THE CHALLENGE OF 1966

THE turn of a year encourages a state of anticipating new enterprises. Anticipation is often greater than realization, but contemplation constitutes a challenge to influence events.

The relationship of the Council of Engineering Institutions with the professional bodies and indeed with individuals invokes a challenge and a change to the existing order. The formal establishment of a Register of Chartered Engineers will be but the first step in a process which must end in strengthening the influence of the engineer. C.E.I. has still to make an impact on the individual engineer and there will be much discussion over the coming months on the extent to which the engineer's eventual specialization can be reflected in the selection of optional subjects in the C.E.I. examination. The ultimate aim of the C.E.I. is to require the engineer of the future to possess a university degree or its equivalent. This poses the question of whether the Higher National Certificate and other hitherto acceptable examinations will continue to have a place in the training of the professional engineer.

The development of the C.E.I. examination to succeed those set by the individual institutions might appear to be principally of interest and concern to members in Great Britain. The influence of C.E.I. must be world wide, however, for there will have to be a process of equating the qualifications of prospective members from overseas with the new schemes—or extending the facilities for training to take the C.E.I. examination. The extension of our own Institution's activities in Divisions and Local Sections will assist in promoting discussion of educational matters with the universities and technical colleges as well as government educational departments in their respective countries.

The Divisions and Local Sections outside Great Britain are also active in promoting the 'learned society' aspect of the Institution and the comments made in the editorial article in the December 1965 issue of this *Journal* have stressed the important links which can lead to the discussion of locally generated problems on a much wider forum. Ambitious programmes of meetings have already been planned by several of these Sections and Divisions.

The two major Institution Conferences in Great Britain this year will be characterized by a strong international flavour in the papers, as well as the attendance. These Conferences will deal with Applications of Thin Films in Electronic Engineering (in London in July) and Electronic Engineering in Oceanography (at Southampton in September). This latter Conference in particular will deal with the impact of electronic techniques on a field notable for international co-operation.

International co-operation is being encouraged by the National Electronics Research Council in its discussions with research organizations in Canada, India and Australia. Such co-operation in connection with the Council's project on selective dissemination of information can considerably improve the effectiveness of the project in promoting a greater awareness of research work in progress.

Technical milestones during 1966 are hardly predictable in advance, but it seems probable that colour television and its introduction will continue to be a controversial subject. Recent developments in solid-state electronics have been legion and devices such as bulk-effect microwave generators, integrated circuits and thin film active devices present fresh challenges to the ingenuity of the engineer in exploiting their possibilities. Communications satellites represent the acceptance of other challenges.

Thus in professional affairs, education, technical meetings and developments, the continuing theme of challenge which is inherent in science and engineering calls for the maximum response of every chartered engineer in 1966.

F. W. S.

INSTITUTION NOTICES

Secretary to Visit Overseas Divisions and Sections

On 8th February Mr. Graham D. Clifford, Secretary of the Institution, will leave London by air on a two-months' visit to Overseas Divisions and Sections and groups of members in centres round the world.

His itinerary will include several days in Israel, Pakistan, India (where he will visit all five Sections), New Zealand and Australia. He will then visit Ottawa again in April.

Members in these centres will be sent details of meetings arranged during Mr. Clifford's visit by the Local Honorary Secretary or Local Representative.

Civil Service Recognition of Membership

The Civil Service Commissioners in Great Britain have notified the Institution that the I.E.R.E. Graduateship Examination is an acceptable qualification for appointment of mechanical and electrical engineers in Government departments, irrespective of the date at which the Graduateship Examination was passed or when exemption from it was obtained.

This recent decision of the Commissioners therefore modifies the stipulation, published in the January 1965 issue of *The Radio and Electronic Engineer*, that a candidate's Graduateship Examination pass or exemption should have been based on the 1962 syllabus.

Canadian Division

Following the recent visit of Mr. Graham D. Clifford to Canada, a third Canadian Local Section has been formed to serve the interests of members in Ottawa. Consequently, the constitution of the Canadian Divisional Council has been amended to incorporate representatives from all three Canadian Sections. The Council now comprises:

Chairman: Professor A. D. Booth (a Vice-President of the Institution);

J. H. Beardall (Hon. Secretary, Toronto Section);

K. N. Coppack (Hon. Secretary, Montreal Section);

R. G. Hall (Chairman, Toronto Section);

H. Rondeau (Chairman, Ottawa Section);

E. F. Wale (Chairman, Montreal Section);

R. W. Wray (Hon. Secretary, Ottawa Section).

Honorary Secretary to the Council: H. Rondeau.

Honorary Treasurer to the Council: R. W. Wray.

In addition to establishing the Institution's Canadian office in Ottawa, Mr. Clifford attended meetings of members in Ottawa, Montreal and Toronto. Each Section is producing an interesting programme of meetings and arranging to secure contributions to the *Canadian Proceedings*.

Arrangements have been made for the *Canadian Proceedings* of the Institution to be published in Canada four times a year. The Editor is Professor A. D. Booth and papers, contributions and news items for the *Canadian Proceedings* should be sent to him at the Faculty of Engineering, The University of Saskatchewan, Saskatoon, Canada.

At the conclusion of his visit, Mr. Clifford visited Baltimore, Washington and New York, and met Institution members in these cities.

I.F.A.C. Congress, London, 1966

The Third Congress of the International Federation of Automatic Control is to be held in London from 20th to 25th June, 1966. The Congress is being organized on behalf of the I.F.A.C. by the United Kingdom Automation Council, which is responsible for co-ordinating the programme of technical sessions and informal discussion groups and for all aspects of the arrangements, including accommodation for delegates, booking conference halls and lecture theatres. The Institution is a member of the U.K.A.C.

Further information may be obtained from I.F.A.C. Secretariat, U.K.A.C., c/o Institution of Electrical Engineers, Savoy Place, London, W.C.2.

Canadian Division Dinner

Earl Mountbatten of Burma, K.G., will be the Guest of Honour at a dinner to be given by the Canadian Division of the Institution at the Chateau Laurier Hotel, Ottawa, on Thursday, 14th April.

On the afternoon of the same day Lord Mountbatten is to deliver a lecture, on the work of the National Electronics Research Council, in the Lecture Theatre of the National Research Council of Canada.

Lord Mountbatten will be visiting Canada to discuss, among other matters, the possibility of closer ties between N.E.R.C. and Canadian research organizations.

Engineering Institute of Canada

A reciprocal arrangement has been made with the Engineering Institute of Canada to enable members of either Institution to obtain the Journal of the other at a specially reduced rate. I.E.R.E. members may obtain *The Engineering Journal* at a charge of £1 9s. per year.

Subscription orders should be sent to the I.E.R.E., 9 Bedford Square, London, W.C.1. Members in Canada and the United States of America should send their orders to the Administrative Secretary of the Institution's Canadian Division, 504 Royal Trust Building, Albert Street, Ottawa 4, Ontario.

Environmental Realism in Flight Simulators

By

A. E. CUTLER, Ph.D.†

Summary: In flight simulators great care is taken to display to the pilot under training, as accurately as possible, the instrument readings appropriate to his handling of the aircraft. The pilot cannot make full use of this information, however, unless it is presented in correct relationship with other data derived from his simulated environment. The paper traces the development of environmental simulation, with particular reference to control loading, fuselage movement and visual systems.

1. Introduction

During the last decade, electronic flight simulators have been adopted by nearly all the world's major airlines for aircrew training. Although expensive in themselves, they are so much cheaper to use than actual aircraft that the simulator manufacturer is under constant pressure to extend the range and validity of his product as a training device. At present the flight simulator can replace a large proportion of aircraft hours, particularly in cockpit familiarization, instrument flying, and emergency procedures, without the problems of weather, availability and expense attendant upon operating real aircraft, and in far greater safety.

All manufacturers until very recently based their equipments on specially designed but otherwise conventional electro-mechanical analogue computing techniques.[1-3] These have coped handsomely with the computing task but they require a fair amount of maintenance and are a problem to keep in an up-to-date modification state as aircraft changes are made.

The possibility of using digital computers in a similar role to avoid these problems has been recognized for a number of years[4, 5] but when they were first proposed there were a number of good reasons why the analogue equipment was preferred. The development of digital computers, spurred on by the availability of cheap and reliable semiconductors, has been extremely rapid whereas analogue computation has developed much more slowly, so that the gap between the two techniques has now closed and both types of computers are equally capable of meeting the requirement, the digital equipment still being rather more expensive. The difference in cost, however, is now such that it may be regarded as a reasonable payment for the improved reliability and flexibility obtained. In consequence, a number of simulator manufacturers are offering digital equipments and several digital simulators have already been put into service. It is likely that, for some time to come, procedure trainers, type trainers and the simpler

flight simulators will continue to use analogue computation.

At this present time, when the simulator industry is undergoing a major technical revolution, it is interesting to examine the state of the art to determine to what extent the developments associated with the analogue flight simulator are passing over into the new era of technology. By definition, these are inevitably concerned with those aspects of simulation which are distinct from the computer itself. In the broadest sense these equipments may be described as interface equipments because they are all concerned with the process of translating the digital computer outputs into analogue form as required to synthesize the training environment. They include instrument drives and systems for noise generation, control loading, fuselage movement and visual simulation.

In most of these areas there have been major developments during the last few years as the flight simulator has been brought from the basic form in which it was originally conceived, when the emphasis was on flying by instruments alone, to its present state where it is recognized that all forms of data, whether derived from instruments, noise, fuselage motions, control feel or visual displays, are significant to the pilot. The view is nowadays widely held that unless these sources of data are available in combinations such that the pilot recognizes the situation which they create instinctively as corresponding to flying an actual aircraft, then the psychological involvement required for the perfection of simulation is impaired and it is possible that bad habits may be formed. This may be termed 'negative transfer'.

The development of the psychological approach to training and its influence on simulators has not been extensively dealt with in the literature and some pointers are given in this paper to the more important examples of this approach.

2. The Man-Machine System

Although during the last war there were several experiments designed to establish the relationship between human beings and their environment, this

---

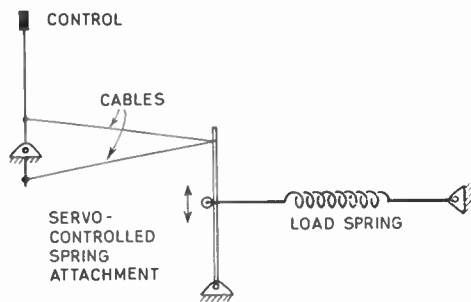† Redifon Ltd., Flight Simulator Division, London, S.W.18.

Fig. 1. Schematic diagram of the variable lever control loading arrangement.

work did not seem to catch the popular imagination until the middle 50's when, largely as a result of work carried out in the U.S.A., the concept of the man-machine system was proposed and elaborated, mainly in connection with the complex problems of military aircraft and space vehicles. A by-product of the interest which centred on the solution of the immediate problems was an increased awareness that pilots do not fly their aircraft merely by applying their analytical skill to the indications of their flight instruments, but that they use instrument data in conjunction with other data derived from their own bodily sensors to create for themselves a sensation of their relationship with the outside world, by virtue of which they are able to maintain refined spatial control of their aircraft of a kind which is not possible by analysis of the instrument data alone. Failure to recognize the importance of these factors can be seen in retrospect to have been responsible for a number of complaints relating to the handling qualities of early simulators.

Whilst there has been over the years a steady improvement of technique in every area of simulation, the greatest yields, measured in terms of increased realism, have come from developments in:

control loading,

fuselage movement,

and visual simulation,

and these will be considered in detail.

3. Control Loading

The first electronic flight simulators were manufactured in 1949 and 1950 by Curtiss-Wright and Redifon for Pan-American and B.O.A.C. respectively. The devices which were used to reproduce the forces on the pilot's controls in these simulators were based upon the variable lever principle (Fig. 1), in which an electro-mechanical servo is used to change the leverage between the control and a spring in simulation of the increase of control stiffness with airspeed. Statically, the force-deflection characteristics of this type of device are very good, particularly when the variation is linear, and developed versions of the system are still used extensively by Redifon on type and procedure trainers where simplicity and robustness are important (Fig. 2). The technique was soon criticized in full flight simulator use because it could not reproduce the precise characteristics of the aircraft control systems, in particular such effects as the inertia of the control surfaces, mechanical and aerodynamic damping, control run stretch, lost motion, shock
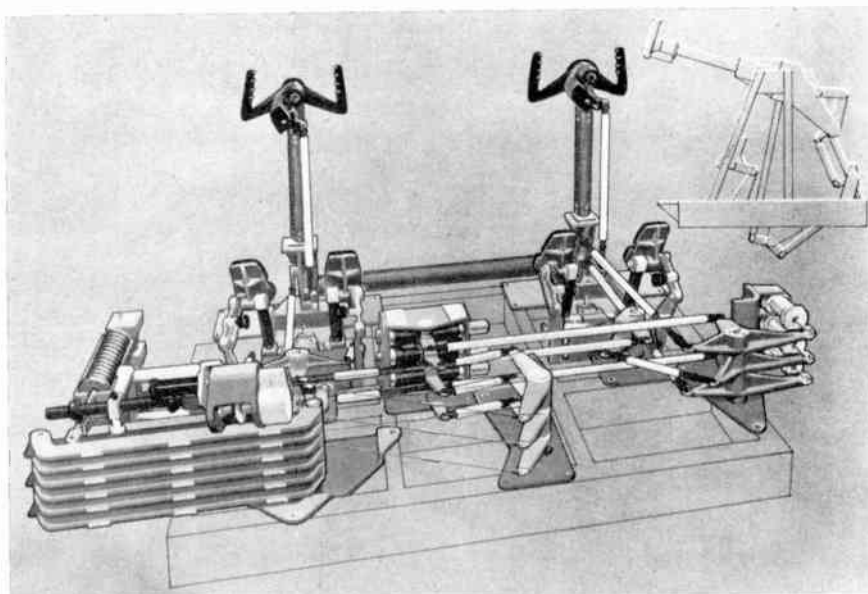

Fig. 2. Practical embodiment of variable lever control loading including the variable lever assembly (left), motorized trim changing unit (centre) and dampers (right).
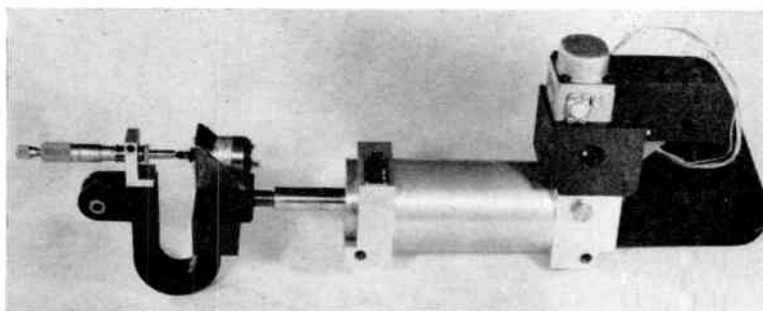
Fig. 3. Hydraulic force jack control loading unit. The U-beam and displacement pick-off form the force sensing element.

loads, autopilot coupling and miscellaneous knocks and non-linearities.

It was natural to seek a more direct way in which to convert the electrical signals which were computed to represent the required effects into the actual forces on the controls, and in 1954 such a device based upon stalled motors was produced and later installed in a flight simulator for the British Ministry of Supply (now Ministry of Aviation). This device was designed to produce shock loads of up to 200 lb (90 kg) at the control handwheels to simulate automatic reversion of the control system from power assistance to manual operation when the hydraulic supply was cut off. In this device the reflected inertias were approximately matched to those of the aircraft and electrical damping was used both for stability and to simulate the actual control damping. The 2 kW or so of controlled power required to drive the motor was produced by thyratrons in a split-field arrangement, the phase of the control signals being varied differentially to obtain d.c. drive to the motor. The inertia of the motor provided the necessary smoothing. This equipment was very realistic in operation and valuable simulation studies were carried out on it to develop emergency handling techniques which were later adopted in the aircraft. Attempts to apply this method on a more moderate scale met with a variety of problems. Electro-hydraulic systems essentially of the variable lever type were also tried and for a time were accepted, although the pressure to produce more faithful simulation was mounting continuously. Curtiss-Wright were able to apply the stalled torque motor technique with fair success by gearing the motor to produce the required torque from moderate input power and, since this increased the reflected inertia excessively, using positive acceleration feedback to reduce the effective inertia. A.c. induction motors were used for smoothness of control driven by conventional 200-watt class AB valve amplifiers but the falling off of their torque/speed characteristic at high speeds made it possible to beat the motors and the controls hardened unrealistically if moved quickly.

Nevertheless, this method went some way towards giving the required performance because it permitted many of the effects mentioned above to be obtained.

It had been realized for some time that the force which the pilot felt was often not in accordance with the computed signal but was modified by the characteristics of the force producing device. Every effort had been expended on removing these obtrusive characteristics, the only approach available in an open-loop system. About 1958 a number of high performance electro-hydraulic servo valves began to appear on the market and attention was once more turned to the possibility of achieving the desired results with a hydraulic system. The very wide frequency response available from these new servo valves made it feasible to consider the possibility of using negative feedback to reduce the distortions and force transducers were placed in the linkages to produce the required feedback and to ensure that the forces applied to the control column were in accordance with the demand. As the best pick-offs for small displacements are a.c.-operated, and the servo valves require d.c. drive, the amplifier contains a phase-sensitive rectifier and considerable smoothing, since the servo valves still have good response at hum frequencies. This type of device was found to work very successfully, making a very compact and convenient mechanism, becoming known as the force jack (Fig. 3). The frequency response of a high grade hydraulic servo valve is shown in Fig. 4. The force
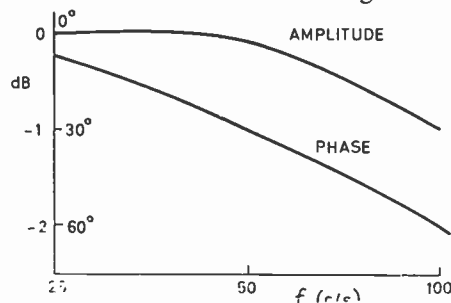

Fig. 4. Phase and amplitude plots for a typical hydraulic servo valve.

applied to the control column in a typical system for inputs of various frequencies is plotted as a frequency response for the two extreme conditions of stick held elastically and stick rigidly held in Figs. 5 and 6.
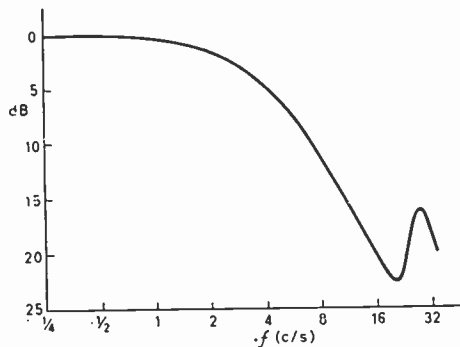

Fig. 5. Frequency response of force jack control loading unit with the stick loaded elastically.
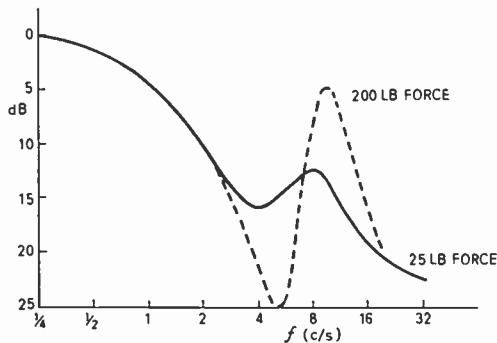

Fig. 6. Frequency response of force jack control loading unit with the stick clamped, shown for different peak force demands.

The use of force feedback ensures that the force produced is largely independent of the characteristics of the jack and the yielding of the support structure. In this system the force is representative of the computed signal within an acceptable tolerance over a pass-band which includes all frequencies at which a human operator is capable of discrimination. A great variety of effects can be achieved simply by providing the appropriate electrical inputs and the force jack has raised the standard of control loading simulation to a point where the pilot is at last completely unaware of any differences between the simulated and actual systems. Except for the inevitable technical refinements it seems very unlikely that the force jack will be found wanting when used with digital simulators.

4. Fuselage Movement

The wartime Link trainers had a simple pneumatic fuselage movement system. This system, although in keeping with the simplicity of the trainer of which it was a part, did not produce realistic effects and had a marked influence on the later acceptance of fuselage

motion in training equipments. At the time when simulators were first introduced, flying 'by the seat of the pants' was also very much out of favour and as a result, flight simulators were manufactured without fuselage motion for a number of years. Not an insignificant factor in this decision was the obvious mechanical difficulty of moving realistically a 4,000 lb flight deck in accordance with the computed flight manoeuvres. It was found that by an *ad hoc* process the control loading forces, the control effectiveness and the simulated aircraft dynamics could be adjusted to produce a combined effect on the pilot which in some way compensated for the absence of motion cues, but it was nevertheless quite obvious to the pilot under training, if he gave it thought, that he was still very firmly on the ground, so that the required degree of involvement could not always be achieved.

The pilot's instinctive feeling that some data were missing manifested itself in excessive scrutiny of the instrument indications and to avoid the criticisms which arose if these did not appear to be sufficiently refined a great deal of development was carried out to perfect both the computers generating the information and the instruments displaying it. One particular controversy which arose out of this need for computing refinement was whether a.c. or d.c. electro-mechanical computing gave the best accuracies.

This question is mainly of historical interest now that digital techniques are entering so strongly into the contest and it suffices to point out that the majority of analogue simulators made today in this country use both techniques. D.c. computing is used in areas where accurate dynamic performance and good small-signal response are essential; a.c. computing is used particularly for trigonometrical computing but also in many other areas where its use confers reliability, simplicity and ease of maintenance, factors which in flight simulators are often as important as performance.

For several years in the middle to late 50's simulator manufacturers and airline operators discussed the pros and cons of fuselage movement, most manufacturers making technical proposals, but the airline operators, being either not convinced of the need or not anxious to undertake a costly experiment, were slow to respond. Pilots whose opinions were sought in this matter were not very helpful as few of them had any clear idea which cues they made use of in flying and even those who favoured movement often did so for the wrong reasons. The slowness of pilots and therefore operators to accept movement systems had already cost the manufacturers dear in the excessive 'calibration' of fixed based simulators mentioned above, and their lack of understanding of the role of motion cues was also to prove costly during the early days of simulators with movement.
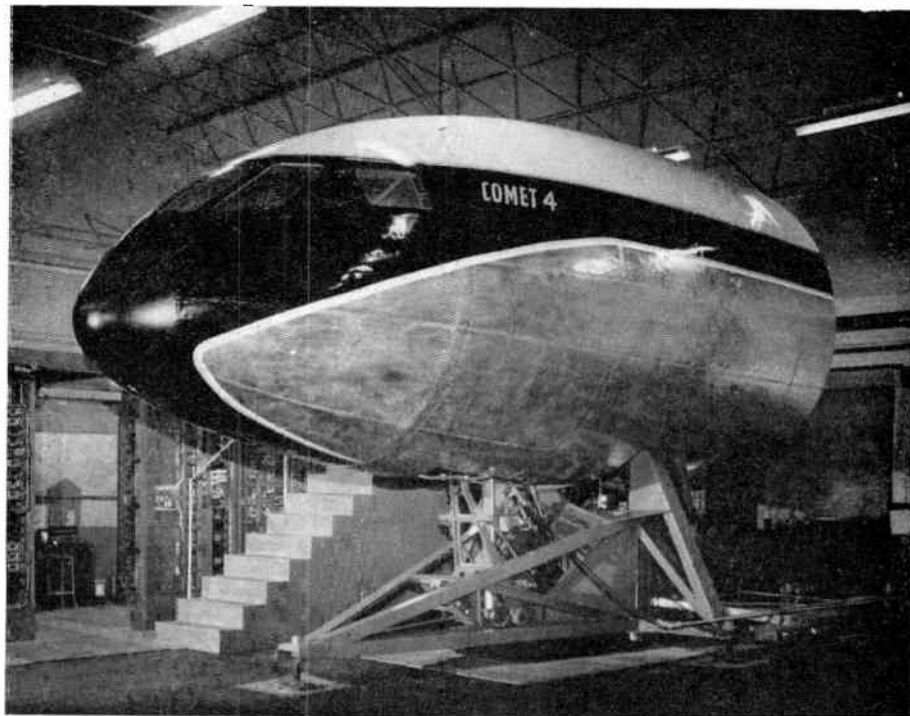
Fig. 7. The Redifon *Comet 4* flight simulator for B.O.A.C.; the fuselage is pivoted for pitch motion.

The breakthrough came when the B.O.A.C., in the course of constructing a simple procedure trainer, adapted some spare Link trainer parts to provide limited pitch motion about an axis a few feet behind the pilot. Their observation was that this limited simulation enormously enhanced the feeling of flying and in 1958 they commissioned Redifon to incorporate a pitch motion system into the *Comet 4* simulator. A number of other airlines reached their own decision at about that time and movement systems were soon on order from all the leading simulator manufacturers. In the Redifon equipment, a forward fuselage section from an earlier mark of *Comet* was available and this was adapted as shown in Fig. 7. The system was counterweighted by concrete blocks and the fulcrum was about level with the pilot's seat and 10 feet behind.

Pitch movement was achieved by the use of a hydraulic ram with fluid pumped from one side of the piston to the other using variable speed pumps positioned according to the rate of pitch required. A number of safety features were incorporated. This equipment proved very satisfactory in service and is still in continuous use at the B.O.A.C. Central Training Unit.

Whilst the pitch motion alone made a very valuable contribution to realism, the development was completed at a time when there was a world-wide awakening to the importance of motion cues in flying, and operators began to ask for much more faithful simulation. In fact, if the price were right, operators would like the simulator to fly! The suggestion has been made more than once that the simulator should be mounted in a freight aircraft, with the simulator controls so coupled to the aircraft controls through a computer that the effective response is representative of the aircraft simulated. Airborne simulation of this type is in fact used in the U.S.A. for research into the handling properties of proposed new aircraft.[6] Clearly, although it is excellent from the point of view of realism, the airborne approach defeats some of the main objects of simulation, in particular economy, safety, and independence of weather, and therefore some compromise must be sought.

The motions of a solid body consist classically of a linear velocity and an angular velocity about the centre of gravity. For convenience it is customary in aircraft usage to resolve these velocities into components along orthogonal axes through the centre of gravity. The x axis is parallel to the aircraft centre line looking forwards, the y axis at right angles to this and parallel to the wings looking right and the z axis is at right angles to the other two looking downwards, forming a right-handed set. On these axes the linear motions may be called surge, sway and heave; the angular motions are conventionally called roll, pitch and yaw.
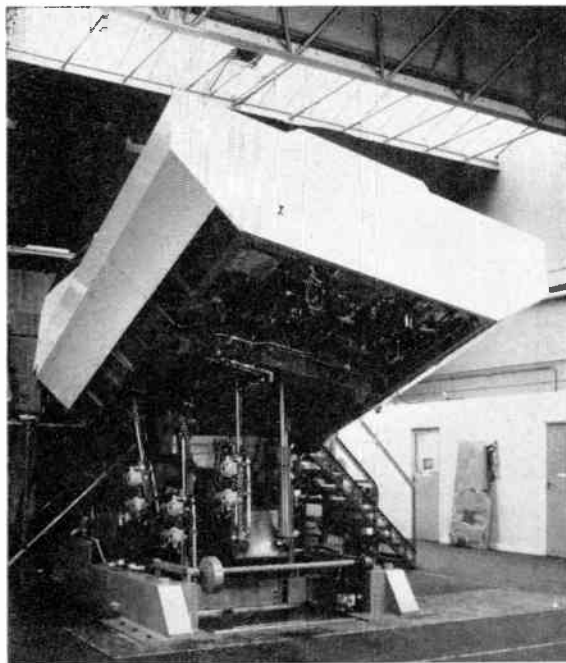
Fig. 8. The Redifon Boeing 727 flight simulator for Lufthansa in the extremes of pitch and heave motion. The hydraulic actuator packages are clearly visible.
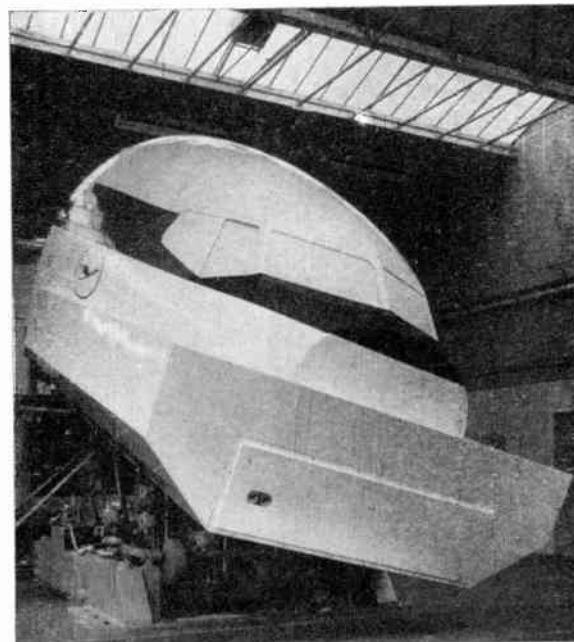

Fig. 9. The Redifon Boeing 727 flight simulator for Lufthansa in a diving attitude.

