

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 38 No. 1

JULY 1969

The Exhibition Season

DURING the months of March to June each year the electronic engineer is faced with an almost embarrassing number of exhibitions, 'open days', etc., in Great Britain and in Western Europe. Exhibitions of course have both commercial and technical objects, but their full potential value as sources of technical information is perhaps difficult to realize. Even when the decision on *which* exhibitions to visit has been made, all too often the time available means that the further selection of the exhibits to be examined tends to be a rather restrictive exercise: the equivalent of 'browsing' through the pages of a book or journal can be a very lengthy process for the engineer with broad interests. It is, however, entirely practicable for a single-minded tour of an exhibition to be contrived in order to find out the 'state-of-the-art' in a specific, limited field, for instance, phase meters, v.h.f. receivers, Zener diodes, to quote three headings at random from exhibition catalogues.

A very few years ago integrated circuits could have formed the object of a day's visit to any of the major components shows. Now, however, both in Paris and London, the rapid extension of these techniques is everywhere apparent, fostered in many applications by the versatility of metal oxide semiconductor technology. Perhaps the most significant development, and one easier to assess than the 'fashions' of logic circuits and the like, is the introduction of printed cabling in a really compact form which will match the inherent compactness and reliability of monolithic and hybrid integrated circuits. Indeed, constructional techniques generally are undergoing a minor revolution to achieve greater automation and reduce calls on skilled manpower: the electro-hydraulic cable jointer developed for the British Post Office by Plessey shown at both exhibitions is an interesting technique noted this year.

Although primarily concerned with components, both the Paris and London exhibitions included a fair number of stands carrying instruments and other professional equipments. Here there was an evident overlap both generally and in detail with other exhibitions, for instance, INEL in Basle, the German exhibition in Hanover, and the S.I.M.A. exhibition in London. The last named was held in association with the Physics Exhibition organized by the Institute of Physics and the Physical Society which represents a type of venture that is unusual in that the emphasis is on novelty of concept rather than on fully engineered commercial equipment. Nevertheless, the Physics Exhibition is almost 'required viewing' for the engineer since its presentation of new ideas has on many occasions led to highly successful commercial exploitation. Laser devices, Gunn-effect microwave generators and precision measuring techniques are cases in point that have been of interest to the electronic engineer in recent years.

Space does not permit more than a mention of such events as the 'open days' at Government research laboratories in Great Britain, for instance the National Physical Laboratory, Teddington, and the Royal Radar Establishment, Malvern, which have the character of a rather special kind of exhibition. Or of the extensive exhibition which supported the Joint I.E.E.-I.E.R.E. Conference on 'Computer Aided Design' in Southampton.

These then are some of the ways in which an engineer could have profitably occupied himself during the past months. Exhibitions were referred to in this *Journal* some years ago as providing a 'mirror of progress', and this is certainly an apt simile. Even the truest mirror, however, can only reflect and any inspiration derived from it must come from the viewer himself: but put in more down-to-earth terms the exhibition season does provide a review of achievement and an incentive to further progress.

F. W. S.

INSTITUTION NOTICES

Birthday Honours

The Council has congratulated the following members whose names appear in Her Majesty's Birthday Honours List.

On the conferment of the Honour of Knighthood (Knight Bachelor):

Raymond F. Brown, O.B.E. (Companion)
(Head of Defence Sales, Ministry of Defence).

On his appointment to be an Ordinary Member of the Most Excellent Order of the British Empire (M.B.E.):

Oswald B. Kellett (Fellow)
(Regional Wireless Engineer, Home Office).

Chartered Engineers in Canada

On 7th February last, a Dinner was held in Montreal by the Chartered Engineering Institutions Information Committee (CEIIC). It was attended by 280 engineers originally from the United Kingdom or the Commonwealth, an impressive proportion of the 1300 chartered engineers in Quebec who have retained their membership of various U.K. Chartered Engineering Institutions.

The CEIIC is composed of senior members and representatives of membership groups of some of the fourteen Chartered Institutions which together represent all engineering disciplines. It was formed in 1966 to work with the Corporation of Engineers of Quebec (CEQ) Commission on Admission Standards. The work done by the Committee at that time indicated that adequate representation by specific engineering groups was not only desirable but essential.

The CEIIC Committee is made up of Canadians with original backgrounds of education and training in the U.K. and Commonwealth. The present Chairman is Mr. E. F. Wale, C.Eng., F.I.E.R.E., and Mr. K. N. Coppack, C.Eng., M.I.E.R.E., is also a member.

Whilst it is representative of members of the British Chartered Institutions and immigrant engineers, the Committee wishes to reflect a 'Canadian' public image. The recent dinner meeting was a reflection of this aim. The theme of the meeting was 'The responsibility of the engineer in Canada's development', and the guest speakers were the Hon. Bryce MacKasey, Federal Minister of Labour, introduced by Brig.-Gen. J. P. Carrière, President of the Engineering Institute of Canada, followed by CEQ President, Charles Laferrière.

The CEIIC has been concerned with the problem of professional recognition for chartered engineers, particularly certain aspects of CEQ admission standards and examination criteria. The Canadian citizen-

ship requirement has long been a problem and whilst this is now being abolished, the problem of examination criteria still exists and will require more study and follow-up by the CEQ and CEIIC. It is felt that the standardization of the C.E.I. examination requirements in Great Britain should provide a better basis for registration of immigrant engineers and should remove much of the confusion and frustration that existed in the past.

Institution Dinner

The 1969 Institution Dinner will take place at the Savoy Hotel, London, on Thursday, 27th November. Members of all grades can attend and a limited number of male guests may be invited. A form of application for tickets will be included in the next issue of the *Journal*. The Dinner will be under the chairmanship of the President, Mr. Harvey F. Schwarz.

Forthcoming I.E.R.E. Conferences

'Industrial Ultrasonics'—(University of Loughborough, 23rd–25th September, 1969). The outline programme will be published in the August issue of the *Journal*.

'Automatic Test Systems'—(University of Birmingham, 14th–16th April, 1970). The proposed scope of this Conference will be published, with a 'call for papers', in the August issue.

Requests for registration forms for both these Conferences may be made on the application form on page (xi) of this issue.

List of Members

Members are advised that the new List of Members is now ready and that copies may be obtained from the Institution price 10s. 6d. each post free. This edition, the 12th, is corrected up to March 1969 and contains the names of all Corporate Members and Companions, as well as the membership of Standing and Group Committees. An order form is given on page (ix) of this issue of the *Journal*.

Courses in Electronics

This issue of *The Radio and Electronic Engineer* is accompanied by an advertisement supplement containing details of courses in electronic engineering and allied subjects which are being arranged by universities, polytechnics, and colleges of technology in the British Isles.

Although copies are only being sent to members and subscribers in Great Britain, Northern Ireland and the Republic of Ireland, members in other countries may obtain the supplement at a charge of 3s. 6d. post free; extra copies may be obtained at the same rate.

Current Developments in Computer Hierarchies for Industrial Control

By

S. L. H. CLARKE,
B.A., C.Eng., F.I.E.R.E.†

AND

C. AYERS, B.Sc. (Eng.),
C. Eng., M.I.C.E., M.I. Mech. E.,
M.I.E.E.‡

Reprinted from the Proceedings of the I.E.R.E. Convention on 'Electronics in the 1970s' held in Cambridge on 2nd to 5th July 1968.

Summary: The paper reviews the present trends in the use of computers for process control through the combined experiences of the activities of the authors' organizations in the field. Examples quoted showing the wide range of application which has to be covered include integrated production control, hierarchical control and direct digital control. Conclusions are drawn about the effect which the requirements will have on the design of suitable equipment in the next few years.

1. Computer Hierarchies

Hierarchies have been proposed for computer control for many years. The ARCH philosophy was expounded in 1959¹ and the computer system at Park Gate Iron and Steel² and Richard Thomas and Baldwin³ were early examples of hierarchies in practice. Large monolithic installations have also been proposed which deal with all operations from direct control of plant to the payroll of the staff, including complex computations to optimize the operation of individual processes and of the complete plant.

In practice the application of simple direct control of plant has proved difficult enough to constitute a field of its own. Optimization has proved to be even more difficult and only a disappointingly small number of systems have been successfully completed. This is mainly due to the lack of knowledge of plant dynamics and to underestimating the difficulties in measuring vital plant parameters and other interface problems. These problems are slowly being overcome and the gap between elegant theoretical optimization methods and practice in the hard environment of process plant is closing, but still the difficulties are such that they must be kept organizationally separate. This, in itself, does not preclude the use of a single computer system for multiple tasks each isolated in the software sense. The need for system security coupled with the reducing costs of individual processor units and the predominance of peripheral costs will however lead progressively to a hierarchy system of control.

In a hierarchy system there is a tendency to utilize small computers at the lowest level, replacing conventional instrument and relay control schemes. These computers require short word-lengths in order

to minimize the number of components in the system. This fulfils the dual necessity of reducing cost and increasing reliability. At the upper levels in the hierarchy the central processors of the system become increasingly like the conventional business machine particularly when tape files, line printers and all the commercial software edifice are necessary.

At the present time there are very few installations where all levels of a hierarchy are in simultaneous operation. This paper describes applications where various levels of control are being performed in conjunction with manual, or conventional automatic controls at the other levels.

The first group of examples are primarily concerned with the levels above the control of individual loops and process units, whereas the second group are confined to the lowest level of control and monitoring.

2. Plant Co-ordination Level

In considering these systems the control hierarchy may be considered as consisting of three basic levels. Adjacent to the plant is the base level of process control; above this is the area of co-ordination control, and the uppermost level may be regarded as the strategic level of control.

The overall strategy of control as defined by the strategic control section of the system passes the desired operational plan to the co-ordination control level. This overall strategic plan may refer to a single industrial unit or to a group of such units such that the planned instructions are given to different sections of the same process plant or to separate process plants. The instructions received by the co-ordination control level are actioned in accordance with the rules and requirements of the complex in order that the process plant may perform the desired plan. The base level of control, process control, is charged with maintaining the various process plant items in correct operation to the desired pattern given by the co-ordination control.

† A.E.I.-Elliott Process Automation Ltd., Elstree Way, Borehamwood, Herts.

‡ Power and Marine Division, English Electric-A.E.I. Ltd., Kidsgrove, Staffordshire.

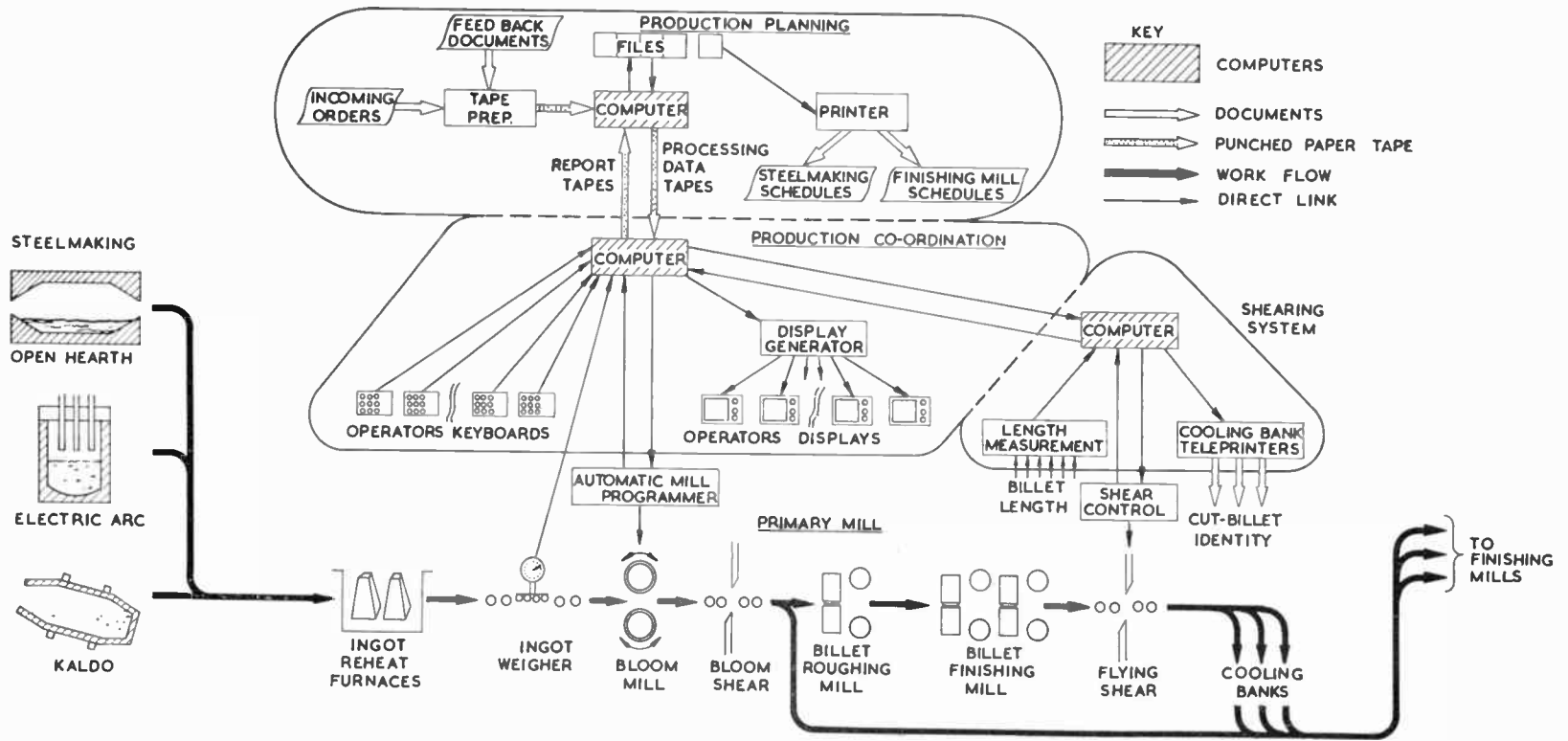


Fig. 1. Integrated production control at the Park Gate Iron and Steel plant.

Examples from two basic industries will be given and in the case of one of these industries two solutions are cited.

The first example is taken from the steel industry in which the control hierarchy takes the following form. (See Fig. 1).

The strategic control is given the task of production planning whose aim is to organize the available plant material and manpower resources to meet the customer's order requirement in the most effective manner. In doing this it is charged with the examination of the incoming orders, the preparation of a production plan and the adjustment of this plan to cover changes in the production situation within the plant based on information received from the lower levels of control.

The co-ordination control is usually referred to as production co-ordination and this is concerned with the various plant areas to ensure that the operation of each individual area is in accordance with the strategic plan and that continuous operation by the process plant is possible and performed effectively. This is necessary as the output from one part of the process forms the input of the next section of the process etc.

The process control level is concerned with the current operation of the various items of plant to ensure that the operational requirements are met.

It is obvious that there has to be a two-way flow of information from each level of control in the hierarchy as decisions have to be fed from a higher level to a lower level and information has to pass in the reverse direction indicating whether or not the decisions have been correctly performed. Such information is then used to modify the tactics or strategy of control.

As an example of such a system consider the Dynamic Production Control System at the N.H.K.G. Steelworks in Czechoslovakia which is a large integrated steelworks with an output approaching 3 million tons per annum. The works comprise blast furnaces, open-hearth steel making shop with ten furnaces, primary mill area with two reversing bloom mills and three continuous billet mills and a variety of finishing mills producing narrow and medium strips, sections, rods and wire.

The control system comprises two separate but linked digital computer systems, a LEO 360 for production planning and other data processing work and a KDF 7 for real-time, production co-ordination and control in the roughing mill area of the plant. The production planning function involves the receipt and recording of orders, their assembly into a forward plan and the calculation of deliveries and forecasting of raw material requirements. The planning schedules

are transmitted to the production co-ordination computer by a direct link and to the steel plant, semi-finished stockyard, and finishing mills by printed documents. Information feed back from the co-ordination system is used by the planning system to maintain an up-to-date picture of the progress of each order and to enable modifications to the planning schedule to be made. In addition to the above the production planning computer is used for payroll, sales invoicing, material stock control, accounting, quality control and general management statistics.

The on-line production co-ordination system receives the detailed schedules proposed for the primary mill area covering 24 hours' production prepared by the planning computer. It also receives a copy of the cast lists prepared by this computer and allocates casts to specific orders. Following completion of casting the on-line computer checks the allocation and re-schedules as necessary. It then schedules the order in which ingots are to be charged to the soaking pits and controls the heat input to each pit to ensure that ingots reach the correct temperature when required by the bloom mill. It also prepares the rolling programme for the primary mill for 2 hours' production and tracks and records the progress of pieces through this area. This is of great importance as there are 65 different routes through the roughing mill area and a total of 865 possible combinations of product and route. The computer calculates and displays the optimum lengths to be cut at the shears and saw and directs the control of the flying shear to achieve maximum yield. Throughout the production area advance plans are typed out and processing instructions are displayed on tabular display units. At any stage should production not go according to plan either because of human error or equipment failure, the computer is informed and it re-schedules the rolling program etc.

The on-line computer passes to the planning computer a record of production actually achieved to allow future plans to be considered. The entire system operates under the general supervision of a human production controller who may intervene should any special decision be required or change of plan enforced.

The second basic industry cited as an example is that of electricity supply and in this industry also the control hierarchy can be seen to be utilized. The strategic control may be considered to be concerned with the calculation or measurement of load to be met by the generating stations and the calculation of the individual generating station contributions taking into account all the factors that affect economic load dispatch and transmission system utilization. The co-ordination control may be considered segregated and as being located at each generating station.

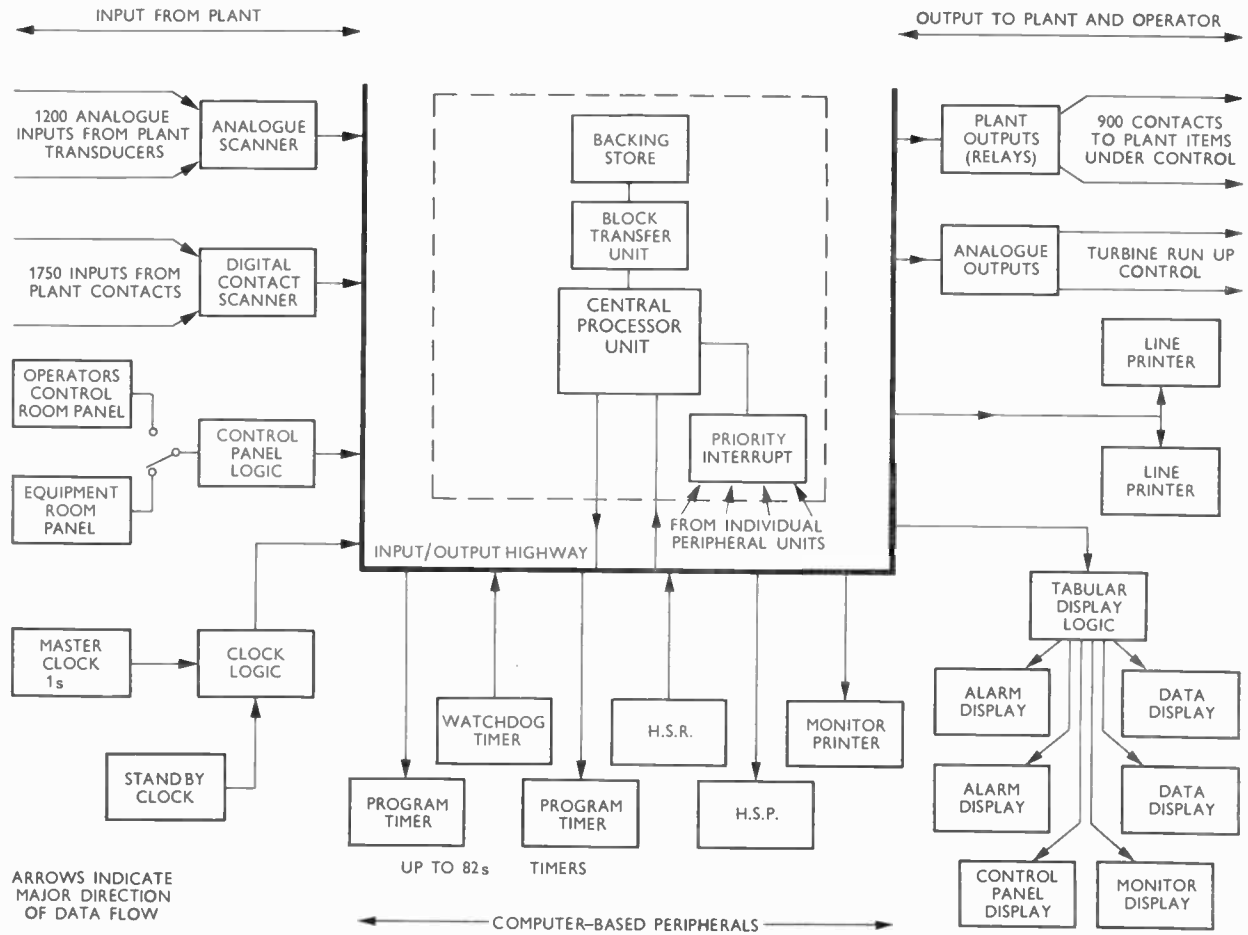


Fig. 2. Plant control hierarchy at the Fawley Electricity Generating Station.

receiving its instructed load and arranging that this is met by the units and boilers sited at the station. The process control level may be regarded as the controls etc., associated with the individual plant items necessary for the operation of the generating station units and services. The two examples to be quoted cover two approaches to the solution of the co-ordination control problem.

The first concerns the KDF 7 computer control system applied to each of the 500 MW generating units at the Fawley Power Station (Fig. 2). This computer system is used to control the generating unit throughout its operational regime from start-up through load operation to shut-down. The system is used in an on-line mode to effect the control of the various auxiliary plant items and to co-ordinate the operation of these plant items in the event of faults arising in the plant. The system gathers information from the plant by means of an analogue scanner, covering 1200 points and a digital scanner covering 1800 points. Control information is fed back to the

plant by an output suite covering 900 individual outputs. The operation of the system is basically as follows.

Assuming the plant is shut down after having operated for some time and has been shut down for a period of some 30 to 60 hours, the operator requests the system to take the plant from this state to a pre-determined load at a stated time. Following these instructions, the system gathers information on the current state of the plant, analyses this information, and based on its findings commences to start up the various auxiliary plant systems such as the boiler fans, turbine oil pumps, extraction pumps, vacuum pumps etc. It does this in the correct order to enable firstly the boiler to be started and put through its pressure-raising procedures. Next the turbine is prepared to accept steam, and at the appropriate time the boiler and turbine are matched and the turbine acceleration program commenced, resulting eventually in the unit reaching the pre-determined load at the appointed time. During the whole of

these operations, any abnormality or fault condition occurring on the main plant is immediately recognized by the system, analysed and the necessary corrective or remedial action taken to ensure that the start-up procedures are completed correctly. In such cases the change of operational pattern is advised to the operator by means of cathode-ray displays, and printed so that he is fully aware of the state of the plant and can advise the maintenance engineers to repair the defective plant. Having reached the operational load required, the system continues to interrogate the plant, looking for trouble, and if faults or abnormalities arise it analyses these to determine the prime cause of the trouble and then sends instructions to the plant to correct the fault so as to maintain generation at the desired load. The operator is advised that changes have occurred by the appropriate displays and printed records. Should the fault be serious, such that further operation of the plant may constitute a danger, the systems shuts down the plant in a controlled manner and advises the operator that this has been done.

At pre-determined intervals of time, the system calculates the efficiency of generation based on 15-minute averages of plant data, and compares these efficiencies with test figures. Any discrepancy between the two figures is then further analysed by considering each individual plant item in turn to determine the cause of the inefficiency, the operational cost of this inefficiency and to advise the operator of the action necessary, such as cleaning of condensers, boiler plant, or the modification of the operational pattern such as by-passing various stages of the feed heating train to allow cleaning operations and routine maintenance to be carried out to improve the operational efficiency. In addition, the operator may request operational data from the system at any time, and these data will be presented to him in a correlated fashion to enable him to control the plant more effectively and to allow him to optimize the plant state. From the above it will be seen that the computer system employed at Fawley takes over some of the facets of plant control in the process control section of the control hierarchy. This is to be expected due to the type of process involved where an integrated plant is used to manufacture one product (electricity) from one or two raw materials (fuel and water) so that the process is a single continuous process.

The second approach to this problem indicates that it is possible to consider these problems from more than one standpoint. This example concerns the solution adopted at the Pembroke Generating Station housing four 500 MW generating units. In this station one control computer system is used to monitor the operation of all four generating units. It does this by scanning continuously some 10 000

measurements (2500 per unit), comparing them with desired or alarm values. Should any abnormality or fault condition be detected this condition is immediately analysed to determine its cause. This cause is then displayed to the operator of the unit concerned by means of a tabular display system together with a statement of the action he should take to correct the faulty condition. This means that the system is mainly a monitoring and advisory system to the operator.

In addition, it prepared the operational and management statistics to aid the efficient operation of the units and station. In order to control the start-up and shut-down of the individual units a separate system for each unit is provided. This is a sequential logic system covering the preparation, start-up and shut-down of the plant and its auxiliaries. In addition the logical system provides certain facilities covering standby plant starting functions, protective system and safety interlocks and intertrips. The whole of this system is operated by a small number of push buttons allowing the operator to determine the basic start-up pattern.

It can readily be seen that the difference between this system and that previously described for Fawley is the removal of start-up features from the computer system and the provision of them by sequential logics which can rightly be identified with the process control level of control activity.

Considering the schemes described it can be seen that they both employ fairly large computer systems and that when this is the case the division lines between the various control activities exist only in the system designers philosophy. One computer may therefore do related functions in all control regimes. The actual amount of control vested in one system or one computer requires careful considerations in the light of the reliability, maintainability criteria etc., peculiar to the project being considered.

3. Process Control Level

The second group of examples are taken from installations where the operation of a single part of the process plant has been improved by the direct connection of a digital computer. Direct digital control (d.d.c.), replacing conventional instruments, flexible and reliable adjustment of a unit process, and the control of a telemetry system all represent the lower levels of the control hierarchy introduced step-by-step from the bottom.

