FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

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

THE RADIO
AND ELECTRONIC
ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 34 No. 4 OCTOBER 1967

The Higher Technician

THE increasing diversity and complexity of industry since the second World War has created new specialist occupations calling for men and women of good education, sound training and considerable experience, to stand between the two former industrial groups—the professional man (qualified in science and technology) and the skilled craftsman or artisan. The professional man or technologist who organizes and directs research, design and production, or undertakes general management, needs to be free to look ahead, to plan and to co-ordinate. To do so, he must be able to delegate responsibilities in limited spheres to capable, experienced technicians whose knowledge of science and engineering qualifies them to work in the technologist's field, applying established engineering techniques under the general guidance of the professional engineer. This division of technological effort has been recognized in reports by British† and overseas governments.

Statistical reports suggest that for every professional engineer, three 'higher technicians' and five general technicians are required. Definitions of technicians, as a class, and the higher technician in particular, have filled many books and articles. The only sound ones so far have been in terms of desirable educational attainment. It is becoming generally accepted that the higher technician should hold the Higher National Certificate in his field of technology, and the general technician should have an Ordinary National Certificate or a technician's final certificate of the City and Guilds of London Institute.

The position in reality is not so straightforward. A Ministry of Labour survey of technicians in 1960 showed that 14% held the Higher National Diploma, the Higher National Certificate or a qualification of a professional Institution, 13% had the O.N.D. or O.N.C., and 13% held City and Guilds or other certificates, leaving the balance of 60% without any formal qualification. Furthermore it is difficult to draw a clear line between the type of work and level of responsibility of higher technicians and technologists. Industry tends to take the view that a man's talents are used up to the level he is capable of reaching which may be the Board Room. There is also a tendency by firms competing for manpower to offer attractive titles in their advertisements. 'Assistant Engineer' may suggest a higher social if not real status than 'Higher Technician'.

It is hoped that the current studies of the necessary education, training and experience for the technician, particularly the so-called higher technician, which are being undertaken by the Council of Engineering Institutions and the professional institutions, by the National Advisory Council on Education for Industry and Commerce and by the Engineering Industries Training Board, will produce a clear pattern of training, development and status for the technician.

The I.E.R.E. is quite clear in its own views. It subscribes to the common ideal of the C.E.I. constituent institutions that the chartered engineer must be educated to degree level. It also believes that the academic achievement and industrial importance of the H.N.C. holder should receive proper recognition. Whatever identity tag is put on him, he should be associated in the same institution with the chartered engineer, in order that his knowledge and responsibility should receive appropriate acknowledgment and that he may share fully in the 'learned society' activities of the professional institution. To achieve this, the Institution intends to seek approval for an amendment to its Bye-Laws to allow the higher technician who holds an H.N.C. in Electrical/Electronic Engineering or its equivalent to be admitted to non-corporate membership in the class of Associate. A. M. R.

---

† See, for example, 'Better Opportunities in Technical Education', Cmd. 1254. (H.M.S.O., London, January 1961.)

INSTITUTION NOTICES

Conference on R.F. Measurements and Standards

The programme arrangements for this joint conference, to be held at the National Physical Laboratory, Teddington, Middlesex, from Tuesday, 14th to Thursday, 16th November 1967 were published in the September issue of *The Radio and Electronic Engineer*, together with synopses of the papers. The majority of the papers will be available in preprint form at the Conference; the complete Conference Proceedings will be published in early 1968, price £5.

Requests for registration forms for the Conference may be made using the form at the end of this issue, or by telephone (01-580 8443, Extension 11). The charge for registration (which includes the preprint volume and lunch and refreshments on the three days of the Conference), is £10 for members of the sponsoring institutions (I.E.R.E. and I.E.E.) and £13 for those who are not members of either institution.

The Institution's 1968 Convention

Next year, in the first week of July, the Institution will hold a Summer Convention at Cambridge University, which was the venue of three previous Institution Conventions, in 1951, 1957 and 1959. The technical sessions of the Convention will take place in the lecture theatres of the Cavendish Laboratory, and there will be residential accommodation in King's College.

The form of the Convention will differ from the more specialized Conferences held by the Institution in recent years, which have usually dealt with a single theme in some depth. Instead, the Convention will cover new developments over a wide field of radio and electronic engineering, probably by devoting one or two days to each individual subject. The full range of subjects to be covered has not yet finally been decided, but further information will be published shortly.

The Convention programme will also include the presentation of the Sixth Clerk Maxwell Memorial Lecture.

Joint Conferences in 1968

The Institution is co-sponsor in the **International Television Conference**, to be held at the London headquarters of the Institution of Electrical Engineers from 9th to 13th September 1968. The other sponsors are the I.E.E. Electronics Division, the I.E.E.E. (United Kingdom and Republic of Ireland Section) the Royal Television Society and the Society of Motion Picture and Television Engineers. The I.E.R.E. representatives on the Joint Organizing Committee are A. J. Bennett, A. A. Kay, P. L. Mothersole and R. S. Roberts (Members).

The Conference will cover all engineering aspects of television, and in particular: systems; studio and originating equipment; recording; signal distribution, including wire relay; transmitters and transposers; aerials; receivers; propagation.

The Organizing Committee invites offers of papers of up to 1500 words. Synopses of approximately 250 words should be sent to the Conference Department, Institution of Electrical Engineers, Savoy Place, London, W.C.2, by 5th January 1968. Complete manuscripts will be required by 8th June.

The Institution is co-operating with the I.E.E. Electronics Division and the I.E.E.E. (U.K. and Republic of Ireland Section) in another Joint Conference next year. This is the Conference on **Tropospheric Wave Propagation**, which will take place at the I.E.E. from 30th September to 2nd October. The I.E.R.E. representative on the Joint Organizing Committee is L. W. Barclay (Member).

The aim of the Conference is to survey knowledge gained in tropospheric wave propagation since the Symposium on the same subject held in 1958, including the effect of the lower atmosphere on space communications. There will be sessions on: beyond-the-horizon propagation, including tropospheric scatter; v.h.f. and u.h.f. broadcasting; line-of-sight (including space) links and special problems of centimetre, millimetre and sub-millimetre waves; thermal noise and interference problems; navigational aids; optical communications.

Offers of papers of approximately 2000 words are invited and intending contributors are asked to submit synopses of up to 250 words to the Conference Department, I.E.E., Savoy Place, London, W.C.2, before 12th January 1968. Completed manuscripts should be sent in by 15th June.

'The PAL Colour Television System'

In view of the considerable interest in the PAL colour television system which has been adopted for television broadcasting in colour in Great Britain (as well as in many European and other countries), the Institution has reprinted from the March 1967 issue of *The Radio and Electronic Engineer* the above paper by B. J. Rogers. It will be recalled that this paper formed the basis of a lecture at the Joint Easter School on PAL Colour Television held at Nottingham in April.

Copies of this reprint may be obtained from the Publications Department, I.E.R.E., 9 Bedford Square, London, W.C.1, price 5s. each. A special rate, depending on the number required, can be quoted for orders of 24 or more copies by educational organizations.

U.D.C. 621.37/9 : 004.6

Design for Maintainability with particular reference to Domestic Radio and Television Receivers

By

D. W. HEIGHTMAN,
C.Eng,, M.I.E.R.E.†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.Prod.E.–I.E.E. Conference on 'The Integration of Design and Production in the Electronics Industry', held at the University of Nottingham on 10th to 13th July 1967.

Summary: The main theme of this paper is that of designing good maintainability into a product from the viewpoints of (a) basic component reliability, (b) readily accessible test/monitoring points and (c) ready access and speedy changing of components or modules in cases of failure. The coming of colour television, with its greater complexity, emphasizes the need for attention to these points.

1. Introduction

Those of us who have occasionally endeavoured to gain access to certain troublesome parts of a motor car will know that Design for Production does not necessarily mean Design for Maintainability! The same comment is true of electronic equipment, and particularly domestic radio and television equipment, with which the author is concerned.

The ideal products of design would require no maintenance! Few, if any, electronic engineering product designs have reached that happy state. The subject matter of this paper is not claimed to be new but the lessons of past experience are frequently overlooked in new product designs. If the paper does no more than produce a check-list to which designers refer before passing designs to production it will have achieved some purpose.

Having accepted that some maintenance will be necessary, the aim should be to design the product so that with *minimum time and skill* and with a *minimum of inconvenience* to the user:

(1) Causes of failure or sub-normal performance can rapidly be diagnosed.

(2) Corrective action can be taken.

In other words, good design ensures *rapid* and effective repair.

For various reasons which need not be discussed here, maintenance or service technicians invariably seem to be in short supply, therefore attention in the design stages to the main points made above will also make service possible with less skilled staff.

It should also be borne in mind that it is not only in subsequent field service that good maintainability is required. Prior to leaving the factory the product will proceed through inspection and test functions and as a result some defects requiring 'trouble-shooting' will occur. This trouble-shooting and correction will be more efficiently carried out in a product with good maintainability 'built-in'.

2. Reasons for Maintenance

What are the main reasons that allowance in product design must be made for subsequent maintenance? They are generally:

(1) Economic: The initial cost of the product is sufficiently high that it cannot be considered 'expendable' when it fails and economic factors dictate that it should be maintainable.

(2) Component life: Certain component parts are expendable, i.e. have a definite (or indefinite) working life and need replacement.

(3) Mechanical wear: Some component parts, generally of types involving mechanical movement, require periodical treatment, i.e. cleaning, lubrication, etc.

(4) Environment: The products are sometimes submitted to environmental conditions for which they were not designed, e.g. mechanical shock, excessive temperature or humidity, corrosive or polluted atmosphere.

(5) Cost-reliability factors (components): Almost inevitably there will be some compromises in design between reliability and cost (including cost of developmental period). Dependent on the particular type of product, there is generally an acceptable degree of reliability, against random failure, to give an economic balance between cost of product and cost of subsequent maintenance.

(6) Consequential damage: Generally as a result of (4) or (5) above, in addition to component failure, further consequential self-generated failures or damage by excessive heat, fire, fumes, etc., can follow.

(7) Design weaknesses: No design team is infallible and certain points are overlooked which call for subsequent modifications.

---

† Radio Rentals Limited, Wigmore Street, London, W.1.

3. Field Problems

Attempted maintenance under the above headings very often produces the following main grumbles by servicemen:

(a) Lack of product technical information

There is insufficient (or no) technical and general service information available. In practice it is not known where a suspected faulty component is located in a maze of other parts, and when eventually found its operating conditions are also not known.

(b) Inaccessibility

Having located a faulty component or sub-assembly, or one requiring maintenance, it is difficult and time-involving to remove and replace (i) because the designer forgot to make it accessible, or (ii) the fixings were not designed for ready removal and replacement, special tools being needed to gain access which are seldom available.

(c) Insufficient test points

Insufficient circuit test or monitoring points are available to check performance at various parts of the circuit.

(d) Inadequate safety

Safety precautions are overlooked in the design, i.e. in carrying out internal adjustments or fault diagnosis, it is difficult to avoid getting electrical shocks from exposed tags and components.

(e) Non-standard parts

Non-standard parts are used more by whim of the designer than for good common-sense reasons.

(f) Difficult pre-set adjustments

Pre-set adjustments are difficult to make in setting up equipment.

(g) Modifications

Due to oversight in the design stages various modifications have to be dealt with in the field.

4. Design Requirements

In designing for good maintainability, which will in turn ensure good design for production, the designer should ensure:

(1) Adequate availability of performance and maintenance information.
(2) Adequate monitoring (or test) points.
(3) Good component, cable, connector and control identification.
(4) Good access to all components or sub-assemblies for functional checking, maintenance and/or removal.
(5) Readily removable standard fixings and interconnections which can be re-used.
(6) Reasonable internal precautions against electric shock or heat (as well as compliance with standards of safety externally, including X-rays).
(7) Use of standard, well-tried, parts wherever possible and variety reduction.
(8) Ease of making pre-set or setting up adjustments.
(9) 'Fail safe' arrangements which avoid consequential damage to other parts when one component fails.
(10) Easily handled and mechanically robust chassis or similar main mounting structures.
(11) An adequate period of pre-production field test and maintainability meetings with service organization during design stages.

The points made in this paper so far are applicable to almost all or any electronic or mechanical product. It is now proposed to review in greater detail the application of these good 'maintainability' techniques to domestic radio and television receivers with which the writer has been mainly concerned.

5. Design of the Main Structure: Cabinet, Chassis, etc.

Functional and aesthetic considerations as well as the size of the major components will decide the external shape, styling and dimensions of the product, e.g. in a television receiver the mechanical design is largely dictated by the shape and size of the cathode-ray tube and to a lesser degree by the loudspeaker and main controls. However, good maintainability can also be designed into the product even in considering the general form. Included under the heading of 'maintainability' should be the requirement that the product will withstand reasonable transport hazards such as vibration, mechanical shock, etc. The cabinet should not have vulnerable projections (control knobs, decor features, plastic mouldings, etc.) which will suffer in transit unless very careful and expensive precautions are taken in the packing. Dependent on its volume the structure should be sufficiently strong to withstand stacking (in cartons) of reasonable numbers of the product in stores or vans.

Good designs, before being passed to production, will have been subjected to drop and vibration tests in their normal packing to ensure adequate robustness.

Generally, the majority of components and sub-assemblies will be mounted on a main chassis or framework which can be inserted or withdrawn from the cabinet for production or maintenance purposes. Both from the viewpoint of handling in production and also subsequent maintenance, considerable thought needs to be put into the design of the chassis. In recent years television designers have become conscious of the need to make partial withdrawal, or complete removal, of the chassis from the cabinet for service purposes, a straightforward process.

It is generally possible to make most of the components on the chassis accessible by using a hinging arrangement for the chassis mounting. Such an arrangement saves considerable time in field maintenance as well as avoiding awkward lifting, handling, etc.

The coming of integrated circuits and almost universal use of printed board assemblies has meant that receivers and other equipment are becoming more compact. However, the basic conception of the main chassis or framework as the mounting and combining base for the various sub-assemblies looks as though it will last for a long time. Maintenance or fault repair will be done increasingly on a *sub-assembly exchange basis* rather than by the finding and changing of individual faulty components. Obviously this trend calls for considerable study of the best mechanical and electrical arrangements for mounting and interconnection of the individual modules, sub-assemblies, etc. Plug-socket arrangements almost become a necessity for maintenance and test purposes because of the large number of interconnections between units.

Whatever the exact form taken by the main chassis, good maintainability will require that it can be withdrawn from the housing cabinet by removal of as few fixing devices as possible (i.e. screws, 'speed-fixes' or similar devices). The fixings should be removable with normally available standard tools and should be accessible. The same comments apply to detachable backs or other detachable sections of the cabinet structure. If there are a dozen screws fixing a television receiver-back on its cabinet the serviceman will probably only replace four!

Main components, e.g. cathode-ray tube, scan coil assemblies and other tube neck components, the loudspeaker and the tuner unit, are preferably fixed in the cabinet *separately* from the main chassis. With this arrangement chassis maintenance or exchange can be carried out without the necessity for handling these bulkier items. Such an arrangement, however, also calls for plug-socket interconnection between the chassis and the main separate components. It is also desirable that should one or other of the separate main components or sub-assemblies require replacement, it can be removed and replaced without the necessity for taking the chassis and other parts out of the cabinet. Apart from fixings, removal of the chassis or other main components generally involves withdrawal of front control knobs and similar items. Careful attention to design aspects of these controls or mountings will often show that they can be so shaped that in removal of the main chassis the knobs pass *through* the cabinet front *without* actual removal, i.e. are permanent parts of the chassis assembly.

Once out of the cabinet, the main chassis assembly should on its own be of a general form and strength which makes handling both on the production line and in the field a straightforward procedure:

(i) there should be no awkward projections,
(ii) it should stand without special support jigs,
(iii) preferably it should be rotatable and stand on its various surfaces without mechanical damage so that access can be gained to all parts,
(iv) the mechanical strength should be adequate without relying on the cabinet for support.

6. Running Temperatures and Ventilation

Many field failures result from oversight in the design stages in not providing adequate ventilation or, worse still, in checking that component running temperatures are within limits. Alternatively, certain components, notably resistors or valves, run hot and are placed too close or inappropriately in relation to other components which can thus be damaged due to excessive heat. No new product design should be passed for production until it has been tested for running periods in excess of what would be met in practice and in ambient temperatures 10% above the highest likely to be met in the field.

Where the receiver relies on air-flow through louvers, slots or similar openings in the cabinet for cooling, it should be made quite clear to the user, or installer, that the set should not be placed so that the air-flow would be restricted.

One of the maintenance problems related to air-flow, particularly when the product is used in industrially-polluted atmospheres, is caused by the fine dust and other deposits which gradually build up within the unit over a period of use. Such deposits can have bad effects on components running at extra-high voltages, leading eventually to insulation breakdown. Switches with exposed contact surfaces, such as the rotary wafer type and those used in tuners also suffer badly in this respect, particularly in sulphurous atmospheres resulting from coal fires and similar pollution. Most of these switches rely on silver plating for protection of the contact surfaces, but in such conditions high contact resistance develops after a period of use and the switches become unsatisfactory, particularly in u.h.f. oscillator and similar circuits.

Attention in the design stages to diverting air-flow away from such components and, where necessary, the use of easily-removable dust-covers, ensures good maintainability. Where cost permits, the use of gold instead of silver plating on critical switch contact surfaces will give greater reliability in polluted atmospheres. Switches should in any case be placed where ready access for maintenance can be obtained.

7. Consequential Damage and 'Fail-Safe' Design

Most consequential damage seen in electronic equipment in the field results from excess of heat, in some cases causing fire. Generally one component fails in a short-circuit condition and another related component overheats due to the resulting current flow; thus frequently a by-pass capacitor will short-circuit and the resistor feeding power to that part of the circuit as a result is overloaded. Depending on the placing of such resistors and similar components, the heat generated will either be dissipated without further damage to adjacent components, etc., or a serious conflagration may result. It is not unusual to see a whole printed circuit board containing many components so badly burnt as to be not worth repair. A little thought and suitable tests in the design stages will avoid such catastrophic happenings. Some aspects of wax and similar insulating impregnants and the bad effects of overheating unless suitable precautions are taken in the design stages have been dealt with in a previous paper.†

8. Standard Parts—Variety Reduction

One of the major problems in providing an efficient maintenance service is that of obtaining and stocking the required wide range of spare parts—components, valves, transistors, etc. In seeking to ensure good maintainability the designer should use every effort to reduce to a minimum the number of different parts used in a product, and should select industry standard parts wherever possible.

The use of a little ingenuity will show that it is often possible to use the same type of component, transistor, etc., in various stages in a circuit and thus reduce the variety of parts within a particular model. There is no virtue at all in 'being different just for the sake of being different'. The size of the inventory of spares which a large service organization needs to maintain is evidence of insufficient attention to this important point in the past.

9. Component Ratings

Much maintenance work results from the failure of components—resistors, capacitors, valves particularly—which are run too close to their maximum rating. Generally for cost reasons, but sometimes with size considerations, optimistically-rated components are used in a product. These components invariably have a higher failure rate than those which are run well within rating. During the development stage the designer should check that all components are being run within their limits, also that normal tolerance spreads will not produce overrunning or other detrimental effects. This condition should apply both with or without the application of normal signal and synchronization waveforms, etc. Pre-set controls, mostly resistors or potentiometers, should also be checked to ensure that where only sections of their tracks are in use these sections are not overrun.

Many aspects of component reliability were covered in a paper published in 1961.‡ Since that paper was written transistors have come into more general use in radio and television receivers and have brought about an improvement in reliability. In television receivers particularly, however, the ease with which transistors may be irreparably damaged by accidently-created voltage surges has made it necessary to take precautions to avoid this type of failure. Such voltages particularly occur around receiver circuits when the picture tube momentarily arcs over. (The colour picture tubes with e.h.t. of the order of 25 kV are particularly prone to this defect.) Suitable protective spark gaps and similar precautions should be provided at the appropriate points to avoid these bad effects.

10. Colour Television

The introduction of the first mass-produced colour television receivers in 1967 can be expected to bring special maintenance problems unless great care has been applied to the foregoing points in the design stages. The greater number of components used will mean that for overall reliability similar to that of current monochrome receivers, higher individual component reliability will have to be obtained. The critical points will undoubtedly be those peculiar to the colour receiver:

(a) the special scanning, convergence and other scan correction components and circuits,

(b) the e.h.t. supplies of the order of 25 kV,

(c) colour drive-circuits and pre-sets related to the shadow-mask tube

and

(d) while dual-standard conditions apply (i.e. 405/625 lines), there will also be potential 'built-in' unreliability resulting from the multiplicity of ganged switches and circuit compromises necessary in the dual-standard colour set.

The various setting-up conditions for the tube are, of course, critical and the designer should check that these do not change over running periods, due, for instance, to thermal drift variations both external and internal to the tube.

Manuscript received by the Institution on 26th May 1967. (Paper No. 1143.)

---

† D. W. Heightman, 'Printed circuit reliability and flammability', *J. Brit. Instn Radio Engrs*, **20**, p. 281, April 1960.

‡ D. W. Heightman, 'Component and valve reliability in domestic radio and television receivers', *J. Brit. Instn Radio Engrs*, **21**, No. 5, p. 401, May 1961.

U.D.C. 621.396.946

A Review of Methods of Measuring System Parameters in Time-variant Space Communication Channels

By

T. KALISZEWSKI,
M.A., B.S.†

Summary: The objective and methods of measurement on linear, time-variant communication channels are discussed with a view of assessing their applicability for the diagnostics of certain unconventional channels of interest to space communication. These may include channels formed during the early, powered phase of the space vehicle flight and during its re-entry into the Earth's atmosphere. No attempt is made here to elucidate the physical phenomena which are responsible for the formation and properties of such channels and the problem of their measurement is approached from the input–output point of view. It is assumed that the channels in question have been rendered semi-transparent through a suitable choice of carrier frequencies, but that the transmitted waveforms are impressed with identifiable channel characteristics, such as the multi-path and Doppler spread. It is concluded that due to the uniqueness and complexity of such channels only the most rudimentary measuring methods, leading to the estimates of gross channel parameters, can be contemplated at this time. The need for further research is indicated.

1. Introduction

Many of the communication channels in use during an aerospace mission are known to exhibit certain characteristics which suggest that the channels are not only noisy but are also *dispersive*, in both time and frequency. Time dispersion, referred to as *multi-path spread*, manifests itself when a narrow pulse at the channel input results in a multiplicity of pulses at the output, spread over a time L, or in a continuum of comparable length. Frequency dispersion, known also as a *Doppler spread*, is observed when a monochromatic input is converted by the channel into a narrow band of frequencies of width B centred at the frequency of the input signal. Both of these effects are suggestive of the random character of the channel and both contribute to the distortion of the transmitted waveforms. These effects are especially deleterious in the case of digital communication systems, where they set a definite limit to the achievable reliability regardless of the signal/noise ratio.[1]

An important feature of most dispersive channels is that they are time-variant, i.e. are not invariant under translations in time. Thus, the system functions such as the impulse response function or its Fourier transform, the frequency response function, depend on time. Generally, these functions are sample functions of a random process, which may or may not be stationary, even over the period of measurement. In many channels the superposition principle, which makes the characterization of the channel by its impulse response function possible, may not obtain. However, neither of these properties can be readily ascertained in new and unexplored channels. As a rule, both the linearity and the stationary nature are assumed if for no other reason than to facilitate a discussion of what is otherwise an exceedingly complex matter.

The characterization of a time-variant channel in terms of its system functions renders it accessible to input–output measurements similar, in many ways, to the customary measurements on time-invariant filters. There are three types of measurements possible: (i) measurements of the instantaneous value of the impulse response function; (ii) measurements of the statistical averages, such as the mean values and the correlation functions; (iii) measurements of the r.m.s. widths of the impulse and frequency responses of the channel. These measurements vary as to their objectives and complexity. Some are more suitable for diagnostic studies of the channels (example: Moon scatter, ionosphere), others are adequate for gross estimation of the parameters of importance to the design of communication systems (spread factor, coherence bandwidth, etc.).

It is the objective of this paper to review, briefly, the principal methods of measurement of time-variant channels and to ascertain, if possible, which of these is suitable or desirable in a programme involving such composite and transient channels as are encountered in some aerospace applications.

2. Physical Situation in Some Channels of Interest

Apart from the detailed characterization through its system functions, the channel can also be characterized by the multi-path and Doppler spread, in

---

† Raytheon Company, Spencer Laboratory, Burlington, Massachusetts.

addition to the signal/noise ratio. These parameters constitute what may be termed the basic and minimum channel data needed to ascertain the performance of digital communication systems. Before reviewing the methods for securing these data it may be instructive to note what is already known about such parameters in certain, better documented channels and to observe that, occasionally, the gross parameters of the channel can be obtained from physical arguments. Thus, for instance, for the Moon scatter channel, both multi-path and Doppler spread have long been known from astronomical observations and further statistical information (frequency correlation functions) has recently been secured by, among others, Hagfors.[2] Similar information has been available for other celestial scatterers (Venus, Mars), as well as for the more customary communication channels, such as the ionosphere and, recently, for the orbiting dipole and chaff channels. To get an idea as to the magnitudes involved, Table 1 shows the estimates of B, L and k (the spread factor) for several channels of interest. It can be seen that only one channel (incoherent, electron scatter), appears to be overspread, i.e. $k \geq 1$. However, other channels can also become overspread, since the Doppler spread is a function of frequency and in some media (notably, turbulent plasmas) the multi-path spread will be a function of frequency as well. This is not difficult to see, as the following brief consideration of the multi-path introduced by the turbulent rocket exhaust shows (see Fig. 1). Multi-path in this case originates with the turbulent plasma inhomogeneities and with the internal and external boundaries of the rocket plumes. Both the boundaries and the individual 'blobs' scatter, or reflect, because the plasma frequencies associated with the local plasma densities are either higher than or comparable to the carrier frequency of the transmission. Clearly, the effectiveness of these scatterers as a multi-path generating mechanism depends strongly on the frequency. This is not unlike the case of scatterings from the ionosphere, where it can be shown that the effective power scattered by the turbulent inhomogeneities depends not only on the density gradients and scattering angles, but also is *inversely* proportional to the carrier frequency.[3] Thus, by proper choice of the latter parameter, the effectiveness of the multi-path generating mechanism can be minimized and, perhaps, even eliminated.

Table 1

Estimates of L, B and k for some channels

Channel	L (seconds)	B (s^{-1})	$k = BL$
Ionosphere (h.f.)	5×10^{-3}	3	$1{\cdot}5 \times 10^{-2}$
Ionoscatter (v.h.f.)	10^{-4}	10	10^{-3}
Troposcatter (u.h.f.)	10^{-6}	10	10^{-5}
Incoherent electron scatter (u.h.f.)	10^{-4}	10^4	1
Orbiting dipoles (s.h.f.)	10^{-4}	10^3	10^{-1}
Chaff ($f = ?$)	5×10^{-6}	10^2	5×10^{-4}
Moon (u.h.f.)	10^{-2}	10	10^{-1}
Telemetry (powered flight)	?	?	?
Telemetry (re-entry)	?	?	?

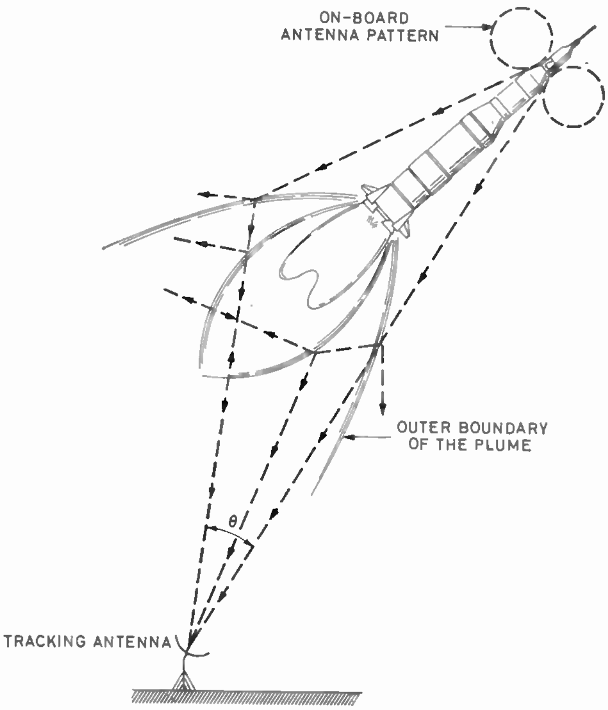


Fig. 1. Origins of multi-path spread in a channel incorporating rocket exhaust.

Unfortunately, estimates of the type quoted in Table 1 cannot be readily obtained for most of the channels encountered in aerospace missions. There is simply not enough information even to ascertain the opacity regions of the rocket plumes, not to speak of the density gradients, size or motion of inhomogeneities, etc. Recently, some attempts have been made to estimate the upper bound on the multi-path spread of such channels, by using the information obtained from optical tracking and by relating it to the simple, forward-scatter relation between the angular spread of transmission, its frequency and the size of channel inhomogeneities.[4] Needless to say, the validity of these estimates has not as yet been verified by experiment. No serious attempts have been made to date to construct, analytically, either the frequency correlation function or, what amounts to the same, the scattering functions for such channels and no spectacular developments can be expected in this area in the immediate future. Progress in that direction can be made, however, by instituting a concerted and purposeful measuring programme.

3. Objectives and Methods of Channel Measurements

An electromagnetic signal transmitted through or reflected from a randomly inhomogeneous, time-variant channel is characterized by certain parameters, which vary as a result of signal-channel interaction.[5] Thus, the amplitude, phase, frequency and polarization of the received signal are potential carriers of information concerning the nature of the channel and the extent of its interaction with the signal. In principle, it should be possible to unravel this information through the comparison of the received signal with its unaltered copy. Signal delay (multi-path), frequency dispersion (Doppler spread), depolarization and amplitude changes can be measured this way either as instantaneous or average quantities. Frequently, it is possible to relate the measured quantities to the physical features of the channel, such as the roughness of the reflecting surface, the scale of density inhomogeneities, density gradients, etc. Examples of such extensive diagnostic measurements can be found in both radio astronomy and communication.

The methods employed for channel measurements will differ as a rule with the objectives. In communication, it is rather customary to depend on transmission, even though the well-known technique of ionospheric sounding is better classified as reflection (or scattering). Other techniques used include interferometers and polarimeters. In either case we are faced with the problem of comparing the signals at the input–output terminals and defining those functions and parameters which can meaningfully characterize the 'black-box' (i.e. the channel) in between. To see how this comparison can be effected let us review some of the definitions, models and relations used in connection with linear, time-variant channels.

3.1 *Some Models, Definitions and Relations*

Consider the block diagram of a general communication channel as shown in Fig. 2. Here, the additive noise $n(t)$ is usually thought to originate in the receiver, but may also represent an interference originating in the channel. The filter, A, can be specified by two system functions, $h(\tau, t)$ and its Fourier transform, $H(f, t)$; here we define[6]:

$h(\tau, t)$ = response of the filter, measured at t to a unit impulse input at $t-\tau$,

$$H(f, t) = \mathscr{F}[h(\tau, t)] = \int_{-\infty}^{+\infty} h(\tau, t)\, \mathrm{e}^{-\mathrm{j}2\pi f\tau}\, \mathrm{d}\tau \qquad \ldots\ldots(1)$$

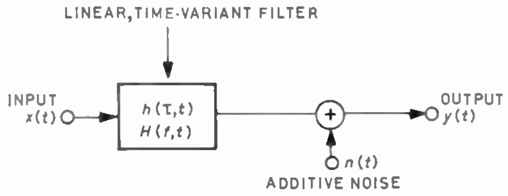


Fig. 2. Block diagram of a general communication channel.

Functions $h(\tau, t)$ and $H(f, t)$ are not the only ones possible (see Kailath[7]) but have the virtue of being directly related to each other by a Fourier transform; the realizability condition for $h(\tau, t)$, as defined here, is simply, $h(\tau, t) = 0$ for $\tau < 0$. The meaning of variables, τ, t, and f is as follows:

t variable corresponding to the instant of measurement at the output,

τ variable corresponding to the age of input,

f variable corresponding to the input frequencies.

With the impulse response function known it is possible, by the super-position principle, to obtain the filter output for an arbitrary input, i.e.

$$y(t) = \int_{-\infty}^{+\infty} x(t-\tau) h(\tau, t)\, \mathrm{d}\tau \qquad \ldots\ldots(2)$$

The convolution integral (2) suggests a representation of a linear, time-variant channel by a densely tapped delay line as shown in Fig. 3. The tap function at the delay τ' is $h(\tau', t) = h_{\tau'}(t)$; the length of the line L corresponds to the maximum spreading of an impulse input by the channel and the bandwidth of the taps is B, the maximum frequency spreading of a sinusoidal input. Justification and details of this modelling of the channel are given by Kailath.[7]

Additional system functions appropriate to the characterization of the statistical properties of the channel have been introduced by Kailath,[8] Hagfors,[2] and Price and Green.[9] They are the tap-gain correlation function $R(\tau, \Delta t)$, the two-frequency correlation function $\mathscr{R}(\Delta f, \Delta t)$ and the scattering function $\sigma(f, \tau)$.