Now the human body is not sensitive to linear velocity but only to linear accelerations, through the medium of the resulting forces and pressures. It is also not apparently very sensitive to attitude but appears to react both to angular velocity and angular accelera-tion. These motions are detected by the semi-circular canals of the ear in a way which is not fully under-stood. A further factor is that the bodily sensors seem mainly to monitor change, the reaction to steady stimuli being apparently turned off by the brain. Hence we can sit down on to a chair using our pressure sensors to tell us sensitively when we touch, but may remain seated for long periods thereafter without being aware of the support given by the chair.

Fortunately, these observations permit some sim-plification in the design of motion systems without perceptible loss of realism. For example, in the heave direction there is a background level of reaction on the body equal to its own weight, compared with which small steady increments of force are rapidly forgotten, so that cueing information based upon acceleration transients gives good simulation. Surge and sway are quite important but fortunately they can be simulated approximately through the roll and pitch system. Detectable roll and pitch rates normally only persist for short periods, because normal changes in attitude are accomplished very quickly, but the new pattern of linear accelerations

Table 1

Characteristics of Redifon Fuselage Motion Systems

Maximum permitted load	9000 lb (4100 kg)	16 000 slug* ft² (22 000 kgm²)	8000 slug ft² (11 000 kgm²)
Range of movement	± 12 in (30·5 cm)	+ 15 deg − 10 deg	± 10 deg
Peak velocity	± 12 in/s (30·5 cm/s)	± 10 deg/s	± 10 deg/s
Peak acceleration	± 0·6g	± 10 rad/s²	± 10 rad/s²
Peak amplitude at 1/3 c/s	± 6 in (15·2 cm)	± 5 deg	± 5 deg
Peak amplitude at 2 c/s	± ½ in (1·25 cm)	± ½ deg	± ½ deg

* 1 slug = 32·174 lb

which results is related to the new attitude and persists indefinitely. The roll and pitch systems may without noticeable interaction be used to create both effects. They may be operated by roll and pitch accelerations to cue the onset of manoeuvres, being returned not to zero as the simulated attitude stabilizes but to a fuselage attitude such that the gravity vector representing the pull of the pilot's weight is in the same direction with respect to the fuselage as the resultant acceleration vector which would act on the pilot in the actual aircraft in a similar attitude. In this way a system of components along the fuselage axes is created in the required proportions but with the restriction that the resultant is always exactly equivalent to $1g$ acceleration.[7]

Yaw effects require a separate actuation system but are not regarded as being very important, since they seldom occur except in combination with effects to which the body is very much more sensitive. Hence it is found that very faithful simulation may be achieved by the use of pitch, roll and heave motions only, in what is called a three-axis system.

There are two basic ways for providing the necessary actuation. In the first of these the motions are generated separately along or about the appropriate axes, and in the second the resultant effects on the fuselage are resolved into the displacements of three supporting actuators, appropriately positioned. The latter system appears at first sight to be the simpler, the complexity being mainly in the control circuits rather than in the mechanism, but considerable benefit is derived from the fact that in the former system, the jacks are completely non-interacting, and furthermore that the maximum limit of all movements is available simultaneously. This latter point is shown very clearly by the dramatic photographs in Figs. 8 and 9 taken when the Redifon Lufthansa 727 simulator was being used by the Lufthansa crews for training prior to delivery. The system shown uses an ingenious structure of immense strength (Fig. 10) and provides faithful control of movement over an exceptionally wide range of motions and frequencies, as will be obvious from Table 1.

Over 20 simulators have been supplied using this type of system and its antecedents. The heart of the present system is the two-stage servo actuator package visible in the photographs in which the hydraulic ram, control valve, hydraulic snubbers, locking devices, over-run trips and pick-off potentiometers are built into a single unit to which it is only necessary to supply hydraulic pressure and electrical control signals. The flow control valves are high performance devices of basically the same type as are used in the control loading application, but to ease the stabilization of the system, d.c. control signals are used.

With equipments of this type the deadness of the

fixed base simulator is a thing of the past, and though there may occasionally be a misgiving on the absence of yaw motion, up to the present no operator has considered it important enough to warrant the expense.

With the control loading and fuselage systems outlined, environmental simulation for blind flying is essentially complete. The coast is therefore clear for
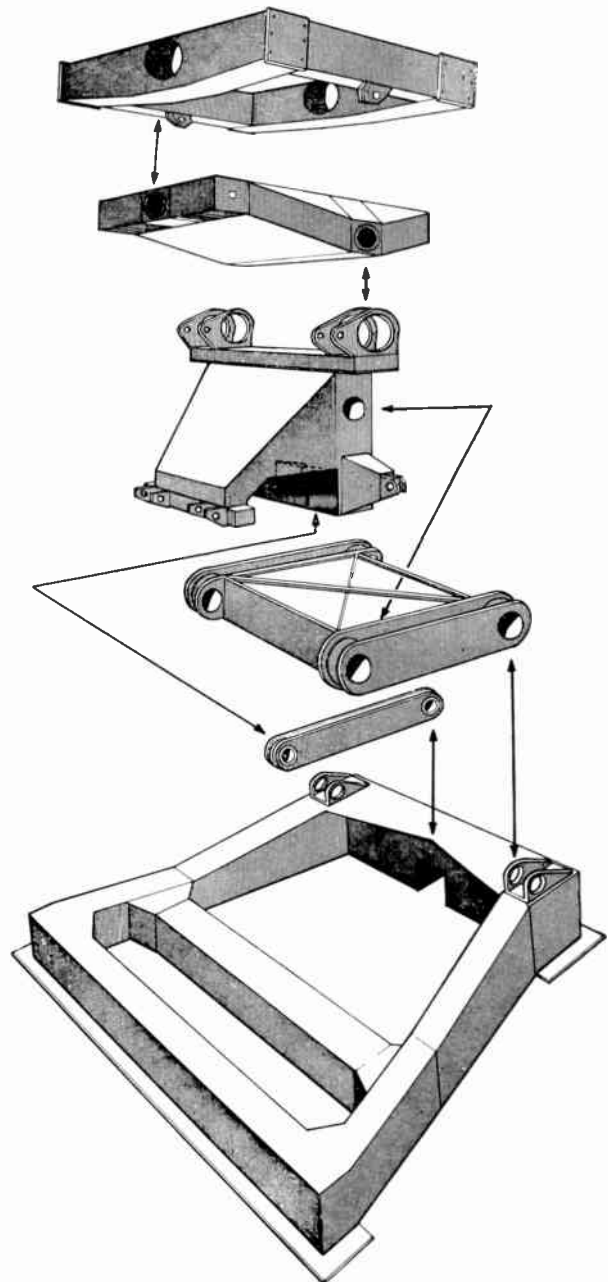

Fig. 10. Exploded view of the 3-axis motion platform mechanical assembly similar to that visible in Fig. 8. The arrows indicate how the assemblies are hinged together.

the next stage of development which is an attack on the operational problems of computer reliability and flexibility, in some cases using the digital equipment.

On the aircraft research side it should be mentioned that a very comprehensive five-axis system, leaving out sway, has been manufactured in this country by Redifon and a six-axis system of gargantuan proportions has been acquired by N.A.S.A. in the U.S.A. It is always possible that the programmes based upon these and similar machines may establish that motions additional to those provided by the present three-axis motions are required in some circumstances and if this happens the flight simulator industry, now extremely conscious of the importance of psychological cues, will adapt their designs accordingly.

5. Visual Systems

The foregoing parts of this paper have dealt with the problems of imparting realism to what has been regarded as the traditional role of the flight simulator, blind flying training. The most recent development, the provision of visual displays for low level flying and landing, has now made it possible for the flight simulator to contribute to every phase of aircraft training.

There has always been a requirement for visual displays and there has been a steady flow of equipments offering solutions, many successful in their own field, but none until recently has had the realism which is necessary to ensure good training. Most early equipments, including the well-known systems by Bell in the U.S.A., Giravions Dorand in France and G.P.S. and Shorts in the U.K., were based upon direct optical projection using point light sources and transparent models which were moved to change the position, scale and aspect of the display, and were comparatively inexpensive.

Although visual landing has always created a problem in air crew training the absence of an appropriate training aid in the propeller era was not serious because landing aircraft was much easier. This was because the approach speeds were lower and because the engine controls could be used to correct errors in the approach. The most dangerous error in landing is excessive sink rate leading to undershooting, but if detected soon enough this could fairly easily be checked in propeller driven aircraft by increasing the propeller thrust which gave an increased airflow over the wings and lifted the aircraft out of danger. This does not occur with jet engines and when jets became established there was immediately a need for more careful training. The need was accentuated by the introduction in 1960 of the 'big jets', whose approach to landing has been described as a trajectory rather than a flight path. Basically the requirement in

descent through cloud is to align the aircraft precisely with the i.l.s. glide path beam so that the pilot can go over to visual and land as soon as he breaks out of cloud. With present minimums there are only 15 seconds or so from breaking out of cloud to touchdown and if the approach is not quite right the pilot has to work very fast if he is to recover the situation, and the rapid acquisition of appropriate visual data is essential.

The ideal requirement of a visual system is therefore quite obviously to present a picture to the pilot throughout the approach to landing such as he would see from the actual aircraft in the same situation. This is plainly impossible so that the next question to ask is: what are the essentials of a landing picture? The first of these is instinctive, spatial orientation. Gibson[8] has written extensively on this subject; the picture must be of reasonable extent, lifesize, properly orientated, correct in both horizontal and vertical perspective and having familiar objects immediately recognizable in size, shape, texture, orientation and colour. Spatial orientation can be obtained under significantly less than ideal conditions but the rate of acquisition of data is slower, which of course, with the short time available, can make a crucial difference to the value of the training.[9] An unrealistic presentation may also cause the pilot to form slow habits on the approach and to rearrange adversely his natural processes of perception.

The second major factor is that the pilot must be able to perceive fully the data which are put there to assist him in landing, such as the amber approach lights, the green and red threshold bars, runway red, amber and green lights, the red obstruction lights, blue taxi strip lights and most important of all, the red/white VASI landing aid.

If all these facilities are provided, even though the field of view is relatively narrow, (50 deg horizontal × 40 deg vertical), and the detail is somewhat below the peak resolution of the eye, complete involvement may be obtained and transfer of training efficiency is high—the pilot is to all intents and purposes actually flying.

These conclusions, particularly on the importance of colour, were only reached after a great deal of pioneering investigation by Redifon into the psychology of vision as applied to visual simulation and the synthetic generation of correct spatial perception. In this work valuable assistance was given by the psychologists of the Institute of Aviation Medicine at Farnborough. The conclusions reached were subsequently made the basis of the first full-colour visual system to be developed for flight simulation.

Earlier equipments using television which attempted to meet the requirement for a realistic visual simula-

tion device were made in the U.S.A. where Curtiss-Wright and General Precision (Link) both began development of monochrome closed-circuit television visual systems in about 1957, the former using a three-dimensional model to a scale of 1 : 2400 whilst the latter used a mainly two-dimensional model to a scale of 1 : 600. The latter equipment required a model 75 feet or so in length. Some equipments were delivered but little is known of the extent to which they are being used and it is now understood that both developments have been discontinued. Classified work was also carried out by G.P.S. in the U.K. from about 1955 onwards. The I.A.P. Department at the Royal Aircraft Establishment (J. M. Naish and others)[10] started development of a monochrome system about the same time, in which a television camera takes a picture of an image projected from a transparent aerial photograph on to an intermediate screen. An experimental study of this technique was carried out by Redifon in 1959 and it was on the basis of this study and the investigations mentioned above that the decision was taken that vertical perspective and colour were essential components of the visual scene. An extensive programme of development was initiated to produce such a system. This culminated in a long series of demonstrations in 1962 of the first full colour system ever to be produced. Early attempts to convince the airlines and air forces of the need for colour met with very limited success but the demonstrations proved far better than any verbal arguments could ever do that full colour is indispensable to true simulation. At the time of writing the Redifon system has been successfully installed on the B.O.A.C. *VC10* simulators at Cranebank, one at N.A.S.A. Ames Research Laboratory and one at Western Airlines in the U.S. with several more being prepared for delivery. All other manufacturers offering equipment for visual simulation are now developing colour equipments.

In the meantime, General Precision Systems developed a monochrome system including a continuous-belt-type model and a number of these have been delivered to the British Ministry of Aviation

and one to B.E.A. This equipment has now been superseded by a colour system which they are developing. A number of very simple devices have also been manufactured by the Dalto Corporation in the U.S. but they have not been fitted to simulators.

The problems of applying television to visual simulation stem almost entirely from the optical requirements. A pilot sitting in an aircraft on the ground has his eye about 15–20 feet above the ground and can see the runway from about 100 feet ahead to the horizon. The airspace required for approach and take-off is about 6 miles (36,000 feet). Physical space requirements enforce scales of the order 1 foot = 1000–2000 feet so that the eye height which limits the inlet pupil diameter scales to about 0·1 inches and the closest range visible which limits the focal length to about half an inch. It is not possible to obtain a well resolved television picture with a pupil diameter less than about 0·020 inches because of diffraction effects. Using conventional lens arrangements the depth of focus increases as the aperture is reduced but to meet the full requirement the diameter falls below the diffraction limit and a compromise has to be made. Using inclinable lenses, as in some older bellows cameras, the focus may be made perfect along a plane inclined to the lens axis which permits the use of greater aperture but the correction of lens aberrations becomes more important. To design a lens system which will work over a 60-deg field and from half an inch to infinity in colour is obviously a very considerable problem. G.P.S. use the first of the above systems and Redifon the inclined lens arrangement. In both cases the resulting usable aperture is very small and an immense amount of scene lighting ($\sim$ 2000 ft-candles) must be used to give a satisfactory low-noise picture. Recently it has become possible using Plumbicon tubes to work with lower light levels and still avoid the effects of dark current and shading which make vidicon cameras so difficult to use in low light, particularly in colour applications.

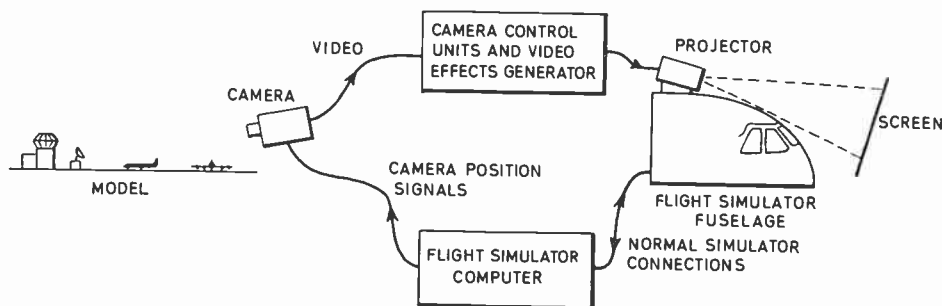A block diagram of the Redifon Full Colour Visual System is shown in Fig. 11 and photographs of the

Fig. 11. Schematic diagram of a flight simulator visual display system using closed circuit television.
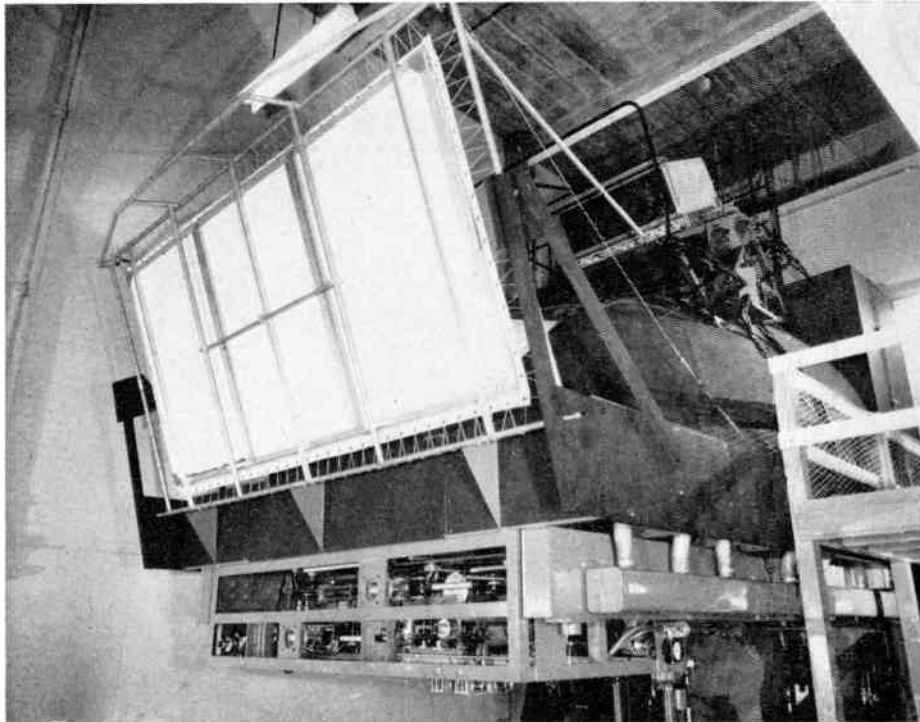
Fig. 12. The Redifon *VC10* simulator colour visual system supplied to B.O.A.C. Exterior view of the flight deck with the colour projector located on top of the cabin and the screen mounted in front.

Fig. 13. The Redifon *VC10* simulator colour visual system supplied to B.O.A.C. Pilot's-eye view from the flight deck during a simulated approach and landing.

VC10 installation at B.O.A.C. in Figs. 12 and 13. The picture is taken from a 1200 : 1 scale, full colour model by an E.M.I. type vidicon colour camera and the display is presented by a Cintel 3-barrel Schmidt projector system.[11]   A picture is presented to the pilot from 1200 feet to touchdown with cloud and visibility effects as required.   Some leading characteristics of the camera, projector and the system are tabulated below.

Although this equipment was a major step forward in the completion of the training environment, it is recognized that it is still far from ideal and second-generation colour equipments are being produced using Plumbicon tubes and improved lens systems. The years to come will see further improvements in the designs of the lens systems used, higher sensitivities and greater linearity in the pick-up tubes, lower noise in the head amplifiers, and a general refinement of performance all round.   At present the greatest single contribution which could be made towards the realism of the visual systems is the provision of a significantly wider picture, wide enough in

Table 2
Characteristics of the Redifon Full Colour Visual System

System	
Scale of model	1 in 1200 or 1 in 1800
Size of model	36 ft long × 10 ft high (bigger models
Area represented	7·2nm × 2nm × 1500 ft or available as
	10·8nm × 3nm × 2250 ft required)
Model characteristics—general	Accurate, 3-dimensional, in full colour with illuminated sky and side mirrors to extend apparent width
—runway	Full colour, threshold bars, runway lighting, red/white VASI, approach lights on poles, taxi lights, etc.
Angular freedom of camera	Roll—continuous
	Pitch—± 20 deg
	Heading—continuous
Ground speed	0–240 knots
Environmental effects	Day/dusk/night operation with cloud base 0–1500 feet and visibility 100 yards to 4 miles
Camera	
Type	Simultaneous, vidicon or plumbicon, red/green/blue
No. of lines per frame	405, 525 or 625
Fields per second	50/60
Interlace	2 : 1
Aspect ratio	4 : 3
Bandwidth,	405 lines, 2·8 Mc/s at 4 dB
each of three channels	525 lines, 5 Mc/s at 8 dB
	625 lines, 5 Mc/s at 8 dB
Projector	
Type	Natural colour simultaneous—Schmidt, red/green/blue
Screen size	6 ft wide × 4 ft 6 in high
Angle subtended at pilot	50 deg × 38 deg
Picture brightness	7 ft lamberts peak
Overall Performance	
Height of pilot's eye above ground	15–1500 feet
Average resolution	On approach 16 elements per degree
	Over threshold 8 elements per degree
	At touchdown 6 elements per degree
Signal/noise ratio	Better than 28 dB

fact to cover the entire field of view from the pilot's position, but this at present would cost more to achieve than it could possibly repay in increased training value.

6. Conclusions

The simulator of today is a highly professional equipment and well adapted to meet the complex needs of aircrew training. Whether the digital computer ultimately displaces its analogue counterpart entirely or whether a compromise is reached, the peripheral systems which have been developed over the last 10 years to perfect the training environment will continue to be used with only such small modifications as are necessary as the computing system changes.

7. Acknowledgment

The author is indebted to the directors of Redifon Ltd. for permission to publish this paper and would like to thank all other members of the organization who, directly or indirectly, have made contributions to it.

8. References

1. K. H. Simpkin and E. T. Emms, 'Flight simulators', *Electronic Engng*, 25, No. 305, p. 270, July 1953.
2. G. B. Ringham and A. E. Cutler, 'Flight simulators', *J. Roy. Aero. Soc.*, 58, No. 519, p. 152, March 1954.
3. A. E. Cutler, 'Electronic flight simulators', *Brit. Commun. Electronics*, 3, No. 2, p. 70, February 1956.
4. A. E. Cutler, 'Analogue computers in aircrew training apparatus', *J. Brit. I.R.E.*, 14, No. 8, p. 351, August 1954.
5. F. J. W. Lager, U.K. Patent No. 768,994.
6. W. O. Brehaus, 'Airborne Simulation for the SST', ASME Second International Simulation Conference, March 1963.
7. A. E. Cutler, U.K. Patent No. 749,048.
8. J. J. Gibson, 'The Perception of the Visual World' (Houghton-Mifflin, Boston, Mass., 1951).
9. W. A. Lybrand and others, 'Simulation of Extra-Cockpit Visual Cues in Contact Flight Transition Trainers', ASTIA Document No. AD 152 123. (February 1958).
10. D. H. Perry, and J. M. Naish, 'Flight simulation for research', *J. Roy. Aero. Soc.*, 68, No. 646, p. 645, October 1964.
11. P. Lowry, 'A colour television projector for medium screen applications', *J. Brit. I.R.E.*, 25, No. 4, p. 305, April 1963.
12. P. M. Carey, 'Visual simulation for aircraft and space flight trainers', *Canadian Proc. I.E.R.E.*, 2, No. 1, pp. 11–19, November 1965. (To be reprinted in *Proceedings of the I.E.R.E.*).

Increased Area Density in Magnetic Recording

By

W. T. FROST †

AND

B. H. SINGER ‡

Based on a paper presented at the International Conference on Magnetic Recording in London in June 1964.

Summary: Reliable digital storage at linear densities as high as 20 000 bits per inch has resulted from an investigation establishing means of providing optimum recording and reproducing resolution. High linear densities make greater lateral densities practical by reducing interference between tracks. High lateral densities have been demonstrated: 1000 tracks per inch with special head structures, and 4000 tracks per inch with selective-erasure techniques. An area density as high as 32 million bits per square inch has been achieved.

These developments remove some of the apparent limits of area density in magnetic recording, and place the emphasis on the necessity for improving reproducing techniques.

1. Introduction

Magnetic recording has helped to make mass information storage possible, and now is called upon to meet demands which tax the most advanced recording techniques and the most highly developed materials. Not only are conventional document files being converted from paper to tape, but new kinds of information are being generated rapidly in great quantity: continuous instrument readings are automatically recorded; computer data are amassed and processed; audio and television programmes are taped for repeated listening and viewing. In such applications, which have developed in parallel with magnetic recording technology over the past 15 years, the goal is ready availability, under control, of all records stored in growing libraries and files. To meet this goal, given the bulk of information now being handled and the level of detail involved, information must be stored as compactly as is compatible with reliable read-out, that is, higher storage densities must be provided.

Storage densities can be described in terms of the dimensions of the recording surface. Linear density represents the number of information-carrying bits (recorded flux changes) per inch length; lateral density —the number of tracks recorded per unit width. Linear storage densities have been increased over the years so that now, as will be shown, 20 000 bits (or recorded flux transitions) per inch can be reliably achieved. Lateral storage densities have also been increased, if somewhat less spectacularly. The

potential for high area density magnetic recording was not realized to any extent until a practical system was developed for television in 1955.[1] Equipment subsequently developed for television can store as many as 750 000 bits per square inch. Computer systems in general demand a higher degree of reliability, which can be provided by present equipment but only at lower storage densities—25 000 to 100 000 bits per square inch.

Limitations in high linear density magnetic recording are set by both the recording and reproducing processes. It will be shown that the reproducing process establishes a limited resolution which is due to physical factors that are difficult to control beyond a certain point. It is necessary, however, to determine the resolution-limiting properties of the recording process so that they can be controlled to provide optimum results.

Increased tracking densities necessitate consideration of head design and novel recording means whereby narrow tracks can be written by wide-track heads. Means of mitigating crosstalk between narrowly separated tracks must be developed.

2. Improving Reproducing Resolution

If a step change in magnetization is recorded perfectly on an infinitesimally thin medium of unit permeability and reproduced with a head of infinite permeability, the output is, as Hoagland[2] has shown, a pulse identical in shape with the field configuration of the head when used in a recording mode, both represented as a function of longitudinal distance.

The variation of longitudinal head field, H_x, with distance x at a separation from the head, y, is (from Karlqvist[3])

† Advanced Systems Development Division, International Business Machines Corporation, Los Gatos, California.

‡ Stanford University, Stanford, California.

$$H_x = H_0 \frac{1}{\pi} \left[\tan^{-1} \left(\frac{1 + \dfrac{x}{l/2}}{\dfrac{y}{l/2}} \right) + \tan^{-1} \left(\frac{1 - \dfrac{x}{l/2}}{\dfrac{y}{l/2}} \right) \right]$$
$$......(1)$$

where H_0 is the maximum field at the centre of the head gap of length l.

Figure 1 shows how the reproduced amplitude varies with the linear density of an ideally recorded pattern of alternate flux changes. These results were obtained with various head-to-medium separations by superimposing pulses calculated from eqn. (1).

It can be seen from these results that a reproducing head whose gap length is 40 microinches resolves 40 000 bits per inch with a relative amplitude loss of 6 dB at a head-to-medium separation of 5 micro-inches. At a separation of 40 microinches, the same head resolves only 16 500 bits per inch with a relative amplitude loss of 6 dB. At this density, the loss relative to the amplitude obtained with a separation of 5 microinches is 15 dB.

Heads are presently available having gaps as small as 40 microinches and finishes of 5 microinches or less on the pole surfaces. Such dimensions have been achieved in various ways—by making heads of dense ferrite with gaps defined by glass-bonding processes[4] or by facing a ferrite core with a thin lamination of a hard-wearing magnetic alloy such as Alfenol which contains a gap formed by deposition of a non-magnetic material such as silicon monoxide.[5] Comparable surface finishes (on the order of 5 microinches) can also be obtained on magnetic oxide coated media and thin metallic magnetic films.
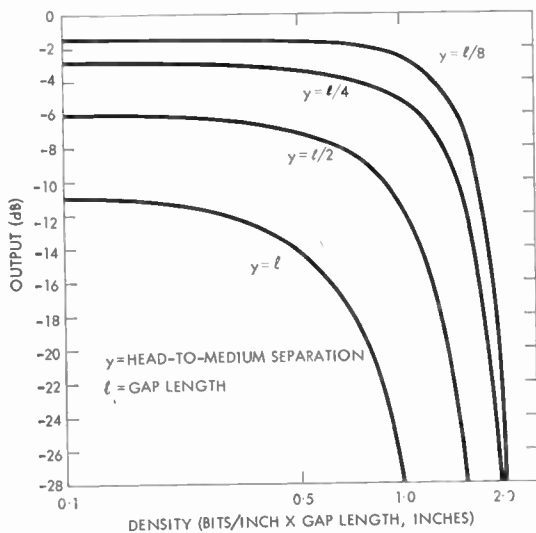

Fig. 1. Reproducing resolution.

The response of such narrow-gapped reproducing heads under practical dynamic conditions might correspond to that shown in Fig. 1 when $y = l/2$. Reliable resolution at densities above 20 000 bits per inch would depend, therefore, on maintaining a smaller separation between head and medium, a difficult task to achieve.

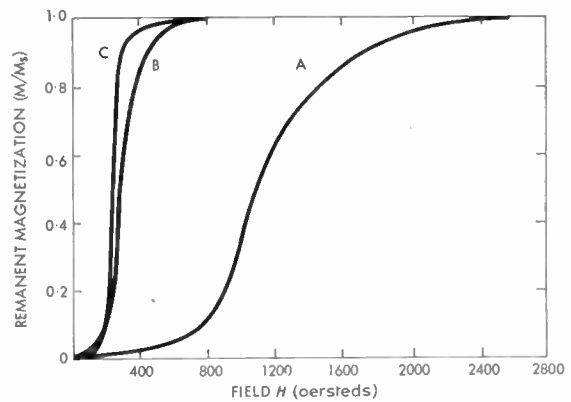| Medium	Thickness	H_r (oersted)	M_s (e.m.u./cm^3)
A	$< 10\ \mu$in	1100	550
B	0·4 mil	280	80
C	$< 10\ \mu$in	240	725

Fig. 2. Magnetization characteristics of three recording media.

3. Improving Recording Resolution

If an ideal reproducing head reads a recorded flux transition, the response (when no demagnetization is present) can be represented by

$$E \propto \frac{\partial M_x}{\partial x} \qquad ......(2)$$

where M_x is the magnetization in the longitudinal direction, x, of the medium. This can be written:

$$E \propto \frac{\partial M_x}{\partial H} \cdot \frac{\partial H}{\partial x} \qquad ......(3)$$

where H is the longitudinal field from the recording head gap as given by eqn. (1).