The example of direct digital control is taken from the batch fermentation process of Dista Products making antibiotics (Fig. 3). The critical factors governing the yield of antibiotic material are the acidity, the rate of oxygen supply to the organism, the maintenance

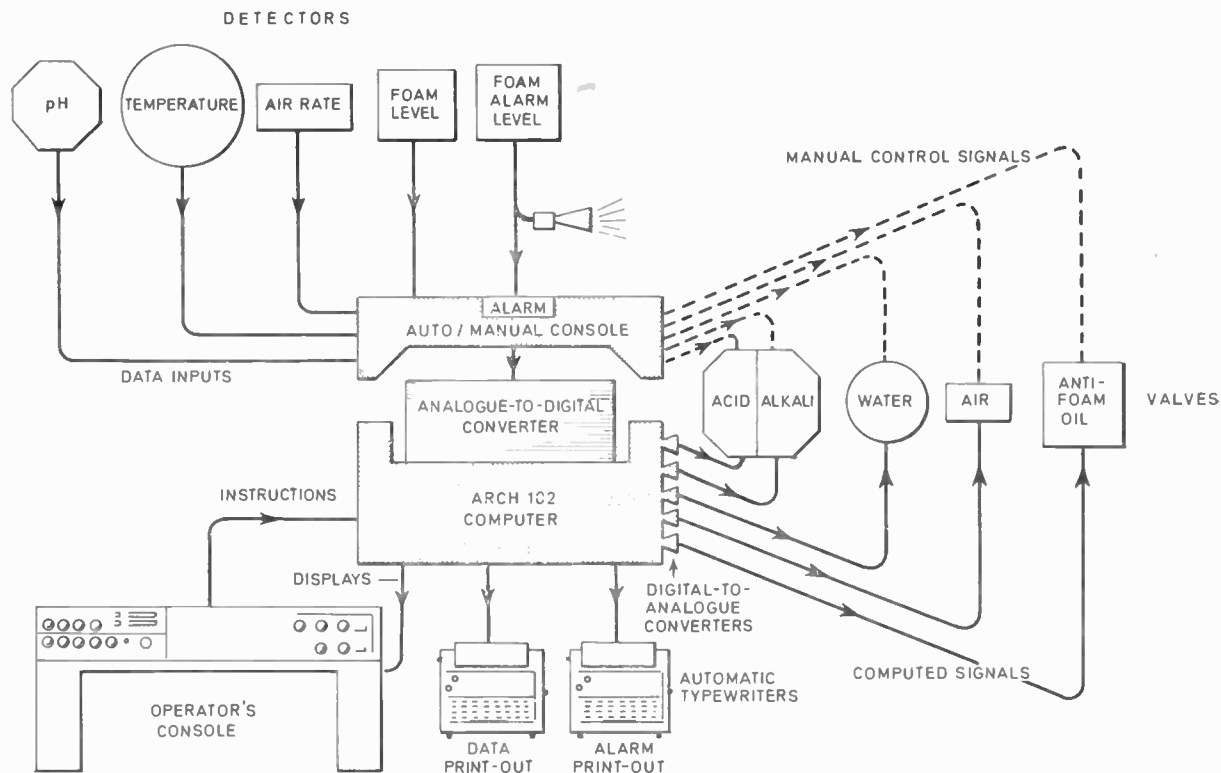


Fig. 3. Outline of the direct digital control structure for Dista Products fermentation process.

of correct temperature conditions, and the reduction of foam in the vessel. Thus in the major fermentation vessels all four factors are controlled, while in all fifty vessels, including the smaller seed vessels, the temperature is controlled and alarm levels are monitored. The computer in this case is an ARCH 102, very much smaller than those mentioned earlier in the paper, with only 13-bit word length. The advantages gained from the introduction of such a system do not lie directly in the economic justification of lower capital cost as was at one time hoped. These savings do go a long way to paying for the computer equipment and programming, but in the case of Dista Products, and most others, the final justification is going to come from additional facilities, improved control, or, where the intention is to proceed to optimization at a later date, the greatly reduced cost of this operation since the necessary inputs and outputs are already available in digital form.

Unlike the more sophisticated automation projects, the software for a d.d.c. application need not be special, and thus its cost can be kept low. However, to date the advantages of increased flexibility or better control have tended to mean a number of specialized algorithms rather than the straightforward two- or three-term control derived from conventional instruments. In this example neither the control of foaming,

nor acidity are normal control loops, whereas the air and temperature are. The control of temperature to a constancy of $\pm 0.1 \text{ degF}$ ($\pm 0.06 \text{ degC}$) have been described by process operators as 'uncanny' so great is the improvement on conventional electronic or pneumatic control. In addition to these aspects of control, the computers compare the alarm levels constantly giving the appropriate warning when alarm conditions occur, and also both routine and alarm logs are produced. The final benefit will come in the optimization of running conditions to increase yield, and reduced fuel costs. The relevant plant parameters are already fed into the computer, and changes in plant conditions can be made by changing values of loop set points and other loop constants in the computer store. Thus the extension of the system to include this optimization only involves extra storage and programming. If the optimization is to be considered as a separate level in the hierarchy it will be carried out in a second processor connected directly to the existing machine.

The capital cost of these extensions will be less than half the cost to embark on the optimization of a conventionally-instrumented and controlled plant.

One of the advantages the digital system will show in this second phase is that it is possible to schedule changes in the control constants at different stages of

each batch in the fermentation vessel. This kind of flexibility is needed in many kinds of plant. The changing of paper grade at the Wolvercote paper mill,⁴ and the 'turn-up' control of a naphtha reformer producing 'town's gas' for the East Midlands Gas Board, are operating examples of this kind of flexibility. Reformers of this type have given troubles in operation, and the E.M.G.B., using a new version for the first time, were most anxious to maintain the reliability of the plant under all circumstances. They accordingly placed first priority on using their ARCH 9000 for data reduction and presentation to the operating staff. However, another feature of reformers is their high efficiency when running at full output, but low efficiency when turned down. This coupled with dangers encountered during 'turn-up' led to the second phase of the operation of achieving higher efficiency at reduced throughput and more rapid, safe load changes. This necessitates more calculation than the basic d.d.c. application, and the 13 bits of the word in the simple machine was no longer adequate and the 18-bit version was chosen. The requirements for analogue input and output, the needs for operator panels and displays are the same in kind as those for Dista Products, and so can be met from the same range of process peripherals, the whole being driven by a different central processor.

A somewhat different situation is encountered when one considers the problem of distribution management. The supply of gas or water over the areas currently administered by Boards in Britain or municipality undertakings in Europe demands the gathering of data from instruments spread over hundreds and often thousands of square miles. These instruments are usually grouped together in pumping stations, gas holder stations or water towers, and the information transmitted to and from a master station over telephone wires or radio links. The transmission of analogue values over such links by different kinds of modulation has been practised for many years. This has required a one-to-one link with a receiving instrument with separate transmission for each instrument. More recently digital coding, first of all of tone channels, was introduced to switch between analogue values at an out-station, and thus to remove the need for one-to-one links. However, with the introduction of this digital coding it was only a small step to introduce truly 'digital telemetry' with the plant values digitized at the outstation. This meant that loss of accuracy could be contained by appropriate coding and error checking techniques in the supervisory equipment. If a transmission error occurred, the master station asks for a re-transmission of the affected measurement. Equally, a control message is not acted on until checks have been carried out to ensure that it was received correctly.

The task which the equipment at the master station has to carry out in conjunction with the outstations is almost exactly the same as that of a process control data logger and remote control panel. The only differences are that instead of one or two independent scanning units there are now many, and that a considerable amount of checking and coding has been added to the system for security of data acquisition and control. The state of the network at each outstation must be presented to the controller and he must be able to effect such changes as he wishes directly from his panel. In many cases it will be necessary to integrate flows and very rapidly it becomes evident that a digital computer may be the answer to the master station equipment. From the manufacturers' viewpoint, the ability to meet many customers' needs with the same equipment is a desirable move. The customer's ability to evolve his distribution network both physically as the supply and demand pattern varies and also for improved control, becomes vastly greater by the introduction of a stored-program device. This technique is being used in gas, water, oil and electricity undertakings. For this lowest level of telemetry control the smallest machine is again quite suitable but, as in the case of normal process control, the presence of the digital computer in the system opens up the possibility of further control with only small incremental cost. Two Water Boards, Doncaster and Lee Valley, are both going further into the improvement of control and reduction of pumping costs, together with a reduction in the staffing requirement which is vital due to staff shortages. Doncaster is using a larger ARCH 2000 both to control the telemetry system and also to do the strategic calculations, whereas Lee Valley is using two ARCH 102 machines in tandem, one for strategy and one for tactics. A similar large system for the control of electricity distribution is being provided for the Midlands Electricity Board in the Birmingham area.

4. Effect on Equipment Design

The pattern of equipment which emerges from these applications is of four basic types. The first is a range of central processors which vary greatly in power. At the levels of the hierarchy where control is exercised these processors must be designed for the appropriate environment with on-line operation in mind, but at the higher levels the normal multi-program facilities of the present day commercial machine will suffice to link with the machines in more direct contact with the process as such machines will only be called on to work in an 'office' environment. The on-line computers will be common to as many processes as possible in order to minimize the effect of costly development and also to enable the maximum investment both during design and production in order to achieve the highest

reliability. As large-scale integration (l.s.i.) becomes an economic possibility the proportion of design costs will increase and accentuate this consideration. This will doubtless affect the second and third categories of equipment respectively for connecting the processor to the human operator and discrete instrumentation units (which are themselves the fourth group).

In considering such equipment modules the question of standard interfaces and the ability to produce system hardware configurations from standard production items becomes a necessity. This concept, however, presents the specification experts with substantial headaches.

In the category of discrete instrumentation units, there is less scope for l.s.i. in its widest sense, but even here the advent of more complex integrated circuits, particularly in the linear field, will demand a high degree of standardization and discipline. Again the specifiers face a formidable problem of maintaining the balance between standardization as a result of

the need for economy and reliability and flexibility as a result of the wide variety of applications, but the experience to date in the use of computers and applying discipline to system designers points the way.

5. References

1. Cole, G. B. and Clarke, S. L. H., 'The development of ARCH—a hybrid analogue-digital system of computers for industrial control', *J. Brit. Instn Radio Engrs*, **26**, No. 1, p. 45, July 1963.
2. Jones, J. T. and Williams, N. J., 'Computer control of steel-works production', *Proc. Instn Elect. Engrs*, **111**, No. 6, p. 1183, June 1964.
3. Massey, R. and Hersom, S. E., I.F.A.C. 1963 Basle.
4. Sidebottom, A. W. 'Computer control for paper-making plant', I.E.E. Conference on 'Automation for Productivity', May 1968. (I.E.E. Conference Proceedings No. 40).

Manuscript received by the Institution on 5th June 1968. (Paper No. 1246/1C6).

© The Institution of Electronic and Radio Engineers 1969.

Study Tour of Japan and Expo '70'

The Institution has received proposals for a combined study tour of the Japanese electronics industry and visit to EXPO 70 in Osaka in March/April next year. The suggestions are naturally tentative at this stage but, assuming a minimum number of 25 in the party, a 10–14 day stay in Japan is envisaged, primarily devoted to visits to leading Japanese companies and including two free days in which to visit the Exposition.

Travel would be by scheduled services of Japan Air Lines and the cost between £300–£350 per person; this would cover return economy class air fares from London to Tokyo/Osaka, first-class hotel accommodation, ground transportation and the services of an English-speaking guide throughout the stay (except during leisure time), gratuities, service charges and entrance fees. (Normal economy-class return fare to Tokyo is £564.) Regarding foreign exchange, a special

business allowance can be obtained up to £20 per day, comfortably in excess of the V-Form amount of about £80 required for the trip.

Further exploration of the viability of such a tour must depend on the evidence, as soon as possible, of a sufficiently large number of interested members. Members who are interested are therefore urged to advise the Institution informally, *before mid-September*—a postcard will be quite adequate for this purpose. This action will not commit either party, but it will enable further planning to proceed, and those who have inquired will then receive details of the proposed arrangements.

Please write to: Japan Study Tour, c/o I.E.R.E., 9 Bedford Square, London W.C.1. At this stage it will not be possible to give any more information than the above outline details.

A Novel Method for Analysing Singly-occurring Pulses with Nanosecond Resolution

By

P. M. CASHIN,
M.E., B.E. (Hons.)†

AND

D. A. H. JOHNSON,
M.Sc., M.N.Z.I.E., C.Eng., M.I.E.E.‡

Presented at the Second New Zealand Electronics Conference (NELCON II) organized by the New Zealand Section of the I.E.R.E. and the New Zealand Electronics Institute and held in Auckland in August 1968.

Summary: The system is designed to record pulse-shape information of singly occurring pulses with rise-times down to 1 ns. The basic theory of the 'interrogate pulse' sampling scheme is described and a system to implement this scheme is outlined. The proposed system employs short term analogue storage of the sampled input, followed by a fast multiplexed analogue to digital converter and digital storage. Some results from an experimental instrument using this system are given.

1. Introduction

This paper describes a technique we have called the 'interrogate pulse' technique, which permits an amplitude analysis of extremely short (nanosecond) duration 'single-shot' impulses.

In the interrogate pulse system the signal to be analysed is applied to one end of a tapped delay line and an interrogate pulse is applied to the other end. The interrogate pulse enables samples to be recorded at each tap point without the need for any active electronic gating. The samples are temporarily stored on capacitors before being converted to digital form and stored in a shift register.

The control and digital section make extensive use of integrated circuits and result in extremely versatile digital and analogue read-out facilities. The method, as demonstrated by the experimental equipment, has considerable economic advantages over the travelling-wave-tube oscilloscope and camera. Some of the recordings made with the experimental equipment are included in this paper.

2. The 'State of the Art'

The need for singly-occurring pulse recording, in the nanosecond range, arises in such fields as communication channel interference measurement (particularly on the new wideband pulse code modulated channels), nuclear physics research, and corona discharge investigations.

At present the only widely used method of recording single nanosecond-pulses is to photograph the trace on a wideband oscilloscope. For rise-times less than 2 or 3 ns a special travelling-wave oscilloscope and

ultra fast film are required. In contrast, the recording of repetitive waveforms, even with sub-nanosecond rise-times, is quite easy with a sampling type oscilloscope.

Current research is underway to add a fibre optics screen to a travelling-wave tube oscilloscope in order to increase the sensitivity of the system.¹ Other workers² are developing a superconducting delay line. By circulating the single nanosecond pulse around this delay line, and taking samples from a fixed point in the loop as the waveform circulates, they are essentially synthesizing a repetitive waveform from the singly-occurring input. Another approach uses 50 cascaded samplers to achieve a bandwidth of 500 MHz.³ An excellent review of the 'state of the art' up to 1967 has been given by N. S. Nahman.⁴

3. The Interrogate Pulse System

This system uses an 'interrogate pulse' of large negative magnitude, and narrow width, fired down a delay line in the opposite direction to the signal to be recorded. Samples of the input signal are taken by recording the peak negative height of the interrogate pulse as it passes sample points spaced along the delay line. Since the interrogate pulse sums on the delay line with the input signal the peak height at any point will be a measure of the height of the signal, at the point in time when the interrogate pulse passes the sample point. Figure 1 shows the general arrangement.

To record a signal $f(t)$ entering the delay line from the left-hand end at time $t = 0$ (refer to Fig. 2), the interrogate pulse must be fired into the right-hand end of the delay line at time $t = NT$, $N+1$ is the total number of sample points along the delay line and T is the cable delay between them. The sampling window is determined by the width of the interrogate pulse.

† Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand.

‡ Formerly at the University of Canterbury; now at the University of Melbourne.

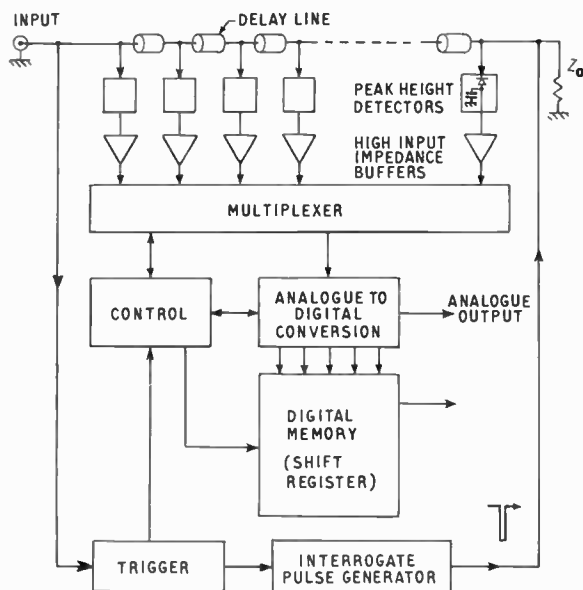


Fig. 3. Diagram of sampling system.

time. The voltage output in this case is correspondingly less.

The problem of the relatively short memory of the diode capacitor circuit is overcome in the instrument described by performing analogue-to-digital conversion, and digital storage, of the peak height information. Only one converter, multiplexed to each memory unit in turn, is used. A discharge time-constant of say 100 ms for each of the peak height recorders can be tolerated by this system since the voltage on each capacitor can be converted to digital form, and stored, in well under 100 μ s. The system design is shown in Fig. 3.

It is possible to generate interrogate pulses as short as a few hundred picoseconds duration, but the energy that can be tapped off at each tap point becomes extremely small. This in turn requires more elaborate, and expensive, peak height detector units. It should be pointed out however that both gain and d.c. offset correction can, if necessary, be incorporated in each sample channel between the storage capacitor and the multiplexer. In this way degradation of the interrogate pulse, caused by energy being extracted at previous tap points, can be compensated for.

5. Advantages of the System Design

Since the instrument has to have automatic control for digital conversion of the input samples, it is a simple matter to provide full automatic control. This allows the instrument to be set up to record impulses and then automatically to read out the samples in digital or analogue form, either once or

repetitively, before resetting back to be ready to record the next input signal.

By using a shift register for the digital store the samples can be read in as fast as the analogue-to-digital converter will allow. Once stored however, the shift register can be clocked around at a speed which suits the output device. For example, it could be advanced on command from a computer, or circulated around to provide a repetitive output waveform for an oscilloscope.

For analogue outputs a digital-to-analogue converter is required, but many forms of analogue-to-digital converter already embody this, so that a separate converter is not required. The provision of a direct digital output that can be used for punching paper tape, or for direct input to a computer, is becoming an extremely desirable feature for special-purpose recording instruments such as this.

6. Experimental Results

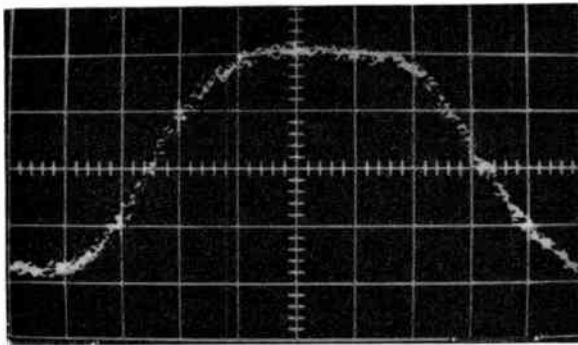
An experimental instrument was built⁶ to record 10 samples at 2 ns intervals. The samples were stored in a 4-bit 10-stage shift register (16 levels of quantization). The instrument was triggered from the input pulse, and after storing the samples in the shift register at maximum speed, the instrument automatically set the shift register circulating at a slower speed so as to give a repetitive reconstruction of the input waveform. This slow-speed reconstruction was displayed on a normal oscilloscope but it could just as easily have been recorded by a pen recorder or sent on as digital information. Figure 3 shows the general arrangement of the instrument.

The dead-time of the system before another pulse can be recorded is dependent on the output device used to accept the samples. If a direct link to a fast computer is used, a sample may be required every one or two microseconds. A very expensive multiplexer and analogue-to-digital converter would be needed to match this speed and keep the dead-time to a minimum. In the experimental equipment the minimum dead-time was about 1 ms.

Test waveforms were generated repetitively to enable them to be displayed on a sampling oscilloscope. The same generator was then fired single-shot and the waveform recorded by the experimental instrument. Figures 4 and 5 show comparisons of the sampling oscilloscope trace and the trace on a normal oscilloscope as reconstructed (without filtering) from a single-shot recording.

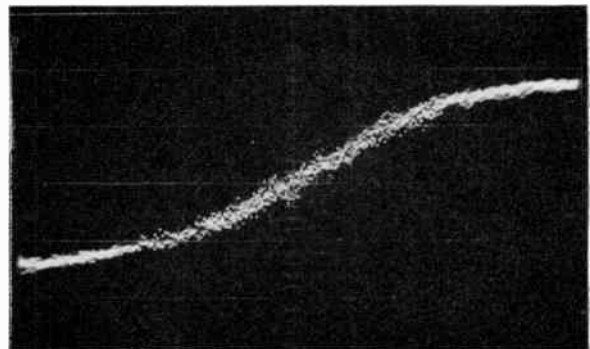
7. Conclusion

The interrogate pulse method of recording single nanosecond pulses has been described. This method does not require active electronic gating of the input



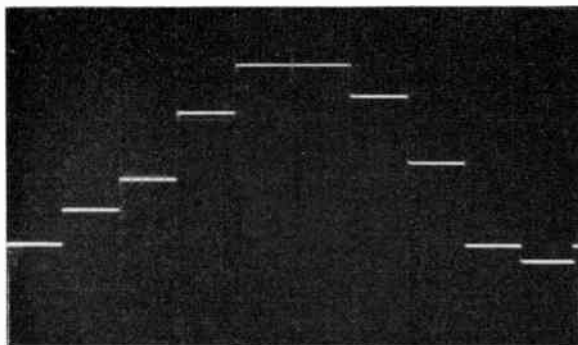
(a) Sampling oscilloscope

Repetition rate 300 kHz
 Time scale 2 nS/div.
 Amplitude 200 mV/div.



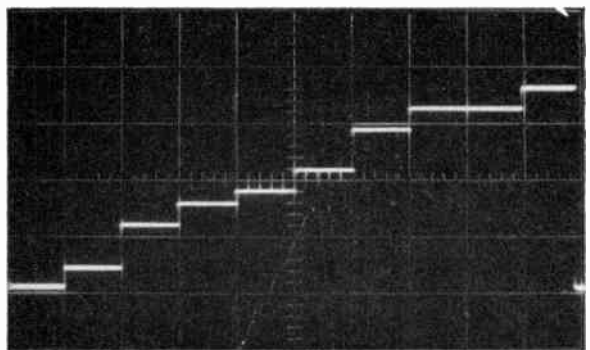
(a) Sampling oscilloscope

Repetition rate 300 kHz
 Time scale 2 nS/div.
 Amplitude 300 mV/div.



(b) 'Interrogate pulse' recorder. The same waveform as in (a) but recorded from a 'single shot'.

Fig. 4.



(b) 'Interrogate pulse' recorder. The same waveform as in (a) but recorded from a 'single shot'.

Fig. 5.

signal, and it supplies samples of the input to separate channels.

The method has been extended by the authors and variations have led to suggestions for gigahertz pulse-code-multiplexing and de-multiplexing,⁷ as well as for binary multiplication and fast pulse counting.⁶

The design of the instrument relies on the use of a high-speed multiplexed analogue-to-digital converter and digital storage. It is only the availability of low-cost integrated circuits (both digital and analogue) that have made such a design practicable and inexpensive.

8. References

1. Katzmann, F. L., 'Improving ultra fast transient recording using fibre-optics crt's', *Electronic Instrum. Digest*, pp. 41-7, October 1966.
2. Cummings, A. J. and Wilson, A. R., 'Cryogenic nanosecond pulse recirculator', *Proc. Inst. Elect. Electronics Engrs*, 52, No. 12, p. 1749, December 1964 (Letters).

3. 'The Mark 2—Single Transient System.' General Applied Science Labs. Inc., Westbury, N.Y., 1964.
4. Nahman, N. S., 'The measurement of baseband pulse rise times less than 10⁻⁹ seconds', *Proc. I.E.E.E.*, 55, No. 6, pp. 855-64, June 1967.
5. Bray, D. N. and Jensen, H. J., 'Simple submicrosecond pulse recording', *I.E.E.E. Int. Conv. Rec.*, 12, Pt. 8 (Instrumentation), pp. 111-28, March 1964.
6. Cashin, P. M., 'Investigation into the Recording of Singly Occurring Submicrosecond Pulses.' Thesis submitted for M.E. Degree at University of Canterbury, New Zealand, 1968.
7. Cashin, P. M. and Johnson, D. A. H., 'Time-division multiplexer/demultiplexer for digital transmission in the gigabit per second range', *Electronics Letters*, 3, No. 6, pp. 282-3, June 1967.

Manuscript first received by the Institution on 23rd October 1968 and in final form on 24th February 1969. (Short contribution No. 118/IC7.)

© The Institution of Electronic and Radio Engineers, 1969.

The Physical Basis of Current Noise

By

G. G. BLOODWORTH,
M.A., D.U.S., C.Eng., M.I.E.R.E.†

AND

R. J. HAWKINS,
B.Sc.†

Summary: The characteristics of typical $1/f$ noise spectra which have been observed, and the physical models proposed for their explanation, are reviewed. In m.o.s. transistors the $1/f$ spectrum extends over at least ten decades, and the effect of the gate field on the frequency index has been investigated. It is argued that these measurements support the theory that the conductivity of the inversion layer fluctuates due to electrons tunnelling to and from traps in the silicon oxide dielectric. It is shown that the tunnelling model satisfies Halford's statistical criterion.

1. Introduction

In virtually all semiconductor devices the dominant noise component at very low frequencies has a spectral density which is approximately inversely proportional to frequency. The phenomenon of ' $1/f$ ' noise (otherwise current, flicker, or excess noise) has several intriguing features. The difficulties are not practical but theoretical, being concerned with spectral limits not observed and physical models not generally accepted. There is no difficulty, for example, in calculating the noise output of a transistor in a given bandwidth if the general statistical nature of the noise is known.

The greatest problem is of course the choice of physical model, which has been a controversial topic for many years. Therefore it seems appropriate to write under the headings of relevant questions, to emphasize that we are giving our own answers and explanations. The main purpose of our contribution is to argue that recent work on m.o.s. transistors gives considerable weight to the 'tunnelling theory' for the origin of the noise, at least in devices fabricated of silicon passivated with silicon oxide.

Before discussing physical causes, a summary will be attempted of the basic characteristics of the noise, and of the mathematical and philosophical difficulties they produce.

2. What Noise Spectra Have Been Observed?

Current-noise measurements in semiconductor materials and devices have shown that the power spectral density $G(f)$ is generally of the form

$$G(f) = D \frac{I^\beta}{f^\alpha}$$

over a wide range of frequency f . Here I is the steady current flowing, the index $\alpha \simeq 1$, and (except for non-linear devices), $\beta \simeq 2$. The coefficient D is, in general, a weak function of temperature,¹ but this aspect of current noise has received insufficient attention to

provide even an empirical expression for the temperature dependence in most cases of practical importance. It is often difficult to separate the temperature dependence of the basic noise mechanism from that of other parameters such as carrier concentration and mobility.

Table 1 summarizes a small sample of measurements by several workers on a variety of materials.

Table 1
Basic features of some typical current noise measurements

Sample	Frequency range, Hz	Frequency index, α	Current index, β
carbon resistors ²	20-500 k	1-1.6	~ 2
carbon resistors ³	2.5×10^{-4} -7.5	~ 1	~ 2
n-type Ge crystal ⁴	10-10 k	1.25	
n-type Ge crystal ⁵	7-1 k	1.4	2.3
PbS film ⁶	20-20 k	1.1	2.0
CdS crystal ⁷	< 1 k	1.0	2.7
m.o.s. transistors ⁸	20-1 M	0.85-1	
m.o.s. transistors ⁹	20-100 k	~ 1	
m.o.s. transistors ¹⁰	5×10^{-8} -1	1-1.2	

It will be seen that the frequency index α is typically close to, but greater than, unity. This departure from the simple ' $1/f$ ' law imposes additional complexity on any satisfactory theory of current noise.

The value of approximately 2 for the current index β implies that the fluctuations, whatever their cause, are manifested as variations in conductivity, and are independent of the steady current flowing. This is in contrast to shot noise, where the current index is unity.

Fluctuations in conductivity could, in principle, be caused by changes in carrier concentration or mobility. It seems generally accepted, however, that fluctuations in carrier concentration caused by some form of

† Department of Electronics, University of Southampton.

trapping process provide the more likely explanation.^{5,11} The synthesis of a broad $1/f^\alpha$ spectrum requires, as we shall see, a wide distribution of electron lifetimes. The possible origins of such a distribution will be considered in Section 4.

3. Can the Spectrum Extend without Limit?

Every reduction in measurement frequency has been accompanied by a steady increase in spectral density, and no conclusive experimental evidence of a low-frequency 'turnover point' has been reported. However, it is generally accepted that such a point must exist to avoid infinite power density at zero frequency.

Assuming $\alpha = 1$ for simplicity and integrating the spectral power density between upper and lower frequency limits f_1 and f_2 , the total noise power is proportional to

$$\int_{f_2}^{f_1} \frac{df}{f} = \ln \left(\frac{f_1}{f_2} \right). \quad \dots\dots(1)$$

Now, any physical process of the sort we shall consider must have a minimum time-constant τ_1 and hence an effective upper frequency limit ($1/2\pi\tau_1$). Above this frequency the power spectrum decays at least as fast as $1/f^2$, and therefore the integrated noise power remains finite as $f_1 \rightarrow \infty$.¹² However, eqn. (1) suggests that the noise power becomes infinite as $f_2 \rightarrow 0$.

Fluctuations at frequencies close to zero can only be detected by measurements taken over an extremely long time. A finite measuring time T effectively introduces a high-pass filter into the system, which severely attenuates power at frequencies much less than $1/2\pi T$. Thus the predicted infinite power at zero frequency can only be investigated if we can afford to observe the noise over infinite time.