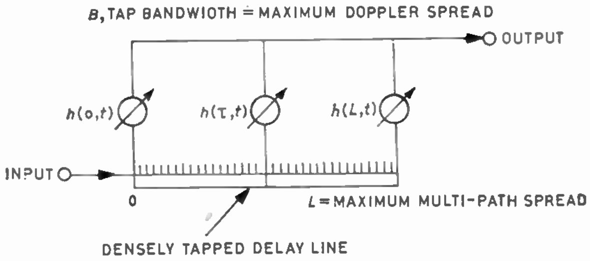


Fig. 3. A delay line model for a linear, time-variant channel.

Their definitions follow:

$$\mathscr{R}(\Delta f, \Delta t) = \tfrac{1}{2}\overline{H^*(-\Delta f/2, t)H(\Delta f/2, t+\Delta t)} \quad \ldots\ldots(3)\dagger$$

$$R(\tau, \Delta t) = \int_{-\infty}^{+\infty} \mathscr{R}(\Delta f, \Delta t)\, e^{j2\pi\Delta f\tau}\, d(\Delta f) \quad \ldots\ldots(4)$$

$$\sigma(\tau, f) = \int_{-\infty}^{+\infty} R(\tau, \Delta t)\, e^{-j2\pi f\Delta t}\, d(\Delta t) \quad \ldots\ldots(5)$$

All of these functions are defined here on the assumption that (i) the taps are uncorrelated, and (ii) random processes controlling the tap functions are stationary, i.e. the tap-gain correlation function is independent of the variable t.

Before discussing further these functions in the context of their measurements, it may be instructive to dwell briefly on some of their properties and to introduce certain specialized terms. Since all of them are interrelated, it will be sufficient to discuss only one, say, the two-frequency correlation function. We will note in this connection that, in general, both $\mathscr{R}(\Delta f, \Delta t)$ and $R(\tau, \Delta t)$ are complex while the scattering function $\sigma(\tau, f)$ is a real quantity related, under proper receiving arrangements, to the power scattered by the channel (target).

Consider first the correlation function $\mathscr{R}(0, \Delta t)$, which is simply the autocorrelation of the channel response to a sinusoid of frequency f_0. We define the *coherence time of the channel* T_c as the interval in Δt over which $\mathscr{R}(0, \Delta t) > 0$. Now, consider the spaced-frequency correlation function $\mathscr{R}(\Delta f, 0)$; here we define *the coherence bandwidth of the channel* F_c, as the frequency separation at which $\mathscr{R}(\Delta f, 0)$ differs significantly from $\mathscr{R}(0,0)$ (by, say, 3 decibels, etc.). Despite appearances, these two parameters *are not related to each other!* Thus, it can be shown that the inverse of T_c is approximately equal to the Doppler spread, i.e. $1/T_c \propto B$ and the inverse of F_c is approximately equal to the multi-path spread, i.e. $1/F_c \propto L$. That this is so can be seen from the alternative definition of $R(\tau, \Delta t)$,

$$R(\tau, \Delta t) = \overline{h(\tau, t)h(\tau, t+\Delta t)} \quad \ldots\ldots(6)$$

Clearly, the duration in τ over which $R(\tau, 0)$ is non-zero is effectively the same as the duration of the impulse response $h(\tau, t)$ or L. Now, since $R(\tau, \Delta t)$ is the Fourier transform of $\mathscr{R}(\Delta f, \Delta t)$, L will have the order of magnitude of $1/F_c$. The relation of T_c to the Doppler spread, B, can be established in a similar manner using the definition of $\mathscr{R}(\Delta f, \Delta t)$ (eqn. (3)) and examining the frequency response of the channel via the Fourier transform of $H(f, t)$, also known as the bi-frequency function (see Kailath[7]).

3.2 *A Condition for Unambiguous Measurements*

The spread-factor $k = BL$, introduced in Section 3, plays an important role in channel measurements, inasmuch as the condition $BL < 1$ (for underspread channels) is both necessary and sufficient[8] for unambiguous measurement of instantaneous quantities such as $h(\tau, t)$ when no *a priori* knowledge exists about the channel.

As pointed out by Kailath,[8] the densely tapped delay line model of a linear, time-variant channel can be represented by a delay line of length L with taps $1/2W_i$ apart, if the input signal is constrained to a bandwidth $(-W_i, W_i)$. Now, since each tap is band-limited to $(-W, W)$ where $B = 2W$, there are $(2LW_i + 1)$ such taps, each of which requires $(2TW + 1)$ values for its specification (here T is the period of the tap function, that is, the duration of $h(\tau, t) > 0$ for a fixed delay τ). The output waveform is limited to a band of $2(W_i + W)$ and a time interval of $L + T$, thus providing $2(T+L)(W_i + W) + 1$ independent values. For unambiguous measurements it is required that

$$\{2(T+L)(W_i + W) + 1\} \geq (2LW_i + 1)(2TW + 1) \quad \ldots\ldots(7)$$

or

$$L \leq \frac{1}{2W}\left(\frac{1}{1 - 1/2TW_i}\right) \quad \ldots\ldots(8)$$

If we now assume a very wide bandwidth for the input signal, we obtain

$$BL \leq 1 \quad \ldots\ldots(9)$$

The condition (9) assures that, for instance, impulsive signals applied to the channel at intervals $i/2W, i = 0, 1 \ldots$ will result in non-overlapping responses of the channel and consequently, in their unambiguous determination.

The constraint $BL < 1$ appears to be applicable to the instantaneous measurement only. Recent work by Gallager[10] and Hagfors[11] shows that it is of little consequence to the measurement of average quantities.

3.3 *Random Signals versus Deterministic Probing Signals*

In our discussion thus far we have encountered definitions and operations involving time and frequency impulsive input waveforms. These are, however, not the only signals proposed or used in the channel measurements. As we will see later, other waveforms, such as pulse-trains, 'chirp' signals, etc., can also be used. In general, the signals proposed for the channel measurement can be divided into two groups, random (or pseudo-random) and deterministic. Thus, in the fourth-moment method, to be discussed shortly, a white Gaussian noise is used as an input. On the other hand, to measure the scattering function, a deterministic signal consisting of a long duration

---

† This is a complex, low-pass form; for other forms of H, see Gallager.[10]

train of short pulses is used. The criteria used in selecting a proper waveform depend not only on the resolution required of the measurements, in time and frequency, but also on the convenience with which the input signal can be reproduced or stored and delayed at the receiver where it is used for signal processing, generally correlation reception. Other considerations involved in the selection of the input waveform are the duration and time required for its processing; these must be compatible not only with the required resolution but also with the duration of the channel processes, over which the latter can be assumed stationary. In general, there are no firm rules governing the choice of the input waveforms, short of the guidelines implied in the above remarks and experience with a particular channel.

3.4 *Kailath's Method of Measuring the Tap-Gain Correlation Function*

In this method, proposed by Kailath,[8] use is made of a white Gaussian noise and the cross- or autocorrelation functions are obtained from the following fourth cross-moment.

$$I(\tau,\tau',\Delta t) = \overline{\overline{y(t)x(t-\tau)y(t+\Delta t)x(t-\tau'+\Delta t)}} \quad ......(10)$$

where $x(t)$, $y(t)$ are input, output functions, respectively, and the double bar denotes averaging over the input and channel ensemble, the two being assumed independent. After some manipulation, Kailath shows that

$$\begin{aligned} I(\tau,\tau',\Delta t) &= \overline{h(\tau,t)h(\tau',t+\Delta t)}+ \\ &\quad +\overline{A(\tau'-\Delta t,t)A(\tau+\Delta t,t+\Delta t)} \\ &= R(\tau,\tau',t,t+\Delta t)+ \\ &\quad +R(\tau'-\Delta t,\tau+\Delta t,t,t+\Delta t) \quad ......(11) \end{aligned}$$

Here, the first term is the desired cross-correlation function of the taps (τ,τ'), and the second term can be viewed as an interference. An especially simple form is obtained when the taps are uncorrelated and the random processes controlling them are stationary. We have then,

$$I(\tau,\Delta t) = R(\tau,\Delta t) \quad ......(12)$$

and the operation can be implemented either through time or ensemble averaging.

The use of a white Gaussian noise in the fourth-moment method is convenient (from an analytical point of view) but not necessary. Using similar arguments, and a non-white Gaussian noise, Gallager[10] obtained results comparable to eqn. (11) and, in addition, has shown that the error is, for a wideband input noise, impulsive in character at $\Delta t \simeq 0$. As such, it need not interfere with the measurements of $R(\tau,\Delta t)$ for all values of Δt, except at $\Delta t \simeq 0$, where extrapolation procedures can be used to secure its true value.

It has also been established by Gallager that this method is not restricted to underspread channels, even when the taps are correlated; previously, this condition has been established only for the case of uncorrelated taps.

3.5 *Green's Method of Measuring the Scattering Function*

It was shown at the beginning of this Section that the real character of the scattering function $\sigma(t,f)$ and its relation to the geometric and physical features of the channel makes it somewhat of a fundamental quantity. Formally, we could obtain $\sigma(\tau,f)$ by taking a Fourier transform of $R(\tau,\Delta t)$ with respect to Δt. However, a more direct method has been proposed by Green.[9] In that method, an input signal $x(t)$ (still to be further specified) is used in connection with a filter matched to it or with an equivalent correlator. It can be shown that the power out of the matched filter as a function of the time and frequency offsets is proportional to the two-dimensional convolution of the scattering function with the ambiguity function of the input waveform, i.e.

$$P(\tau,f) = \iint\limits_{-\infty}^{+\infty} \sigma(\tau',f')\Psi^2(\tau'-\tau,f-f')\,\mathrm{d}\tau'\,\mathrm{d}f' \quad ......(13)$$

where

$$\Psi^2(\tau,f) = \left| \int\limits_{-\infty}^{+\infty} \chi^*(t)\chi(t+\tau)\,\mathrm{e}^{\mathrm{j}2\pi ft}\,\mathrm{d}t \right|^2 \quad ......(14)$$

and

$$x(t) = \mathrm{Re}[\chi(t)\,\mathrm{e}^{\mathrm{j}2\pi ft}] \quad ......(15)$$

The quantity $\chi(t)$ is called the complex envelope of the signal, the real part of which is the input signal proper; $\chi(t)$ is independent of the carrier frequency.

The significance of eqn. (13) lies in the fact that, by properly choosing the input waveform, it may be possible to obtain an ambiguity function having a sharp peak at the origin and a negligible value outside of the area defined by the rectangle $1/TW$, where T is the duration of the waveform, equal to the reciprocal of the desired frequency resolution and W is the bandwidth of the signal equal to the reciprocal of the desired range resolution. With an impulse-like ambiguity function, the power out of the matched filter will, approximately, be equal to the scattering function $\sigma(\tau,f)$. Thus, the ambiguity function $\Psi(\tau,f)$ can be looked upon as a sort of window through which the scattering function can be viewed. Clearly, the smaller this window the greater the resolution of the scattering function. However, there are some complications since an ideal, impulsive ambiguity function

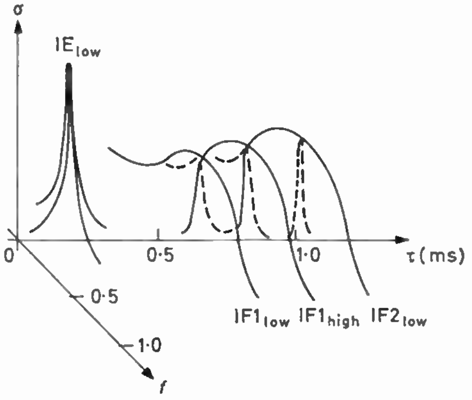


Fig. 4. An example of the scattering function for h.f. ionospheric channel (after Green[12]).

is difficult if not impossible to realize. Most large TW waveforms give rise to secondary peak in the ambiguity function, which will contribute to the output of the matched filter a sort of self-noise. This self-noise will be negligible if the scattering function does not spread outside the area defined by $BL < 1$. For overspread channels, $BL > 1$, a waveform having $TW = 1$ will also assure negligible self-noise, since in that case the ambiguity function is almost entirely concentrated in the central peak. For channels for which $BL \simeq 1$, the self-noise cannot be avoided, but recent work by Gallager[10] indicates that the error introduced by it can be estimated and subtracted from the measurements.

We close this discussion of Green's method by showing examples of the scattering functions for the h.f. ionospheric (Fig. 4)[12] and the u.h.f. Moon scatter channel (Fig. 5), the latter approximated by a uniform, rotating sphere.[9]

3.6 *Hagfors' Method of Measuring the Two-Frequency Correlation Function*

Before discussing the method of measuring $\mathscr{R}(\Delta f \Delta t)$, first proposed by Hagfors,[2] let us rewrite the input–output relation of eqn. (2) and consider it in the frequency domain. If $X(f)$ be the Fourier transform of $x(t)$, the input signal, then eqn. (2) can also be written as

$$y(t) = \int_{-\infty}^{+\infty} X(f)H(f,t)\,\mathrm{e}^{\mathrm{j}2\pi ft}\,\mathrm{d}f \qquad \text{......(16)}$$

Suppose that the input is a sinusoid of frequency f_0 so that $X(f) = \delta(f-f_0)$, where δ is the *frequency* impulse function. As a result of the sifting properties of $\delta(f-f_0)$, eqn. (16) yields,

$$y(t) = H(f_0,t)\,\mathrm{e}^{\mathrm{j}2\pi f_0 t} \qquad \text{......(16a)}$$

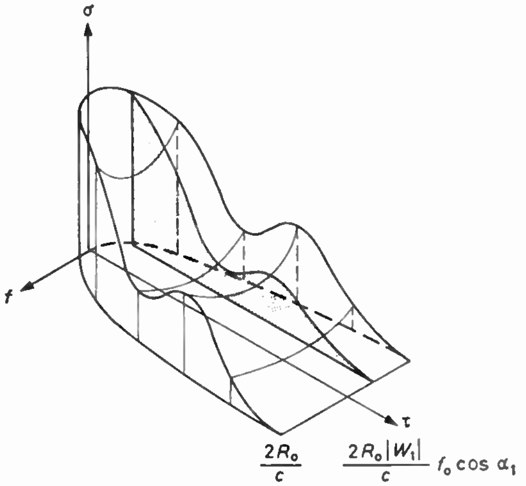


Fig. 5. An example of the scattering function for u.h.f. Moon scatter channel (after Price and Green[9]).

R_0 = radius of the sphere
$|W_1|$ = velocity of rotation
α_1 = tilt angle
c = velocity of light
f_0 = carrier frequency

For the mean-square value of $y(t)$ we have

$$\overline{|y(t)|^2} = \overline{H(f_0,t)H^*(f_0,t)}$$

$$= \left[\overline{H^*\left(f_0-\frac{\Delta f}{2},t\right)H\left(f_0+\frac{\Delta f}{2},t+\Delta t\right)}\right]_{\substack{\Delta f=0\\ \Delta t=0}}$$

$$= [\mathscr{R}(\Delta f,\Delta t)]_{\substack{\Delta f=0\\ \Delta t=0}} \qquad \text{......(17)}$$

Thus, the mean power output from the channel is equal to a two-frequency correlation function evaluated at $\Delta f = 0$, $\Delta t = 0$. Equation (17) contains the simple idea employed in the measurement of $\mathscr{R}(\Delta f,\Delta t)$. We transmit two sinusoids spaced in frequency by Δf and cross-correlate, as a function of Δt, the received waveforms; we then repeat for other values of Δf, to provide for the full coverage of the response. The required frequency and delay excursions Δf and Δt are determined by the maximum Doppler and multi-path spread B and L; that is, we want to measure $\mathscr{R}(\Delta f,\Delta t)$ at spacings Δf as low as $1/L$. However, since each sinusoid is corrupted by a Doppler spread of magnitude B, we must have

$$\frac{1}{L} \simeq \Delta$$

$$\Delta f > B$$

hence

$$\frac{1}{L} > B \quad \text{and} \quad BL < 1 \qquad \text{......(18)}$$

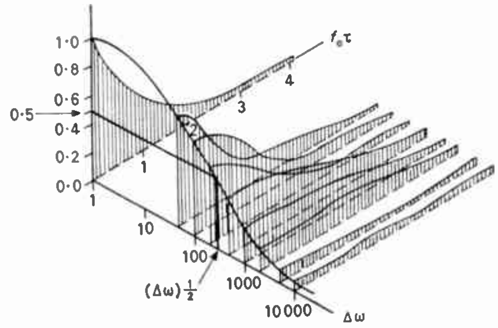


Fig. 6. Spaced-frequency (amplitude) correlation function for u.h.f. Moon scatter channel (after Ingalls[13]).

It would appear from eqn. (18) that the two-frequency measurements are subject to the constraints discussed in Section 3.2. However, recent work by Hagfors[11] indicates again that the self-noise can be eliminated by a fortuitous choice of the integration time, thus making the two-frequency method applicable to both under- and overspread channels. This conclusion, however, applies only to channels which are stationary and in which the taps are uncorrelated. An example of a spaced-frequency (amplitude) correlation function obtained by Ingalls[13] for the Moon scatter channel is shown in Fig. 6.

3.7 *Measurements of the Multi-path and Doppler Spread*

Although a complete determination of the system functions is a desirable and essential step toward the channel characterization, frequently a less ambitious goal may be both sufficient and more realistic. We have pointed out earlier that the spread factor is a parameter of great importance to the design and evaluation of the digital communication system. Consequently, its determination should be the minimal objective in channel measurements. What is involved in such a minimal-objective measuring programme is the estimation of the widths, suitably defined, of either the two-frequency correlation function $\mathscr{R}(\Delta f, \Delta t)$ or of its Fourier transform, the scattering function $\sigma(\tau, f)$. In a recent paper, Bello[14] described a remarkably simple method of measuring the r.m.s. widths of the Doppler and multi-path spread. This method is predicated on the assumption that the channel is essentially stationary and characterized by an uncorrelated scattering; also, on the assumption that the transmission of a sinusoid through the channel results in a narrow-band Gaussian process. The latter restriction can, however, be relaxed by the use of coherent processing. In what follows, only the envelope approach is outlined.

Following Bello, we define the r.m.s. Doppler, $B_{\text{r.m.s.}}$, as follows:

$$B_{\text{r.m.s.}} = 2\sqrt{\frac{\int_{-\infty}^{+\infty} (f-\bar{f})^2 \sigma(f)\,df}{\int_{-\infty}^{+\infty} \sigma(f)\,df}} \qquad \ldots\ldots(19)$$

where, as before

$$\sigma(f) = \int R(0, \Delta t)\, e^{-j2\pi f \Delta t}\, d(\Delta t) \qquad \ldots\ldots(20)$$

and

$$\bar{f} = \frac{\int f\sigma(f)\,df}{\int \sigma(f)\,df} \qquad \ldots\ldots(21)$$

Similarly, we define r.m.s. multi-path spread $L_{\text{r.m.s.}}$ as

$$L_{\text{r.m.s.}} = 2\sqrt{\frac{\int_{-\infty}^{+\infty} (\tau-\bar{\tau})^2 \sigma(\tau)\,d\tau}{\int_{-\infty}^{+\infty} \sigma(\tau)\,d\tau}} \qquad \ldots\ldots(22)$$

To measure $B_{\text{r.m.s.}}$, $L_{\text{r.m.s.}}$, Bello proves[15] that the ratio of the r.m.s. bandwidth of the output from an envelope detector to the r.m.s. bandwidth of a narrow-band Gaussian input is *a constant* depending only on the detector and not on the shape of the input spectrum. Thus, he shows that in order to measure B, it is necessary only to measure the r.m.s. bandwidth of the envelope of a sinusoid transmitted through the channel and to determine the constant for the detector in use. This constant is given by

$$\alpha = \sqrt{\frac{j}{2}}, \quad j = 1, 2 \ldots \qquad \ldots\ldots(23)$$

where j refers to the power law characterizing the detector. Obviously, for a square detector, $\alpha = 1{\cdot}0$ and for a linear detector $\alpha = 0{\cdot}707$.

If now $\sigma_e(f)$ refers to the power spectrum of the received envelope $e_f(t)$ we have

$$B_{\text{r.m.s.}} = \frac{2}{\alpha}\sqrt{\frac{\int_{-\infty}^{+\infty} f^2 \sigma_e(f)\,df}{\int_{-\infty}^{+\infty} \sigma_e(f)\,df}} \qquad \ldots\ldots(24)$$

Also since

$$\overline{e_f^2(t)} = \int \sigma_e(f)\,df \qquad \ldots\ldots(25)$$

and

$$\overline{\left(\frac{de}{dt}\right)^2} = (2\pi)^2 \int_{-\infty}^{+\infty} f^2 \sigma_e(f)\,df \qquad \ldots\ldots(26)$$

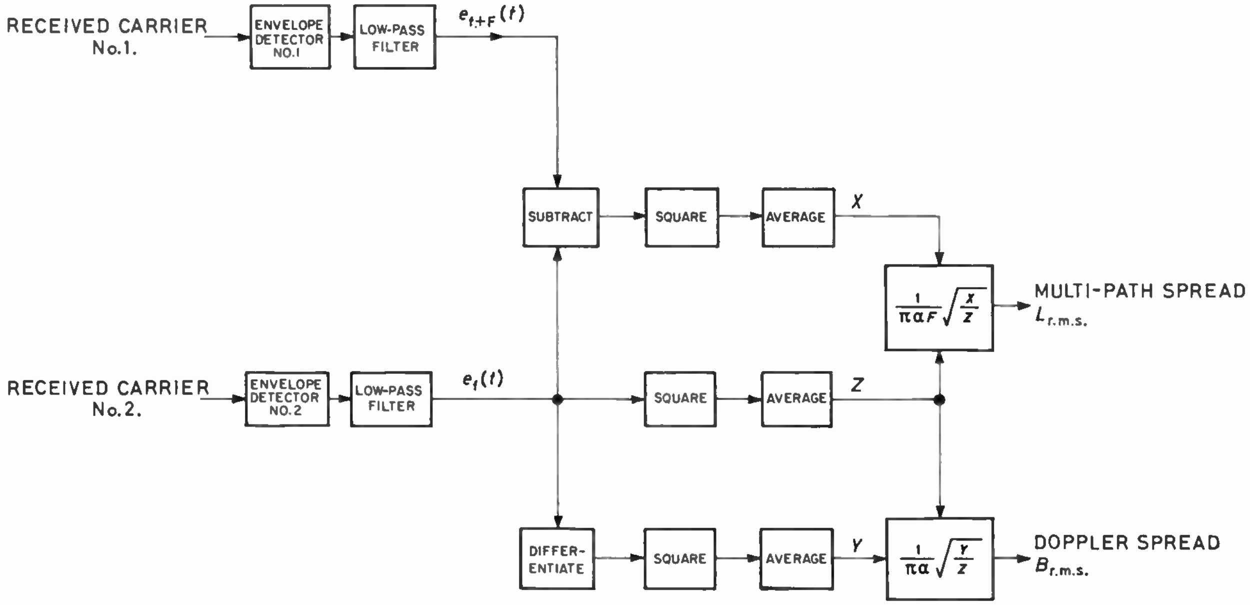


Fig. 7. An implementation of the envelope method of measuring the r.m.s. Doppler and multi-path spread (after Bello[14]).

we can write

$$B_{\text{r.m.s.}} = \frac{1}{\pi\alpha}\sqrt{\frac{\overline{(\mathrm{d}e/\mathrm{d}t)^2}}{\overline{e_f^2(t)}}} \qquad \ldots\ldots(27)$$

Similar expressions are obtained for $L_{\text{r.m.s.}}$:

$$L_{\text{r.m.s.}} = \frac{1}{\pi\alpha}\sqrt{\frac{\overline{(\mathrm{d}e/\mathrm{d}f)^2}}{\overline{e_f(t)^2}}} \qquad \ldots\ldots(28)$$

One possible implementation of the measurements of $B_{\text{r.m.s.}}$ and $L_{\text{r.m.s.}}$, based on the two-frequencies transmission, is shown in Fig. 7. Other forms utilizing the f.m.-probing signal and also the complex envelope measurements are discussed by Bello.[14]

3.8 *Summary of Measuring Methods*

In discussing the various measuring methods attempt has been made to emphasize the philosophy

Table 2

Summary of measuring methods†

Function and parameter	Method of measurements	References and remarks
$R(\tau, \Delta t)$	*Tap-gain Correlation Function:* use a long sample of white Gaussian noise and form the fourth cross moment between the delayed versions of input and output functions; for a stationary channel with uncorrelated taps, the fourth cross moment is equal, identically, to $R(\tau, \Delta t)$.	Kailath[8]: method applicable to both under and over spread channels; $R(\tau, \Delta t)$—generally a complex function.
$\mathscr{R}(\Delta f, \Delta t)$	*Two-frequency Correlation Function:* transmit two sinusoidal waves of known frequency spacing; receive each input separately, insert a delay in one and measure average amplitude and phase of the difference tone; repeat for other values of Δf.	Hagfors[2, 11]: self-noise free measurement constrained by $BL < 1$; extension to overspread channels possible; $\mathscr{R}(\Delta f, \Delta t)$—complex function.
$\sigma(f, \tau)$	*Scattering function:* transmit a waveform $x(t)$ of duration T and bandwidth W such that $TW > 1$; receive through a filter matched to $x(t)$, by observing the average envelope of the output, squared, at various delays τ and frequency offsets f.	Price and Green[9]: self-noise free measurement constrained by $BL < 1$, $\sigma(f, \tau)$—real function.
$B_{\text{r.m.s.}}, L_{\text{r.m.s.}}$	*Doppler and Multi-path Spreads:* transmit two sinusoids of small frequency spacing; arrange to receive each through an envelope detector, filter and perform operations indicated in Fig. 7; repeat at various times to determine probability distribution functions of the measured parameters.	Bello[14]: two-frequency implementation of this method, as illustrated in Fig. 7, constrained by $BL < 1$; validity of this constraint for other waveforms and processing not established.

† Methods for obtaining statistical averages only.

of each rather than the details of their possible implementation. Thus, with the exception of the Doppler and multi-path spread measurements, for which the block diagram has been reproduced, no other schemes of reception or alternatives have been mentioned. However, certain guide-lines for the implementation of each measuring method are implied in the present discussion and it may serve a useful purpose to summarize them here, together with the system function or parameters which they measure, key references and observations pertaining to their applicability criteria. Such a summary is presented in Table 2. It should be emphasized again that alternative implementations of each of the measuring methods are possible and that the details of these can be found in the quoted references.

4. A Note on Some Measurement Problems in Certain Space Channels

Space communication channels, to which reference has been made in Section 2, present a number of special problems not encountered in the more conventional channels such as the ionosphere. Perhaps the most characteristic features of the space channels are their *impermanency* and *uniqueness*. Needless to add, these are also very complex channels, difficult to characterize and, for a number of reasons, difficult to measure.

The implementation of most of the measuring methods implies a certain prior knowledge of the gross channel characteristics. In most of the conventional channels, these characteristics can be established by a crude physical argument; in most of the space channels they must be measured. This fact alone is sufficiently decisive to restrict any initial measuring programme to those methods which require the least *a priori* characterization and whose primary objectives are the diagnostics, so to say, of the channel. There are, of course, other factors which discourage more detailed and complex measurements. Some of the more obvious of these factors will now be considered.

At the outset it must be recognized that most of the space channels do not lend themselves to measurements which depend on back scattering. Consequently, transmission measurements must be employed in which one of the terminals will be rocket or vehicle borne. This complicates the matter, since the measuring methods which depend on correlation must be provided with an undistorted copy of the transmitted signal and the events must be synchronized for proper time delays, integration, etc. This can be accomplished, presumably, by storing the probing signal and by transmitting a pilot reference tone, preferably at a frequency at which its fidelity is least affected by the channel. Synchronization can also be secured by the frequency standards but their effective use in a space environment is rather doubtful or, at least, undocumented. In view of these complications, it appears that early attempts at measuring the correlation functions $R(\tau, \Delta t)$ or $R(\Delta f, \Delta t)$ should be postponed in favour of some less demanding measuring schemes.

It has also been observed that a typical space channel cannot be assumed to be stationary, except perhaps over short periods of time over which no significant changes in the structure of the channel take place. What is meant by short periods of time in the context of such a transient channel as the rocket exhaust is not quite clear, but here again, any measuring method which minimizes the processing time must be looked upon with favour.

Finally, there is the question of isolating the various events during the measurements so that the true characteristics of the channel, as discussed here, are being measured. Thus, for instance, the isolation of the *Doppler shift* from the Doppler spread is essential since only the latter has anything to do with the coherence time of the channel. Similarly, signal fluctuations due to the vehicle's motion ought to be identified and filtered out from the true multi-path facing. Of course, the totality of these events contributes to the distortion of the transmitted signals and should, properly, be considered as legitimate channel events. However, an attempt at modelling a channel which also incorporates terminal variations may prove to be considerably more difficult than the pure medium channels discussed here. In any case, we are not prepared to cope with this situation as yet, except by attempting to isolate the various identifiable events and interpreting them separately.

5. Tentative Conclusions and Recommendations

As this brief review shows, the subject of linear, time-variant channel characterization and measurement is of paramount importance to the design and evaluation of modern communication systems. Much of the motivation for the development of this subject originates in the ever-increasing importance of digital communications over natural or man-made channels and in the current interests in astronomy and space exploration. The characterization and measurement of conventional communication channels appear to be fairly complete and meaningfully related to either the physical features of the channel or the performance criteria of the communication systems. There are now available a number of measuring methods through which this characterization can be secured. Examples of gross parameters and of complete system functions for such channels as the ionosphere Moon scatter, etc., were given earlier in this report, as were the procedures for their experimental acquisition.

The situation is not as simple or complete in the case of unconventional channels which are transient and irreproducible, even on a statistical basis. We would like to believe that the characterization and measurements developed for the conventional channels are appropriate and meaningful in the case of, say, a channel formed by re-entry ionization or rocket exhaust. A brief inquiry into this subject leaves, however, many questions unanswered. For instance, it has been concluded that it is difficult and perhaps even impossible to deduce the gross parameters of such channels from strictly physical arguments. It has not been possible, for the time being, to account for or to isolate those effects which originate with the terminal or are due to its interaction with the 'pure' medium channels.

An examination of the existing measuring methods leads to the conclusion that only a most rudimentary measuring programme can be contemplated for such channels at this time and, in fact, even that programme ought to be preceded by (i) further inquiry into the special measurement problems peculiar to the unconventional channels, (ii) examination of the existing records, spectral or otherwise, for clues as to the magnitude of the gross channel parameters. Only then should diagnostic measurements be undertaken. For reasons pointed out in Section 4, these are limited to impulsive testing, in either the time or frequency domain. Finally, a serious consideration should be given to Bello's method of measuring the r.m.s. values of the Doppler and multi-path spread, but beyond that, proposals to measure the channel system functions must be looked upon, for the time being, as too ambitious and unrealistic.

6. Acknowledgment

The author gratefully acknowledges the helpful advice of Dr. Phillip A. Bello of Signatron, Inc., Lexington, Massachusetts.

