voltage V_B can be increased without impairing the performance of the system provided that it does not thereby pose any loading problem on the decoder and other associated circuitry.

The load supply V_{cc} can be used for biasing in place of V_{BB} if it is stabilized.

If the driving transistors are of pnp type, the logic voltages at input terminals should be in positive logic and vice versa. If the threshold voltage at input terminals is zero or slightly positive as is generally the case with positive logic, it will feed a certain amount of current to the bases of the driving transistors connected to the input terminals via the diodes. This amount of current will be sufficient to neutralize the bias of the transistors and they will be held below cut-off. Thus, all the lamps will remain extinguished. Therefore, to avoid this undesirable phenomenon, it is essential to keep the threshold voltage and hence, the voltage at logic state '0' more negative than the base saturation voltage of the driving transistors. For germanium transistors it can be kept approximately equal to $-0{\cdot}5$ V and for silicon transistors approximately $-1{\cdot}0$ V.

4. Acknowledgments

The author is grateful to Brigadier B. J. Shahaney, Managing Director, Instrumentation Limited, and Shri P. P. Dwivedi, Manager (Research and Training) for providing necessary facilities, encouragement and inspiration for this development. Thanks are due to Shri A. K. Jain for assistance and Shri R. N. Mehra for making necessary drawings and printed circuits.

Based on a manuscript first received by the Institution on 27th October 1969. (Short Contribution No. 144/CC100.)

STANDARD FREQUENCY TRANSMISSIONS—April 1971

(Communication from the National Physical Laboratory)

April 1971	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)		April 1971	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	627	621·8	**17**	− 300·0	0	+ 0·1	613	610·4
2	− 299·9	+ 0·1	0	626	620·9	**18**	− 299·9	+ 0·1	+ 0·1	612	610·8
3	− 300·1	+ 0·1	+ 0·1	627	620·1	**19**	− 299·9	+ 0·1	+ 0·1	611	610·0
4	− 300·0	0	0	627	620·4	**20**	− 299·9	0	+ 0·1	610	609·7
5	− 299·9	+ 0·1	+ 0·1	626	619·7	**21**	− 299·9	+ 0·1	+ 0·1	609	609·2
6	− 299·8	+ 0·1	+ 0·1	624	618·8	**22**	− 299·9	0	+ 0·1	608	609·0
7	− 299·9	+ 0·1	+ 0·1	623	616·8	**23**	− 299·9	+ 0·1	0	607	606·3
8	− 299·9	+ 0·1	+ 0·1	622	615·9	**24**	− 299·9	+ 0·1	0	606	604·8
9	− 299·9	+ 0·1	+ 0·1	621	614·6	**25**	− 300·0	+ 0·1	+ 0·1	606	604·1
10	− 299·8	0	+ 0·2	619	613·8	**26**	− 300·0	0	+ 0·1	606	606·8
11	− 299·9	+ 0·1	+ 0·1	618	613·2	**27**	− 299·9	0	+ 0·1	605	607·9
12	− 299·9	+ 0·1	+ 0·2	617	612·5	**28**	− 299·9	+ 0·1	+ 0·1	604	607·2
13	− 299·9	0	+ 0·1	616	612·1	**29**	− 299·9	+ 0·1	+ 0·1	603	605·3
14	− 299·9	+ 0·1	+ 0·1	615	611·7	**30**	− 299·8	+ 0·1	+ 0·1	601	604·3
15	− 299·9	0	+ 0·1	614	611·1						
16	− 299·9	+ 0·1	+ 0·1	613	610·9						

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.

U.D.C. 621.327/8: 629.127: 626.02

Future Electronic Instrumentation for Submersibles, Habitats and Divers

By

KENNETH R. HAIGH,

C.Eng., F.I.Mech.E., M.I.E.E., M.I.E.R.E.†

Reprinted from the Proceedings of the Conference on Electronic Engineering in Ocean Technology held in Swansea from 21st to 24th September 1970.

A review is made of current electronic instrumentation in use in submersibles, habitats and by divers. The design constraints of the underwater environment and their effect on electronics is discussed. A prediction is made of future requirements in the three applications considered.

1. Introduction

In the last two decades the demands to know more about the oceans for military, economic or scientific reasons has reached such a pitch that instrumentation to fulfil these demands has in many instances been hurriedly conceived, designed and built. Catastrophic results have occurred many times. Demand has pushed technology to the limits in what has always been a very conservative discipline. The technological manpower influx to meet the demands has in many instances been unaware of the vagaries of the ocean and design reminiscent of 're-inventing of the wheel' has been a common occurrence.

The problem has been accentuated by demands to work in the water column or on the sea bed, as distinct from working by remote control from the sea surface. In such instances completely new techniques are being evolved or old ones modified, but nevertheless, in some areas, such as underwater navigation, there are still very real unsolved problems. Electronics is playing a very big role in carrying man into the ocean, but there is a lack of appreciation in electronic instrument designers of the problems presented by working there; the ocean is inhospitable to man and hostile to electronics, not the tranquil, scenic milieu depicted in a popular television series.

In a recent article Professor R. E. D. Bishop[1] emphasized the lack of systematic training for ocean engineering in British Universities. There is not a single chair of ocean engineering or even a reader in the subject at any British University. The closest formal approach to oceanographic instrumentation training is that presented at post-graduate level at University College London and at Birmingham by Professor D. G. Tucker with his specialized interest in sonar. The special electronic instrumentation required and the problems encountered by man below the air/water interface are relatively new and not widely known, some of them are therefore presented here.

2. Sub-surface Working

Considering the whole of the world's oceans, 10% of the bottom area is less than 610 metres (2000 ft) deep; 13% of the bottom area lies between 610 m (2000 ft) and 3050 m (10 000 ft); the abyssal plains from 3050 m (10 000 ft) to 6100 m (20 000 ft) account for 75% of the total area, whilst the remaining 2% is accounted for by the deep ocean trenches which equal the depth of 10 900 m (35 800 ft), reached by the bathyscaphe *Trieste* in 1960.

Currently, the maximum depth at which divers are working regularly for commercial purposes is 107 m (350 ft),[2] but this will very soon be overtaken as offshore petroleum wells go deeper. The deepest open ocean dive to date for experimental purposes is 305 m (1000 ft),[3] whilst the deepest simulated dive has been to 457 m (1500 ft).[4] Research monkeys have successfully been subjected to simulated depths of 915 m (3000 ft).[5]

The submersible *Deep Quest* (Fig. 1) holds the record depth of slightly over 2440 m (8000 ft) for a vessel of this type,[6] but is likely to be supplanted by *Deepstar 20000* in the near future[7]; this submersible has a design depth of 6100 m (20 000 ft). Bathyscaphes such as *Trieste*[8] and *Archimede*[9] although capable of reaching the floors of the ocean trenches have been excluded from this paper as they are virtually underwater elevators with very little manœuvring power, payload, or future potential.

The pressure resistance properties of equipment for external use on submersibles should be in one of two depth ranges, to the depth of the break in the continental shelf (200 m) or to the abyssal plain depth. For divers, the ability to work at ease for extended periods on the continental shelf is the first target; once this has been achieved we can proceed down the continental slope as the demand arises. Sea-bed habitats at single or multi-atmosphere pressure, air or mixed gas environment are now in commercial service (Fig. 2),[10, 11] with the possibility of a rapid extension of their use, particularly in the petroleum industry.[12]

It is not sufficient merely to establish the pressure range and atmospheric conditions likely to be encountered by sub-surface instrumentation, there are many other problems such as payload weight and volume available, corrosion, marine fouling, metal fatigue, excessive accelerations caused by cable snatch, problems in the nature of power supplies, mutual interference between equipments, and the fundamental problem of keeping the sea out of the electronics.

No matter what the reason be for venturing below the surface, be it recreative, scientific, military or economic in submersible, habitat or diving dress, many of these problems will be encountered and in looking to the future new ones are bound to arise.

† Admiralty Underwater Weapons Establishment, Portland, Dorset; late Exchange Scientist U.S. Naval Oceanographic Office.