Equation (1) shows that the $1/f$ noise in a bandwidth ($f_1 - f_2$) depends only on the ratio of the limiting frequencies. Thus the noise power per frequency decade is constant. Typically, the current noise power from ten frequency decades represents approximately 1 part in 10^{10} of the steady power dissipated in a specimen. To double this noise power would require an extension of the measuring time by a factor of 10^{10} , and therefore an enormously long time would be required for the observed noise to become a significant fraction of the steady power supplied. Thus there is no thermodynamic problem about the increased noise power at low frequencies in real devices operated for, say, 50 years, since this time corresponds to about 10^{-10} Hz—an extension of less than six decades beyond the frequency range of noise already analysed.

The necessity or otherwise of a low-frequency spectral turn-over point *in the semiconductor device itself*, regardless of the low-frequency cut-off due to the finite measuring time, is therefore an academic question—but not a trivial one. It appears that all physical models so far proposed involve turn-over points, and it is only necessary for the predicted point to be lower than the limit of measurements so far made. This criterion will be applied to various theories in the next Section. If a likely theory predicts a turn-over point much lower than, say, 10^{-10} Hz, further low-frequency measurements will hardly be worth while; and we shall argue that this is indeed the case. If there was a physical theory which could be verified in 'real time', and yet implied a spectrum without a low frequency limit—then we should have an interesting academic question.

4. What Physical Models have been Proposed?

In almost all published theories of current noise, the required range of lifetimes arises from a trapping process with a wide distribution of characteristic time-constants. The choice of physical process differs widely among the suggested theories, as does the range of time-constants associated with each mechanism. The theories fall broadly into three classes, relying on electronic, ionic, and breakdown processes.

There is a great deal of evidence that the current noise in a semiconductor specimen is strongly dependent on surface properties. This evidence comes mainly from the effects of etching and ambient conditions on current noise,¹³ and from field-effect studies.¹⁴

The 'ionic' models of Macfarlane,¹⁵ Richardson,¹⁶ Bess¹⁷ and Jäntschi¹⁸ take for the long time-constant process the diffusion of atoms at an interface or surface. The theories of Macfarlane and Richardson do not give $1/f$ spectra over a wide enough frequency range and may therefore be discarded.¹⁹ The more sophisticated theories of Bess and Jäntschi predict spectra of $1/f$ form over $8\frac{1}{2}$ and 4 frequency decades respectively, and the former approaches the observed extent of the $1/f$ spectrum in m.o.s. transistors. This theory suggests that ions diffuse along dislocations to the semiconductor surface, where each ion captures an electron. Although there is some experimental evidence²⁰ to support this mechanism in semiconductor filaments, it does not appear to be directly applicable to the m.o.s. transistor where there is a thick ($\sim 1000 \text{ \AA} = 0.1 \text{ \mu m}$) surface oxide layer.

Petriz²¹ assumes that at a semiconductor surface or contact an irregular oxide layer will exist, in which there are localized high electric fields when current flows. A process of successive breakdown and healing is proposed, leading to a $1/f$ spectrum if there is a suitable distribution in oxide thickness. Watkins¹³ has

suggested a modification of this theory, in which high field regions exist where edge dislocations meet the semiconductor surface.

It may, perhaps, have been the apparent inability of purely electronic processes to give very long time-constants which prompted the ionic and breakdown theories described above. Nicollian and Melchior²² have published a theory in which a distribution of trapping time-constants arises from a random distribution in surface potential at a silicon-silicon oxide interface in the accumulation condition. The variation in surface potential over the interface results from a random distribution of fixed charges in the oxide, and can produce noise spectra with a frequency index of approximately unity over several frequency decades for the m.o.s. structure. This theory does not explain the spectra observed at very low frequencies in the m.o.s. transistor, which is, of course, operated with inversion, not accumulation, at the interface.

In another theory of current noise at the silicon-silicon oxide interface, Leventhal²³ ascribes the required range of time-constants to a continuous distribution of trap energies throughout the energy bandgap. His theory assumes that the surface traps interact to produce a conduction band, where carriers have a much lower mobility than in the adjacent inversion layer at the silicon surface. Where an electron is trapped, the conductivity of this band increases. In this model the long time-constant traps are those with energies close to the valence band, and a maximum time-constant of the order of 10^7 s is claimed. There is some doubt, however, whether time-constants of this order will be realized if the full equations of carrier transfer between the traps and the conduction and valence bands are considered.²⁴

A different mechanism which can give rise to very long time-constants, suggested by McWhorter²⁵ in 1956, is quantum-mechanical tunnelling. Modifying his theory for application to the m.o.s. transistor, we assume that electrons and holes from the conducting channel tunnel to and fro across the interface. Of these, a small fraction is captured by fixed traps in the oxide, losing energy by the emission of phonons and causing fluctuations in the channel conductivity. With increasing distance the transition probability decreases, and the more infrequent pulses produce conductivity fluctuations at lower frequencies. It can be shown²⁶ that a uniform spatial distribution of traps should generate a $1/f$ spectrum, and a departure of the frequency index from unity (e.g. to 1.2) can arise simply from a non-uniform distribution. The time constants increase exponentially with distance z from the interface, in proportion to $\exp(2Kz)$, where K is the decay constant of the electron wave function in the oxide. Taking the typical value of 1 \AA

(10^{-10} m) for $1/2K$, traps distributed from $z = 0$ to $z = 50 \text{ \AA}$ can produce a range of time-constants extending over about 18 decades, with a lower limit outside the measured range. If the active traps extend to a distance of the order of 1000 \AA , a typical value for the thickness of the gate insulation in m.o.s. transistors, the $1/f$ spectrum will extend to an extremely low *but non-zero* frequency. This model therefore satisfies the requirements outlined previously, and will be discussed further in the following sections.

Bell²⁷ suggested that current noise might be associated with the general statistical properties of semiconductors, and developed a 'queueing model'. He compared the problem to that of a telephone exchange, where subscribers (electrons) are served at random from a limited number of outlets (traps). The electrons have long waiting times in the conduction band because their number greatly exceeds the number of traps. Bell has shown that the consequent distribution of lifetimes of the conducting electrons gives a frequency index of roughly 1.5, which is somewhat greater than observed values. A feature of this model is that, in order to assign waiting times to individual electrons, one must apparently be able to identify them. While this is reasonable for the electrons trapped at particular localized sites, it does not seem valid for the mobile electrons. Once the identity of each electron is lost, the concept of separate waiting times has no physical justification, and it does not seem legitimate to treat these waiting times as long pulses contributing very low frequency components to the conductivity fluctuations.

Handel²⁸ has shown that current noise might arise from instabilities in the semiconductor current, by considering the equations for turbulence in an electron-hole plasma. Both his theory, and that of Bell, are unspecialized models which could be applied to a wide variety of semiconductor materials and devices, both in the bulk and at the surface. However, little experimental evidence has been published to support them.

5. Why is the M.O.S. Transistor an Important Device for Investigating Current Noise?

(1) There is strong experimental evidence that current noise in semiconductors originates mainly at the surface. In the m.o.s. transistor, the channel current flows at the silicon-silicon oxide interface, and the current noise exhibited by this device is comparatively large. Also, the properties of this interface have been extensively investigated and are relatively well understood.

(2) The structure of the m.o.s. transistor allows control of the surface charge by the gate voltage, independently of the drain voltage causing current flow. Therefore the conductivity fluctuations can be

studied as a function of the density of carriers in the channel of a single device. Earlier workers using two-terminal semiconductor samples could not change the charge present, for example by adding donor impurities, without affecting the crystal imperfections (dislocations, surface traps, etc.) which cause the noise.

More recently, several workers^{9, 29} have investigated the variation with bias conditions of the *magnitude* of the noise generated in m.o.s. transistors. For each element of the channel, the dependence of the fluctuations in carrier population on the mean value is primarily a function of the statistics of the trapping mechanism, rather than of the physical process assumed. Consequently, the variation of noise magnitude with bias conditions does not distinguish between the tunnelling model and some other theories of current noise. However, if the tunnelling model is taken, such measurements give useful information about the number of traps and their energy levels.

(3) We believe that an investigation of the effect of the gate voltage on the *frequency index* of the noise spectrum makes possible a direct check of the tunnelling model. This last aspect will be considered in more detail in the next Section.

6. What Conclusion is Suggested by Recent Measurements on M.O.S. Transistors?

Many investigations of current noise have examined the overall statistical nature of the problem rather than the particular physical process involved. Halford³⁰ has recently derived a mathematical criterion which must be satisfied by perturbations generating a $1/f^a$ spectrum over an arbitrarily wide range, and in the Appendix it is shown that the tunnelling model meets this criterion.

Very high fields are applied to the gate insulator in an m.o.s. transistor in normal operating conditions, and the tunnelling model theory²⁴ shows that these fields can alter the frequency index of the noise. The characteristic time-constants of the tunnelling process depend on the position of the traps in space and energy. The applied field modifies the trap energy levels, and alters the time-constants. The frequency index should vary with the applied field if the traps are non-uniformly distributed in energy.

Measurements of current noise³¹ at very low frequencies in n-channel m.o.s. transistors have indeed shown a change of frequency index with gate voltage, which is consistent with the theory if the trap energy levels in these devices were concentrated just below the conduction band. However, no change in index with gate voltage was observed in p-channel samples, indicating (according to the theory) that the trap distribution in energy was essentially uniform.

It is suggested that the observed change in frequency index with applied oxide field is strong evidence for the tunnelling model, since the field could only have a differential effect on active traps which were distributed in space, the basic assumption of the model.

Although current noise appears in a wide range of materials, it is tempting to believe that a common origin exists, a single ubiquitous physical mechanism. It is thought that the tunnelling model is applicable to any semiconductor device with an insulating surface layer. For example, current noise in planar bipolar transistors (fabricated like m.o.s. transistors of silicon passivated with silicon oxide) may arise from the tunnelling of majority carriers in the base to and from traps in the surface oxide. We have seen, however, that a number of theories can explain the wide distribution of time-constants required for the $1/f$ spectrum observed in various cases. Some theories apply specifically to the material involved and some are more general, and it may be that no one process can ever account for the observed behaviour of current noise in all materials and devices. However, for the important class of silicon planar devices we believe the tunnelling model is the strongest candidate.

7. Acknowledgments

We wish to acknowledge the contributions of our colleague Mr. I. R. M. Mansour, who first analysed noise in m.o.s. transistors below 10^{-1} Hz, worked out in detail the effect of the gate field in the tunnelling model, and took part in many stimulating discussions. We are indebted to the Science Research Council and the U.K.A.E.A. Establishment at Winfrith Heath for financial assistance.

8. References

1. Kornfel'd, M. I. and Mirlin, D. N., 'Temperature dependence of the low-frequency fluctuations of conductivity in germanium', Transl. from *Fizika Tverdogo Tela*, **1**, pp. 1866-8, 1959.
2. Campbell, R. H. and Chipman, R. A., 'Noise from current carrying resistors, 20 to 500 Kc', *Proc. Inst. Radio Engrs*, **37**, pp. 938-42, 1949.
3. Rollin, B. V. and Templeton, I. M., 'Noise in semiconductors at very low frequencies', *Proc. Phys. Soc. B*, **66**, pp. 259-61, 1953.
4. Brophy, J. J., 'Excess noise in n-type germanium', *Phys. Rev.*, **106**, pp. 675-8, 1957.
5. Brophy, J. J., 'Seebeck effect fluctuations in germanium', *Phys. Rev.*, **111**, pp. 1050-2, 1958.
6. Barber, D., 'Measurements of current noise in lead sulphide at audio frequencies', *Proc. Phys. Soc. B*, **68**, pp. 898-907, 1955.
7. Zijlstra, R. J. J. and van der Ziel, A., 'Noise in thin n-type CdS layers on an insulating CdS substrate', *Physica*, **29**, pp. 851-6, 1963.
8. Bozic, S. M., 'Noise in the metal-oxide-semiconductor transistor', *Electronic Engng*, **38**, pp. 40-1, 1966.

9. Flinn, I., Bew, G., and Berz, F., 'Low frequency noise in m.o.s. field effect transistors', *Solid State Electronics*, **10**, pp. 833-45, 1967.
10. Mansour, I. R. M., Hawkins, R. J. and Bloodworth, G. G., 'Measurement of current noise in m.o.s. transistors from 5×10^{-5} to 1 Hz', *The Radio and Electronic Engineer*, **35**, pp. 212-16, 1968.
11. Brophy, J. J. and Rostoker, N., 'Hall effect noise', *Phys. Rev.*, **100**, pp. 754-6, 1955.
12. Bell, D. A., 'Electrical Noise' (Van Nostrand, New York, 1960).
13. Watkins, T. B., '1/f noise in germanium devices', *Proc. Phys. Soc.*, **73**, pp. 59-68, 1959.
14. Hsu, S. T., Fitzgerald, D. J. and Grove, A. S., 'Surface-state related 1/f noise in p-n junctions and m.o.s. transistors', *Appl. Phys. Letters*, **12**, pp. 287-9, 1968.
15. Macfarlane, G. G., 'A theory of contact noise in semiconductors', *Proc. Phys. Soc. B*, **63**, pp. 807-14, 1951.
16. Richardson, J. M., 'The linear theory of fluctuations arising from diffusional mechanisms—an attempt at a theory of contact noise', *Bell Syst. Tech. J.*, **29**, pp. 117-41, 1950.
17. Bess, L., 'Study of 1/f noise in semiconductor filaments', *Phys. Rev.*, **103**, pp. 72-82, 1955.
18. Jäntschi, O., 'A theory of 1/f noise at semiconductor surfaces', *Solid State Electronics*, **11**, pp. 267-72, 1968.
19. van Vliet, K. M. and van der Ziel, A., 'On the noise generated by diffusion mechanisms', *Physica*, **24**, pp. 415-21, 1958.
20. Bess, L. and Kisner, L. S., 'Investigation of 1/f noise spectra', *J. Appl. Phys.*, **37**, pp. 3458-62, 1966.
21. Petritz, R. L., 'Semiconductor Surface Physics', p. 226. Ed. R. H. Kingston (University of Pennsylvania Press, Philadelphia, 1957).
22. Nicollian, E. H. and Melchior, H., 'A quantitative theory of 1/f type noise due to interface states in thermally oxidized silicon', *Bell Syst. Tech. J.*, **46**, pp. 2019-33, 1967.
23. Leventhal, E. A., 'Derivation of 1/f noise in silicon inversion layers from carrier motion in a surface band', *Solid State Electronics*, **11**, pp. 621-7, 1968.
24. Mansour, I. R. M., Hawkins, R. J. and Bloodworth, G. G., 'Physical model for the current noise spectrum of m.o.s. transistors', *British J. Appl. Phys. (J. Phys. D)* (G.B.), Ser. 2, **2**, No. 8, pp. 1063-82, August 1969.
25. McWhorter, A. L., 'Semiconductor Surface Physics', Ed. R. H. Kingston, (University of Pennsylvania Press, Philadelphia, 1957).
26. van der Ziel, A., 'Fluctuation Phenomena in Semiconductors' (Butterworths, London, 1959).
27. Bell, D. A., 'Semiconductor noise as a queuing phenomenon', *Proc. Phys. Soc.*, **82**, Pt. 1, pp. 117-20, 1963.
28. Handel, P. H., 'On the theory of 1/f noise', *Phys. Letters*, **18**, pp. 224-5, 1965.
29. Christensson, S., Lundström, I. and Svensson, C., 'Low frequency noise in m.o.s. transistors—I', *Solid State Electronics*, **11**, pp. 797-812, 1968.
30. Halford, D., 'A general mechanical model for $|f|^{\alpha}$ spectral density random noise with special reference to flicker noise $1/|f|$ ', *Proc. Inst. Elect. Electronics Engrs*, **56**, pp. 251-8, 1968.
31. Hawkins, R. J., Mansour, I. R. M. and Bloodworth, G. G., 'The spectrum of current noise in m.o.s. transistors at very low frequencies', *Brit. J. Appl. Phys. (J. Phys. D)* (G.B.), Ser. 2, **2**, No. 8, pp. 1059-62, August 1969.

9. Appendix: Mathematical Form for the Physical Model

Halford³⁰ has recently discussed the mathematical requirements of current noise models which could be physically realizable, with a spectrum of the form $1/f^{\alpha}$. We shall briefly review his method and conclusions, and show that the tunnelling model is consistent with them.

Halford considered a system perturbed by a succession of similar events, denoted by $H(a, s)$. Here a is the amplitude of H , and s is the lifetime, defined such that half of the total energy of a perturbation lies within the frequency range $0 \rightarrow 1/2\pi s$. The perturbations have the same shape, i.e. a change in amplitude or lifetime simply implies a scaling in magnitude or time. Consequently, the Fourier transforms of the perturbations have the same shape, again with the appropriate scaling in magnitude and frequency. He denoted the mean square amplitude of perturbations with the same lifetime by $A^2(s)$, and by $P(s)$ the probability that perturbations, occurring at a rate R , have lifetime s .

The energy of a perturbation is proportional to the square of its amplitude and to its lifetime, and thus the power $W(s)ds$ associated with perturbations in the range $s \rightarrow s+ds$ is given by

$$W(s)ds = sRP(s)A^2(s)ds \quad \dots\dots(2)$$

Using the definition of s in terms of the median frequency of the spectral energy distribution, half of the power $W(s)ds$ is below the frequency $1/2\pi s$. For example, if s is doubled each pulse contributes twice as much power into half the bandwidth. Thus the power spectral density produced by these perturbations is proportional to $s^2RP(s)A^2(s)ds$. Integration of the power spectral density over all lifetimes yields a 1/f spectrum if

$$P(s)A^2(s) = B/s^2$$

where B is a constant. This result is valid in the frequency range $\omega s_2 \gg 1 \gg \omega s_1$ (where s_2 and s_1 are the maximum and minimum lifetimes) if $P(s)A^2(s) < B/s^2$ for $s > s_2$ and $s < s_1$. Halford showed that $P(s)A^2(s) = B/s^2$ is a necessary and sufficient condition for a class of perturbations to give a 1/f spectrum, independent of the shape of the Fourier transform of each individual event.

From eqn. (2) the requirement $P(s)A^2(s) = B/s^2$ implies that $W(s) = RB/s$. In other words, the power contributed by perturbations of lifetime s must be inversely proportional to s . This may at first seem surprising, since the noise power spectral density increases as the frequency is reduced, and low frequency fluctuations correspond to long lifetimes. However, the noise power associated with long life-

times is concentrated at low frequencies, negligible power being produced at high frequencies, whereas processes with very short lifetimes produce power at both high and low frequencies.

We now consider the model²⁴ in which electrons and holes in the conducting channel of an m.o.s. transistor tunnel to and from traps in the adjacent oxide layer. Any fluctuation in the occupancy of traps at a distance z from the interface decays exponentially in time, with a time-constant $\tau(z)$. It is possible to treat the problem by identifying a perturbation with the combined behaviour of all electrons trapped at distance z . In this case, the perturbing events are exponential decays in time, and the analysis corresponds mathematically to an example given by Halford.

Alternatively, we can consider the effect on the channel conductivity of individual electrons, represented by rectangular pulses of constant height but with a distribution of lifetimes s for each time-constant τ . The exponential decay in time t of a fluctuation in the trap occupancy at z is governed by the factor $\exp(-t/\tau)$. The probability $P_z(s)$ that an individual electron trapped at z is released in the time interval $s \rightarrow s+ds$ is given by differentiating this factor with respect to t . Thus

$$P_z(s) = \frac{\exp(-s/\tau)}{\tau} \dots\dots(3)$$

While an electron is trapped the channel conductivity is decreased by μq for a time s , where μ and q are the electron mobility and charge.

In this analysis, R is the total number of electrons leaving the channel, per unit area and unit time. The magnitude of the electron wave function decreases exponentially with increasing distance z , having a decay constant K . Thus the rate of arrival of electrons from the channel at distance z is $R \exp(-2Kz)$, per unit area. The time-constant τ therefore increases exponentially with z , and is given by

$$\tau(z) = \tau_1 \exp(2Kz), \dots\dots(4)$$

where τ_1 is the minimum time-constant, which occurs at the interface ($z = 0$).

The probability $P_z(\tau)$ that an electron becomes trapped at z is proportional to the probability of the electron arriving at z , to the capture cross-section σ , to the density of traps N_t , and to the fraction of these which are empty. Thus

$$P_z(\tau) d\tau = 2K \exp(-2Kz) \sigma N_t (1-f_{i0}) dz,$$

where f_{i0} is the average fraction of traps occupied by electrons. Changing the variable from z to τ , using

eqn. (4), the probability becomes

$$P_z(\tau) d\tau = \frac{\sigma N_t (1-f_{i0}) \tau_1 d\tau}{\tau^2} \dots\dots(5)$$

If all the traps have a uniform spatial distribution and identical capture cross-sections, N_t and σ are constants. The fraction of filled traps depends only on the position of the trap energy levels relative to the conduction and valence bands of the semiconductor. If all traps have the same energy level, f_{i0} is also constant and $P_z(\tau)$ is then simply proportional to $1/\tau^2$.

The total probability $P(s)$ of an electron having a lifetime s in the oxide is given by integrating the product $P_z(s)P_z(\tau)$ over all values of τ , using eqns. (3) and (5). Therefore,

$$P(s) = \sigma N_t (1-f_{i0}) \tau_1 \int_{\tau_1}^{\tau_2} \frac{\exp(-s/\tau)}{\tau^3} d\tau$$

$$\equiv \frac{\sigma N_t (1-f_{i0}) \tau_1}{s^2} \int_{s/\tau_2}^{s/\tau_1} \eta \exp(-\eta) d\eta \dots\dots(6)$$

where $\eta \equiv s/\tau$. For $\tau_1 \ll s \ll \tau_2$,

$$\int_{s/\tau_2}^{s/\tau_1} \eta \exp(-\eta) d\eta \simeq \int_0^{\infty} \eta \exp(-\eta) d\eta = 1.$$

Therefore from eqn. (6) Halford's basic criterion $A^2(s)P(s) \propto s^{-2}$, is satisfied, since $A^2 (= \mu^2 q^2)$ is constant in this case. The integrand in eqn. (6) is always positive, so that when s approaches or exceeds τ_2 , or approaches or falls below τ_1 , $P(s) < \sigma N_t (1-f_{i0}) \tau_1 / s^2$, as also required by his theory.

As mentioned earlier, we cannot identify individual mobile electrons in the channel. Therefore it would be erroneous to visualize the rectangular perturbations as short pulses separated by much longer periods of time when the individual electrons are free. Instead, we must analyse the distribution of the lengths of time for which individual traps are empty and full. Corresponding to these two states there are two distinct time-constants, one determined by the rate of trapping and the other by the rate of excitation of trapped electrons. For simplicity, we have considered above only the distribution of the lifetimes of electrons in the traps, but a similar analysis applies for the times during which traps are empty and gives the same distribution of lifetimes s .

Manuscript first received by the Institution on 27th January 1969 and in final form on 25th February 1969. (Paper No. 1265/CC88).

© The Institution of Electronic and Radio Engineers, 1969

A Linear Temperature Control System using Thyristors

By

P. ATKINSON,

B.Sc.(Eng.), A.C.G.I., C.Eng.,

M.I.E.E., M.I.E.R.E.,†

AND

A. E. GEE†

Summary: The problem of the precise control of the temperature of liquid flowing through a tank containing a heater is one which is important and familiar in the process-control industry. When the heater is electrical the thyristor is a natural choice for the main controller. A simple and novel method of triggering the thyristors is described in which the power dissipated in the heater is a linear function of the control signal. The method is compared with the conventional non-linear and linearized arrangements and its electronic implementation is described. A 0.75 kW model has been designed to demonstrate the principle. With the same basic circuitry but with more powerful gating circuits the arrangement could be used to control very large powers.

List of Symbols

θ_i	set value of temperature
θ_o	measured value of temperature
θ	deviation
$\theta = \theta_o - \theta_i$	
$f(\theta)$	control signal
ω	angular frequency of mains power supply
t	time
φ	thyristor firing angle (conventional)
ψ	thyristor firing angle (novel)
I_m	maximum value of heater current
i	instantaneous value of heater current
i^2	mean square value of heater current
a, b, c	binary signals
\bar{a}	inverse of a
\bar{b}	inverse of b
a'	differentiated and rectified version of a

1. Introduction

The problem of the precise control of the temperature of liquid flowing through a tank containing a heater is one which is important and familiar in the process control industry. When the heater is electrical, the thyristor is a natural choice for the main controller. A scheme in which it is intrinsically possible to achieve a steady-state accuracy of about ± 0.01 degC is shown in Fig. 1.

The temperature θ_o of the emerging liquid is monitored by means of a thermistor connected in an a.c. bridge network excited by a sinusoidal e.m.f. generated by an oscillator of suitable frequency (e.g. 2 kHz). The set value of the temperature θ_i is adjusted by means of a variable resistor connected in the opposite arm of the bridge. The voltage appearing across the detector terminals of the bridge is then amplified and detected by a phase-sensitive detector; a d.c. signal proportional to the deviation is thus available. This signal is processed by a single, two- or three-term electronic controller to form the control

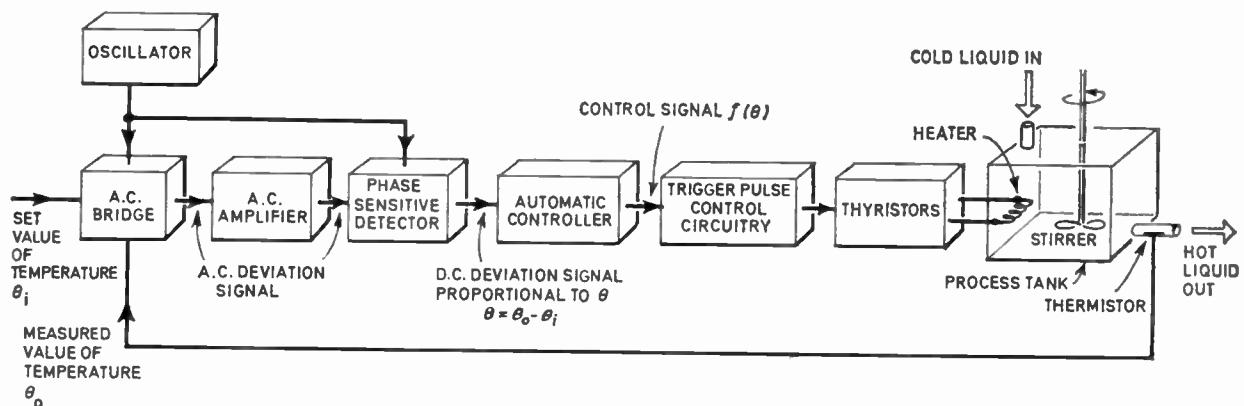


Fig. 1. Schematic diagram of liquid temperature control system.

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

signal $f(\theta)$ and then is applied to a trigger pulse control circuit. The trigger pulses are then routed into the gates of the thyristors.

The object of this paper is to describe a simple and, the authors believe, novel method of triggering the thyristors such that the power dissipated in the heater is a linear function of the control signal $f(\theta)$.

2. The Linear Control

Most workers in the field of process control will agree that if it is convenient and cheap then linear control is preferable to non-linear control, although so long as it is possible to keep the loop-gain high the undesirable effects introduced by non-linearities in a feedback control system can usually be virtually eliminated. Unfortunately, for reasons of stability of operation of closed-loop systems, particularly in process control where the time-constants tend to be large and distance/velocity lag is prevalent, it is not usually possible to keep the loop-gain very high. This will often lead to the emergence of the undesirable side-effects of the non-linearities. For instance, in a temperature control system, if the loop-gain decreases with the magnitude of the deviation (saturation effect) the system will respond relatively sluggishly for the larger deviations. Also, when such a system is without integral control, the static accuracy will vary with the set value.