7. References

1. H. B. Voelcker, 'Phase shift keying in fading channels', *Proc. Instn Elect. Engrs*, **107B**, pp. 31–8, January 1960.
2. T. Hagfors, 'Some properties of radio waves reflected from the Moon and their relations to the lunar surface', *J. Geophys. Res.*, **66**, pp. 777–85, March 1961.
3. F. Villars and V. H. Weisskopf, 'On the scattering of radio waves by turbulent fluctuations of the atmosphere', *Proc. Inst. Radio Engrs*, **43**, pp. 1232–9, October 1955.
4. E. J. Baghdady, Olen P. Ely and Parley L. Howell, 'Radio propagation effects of rocket exhaust plasma', Paper presented at the I.E.E.E. International Space Electronics Symposium, Miami Beach, Florida, 2nd November 1965.
5. T. Kaliszewski, 'A discussion of factors affecting the performance of microwave systems in certain plasma channels', *The Radio and Electronic Engineer*, **31**, pp. 117–27, February 1966.
6. L. A. Zadeh, 'Frequency analysis of variable networks', *Proc. I.R.E.*, **38**, pp. 291–9, March 1950.
7. T. Kailath, 'Sampling Models for Linear, Time-Variant Filters', Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, Mass., Tech. Rept. No. 352, 25th May 1959.
8. T. Kailath, 'Measurements on time-variant communication channels', *Trans. I.R.E. on Information Theory*, **IT-8**, pp. S229–S239, September 1962.
9. R. Price and P. E. Green, Jr., 'Signal Processing in Radar Astronomy-Communication via Fluctuating Multipath Media', Lincoln Laboratory, M.I.T., Cambridge, Mass., Tech. Rept. No. 234, 6th October 1960.
10. R. G. Gallager, 'Characterization and Measurement of Time and Frequency Spread Channels', Lincoln Laboratory, M.I.T., Cambridge, Mass., Tech. Rept. No. 352, 30th April 1964.
11. T. Hagfors, 'Measurement of Properties of Spread Channels by the Two-Frequency Method with Application to Radar Astronomy', Lincoln Laboratory, M.I.T., Cambridge, Mass., Tech. Rept. No. 372, 11th January 1965.
12. P. E. Green, Jr., 'Time varying channels with delay spread', *Acta tech. Hungar.*, **42**, No. 1–3, pp. 85–97 (1963).
13. R. P. Ingalls. Unpublished measurements.
14. P. Bello, 'Some techniques for the instantaneous, real-time measurement of multipath and Doppler spread', *Trans. Inst. Elect. Electronics Engrs on Communication Technology*, **COM-13**, pp. 285–92, September 1965.
15. P. Bello, 'On the r.m.s. bandwidth of nonlinearly envelope detected narrow band Gaussian noise', *Trans. I.E.E.E. on Information Theory*, **IT-11**, pp. 236–9, April 1965.

Manuscript received by the Institution on 23rd November 1966. (Paper No. 1144/RNA 75.)

U.D.C. 62.001/002

Handing Over the Design to the Works

By

S. C. DUNN,
M.Sc.(Eng.), C.Eng., F.I.E.E.†

AND

F. J. ADAMS,
C.Eng., A.M.I.Mech.E.†

Reprinted from the Proceedings of the Joint I.E.R.E–I.Prod.E.–I.E.E. Conference on 'The Integration of Design and Production in the Electronics Industry', held at the University of Nottingham on 10th to 13th July 1967.

Summary: In this paper the importance of effective communications is stressed. There are three stages in the handover: preparation, transfer and development. While there is still sufficient time to explore the reasons for disagreement between the parties each should be preparing the other to deal with problems which will be urgent later on. The formal transmission of design documents must take into account the mechanics of issuing different versions of the same assembly to different groups of users and coping both with the need for advance information to some clients and historical information to others. Development is the most hectic process. The differing objectives of the principal agents, project, design, inspection and works are described. There are several characteristic features of the manufacturing department which strongly influence this stage of handover.

1. Introduction

Handing over the design to the works may appear on the simpler P.E.R.T. charts as a single event. This may be a convenient fiction for planning purposes. It is our experience that the conditions under which a large number of engineers works together so that thoughts become things must be arranged in such a way that not a single event but a whole series of planned events take place, blending into a continuous shared experience.

Even when one particular design is considered there are more likely to be three stages to the transfer, a preparatory stage, an information transfer and a long aftermath.

In what follows we shall discuss these stages and of the influence upon them of the organization and the aims of different departments. What is described applies particularly to a project organization engaged in small and large-batch manufacture of military and aerospace equipment but many of the features have quite general implications.

The nature of our work is such that each major order given to the works is for designs which may make little use of what was made previously so that a significant cause of manufacturing difficulty is the high rate of innovation. Nevertheless the 'steady-state' component of manufacture has its own special difficulties which will be mentioned.

2. Effective Communication

The realization of ideas involves all aspects of human communication. The handover across the interface between the designer and the maker will only be successful if the proper steps have been taken to safeguard the *effectiveness* of the communication. Since the traffic of ideas is two-way the actions which those ideas lead to will only be what is wanted if each party understands both its own and the other's language, objectives, organization and capabilities. Such mutual understanding is only an extension, on a grand scale, of what two people should know of one another if they are to converse intelligently and persuade each other to courses of action. The exchange of confidence which it implies necessitates careful preparatory cultivation before the formal transfer of information and equally assiduous attention afterwards.

The delivery of a bundle of drawings and specifications from the printing department to production control is conventionally regarded as the occasion of handover. This is a pure formality. The hard work is already half-completed. The tasks which should have been completed by now are:

(a) The choice and detailing of solutions which satisfy the customer's needs and which the works have already decided are best made in a particular way.

(b) Familiarization of all concerned with any new components to be used.

(c) Establishment of communication links between those who can detect difficulties and those who can sort them out.

(d) Proving new test methods and the interpretation of non-standard performance.

(e) Instruction of inspectors as to what is expected of the design.

(f) The development by the works of manufacturing methods capable of realizing the designs.

---

† British Aircraft Corporation (Operating) Ltd., Guided Weapons Division, Stevenage, Herts.

Although the formal transmission of the paperwork during the handover phase proper does not involve the semantic and social problems which come before and after it can be a complex task demanding considerable planning. This is especially so if, as in the case of guided missile programmes, there are regular issues of modified information for successive versions of the product.

Among the topics which are relevant the following have been found to be important.

(a) A paperwork system which satisfies both publisher and reader.

(b) Limitation of distribution.

(c) Storing, retrieval and printing appropriate to user's demands.

While the works is absorbing the instructions and is beginning to make a few parts and put them together it is also discovering reasons why it cannot do so. Or, to be strictly accurate, it is discovering that it cannot do so and believes it knows the reasons. This irregularity can happen in both development and production manufacture. In the former the manufacturing conditions can be modified to take special account of the troubles to be expected but in the latter the exigencies of production are usually dominant.

In the next three sections we will review in more detail what can be done in the three phases of handover. We will call these three phases preparation, transfer and development.

3. Preparation

Activities will be described in the approximate order in which they should occur in order to make the best use of time.

The most important commonalty is to be of one mind as near as possible. This may mean changing of minds on either side. It can be a slow process and may need the experience of several projects to bring conviction. Among the basic issues, old diehards can be recognized. For uniformity they will be presented as prime requirements by the works.

3.1 *Ease of Manufacture*

What is meant by 'easy'? Does it mean the expenditure of the least possible effort or ingenuity to get it made? Or does it mean that ultimately unskilled or semi-skilled labour can be used? This objective is suspect unless carefully qualified. It is usually not basic enough and can lead, if pursued unthinkingly, against the direction of general industrial progress.

3.2 *Lowest Possible Cost*

This is usually the most contentious topic. We believe that the following propositions are valid, especially where a new product is concerned.

(a) It must be made to work first, then trimmed for cheapness afterwards.

(b) Where the product is very new it must be established that it can survive its journey through the works in the first place! Performance may depend on careful attention to details which will feature in almost any design.

(c) Within a short time one man cannot both create and cheapen a new product. Especially where an inventive step has occurred it usually takes a subsequent team, including works participation, to modify the original arrangement of parts while preserving the essential function.

(d) The most spectacular cost reductions occur by not giving the customer what he does not need.

3.3 *Use of Existing Plant and Processes*

This requirement may imply two quite different consequences. If the existing facilities are not capable of making what the designers call for and what the rest of the industry is turning out, the whole operation is under a penalty. On the other hand the works may be equipped with the latest tools but has not explained to designers what it is now capable of. An example of the first situation is the gradual obsolescence of the plant good enough to make printed circuit boards to the standards of a few years ago but which now causes an unreasonably high proportion of scrap. On the other hand the introduction of numerically controlled machine tools may have left quite unaffected the way in which the design office conceives and draws its piece-parts.

3.4 *In-house Manufacture*

This is a delicate question. In a project-oriented company it is often advisable, in order to attract custom, to be able to demonstrate a 'total capability'. As the scale of investment in new processes increases this becomes more difficult. A choice which is of current interest in the electronics-based industries is that between buying and making the appropriate forms of microminiature circuits. We have found it helpful to make visits to semiconductor plants in groups which include both design and works representation.

There is a great deal to be said for shared experiences in preventing misunderstanding. Positive attempts to arrive at better working arrangements are to be encouraged. We can give three instances of ways in which design and works can co-operate in addition to the occasions provided by particular projects.

3.5 *Handbook Compilation*

It is a worthwhile aim of management to replace the exercise of individual discretion by a standard procedure wherever possible. To this end we are compiling handbooks, for use in the design office and elsewhere, which guide the user and creator of designs. Regular discussions on restricted topics force attention on issues which are otherwise in day-to-day controversy. There is a convenient record of what works and design agree are the sensible bases of choice of manufacturing method.

3.6 *Total Process Design*

There are becoming available a number of machines which carry out such diverse tasks as drawing diagrams, shaping metal, inspecting hardware, processing data and so on which can be instructed in codes and languages which, though they may not be the same, are themselves mutually convertible by machines. One can already in a few cases envisage a complete cycle of operations in which the part played by human beings is considerably reduced. It is then possible not only to reduce the time taken to realize an idea but, much more important, to make that time more certain. The design of such routines, which virtually eliminate the handover problem, is already proceeding for printed circuit boards and parts which can be made on numerically controlled machines. It is notable also that there is greater enthusiasm for this kind of collaboration because the chances of success seem greater.

3.7 *Planned Evolution of Future Designs*

By far the most effective preparation for manufacture is to plan jointly to make a product whose performance, manufacturability and cost are intended to evolve progressively towards a level which is a significant improvement on current values. The steps in that progress should be so chosen that achievements short of the final aim are definable and worth having. In this way the disparity between gestation times of projects and long-term technique improvements can be reduced.

We have a programme for developing a standard electronic assembly scheme which is intended to put those constraints on the designer which will not unduly hamper his freedom to package a wide variety of circuits and yet limit the possible variety of packages so that the shops can optimize process and tool design.

4. Transfer

Engineers tend to be impatient of what they think are unnecessarily elaborate arrangements for printing off some drawings for the shop. It is instructive to review the reasons for a typical issuing system.

(a) Despite the preparation by a substantial number of engineers, most of the manufacturing and associated staff will be seeing the documents for the first time. As in any other publishing business, explanation and advertisement in advance of the release of the bulk of material is advisable.

(b) The dispersion in reading time over the whole range of items described obliges much information to be made available before the set of drawings have been completed.

(c) No matter how soon and how well parts of the design have been specified, this advantage has often to be sacrificed in order to communicate an even more effective idea or perhaps just one that actually works. Successive refinements in design have to be identified since they may follow one another quite quickly. The consequent actions (planning, purchasing, tooling, etc.) may be quite sensitive to these changes.

(d) Close control must be exercised over the planned variations in design. Successive models may have to possess special features and yet embody most unchanged parts.

(e) Apart from variations in build standard, the inspection and testing of certain models which may not appear for several months has to be determined in advance.

(f) Although a part or an assembly may undergo a series of changes useful quantities of hardware may lie in stores because they were completed before the changed order reached the shop. This hardware may be redeemable if the historical design record is available in its entirety. It may be necessary, therefore, to print from any stage of modification.

(g) The difficulties of handling, storing and distributing large numbers of prints and original drawings and modification states forces the use of microfilm for the propagation of design documents.

One example only will be given of the particular issuing system in use at Stevenage. The need to introduce new documents, to modify them, to record their birth, marriage and death and in a generally flexible way to inform the readership is satisfied by a MAPIN (modification, amendment, print issue note), so called from its three chief modes of use. The design office is thus provided with an instrument of communication which the distributing network handles without fuss and whose application is limited only by ingenuity.

5. Development

Our experience of what we call development is probably common to all those engaged in similar industries but may seem rather odd to many others. Our products are guided missiles, satellites, their associated ground equipment and a variety of industrial and military by-products. In the consumer-durable industries the development process tends to be slower and is seen as model-to-model variations. With us it is more appropriate to define development as the controlled modification of drawings within a limited time and budget. We probably experience all the difficulties encountered in other industries in an acute, accelerated form.

The principal agents in this last act of our industrial drama are 'the project' (in fact a number of engineers whose ultimate interest is the satisfaction of the customer), the design department (again, several engineers, representing distinct techniques such as circuit design, structures, control), the inspection department and the manufacturing department.

The action chiefly concerns the solution of problems which appear to have their origin in the works. We say 'appear' because although the message which signals 'trouble' comes from that area this is often only because the drawings are being scrutinized minutely for the first time under the conditions in which they are intended for use.

The technical and inter-personal difficulties which arise are due more to the differences in the objectives of those taking part than to technology. It is helpful to examine a little more closely what the agents are trying to do.

5.1 *The Project Engineer*

The project engineer has by now spent more than he intended in getting a set of allegedly checked drawings and specifications into the shop. Although his programme allows for unforeseen difficulties and delays, when these do occur he cannot agree to remedial action which is too radical or takes too long because further trouble may be ahead. There is one way, however, in which he is always willing to spend money. It is usually judged better, in the long run, to retain a selected band of engineers with long experience of the project even if their effective utilization may at times be rather low. Continuity is always worth paying for. Not only is the introduction of a new engineer an extra expense but the previous man has departed with a great deal of useful information which the project has paid for but never seen in a communicable form. The next important programme event is the completion of the prototype, followed by a practical demonstration of its performance.

5.2 *The Designer*

Unless he is experienced, the designer will find himself being drawn into an environment which is unfamiliar. Electronics, especially, tends to be a young man's game and circuit designers soon become involved in situations very remote from those which inspired them in the first place. At this time it often becomes clear that the organization itself is hindering the development process. The nature of the hindrance can be understood by asking the simple question—who is the designer? The replies are often evasive. But to the question—'whom must I ask if I wish to modify the design?'—the answer could be—three people, the detail draughtsman, the mechanical assembly designer and the circuit designer. It happens all too frequently that the project engineer is not only the final arbiter but also acts as designer when the design is the work of many hands.

The designer who is in no doubt that he is responsible can at this time be involved in months of work, spending long hours in the test department and the drawing office. This is the real post-graduate engineering experience for which higher degrees are unlikely to be awarded and about which learned society papers are seldom written. Many find it so unpalatable that they seek employment as salesmen or college lecturers at the earliest opportunity. During development the designer corrects those errors which he could not foresee earlier, the errors which the works may have made in the tools, the errors newly-arisen from the customer's change of mind, the errors caused by other designers. He also makes those choices for which there was formerly insufficient information and which the development programme now makes available. As new problems appear and the programme drags, the management is impatient to begin the design of the next project and finally does so at the expense of the original one by transferring the designer. One reason why this unprincipled horse-trading does not bring all projects to a premature end is that some dedicated souls are development engineers by temperament. A well-balanced engineering department has the appropriate mixture of 'blank paper' designers and development engineers.

5.3 *The Inspection Department*

The inspection department must often wonder what it is supposed to be doing. The designers believe it should make certain that the works are actually doing what they have been told. The quality manager regards them as an instrument measuring the defects. At times they act as agents on behalf of the customer's inspectors. Occasionally both design department and works feel that the design has slipped from their grasp and has been handed over to the inspection department. What the inspection department are actually doing is to detect the discrepancies between what is stated, on documents, to be wanted, and what turns out to be acceptable. They are often castigated as obstructive, but if they are persuaded too often to bow to the discretion of the designer they cease to be of any use at all.

The works is the management's chief instrument for making money. Its aim is to expend the greatest possible amount of effort profitably. Because their survival conditions are different, works and design may become estranged. Only by reconciling their very different aims can designs be handed over successfully.

It could be that the finest course in industrial management is that undertaken by a development engineer on a large project. Trouble-shooting involves him in trying to understand the works organization, its communication system and procedures and the nature of the decision-taking processes. We have judged it inappropriate here to describe in detail a complete works organization because the details, though fascinating to anyone caught in its web, would tend to obscure our main theme. Instead we will touch upon some of the more significant features of that organization as they affect a development engineer.

5.4 *Manufacturing Instructions*

If the engineer believed that the drawings only were necessary to make the hardware a brief acquaintance with production control will put the design in perspective. The project and contract requirements together with the drawings have now generated perhaps five new families of paper with names like manufacturing order, works programme, batch plan, schedule and process sheet. In some areas, where the drawing might be thought to be rather useful, the drawing has disappeared altogether and the operative seems quite content with a planning sheet.

5.5 *Organization*

There is a much finer division of labour in the works than in the engineering department and the paperwork system is more disciplined. Near the shop floor we are now far downstream in the information flow and many tributary streams have joined that flow. Back in the design office, months ago, it was possible to stem or briefly reverse the progress of design. Here, now, there is relentless pressure to keep on doing whatever can be done and to put on one side anything which gives trouble.

Although there is an extensive rule-book, like any other human organization it works smoothly because there are more effective social contacts governing the relations between the members. The engineer called in to sort out trouble only becomes effective when he is admitted to honorary membership of this social network.

The control system is sometimes difficult to identify because the people encountered do not always have a wide view of the complete organization. The assistance they can render an outsider is limited by the role they play. That role is usually quite sharply defined.

5.6 *Constraints*

Called in to sort out a difficulty the engineer finds several features which continually influence what he can do.

(a) Although the problem has arisen from a single measurement on one part any action affects the fate of an entire batch of parts. The extent to which he can devote resources to solving the problem will depend on how seriously the programme is affected by delaying the batch.

(b) If the reason for the difficulty is found it may then be sensible to change something. It usually appears that nothing whatever can be altered, or rather there is so much doubt about the further consequences of some of the possibilities that they are left untouched.

(c) The kind of solution depends on whether the problem will recur (design fault) or may not (manufacturing fault). In either case any change in the anatomy of an assembly must be adequately recorded since the results of subsequent tests may not otherwise be interpretable.

(d) It may be sensible to ask the indulgence of inspection. Depending on the magnitude of the misdemeanour and its time of occurrence the exercise of discretion may take the form of an entry in the grant book or the favour of a manufacturing permit or a concession.

(e) A characteristic feature of manufacturing instructions in the conventional form of drawings and specifications is that, in formal terms, their semantic content outweighs their syntactic content. By this we mean that people do not necessarily interpret the instructions literally but claim to know intuitively what is wanted. This delusion may lie hidden as a trap until manufacture is transferred to another place. An object which has been made without comment for years is suddenly declared unmakeable.

5.7 *Development as a Learning Process*

After months of working only on paper attention is now concentrated on the creation of hardware. The true nature of the initial stages of the development phase can easily be lost sight of. It is a learning process. As successive batches are made, even with identical drawings, the stimulus they provide yields a response which is quicker and nearer what is wanted. Like more conventional learning situations it displays the characteristic faults of those situations. In particular it tends to be 'teacher-oriented' rather than 'learner-oriented'. The design department is so often the source of new ideas that the works are likely to suffer a feeling of chronic inferiority. This is regrettable because the works should be playing its part in educating the designers to know about its capabilities. In most firms much more money is spent on machine tools and processes than on improving the means whereby designers produce their ideas. The other typical fault is inadequate provision for feedback from learner to teacher. It often seems to the outsider that the works is well designed to run forwards in the steady-state but behaves poorly in feeding back information in time of crises.

For this reason it is common to appoint a liaison engineer who can smell out trouble when it still has only a delicate aroma. He must be supported by an adequate reporting system. It is convenient to digress a little on the general communication problem.

5.8 *Communications*

It is sometimes said that drawings seem to be produced by a draughtsman for other draughtsmen. A distinction should be made in the three uses of drawings. These are:

(a) Originally, by designers, to record decisions and to aid the imagination.

(b) Eventually, by the customer, to make further numbers of the hardware from other sources of manufacture.

(c) Now, to tell the works what to do.

In handing over the design every effort should be made to ensure that the manufacturing staff—all of them—are in no doubt as to what is wanted. This is usually appreciated well enough in the planning department even to the extent of substituting for the drawings instructions more suitable for the operative. Most of the changes taking place in the media at the present time are concerned with machine-compatible instructions.

Special problems face those companies whose establishments are scattered over the country or who have formed associations with foreign companies. When there are several design offices and several works all in different companies and different places the problem of collaboration ceases to be merely technical and can only be handled satisfactorily as a senior management responsibility. Even within one company we have found advantages in using special means of communication to enable works and design difficulties to be reported and advised on as quickly as possible while still compiling a record of action taken. Closed-circuit television, facsimile and telex may all be appropriate.

6. Conclusions

The works is gradually put in full possession of a new design after three stages in communication. The second stage, formal transfer of documents, is special to a particular project. The other two, preparation and development, grow out of the general arrangements for co-operation between manufacturing and design departments. Of the various difficulties which arise two may be singled out as more important than all the rest.

In the manufacturing department it is difficult to proliferate responsibility among all concerned with the fine divisions of the total task. This is basically a social problem (see Appendix 2), but until it is solved there may be no true meeting of minds.

The disparity between the formal organization and the social network which keeps work moving is often remedied by the appointment of liaison engineers. There is, however, a need to recognize the existence of an interface problem so great that it merits solution at management, engineering and technician levels.

7. Acknowledgments

We would like to thank our colleagues at Stevenage Works of B.A.C. (Operating) Ltd., G.W. Division for much stimulating discussion and the Chief Engineer for permission to publish this paper.

8. References

1. Colin Cherry, 'On Human Communication', p. 217, 1st Edition (Chapman and Hall, London, 1957).
2. W. Ross Ashby, 'An Introduction to Cybernetics', p. 140, 1st Edition (Chapman and Hall, London, 1957).
3. N. Wiener, 'Cybernetics' (John Wiley, New York, 1948).

9. Appendix 1:

Design as a Cybernetic Process

Following Ashby's restatement[2] of Wiener,[3] the communication of design intent is a cybernetic process since it is the 'transmission of variety'. It would thus be reasonable to expect to see the characteristic features of such a process:

(a) The presence of information, characterized by syntactic, semantic and pragmatic content.[1]

(b) Problems of noise and the tactics of coding and de-coding.

(c) Feedback to oblige purposeful behaviour.

(d) System 'diseases' such as distortion, delay, attenuation, instability.

Information: compared with the situations more often studied, the pragmatic aspect of the transmitted information is of relatively greater interest. Also because features such as distance and bandwidth and time are all much less pressing than in the much-studied telecommunications systems, the construction of the message is a less important task than the recording of design decisions.

Coding: the idea originally present in the designer's mind is converted into several forms and broken in several parts before it takes shape as a real object. The idea is relayed through a branching chain whose links are sometimes men and sometimes machines. The form taken by the fragmented idea has to suit the agent for whom it is intended.

The number of different forms taken by the information can be quite large since many agents are involved. Although different companies will have different names for them the following list can usually be identified: parts list, material in advance, stores demand note, purchase requisition, purchase order, drawing.

The nature of the agent influences the form. The drawing board, the numerically controlled machine tool, the unskilled assembler, the process operative all wish to be spoken to in the language they best understand.

Feedback: error correcting loops abound in the design process. Some are local (for instance confined to the designer and his immediate colleagues), some are more extensive (embracing the customer and sources of supply). Others are active in real time and involve the designer in current manufacturing difficulties, others are institutional in nature and intended to keep the designer aware of the general capabilities of the works.

9.1 *System Response*

The complexity of what is transmitted and what is to be made is reflected in the intricacy of the organization and procedure employed. The consequence is that the arrangements made to handle information conflict with the need for adequate control. This organic conflict leads to a very common organizational disease whereby difficulties arising in manufacture are solved by giving one man or a very restricted group of people special powers to get things moving. The difficulties arise in the first place because the organization and its information transfer properties have been poorly designed. For example, it is common to find that the paperchain works quite well in the forward direction so that abstractions eventually and progressively become hardware but hardware which develops defects at any stage does not effectively signal that fact back to those who can find remedies.

9.2 *The Project-oriented Organization*

It is not surprising that the project-oriented company adapts its structure to favour its ultimate sources of profit. The arrangements for communication and control tend to be scaled to the size of a project team and to have time constants which match the duration of the project. System diseases are seen as acute and are dealt with quickly, therapeutically. Although the need for long-term schemes for improving techniques is admitted, prophylaxis is at a discount. Eventually the organization polarizes into project and technique groupings. Unless information produced by the quick-acting project feedback loops is used as a basis for the slow-acting controls, project achievement can be no better than what is possible within its own limited time-scale and budget.

10. Appendix 2:

The 'Apartheid' Between Designing and Making

If, for our present purpose, we define designing as the period of mental preparation which comes before the modification and rearrangement of materials which we call making, then in engineering these two activities have grown so far apart that they challenge each other in a way essentially different from, say, in painting or sculpture.

For a host of reasons they have been encouraged to seek their own fortunes separately. There are manufacturing plants which have no associated design office, design offices remote from any means of realizing their output. As a result the preparation of drawings becomes an end in itself, the exploitation of existing machines, tools and processes the main aim.

Besides division there is also a lack of balance between the activities due to the different ways in which the employees are involved in their work. In the works, the lesser interest in a career and the greater discipline lead to a more mechanical operation which, coupled with the progressive replacement of muscle by machine, result in a considerable dilution of intellect.

The identification of the works with social unrest, political extremism and unprogressive ideas coupled with a mixture of snobbery, prejudice and lack of vision has consistently loaded the scales against the works. The learned societies, the universities and the schools have all at one time or another helped to poison the mainsprings of the industrial wealth on which they ultimately depend.

It is a common complaint by rejected applicants for transfer to a higher grade of membership in an engineering institution that their too close association with the works has been detected. Apparently it is entirely acceptable to set problems for the works but damaging to help them solve those problems! A serious consequence of this has been a manifest reluctance on the part of graduates to embark on a career outside design and development. This is already hindering the deployment of the appropriate talent where it is most needed.

The formal education of engineers no longer reflects faithfully the practice of engineering. First degrees are now granted in narrower and more specialized aspects of that practice than is healthy for the future development of the profession as a whole.

The part played by the schools in laying the foundations of opinion and attitude is also important. Their role in moderating the influence of a materialistic society upon the next generation of salary earners must be respected. Yet the place of technology in that society does seem to be misrepresented to children. While science appears in school as a subject worthy of study, engineering often takes the form of a restricted set of craft skills.

The expansion in education has led to a more careful matching of the output from schools to the input requirements of universities and colleges. It follows that not only are those recruited directly to the works as craft apprentices less promotable than the previous generations but those who become designers by the new route lack the extensive background of works experience of their predecessors.

Manuscript received by the Institution on 3rd May 1967. (Paper No. 1145.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **September 1967**

September 1967	24-hour mean centred on 0300 U.T.		
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz
1	− 299·9	0	0
2	− 300·0	0	0
3	− 299·9	0	0
4	− 299·9	+ 0·1	0
5	− 300·0	0 *	0
6	− 300·0	0 *	0
7	− 299·7	+ 0·1	+ 0·1
8	− 299·9	+ 0·1	+ 0·1
9	− 299·9	+ 0·2	+ 0·2
10	− 299·9	+ 0·2	+ 0·2
11	− 299·9	+ 0·1	+ 0·1
12	− 299·8	+ 0·1	+ 0·1
13	− 299·8	+ 0·3	+ 0·2
14	− 299·8	+ 0·2	+ 0·2
15	− 300·0	+ 0·1	+ 0·1
16	− 299·7	+ 0·2	+ 0·2
17	− 299·9	+ 0·1	+ 0·1
18	− 299·8	+ 0·1	+ 0·1
19	− 300·0	+ 0·1	+ 0·1
20	− 300·0	+ 0·1	+ 0·1
21	− 299·9	+ 0·1	+ 0·1
22	− 299·8	0 *	+ 0·1
23	− 300·0	+ 0·1	+ 0·1
24	− 300·0	+ 0·1	+ 0·1
25	− 300·0	+ 0·1	+ 0·1
26	− 300·0	0 *	+ 0·2
27	− 299·8	+ 0·2	+ 0·2
28	− 299·9	+ 0·1	+ 0·2
29	− 299·9	+ 0·1	+ 0·2
30	—	+ 0·2	+ 0·2

Nominal frequency corresponds to a value of 9 192 631 770·0 Hz for the caesium F,m (4,0)–F,m (3,0) transition at zero field.

1. The measurements were made in terms of H.P. caesium standard No. 134 which agrees with the N.P.L. caesium standard to 1 part in 10^{11}.
2. *Phase variations occurred at the transmitter in the afternoon of these days.

U.D.C. 681.32:621.37/9

Some Applications of Computers in Electronics Design

By

N. E. WISEMAN,
M.A., M.S., B.Sc.(Eng.)†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.Prod.E.–I.E.E. Conference on 'The Integration of Design and Production in the Electronics Industry', held at the University of Nottingham on 10th to 13th July 1967.

Summary: Design is seen as a complex activity requiring intellectual effort at several levels. The highest level is what we, rather loosely, call creative and corresponds with the selection ('invention') of potential schemes. There follows an analytical activity ('how does it work?') and finally a clerical operation ('how good is it?') which yields quantitative data on which the acceptability of the scheme is judged. Interactions occur between activities at the several levels until a satisfactory solution is achieved (or hope is given up of ever finding one). Some of the operations, primarily from the lower intellectual levels, may be handled by computer methods but very few applications are known in which the entire design process has been performed by the computer. Typical applications at the present time necessitate constant interplay between man and computer and this requirement for a man-machine partnership raises special problems in hardware and software departments.

The paper identifies the roles of man and computer in typical design applications and outlines some of the special hardware, software and system techniques which have been developed. Particular computer-aided design (c.a.d.) topics in circuit design, logic design, placement and conductor layout are discussed. Finally the generality of total systems and the interfaces between the c.a.d. office and other plant activities are considered.

1. A Model of the Design Process

If we ask, say, a professional circuit designer how he goes about his work he will probably answer something like this: 'First I think about the problem. Possible circuit topologies will occur to me, sometimes out of the blue, sometimes out of my past experience with similar situations. I experiment with each arrangement and generally doodle about until something tells me I have a potentially viable scheme. This I investigate in greater detail, first by writing inequalities which relate the required performance with analytic expressions for the actual behaviour, second by seeking numerical values for the parameters of the circuit such that all the inequalities are satisfied. Sometimes I succeed; sometimes not, in which case I go back and start doodling again.'

In order to carry out such a task on the computer we should have to analyse and quantify all this. We should have to decide, for example, what 'thinking about the problem' could mean, how possible topologies suggest themselves 'out of the blue' and what it is that indicates a 'potentially viable scheme'. It is unlikely that this would achieve much success! The designer does not know with any precision why he does what he does and his words, parodied above, betray this fact only too clearly. However, note that not all his explanation is vague, and the last few steps, in particular, may be stated clearly, unambiguously and even algorithmically. Some parts we conclude are immediately suited for attack by computer, others are rather difficult and some are (at present) beyond comprehension. Pictorially we envisage Fig. 1. There are, of course, instances of design activities which are not sensibly represented by Fig. 1. The design of a power transformer, for example, is very probably an activity which suffers only slightly from the intuitive hunches of the engineer and can therefore be handled almost entirely by computer. A one-of-a-kind lady's evening dress on the other hand will be designed from objectives which are neither functional, practical nor economic and only the top box in Fig. 1 will have any works inside it. Nevertheless, extreme cases are fairly rare and a considerable amount of design activity is reasonably described by Fig. 1. Applications are known in, for example, ship-building, mechanical engineering, building, printing, chemical plant, aerospace and electronics industries. Examples from the latter will be used here to typify c.a.d. and expose some of the underlying problems.

---

† University Mathematical Laboratory, Cambridge.

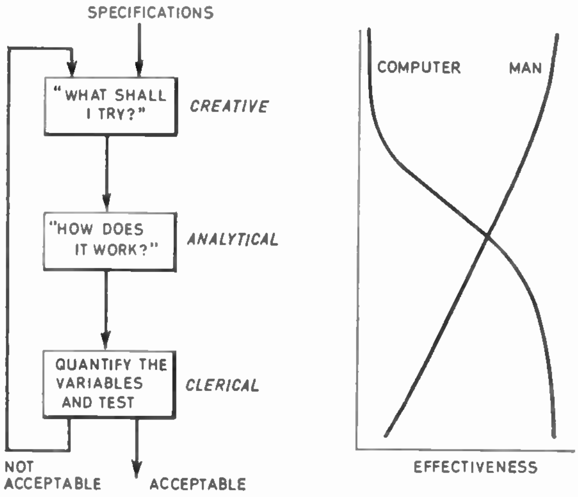


Fig. 1. A model of the design process.

2. Role of the Computer in Design

It is clear that those operations within the design process which are largely clerical are immediate candidates for handling by the computer. It is equally clear that certain of the so-called creative functions should be performed by the man, at least until our understanding of these functions is drastically improved. Wherever the precise division occurs between the tasks for each, we are faced with an interface and signalling problem between the man and the machine. The success of the entire process may, in fact, depend on the efficiency of communication across the man–machine interface and special consideration to this aspect will be given in later sections of this paper. The administration of the interface and format translations of the data which cross it are examples of further tasks which are obviously best carried out by the computer, for it is important to remember that the designer will not ordinarily be a professional programmer and will particularly not want to bother with machine-oriented data formats.