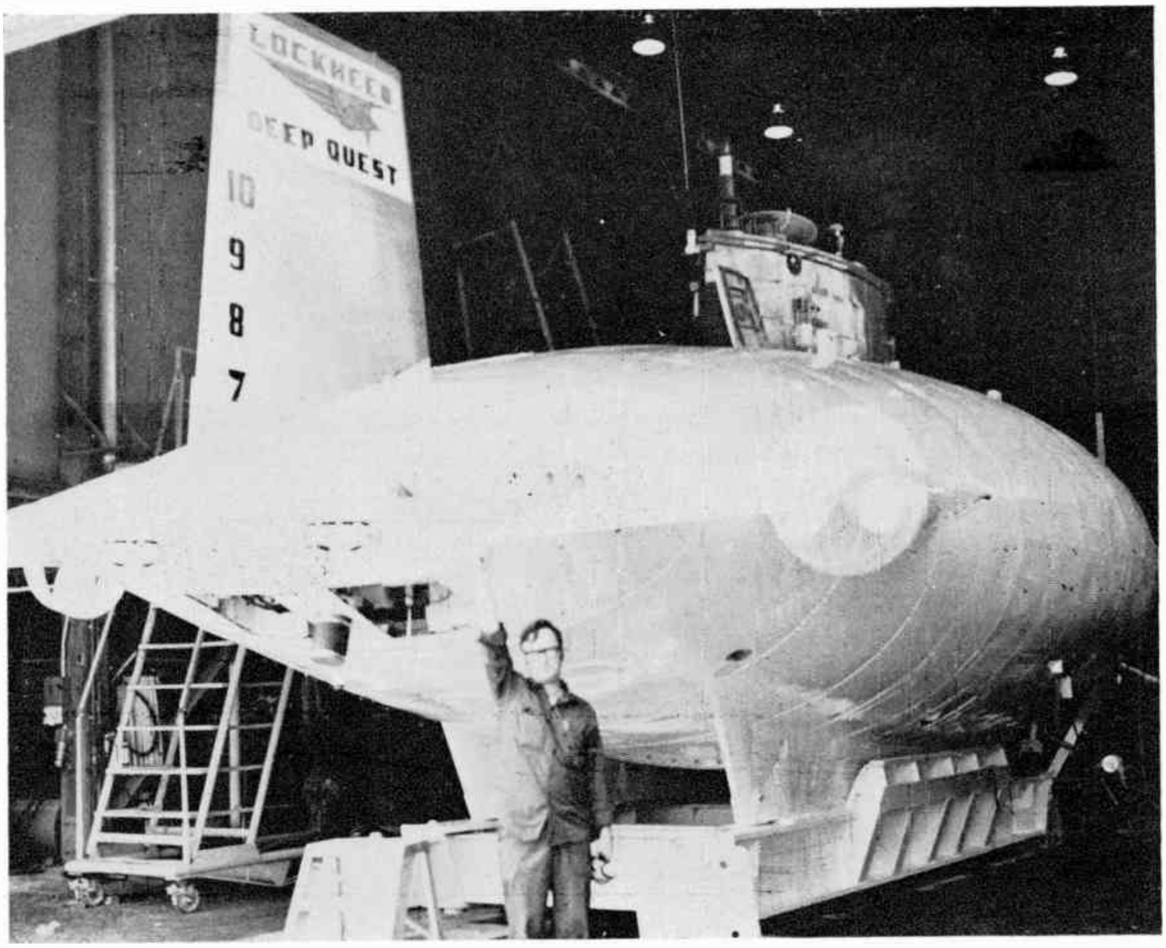


Fig. 1. *Deep Quest.*

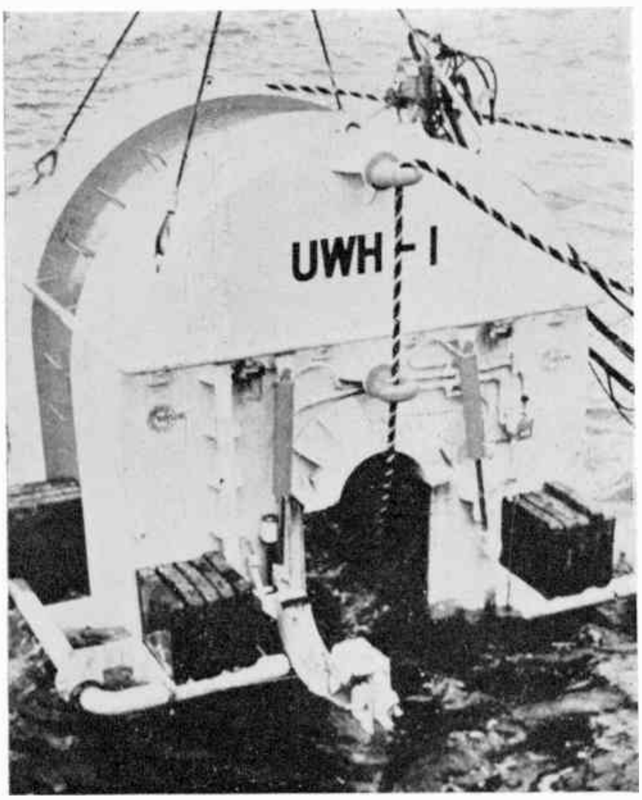


Fig. 2. Habitat for pipeline welding.

3. Submersibles

We are now in the third generation of submersibles; the first generation includes the bathyscaphes, the second, boats such as *Alvin* and *Deepstar 4000*, while *Deep Quest*, *DSRV-1* (deep submergence rescue vehicle) and *Deepstar 20000* exemplify the third generation. The original bathyscaphes were practically void of electronic instrumentation whilst in current vehicles a large proportion of the displacement is made up of electronic equipment.

No matter what the type of submersible, manned free swimming, manned tethered, unmanned free swimming, unmanned tethered, manned bottom crawler, unmanned bottom crawler, there are common problems. In the electronic field these problems are more complex than in any other part of the submersible. In the early days of submersible design there was a tendency to build a vehicle without determining what it was to be used for other than visual observation. Consequently there has been a tendency to fit the boats after launch with instrumentation which was totally inadequate for the task. The desire to produce a general-purpose boat and use it for specialized tasks has not met with a great deal of success. Fortunately this tendency appears to have lost momentum as specialized boats such as *Ben Franklin*,[13] *Beaver*,[14] the *DSRV*[15] appeared, however, even these are being provided with secondary roles. *Ben Franklin* is being considered for pollution studies whilst the *DSRV* will also have an oceanographic

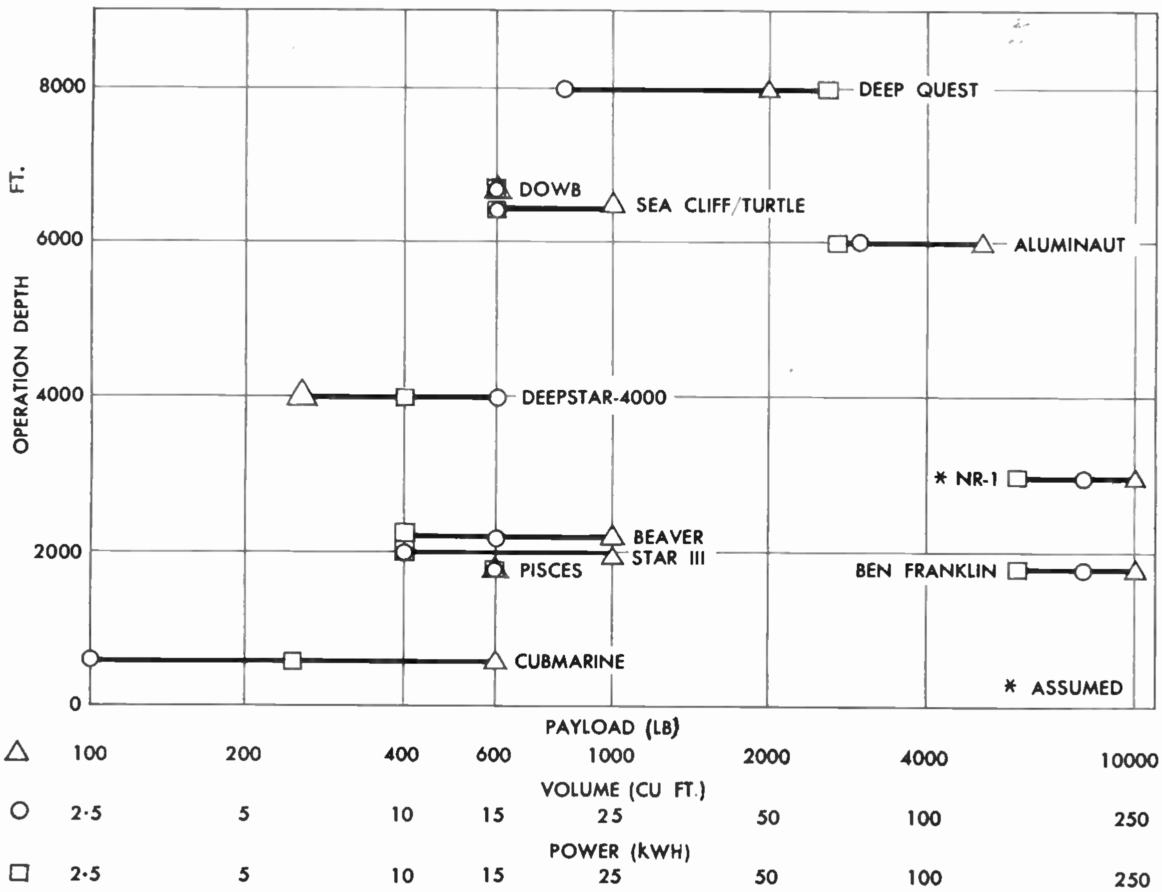


Fig. 3. Submersible characteristics.
1000 lb = 453·6 kg 10 cu ft = 0·28 m³.

facility. If a secondary role is considered in the design stage then many of the problems of the past may be avoided.

3.1 *Instrumentation Design Limitations*

With one exception all submersibles are supported by a parent ship, which in the majority of cases removes the submersible from the water for maintenance purposes such as battery charging, and also for quick transit on the surface. Weight of the submersible is therefore very important and this in turn determines the volume of the submersible, as a weight/displacement ratio of between 0·5 to 0·7 must be aimed for if the boat is to have an effective pay load. In consequence the miniaturization of instrumentation is vital, every kilogramme inside the hull means an overall increase in weight of 2–3 kg because of the need for the addition of buoyant material externally. Instrument volume is also important; the strongest pressure hull configuration is a sphere but this is the worst possible for utilization of space. Instrument power consumption should be as low as possible as the demands of propulsion leave very little for other purposes. Figure 3 illustrates the payload, volume and power available for a range of manned submersibles.

Some of the limitations of weight and volume can be negated by placing instrumentation in pressure tight or pressure balanced compartments external to the pressure hull, however, this technique introduces additional problems of electrical cable glands and difficult access for maintenance. Figure 4 illustrates the technique of carrying extra payload external to the pressure hull and counteracting its weight with additional buoyancy material.

Instrumentation for submersibles must be absolutely reliable, seldom is there any opportunity to carry out the simplest of repair or maintenance task whilst the boat is in the water. Carrying of spares is limited by the lack of space. The close proximity of instruments leads to acoustic and electrical interference between circuits.[16] Extensive use of ill-conceived electrical plugs and sockets external to the pressure hull has resulted in many pieces of equipment flooding; it is only recently that a cabling system to overcome these deficiencies[17] has been introduced into submersibles. Excessive heat and 100% humidity are common. Such are some of the design limitations.