In some systems it is necessary to have a wide range of set values and therefore linear control for large step-inputs. In others, whilst the set value may normally be constant, the system must cater for a large range of disturbances (e.g. changes in fluid flow rate, ambient temperature and power supply voltage); the magnitude of the control signals will vary considerably and again it is desirable to have linear control over a wide range.

3. Conventional Methods of Thyristor Triggering

A highly suitable arrangement of thyristors and diodes for controlling the mains power dissipated in a heater element is shown in Fig. 2. The most obvious method of triggering the thyristors is to arrange for the trigger pulses 1 and 2 to appear at firing angles ϕ and $\phi + \pi$ respectively as shown in Fig. 3.

The power dissipated in the heater element is proportional to the mean square current \bar{i}^2 which is given by

$$\begin{aligned} \bar{i}^2 &= \frac{I_m^2}{2\pi} \int_{\phi}^{\pi} \sin^2 \omega t \, d(\omega t) + \frac{I_m^2}{2\pi} \int_{\pi+\phi}^{2\pi} \sin^2 \omega t \, d(\omega t) \\ &= \frac{I_m^2}{2} \left\{ 1 - \frac{\phi}{2\pi} + \frac{\sin 2\phi}{2\pi} \right\} \end{aligned}$$

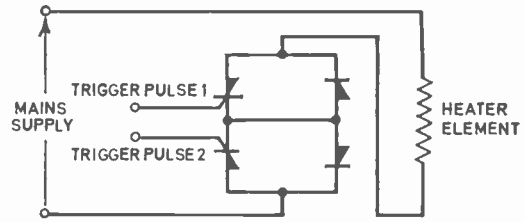


Fig. 2. Circuit for controlling electrical heater.

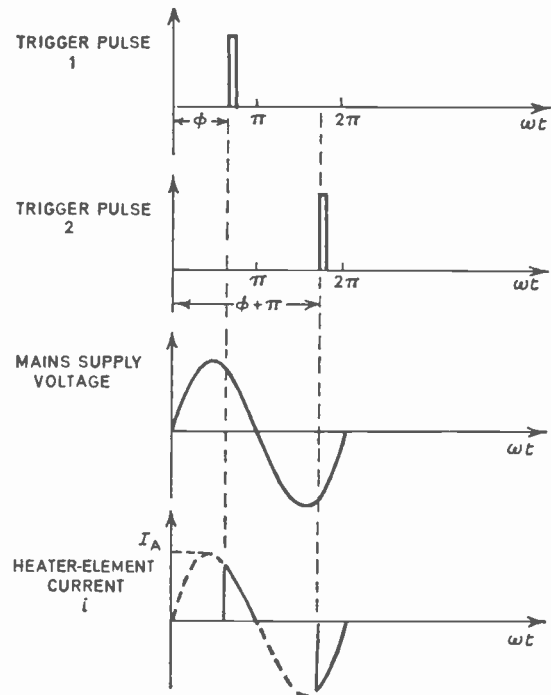


Fig. 3. Waveforms of thyristor trigger action.

A plot of \bar{i}^2 as a function of ϕ is given in Fig. 4(a) The output of the controller (Fig. 1) is the control signal $f(\theta)$, a linear function of the deviation θ . If the firing angle ϕ is made to vary linearly with $f(\theta)$ as shown in Fig. 4(b), then the overall control characteristic resulting by combining (a) and (b) is as shown in Fig. 4(c). The resultant control characteristic is thus non-linear, the gain being very low for small or large control signals. One method of overcoming this non-linearity is to introduce a second non-linearity to cancel exactly the original one as indicated in Fig. 4(d).

This technique (termed cosine-modified ramp control) has been described elsewhere.[†] It has the advantage that it gives linear control over the whole

[†] 'G.E. S.C.R. Manual', 4th ed. (General Electric Co. of America, New York, 1967).

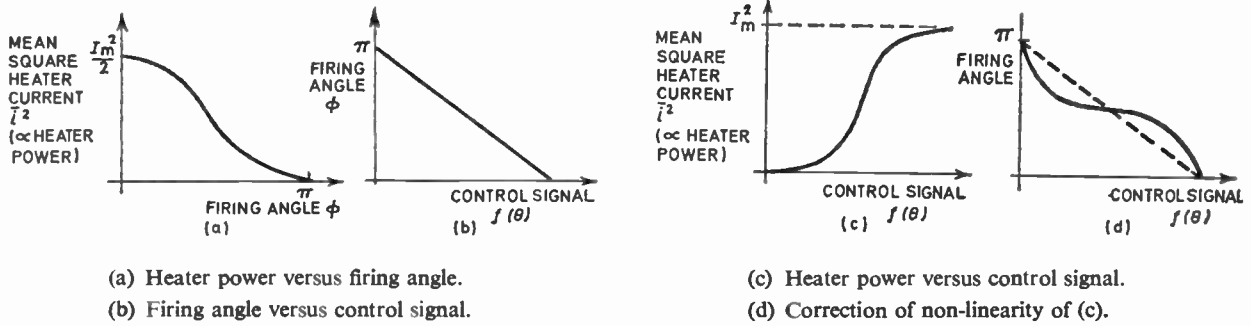


Fig. 4. Characteristics for conventional thyristor control.

of the range of power output. Its disadvantage is that the circuitry required to achieve this modification is fairly involved and must generate the required non-linearity with good accuracy to be of real use. It may also be argued that for a substantial range of control signals, Fig. 4(c) is fairly linear anyway. This is of course true and all the more valid in the present context since power is continuously being absorbed by the process when in normal operation so that under steady-state conditions the control signal $f(\theta)$ would be about half its maximum value for a given normal liquid temperature and rate of flow.

The authors feel, however, that regardless of the validity of these arguments, if it is possible to achieve completely linear control neatly and without substantially increasing the cost it is well worth doing so.

4. The New Method of Achieving Linear Control

The new method of control developed by the authors takes note of the fact that in a system designed to control the temperature of liquid flowing through

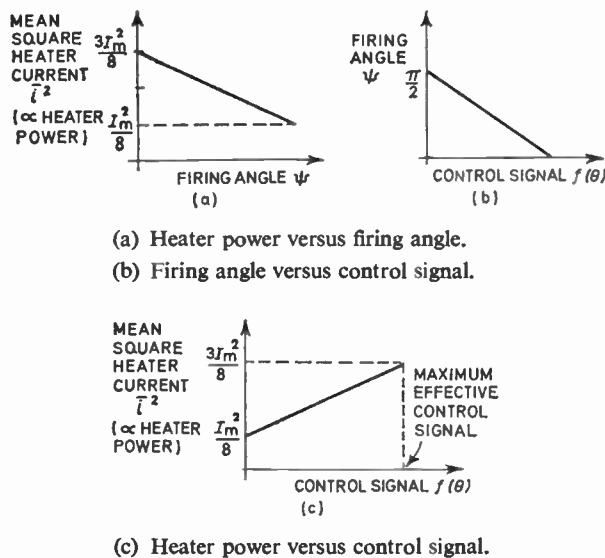


Fig. 6. Characteristics for new method of thyristor control.

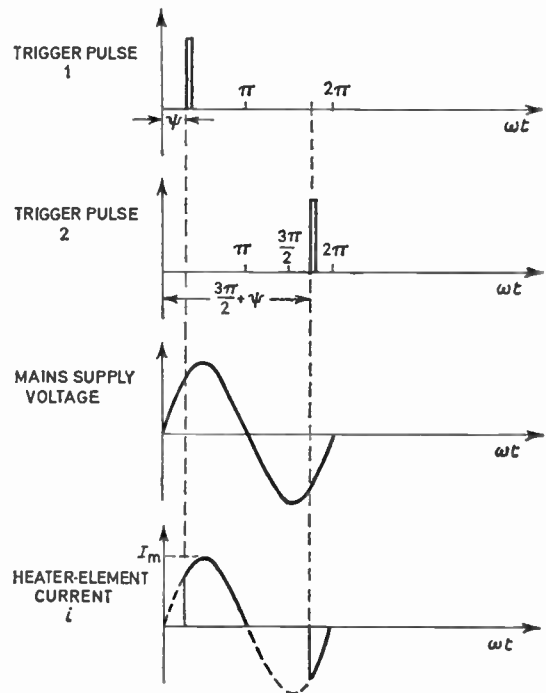


Fig. 5. Waveforms for thyristor trigger action with zero power not associated with zero control signal.

a process tank, it is implicit that it is quite unnecessary to associate zero power with zero control signal. With this in mind the authors defined a new firing angle ψ as shown in Fig. 5. The firing angle varies between 0 and $\pi/2$ for positive half-cycles and $3\pi/2$ and 2π for negative half-cycles. The power dissipated in the heater element is again proportional to the mean square current i^2 which is given by:

$$i^2 = \frac{I_m^2}{2\pi} \int_{\psi}^{\pi} \sin^2 \omega t \, d(\omega t) + \frac{I_m^2}{2\pi} \int_{\frac{3\pi}{2} + \psi}^{2\pi} \sin^2 \omega t \, d(\omega t)$$

$$= I_m^2 \left(\frac{3}{8} - \frac{\psi}{2\pi} \right)$$

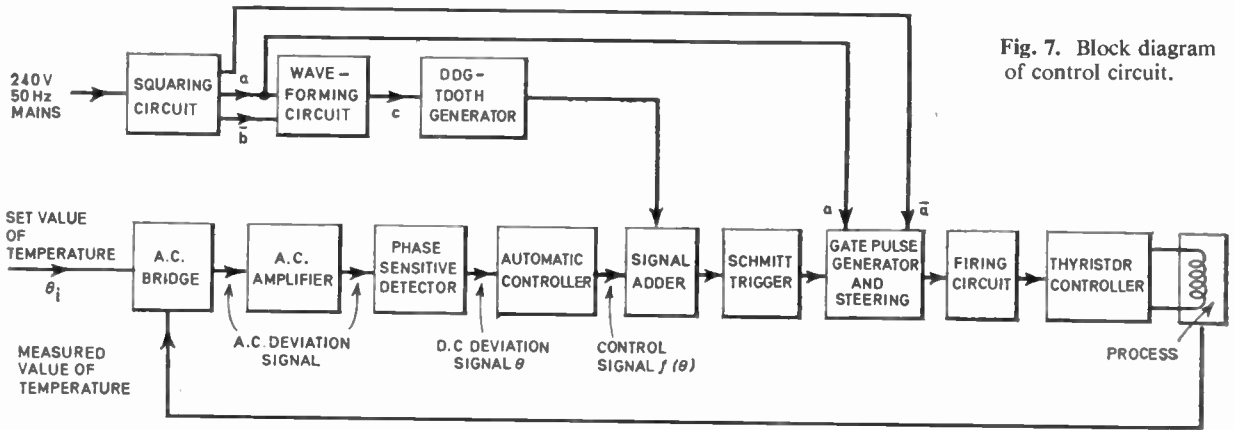


Fig. 7. Block diagram of control circuit.

A plot of i^2 as a function of ψ is given in Fig. 6(a). The new firing angle ψ is made to vary linearly with the control signal $f(\theta)$ as shown in Fig. 6(b). The overall control characteristic resulting from a combination of Fig. 6(a) and Fig. 6(b) is thus as shown in Fig. 6(c).

5. Electronic Implementation of the New Method of Control

The advantage of the new linear method of control lies chiefly in the fact that it can be implemented simply and economically by means of a few conventional electronic circuits. The basic block diagram of Fig. 1 is expanded to give the arrangement of the electronic circuitry suggested by the authors and shown in Fig. 7. The operation of the circuit may be understood by reference to the waveforms shown in Fig. 8.

The 50 Hz, 240 V mains supply is converted to a square-wave and inverted to form signals a and \bar{a} respectively. The phase of the mains is also retarded by 90° and subjected to the same squaring and inverting process to form b and \bar{b} . The signal a is differentiated and half-wave rectified to form the signal a' . A logical OR process is used to produce the signal c which is given in Boolean notation by the expression

$$c = a' + \bar{b}$$

The signal c is converted into the 'dog-tooth' waveform by a simple ramp generator with a 'reset to zero' facility, controlled by the input. The authors have successfully used the circuit of Fig. 9 in this application.

The dog-tooth waveform is added to the control signal $f(\theta)$ which under steady-state conditions is a d.c. level and the resultant is used to control the firing time of a Schmitt trigger. The Schmitt trigger circuit is biased so that for the minimum control signal it fires at the peak of each dog-tooth ramp.

For larger control signals the Schmitt circuit fires progressively earlier on the dog-tooth ramp until at maximum control signal it fires at the beginning of each ramp.

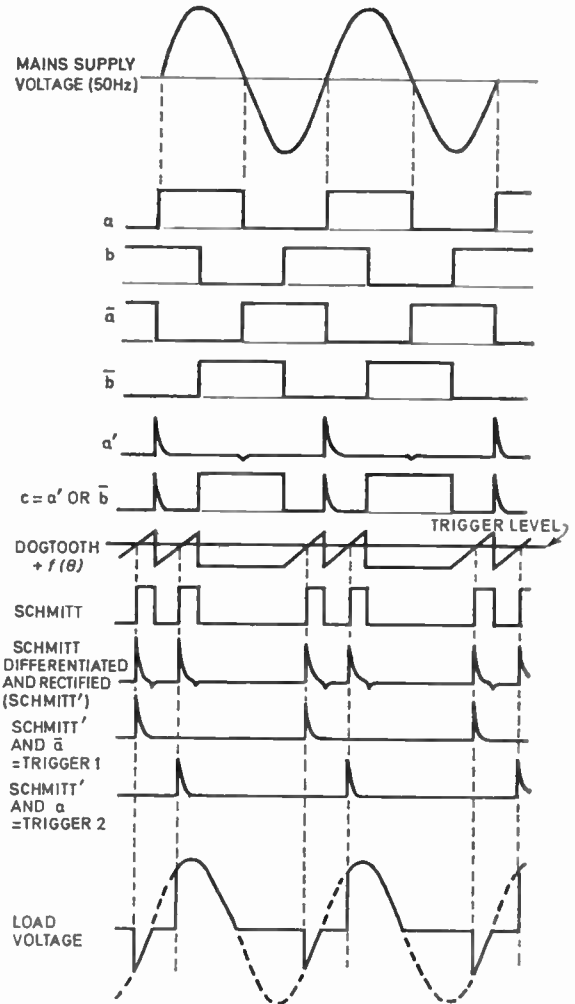


Fig. 8. Waveforms in control circuit.

The output signal from the Schmitt circuit is differentiated and half-wave rectified. Finally this signal is combined in logical AND gates with a and \bar{a} to form the trigger pulses 1 and 2 respectively. These pulses are steered on to the firing circuits of the thyristors.

The factors which limit the linearity of the system are principally:

- (i) The characteristic of the thermistor which is linear to about 2% over a temperature range of about 30 degC but can be greatly improved by the use of a 'thermi-linear' thermistor. Such a device is linear to within about 0.1%.
- (ii) The characteristic of the bridge circuit, the non-linearity of which can be minimized by the correct selection of components.†
- (iii) The linearity of the ramp waveform which can be made better than 1% with the circuit shown in Fig. 9.
- (iv) The characteristics of the phase-sensitive detector and a.c. amplifier which can be made extremely linear by correct design.

The accuracy of control depends on how high the loop-gain may be set and is governed ultimately by the loop-stability of the complete system. Accuracy is also dependent on the drift associated with the phase-sensitive detector and the automatic controller which can be minimized by correct design. The ultimate limit of accuracy is set by the signal/noise ratio of the thermistor.

6. Practical Verification of the Method

In order to verify the method practically, a model of the system has been constructed by using a 1 kW heater working from the 240 V, 50 Hz mains supply. The process tank is supplied with water from a constant head apparatus. The controller used in the model is a simple proportional controller and the range of control temperatures is from about 25°C to about 55°C.

The open- and closed-loop step response tests which have been performed indicate that the dynamic behaviour of the system can be predicted accurately using conventional linear theory. The tests also indicate that it is possible to achieve steady-state accuracy of temperature control to about ± 0.01 degC fairly easily.

Whilst in the model the maximum power dissipated is limited to 0.75 kW (i.e. three-quarters of the maximum possible with the 1 kW element operating

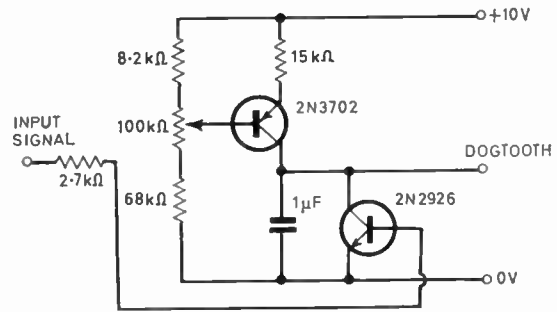


Fig. 9. Ramp generator.

on the mains supply), the same circuitry with more powerful gating circuits could be used to control hundreds of kilowatts.

Particularly important is the economical implementation of the method; component costs for the entire electronic control system of the model excluding the thyristors was about £15.

7. Conclusions

A new method of controlling thyristors for which the power controlled is to vary as a linear function of the control signal has been described. Its implementation using conventional electronic circuitry has been outlined. A temperature control system in which the new method has been used has been constructed and successfully tested. It has been demonstrated that the new method is simple and economical to implement and may be preferred to the conventional methods of non-linear and linearized non-linear control which have been used hitherto.

Although the method has so far only been used to control powers up to 0.75 kW, the same circuitry but with more powerful gating circuits could be used to control very large powers (hundreds of kilowatts).

8. Acknowledgments

The work of detailed designing, building and constructing the model temperature control system described in this paper was carried out by a team of second-year undergraduates in the Department of Applied Physical Sciences of the University of Reading in the Spring and Summer terms of 1968. The team consisted of Miss G. Midgely, R. Arnott, H. Breen and H. Hovland and was led by A. Gee (himself a second-year undergraduate); the team was supervised by P. Atkinson, lecturer in the Department.

Manuscript first received by the Institution on 25th November 1968 and in final form on 17th February 1969. (Paper No. 1266/1C8).

© The Institution of Electronic and Radio Engineers, 1969

† Bell, E. C. and Hulley, L. N., 'Precision temperature control', *Proc. Instn Elect. Engrs*, 113, No. 10, pp. 1671-7, October 1966 (I.E.E. Paper No. 5083C).

Simple Multi-level Logic Circuit Design

By

I. ALEKSANDER, Ph.D.†

Presented at the Conference on 'Electronic Switching and Logic Circuit Design', organized by the College of Technology, Letchworth, with the support of the I.E.R.E., and held at Letchworth on 24th October 1968.

Summary: A pencil-and-paper method of designing multi-level logic circuits is presented. The approach is intuitive rather than systematic. It bears the same relation to the design of circuits with gates of limited fan-in (as is the case with most microcircuit families) as does the Karnaugh-map method to two-level circuits and gates with unlimited fan-in. NAND/NOR systems are considered, and a graphical procedure, based on Karnaugh-maps, is followed.

1. Introduction

Multi-level logic circuit design has often been treated in the literature. However, one finds that this is a topic which becomes rather complex if tackled rigorously and thus the available approaches tend to be far too cumbersome to be of much practical use. Nevertheless, most practical logic circuits are of the multi-level type since the fan-in of gates is generally limited. This paper therefore, sets out to describe a method of multi-level circuit design largely based on Karnaugh maps. Central to the technique is the division of a Karnaugh map into sub-maps so as to produce the lowest number of gates in the final circuit. NAND/NOR circuits are assumed throughout. The treatment is exhaustive up to four variables and provides a basis for proceeding to more variables.

2. Method of Presentation

A somewhat unusual method of presentation has been chosen for this paper. Since the approach is largely graphical, most of the subject matter is expressed graphically and the text is restricted merely to brief notes.

The main objective is to establish a correspondence between Karnaugh maps (K-maps) and multi-level circuit structures in which the components are all of the NAND or the NOR type. It is assumed that a circuit cannot contain *both* NAND and NOR elements.

3. NAND/NOR Circuits from K-Maps

In Fig. 1(a) we have a typical K-map from which Fig. 1(b) is the AND/OR circuit obtained by 'looping' the logical 1's. If all the gates were replaced by NAND elements the same function would be obtained. An alternative is obtained by 'looping' the 0's as in Fig. 1(c). This may be turned into an all-NOR circuit simply by making all the gates NOR's. We call the above implementations *direct*.

† The Electronics Laboratory, University of Kent at Canterbury.

Indirect implementations may be obtained by using NOR gates for Fig. 1(b) (see Fig. 2(b)) and NAND gates for Fig. 1(c) (see Fig. 2(c)). All these may be derived using the transformations in Fig. 2(a).

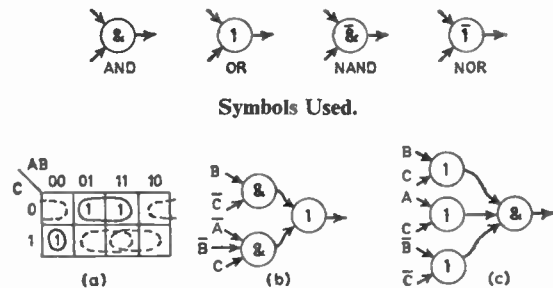


Fig. 1.

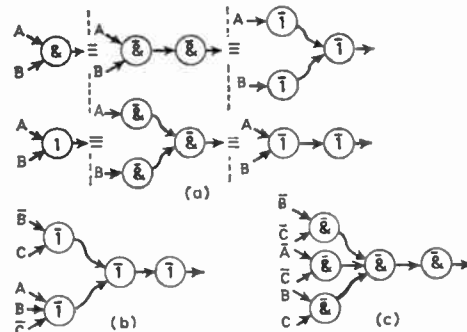


Fig. 2.

Indirect circuit structures often introduce unnecessary levels. Clearly, every K-map has a *direct* implementation either into NAND form or NOR form. It is important to note that it is the relationship between a pattern on the K-map and the corresponding circuit structure that one wishes to establish. If the pattern happens to be of 0's, the direct circuit will be a NOR structure whereas if the pattern is of 1's, the structure will be a NAND one. However, a pattern

will always correspond topologically to a structure. Hence in the rest of this paper we will deal with patterns of 1's and their corresponding NAND structures, recalling that identical results hold for NOR's and 0's. If the designer can choose whether he uses all-NOR circuits or all-NAND circuits, he must carry out the design twice and choose the cheaper solution.

4. Two-variable Circuits

Figures 3(a) and (b) represent functions independent of A and B. We take $f = 0$ for Fig. 3(a) and $f = 1$ for Fig. 3(b).

Figure 3(c) shows typical functions of one variable. At most, a single-input single-output NAND or NOR gate is required. If negated variables are available, no gate is required.

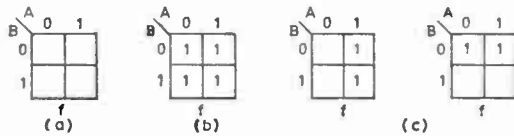


Fig. 3.

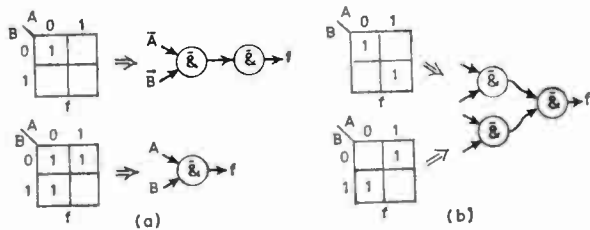


Fig. 4.

The function patterns in Fig. 4(a) are typical of two-variable functions that require one or two gates as shown.

The only two, two-level functions that require three gates are shown in Fig. 4(b). The same structure is required for either function.

4.1. Two-output Functions

It is important to recognize complementary functions as shown in Figs. 5(a) and (b). These are useful when comparing sub-maps (see later).

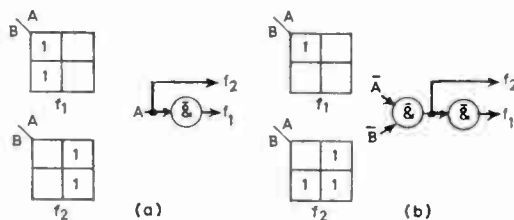


Fig. 5.

5. Three-variable Circuits

The normal three-variable two-level circuit cannot have more than three-inputs per gate at the input level and up to four inputs at the output gate. We assume that a multi-level circuit is required only if the fan-in is limited to two inputs/gate. (If the limitation is three inputs, only two out of the possible 256 logic functions cannot be implemented.)

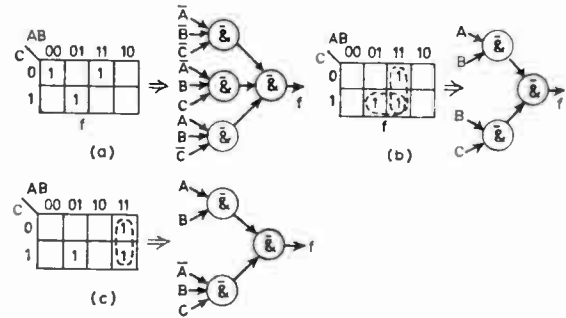


Fig. 6.

The examples in Fig. 6(a), (b) and (c) show three-variable functions and their two-level implementations. Clearly (a) and (c) have to be modified to a multi-level version. The rules are:

Expand to multi-level if,

- (a) the K-map has more than two 'loops' and/or
- (b) if there are any single-square 'loops'.

5.1. The Expansion Procedure

There are many ways of replacing a gate with many inputs by several gates with fewer inputs. One can, in many cases, strap the output of gates together or cascade AND or OR gates. The former is sometimes not recommended by the manufacturer of the gates and the latter is not economical when NAND and NOR elements are used. In this paper we concentrate on expanding functions of many variables into functions of fewer variables.

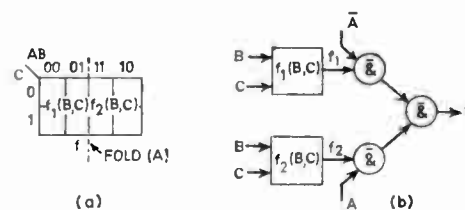


Fig. 7.

A three-variable function may be taken as two, two variable functions as shown in example of Fig. 7(a):

$$f = f_1 \bar{A} + f_2 A$$

f_1 and f_2 are functions of B and C only.

The circuit implication of this expression is shown in Fig. 7(b).

The examples of Figs. 6(a) and (c) are expanded in the above manner in Fig. 8(a) and (b) respectively.

We note the use of the function patterns of Fig. 3(c) and Fig. 4(a) and (b) in forming the sub-functions f_1 and f_2 .

We call this separation into two K-maps a *fold* since it represents a line along which the map may be folded. In particular in Fig. 7(a) we have an *A-fold* because it is the A variable that has been allocated to the second level. The concept of *folding* becomes clearer if we consider the corresponding squares of the K-map of the *A-fold* in Fig. 9(a). We note that

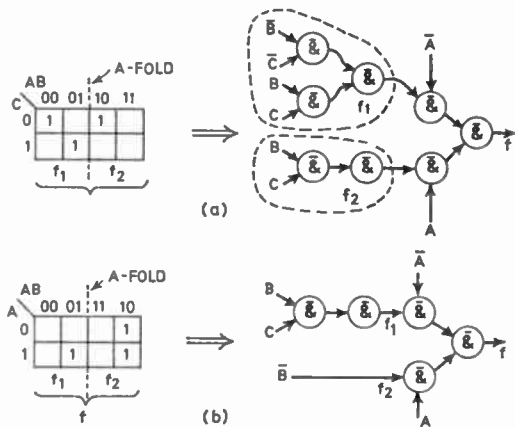


Fig. 8.

upon folding, the corresponding squares coincide. As a result, we note that in Fig. 9(b), $f_1(B, C)$ and $f_2(B, C)$ are inverses of one another yielding the circuit shown (using Fig. 4(b)). Note that this is a circuit that performs the binary sum of the three input variables. B and C folds are shown in Fig. 10(a) and (b) respectively.

Clearly design is a *process of choosing the right fold*.

Rule: Do not place a fold across a 'loop' if possible.