Having selected an application to be implemented, there is very considerable scope through proper programming for helping the man do his part of the job effectively. It is possible, for example, to arrange that the machine leads the man 'by the nose' through certain steps, making sure that no casual decisions are taken which might invalidate the result. It may show him catalogues of standard components and remind him about the availability of existing solutions, procedures or consultant experts. It can provide the data base in which a number of designers might work in collaboration, providing rapid and convenient filing and retrieval facilities for their programs and data. To do these jobs well, if at all, a large time-sharing computer with multiple access facilities is required. Additional special features to do with the use of such a machine in computer-aided design form the topic of the next section.

3. System Requirements

The hardware interface between the man and the machine may take many forms and an appropriate choice for any particular application should enable communication in a language 'natural' to the application. This may involve the sensing and transmission of words, pictures, sounds, movements and what have you. An ideal interface for general purpose use has not yet been invented or, come to that, even specified although many devices of restricted capability are available. Besides providing a suitable language capability the interface must pass data at a suitable rate and within a specified delay. There are thus bandwidth and signalling speeds to be considered. A typical batch processing installation may have interface devices of high bandwidth (say, line printers) but the signalling delay (turn-round time) would rarely be less than an hour—and may even be a day or more. On the other hand a multiple-access installation may provide the user with an interface of low bandwidth (e.g. a typewriter) but with very short signalling delays, say a few seconds. Some design applications may be done satisfactorily on one installation, some need the other, some demand something more. It may be that both a wide bandwidth and a short signalling delay is required, for example, in make-up planning for page composition in the printing industry. It may also be that the interface data do not fit conveniently into the format of verbal language. They may, for example, involve the description of shapes, or connected networks or mechanisms in motion and in such cases a more sophisticated interface is required.

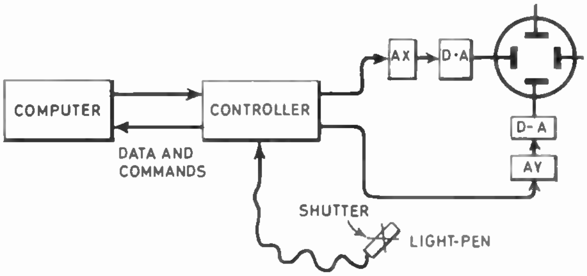


Fig. 2. Cathode-ray tube display and light-pen.

A device which is becoming common is a cathode-ray tube display coupled with some form of position indicator. With it we can communicate graphical and pictorial information in both directions with comparative ease. Suitable position indicators include the light-pen, joystick, tracker ball and various stylus-tablet devices (e.g. the RAND tablet). In the case of, for example, the light-pen, the operation of the interface is as follows (see Fig. 2). Two registers AX and AY drive digital-to-analogue converters which determine the beam position on the screen. In response to position commands sent by the computer, the controller may change these registers so as to move the beam to any desired location. A succession of such commands can be issued to form pictures of arbitrary complexity and in arbitrary formats. The light-pen is an electronic eye. It simply senses the presence of light within its field of view when the switch or shutter is operated and informs the controller. Since the computer knows, or can find out, what it was doing when the pen registered a 'hit' it can determine whereabouts in the picture the pen is pointing. To point at an arbitrary position on the screen, where there may be no light, it is necessary for the computer to track the pen by probing its field of view. In order to maintain the picture on the screen the controller must repeatedly send the data to the display. It may be necessary to refresh (repeat) the display 30 times per second to reduce flicker and even a moderately simple picture may contain over 10^4 points so that a data rate of 3×10^5 points/s may be necessary. This would use up most of the time of even a very fast computer and is clearly very inefficient. The situation may be improved by adopting one or both of the following devices:

(a) Line, curve and character generators may be incorporated in the controller to expand the data sent from the computer.

(b) The controller may contain a private buffer used solely to refresh the display. The controller may in fact be a small computer operating as a satellite to the main machine in which case some of the administration of the interface can be carried out without attention from the main computer. Which jobs are done in each machine is not easily decided. The main machine is a multi-access computer and the signalling delay will depend on its work load. The satellite on the other hand is limited in computing power and core size and cannot do everything. Servicing terminal interrupts (display, pen, keyboards, etc.) would obviously be done by the satellite, while the main computer would, of course, run the design programs themselves. In between is a substantial software interface to be shared between the two machines in some way. This is discussed in the next section.

4. Software Considerations

The data flow within the design process is outlined in Fig. 3. A description of a proposed system is first constructed in formal terms suitable for input to the design-aiding programs.

This description ('model') may then be filed as part of the designer's work-in-hand. The model will then be prodded and examined by the man and by design-aiding programs. It will be edited from time to time and in due course, hopefully, will come to represent a satisfactory solution. The command flow, not shown in Fig. 3, will in general issue sometimes from the man, sometimes from the machine and the man–computer interface may thus have to couple with any, or all, boxes in the figure. We shall therefore consider the operation in each box as a computer-aided activity when discussing the software techniques involved.

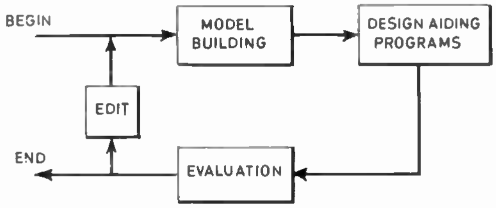


Fig. 3. Data flow through the design process.

4.1 *Model Building*

Any physical system may in principle, be represented in the computer with arbitrary precision by a suitable analytic model. The model can show the behaviour of, and the relationships between, the component parts of the system and exhibit appropriate responses from interactions with a simulated environment. For a particular investigation or under particular operational conditions large parts of such a model will be redundant and wasteful of modelling space (core space is the clay out of which such models are built). For example, a model for an electronic circuit may sometimes require to exhibit geometrical features, sometimes topological features, but rarely both at once. Sometimes we shall want to know the temperature characteristics of some component, sometimes its acoustical resistance, but rarely both at once. Our modelling technique must be flexible and easy to adapt to different circumstances. If we choose to construct models of great precision for filing purposes, it should be possible to distil out of them simplified versions appropriate to particular applications. The behaviour of component parts of the system would be modelled by characteristic equations, or tables, relating response to stimulus. This information could be stored in the form of subroutine calls to routines which evaluate the behaviour. The relationships between components could be modelled with a topological structure carrying the components on its nodes. It could be stored conveniently as some kind of list structure. List structures, such as those employed by list processing languages

(LISP, SLIP, WISP, etc.) can represent with complete generality multi-dimensional relationships between symbols, although more rigid structure formats are sometimes used, such as in the associative ring structure language CORAL, in order to speed the traversing of a given model and to conserve slightly on storage space. A possible list representation for a simple RC network is shown in Fig. 4.

The boxes represent list cells and contain two address-length fields, known in LISP (and other languages) as CAR and CDR. These fields contain either pointers to other list cells or symbols (atoms) or terminators (shown as /). Note that this structure is only one out of several possible representations for the topological model of the RC network shown, and of course there are other models of the same RC network which may be needed—geometrical, thermal, etc. In the case shown the pointers (addresses of list cells) form closed rings which associate the connective features of a component with the component. Thus the two rings starting on node 'a' pass through the branches R and V, revealing the fact that R and V connect with node 'a'. Connection is an all or nothing affair so the association between nodes and branches is simply that one is on a ring of the other or it is not. This is a slightly special view of class membership which may in general involve degrees of association and in such cases a ring passing through an element would carry an extra word (or group of words) containing the value of the association. The number of rings and ring-pointers may be reduced somewhat by passing each ring through all the associations for a given feature of an element in succession. The ring would then need to have a definite start such that the associators could relate without ambiguity one element (a ring member) with another (the ring start element). Thus a ring from node 'a' in Fig. 4 could pass through branches R and V in turn to indicate that R and V connect with node 'a'. It may also be necessary to indicate different kinds of relationships between elements—the example of Fig. 4 illustrates only connective relationships—and in this case each ring would need a name or type descriptor. Data structures having all these properties can be built quite readily from these simple two-field list cells but in some cases a different format may be advantageous. For example, to traverse a structure from ringpointer to ringstart requires that the program chains forwards through all intervening ringpointers. In a complex structure the time penalty for this may be unacceptable and to improve matters the rings are sometimes embellished with further pointers enabling traverse in both directions coupled with a quick hop to ringstart. A further inefficiency may occur when an element of many words is set up from consecutive registers because the CDR pointers which are then unnecessary become essentially deadwood in the structure. Examples of data structures designed to overcome one or more of these problems are found in SKETCHPAD, CORAL, ASP, RSP, etc., but further detail is beyond the scope of this paper.

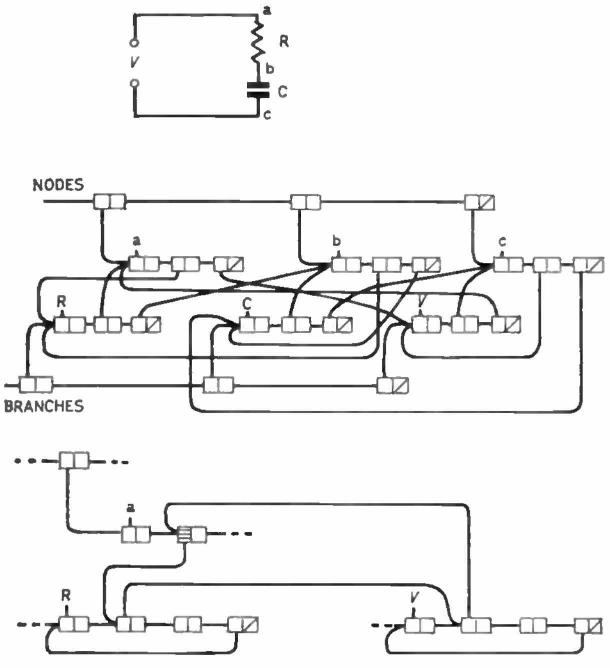


Fig. 4. Simple data structure representations of an RC network.

How would a model, such as has been described, be put together? It embodies the ideas of the designer and he has thus in some way to create it. As a structural entity, however, it contains altogether too much fine detail for him to control or even apprehend, and he will need a good deal of assistance from the computer. Here is where one of the format translation tasks at the man–machine interface is relevant. Programs can be built which enable, say, a circuit diagram to be put together on the screen by the man while a structural model of the circuit is assembled in core. A particular program which does just this is part of a computer-aided circuit design project at Cambridge. The main data paths are shown in Fig. 5.

The real time program services all the terminal interrupts. When tracking the light-pen this routine builds segments of display file showing the user what interpretation is being put to his actions. For example, when his hand is moved along roughly a NS or EW direction straight line segments exactly along NS or EW directions are inserted. A deliberate deflection from such a path causes entry to a recognition routine which selects and attaches a symbol from a catalogue of standard electronic symbols. Lines and symbols may be 'undrawn' (erased) by simply retracing over them. From time to time (actually when a new start

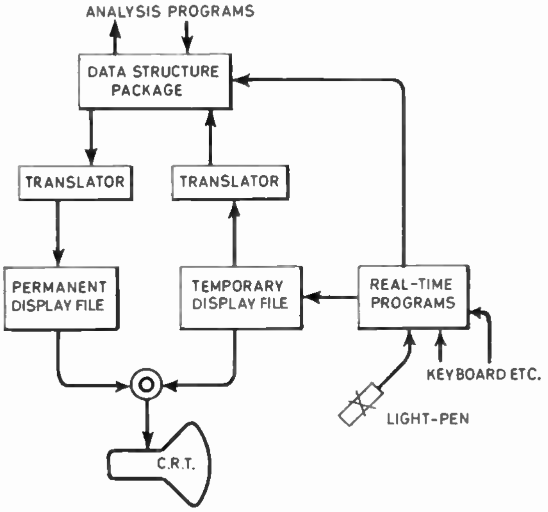


Fig. 5. Model building and format translation.

point is selected) the temporary display file is translated and appended to the data structure. The permanent display file is then recompiled from the data structure so that it carries the new features, after which the temporary file is reset. The reason for coupling the real time programs with a temporary display file rather than directly with the data structure is to improve the 'naturalness' of the drawing action by providing a very rapid visual feedback from the screen to the man. This is especially important if the data structure package runs in the main computer rather than in the satellite since a signalling delay of only a few seconds would make the drawing action seem quite unnatural. In the circuit design project mentioned, it is expected that a copy of the structure processor and a large part of the data structure itself will be in the satellite at input time, so the delay time to get an item into the data structure will usually be quite small. In this way model building can, it is hoped, be made relatively easy for the man; perhaps not too much harder than drawing schematic diagrams with pencil and paper.

4.2 *Design-aiding Programs*

There is already a substantial literature on the application of computers to problems in electronics design and useful non-interactive programs exist for circuit analysis, package layout, placement, wire routing, system and logic simulation. To connect these programs into a conversational environment such as has been described is quite straightforward and the advantage of rapid communication between such a program and the user have already been stressed. There are, however, additional benefits to be had. Particular problems in electronics design are known for which useful programs have proved elusive. For example, no non-interactive program is known which can plan layouts for large-scale microcircuits. The author's attempt at this particular problem some years ago resulted in a program which could do perhaps 50% of all circuits presented to it. It was thus only marginally useful. Its clerical handling of the layout data was, of course, faultless (by contrast with a draughtsman's) and what was needed was a rapid method for sending hints and suggestions to the program when it found itself unable to proceed. The implication is obvious.

It is still a general observation that package design and placement programs are among the least successful in electronic design applications. They work slowly and generate unspectacular solutions. Any benefit from the use of such programs comes more from their low fault rate (in terms of forgotten components and so on) than from their elegant results. Great improvements may be expected from the adoption of interactive methods of solution for problems of this sort.

4.3 *Editing*

The main computer will presumably offer the user some form of filing system in which his programs and data can reside during the course of a project. The designer will use such a facility to hold intermediate results, structural models, catalogues and so on and will be able to do some of his work through the usual systems of filing and editing facilities. For editing his structural models, however, he will need the same sort of assistance he required to construct the models, probably some form of graphical language operating through the screen of his c.r.t. display. In fact structure building and structure editing are inextricably tied together. During building the designer will want to edit out any mistakes which he notices; during editing he will want to build new parts to the structure. The scheme shown in Fig. 5 will thus connect with the filing system at some point and picture editing facilities will be provided through the same mechanism. Picture editing requires operations like *move* and *delete* in addition to the *create* type operations used for building. Pointing the light-pen at some object on the screen will give an interrupt from some part of the 'permanent' display file which the edit program will have to associate with an entity in the data structure before the required edit function can be applied. This may be done by carrying identifiers in the display file which reference print-names in the data structure. The print-names in turn reference appropriate elements in the data structure, so that the real time program can issue interpretations of a pen hit which are meaningful to the man (in the form of a print-name, say for example R16) and to the program (in the form of an address in the data structure, for example the address of the model for the resistor of serial number sixteen).

5. Applications

A few further words about applications seem justified. A review of what a few specific systems should do will serve to point out some desirable operational characteristics of c.a.d. systems in general.

5.1 *Circuit Design*

The designer draws on the screen, more or less freehand, sketches of some proposed circuit. He names components and specifies values, sends the circuit to the file and subsequently applies an analysis program to it. The results of the analysis are displayed perhaps as phase planes or time domain waveforms and, after inspecting them, the designer probably recalls the circuit model from the file, edits it, files it, and tries again. When drawing the circuit certain components and sub-circuits would be required and these also reside in the file. To incorporate, for example, a transistor into a circuit the designer expects the conventional symbol for a transistor to appear on his screen but some equivalent circuit for it to be added to the structural model. This equivalent circuit is an editable thing and a catalogue of different versions, appropriate to different circumstances, may be built up by the designer. One is led to the idea of some hierarchical arrangement whereby high-level picture parts are defined in terms of lower-level parts, the lowest level comprising the basic circuit elements, resistor, inductor, capacitor, switch, voltage source and current source. Dynamic conditions may be set up to simulate the behaviour of non-linear components and to operate switches in such a way that any electronic system may be represented with good precision.

5.2 *Layout*

In this case the designer draws with a different sort of accuracy. He has control over those geometrical factors which affect placement and the topological features are left to be administered by the program. Both the designer and the applications program can edit the model on the file and each encounter with the model may therefore present the designer with a picture he has not seen before. Certain features in it which hampered the program are brought to the attention of the designer by being marked in some way. It is possible to display these parts at increased brightness, or to make them blink on and off, or some such thing. In general, layout involves patterns which overlay each other. Layout of a printed circuit package, for example, may involve two or more layers of conductor patterns with an array of components superimposed. When displayed simultaneously on the screen these patterns may sometimes become too jumbled to be comprehensible to the designer, and in this case it may be desirable to regard each pattern temporarily as a separate entity. Any or all of the patterns may then be displayed independently of each other to aid the designer's appreciation of the situation.

5.3 *Logic Design*

Interplay between computer and the designer is rather similar to that in circuit design. However, the logic system being handled may be extremely large (compared with an electronic circuit); perhaps many thousands of elements. Several designers may work in succession on different parts of the system and they will not necessarily be aware of one another's work in detail, although the interface between the different parts is well defined. The computer can help to reduce the confusion. When, for example, a designer sends a logic simulator program to work on his subsystem, the program can retrieve all relevant sections of logic from the file, irrespective of who designed them. The linking between the different sections can be automatically updated as the sections are edited so that the requirement for one designer to know the status of another's work is minimized.

A complete picture-editing system oriented to a logic design application has been developed at the author's laboratory. The screen serves as a window on to a large drawing board and can be positioned anywhere on it. Logic elements and connecting wires can be put together in the window while the program builds a corresponding structural model. The picture, and hence the structure, may be edited freely, wires and logic elements may be labelled and models may be sent to and from the program. Simple filing facilities are being planned and a logic simulator will be added later.

6. Communications with other Plant Activities

Design of a product is but one of many activities required to manufacture and sell it. Once the design office starts doing its work with computer assistance it is important to enquire

(a) whether any of the other plant activities are affected;

(b) what the computer could do for them.

Planning, scheduling, production, ordering, invoicing, accounting, and other departments may all be affected —and not necessarily for the better unless proper care is taken. Computer-aided management, administration, production are topics beyond the scope of this paper (and the experience of the author!) but some general comments will not seem out of place.

At the end of a design phase the file will contain a description of a 'working model' corresponding in some respects with a physical prototype for the product. From this model it should be possible to extract a certain amount of manufacturing and documentation information directly. Stock control programs can take possession of the model at some stage to prepare ordering and scheduling plans. In the case of a large electronic assembly (e.g. a computing system), wire lists, waveform dictionaries and fault characteristics can be prepared to assist in commissioning and maintaining the equipment. Record keeping itself can be regarded as a design activity and computer-aided documentation may bring substantial benefits to all departments in the plant.

7. Acknowledgments

The author acknowledges the unpublished work of his colleagues at Cambridge. The circuit design project is being carried out enthusiastically by Joe Hiles and an experimental real-time program for it has been written by Steven Parry. The logic design project is in the capable hands of Philip Cross. His own ring structure processor owes a lot to the ideas of Crispin Gray and his data structure package ASP.

8. References

1. D. G. Borrow and B. Raphael, 'A comparison of list-processing computer languages', *Commun. Assoc. Computing Machinery*, **7**, April 1964.
2. S. H. Chasen, 'The introduction of man-computer graphics into the aerospace industry', AFIPS Conference Proceedings, **27**, 1965 Fall Joint Computer Conference (Spartan Books, Inc., 1965).
3. S. A. Coons, 'An outline of the requirements for a computer-aided design system', AFIPS Conference Proceedings, Spring Joint Computer Conference, 1963.
4. M. R. Davis and T. O. Ellis, 'The RAND tablet: a man-machine graphical communication device', AFIPS Conference Proceedings, Fall Joint Computer Conference 1964.
5. M. J. Dertouzos, 'CIRCAL; On-line circuit design', *Proc. Inst. Elect. Electronics Engrs*, **55**, pp. 637–54, May 1967.
6. E. L. Jacks, 'A laboratory for the study of graphical man-machine communications', AFIPS Conference Proceedings, Fall Joint Computer Conference 1964.
7. J. Katzenelson, 'AEDNET: A simulation for non-linear networks', *Proc. Inst. Elect. Electronics Engrs*, **54**, pp. 1536–52, November 1966.
8. J. S. Koford, P. R. Strickland, G. A. Sporzynski and E. M. Haubacher, 'Using a graphic data processing system to design artwork for manufacturing hybrid circuits', AFIPS Conference Proceedings, Fall Joint Computer Conference, **29**, pp. 229–46, November 1966.
9. J. McCarthy, 'Recursive functions of symbolic expressions and their computation by machine, Pt. 1', *J. Assoc. Computing Machinery*, **3**, No. 4, p. 84, 1960.
10. D. E. Rippy, D. E. Humphies and J. A. Cunningham, 'Magic: a machine for automatic graphics interface to a computer', AFIPS Conference Proceedings, **27**, 1965 Fall Joint Computer Conference (Spartan Books, Inc., 1965).
11. L. G. Roberts, 'Graphical communications and control languages, Lincoln Laboratory', M.I.T. Reprints MS1173, November 1964.
12. D. T. Ross and J. E. Rodrigues, 'Theoretical foundations for the computer-aided design system', AFIPS Conference Proceedings, Spring Joint Computer Conference 1963.
13. I. E. Sutherland, 'Sketchpad, a man-machine communication system', AFIPS Conference Proceedings, Spring Joint Computer Conference 1963.
14. R. Waxman, H. T. McMahon, B. J. Crawford and A. B. De Andrade, 'Automated logic design applicable to integrated circuit technology', AFIPS Conference Proceedings, **29**, pp. 247–65, November 1966.
15. J. Weizenbaum, 'Symmetric list processor', *Commun. Assoc. Computing Machinery*, **6**, September 1963.
16. M. V. Wilkes, 'An experiment with a self-compiling compiler for a simple list processing language', University Mathematical Laboratory, Cambridge. Tech. Memo 63/1.
17. N. E. Wiseman, 'Application of list processing methods to the design of interconnections for a fast logic system', *Computer J.*, **6**, January 1964.
18. N. E. Wiseman, 'A simple list processing package for the PDP7', Proc. 2nd European Seminar, Digital Equipment Corporation Users Society, 1966.

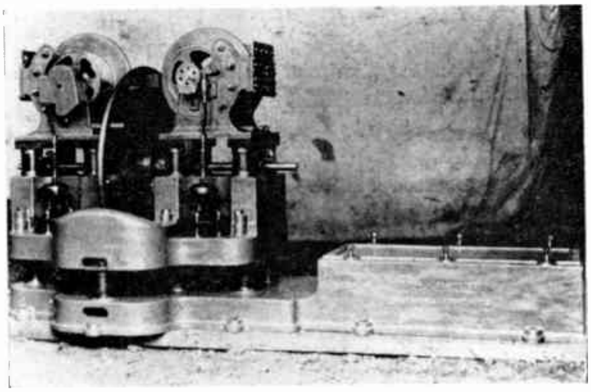
Manuscript received by the Institution on 19th May 1967. (*Paper No. 1146/CG 7.*)

Radio Engineering Sixty Years Ago

An interesting instance of 'sub-contracting' in the early days of radio by the old-established engineering company of Armfield's of Ringwood, Hampshire, is described in a recent article in *Industrial Archaeology*†.

It appears that Marconi had despaired of getting made in England the special rotary-disk spark dischargers for his new transmitter being installed at Clifden in Ireland in 1907, and was on his way to Poole to take a boat to France when, passing through Ringwood, he happened to notice the Vale of Avon Iron Works. He met and discussed the problem with Joseph Armfield who agreed to attempt the making of the high-speed steel disk with studs which was an essential part of the apparatus.

The rotary-disk discharger essentially consisted of a metal disk rotated at high speed by means of a central shaft mounted between end bearings. The periphery of the wheel carried a number of metal studs equally spaced circumferentially and radially. This circle of studs passed between two stationary electrodes and in so doing momentarily reduced the air-gap between the electrodes thus causing a highly charged capacitor connected across the gap to discharge. The capacitor was recharged in the interval before the next stud closed the gap to repeat the process. There are various refinements of the process and the apparatus, and there were means whereby the air-gap could be adjusted by hand.



Copyright: D. A. E. Cross

Rotary spark-discharge apparatus made by J. J. Armfield for Marconi in 1907.

The speed of the disk was such that the number of studs passing the gap per second bore a direct relation to the frequency of the alternator which provided the charging current, and also a sub-harmonic relation with the natural oscillation frequency of the associated tuned circuit and aerial system.

In an earlier version of the rotary-disk spark discharger a belt-driven disk was used in which the timing of the studs was not locked to an a.c. source of power as in this case, which is a small but fairly advanced type.‡

Marconi stayed in a Poole hotel while the work was in progress, and returned to Ringwood for the tests of the disk and its bearings. The disk had to be run at a very high speed and was coupled by belting to a stationary steam engine at the works. The illustration of the completed apparatus proudly bearing the cast inscription on its base, 'Joseph J. Armfield & Company, Engineers, Ringwood' was recently found amongst the records of the Company, together with other photographs of the disk. In one of these, Marconi himself (second from right) is seen watching the test.

Copyright: D. A. E. Cross

G. Marconi watching test of rotary spark-discharger disk at Armfield's Works, Ringwood, Hampshire, 1907.

There are no documentary records held by the Marconi Company to establish this sub-contracting of the manufacture to Armfield, but a book of reference numbers§ to the piles of technical drawings found in the Ringwood loft quotes, in manuscript under 'Miscellaneous drawings, Mark X' that Drawings Nos. 418 to 423 are 'Marconi-Discs'. Unfortunately, in searching through the pile of drawings, it was found that, of all the drawings, only these most fascinating records had been at some time removed and were missing. Such are the frustrations of the industrial historian!

---

† Donald A. E. Cross, 'Armfields of Ringwood', *Industrial Archaeology*, 4, No. 2, pp. 118–120, 161–4, 1967. (The present note is based on Mr. Cross's original article with additional material also contributed by him.)

‡ Letters from the Marconi historian to Mr. Cross.

§ In Mr. Cross's possession.

U.D.C. 681.39 : 621.395

A Multi-parameter Data Recording and Processing System using Magnetic Tape

By

J. SEARS,
C.Eng., A.M.I.E.R.E.†

Summary: This paper describes a flexible digital data recording and recovery system which will accept information of up to 76 bits per event at a rate variable from some 375 to some 48×10^3 bits per second (b/s), and replay at a rate of from 3×10^3 to 48×10^3 b/s. The total tape storage capacity is some 55×10^7 bits. The number of parameters recorded is limited only by the accuracy required and by the total event size (in relation to the maximum number of bits per event), provision is made within the equipment for correlation determination between selected parameters and for the selective examination of recorded data. The tape transport control system permits remote control which facilitates its integration into a computer complex.

The system is illustrated by examples of its application to the acquisition of data from nuclear experiments.

1. Introduction

In the paper by Wells *et al.*[1] virtues advanced for the direct magnetic tape storage of nuclear experimental data include its large information capacity and compatibility with the data rates involved. Also the digital form of data recording can meet the accuracy demands of the experiments and is compatible with following processing equipment. There is in addition a reduction of overall complexity of the experimental instrumentation, with a consequent increase in reliability, by the separation of measurement and computation. The role of the tape recorder is now largely a supporting one to that of the 'on-line' computer. It does, however, still provide a valuable safeguard during the actual experiment particularly where the subject is of a unique event, and is of value in preliminary setting up work[2] and as a source of typical data thereby relieving the pressure on the availability time of the experimental signal source; a significant economy in the case of particle accelerators. An economy of analysis time may also be obtained by replaying the data in a shorter time than that used for recording. The existence of a permanent record also enables the experimenter to have 'second thoughts' without the necessity of actually repeating the experiment; a further field of application should also be in making available real experimental data for instructional purposes. The choice of a digital system permits great freedom and flexibility in the selection of data channels, and the potential accuracy of associated equipment, detectors, converters and analysers may be more fully realized than with analogue recording systems.[3, 4]

2. Application

One example of the application of the system which illustrates some of its features is in the recording of information from a nuclear reaction experiment in which information is available at the same time from a number of reaction products and where it is required to be able to separate this information and to reject spurious signals.

Such an experiment[2, 5] is that in which nuclei are bombarded with medium energy (30 to 50 MeV) protons. A number of particles having differing mass and/or charge result, the parameters recorded being charge, total energy and time of flight. The system described enables the particles to be separated on a charge basis and for much of the unwanted background to be removed, the energy and time of flight information is then available for the computation of individual energy spectra separated by mass determination. Thus Fig. 1 shows the 'raw' recorded information while Fig. 2 is the same, subject to charge selection in which the uncharged particle information has been substantially rejected. Again Fig. 3 shows the selection of singly-charged particles only, and Fig. 4 shows the selection of doubly-charged particles only. In these figures the photographic density is an indication of total numbers of particles or events for a given combination of values of E and T so that, for example in Fig. 3, if one considered a particular value of T, and plotted event frequency as a function of E, three peaks would occur corresponding to protons, deuterons and tritons, this would then constitute a spectrum of particle density as a function of energy, the spectrum width corresponding to the range of energy considered.

The correlating facility then enables a user to select one variable as a function of another (in the example

---

† Formerly at the Atomic Energy Research Establishment, Harwell; now with the Science Research Council, London.

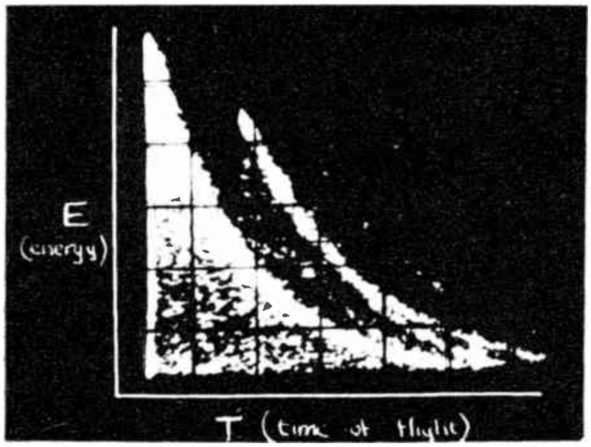


Fig. 1. All recorded events.

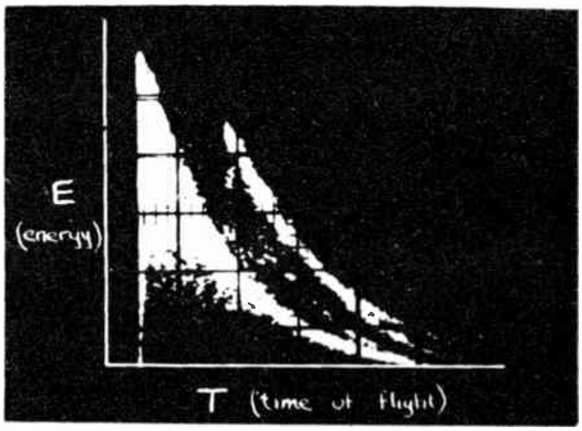


Fig. 2. All recorded data except uncharged particles.

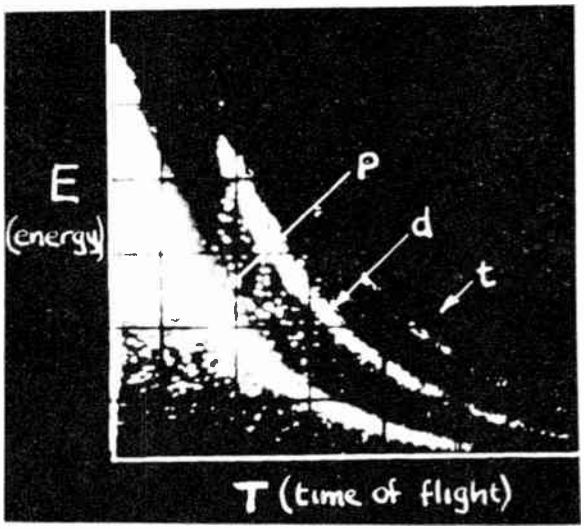


Fig. 3. Singly-charged particles only, selected by intermediate size dE pulses.

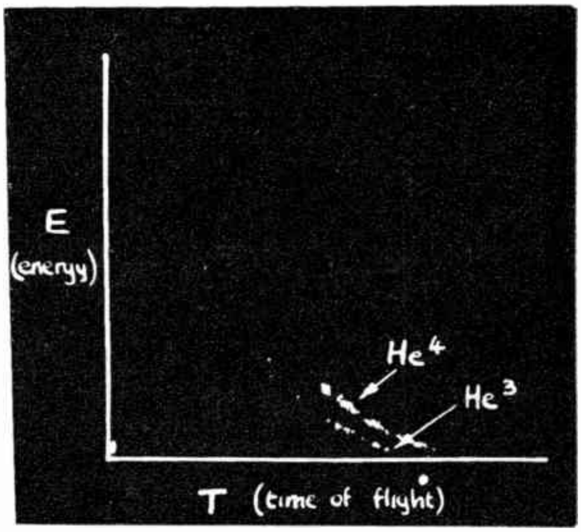


Fig. 4. Doubly-charged particles only, selected by large dE pulses.

used as illustration as a function of charge) and for the cases of charge being below, within or above specified limits, each selection condition requiring one pass of the tape to recover the required information. Although the example showed only one selection condition specified, the equipment in fact makes provision for a double limitation if required. It is further clear in general that by associating specified parameters while recording, the variables may be recovered selectively from the whole recorded ensemble.