3.2 *Control and Command*

The six degrees of freedom possessed by a submersible are achievable using horizontal and vertical thrusters, hydroplanes, rudders, trim tanks, variable and main ballast. Control of these functions has been direct, with visual instrument feedback to the pilot of their effectiveness. Autopilots are now coming into use where a

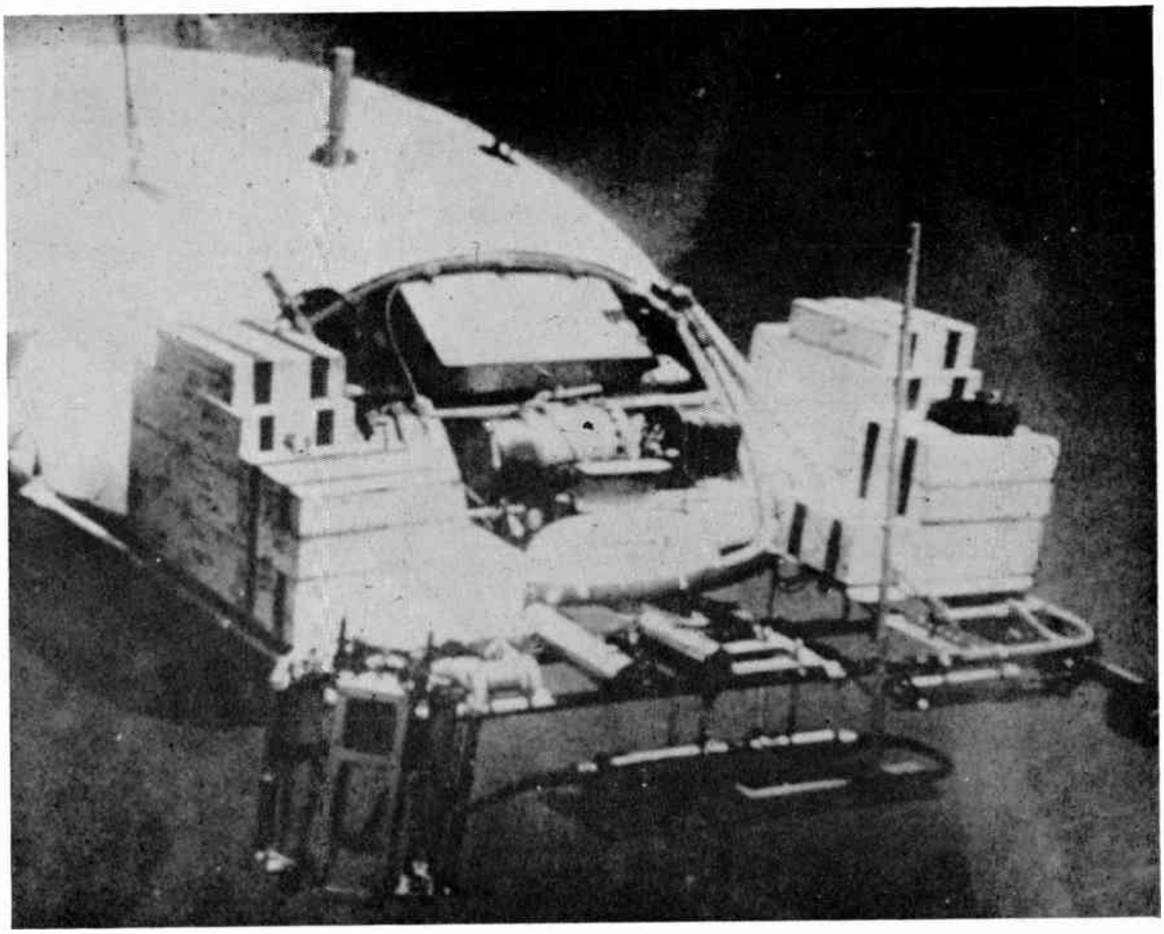


Fig. 4. *Deepstar 4000.*

prescribed course and depth can be maintained. It is claimed that the *DSRV* will be able to hover at any attitude within a solid angle of 45° in a cross-current of 3 knots in order to mate with a distressed submarine. The need for an autopilot is questionable in a small, short duration boat.

3.3 *Pilotage and Navigation*

Pilotage and navigation are dominated by the use of acoustics. Very few boats carry a gyro compass and only one commercial vessel *Deep Quest*, has an autopilot and X–Y plotter. The *DSRV* will have all these plus a miniaturized inertial system.

Typical submersible acoustic installations are shown in Fig. 5 where it will be seen that systems 2–7 and 10–12 are all concerned with pilotage and navigation.

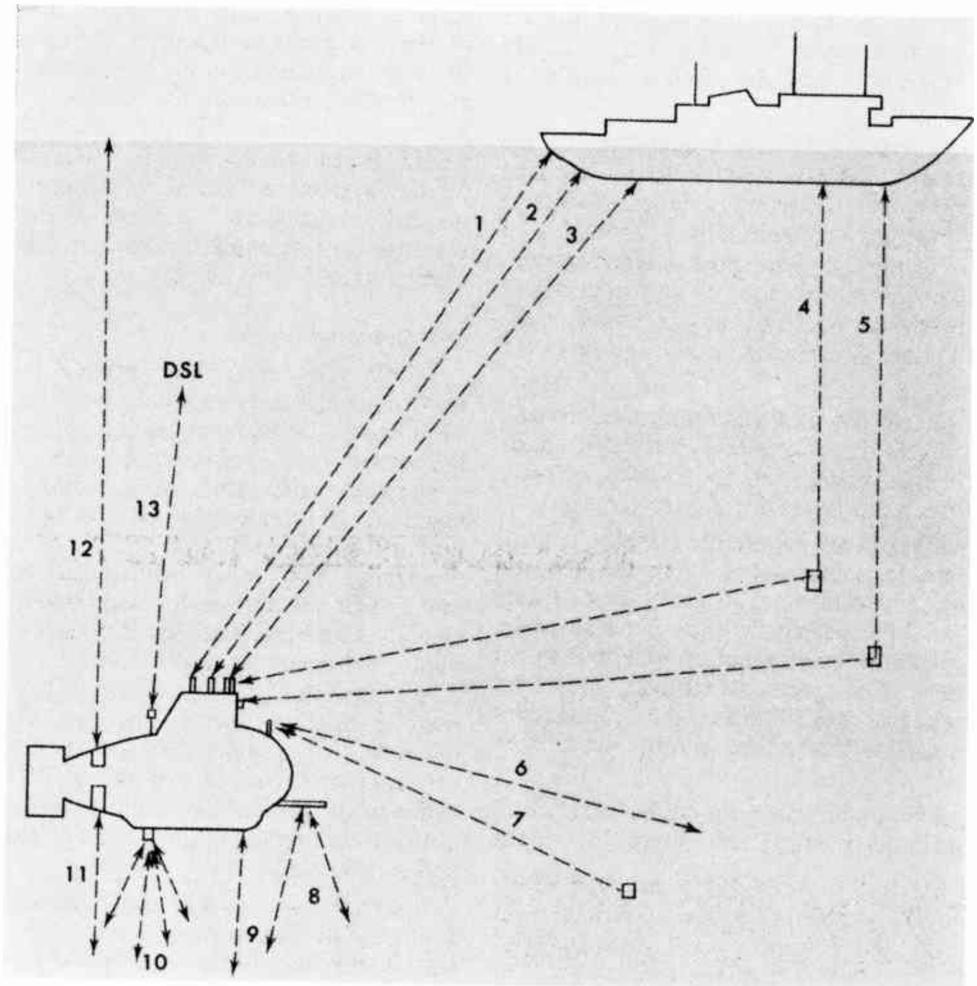
Vehicle depth from the surface and elevation from the bottom are obtained by conventional echo sounding equipments, sometimes one set of electronics switched between up and down transducers. Displays are usually graphic but digital displays are now coming into service. One equipment employs a one-shot digital display, manually triggered, which pilots find most attractive. Acoustic depth measurement is sometimes backed up by strain gauge or Bourdon tube type equipments. One pressing requirement is a high accuracy differential pressure gauge which will allow the submersible to fly a constant height above the bottom.

Obstacle-avoidance sonars are fitted in many boats; continuous transmission frequency-modulated equipments[18, 19] are the most popular although a beam swinging pulse system is also in service. In the smaller boats the weight and volume of these equipments can be a real problem and there is a requirement to reduce their size. Work is proceeding at Birmingham University and Christchurch University, New Zealand, which will be directly applicable to sonars for submersibles.

The problems with submersible navigation systems arise from the fact that there is no natural reference point underwater, nor when an artificial one is created can it be used over long distances. On the surface if all electronic systems fail celestial fixes can be used, or if visibility excludes these, dead reckoning is at hand. Below the surface there is nothing until a reference point is created, even an inertial system requires surface updating from time to time.

The simplest form of reference is a repetitive beacon or transponder carried by the submersible and tracked in bearing and range by the surface ship.[20] Vectors passed by voice communication from the surface to the submersible direct the underwater vehicle's path.

Next in simplicity comes the single beacon, transponder, or timed pinger on the ocean bed, laid from the surface where the point of entry into the water is taken to be the same geodetic position on the bottom. A single beacon can be used to obtain bearing only using a



SYSTEM	FREQ (kHz)	USES
1. U/W TELEPHONE	8·5–11	VOICE COMMUNICATION AND DATA TELEMETRY
2. VEHICLE PINGER	10	BEARING FOR TRACKING
3. VEHICLE TRANSPONDER	32 & 28	RANGE FOR TRACKING
4. TIMED-PINGER MARKERS	3·5-4·5	BOTTOM NAVIGATION (MULTI-VEHICLE & RUGGED TOPOGRAPHY)
5. TRANSPONDER MARKERS	18 & 14-16	BOTTOM NAVIGATION (LONG DURATION)
6. CTFM SONAR	72-87	OBSTACLE AVOIDANCE, OBJECT LOCATION
7. PINGER MARKER	37	OBJECT RELOCATION
8. SIDE-SCAN SONAR	120	MAPPING
9. SUB-BOTTOM PROFILER	6	DELINEATE SUB-STRATA
10. DOPPLER SYSTEM	300	DEAD RECKONING NAVIGATION
11. DOWNWARD SONAR	24	ALTITUDE OF VEHICLE
12. UPWARD SONAR	12	VEHICLE DEPTH, WAVE HEIGHT, ICE THICKNESS
13. BROAD-BAND, TRAINABLE SOURCE & HYDROPHONE		VOLUME REVERBERATION (DSL), BOTTOM REFLECTIVITY

Fig. 5. Submersible acoustic systems.

trainable receiver on the submersible. Range can be added by using a transponder[21] or by using a highly accurate timed pinger with a synchronized receiver in the submersible.[22, 23]

A great many papers[e.g. 24-28] have been written on the fixing of submersible vehicles relative to an acoustic network, built up by the use of multiple transducers in long or short baseline configurations. Examples of typical geometry of the long and short baseline systems are shown in Fig. 6 where the sea bed and submersible units can be either synchronized pingers or transponders. Processing of the time of arrival data takes place in the surface ship.[29]

Instrumentation of the beacons, transponders and synchronized pingers is well established with many different commercial versions on the market. Power sources range from a few hours duration to ten years.