The reason for this is illustrated for the function in Fig. 11(a) where an *A-fold* yields the circuit in Fig. 11(b) where a *B-fold* would provide a two-gate circuit. This is introduced only by way of an illustration since the function does not really need expanding.

Example: find the best fold for the function shown in Fig. 12(a).

Both the *A* and *B* folds yield circuits with the structure shown in Fig. 12(b). The *C-fold* is more economical, giving the structure in Fig. 12(c). This is due to the fact that the *C-fold* does not cut through any 'loops' and has f_1 as the inverse of f_2 .

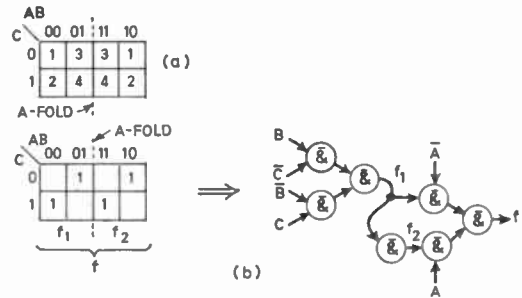


Fig. 9.

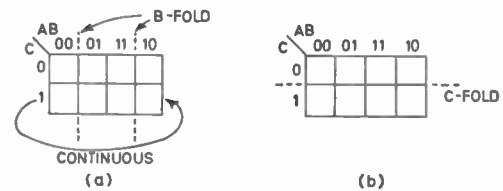


Fig. 10.

6. Four-variable Circuits

A two-level, four-variable circuit may require up to eight inputs at the second level. A single fold of the K-map will reduce the maximum fan-in to three, and should be used for fan-in limitations between three and eight. A double fold is required if the limitation is two inputs.

6.1. Single Folds

Four folds are possible, one for each variable. The rules are much the same as in the last case: the fold should not cut through a 'loop', and inverse functions should be found where possible. We note that f_1 and f_2 are now functions of *three* variables.

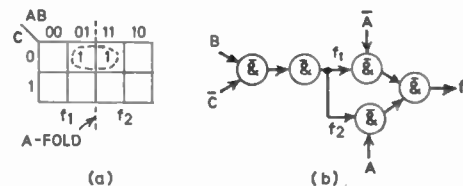


Fig. 11.

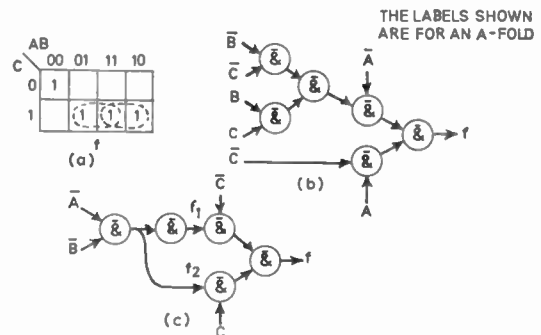


Fig. 12.

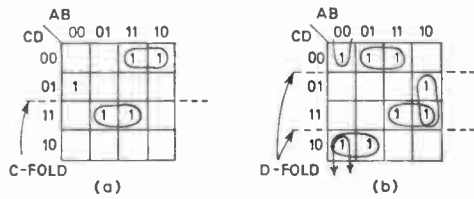


Fig. 13.

We note that in Fig. 13(a) the C-fold and the D-fold are preferable to an A or a B fold, due to the loop-splitting argument; whereas in Fig. 13(b), the D-fold is optimal on both accounts.

6.2. Double Folds

When two-input gates are used, it may be necessary to extend the circuit as shown in Fig. 14.

This represents a first fold for the variable fed in at X_4 and further folds for X_3 and X_5 . Note that X_3 and X_5 could accept the same variable, but they need not necessarily do so.

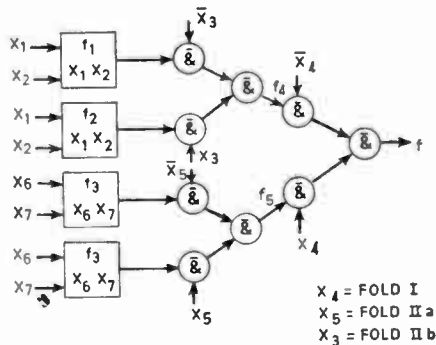


Fig. 14.

It is difficult to look at the folding exhaustively since there are four choices for the first fold and nine for the second, making a total of thirty-six. This number grows alarmingly with the number of variables. It is recommended therefore that the folds be made according to simple rules which may be illustrated by a few examples.

The function in Fig. 15(a) has first been folded with respect to A and second with respect to C. Figure 15(b) is the resulting circuit structure. The only rule

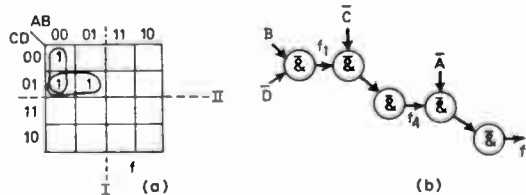


Fig. 15.

invoked here is that the fold should not be made across a 'loop'.

Three methods of folding Fig. 16(a) are now presented.

- (a) B followed by D.
- (b) C followed by B.
- (c) B followed by C.

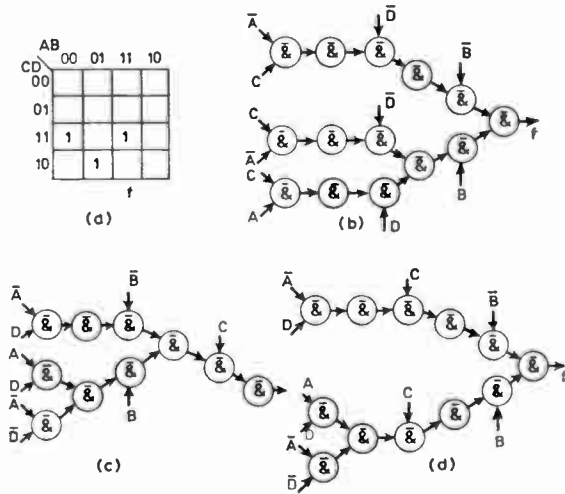


Fig. 16.

The resulting structures are shown in (b), (c) and (d) respectively; of which (c) is the most successful. Fold C, as the first fold, ensures that only one 'branch' grows backward from the output gate (c). That is, the output level is made independent of C. This does not happen if C is the second fold (d), or if D is chosen instead of C (b).

Hence the rule: The first fold should create an independence if possible.

As a final example, consider Fig. 17(a). The first and second folds are shown and the resulting circuit structure is in Fig. 17(b).

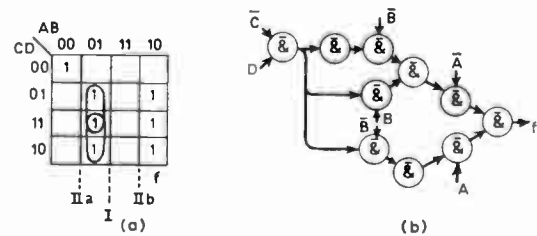


Fig. 17.

It is noted that these folds make use of the inverse and also of the similarity of the functions on the two sides of the first fold. This similarity search should be added to our list of intuitive rules.

7. Summary of the Method and Conclusions

The method is largely intuitive. It is intended to give the designer who has learned to translate a K-map into a two-level circuit, a 'feel' for doing the same if the gates he uses have a limited fan-in. Summarizing the rules we have:

- (1) Do not fold a map through a 'loop' if possible.
- (2) Fold a map so as to group the 1's (or 0's) together (i.e., do not split groups of 1's even if they do not form 'loops'). This creates independences.
- (3) Remember that the *order* of folding is important. If a fold can create independences as in (2), it should be made first. This reduces 'branching' near the output levels of the circuit and avoids duplication.
- (4) Always look for inverses or similarity between the functions separated by a fold.

In general, it should be pointed out that the work by Holman and Barnard¹ and Prather² probably covers the sentiments expressed in this paper. However, these papers are rigorous and aimed at mechanization in the former and the setting-up of a mathematical structure in the latter. Our work is aimed at the 'pencil-and-paper' designer. In this class there are other excellent papers, particularly one by Hellerman³ in which a complete catalogue of three-variable NAND/NOR circuits is presented, and one by Zissos and

Copperwhite⁴ which deals with minimal structures (irrespective of fan-in limitation). On the more abstract topic of multi-level design by decomposition techniques the classical papers by Ashenurst⁵ and Roth⁶ are recommended.

8. References

1. Holman, D. F. and Barnard, D., 'The Use of Roth's Decomposition Algorithm in Multi-level Design of Circuits', 1st British Computer Society Symposium on Logic Design, 1967. (Available from the B.C.S. Logic Design Study group document repository.)
2. Prather, R. E., 'Tree circuits', *Trans. Inst. Elect. Electronics Engrs. on Electronic Computers*, EC-14, pp. 841-51, December 1965.
3. Hellerman, L., 'A catalogue of three-variable OR-invert and AND-invert logical circuits', *Trans. I.E.E.E.*, EC-12, pp. 198-223, June 1963.
4. Zissos, D. and Copperwhite, G. W., 'The design of minimal NOR/NAND logical circuits', *Electronic Engineering*, 37, pp. 592-7, September 1965.
5. Ashenurst, R., 'The Decomposition of Switching Functions'. Harvard Symposium on the Theory of Switching (Annals 29), 1959.
6. Roth, J. P. and Karp, R. M., 'Minimization over Boolean graphs', *IBM J. Res. Developm.*, 6, No. 2, pp. 227-38, April 1962.

Manuscript first received by the Institution on 1st November 1968 and in final form on 28th January 1969. (Short Contribution No. 119/Comp. 119.)

© The Institution of Electronic and Radio Engineers, 1969.

On the Theory of Logarithmic Silicon Diodes

By

M. J. BUCKINGHAM,

B.Sc.†

AND

E. A. FAULKNER,

M.A., Ph.D., C.Eng., F.I.E.R.E.,

M.I.E.E.†

Summary: A logarithmic diode is defined as one showing a forward characteristic $I \propto \exp(eV/mkT)$ where the factor m is independent of voltage (though not necessarily of temperature) over at least four decades of current. Many commercial diodes show such a characteristic with m in the region of 1.5. It is shown that this behaviour cannot be explained by the theory of Sah, Noyce and Shockley and a modified theory is proposed, based on the assumption that the diode current is predominantly due to Shockley-Read-Hall traps non-uniformly distributed in the depletion layer. The theory is shown to be consistent with observed current-voltage-temperature characteristics of a number of specimens.

List of Symbols

n	density of electrons in conduction band
p	density of holes in valence band
n_B	density of electrons in bulk n-type material
p_B	density of holes in bulk p-type material
e	magnitude of electronic charge
k	Boltzmann's constant
T	absolute temperature
V	voltage applied at leads of diode
I	diode current
$m = (e/kT) \cdot dV/d(\ln I)$	
E_T	energy level of Shockley-Read-Hall traps
E_i	intrinsic Fermi level
$V_x = -E_i/e$	electrostatic potential
T_n	lifetime for electrons injected into p-type material
T_p	lifetime for holes injected into n-type material
$\epsilon_T = (E_T - E_i)/kT - (1/2) \cdot \ln(T_n/T_p)$	
U	steady-state electron or hole recombination rate
N_A	density of acceptor ions
N_D	density of donor ions
V_B	built-in voltage
E_0	electric field at $x = 0$ in depletion layer
V_{G0}	zero temperature band gap = 1.2 V for Si
β	temperature coefficient of band gap
ϵ_0	permittivity of free space
ϵ_r	dielectric constant

1. Introduction

The classical equation for the current-voltage characteristic of a p-n junction is the Shockley relationship¹

$$I = I_s(T) [\exp(eV/kT) - 1] \quad \dots\dots(1)$$

† J. J. Thomson Laboratory, University of Reading, Reading, RG6 2AF.

which, for $V \gg kT/e$, can be expressed in the form

$$I \approx I_s(T) [\exp(eV/kT)] \quad \dots\dots(2)$$

This equation was derived for low injection levels, on the assumption that the effect of recombination in the depletion layer is negligible.

Experimentally it is found that in the case of most commercial silicon diodes the current-voltage relationship does not conform to eqn. (2). We may describe the relationship in terms of a parameter m defined by the equation

$$m = (e/kT) \cdot dV/d(\ln I) \quad \dots\dots(3)$$

Obviously eqn. (2), valid for low injection levels, shows $m = 1$ and it has been shown by Hall² that, in the limit of high injection m should be expected to approach 2. However, some diodes give values of m nearer to 2 than 1 even at low injection levels. In a later paper, Sah, Noyce and Shockley³ showed that in silicon diodes the current may be dominated at low injection levels by the effect of recombination at Shockley-Read-Hall^{4, 5} traps uniformly distributed in the depletion layer. On this basis they showed that a semi-qualitative argument led to a value of $m = 2$, while a detailed calculation led to a rather complicated formula in which m varies with V , the exact form of the result depending on the assumptions made about the built-in voltage V_B and the quantity ϵ_T defined as

$$\epsilon_T = [(E_T - E_i)/kT - \frac{1}{2} \ln(T_n/T_p)]$$

where E_T is the energy level of the traps, E_i is the intrinsic Fermi level and T_n and T_p are the lifetimes of the minority carriers in the bulk n and p-type regions respectively. Various curves of m against V from the Sah, Noyce and Shockley theory are plotted in Fig. 1, from which it can be seen that with $V_B = 1$ V and $\epsilon_T = 0$, the theory predicts $m = 1.9 \pm 0.1$ over a voltage range of 100–450 mV, but that the theory does not predict other values of m substantially independent of applied voltage. From a physical point of view values of V_B and ϵ_T in the region of 1000 mV and zero are the most reasonable choice; both for this

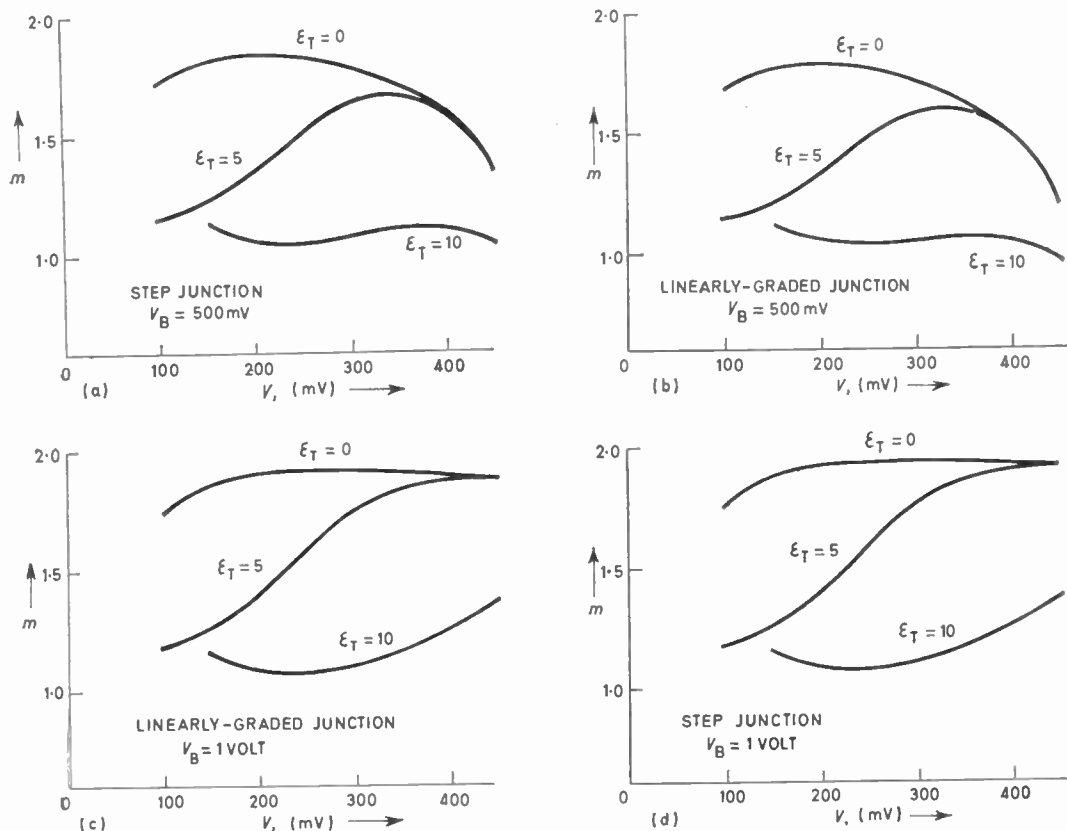


Fig. 1. Curves of m against applied voltage V calculated from Sah, Noyce and Shockley theory for various values of the parameters ϵ_T and V_B .

reason and on the basis of the semi-qualitative argument it is generally accepted that the Sah, Noyce and Shockley theory predicts an m value of about 2.

Although Sah, Noyce and Shockley produced experimental evidence that some junctions show an excellent agreement with their theory, the fact remains that many junctions do not. A more recent paper by Sah⁶ gave experimental results showing a variation of m between 1.10 and 1.30 with applied voltage over the same voltage range as above and suggested that this was due to a combination of the mechanisms discussed above.

2. Logarithmic Diodes

A logarithmic diode is one which shows a forward current-voltage characteristic which is accurately exponential over many decades of current. It is well known that many commercial diodes have this property and several firms have produced selected logarithmic devices for use in analogue computers and logarithmic voltmeters. For the purposes of this investigation we shall define a logarithmic diode as one showing a departure from logarithmic behaviour of not more than $\pm 10\%$ in current over four decades;

on the assumption that m varies steadily with applied voltage (i.e. that any curvature of the $\ln I$ vs. V characteristic is uniform) we can easily calculate that this degree of accuracy requires that m varies by not more than $\pm 8\%$ over four decades of current.

The feature of logarithmic diodes which is of especial interest here is that they often show an m -value in the region around 1.5 (Fig. 2). We have found only a few passing references to this point in the literature and on each occasion the author assumes that the mechanism is explained by Sah.⁶ This misconception appears to arise from the statement made by Sah⁶ (and also given in Ref. 3 in the abstract and the summary, though not in the main part of the paper) that the given theory predicts a diode characteristic which can be written in the form

$$I \propto \exp(eV/mkT) \text{ where } 1 < m < 4$$

This statement is very misleading because, taken in conjunction with eqn. (3), it implies that m is independent of V ; on the other hand, Sah⁶ proposes no mechanism giving m independent of V apart from those mentioned above, which are restricted to $m = 1$ (Shockley) and $m = 1.9$ (Sah, Noyce and Shockley). It is not possible to derive the logarithmic characteristic

on the assumption that the current is the sum of two exponential terms with these values of m ; this fact is illustrated in Fig. 3, from which it is clear that the variation in m is far too rapid in the region where its value is close to 1.5. We may generalize the problem of combining two exponential characteristics by looking at the function

$$y = (1/2)\{\exp[(1+\alpha)z] + \exp[(1-\alpha)z]\}$$

$$= \exp(z) \cdot \cosh(\alpha z)$$

$$d(\ln y)/dz = 1 + \alpha \tanh(\alpha z) \quad \dots\dots(4)$$

The departure of y from the perfectly logarithmic characteristic $y = \exp(z)$ is given by the multiplying factor $\cosh(\alpha z)$. For the required range of four decades z must vary between ± 4.6 ; and for the required extreme error of $\leq 10\%$ in y , we put $\cosh(\alpha z) \leq 1.22$, giving $(\alpha z) \leq 0.65$. We thus reach the conclusion that the value of α must not be greater than $0.65/4.6 = 0.14$. The corresponding extreme variation in $d(\ln y)/dz$ is $0.14 \tanh(0.65) = 8\%$.

3. Localized Traps

It is clear from the foregoing that in order to explain the action of logarithmic diodes we must look for some further mechanism which is inherently

capable of giving a logarithmic current-voltage characteristic, with $1 < m < 2$. Such a model should contain at least one structure-dependent parameter to account for the undeniable differences in characteristic between one diode and another. In a recent note⁷ we suggested a model in which the current is predominantly due to recombination at Shockley-Read-Hall traps, non-uniformly distributed in the depletion layer. A model of this kind was briefly mentioned by Shockley^{8,9} but its implications were not pursued. Here we shall develop this theory (which was originally proposed on the simple basis of an ideal step junction, giving a value of m between 1 and 2 independent of temperature and applied voltage) to the more realistic case of a diffused junction and extend it to account for the observed variation of m with temperature.

Figure 4 illustrates the variation in the excess impurity concentration with distance across the junction. The concentration gradient is not constant but increases as the surface through which the acceptor impurity was diffused, is approached. For our purposes it is sufficiently accurate to approximate this distribution by two straight lines each having a slope equal to the mean concentration gradient on the appropriate side of the plane ($x = 0$) at which the acceptor concentration N_A equals the donor concentration N_D and which we shall take as our plane of

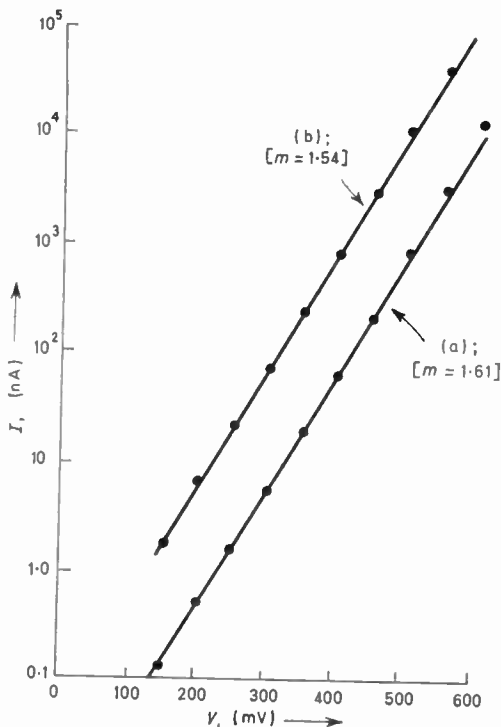


Fig. 2. Two examples of commercial diodes showing an exponential current-voltage dependence with m values in the region of 1.5. (a) Hewlett Packard step-recovery diode, (b) Texas 1S922 diode.

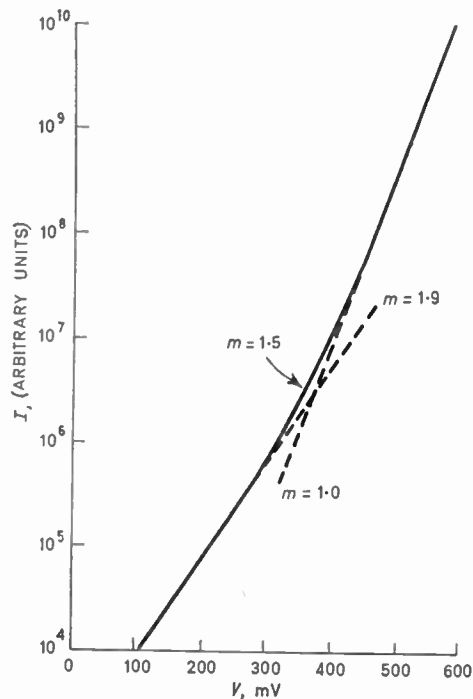


Fig. 3. The function $I = [\exp(eV/kT) + 1000 \exp(eV/1.9 kT)]$. Clearly, the variation of m in the region where its value is about 1.5 is far too rapid for the curve to be used as an approximation to an exponential.

zero potential. We then have

$$\left. \begin{aligned} N_A - N_D &= K_1 x & (x > 0) \\ N_A - N_D &= K_2 x & (x < 0) \end{aligned} \right\} \dots\dots(5)$$

where $K_1 > K_2$.

This approximation shows a discontinuity in the gradient at the origin but this does not introduce appreciable errors because the field distribution is very insensitive to the details of the charge distribution near the origin.

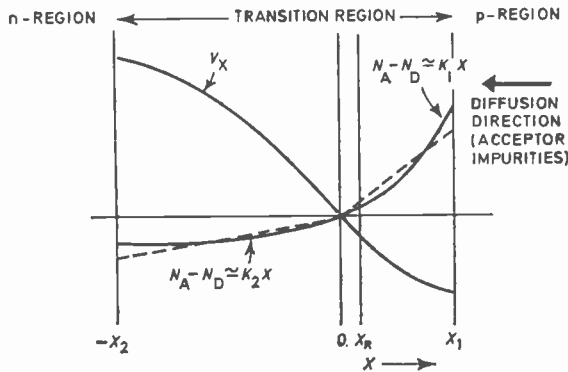


Fig. 4. Sketch of the depletion layer showing the approximation used to describe the net impurity distribution and the corresponding electrostatic potential variation across the junction. The region defined by $x > x_R$ contains a uniform distribution of Shockley-Read-Hall traps.

We now calculate the potential V_x as a function of x , on the usual assumption that the depletion region is sharply bounded by planes at x_1 and $-x_2$ at which dV_x/dx is zero. A straightforward application of Poisson's equation gives

$$\left. \begin{aligned} V_x &= eK_1 x(x^2 - 3x_1^2)/6\epsilon_r \epsilon_0 & (x > 0) \\ V_x &= eK_2 x(x^2 - 3x_2^2)/6\epsilon_r \epsilon_0 & (x < 0) \end{aligned} \right\} \dots\dots(6)$$

and the condition for total charge neutrality in the junction (equivalent in this case to the requirement that the potential gradient has no discontinuity at $x = 0$) gives the further equation

$$x_1^2/x_2^2 = K_2/K_1 \dots\dots(7)$$

Combining eqns. (6) and (7) we find the following expression for the ratio of the potentials V_P and V_N of the p-type and the n-type bulk material at the edges of the depletion layer

$$V_P/V_N = -(K_2/K_1)^{1/2} \dots\dots(8)$$

Now the carrier concentrations p_x and n_x at any plane x in the depletion layer are related to the potential V_x by the equation

$$\left. \begin{aligned} p_x &= p_B \cdot \exp[e(V_P - V_x)/kT] \\ n_x &= n_B \cdot \exp[-e(V_N - V_x)/kT] \end{aligned} \right\} \dots\dots(9)$$

where p_B and n_B are the majority carrier concentrations at the edges of the depletion layer. Assuming complete ionization of the impurities we may write

$$p_B/n_B = K_1 x_1/K_2 x_2 = (K_1/K_2)^{1/2} \dots\dots(10)$$

Bearing in mind that we are assuming that $K_1 > K_2$ we see from eqns. (8), (9) and (10) that $p > n$ for the whole of the region for which $x > 0$.

We now introduce the assumption of localized Shockley-Read-Hall traps. In the simplified treatment described in Ref. 7 which dealt with an idealized step junction, it was assumed that these traps were associated with impurity ions of only one type; this is not a physically reasonable hypothesis for a diffused junction where impurity ions of both types are present on both sides of the junction. Here we propose a model in which the region of the depletion layer defined by $x > x_R$ (see Fig. 4) contains a substantially uniform distribution of trapping centres; the physical mechanism by which such a distribution could occur will be discussed later.

We have already seen that throughout the region $x > 0$ the minority carriers are of the negative type. As shown in Ref. 7 we may write for the recombination rate at the traps

$$U = n/T_r \dots\dots(11)$$

and the total recombination current I is the integral of this expression over the region containing the traps. Thus

$$I = (e/T_r) \cdot \int_{x_R}^{\infty} n \cdot dx \dots\dots(12)$$

Now if x_R is a relatively small fraction of the total width of the depletion layer, the electric field in the vicinity of the $x = x_R$ plane has its maximum value E_0 and n falls off rapidly as x increases. For the purpose of our integration it is sufficiently accurate to rewrite eqn. (11) in the form

$$U = (n_0/T_r) \cdot \exp(-eE_0 x/kT) \dots\dots(13)$$

where n_0 is the electron concentration at the plane $x = 0$, and using this expression in conjunction with eqns. (12) and (9) we obtain

$$\begin{aligned} I &\propto n_R T/E_0 \\ &\propto (T/E_0) \cdot \exp[e(V_R - V_N)/kT] \end{aligned} \dots\dots(14)$$

where the constant of proportionality includes all factors which are not dependent on the applied voltage or temperature.