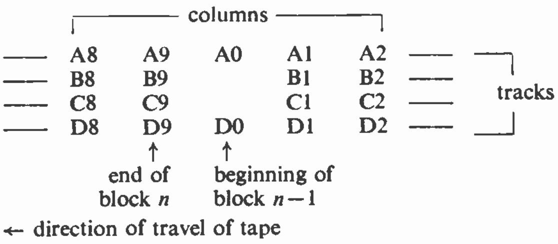
A further facility is available in the capacity of the equipment to manipulate the recorded data; occasions may arise, for example, where the width of a recorded spectrum is greater than that of the available analyser, or where it is required to examine in detail only certain ranges. In the former case it is possible to compress the spectrum and in the latter case selection conditions may be imposed, by combining these two features it is possible to firstly select ranges of interest by compressing a spectrum for a general examination and then to select and analyse areas of interest to a degree limited only by the resolution of the input data.

3. General Description of the Equipment

The information from nuclear reactions is usually of a statistical nature so that some loss, provided that it is not systematic, is acceptable; it is, however, very desirable and sometimes essential to reduce false signals, or background, to a minimum, so that while occasional events may be randomly lost, numerical errors must be minimized. Again, the time of arrival of information is often uncertain but for the economic use of data processing and analysing equipment a smooth information rate is desirable so that some degree of 'de-randomizing' is helpful.

The method used to realize these and the other features earlier described is now outlined in Fig. 5. As may be seen, the system comprises essentially two-stage input and output stores coupled by a permanent tape record; tape recorded on one equipment may be replayed on another.

Before describing the record and replay systems it is convenient to consider the tape record itself. To provide maximum flexibility the equipment was designed to be usable with a minimum tape width of a quarter of an inch, using standard four-track heads; with this limitation a track could not be afforded to provide 'clock' information, nor could space be provided for checks adequate to meet the required criterion for the limit of false information. This situation was met by using return-to-zero recording[6] which provides three states, 'positive' magnetization corresponding to a logical '1', 'negative' magnetization to a logical '0', and absence of magnetization to a fault condition. By assembling the recorded bits to form a block, the start of the block being identified by omission of the bits on the two centre tracks, the tape record format is as follows:



A8	A9	A0	A1	A2
B8	B9		B1	B2
C8	C9		C1	C2
D8	D9	D0	D1	D2

The number of columns comprising a block may be determined by adjustment of the recording system to be between 2 and 10, the size being determined by the experimental requirements. On replay, the number of columns from one 'start of block' ident (B0 and C0 missing and A0 and D0 present, as '1's or '0's) to the next must correspond to the predetermined number, if not the block is rejected; a 'clock' is generated from signal present coincidence ('1's or '0's) outputs from the four bits comprising each column after the first.

3.1 *The Record System*

The recording system consists of the input store which accepts the experimental data, followed by the writing store from which information is transferred sequentially to the tape via the recording head of the tape transport. The data transfer is under a logical control which orders the sequence of events in response to the Start and Write commands which must respectively precede and follow the input data.

Both the input and writing stores correspond physically bit for bit to the information assembled on the tape. Thus if bistable element B1 in the input store is in a '1' state bit position B1 on the corresponding tape block will become 'positively' magnetized.

The input store consists of 38 identical independent bistable elements which may be fed in parallel or connected by means of patch leads in series to form scalers. Access to these input store elements for purposes of interconnection is by means of a patch-board physically disposed in facsimile of the tape record format. The minimum number of bits usable is six, corresponding to a format as:

A0	A1
	B1
	C1
D0	D1

and the number may be increased in steps of one column up to a total of ten columns. Parallel and serial feeds may be used simultaneously if required, for example in the minimum format above elements A0, A1, B1 and C1, could be connected serially to form a four-bit scaler while the remaining elements D0 and D1, could be fed independently in parallel or be patched to form a separate two-bit scaler; any other arrangement is possible.

On the arrival of a Write command, and provided the contents of the writing store have been transferred to the tape, the contents of the input store are transferred in parallel to the writing store and at the same time the generation of a pulse sequence is started by the control logic which transfers the contents of the writing store column by column to the tape, the rate of transfer being in sympathy with the tape speed. The number of sequential transfer pulses corresponds to the selected block length. As soon as the contents of the input store have been transferred to the writing store, it is available for a further input, but if, on the completion of this input, transfer from the writing store to the tape has not been completed, further Start commands are rejected and an output is delivered which may be used to inhibit data inputs until the input store is again empty. The proportion of data signals lost is displayed so that the tape speed may be adjusted relative to the input data rate to obtain an acceptable loss figure.

3.2 *The Replay System*

The information on the tape is recovered by effectively reversing the recording procedure; in addition the replay system includes data block recognition and error detection and rejection logic, and the logic for conditional output.

At certain tape speeds recording and replay are simultaneously possible so that the performance of the

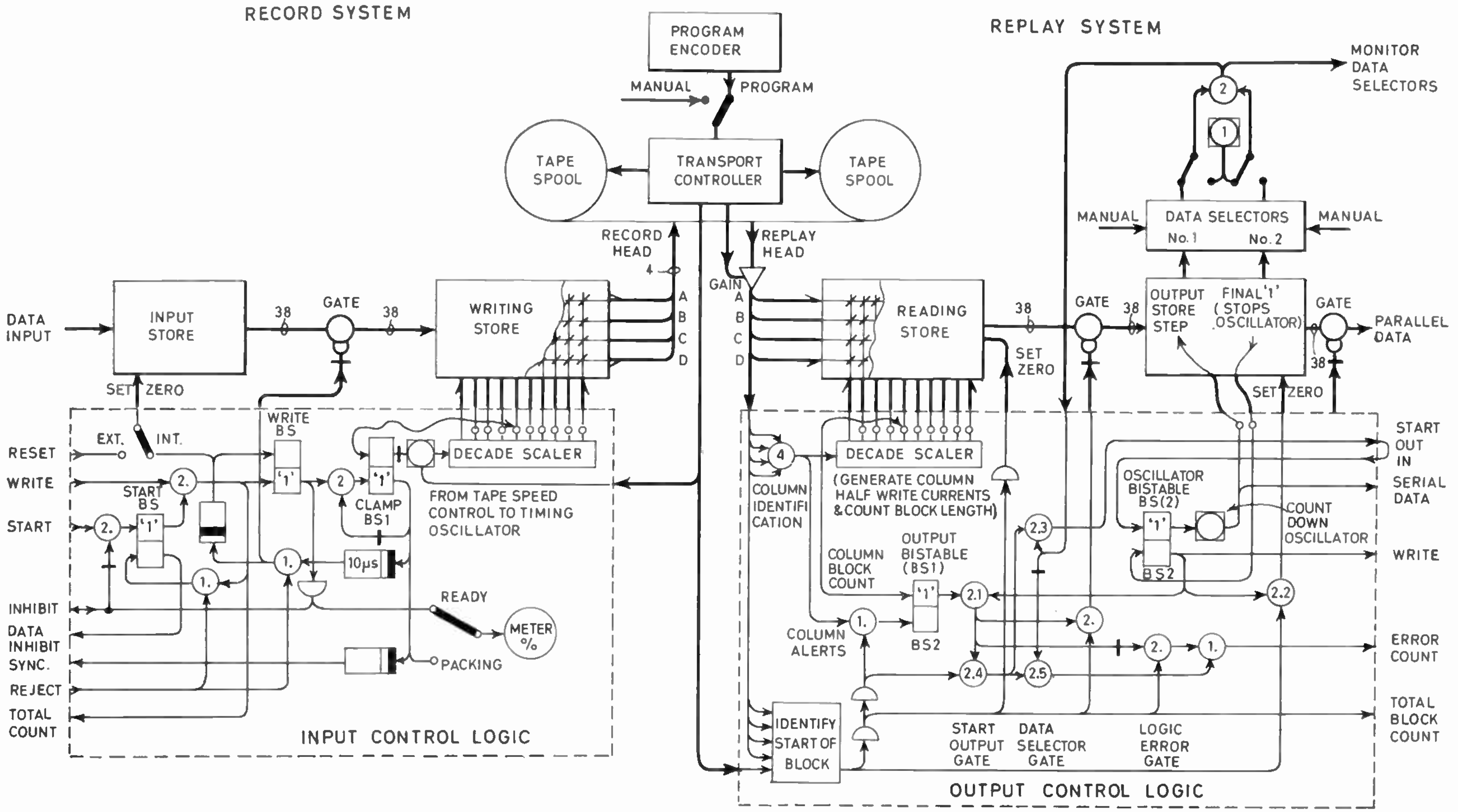


Fig. 5. Record and replay system.

whole equipment and the progress of the experiment may be monitored during operation but in this mode the replayed data suffers some deterioration although the quality of the recorded data is unaffected.

Imperfections resulting in signal loss may occur due to an adjustment of the replay system which, while the bulk of signals are accommodated, permits occasional mismatch between the data rate from the tape and the output data rate, and also to the drop out of signals from the tape due to mechanical imperfections. In both cases information is lost but no numerical errors are introduced.

3.3 *The Tape Transports*

The tape transport used so far with the equipment has carried a $\frac{1}{4}$-in tape providing a total storage of some 5×10^7 bits at 400 bits/in. This relatively cheap transport provided a choice of recording speed of from $\frac{15}{16}$ in/s to 30 in/s in binary ratios, and replay speeds of from $7\frac{1}{2}$ to 30 in/s. To extend the range and flexibility of the system, however, a design study has been made for its use with the Harwell 1-in tape transport to provide a total capacity of some 55×10^7 bits at 200 bits/in and a speed range from $\frac{15}{16}$ in/s upwards in binary ratios to 60 in/s on recording, and from $7\frac{1}{2}$ to 60 in/s on replay.

4. Functional Description of the Sub-sections

4.1 *The Record System*

A somewhat simplified version of the record system is shown in Fig. 5. The stores and logic are implemented by conventional techniques. The input store provides, using transistor bistables, fast acceptance of input data and the writing store uses magnetic cores as a matter of circuit convenience, the cores combining within themselves the characteristics of bistable elements and the OR outputs which are demanded of each row of cores associated with a particular track on the tape.

The conditions for the acceptance of input information are now considered in rather more detail. The first requirement is that the input shall comprise a complete ensemble and to define this it is punctuated by two logic command signals, the Start heralding, and the Write concluding; these must both be present and arrive in the correct order. Secondly, the input store must be empty. It may be noted that if an ensemble is accepted on these conditions the completion of its acceptance cycle (i.e. the implementation of the Write command) may be delayed if the writing store transfer to tape cycle is incomplete. The mode of operation is shown in Fig. 6 in which the logical states appear in their related time sequences. As previously stated the contents of the input store are transferred in parallel to the writing store and the contents of the writing store are transferred in successive columns to the tape. The oscillator clamp bistable BS1 provides the inter-store gating control, zero set to the input store, and starts the timing oscillator. The timing oscillator drives the decade scaler which in its turn empties the writing store column by column, the scaler output corresponding to the last column of a block being used to reset bistable BS1 and thus to terminate the data writing cycle. The oscillator frequency is in sympathy with the tape speed so that the number of bits per inch recorded on the tape is constant.

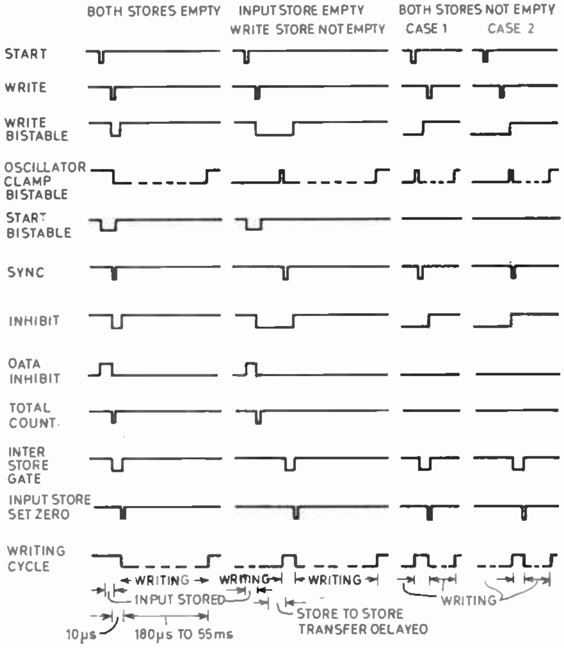


Fig. 6. Record unit logic sequences.

In addition to meeting the minimum control requirements set out above the logic elements provide certain other information and control facilities, notably Sync., Reject, Inhibit, Data Inhibit and Total Count. These are also shown as time functions in Fig. 6. Reject may be used for an interrupt termination of an input cycle and behaves like the Write command but without any transfer from the input to the writing store. Inhibit indicates when a Start cannot be accepted either during the store to store transfer period or because transfer (and hence clearance of the input store) cannot be accepted due to the incompletion of transfer from the writing store to the tape. Choice of External or Internal reset determines whether the input store is zero-set as part of the data transfer cycle or not, if not, and if no external reset is applied successive data cycles will be superimposed in the input store. For serial inputs these would be added numerically.

4.2 *The Replay System*

4.2.1 General logics

The control requirements are determined firstly by the need to identify valid information ensembles, and secondly by the states of the stores. As stated, the validity of ensembles of information is determined by comparing the length of the blocks delivered with their length as determined by the record system. The logical requirement for this is then the detection of the 'start of block code', the counting of the number of complete columns from this to the next block identity output, and then comparing this number with the correct value, accepting or rejecting the block accordingly.

Turning now to the general replay logic diagram (Fig. 5) the responses to the various input conditions will be considered. Assuming firstly that the blocks of information are fault free the first two logic sequences of Fig. 7 show the principal state/time relations. If both stores are empty, it can be seen that at the time of arrival of a block identification signal the output bistable will be in the '1' state having been set by the output of the column counter corresponding in position to the known column length of the block, and that a '1' will be present at the output of AND element 2.1.† Now following the block identification output, firstly a zero set signal is applied to the output store to ensure that it is empty, secondly the inter-store gate is opened permitting store to store transfer, at the same time a Total Count signal output is delivered, and shortly afterwards the reading store is reset to zero, thirdly an output cycle (parallel and/or serial) is initiated for the output store, and finally the output bistable is set zero; this sequence of events occupies some 5 μs as indicated relatively in Fig. 7. If the output store is not empty the behaviour of the control depends on whether a parallel and/or serial output mode is being employed. If a serial output is incomplete at the time of arrival of the start of block ident signal, examination of the logic diagram shows that the AND elements 2.1 and 2.2 are inhibited by the count-down oscillator bistable (BS2) which controls the serial read-out; in this case the serial read-out cycle will continue, inter-store transfer is inhibited, the contents of the reading store are destroyed and in addition to the total count output an error count output signal is generated indicating the loss of a block of information.

Column count error is most frequently due to tape imperfections causing signal loss, but skewing of the tape relative to the replay head can produce bit misalignment and consequent count, or more seriously, number error. A skew detection feature is included in the start of block identification circuit which produces shutdown if the skew becomes excessive.

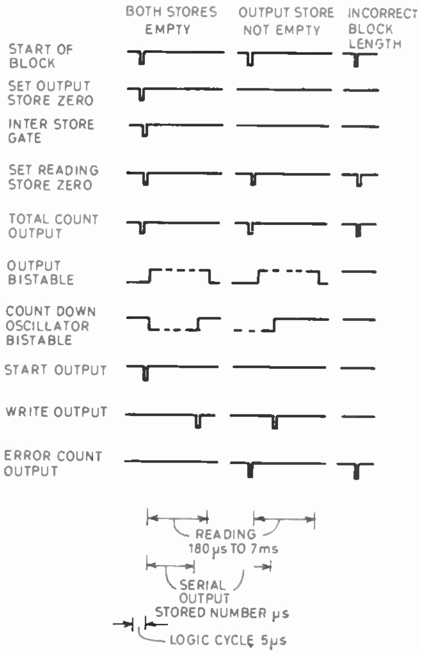


Fig. 7. Replay unit logic sequences.

The reading store like the writing store on the record side uses magnetic cores which now conveniently combine bistable and AND characteristics, each row of cores (Fig. 5) being fed from one tape track respectively. The column counter is formed by a decade scaler, each output corresponding to a column and providing a half-magnetization current to its column to 'AND' with half-magnetization currents from the tape signals, absence of such signals leaving the cores concerned in the reset state. Transfer of the contents of the reading store to the output store is in parallel, each core coupling to its corresponding output store bistable via a gating element, all these elements being enabled simultaneously. To facilitate serial read-out the physical disposition of the binaries is similar to that adopted on the record side, their step and carry connections being brought out to a patch board. This permits interconnection of the binaries to correspond appropriately to their equivalent elements in the input store. Now the output store contains the complement of the input store so that the serial output is recovered by applying a pulse train to the least significant element of the chosen group of binaries and using the carry output of the most significant to terminate it. The number of oscillator pulses is then the '1's complement of the number stored, that is the input number plus one.

---

† Logic elements are denoted as follows:

AND	N*n*	Pulse generators	P*n*
OR	1*n*	Delays	D*n*
Bistables	BS*n*	Inverters	I*n*
Binaries	BN*n*		

4.2.2 Conditional data read-out

Data output dependence is controlled by AND elements 2.3 and 2.5, via 2.3 when the data are to be put out, or via 2.5 enabling an error count output when the data is rejected. Two selectors are available, each providing an independent choice of upper and lower limit numbers, their values being manually selected by binarily weighted switches (providing a range of ten bits for each limit). Connection from the binaries of the output store to the data selectors is by means of patch leads so that up to 20 binaries selected in any way may be associated with them, a maximum of 10 to each selector. The logical arrangement of the data selectors is shown in Fig. 5 from which it can be seen that the selection may be dependent on either, neither or both selectors. These are described in detail elsewhere.[7]

4.2.3 Spectrum manipulation

Occasions may arise where the width of a recorded spectrum is greater than that of the available analyser, or where it is required to examine in detail only certain ranges.

Where the whole spectrum is to be observed on an analyser of restricted range the width may be effectively reduced by delivering only the output corresponding to the more significant bits of the output store, the number of bits selected being such that their total capacity is equal to or less than that of the largest analyser address. For the parallel output case this means merely the selection of the appropriate number of more significant bit outputs, for serial output the final carry (from the most significant binary) is used (as before) to stop the oscillator, but the oscillator output, previously applied to the least significant binary of the chosen group, is now applied to one of an appropriately higher-order so that again the highest count number is within the limits of the associated equipment. The effect of these operations is that while the total count stored in the analyser remains the same the width is compressed by a factor equal to the binary power of the lower order bits discarded.

While this procedure enables the whole spectrum to be dealt with, resolution is lost. It is possible, however, to analyse any part or the whole of the spectrum to the full resolution of the information store set up in the system, even with an analyser of restricted range. The method is to restrict the range of numbers read out to be equal to or less than the range of the analyser by using the data selective output provision to define the width and position of the acceptable range. This permits the examination step by step (each step requiring one pass of the tape) of a spectrum, or the selection for full resolution of a particular range of interest.

4.3 *The Tape Control System*

The tape transport control system now described is intended to be associated with the Harwell 1-in tape transport type 1824 to enable the record and replay systems to be associated with it either as a single pair providing up to 38 bits of information per event or as

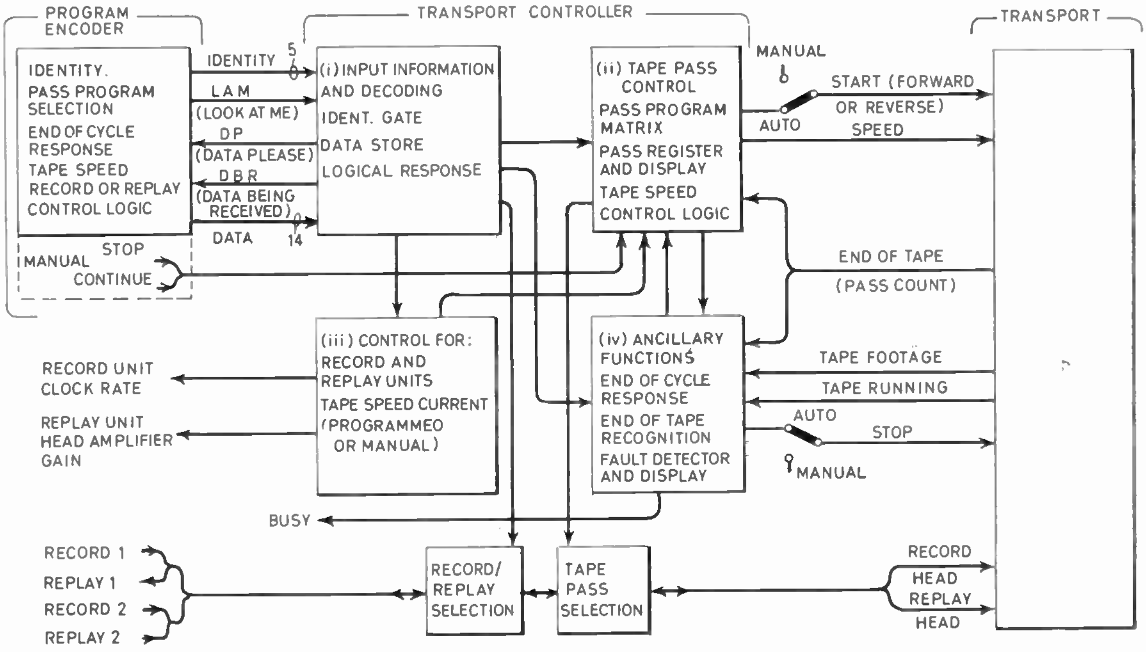


Fig. 8. 1-inch tape transport control system.

a double pair giving a maximum event capacity of 76 bits, although the user may wish to use some of the available capacity for cross-identification coding rather than for data.

As the same data format is proposed for the 1-in tape as was used for the ¼-in, the 1-in tape is therefore to be utilized in ¼-in strips, these being used in four complete sequential passes, or in pairs, or in any order starting and stopping at the beginning or end of any strip. The pass order may be subject to a program set into the controller and generated by a separate program unit or from a computer, the program may be interrupted and continued or returned to start, or may be repeated once or cycled. Manual control may be applied and a detector is provided which guards against tape breakage and other transport faults and gives automatic shut down and alarm. The design is intended to be largely implemented with Harwell 5000 series logic elements.[8]

The system is illustrated schematically in Fig. 8. It falls into two principal divisions, the program encoder, and the transport controller. The program encoder (Figs. 9 and 10) enables the predetermination of a sequence of events and conditions for the transport, the selection of passes and repeats, tape speed and record or replay mode, and presents them to the transport controller in the form of a 14-bit binary code together with an identifying code which permits the use of a common command link for several controlled equipments. With this information the program encoder supplies and receives signals permitting it to be associated with a computer, i.e. to supply its program to a computer for inclusion in a wider field of control, the computer would then issue the program as required to the transport controllers. The transport controller (Figs. 10 and 11) may be used for manual control of the transport or automatic control in response to the encoded information, this information could be accepted and stored subject, if required, to the presence of the identifying code, leaving the encoder free to meet further requirements.

Treating now these two components in more detail the tape transport controller will first be considered. The controller is to comprise a number of sub-sections as set out in block schematic form in Fig. 8. The programmed control is encoded in the 14-bit binary word and the optional identifying code in the 5-bit word. To facilitate computer control LAM ('Look at Me'), DP ('Data Present') and DBR ('Data Being Received') connections are also available.

Functionally the controller may be considered under four headings:

(i) Input information and decoding
(ii) Tape pass control
(iii) Control for the record and replay units
(iv) Ancillary functions.

Under (i) appear the input stores for the codes carrying the tape pass program, the end of cycle response, record or replay selection, and tape speed, together with the identifying code recognition comparator and inputs and outputs for computer dialogue.

(ii) includes the program control matrix, control logic, state indication (display), and tape speed.

(iii) determines the operating conditions for the record and replay units, the clock rate of the record unit appropriate to the tape speed, the replay head output amplifier gain for the replay unit, again as a function of tape speed, and the routing of the head signals appropriate to record or replay modes and single or double unit working. These provisions are shown diagrammatically in Fig. 11.

(iv) comprises ancillary functions as end of cycle response, end of tape recognition, and fault detector (broken tape) shutdown and alarm.

Referring again to Fig. 11, the program control matrix determines the order and number of tape passes while the control logic implements its commands. The program matrix is essentially a rectangular interconnecting patchboard, that on the program unit being set up manually while that on the controller follows it electronically to form a facsimile. The four by four part of the matrix on the control unit accepts inputs from the pass register corresponding in sequential order to the (maximum of) four tape passes per complete cycle and delivers outputs which may summon any of the four ¼-in strips. A fifth output is available to select the end of cycle position. Thus when the program is started the first input line is alerted, and by suitable setting of the matrix patching it is connected to one of the four possible outputs depending on the choice of the first tape strip required. At the conclusion of the pass the second input line is alerted and one of the three remaining strips is run or the end of cycle is initiated.

The functioning of the control logic is conveniently illustrated by considering the response to various programs. Thus, if a single record unit was used to fill the whole tape, the most likely mode would be to record strip A with the tape running in a 'forward' direction followed by strip (pass) B in a 'reverse' direction and so on, concluding with a stop command at the completion of pass C. While the passes may be recorded and replayed in any order it should be noted that the arbitrary definition of direction will always be followed, i.e. for passes A and C the tape is running 'forward' and for passes B and D in 'reverse'.

Now suppose that the tape transport is quiescent with the tape positioned so that the end of tape lights are on at the beginning of pass A, and that the program calls for A–B–C–D–Stop. Data available on the program encoder (Figs. 9 and 10) is alerted manually

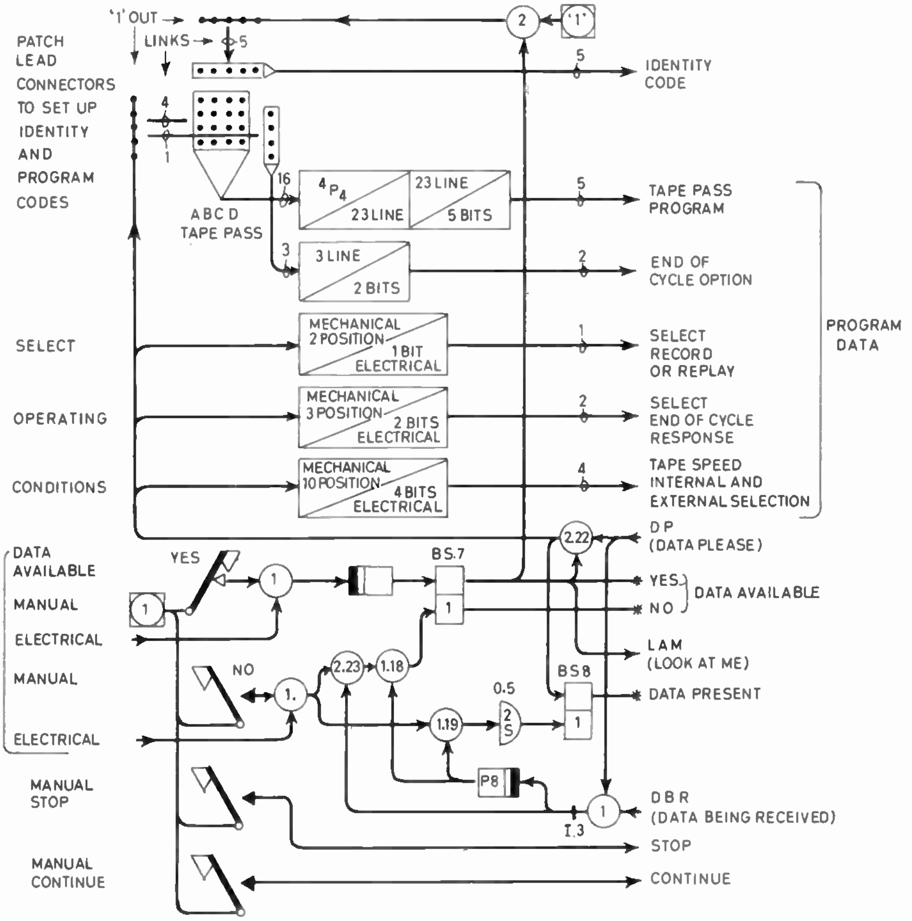


Fig. 9. Program encoder.

or by application of a pulse input so that an identity code is transmitted which, if it is accepted by the identity comparator, enables the LAM signal which is also generated in the program encoder to be accepted via AND element 2.1 (Fig. 11). The access sequence is shown in Fig. 10, i.e. the control logic is set to an initial state via P1 as indicated on the diagram. Stop is transmitted to the tape transport from P1 via 1.14 and 1.18, the input gate (2.2) to the program store is opened, program data are requested from the program encoder, and a DBR signal is sent to the program encoder until its data output has been absorbed, the DBR signal releases the program encoder which removes LAM and a Start signal is generated from its back edge via P2. The pass register has been initially set to zero so that the event sequence line 1 of the program matrix is alerted, also the decoded pass order program has alerted event-sequence/tape-pass-command-output-lines intersections in the matrix as

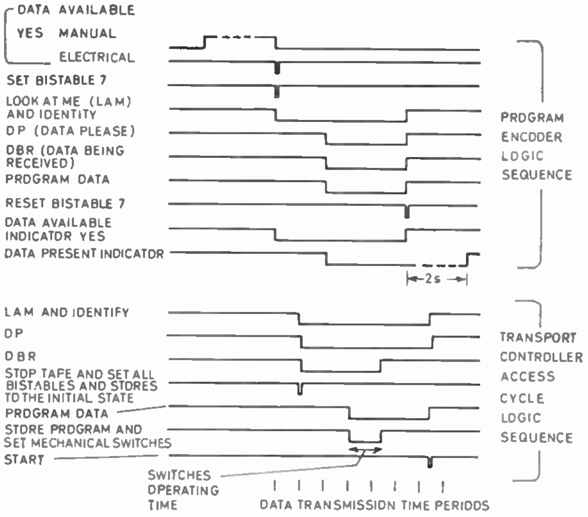


Fig. 10. Tape control access logic sequence.

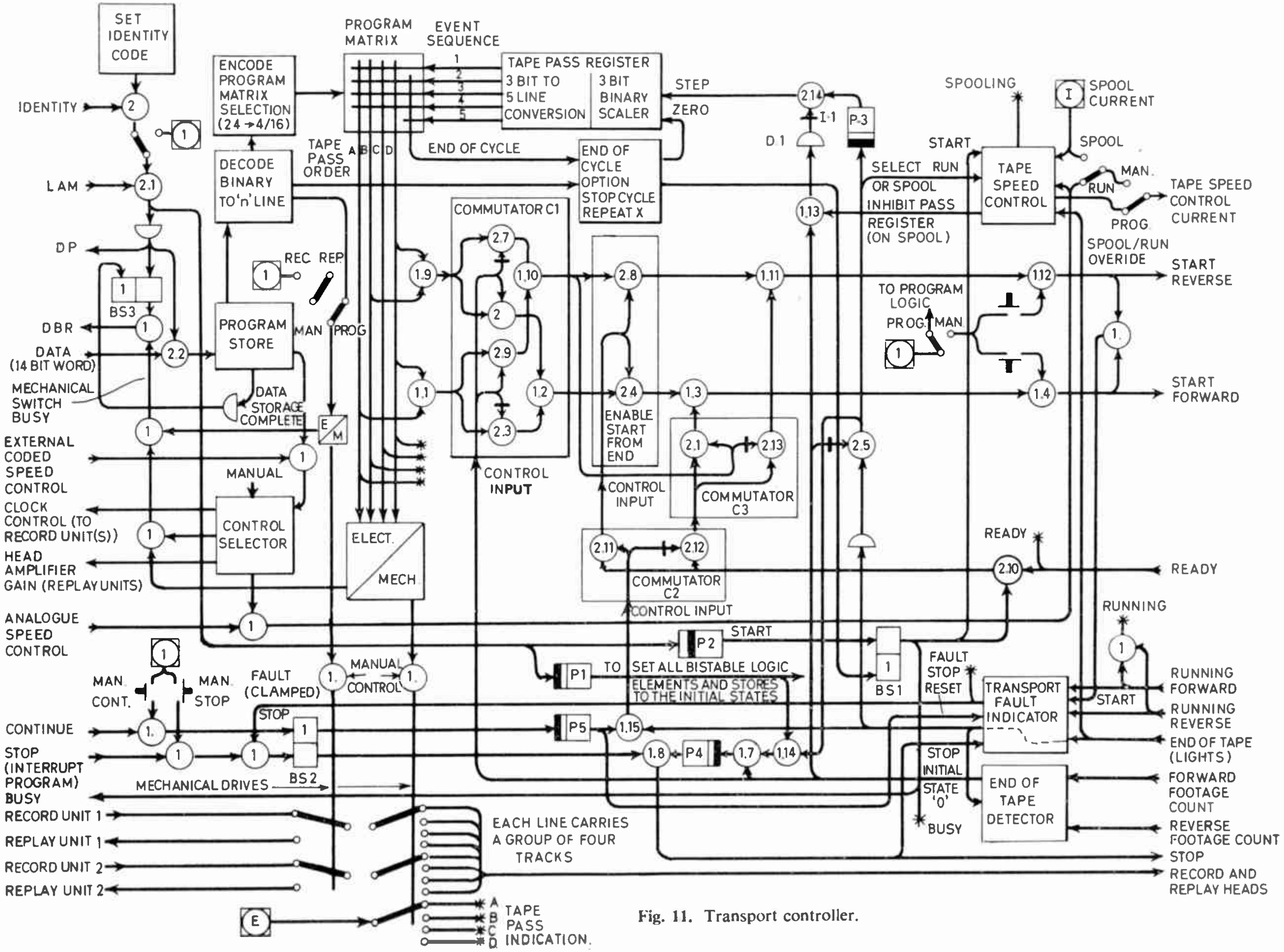


Fig. 11. Transport controller.