All the foregoing systems require the establishment of an acoustic reference point; not so with the acoustic Doppler navigation system.[30] The direction of travel and the velocity of the submersible are determined by processing Doppler returns of a bottom-reflected signal in four quadrants. The system's usefulness is limited due to its short-range capability and the submersibles inability to return from the surface to a pre-determined bottom site.

The ideal system underwater would be one which presented its data in the way the Decca system does on the surface, together with the use of an automatic track plotter. A phase comparison system is difficult due to the short wavelength of the acoustic wave in water but a range/range system with the computation facilities in the submersible instead of in the surface ship as at present, would appear to be feasible. The very high cost of inertial systems currently rules out their use in commercial submersibles but many of the difficulties with acoustic systems could be eliminated if a cheap inertial system could be produced.

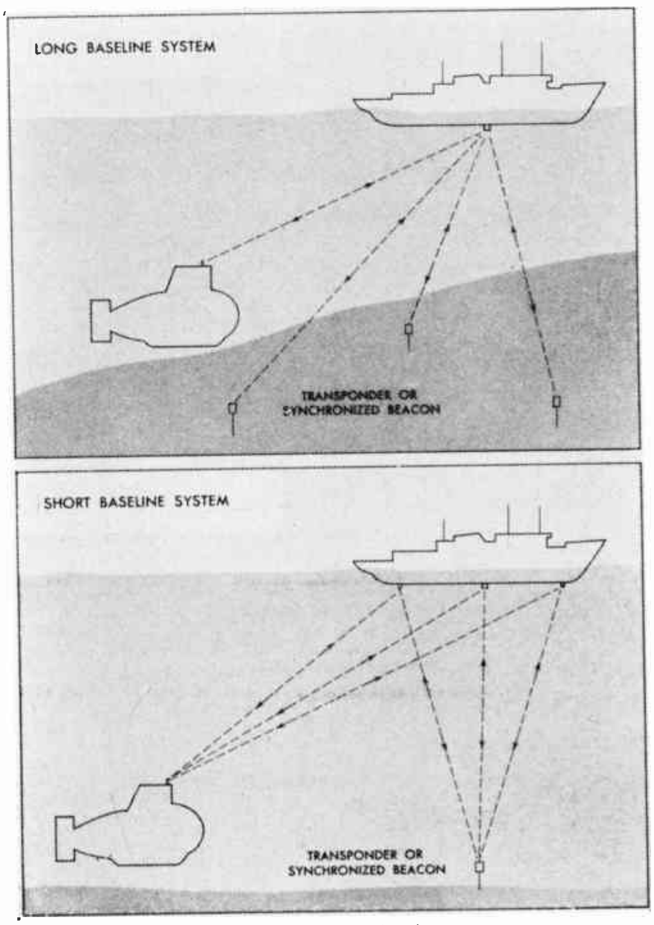


Fig. 6. Submersible navigation systems.

3.4 *Communication*

Within the limitations imposed by the sea water environment, voice communication between surface and submersible, or between submersibles, has been established for some considerable time. Single-sideband suppressed-carrier techniques, usually operating on a standard frequency are in common use in NATO military submarines and many civilian submersibles. Generally, very wide beamwidths are required for underwater telephones to avoid the necessity of continuous searching, but this introduces multi-path links between transmitter and receiver. Consequently an operating technique must be developed to obtain prime performance from the system; slow, deliberate speech is the secret. Improving the acoustic link is one of the principal problems in the sonar world; until this is achieved the present electronic techniques are adequate, notwithstanding the requirements of increased reliability and smaller size.

The question of multi-paths becomes a problem in the telemetry of data over long ranges by means of pulse signals of binary form. A proposal has been made[31] to overcome this difficulty by means of a highly-directional beam. However, in considering telemetry systems for free submersibles the complexity, volume, weight and power requirements of the equipment must be balanced against restricting the vessels operating period and total payload. To date no case is known of telemetry of data from a manned free swimming submersible. In the case of a tethered unmanned submersible with a hard wire connection, telemetry of data presents no undue instrumentation problems and systems developed for trawls can be used.[32]

3.5 *Life Support*

The maintenance of a one-atmosphere air environment in submersibles is straightforward. Instrumentation for monitoring the CO_2 and O_2 content is described under 'Habitats' (Sect. 4).

3.6 *Instrumentation for Specialized Submersible Tasks*

3.6.1 Observation

Under good natural lighting conditions the range of the human eye under water can extend as far as 60 m (200 ft). Normally the range is a fraction of this and means are sought to improve it. External lights on the submersible can be switched on but in many cases this does not help due to the backscatter of light from suspended matter in the water. Swimming the lights in front of the submersible by means of a small remotely controlled vehicle will eliminate most of the back-scattering.[33] Television is affected in a similar manner but the use of image intensifier tubes helps substantially.[34] Gated laser systems have also been suggested as a means of improving the range of visibility.[35] For observation of small targets in extremely turbid water acoustic imaging[36] and acoustic holography[37] are under development.

The most likely fields for development are those of image intensifier tubes and acoustic holography.

3.6.2 Physical oceanography

Every dive that a submersible makes has the potential of providing new information on the ocean. It should be a routine task to collect data on depth, temperature, salinity, chlorophyll content, light transmission, oxygen content, current direction and speed.[38] All these parameters can now be sensed by electronic methods, and a system requires to be engineered and miniaturized as a standard package.

3.6.3 Photogrammetry

Underwater photogrammetry is a new science which has made rapid strides in the last few years.[39–41] It is now possible to produce stereoscopic charts with 1 cm contours. The use of these techniques is likely to expand for scientific and underwater engineering purposes.

3.6.4 Geophysical surveys

The use of magnetometers and gravimeters in the larger submersibles is established.[38, 42–43] Present-day instruments were designed for surface ships where space and weight were not important. For submersibles miniaturized equipments are required.

3.6.5 Geological surveys

Acoustics is an important tool for the marine geologist, with it he can probe the ocean floor with a sub-bottom profiler or examine outcrops with a side scan sonar. Operating near the ocean floor is a great advantage for such equipments as the water column transmission loss is removed which allows smaller source levels and greater definition, a fact which appears to have escaped general notice. There is a requirement to miniaturize side-scan sonar and sub-bottom profilers in order that they may be used in the smaller submersibles.

Measurement of the *in situ* properties of sediments is another requirement. The parameters of interest are sound speed, acoustic impedance, bearing strength, temperature, bulk density and pH.

3.6.6 Biological surveys

The greatest benefit that the submersible brings to the biologist is the ability to observe marine creatures in their natural habitat. An extension of the visible observation is possible by the use of specialized acoustic equipment for probing the deep scattering layer.[38]

4. Habitats

A submersible may be considered to be a 1 atmosphere air environment habitat; so may any sea-bed capsule which internally maintains the sea surface pressure conditions. Instrumentation for such sea-bed living imposes no restrictions above those discussed for submersibles.

In the context of this paper the term 'habitat' is defined as any enclosed structure where the environmental pressure rises above 1 atmosphere; included in this are the diver lockout chambers on submersibles,[44] deck and submerged pressure chambers, pressure transfer capsules and sea-bed dwellings operating at ambient pressure. Any electronic instrumentation included in such habitats must be capable of resisting the maximum working pressure or be enclosed in small pressure-tight containers. Fortunately the use of integrated and solid-state circuitry has eliminated many of the possible danger areas and pressure sensitive components. Sealed air spaces in components should be completely eliminated.

When nitrogen in the breathing gas is replaced by helium, two additional design factors are introduced. Firstly the thermal conductivity of helium is roughly seven times that of air, and secondly the gas will permeate glass, rubber seals, plastics, etc. If advantage can be taken of the increased thermal conductivity, equipment can be run at elevated ratings, or if a helium environment will be present constantly, cognizance of the fact should be taken in the design stage.

Equipment which can withstand air or nitrogen under pressure may not necessarily withstand helium. The elimination of sealed pockets in transistors, capacitors and inductors is vital; their presence in components surrounded by an air or nitrogen pressurized environment may be acceptable, but not so with helium as the main constituent of the breathing mixture. Helium may penetrate to any cavities present, build up to the ambient pressure and, when the pressure is released, shatter the cavity shell. It may even be necessary to decompress the equipment slowly to avoid an explosion. Glass envelope electron tubes should be avoided as far as possible, but where this is impractical, as with television equipment, precautions should be taken.

There is a very real hazard of fire in habitats, particularly in those with oxygen enriched atmospheres. Unfortunately a number of fatalities have already occurred. Any electronic equipment placed in a habitat must be absolutely free from the possibility of electric arcs, even under fault conditions.

4.1 *Habitat Life Support Instrumentation*

In maintaining safety levels of CO_2 and O_2 in habitats there is the additional complication of the partial pressure of these gases. The level of CO_2 must not exceed the equivalent of 1·5% at 1 atmosphere because, e.g., 1·5% CO_2 at 40 metres (5 atm. abs.) will have the same physiological effect as 10% CO_2 breathed at the surface. Similarly with O_2, the partial pressure at 40 m is 1 atmosphere, the limit at which oxygen toxicity begins to appear.

In cases where the habitat environment can be sampled externally and directly by bleeding off a sample of gas, e.g. in a deck decompression chamber, standard gas analysis techniques can be used, such as gas chromatography, mass spectrometry or similar methods. Manual

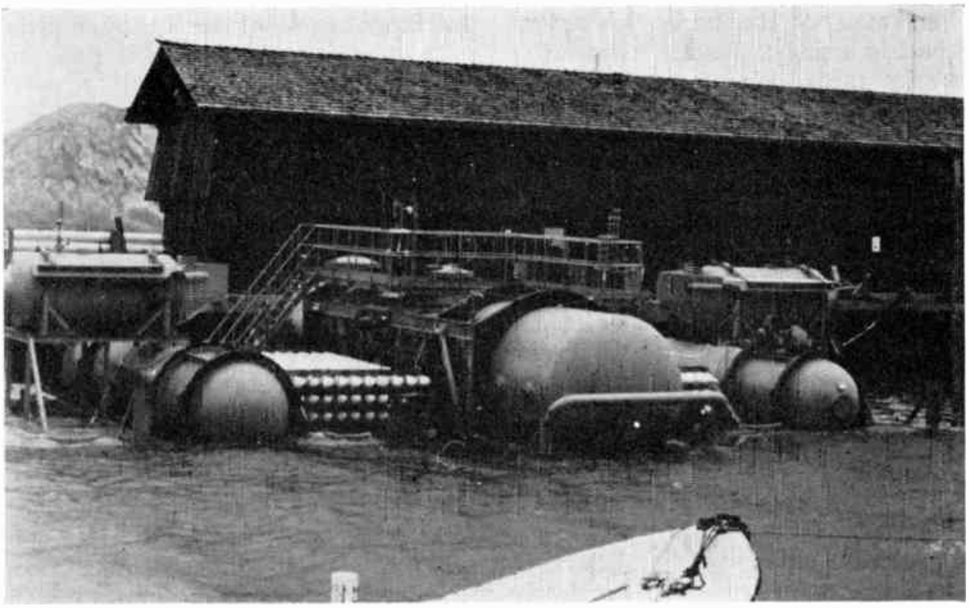


Fig. 7. Habitat *Aegir*.

adjustment of the gas mixture is then made. In the case of a mobile and self-contained habitat like the *Aegir* (Fig. 7),[45] a self-contained and fully automatic gas control system is desirable.