We must now express eqn. (14) in terms of the external voltage V applied to the leads of the diode. The applied voltage will only be varied over a limited range, say between V_1 and V_2 , about a mean value $\bar{V} = (1/2) \cdot (V_1 + V_2)$. The total potential drop ($V_N - V_P$) across the depletion layer will accordingly

vary between $(V_B - V_1)$ and $(V_B - V_2)$ where V_B is the built-in voltage, i.e. the equilibrium potential difference between the intrinsic Fermi levels on the two sides of the junction, which can be assumed to have a typical value of about 1 V. Introducing the new variables

$$V_0 = V_B - \bar{V} \quad \dots\dots(15)$$

and

$$V' = V - \bar{V} \quad \dots\dots(16)$$

we can write

$$V_N - V_P = V_0 - V' \quad \dots\dots(17)$$

(In a typical experimental situation V_0 may be 700 mV and V' may be varied between ± 200 mV.) We also introduce the distance x_0 which is the value of x_1 corresponding to $V' = 0$, that is to the value of the applied voltage V in the middle of the experimental range. From eqn. (6) in conjunction with eqns. (8) and (17) we see that

$$x_1/x_0 = (1 - V'/V_0)^{1/3} \quad \dots\dots(18)$$

and this expression will provide a convenient means of expressing the variation of the potential $(V_R - V_N)$, in eqn. (14), with the applied voltage. Again using eqns. (6) and (7) we have

$$\begin{aligned} (V_N - V_R) &= e(2K_2 x_2^3 - K_1 x_R^3 + 3K_1 x_1^2 x_R)/6\epsilon_r \epsilon_0 \\ &= [2\{1 + (K_1/K_2)^{1/2}\}]^{-1} \times \\ &\quad \times \{2(K_1/K_2)^{1/2} - (x_R/x_0)^3 + 3(x_R/x_0)\} V_0 - \\ &\quad - \{2(K_1/K_2)^{1/2} \cdot V' + (x_R/x_0) \times \\ &\quad \times (2V' + V'^2/3V_0 + \dots)\} \quad \dots\dots(19) \end{aligned}$$

where a binomial expansion has been performed on the voltage term derived from eqn. (18). The first term in this result is independent of V' so the variation of $(V_N - V_R)$ with V' is given entirely by the second term. This dependence is seen to be linear within 1% ($(K_1/K_2)^{1/2}$ being greater than unity) for a maximum value of V'/V_0 given by 200 mV/700 mV, provided that the ratio x_R/x_0 is less than 1/5. If this condition is satisfied we can write

$$V_N - V_R = V_c - (V'/\mu)$$

where

$$\mu = [1 + (K_1/K_2)^{1/2}]/[(K_1/K_2)^{1/2} + x_R/x_0] \quad \dots\dots(20)$$

and V_c is the term independent of V' in eqn. (19). Substituting this value of $V_N - V_R$ in eqn. (14) we finally obtain for the current-voltage relationship at a fixed temperature

$$I \propto (1/E_0) \cdot \exp(eV'/\mu kT) \quad \dots\dots(21)$$

From eqn. (6) we see that at $x = 0$, the value of dV/dx is $-eK_1 x_1^2/2\epsilon_0 \epsilon_r$. Accordingly we have for the value of E_0 as a function of V

$$E_0 = (1/2) \cdot \{3V_0(eK_1/\epsilon_r \epsilon_0)^{1/2}\} / \{1 + (K_1/K_2)^{1/2}\}^{2/3} \cdot (1 - V'/V_0)^{2/3} \quad \dots\dots(22)$$

and the value of m is given by the relation

$$\begin{aligned} 1/m &= kT/e \cdot d(\ln I)/dV' \\ &= 1/\mu + (2kT/3eV_0) \cdot (1 - V'/V_0)^{-1} \\ &= 1/\mu + (2kT/3eV_0) \cdot (1 + V'/V_0 + \dots) \quad \dots\dots(23) \end{aligned}$$

For relevant values of the parameters we find that the term in V'/V_0 has negligible (<1%) effect and we may write to a good approximation

$$1/m = 1/\mu + 2kT/3eV_0 \quad \dots\dots(24)$$

Now in eqn. (20) we know that $0 < K_2/K_1 < 1$ and $0 < x_R/x_0 < 0.2$, from which it follows that $1 < \mu < 2$. The second term in eqn. (24) will have little effect on the value of m , having a typical value of about 1/30 at 400°K. We may therefore conclude that our localized trap model does predict a logarithmic type of characteristic in which m may have values lying between 1 and 2.

4. Temperature Dependence of m

Equation (24) gives a value of m which is only slightly temperature dependent; although the second term is directly dependent on temperature in the numerator and also has a temperature dependence contained in V_0 in the denominator, it only contributes 1 or 2% to the total value, and the first term only shows a temperature dependence through the variation of x_0 in eqn. (20). Typically this will only be of the order of 0.01% per degK and, as we know $(K_1/K_2)^{1/2}$ always exceeds unity, while we must assume $x_R/x_0 < 1/5$ in order to obtain a logarithmic characteristic of sufficient accuracy, we see that the latter term constitutes only a minor part of the denominator of eqn. (20). It is true that in deriving eqn. (24) we have assumed a perfect depletion approximation, although in practice the edges of the depletion layer will be 'smeared out' over a voltage range of the order of kT/e ; however, a detailed calculation shows that this effect will make a negligibly small contribution to the temperature coefficient of m .

On the other hand, experimental measurements show that m may in practice have quite a large temperature coefficient, of the order of -0.1% per degK at $T = 300^\circ\text{K}$ (Fig. 5). Two notable facts are that the temperature coefficient is always negative and that it is less for diodes showing a value of m approaching 2 than in other cases.

We may account for this variation of m with temperature by recognizing that a practical junction is in fact inhomogeneous, with the impurity distribution varying from one part to another. The observed current-voltage relationship will be equivalent to that of an array of junctions connected in parallel, with m varying over a range of values. We may make an approximate investigation of a system of this kind by taking the simplest possible model in which the

current is carried by two junctions in parallel, with m -values of $m_0/(1+\alpha)$ and $m_0/(1-\alpha)$ respectively. We further simplify the model by putting $x_R = 0$ in eqn. (19) and ignoring the second term in eqn. (24), so that the same parameter appears in both the voltage and the temperature dependence. If we denote all the parameters related to each of the two junctions by the subscripts + and - respectively, we can then write

$$I_{\pm} \propto \exp[-e(V_0 - V')/m_{\pm} kT] \quad \dots\dots(25)$$

where

$$m_{\pm} = [1 + (K_{1\pm}/K_{2\pm})^{1/2}]/(K_{1\pm}/K_{2\pm})^{1/2} = m_0/(1 \pm \alpha)$$

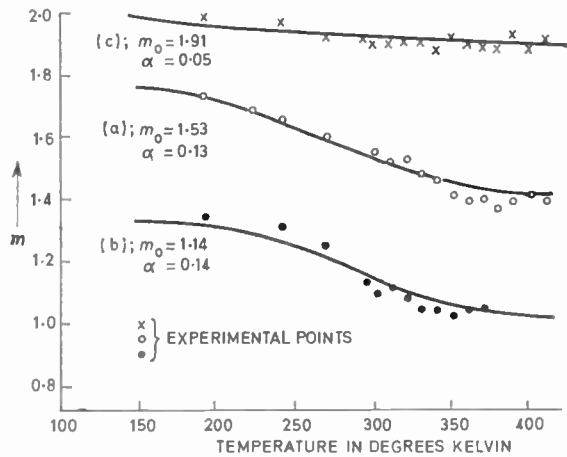


Fig. 5. Variation of $(e/kT) \cdot dV/d(\ln I)$ with temperature for three specimens: (a) Texas 1S922 diode, (b) Mullard OA202 diode, and (c) Hughes HS9008 diode. The full lines are the values of m_0 calculated from eqn. (30) using values of m_0 and α as shown.

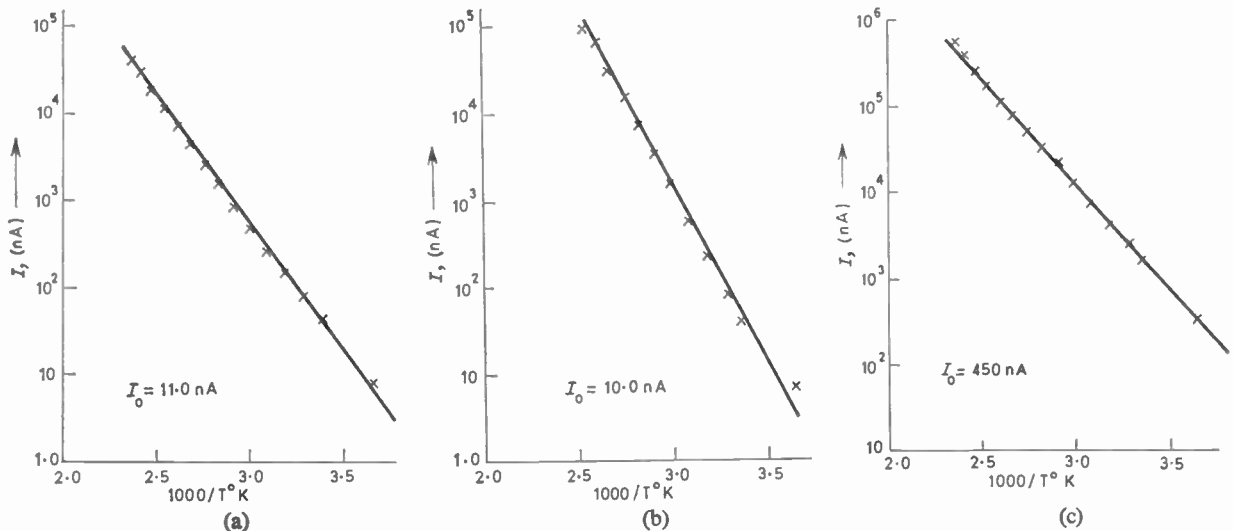


Fig. 6. The discrete points show the variation of $\ln I$ with $1/T$ degK for the same three specimens as in Fig. 5. The full lines are calculated from eqn. (29) using the same values of m_0 and α as in Fig. 5.

Now V_0 is defined by eqn. (15) as the difference between the built-in voltage V_B and the mean value \bar{V} of the applied voltage. V_B varies with temperature according to the relation

$$V_B = V_{G0} - [\beta + (k/e) \cdot \ln(N_C N_V / p_B n_B)] T \quad \dots\dots(26)$$

where V_{G0} is the zero temperature band gap, β is the temperature coefficient of the band gap (assuming the variation to be linear with temperature) and N_C and N_V are the effective densities of states in the conduction and valence bands respectively. From eqns. (25) and (26) we can therefore write

$$I_{\pm} \propto \exp[-e(V_{G0} - \bar{V} - V')/m_{\pm} kT] \quad \dots\dots(27)$$

where the constants of proportionality are independent of temperature and applied voltage. In order to give these constants an explicit and convenient form we introduce two new parameters I_0 and T_0 , which are defined so that I_0 is the current carried by either of the junctions at a temperature T_0 for $V' = 0$, i.e. for an applied voltage \bar{V} . We then have

$$I_{\pm} = I_0 \cdot \exp\{(1 \pm \alpha)z_0\} \quad \dots\dots(28)$$

where

$$z_0 = [-e(V_{G0} - \bar{V}) \cdot (1/T - 1/T_0)/m_0 k + eV'/m_0 kT]$$

We now set up the equation for our composite junction, the value of T_0 being defined as the temperature at which the contributions from the two current components are equal. We thus obtain the expression

$$I = I_0 \cdot [\exp\{(1 + \alpha)z_0\} + \exp\{(1 - \alpha)z_0\}] \quad \dots\dots(29)$$

As with eqn. (4), this result will give an exponential relation of the required accuracy at all temperatures provided that the value of α does not exceed 0.14. For the composite junction we define m_c as the

coefficient $(e/kT) \cdot dV'/d(\ln I)$ where I is given by eqn. (29) and hence we finally obtain

$$m_c = m_0/[1 + \alpha \tanh(\alpha z_0)] \quad \dots\dots(30)$$

Figure 5 shows eqn. (30) fitted to the observed variation of m with temperature, for several diodes. In these cases, T_0 is taken to be 300°K and \bar{V} is 300 mV giving $(V_{G0} - \bar{V}) = 900$ mV.

In Fig. 6 is shown the variation of $\ln I$ with $1/T$ in the same specimens, with a constant applied voltage of 300 mV. The solid line is calculated from eqn. (29) using the same values of m_0 and α as in Fig. 5.

It is clear that despite the crude nature of the model eqns. (28) and (30) are sufficiently accurate to describe the general behaviour of actual diodes within reasonable limits. We have looked at some more realistic models based on a continuous distribution of m -values in the junction, but the resulting expressions for the temperature variation of m_c are not experimentally distinguishable from eqn. (30) so an increase in the complexity of the model does not seem to be justified at this stage.

The fact that m_c cannot exceed 2 places an additional restriction on the value of α which can be used in eqn. (30) when m_0 exceeds 1.75 and this feature of the theory is consistent with the observation that the temperature coefficient of m is always comparatively low for values of m in this region.

5. Physical Basis of Trap Distribution

The theory we have presented here is based on the assumption that Shockley-Read-Hall traps are localized in a certain region of the junction, defined as the region on the right-hand side of the x_R plane in Fig. 4, x_R having some value between zero and $x_0/5$. We must now look for a physical mechanism which could account for such a distribution. It is important to bear in mind that the x_R plane is fixed in the material and does not move when a voltage is applied—this fact excludes any mechanism based on the relative position of the quasi-Fermi level and the intrinsic Fermi level.

A possible mechanism has been suggested by Dr. R. C. Newman of this Laboratory. He points out that there are some commonly-occurring impurities, notably iron, which in p-type material are ionized and therefore will readily form ion pairs with the acceptor impurity ions at diffusion temperatures.¹⁰

The same impurity in n-type material is not ionized and the formation of ion pairs with acceptor impurities in n-type material is therefore not a likely process. It follows that a diffused junction may be expected to contain ion pairs restricted to the region which was p-type at the end of the diffusion process; this provides us with a physical basis for our proposed trap distribution if we suppose that the ion pairs are the predominant trapping centres in the junction.

6. Acknowledgments

The authors are much indebted to Dr. P. C. Banbury of this Laboratory for many valuable discussions over a period of several months. This work was supported by a grant from the Atomic Energy Research Establishment, Harwell, England.

7. References

- Shockley, W., 'The theory of p-n junctions in semiconductors and p-n junction transistors', *Bell Syst. Tech. J.*, **28**, p. 435, July 1949.
- Hall, R. N., 'Power rectifiers and transistors', *Proc. Inst. Radio Engrs*, **40**, p. 1512, November 1952.
- Sah, C. T., Noyce, R. N. and Shockley, W., 'Carrier generation and recombination in p-n junctions and p-n junction characteristics', *Proc. I.R.E.*, **45**, p. 1228, September 1957.
- Shockley, W. and Read, W. T. Jr., 'Statistics of the recombinations of holes and electrons', *Phys. Rev.*, **87**, p. 835, 1st September 1952.
- Hall, R. N., 'Electron-hole recombination in germanium', *Phys. Rev.*, **87**, p. 387, 15th July 1952.
- Sah, C. T., 'Effect of surface recombination and channel on p-n junction and transistor characteristics', *Trans. I.R.E. on Electron Devices*, **ED-9**, p. 94, January 1962.
- Faulkner, E. A. and Buckingham, M. J., 'Modified theory of the current-voltage relation in silicon p-n junctions', *Electronics Letters*, **4**, No. 17, p. 359, 1968.
- Shockley, W. and Henley, R., 'Recombination currents in p-n junctions and distribution of deathnium', *Bull. Amer. Phys. Soc.*, **6**, p. 106, 1961 (Paper No. B-9).
- Shockley, W. and Queisser, H. J., 'Detailed balance limit of efficiency of p-n junction solar cells', *J. Appl. Phys.*, **32**, p. 510, March 1961.
- Collins, C. B. and Carlson, R. O., 'Properties of silicon doped with iron or copper', *Phys. Rev.*, **108**, p. 1409, 15th December 1957.

Manuscript received by the Institution on 12th March 1969. (Paper No. 1267/CC49.)

© The Institution of Electronic and Radio Engineers, 1969

Contributors to this Issue



S. L. H. Clarke (F. 1964, A.M. 1959) read mathematics at Trinity College, Cambridge, and after graduating joined Elliott Brothers. He subsequently became head of Computer Research Laboratory, Elliott Automation, being responsible for design and development of computer systems. He has been a director of A.E.I.-Elliott Process Automation Ltd. since G.E.C.-

English Electric amalgamation, and was formerly director of Elliott Process Automation Ltd. Mr. Clarke has been a member of the Institution's Instrumentation and Control Group Committee since its formation and he will shortly assume the Chairmanship. He has served on the Membership Committee and is author of several papers published in *The Radio and Electronic Engineer*. He is a Vice-President of the Institute of Measurement and Control.



G. G. Bloodworth (M. 1960) is senior lecturer in Electronics at the University of Southampton. After graduating in physics at Oxford he worked on aircraft control systems with Vickers-Armstrong Ltd. and transistor circuits with Mullard Ltd., before joining the University. Currently he is responsible for the organization of the 'Bosworth' M.Sc. course in Semiconductor Technology and Microelectronics which is run jointly by the University and A.S.M. Ltd. Mr. Bloodworth was Secretary of the I.E.R.E. Southern Section from 1962 to 1964 and Chairman from 1966 to 1968.

University and A.S.M. Ltd. Mr. Bloodworth was Secretary of the I.E.R.E. Southern Section from 1962 to 1964 and Chairman from 1966 to 1968.



R. J. Hawkins graduated with first class honours in physics at the University of Bristol in 1966. Since then he has been engaged in research on current noise in field-effect transistors in the Department of Electronics at the University of Southampton.



Colin Ayers is Chief Engineer of English Electric's Power and Marine Division, Kidsgrove. He joined the company in 1960 as a senior process consultant with English Electric-Leo Computers and was appointed to his present position in 1965, being responsible for the design of automation schemes, and marine power transmission and distribution schemes. Mr. Ayers

was educated at Bradford Technical College and the Central Technical College, Birmingham, gaining an external engineering degree of the University of London. He started his engineering career as a graduate engineer apprentice with General Electric in 1945 and subsequently held appointments with Merz & McLellan and with Kennedy & Donkin as an electrical power engineer. Mr. Ayers is Chairman of the B.E.A.M.A. Power Station Applications Committee and a member of the Institution of Mechanical Engineers' Automatic Control Group Committee.



Dr. E. A. Faulkner (F. 1966, A.M. 1964) graduated in physics from University College, Oxford. He worked at G.E.C. Research Laboratories, Wembley and subsequently at C.S.I.R.O. Division of Tribophysics, Melbourne, specializing in radio frequency measurements in solids. In 1960 he was appointed to the J. J. Thomson Laboratory, University of

Reading, in charge of teaching and research work in electronics. A strong advocate of university-industry collaboration, Dr. Faulkner works as consultant in linear circuit design for a number of leading electronics firms and is a director of Brookdeal Electronics Ltd. He has published around 50 papers in scientific and engineering journals and is the author of two books. He is a member of the Institution's Programme and Papers Committee and of the Thames Valley Section Committee.



M. J. Buckingham has been working, since he graduated in physics from the University of Reading in 1967, as a Ph.D. student in the J. J. Thomson Laboratory. He is primarily interested in the properties and applications of semiconductor devices. He recently became an Associate Member of the Institution of Electrical Engineers.

Higher-order Active Networks with One Active Element

By

Professor JIRI VLACH,
M.S., Ph.D.

Summary: Higher-order active networks using the Horowitz decomposition and only one active element (amplifier, negative impedance converter) are considered. Such networks are particularly sensitive to the changes of the active element and can eventually start oscillations. The stability of these networks is studied and it is shown that a special design can virtually eliminate the danger of oscillations for functions up to the sixth degree. A unique criterion is given for tests of the functions and for the stability margin.

1. Introduction

Active networks are attractive for miniature constructions because transfer functions with complex poles can be realized using only R-C elements and some combination of active elements. The price paid for the convenience of avoiding the inductances is a greater sensitivity to element changes, particularly to changes of the active elements.

It was discovered very early that the sensitivity increases very rapidly with the degree of the functions to be realized. As a logical remedy, more complicated transfer functions are usually decomposed into the products of two-pole transfer functions, these are realized separately and the networks connected in cascade. The question arises whether or not it is really necessary to break the function into these elementary two-pole blocks or whether it would be possible in practice to use more complicated functions.

It is shown in this paper that active filters using the Horowitz decomposition and only one active element can realize safely some pole-patterns with four or even six poles. A unique criterion is given for tests of any function and some techniques which may improve the behaviour are indicated. The computed results were confirmed in two cases experimentally.

2. Horowitz Decomposition and Oscillating Conditions

The Horowitz decomposition¹ was designed originally for networks to be realized by means of negative impedance converters (n.i.c.) but, as was shown recently,² this decomposition can also be used for active networks using simple low-gain amplifiers. These amplifiers can be realized much easier than n.i.c.s and thus the practical applicability of this type of active networks synthesis was increased.

The Horowitz decomposition is applied to functions having an even number of poles. The polynomial of

the denominator, with the zeros ($a_i, \pm jb_i$), is decomposed into

$$Q_n(p) = \prod_{i=1}^{\frac{n}{2}} [p^2 - 2a_i p + a_i^2 + b_i^2] \\ = \prod_{i=1}^{\frac{n}{2}} (p + \alpha_i)^2 - \frac{B}{k} p \prod_{i=1}^{\frac{n}{2}-1} (p + \beta_i)^2 \quad \dots\dots(1)$$

The ratio B/k is the coefficient of the highest degree of the second polynomial and k is the design value of the n.i.c.-constant or of the amplifier gain. Usually k is close to 1. The result is a difference of two polynomials, and it is clear that such a decomposition may be equal to zero for some value of k and for $p = j\omega$, and thus cause oscillations of the system. This is certainly the worst possible change of the active element and such a change can be used as a measure of the reliability of the network.

The question arises whether or not the Horowitz decomposition, designed for zero sensitivity (e.g. for least possible changes in the differential vicinity of the desired solution), is really the best one, if large changes of k take place. It could be expected that perhaps another difference of two polynomials would behave better for large changes of k , although the sensitivity for differential changes would not be the minimal one. It is shown in Appendix 1 that the Horowitz decomposition is the best one for both differential and large changes in the case of two poles. The situation cannot be too different for higher-order functions and, therefore, the decomposition is used as the basis for all further considerations. For $n = 2$ and for the oscillating conditions, eqn. (1) reduces to

$$Q_2(j\omega) = (j\omega + T_1)^2 - \frac{R_1^2}{k} j\omega = 0 \quad \dots\dots(2)$$

where

$$T_1 = \sqrt{a_1^2 + b_1^2} \\ R_1 = \sqrt{2(a_1 + T_1)} \quad \dots\dots(3)$$

as follows from the equations of Appendix 2. The

† Department of Electrical Engineering, University of Illinois, Urbana, Illinois 61801, U.S.A.

solution is

$$\omega_{osc} = T_1 \quad \dots\dots(4)$$

$$k_{osc} = \frac{R_1^2}{2T_1} \quad \dots\dots(5)$$

Introduce for the pole co-ordinates the relationship

$$b = H \cdot a \quad \dots\dots(6)$$

then

$$\frac{\omega_{osc}}{|a_1|} = \sqrt{1+H^2} \quad \dots\dots(7)$$

$$k_{osc} = 1 + \frac{a_1}{|a_1|\sqrt{1+H^2}} = 1 - \frac{1}{\sqrt{1+H^2}} \quad \dots\dots(8)$$

The frequency of oscillations is not very important, but it can be seen that it is nearly equal to b for large values of H . The relationship between k_{osc} and H is shown in Fig. 1. It illustrates the fundamental difficulty if an attempt is made to use this decomposition for band-pass filters. In these cases, H must be large and the system is more likely to start oscillations. The network does not oscillate as long as the value of k is above the curve, the design value being $k = 1$.

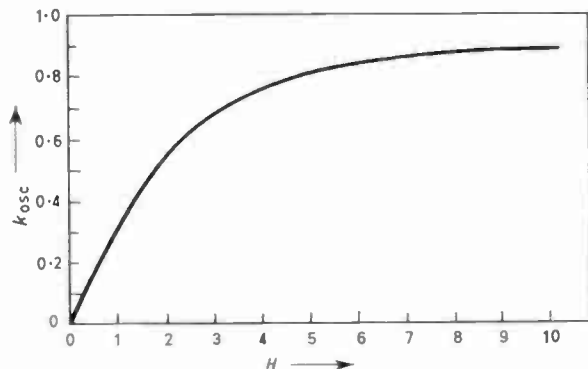


Fig. 1. Oscillating value for second-order filters. Values above the curve indicate stable behaviour.

Suppose now that a substitution $z_i = 1/p_i$ is introduced, transforming the original low-pass filter into a high-pass. Every pole of the original transfer function $p_i = a_i \pm jb_i$ changes into a pole

$$z_i = \frac{a_i \mp jb_i}{a_i^2 + b_i^2}$$

resulting in the same H for a high-pass. It follows from eqn. (8) that k_{osc} remains unchanged and only the oscillating frequency changes. It can be shown that this property (discovered and confirmed in all cases by the computer) remains unchanged for filters of any degree and that band-pass filters can be realized by cascade connection of a low-pass and a high-pass filter, both of them having the oscillating conditions given by the low-pass equivalents.

The value of k_{osc} is independent of the actual frequency band of the filter under test and gives comparative results without the necessity of normalizing the filters to some commensurable frequency bands. One number indicates the practical realizability of the given transfer function and eliminates, at least in the first stages of the design, the necessity to consider a particular circuit or synthesis technique. To show this, consider for instance, the techniques devised by Yanagisawa³ and Linvill.⁴ Linvill's synthesis uses cascade connection of circuits and any transmission zeros are determined entirely by the passive parts of the circuit. Yanagisawa's method connects two circuits in parallel and the n.i.c. constant appears in the denominator as well as in the numerator, thus making the transmission zeros dependent on k . However, both of these realizations have the same k_{osc} , which depends only on the denominator. For small changes, the Linvill circuit is better; for large changes of k , they behave similarly.

3. Polynomial Filters and their Practical Realizability

To get information on the behaviour of filters of fourth and higher degrees the program based on the equations of Appendix 2 was used and the results are collected in tables. The k_{osc} indicated in these tables is always that used in the denominator of eqn. (1) and the nominal value of k was chosen $k = 1$. Should the k be in the numerator, as would result, for instance, if a voltage-inversion type of n.i.c. were used, then the result is the inverse of the value indicated in the tables.

Table 1
Classical polynomial filters

Type of filter	Degree n	k_{osc}	% change of k to oscillate	ω_{osc}
Maximally flat	4	0.7612	23.88	1
	6	0.9405	5.95	1
	8	0.9864	1.36	1
Chebyshev 0.5 dB	4	0.9341	6.59	0.971
	6	0.9947	0.53	0.916
Chebyshev 1 dB	4	0.9489	5.11	0.9296
	6	0.9956	0.44	0.8858
Chebyshev 2 dB	4	0.9622	3.78	0.898
	6	0.996	0.4	0.846
Chebyshev 3 dB	4	0.969	3.1	0.886
	6	0.99669	0.33	0.8149
Thomson	4	0.53058	46.94	3.3215
	6	0.8168	18.32	4.942
	8	0.93817	6.183	6.579

Table 2
Equidistant poles

Poles	Degree	k_{oso}	% change of k	ω_{oso}
$-1 \pm j$ $-1 \pm j0.33333$	4	0.463	53.7	1.32
$-1 \pm j1.5$ $-1 \pm j0.5$	4	0.65	34.99	1.622
$-1 \pm j2$ $-1 \pm j0.666666$	4	0.756	24.4	1.963
$-1 \pm j1$ $-1 \pm j0.6$ $-1 \pm j0.2$	6	0.6607	33.9	1.28
$-1 \pm j1.5$ $-1 \pm j0.9$ $-1 \pm j0.3$	6	0.832	16.84	1.55
$-1 \pm j2$ $-1 \pm j1.2$ $-1 \pm j0.4$	6	0.904	9.5	1.854
$-1 \pm j1$ $-1 \pm j0.7143$ $-1 \pm j0.4286$ $-1 \pm j1.429$	8	0.8028	19.71	1.265
$-1 \pm j1.5$ $-1 \pm j1.071$ $-1 \pm j0.6429$ $-1 \pm j0.2143$	8	0.926	7.4	1.519

Data for classical polynomial filters are shown in Table 1. It can be seen that Chebyshev filters with a ripple of 0.5 dB or more cannot be realized in one active network. Fourth-order maximally flat filters can be practically used, but not higher degrees. The only filters useful, practically, for a degree as high as 6 are Thomson filters, designed for the maximally flat group-delay.