Three samples of recording media were chosen having widely differing properties. Samples A and C are metal films of cobalt-phosphorous produced by chemical deposition; sample B is a gamma ferric-oxide-coated tape. Various properties and the measured remanent magnetization characteristics of each are shown in Fig. 2. The magnetization characteritsics of sample A can be approximated by

$$\frac{M}{M_s} = 0.318 \tan^{-1} \left[4 \left(\frac{H}{H_r} - 1 \right) \right] + 0.5 \ ......(4)$$

where H_r is the remanent coercivity and M_s is the saturation magnetization of the medium.

If eqn. (1) is differentiated with respect to x, and eqn. (4) with respect to H, expressions for $\partial H/\partial x$ and $\partial M/\partial H$ are obtained. Then if (1) is substituted in the expressions for $\partial M/\partial H$ when $y = l/8$, eqn. (3) can be written as

$$E \propto M_s \frac{H_0}{H_r} \left[\frac{1}{1+16\left(1+\dfrac{x}{l/2}\right)^2} - \frac{1}{1+16\left(1-\dfrac{x}{l/2}\right)^2} \right] \times$$

$$\times \left[\frac{1}{1+\left\{4\left[\dfrac{H_0}{H_r}\cdot\dfrac{1}{\pi}\left(\tan^{-1}4\left(1+\dfrac{x}{l/2}\right)+ \right.\right.\right.}{ \left.\left.\left. +\tan^{-1}4\left(1-\dfrac{x}{l/2}\right)\right)-1\right]\right\}^2} \right]$$

......(5)

Output pulse widths at half-amplitude points and maximum amplitudes, computed from this expression, are plotted in Fig. 3 with respect to H_0/H_r.
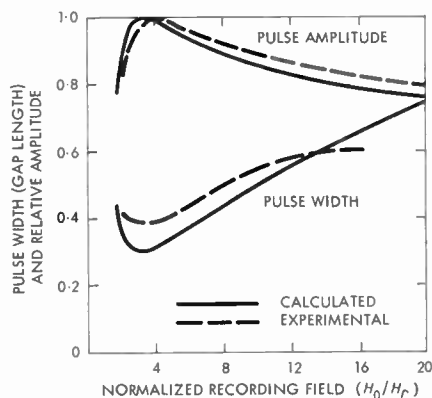
Fig. 3. Recording resolution.

Experiments using the high coercivity sample (A) showed how the resolution varies with changing recording head current. The medium was initially in a saturated remanent condition and flux transitions were recorded with the medium stationary. The gap length in the recording head was 0·001 inch. When the recorded medium was moved at constant velocity past a reproducing head (gap length 40 micro-inches), the width and amplitudes of the resolved pulses changed with the measured recording-head current as shown in Fig. 3. The difference between the calculated and experimental results is due primarily to the weighting function of the reproducing process.

Density responses of the samples (Fig. 4) show that the responses of samples A and B are similar, although these media differ substantially in thickness
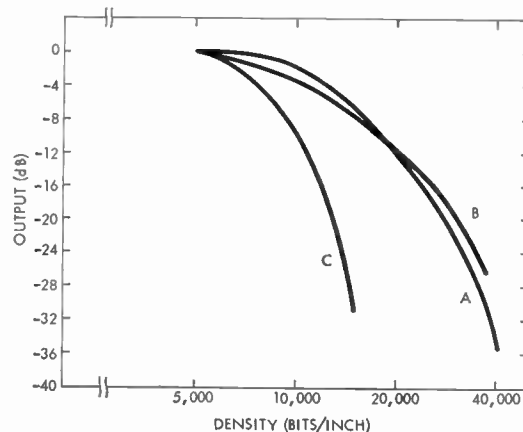
Fig. 4. Density response of three recording media.

and magnetic characteristics. The narrow-gapped recording head used with B, the thick sample, restricts the recording to the surface, which minimizes demagnetization effects. The responses of samples A and B might be explained by the existence of an effective separation, small yet significant, between the head-gap and tape surface. The relatively poor response of sample C is due to the relatively high ratio of saturation remanence to coercivity, which would make the medium prone to demagnetization.

4. Reliable High Density Recording

Experiments have been performed to show the feasibility of a serial high density magnetic recording channel of high reliability. The experimentally produced medium, A, was transported by a simple tape-loop device at 15 inches per second past a write–read head having a 40-microinch gap. The reliability of the system, shown diagrammatically in Fig. 5, was evaluated by phase-encoding a digital pattern and recording a predetermined record length on the tape loop. During the read process, the data from the tape were synchronously detected[6] and arranged to be in phase with that from the pattern generator. The two wave trains were compared and any difference
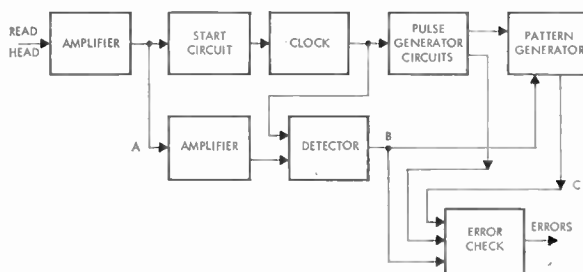
Fig. 5. Detection and error checking at 10 000 and 20 000 bits/inch.

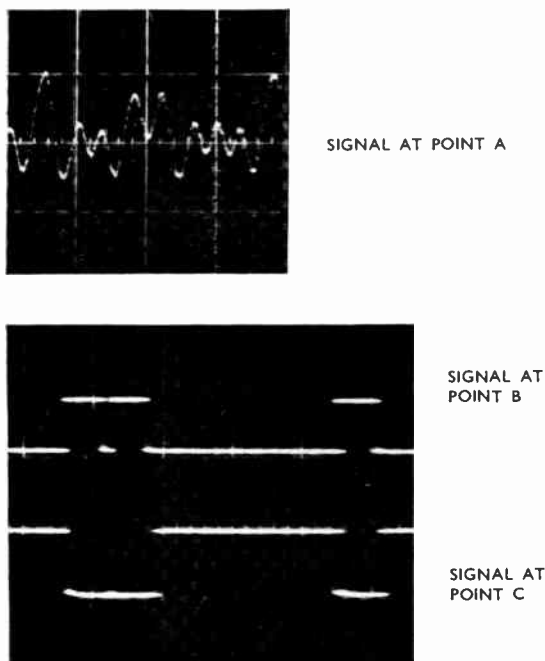SIGNAL AT POINT A

SIGNAL AT POINT B

SIGNAL AT POINT C

Fig. 6. Signal waveforms in detection and error checking.

between them resulted in the generation of an error pulse. Signal waveforms, taken at specified points, are also shown in Fig. 6. Operating at a maximum density of 10 000 bits per inch (5000 information bits per inch), $5 \cdot 8 \times 10^9$ data bits were recovered, and at 20 000 bits per inch (10 000 information bits per inch) $0 \cdot 5 \times 10^9$ data bits were recovered without permanent errors occurring in either case.

5. High Tracking Density

In the past, increased lateral density in magnetic recording has been limited by the apparent difficulties in reducing the head tracking width much below 0·005 inch. A method has been developed[7] whereby a recording head of normal tracking width—say 0·050 inch—can record extremely narrow tracks. When the head completes one writing pass, it is shifted in a direction perpendicular to the track, and erases the track just recorded except for the narrow region remaining as a result of the shift. The head then moves to a new track position and repeats the process. In this way, it has been found possible to record tracks as narrow as 125 microinches. Information thus recorded cannot, at present be read out, by conventional means, but the recorded pattern can be made visible by Bitter techniques.[8] Each of the 125-microinch tracks displayed in Fig. 7(a) contains 8000 bits per linear inch—an area density of 32 million bits per square inch. To eliminate the separate erasure pass, the signal can erase the excess width of one track as it records the next. Such an overlay process produced the recording shown in Fig. 7(b).

To read and write narrow tracks directly, special heads have been fabricated. One of these heads recorded the 0·8-mil track shown in Fig. 8. Here, the resolved bit width (approximately 0·1 mil) establishes a potential linear recording density of at least 10 000 bits per inch at a tracking of 1000 per inch.

6. Some System Implications of High Area Density Recording

One major factor limiting the use of narrow-track recording is crosstalk between tracks placed very close together. Such interference is most troublesome at long wavelengths, i.e. low densities.[9] One way to reduce these effects is to make use of the available very high linear densities to employ modulation techniques so that recording takes place in a restricted range of high densities. As an example, in an analogue system designed to operate between 50 c/s and 5 kc/s at a tape speed of 1·25 inches per second, a tracking density of 30 per inch would be the maximum allowable in order to keep crosstalk at the low frequencies to an acceptably low level. With modulation techniques, an operating spectrum around 20 kc/s, at a tape speed of 5 inches per second, could restrict the

0·000125 in.

Fig. 7(a). Recording by selective erasure: 32 000 000 bits/square inch.
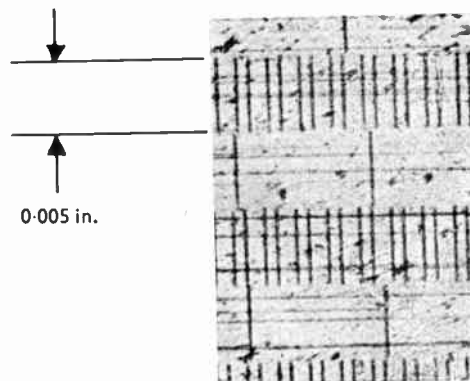

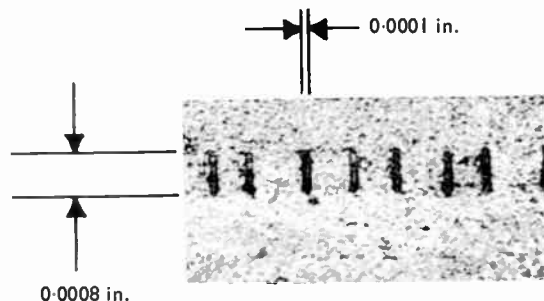0·005 in.

Fig. 7(b). Recording by overlay process.

Fig. 8. Direct narrow-track recording.

density range to be high enough to allow 1000 tracks per inch with negligible crosstalk. The result is a net increase of $8\frac{1}{3}$ to 1 in area density.

With the selective erasure and overlay methods of recording narrow tracks, new concepts in magnetic recording can be considered. Very high area density recordings could be developed by high-resolution Bitter techniques for display, bringing magnetic recording somewhat into line with thermoplastic and photographic recording, where exposure is followed by development and display.

7. Conclusions

The continuing need to increase the area density capabilities of magnetic recording systems is satisfied by increasing both tracking and linear densities. Narrow tracks can operate close together without interference only if the information within the tracks is confined to a limited range of high linear densities. It is doubly important, therefore, to increase the linear density of magnetic recording systems. Practical limitations in the reproducing process appear to restrict reliable operation to linear densities up to about 20 000 bits per inch.

The analysis and experimental results presented show that where demagnetization is negligible, high density magnetic recording depends on the inter-action of the recording head field and the magnetization characteristic of the medium. A thick medium can be used successfully with a narrow-gapped recording head; such a gap restricts the recording to a thin surface layer and demagnetization is minimized.

As has been shown, reliable operation at linear densities as high as 20 000 bits per inch is feasible and a significant increase in tracking densities is possible.

Very narrow tracks can be recorded by selective erasure techniques, and resolved, at present, by developing the magnetically recorded information for visual display. The possibility of combining the versatile high-resolution reading properties of optics in two dimensions with high area density magnetic recording is of obvious future interest.

8. Acknowledgments

9. References

1. C. P. Ginsberg, 'A new magnetic video recording system', *J. Soc. Motion Pict. Television Engrs*, **65**, pp. 302–4, May 1956.

2. A. S. Hoagland, 'Magnetic data recording theory: head design', *Trans. Amer. Inst. Elect. Engrs*, I, **75**, pp. 506–12, 1956; *Communication and Electronics*, No. 27, November 1956.

3. O. Karlqvist, 'Calculation of the magnetic field in the ferro-magnetic layer of a magnetic drum', *Trans. Roy. Inst. Tech.*, Stockholm, No. 86, pp. 3–27, 1954.

4. S. Duinker, 'Durable high resolution ferrite transducer heads employing bonding glass spacers', *Philips Res. Repts*, **15**, No. 4, pp. 342–367, August 1960.

5. O. Kornei, 'Magnetic transducer head for high-frequency signals', U.S. Patent No. 2,711,945, 1955.

6. E. Hopner, 'An experimental modulation-demodulation scheme for high-speed data transmission', *IBM J. Research and Development*, **3**, No. 1, pp. 74–85, January 1959.

7. B. I. Bertelsen and H. J. Kump, 'Magnetic recording', U.S. Patent No. 3,058,112.

8. F. Bitter, 'On inhomogeneities in the magnetization of ferromagnetic materials', *Phys. Rev.*, **38**, pp. 1903–5, November 1931 (letter).

9. D. G. Eldridge and A. Baaba, 'The effects of track width in magnetic recording', *I.R.E. Internat. Conv. Rec.*, **8**, Pt. 9, pp. 145–155, 1960.

p-i-n Diode Modulators for the K and Q Frequency Bands

By

L. J. T. HINTON, B.Sc.(Eng.)

A.C.G.I.†

AND

L. F. BURRY †

Summary: An experimental investigation into extending the operation of p-i-n diode modulators from 18 Gc/s to 40 Gc/s is described. Wafers of stacked p-i-n diodes are considered but most success is achieved with arrays of unencapsulated post-mounted diodes. By reducing the dimensions of diodes of this type used at lower frequencies, broad-band modulators can be designed to operate at frequencies up to 40 Gc/s. A minimum on/off ratio of about 10 dB and an insertion loss of < 2 dB can be obtained over a 1·5 to 1 band of frequencies.

1. Introduction

It has been shown[1,2,3] that modulators using two or more p-i-n diodes suitably spaced as shunt elements along the gap of a ridge waveguide can be designed to meet broad bandwidth requirements at frequencies up to 18 Gc/s. Although no encapsulation of the diodes is used, there are parasitic reactances associated with a diode in its microwave circuit which increase in their importance at higher frequencies. Ultimately, it has been suggested that these reactances will determine the practical upper limit to frequency of operation.

However, it is not yet clear where this upper limit might be, and the purpose of this paper is to describe an experimental investigation into extending operation from 18 Gc/s to 40 Gc/s. Designs of diode switch modulator have been attempted using attenuator elements formed from relatively large wafers of stacked p-i-n diodes, but most success has been achieved with arrays of very small p-i-n diodes mounted on metal posts. To make such arrays feasible at the highest frequencies considered, both the silicon slices and their mounting posts have to be considerably smaller in diameter than is necessary for lower frequency applications.

For switch modulation it was the aim to obtain a minimum on/off ratio of about 10 dB with an insertion loss of not greater than 2 dB over the frequency bands 18 to 26 Gc/s (K-band), and 26 to 40 Gc/s (Q-band). It was desirable to have a reasonable match under both switch conditions.

† E.M.I. Electronics Ltd., Hayes, Middlesex.

2. Design Considerations

If the array of diodes is considered as a network of shunt resistances, it has been shown[3] that, for a broad band performance, the diodes should be spaced a quarter of a wavelength apart at mid-band. This assumes that parasitic reactances associated with a diode in its microwave circuit may be neglected. However at Q-band the diode spacing will be only about 0·1 in, and the mounting posts for the silicon slices used at J-band and lower frequencies are too large in diameter for this requirement to be met. Consequently a different approach to the design of a modulator was initially considered. By placing a suitably shaped wafer of stacked p-i-n diodes along the gap of a ridge waveguide and making d.c. connections, attenuation that is a function of bias current should be obtained (Fig. 1). The heavily doped regions of the silicon and the metallic boundaries between the diodes will be normal to the electric field and were not expected to present a very significant discontinuity.
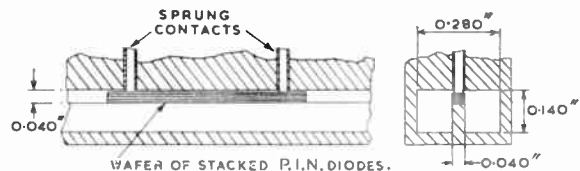

Fig. 1. Diode wafer in ridge waveguide.

Wafers of diodes 0·040 in square in section and varying from ½ in to 1 in long were tested at Q-band, and it was found that the attenuation could be varied

with bias in the expected manner. However there was a drawback in that the insertion loss was at least 6 dB in the zero bias condition. Also the method of applying the bias was not very satisfactory due to the d.c. connections being in the microwave field. Figure 2 shows an alternative arrangement in which the semiconductor material is placed against the narrow wall of a rectangular waveguide thus allowing low impedance decoupling circuits to be incorporated. Unfortunately, if the insertion loss was kept below 2 dB, the variation of attenuation with bias was too small to be of much interest.

The disappointing results with wafers of p-i-n diodes are probably due to the high dielectric loss of the intrinsic regions. It has been suggested that the loss tangent of the intrinsic silicon may be 0·1 or greater. If this is so, then the results obtained are not unexpected.

The possibility of reducing the dimensions of postmounted diodes was next considered. In addition to reducing the diameter of the metal post itself, the parasitic reactances associated with lower frequency diode configurations have to be reduced. With the VX4190, for example, the p^+-n junction is·formed by alloying a pellet of aluminium to a slice of intrinsic silicon, or more correctly high resistivity n-type silicon, and a hemispherical pip of about 0·010 in diameter remains at the end of the manufacturing process. To meet the requirement, the manufacturers changed the alloying method so as to dispense with this pip. The diodes which were supplied for experimental purposes comprised silicon slices about 0·020 in diameter by 0·008 in thick mounted on metal posts 0·020 in diameter.
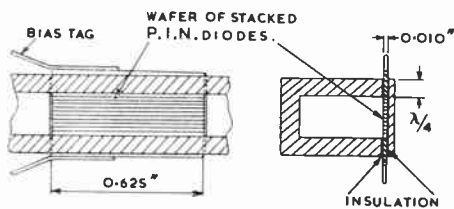

Fig. 2.   Diode wafer in W.G.22 waveguide.

Based on these diodes, ridge waveguide mounts were designed for K-band and for Q-band. Each mount comprised an array of five diodes spaced a quarter of a wavelength apart at mid-band. The dimensions of the ridge were chosen to give an impedance of about 50 ohms and care was taken to avoid supporting higher order modes in the guide. High-low impedance decoupling circuits were incorporated for each shunt element. Details of a Q-band mount are shown in Fig. 3.
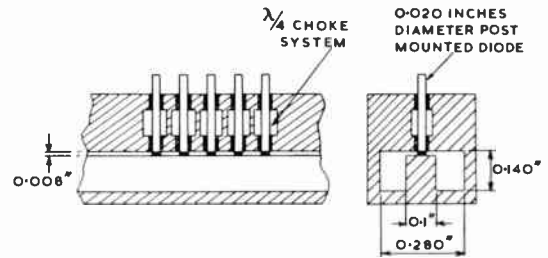

Fig. 3.   Five-diode Q-band ridge waveguide mount.

3. Experimental Work

Figure 4 shows the insertion loss and v.s.w.r. of a K-band mount as a function of frequency for different forward bias current levels. Work at lower frequencies has shown that an improvement in v.s.w.r. may be obtained with higher impedance end elements in the attenuating array, but for this experimental investigation tapering of bias currents was not employed. Switching between zero bias and 2 mA, an on/off ratio of 8 dB is obtained at 18 Gc/s which increases to 21 dB at 26 Gc/s.

Figure 5 shows the same characteristics for a similar Q-band mount. The outstanding feature is the appearance of a low Q resonance which produces a marked peak in zero bias loss at about 38 Gc/s. The resonance still appears to be present when forward bias is applied, the resonant frequency decreasing as the current increases. It was found that changing the diode elements did not alter the characteristics materially. The resonance is also observable with a single shunt diode element.

The similarity of the mount response to that of an m-derived low-pass filter was noted, maximum loss occurring at a frequency at which the shunt arms of the filter sections are resonant. However, an absorbent loss mechanism has to be postulated to account for the fact that the resonance is not accompanied by a large increase in v.s.w.r.

Although experiments showed fairly conclusively that power was not being absorbed in the decoupling circuits of the Q-band mount, it was decided to remove any doubt about this by confining investigation of the resonance to the zero bias condition. These circuits are not then required if unmounted silicon slices are used. In the first experiment, five p-i-n slices, 0·020 in square by 0·008 in thick were clamped in the gap of a ridge waveguide a quarter of a wavelength apart. These slices had to be selected for equality in thickness, and a thin layer of tin foil on the ridge of the guide was necessary to allow for residual tolerances when the mount was assembled. Figure 6(a) shows the insertion loss and power absorption of the array of diodes as a function of frequency. It is seen that the performance
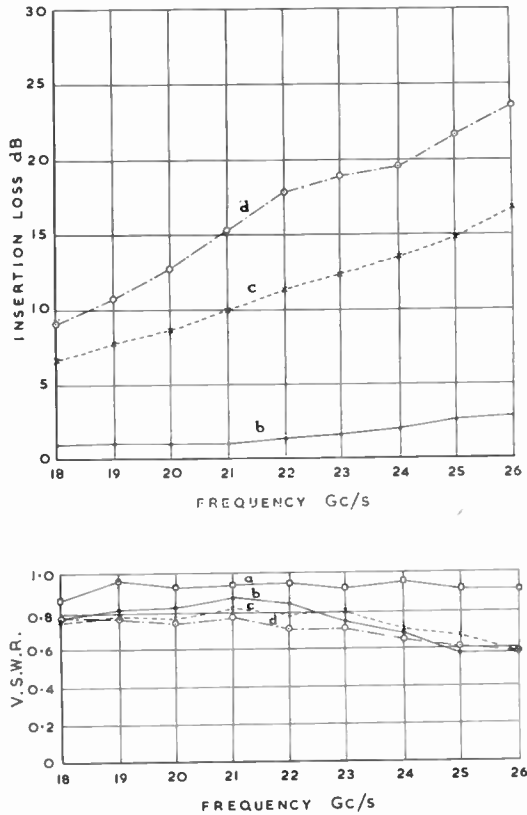
Fig. 4. Loss and v.s.w.r. as a function of frequency for a K-band mount comprising an array of 5 diodes 0·020 in diameter.

a Mount alone. c 1 mA per diode.
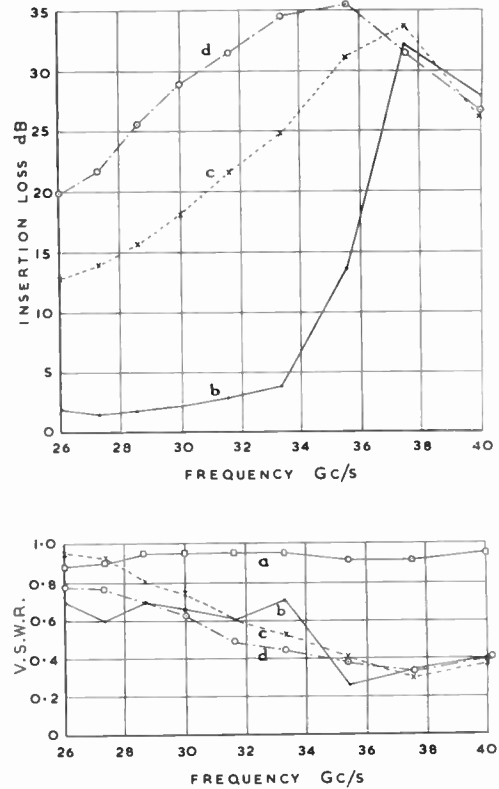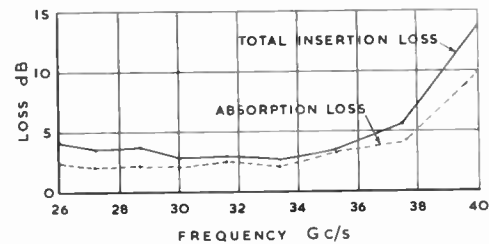b Zero bias. d 2 mA per diode.

Fig. 5. Loss and v.s.w.r. as a function of frequency for a Q-band mount comprising an array of 5 diodes 0·020 in diameter.

a Mount alone. c 1 mA per diode.
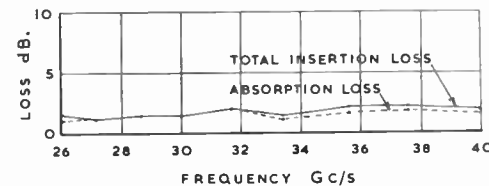b Zero bias. d 2 mA per diode.

is similar to that of the Q-band modulator in that there is a rapid increase in loss at the high frequency end of the band, indicating the presence of a resonance.

The p-i-n slices were next replaced with slices of intrinsic silicon of the same dimensions, and Fig. 6(b) shows that the loss is less and that it is uniform over the frequency band. Thus it was deduced that the resonant condition must be associated with the doping of the silicon. As Fig. 5 shows that the resonant frequency decreases with increasing bias current, there is a basis for suggesting that the resonant condition may be due to the depletion layer capacitance at the p^+-n junction going into series resonance with an inductance.

Measurements of depletion layer capacitance at zero bias indicate that for 0·020 in slices this capacitance has a value of about 2 pF. By taking a slice of intrinsic silicon with nickel-plated surfaces, a parallel-plate capacitance of 2 pF can be placed in series with the silicon slice to form a model of an unbiased p-i-n diode as shown in Fig. 7(a). The variation of loss with frequency of an array of five of these elements spaced

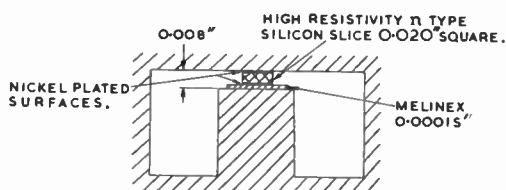(a) Loss of an array of 5 unmounted p-i-n 0·020 in squares as a function of frequency.

(b) Loss of an array of 5 unmounted undoped silicon 0·020 in squares as a function of frequency.
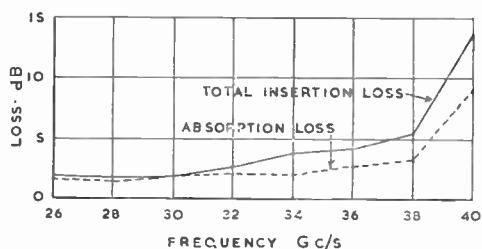
Fig. 6.

a quarter wavelength apart in ridge waveguide was measured, and Fig. 7(b) shows the close similarity between this characteristic and that for the array of p-i-n diodes (Fig. 6(a)). When the series capacitance is reduced by a factor of about ten, the resonance is moved well beyond 40 Gc/s, and a flat characteristic is obtained over Q-band (Fig. 7(c)).
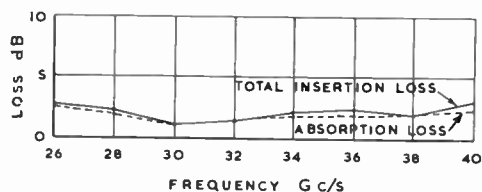
Although this model of the p-i-n diode has limitations, the results indicated there might be an advantage in reducing the diode depletion layer capacitance. Consequently some diodes were reduced in diameter from 0·020 in to about 0·010 in and then mounted on metal posts 0·020 in diameter. Figure 8 shows the insertion loss and v.s.w.r. as a function of frequency for an array of five diodes. The increase in zero bias loss at the high frequencies still shows the presence of a resonance, but the resonant frequency is now beyond 40 Gc/s. Switching action can be obtained over the whole frequency band, but the zero bias loss at the higher frequencies restricts its usefulness to the low frequency part of the band.

(a) Simulated diode configuration.

(b) Loss of an array of 5 simulated diodes as a function of frequency (capacitance of about 2 pF).

(c) Loss of an array of 5 simulated diodes as a function of frequency (capacitance of about 0·2 pF).
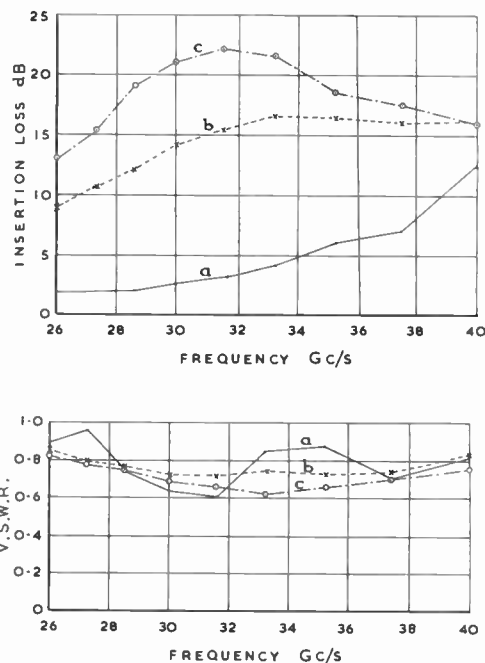
Fig. 7.

Fig. 8. Loss and v.s.w.r. as a function of frequency for a Q-band mount comprising an array of 5 diodes 0·010 in diameter.

a Zero bias.   b 1 mA per diode.

c 2 mA per diode.