Instruments are available which will measure directly pCO_2 and pO_2;[46] control commands from these instruments can be made to operate the CO_2 scrubbers or bleed in O_2 as required.

Measurement of the pCO_2 is independent of the total sample pressure. A pH electrode together with a reference electrode are separated from the gas mixture by a membrane permeable to CO_2. The electrolyte pH will change as a function of exposure to CO_2, the resulting electrode potential being proportional to the logarithm of the pCO_2.

Measurement of the pO_2 is also carried out electrochemically.[47] The sensor contains a gold cathode and silver anode immersed in a gel electrolyte in contact with a Teflon membrane. Oxygen permeates the membrane and is reduced electrochemically at the cathode, producing a small current which is directly proportional to the pO_2.

In habitats subject to long-term occupancy the build-up of other gases can also create problems, e.g. carbon monoxide and nitrogen dioxide. Monitoring of up to forty different gases by gas chromatograph has been a routine chore where the sample can be expanded to 1 atmosphere. In bottom dwellings at ambient pressure isolated from the surface, no means currently exist for monitoring these contaminants. A fully automatic system with built-in alarms is required.

4.2 *Communications*

Voice communication between the pressurized environment and the control post at atmospheric pressure has in the past been by a wire system, however, with future systems isolated from the surface an acoustic link will be a necessity. Problems in addition to those pointed out for submersibles arise due to the pressurized environment and the presence of helium in the breathing mixture. Even when breathing air under pressure, distortion occurs but this can be corrected by the use of frequency compensation in the receiver. When breathing a mixture of oxygen and helium considerable distortion occurs leading to the well-publicized 'Donald Duck' effect, where the intelligibility may be as low as 10%. There are three techniques currently employed to improve the intelligibility: vocoder methods,[48] band shifting[49] and time domain systems.[50]

In the vocoder system a continuous real-time analysis is performed on the input speech signal to extract information about the time-varying characteristics of the sound energy, termed the source function, and also the time-varying emphasis of frequencies within the regions of resonance of the vocal track, termed the transmission function. A typical vocoder may have 16 channels 200 Hz wide to describe the transmission function, plus a further channel to describe the source function. The output of the analyses consists of slowly varying d.c. signals with a bandwidth of 25 Hz which, after transmission are used to control the operation of a synthesizer which remakes the speech.

In the bandshifting technique a simple heterodyne system can be used where the speech band is split into three or four channels by bandpass filters, and each band is shifted downward in frequency by single sideband modulation and demodulation. The bands are then combined to remake the speech which is more intelligible but still distorted.

The time-domain method is very new and shows exceptional promise. This method was in use during the recent record simulated dive to 510 m at the Royal Naval Physiological Laboratory[4] where for the first time the medical staff were able to hold conversations with the divers with almost no loss of intelligence.[51]

5. Divers

The first successful diving equipment was that produced by Augustus Siebe in 1837; it took the familiar form, little unchanged today, of metal helmet, closed flexible dress, breastplate and heavy boots. Air was supplied by hose from the surface and vented to the sea after use. The only overture to electronic instrumentation in latter years was the incorporation of a telephone line in the diver's lifeline with microphone and earpiece built into the helmet.

Self-contained regenerative type of diving equipment is also British in its origin; Henry Fleus produced the first equipment in 1878. A CO_2 absorbent was carried in a canister on the diver's back, his exhalations being passed through the absorbent to a breathing bag where additional oxygen was added from another cylinder. In this type of equipment the control and mixing of gases has, until recently, been manual or mechanical; electronic gas monitoring and control has now been introduced. This type of equipment requires no supply of gas from the surface, but nevertheless, physical contact with the surface is usually maintained by means of a safety line.

Hard-hat diving restricts the diver's movements considerably due to the air hose, lifeline and cumbersome equipment. Divers with compressed-air scuba gear can be free of the surface and weightless. Working divers below the compressed air working level currently are not free. Even if they are using self-contained regenerative equipment, a lifeline and telephone are usually used to connect them either to the surface or to the pressure transfer capsule which brought them down. In cold water the diver may also be supplied by hose with hot water to keep him warm.

It takes a tremendous amount of courage to sever the final umbilical and leave yourself in cold darkness in a weightless condition, with no knowledge of your orientation and aware that, if any equipment fails or you rise to the surface, the results could be fatal. Until we can provide the diver with utterly reliable instrumentation to control his gas mixture, to navigate, to see, to communicate, to keep him warm, to control his buoyancy, and to do all these things in the way that his natural faculties do, we will always have the final lifeline. All this instrumentation must be packaged around the diver's person in such a manner that leaves him free to move and work. For man to be really free in the sea that lifeline is a barrier.

5.1 *Breathing Apparatus*

In deep diving, nitrogen is replaced by helium in the breathing mixture to avoid nitrogen narcosis and oxygen must be kept at a level to avoid oxygen toxicity. If the diver is supplied from the surface or via a hose from a pressure transfer capsule the gas mixing is controlled on the surface, depending upon the diver's depth. If he is using a modern closed-cycle regenerative equipment his gas mixture is controlled automatically.

5.1.1 Closed-cycle breathing system

The principle of operation of a closed-cycle breathing system is shown in the schematic of Fig. 8. Exhaled gas from the diver passes through a CO_2 absorbent into a diaphragm assembly. An O_2 sensor and associated electronics measures the partial pressure of the O_2 of the exhaled gas every five seconds and compares this value with the set control point which can vary between 0·2 and 1·2 atmospheres of oxygen. If the measured value is lower than the set value a solenoid in the oxygen supply line operates to admit a pulse of oxygen into the diaphragm assembly. The value of the partial pressure O_2 is displayed on a wrist mounted instrument along with alarm circuits for electronic failure. Redundancy is

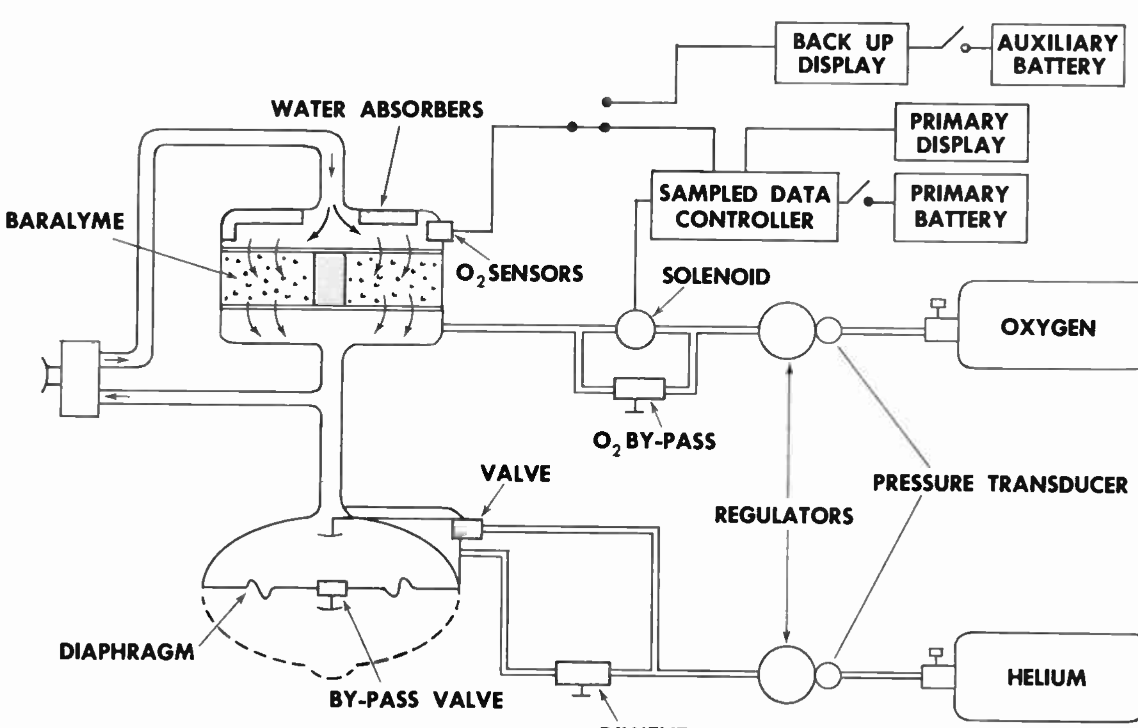


Fig. 8. Closed-cycle breathing equipment.

built into the equipment with a secondary display worn around the neck; a manual operating procedure is also incorporated. In addition to the displays the diver is provided with pressure gauges on his O_2 and He supplies; data from these displays and gauges must be routinely checked.

The above is typical of current design in closed-cycle rebreathing apparatus where the diluent gas can be air, helium or a mixture of helium and oxygen. Other diluent gases such as argon, neon and hydrogen are being considered and these could possibly require a new O_2 sensing system.

5.1.2 Research in breathing apparatus

Most current research on underwater breathing centres around the diver carrying his breathing gases with him in gaseous or liquid form but two revolutionary systems have been demonstrated. The first extracts O_2 directly from the water through a silicone rubber membrane,[52] the other eliminates completely the gas phase in the lungs, and uses a fluid as the respiratory medium.[53] By using a balanced ion, pressure oxygenated, isotonic solution, mice and dogs have been kept alive for long periods of time. If these two methods can be developed into practical systems they will undoubtedly require instrumentation for gas pressures, flow rates and monitoring of the physiological parameters of the diver.