Table 1 indicates that the oscillating frequency is always near the outermost pole of the original transfer function and simultaneously shows the frequency where the peak of the characteristic, caused by changes of k , will form. This means that the denominator changes in such a way that it is always the outermost pole which moves towards the imaginary axis and indicates that any realization with the outermost pole relatively far from the imaginary axis should secure a greater margin of safety. Such filters cannot secure great selectivities, but they may be useful for other purposes; for instance, transmission of impulses.

This possibility was checked on transfer functions with all poles arranged on a straight line parallel to the imaginary axis and equidistant to each other. It

can be seen from Table 2 that fourth-order transfer functions of this type may be realized with a considerable margin of safety, as long as H of the outermost pole is $H \leq 2$. A sixth-order transfer function of this type is useful as long as $H \leq 1.5$. An eighth-order transfer function can be used only for $H \leq 1$, but this is already a rather impractical possibility.

Equidistant poles are not particularly suitable for good selectivity conditions, but the idea of keeping poles on the straight line parallel to the imaginary axis is attractive because such filters have a good safety margin, if realized as active filters. An improvement in selectivity can be achieved if the inner poles are shifted to the outer ones. One of many such possibilities was tested and the results are given in Table 3. These filters have been described elsewhere⁵ and are actually low-frequency equivalents to band-pass filters realized by equal tuned circuits, all of them coupled either inductively or capacitively by the same coupling coefficients. As can be seen, this slightly reduces the safety margin, as compared with equidistant poles, but selectivity is improved.

Another possibility lies in the use of other than Chebyshev polynomials for the approximation. The whole class of classical orthogonal polynomials may be used. If the amplitude curves of filters are normalized to the frequency $\omega = 1$ for 1 dB attenuation relative to their maximal transmission, then the poles and the safety margins are as given in Table 4.⁶ Fourth-order filters may still be used if the active element (amplifier, n.i.c.) is carefully designed.

Table 3

Poles on straight line parallel to $j\omega$ -axis improved selectivity

Poles are

$$\begin{aligned} & -1 \pm j1\alpha \\ & -1 \pm j0.5\alpha \end{aligned} \text{ for } N = 4$$

and

$$\begin{aligned} & -1 \pm j1\alpha \\ & -1 \pm j0.809\alpha \\ & -1 \pm j0.309\alpha \end{aligned} \text{ for } N = 6$$

α	Degree	k_{oso}	% change of k	ω_{oso}
1	4	0.538	46.13	1.312
	6	0.754	24.6	1.31
1.5	4	0.726	27.4	1.61
	6	0.895	10.52	1.605
2	4	0.822	17.79	1.94
	6	0.947	5.3	1.94
2.5	4	0.874	12.63	2.3

Table 4
Orthogonal polynomial filters

1 dB attenuation for $\omega = 1$

Type	Degree	k_{osb}	% change of k	ω_{osb}
Ultraspherical	4	0.865	13.44	1.03
	6	0.985	1.529	0.976
Chebyshev Class-2	4	0.884	11.62	1.01
	6	9.988	1.18	0.958
Legendre	4	0.909	9.02	0.97
	6	0.992	0.802	0.933

4. Rational Function Filters

The oscillating conditions are entirely given by the decomposition of the denominator. The transmission zeros do not influence this condition, but they may considerably improve the characteristics of the filter. Some of the possible ways of improvement and some results will be mentioned.

Suppose the denominator with a given k_{osb} was chosen as satisfactory. The amplitude characteristic

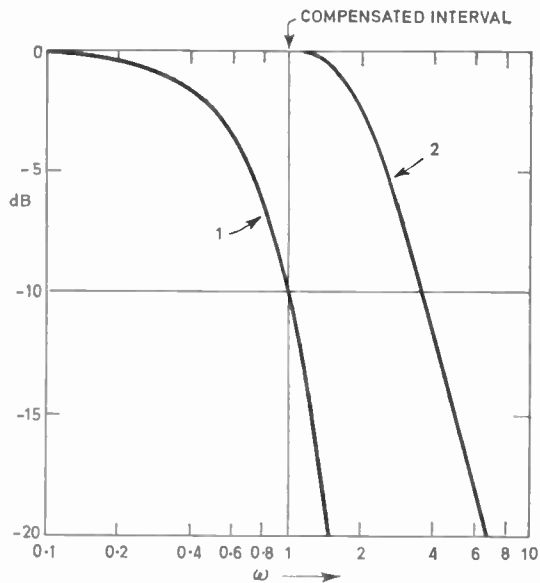


Fig. 2. Filter with six poles, realized with one active element. Safety margin is 33.9%.

1. Polynomial filter with poles
 - $-1 \pm j1$
 - $-1 \pm j0.6$
 - $-1 \pm j0.2$
2. Filter with the same poles and zeros
 - $-0.66640358 \pm j0.27474576$
 - $-0.68515841 \pm j0.83878779$
 compensating for the drop. Residual ripple is 0.01 dB.

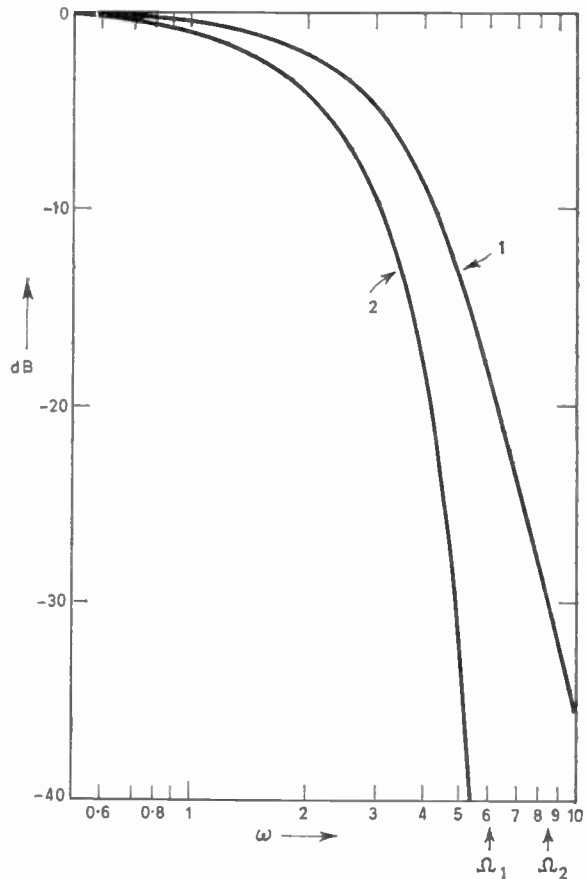


Fig. 3. Filter with six poles of the maximally-flat group-delay approximation. Safety margin 18.32%.

1. Polynomial filter with poles
 - $-4.2483594 \pm j0.8675097$
 - $-3.7357084 \pm j2.6262723$
 - $-2.5159322 \pm j4.492673$
2. Filter with the same poles and zeros
 - $0 \pm j6.090$
 - $0 \pm j8.99$

due to these poles will generally fall very early, because the outermost pole will be placed relatively far from the imaginary axis, and the selectivity will not be great. If constant (or any other) transmission is important in the pass-band, then it is possible to compensate for the fall of the response by proper choice of the zeros. Any method of polynomial approximation may be used. The example shown in Fig. 2 was approximated in a discrete minimax sense, using 21 points of the interval $<0, 1>$ and the method proposed by Stiefel.⁷ The choice of equidistant poles on the line parallel to the imaginary axis gives a large oscillatory safety margin of 33.9% (see Table 2). The amplitude characteristic of the polynomial filter (curve 1) is 10.13 dB down at the frequency $\omega = 1$, compared with the transmission at the origin. If four zeros are used to compensate for this fall, the ripple is reduced

to 0.01 dB in the same band, without influencing the safety margin (curve 2).

If selectivity is important, then purely imaginary zeros may be used for improvement. Such zeros may be realized by R-C twin-T circuits, or by other methods. This possibility is described,^{8,9} with all the necessary formulas. Synthesis of these active filters is also given.⁹ Thomson poles were used and the zeros are chosen in such a way that they cause Chebyshev behaviour in the stop-band. Figure 3 shows the selectivity curve of a filter with six poles, four zeros and a safety margin 18.32%. For comparison, the selectivity curve of the Thomson filter with only poles is shown. Other pole configurations may be used and zeros computed using formulas already published.^{8,9}

Approximation in the time-domain can be mentioned briefly. Suppose that the response to the Dirac impulse is given and is to be approximated. Because the pole-positions are chosen in order to fulfil the safety margin of k_{osc} , the natural frequencies and the damping factors of the filters are given and only initial amplitudes of the damped sinusoids must be determined. This reduces the problem to a linear one, and many methods, for instance the classical one of Kautz¹⁰, can be used. The results are generally good if the poles are chosen on straight lines passing through the origin and if the poles are equidistant on these lines. Results for some of these cases are shown in Table 5.

5. Conclusions

A safety margin for Horowitz-decomposition active filters was defined as a per cent change of the nominal constant, causing the oscillations of the filter. Many polynomial-type approximations were tested on this stability condition and it was shown that some realizations of the fourth and sixth degrees are quite safe as far as sensitivity to changes of the active-element constant are concerned. From the results shown above the design procedure for the given conditions follows immediately:

(i) If a given more complicated function is to be realized, then the outermost pole is tested for safety margin using the curve of Fig. 1. The safety margin of the remaining poles is then computed using a program based on the equations of Appendix 2 and if it is better than the result of the first pole, the remainder can be realized in one network. Otherwise the next outermost pole is realized by the elementary building block and the remainder tested on safety margin.

(ii) If the problem is given in terms of characteristics, then the design may use the following steps:

(a) Pole positions are selected and the polynomial-type filter is tested on safety margin. The

admissible lowest margin depends on the quality of the active element and better selectivity can be achieved by proper placement of poles, if a good stable active element is available.

(b) The results are improved by proper approximation of the transfer-function zeros. Some of the possible methods were mentioned above.

(iii) Should a band-pass filter be realized, then it is necessary to realize it as two filters, one low-pass and the other high-pass. The safety margin of the high-pass filter with poles p_i is equal to the safety margin of the low-pass with the poles $z_i = 1/p_i$.

Table 5
Poles on straight line passing through the origin

Type	Degree	k_{osc}	% change of k	ω_{osc}
$-0.5 \pm j0.5$ $-1 \pm j1$	4	0.658	34.1	1
$-0.5 \pm j0.5$ $-1 \pm j1$ $-1.5 \pm j1.5$	6	0.859	14.0	1.32
$-0.5 \pm j0.5$ $-1 \pm j1$ $-1.5 \pm j1.5$ $-2 \pm j2$	8	0.946	5.38	1.64
$-0.5 \pm j1$ $-1 \pm j2$	4	0.875	12.5	1.58
$-0.5 \pm j1$ $-1 \pm j2$ $-1.5 \pm j3$	6	0.971	2.93	2.13
$-0.5 \pm j1.5$ $-1 \pm j3$	4	0.926	7.39	2.24

These steps secure practical applicability of the circuit before the particular type of filter realization is chosen. The behaviour is known for large changes of k and may be checked additionally for small changes using sensitivity methods. This, however, is generally not necessary because enough information is contained in the number k_{osc} .

Transfer function with as many as six poles can be realized safely in one circuit, if the proper approximation method is used.

6. References

- Horowitz, J. M., 'Optimization of negative-impedance conversion methods of active RC synthesis', *Trans. Inst. Radio Engrs on Circuit Theory*, CT-6, pp. 296-303, September 1959.
- Hakim, S. S., 'RC active filters using an amplifier as the active element', *Proc. Instn Elect. Engrs*, 112, pp. 901-912, May 1965.

3. Yanagisawa, T., 'RC active networks using current inversion type negative impedance converters', *Trans. I.R.E.*, CT-4, pp. 140-144, September 1957.
4. Linvill, J. G., 'RC active filters', *Proc. I.R.E.*, 42, pp. 555-64, March 1954.
5. Vlach, J., 'Bandfilter mit gekoppelten Schwingkreisen von endlicher Güte', *Archiv Elektr. Übertrag.*, 17, pp. 537-46, August 1963.
6. Vlach, J., 'Approximation of low-pass filters with classical orthogonal polynomials', *Slaboproudý Obzor.*, 26, No. 3, pp. 50-58, March 1965.
7. Stiefel, E. L., 'Numerical Methods of Chebychev Approximation', in 'On Numerical Approximation', R. Elanger (Ed.) (University of Wisconsin Press, 1953).
8. Feistel, K. H. and Unbehauen, R., 'Tiefpässe mit Tschebyscheffcharakter der Betriebsdämpfung im Sperrbereich und maximal gegebener Laufzeit', *Frequenz*, 19, pp. 265-82, August 1965.
9. Vlach, J. and Bendik, J., 'Active filters with low sensitivity to element changes', *The Radio and Electronic Engineer*, 33, pp. 305-16, May 1967.
10. Kautz, W. H., 'Transient synthesis in the time domain', *Trans. I.R.E.*, CT-1, pp. 29-39, September 1954.

7. Appendix 1

Optimal Decomposition of Second-order Functions for Large Changes of *k*

Second-order transfer function

$$F(p) = \frac{N(p)}{Q_2(p)}$$

with poles ($a \pm jb$) is decomposed by means of an auxiliary polynomial $D(p)$ with distinct non-positive real roots

$$F(p) = \frac{N(p)/D(p)}{Q_2(p)/D(p)}$$

so that

$$\frac{Q_2(p)}{D(p)} = \frac{p^2 - 2ap + a^2 + b^2}{(p + \alpha)(p + \beta)} \dots\dots(9)$$

For $\beta > \alpha$, partial fraction decomposition results in

$$\frac{Q_2(p)}{D(p)} = \left(1 + \frac{A}{p + \alpha}\right) - \frac{1}{k} \frac{B}{p + \beta} \dots\dots(10)$$

where the constant k , expressing the behaviour of the active element, was incorporated in the denominator similarly as in eqn. (1) and the residues are

$$\left. \begin{aligned} A &= \frac{\alpha^2 + 2a\alpha + a^2 + b^2}{\beta - \alpha} \\ B &= \frac{\beta^2 + 2a\beta + a^2 + b^2}{\beta - \alpha} \end{aligned} \right\} \dots\dots(11)$$

Oscillations can build up for such a change of k that $Q_2(j\omega)/D(j\omega) = 0$. Insertion of $p = j\omega$ into eqn. (10) and use of equations (11) results in

$$\omega_{osc}^2 = a^2 + b^2 + 2a\alpha \dots\dots(12)$$

$$k_{osc} = \frac{\beta^2 + a^2 + b^2 + 2a\beta}{\beta^2 + a^2 + b^2 + 2a\alpha} \dots\dots(13)$$

The choice of α, β should secure the minimum value of k_{osc} , thus optimizing the safety margin. It can be easily shown that there exists no minimum with respect to α . The minimum with respect to β , e.g.

$$\frac{\partial k_{osc}}{\partial \beta} = 0$$

results in the condition

$$\alpha = \frac{\beta^2 - (a^2 + b^2)}{2(a + \beta)} \dots\dots(14)$$

so that eqns. (12) and (13) acquire the forms

$$\omega_{osc(min)}^2 = \frac{\beta^2 a + (a^2 + b^2)\beta}{\beta + a} \dots\dots(15)$$

$$k_{osc(min)} = 1 + \frac{a}{\beta} \dots\dots(16)$$

$k_{osc(min)}$ will reach the smallest value if β is small, because a is actually negative. As follows from eqn. (14), β is smallest for $\alpha = 0$, resulting in

$$\beta = \sqrt{a^2 + b^2}$$

and

$$k_{osc(min)} = 1 + \frac{a}{\sqrt{a^2 + b^2}} = 1 - \frac{1}{\sqrt{1 + H^2}}$$

This is, however, exactly the formula derived for the Horowitz decomposition. Therefore, in the case of second-order functions, this decomposition is the best one even for large changes of k .

8. Appendix 2

Equations for ω_{osc} and k_{osc}

Equations for programming are collected using the steps for Horowitz decomposition. Let the poles be ($a_i \pm jb_i$) and their number be even. Then the denominator is

$$Q_n(p) = \prod_{i=1}^{\frac{n}{2}} [p^2 - 2a_i p + a_i^2 + b_i^2]$$

1. Compute

$$T_i = \sqrt{a_i^2 + b_i^2} \dots\dots(17)$$

$$R_i = \sqrt{2(a_i + T_i)} \dots\dots(18)$$

and evaluate the polynomial

$$Q_H(P) = \prod_{i=1}^{\frac{n}{2}} [P^2 + R_i P + T_i] \dots\dots(19)$$

where the formal change of the variable $p = P^2$, based on the theory, can be disregarded in numerical computations.

2. Decompose $Q_H(P)$ into even and odd part

$$Q_H(P) = Q_{HE}(P) + Q_{HO}(P) \quad \dots\dots(20)$$

(Zeros of $Q_{HE}(P)$ are α_i , zeros of $Q_{HO}(P)$ are β_i and are simple. The square of the highest-power coefficient of $Q_{HO}(P)$ is equal to B/k of eqn. (1). The root-finding procedure would be applied to these polynomials and results inserted into eqn. (1).)

3. It is not necessary to know the roots for evaluation of k_{osc} and ω_{osc} and eqn. (20) may be used directly. Determine

$$Q_{HE}^2(p) = \sum_{i=0}^n E_i p^i \quad \dots\dots(21)$$

$$Q_{HO}^2(p) = \sum_{i=0}^n F_i p^i \quad \dots\dots(22)$$

and set

$$F_n = F_0 = 0 \quad \dots\dots(23)$$

which secures the special form of eqn. (1) and makes the following programming easier. For oscillations

$$Q_n(j\omega) = \sum_{i=0}^n E_i(j\omega)^i - \frac{1}{k} \sum_{i=0}^n F_i(j\omega)^i = 0$$

This equation splits into real and imaginary part, both to be equal to zero simultaneously

$$f_1(k, \omega) = \sum_{i=0}^{\frac{n}{2}} (kE_{2i} - F_{2i})(-1)^i \omega^{2i} = 0$$

$$f_2(k, \omega) = \sum_{i=0}^{\frac{n}{2}-1} (kE_{2i+1} - F_{2i+1})(-1)^i \omega^{2i} = 0 \quad \dots\dots(24)$$

Here, $f_2(k, \omega)$ was divided by ω to remove the trivial solution $\omega = k = 0$. This system of two non-linear equations can be solved by the Newton iterative method using matrix inversion. It was found that the convergence is always good and rapid for almost arbitrary initial values of $k^{(1)}$ and $\omega^{(1)}$. In all cases, these values were set equal to 1, if usual transfer functions with the pass-band approximately equal to 1 were checked.

Manuscript first received by the Institution on 9th January 1969 and in final form on 22nd May 1969. (Paper No. 1268/CC50.)

© The Institution of Electronic and Radio Engineers, 1969.

The Author



Jiri Vlach studied in Prague and received his M.S. and Ph.D. degrees in electrical engineering from the Prague Technical University in 1947 and 1957 respectively. In 1948 he joined the Research Institute of Radio-communications where he worked on antennas, direction finding, pulse communications, and later on network theory and the application of computers to network design.

He taught also at the Prague Technical University, published in Czech, German and English, and translated several books. Since 1967, he has been a Visiting Professor at the University of Illinois, U.S.A., and has published a book on computerized approximation and synthesis of linear networks. He is a Senior Member of the American Institute of Electrical and Electronics Engineers.

Contributors to this Issue



Dr. I. Aleksander obtained his first degree in electrical engineering at the University of Witwatersrand, South Africa, and his Ph.D. at the University of London. In 1959 he joined STC at Footscray, where he worked on the application of transistors and tunnel diodes to computer circuitry. Since 1962 he has held lecturer appointments at West Ham College of Technology and Queen Mary

College, London, and he is at present Reader in Electronics at the University of Kent at Canterbury. During this time he has pursued research on the electronics of artificial intelligence systems, switching theory, automata theory and control systems. He is the author of several papers and a book on these subjects. In 1962 he was joint recipient of the Institution's Charles Babbage Award.



Peter M. Cashin graduated B.E.(Hons.) in electrical engineering from the University of Canterbury, New Zealand, in 1966. He received his M.E. degree in 1968 after working on singly occurring nanosecond pulse recording techniques. A patent is held on a binary multiplier scheme arising from this work. He is currently studying for a Ph.D. degree in electrical

engineering at Canterbury in the field of learning machines (artificial intelligence).



Peter Atkinson (M. 1962) graduated with honours in electrical engineering from Imperial College, London, in 1955. He served a post-graduate apprenticeship and had further industrial experience with the G.W. division of English Electric. From 1959 to 1962 he was a lecturer in electrical engineering at North Hertfordshire Technical College and, from 1962 to

1964, a senior lecturer in control engineering at the College of Technology, Letchworth. In 1964 Mr. Atkinson became a lecturer in control engineering at Reading University and a founder member of the Department of Applied Physical Sciences. Apart from his research and teaching interests he has written a book on control engineering and is an industrial consultant. He is honorary treasurer of the Thames Valley Section of the I.E.R.E. and a member of the Institution's Examinations Committee.



D. A. H. Johnson, is a Senior Lecturer in the Department of Electrical Engineering at the University of Melbourne, Australia. After graduating from the University of Canterbury, New Zealand with B.Sc. and M.Sc. degrees in physics, he joined the New Zealand Post Office Engineering branch in 1953, where he worked on radio communication systems. In 1956 Mr.

Johnson took up a two-year C.B.I. Scholarship to study design and construction of radio communication systems in British industry. He was a New Zealand member of the Commonwealth team which worked with the British Post Office on *Telstar* and *Relay* communication satellite experiments in 1962-63. From 1966 until his present appointment in 1969, he was Lecturer in Electrical Engineering at the University of Canterbury, where his principal teaching and research interests were in the fields of communication and signal processing.



A. E. Gee was educated at Chippenham Grammar School and subsequently worked at the Royal Aircraft Establishment, Farnborough. He obtained a H.N.C. in Applied Physics in 1966 at Farnborough Technical College and is at present studying for a B.Sc. in cybernetics and instrument physics in the Department of Applied Physical Sciences, Reading University.



P. W. Philo graduated from Bristol University in 1952, where he gained an honours degree in pure and applied mathematics. From 1954 to 1959 he was with the Admiralty Education Service and in 1959 Mr. Philo was appointed a lecturer at Letchworth College of Technology, where he eventually specialized in mathematical statistics. Now a senior lecturer, he also acts as

a consultant for a local engineering company on the stability of cranes.

An Algorithm to Evaluate the Logarithm of a Number to Base 2, in Binary Form

By
 P. W. PHILO,
 B.Sc.†

Presented at a Conference on 'Electronic Switching and Logic Circuit Design' organized by the College of Technology, Letchworth with the support of the I.E.R.E. and held at Letchworth on 24th October 1968.

Summary: This algorithm evaluates, in binary form, $\log_2 N$ for a given value of $N > 0$ using steps which are easily reproduced by logic blocks. It is also possible to extend this to obtain 2^x , the antilogarithm to base 2 of a given number x .

1. Introduction

The impetus for this work came from a problem in the field of logic design.¹ This problem generated an interest in obtaining the logarithm of a number by a mathematical method which could be reproduced using appropriate circuitry. It quickly became clear that the logarithm to base 2 in binary form would best suit the requirement and the algorithm described in this paper was formulated. The algorithm may be implemented using a cellular logic array interconnected in an iterative manner so that it can carry out the operations of addition and multiplication.² The logarithm is obtained by using the array with a register consisting of master-slave flip-flops in a feedback loop.³

2. General Description

It is only necessary to consider numbers between 1 and 2 so that for X_1 , $1 < X_1 < 2$, the logarithm in binary form may be written as

$$\log_2 X_1 = 0 \cdot b_1 b_2 b_3 b_4 \dots$$

where b_i is the i th binary decimal place and takes the value 0 or 1.

The above may be rewritten in the form

$$\begin{aligned} \log_2 X_1 &= b_1 \times 0 \cdot 1 + b_2 \times 0 \cdot 01 + b_3 \times 0 \cdot 001 + \dots \\ &= b_1 \log_2 (2^{\frac{1}{2}}) + b_2 \log_2 (2^{\frac{1}{4}}) + \\ &\quad + b_3 \log_2 (2^{\frac{1}{8}}) + \dots \end{aligned}$$

as $0 \cdot 1$ is the binary form of $\log_2 (2^{\frac{1}{2}})$ etc.

This leads to the indicial representation of X_1 as

$$X_1 = 2^{\frac{1}{2}b_1} \cdot 2^{\frac{1}{4}b_2} \cdot 2^{\frac{1}{8}b_3} \dots \dots \dots (1)$$

If $X_1 \geq 2^{\frac{1}{2}}$ then $b_1 = 1$, otherwise $b_1 = 0$. To determine the value of b_1 it is necessary to test whether $X_1 \geq 2^{\frac{1}{2}}$ but it is easier to test whether $X_1^2 \geq 2$ and so the first step in the procedure is to square X_1 , the given number.

Squaring both sides of eqn. (1) we get,

$$X_1^2 = 2^{b_1} \cdot 2^{\frac{1}{2}b_2} \cdot 2^{\frac{1}{4}b_3} \dots \dots \dots (2)$$

Now define a new number X_2 as follows:

$$X_2 = \begin{cases} X_1^2 & \text{if } X_1^2 < 2 & (b_1 = 0), \\ X_1^2 \div 2 & \text{if } X_1^2 \geq 2 & (b_1 = 1). \end{cases}$$

Equation (2) becomes

$$X_2 = 2^{\frac{1}{2}b_1} \cdot 2^{\frac{1}{4}b_2} \dots \dots \dots (3)$$

This is similar to eqn. (1) and squaring X_2 and comparing X_2^2 with 2 determines the value of b_2 and introduces the number X_3 defined as

$$X_3 = \begin{cases} X_2^2 & \text{if } X_2^2 < 2 & (b_2 = 0) \\ X_2^2 \div 2 & \text{if } X_2^2 \geq 2 & (b_2 = 1) \end{cases}$$

$$X_3 = 2^{\frac{1}{4}b_1} \cdot 2^{\frac{1}{8}b_2} \cdot 2^{\frac{1}{8}b_3} \dots \dots \dots (4)$$

Repetition of the above process builds up $\log_2 X_1$ in binary form, digit by digit.

The following numerical work demonstrates the procedure by calculating the first six binary decimal places for $\log_2 1 \cdot 27$.

$X_1 = 1 \cdot 27$		
$X_1^2 = 1 \cdot 6129$	$= X_2$	$b_1 = 0$
$X_2^2 = 2 \cdot 601446$	$(\div 2 = X_3)$	$b_2 = 1$
$X_3 = 1 \cdot 300723$		
$X_3^2 = 1 \cdot 691880$	$= X_4$	$b_3 = 0$
$X_4^2 = 2 \cdot 862458$	$(\div 2 = X_5)$	$b_4 = 1$
$X_5 = 1 \cdot 431229$		
$X_5^2 = 2 \cdot 048416$	$(\div 2 = X_6)$	$b_5 = 1$
$X_6 = 1 \cdot 024208$		
$X_6^2 = 1 \cdot 049002$	$= X_7$	$b_6 = 0$
$\log_2 1 \cdot 27 = 0 \cdot 010110$ etc.		