A1, B2, C3, D4, End of Cycle 5, so that command output A is in the '1' state, the remaining command outputs being zero. Ready and End of Tape inputs from the transport are '1' so that on the start command (from bistable BS1) the tape speed control unit is alerted and a start forward command is delivered to the transport Start Forward input via elements 1.1, 2.3, 1.2, 2.4, 1.3 and 1.4. The tape then starts forward, the End of Tape and Ready indicators go '0'. It should be noted that the initial state setting action sets the output of the End of Tape Detector '0' so that a '1' input is present (via 2.5) at the 'run or spool' input of the tape speed control, setting it to the 'run' condition. When the end of the pass is reached the end of tape signal reappears and causes the event sequence to shift from line 1 to line 2 (via 2.5, P3, 2.14 and the pass register), a stop command is delivered to the transport via 2.5, 1.14, 1.7, P3 and 1.8. As soon as the Ready signal from the transport reappears the matrix output, now on command line B, develops a start reverse instruction via 1.9, 2.7, 1.10, 2.8, 1.11 and 1.12. The cycle then proceeds similarly until event sequence line 5 alerts the end of cycle command.

The operation may be similarly considered for other cases and conditions including start from the middle of the tape or from the 'wrong' end. In these cases the tape automatically spools to the correct start end before the program cycle is initiated, also where alternate strips are called the intervening return is spooled. A program may be subject to an 'interrupt' stop (and continue), or a new program may be imposed by acceptance of a LAM signal.

While the detailed functioning of the program encoder will by this point have been substantially inferred, it is however briefly set out for the sake of completeness (Figs. 9 and 10).

The program patchboard enables any one of the 24 possible complete tape pass programs (programs comprising each of the four tape strips) to be selected by patching '1' signals (Fig. 9) to the required tape-pass/sequence points, the four selected points being converted to a five-bit binary code as indicated using rectangular diode matrices. The programs may be modified by appropriate patching of the end of cycle signal relative to the main sequence patching and by varied selection of end of cycle option.

The control logic provides the LAM signal and accepts the DP and DBR signals. Initially BS7 and BS8 are set as indicated when the equipment is switched on. When BS7 is set '1' a LAM signal is generated, if it is accepted a DP signal is returned which enables 2.22 so that the '1' sources associated with the program patchboard ('1' OUT) and elsewhere are alerted, in addition BS8 is set '1' so that the 'Data Present' indicator (Fig. 9) is illuminated. Starting at the same time as the DP signal is the DBR which inhibits via I3 and 2.23 the resetting of BS7 and thus ensures the continued presence of the data output. When the transmitted data has been assimilated DBR returns to '0' and in doing so resets BS7 via I3, P8 and 1.18, and resets BS8 via 1.19 and D5. This two-second delay is provided because of the normally very short period of the data transfer cycle, to ensure that the Data Present indicator light is on long enough to confirm the cycle.

5. Conclusion

Although there has been a move towards the replacement of magnetic tape by 'on-line' computation for the direct processing of data from large nuclear experimental installations, nevertheless there are fields in which the recording of 'raw' data still plays a direct and prominent part, apart from the more secondary functions. The system described fulfils these various roles and by its flexibility offers potential for development and integration with evolving techniques.

6. Acknowledgments

Thanks are due to Mr. I. N. Hooton who was responsible for the initial design of the record and replay system, to Mr. A. Lewis who was responsible for the 1-inch tape transport, and to the Director of the Atomic Energy Research Establishment, Harwell, for permission to publish this paper.

7. References

1. F. H. Wells, I. N. Hooton and J. G. Page, 'Magnetic tape digital recording for nuclear research', *J. Brit. Instn Radio Engrs*, **20**, No. 10, p. 749, October 1960.
2. P. E. Cavanagh, C. F. Coleman, G. A. Gard, B. W. Ridley and J. F. Turner, 'The neutron transfer reaction Ca^{40} (p, d) Ca^{39} at 30 MeV proton energy', *Nuclear Physics*, **50**, pp. 49–60, 1964.
3. P. E. Cavanagh, 'The recording of nuclear pulse data', *Nuclear Engineering*, **5**, No. 49, pp. 255–7, June 1960.
4. I. N. Hooton, 'A Magnetic Tape Recording System for Nuclear Physics Research', Atomic Energy Research Establishment (Harwell) Report No. R.3422 (H.M.S.O., 1960).
5. K. Kandiah, Magnetic tape systems at Harwell, in 'Instrumentation Techniques in Nuclear Pulse Analysis', National Academy of Sciences—National Research Council Publication No. 1184, p. 296, 1964.
6. Gomer L. Davies, 'Magnetic Tape Instrumentation' (McGraw Hill, New York, 1961).
7. I. N. Hooton, 'A Digital Discriminator and Single Channel Analyser', Atomic Energy Research Establishment (Harwell) Report No. R3402 (H.M.S.O., 1960).
8. G. C. Best, 'A General Purpose Logic System', Atomic Energy Research Establishment (Harwell) Report No. M1514, 1964.

Manuscript first received by the Institution on 10th August 1966 and in final form on 12th April 1967. (Paper No. 1147/C98.)

U.D.C. 621.377

A Rigorous Theory of Electromechanical Delay Lines

By

M. BÖRNER, Dr. ter. nat.†

Summary: Introduction of some insignificant restrictions into the theory of electromechanical delay lines gives rigorous solutions of great simplicity. These restrictions are: (a) the line shall be described by a two-row two-line matrix; (b) the transducers are represented by equivalent circuits with concentrated elements; (c) the transducers shall be electrically tuned to obtain the widest possible bandwidth; (d) the delay line shall be symmetrical; (e) the losses shall be restricted to the electromechanical transducer resonators, their mechanical terminating resistors and the electrical input and output circuits. The solution of the equations obtained solves an optimization problem and allows, among other things, the determination of the ripple W of the transfer at the band-edge as a function of the normalized attenuation in the middle of the pass-band.

1. Introduction

A rigorous theory of electromechanical delay lines is set out in this paper which is based on certain assumptions. (More general theories are presented in Refs. 1, 2 and 3.) These assumptions are as follows:

(a) The delay line is described by a 2×2 matrix. The problem of radiation characteristics and mode transformations in inhomogeneities or on reflexion surfaces are therefore excluded. We confine ourselves to simply enumerating other limiting factors:

(b) The behaviour of the delay lines is described with a certain accuracy only up to bandwidths of 10%. The transducers in fact can then be represented by equivalent circuits with concentrated elements.

(c) The electromechanical transducers are tuned to enable the greatest possible bandwidth to be achieved. It follows from this (and from (b) as well) that neither the magnetostrictive nor the piezoelectric transducer differ with respect to the equivalent circuit.

(d) The symmetry of the delay line can furthermore be assumed without any real restrictions. This is advantageous as one then only has to calculate one delay line half.

(e) The losses are restricted to the electromechanical transducer resonators, their mechanical backings and the input and output circuits.

The losses within the connecting layer between the transducer and delay medium are always very small provided that the design is in order and that the thickness of the connecting materials (i.e. indium, solder, etc.) does not exceed half the wavelength. The losses within the actual delay medium are not dealt with here, so that the operating characteristics of the transducers and junctions stand out even more clearly.

Despite the limitations imposed by points (a) to (e), the major number of all delay lines built so far can nevertheless be covered. Only the dispersive lines and the untuned ones are excluded. The description of the very broad-banded lines, which are driven by tuned piezo-ceramic transducers (coupling factor $K \sim 0{\cdot}6\ldots0{\cdot}7$) by this theory represents only an approximation but one that is only possible at all by using concentrated elements of the equivalent circuit of the transducer.

2. Circuit Equations of Tuned Delay Lines

While bearing the restrictions outlined in the previous section in mind, one can still distinguish the two main types of delay lines in accordance with the way in which the mechanical matching between the electromechanical transducer and the delay time medium is established. The equivalent circuit of both types of delay lines is shown in Figs. 1 and 2.

As will be shown, the direct contact (Fig. 2) between the line and the ($\lambda/2$ thick or long respectively) transducer is only advantageous under certain conditions; the preferred (but not essential) $\lambda/4$-thick matching layer shown in Fig. 1 (mechanical coupling medium K_{23}) permits greater liberty in the design of such delay lines where the choice of the material and its characteristic impedance actually impose their own restrictions. In Fig. 1 R_1 represents the feed or load resistance respectively. L_1, C_1 constitute the tuned input and output circuit of the transducer (magnetostrictive or piezoelectric). The electromechanical

---

† AEG-Telefunken Research Institute, Ulm, Western Germany.

coupling factor is characterized by the coupling quadripole $[K_{12}]$ and the mechanical transducer resonator by the equivalent circuit L_2, C_2 (longitudinal resonator, torsional resonator, shear resonator, etc.). The mentioned coupling medium (characteristic impedance Z_{K2}) establishes the connection to the mechanical delay line (characteristic impedance Z_L). R_2 not only represents the internal losses of the transducer but also those which arise from fitting mechanical backing to the transducer on the side averted from the mechanical delay line. With an appropriate ratio of $Z_L/\omega_0 L_2$ this coupling medium can be omitted or rendered very thin compared with the wavelength (Fig. 2). Representing electromechanical transducers one usually uses ideal transformers characterizing the electromechanical coupling effect, which were, however, in view of the limitation expressed in (d) (Sec. 1), shifted within the line so far, that both transformers of both the transducers cancel each other.

3. Method of Calculations

We shall first consider a circuit as shown in Fig. 1. By following the equivalent circuit, one can then formulate the transmission matrix of the delay line, which is fed by the feed impedance R_1. This nevertheless leads to a rather complicated configuration. We shall thus first simplify the matter by applying limitation (d). For that purpose we start by considering the equivalent circuit shown in Fig. 3, which has been slightly modified.

One can see at once that this circuit—including the resistance R_1—is symmetric to the geometrical and topological centre of the mechanical line.

If

$$[\overline{A}] = \begin{bmatrix} \overline{A}_{11} & \overline{A}_{12} \\ \overline{A}_{21} & \overline{A}_{22} \end{bmatrix} \qquad \text{......(1)}$$

is the chain matrix for the left half (including R_1),

$$[\overline{B}] = \begin{bmatrix} \overline{A}_{22} & \overline{A}_{12} \\ \overline{A}_{21} & \overline{A}_{11} \end{bmatrix} \qquad \text{......(2)}$$

stands for the chain matrix of the right half, condition $|\overline{A}| = 1$ thereby being fulfilled. The transmission factor U of the input current I_1 into the output voltage V_2 (Fig. 3)

$$U = \left(\frac{I_1}{V_2}\right)_{I_2=0} \qquad \text{......(3)}$$

can easily be calculated from the elements of the matrix. This results in

$$U = 2\overline{A}_{21}\,\overline{A}_{22} \qquad \text{......(4)}$$

The effective transmission factor θ of the quadripole in Fig. 1 between both resistances R_1 then obtained is

$$\theta = \frac{R_1}{2}\,.\,U = R_1\,\overline{A}_{21}\,\overline{A}_{22} \qquad \text{......(5)}$$

Using this method of calculating, one only has to calculate the product of two elements of the matrix of half the quadripole from Fig. 3. The matrix, as the product of the parts marked in Fig. 3 then reads

$$[\overline{A}] = [\overline{A}_1]\cdot[\overline{A}_2]\cdot[\overline{A}_3]\cdot[\overline{A}_4]\cdot[\overline{A}_5]$$

$$= \begin{bmatrix} 1 & 0 \\ \mathrm{j}\dfrac{1}{Z_1}\,v + \dfrac{1}{R_1} & 1 \end{bmatrix} \begin{bmatrix} 0 & \mathrm{j}Z_{K1} \\ \mathrm{j}\dfrac{1}{Z_{K1}} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \mathrm{j}\dfrac{1}{Z_2}\dfrac{\pi}{2}\,v + \dfrac{1}{R_2} & 1 \end{bmatrix} \begin{bmatrix} 0 & iZ_{K2} \\ \mathrm{j}\dfrac{1}{Z_{K2}} & 0 \end{bmatrix} \begin{bmatrix} \cos K\dfrac{l}{2} & \mathrm{j}Z_l \sin K\dfrac{l}{2} \\ \mathrm{j}\dfrac{1}{Z_l}\sin K\dfrac{l}{2} & \cos K\dfrac{l}{2} \end{bmatrix} \qquad \text{......(6)}$$

The matrices $[\overline{A}_2]$ and $[\overline{A}_4]$ are the narrow-band approximations; in particular, the representation of $[\overline{A}_4]$ is valid only with a $\lambda_0/4$-coupling medium. λ_0 = the wavelength in the middle of the transmission band, i.e. at

$$v = \frac{2\Delta_f}{f_0} = 0$$

(v = detuning of the frequency f_0 by Δf). $K = 2\pi/\lambda$ stands for the wave number. All the other values of eqn. (6) will be readily understood by employing the equivalent circuit 3. The factor $\pi/2$ in matrix $[\overline{A}_3]$ results from the fact that the described parallel circuit is a line one (for example a shear resonator with a thickness of $\lambda/2$), which has been approximated for narrow bandwidths.

For the two pertinent values of $\overline{A}_{21}$ and $\overline{A}_{22}$ one arrives at the following after a certain amount of calculation:

$$\overline{A}_{21} = -\mathrm{j}\frac{1}{Z_1}\left(\frac{K_{23}}{K_{12}}(v-\mathrm{j}d_1)\cos K\frac{l}{2} + \left(\frac{K_{12}}{K_{23}} + \frac{d_1 d_2}{K_{12}K_{23}} - \frac{v^2}{K_{12}K_{23}} + \mathrm{j}v\frac{d_1+d_2}{K_{12}K_{23}}\right)\sin K\frac{l}{2}\right) \qquad \text{......(7)}$$

$$\overline{A}_{22} = \frac{K_{23}}{K_{12}}(v-\mathrm{j}d_1)\sin K\frac{l}{2} - \frac{K_{12}}{K_{23}}\left(1 + \frac{d_1 d_2}{K_{12}^2} - \frac{v^2}{K_{12}^2} + \mathrm{j}v\frac{d_1+d_2}{K_{12}^2}\cos K\frac{l}{2}\right) \qquad \text{......(8)}$$

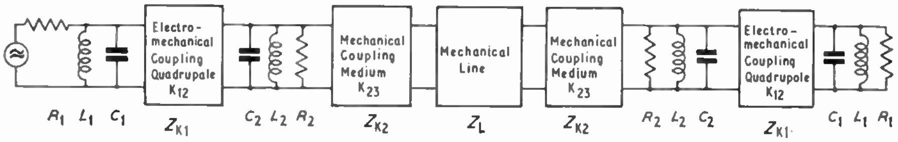


Fig. 1. Equivalent circuit of an electromechanical delay line with λ/4-thick coupling medium between transducer resonator and mechanical line.

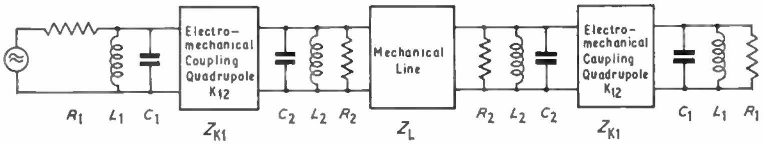


Fig. 2. Equivalent circuit of an electromechanical delay line and direct connection between transducer resonator and mechanical line.

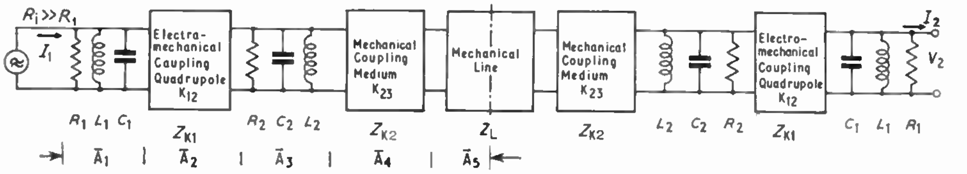


Fig. 3. Symmetrized equivalent circuit of an electromechanical delay line with λ/4 coupling medium between transducer resonator and mechanical line.

where, abbreviating:

$$\frac{Z_1}{R_1} = d_1 \qquad \text{......(9)}$$

$$\frac{\sqrt{\frac{2}{\pi} Z_1 Z_2}}{Z_{K1}} = K_{12} \qquad \text{......(10)}$$

$$\frac{2}{\pi}\,\frac{Z_2}{R_2} = d_2 \qquad \text{......(11)}$$

$$\frac{\sqrt{\frac{2}{\pi} Z_2 . Z_l}}{Z_{K2}} = K_{23} \qquad \text{......(12)}$$

K_{12} stands for the electromechanical coupling factor and its maximum value is actually a material constant. K_{23} is a measure for the mechanical coupling between the electromechanical transducer resonator and the delay time medium. d_1 and d_2 are attenuation constants.

Without in any way restricting the general validity one can moreover say:

$$Z_1 = \frac{2}{\pi} Z_2 = Z_l \qquad \text{......(13)}$$

These magnitudes, if assumed to be different, can be represented by means of ideal transformers. They may then be shifted towards the centre of the symmetric quadripole, where they cancel each other. The following then results from eqn. (5) by utilizing eqns. (7) and (8):

$$\theta = \sin Kl\left[-\frac{v}{K_{12}}\left(\frac{K_{23}^2}{K_{12}} + \frac{d_1+d_2}{K_{23}^2}\left(\frac{K_{12}}{d_1} + \frac{d_2}{K_{12}}\right)\right) + \frac{v^3(d_1+d_2)}{d_1 K_{12}^2 K_{23}^2}\right] + \cos Kl\left[1 + \frac{d_1 d_2}{K_{12}^2} - \frac{v^2}{K_{12}^2}\,\frac{2d_1+d_2}{d_1}\right] +$$

$$+\mathrm{j}\sin Kl\left[\frac{K_{23}^2 d_1}{2K_{12}^2} + \frac{K_{12}^2}{2d_1 K_{23}^2}\left(1 + \frac{d_1 d_2}{K_{12}^2}\right)^2 - \frac{v^2}{K_{12}^2}\left(\frac{K_{23}^2}{2d_1} + \frac{K_{12}^2}{d_1 K_{23}^2}\left(1 + \frac{d_1 d_2}{K_{12}^2}\right) + \frac{(d_1+d_2)^2}{2d_1 K_{23}^2}\right) + \frac{v^4}{2d_1 K_{23}^2 K_{12}^2}\right] +$$

$$+\mathrm{j}\cos Kl\left[\frac{v}{K_{12}}\left(\frac{K_{12}}{d_1} + \frac{d_2}{K_{12}} + \frac{d_1+d_2}{K_{12}}\right) - \frac{v^3}{d_1 K_{12}^2}\right] \qquad \text{......(14)}$$

That is actually the circuit equation of the equivalent circuit (Fig. 1) of a delay line with the coupling medium between the transducer resonator and the delay line which we were looking for.

The problem now consists of choosing the values of K_{12}, K_{23}, d_1 and d_2 so that the effective transmission factor is optimized in a certain way. There are various means of doing this: one can either avail oneself of the phase characteristics or of the amplitude characteristics. As this is a network of minimum phase shift, one characteristic is determined by the other. For a reasonably smooth phase shift it will suffice for the amplitude characteristic to attain maximum flatness in the centre of the band in the transmission range. Detailed calculations for different types of delay lines are given elsewhere.[4-7] This paper is limited to the results so far achieved and describes the method of calculations.

$|\theta|^2$ is derived from eqn. (14) and determines the values K_{12}, K_{23}, d_1 and d_2 in such a way that the lower powers of the detuning v are eliminated. That finally results in the very simple equation:

$$d_1 = K_{12} = K_{23}^2 - d_2 \qquad \ldots\ldots(15)$$

If one establishes

$$\frac{d_2}{K_{12}} = D_2 \qquad \ldots\ldots(16)$$

and

$$\frac{v}{K_{12}} = \Omega \qquad \ldots\ldots(17)$$

the result will be

$$|\theta|^2 = (1+D_2)^2\left[1+\Omega^2\frac{2D_2}{1+D_2}+\Omega^4\frac{1+2D_2}{2(1+D_2)^2}+\Omega^6\frac{D_2}{2(1+D_2)^3}+\Omega^8\frac{1}{8(1+D_2)^4}\right]+$$
$$+\tfrac{1}{2}\Omega^4\left[1+\Omega^2\frac{D_2^2}{1+D_2}+\Omega^4\frac{1}{4(1+D_2)^2}\right].\sin\,(2\,Kl+\Psi(\Omega)) \qquad \ldots\ldots(18)$$

where $\Psi(\Omega)$ is given by

$$\Psi(\Omega) = \arctan\frac{-\tfrac{1}{2}+\tfrac{1}{2}\Omega^2\dfrac{2+D_2}{1+D_2}-\tfrac{1}{8}\Omega^4\dfrac{1}{(1+D_2)^2}}{\Omega-\tfrac{1}{2}\Omega^3\dfrac{1}{1+D_2}} \qquad \ldots\ldots(19)$$

Besides a slow variation of the effective transmission factor by means of the frequency f (with $v = 2\Delta f/f_0$), there is a very fast ripple which is delineated by the second term of eqn. (18), i.e. predominantly by the $\sin\,(2Kl+\Psi(\Omega))$ (multiple reflection). The wave number

$$K = \frac{2\pi}{\lambda} = \frac{2\pi}{c}\,.f$$

again gives the main component (c = sound velocity of the excited waves).

4. Results

One obtains an excellent overall view of the solution of eqn. (18) if one considers only the 'envelope' of the fast ripple. With $l \to \infty$ the points of the curve, where

$$\sin\,(2Kl+\Psi(\Omega) = \pm 1$$

are actually the true envelopes. If we take $\sin\,(2Kl+\Psi(\Omega))$ to equal ± 1, then the following equations result instead of eqn. (18) (the roots can be extracted!):

$$|\theta|_{\text{max}} = 1+D_2+\Omega^2 D_2+\Omega^4\frac{1}{2(1+D_2)} \qquad \ldots\ldots(20)$$

$$|\theta|_{\text{min}} = 1+D_2+\Omega^2 D_2 \qquad \ldots\ldots(21)$$

Figure 4 depicts the result for three values of D_2.

The result is remarkable. In electromechanical delay lines such as those described it is not possible to nullify the ripple of the transmission factor over the whole transmission range. The addition of a certain attenuation is always necessary. The definition of the bandwidth of such delay lines is imposed by Fig. 4:

$$(|\theta|_{\text{max}})_{\Omega=\Omega_g} = \sqrt{2}(|\theta|_{\text{max}})_{\Omega=0} = \sqrt{2}(|\theta|_{\text{min}})_{\Omega=0} \quad \ldots(22)$$

By means of eqn. (20) one obtains for the detuning at the edge of the band the following:

$$\Omega_g = [(1+D_2)(D_2^2+2(\sqrt{2}-1))^{\frac{1}{2}}-D_2(1+D_2)]^{\frac{1}{2}} \qquad \ldots\ldots(23)$$

This term has two critical values:

$$\Omega_g = [2(\sqrt{2}-1)]^{\frac{1}{4}} \quad (D_2 = 0) \qquad \ldots\ldots(24)$$

or with eqn. (17)

$$v_g \simeq 0{\cdot}954\,.K_{12} \qquad \ldots\ldots(24a)$$

For $D_2 \to \infty$ one obtains

$$\Omega_g = (\sqrt{2}-1)^{\frac{1}{2}} \qquad \ldots\ldots(25)$$

or

$$v_g \simeq 0{\cdot}643\,.K_{12} \qquad \ldots\ldots(25a)$$

The intermediate values may be obtained from Fig. 5.

In dimensioning a delay line an important value for the choice of D_2 is obviously the admissible ripple W at the band-edge $\Omega = \Omega_g$, in accordance with Fig. 4. This ripple W is defined by

$$W = \left[\frac{|\theta|_{\text{max}}-|\theta|_{\text{min}}}{|\theta|_{\text{min}}}\right]_{\Omega=\Omega_g} \qquad \ldots\ldots(26)$$

which results in

$$W = \frac{\sqrt{2}}{1-D_2+D_2(D_2^2+2\sqrt{2}-2)^{\frac{1}{2}}} \qquad \ldots\ldots(27)$$

This result is shown in Fig. 6. It can be clearly seen that the ripple W of the transmission curve of the delay line disappears only when the attenuation is almost >6 dB in the centre of the band ($W \leq 4{\cdot}4\%$). If the method of calculations (Sect. 3) is also applied to a delay line in accordance with Fig. 3, it is astonishing to find that with the solution

$$d_1 = K_{12} = t-d_2 \qquad \ldots\ldots(28)$$

whereby d_1, K_{12}, d_2, as above,

$$t = \frac{2}{\pi} \cdot \frac{Z_2}{Z_l} \qquad \ldots\ldots(29)$$

the function $|\theta|^2$ again has the same configuration as in the first case.

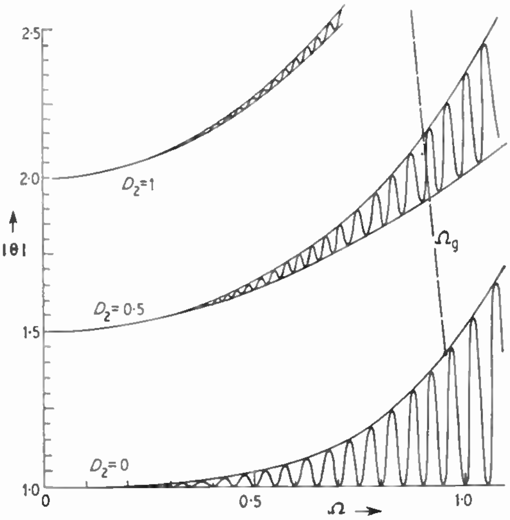


Fig. 4. Limiting curves of effective transmission loss of electromechanical delay lines.

5. Conclusions

All the above results are thus valid for a delay line without a coupling medium if one ensures that—in accordance with eqn. (29)—the material of the electromechanical transducer is properly matched to that of the delay line by a characteristic impedance variation as shown in eqns. (28) and (29). In practice a close approximation will be achieved.

With the derived relations, an overall view of the function of the individual elements of an electromechanical delay line can easily be gained. Moreover a good inside view into the mechanism of suppression of the disturbing multiple reflections is obtained.

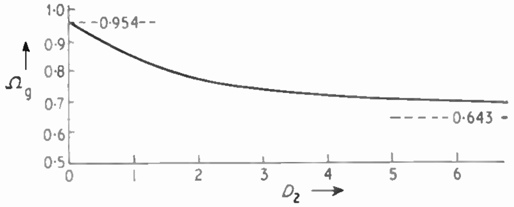


Fig. 5. Normalized band edge detuning Ωg in dependence of the normalized mechanical damping D_2.

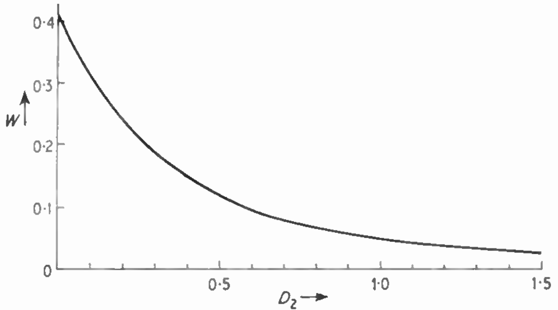


Fig. 6. Ripple W of an electromechanical delay line (fast ripple), i.e. the effective transmission loss, in the middle of the band.

6. References

1. John E. May, Jr., 'Guided wave ultrasonic delay lines', in 'Physical Acoustics I', ed. W. P. Mason, Part 17, Sect. 6, (Academic Press, New York and London, 1964).
2. R. N. Thurston, 'Effect of electrical and mechanical terminating resistances on loss and bandwidth according to the conventional equivalent circuit of a piezoelectric transducer', *Trans. Inst. Radio Engrs on Ultrasonic Engineering*, **UE-7**, p. 16, February 1960.
3. D. F. Brockelsby, J. S. Palfreeman and R. W. Gibson, 'Ultrasonic Delay Lines' (Iliffe, London, 1963).
4. M. Börner, 'Reaktanztheorie von elektromechanischen Laufzeitleitungen mit Koppelmedium zwischen Wandlern und Leitung', *Archiv der Elektrischen Übertragung*, **20,** No. 9, pp. 489–93, September 1966.
5. M. Börner, 'Reaktanztheorie von elektromechanischen Laufzeitleitungen bei direkter Kopplung zwischen Wandlern und Leitung', *A.E.Ü.*, **20,** No. 10, pp. 566–8, October 1966.
6. M. Börner, 'Strenge Theorie elektromechanischer Laufzeitleitungen mit Verlusten und einem Koppelmedium zwischen Wandlern und Leitung', *A.E.Ü.*, **20,** No. 11, pp. 605–10, November 1966.
7. M. Börner, 'Strenge Theorie elektromechanischer Laufzeitleitungen mit Verlusten bei direkter Kopplung zwischen Wandlern und Leitung', *A.E.Ü.*, **20,** No. 12, pp. 700–4, December 1966.

Manuscript received by the Institution on 22nd February 1967. (Paper No. 1148/EA39.)

U.D.C. 681.33 : 621.382.23

A Silicon Diode Analogue Multiplier

By

G. S. DEEP, M.E.†

AND

T. R. VISWANATHAN,
Ph.D.†

Summary: The logarithmic nature of the voltage–current characteristics of silicon junction diodes is exploited to design an analogue multiplier. Four silicon junction diodes are used in conjunction with a differential operational amplifier in the multiplier circuit. The principle of operation, design considerations and techniques to compensate for temperature effects and unequal diode parameters are discussed. Methods of performing division and generation of special types of functions are outlined. Measurements on an experimental single quadrant multiplier show that an accuracy of multiplication of the order of 1% can be achieved. Long term temperature drift at the output is measured to be 2% per 10 degC change in ambient temperature. The bandwidth of the multiplier is directly dependent on the operational amplifier employed. With a Donner (Model 3400) amplifier, the bandwidth is measured to be about 200 Hz. An attractive feature of the circuit is that it lends itself for miniaturization with the advent of differential operational amplifiers in the integrated circuit form.

1. Introduction

There are two distinct ways of using the exponential nature of diode characteristics to perform analogue multiplication and division. The first[1–5] utilizes the nature of the variation of the incremental conductance of a diode (or base-emitter junction of a transistor) as a function of the quiescent direct current through the junction. The second method[6, 7] consists in reducing the operation of multiplication to one of addition of the logarithms and a subsequent determination of the antilogarithm; and these two non-linear operations can be performed by the silicon diodes.

Since the characteristics of silicon diodes are temperature dependent, circuit functions based on these characteristics will, in general, be dependent on temperature unless suitable compensation techniques are incorporated in the design. This paper describes the principle of operation, the design considerations and various compensation techniques of a multiplier based on the second method.