5.2 *Diver Heating*

A diver breathing helium loses his body heat at a rate seven times faster than he does breathing air or nitrogen and he therefore requires to be heated. Electrically and hot-water heated suits are now in use. Chemical and radioisotope suits have been built.[54] The feasibility of radio-frequency heating for divers still requires to be explored.

5.3 *Communications*

The solution of voice distortion when breathing helium in habitats will not, unfortunately, solve all the diver's communication worries, helium speech being only one facet of a complex problem. The ability to produce intelligible speech whilst wearing an oral/nasal mask, a regulator, a muzzle or a full face mask, requires research. What happens to the vocal tract when it is loaded by such devices? Special microphones tailored for particular breathing apparatus are a distinct possibility. Hearing acuity is little affected by ambient pressure, but on immersing the head in water a substantial hearing loss is encountered.[55] Binaural listening underwater is in its infancy as the hearing mechanism is not yet completely understood. When the head is immersed, undoubtedly bone conduction is a very active component of the perception of sound, as the cochlea is excited directly through the skull.

In addition to the diver instrumentation problems, the nature of the propagation path between transmitter and receiver and its ability to transmit intelligible sound is another problem. Currently, single-sideband suppressed carrier systems working at a standard frequency are the most commonly used as they are compatible with submarine equipments, however, other frequencies have been used.[56] Direct audio transmittal into the water is also used over short distances.[57] Ionic conduction of the water path is another method currently available.[58]

There are many instrumentation problems to be overcome before the diver can communicate to such a degree as to make him really free.

5.4 *Orientation and Navigation*

Free of physical or visible contact with the surface or bottom, a diver never has the confidence of his terrestrial environment in determining whether he is on his head or his heels, it requires a reference such as a gas bubble ascending, plus a thought process to determine such a simple thing, such is the problem of orientation and navigation underwater. A method is required to feed the diver information on his vertical orientation without him having to perform a task to determine this. Optical projection onto his faceplate of an artificial horizon is one possible method of achieving this. Acoustic methods are possible but undesirable due to the extensive use of acoustics for other purposes.

Several hand-held sonar sets have been developed over the past two decades to enable a diver on the sea-bed to move around and locate obstacles or targets. Aural and visual display methods have been employed, the latter being limited in cases of poor visibility. Such sonar equipments do give the diver some extension of his visual perception but their operation is cumbersome as they occupy the diver's hands. An acoustic location system for the blind being developed by Professor L. Kay of Christchurch University, New Zealand, has the potential of being developed into a system suitable for divers.

Fig. 9. Guidance system for the blind.

In its terrestrial form the apparatus is built into spectacle frames, with a binaural display (Fig. 9); it is a f.m. system. Engineering of this equipment into a diver's mask or faceplate should present few problems. Such a system will at least give the diver initial confidence to move around from a given datum but he still requires to navigate from place to place.

The establishment of an acoustic net in the manner described for submersibles, using synchronized beacons or simple transponders, is one method of approach. Acoustic interpretation of the received data is the only method considered practical. A monaural receiver measuring time differences from synchronized pingers will give a simple range/range solution, but a binaural system could also produce bearing.

5.5 *Physiological Data*

The human physiology problems of deep diving cannot all be solved in simulated dives in a dry chamber, where heart rate, brain wave, respiration rate and body temperature data can all be obtained without difficulty; it is vital to obtain such data under actual operating conditions. Multi-core cables are possible but cumbersome. A self-contained multi-channel recorder carried by the diver has been used,[59] but real time data would be more useful. What is required is a diver telemetry system of the highest capacity. Research is currently in hand to determine by acoustic means the content of gas bubbles in the bloodstream. If such a measurement could be developed to the stage where every decompressing diver was examined, the incidence of the bends would decline dramatically.

5.6 *Diver's Faculty System*

There are many similar features of the stated requirements for navigation, communication, obstacle avoidance and orientation, features which could be combined in a 'diver's faculty system', powered from a common source. All the transducers could be carried in the helmet or facemask in such a way as to duplicate as far as possible, their natural counterpart. Much of the electronics could also be carried in microelectronic packages carried in the helmet or in special skull caps.

Similarly with all the control and display information from the life support equipment, centralized grouping is required with automatic monitoring of the parameters. The aim must be to free the diver's mind from worries and his hands for work.

Diver instrumentation probably presents the biggest challenge to the electronic engineer, it is certainly the most fascinating.

6. Acknowledgments

Permission to publish this paper was given by the Scientific and Technical Director, U.S. Naval Oceanographic Office, Washington, D.C., and the Director, Admiralty Underwater Weapons Establishment, Portland, Dorset. The views expressed should not be taken as official policy of either establishment, but rather the personal opinions of the author.

7. References

1. Bishop, R. E. D., 'Ocean engineering', *Chartered Mechanical Engineer*, pp. 142–8, April 1969.
2. 'The limits of man', *Offshore*, **29**, No. 9, p. 51, August 1969.
3. Searle, W. F., 'How deep sea work capability is growing', *Ocean Industry*, **4**, No. 4, p. 36, April 1969.
4. 'World record chamber dive at RNPL', *Hydrospace*, **3**, No. 2, April 1970.
5. Delauze, H., 'French divers start deep tests', *Offshore*, **29**, No. 9, p. 65, August 1969.
6. '*Deep Quest:* a deep submergence research system', *Lockheed Star*, **14**, No. 33, 29th March 1968 (Lockheed Missile and Space Company).
7. Pritzlaff, J. A., 'The development of *Deepstar 20000*', Marine Technology Society Sixth Annual Conference, July 1970.
8. Piccard, J., Dietz, R. S., 'Seven Miles Down'. (Putnam, New York, 1961).
9. Terry, R. D., 'The Case for the Deep Submersible' (Western Periodicals Co., Hollywood, California, 1965).
10. Landry, O., 'Old man in the sea', *Undercurrents*, **3**, No. 1, pp. 4–15, May 1970.
11. 'Now–cold welds with new underwater welding technique', *Ocean Systems News*, **1**, No. 2, 1969.
12. 'Subsea oil system on test', *Hydrospace*, **3**, No. 2, p. 18, April 1970.
13. Dolan, F. J., 'The Gulf Stream drift mission', Marine Technology Society Sixth Annual Conference, July 1970.
14. '*Beaver Mark IV* Submarine Work Boat', North American Rockwell Corporation publicity booklet.
15. Feldman, S., 'Developing a new breed of deep-submergence vehicles', *Astronautics and Aeronautics*, pp. 44–8, July 1967.
16. Haigh, K. R., 'Instrumentation interference in submersibles', Marine Technology Society Sixth Annual Conference, July 1970.
17. Haigh, K. R., 'Deep sea cable gland system for underwater vehicles and oceanographic equipment', *Proc. Instn Elect. Engrs*, **115**, No. 1, pp. 153–7, 1968.
18. 'CTFM sonar goes deep', *Underseas Technology*, Reprint R201, May 1967.
19. Kay, L., 'An experimental comparison between pulse and frequency modulation echo-ranging systems', *J. Brit. I.R.E.*, **20**, p. 785, 1960.
20. Merrifield, R. and Fagot, M., 'Tracking the *Ben Franklin* during the Gulf Stream Drift', Proc. Inst. of Nav. Second Symposium on Manned Deep Submergence Vehicles, November 1969.
21. Mackenzie, K., 'Early history of deep submergence navigation aboard *Trieste*', Proc. Inst. of Nav. Second Symposium on Manned Deep Submergence Vehicles, November 1969.
22. Spiess, F. N., 'Underwater acoustic positioning applications', Proc. First Marine Geodesy Symposium, Columbus, Ohio, 1966, pp. 93–101 (U.S. Govt. Printing Office, Washington D.C., 1967).
23. Marquet, W. M., Webb, D. C. and Fairhurst, K. D., 'A recoverable deep ocean navigation beacon', 4th National ISA Marine Sciences Instrumentation Symposium, 1968, Plenum Press.
24. Campbell, D. E., 'Precise acoustic navigation and position keeping', Proc. Inst. of Nav. Second Symposium on Manned Deep Submergence Vehicles, pp. 31–51, November 1969.
25. Lowenstein, C. D. and Mudie, J. D., 'On the optimization of transponder spacing for range-range navigation', *J. Ocean Technology*, **1**, No. 2, 1967.
26. Van Calcar, H., 'Precision subsea navigation system', Proc. Inst. of Nav. Second Symposium on Manned Deep Submergence Vehicles, pp. 155–72, 1969.

27. Merrifield, R. and Delort, R. R., 'Results with a deep research vehicle transponder navigation system', 4th National ISA Marine Sciences Instrumentation Symposium, 1968, Plenum Press.

28. Fain, G., 'A bench mark navigation system for surface and sub-surface vessels'. Inst. of Nav. 24th Annual meeting, June 1968, Monterey, California.

29. Mudie, J. D., 'Applications of sea going computers', Undersea Technology Handbook Directory, pp. A3–A7, Compass Publications, 1969.

30. Goulet, T. A., 'The use of pulsed doppler sonar for navigation of manned deep submergence vehicles', Proc. Inst. of Nav. Second Symposium on Manned Deep Submergence Vehicles, 1969.

31. Berktay, H. O., 'Self-aligning arrays for communication in multipath conditions', *Electronics Letters*, **3**, p. 441, 1967.

32. Mowat, M. J. D., 'A digital acoustic telemeter for fishing gear research', Proceedings of I.E.R.E. Conference on Electronic Engineering in Oceanography, September 1966, Paper No. 20.

33. Rebikoff, D. I., 'The case for unmanned underwater systems', *Sea Frontiers*, **13**, No. 3, May/June 1970.

34. 'Poor light tv tube', *Ocean Product News*, May/June/July 1970.

35. McManus, R. G. and Chabot, A., 'Scan-synchronized receiver optimizes laser ranging', *Laser Focus*, October 1966.

36. Haslett, R. W. G., Pearce, G., Welsh, A. and Hussey, K., 'The underwater acoustic camera', *Acustica*, **17**, p. 187, 1966.