The above calculation was carried out on an electronic desk-calculator. The accuracy of the result is not easy to assess theoretically but empirical tests

† Department of Mathematics, Science and Computing, College of Technology, Letchworth, Hertfordshire.

have shown a very good agreement with calculations made by more conventional methods. Three such comparisons are listed below, these were obtained by using an ALGOL program on an Elliott 803 computer.

Approximately a hundred such comparisons have been made and these indicate a discrepancy of at most 1 in the 21st binary decimal place which is

N	1.09						
$\log_2 N$ (algorithm)	0.000	111	111	101	001	111	110
$\log_2 N$ (conventional)	0.000	111	111	101	001	111	111
N	1.50						
$\log_2 N$ (algorithm)	0.100	101	011	100	000	000	011
$\log_2 N$ (conventional)	0.100	101	011	100	000	000	011
N	1.83						
$\log_2 N$ (algorithm)	0.110	111	110	011	000	100	100
$\log_2 N$ (conventional)	0.110	111	110	011	000	100	101

equivalent to about 0.0000005 in the decimal value of the logarithm. The round-off errors introduced by squaring do not have a significant effect on the early values of b_i because the precise values of X_i are not needed.

When the logarithm of a number X , not between 1 and 2, is required it may be expressed in the form $X = X_1 \cdot 2^k$ where $1 < X_1 < 2$ and k is a positive or negative integer, so that $\log_2 X = \log_2 X_1 + k$ and k becomes the characteristic of $\log_2 X$.

3. Conclusion

Hardware capable of implementing the algorithm has already been developed using integrated circuits in a form suitable for large scale integration. This system could be used to provide a 'log' facility on an electronic desk-calculator or as part of a larger data processing unit.

4. References

1. Dean, K. J., 'Design for a full multiplier', *Proc. Instn Elect. Engrs*, **115**, pp. 1592-4, November 1968.
2. Dean, K. J., 'Some applications of cellular logic arithmetic arrays', *The Radio and Electronic Engineer*, **37**, pp. 225-7, April 1969.
3. Dean, K. J., 'A fresh approach to logarithmic computation', *Electronic Engineering*, to be published.

Manuscript first received by the Institution on 1st November 1968 and in final form on 18th February 1969. (Short Contribution No. 120/Comp. 120).

© The Institution of Electronic and Radio Engineers, 1969.

Frequency-dependence of a Balun Evaluated by a Frequency Transformation

By

Professor Dr. ir.
M. C. VANWORMHOUDT†

AND

D. H. J. BAERT, M.Sc.

Summary: By using a network transformation the frequency dependence of the balancing and matching properties of a transmission-line balun can be evaluated by studying rational network functions.

1. Introduction

There exist, in network theory, a group of well-known transformations,¹ by which one can change a given RLC-network into another network, whose properties are simply related to those of the original network. A fairly general way to describe these frequency transformations is as follows:

From any given RLC-network one obtains a new network by:

- (a) Leaving the resistances unaltered.
- (b) Replacing every inductor (inductance L_k) by a one-port network whose impedance is given by

$$Z_k(s) = L_k \omega_0 f(s/\omega_1) \quad \dots\dots(1)$$

In this expression ω_0 and ω_1 are two normalization constants which have the dimensions of an angular frequency and s represents the complex frequency.

The function $f(s)$ is a dimensionless positive real function. All one-ports that are needed for replacing the various inductors are equal except for a change in impedance level.

- (c) Replacing every capacitor (capacitance C_n) by a one-port network whose admittance is given by

$$Y_n(s) = C_n \omega_0 f(s/\omega_1) \quad \dots\dots(2)$$

Again all one-ports needed are identical except for a change in impedance level. The one-ports replacing the inductors and the capacitors are obviously inverse networks.²

Any network function $N(s)$ pertaining to the original network is transformed into a corresponding network function $N'(s)$ belonging to the new network. The relation between $N'(s)$ and $N(s)$ is given by

$$N'(s) = N[\omega_0 f(s/\omega_1)] \quad \dots\dots(3)$$

It follows that the value of a network function of the transformed network at a given point s of the complex-frequency plane can be found by calculating the value of the corresponding network function of the original

network at the point $\omega_0 f(s/\omega_1)$ of the frequency-plane. For real frequencies we have

$$N'(j\omega) = N(j\omega') \quad \text{where} \quad j\omega' = \omega_0 f\left(\frac{j\omega}{\omega_1}\right) \quad \dots\dots(4)$$

Several of these frequency transformations have been studied extensively, e.g. the low-pass to high-pass, band-pass or band-stop filter transformations. A transformation not so widely appreciated is that in which every inductor is replaced by a short-circuited lossless transmission line with a characteristic impedance $\omega_0 L_k$ and with a given electrical length.

The capacitors C_n must then be replaced by an open lossless transmission line of the same electrical length and with a characteristic admittance $\omega_0 C_n$. This can be done because the open and short-circuited impedances of a transmission line of a given length and with unity characteristic impedance, are inverse quantities. This transformation has been used from a purely mathematical standpoint in different publications⁴ concerning lossless transmission line filters. It is thought that its interpretation as a special case of the well-known frequency transformations of networks will add to an intuitive understanding of its usefulness. Although cases could be considered in which lossy transmission lines would appear, we will restrict ourselves to the lossless case.

As will be shown below, the transformation can be used inversely to transform a class of transmission line networks into lumped-element networks. This class is restricted to networks containing only:

- (1) Resistors.
- (2) Open- and short-circuited segments of lossless transmission line having different characteristic impedances and all having an equal electrical length.
- (3) Segments of lossless transmission-line that are neither open- nor short-circuited, having different characteristic impedances and having lengths that are even multiples of the basic length of the open- and short-circuited segments.

† Laboratory of Electronics and Metrology, University of Ghent, Ghent, Belgium.

It will be shown below that these segments are indeed equivalent to networks containing only open- and short-circuited segments of basic length.

2. An Example

Out of many possible applications we shall consider the broad-band matching and balancing properties of the very commonly-used balun depicted in Fig. 1. A lossless and dispersionless transmission line with length l and characteristic impedance Z_1 replaces, at higher frequencies, the inverting transformer used in lower frequency baluns of similar type. The balancing and matching is exactly realized at that frequency where l equals one half-wavelength, and this is independent of the characteristic impedance Z_1 . We note that a two-port consisting of a transmission line of some length l can be transformed, by using Bartlett's bisection procedure,³ into a lattice consisting of open- and short-circuited sections of transmission line with length $l/2$. This network can be derived from the lumped-element network of Fig. 2 by a frequency transformation of the kind discussed before. In this case the function $f(s/\omega_1)$ must be chosen equal to $\tanh(\pi s/\omega_1)$. The ideal, 1 : 1 transformers have to be introduced because of the impossibility of interconnecting the nodes P and Q, without disturbing the lattice-network. It has been assumed implicitly that the twin-lead transmission line has an impedance with

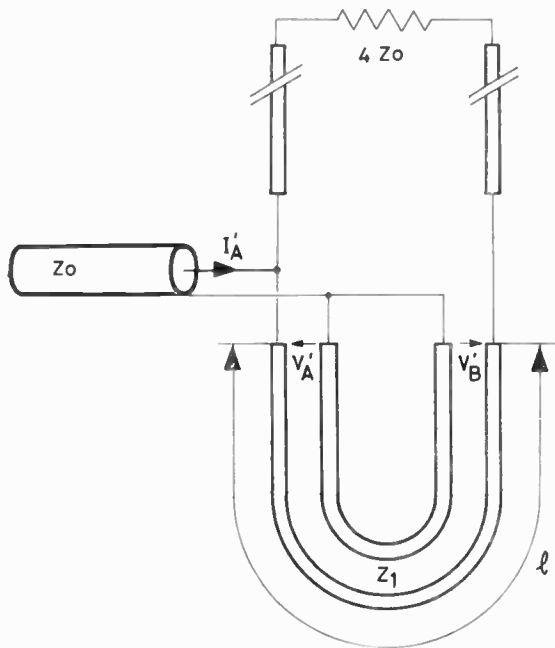


Fig. 1. The balun is represented by a transmission line of length l and characteristic impedance Z_1 .

respect to excitations in the symmetric mode that is much higher than that with respect to balanced excitation over the entire frequency band of interest. The values of L , C and ω_1 are given by

$$\omega_0 L = \frac{1}{\omega_0 C} = Z_1 \dots\dots(5)$$

$$\omega_1 = 2\pi \frac{c}{l}$$

where c denotes the phase velocity of the balun.

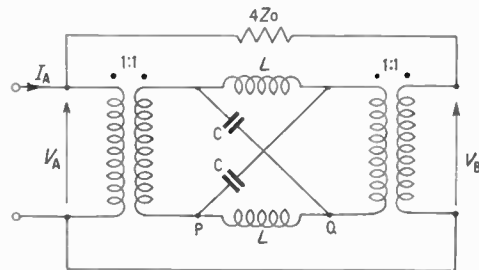


Fig. 2. Lumped element circuit derived from the balun of Fig. 1.

The normalizing constant ω_0 can be chosen at will. We easily obtain the voltage transfer function of the LC-network of Fig. 2.

$$T \triangleq \frac{V_B}{V_A} = - \frac{(s/\omega_0)^2 - 2k(s/\omega_0) - 1}{(s/\omega_0)^2 + 2k(s/\omega_0) + 1} \dots\dots(6)$$

In this formula k denotes the ratio $Z_1/4Z_0$. The matching properties are described by the normalized input admittance function:

$$\frac{Y}{Y_0} \triangleq Z_0 \frac{I_A}{V_A} = \frac{(s/\omega_0)^2 + \frac{1}{2k}(s/\omega_0)}{(s/\omega_0)^2 + 2k(s/\omega_0) + 1} \dots\dots(7)$$

where $Y_0 = Z^{-1}$.

The network functions (6) and (7) are rational in $s = j\omega$, and can be studied by standard methods, in order to find their values as a function of ω .

The absolute value and the phase angle of T and Y/Y_0 , are given by the expressions (8), (9), (10) and (11). Here x denotes the ratio ω'/ω_0 :

$$\text{mod } T = |T| = \left[\frac{(1+x^2)^2 + 4k^2 x^2}{(1-x^2)^2 + 4k^2 x^2} \right]^{-\frac{1}{2}} \dots\dots(8)$$

$$\text{arg } T = \tan^{-1} \frac{2kx}{1+x^2} - \tan^{-1} \frac{2kx}{1-x^2} \dots\dots(9)$$

$$\text{mod } \frac{Y}{Y_0} = \left| Z_0 \frac{I_A}{V_A} \right| = \left[\frac{x^4 + \left(\frac{x}{2k}\right)^2}{(1-x^2)^2 + 4k^2 x^2} \right]^{-\frac{1}{2}} \dots\dots(10)$$

$$\arg \frac{Y}{Y_0} = \tan^{-1} \frac{-1}{2kx} - \tan^{-1} \frac{2kx}{1-x^2} \dots\dots(11)$$

It is sufficient to draw the curves only for positive values of x because $\text{mod } T$ and $\text{mod } Y/Y_0$ are even in x and $\arg T$ and $\arg (Y/Y_0)$ are odd functions of x .

The results must finally be interpreted as properties of the balun of Fig. 1, and this is accomplished by performing the frequency transformation, e.g.

$$\frac{V'_B}{V'_A}(\omega) = \frac{V_B}{V_A}(\omega')$$

where

$$\begin{aligned} \omega' &= -j\omega_0 \tanh \left(j \frac{\pi\omega}{\omega_1} \right) \\ &= \omega_0 \tan \left(\frac{\pi\omega}{\omega_1} \right) \end{aligned}$$

The periodic nature of the frequency transformation $x = \tan \pi\omega/\omega_1$ results into the periodic behaviour of the balun as a function of frequency (Fig. 3).

A study of Figs. 4 and 5 reveals not only the known fact that exact balancing is obtained at $\omega' = \infty$, that

Fig. 3. The transformation function $x = \tan \pi\omega/\omega_1$.

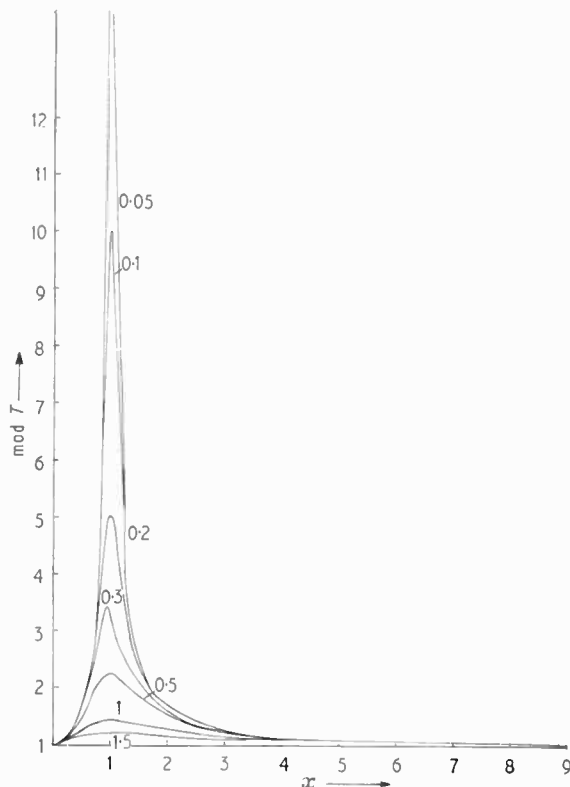
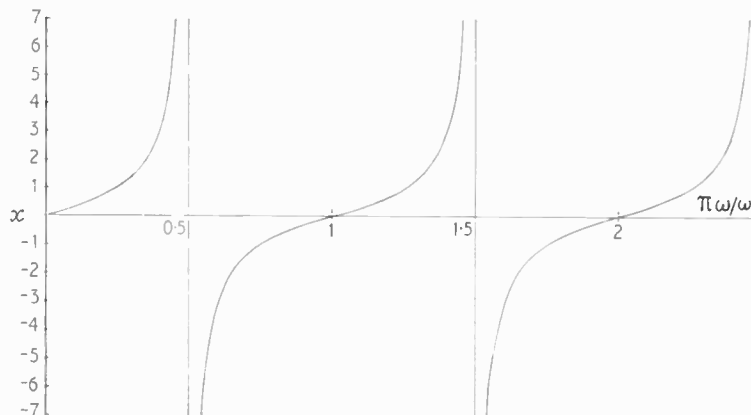


Fig. 4. $\text{mod } T$ as a function of $x = \omega'/\omega_0$. The parameter is $k = Z_1/4Z_0$.

is at $\omega = \frac{1}{2}\omega_1$ or $l = \lambda/2$, but it also shows the deviations from perfect balancing when ω changes. Balancing can only be achieved approximately within the range $|\omega'/\omega_0| > 1$. The approximation is best when k is much smaller than 1. If one chooses, for example, $k = 0.1$ and tolerates magnitude errors up to 25% and phase errors up to 8° in the balancing, $|\omega'/\omega_0|$ must be larger than 3. This then gives, according to Fig. 3, a corresponding frequency range for the balun extending between $0.815(\frac{1}{2}\omega_1)$ and $1.185(\frac{1}{2}\omega_1)$. Reducing k further has only a small effect on the balancing in the range $|\omega'/\omega_0| > 3$.

The situation is different for the matching properties. Figures 6 and 7 indicate that the input mismatch increases when k decreases. The functions $\text{mod} (Y/Y_0)$

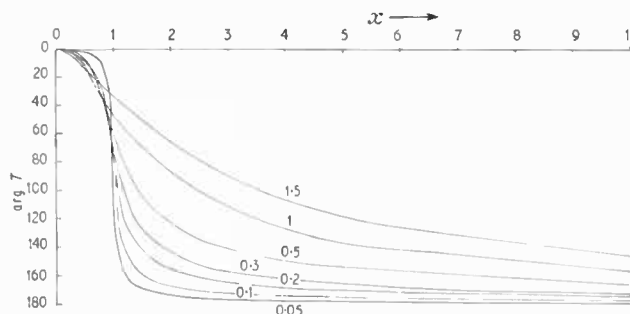


Fig. 5. $\arg T$ as a function of $x = \omega'/\omega_0$, parameter k .

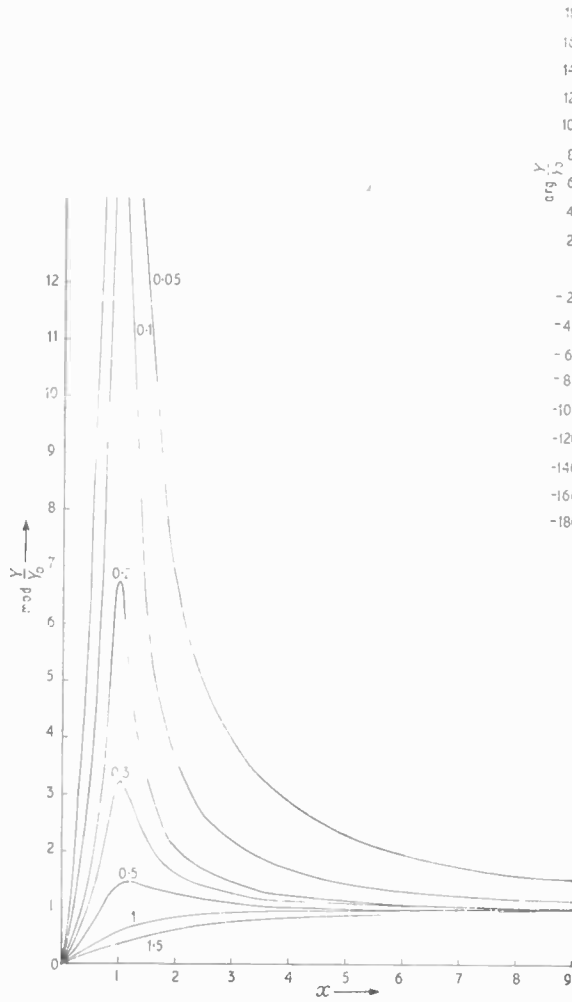


Fig. 6. $\text{mod } Y/Y_0$ as a function of $x = \omega'/\omega_0$; parameter k .

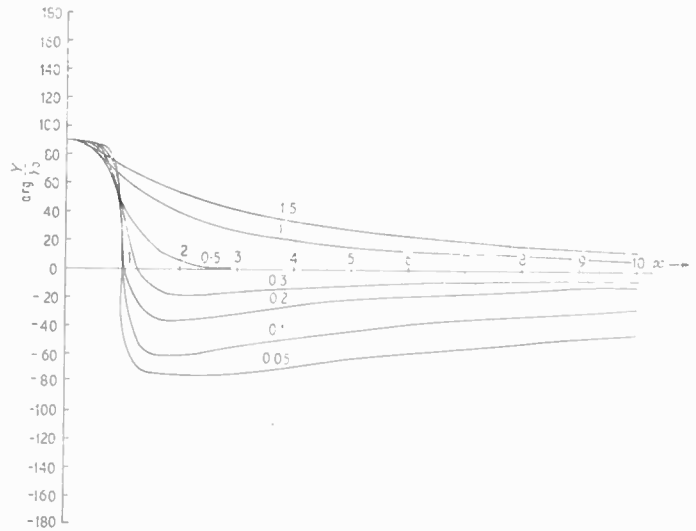
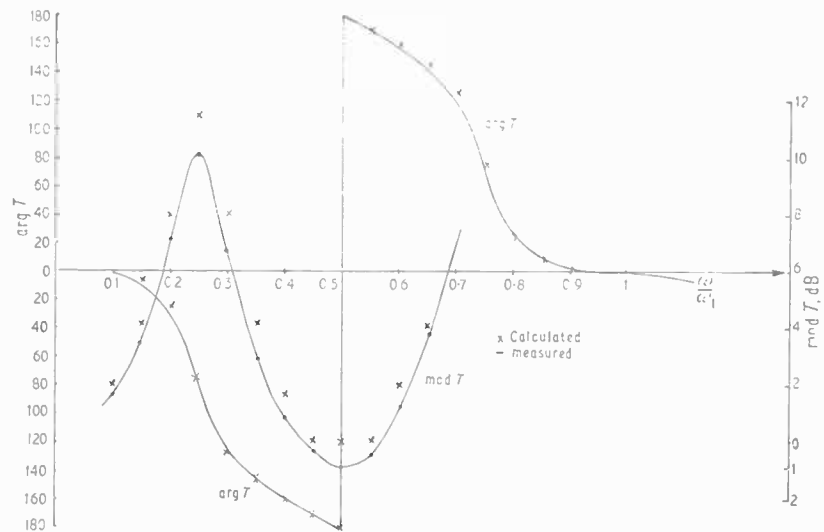


Fig. 7. $\text{arg } Y/Y_0$ as a function of $x = \omega'/\omega_0$; parameter k .

Fig. 8. Experimental results: the discrete points show the calculated values of $\text{mod } T$ (dB) = $20 \log T$ and $\text{arg } T$.



and $\arg(Y/Y_0)$ are very sensitive with regard to the parameter k . The design of a balun will therefore involve a compromise between the matching and balancing requirements within a given bandwidth.

2.1. Limiting Values

For small values of k ($k < 0.3$), it is possible to approximate the expressions (8), (9) and (11):

$$\begin{aligned} \text{mod } T &\simeq \left| \frac{1+x^2}{1-x^2} \right|; & (x > 1.5) \\ \arg T &\simeq \frac{4x^3}{x^4-1} k - \pi; & (x > 1.2) \\ \arg Y/Y_0 &= \frac{2x^3}{x^2-1} k - \frac{\pi}{2}; & (x > 1.5) \end{aligned}$$

2.2. Experimental Results

Figure 8 shows $\text{mod } T$ and $\arg T$ of a balun for which $4Z_0 = 253 \Omega$, $\omega_1 = 98 \text{ MHz}$ and $Z_1 = 70 \Omega$, i.e. $k = 0.275$. It is found that the measurements agree fairly well with the calculated values, especially when considering that line losses were not taken into account.

3. References

1. Tuttle, D. F., Jr., 'Electric Networks', p. 306 (McGraw-Hill, New York, 1965).
2. *ibid.*, p. 102.
3. *ibid.*, p. 185.
4. Kinariwala, B. K. 'Theory of cascaded structures: lossless transmission lines', *Bell Syst. Tech. J.*, 45, No. 4, pp. 631-49, April 1966.

Manuscript first received by the Institution on 28th October 1968 and in final form on 4th February 1969. (Paper No. 1269/CC51.)

© The Institution of Electronic and Radio Engineers, 1969.

STANDARD FREQUENCY TRANSMISSIONS—June 1969

(Communication from the National Physical Laboratory)

June 1969	Deviation from nominal frequency in parts in 10 ¹⁰ (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		June 1969	Deviation from nominal frequency in parts in 10 ¹⁰ (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	-299.9	—	0	567	—	17	-300.0	+0.1	0	558	478.9
2	-300.0	—	0	567	—	18	-299.9	0	0	557	478.8
3	-300.0	+0.1	0	567	482.5	19	-300.1	+0.1	-0.1	558	477.7
4	-299.9	-0.1	0	566	483.1	20	-300.0	+0.1	-0.1	558	476.2
5	-300.1	0	0	567	482.9	21	-299.9	0	-0.1	557	475.9
6	-299.9	0	0	566	483.0	22	-299.8	0	0	555	475.1
7	-300.0	-0.1	0	566	483.8	23	-299.9	0	0	554	474.9
8	-299.9	+0.1	0	565	483.2	24	-299.9	+0.1	0	553	474.3
9	-300.0	0	0	565	483.5	25	-300.0	+0.1	0	553	473.6
10	-300.0	0	0	565	483.9	26	-299.9	+0.1	0	552	473.0
11	-300.0	0	0	565	484.3	27	-299.9	+0.1	0	551	472.0
12	-299.9	+0.1	0	564	483.1	28	-299.9	+0.1	0	550	471.3
13	-299.7	+0.1	0	561	482.5	29	-299.9	+0.1	0	549	470.3
14	-299.8	0	0	559	482.1	30	-299.9	+0.1	0	549	469.3
15	-299.9	0	0	558	480.1						
16	-300.0	0	0	558	479.7						

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

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

† Relative to AT Scale; (AT_{NPL} - Station) = + 468.6 at 1500 UT 31st December 1968.

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.

BACKFIRE ANTENNA

The 'Backfire' is a novel antenna type that operates on a combination of surface-wave and open-cavity principles and produces gains ranging from 15 to 30 dB. Its side- and back-lobes in both planes are much lower and its construction is far more compact than that of conventional antennas in that gain range. It may be used as a single radiator as well as an array element for large antenna arrays.

In a German paper, after a brief introduction to the backfire principle, models of backfire antennas designed for ground-to-ground and air-to-ground communications, u.h.f. reception, telemetry, and automatic satellite tracking, are described and a special model is discussed which lends itself to flush installation in aircraft, missiles and space vehicles.

'"Backfire" antennas', H. W. Ehrenspeck, *Nachrichtentechnische Zeitschrift*, 22, No. 5, pp. 286-92, May 1969.

AN 'ELECTRONIC' D.C. MOTOR

The commutator of a conventional d.c. motor can be regarded as a system of controlled switches which reverses the voltage across the armature coils at certain positions of the rotor. Since the commutator segments have a fixed position in relation to the coils of the rotating armature, and the brushes have a fixed position in relation to the poles of the stator, the switching is always bound to occur at the right moments.

In small, fast-running machines the wear caused by friction and sparking limits the life of the commutator to a few thousand hours under normal conditions and possibly to a few minutes at very low pressure. A longer life without maintenance becomes possible if electronic switching elements are used instead of the mechanical commutator. To control these switches in the correct phase, however, the motor must be provided with a device which senses the position of the rotor, without using mechanical contacts, and which gives a position signal from which a suitable control signal is then obtained.

Switching systems of this kind, known as electronic commutators, and based, for example, on optical or magnetic methods of sensing the position of the rotor, have been known for some time. In all such systems there is a clear distinction between two functions: position sensing with information processing, and the actual power

switching by the commutator. This division of function has the fundamental consequence that the power switching can be carried out instead by a position sensor, i.e. depending on the position of the rotor, by a switching device whose operation can be controlled by any variable that one may choose. The special problems related to this approach have been studied intensively at the Philips laboratory at Aachen and are discussed in a recent paper.

The paper describes a motor which gives a rated power of up to about 20 W and in which the commutation is controlled by a magnetic position sensor that takes very little power. The state of the sensor is scanned at a high frequency. The armature coils are energized by means of power transistors, which are controlled by the position sensor in a switching pattern that gives virtually no harmonics of order divisible by 3 in the armature current. As a result of this and other features of the design the overall efficiency of such a motor can reach 80%.

'A highly efficient small brushless d.c. motor', W. Radziwill *Philips Technical Review*, 30, No. 1, pp. 7-12, 1969. (In English)

PARAMETRIC AMPLIFIER AS A TRACKING FILTER

The resonant frequency of a parametric amplifier changes according to the pumping frequency. Hence a parametric amplifier can be used as a time-varied tracking filter.

A paper from the Faculty of Engineering at Tokyo University discusses the fundamental characteristics of a parametric amplifier which operates as a tracking filter by changing the resonant frequency by variation of the pumping frequency. The theoretical equation for the practical variation range of the resonant frequency, the relation of the pump frequency and the resonant frequency, the gain, the bandwidth, etc., are introduced and are subjected to experimental checks. Thus, the variation range of the resonant frequency and the bandwidth can be designed to have desired values. Graphical analysis by immittance loci is shown. By this method, a parametric amplifier which has a complex circuit is easily analysed. A circuit arrangement to widen the practical variable range is also discussed.

'A study of a parametric amplifier as a tracking filter', T. Osatake, A. Fujii and T. Akutagawa, *Electronics and Communications in Japan* (English language edition of *Denshi Tsushin Gakkai Ronbunshi*), 51, No. 6, pp. 63-9, June 1968.