2. Principle of Operation

The forward voltage V across a junction diode and the forward current I passing through it satisfy the following relation

$$I = I_s(e^{\gamma V} - 1) \qquad \ldots\ldots(1)$$

where I_s is the theoretical value of the reverse saturation current and

$$\gamma = \frac{q}{nKT}$$

where q is electronic charge

K is Boltzmann's constant

T is the absolute temperature, and

n is a constant nearly equal to unity[8] (which varies from diode to diode)

For values of V for which $e^{\gamma V} \gg 1$, eqn. (1) reduces to[9]:

$$I = I_s\, e^{\gamma V} \qquad \ldots\ldots(2)$$

or

$$V = \frac{1}{\gamma} \ln \left(\frac{I}{I_s}\right) \qquad \ldots\ldots(3)$$

Thus, the diode performs the operations of obtaining the logarithm and antilogarithm.

Consider the circuit shown in Fig. 1. The analogue voltages E_x and E_y to be multiplied are converted into currents I_x and I_y which are passed through the diodes D_x and D_y respectively. The voltage drops V_x and V_y across the diodes are given by

$$V_x = \frac{1}{\gamma_x} \ln \left(\frac{I_x}{I_{sx}}\right) \qquad \ldots\ldots(4)$$

$$V_y = \frac{1}{\gamma_y} \ln \left(\frac{I_y}{I_{sy}}\right) \qquad \ldots\ldots(5)$$

where the subscripts x and y refer to the diodes D_x and D_y respectively. The voltages V_x and V_y are added using an adder whose output is considered to be an ideal voltage source of magnitude V_z given by

$$V_z = V_x + V_y = \ln \left\{ \left(\frac{I_x}{I_{sx}}\right)^{\frac{1}{\gamma_x}} \left(\frac{I_y}{I_{sy}}\right)^{\frac{1}{\gamma_y}} \right\} \qquad \ldots\ldots(6)$$

The voltage V_z is impressed upon a diode D_z, the resultant current through which is given by

$$I_z = I_{sz}\, e^{\gamma_z V_z} \qquad (7)$$

---

† Department of Electrical Engineering, Indian Institute of Technology, Kanpur.

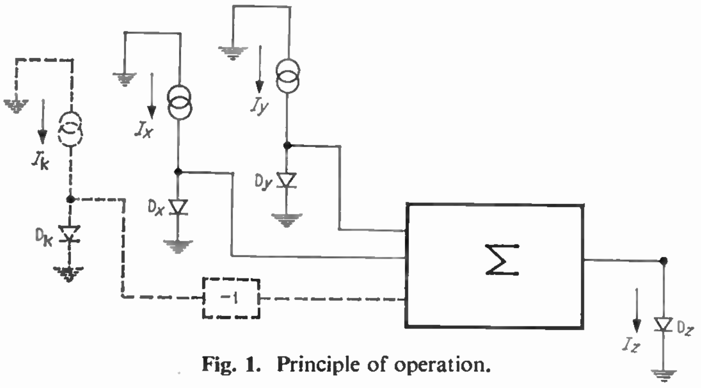


Fig. 1. Principle of operation.

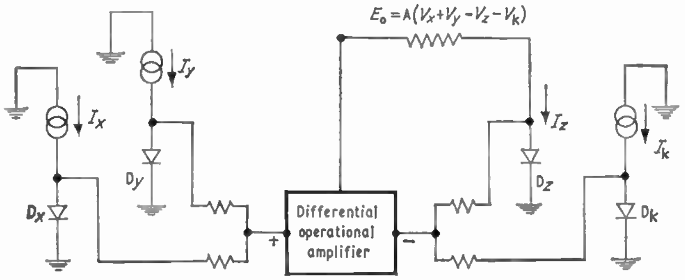


Fig. 2. The feedback arrangement.

Equations (6) and (7) yield

$$I_z = I_{sz}\left(\frac{I_x}{I_{sx}}\right)^{\frac{\gamma_z}{\gamma_x}}\left(\frac{I_y}{I_{sy}}\right)^{\frac{\gamma_z}{\gamma_y}} \qquad \ldots\ldots(8)$$

and if $\gamma_x = \gamma_y = \gamma_z$

$$I_z = \frac{I_{sz}}{I_{sx}I_{sy}}\, I_x\, I_y \qquad \ldots\ldots(9)$$

Thus, the measurement of current I_z gives the product of I_x and I_y, the constant of proportionality being

$$\frac{I_{sz}}{I_{sx}\; I_{sy}}$$

The reverse saturation current of a silicon diode is a function of ambient temperature and is expressed as[9]

$$I_s = (\text{constant})T^3\exp(E_gKT) \qquad \ldots\ldots(10)$$

where E_g is the gap energy and other symbols have their usual meaning. The constant of proportionality in eqn. (9) is therefore temperature dependent and this renders the multiplication scheme useless, unless suitable compensation is incorporated. In order to make this constant independent of temperature, another diode D_k is employed (shown dotted in Fig. 1). A constant current I_k is passed through the diode D_k, and the voltage drop V_k across it is inverted and added to V_x and V_y.

Now

$$V_z = V_x + V_y - V_k \qquad \ldots\ldots(11)$$

or

$$\ln\left(\frac{I_z}{I_{sz}}\right)^{\frac{1}{\gamma_z}} = \ln\left\{\left(\frac{I_x}{I_{sx}}\right)^{\frac{1}{\gamma_x}}\left(\frac{I_y}{I_{sy}}\right)^{\frac{1}{\gamma_y}}\left(\frac{I_{sk}}{I_k}\right)^{\frac{1}{\gamma_k}}\right\} \qquad \ldots\ldots(12)$$

Thus

$$I_z = I_{sz}\left(\frac{I_x}{I_{sx}}\right)^{\frac{\gamma_z}{\gamma_x}}\left(\frac{I_y}{I_{sy}}\right)^{\frac{\gamma_z}{\gamma_y}}\left(\frac{I_{sk}}{I_k}\right)^{\frac{\gamma_z}{\gamma_k}} \qquad \ldots\ldots(13)$$

and setting

$$\gamma_x = \gamma_y = \gamma_z = \gamma_k \qquad \ldots\ldots(14)$$

$$I_z = \frac{I_{sz}I_{sk}}{I_{sx}I_{sy}}\,\frac{I_x\;I_y}{I_k} \qquad \ldots\ldots(15)$$

It was stipulated that the output of the adding circuit should be an ideal voltage source. In practice a voltage source whose source resistance is small enough (1 ohm) for the purpose is hard to achieve. Moreover, the determination of the product involves measurement of the current I_z. These difficulties along with the need for an inverter can be overcome by using the negative feedback configuration shown in Fig. 2, which employs a differential operational amplifier. For values of gain A of the amplifier tending to infinity, the arrangement forces the following equality:

$$V_x + V_y = V_z + V_k \qquad \ldots\ldots(16)$$

or

$$V_z = V_x + V_y - V_k \qquad \ldots\ldots(17)$$

which is the same as eqn. (11). If the voltage E_o is very much greater than diode drop V_z, then E_o together with the resistance R_f forms the current source I_z. Thus, the output voltage E_o is proportional to the product of the input voltages E_x and E_y. The design considerations of the multiplier will now be discussed.

3. Diode Characteristics

The first step in designing a multiplier based on the above principle would be to study the range of voltages and currents over which the simplified diode of eqn. (2) is valid at room temperature. The experimental investigations indicate that the actual diode characteristics depart from eqn. (2) at higher values of I (about 1 mA) due to the spreading resistance. Again eqn. (2) is not satisfied at lower values of I (comparable to I_s) due to the approximation involved in deriving the equation itself. Experimentally it is found that many diodes satisfy the diode equation (e.g. type T.I. 601) in the current range of 1 μA to a few hundred microamperes within the limits of experimental error (1 %). The spreading resistance is found (graphically) to be of the order of 10 Ω and the typical magnitudes of I_s and γ are 10^{-9} A and $(40\,\text{mV})^{-1}$ respectively. γ varies from sample to sample by as much as 10%.

4. One Quadrant Multiplier

In the simulation shown in Fig. 3, two operational amplifiers (Donner type 3400) are used instead of a differential operational amplifier. The purpose of the simulation is to demonstrate the principle of operation and to get an idea of the attainable accuracy of multiplication. Inputs are obtained from variable direct voltage sources and relation between input and output is studied.

The input voltages E_x and E_y (0–100 V) in series with resistances of 100 kΩ form the current sources which pass currents (0–1 mA) through the diodes D_x and D_y respectively. The voltage drops across the diodes namely V_x and V_y are added using the operational amplifier A1. Each of the input resistances R_{x1}, R_{y1}, R_{x1} and R_{z1} are chosen to be 1 MΩ in order to limit their shunting effect on the diodes to be less than 1 μA. The amplifier A1 is operated with unity gain.

A constant current of 1 mA is passed through the diode D_k. The voltages $[-(V_x+V_y)]$, V_k and V_z form inputs to amplifier A2. Little examination will reveal that the resistance R_{b2} controls the overall scale factor of the multiplier. Since the diode D_z is to be operated with a maximum current of 1 mA, choosing R_{b2} as 100 kΩ fixes the maximum output voltage at 100 volts (i.e. when $E_x = E_y = 100$ V). Thus, the scaling employed in the simulation yields

$$E_o = \frac{E_x E_y}{100}$$

The $V-I$ characteristic of an ideal diode is shown in Fig. 4. If the diodes D_x and D_y are operated over six decades of currents (1 nA to 1 mA), then it can be seen that the diode D_z will have to operate over twelve decades of current which involves practical difficulties. Thus, the diodes D_x and D_y are operated only over a range of three decades of current (1 μA to 1 mA), and the diode D_z will now operate over a range of six decades of current (1 nA to 1 mA).

The range of operation of the current in the diodes D_x and D_y is restricted from 1 μA to 1 mA by passing an independent current (referred to as zeroing current) of 1·0 μA through each of the diodes D_x and D_y. This is achieved by using the resistances R_{xo} and R_{yo}. The sum of the voltage drops across the diodes D_x and D_y, due to zeroing currents alone is made equal to the voltage drop across the diode D_k which makes V_z and E_o equal to zero. Now the range of variation of current through the diode D_z will be 0 to 1 mA.

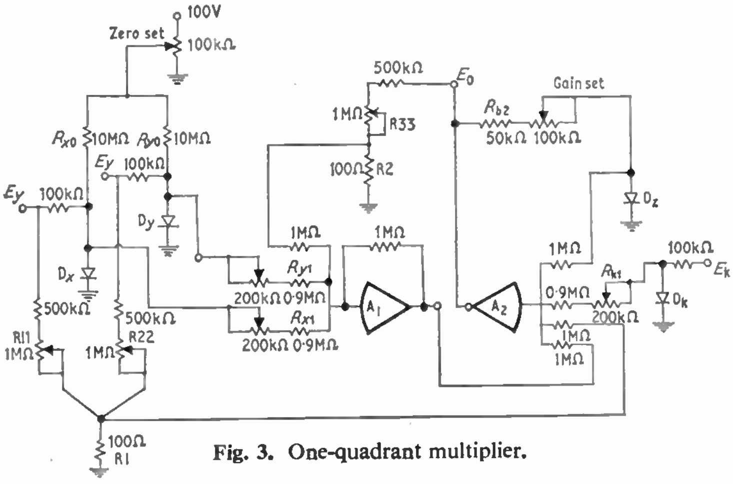


Fig. 3. One-quadrant multiplier.

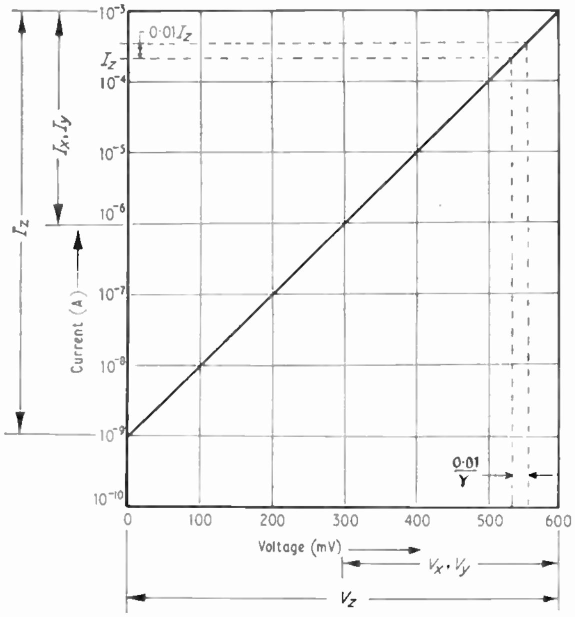


Fig. 4. Voltage and current ranges employed in the diodes.

Since the zeroing current is 0·1% of the full scale value, the maximum error introduced in the product due to the presence of the zeroing currents is only 0·1% (this occurs when one of the inputs is maximum and the other is zero).

If the gain A of the amplifier (Fig. 2) is finite, the sum $V_x + V_y$ will not be equal to $V_z + V_k$. In fact, it is this small difference between the input voltages which causes the output voltage of the difference amplifier. In order to achieve a multiplication accuracy of 1%, the current I_z or the output voltage E_o can depart from the actual value only by 1%. It can be seen from Fig. 4 that the corresponding difference between the input voltages can only be $0{\cdot}01/\gamma$ ($\sim$0·4 mV). Thus the minimum open-loop voltage gain of the amplifier necessary is estimated to be of the order of 250 000 (under full scale conditions).

The maximum closed-loop gain of the feedback amplifier in this configuration, measured between the input voltage V_x and the output voltage E_o (with V_y held constant at its maximum value) is of the order of 300. With an amplifier whose gain-bandwidth product is 60 kHz, the bandwidth obtained for the multiplier is about 200 Hz.

5. Compensations

5.1 *Compensation for Unequal γ Values*

Equation (15) is valid only if

$$\frac{\gamma_z}{\gamma_x} = \frac{\gamma_z}{\gamma_y} = \frac{\gamma_z}{\gamma_k} = 1 \qquad \ldots\ldots(18)$$

In practice this relation is seldom satisfied unless the diodes are chosen carefully. To force the equality the coefficients C_x, C_y and C_k are introduced such that

$$C_x \frac{\gamma_z}{\gamma_x} = C_y \frac{\gamma_z}{\gamma_y} = C_k \frac{\gamma_z}{\gamma_k} = 1 \qquad \ldots\ldots(19)$$

Thus the eqn. (16) is modified as

$$C_x V_x + C_y V_y = V_z + C_k V_k \qquad \ldots\ldots(20)$$

In practice this is conveniently achieved by adjusting the value of input resistances R_{x1}, R_{y1} and R_{k1}. The compensations for inequality of γ'_s can be affected by the following experimental procedure. With $E_x = 10$ V and $E_y = 100$ V the voltage E_o is set at 10 V using gain set control. A plot is obtained on the X–Y recorder between E_x and E_o. The values of E_x and E_o are kept low (0–10 V) in order to operate the diode D_x and D_z in the current range where the effect of their spreading resistance is negligible. In case the trace differs from a straight line, it can be made linear by adjusting the value of resistance R_{x1}. A similar procedure is repeated for linearizing the relationship between E_y and E_o keeping E_x constant. The method of making $C_k(\gamma_z/\gamma_k)$ equal to unity is described subsequently under temperature compensation.

5.2 *Spreading Resistance Compensation*

The effect of the spreading resistance in the diodes is to modify eqn. (11) as follows:

$$\ln\left(\frac{I_z}{I_{sz}}\right)^{\frac{1}{\gamma z}} + I_z R_z = \ln\left(\frac{I_x}{I_{sx}}\right)^{\frac{1}{\gamma_x}} + I_x R_x + \ln\left(\frac{I_y}{I_{sy}}\right)^{\frac{1}{\gamma_y}} +$$

$$+ I_y R_y - \ln\left(\frac{I_k}{I_{sk}}\right)^{\frac{1}{\gamma_k}} - I_k R_k \qquad \ldots\ldots(21)$$

R_x, R_y, R_z, R_k being the spreading resistances of the diodes D_x, D_y, D_z and D_k respectively. A simple way to reduce the above relation to that given by eqn. (12) will be to cancel the linear terms involving I_x, I_y and I_z and this can be achieved by incorporating resistances R11, R22, R33, R1 and R2 as shown in Fig. 3. The values of R11, R22 and R33 are adjusted to requisite magnitudes by the following experimental procedure. The principle underlying the procedure is to bring the various diodes in the region of their characteristics in which their spreading resistances come into play, on a 'one at a time' basis.

With $E_x = 10$ V and $E_y = 100$ V, E_o is set at 100 V by adjusting the current in the diode D_k. A plot between E_x and E_o is obtained on the X–Y recorder, by varying E_x from 0 to 10 V. The departure of this curve from a straight line is only due to the presence of the spreading resistance R_z of the diode D_z, since the current in the diode D_x varies only from 1 μA to 100 μA and the current in the diode D_y does not vary. The above curve can be linearized by adjusting the value of the resistance R33. By a similar procedure, compensation is achieved for R_x and R_y.

5.3 *Temperature Compensation*

A first order temperature compensation is achieved by the use of the diode D_k itself. Better compensation is obtained if $C_k(\gamma_z/\gamma_k)$ is made equal to unity (refer to eqn. (18)). This is done by adjusting the value of resistance R_{k1}. It may be mentioned that all the four diodes should be placed close to each other (preferably the four silicon junctions should be grown on the same silicon chip) to avoid any differential heating. The diodes are then heated in a constant temperature bath and the voltage E_o is observed. If E_o increases with temperature, the resistance R_{k1} is decreased, otherwise it is increased. Two or three runs are made until a good temperature compensation is obtained. Experimental results show that the drift after compensation is of the order of 2% per ±10 degC change in ambient temperature.

6. Four Quadrant Operation

It can be seen that the circuit shown in Fig. 2 will perform multiplication only if the currents I_x, I_y and I_z are in the forward directions of diodes D_x, D_y and D_z. Thus the input voltages cannot change polarity. In order to make the multiplier circuit accept input voltages of both polarities (for four quadrant operation), independent quiescent currents I_x' and I_y' are set in the diodes D_x and D_y respectively, as shown in Fig. 5. Now, I_x and I_y can change polarity, still keeping the diodes forward biased provided

$$|I_{x(\max)}| \leqslant I_x' \quad \text{and} \quad |I_{y(\max)}| \leqslant I_y'$$

With reference to the eqn. (15), the new relationship between I_z, I_x and I_y is given by:

$$I_z = C(I_x + I_{x'})(I_y + I_{y'}) \qquad \text{......(22)}$$

where C is a constant.

Thus, in addition to the desired output, the expression for I_z contains unwanted linear terms in I_x and I_y and a constant term $C(I_x'I_y')$. These unwanted terms must be subtracted from the output voltage in order to obtain the desired product of the input voltages. This method is widely used in the conversion of a one-quadrant multiplier into a four-quadrant one and this involves the use of an additional operational amplifier.

In the circuit shown in Fig. 5, the subtraction of the unwanted terms is achieved by setting a constant independent quiescent current I_z' and in addition, passing currents proportional to I_x and I_y through the diode D_z. Thus the balance equation for the amplifier becomes

$$I_z = i_z + I_{z'} + K_1 I_x + K_2 I_y = C(I_x + I_{x'})(I_y + I_{y'}) \qquad \text{......(23)}$$

where K_1 and K_2 are chosen, such as

$$K_1 = CI_{y'} \qquad \text{......(24)}$$

$$K_2 = CI_{x'} \qquad \text{......(25)}$$

and

$$I_{z'} = CI_{x'}I_{y'} \qquad \text{......(26)}$$

From the above equations, i_z (which is proportional to E_o) is given by

$$i_z = CI_x I_y \qquad \text{......(27)}$$

thus achieving four quadrant multiplication.

7. Division and Function Generation

It was stipulated earlier that the current I_k through the diode D_k should be constant. Referring to eqn. (15), it is easily seen that if the current I_k is made proportional to a variable, it is possible to achieve division.

Little examination of the circuit shown in Fig. 2 will reveal that if fractions δ_1, δ_2 and δ_3 of voltages V_x, V_y and V_k respectively are introduced at the inputs of the amplifier, then

$$\delta_1 V_x + \delta_2 V_y = V_z + \delta_3 V_k \qquad \text{......(28)}$$

and this yields

$$I_z = I_{sz}\left(\frac{I_x}{I_{sx}}\right)^{\delta_1 \frac{\gamma_z}{\gamma_x}} \left(\frac{I_y}{I_{sy}}\right)^{\delta_2 \frac{\gamma_z}{\gamma_y}} \left(\frac{I_{sk}}{I_k}\right)^{\delta_3 \frac{\gamma_z}{\gamma_k}} \qquad \text{......(29)}$$

and if

$$\delta_1 \frac{\gamma_z}{\gamma_x} = m, \quad \delta_2 \frac{\gamma_z}{\gamma_y} = n, \quad \text{and} \quad \delta_3 \frac{\gamma_z}{\gamma_k} = p$$

$$I_z = \frac{I_{sz} I_{sk}^p}{I_{sx}^m I_{sy}^n} \; \frac{I_x^m I_y^n}{I_k^p} \qquad \text{......(30)}$$

Thus it is possible to obtain functions of the form

$$Z \propto \frac{X^m . Y^n}{K^p}.$$

In order that temperature compensation is preserved, the choice of p is restricted to $(m+n-1)$.

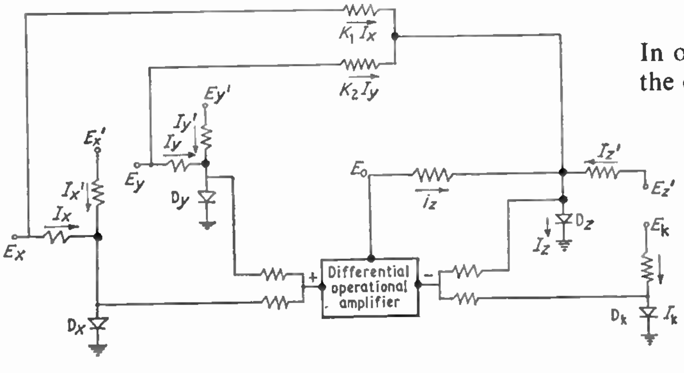


Fig. 5. Four-quadrant operation.

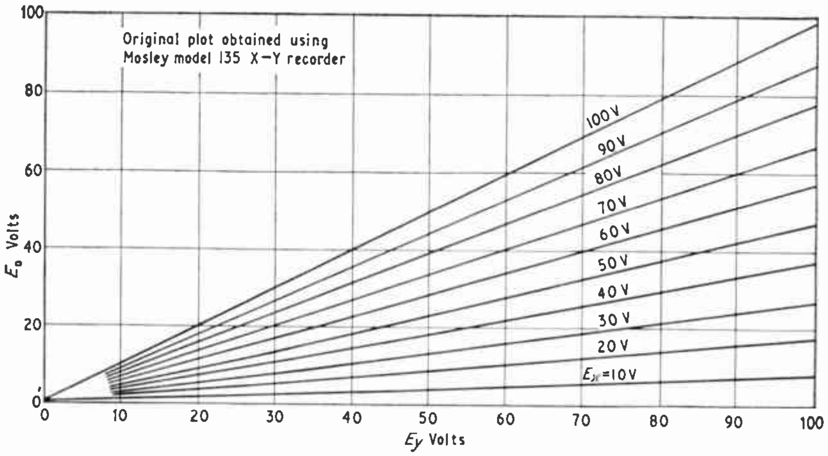


Fig. 6. Experimental results.

8. Experimental Results and Conclusions

The results obtained from experiments on one quadrant multiplier are presented in Fig. 6. The curves show the relationship between the product E_o and E_y for various values of E_x. The plot was done using X–Y recorder (Mosley Model 135). The linearity is estimated to be better than 1% of f.s.d.

In the experience of the authors,[5] the multiplication scheme based on method 1, although it obviates the need for a d.c. amplifier, involves modulation and demodulation procedures, which bring down the overall accuracy to 2–3%.[2, 3, 5] However, it is most suited when the signals to be multiplied are widely separated in frequency; a situation which naturally arises in amplitude modulation problems.

Although the logarithmic nature of the characteristics of a silicon junction has been used in a variety of ways to achieve multiplication, the method indicated above has some distinct advantages. The balanced configuration centred around a differential amplifier obviates the need for current measurement and naturally adapts itself for four-quadrant operation, division and generation of certain types of functions such as square rooting, etc. Although the bandwidth (the frequency at which the output falls by 3 dB below its low frequency value) of the experimental multiplier is restricted to 200 Hz, it can be improved considerably by using an amplifier with larger gain-bandwidth product. The most attractive aspect of the circuit is that it lends itself for miniaturization with the advent of integrated differential operational amplifier. In fact the four silicon junctions needed can be grown on the same chip in close proximity so that their properties track nicely with temperature. Under these conditions, it is reasonable to expect better temperature compensation. It may be mentioned that presently available integrated operational amplifiers have a dynamic range of the order of 10 or 20 V as compared to the Donner amplifier (100 V) used in simulation. Hence loss of linearity at the input current generating circuit is inevitable unless suitable input circuits are incorporated.

9. Acknowledgments

The authors wish to thank Dr. H. N. Mahabala for the many interesting discussions and Dr. M. A. Pai for his continued interest and encouragement during the course of this work.

10. References

1. G. B. B. Chaplin and A. R. Owens, 'Some transistor input stages for high-gain d.c. amplifiers', *Proc. Instn Elect. Engrs*, **105**, Part B, pp. 249–57, 1958. (I.E.E. Monograph No. 2832, July 1957.)
2. S. Deb and J. K. Sen, 'Variation of input conductance of a grounded base junction transistor', *Electronic Engng*, **31**, pp. 753–5, December 1959.
3. S. Deb and J. K. Sen, 'Transistorised analog multiplier', *Rev. Sci. Instrum.*, **32**, No. 2, pp. 189–92, February 1961.
4. Herbert L. Kahn, 'Multiplication and division using silicon diodes', *Rev. Sci. Instrum.*, **33**, No. 2, pp. 235–8, February 1962.
5. T. R. Viswanathan, 'An Analog Multiplier based on Semiconductor Properties', M.Sc. Thesis, University of Saskatchewan, Saskatoon, Canada, 1961.
6. George E. Platzer, Jr., 'Using transistor circuits to multiply and divide', *Electronics*, **39**, No. 7, pp. 109–15, 4th April 1966.
7. P. Kundu and S. Banerji, 'Transistorized multiplier and divider and its applications', *Trans. Inst. Elect. Electronics Engrs on Electronic Computers*, EC-13, No. 3, pp. 288–95, June 1964.
8. C. T. Sah, R. N. Noyce and W. Shockley, 'Carrier generation and recombination in p–n junctions', *Proc. Inst. Radio Engrs*, **45**, No. 9, pp. 1228–43, September 1957.
9. David Dewitt and Arthur L. Rossoff, 'Transistor Electronics', pp. 61, 76–77 (McGraw-Hill, New York, 1957).

Manuscript first received by the Institution on 18th November 1966, and in final form on 28th July 1967. (Paper No. 1149.)

U.D.C. 621.37/9.004.6

Major Causes of Equipment Unreliability

By

J. PATERSON†

Presented at the Joint Letchworth College of Technology–I.E.R.E. Conference on 'Reliability and Environmental Testing of Electronic Components and Equipment' held at The College of Technology, Letchworth, on 3rd May 1967.

Summary: Defect statistics on guided weapons and other equipment clearly indicate that the more unusual components and the 'special to project' innovations show the highest failure rates. Electromechanical components are examples and such failures are emphasized in guided weapons as the total running time may be a few hours with only a few minutes in flight.

This paper shows that since the design and shape of most electrical/electronic components is primarily to facilitate their functions and ease of production, a large variety of shapes and sizes for different types of components is a major problem in equipment design. It is suggested that improved customer/vendor relations and national standardization of new styles of components might help to ease the situation.

1. Introduction

The following discussion emphasizes certain areas of electronic design and construction that contribute largely to the unreliability or 'down time' of guided weapons.

It is thought that many of the observations made could equally apply to many similar complex electromechanical systems as the information is extracted from field defects on both missile and the associated ground equipment (vehicle borne). Missile defects are those which occur during ground testing and not flight.

As a rough comparison of complexity, missile electronics require some few thousand components with a ground running time of 50 hours and the ground equipment may total some 25 000 components with perhaps 8 000 or 10 000 hours running. The missile is, therefore, still in the 'infant mortality' phase of the component life curve (bath-tub curve) although the actual age may be a few years.

These qualifications are made as it is important to remember that, although guided weapons are not the main subject of the discussion, they are the source of the defect statistics and *failure rates* may, therefore, differ radically from those one might obtain from a computer, say, running for many thousands of hours in a very stable environment.

Analysis of failure data with a view to improving equipment unreliability is not a new topic and the task has been neither as easy nor as successful as one might expect. This discussion attempts to shed some fresh light on the subject without introducing statistical tabulations or the conventional reliability platitudes: it is hoped that the more practical approach will be of some value.

2. Defect Analysis

Clearly, no human being is infallible and one cannot simply issue an edict demanding perfection at all times, but one can help by the intelligent analysis and use of existing defect statistics so that the crux of the problem is understood and the activities channelled accordingly. This is not so easy as it sounds as the prime requirement to establish a defect reporting system and correctly to establish the *true cause* of many electronic defects is in itself a difficult and costly business. A resistor defect, for example, without a conclusive cause is a useless statistic for such an exercise, and in the writer's experience the majority of electronic component faults that occur *after final factory test* are not due to the component itself, but to some secondary feature. A thorough knowledge of the component, the equipment, and conditions of use is often essential to determine the true cause.

Marshalling this defect evidence to make best use of what slender trends there are introduces further problems and since the cause is the real key we must devise a means of rationalizing this aspect. This may best be illustrated by some examples:

(1) Relays, toggle and wafer switches, potentiometers, resistive transducers and commutated machinery have a common problem in that wiping or moving contact surfaces are involved. Common faults occur, due to oxidation, wear, pitting, etc., of the contacts. The actual component ceases to be important here as the common action or advice to engineers concerns wiping contacts and may further reduce to plating, hermetic sealing or arc suppression problems.

(2) Transformers, chokes, relays, wire-wound potentiometers, motors, synchros, tachogenerators and many others have a common problem because they all require many turns of fine-gauge wire and a

---

† British Aircraft Corporation, Guided Weapons Division, Stevenage, Hertfordshire.

common fault is an open-circuit at the junction of the lead-out wire to the winding wire.

Hence, what might appear to be a random fault on a variety of components, and therefore difficult to control, can be reduced in the final analysis to one specific mode of failure—providing we are prepared to diagnose each defect to the required depth and then, with inventiveness determine what the real mode is. It will be apparent that this is not a task for unskilled labour or junior clerks. It will also be apparent that simply rationalizing groups of defects does not necessarily suggest a cure, but our design ingenuity or the appropriate manufacturing method can now be applied to a more specific area.

3. Electromechanical Components

When these techniques are applied to guided weapon defects it can be seen that the electromechanical components, those with moving, wiping or rotating contacts and parts, give much higher defect *rates* than the more conventional electronic component. This is not due to wear-out characteristics as the running time of the missile is within the infant mortality phase of the life curve and the ground equipment is within the useful life plateau of most components used.

Prime responsibility for failure is partly due to the following causes:

(1) Component manufacturer
  (a) Manufacturing variations and errors.
  (b) A design which may not suit all conditions of use.
  (c) Complexity introduced by tolerance, friction, springs and adjustments.

(2) Equipment manufacturer
  (a) Incorrect mounting/assembly or damage during mounting.
  (b) Lack of experience in the correct/safe function of a less familiar component.

(3) Combination of component weakness and unusual equipment use.

For example, an ingress of encapsulation/brush coating in a partially sealed component—particularly troublesome with vacuum encapsulation.

The failure rate is probably proportional to the number of discrete parts used in the component and also on the degree of miniaturization and is therefore no reflection on specific manufacturers. There is no simple solution—just painstaking attention to every small detail. A careful scrutiny of each design feature with a receptive attitude to 'What might fail?' is also a good design discipline.

Similar comments apply to most of the special to project components as they stand less chance of long development and are made in small quantities. General guidance on such components is not easy but it would appear that engineers tend to concentrate on the functional parameters (or only those which they require) and tend to ignore the more practical aspects of mounting, wiring, etc. We are less familiar with these more complex parts and so mistakes are made and we learn the hard way.

It is doubtful if we can accelerate experience but we can be much more inquisitive and sceptical of our basic building bricks and engineers must dissect and familiarize themselves with each new part they design into equipment. Proprietary manufacturers will readily co-operate as they accept that most electromechanical devices require to be tailored to the job, but the responsibility for establishing such liaison rests with the user. It is the writer's opinion that time so spent pays better dividends than attempting to choose a best buy or the most reliable component by submitting *small batches* to environmental testing. Experience shows that environmental tests can best be applied after a critical examination has suggested possible weak areas, rather than a slavish repetition of standard tests. It is quite surprising the number of defects that elude formal specifications and artificial environments: a quality approved resistor, for example, has recently been modified because it was prone to body fracture and yet had escaped all the DEF approval tests as no one had ever thought of including a body strength test.