4. Conclusions

Using 0·020 in p-i-n diodes mounted on metal posts of the same diameter, it may be concluded that it is quite practical to design a switch modulator to operate over the frequency band 18 to 26 Gc/s (K-band). The minimum on/off ratio of 8 dB obtained with an experimental mount may be increased with a higher forward bias. The theoretical limit, assuming good thermal contact over the whole cross-section of the diode, is in the region of 100 mA. A disadvantage of increasing the bias current is that it degrades the input v.s.w.r., but this may be overcome by tapering the impedances of the end shunt elements as at lower frequencies.

For the frequency band 26 to 40 Gc/s (Q-band), there has been only partial success, so far, in establishing the feasibility of a design to cover the whole band. The experimental investigation of the resonant condition suggests a series resonance between the depletion layer capacitance and an inductance, accompanied by an absorbent loss mechanism. As the intrinsic region of the p-i-n diode has a finite resistivity even under the zero bias condition, it may be possible to associate this inductance with the silicon slice. In this connection, a theoretical analysis of an obstacle comprising a lossy dielectric circular post of variable height in ridge or rectangular waveguide might be rewarding.

On the other hand, the fact that the resonant condition is not associated with a large increase in reflection loss supports the view that it is a more complex mechanism of power absorption in the doped crystal lattice.

The possibility of designing a broad band microwave circuit to reduce the series capacitance of the shunt elements of the p-i-n diode modulator is attractive and the study of the design of a suitable circuit is being undertaken.

5. Acknowledgments

This paper, which is published by the permission of the Ministry of Aviation and the Directors of E.M.I. Electronics Ltd., describes work which was part of a programme undertaken on behalf of the Radio Department of the Royal Aircraft Establishment. The authors gratefully acknowledge the co-operation of A.E.I. Ltd., Rugby, in supplying the experimental p-i-n diodes.

6. References

1. T. H. B. Baker, 'Junction-diode microwave modulators', I.E.E. Conference on Components for Microwave Circuits, September 1962. Conference Digest, p. 40.

2. N. A. D. Pavey, 'Broad-band p-i-n diode switches for 8–16 Gc/s range', Proceedings of Symposium on Microwave Applications of Semiconductors, London, June 1965. Paper No. 6 (I.E.R.E. Conference Proceedings No. 5).

3. J. K. Hunton and A. G. Ryals, 'Microwave variable attenuators and modulators using p-i-n diodes', *Trans. Inst. Radio Engrs on Microwave Theory and Techniques*, MTT–10, pp. 262–73, July 1962.

7. Appendix

Subsequent to the preparation of this paper, the authors have found that, by modifying the microwave circuit of the Q-band modulator, a uniform performance can now be obtained over the whole frequency
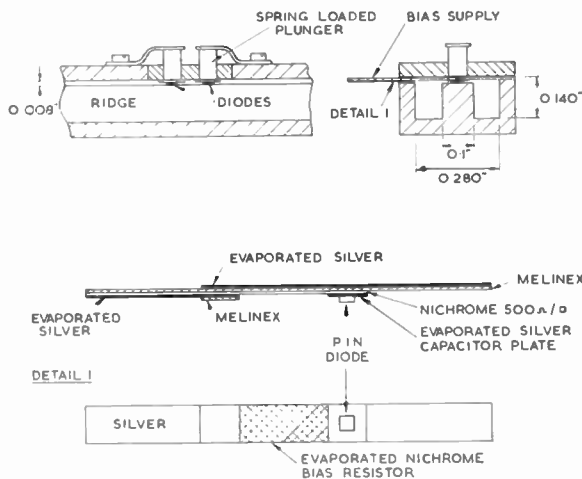

Fig. 9. Two-diode Q-band ridge waveguide mount.

band 26 Gc/s to 40 Gc/s using p-i-n diodes 0·020 in square. The experiment with unmounted slices of intrinsic silicon (Section 3) suggested that more success in avoiding a resonant condition in this frequency range might be achieved if a lumped capacitance could be placed in each shunt arm of the attenuating array of diodes. The difficulty is to form capacitances of about the right value and at the same time make the

Fig. 10. Loss and v.s.w.r. as a function of frequency for a Q-band modulator comprising two p-i-n diodes 0·020 in square

a Zero bias.

b 20 mA total bias.

necessary bias connections to the diodes. Figure 9 shows an experimental arrangement which has been found to be satisfactory. A parallel-plate capacitance of about 0·8 pF is formed in series with the diode with a high resistance nichrome film on the Melinex insulation to determine the capacitance plate area and to act as the control resistor for the bias supply.

Figure 10 shows the insertion loss and v.s.w.r. as functions of frequency for a two diode mount. An on/off ratio of greater than 10 dB with an insertion loss of 1 to 2 dB is obtained over the whole 1·5 to 1 frequency band. The satisfactory uniformity of the performance is most marked and clearly indicates that very wide band designs of p-i-n diode modulator up to at least 40 Gc/s are possible.

Design of a 4 Gc/s Nitrogen-cooled Non-degenerate Parametric Amplifier

By

D. CHAKRABORTY, B.Sc.,†

G. F. D. MILLWARD, B.Sc.(Eng.)†

AND

D. GEDEN, B.Sc.†

Summary: The design of a 4 Gc/s nitrogen-cooled non-degenerate parametric amplifier suitable for installation at the focus of a dish aerial in a communication-satellite-system earth-station is described. Two single-diode amplifiers, each of 15 dB gain and 70 Mc/s bandwidth to the − 3 dB points, are cascaded to provide a minimum gain of 30 dB and a − 3 dB staggered tuned bandwidth of 60 Mc/s with 1 dB ripple in the passband. The noise temperature of the amplifier including a cooled circulator when cooled to 77° K is 35 ± 5°K (with contributions from the waveguide connexions and external circuitry increasing the overall noise temperature to 46 ± 5°K).

List of Principal Symbols

a transmission loss ratio

γ coefficient of capacitance modulation
$$= \frac{(C_{1\mu A} - C_{VR})}{2(C_{1\mu A} + C_{VR})}$$

$f_s = \dfrac{\omega_s}{2\pi}$ signal frequency

$f_i = \dfrac{\omega_i}{2\pi}$ idler frequency

g power gain of the amplifier

$Q(0)$ Q-factor of the varactor diode at zero-volt bias $= \dfrac{1}{\omega_s C_J(0) R_s}$

Q_1 Q-factor of the signal circuit $= \dfrac{1}{\omega_s C(R_s + R_g)}$

C mean value of the junction capacitance under pumped condition $= C_0(1 - \gamma^2) \simeq C_0$

C_s stray capacitance

C_0 first term of the Fourier Series representation of the time varying capacitance

R_g source impedance as seen from the diode junction

R_s spreading resistance of the varactor diode junction

R magnitude of the negative resistance as seen by the passive signal circuit

T_d physical temperature of the diode

T_e effective noise temperature of the amplifier

† Post Office Research Station, Dollis Hill, London, N.W.2.

1. Introduction

In a satellite communication system, satisfactory reception requires a ratio of received carrier power to system noise in excess of a characteristic threshold value. The critical value of the threshold depends upon the type of demodulators in use. The signal level received from the *Telstar* and *Relay* type satellites is extremely low, and therefore a very low system noise is essential. For example, at Goonhilly satellite system earth station, the system noise temperature at the zenith was of the order of 70°K during the *Telstar* and *Relay* experiments. This was achieved by using a liquid helium-cooled travelling wave maser[1] installed in a cabin behind the parabolic dish at some distance away from the focus. The additional noise contribution from the interconnecting waveguide was about 45 deg K. Alternatively, a similar performance could be achieved by using an amplifier of higher inherent noise temperature installed closer to the focus of the aerial. The noise temperature of a non-degenerate parametric amplifier cooled to liquid nitrogen temperature is low enough for such an amplifier to be considered for this purpose.[2] The higher noise temperature of this type of amplifier would be acceptable due to the fact that the finite length of the transmission line from the focus would come after the head amplifier and consequently its contribution towards the overall system noise temperature become negligible.

Although it was not practicable to replace the maser for normal operations, an experimental nitrogen-cooled amplifier was made as part of a study of cooled microwave parametric amplifiers. In order to reduce the noise contribution of the second stage, a minimum gain of 30 dB is required and this was achieved by cascading two single-diode parametric amplifiers and cooling them in the same cryostat. A single-stage gain of 15 dB with a 3 dB bandwidth of

70 Mc/s was achieved and the amplifiers were stagger-tuned to give a 3 dB bandwidth of 60 Mc/s with 1 dB ripple in the signal passband. The measured overall noise temperature was 46°K of which 10 deg K was contributed from two four-port waveguide switches, two waveguide corners and a short length of stainless-steel waveguide required for thermal insulation of the amplifier from the ambient temperature waveguide run.

The contribution from the cooled circulator was estimated as about 3 deg K, and that from the second stage about 1 deg K.

The noise temperature of the first stage amplifier alone was thus estimated as 32°K, which is in fair agreement with the theoretical calculation taking into account the noise contributed from the coolable circulator used in conjunction with the amplifier. The design procedure employed is considerably different from that employed by Uenohara et al.[2] in designing a cooled amplifier for use on aerials at Holmdel and Andover.

2. Theoretical Considerations

The basic theory of the gain, bandwidth and noise temperature of a non-degenerate parametric amplifier has been described in a number of publications[3, 4] and therefore only the main formulae will be quoted below.

The power gain[3] of a single-port reflection type non-degenerate parametric amplifier (using a circulator) is:

$$g = \frac{4R_g^2}{\left[(R_g+R_s) - \dfrac{\gamma^2}{\omega_i\omega_s C^2 R_s}\right]^2} \qquad \text{......(1)}$$

and the equivalent circuit representation of this type of amplifier is shown in Fig. 1 where:

R_g is the equivalent source impedance,

R_s is the varactor series (spreading) resistance,

ω_s is the angular signal frequency,

$\omega_i = (\omega_p - \omega_s)$; ω_p is the angular frequency of the pump supply,

$C \simeq C_0(1-\gamma^2)$ and represents the mean value of junction capacitance under pumped condition,

γ is the first-order coefficient of capacitance modulation and is derived from a Fourier Series:

$$C(t) = C_0 + \gamma C_0[\exp(j\omega_p t) + \exp(-j\omega_p t)],$$

C_0 represents the lumped effect due to the static junction capacitance $\simeq C$ for small value of γ.

In eqn. (1), the quantity appearing with a negative sign has the dimensions of resistance and represents the magnitude of the negative-resistance (R), as seen by the passive signal circuit looking towards the passive idling circuit via the varactor diode as the mutual coupling element. Examining eqn. (1) it will be seen that with no pump power, R tends to zero, and with the pump power applied, R tends to a finite value depending on the magnitude of the pump power and the circuit parameters. In the limit, the denominator may tend to zero giving rise to self-oscillation.
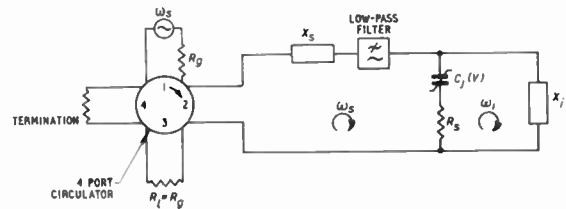

Fig. 1. Simplified circuit representation of a single-port negative-resistance non-degenerate parametric amplifier.

With single-tuned resonant circuits at the signal and the idler frequency, the approximate gain-bandwidth product of a non-degenerate parametric amplifier can be written as follows, using small signal approximations:

$$\sqrt{g} \cdot b \simeq \frac{1}{\dfrac{1}{2}\left(Q_1 + \dfrac{\omega_s}{\omega_i}\dfrac{1}{\gamma^2 Q_1}\right)} \qquad \text{......(2)}$$

where

$$Q_1 = \frac{1}{\omega_s C(R_s + R_g)},$$

b = 3 dB bandwidth normalized with respect to the centre of the signal frequency band.

Differentiating eqn. (2) with respect to Q_1 and equating to zero yields the maximum value of,

$$Q_1 = \frac{1}{\gamma}\sqrt{\frac{\omega_s}{\omega_i}}$$

and therefore,

$$(\sqrt{g} \cdot b)_{\max} = \gamma\sqrt{\frac{\omega_i}{\omega_s}} \qquad \text{......(3)}$$

Equation (3) indicates that the optimum gain-bandwidth product will be obtained when the maximum value for idler resonant frequency has been achieved. The idler circuit can be brought to resonance in two different ways which can be explained by considering the equivalent circuit of a varactor diode as shown in Fig. 2, where $C_J(V)$ is the voltage variable capacitance, L_s is the stray inductance, C_s is the stray capacitance, and R_s is the spreading resistance. If a lossless short circuited line is connected across the terminal BB′ at a distance, $l = \lambda_i/2$, away from the

reference plane of the diode, the idler resonance will be confined across $C_J(V)$, L_s and R_s, and consequently,

$$\omega_i \simeq \frac{1}{\sqrt{L_s C_J(V)}}$$

which we shall define as diode self-resonance. When $l = \lambda_i/4$, the idler resonance will be confined in L_s, $C_J(V)$, R_s and C_s and therefore:

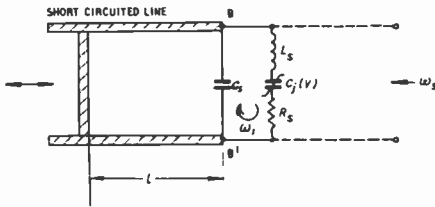$$\omega_i \simeq \frac{1}{\sqrt{\dfrac{L_s C_J(V) C_s}{C_J(V) + C_s}}}$$

Fig. 2. Circuit representation of idler resonance in the varactor diode.

which we shall define as parallel resonance. In a high quality varactor the stray capacitance is normally very small in comparison with the static junction capacitance and the parallel resonance configuration will therefore yield the highest idler resonant frequency. In a typical varactor the measured ratio of these two resonant frequencies has been found to be as high as a factor of two. This suggests that a considerable increase in bandwidth will be achieved by maintaining the idler resonance in parallel configuration when compared with the case of diode self-resonance. The simple theoretical values of bandwidth are not realizable in practice due to storage of signal and idler energy in the additional stray capacitances inevitable in a practical amplifier. However a moderate gain-bandwidth product can be achieved by optimizing the circuit design and using an appropriate isolating filter between the two reactive networks. More elaborate circuits[5] to achieve a very wide bandwidth were not considered for the initial experiment, although these may be necessary in amplifiers for operational use.

The effective noise temperature of a non-degenerate parametric amplifier can be written as:

$$T_e \simeq T_d \left(\frac{R_s}{R_g} + A \frac{\omega_s}{\omega_i} \right) \qquad \dots\dots(4)$$

where T_d is the physical temperature of the diode,

A is a function of the diode and circuit parameters and also depends upon the pump power, and in the limiting case tends to unity.

In eqn. (4) it has been assumed that the signal and the idler circuits are lossless and it is also assumed that the shot noise in the diode is negligible.

3. Design of the Amplifier

The specification for the amplifier was as follows:

Two-stage gain   > 27 dB over the signal pass-band,

Bandwidth (-3 dB) > 40 Mc/s,

Noise temperature $\simeq 40°$K.

With these specifications a diffused-junction mesa-type gallium-arsenide varactor was chosen having the following parameters:

$C_J(0) = 0.5$ pF average,

$C_s \quad = 0.18$ pF average,

$L_s \quad = 0.4$ mμH

$R_s \quad = 2.5$ Ω

$Q(0) = 21$ (measured at $f_s = 4.17$ Gc/s at 290°K)

A value of $\gamma = 0.3$ was assumed to calculate the optimum value of R_g:

$$R_g = R_s \sqrt{1 + [\gamma Q(0)]^2} \simeq 16 \quad \Omega$$

An experimental measurement of $Q(0)$, the diode quality factor,[6] at zero-volt bias and at the signal frequency is desirable as the parameters of individual diodes may differ from the manufacturer's specifications.

From the foregoing theoretical consideration it is apparent that in order to achieve optimum bandwidth and minimum noise temperature the idler circuit should be resonated in parallel configuration. From the diode parameters the calculated value for the idler frequency f_i is given by:

$$\frac{1}{2\pi \sqrt{L_s \dfrac{C_J(0) C_s}{C_J(0) + C_s}}} \simeq 22 \quad \text{Gc/s}$$

Based on these calculations the amplifier was fabricated as shown in Fig. 3. The signal circuit is constructed in 50 Ω coaxial line and the appropriate R_g is obtained by a linearly-tapered transformer. A low-pass filter[7] is incorporated at the end of the signal circuit to isolate the signal and idler circuits. The diode is held by a spring so that the actual junction is in the centre of the reduced-height pump waveguide. The pump power is supplied through a stepped transformer from waveguide WG 20. A step in the wide dimension of the reduced-height guide (not shown) places the idler frequency beyond cut-off in the pump waveguide but transmits the pump power with some additional loss and therefore confines the idler power to the cavity formed between the diode and a moveable non-contact type short-circuit. The

screw-tuner in the signal circuit provided additional flexibility for trimming the signal pass-band. It is imperative to place the circulator as near as possible to the amplifier to optimize the gain-bandwidth and noise temperature. This was possible because a suitable coolable circulator was commercially available.
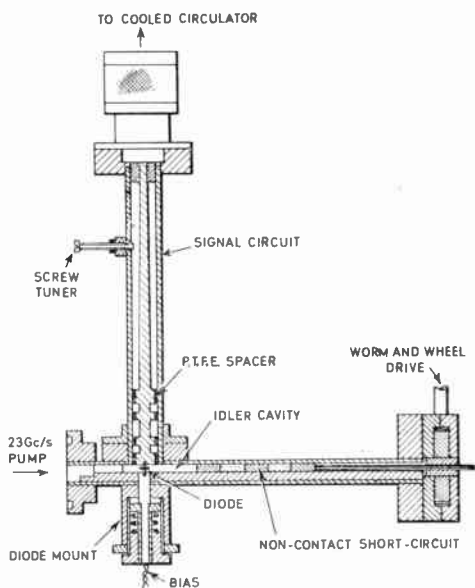

Fig. 3. Sectional view of the parametric amplifier.

Two amplifiers with coolable coaxial circulators were mounted inside a copper canister which was in turn immersed in liquid nitrogen in an all-metal Dewar vessel. Initially, this amplifier canister was continuously evacuated. The shape and size of the Dewar vessel were determined largely by the space available on the Goonhilly aerial structure. This restriction on size and weight precluded the use of a deep Dewar vessel and the resulting compromise design is thermally inefficient, and a single charge of liquid nitrogen therefore only lasts some four hours. For operational use, the amplifier would have to be redesigned with a more efficient cryogenic system. The signal circuit input and output were taken through 'endfire' type waveguide to coaxial transformers and those sections of the signal and the pump wave-guides which were inside the Dewar were made of thin-walled stainless steel to provide thermal insulation.

Although it would have been preferable for the varactor to be used with zero applied bias, it was convenient for experimental purposes to provide additional tuning by arranging an external variable d.c. bias supply.

The measured optimum idler frequency was somewhat lower than the theoretical prediction and two individual K-band klystrons centered around 23 Gc/s (minimum power 40 mW) were used as pump sources. The available ranges of applied bias and pump tuning were found to be inadequate to counteract the change in junction capacitance produced by cooling and hence additional remote tuning of the idler cavities by worm and wheel drive was incorporated. The connecting rods of the tuning mechanisms were taken through sealed holes on the top of the evacuated canister. This considerably increased the leak rate of the canister which was therefore pressurized by dry helium gas at some 2 lb/in^2.

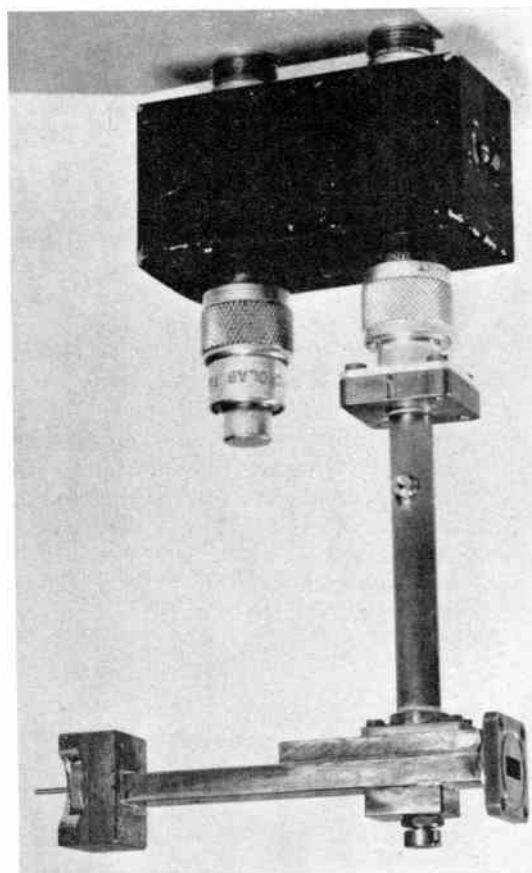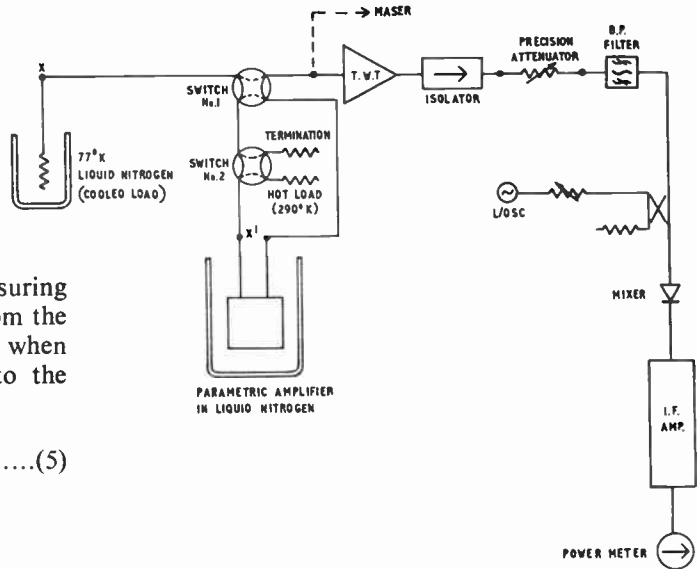Fig. 4. Laboratory model of a single stage of the amplifier.

The measurements described below were all carried out in the laboratory, the measured gain-bandwidth response of a single stage and the cascaded amplifier being shown in Fig. 5.

4. Noise Temperature Measurement

The noise temperature of the amplifier was measured by the standard hot and cold load technique.[8] The schematic arrangement is shown in Fig. 6. The

Fig. 6. Block schematic of the arrangement for noise temperature measurement.

accuracy of the technique is reduced when measuring very low noise temperature. This is evident from the fact that the ratio of the noise power available when the amplifier is switched from hot (290°K) to the cooled (77°K) load is given by:

$$n = \frac{290 + T_e}{\dfrac{77}{a} + \left(1 - \dfrac{1}{a}\right)290 + T_e} \qquad \ldots\ldots(5)$$

where

$T_e = T_1 + \dfrac{T_2}{g_1} + \dfrac{T_3}{g_1 g_2}$ = overall noise temperature of the cascaded amplifier including the noise contribution from the t.w.t.

T_1 = noise temperature of the first cooled-amplifier,

T_2 = noise temperature of the second cooled-amplifier,

T_3 = noise temperature of the t.w.t. amplifier,

g_1 = gain of the first cooled amplifier,

g_2 = gain of the second cooled amplifier,

$\dfrac{T_3}{g_1 g_2} \simeq 3°K$

a = loss ratio between XX'.

The tolerance of the overall measured noise temperature of the amplifier therefore depends upon the accuracy of reading n from the attenuator calibration and evaluating a however small it may be; the

Fig. 5. Gain-bandwidth response of the amplifier.

small gain variation of the amplifier of the order of 0·05 dB makes this error completely random in nature. A set of six measurements of noise ratios n has been performed in quarter of an hour and the average overall noise temperature of the amplifier was calculated to be $46 \pm 5°K$. Most of this estimated error was a statistical variation of the attenuator readings. Out of this total noise temperature 10 deg K was contributed from two four-port waveguide switches, two waveguide corners and a section of about 12 inches of thin-walled stainless steel waveguide WG11 in front of the first amplifier. One of the waveguide switches was required for noise measurement and the second one was intended to by-pass the amplifier to permit normal operation of the maser. The cooled coaxial circulator contributed some 3 deg K and the second cooled amplifier contribution was about 1 deg K. The absolute noise temperature of the amplifier was thus 32°K and the theoretical value based on high-gain approximation (eqn. (4)) was 29°K.

5. Conclusions

A low-noise nitrogen-cooled non-degenerate 4 Gc/s parametric amplifier, giving an overall noise temperature of $46 \pm 5°K$ and a bandwidth of 60 Mc/s, suitable for communication via artificial earth satellites has been described.

6. Acknowledgments

D. CHAKRABORTY, G. F. D. MILLWARD and D. GEDEN

7. References

1. J. C. Walling and F. W. Smith, 'The travelling-wave maser amplifier in the Goonhilly radio station', Paper presented in the International Conference on Satellite Communication, I.E.E. London, November 1962.

2. M. Uenohara, M. Churney, K. M. Eisele, D. C. Hasan and A. L. Stillwell, '4 Gc/s parametric amplifier for satellite communication ground station receiver', Paper presented to the Western Electronic Show and Convention, August 1962. 1962 WESCON Convention Record, Part 1.

3. L. A. Blackwell and K. L. Kotzebue, 'Semiconductor-diode Parametric Amplifiers', pp. 49–66, (Prentice Hall, London, 1961).

4. G. L. Matthaei, 'A study of the optimum design of wideband parametric amplifier and upconverters', *Trans. Inst. Radio Engrs on Microwave Theory and Techniques*, MTT–9, No. 1, pp. 23–28, January 1961.

5. W. J. Getsinger and G. L. Matthaei, 'Some aspects of the design of wideband upconverters and non-degenerate parametric amplifiers', *Trans. Inst. Elect. Electronics Engrs on Microwave Theory and Techniques*, MTT-12, No. 1, pp. 77–87, January 1964.

6. D. Chakraborty and G. F. D. Millward, 'A nitrogen-cooled degenerate parametric amplifier for a radiometric application', *Electronic Engineering*, 37, No. 450, pp. 532–5, August 1965.

7. Radio Research Laboratory Staff, Harvard University, 'V.H.F. Techniques', (McGraw-Hill, New York and London, 1947).

8. E. O. DeGrasse, E. O. Schulz-DuBois and H. E. D. Scovil, 'The three-level solid state travelling-wave maser', *Bell Syst. Tech. J.*, 38, No. 2, pp. 305–34, March 1959.

9. H. N. Daglish and D. Chakraborty, "Applications of parametric amplifiers in satellite communications", Proceedings of Symposium on Microwave Applications of Semiconductors, Paper No. 10 (Institution of Electronic and Radio Engineers, Conference Proceedings No. 5, London, 1965).

Prediction and Engineering Assessment in Early Design

By

W. P. COLE, B.Sc.†

Summary: The paper discusses the steps which could be taken in the early design stages to predict, assess and then verify the reliability of the component parts of the system. Prediction of 'mean time between failures' of electronic equipment is dealt with in particular. The A.G.R.E.E. type of testing carried out by the manufacturer to determine the reliability of his final product is described in some detail in an Appendix.

1. Introduction

It is often said 'Reliability can and must be designed into an equipment from its conception'. This statement implies that it is possible to do something about reliability in the early stages of development. The purpose of this paper is to discuss what can be done and what action must be taken. In particular, it deals with the prediction of the mean time between failures (m.t.b.f.) of electronic equipment and the uses to which this technique may be put. A further section describes how the predicted figures may be verified by 'reliability tests' and discusses, in some detail, the problems involved in fitting such tests into an equipment development programme.

2. The Development Programme

Actions taken at the beginning of the development programme will have their repercussions on the later stages of development, and one of the very first steps that must be taken is to obtain a full understanding of the implications of all the stages in the design of the equipment.

Many will be sure that they already know these but, unless a chart which shows the stages is prepared, it is difficult to appreciate fully the logical flow of information, or to see how the various stages are interlinked.

Figure 1 shows, in diagrammatic form, the stages in the development of a piece of electronic equipment where reliability is of prime importance.

The main 'leg' of this diagram shows the stages in circuit development from the initial requirements through to the final stage where information is available for production. The 'leg' above shows the stages in mechanical design, and the one below that of the

---

† Applied Electronics Laboratories, G.E.C. (Electronics) Ltd., Stanmore, Middlesex.

development of test specifications and, where necessary, the development of special test equipment.

During the course of development it is essential to have positive control of reliability at all stages. As part of this process of reliability control it is desirable to introduce various surveillance, or check, points, at which note can be taken of the progress of the design from a reliability viewpoint. Some of the major check points, reliability assessment and mechanical scrutiny, for example, are shown in Fig. 1. At these check points it is possible to make use of the accumulated experience of the design establishment to ensure that design weaknesses noted in previous projects are not perpetuated.

This diagrammatic programme can be divided into three main stages which from a reliability point of view must, in some ways, be treated separately. At the beginning, the details are less well defined and modifications to improve reliability can be readily introduced. Towards the end, details are firmer and the equipment reliability can be more surely established, but the design is less amenable to modification.

The three main stages are as follows:

System Design

This is defined as being from the conception of the equipment when the requirement is first known, through the early feasibility and design studies and system thinking, up to the time when a system specification is evolved and a breadboard model is produced.