37. Greene, D. C. and Hildebrand, B. P., 'Acoustic holograms for underwater viewing', Parts 1 and 2, *Ocean Industry*, **4**, No. 9, September 1969 and October 1969.

38. Haigh, K. R., 'Oceanographic instrumentation for *Ben Franklin*', *Underwater Journal*, **2**, No. 1, pp. 11–16, March 1970.

39. Pollio, J., 'Applications of Underwater Photogrammetry', U.S. Naval Oceanographic Office, Report IR 68–52.

40. 'Underwater photo optics' Seminar Proceedings, Society of Photo-optical Instrumentation Engineers, 7, 1966.

41. 'Underwater Photo-optical Instrumentation Applications' Seminar Proceedings, Society of Photo-optical Instrumentation Engineers, 12, 1968.

42. Higgs, R. H. and Carroll, J. C., '*Aluminaut* Magnetometer Operations St. Croix, Virgin Islands, 1966'. U.S. Naval Oceanographic Office Report IR 67–33.

43. Thompson, L. G. D., 'Gravity and Magnetic Instrumentation Systems on Deep Oceanographic Survey Vehicles', GCA Corporation Report prepared for U.S. Naval Oceanographic Office.

44. MacLersh, K., 'A taxi for the deep frontier', *National Geographic Magazine*, pp. 139–50, June 1968.

45. 'Makai Range habitat gives faultless performance', *Ocean Industry*, **5**, No. 2, pp. 42–3, February 1970.

46. ATO Bulletins 2004, 2008, 2009. Beckman Instruments Inc.

47. U.S. Patent 2,913,386.

48. Roworth, D., 'Helium speech', *Hydrospace*, **2**, No. 2, pp. 26–8, June 1969.

49. Copel, M., 'Helium voice unscrambling', *I.E.E.E. Trans. on Audio and Electroacoustics*, **AU-14**, pp. 122–6, 1966.

50. Gill, J. S., Morris, R. J. and Edwards, M. G., 'A helium speech processor operating in the time domain', Proceedings of I.E.R.E. Conference on Electronic Engineering in Ocean Technology, September 1970.

51. Personal Communication, Surgeon Commander D. Elliot, R.N.

52. Williams, P., 'Breathing under water', *Weekend Telegraph*, p. 20, 12th August 1966.

53. Klystra, J. A., 'Survival of submerged mammals', *New England J. Medicine*, No. 272, pp. 198–200, 1965.

54. Majendie, J. L. A., 'Diver heating', Proceedings Symposium on Equipment for the Working Diver, Marine Technology Society, February 1970.

55. Brandt, J. F. and Hollien, H., 'Underwater hearing thresholds in man as a function of water depth", *J. Acoust. Soc. Amer.*, **46**, pp. 893–7, October 1969.

56. Gazey, B. K. and Morris, J. C., 'An underwater acoustic telephone for free swimming divers', *Electronic Engineering*, **36**, p. 364, 1964.

57. 'Diver's Intercom System', CBS Laboratories, Stamford, Connecticut, Project 6052.

58. Hoe, R. F., 'Underwater wireless communication', Proceedings of Symposium on Equipment for the Working Diver, Marine Technology Society, February 1970.

59. 'Monitoring man-in-the-sea', *Ocean Systems News*, **1**, No. 2, 1969 (Ocean Systems Inc., New York).

Manuscript received by the Institution on 14th July 1970. (Paper No. 1384/AMMS 37.)

U.D.C. 535.653.8 : 621.397.331.2

B.B.C. Test Chart 57, A New Grey-Scale Reflectance Chart for Colour Cameras

By

S. J. LENT, C.Eng., M.I.E.R.E.†

Reprinted from the Proceedings of the Conference on Television Measuring Techniques held in London from 12th to 14th May 1970.

The design considerations and use of a new grey-scale reflectance chart for colour cameras are discussed in relation to some of the parameters of a colour television system. The construction of a prototype chart based on the new design specification is described.

1. Introduction

In order to achieve the optimum colour fidelity in the existing colour television systems it is necessary for the colour signal source to produce equal amplitude colour separation video signals at the different levels representing white, greys and black in the original scene. In the case of a colour camera this is normally obtained by correctly exposing the camera to a neutral grey-scale produced either by a rear-illuminated transparency or by a front-illuminated reflectance chart and then adjusting the relevant camera controls to give the balanced red, green and blue signals as described above. This type of test chart or transparency is normally intended to have additional functions to aid camera set-up and enables the transfer or gamma characteristic of the camera to be checked and the signal level corresponding to black to be set.

Test charts and transparencies of this type have been in existence for a considerable time but the now much-improved performances of modern colour cameras have created a need for a new grey-scale with improved neutrality, increased contrast range and with modified contrast law characteristics.

This paper discusses the considerations given to the design and use of a new reflectance chart, B.B.C. Test Chart 57, to meet these requirements and also describes the construction of a prototype based on the new design.

2. The Scene Illuminant and the White Point of a Colour Television System

When a reflectance grey-scale is used for the camera colour-balancing procedure it is useful to consider the relationship between the scene illuminant and the white point in a colour television system. In specifying the colour analysis of a colour television system, account is taken of the chromaticities of the shadow-mask tube phosphors on which the colour-separation pictures will be reproduced. It is theoretically possible to calculate for one particular chromaticity of the scene illuminant and for a particular set of primary chromaticities, ideal camera analysis characteristics which will give (within the colour gamut defined by the primary chromaticities) exact chromatic reproduction of the original scene. In this situation the display white-point and the scene illuminant would have the same chromaticity. In present-day practice, however, the white point of the display is set to an agreed chromaticity, Illuminant D_{6500}, and the colour camera is balanced to suit the scene illuminant, which may vary over a wide range. When the scene illuminant is not D_{6500} the chromaticities and relative luminance values reproduced by the television system will inevitably differ from those visually observed in the original scene.

Optimum colorimetry could be taken to be the case in which the colour reproduction of the scene is identical with that which the scene would have if illuminated with D_{6500}. Linear matrix techniques[1] can be used to obtain a performance which is reasonably close to this optimum.

3. The Neutrality of the Grey-Scale and the Reflectance Chart Background

An investigation[2] has shown that a 4% change in the unity ratio of red, green and blue signals in a linear system can produce a noticeable colour change in a receiver display balanced at D_{6500}. As the overall gamma of current television systems tends to be higher than unity, an even smaller tolerance in this ratio is required in order to preserve a good colour balance and in particular to obtain the best reproduction of facial tones.

In order to minimize colour balance errors the reflectance grey-scale must be an excellent neutral and should if possible be placed in the scene to reflect the scene illuminant. A difficulty often found with rear-illuminated transparencies lies in achieving precisely the same illuminant as in the scene to be televised.

In terms of the grey-scale neutrality it was estimated from these considerations that a uniformity of reflectance of at least $\pm 4\%$ over the spectral range from 400 nm to 700 nm would be required and the spectral characteristic should not contain any abrupt changes in the central region of this band. Although the neutrality of the background to the chart is of slightly less importance, it must nevertheless be good. The final assessment of colour balance is always made subjectively from a colour display monitor and the assessment can be made more accurately if the grey-scale and background are of equal neutrality. It is also useful if the background is sufficiently neutral and of uniform reflectance so that any significant shading effects introduced by the camera can be revealed.

4. White Reflectance and Contrast Range

In most television studios it is current practice to limit the light reflected from the bright areas in a scene to about 60% of the incident light so that with normal camera exposure, facial tones are displayed on the

† B.B.C. Research Department, Kingswood Warren, Tadworth, Surrey.

receiver at subjectively acceptable luminance levels. If the maximum reflectances in the scene were higher, the high overall gamma of the system would cause faces to be reproduced too darkly.

The white step of the grey-scale in the test chart should therefore have a reflectance of 60% to represent the brightest area in a scene and the camera, when fully exposed to this, will produce maximum amplitude signals. Any specular reflexions subsequently encountered in a scene will be reduced to 'white' level by the action of the peak white clippers in the camera. In television studios, the light levels in the scene in excess of 60% 'white' can be to a large extent controlled but this is not possible with outside lighting. Here the test chart is used for colour-balancing as in the studio and then if necessary the camera exposure is adjusted to accommodate that part of the scene contrast range in which the detail is required.

The contrast ratio encountered in a studio scene can be as much as 100 : 1 and 1000 : 1 in an outdoor scene. Until recent years the performances of cameras and receiver display tubes limited the overall system contrast ratio to about 20 : 1. However, the modern colour camera equipped with electronic flare correction is capable of handling a contrast ratio approaching 100 : 1. Similarly modern colour receivers viewed under favourable ambient lighting conditions are capable of displaying a similar contrast ratio but for the average range of studio scenes and receiver display conditions a contrast ratio of 40 : 1 is regarded as more typical.

This value was therefore considered as the smallest range over which accurate camera colour balance tracking should be achieved and was therefore adopted as a useful working contrast ratio for a reflectance grey-scale. With the white step already set at 60% this would mean that the darkest step is required to have a reflectance of $1\frac{1}{2}$%.

5. Grey-Scale Contrast Law

If the steps of the grey-scale are equally spaced and arranged to run horizontally across the reflectance chart, the resultant line waveform before gamma correction, from a camera exposed to the chart describes the contrast law of the grey-scale. The steps could be made to run vertically and reference made to the field waveform but the dimensions of the grey-scale would then be limited due to the aspect ratio of the chart.

The most suitable contrast law for a test card grey-scale would have a gamma exponent equal to the reciprocal of that of the output of the camera. The resultant output video waveform would then appear as a linear step scale when the gamma correction was correct. This is a convenient waveform as the equal step intervals facilitate the operational aspect of the tracking procedure and the linear display provides a quick check of the uniform gamma correction of different cameras. If two horizontal grey-scales are arranged on a uniform grey background in the top and bottom halves of the chart a distinction can then be made readily between an incorrect camera gamma characteristic and shading effects.