4. Module/Packaging Problems

The next general area of random failures concerns the practical aspects of electronic modules. The large variety of packaging methods permits each designer or draughtsman a great deal of scope and ingenuity in accommodating components and naturally mistakes are made and we have teething troubles. Even the most simple expedient of printed circuit board construction introduces many problems. These can be listed as follows:

What is the best type of printed circuit board for a given job?

What is the minimum width of track?

Since component leads can vary in diameter and may have solder blobs and spikes, what is the correct hole diameter for best soldering?

Should component leads be bent along the copper track or project straight through the hole (normal to the copper) for good soldering?

What is the minimum length of component lead before bending and soldering?

How do we apply heat sinks—or are they really necessary?

What is the best size of soldering iron and the best bit temperature? (It can be anything from 245°C to 375°C.)

How do we mark or stencil the board for ease of assembly and polarity of components, etc.?

How do we make external connections to the board?

How do we protect the track before and after components have been assembled?

How do we accommodate small transformers, Vinkor coils, trimpots or other adjustable devices and irregular-shaped components?

What is the best form of encapsulation?

Can we readily test the board as a sub-assembly?

The scope and variety of answers to some of these questions is considerable and it seems that once some degree of standardization has been achieved in the industry we move on to integrated circuits, microelectronics or some new technique with new attendant problems. It is probable that packaging and production problems account for more unreliability than either circuit design defects or true defects of the actual components.

It should be noted here that our National Certificate courses and electronic training programmes do not include any guidance on this subject and the circuit designer and draughtsman must learn from practical experience.

Employing our previous logic to rationalize 'electronic module' faults, we find that many wiring and so-called component faults reduce in fact to cramped packaging and to the indiscriminate insertion of irregular shapes and sizes, such that short circuits, pierced insulation, deformed and broken components and strained connections result. A clean orderly layout of similar-shaped components (usually cylindrical) with the minimum of flying leads and cross connection is the ideal. Mechanical supports, spacers and other large or irregular metallic parts should be avoided in and around connections to components—it should also be remembered that the assembly and wiring of electronic components is not an exact science and what appears as ample tolerance on an assembly drawing can reduce to zero due to bowing of a printed board or a component lead which is not bent quite at right angles. It is suggested that the draughtsman must look more critically at the actual hardware produced, and how it is produced, including the actual components. Outline drawings and brochure dimensions are not necessarily accurate or up-to-date and can obscure relevant detail—familiarization and job knowledge goes a long way on any job.

4.1 *Need for Standardization*

Although the component manufacturers might not agree, it is apparent that most electronic components are designed to facilitate their particular function and for ease of manufacture (most things are) and not to suit the equipment constructor. Resistors, capacitors and diodes with axial leads present no great problem as their shape just happens to suit most requirements, but anything which departs from this basic shape or has a different aspect ratio presents the equipment constructor with an immediate problem of assembly and hence, the possibility of reduced reliability. In effect, we are attempting to build a house with dissimilar shapes and sizes of bricks! It is obvious that component manufacturers cannot design for every type of construction and although there is some attempt to cater for printed board assembly by suitable lead configuration and by miniaturization, there is little progress with respect to shape and compatibility of build with many of our standard components.

One tends to take components very much for granted but, in fact, the industry is very young and reasonably advanced design and attempts to miniaturize has only occurred in the past decade. We have come a long way in less than 50 years and some of the problems discussed may well be resolved with the increasing use of integrated circuits and the automatic standardization which must eventually follow.

In the meantime, one can reduce many so-called random defects by recognizing and accepting the limitations and problems of current constructional techniques. Printed circuit board construction with conventional components will continue for several years and there is ample scope for improvement. It may be impracticable to consider a standard modular construction but some restriction or engineering discipline is still essential. The use of eyelets, turret lugs plated through holes and other intermediary printed circuit board connections can introduce soldering and insulation hazards when used indiscriminately.

Component manufacturers can help by being more specific on the basic assembly and wiring problems and typical layout sketches, showing minimum length leads, bending angles, and other recommendations on maximum soldering temperatures and times would help. There is no reticence in supplying this information when complaints are made on faulty components.

5. Importance of Attention to Small Detail

It is evident from guided weapon defect statistics that the majority of unreliability is caused by lack of attention to small detail, whether in components or equipment and is the concern of both the design engineer and everyone involved in every facet of production. Major design/manufacturing problems are normally self-evident and consequently receive a great deal of attention and are usually resolved, possibly to the extent of over-design, whereas minor detail tends occasionally to be overlooked and give rise to a variety of random defects. There are, of course, exceptions to this when new techniques and ideas are used and we push beyond the frontiers of existing knowledge.

It is the elusive nature of the random defect which prohibits effective cure and we are reminded of the colander analogy where we never have quite sufficient fingers to block each hole. It is stated by the Royal Air Force defect statistics department that defect rates do not materially reduce throughout the life of an aircraft regardless of extensive defect investigation and the various corrective measures which ensue.

This may well be an exceptional case, but the point to be established is that the majority of failures in a complex electro-mechanical system occur in a fairly random manner and usually reduce in the final analysis to some fairly simple part. The following are some of the examples:

(i) Chafed insulation on a hinged rack/module.
(ii) Excessively long screw fouling or causing short-circuit.
(iii) Wiring debris or spare washers causing short-circuit.
(iv) Ingress of moisture.
(v) Overheating due to poor ventilation or bad siting of components, etc.

None of these faults are brand new, they just keep popping up in different guises. None of them are true component faults and they do not require elaborate tests, sampling plans or engineering genius to prevent or correct them. They do require painstaking, methodical attention to every small detail from the drawing board to the final assembly or shipment.

Each detail part must be evaluated for possible failure modes—preferably by someone other than the original draughtsman. Deliberate criticism and imagination of worst case circumstances, environment, handling, circuit conditions or even secondary effects, such as overload conditions, can reveal weaknesses. The 'inventor' tends to see a perfect creation or the most favourable conditions for success, whereas the critic is much more sceptical.

A similar philosophy can be applied to inspection techniques such that in addition to checking compliance to drawing a conscious effort should be made for general unsatisfactory features. In certain electronic assemblies the functional test is often the best method of checking drawing compliance and visual inspection can be reserved for the more general state. We can rarely find the time for any task that is not mandatory and a strong discipline is essential if these measures are to be successful.

6. Conclusion

It would be naïve to suggest that there is an effective remedy to this problem and the topic is introduced as a means of placing it in perspective rather than providing an answer to it. However, it is felt that progress is possible if one can accept that 'electronic circuit design' and 'random component faults' are *not* necessarily the chief sources of unreliability and this view may well apply to most equipment and not just guided weapons.

Manuscript received by the Institution on 3rd May 1967. (Paper No. 1150.)

U.D.C.621.396.677

Superdirective Arrays:
The Use of Decoupling between Elements to Ease Design and Increase Bandwidth

By

Professor D. G. TUCKER,
D.Sc., C.Eng., F.I.E.E., M.I.E.R.E.†

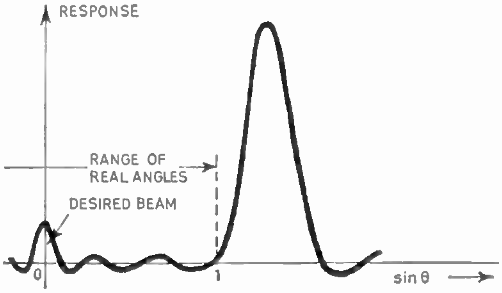
Summary: The design and achievement of superdirective receiving arrays for radio or acoustic waves can be greatly eased by the use of decoupling arrangements between the elements, such as buffer amplifiers in the output of each element before the outputs are combined to form the directional pattern. By this means, too, the severe bandwidth restrictions normally associated with superdirectivity can be removed.

1. Introduction

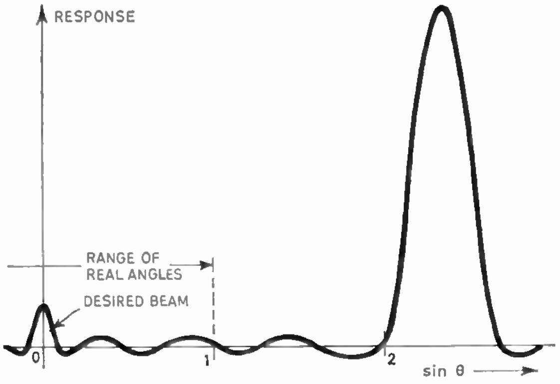
The concept of superdirectivity is quite old, discussions of it occurring in numerous text-books on radio antennas,[1] and the literature on the subject is vast.[2] Nevertheless little practical use has been made of superdirective arrays either in radio or in acoustics and experimental results are conspicuously sparse in the literature. Indeed, no very large degree of superdirectivity appears ever to have been achieved. Some of the practical difficulties of achieving a large superdirective gain are fairly easy to see, but some are rather obscure. Even when obtained, superdirectivity is usually assumed to be of little value because of the very narrow bandwidths apparently inherent in it. This note is intended to show that some of the difficulties normally associated with superdirectivity can be eliminated (or at any rate largely eliminated) by the use of decoupling (or buffer amplifiers) between the elements of the array.

A formal and precise definition of superdirectivity is difficult to give, but there seems no doubt that what is usually meant by the term is the property of obtaining a narrower beam (or higher directivity factor) from a linear array than is obtained with uniform or smoothly-tapered excitation, by crowding zeros (or null responses) into its directional pattern in the range of real angles at the expense of very large responses outside the range of real angles. Naturally, such responses outside the range of real angles cannot be directly observed, but what is meant becomes clear if the response of the array is plotted as a function, not of θ, the direction in space, but of $\sin\theta$. This is done in Fig. 1 for a broadside array, where $\theta = 0$ in the broadside direction. The directional pattern which can be measured is evidently that corresponding to angles for which $|\sin\theta| \leq 1$, so that $\sin\theta = \pm 1$ defines the limits of the range of real angles. For values of θ where $|\sin\theta| > 1$, i.e. where θ is not a real angle, the response can still be calculated although it cannot be directly measured.

Methods of calculating the required excitation (for transmitting) or sensitivity (for receiving) of the various elements of the array in order to obtain a desirable directional pattern are adequately described in the literature[3-5] and need not be discussed here. One feature which is fundamental to superdirectivity needs to be mentioned however. This is that the element spacing must be less than one half-wavelength and that alternate elements (or occasionally alternate groups of elements) must be connected in opposite phase, or at any rate with an antiphase component. It is not difficult to see that this is associated with the placing of large responses in the range of angles for which $|\sin\theta| > 1$. It is also easy to see that if alter-



(a) Major response in the range $|\sin\theta| > 1$ but as near as possible to the range of real angles.



(b) Closer element spacing is used, giving a large margin for increase in frequency.

Fig. 1. Superdirective directional responses plotted against $\sin\theta$, where θ is the angle of a particular wave direction relative to the broadside direction.

---

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

nate elements are in phase opposition, then whatever magnitudes of response are given to the individual elements (and these are usually unequal), the resultant far-field peak response in the broadside direction must be considerably less that it would be for the same elements used with the same currents in the more normal co-phasal manner. This leads to some important practical objections to superdirective arrays, as follows:

(i) Since for given currents in the elements the losses are fixed, a superdirective transmitting array gives a smaller peak transmission for the same losses as a normal array. It is therefore less efficient.

(ii) Since for given signal levels on the individual elements, the resultant output of a superdirective receiving array is smaller than for a normal array, its noise figure is worse. (Noise figure is a measure of the performance of an array in respect to thermal noise generated in the loss resistances of its elements and has the same general significance as the noise figure of an amplifier.)

(iii) Since the peak response is formed by differences between elements, and not by the additions as in a normal array, the response is very sensitive to small errors or variations in the individual elements or their associated circuits.

There may be circumstances where these objections are unimportant. For example, in a sonar system operating at low acoustic frequencies, say below 50 kHz, the acoustic noise level in the water due to wave action, reflected ship's noise, etc., usually so greatly exceeds any thermal noise in the receiver that a very considerable worsening of the noise figure of the array is quite acceptable. Special circumstances, or even careful design in ordinary circumstances, could reduce the importance of objection (iii).

There are other difficulties normally associated with superdirectivity which are usually regarded as fundamental. These are:

(a) the effect of the mutual coupling between elements in preventing the desired operation of the array from being achieved, and

(b) the very narrow bandwidth of a superdirective array as compared with the same array and elements used normally.

It must be obvious that the mutual coupling of elements spaced at fractions of a wavelength is considerable. In a normal array, where the elements are co-phasal, this can have a serious enough effect on the operation of the array,[6,7] but with superdirective operation the effect is clearly very great. This makes the design and setting-up of a superdirective array very difficult, although with care and in simple cases it can be successfully achieved.[8]

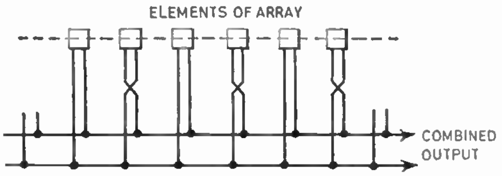
The narrow bandwidth associated with superdirectivity is perhaps harder to understand. It arises because the radiation resistance is very low and the reactance large, thus making tuning-out of the reactance necessary; the system then becomes a high-Q resonant system with consequently narrow bandwidth. We need, therefore, first to see why the radiation resistance is low. It is easier to see this for a transmitting array, and then invoke the principle of reciprocity to justify the same low resistance for the same array used for reception. On transmission we have alternate elements caused to carry currents (if they are radio antennas) or to vibrate (if they are electro-acoustic transducers) in opposite phase; yet they are very close together and thus have large mutual coupling. This means that even in the vicinity of one element the resultant field intensity or acoustic pressure is partially cancelled and shifted in phase, so that its in-phase component is small in relation to the current in the element (or to its velocity if it is acoustic). This explains the low radiation resistance, and also suggests a considerable radiation reactance. Most authors (e.g. Woodward and Lawson[4]) refer to the large reactive field in the vicinity of the aperture; Schelkunoff and Friis[1] describe it as resonant, but Woodward and Lawson state that 'a large balancing reactance would have to be incorporated behind the aperture plane, thus forming a highly-resonant and therefore frequency-sensitive arrangement'.

2. The Method of Overcoming the Limitations of Mutual Coupling and Narrow Bandwidth in Superdirective Arrays

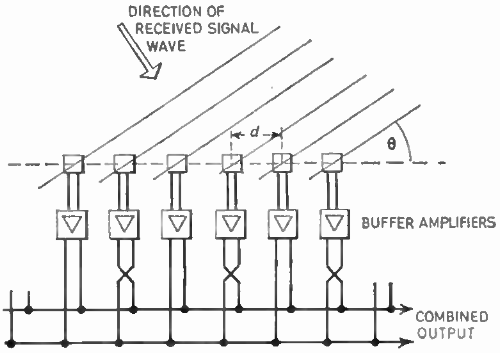
It seems that neither the mutual coupling effect nor the narrow-bandwidth effect discussed above are really fundamental in the design and operation of superdirective receiving arrays, since both can be kept exactly the same as in the corresponding non-superdirective array merely by interposing buffer amplifiers (or any other suitable decoupling arrangements) between the elements and their electrical interconnections as shown in Fig. 2(b). The taper function and its phase reversals can be inserted without affecting the mutual couplings and radiation impedances at all. The bandwidth of the superdirective array is then exactly the same as that of the same array used without superdirectivity.† The large reactive field in the vicinity of the array does not now exist, since the currents or vibrations in the elements are now co-phasal. Yet the directional response of the array is undoubtedly the same as that of the super-

---

† This is not to say, of course, that non-superdirective arrays with very close element spacings of only a fraction of a wavelength do not have complicated performance phenomena. Rusby[6] gives an interesting example, where mutual couplings in a closely-spaced rectangular array cause one element to have a negative radiation resistance.



(a) With direct electrical coupling.



(b) With buffer amplifiers.

Fig. 2. Superdirective array connections.

directive array of the conventional type (if the latter could be successfully constructed).

The writer cannot see any rigorously corresponding system or reasoning for a transmitting array and it is possible that reciprocity fails here. But an approach to a reciprocal system can be obtained by using a separate transmitting amplifier for each element, with each amplifier output approximating to a constant-current drive, so that each element has an amplitude largely independent of the oscillation of other elements.

It is interesting to note that another system for overcoming the effects of mutual coupling (and incidentally also the reactive field) in receiving arrays has been proposed[9] and tried experimentally.[10] This system is not of general application, however, as it uses different frequencies for the signals on the different elements, and thus requires the 'co-operation' of the sending end.

3. Analysis

A simple analysis of the system of Fig. 2(b) is as follows: Assume that the number (r) of elements in the array is fairly large, and that the complex coupling coefficient (expressed as an equivalent voltage-transfer ratio) is m_1 between adjacent elements, m_2 between elements spaced by two inter-element distances, m_3 by three, and so on. Let v_n be the voltage output of the nth element when the incoming wave is not allowed to excite any other element, and v'_n the voltage when all elements are excited. Then when all elements receive the wave, their outputs are each given by

$$v'_n = v_n + m_1(v'_{n-1}+v'_{n+1}) + m_2(v'_{n-2}+v'_{n+2}) + \\ + m_3(v'_{n-3}+v'_{n+3}) + \ldots\ldots(1)$$

where the terms continue until the subscripts of the v' terms reach 1 or r. It is easiest to assume at this stage that all elements have equal sensitivity and that the taper function (T_n for element n) is imposed after the buffer amplifiers.

Now a rigorous solution of the set of n equations represented by eqn. (1) is evidently very complicated, and it is quite unnecessary for our present argument. We assume that the signal is a coherent plane wave coming from a direction making an angle θ with the normal to the array, and we put

$$\phi = (2\pi d/\lambda)\sin\theta \qquad \ldots\ldots(2)$$

where d is the spacing between the centres of the elements and λ is the wavelength. Let the magnitude of v_n resulting from this signal be v_S for every element. Then just to see what form the solution of eqn. (1) should take, assume for a moment that the modulus of v'_n is the same for all values of n, and that the phase interval ϕ applies equally to adjacent values of v' as to adjacent values of v. Then as a first approach to the solution we may derive from eqn. (1) this relationship:

$$v'_n = v_S \exp(jn\phi)/ \\ [1-2m_1\cos\phi - 2m_2\cos 2\phi - 2m_3\cos 3\phi - \ldots]\ldots(3)$$

where the series in m_1, m_2, etc. continues only according to the relation of n to the end elements 1 and r; residual terms in the series are of the form $m_k \exp(\pm jk\theta)$. We see, therefore, that the true solution is of the form:

$$v'_n = v_S \exp(jn\phi) . A(n,m,\phi) \qquad \ldots\ldots(4)$$

where $A(n,m,\phi)$ is a function of the m-values and of ϕ which is different for each value of n.

The total output after combining the signals from the buffer amplifiers, with the polarity reversals as indicated in Fig. 2(b), is proportional to

$$v_S \sum_{n=1}^{r} A(n,m,\phi) . (-1)^n . T_n \exp(jn\phi) \qquad \ldots\ldots(5)$$

Now if the number of elements is infinite, A is independent of n and is

$$A(m,\phi) = 1/[1-2\sum_{k=1}^{\infty} m_k \cos k\phi] \qquad \ldots\ldots(6)$$

In other words, the asymmetry of the mutual coupling towards the ends of the array is ignored when the array is infinite. Since in practice the values m_1, m_2, ... m_k, ... can be expected to diminish fairly rapidly as k increases, eqn. (6) is probably an adequate approximate representation for arrays with as few as six or seven elements. We can then write the total output as

$$v_S . A(m,\phi) . D(\phi) \qquad \ldots\ldots(7)$$

where $D(\phi)$ is the superdirective directional function,

$$D(\phi) = \sum_{n=1}^{r} (-1)^n . T_n \exp(jn\phi) \qquad \ldots\ldots(8)$$

It should be noted that, since $d < \lambda/2$ in a superdirective array, the term $A(m,\phi)$ may well have little effect on the overall directivity of the array; we would normally expect $D(\phi)$ to be the dominant directional response.

It is clear that $D(\phi)$ is the only part of the response shown in eqn. (7) which is affected by making the array superdirective, and there is no basic frequency limitation in $D(\phi)$. It is, of course, necessary that, over the frequency range in which operation is intended, the large responses (shown in Fig. 1) which occur for $|\sin\theta| > 1$ should not inadvertently enter the range of real angles. This can be avoided by suitable design as shown in exaggerated form in (b) in Fig. 1, where a large margin for increase in frequency has been allowed. The condition of (a) in Fig. 1 should occur only at the lowest frequency.

It is probable that the array using buffer amplifiers, while enabling the desired superdirective directional pattern to be obtained, actually gives a lower gain. Here 'gain' is distinguished from 'directional gain' or directivity as being the relation between output from the array and the strength of the incident field, for normal incidence. For the conventional superdirective array, but with loss-free elements, Schelkunoff and Friis[1] state that the resonance effect and low radiation resistance 'enable the antenna to create a strong reactive field extending to large distances from the antenna which re-directs the power passing through a large area of the incoming plane wave and forces it to flow toward the antenna. Detuned superdirective antennas intercept but little power'. This explains how, in a loss-free array, superdirectivity leads to an actual gain corresponding to the directional gain (or, in acoustic terms, to the increase in directivity index) over the same array used co-phasally. Thus the terms 'superdirectivity' and 'supergain' have tended to become synonymous. Under practical conditions the reactive field will not be so strong and effective, since the resistance losses in the elements are arithmetically additive and will tend to swamp the other effects. This is especially true of acoustic arrays, where efficiencies are usually much lower than in radio arrays. Experimental results are available for end-fire arrays, but the author knows of none for broadside arrays. In a radio end-fire array operating at 75 MHz, an actual gain (relative to a non-superdirective arrangement) of over 4 dB has been reported by Bloch, Medhurst and Pool[2] for a four-element array.

There is no doubt that the decoupled superdirective array cannot give an actual gain as compared with the same array used co-phasally, and indeed must give a loss and cannot therefore be called a supergain array, even though its directivity index is the same as that of the normal superdirective arrangement. It is open to argument whether the decoupled array gives less or

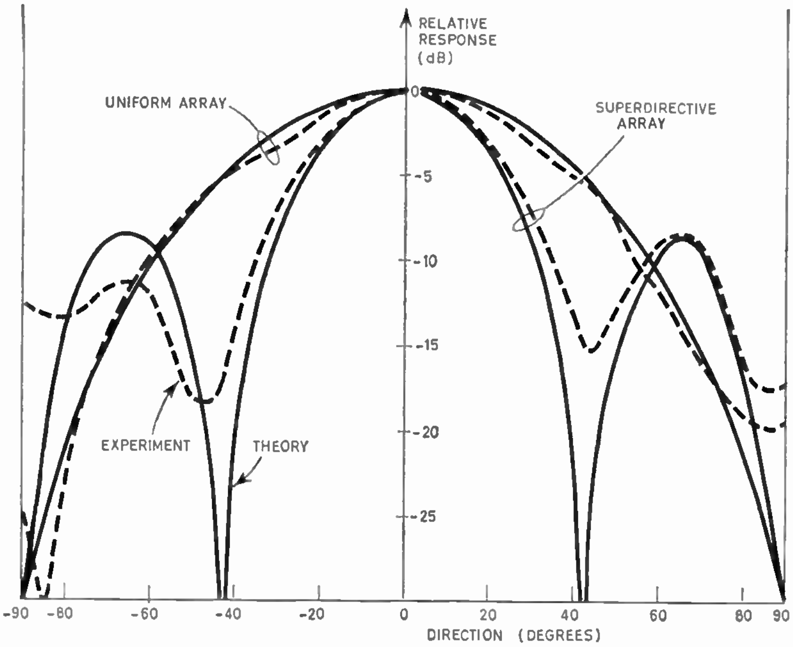


Fig. 3. Comparison of experimental and theoretical results for a 420 MHz array of three radio dipoles spaced at quarter-wave intervals. Buffer amplifiers were used as in Fig. 2(b). The superdirective array merely had the phase of the output from the central dipole reversed before adding to the others.

more gain than the normal superdirective array used without the reactive field effect being tuned out, thus also having a less restricted bandwidth. However, the use of decoupling certainly appears to simplify the design problem. The noise figure of the decoupled superdirective array may be somewhat worse than that of the normal superdirective array, but it is unlikely that the increase will be very significant in practice.

4. Experimental Results

No full experimental investigation of the proposals discussed above has yet been made and no experimental verification of the wider bandwidth which is predicted for the superdirective array is yet available. Some preliminary experiments have, however, been made by Mr. M. S. Pollard,[11] an undergraduate student in the author's department, and the results show that, at any rate, a simple superdirective array using buffering arrangements as in Fig. 2(b) does give the predicted directional pattern. The experimental arrangement comprised the central three half-wave radio dipoles in an array of seven (the others being terminated in dummy loads), spaced at one quarter-wavelength intervals, and operating at 420 MHz. The array was backed by a ground plane. The dipoles were connected via baluns, buffer amplifiers and hybrid arrangements to correspond in effect exactly to Fig. 2(b); the phase reversals could be removed when desired by suitable connections to the hybrids.

The three active elements had equal gains in the amplifier and hybrid systems, and if the outputs were connected *in phase*, a uniform normal array should be obtained. Allowing for the ground plane, the directional pattern should theoretically be as shown in the appropriate full-line curve in Fig. 3. The experimentally-measured pattern shown in the dashed curve agrees very well.

If the phase of the output of the central element is reversed, a simple superdirective system should be obtained which theoretically (allowing again for the ground plane) should have the directional response shown in Fig. 3 by the other full-line curve. This pattern is very much narrower in the main beam than that of the uniform array, but has bad secondary responses that would probably render it unattractive in practice. Nevertheless it makes a suitable basis for a test of the principle. The experimental pattern is shown in the appropriate dashed curve, and can be seen to be in reasonable agreement with the theoretical curve except that it has no real nulls. Thus there is no doubt that a superdirective response can be obtained this way and the discrepancies between theoretical and experimental patterns are not appreciably worse than in the case of the uniform array in spite of the very large mutual couplings. (The mutual impedances appeared, from indirect measurements, to be of the order of one-half the self-impedances.)

5. Conclusions

It has been shown that the design and achievement of superdirective receiving arrays can be greatly eased by the use of decoupling arrangements between the elements, and that by this means the severe bandwidth restrictions normally associated with superdirectivity can be removed. It is believed that the arguments are equally valid for radio or acoustic arrays. A distinction between superdirectivity and supergain has been made, and the decoupled array, while having the former, cannot have the latter.

6. Acknowledgments

The author acknowledges the help received in discussions with Dr. D. E. N. Davies and Mr. M. Mellors and from the experimental work of Mr. M. S. Pollard.

7. References

1. See, for example, S. A. Schelkunoff and H. T. Friis, 'Antennas', (John Wiley, New York, 1952).
2. A good bibliography on superdirectivity is given in A. Bloch, R. G. Medhurst and S. D. Pool, 'A new approach to the design of superdirective aerial arrays', *Proc. Instn Elect. Engrs*, **100**, Part III, p. 303, 1953. A supplementary bibliography was given by the same authors in 'Superdirectivity', *Proc. Inst. Radio Engrs*, **48**, p. 1164, 1960.
3. S. A. Schelkunoff, 'A mathematical theory of linear arrays', *Bell Syst. Tech. J.* **22**, p. 80–107, January 1943.
4. P. M. Woodward and J. Lawson, 'The theoretical precision with which an arbitrary radiation pattern may be obtained from a source of finite size, *J. Instn Elect. Engrs*, **95**, Part III, p. 363, 1948.
5. D. G. Tucker, 'Signal/noise performance of superdirective arrays', *Acustica*, **8**, p. 112, 1958.
6. J. S. M. Rusby, 'Investigations of an interaction effect between sound projectors mounted in an array', *J. Brit. Instn Radio Engrs*, **25**, p. 295, April 1963.
7. C. H. Sherman, 'Analysis of acoustic interactions in transducer arrays', *Trans. Inst. Elect. Electronics Engrs on Sonics and Ultrasonics*, **SU-13**, No. 1, p. 9, March 1966.
8. See, for example, the first paper cited in Ref. 2.
9. D. G. Tucker, 'Space-frequency equivalence in directional arrays', *Proc. I.E.E.*, **109** Part C, p. 191, 1962. (I.E.E. Monograph No. 479E, November 1961.)
10. B. S. McCartney, 'Theoretical and experimental properties of two-element multiplicative multi-frequency receiving arrays including superdirectivity', *The Radio and Electronic Engineer*, **28**, p. 129, August 1964.
11. M. S. Pollard, Unpublished report, Department of Electronic and Electrical Engineering, University of Birmingham, April 1967.

Note added in proof:

A referee has drawn attention to the following very interesting and closely related report (apparently still unpublished, but available through recognized channels):

> E. Spitz, 'Supergain and Volumetric Antennas', Compagnie Générale de Télégraphie Sans Fil (CSF), Paris, 15th June 1959. Contractors to Air Force Cambridge Research Center, U.S.A., ref. AFCRC-TR-59-194.

Manuscript first received by the Institution on 1st February 1967 and in final form on 1st June 1967. (Paper No. 1151/RNA76.)

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied.

ECHO FORMATION OF RADAR AND SONAR TARGETS

The echoing effect of a sonar or radar target is described by its echoing area. This is a measure of the integrated effect of all partial echoes from a target of finite extension, but does not provide a statement on the transient series of partial echoes. Modern sonar or radar equipment employing coded transmission signals and pulse compression (matched filter operation) after the reception may under certain circumstances permit the resolution of individual partial echoes from a target with respect to magnitude and time of occurrence. Thus the generally accepted term 'echoing area' loses its significance and the problem is to investigate the formation of the target echo. This formation of the target echo can be explained as a convolution of the transmission signal used and the weighting function of the target. One essential task of the theory of echo formation is therefore the derivation of the weighting function of various targets.

In a German paper the weighting function of the target is presumed to be known and the convolution integral for an arbitrary transmission time-function is solved. This provides a relationship between the received time function and the transmitted time function, the so-called equation of the echo centres in analogy to the radar equation but with a much higher information current.

Specifying the equation of the echo centres for the case of monochromatic transmission signals and reverting to power in the calculation of the time functions provides a relationship between the weighting function and the echoing area of the target.

A few simple examples confirm the validity of these considerations and form a link with known facts.

'The relationship between echoing area, target coefficient and weighting function for sonar and radar targets of finite extension', G. Ziehm, *Nachrichtentechnische Zeitschrift*, **20**, No. 8, pp. 441–48, August 1967.

MEASUREMENT OF HALL E.M.F.

A method for determination of the Hall e.m.f. which is based on the measurement of current in the Hall circuit, is proposed in a Soviet paper.

The measurement of the Hall e.m.f. is carried out, as a rule, using compensating circuits or with the aid of electrometer amplifiers having a high input resistance. However difficulties due to the existence of a potential drop on the resistance of the sample appear, even for a small selected current, when the Hall e.m.f. measurement is carried out on high resistance samples. To circumvent these difficulties special high resistance potentiometers and tube type electrometers are devised, allowing the measurement of small potentials on high resistance samples with a small selected current. Such schemes are somewhat complicated and inconvenient for use in work. In addition, for fast measurements with the recording of measurements, these methods are either useless or lead to a significant complication of the measuring circuits. It should also be noted that these methods are of no use in most cases for the measurement of the Hall e.m.f. in the pulse régime.

'Several methods of measurement of the Hall e.m.f.', Y. F. Ogrin and V. N. Lutskiy, *Radio Engineering and Electronic Physics*, (English language edition of *Radiotekhnika i Elektronika*), pp. 276–84, No. 2, February 1967.

REMOTE CONTROL OF COMMUNICATION SYSTEMS

For carrier-frequency long distance communication systems it is an advantage to be able to monitor the operational condition from a centre. The occurrence of a fault should be signalled automatically to the centre and a fast fault location and detection should be possible by means of supervisory equipment. The conditions to be monitored are system faults such as failures of power supplies for racks or remotely fed repeaters, pilot-tone failure due to a fault in the high-frequency transmission through amplifiers and other equipment as well as faults outside the system such as mains failure or unauthorized opening of doors in unattended repeater stations.

During the monitoring process the monitored sub-stations are called up by pulses from the centre and made to transmit an answer signal, which includes a statement in coded form of the type and location of the fault. The call and the answer process is carried out by means of both receiver counters which are installed in all sub-stations and also a comparing counter in the centre.

A German paper describes the equipment in which the fault location is visually indicated at the centre by means of a few panel lights arranged in the manner of a co-ordinate system. Also the type of fault is automatically indicated by panel lights. The supervisory equipment contains no electro-mechanical circuit elements and thus permits very fast fault recognition.

'A proposal for remotely monitoring carrier-frequency long distance communication systems', J. Korn, *Nachrichtentechnische Zeitschrift*, **20**, No. 8, pp. 453–61, August 1967.