Development Models (D models)

By the time stage 1 is completed, the engineering phase has already started and the programme is proceeding along three lines, mechanical design, circuit design and test procedures, which have been previously defined.
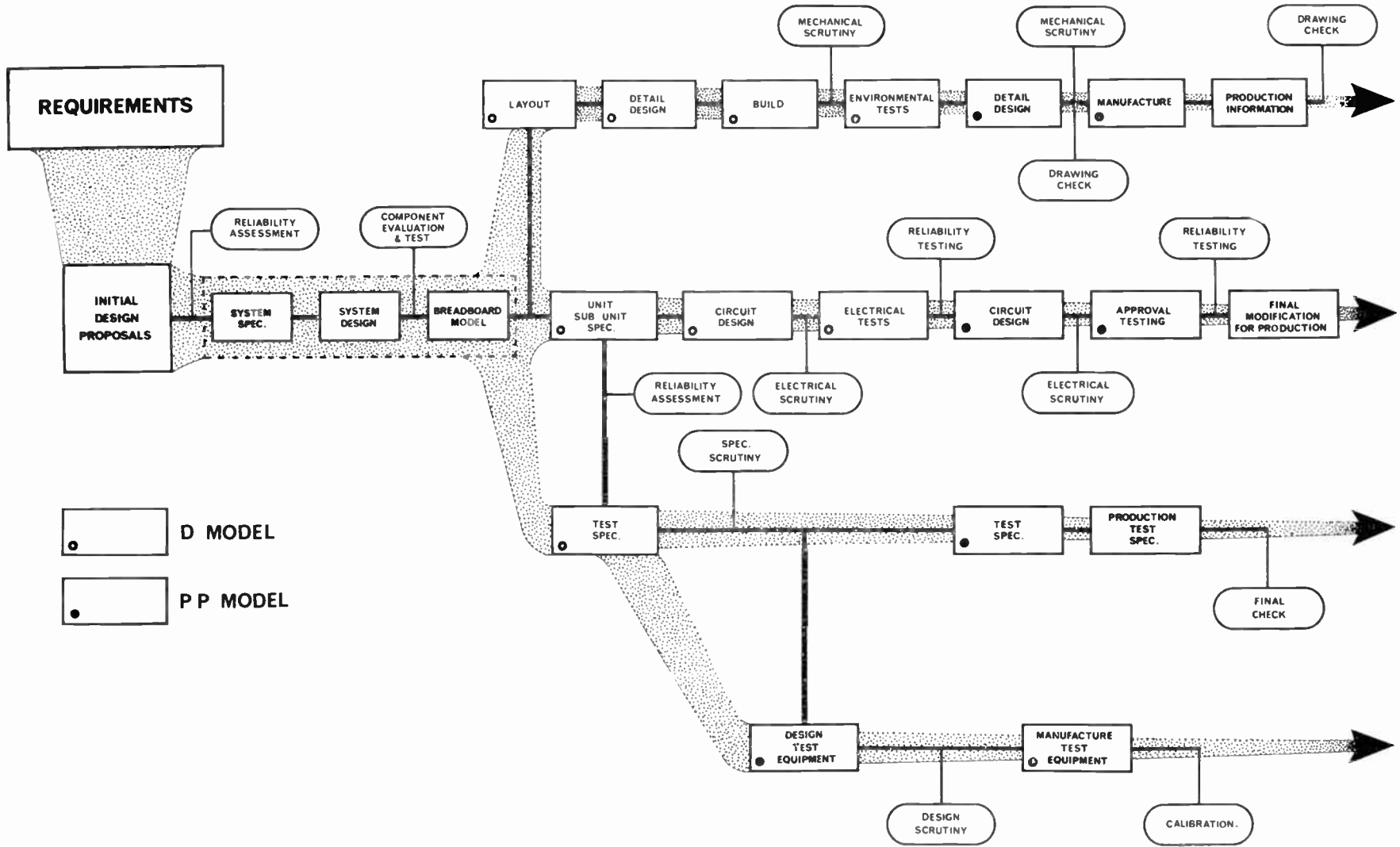
Fig. 1. Development of a piece of electronic equipment.

This stage is concluded when one or more development models are produced and have been subjected to the necessary electrical and mechanical testing.

Design of Pre-Production Models (PP models)

This is the final stage and finishes with a proven design of equipment and with information fully documented so that models made in production will have at least the same performance as the pre-production models

As mentioned previously, it is necessary to see how all these stages fit together.

One major point about these three stages which should be noted is that they all have a distinct beginning and ending. The beginning and end of each stage must be fully defined and there should be detailed information passed from one stage to another. For example, it is clearly impossible for test equipment specifications and hence test equipment to be sorted out unless unit and sub-unit specifications are available from which detailed information on the design parameters can be obtained. Nevertheless, it is certain that such a difficulty frequently arises; it happens because by the time test specifications are due to be written the equipment specifications are not complete. They are not completed probably because it was not appreciated that they would be required at this time. This example serves to emphasize the point that one must know in advance what stage is dependent on another.

It must be noted that the whole emphasis in this argument is being placed on reliability aspects. These points are directed at the best way to lay out a development programme with reliability in mind. Other factors in design tend, naturally, to limit the way that this can be done. This prevents the optimum solution being reached—it is, therefore, all the more necessary to clearly appreciate the layout of Fig. 1.

The one major point not shown on this chart is time. Obviously, each stage can take a variety of different times, dependent on individual circumstances. It is probable that a much better appreciation of the whole problem could be obtained by applying time to the diagram which would then take the form of a PERT (Programme Evaluation and Review Techniques) chart.

3. Early Design

Having explained the development programme, it is now possible to define the meaning of the words 'early design' used in the title of this paper. It will be defined as covering that same period of the development programme as System Design.

During this stage it is necessary to carry out reliability assessment work with the objective of determining what work will be required later in development to ensure that the reliability figure is met. It is necessary to determine how this work can be fitted into the development programme without introducing undue delay; it is also necessary to make some estimate of the probable cost of such an exercise.

Comparing the reliability of alternative methods of tackling a problem can also play a part in system thinking, by providing information which will later help to make decisions. It will also help in the evaluation of various methods of construction and design, and indicate where investigational work must be undertaken. This kind of thinking on reliability aspects can, and should, be done from the earliest stages of the design. These, in very general terms, are the points of which note must be taken as they are the actions which will help the whole programme.

It is widely believed that much of the assessment work which will be mentioned is too inaccurate to have practical value. This is not true; certainly one would like it to be more accurate, it would then possibly be of more value but, even so, by adopting the discipline which is implied in the acceptance of the philosophy, one takes a large stride forward and a positive attitude to reliability. This, it is hoped to demonstrate, has considerable value; one starts with very little information and as this is gradually built up, so is one's confidence in the practical value of the approach.

4. Prediction of M.T.B.F.

4.1. The Meaning of Prediction

The Concise Oxford Dictionary defines 'predict' as follows:

Predict; to prophesy or forecast.

In our context prediction implies deduction from facts already known and the use of mathematical calculation to determine future events or conditions. At any stage in the evaluation the results may not be as accurate as those which may be determined later, but one does have the advantage that it is possible to take corrective action if the indications of the prediction are that the performance will be unsatisfactory.

Prediction can, and should, be applied to many aspects of the proposed design, such as:

(i) Performance

(ii) Maintenance requirements

(iii) System failure rate

(iv) Economics, development cost and time

(v) Design characteristics—construction—components.

It must be realized that prediction is not a 'one shot' tool; it is a procedure of continual refinement. In the

early stages, during the feasibility study, complex predictions are unwarranted because of the lack of detailed information. It is during the stages leading up to system specification and system design that the most detailed work may be done. During this period the design is still fluid but there is sufficient information available to enable a reasonably accurate prediction to be made. At the same time, any changes shown to be required can be readily incorporated.

It has been shown that prediction is a valuable tool which can be used in a variety of ways. In the field of reliability, prediction has more limited connotation; it refers to the calculation of the fault rate of an equipment.

Calculations of this type are based on the assumption that over most of the operating life of the equipment the probability of failure in any given period is constant. This assumption of a constant failure rate, or more accurately, constant hazard rate, is sometimes referred to as the exponential model.

Thus $R = e^{-\lambda t}$ $\ldots\ldots(1)$

where R is the probability that the equipment will operate without failure for a period of time t.

λ = failure rate of equipment.

The m.t.b.f. of the equipment

$$m = \frac{1}{\lambda}$$

Therefore $R = e^{-t/m}$ $\ldots\ldots(2)$

When making predictions of equipment reliability it is customary to work in failure rates, since these may be added or subtracted arithmetically. When the final answer is obtained, the failure rate is inverted to obtain the m.t.b.f.

Thus $(m.t.b.f.)^{-1} = n_1\lambda_1 + n_2\lambda_2 + \ldots$ $\ldots\ldots(3)$

where n_1, n_2 = number of components of each type

λ_1, λ_2 = failure rate of each type of component.

A prediction of the m.t.b.f. of any equipment demands a knowledge of the number of components of each type and the appropriate failure rate for each type.

4.2. Component Fault Data

The most reliable information on the performance of components comes from equipment in actual operational use. Some lists of failure rates are available and provide a very useful basis on which to work but there are, in general, insufficient data defining where and how the information was obtained and the actual conditions of operation of the components.

The failure rate of any component is dependent mainly on its mechanical and electrical environment. The former includes things such as shock, humidity, and temperature. The latter includes the rating of the component and this must, of course, be related to the ambient temperature. A most important aspect of the environment of a semiconductor is the characteristics of the power line which supplies it; semiconductors are very prone to failure due to 'spikes' and surges, and their failure rates are therefore dependent on the freedom of the power lines from such irregularities.

In recent years, most equipment has been designed around semiconductors and, as these run at lower voltages and consume less current, the electrical environment of all associated components, such as resistors and capacitors has been considerably eased since the days when valves were in more common usage. It is therefore undesirable to employ data collected from valve equipment for the prediction of the performance of semiconductor equipment.

It is preferable for individual firms to measure the performance of components in equipment of their own design. This enables a list to be prepared which is representative of modern components, modern design and, perhaps more important, their own design, as each company will generally have sufficient differences in both electrical and mechanical design to modify component fault rates.

Such a list can be continually compared with other and perhaps more comprehensive lists from establishments like the Royal Radar Establishment (R.R.E.); by these means values for components which one has not actually measured, may be extrapolated.

4.3. List of Component Failure Rates

It was stated earlier that it is desirable to collect information on the performance of components in one's own equipment and apply this to the prediction of future developments. A typical example will be considered.

The equipment concerned was the ground part of a ground-to-air communications system; a relatively large digital equipment operated under conditions not dissimilar from those normally found in a laboratory. The details were obtained from one equipment but have since been confirmed from others; this particular equipment was operated for a total period of nearly 30 000 hours and details of the component failures are shown in Table 1.

It can be seen from the Table that quite large quantities of components were used—nearly three thousand transistors, nearly four thousand diodes, eleven and a half thousand resistors, nearly five thousand capacitors —and that there were seventy-four thousand soldered

Table 1

Component failure rates

Component causing failure	Number in one system	Failures: No. and rate per 10^6 hr					
		0–10 000 hr		10–20 000 hr		20–29 000 hr	
Transistors	2912	9	0·31	1	0·034	1	0·038
Diodes	3883	1	0·026		nil		nil
Resistors	11 500		nil		nil		nil
Capacitors	4715		nil		nil		nil
Transformers	110		nil		nil		nil
Printed circuit connector contacts	11 160	3	0·027	3	0·027		nil
Soldered joints	74 000	5	0·007	3	0·004	5	0·0075
Miscellaneous		2		nil		nil	

joints and eleven thousand other joints, such as connectors and printed circuit connectors.

The total number of component-hours involved is fairly high, of the order of 10^9 component-hours. There were nine transistor failures in the first ten thousand hours, whereas there was only one failure in each of the two remaining periods. Of these nine, six or seven occurred in the first few hundred hours of operation of the equipment and can therefore be regarded as non-random faults. The faults that occurred at a later period can be regarded as random faults.

It is significant that there were virtually no failures of diodes, resistors or capacitors and the failure rates were therefore not measurable.

The soldered joint failure rates were also very low, slightly lower than the value generally accepted for

Table 2

Component	Failures per 10^6 hours
Transistors: Germanium alloy/diffused	0·1
Germanium mesa	0·03
Silicon alloy/diffused	0·05
Silicon mesa	0·02
Silicon planar	0·005
Diodes: Germanium point contact/gold bond	0·05
Germanium junction	0·03
Germanium mesa	0·015
Silicon junction	0·02
Silicon mesa	0·01
Silicon planar	0·002
Silicon carbide planar	0·001
Integrated circuit solid state networks	0·03

soldered joints; this is also true of the failure rate of the printed circuit connectors.

4.4. *Semiconductor Fault Rates*

This particular equipment is now rather old: the design was 'frozen' in 1959, and the transistors and diodes used were germanium types. Since that time enormous strides have, of course, been made in the design of transistors. As this has resulted in improved methods of construction as well as in improved performance, the reliability of semiconductor devices has improved. Information is now available from several sources, mainly in the U.S.A. and R.R.E. Malvern, which make it possible to estimate the percentage improvement of these new types over those from which the initial information was obtained. The failure rates in Table 2 are based on our original figures with the appropriate multiplying factors.

It may be observed that a figure is quoted for integrated circuits. There is very little reliability experience of the use of these devices. Such information as there is stems from the U.S.A. where first results suggest that semiconductor integrated circuits will eventually have the same reliability as a single transistor. However, since this figure is a prediction of future performance and one for which there is little evidence to date, the figure chosen is 6 times that of a silicon planar transistor.

4.5. *Semiconductor Application Factor K_A*

The fault rate of all components is dependent on the total power they are dissipating. Failures can also be catastrophic or parametric.

A semiconductor in a digital application will be running with very low dissipation and its circuit is relatively insensitive to parameter changes. On the other hand, linear amplifiers tend to run at a higher level and are more sensitive to parameter changes.

out a simple switching operation will be run at a very low percentage of maximum rating and the circuit will be relatively insensitive to parametric changes. The transistor in a power supply application will be run at higher dissipation. The transistor in an analogue circuit probably will also be run with higher dissipation than the digital switch and circuit performance will be more dependent on parameter changes.

From the experience of the author's organization and that of other people, values of K_A shown in Table 5 for use in the prediction of the computer were obtained. The particular values given were derived for this application only and should not be applied directly to Table 2.

Table 5

Application	K_A
Power supplies	1·0
Store drive	0·5
Switching circuits	0·3
Analogue circuits	3·0

There was insufficient experience on the failure rate of silicon planar transistors so the figure of 0·05 faults per 10^6 hours quoted by R.R.E. was used. This figure relates to transistors operating at 100% of rated dissipation in ground equipment with an ambient temperature of 25°C. With the power supplies in the airborne computer, on the other hand, the transistors dissipate only 20% of their rated power and the maximum design temperature of the air within the equipment is 85°C. Clearly due allowance must be made for these differences before the failure rate quoted by R.R.E. can be used.

To do this, graphs similar to Fig. 3, but which refer to transistors, are employed. From these it is found that silicon transistors at 20% dissipation in an ambient temperature of 85°C are about 10 times more reliable than those at 100% dissipation at 25°C.

Thus the failure rate under these conditions is 0·005 faults per 10^6 hours.

The graphs also show that, in the range being considered, the failure rate of silicon transistors is roughly doubled for every 10°C rise in ambient

Table 6

Ambient temperature maximum	K_θ
65°C	0·3
75°C	0·5
85°C	1·0
100°C	3

temperature; Table 6 can therefore be produced for K_θ.

Three other types of transistor circuit have still to be considered and for this the failure rate established for the power transistor is taken as a basis. Therefore the failure rate for any other transistor in this equipment will be:

$$K_A \times K_E \times K_\theta \times 0 \cdot 005 \text{ faults per } 10^6 \text{ hours}$$

This process can be continued for other components of the system to obtain an overall m.t.b.f.

6.1. Variation of M.T.B.F. with Temperature

It has been shown that it is possible and desirable to allow for the effects of temperature in determining the fault rates of individual components and it is interesting to see what the overall effect of temperature is on the complete equipment.

Figure 4 shows the predicted variation in m.t.b.f. of the same proposed airborne digital equipment. This unit employs some 30 000 components of various types. It can be seen that whilst, as one might expect, the reliability falls quite rapidly beyond 80°C, even between 65–85°C there is a predicted drop of nearly 20% in m.t.b.f. It is clear therefore that it is most desirable to keep the ambient temperature below, say, 65°C.
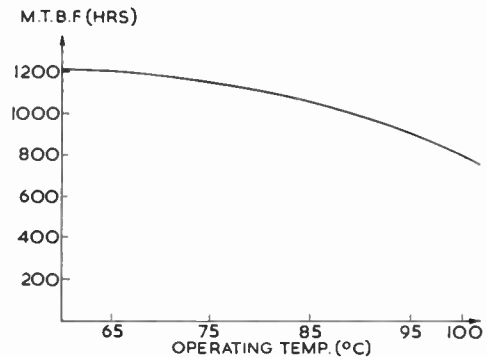

Fig. 4. Predicted variation in m.t.b.f. with operating temperature

Naturally this is only one piece of information which helps the designer but it does enable him to decide whether the disadvantages of more powerful cooling systems are justified by the improved system reliability resulting from their use.

6.2. Choice of Component

These prediction techniques can help to decide which components are best used in equipment and as an illustration we shall use an airborne digital converter. Let us assume a typical question, 'What are the advantages to be gained by the use of integrated circuits?'

Table 7
Airborne converter

	Fault rate per 10⁶ hours	Proportion of total fault rate
(i) With integrated circuits		
Power unit	15	5%
Selector unit	100	30%
Analogue circuits	200	60%
Logic unit	15	5%
Total	330	100%
M.T.B.F. = 3000 hours		
(ii) With discrete components		
Power unit	15	4%
Selector unit	100	25%
Analogue circuits	200	51%
Logic unit	75	20%
Total	390	100%
M.T.B.F. = 2500 hours		

It is often assumed that the answer to this is increased reliability but, considering the figures given in Table 7, it can be seen that the improvement in the fault rate is of the order of 20%, nothing like as much as one might perhaps have expected. The reasons for this will be dealt with later—let us for the time being examine the lessons to be learnt from Table 7.

The main point is that, although the overall failure rate does not alter significantly by changing discrete components for integrated circuits, the fault rate of those sections using integrated circuits is very low indeed. If a section of an equipment has a low probability of failure over a long period, consideration can be given to designing this section in such a way that no maintenance is possible, except possibly in the factory. This, apart from a reduction in maintenance, often results in a reduced number of external connections to this unit and the need for test points; such design features improve further still the reliability of that section.

Table 8

	Fault rate per 10⁶ hours	
	Existing design	New design
Circuit components	2800	200
Relays, switches, etc.	600	300
Rotating machines	2000	1000
Connections	1400	200
Total	6800	1700
M.T.B.F. (hours)	150	600

Decisions on the size of such non-repairable sections are vexed ones and are dependent on the m.t.b.f. of the section in question. As such decisions have to be taken in 'early design', prediction is of particular value. So in our particular example one can see that those sections of the logic units which use integrated circuits may well be designed to be non-repairable.

One further example of the value of prediction in the choice of components concerns a piece of mobile equipment for the army. Details are given in Table 8. The existing equipment has an overall fault rate of nearly 7000 faults per 10⁶ hours (m.t.b.f. 150 hours). By making use of modern components and modern connections it is possible to reduce the fault rate of components by 14 to 1 and of connections by 7 to 1. Yet the overall fault rate is only improved by 4 to 1 and further improvement will be extremely difficult unless the rotating machines used to generate power supplies are replaced by more reliable sources of power.

6.3. Choice of Construction

Earlier, when considering the choice of components, it was noted that the use of integrated circuits instead of discrete components improved the reliability of the equipment by about 20%, when perhaps a greater improvement might have been expected.

The use of integrated circuits often involves the use of more sophisticated forms of construction, such as 'Rational Packaging', in order to obtain the improvements in packing density which such circuits offer. Thus, when comparing the improvement in reliability which is obtained by the use of integrated circuits, it is necessary to stipulate what improvement is due to the use of the components and what is due to the use of improved forms of construction.

In order to demonstrate the improvement in m.t.b.f. with different forms of construction, let us take the example of an airborne digital equipment which is engineered in three ways.

Firstly, consider the design as it might have been some seven years ago in the days when most transistors were germanium devices. A very simple prediction (Table 9(a)) shows that the m.t.b.f. would be approximately 260 hours.

Using modern silicon planar devices the prediction (Table 9(b)) shows that the fault rate of the semiconductors is reduced from 2000 per 10⁶ hours to 100 per 10⁶ hours, an improvement of 20 to 1 and yet the overall improvement in m.t.b.f. is only about 2 to 1. Further examination shows that a large percentage of the total fault rate is attributable to connections of all sorts and hence any further improvement in

reliability can only be achieved by making a substantial reduction in the connection fault rate.

This information is of prime importance in the design of this type of equipment and it has been used extensively in arriving at a form of microminiature construction called 'Rational Packaging' which has been developed at the G.E.C. Applied Electronics Laboratories, Stanmore. It is not intended to discuss this construction but merely to refer to some features which relate to reliability.

Table 9
Airborne Digital Equipment

Design	(a)	(b)	(c)
	Germanium	Silicon	Silicon
6000 semiconductors	2070	106	106
8000 other components	490	186	186
2000 plugs and sockets	520	520	—
40000 soldered joints	800	800	—
2000 wrapped joints	—	—	8
800 soldered joints	—	—	15
40000 welded connections	—	—	43
Faults per 10^6 hours	3880	1612	358
M.T.B.F. (hours)	260	620	2800

The design involved the replacement of plugs and sockets previously used for internal interconnections by wrapped connections. This reduced the predicted failure rate from 520 to 8 per 10^6 hours, thus effectively eliminating one of the major obstacles to the achievement of higher equipment reliability. This left soldered joints as the major cause of unreliability. In 'Rational Packaging' nearly all component connections are welded rather than soldered to provide a suitably small joint for microminiaturized equipment. Originally the connections were made by single welds whose failure rate is estimated as being about four times better than that of soldered joints. This could have reduced the fault rate from 800 to 200 faults per 10^6 hours; still, however, a large part of the total fault rate.

As a result of reliability considerations, the decision was taken to provide, wherever possible, twin welds on all connections. Calculated over a period of, say, 20 years the failure rate for a twin welded joint is negligible. The failure rate for welded joints quoted in (Table 9 (c)), is almost entirely due to certain connections where redundant welds are at present impractical. Here, therefore, is a clue as to how the designer can further improve reliability.

However, the total failure rate for connections is now less than 20% of that for the whole equipment

and it can be said that connection reliability is no longer the major problem.

7. Verification of Predicted M.T.B.F.

The value of the prediction of m.t.b.f. lies not so much in getting an accurate answer but in its value to the designer in determining the correct course of action to be taken. This paper has attempted to demonstrate this. The accuracy with which one can do the calculations is, of course, always of interest and the proof of the pudding is always in the eating, so, in conclusion, what can one say on this subject?

For the purpose of verifying the predicted m.t.b.f. a large analogue computer has been chosen. It is used in our laboratories and employs many tens of thousands of components. The fault rates employed in the 'part count exercise' were obtained from the same digital equipment to which reference was made previously and, because the subject of the prediction is an analogue device, an overall weighting factor of approximately six times was allowed. Fault recording was instituted on the equipment from the beginning of its operation and, over the first period of six thousand hours running time, the fault rate actually measured was twice that which we had predicted. This, it is felt, is a reasonable agreement, particularly as during the period under consideration the equipment had been subjected to modifications. It is worthy of note that, as predicted, connections, particularly plugs and sockets, were a major cause of trouble.

Component failure rates were obtained and a reasonable degree of confidence can be placed on the answers because of the relatively high number of component-hours obtained.

Germanium transistor failure rates were lower than we had predicted being 0·2 faults per 10^6 hours instead of 0·5 faults per 10^6 hours. This may well indicate that the construction of semiconductor devices has been gradually improved.

Semiconductor diodes had a higher failure rate than had been expected but other types of components had observed failure rates which compared quite closely with their predicted value. The main reason for the higher fault rate of the whole equipment was the larger number of connection failures than was predicted.

8. Reliability Testing

The uses of prediction of failure rate at the early stages of a development programme have been reviewed. If the programme is to be successfully concluded from a reliability viewpoint it is essential that a measurement be carried out on pre-production models to determine whether the required failure rate has been achieved.
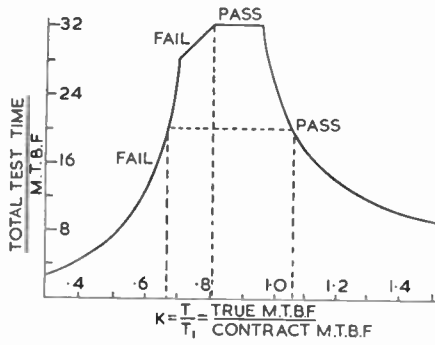
Fig. 5. A.G.R.E.E. sequential sampling test.

This type of testing, carried out to determine the fault rate of the equipment, is often referred to as A.G.R.E.E. testing, so named from the initial letters of the American committee which first proposed the particular tests widely used. Details of these tests are given in the Appendix.
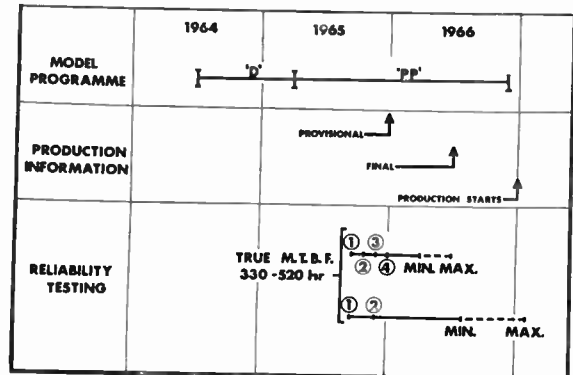
The purpose of the A.G.R.E.E. was to ensure that equipment in service had an m.t.b.f. which was at least as high as the specified value. The tests therefore were designed to show that the equipment met the specification; they give a simple accept or reject decision and not an absolute measurement of the m.t.b.f.

Such reliability testing must inevitably be a somewhat lengthy life test in order to build up a significant number of equipment running hours. It is essential that this testing be fitted into the overall development programme if it is to have the maximum value. It is therefore necessary to be able to predict the number of equipment hours of testing which will be required and to determine from this the number of models which will be required for test.
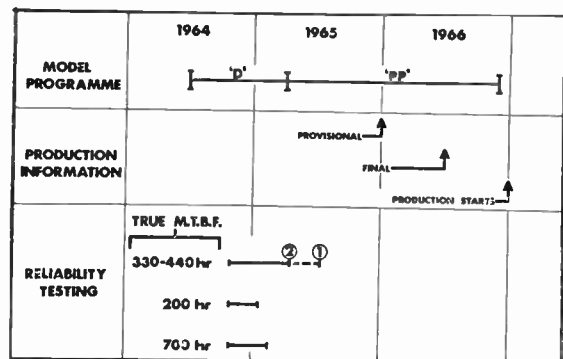
The A.G.R.E.E. sequential sampling test is shown in Fig. 5. This illustrates the standard A.G.R.E.E. production test, with normalized testing time plotted against K which is the actual m.t.b.f. of the equipment divided by the contract m.t.b.f. The first thing to note is that the longest testing time will be when the true m.t.b.f. is between 0.81 and $0.93 \times$ the contract m.t.b.f. As the true m.t.b.f. falls the testing time required to reject an equipment reduces rapidly and it can be seen that when $K = 0.5$ the normalized testing time is 7 compared with a maximum of 33. Similarly, as the true m.t.b.f. increases the testing time again falls. Thus if an equipment m.t.b.f. is much lower or higher than the contract m.t.b.f. this fact will be discovered fairly quickly. As the m.t.b.f. approximates to the required m.t.b.f. the testing time approaches the maximum. Let us consider what this means in terms of the development programme and for this purpose we will take a hypothetical but nevertheless practical programme of model manufacture.

Figure 6(a) shows a period of development (D) model manufacture and a period of pre-production (PP) model manufacture and the times when information is to be confirmed with the production organization. If we first take the case of reliability testing at the pre-production stage, it is assumed that we are demonstrating to the customer that the equipment will pass the specified m.t.b.f. The designer should be confident that the equipment will pass this test; preferably the m.t.b.f. should not be too excessive as it is then possible that the equipment is over-designed and perhaps, therefore, too expensive.

Referring again to our example, let us assume that the contract m.t.b.f. is 500 hours. A.G.R.E.E. quotes that the most probable normalized testing time is 20. In this example a test of this duration would give a result if the true m.t.b.f. was less than 330 or more than 520 hours. If the m.t.b.f. lies between these two figures the testing time will lie between 20 and 33. If we have four models available at a rate of 1 model per month this would require a minimum testing time of 6 months, and a maximum of 9 months. The first of these models would probably be No. 6 in the batch as it is unlikely that earlier models would be available for testing, simply because

(a) PP models.

(b) D models.

Fig. 6. Phased testing.

of prior demands. So it can be seen that in the worst case the tests would be complete at about the time of the freeze of the design information to the production unit. If, however, there are only two models available then the testing time is such that the test will not be complete until after the design freeze or, in the worst case, until after production has started.