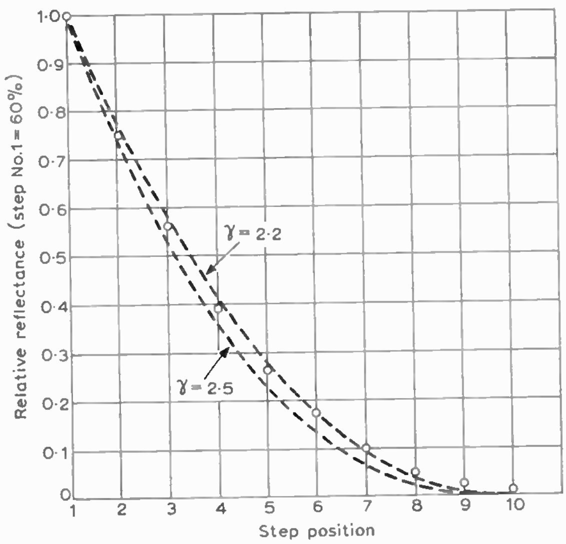


Fig. 1. Reflectance characteristics proposed for grey-scale with equally-spaced step positions on the chart.

Dotted lines show gamma characteristics for comparison. Step position 10 corresponds to the super-black.

The gamma correctors available in the current types of colour camera are not capable however of fully compensating the higher gamma of modern display tubes[3] and also vary somewhat in the contrast range over which the correction follows a pure power law. This does not normally exceed a contrast ratio of 40 : 1 at maximum correction and is limited for practical considerations such as the effect of noise and lag in the camera signals.

From the measurements made of the gamma characteristics of different types of camera it would seem that a practical compromise for a grey-scale is to have a contrast law with an exponent of 2·5 near white, reducing to a value of 2·2 in the darker greys. This then allows the camera to produce a substantially linear waveform over the contrast range of 40 : 1. The reflectance characteristics of the proposed grey-scale are given in Fig. 1, which shows the values of the different steps; nine steps provides a suitable number of signal level samples at which to check colour balance.

6. Super-Black

Most colour cameras are now fitted with forms of electronic compensation for flare produced in the camera optics. This normally operates by measuring the d.c. content of the linear signal corresponding to the scene and it can be adjusted to mitigate the effects of any flare produced in the darker areas of the picture by automatically reducing the black level. A method of setting this compensation is to expose the camera to a small very dark area surrounded by a relatively larger and brighter area. The compensation is then adjusted until the part of the waveform corresponding to the small dark area remains at the same level as the camera exposure is varied from normal to a minimum value by means of the lens iris.

Provision of a special 'super-black' with, if possible, zero reflectance should therefore be made for flare correction, and be positioned in the centre of the chart between the two horizontally placed grey-scales on the background. The background should have a reflectance of about 16% so that the reflectance chart when framed correctly by the camera and used for flare correction would represent the illumination content of an average scene. An additional advantage of the background with this reflectance is that at this level any shading effects due to the camera and revealed by the waveform are neither exaggerated or compressed excessively by the camera gamma corrector.

7. The Use of the Chart in Setting Black Level

When using the chart for setting the camera colour balance it is also necessary to establish a signal level from the gamma corrector which corresponds to the contrast ratio for black in the scene and once set, providing flare correction is in operation, should require no further adjustment irrespective of the scene content, unless special effects were required.

Ideally with a pure-law gamma corrector having, say, an exponent of 0·4 it would be correct to set the level corresponding to the 40 : 1 dark grey step in the gamma-corrected waveform to $0{\cdot}025^{0{\cdot}4}$ (approximately 30%) of white level. This would be rather inefficient as even the maximum contrast ratio likely to be encountered in the studio, of 100 : 1 would result in a transmitted signal equal to 15% of white level. Practically it is difficult to specify the correct level. Typical gamma correctors do not conform to a pure power law in this region and tend to remain at a constant slope to zero. This means that the waveform level corresponding to the 'super-black', with a contrast ratio close to 100 : 1, can be set to represent true black level, i.e. zero level signal from the gamma corrector, and result in only a very small departure from pure power law correction at a contrast ratio of 30 : 1. Figure 2 shows a typical colour camera gamma characteristic and the step waveform after gamma correction,

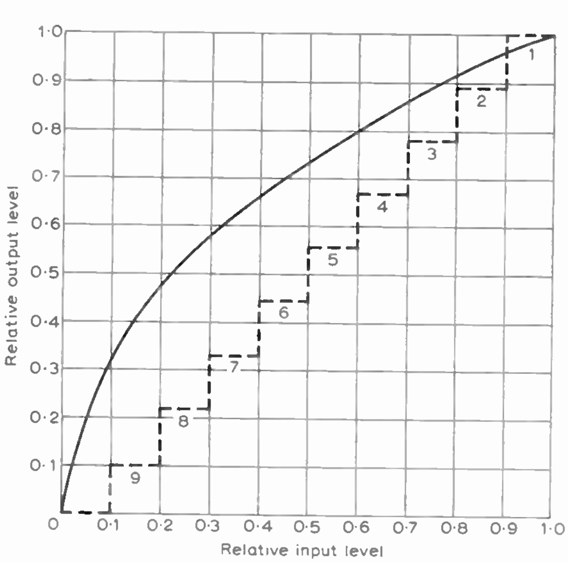


Fig. 2. Transfer characteristics of a typical colour camera gamma corrector nominally set at $\gamma = 0{\cdot}45$.

Dotted line shows calculated waveform obtained from corrector corresponding to reflectance levels of grey-scale step positions given in Fig. 1 and with super-black set to zero.

with 100 : 1 black set at zero, corresponding to the proposed grey-scale reflectances.

Setting black level in this manner also means that any dark area detail in the scene up to 100 : 1 contrast ratio will be transmitted and it is quite possible that some receivers with favourable ambient lighting conditions as previously described will display this detail, with some inevitable compression but with nevertheless some enhancement of the picture contrast.

8. The Prototype B.B.C. Test Chart 57: Colour Camera Grey-Scale

A prototype chart was made to the format shown in Fig. 3 and conforming as closely as possible to the various recommendations made in the previous sections of this paper. An exhaustive search was made for a suitable material from which to make the reflectance grey-scales and after many spectrophotometric tests it was found that one special type of photographic printing paper met the specification. After processing it can have excellent neutrality with a uniform reflectance within ±4% over the required spectral band (400 nm to 700 nm). Many materials were rejected as being too specular in reflectance even though sufficiently neutral, or they were only available in a limited number of reflectance values.

The grey-scale steps with the required reflectances specified in Table 1 were produced from the printing paper by a carefully-controlled exposure and development process. This was followed by rigorous measurement and

Table 1

Nominal reflectance values for Test Chart 57 grey scale

Step	Nominal reflexion densities	Nominal % reflectance relative to magnesium carbonate	Nominal % reflectance relative to step 1
1	0·22	60	100
2	0·35	45·7	76·4
3	0·47	34	56·6
4	0·62	24	40
5	0·8	15·8	26·4
6	1·00	10	16·7
7	1·16	6·9	11·5
8	1·51	3·1	5·2
9	1·82	1·5	2·5
Super-black	2·22	0·6	1·0

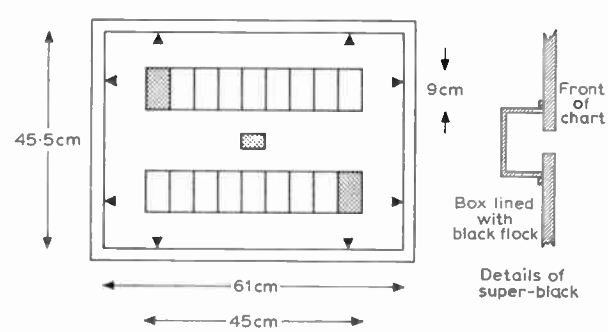


Fig. 3. B.B.C. Test Chart 57 colour camera grey-scale format.

selection to obtain the correct values and uniformity of reflexion density over each step area. The contrast range required for the grey scale however extended beyond the range of the photographic material so that the 40 : 1 low reflectance step was made from a different material having a suitably neutral matt ink-printed surface. Photographic paper was not used for the chart background as it was difficult to obtain a sufficiently small enough tolerance in the reflectance variations over the relatively large area required. Again, an ink-printed paper was used. The 'super-black' was made in the form of an aperture forming the entrance to a small box lined with flock-paper, a type of black paper with a surface consisting of short black fibres. This arrangement reduced the amount of light reflected back out of the aperture to a minimum and produced an effective reflectance as far as could be measured of about 0·6% providing a 100 : 1 contrast ratio with the white steps of the grey scales.

9. Quantity Production of the Test Chart

A number of copies were made of the prototype and tested successfully in operational use at Television Centre over a period of about six months. Discussions were then held with a firm of precision printers, Royle and Son Limited, with regard to development for quantity production under licence. This created a number of technical problems, particularly in respect of maintaining neutrality and reflectance levels accurately. However these problems were overcome and a reasonably economic process which met the requirements was developed, involving both photographic and printing techniques; a photograph of a production version of Test Chart 57 is shown in Fig. 4. A considerable number of these charts have now been manufactured and are already used by many broadcasting authorities both in this country and abroad.

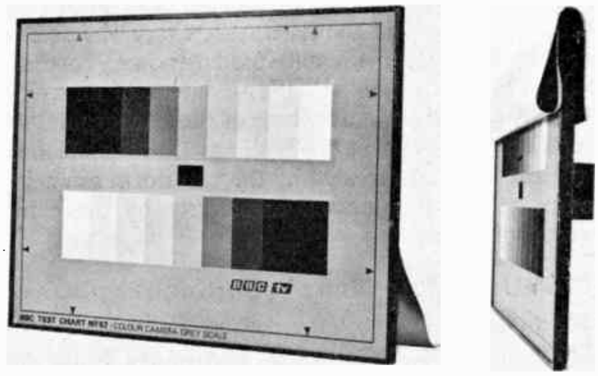


Fig. 4. A production version of Test Chart 57.

10. Acknowledgments

The author wishes to acknowledge the assistance of his colleagues in the Photographic Section of the B.B.C. Engineering Research Department for the preparation of prototype charts. He is indebted to the Director of Engineering for permission to publish this paper.

11. References

1. Jones, A. H., 'Optimum colour analysis characteristics and matrices for colour television cameras with three receptors', *J. Soc. Mot. Pict. Telev. Engrs*, **77**, pp. 108–15, February 1968.
2. 'Dichroic Tilt in Colour Television Cameras', B.B.C. Research Department Technical Memorandum No. T–1089.
3. 'The Contrast Law of Shadow-mask Tubes', B.B.C. Research Department Report No. 1969/30.

Manuscript first received by the Institution on 11th March 1970 and in final form on 21st September 1970. (Paper No. 1385/Com. 44.)