Testing at this stage is a demonstration to the customer that the m.t.b.f. is satisfactory, therefore it is necessary that the designer has a high confidence in the successful outcome of the test. This means, however, that the test will be fairly long, probably lasting for at least a normalized time of 20. It is to be assumed that these tests are part of approval testing and must be completed before Stage B approval is granted and full scale production starts.

It is therefore clear that at some time prior to this, some reliability testing must be carried out to enable these later tests to be conducted with confidence in their successful outcome. Equally is it necessary for there to be a reasonable number of equipments available to complete the testing as quickly as possible?

Now what about testing at an earlier stage: for example, on D models? Difficult though it may be to get pre-production models for reliability testing, obtaining D models for this purpose is even more difficult. In order to ease this situation, A.G.R.E.E. have designed a test for development models which requires far fewer equipment-hours of testing than does the production test. We are thus enabled to reach a decision in a reasonable time despite having only the one or two D models likely to be available for such tests.

If the equipment m.t.b.f. lies within the range 330–440 hours, the testing time with two equipments will be six months. (Fig. 6 (b).) This would tell us that the equipment was capable of achieving the specified m.t.b.f. within the course of subsequent development. With only one equipment the test would take nine months, finishing close to the start of reliability testing of pre-production models. However, let us consider the case of an m.t.b.f. of 200 hours, a period of time considerably below the figure we require. In this case we can, using one equipment only, decide if it is unsatisfactory within three months, just prior to the start of the pre-production model manufacturing programme. This will still leave a reasonable amount of time in which to put things right. Equally, if the true m.t.b.f. is 700 hours we could, again using one equipment only, accept the design as satisfactory after four months, well before the start of pre-production model reliability testing.

As long as the testing time increases without rejection, one's confidence of the successful outcome increases. If the m.t.b.f. is too low, even by a factor of 2 to 1, the fact will be discovered early enough for modifications to be made to at least the later pre-production models which are to be tested.

The number of equipments tested is, of course, very small and there are obvious dangers in trying to read too much into the answers obtained. However, they should be to some extent confirmation of the previous predictions and valuable information on the m.t.b.f. of the equipment will be obtained. It should not be forgotten, however, that one is taking a chance; a chance that is necessary because of the almost certain lack, in any development programme, of enough models for test purposes.

Consideration should therefore always be given to testing all models for whatever purpose they are intended, in order to build up some additional information on the reliability of the equipment.

9. Bibliography

1. "Reliability Engineering for Electronic Systems". Edited by Richard H. Myers, Kam L. Wong and Harold H. Gordy. (John Wiley, New York, 1964.)

2. Richard R. Landers, "Reliability and Product Assurance", (Prentice-Hall, Englewood Cliffs, N.J., 1963.)

3. "Reliability Abstracts and Technical Reviews". Prepared monthly for National Aeronautical Space Administration by Research Triangle Institute.

4. "Reliability Engineering". Edited by William H. von Alven, Arinc Research Corporation. (Prentice-Hall, Englewood Cliffs, N.J., 1964.)

5. Norman H. Roberts, "Mathematical Methods in Reliability Engineering", (McGraw-Hill, New York, 1964).

6. "Semiconductor Reliability", Edited by John E. Shwop and Harold J. Sullivan. (Chapman and Hall, London, 1962.)

7. "R.A.D.C. Reliability Notebook", U.S. Department of Commerce. (McGraw-Hill, New York, 1961.)

8. "Reliability of Military Electronic Equipment" ('AGREE' Report U.S.A.). Advisory Group on Reliability of Electronic Equipment, Office of the Asst. Secretary of Defense (Research and Engineering), 4th June, 1957.

9. "Engineering Electronic Equipment for Reliability", G.E.C. (Electronics) Ltd., Brochure, Reference No. TB 004.

10. "Reliability in Action", G.E.C. (Electronics) Ltd., Brochure, Reference No. TB 010.

11. "Reliability in Action", G.E.C. (Electronics) Ltd., Brochure Reference No. TB 015.

12. "Reliability of Electronic Equipment". Report by the Technical Committee of the British Institution of Radio Engineers. *J.Brit. I.R.E.*, 21, No. 4, pp. 287–90, April 1961.

10. Appendix

A.G.R.E.E. Testing

A.G.R.E.E. testing derives its name from the report published in 1957 by the American Advisory Group on Reliability of Electronic Equipment. It is a form of testing carried out by the manufacturer of an electronic equipment to determine experimentally whether or not his product satisfies certain quantitative standards

of reliability which have been specified by the customer as a contractural requirement. These standards are usually expressed as the mean time between failures, (m.t.b.f.), of the equipment, and A.G.R.E.E. recommended that the m.t.b.f. specified in the contract should be 50% greater than the customer's minimum requirement.

The A.G.R.E.E. test plan, (a truncated sequential test), is designed to ensure that the risk of accepting an equipment whose m.t.b.f. is less than the customer's minimum requirement, and the risk of rejecting an equipment whose m.t.b.f. is greater than the contractual requirement, are both less than 10%. The operating characteristic of the test is shown in Fig. 7. The number of failures experienced on test is plotted against the total number of equipment hours of operation accumulated up to the time of failure, thus producing a 'staircase' type of trace, (see Fig. 8). Two 'decision lines' are drawn on the graph with slopes reflecting the required m.t.b.f., and when the trace meets one of these lines, the test stops and the decision is taken to accept or reject the equipment.
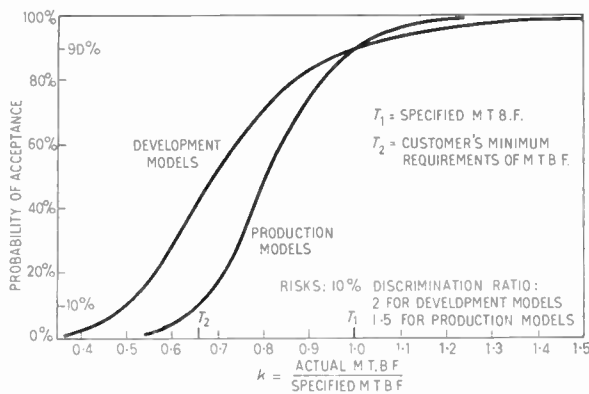

Fig. 7. Operating characteristics of A.G.R.E.E. sampling schemes.

These are clearly specific examples of operating characteristic curves for truncated sequential reliability tests. They have been chosen with a view to obtaining the necessary result in a reasonable time without too heavy a risk being applied to either customer or producer. There is little evidence so far in this country on the operation of tests to such conditions, though these same tests have been current in the U.S.A. for 8 years without modification of these particular parameters.

However it should be remembered that any demonstration which is called for under these conditions can be a contractual liability and as such it may well be desirable to examine closely the risks associated with any particular test.
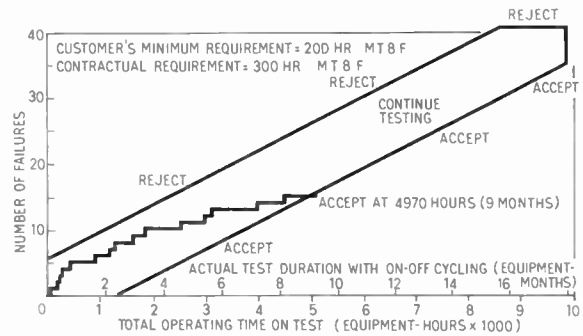

Fig. 8. Example of A.G.R.E.E. production test.

The test is truncated so that the maximum operating time required to reach a decision is 33 times the m.t.b.f. specified in the contract. The truncation of the test slightly modifies the 10% risks mentioned above but, for all practical purposes, this effect can be ignored. Test equipments undergo 'on-off' cycling in the ratio 3:1, so the maximum test time will be 44 times the specified m.t.b.f. However, it is very unlikely that the test will actually take this long; unless the true equipment m.t.b.f. is in the marginal zone between the customer's requirement and the figure specified by contract, a decision should be reached within 30 (rather than 44) times the specified m.t.b.f. Thus, in the example, (Fig. 8), the customer's requirement is 200 hours, giving a contractual m.t.b.f. of 300 hours; the maximum test duration with continuous testing (168 hours/week), is then 18 equipment-months, but a decision will be reached within 12 equipment-months unless the true m.t.b.f. lies between 200 and 300 hours. The greater the difference between the true and the required m.t.b.f., the shorter will be the testing time required to reach a decision.

Since the test duration is expressed in terms of equipment-hours, it is obvious that the period of testing can be varied by altering the number of equipments tested. A.G.R.E.E. specify that at least two equipments shall be used in the production test, and that no decision may be taken until each equipment has operated for at least three times the specified m.t.b.f. This means that between two and eleven equipments may be used in the test. In practice, however, the possible number is often reduced by considerations of the production rate of equipments for test. Again consider our example, in which we know that the greatest possible test duration is 18 equipment-months, and let us now assume that equipments are produced for test at the rate of one a month. If the test begins as soon as the first equipment is ready, and subsequent equipments are put on test when they become available, the maximum test period will be:

9½ months with two equipments

7 months with three equipments

6 months with four equipments

5½ months with five equipments

If five equipments are allocated for testing, the last one will only just achieve its minimum operating period of three times the specified m.t.b.f. before the total test time reaches the maximum. (Clearly, the actual time required to reach a decision may be much less than the maximum, but this can seldom be predicted with sufficient confidence.) Thus the number of equipments allocated for testing should be between two and five; the choice within this range is at the discretion of the manufacturer and will depend on such factors as the cost of testing, the capacity of the test facilities available, the urgency with which the result is wanted, and so on.

So far this appendix has considered only the A.G.R.E.E. test for pilot-production and production equipments, since this is what is generally meant by 'A.G.R.E.E. testing'. However, A.G.R.E.E. have also proposed additional tests at the development and the full production stages. The development test is intended to establish whether or not the equipment is likely to meet the required standard of reliability in the course of subsequent development. This test is designed to have rather wider confidence limits and consequently does not take so much time; the maximum normalized test time is only 10·3 as against 33 for the production test.

The full production test is intended to ensure that equipments from the production line will continue to maintain the required standard of reliability after the equipment type has been accepted by the production test previously discussed. Since this test requires simultaneous testing of at least 22 equipments before any may be delivered to the customer, it is not really applicable to complex equipments with production rates of the order of two or three a month.

Finally, it should again be pointed out that A.G.R.E.E. testing is *not* primarily intended to measure equipment reliability in absolute terms. The tests are designed to compare the observed reliability with a predetermined standard agreed between manufacturer and customer, and to decide whether or not the equipment meets this standard. In addition, it has been shown that reliability testing may take a considerable time even with a significant number of early production equipments. It is, therefore, essential that if reliability testing is called for it should be integrated into the development programme from the start. Customer and manufacturer must agree upon the contractual requirement of reliability before plans for testing can be made and the number of test equipments decided upon. If these arrangements are not settled early in the model programme, it is most unlikely that sufficient equipments will be available for testing in time for the results to be of any practical value.

Transient Analysis of a Varactor Bridge Doubler

By

A. UHLIR, Jr., Ph.D.†

Summary: A direct digital integration of circuit equations was used to study the transient behaviour of a varactor bridge doubler circuit. A suddenly-applied sine wave was assumed for the fundamental drive. A piecewise-linear varactor characteristic was used.

The calculations show that good efficiency and rapid transient response can be obtained for certain ranges of tuning and bias.

Experimentally observed input-output power relations are reproduced remarkably well by theoretically calculated results. These include an abrupt increase from negligible output to high efficiency as output is increased, for fixed bias voltage. Also, instabilities are observed when the power is too high for a given bias. The calculations thus prove that side-effects such as rectification and avalanche multiplication in the varactor diode are not necessary for the presence of spurious oscillations.

1. Introduction

The transient behaviour of varactor multipliers is of interest for generation of pulsed microwave power. A theoretical solution in the time domain is required to study this situation. Another objective of the calculations was to evaluate the time domain solution for its applicability to steady state problems.

One of the advantages of the time domain approach is its conceptual simplicity. Problems can be set up very directly from Kirchhoff's laws and the defining equations of the circuit elements. The amount of arithmetic calculation required for a digital solution in the time domain is probably larger in most cases than in a well-organized frequency-domain analysis to obtain steady-state results. The labours of the computer may be said to have been substituted for some of the thinking required in analysis.
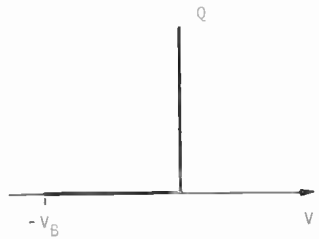
Steady-state multiplier analyses are not always valid for varactor circuits. The frequency-domain analysis is always subject to possible error from this basic assumption. Time-domain analysis is more likely to reveal spurious oscillations and sub-harmonic generation than is the frequency-domain type of analysis. A suddenly-applied sine-wave fundamental generator was assumed in all calculations. The choice of a digital computer for this time-domain solution was primarily dictated by availability.

A piecewise-linear varactor characteristic was used in the calculations, not as a particularly significant

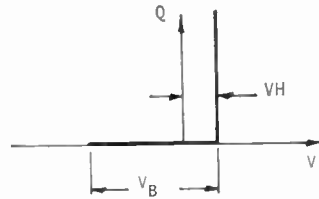† Microwave Associates, Inc., Burlington, Massachusetts, U.S.A.

reduction in computation, but because of the opinion that it is the most desirable idealization of the varactor for preliminary analysis of multiplier circuits. This belief is part of a proposed strategy for obtaining deeper understanding of multiplier circuits. First, find out what is possible and not possible theoretically with ideal varactors having the characteristic shown in Fig. 1. The consideration of all possible circuits using one or more such varactors will include the circuit possibilities for any other monotonic charge-voltage characteristic, since any other characteristic can be approximated by the combination, with suitable bias, of a number of the ideal varactors. Series resistance can also be added as part of the realm of possible circuits. (Non-linear resistance elements, in combination with varactors, would not be encompassed under this classification; the combination deserves special investigation.)

From theoretical studies based on the ideal varactor characteristic, general principles of circuit design may be developed by induction. It is likely that some of these principles will transfer over to the case of varactors with more realistic charge-voltage characteristics or losses such as series resistance. Further design principles will have to be added to achieve optimum circuits in the presence of varactor losses. The regions of search will be significantly narrowed, however, if it is possible to start with the knowledge of the optimum circuits for ideal varactors. Some of the available theoretical work on ideal varactor circuits is given in refs. 1, 2 and 3. The combinations of non-linear resistance elements (varistors) and

(a) Ideal varactor characteristic.

(b) Ideal varactor with built-in bias voltage.

Fig. 1.

varactors deserve special investigation;[4] a comprehensive knowledge of varactor circuit possibilities would help define the unique contributions of the varistors.

2. Circuit Equations

A bridge circuit was chosen because the automatic cancellation of the odd harmonics leads to a simpler circuit than a comparably efficient single varactor circuit, as far as is known. The only passive elements required for good efficiency in the bridge circuit are two inductors. The advantages of using more junctions to reduce the number of passive elements are well known in design of integrated circuits.

The circuit simplicity is thought to favour bandwidth and good transient performance, and this intuitive notion has been borne out by the calculations. Bridge doubler circuits have been mentioned previously[5, 6]; the suppression of odd harmonics makes it possible to contemplate octave bandwidths.

The capacitance of the depletion region under reverse bias is a first-order parameter in nearly all varactors for multiplier use, so that the calculation has been done for the piecewise-linear varactor shown in Fig. 2 (equivalent to an ideal varactor plus a fixed shunt capacitor) with care taken to let the forward elastance S_{min} approach zero. The calculations were quite stable throughout this limiting process, in agreement with prior theoretical and experimental notions that varactors can be specified rather well by S_{max} alone.

The bridge circuit is shown in Fig. 3, with the numerical values used in most of the calculations. The

necessary equations can be established by considering three meshes. The bridge of varactors may be treated as one mesh, giving

$$V_3 + V_4 = V_5 + V_6 \qquad \ldots\ldots(1)$$

The input and output circuits are two more meshes. By expressing the voltage equation around these meshes in terms of the voltage across the inductor appearing in each mesh, one obtains the equations:

$$i_1 = A_1 \int dt [V_g(t) - i_1 R_g + V_5 - V_3] \quad \ldots\ldots(2)$$

$$i_2 = A_2 \int dt [-V_5 - V_6 - i_2 R_L] \qquad \ldots\ldots(3)$$

It seems to be most expeditious to calculate the currents through the varactors and to obtain the charges in the varactors by integration of these currents. To use this approach, we can deal with the varactors in terms of elastances, defined by

$$\begin{aligned} S_3 &= dV_3/dQ_3 \\ S_4 &= dV_4/dQ_4 \\ S_5 &= dV_5/dQ_5 \\ S_6 &= dV_6/dQ_6 \end{aligned} \qquad \ldots\ldots(4)$$

Equations involving elastances and the currents can be derived by differentiation of eqn. (1) with respect to time. Since

$$dV_3/dt = (dV_3/dQ_3)(dQ_3/dt) = S_3 i_3 \qquad \ldots(5)$$

equation (1) becomes

$$S_3 i_3 + S_4 i_4 = S_5 i_5 + S_6 i_6 \qquad \ldots\ldots(6)$$

The equations of continuity are

$$\begin{aligned} i_1 &= i_3 - i_4 \\ i_1 &= i_6 - i_5 \\ i_2 &= i_3 + i_5 \\ i_2 &= i_4 + i_6 \end{aligned} \qquad \ldots\ldots(7)$$
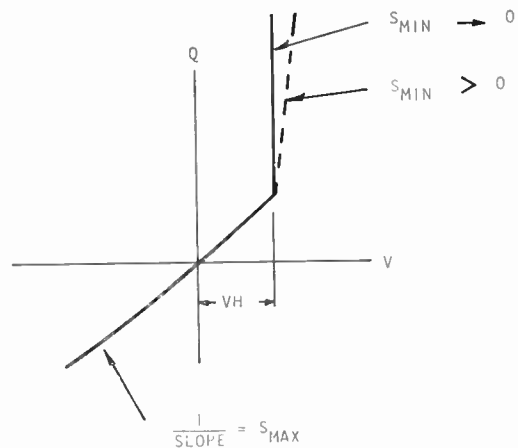

Fig. 2. Piecewise-linear biased varactor characteristic assumed in calculations. Equivalent to ideal varactor with shunt capacitor.

These equations are not independent. Any three of them may be combined with eqn. (6) to give the varactor currents in terms of the input and output currents, e.g.

$$i_4 = \frac{(S_5+S_6)i_2-(S_3+S_5)i_1}{S_3+S_4+S_5+S_6} \qquad ......(8)$$

$$i_3 = i_1 + i_4 \qquad ......(9)$$

$$i_5 = i_2 - i_3 \qquad ......(10)$$

$$i_6 = i_2 - i_4 \qquad ......(11)$$

A set of values for most of the circuit elements was selected by rule-of-thumb as a starting point in the calculations. These values are shown in Fig. 3 and
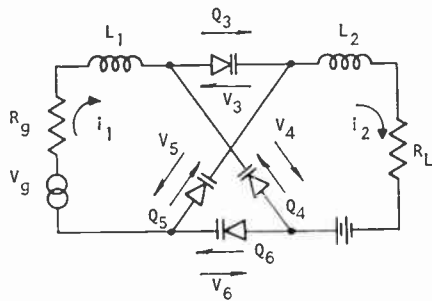

Fig. 3. Bridge doubler circuit with element values.

$S_{max} = 12\cdot5664$	$1/L_2 = A_2 = 25\cdot1327$
$S_{min} = 0\cdot0001257$	$R_g = R_L$
$1/L_1 = A_1 = 6\cdot2832$	$V_g = 100.$

were chosen on the basis of the following reasoning. The frequency and generator voltage were arbitrarily set equal to unity and 100, respectively. Another parameter can be chosen arbitrarily without restricting the generality of the results; the maximum elastance S_{max} was given the value 4π. The average elastance of the varactors depends upon the bias and drive voltage. As a guide for the assignment of inductance values, an average elastance of one-half S_{max} was postulated. Then, the input inductor was chosen for series resonance with this average elastance at the fundamental frequency. Similarly, the output inductor was chosen for series resonance with the average elastance at the output frequency. The finite value of S_{min} was chosen arbitrarily and was determined by trial to be small enough so that its actual value did not matter. This parameter cannot, however, be set equal to zero without modifications of the program to avoid dividing by zero in eqn. (8).

3. Calculation Methods

The procedure used in the calculation is as follows. At a given instant, each varactor has a certain value of charge. The voltages and elastances of the varac-

tors are calculated from the charges and the charge-voltage characteristics.

A special feature of this type of problem is that all of the derivatives of the functions to be integrated can be calculated. Taylor's series of any desired accuracy can thus be used to calculate the increments in the integrals. Just the first derivative of the integrand is used in this program, that is,

$$\int_{x_1}^{x_2} f(x)\,dx \simeq (x_2-x_1)f(x_1)+\tfrac{1}{2}(x_2-x_1)^2 f'(x_1) ...(12)$$

In effect, a parabola is used as an approximation to the current over the integration interval.

The above method of integration may be adapted to problems using any analytically specified varactor capacitance-voltage characteristic. Exact simultaneous integration of the circuit equations can be accomplished for the piecewise linear case, based on the fact that the network behaves in linear fashion between breakpoints, but this method was not adopted because it lacked general applicability.

The varactor voltages, together with the existing values of currents and known value of generator voltage (and its time derivative) are used to evaluate the integrands for eqns. (2) and (3) and the time derivative of the integrands. The input and output currents for the next time interval are obtained by these integrations carried out in accordance with eqn. (12).

The new currents in the varactors are calculated from the input and output currents and the varactor elastances, in accordance with eqns. (8) to (11). The charges in the varactors are then obtained by trapezoidal integration, using both the old and new currents. Next, the old currents are replaced by the new currents. The steps are then repeated.

A calculation involving the above steps solves the basic problem, since it gives the output current as a function of time. The peak inverse voltage applied to each varactor is stored. Other questions require calculation of efficiency. It is also interesting to find out how efficient the multiplier circuit is in generating unwanted harmonics, to indicate how much additional filtering may be required.

4. Fourier Analysis

Strictly speaking, efficiency is defined only for a steady periodic waveform. If the output waveform for each cycle of the fundamental is used to calculate Fourier components, the values of these components will approach asymptotic values if the output settles down to a periodic waveform, as it must for a satisfactory multiplier. If the result is non-periodic, this fact will be apparent in the failure of the cycle-by-cycle Fourier components to approach steady values.

The textbook discussions of the basic method for finding Fourier coefficients prescribe integration of the product of the function together with sine and cosine functions representing the desired harmonics. The analysis could therefore be done by standard methods of numerical integration. However, simple summations can be used and are likely more accurate than integration.

The basis for this approach is to treat the specific problem of calculating Fourier coefficients in a slightly different way from the numerical integration of functions in general. In effect, the place where the approximation is made is transferred from the integration to the representation of the function. That is, one makes the approximation that a function $f(x)$ can be represented by a finite number of harmonics in accordance with eqn. (13).

$$f(x) = \sum_{n=-M}^{M} a_n \exp(jnx) \qquad -\pi < x < \pi \quad ...(13)$$

Once this approximation is accepted, there is no further need for approximation. Let us try to find the coefficients a_n by the summation over a selected set of values of $x = x_l$, $l = 1$ to N

Let

$$b_k = \sum_{l=1}^{N} f(x_l) \exp(-jkx) \qquad ......(14)$$

From eqn. (13) and interchange of the order of summation,

$$b_k = \sum_{n=-M}^{M} a_n \sum_{l=1}^{N} \exp j(n-k)x_l \qquad ......(15)$$

If the x_l are equally spaced through a cycle,

$$\sum_{l=1}^{N} \exp j(n-k)x_l \begin{array}{l} = 0 \quad \text{if } n \neq k \\ = N \quad \text{if } n = k \end{array} \qquad ......(16)$$
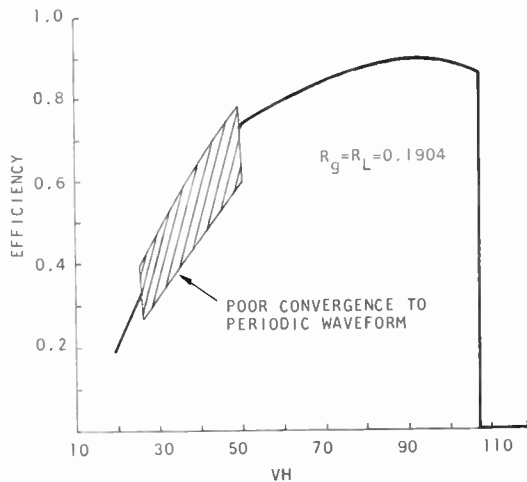

Fig. 4. Efficiency vs. bias of the bridge doubler for relatively high-Q input and output ($R_g = R_L = 0.1904$).

Then

$$a_k = b_k/N \qquad ......(17)$$

By explicit evaluation of the (complex) geometric sum in eqn. (16), one can show that a completely accurate evaluation of the coefficients is obtained from eqn. (17) as long as

$$N \geq 2M + 1 \qquad ......(18)$$

This result is well known as the 'sampling theorem'. For example, most of the calculations were done for one-degree steps so that $N = 360$. Thus, the Fourier analysis will be satisfactory if the output waveform can be 'adequately' represented by 178 harmonics of the fundamental.

The sums required for calculating the first twelve harmonics are accumulated at each step in the integration. In carrying out the calculation, a trigonometric table is initially calculated and stored. This table is used to produce the generator voltage, its derivative, and sine and cosine functions required for all of the harmonics.

5. Results

The fundamental frequency f was set equal to unity except when bandwidth was investigated. The varactor maximum elastance may also be chosen arbitrarily without loss of generality. A value $S_{max} = 4\pi$ was chosen for all calculations except for those involving imbalance. The average reactance at the fundamental frequency is $\frac{1}{2}S_{max}/(2\pi f)$ if the varactor spends equal time in the high elastance and low elastance states. For this particular condition, the average reactance at the fundamental frequency is unity.

For resonance at the fundamental frequency, an input inductance of $L_1 = 1/(2\pi)$ is indicated. Similarly, an output inductance of $L_2 = 1/(8\pi)$ was chosen.

It was originally expected that these element values, which are summarized in Fig. 3, would be altered by trial to optimize the efficiency. However, these values were used for nearly all of the calculations because it was found that the source and load resistances, and the bias, were much more significant variables.

For most of the calculations, the source and load resistances were equal to a single assigned value. Then efficiency was studied as a function of bias voltage (represented by V_H).

Results are shown in Fig. 4 for $R_g = R_L = 0.19$. From calculations with other values of source and load resistance, it is believed that this value of resistance is too low for best operation of the circuit. That is, the Q of the input and output circuits is higher than necessary for good efficiency. This case produces one of the more interesting results

obtained from the calculations. For V_H of 30 to 50 (perhaps also lower), convergence of the cycle-by-cycle efficiency to a steady value does not occur within the 20 cycles used in the standard calculations. In fact, convergence to 0.01% is not found within 30 cycles. The output waveform is shown in Fig. 5 for this high-Q case, for values of bias in and out of the troubled region. It is not certain whether the spurious outputs present when the varactors are under-biased are subharmonic or anharmonic. The distinction does not seem to be important since the behaviour of the multiplier would not be very satisfactory in either case. For excessive reverse bias, the multiplier ceases to operate since the voltage applied to the varactor in the steady state does not reach a value sufficient to reveal the non-linear nature of the varactor.
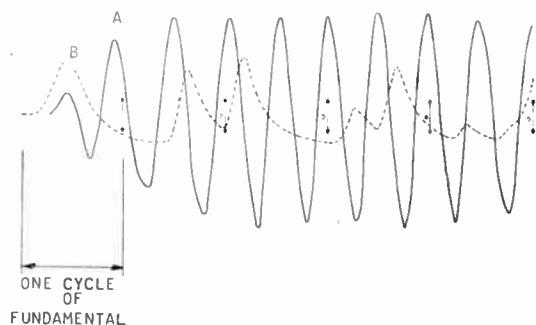

Fig. 5. Output current for $R_g = R_L = 0.19$.

| A | $V_H = 50.0$ |
| B | $V_H = 30.0$ |

Figure 6 shows the input and output waveforms for $V_H = 93$, giving peak efficiency of 0.89. For voltages that are only slightly in excess of the point where efficiency vanishes, some non-linear behaviour is achieved during the transient. An example is the waveform shown in Fig. 7.

The calculations permit one equally well to present the input-output power dependence for fixed bias.

Experimental results very similar to these calculations have frequently been observed for multipliers. The only difference is that in the experiment a hysteresis effect is associated with the over-biased shut-off (equivalent to underdriven condition for fixed bias). Spurious oscillations are noted for underbias or overdrive with fixed bias. The fact that such behaviour has been reproduced by theoretical calculations based on ideal varactors shows that new physical mechanisms in the varactor diode need not be postulated to explain the spurious oscillations. Effects such as

rectification and avalanche multiplication could lead to further complications, but are not necessary for the presence of oscillations. The abrupt fall-off in efficiency for over-bias is slightly exaggerated by the assumption of a piecewise-linear characteristic.
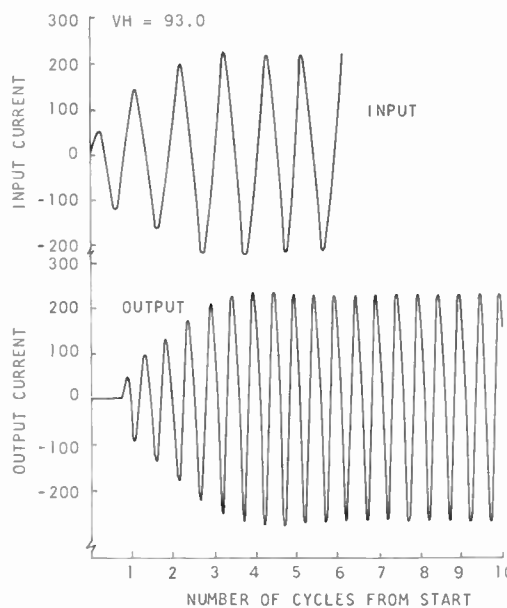

Fig. 6. Input and output currents for high-Q case $(R_g = R_L = 0.19)$.

However, the capacitance in the forward direction of a varactor can increase by as much as a thousand times for voltage change of 0.2 V. This voltage is a small fraction of a breakdown voltage of, say, 100 V or more. The piecewise-linear approximation seems
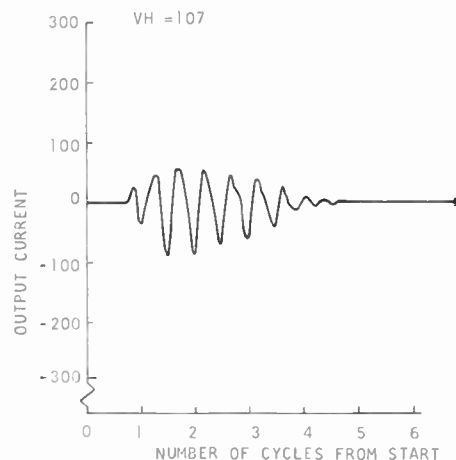

Fig. 7. Output waveform over biased $(R_g = R_L = 0.19)$.

to be the forthright way to face the nearly singular non-linearity possible in varactors.

Before considering frequency response, results for a more moderate value of series resistance will be given. For $R_g = R_L = 0.5$, the efficiency peak is at least 0.9435 at $V_H \simeq 40$. The output waveform is given in Fig. 8. The output waveforms for both this case and the optimally-biased high-Q case were
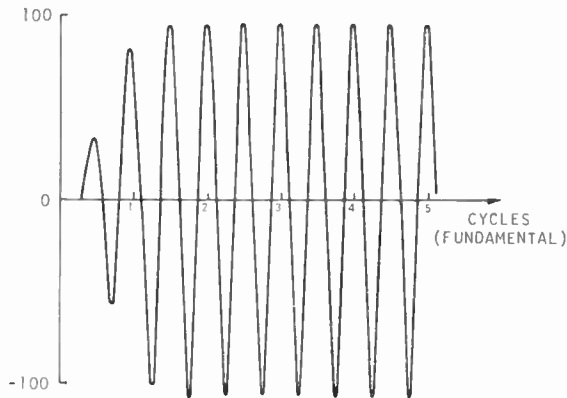

Fig. 8. Output waveform for $R_g = R_L = 0.5$, $V_H = 40$.

quite satisfactory. The distinction between the two cases lies more in the difficulties that are possible for variations in bias in the high-Q case, as well as the hoped-for better bandwidth with decreased Q. The maximum efficiency obtained for each value of R_g is shown in Fig. 9, along with the bias voltage V_H* required to produce this efficiency, as a fraction of the peak inverse voltage (p.i.v.) under steady-state conditions. The coefficient K in the equation[7]

$$P_{in} = \tfrac{1}{2}Kf_{in}C_{min}(V_B)^2 \qquad \ldots\ldots(19)$$

is also given, where the peak inverse voltage is used for V_B.

Some frequency response curves for the high-Q case are shown in Fig. 10 and for low-Q case in Fig. 11. In both cases, the curves are shifted substantially by adjustment of bias. Such behaviour is to be expected since the bias has a strong effect on the average elastance. The peculiar shapes of frequency response obtained in the high-Q case resemble some experimental observations on multipliers using swept-frequency inputs.

It is apparent that efficiency alone is not a meaningful figure of merit for a multiplier. High efficiencies have been obtained in the calculations under circumstances which are obviously not optimum in respect to frequency response and bias dependence. The main design consideration that seems to emerge from these

calculations is that an excessively high Q will lead to difficulties without significantly improving efficiency. Empirical variation of parameters with efficiency as the sole figure of merit will not protect against these difficulties; a deliberate effort to design for a low-Q circuit is recommended.

Suppression of fourth harmonic varies from 23 dB to 16 dB as R_g and R_L vary from 0.3 to 2.0. The sixth harmonic suppression varies from over 30 dB to 26 dB. Additional filtering would be necessary in many applications.

The peak inverse voltage during the transient generally did not exceed the steady state value by more than 10%, even for $R_g = R_L = 0.19$. However, for this high-Q case, shorting the output ($R_L = 0$) produced a peak inverse voltage 2.5 times the value for normal output loading ($R_L = R_g$).

The time required for doing the calculations is of some interest for assessing the feasibility of digital differential analysis. The program was written in *Fortran II*. A typical run of 20 cycles, with an integration interval of one degree at the fundamental frequency and the calculation of twelve output harmonics, requires about 20 seconds on the IBM 7094 Model II.
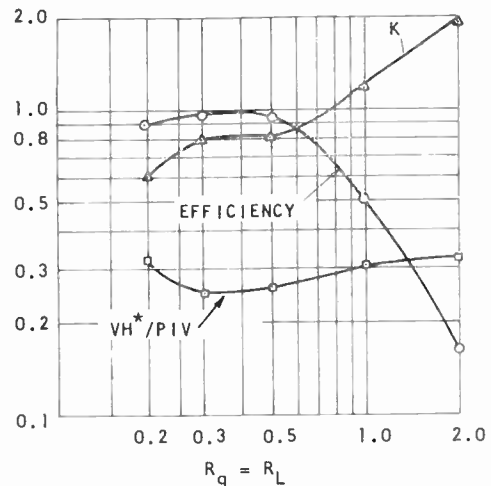

Fig. 9. Peak efficiency, input power constant, and bias fraction as a function of circuit loading.

(The large random-access memory of this machine is mostly idle during the calculation.) Substantial time savings can undoubtedly be made by using a coarser integration interval for most of the 'tune-up' calculations and then confirming the values with a final integration interval only for particularly interesting values of constants. A thorough assessment of accuracy was not undertaken. For most conditions,
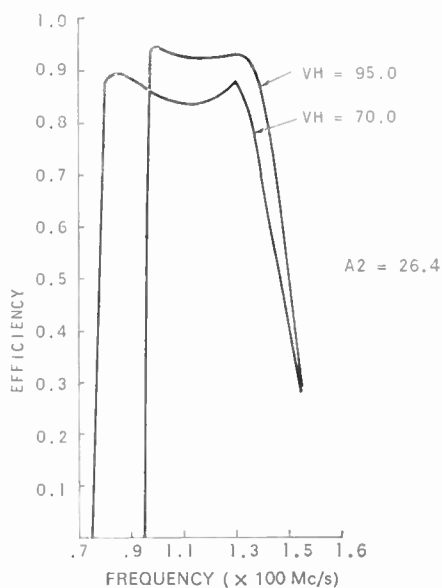
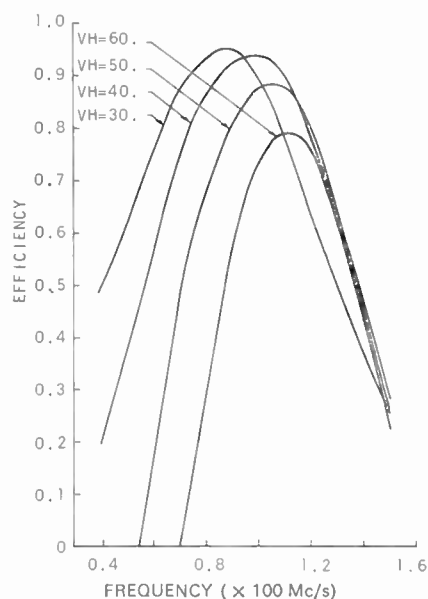**Fig. 10.** Frequency response for high-Q ($R_g = R_L = 0.19$).

Fig. 11. Frequency response for moderate-Q case ($R_g = R_L = 0.5$).

the same efficiency was obtained within 0.1% for integration intervals of one degree and five degrees. However, a few of the high-Q calculations gave results differing by as much as 1% in efficiency for one degree steps and one half degree steps. The accuracy for coarse intervals would undoubtedly be better except for the use, throughout the integration step, of the varactor voltage and elastance at the beginning of the step.

The program as written was designed mainly for the transient solution and was nevertheless satisfactory for studying the steady-state case. Considerable time advantage should be obtainable when only the steady-state case is of importance by using calculations for previous values of the parameters as starting points rather than the application of an initial transient. A programmed sweeping of the bias would be useful.

Before the calculations were carried out, it was thought that steady-state results could be obtained by having the efficiency optimized automatically by programmed instructions. The results have shown, however, that the criterion for optimization cannot be efficiency alone but must be a more sophisticated definition of proper multiplier performance.

6. Experiments

A doubler operating from 90 Mc/s to 180 Mc/s was constructed with four MA 4060AA varactor diodes ($C_0 = 145$ to 160 pF, $V_B = 90$ V, $R_s = 1.0$ ohm). The circuit differed from the one used in the analysis

in having adjustable tuning capacitors in series in both the input and output circuits.

An efficiency of 0.90 was observed with a c.w. input of 10 W and a 0.6 efficiency was measured with an input of 250 W peak one-microsecond pulses.

7. Conclusions

Calculations have shown that good efficiency and rapid transient response are compatible in a bridge doubler. The calculations reveal spurious oscillations for certain values of parameters. These calculations confirm previous beliefs that many experimentally observed difficulties are circuit design problems and do not depend upon effects as rectification and avalanche multiplication in the varactors. Design for the lowest-Q circuitry consistent with good efficiency is recommended. The computed results given here should be useful in design.

Optimization of multiplier design cannot be carried out in the laboratory or on the computer with efficiency as the sole figure of merit to be optimized. Requirements for margins of safety in bias and frequency, and possibly some definition of smoothness of frequency response, must be imposed to obtain non-trivial results.

It is believed that most of the important parameters of varactor multipliers can be effectively investigated by integration in the time domain, with the possible exception of noise. The digital computer seems to be a satisfactorily economical tool for the calculations

but the probable advantages of analogue computation in exploring the effects of many variables cannot be denied.

8. Acknowledgements

This work has been aided by discussions with Prof. R. P. Rafuse and by computer programming by Mrs. B. Chicklis. The multiplier experiments were carried out by Mr. B. Corsetti under the direction of Mr. H. H. Cross.

The investigation was supported by Electronic Technology Laboratory, Aeronautical Systems Div., AFSC, under contract AF33(657)–11655.

9. References

1. D. L. Hedderly, 'An analysis of a circuit for the generation of high-order harmonics using an ideal nonlinear capacitor', *Trans. Inst. Radio Engrs on Electron Devices*, **ED-9**, No. 6, pp. 484–91, November 1962.

2. D. L. Hedderly, 'An exact analysis of a parametric subharmonic oscillator', *Trans. I.R.E.*, **ED-10**, No. 3, pp. 134–42, May 1963.

3. D. Leenov and A. Uhlir, Jr., 'Generation of harmonics and subharmonics at microwave frequencies', *Proc. I.R.E.*, **47**, pp. 1724–9, October 1959.

4. T. Isobe and T. Miyakawa, 'Frequency multipliers consisting of NC and NR elements', *Proc. I.E.E.E.*, **53**, No. 4, p. 394, April 1965.

5. P. Penfield, Jr., and R. P. Rafuse, 'Varactor Applications', p. 337, (M.I.T. Press, Cambridge, Mass., 1962).

6. A. Grayzel, 'The Bandwidth of the Abrupt-Junction Varactor Frequency Doubler', M.S. thesis, M.I.T., Cambridge, Mass., 1963.

7. A. Uhlir, Jr., 'Similarity considerations for varactor multipliers', *Microwave J.*, **5**, pp. 55–9, July 1962).

STANDARD FREQUENCY TRANSMISSIONS
(Communication from the National Physical Laboratory)
Deviations, in parts in 10¹⁰, from nominal frequency for December 1965

December 1965	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 24-hour mean centred on 0300 U.T.	December 1965	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 24-hour mean centred on 3303 U.T.
1	− 150·6	− 150 6	+ 1·1	17	− 150·0	− 150 5	− 2·1
2	− 150 9	− 151·0	+ 1·2	18	− 150·1	—	− 2·2
3	− 150·9	− 151·2	+ 0·9	19	− 151·2	—	− 2·0
4	—	—	+ 1·5	20	− 150·5	− 150·8	− 1·7
5	—	—	—	21	− 150 5	− 148·7	− 1·7
6	—	—	—	22	− 148·7	− 150 6	− 1·3
7	− 151·4	− 151·6	+ 0·7	23	− 150 3	− 150·3	− 1·3
8	− 151·0	− 151·2	− 0·8	24	− 150·9	− 150 6	− 1·2
9	− 150·4	− 151·0	− 1·0	25	− 150·5	− 149 9	− 0·3
10	− 150 4	− 150 8	− 1·9	26	− 150·1	− 150·4	− 0·1
11	− 150 9	− 151·6	− 2·1	27	− 150·1	− 150·1	—
12	− 150·8	− 150·9	− 2·1	28	− 149·8	—	—
13	− 151·1	− 152·7	− 2·2	29	− 149·9	− 151 0	+ 0·1
14	− 151·6	− 150 8	− 2·4	30	− 150·3	− 149·4	+ 0·5
15	− 151·8	− 151·2	− 1·7	31	− 150 4	− 150·6	+ 0·5
16	− 151·1	− 150·1	− 2·0				

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

A Study of Two-terminal Current-controlled Negative Resistance Devices with the Aid of an Analogue

By

R. S. C. COBBOLD, Ph.D.,†

H. N. MAHABALA, M.Sc.,
Ph.D.‡

AND

M. N. S. SWAMY, M.Sc., Ph.D.§

Summary: Some fundamental aspects of current- and voltage-controlled negative resistance devices are considered, in particular the relation of the physical mechanism involved to the equivalent circuit. Further, a unified stability criterion is presented which is applicable to both types of device. A versatile analogue of a current-controlled device is described which is compatible with a voltage-controlled analogue previously described. The high-frequency properties of this analogue are examined both experimentally and theoretically to determine the variation of the equivalent circuit parameters with frequency. With the aid of the analogue two circuit configurations are examined: a bistable trigger circuit, and a four-stable-state circuit obtained by the addition of a voltage-controlled analogue. It is believed that the techniques described enable circuits to be investigated at relatively low speeds, thus facilitating the optimization and understanding of a given configuration.

1. Introduction

Probably the earliest practical exploitation of the properties of negative resistance (n.r.) devices was the use of the Poulsen arc in about 1912 for arc transmitters, but it is only in the last 10 years, with the discovery of various semiconductor and cryogenic devices that these n.r. properties have attracted the interest of the circuit designer. Avalanche injection diodes, tunnel diodes, the cryosar, superconducting tunnel effect devices—to name just a few—are some of the more recent two-terminal n.r. devices whose properties and applications are receiving considerable attention.

In general, negative resistance devices can be classified into two categories,[1,2] voltage controlled and current controlled, according to whether they are multi-valued with respect to current or voltage; an exception being a device which exhibits both voltage and current multi-valuedness, in which case it can usually be synthesized from a current- or voltage-controlled device and some externally appended passive elements. Much work has been done recently on voltage-controlled devices, especially tunnel diodes, and through the principle of duality much of the knowledge can be applied to current-controlled devices. This transfer, and further knowledge, can be facilitated by experimental studies on current-controlled devices. The purpose of this paper is to discuss the properties of two-terminal current-con-

trolled negative resistance devices, to describe the design and application of a versatile analogue, and to discuss its use in conjunction with a voltage controlled analogue.

2. Equivalent Circuits

Any physical device has associated reactances and it is necessary to determine these in order to find the equivalent circuit. This can be done either by the study of the physical properties of the device, or by terminal measurements, or both. In the case of two terminal devices which can be stably biased in the negative resistance region and which contain no active sources, it is shown theoretically in Section 5 that the simplest possible equivalent circuits for voltage- and current-controlled devices are those shown in Figs. 1(a) and 1(b) respectively. They can be deduced in a qualitative manner by proving that stable biasing in the n.r. region is impossible if the reverse were true. Referring to Fig. 1(b), suppose that for the current-controlled device a capacitance is present next to the pure n.r. as shown in the equivalent circuit of Fig. 1(a), and that the device is biased by a current source at the point O. If a small disturbance causes the charge on C to increase, the n.r. current will decrease and this in turn will cause a further increase of the current into C. The device thus moves away from O along the path OA. In a similar manner a decrease in the charge on C causes the device to move along OB, again rendering the bias point unstable. Since we know that the device can be stably biased from a current source it is evident that the equivalent circuit must be of the form shown in Fig. 1(b). In a similar manner it can be shown that for voltage-controlled negative resistance, the simplest equivalent circuit consistent with stable biasing is that shown in Fig. 1(a).

† Department of Electrical Engineering, University of Saskatchewan, Saskatoon, Canada.

‡ Formerly at the University of Saskatchewan; now with the Computer Centre, Indian Institute of Technology, Kanpur, India.

§ Formerly at the University of Saskatchewan; now with the Department of Electrical Engineering, Nova Scotia Technical College, Halifax, Canada.

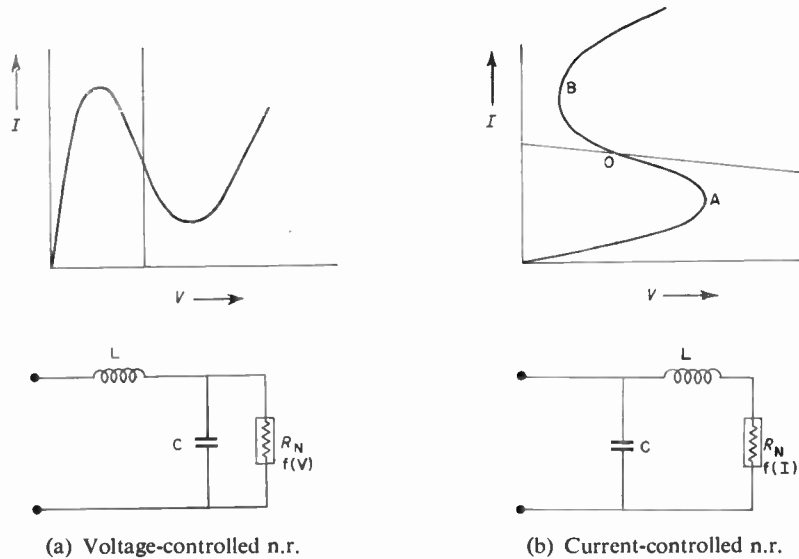(a) Voltage-controlled n.r.

(b) Current-controlled n.r.

Fig. 1. Characteristics and equivalent circuits of negative resistance configurations.

One can perhaps attribute the existence of a series inductance as the first reactance after the negative resistance in the case of a current-controlled device, to the time delays associated with the physical mechanism involved. If one postulates the existence of an ideal current-controlled n.r., the time delay can be thought of as the delay in the external current change reaching the ideal device, and this can most simply be represented by the circuit of Fig. 1(b). For the more general case, one can assume a transmission line ending with a series inductance between the terminals and the ideal n.r., of which Fig. 1(b) is only a special case. Some specific current-controlled devices taken from those listed in Table 1 will now be examined to determine the physical origin of the time delay which gives rise to the inductive first element.

The thermistor[6, 7, 19, 20] demonstrates a n.r. region through ohmic heating of the semiconductor material which causes a decrease in resistivity. The application of a positive current step to the device biased in the n.r. region causes the terminal voltage to increase at a rate governed by the terminal capacitance and iso-thermal resistance at the bias point. This, in turn, produces heating and, due to the thermal lag, eventually results in a terminal voltage less than the initial value. The change in terminal current thus only slowly affects the pure n.r. element.

For the avalanche injection diode,[11, 12] an applied current step causes the avalanche current to increase at a rate dependent on the junction capacitance and bulk resistance at the bias point. The increase in avalanche current injects further minority carriers into the bulk material, thus reducing the resistivity. Because the voltage drop across the space charge region only increases by a small amount, the decrease

in resistivity effectively decreases the terminal voltage thereby generating a n.r. characteristic. Here again the change in terminal current reaches the pure n.r. after a time delay.

Table 1

Some two-terminal direct current- and voltage-controlled n.r. devices and effects

Voltage controlled (S-type)	Current controlled (N-type)
Gas discharge	Gas discharge
Tunnel diode[3]	Thermistor[6, 7]
Superconducting tunnel effect[4]	Cryosar[8, 9, 10]
Impurity barrier electron	Avalanche injection diodes[11,12]
Tunnel effect[5]	Double injection diodes[13]
Thin anodic oxide film effect[14]	Oxide film and ceramic devices[15, 16, 17]
	Electroluminescent diodes[18]

It is also of interest to examine the mechanism of a voltage-controlled n.r. device, such as a tunnel diode,[3] for a time delay in the applied terminal voltage from reaching the ideal device. When a voltage step is applied to the terminal, the junction field can only rise slowly due to the inherent lead inductance and depletion capacitance. Since the tunnelling phenomena involved is extremely fast the reduction in terminal current occurs virtually immediately after the field is set up. Thus, the time delay due to parasitic elements forms a major portion of the total time delay involved in the terminal voltage reaching the junction. This is in contrast to the situation for current controlled devices wherein the mechanism itself limits the speed.

3. A Current-controlled Analogue

Several current-controlled analogues have been reported in the literature,[21, 22] but all lack adequate flexibility. The design of a flexible current-controlled n.r. analogue was undertaken to facilitate an investigation into the circuit properties of current-controlled n.r. devices. Such an analogue, with a bandwidth extending from d.c. to several hundred kilocycles, has a number of advantages for such studies. First, since the bandwidth is limited to fairly low frequencies, the analogue can be readily biased in the n.r. region and difficulties caused by the presence of parasitic oscillations are minimized. Second, the analogue of a given device can be scaled to operate at sufficiently low frequencies so that changes in circuit values can be readily effected, and waveforms readily observed on a low-frequency oscilloscope. Third, since the parameters of the analogue can be easily varied, the tolerance limit of a given circuit configuration, to device parameter changes, can be experimentally determined. The analogue, whose circuit diagram is shown in Fig. 2, was designed for use in conjunction with a voltage-controlled n.r. analogue previously described.[23]

The n.r. region is generated by the complementary transistor pair TR1 and TR2, which are connected in a feedback arrangement. Suppose an increasing current is applied to the input terminals in the direction indicated by the arrow. When the input current is
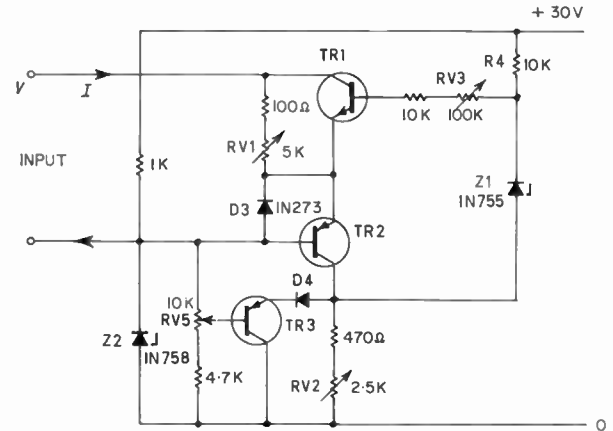

Fig. 2. Circuit diagram of the current controlled analogue. Resistor RV1 controls the positive resistance portion, RV2 the n.r. portion, RV3 and RV5 the peak and valley points respectively.

zero, the emitter-base junction of TR1 is reverse biased by a voltage dependent on the setting of RV2 so that the feedback loop between TR1 and TR2 is open circuit. The input current, for values less than the peak current I_p, flows through resistor RV1, transistor TR2 and out of the Zener diode Z2 so that the slope of the initial positive region can be controlled by RV1. The peak current is reached when the collector of TR2 has risen sufficiently in potential to cause TR1 to enter the active region and therefore
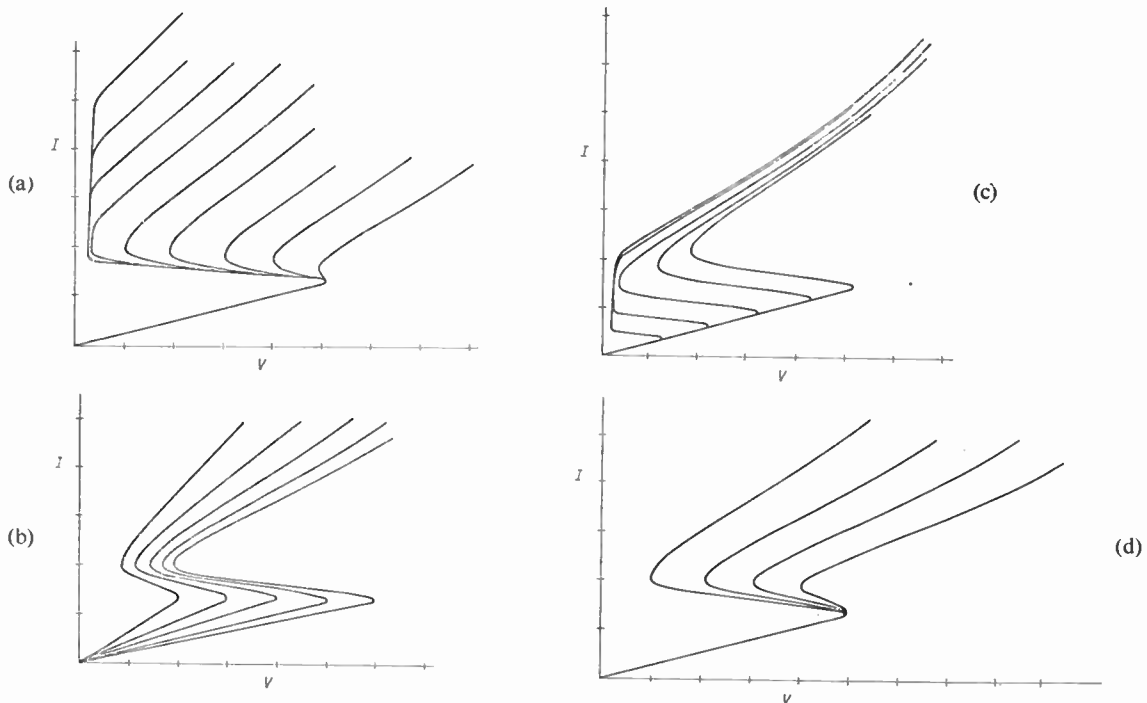

Fig. 3. Families of static characteristics obtainable with the analogue. The controls predominantly responsible for the families of characteristics illustrated are: (a) RV5, (b) RV1, (c) RV2, (d) RV3. Axes: $V - 1 \cdot 0$ V/div., $I - 2 \cdot 0$ mA/div.

close the feedback loop. A further increase in the input current causes TR1 to conduct more heavily, diverting such a portion of the input current away from RV1 as to actually reduce the input voltage and thus produce a negative slope resistance. This incremental negative resistance can be controlled by altering the loop gain by means of resistor RV3. The valley current I_v, is reached when the collector of TR2 is caught by diode D4 and the loop gain is effectively reduced to zero. The magnitude at which this occurs is under the control of RV3 which, together with TR3 form a convenient variable voltage supply. Any further increase in input current now flows through RV1, and this results in a second positive resistance region. The direct current through the Zener diode Z2 governs the maximum input current allowable. When the sense of the input current is reversed, diode D3 provides a conduction path through RV1.

Figure 3 shows some d.c. characteristics of the device obtained with various settings of the parameter controls. It should be pointed out that there is a certain amount of interaction between the various controls and, as such, more than one control has to be altered to change a parameter. The characteristics are also dependent on the supply voltage which consequently should be stabilized.

4. Input Impedance and Equivalent Circuit of the Analogue

For a given bias point, the input impedance of the analogue is frequency-dependent due to the finite transistor cut-off frequency and reactive parasitic elements. It is necessary to know this dependence in order to assess the frequency range over which the elements of analogue equivalent circuit can be considered as frequency-independent.

A number of approximations can be made which serve to considerably simplify the analysis. It should be noted in Fig. 2 that TR1 operates in the common emitter mode and is fed from a high impedance source, while TR2 operates in the common base mode. Hence, for frequencies small compared to the cut-off frequencies of TR1 and TR2, it can be assumed that in the n.r. region, transistor TR1 will determine the input impedance. The collector resistances of TR1 and TR2 are assumed to be large compared to RV1 and the parallel combination of RV2 and R4 respectively. Further, if the short circuit common emitter current gain of TR2 is large, then the circuit of Fig. 2 reduces to the equivalent circuit of Fig. 4, wherein C is the total collector to base capacitance of TR1.

The input impedance Z_{in} is given by:

$$Z_{in} = \frac{i_1 R_1}{i_1 + i_2 + i_3} \qquad \text{......(1)}$$

Now

$$i_3 = j\omega C R_1 i_1 \qquad \text{......(2)}$$

and

$$i_b = \frac{i_2}{\beta} = \frac{R}{R_3}(i_1 + i_2 + i_3) + i_3 \qquad \text{......(3)}$$

where $R = \dfrac{R_2 R_4}{R_2 + R_4}$ and β is the complex short-circuit common-emitter current gain of TR1.

Substituting (2) and (3) into (1) gives,

$$Z_{in} = \frac{R\left(1 + \beta\dfrac{R}{R_3}\right)}{1 + j\omega C R_1(1 + \beta)} \cdot \qquad \text{......(4)}$$

Further if β is expressed in the form

$$\beta = \frac{\beta_0}{1 + j\dfrac{\omega}{\omega_\beta}} \qquad \text{......(5)}$$

where β_0 is the low-frequency short-circuit common-emitter current gain and ω_β is the β cut-off frequency, then eqn. (4) can be written as

$$Z_{in} = R_{in} + j\omega L_{in} \qquad \text{......(6)}$$

in which

$$R_{in} = \frac{R_1}{D}\left[1 - \beta_0\frac{R}{R_3} + \beta_0 C R_1\frac{\omega^2}{\omega_\beta}\left(1 + \frac{R}{R_3}\right) + \left(\frac{\omega}{\omega_\beta}\right)^2\right] \qquad \text{......(7)}$$

and

$$L_{in} = \frac{R_1}{D}\left[C R_1(1 + \beta_0)\left(\frac{R}{R_3}\beta_0 - 1\right) + \frac{1}{\omega_\beta}\beta_0\frac{R}{R_3} - \left(\frac{\omega}{\omega_\beta}\right)^2 C R_1\right] \qquad \text{......(8)}$$

where

$$D = 1 + \frac{2\omega^2}{\omega_\beta}\beta_0 C R_1 + \left(\frac{\omega}{\omega_\beta}\right)^2(1 + \omega^2 C^2 R_1^2) + \omega^2 C^2 R_1^2(\beta_0 + 1)^2 \qquad \text{...(9)}$$
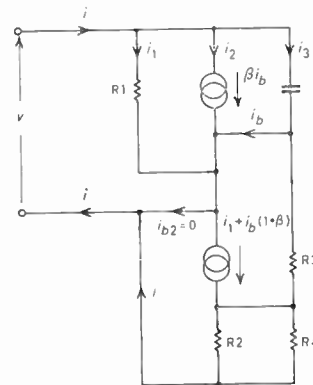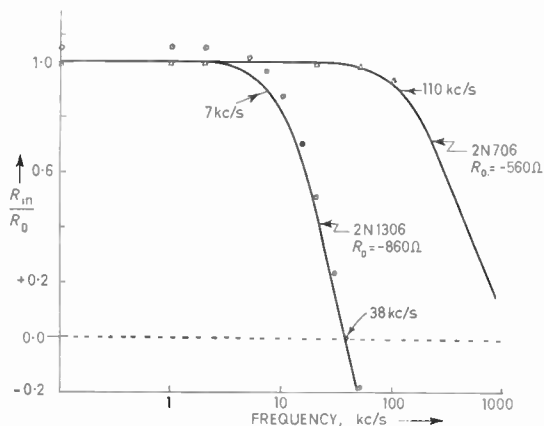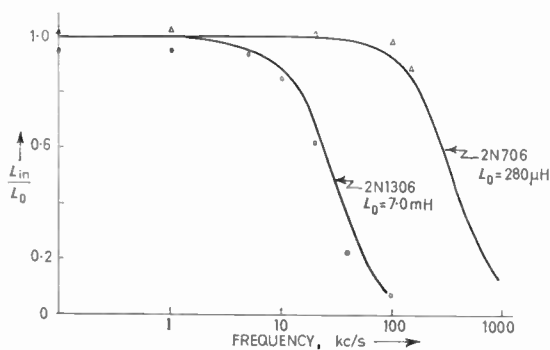

Fig. 4. Approximate equivalent circuit of Fig. 2 in the n.r. region.

(a) Normalized plot of R_{in} as a function of frequency.

(b) Normalized plot of L_{in} as a function of frequency.

Fig. 5. Normalized graphs of L_{in} and R_{in} theoretically computed, compared with experimental results for two transistors (TR1), 2N1306 and 2N706.

The low-frequency values of R_{in} and L_{in}, namely R_0 and L_0, are obtained by neglecting the terms containing ω in eqns. (7) and (8), giving

$$R_0 = R_1\left(1 - \beta_0 \frac{R}{R_3}\right) \qquad \text{......(10)}$$

$$L_0 = R_1\left[CR_1(1+\beta_0)\left(\frac{\beta_0 R}{R_3} - 1\right) + \frac{\beta_0 R}{\omega_\beta R_3}\right] \quad \text{...(11)}$$

It is interesting to note that L_0 depends amongst other things upon ω_β, and that for R_0 to be negative, $\beta_0 R/R_3$ should be greater than unity. Both R_{in} and L_{in} decrease in magnitude with increasing frequency, and the frequency at which R_{in} changes sign is given by

$$\omega_c = \omega_\beta \sqrt{\frac{\beta_0 R/R_3 - 1}{1 + \omega_\beta \beta_0 CR_1(1 + R/R_3)}} \quad \text{......(12)}$$

It is desirable to make the value of ω_c fairly large, which can be achieved by increasing either ω_β or $\beta_0 R/R_3$. However, increasing $\beta_0 R/R_3$ also increases L_{in} which, because of its frequency dependence, is undesirable.

Using a Maxwell bridge, experimental measurements of the variation of L_{in} and R_{in} with frequency were made for transistors of widely differing cut-off frequencies for TR1. In the first case, to facilitate the verification of eqns. (7), (8) and (12), a low-frequency transistor (2N1306) was chosen for TR1. In the second case, a transistor with an alpha cut-off frequency of 700 Mc/s was used, and was retained for subsequent studies using the analogue. The calculated and measured values for R_{in} and L_{in} for the above two cases have been plotted in Figs. 5(a) and 5(b).

Since R_{in} decreases with increasing frequency, it is necessary to limit the highest frequency at which the analogue is operated to a frequency less than that at which $|R_{in}|$ falls by 10%. This limiting frequency, which increases with ω_c, is in the neighbourhood of 100 kc/s for the high frequency transistor used. Even though L_{in} decreases only slightly over this frequency range, it is necessary to include an external series inductance several times larger than L_0 in order to make the equivalent circuit frequency-independent. Under these circumstances, the small signal equivalent circuit of the analogue can be taken as consisting of a series combination of an inductance and a negative resistance, both of which are independent of frequency.

5. Unified Stability Criteria

A unified stability criterion[24] which applies to both the current and voltage controlled devices for a linearized R_N will now be discussed. The common equivalent circuit is shown in Fig. 6 where R_v is positive and R_c negative for a current-controlled device, and R_v is negative and R_c is positive for a voltage-controlled device.
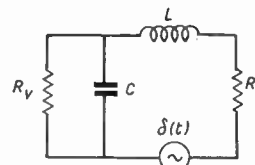

Fig. 6. Equivalent circuit for determining the stability of both current- and voltage-controlled n.r. devices (R_v is negative and R_c positive for voltage-controlled, and R_c is negative and R_v positive for current-controlled).

The voltage response v for a δ-function of voltage in series with R_c is given by

$$v = h_1 e^{\lambda_1 t} + h_2 e^{\lambda_2 t} \qquad \text{......(13)}$$

where λ_1 and λ_2 are given by

$$\lambda_1, \lambda_2 = \frac{1}{2LC}(A \pm \sqrt{B}) \qquad \text{......(14)}$$

where

$$A = -\left(\frac{L}{R_v} + CR_c\right) \qquad \ldots\ldots(15)$$

and

$$B = \left(\frac{L}{R_v} + CR_c\right)^2 - 4LC\left(1 + \frac{R_c}{R_v}\right) \ldots\ldots(16)$$

Replacing $R_c\sqrt{\frac{C}{L}}$ by x and $\frac{R_c}{R_v}$ by y and simplifying (16), it can be shown that B has the same sign as

$$(x^2 - y + 2x)(x^2 - y - 2x) \qquad \ldots\ldots(17)$$

The response is exponential or sinusoidal according as B is positive or negative. In other words, the response is sinusoidal if

$$\begin{array}{ll} x > 0 \quad \text{and} \quad x^2 - 2x < y < x^2 + 2x \\ x < 0 \quad \text{and} \quad x^2 + 2x < y < x^2 - 2x \end{array} \ldots\ldots(18)$$

or

and is exponential otherwise.

The growth or decay of the response depends on the polarities of R_c and R_v and, hence, several cases arise.

(a) $R_c > 0$; $R_v > 0$

From (15) and (16), $A < 0$ and $\sqrt{B} < |A|$ and, hence, the response decays.

(b) $R_c < 0$; $R_v < 0$

From (15), $A > 0$ and, hence, the response grows.

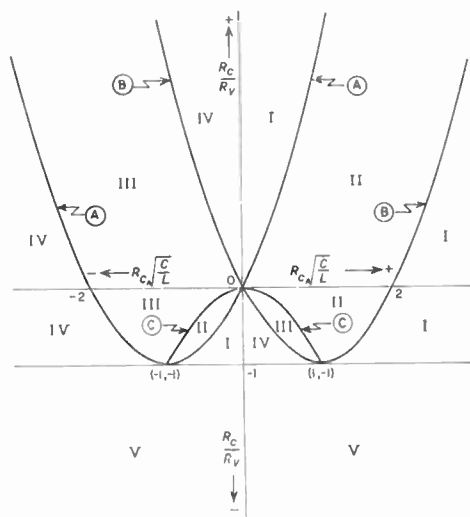(c) $R_c > 0$; $R_v < 0$ (voltage-controlled device).

Fig. 7. Graphical representation of the unified linear stability criterion applied to both current-controlled and voltage-controlled negative resistance.

A Plot of $y = x^2 + 2x$; B Plot of $y = x^2 - 2x$;

C $y = x^2$, where $x = R_c\sqrt{C/L}$ and $y = R_c/R_v$.

I—Region of exponential decay; II—Region of sinusoidal decay; III—Region of sinusoidal growth; IV—Region of exponential growth; V—Region of exponential growth; (switching region).

There are two sub-cases to consider depending on the magnitude of R_c/R_v, which is negative.

(i) $\qquad \left|\dfrac{R_c}{R_v}\right| > 1$

Then, $B > 0$ and $\sqrt{B} > |A|$ and the response grows

(ii) $\qquad \left|\dfrac{R_c}{R_v}\right| < 1$

Then, $|A| > \sqrt{|B|}$ and the response grows or decays according as A is positive or negative. From eqn. (15)

$$A = -\frac{C}{x^2}R_c(y + x^2) \qquad \ldots\ldots(19)$$

Since $R_c > 0$, $A \gtrless 0$ according as $y \lessgtr -x^2$.
Hence the response grows or decays according as

$$y \lessgtr -x^2 \qquad \ldots\ldots(20)$$

(d) $R_c < 0$; $R_v > 0$ (current-controlled device); then both x and R_c/R_v are negative and there are two sub-cases.

(i) $\qquad \left|\dfrac{R_c}{R_v}\right| > 1$

Then $B > 0$ and $\sqrt{B} > |A|$ and the response grows.

(ii) $\qquad \left|\dfrac{R_c}{R_v}\right| < 1$

Then $|A| > \sqrt{B}$ and the response grows or decays according as A is positive or negative. Noting that R_c is negative it follows from eqn. (19), that the response grows or decays according as
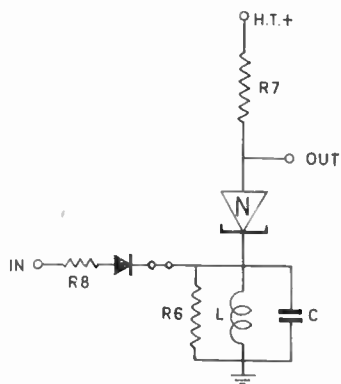
$$y \gtrless -x^2 \qquad \ldots\ldots(21)$$

The various sub-cases and the corresponding condition for the response to be exponential or sinusoidal and to grow or decay can be represented graphically as shown in Fig. 7. The fourth quadrant corresponds to the voltage-controlled case and the third to the current-controlled case.
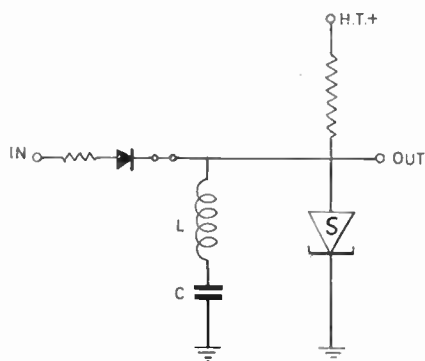
The graphical representation of the criteria in Fig. 7 can also be used to discuss the stability of the various quiescent states, with both R_c and R_v positive, obtained when the n.r. device is biased from either a linear or non-linear load.

It is observed experimentally that the biasing source resistance required to bias a current-controlled device in the n.r. region is higher than the n.r. of the device itself at the bias point. It is evident from Fig. 7 that this is the case only if the equivalent circuit is that shown in Fig. 1(b), thus justifying the existence of an inductance as the first reactance next to the n.r.

The stability criterion for a current-controlled case was verified by the use of the analogue; the method of exciting various regions is similar to that used for the voltage-controlled case.[23]

(a) Basic current controlled binary counter circuit due to Reeves.[25]

(b) Dual of the Reeves circuit.

Fig. 8. The Reeves circuit.

6. Applications

An investigation of some simple switching circuits was undertaken to illustrate some applications of the current controlled analogue. In addition the combination of a current- and a voltage-controlled device to produce four stable states was also demonstrated.
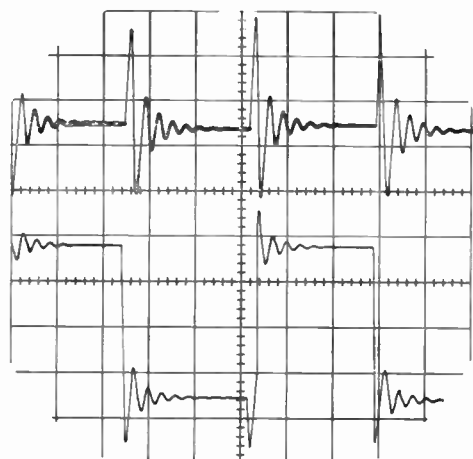
6.1. Trigger Circuits

The multi-valuedness of the *I-V* characteristics of a current-controlled n.r. device can be utilized to perform bistable switching operations in much the same manner as a tunnel diode is used. In fact, it is interesting to note that as early as 1954, Reeves[25] used avalanche injection diodes (positive gap diode) for a variety of switching circuits including a binary counter, the basic circuit of which is shown in Fig. 8(a). The incoming pulse sets up ringing in the parallel tuned circuit formed by *L* and *C*, thereby generating both a positive and negative pulse. If the load line is suitably chosen, then the negative oscillation resulting from the ringing causes the device to switch to the high-current state while, if the device is already in the high-current state, the positive oscillation effects a transition to the low-current state. The circuit

operation was demonstrated with the aid of the analogue and the resulting switching waveforms and trajectories are shown in Fig. 9.
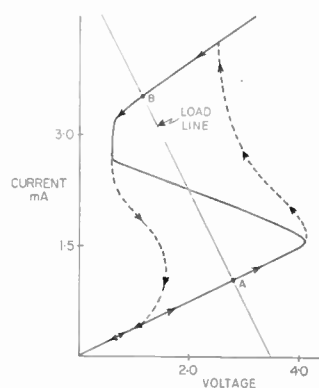
The dual of Reeves' circuit, shown in Fig. 8(b), was also demonstrated using the voltage-controlled analogue,[23] showing that the principle of duality can be very useful in transforming circuit techniques developed for current controlled devices to voltage-controlled devices or vice versa.

6.2. Combination of a Voltage- and Current-controlled Device

Some interesting circuit properties result when combinations of voltage- and current-controlled devices are considered.[26] One such system, shown
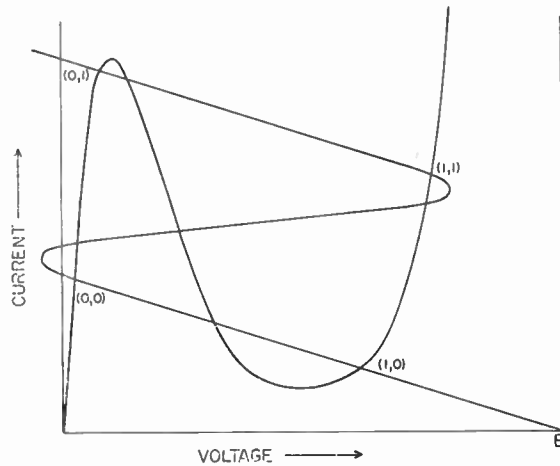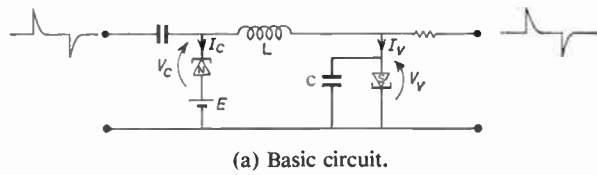

(a) *top* Voltage across C, 1·0 V/div and 1·0 ms/div.
bottom Current through analogue, 0·75 mA/div. and 1·0 ms/div.

(b) Switching trajectories. X-Y display of current through and voltage across the analogue, shown superimposed on the d.c. characteristic. The major portions of the switching trajectories from A to B and vice versa are shown.

Fig. 9. Switching waveforms of the Reeves[25] circuit shown in Fig. 8(a) demonstrated with the aid of the analogue for R7 = 665 Ω, R6 = ∞, R8 = 1·0 kΩ, L = 10 mH, C = 0·22 μF, and a 15-V impulse applied to the input.

(a) Basic circuit.

(b) Static characteristics of both devices shown superimposed to illustrate the nine possible quiescent states of the system, only four of which are stable.

Fig. 10. The combination of a voltage and current controlled device to produce four stable states.

in Fig. 10(a), has nine possible states as illustrated in Fig. 10(b). Each state can be examined for stability using Fig. 7, and it can be shown that the states (0,0), (0,1), (1,0) and (1,1) are unconditionally stable for small disturbances. Two inputs are used, one a voltage trigger changes the state of the current-controlled element; the other, a current trigger, acts

on the voltage-controlled element. Short term buffering between the two inputs is provided by the inductance. By a proper choice of the input pulses, the circuit can be made to move through the four stable states in a cyclic manner. Figures 11(a) and 11(b) show the switching trajectories of both the current- and voltage-controlled analogues.

7. Conclusions

While only two examples of the use of a current-controlled analogue have been given, many more will probably occur to the reader. Such studies, based on the analogue, provide a very good insight into the circuit operation, and enable simple empirical techniques to be developed for optimizing a given circuit configuration. In addition, by using a compatible voltage-controlled analogue, novel circuit arrangements such as the one shown in Fig. 10(a) can also be simply investigated.

The educational value of the analogue is also worth mentioning. Since no high-speed equipment is necessary for the experiments, there is no risk of the students causing damage to expensive high-speed equipment. In fact, the standard pulse generators and dual beam, X-Y display oscilloscopes normally available in undergraduate laboratories, are quite satisfactory for conducting switching studies. Often the use of extensive mathematical analysis, even with the aid of digital computers, cannot provide an insight into the operation of a non-linear circuit that experimentation using an analogue can provide.

8. Acknowledgments
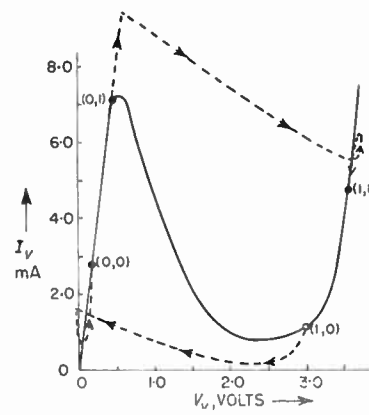
(a) $I_c - V_c$ trajectory,

(b) $I_v - V_v$ trajectory.
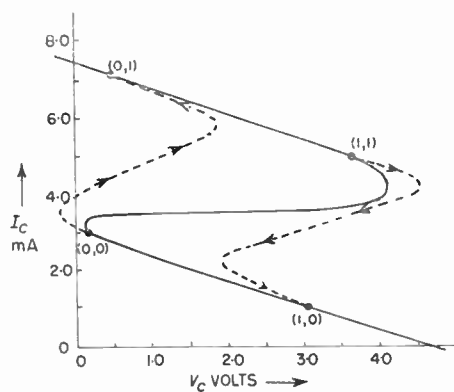
Fig. 11. Switching trajectories of the circuit shown in Fig. 10(a) with L = 400 mH, C = 0·22 μF.

The paper is based, in part, on the Ph.D. thesis submitted to the University of Saskatchewan by H. N. Mahabala, who wishes to thank the National Research Council for the award of a scholarship.

9. References

1. E. W. Herold, 'Negative resistance', *Proc. Inst. Radio Engrs*, **23**, p. 1201, October 1935.

2. G. D. Sims and I. M. Stephenson, 'The negative resistance concept in modern electronics', *J. Electronics Control*, **9**, p. 349, 1960.

3. H. S. Sommers, 'Tunnel diodes as high-frequency devices', *Proc. Inst. Radio Engrs*, July 1959, **47**, p. 1201.

4. I. Giaever and K. Megerle, 'The superconducting tunnel junction as an active device', *Trans. I.R.E. on Electron Devices*, ED-9, p. 459, November 1962.

5. B. K. Ridley and R. G. Pratt, 'A bulk differential negative resistance due to electron tunneling through an impurity potential barrier', *Physics Letters*, **5**, p. 300, 1st May 1963.

6. J. A. Becker, C. B. Green and G. L. Pearson, 'Properties and uses of thermistors—thermally sensitive resistors', *Bell Syst. Tech. J.*, **26**, p. 170, 1947.

7. R. E. Burgess, 'The turnover phenomenon in thermistors and in point-contact germanium rectifiers', *Proc. Phys. Soc.*, B, **68**, p. 908, 1955.

8. A. L. McWhorter and R. H. Rediker, 'The cryosar—a new low-temperature computer component,' *Proc. I.R.E.*, **47**, p. 1207, July 1959.

9. R. H. Rediker and A. L. McWhorter, 'Low temperature semiconducting computing elements', 'Solid State Physics in Electronics and Telecommunications', Vol. 2, p. 939 (Academic Press, New York, 1960).

10. H. Izumi, 'The silicon cryosar', *Proc. I.R.E.*, **49**, p. 1313, August 1961 (Letter).

11. J. B. Gunn, 'Avalanche injection in semiconductors', *Proc. Phys. Soc.*, B., **69**, p. 781, 1956.

12. A. F. Gibson, 'The theory of avalanche injection diodes', *Proc. Instn Elect. Engrs*, **106B**, Supplement No. 15, 1959, (I.E.E. Paper No. 3098E, May 1959).

13. N. Holonyak, 'Double injection diodes and related DI phenomena in semiconductors', *Proc. I.R.E.*, **50**, p. 2421, December 1962.

14. T. W. Hickmott, 'Low-frequency negative resistance in thin anodic oxide films', *J. Appl. Phys.*, **33**, p. 2669, September 1962.

15. D. V. Geppert, 'A new negative-resistance device', *Proc. I.E.E.E.*, **51**, p. 223, January 1963 (Letter).

16. K. L. Chopra, 'Current controlled negative resistance in thin niobium oxide films', *Proc. I.E.E.E.*, **51**, p. 941, June 1963 (Letter).

17. S. Lo, 'Negative resistance in copper-doped ferric oxide ceramic', *Proc. I.E.E.E.* **52**, p. 609, May 1964 (Letter).

18. K. Weiser and R. S. Levitt, 'Electroluminescent gallium arsenide diodes with negative resistance', *J. Appl. Phys.*, **35**, p. 2431, 1964.

19. R. E. Burgess, 'The a.c. admittance of temperature-dependent circuit elements', *Proc. Phys. Soc.*, B, **68**, p. 766, 1955.

20. R. E. Burgess, 'Negative resistance in semiconductor devices', *Canadian J. Phys.*, **38**, p. 369, 1960.

21. C. D. Todd, 'Transistor tunnel diode with N-type negative resistance', *Trans. Amer. Inst. Elect. Engrs*, **81**, Part I, p. 284, 1962.

22. A. H. Marshak, 'A unique current-controlled negative resistance generator', *Electrical Engineering (I.E.E.E.)*, **82**, p. 348, 1963.

23. R. S. C. Cobbold and H. N. Mahabala, 'A tunnel-diode analogue and its application', *Proc. Instn. Elect. Engrs*, **110**, p. 51, 1963 (I.E.E. Paper No. 4061E, January 1963).

24. H. N. Mahabala, 'Oscillation and Switching in Tunnel Diodes', Ph.D. Thesis, University of Saskatchewan, 1964.

25. A. H. Reeves, 'Diode au germanium à espace positif,, *L'Onde Electrique*, **34**, p. 32, 1954.

26. R. A. Johnson and C. O. Harbourt, 'Static characteristics of combinations of negative-resistance devices', *Proc. National Electronics Conference*, Chicago, Ill., **16**, p. 427, 1960.

Radio Engineering Overseas . . .

The following abstracts are selected 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.

MUSICAL INTERVAL GENERATION

A method familiar from digital technique, which enables a given frequency to be divided by a desired whole number, can be exploited for producing musically acceptable intervals. In a Dutch paper the more important tuning systems are briefly reviewed and a description is given of a technique suitable for quick and accurate tuning of keyboard instruments. It produces an equal-tempered semitone, which can be transposed through the gamut at will. The true semitonal interval, whose value is $\sqrt[12]{2}$, can be closely approximated by performing the division 196 : 185, the error being only 5×10^{-6}.

Also described is a procedure for arriving at a complete scale of notes for a melodic instrument, or one affording facilities for harmony. By way of example, numerical values are given for the intervals of natural harmonic scales obtainable by the digital method.

'Generation of musical intervals by a digital method', D. Gossel, *Philips Technical Review*, 26, No. 4/5/6, pp. 170–76, 1965. (In English.)

BLOOD FLOW MEASUREMENT

A plethysmograph records and measures the variation in volume of part of the body due to changes in circulation. It is used in medicine in the study of illness and the effects of drugs.

The author of an Australian paper describes an electronic plethysmograph that is capable of determining volume changes caused by blood flow into the arm to an accuracy of the order of 5% a sensitive high impedance recorder is used.

In addition to providing a voltage–time curve for recording, representing the volume–time changes of the arm, two calibrating positions are provided enabling the volume of the arm to be determined and the ordinate of the chart to be converted into volume units.

'A constant frequency, differential electronic capacitance plethysmograph', E. O. Willoughby, *Proceedings of the Institution of Radio and Electronics Engineers Australia*, 26, No. 8, pp. 264–72, August 1965.

PULSE-TIME MODULATOR USING STEP-RECOVERY DIODES

This Japanese paper presents results obtained when the electric current response of the step recovery diode, driven by a sinusoidal current superimposed on a d.c. bias, was analysed. It was found that for the special case of $\omega\tau = 3 \cdot 644$ (where ω is sinusoidal wave angular frequency, and τ is minority carrier recombination lifetime in the diode) the time when the step recovery occurs changes linearly to the bias current, provided that the bias current is smaller than one-tenth of the sinusoidal current amplitude.

This is applicable to a pulse-time modulation modulator. The sensitivity of modulation and the higher harmonics distortion were calculated.

'Pulse modulation and frequency modulation using the step recovery diode', Kzuo Husimi and Takahiro Inamura, *Review of the Electronics Communications Laboratory*, N.T.T., 13, Nos. 1–2, pp. 44–50, January–February 1965.

ANALOGUE TO DIGITAL CONVERSION

After an account of the various methods of numerical analogue conversion commonly used, the writers of this French paper describe briefly the principle of a conversion mechanism which they call 'modulation by differential duration'.

Two generators of changing voltage (for example, linearly decreasing) are used, one for the voltage which is to be measured and the other as a reference voltage. They are switched on at the same moment. In the first, the changing voltage is stopped by a comparator at the voltage to be measured. For the second, the changing voltage is stopped at zero by another comparator. These two operations appear as two pulses modulated in duration, one variable in relationship to the voltage measured and the other fixed. The difference of these two pulses gives a time which is proportional to the voltage to be measured and its sign. If the two changing voltage generators and the comparator are identical, this arrangement makes it possible to eliminate errors due to distortion products.

'A digital differential analogue convertor of very linear kind using time modulation', J. M. Vauchy, M. Birnbaum, E. Dève, *L'Onde Electronique*, 45, No. 464, pp. 1326–31, November 1965.

ELECTROMAGNETIC RADIATION AND THE HUMAN BODY

A recent Czechoslovakian paper deals with the problem of evaluating correctly the irradiation of the human body by r.f. or u.h.f. energy, with regard to the influence of body-constitution parameters. Using the Smith circle impedance diagram, a qualitative solution is arrived at from models having the form of homogeneous layers of different thicknesses and properties of skin, fat and muscle. Attention is also drawn to the influence of garments.

'The influence of body-constitution on the absorption of electromagnetic waves', J. Musil, *Slaboproudý Obzor*, 26, No. 7, pp. 391–7, July 1965.